STRUCTURAL ANALYSIS OF THE MECHANISMS UNDERLYING RYANODINE RECEPTOR-MEDIATED DYSFUNCTION LEADING TO CARDIAC ARRHYTHMIAS

PhD

2014

Peter Wilson

Acknowledgements

Table of Contents

List of Figures

List of Tables

Summary

The ryanodine receptor (RyR) is a large (563 kDa) tetrameric calcium channel that controls the efflux of calcium ions from the large internal stores within myofibrils called the sarcoplasmic reticulum. The cytoplasmic domain is connected to the pore via an "interacting-domain" (I domain) which plays an important role in the mechanism of channel regulation. Within the I domain is a calcium binding region which resembles a lobe of calmodulin and is thus called the calmodulin-like-domain (CaMLD). CaMLD has a binding partner region, the calmodulin-binding-domain (CaMBD) located in the cytoplasmic domain. Mutations in the cardiac isoform of RyR have been shown to result in the clinical presentation of a lethal disease condition called catecholaminergic polymorphic ventricular tachycardia (CPVT). Thus, mutations proximal to the CaMLD and CaMBD region are of particular interest, since if these mutations affect the binding of calcium or result in any conformational change of the protein, then the role of the mutation could be explained.

Computer simulation methods were employed to investigate the role of CaMLD and CaMBD in the overall biology of RyR. The individual domains and also the entire section of RyR containing both domains were modelled. This latter model underwent a simulated folding process to generate a tool to further investigate the role of CPVT mutations within this region in the context of the structural changes brought about by the changes in amino acid residues of the protein.

CaMLD and CaMBD were produced as recombinant proteins and selected CaMLD CPVT mutants generated. The calcium dependent interaction of these two domains was examined through a range of functional techniques, in particular circular dichroism. All the studied mutants displayed a calcium induced conformational change associated with the functional protein. However, one particular mutation was identified which appeared to significantly weaken the structural stability and have reduced calcium sensitivity when compared to the wild type recombinant protein.

Abbreviations

aa	amino acid
Ach	acetylcholine
apo-CaM	calcium free CaM
ARVD/C	arrhythmogenic right ventricular dysplasia/cardiomyopathy
ATP	adenosine-5'-triphosphate
bp	base pair
BSA	bovine serum albumin
Ca-CaM	calcium bound CaM
CaBP	calcium binding protein
CaM	calmodulin
CaM kinase II	calmodulin kinase II
CaMBD	calmodulin binding domain
CaMBP	calmodulin binding protein
CaMLD	calmodulin like domain
CBD2	calmodulin binding domain 2
CCD	central core disease
CD	circular dichroism
CG	conjugate gradient
CHARMM	chemistry at Harvard macromolecular mechanics
CICR	calcium induced calcium release
CIP	calf intestinal phosphate
CIR	calcium interacting region
cPCR	colony polymerase chain reaction
CPVT	catecholaminergic polymorphic ventricular tachycardia
cryo-EM	cryo-electron microscopy
Da	Dalton
DAD	delayed-after-depolarisation
DHPR	dihydropyridine receptor
DNA	deoxyribonucleic acid
DOPE	discrete optimised protein energy

DTT	dithiothreitol
EC coupling	excitation-contraction coupling
ECC	excitation-contraction coupling
EDTA	ethylenediaminetetraacetic acid
ECG	electrocardiogram
EGTA	ethylene glycol tetraacetic acid
ER	endoplasmic reticulum
FKBP	FK506-binding protein
FKBP12	FK506-binding protein – 12.0 kDa
FKBP12.6	FK506-binding protein – 12.6 kDa (cardiac isoform)
FRET	fluorescence resonance energy transfer
GFP	green fluorescent protein
GndHCl	guanidine hydrochloride
GST	glutathione S-transferase
GUI	graphical user interface
HMJ model	Herzberg, Moult and James model
I domain	"interacting" domain
IP_3R	inositol trisphosphate receptor
IPTG	isopropyl β-D-1-thiogalactopyranoside
LB	lysogeny broth (Luria-Bertani)
LTCC	L-type calcium channel
MALDI-TOF	matrix-assisted laser desorption/ionization – time of flight
MCS	multiple cloning site
MD	molecular dynamics
MH	malignant hyperthermia
MWCO	molecular weight cut off
NCX	sodium/calcium exchanger
Ni-NTA	Ni-nitrilotriacetic acid
NMR	nuclear magnetic resonance
NusA	N utilization substance protein A
OD	optical density
PAGE	polyacrylamide gel electrophoresis

PCR	polymerase chain reaction
PDB	protein data bank
PEG	polyethylene glycol
PKA	protein kinase A
PLB	phospholamban
RIH	RyR and IP$_3$R homology
RMSD	root mean square distance
RNA	ribonucleic acid
ROCC	receptor operated calcium channel
rpm	revolutions per minute
RyR	ryanodine receptor
RyR1	skeletal isoform of the ryanodine receptor
RyR2	cardiac isoform of the ryanodine receptor
SD	steepest descent
SDS	sodium dodecyl sulfate
SERCA	sarco/endoplasmic reticulum Ca^{2+}-ATPase
SOCC	store operated calcium channel
SR	sarcoplasmic reticulum
T tubules	transverse tubules
TAE	tris base, acetic acid and EDTA
TB	terrific broth
TBS	tris buffered saline
TEV	tobacco etch virus
TM	transmembrane
TnC	troponin C
TnI	troponin I
TnT	troponin T
TPM	tropomysin
TRIS	tris(hydroxymethyl)aminomethane
Trx	thioredoxin
UV	ultraviolet
VMD	visual molecular dynamics
VOCC	voltage operated calcium channel
wt	wild type

ΔG   free energy

Chapter 1:
Introduction

The aim of this thesis is to understand the importance of the ryanodine receptor in excitation contraction coupling in normal and pathological disease states of skeletal and cardiac muscle. In particular the role of calcium and calmodulin in regulation of RyR activities. The ryanodine receptor is situated in the sarcoplasmic reticulum (SR) membrane and acts as a calcium "release-channel" for the SR calcium store.

1.1 Calcium and calcium channels

The calcium ion (Ca^{2+}) is a vital signalling agent for a large number of events from a cellular level upwards from the earliest stages of life through to death. These many processes include the activation of oocytes at fertilisation to the death of cells, muscle contraction, synaptic transmission, energy metabolism, activation of transcription factors, and apoptosis and necrosis (Berridge et al. 1998), (Bootman et al. 2001). Although in lower concentrations than other ions, Ca^{2+} required careful treatment by the earliest cells. Ca^{2+} interacts easily with biological molecules and, at high concentrations, will precipitate with phosphates, weaken lipid membranes, and aggregate nucleic acids and proteins. As such, even the most basic of lifeforms maintain Ca^{2+} homoeostasis retaining internal concentration around 100nM, significantly lower than the external micromolar concentration (Case et al. 2007). The maintained difference in internal versus external concentrations generates a gradient which allows the cell to use Ca^{2+} in a signalling role.

To permit the controlled transfer of calcium ions, channels were required in the membrane. It has been proposed that the first of these were not protein based, instead relying on bacterially produced polymers to produce a channel which was sensitive to ion concentration and selective for divalent cations. Voltage-gated protein calcium channels eventually evolved and various types are now found in bacteria and higher organisms. Prokaryotic examples of these channels are simpler in structures involving less domains than eukaryotic versions (Case et al. 2007).

The development of the complexity of the Ca^{2+} channels produced several types of influx channel; these tend to be classified by how they are activated. Mechanically triggered channels have been found in cells that need to respond to physical changes the

cell experiences whereas store-operated Ca^{2+} channels (SOCCs) respond to a decrease in internal Ca^{2+} concentration. The latter are thought to function through binding with a cellular messenger produced with the decrease in Ca^{2+} levels or by direct interaction with internal calcium release stores (Bootman et al. 2001).

As a large variation of cell types display increased calcium entry on depletion of internal stores, it is proposed that SOCCs may be the most common membrane calcium channel. Receptor operated calcium channels (ROCCs) are activated by binding of agonists to their extracellular domains. These are located in large numbers at nerve endings and on secretory cells. In addition, they have a range of structural and functional differences (Bootman et al. 2001).

Voltage operated calcium channels (VOCCs) are activated by depolarisation of the membrane. They are commonly located in muscle cells, neurons and other excitable cells. VOCCs contain several protein subunits which function in the role of pore and its gating. Isoforms of these subunits have been identified permitting a range of combinations as observed in the tissue specific VOCCs that are found. VOCCs are commonly characterised by their response to pharmacological agents and gating behaviours (Bootman et al. 2001). The L-type calcium channel (also referred to as a dihydropyridine receptor) shown in figure 1.11 is a VOCC.

Another factor introduced by the development of eukaryotic cells was the inclusion of organelles that amongst other functions would serve as calcium stores. Mitochondria, with their bacterial heritage contain a calcium channel (calcium uniporter) and a Na^+/Ca^{2+} exchanger (NCX). Hence, they were able to adopt to this role. Increased calcium concentration in the mitochondria promotes activity of enzymes in the citric acid cycle leading to enhanced ATP synthesis (Case et al. 2007).

Another example of an internal calcium store is within the endoplasmic reticulum (ER). Here, they play a key role in protein synthesis, post-translational protein modification and transport of molecules around the cell. The ER maintains a much higher Ca^{2+} concentration; in the range of several hundred micromolar, compared to the intracellular space (where calcium concentration is normally around 1-2 mM). This intracellular store serves as another source of calcium adding a further layer to calcium signalling.

The release of Ca^{2+} from the ER stores is operated by three main sets of intracellular channels; NAADP controlled receptors, inositol 1,4,5-triphosphate $(InsP_3)/Ca^{2+}$ gated $InsP_3$ receptors (IP$_3$Rs), and calcium-gated ryanodine receptors (RyRs). Although RyRs and IP$_3$Rs share similar structure and function, the former is much more advanced down the evolutionary process (Bootman et al. 2001), (Case et al. 2007).

The increased concentration in Ca^{2+} can be used as a single signal per event such as in muscle contraction. However, the rate and/or amount of increase can be controlled to provide greater finesse in signalling. This mechanism of control is important as the increase in the amplitude of the Ca^{2+} signal is thought to become less effective as cells have reduced sensitivity to increasing concentrations of Ca^{2+} (Berridge et al. 1998). Calmodulin dependent protein kinase II (CaMKII) is a commonly cited example of the monitoring of the frequency of Ca^{2+} signal. The subunits of the kinase respond differently to the rate of signal. Therefore, allowing for a graduated response to the rate of Ca^{2+} signal.

Excitation-contraction coupling (ECC) in muscle cells is a prime example of the use of Ca^{2+} for the regulation of a highly specified physiological function. The ECC process involves compartmentalisation and many feedback loops to maintain Ca^{2+} homoeostasis. Ca^{2+} itself directly regulates many of the key proteins and also acts indirectly through the modulatory protein calmodulin (CaM). It should be noted that one of the pivotal proteins involved in the regulation of ECC is RyR (Bers 2002).

1.2 Identification of ryanodine receptors

Ryanodine receptors were first identified from images captured of striated muscle using electron microscopy. The junctions between sarcoplasmic reticulum (SR) and transverse (T) tubular system at the triad were examined. RyRs were seen as part of the "foot" structures at these junctions (Franzini-Armstrong 1970), (Campbell et al. 1980). As discussed above SR, smooth ER located in smooth and striated muscle, was known to act as a calcium store and contain calcium channels. Electrophysiology experiments in which SR membrane extracts were incorporated into lipid bilayers displayed the presence of a channel behaving in the role of a Ca^{2+} release.

RyR was purified from the SR and demonstrated to be part of the previously observed foot structures. This extracted RyR was shown to be sensitive to the plant alkaloid ryanodine when operating in its role as a conductivity channel. Low concentrations (submicromolar) of ryanodine locked the channel in an open state and reduced the current passing through the "open" channel to about 40%. Ruthenium red was known to close the channel but in this ryanodine open locked state this did not occur though there was a further reduction in conductivity. Higher concentrations, 300µM, of ryanodine closed the channel (Nagasaki & Fleischer 1988). Work conducted in rabbit skeletal muscle isolated RyR as an integral SR membrane protein and confirmed its role as a Ca^{2+} release channel (Lai et al. 1988), (Lai et al. 1989), (Meissner et al. 1989).

RyRs were subsequently identified across a range of tissue types and 3 clear types of channel identified. RyR1 was found in rabbit skeletal muscle (Marks et al. 1989) and is mainly found in skeletal muscle though does occur at lower levels in some brain tissues. RyR2 was initially extracted from rabbit cardiac muscle (Otsu et al. 1990), where it is the predominant isoform, and is also found in neurons. RyR3 was initially isolated from rabbit brain (Hakamata et al. 1992) but has been found at low levels across a variety of tissues. Complementary DNA was cloned for the 3 isoforms and gene sequences determined for these 3 isoforms; RyR1 (Takeshima et al. 1989), RyR2 (Nakai et al. 1990), (Otsu et al. 1990) and RyR3 (Hakamata et al. 1992).

It was determined that there was approximately 65% amino acid sequence identity between the mammalian isoforms. Despite the range of tissues in which RyRs occur and the variation between isoforms their major identified role is in ECC in muscle cells, the process of converting an electrical nerve stimulus into the mechanical response of muscle contraction (Endo 2011). This process is discussed in further detail in a later section of this chapter.

1.3 Structure of the ryanodine receptor

Continuing research has revealed more about the structure of the largest known ion channel (Capes et al. 2011). Within SR, RyR was found to be a homotetramer

containing a central pore, each subunit has a mass of ~560 kDa (Lai et al. 1988). The size of the channel has lent itself to study by cryo-electron microscopy. With advancements in microscopes and image processing algorithms the resolution of the produced images have improved from ~30Å (Radermacher et al. 1994), (Serysheva et al. 1995), (Sharma et al. 1998) to ~10Å (Ludtke et al. 2005), (Samsó et al. 2005).
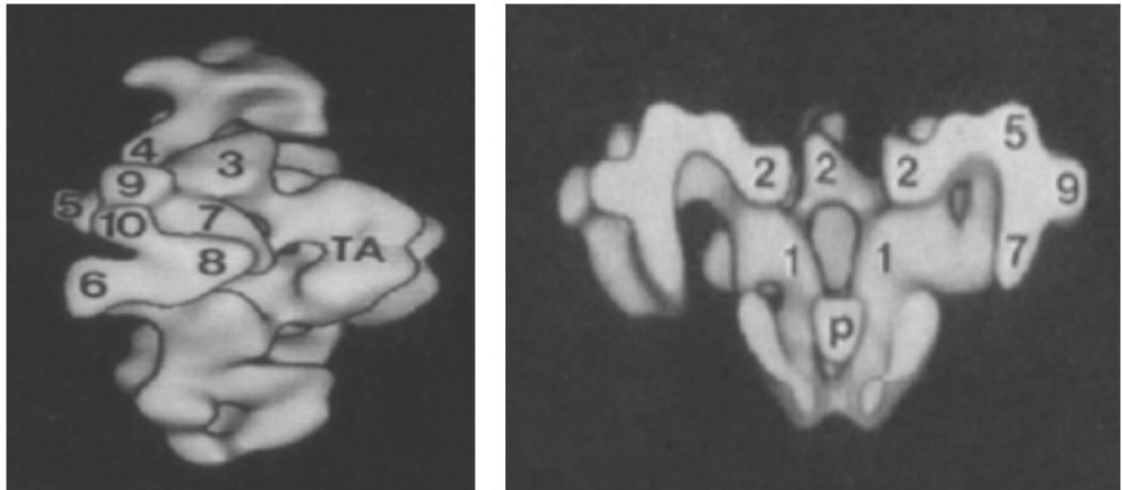
Figure 1.1   3D reconstruction map of RyR1 (Radermacher et al. 1994).
30Å *resolution map processed from cryo-EM data illustrating the "toadstool" like shape of RyR. Also shown is the numbering of putative domains as proposed by the Wagenknecht group. TA relates to the transmembrane assembly and p to a feature of the pore.*

Similar topological surface maps have been produced by the various techniques though the increasing resolution has started to reveal more detail about the overall structure. Analysis of these images has shown the 4-fold symmetry of the homotetramer and led to the "toadstool" analogy model of the structure (see figure 1.1 above). The "stalk" of this structure equates to the transmembrane portion of RyR, which acts as the pore section of the channel in the SR membrane.

The "head" section, represents the region that sits facing the inner cytoplasmic space of the SR. Discrete globular domains separated by solvent-accessible regions have been observed in the cytoplasmic portion. The cytoplasmic head seems to be separated from the transmembrane stalk by 4 columns (Samsó et al. 2005).

There appears to be differences in the transmembrane section if the channel is treated with ryanodine, though the low-resolution of the images makes this difficult to determine. However, domain rearrangement is observed between the open and closed states of RyR (Orlova et al. 1996a). This has lead to the suggestion of an iris lens-style opening and closing mechanisms for the channel (Serysheva et al. 1999). The improving image resolution combined with other techniques such as the incorporation and subsequent monitoring of green-fluorescent-protein (GFP) has allowed for attempts to map specific domains and binding areas onto the overall structure of RyR.

Work conducted mainly on the N-terminal region of RyR has revealed the localisation of domains containing known mutations within the quaternary structure of the protein (Liu et al. 2002), (Liu et al. 2005), (Wang et al. 2007). This approach has illustrated the potential of identifying regulatory regions that are brought together through conformational changes within RyR. However, the disruption caused by the insertion of the GFP tag has to be considered.

The increasing appearance of high resolution structures of regions of RyR has allowed visualisation of mutations within the three-dimensional structure. The techniques of x-ray crystallography and nuclear magnetic resonance (NMR) have been used to produce these structures revealing the nature of the structural impact of the studied mutations. Initially a short stretch of the RyR1 CaM binding site in complex with CaM was observed (Maximciuc et al. 2006). A more detailed discussion of this structure is presented in chapter 3, also the CaM binding site is further explored later in this chapter. Focus has been, primarily, on the N terminal domain region of RyR with a succession of released structures containing more sequence and further studied mutations (Amador et al. 2009), (Lobo et al. 2011), (Tung et al. 2010), (Kimlicka et al. 2013). Other regions of RyR are being investigated by these techniques, recently structures of the central domain phosphorylation sites, including likely CaM kinase II interaction sites, have been published (Sharma et al. 2012), (Yuchi et al. 2012).

Figure 1.2 shows the 10 Angstrom resolution electron microscopy topographical map of the ryanodine receptor. Van Petegem and colleagues have done an *in silico* docking of the recently deduced N-terminal 3D crystallographic structure into this topographical map, this data shows a strong possibility where the N-terminal "repeat-region" should

be located (see figure 1.2). Furthermore, it also shows the putative location of the central regions of phosphorylation. Maps such as these contain important information as to where crucial functional domains may be located and provides clues as to how mutations present on domain interfaces may cause disruption by disruption of domain-domain interactions, thereby destabilising the channel.
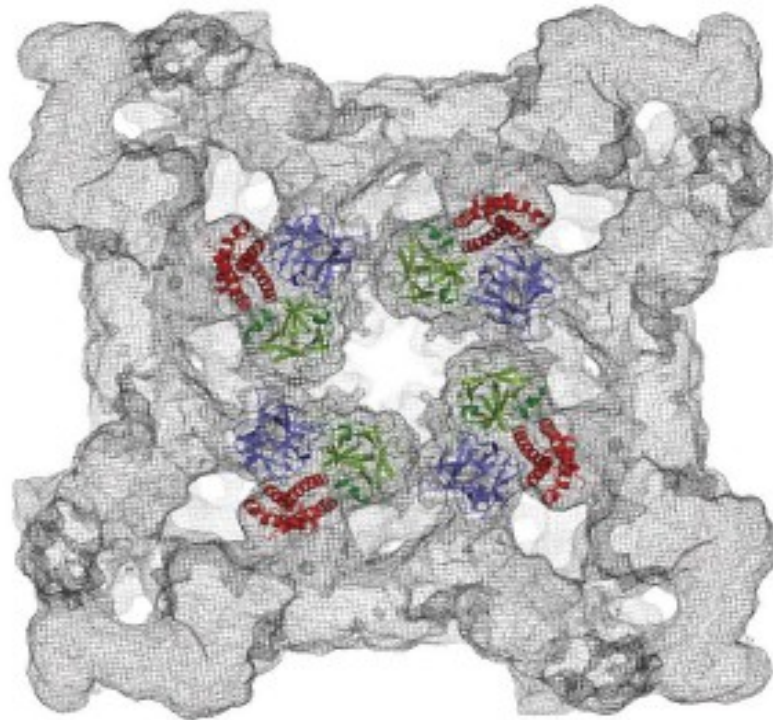
Figure 1.2.   Docked crystal structure of N terminal of RyR1 in cryo-EM map of RyR1 (Tung et al. 2010).

Looking down onto "head" of RyR from cytoplasmic side displaying 4 fold symmetry of RyR tetramer. 2.5Å resolution x-ray crystallography determined structure of RyR1 amino acid residues 1-559 docked in 9.6Å cryo-EM map of RyR1. The N-terminal structure has been split into 3 different coloured predicted domains that interact. Mutations have been observed in small clusters at the interfaces between the domain boundaries.

1.4 Muscle contraction

Each muscle fibre is multinucleate and behaves as a single unit (and therefore it is called a syncytium). It contains bundles of myofibrils, surrounded by sarcoplasmic

reticulum (SR) and invaginated by transverse tubules (T tubules). Each myofibril contains interdigitating thick and thin filaments arranged longitudinally in sarcomeres.
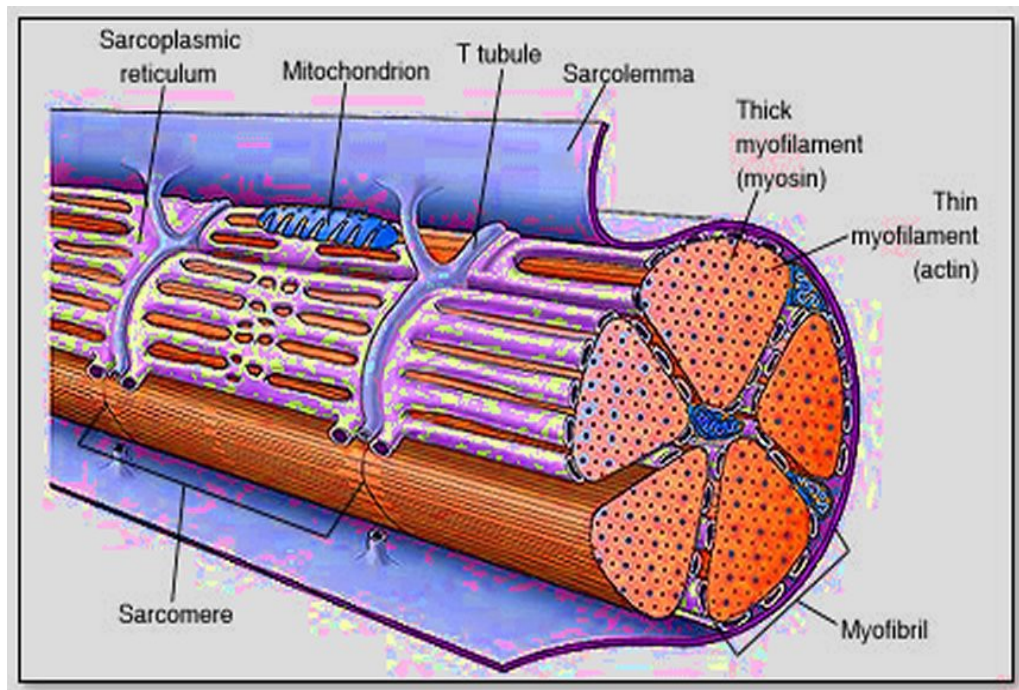
Figure 1.3   Components of muscle tissue (Braunwald 2007).
Schematic diagram showing arrangement of muscle fibres in bundles (myofibrils) surrounded by sarcoplasmic reticulum (SR) and T-tubules.

Skeletal and cardio-myocytes are made up of bundles of sarcomeres. Sarcomeres are bundles of myosin and actin that slide over each other.
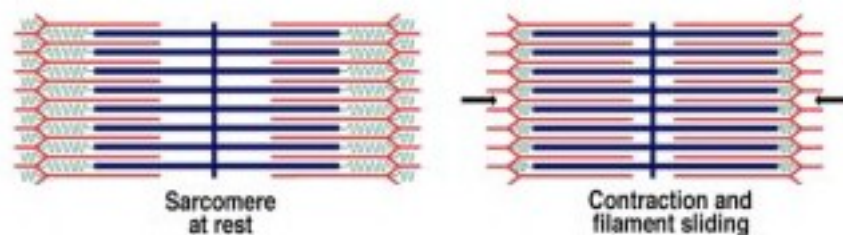
Figure 1.4   The "sliding filament" mechanism of muscle contraction.
The sliding filament model of muscle contraction is where the "thin-filaments" of muscle "slide-over" the "thick filaments".

Briefly, the sliding-filament theory is where actin and myosin filaments slide over each other, thus gradually shortening the length of the sarcomere. This is an active process requiring ATP.

The flood-like massive release of calcium ions from the SR via the RyR1 channels allow calcium ions to flow into the cytoplasm. These large pools of calcium are now capable of binding to Troponin C. This, in turn, causes the troponin complex to dissociate from tropomyosin. The role of tropomyosin is important to shield the myosin-binding site preventing the binding of actin and the formation of cross-bridges in absence of calcium. Myosin can then bind to the uncovered sites on actin, a crucial step in the sliding filament theory", see figure 1.5. Thus, muscle contraction occurs in response to previous electrical events.
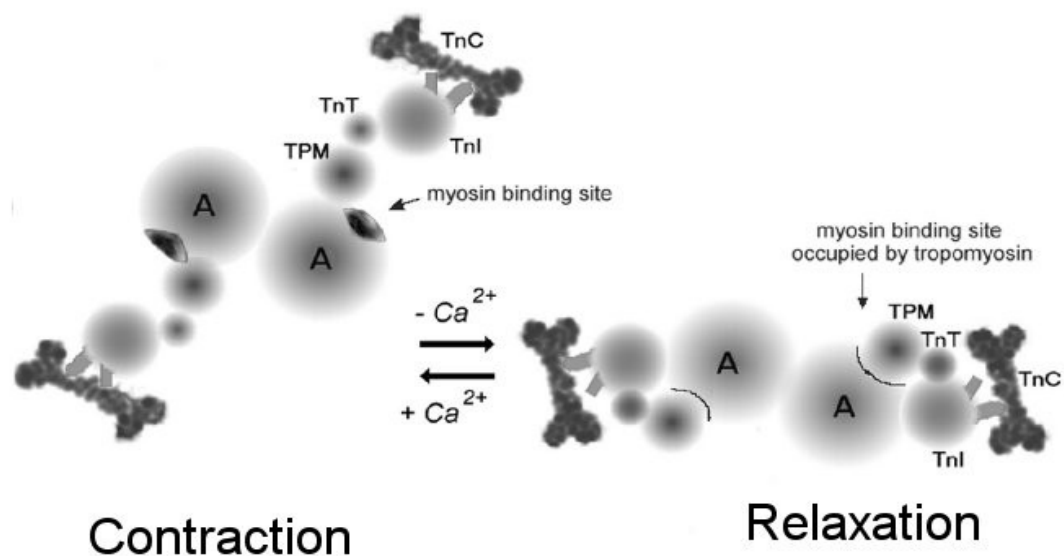
Figure 1.5   Regulation of contraction and relaxation in the muscle (Davison et al. 2000).

Schematic view showing the spatial relationships of actin (A), troponin T (TnT), tropomyosin (TPM), troponin C (TnC) and troponin I (TnI). In the relaxed state (on the right), tropomyosin occupies the myosin binding site on the actin molecule. This therefore prevents actin from forming cross-bridges with myosin. The troponins; T, I and C are associated with each other, in low calcium levels, troponin C loosely associates with troponin I via a single site. When there is a large release of calcium ions, troponin I strongly associates with troponin C via two binding sites, so altering the overall conformation of the troponins and "pulling" tropomyosin out of the binding cleft on actin. This then allows myosin to form cross-bridges with actin and thus commence muscle contraction.

1.5 Excitation-contraction coupling

Excitation-contraction coupling (ECC) is the process by which an electrical signal brings about a mechanical contractile event. The events described for contraction in skeletal muscle are different when it comes to cardiac muscle. Firstly, skeletal muscle is under voluntary control. The intent to move a muscle begins first with a higher cortical response which travels down the cortico-spinal tract and exits at the appropriate vertebral level via the ventral horn, finally ending as a motor end-plate (synaptic bulb) on a muscle where calcium channels control the release of the neurotransmitter acetylcholine that trigger off the rest of the excitatory process (Braunwald 2007).

Cardiac myocytes near the sino-atrial node and the atrio-ventricular nodes have a property called automaticity. This is an important property of these cells that can spontaneously generate periodic oscillations of electrical current and set off an action potential. In the hearts of higher vertebrates, so-called "pacemaker" cells exist, these cells have low contractility but have the capacity to generate periodic electrical oscillations. The generated action potential then spreads via the "Purkinje fibre" system and the bundle of His, down the septum and upwards around the ventricles, spreading the excitatory impulse throughout the myocardium to the individual myocytes (Braunwald 2007).
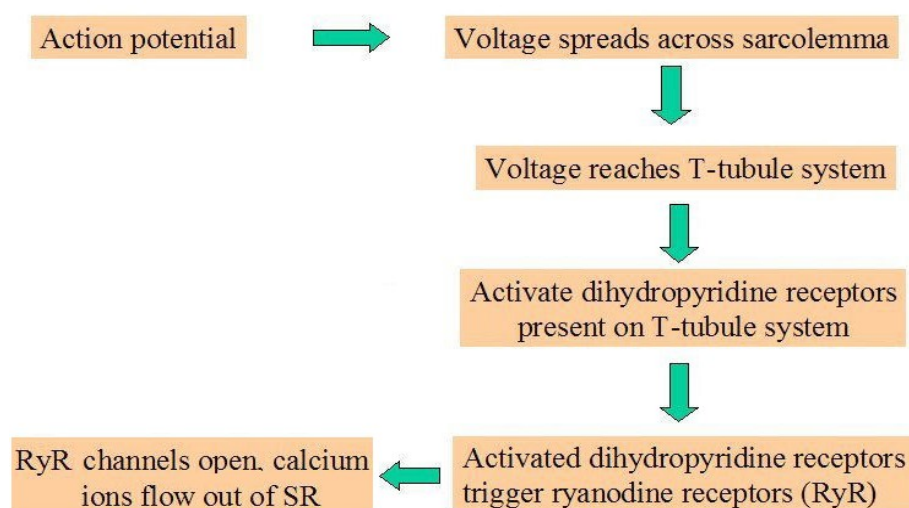
Figure 1.6   Flowchart summarising events leading to opening of RyR channels.
The various events from the propagation of the action potential through to the activation of the dihydropyridine receptors and finally the opening of the SR stores of calcium via the RyR1 receptor.

1.5.1 Action potential
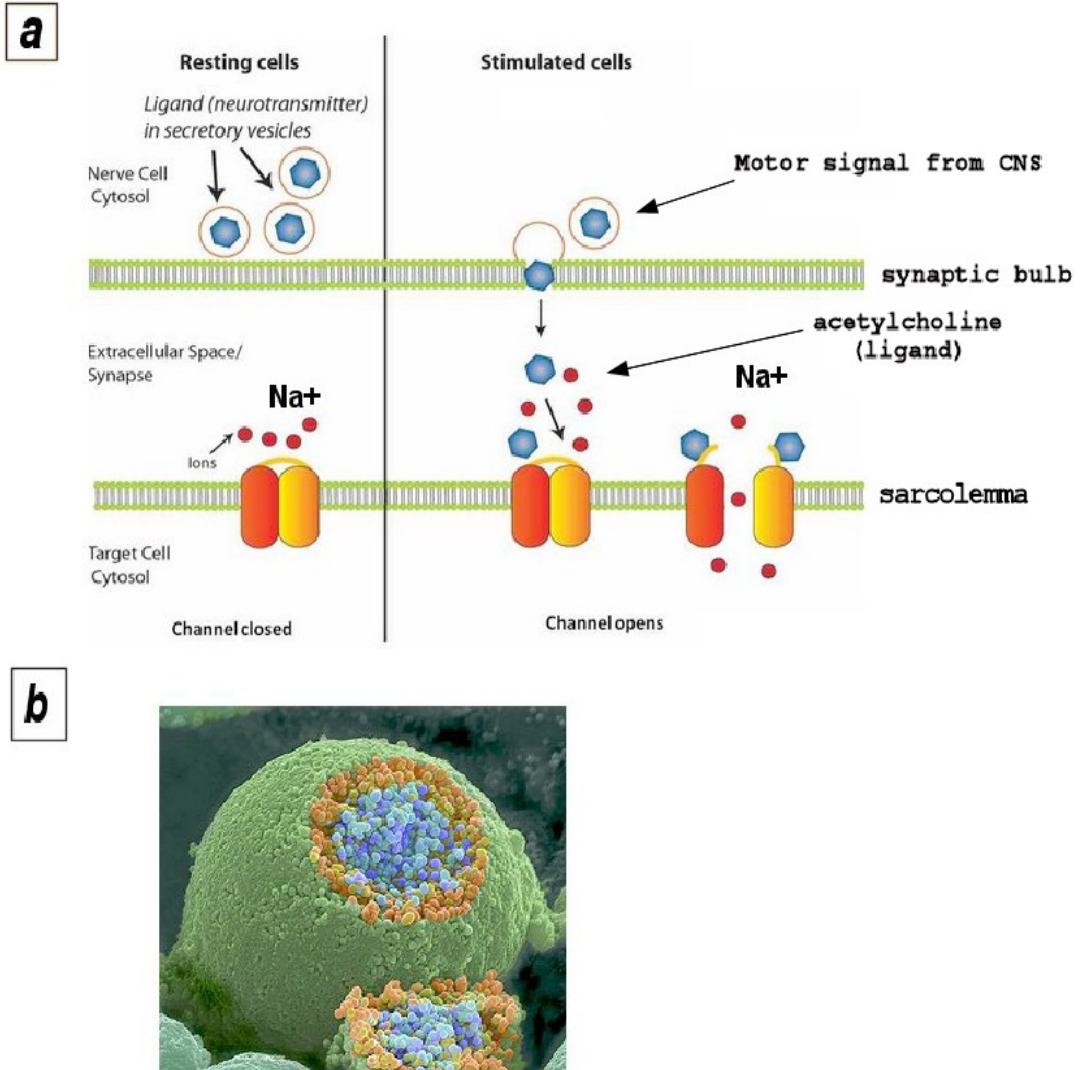
Figure 1.7   Neurotransmission of signals to muscles.

(a) the mechanism by which ligand-gated Na$^+$ channels work; when they bind to ACh, the channels open allowing sodium to rush into the muscle cell, depolarising the sarcolemma.

(b) false colour scanning electron microscope of a presynaptic bulb of neuron (green), ACh molecules are coloured in blue (SEM image from www.brighthub.com, with permission).

A propagating action potential travels down towards the end of a motor neuron. There, as shown in figure 1.7 (panel A), calcium ions travel down the voltage-dependent channels, thus gaining entry into the neuron. The calcium ions then bind to synaptotagmin; a type of sensor protein, which causes the release of the

neurotransmitter acetylcholine (ACh) from the neuron into the synaptic cleft. ACh then travels down the short space of the synapse (approximately 19 -20 nm), binding to the receptors on the membrane of the muscle fibre (Braunwald 2007).

The ACh receptors are in fact, ligand-gated nicotinic receptors. This means that, when ACh (the "ligand") binds to these receptors (which also function as an ionic gate), the gates "open", allowing entry of sodium ions into the muscle cells whilst allowing the concomitant exit of potassium ions out of them. The membrane of muscle cells at rest are approximately between -90 mV to -70 mV. This relatively large negative polarity of the muscle cell membrane is maintained by forcing three sodium ions out of the cell and allowing two potassium ions in via an ATP driven pump leading to a net deficit of one positive charge in the cell. The sudden large influx of sodium ions into the muscle cell causes a shift towards positive values in membrane potential (towards +ve voltages). As the sodium ions rush in, the sarcolemma depolarises rapidly and the membrane potential gradually reach 30 mV (as shown in figure 1.8) (Braunwald 2007).
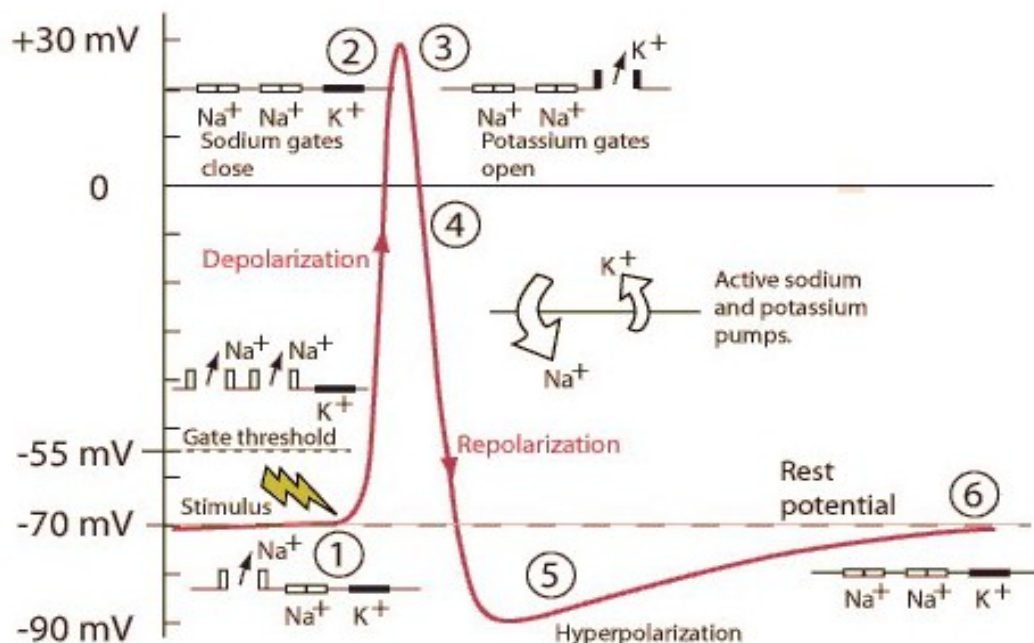
Figure 1.8   Action potential at the neuromuscular junction.

The rush of sodium ions into the myocytes cause a rapid depolarisation of the sarcolemma. The initial sharp rise of depolarisation slows down towards the peak as inward rectifying K^+ channels open. At stage 4, the sodium potassium pumps begin activating with the aid of ATP, restoring the resting membrane polarity.

At the resting membrane potential of –70 mV at the sinoatrial node, there is a steady leak of potassium ions outward, and a slow trickle of sodium ions inward (the so-called funny current of If). The gradual rise in potential is also reached by the opening of voltage-gated calcium channels (L type calcium channels) and this allows the sarcolemma of the cardiac myofibrils to reach a threshold at –40 mV (Braunwald 2007).

The action potential curve of other myocytes has a slightly different shape than that of the pacemaker cells as shown in figure 1.9. The most significant difference is the "plateau phase" in this action potential curve, where phase (1) is due to the deactivation of the fast sodium channels. The notch observed in the plateau phase (figure 1.9 (b)) is due to the efflux flow of potassium ions (I to1) and chloride ions (I to 2) (going out of the cell). The notch is of particular significance, without these transient outward currents, the repolarisation of the myocyte would not occur (Braunwald 2007).
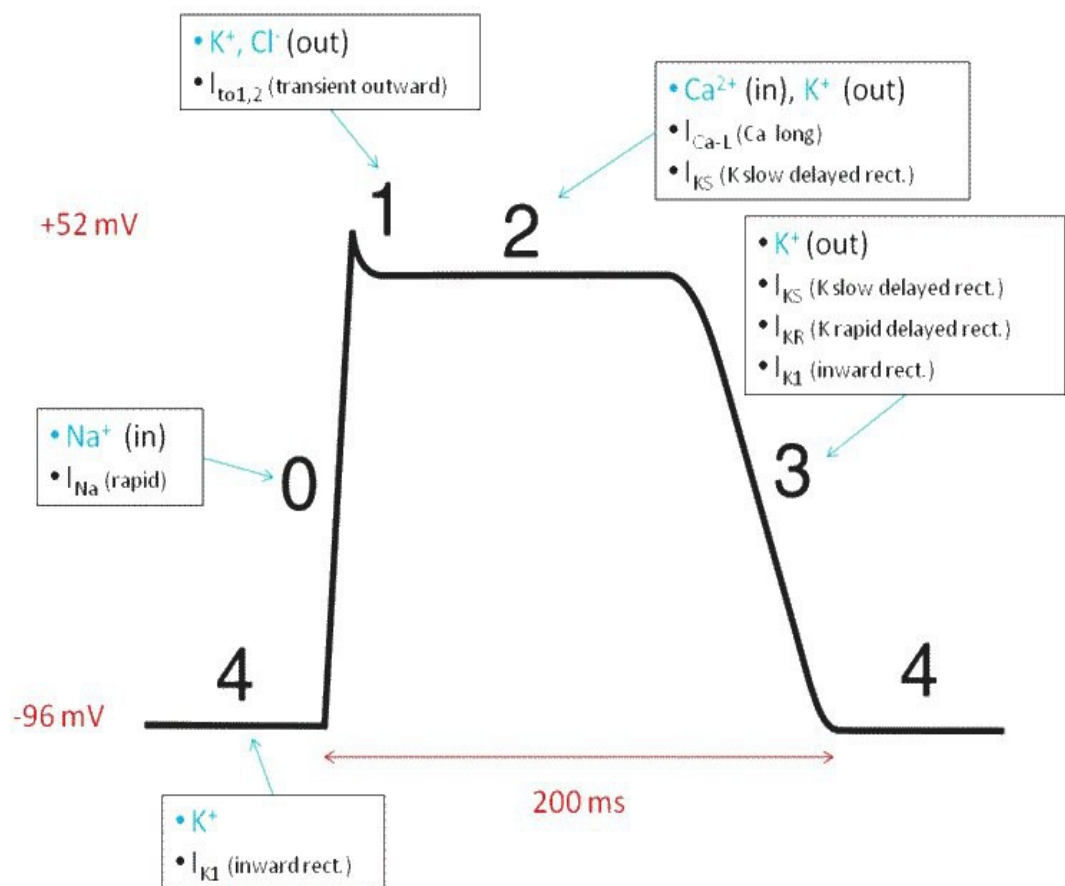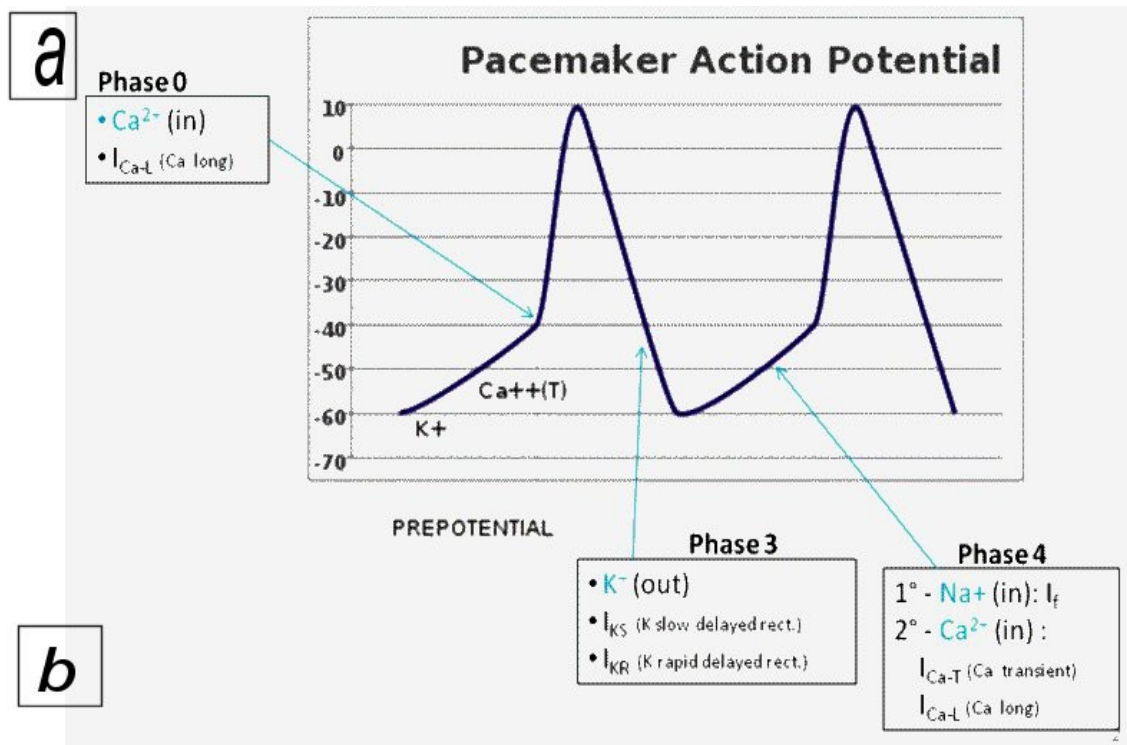
Figure 1.9   Action potential in the heart.

(a) the action potential at the sino-atrial node and (b) the action potential and the various phases at other cardiac myocytes (non-pacemaker cells).

1.5.2 Activation of LTCC/DHPR

The T-tubule system contains voltage-gated Ca^{2+} channels; specifically, these are the dihydropyridine receptors which are also referred to as L type calcium channels (LTCC). The action potential described in figures 1.8 and 1.9, activates these receptors. Dihydropyridines are compounds that contain a pyridine (a C5 heterocyclic) group. Clinically, drugs such as amlodipine, nifedipine and nimodipine are important calcium-channel blockers, which play a role in the control of clinical hypertension. These voltage activated dihydropyridine receptors then trigger and activate the ryanodine receptors that are present on the sarcoplasmic reticulum (SR). Once activated, these ryanodine receptors then open and release a flood of Ca^{2+} ions (Braunwald 2007).
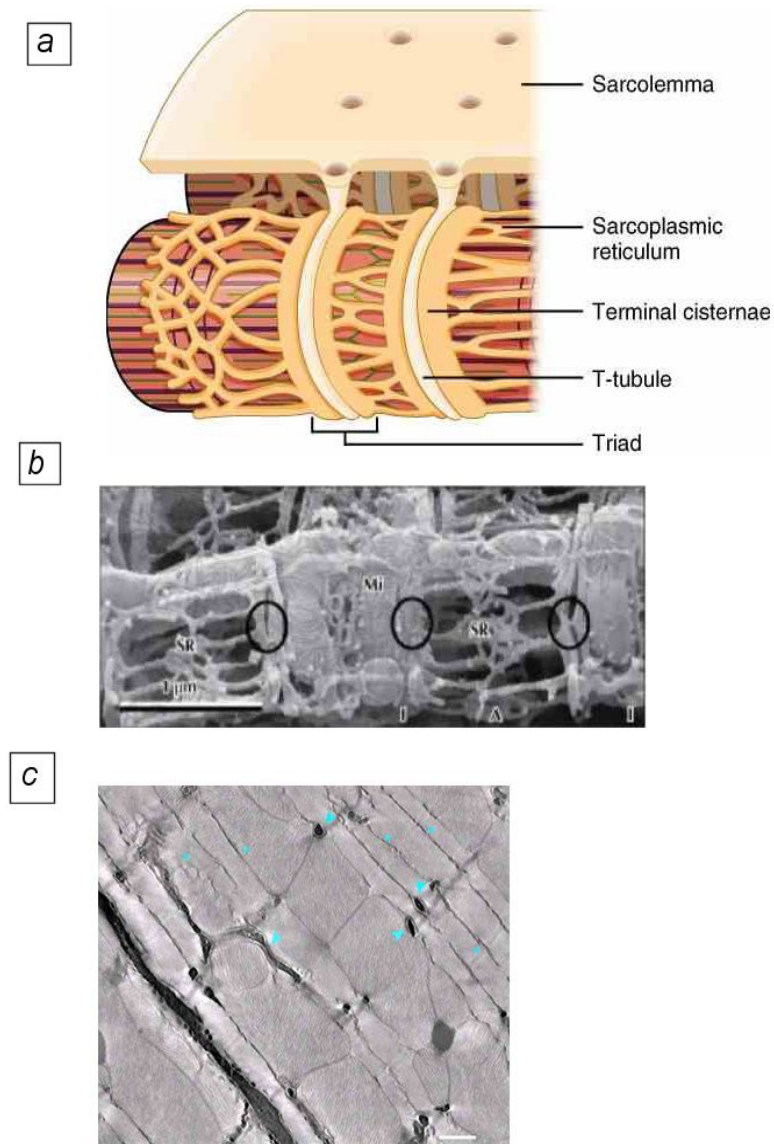
Figure 1.10 Arrangement of myofibrils showing T-tubules (Hayashi et al. 2009).

(a) a schematic diagram showing the components of the myofibril enveloped by the plasma membrane (sarcolemma). Tiny holes are apparent on the sarcolemma, these "invaginations" lead to a tubular system down-beneath, called the T-tubules.

(b) SEM photograph showing the triad (circled) amongst the SR (the sarcolemma has been stripped off here).

(c) SEM photograph of myofibrils with the sarcolemma intact, showing the invaginations on the surface (arrowed in cyan).Scale bars 500 nm.

1.5.3 EC coupling in RyR1

A surge in the voltage causes a conformational change in DHPR which in turn causes an allosteric interaction between DHPR and RyR1 to open the RyR1 calcium-release channel, see fig 1.11 (a). It should be noted that the DHPR is only linked to every other RyR1 and CICR may be responsible, in some way, for triggering these alternate RyR1 channels. RyR1, when isolated from the SR, can also be activated by Ca^{2+} but this is not thought to be the direct physiological mechanism. When the stimulus for contraction ceases, the cytosolic calcium is re-sequestered into the SR-Ca^{2+} store by the SERCA pump (Blayney & Lai 2009).

1.5.4 EC coupling in RyR2

The cardiac DHPR itself, when it is activated by voltage, becomes a Ca^{2+} channel allowing extra cellular Ca^{2+} to flow into the cell and activate RyR2, the process is termed Ca^{2+} induced Ca^{2+} release (CICR). The influx of Ca^{2+} through the DHPR initially flows into a confined space called the dyadic cleft. Due to the small volume of this space a small number of Ca^{2+} ions results in a rapid increase in concentration within the dyadic cleft to reach the threshold of RyR2 activation (10-7umM) (Bers 2002), (Blayney & Lai 2009).

As a result the RyR opens and the SR Ca^{2+} flows out to activate the myofilaments. In order for relaxation to occur the systolic Ca^{2+} is re-sequestered into the SR by the SERCA pump and the proportion which corresponds to Ca^{2+} influx is returned to the extra-cellular space via the Na/Ca exchanger. For this action to run to completion, the RyR channel has to close. There a number of proposed mechanism for the closing of the channel including reduced-luminal-calcium, increased calcium in the dyadic-cleft (which also contributes to the retrograde inhibition of the LTCC and the plateau phase (2) of the action potential shown in figure 1.9), and mechanisms of inactivation within the RyR itself. Thus, normally calcium homoeostasis is restored by returning calcium to the store and extruding calcium (Bers 2002), (Blayney & Lai 2009).
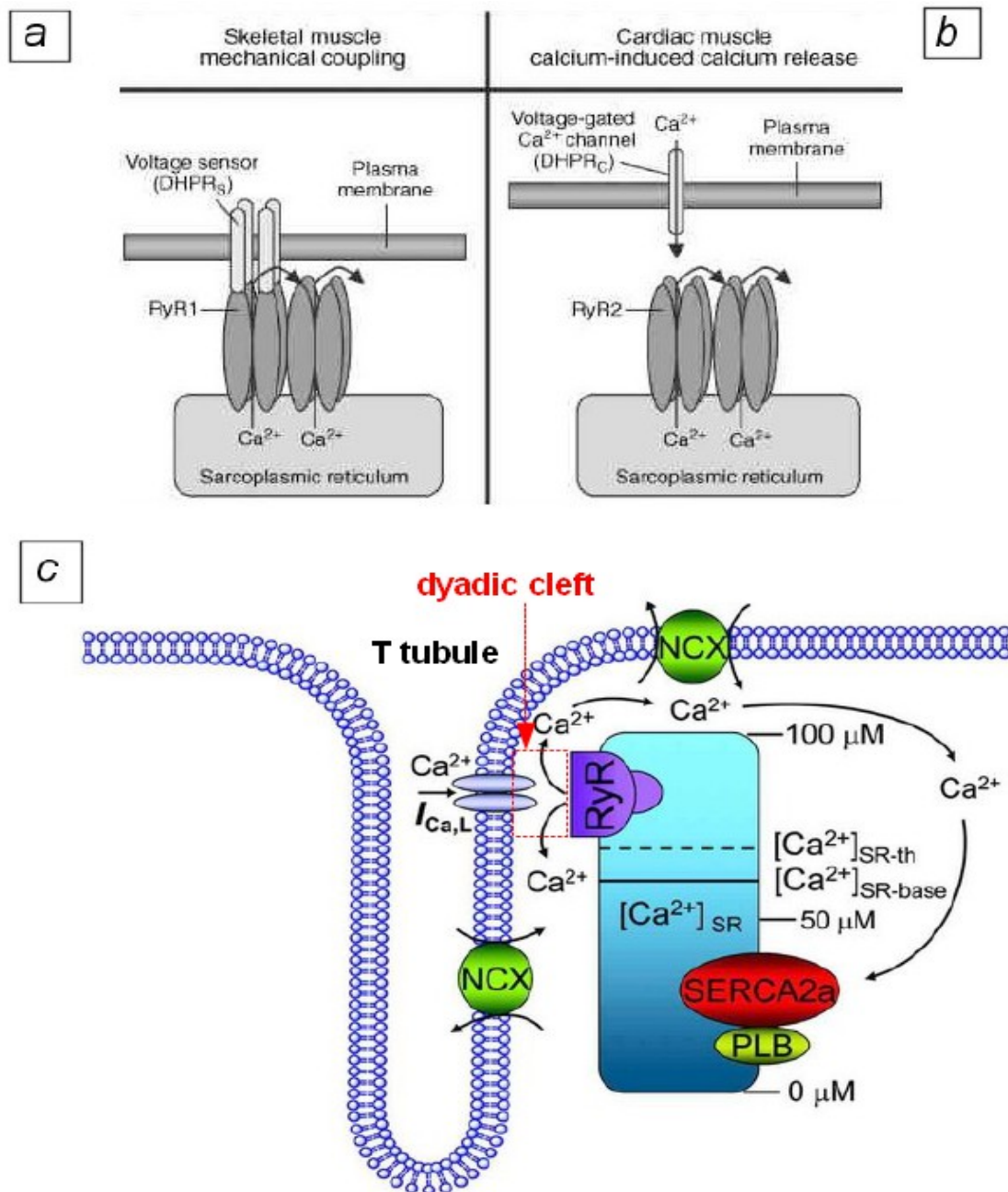
Figure 1.11   The difference between activation of RyR molecules in cardiac and skeletal muscle.

(A & B) In skeletal muscle, RyR1 molecules are coupled to the dihydropyridine receptor (L type Ca^{2+} channel) which then activates the opening of the RyR1 molecule. In cardiac muscle, the influx of calcium ions via the L-type calcium channels drives the opening of the ryanodine receptors via the process known as "calcium induced calcium release" (CICR). The 10-12nm gap at the junction of the SR and T tubules in cardiac muscle in referred to as the dyadic cleft. This is an important microdomain involved in CICR with regards to increasing Ca^{2+} concentration.

(C) The schematic arrangement of channels related to CICR mechanism. These include voltage gated L-type calcium channels present in the T-tubule system, RyR2 in the SR membrane, and also the SERCA and NCX (Na⁺/Ca²⁺ exchangers) associated with repolarisation after CICR has occurred.

1.6 Phosphorylation and the "stress" response

Phosphorylation has a number of roles related to the regulation of protein activity, often a conformational change can be triggered in the target protein resulting in a change in function or continuation of a signalling pathway. Serine is a commonly phosphorylated amino acid with the addition of phosphate to the hydroxyl of the R group of the amino acid. RyR2 has three clearly identified phosphorylation sites, all of which are serines. These sites are acted upon by different combinations of kinase/phosphatase pairs; so allowing for finer control of RyR through phosphorylation.

The phosphorylation sites are usually referred to in conjunction with the kinase that is thought to act upon that site, Ser 2809 (2843 in RyR1) and 2030 are associated with protein kinase A (PKA) and Ser 2815 with Ca^{2+}/CaM-dependent protein kinase II (CaMKII) (Niggli et al. 2013); amino acid numbers here are given for human RyR2. There is still some debate over the specificity of the kinases for these sites and the precise effect that the phosphorylation of each site has on the function of the channel. It is thought likely that there are additional phosphorylation sites throughout RyR which may impact on the physiological activity that has been attributed to individual kinases in some reports.

A range of experiments, involving transgenic animals, indicates that CaMKII activity leads to a general increase in activity from Ca^{2+} signalling through to related downstream events such as EC coupling. This increased response is more Ca^{2+} sensitive and may occur through a response to increased Ca^{2+} levels (MacQuaide et al. 2007) but operates in a less well regulated fashion than the non-phosphorylated state. It should be noted that some of these effects are due to CaMKII having other targets in addition to Ser 2815. These include the L type Ca^{2+} channel, linked to increased Ca^{2+} signalling, and the phospholamban associated with SERCA, found to increase SERCA activity in

preparation for the next excitation-contraction event (Niggli et al. 2013).

PKA appears to have a similar range of targets to CaMKII but its overall effects seen at the cellular and above level maybe more subtle and, as such, are difficult to pinpoint with conflicting data appearing in the literature. Generally there is an increase in heart function activity, as observed with CaMKII activity, as PKA is activated by increasing cAMP levels caused as part of adrenergic stimulation through the flight or fight response. Hyperphosphorylation of Ser 2809 has been linked to heart failure by promoting increased Ca^{2+} leak but analysis of more recent studies are suggesting Ser 2815 as a stronger candidate for this mechanism (Valdivia 2012). Part of this mechanism may be linked to the apparent dissociation of FKBP12.6 from RyR seen in some hyperphosphorylation models (Marx et al. 2000a).

Phosphorylation is linked to up-regulation of ECC in response to beta-adrenergic stimulation also known as the flight or fight response. Essentially, raised circulating adrenaline levels activates the beta 1 adrenegric receptors on the myocyte surface. These G protein linked receptors activate adenyl cyclase which converts ATP to the second messenger cyclic AMP which freely diffuses throughout the myocyte to activate protein PKA. PKA is associated with a number of key proteins involved in ECC, these include PLB which activates SERCA, LTCC, RyR and troponin (as shown in figure 1.11). The net effect is to up regulate calcium fluxes throughout the cell increasing heart rate.

1.7 Arrhythmia

RyR are associated with arrhythmic events which are triggered by phosphorylation. However, arrhythmias can arise from many causes as illustrated in figure 1.12 below.
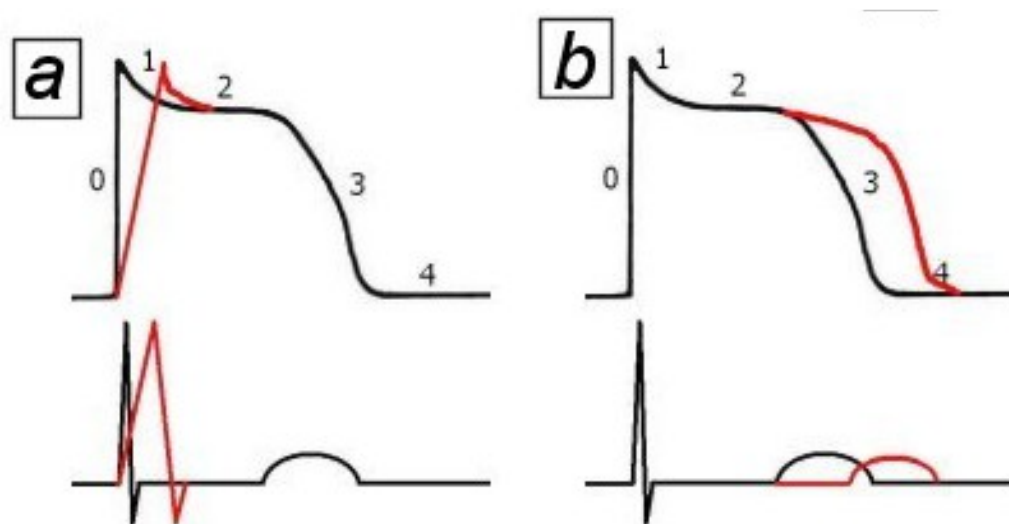
Figure 1.12   Relationship of the cardiac action potential to the ECG.

(a) The effect of sodium channel inhibitors, shows that the slope of phase O is shifted to the right (as shown in the red slope) when sodium channel inhibitors are used.

(b) The lengthening of the ST segment can be achieved by potassium channel inhibitors or even calcium channel inhibitors.

In heart failure, low cardiac output due to weakened pumping activity of the heart results in low blood pressure and impaired tissue perfusion. This is sensed by baroreceptors in the circulation which feedback to the brain, increasing the sympathetic tone by enhanced sympathetic activity.   This also results in enhanced release of adrenaline, and therefore greater release of calcium ions from stores. However, as the contractile apparatus is damaged in heart failure, even this increased calcium wave would not have much of an effect on increasing contractile efficiency, conversely, it would cause depletion of calcium stores in the SR, leading to lower circulating calcium levels (Braunwald 2007). The increased sympathetic tone in damaged muscle, increased circulating catecholamines and the low calcium levels all lead to increased risk of arrhythmias in the myocardium.
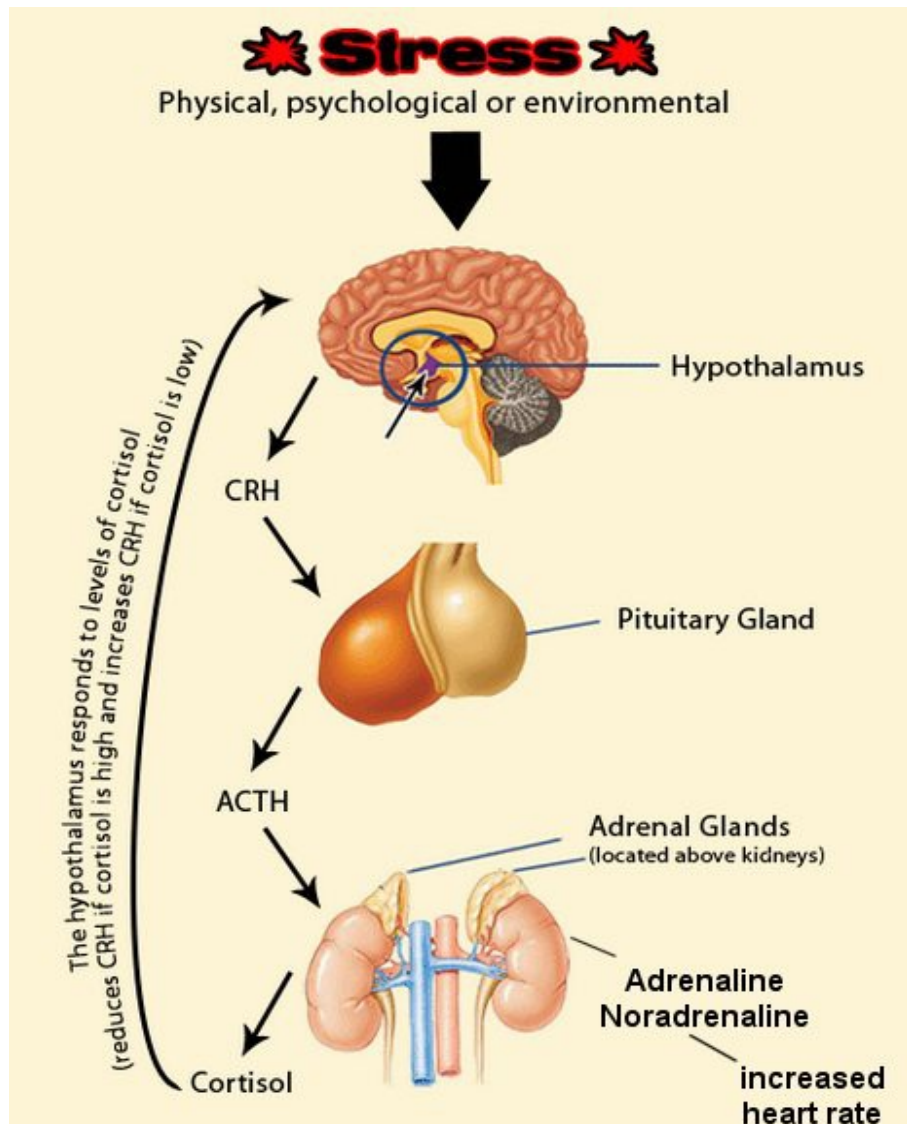
Figure 1.13   The hypothalamic-pituitary axis feedback circuit (Braunwald 2007).

As shown here, increased stress levels will elicit greater release of adrenaline into the systemic circulation which will increase the heart rate, forcing more blood around the circulation.

During intense exercise or when greater cardiac output is needed, feedback circuits to the brain increase adrenaline release, which then encourage enhanced cardiac muscle contractility by enhancing greater release of calcium ions from stores by opening ryanodine receptors.

1.7.1 Mutations in ryanodine receptors

1.7.1.1 RyR1

Mutations in the skeletal isoform of the gene (RyR1 gene) are associated with the clinical manifestation of malignant hyperthermia (MH), encountered when these gene carriers are administered volatile anaesthetics such as the commonly used halothane. Malignant hyperthermia is a potentially fatal condition which if noted during surgery, only a short window of time exists before the administration of life-saving dantrolene (a muscle relaxant) can have an effect and prevent death (Braunwald 2007). Some MH mutations result in central core disease (CCD) which manifests as disordered myofilaments and muscle weakness.

1.7.1.2 RyR2

Similar point mutations in RyR2 cause an arrhythmic condition termed catecholaminergic polymorphic ventricular tachycardia (CPVT) or arrhythmogenic right ventricular dysplasia/cardiomyopathy (ARVD/C) type 2, the latter typified by fibro-fatty streaks in the right atrium (Norman & McKenna 1999). In addition, it should be noted that some CPVT mutations are also found in calsequestrin (Mohamed et al. 2007).

The heart beat is normal at rest but exercise or emotional stress (e.g. flight or fight response i.e. phosphorylation) can result in tachycardia leading to a potentially fatal cardiac arrhythmia. The majority of identified disease causing mutations have been located in 3 cluster regions within RyR as shown in figure 1.14 (below). Sequence analysis conducted on carrier families revealed the first set of known mutations in RyR2 (Priori et al. 2001).

Equivalent clusters of mutations in the N and C terminal, and central regions are seen for RyR2 (CPVT and ARVD2) and RyR1 (MH and CCD). With specific mention to CPVT mutations in RyR2, the C terminal domain region is usually further split into the cytoplasmic I domain and transmembrane domain which also includes the C terminal region itself (Thomas et al. 2006). Selected CPVT mutations studied in the interacting (I) domain were seen to trigger increased signal variability in Ca^{2+} release and

destabilise interdomain interaction (George et al. 2006).
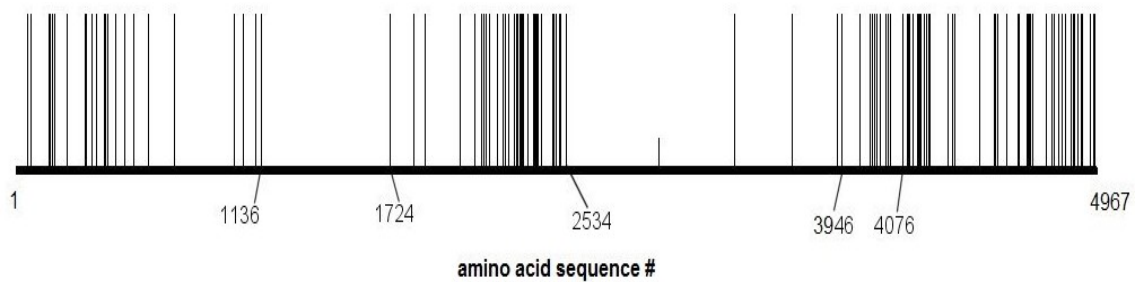
Figure 1.14   Locations of CPVT mutations within RYR2

Each CPVT mutation represented by vertical bar, these mutations can be seen to grouped into clusters. Selected mutations are shown throughout the RyR2 amino acid sequence. 1136 marks the end of the N terminal mutation cluster, 1724-2534 the central region cluster, and 3946 and 4076 are indicated within the I domain section of the C terminal cluster.

1.7.2 Effect of catecholamines on the myocardium

To further explain CPVT, the nature and effects of catecholamines have to be considered. Figure 1.15 illustrates the chemical structure of some of the catecholamines present in the body. In post-infarct patients, the incidence of sudden cardiac death is greatly reduced by regular intake of beta-blocking drugs. This fact highlights that ischemic events induce enhanced sympathetic activation (hence local catecholamine activity) and this leads to increased arrhythmogenic properties of the myocardium (Anon 1981), (Yusuf et al. 1985).
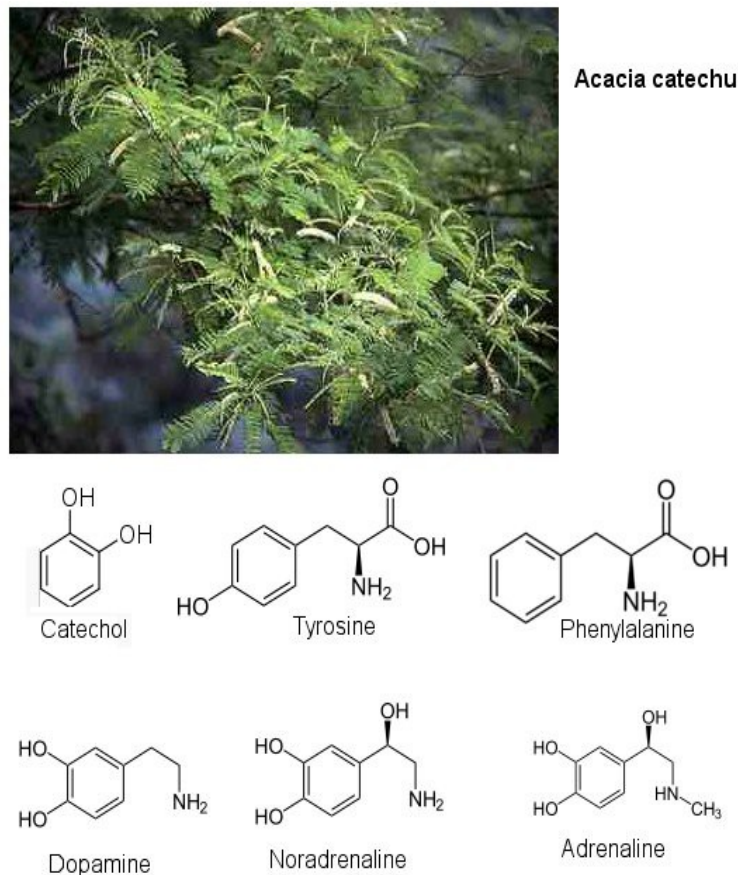
Figure 1.15   Structure of Catecholamines.

Catechols are naturally occurring chemicals (name originating from the extracts of the tree – Acacia catechu) which have an aromatic ring with two hydroxyl groups. Catecholamines are therefore catechols with an amine group. Naturally circulating catecholamines in the body such as dopamine, noradrenaline (norepinephrine) and adrenaline (epinephrine) are synthesised from tyrosine and phenylalanine.

1.7.3 How a faulty RyR2 channel produces cardiac arrhythmias

Inherited or *de novo* mutations in the RyR2 gene or ischemic events, hypertrophy of the heart wall, increased sympathetic tone or cardiac failure; all these mechanisms can cause a diastolic Ca^{2+} leak through the RyR2 channel. This leak of calcium ions can produce transient increases in calcium currents via the heart muscle cells. The wave of calcium ions released at inappropriate timings will reverse NCX thus extruding 2 Ca^{2+} and bringing in 3 Na^{+}. This triggers delayed-after-depolarisations (as shown in figure 1.16). If sufficient DADs are generated within a certain space of time, this can cause the depolarisation of heart muscles above the threshold limits. Since the excitation event is

linked to contraction (EC-coupling), these sudden groups of DADs can result in either a single or multiple premature heartbeats, thus resulting in arrhythmia. The events described above are summarised in figure 1.17 (Iwasaki et al. n.d.).

DADs

Ca^{2+} overload / RyR dysfunction
↓
Abnormal SR Ca^{2+} release (RyR)
↓
I$_{ti}$ through NCX
↓
DAD
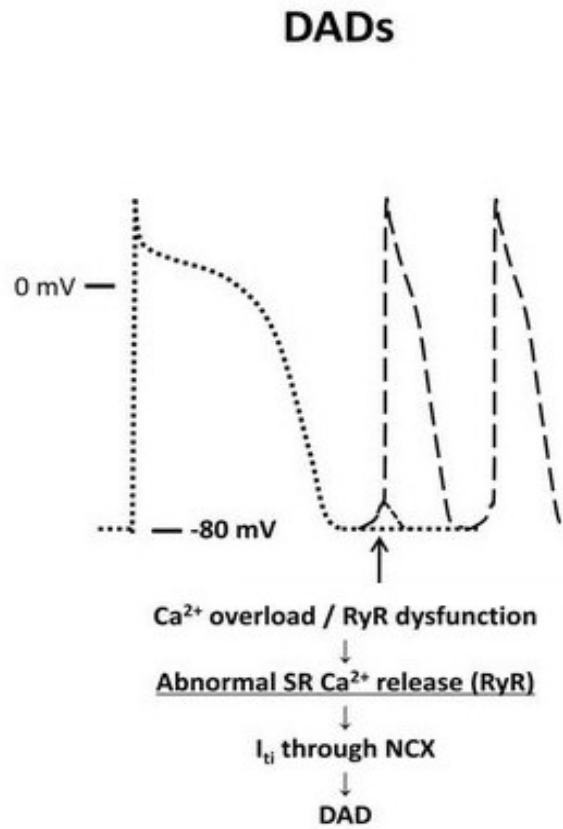
Figure. 1.16 Possible way in which ectopic rhythms contribute to cardiac arrhythmias (Iwasaki et al. n.d.).

The above trace shows how a deviation from the normal action potentials can give rise to an arrhythmogenic pathway, namely delayed-after-depolarisations (DAD). AP = action potential.
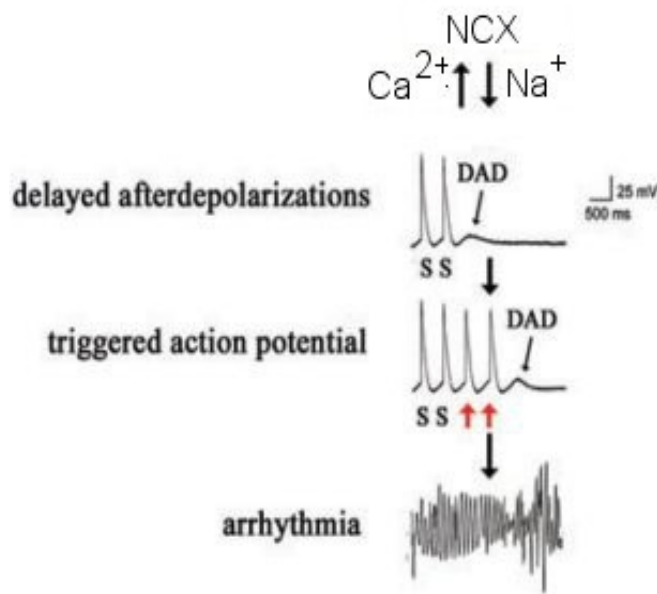
Figure 1.17   How a leaky RyR2 channel leads to arrhythmia (Braunwald 2007).

The untimely wave of calcium ions trigger the inward activating NCX, thus activating multiple DADs. These then trigger an action potential that is asynchronous with the rest of the heartbeats, triggering a premature heartbeat, thus throwing the rest of the heart-beat rhythms into dysryhthmia.

The common arrhythmia caused by faulty RyR2 molecules that gives rise to leaky channels, is a "polymorphic ventricular tachycardia". There are a number of ways ventricular tachycardia can evolve in a heart muscle although a thorough discussion of these are beyond the scope of this thesis. Whenever the heart rate is above 100 beats per minute, a patient is termed clinically to be in tachycardia. This is usually measured on an ECG trace by calculating the distance between subsequent R waves. Ventricular tachycardia usually has a common morphology, broad QRS complexes (wide distances between the Q and S wave) (Braunwald 2007).
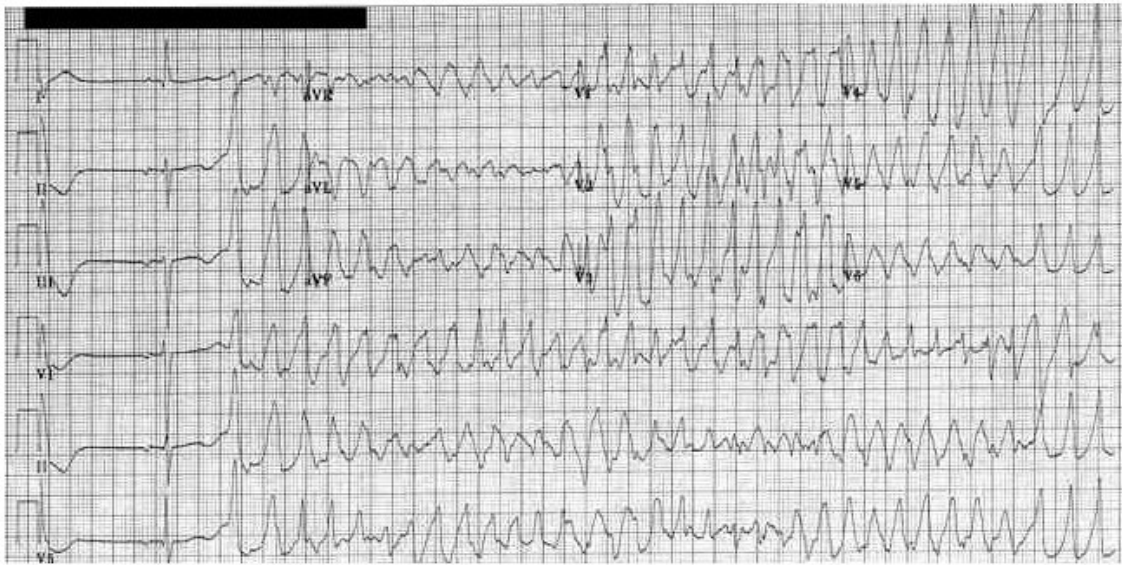
Figure 1.18   ECG displaying polymorphic ventricular tachycardia.

Ventricular tachycardia is any tachy-rhythm that has an R-R rate of more than 100 beats per minute.   Untreated VT can quickly progress to life-threatening ventricular fibrillation and once noticed on a heart trace or cardiac monitor is a medical emergency.

It is clear from the morphology of the trace, that the peaks of the QRS complexes are appearing to "dance" around the ECG-paper, hence giving it the old name of "torsade de pointes" or the polymorphic nature of the QRS complexes. The rhythm is not in sinus, as there are no atrial contraction peaks (P waves) seen at all and there is a clear beat-to-beat variation between each QRS complex. This type of ventricular arrhythmia is commonly seen when patients inherit a faulty RyR2 gene. (note: patient names and details are blackened out to preserve patient confidentiality).

1.8 Overview of the regulation of RyR channel

The previous sections of this chapter have discussed the origin of the electrical impulse, the conversion of that form of energy to mechanical contraction, and subsequently explored the genesis of normal and abnormal rhythms. The focus of the following sections is now shifted onto the regulation of the ryanodine receptor itself.

1.8.1 Domains and interaction sites within RyR

Analysis of the amino acid sequence combined with a range of experimental techniques have located a number of domains and functional sites throughout RyR. In RyR2, Ca^{2+} is the most important physiological regulator. In the absence of Ca^{2+} the channel does not open.

The RyR displays a bi-phasic response to cytoplasmic Ca^{2+}, the threshold for activation is between 10^{-8} and 10^{-7} M. Maximal activation at between 10^{-6} and 10^{-5}M, inhibition at 10^{-3}M. RyR is also regulated by luminal Ca^{2+} concentration. The Ca^{2+} activation site(s) have not been definitively identified, one potential candidate is E 3885 (Chen et al. 1998).

There are other small molecules that a play a physiological regulation role such as ATP which opens the channel and Mg^{2+}; which in millimolar concentrations, closes the channel. There is some suggestion that Mg^{2+} can compete for the Ca^{2+} activation site(s) (Gillespie et al. 2012).

There are a number of small protein modulators which predominantly inhibit the channel such as FKBP12/12.6 and CaM that act upon the cytoplasmic region of the channel. In addition, calsequestrin, triadin and junctin are associated with the luminal portion of RyR (MacKrill 1999). These proteins are discussed below and CaM is examined in greater detail later in this chapter (see section 1.8.2). In addition to these sites of interaction there are organisational domains throughout RyR. Some of these many sites and domains are shown in figure 1.19 and are discussed in the following subsections.
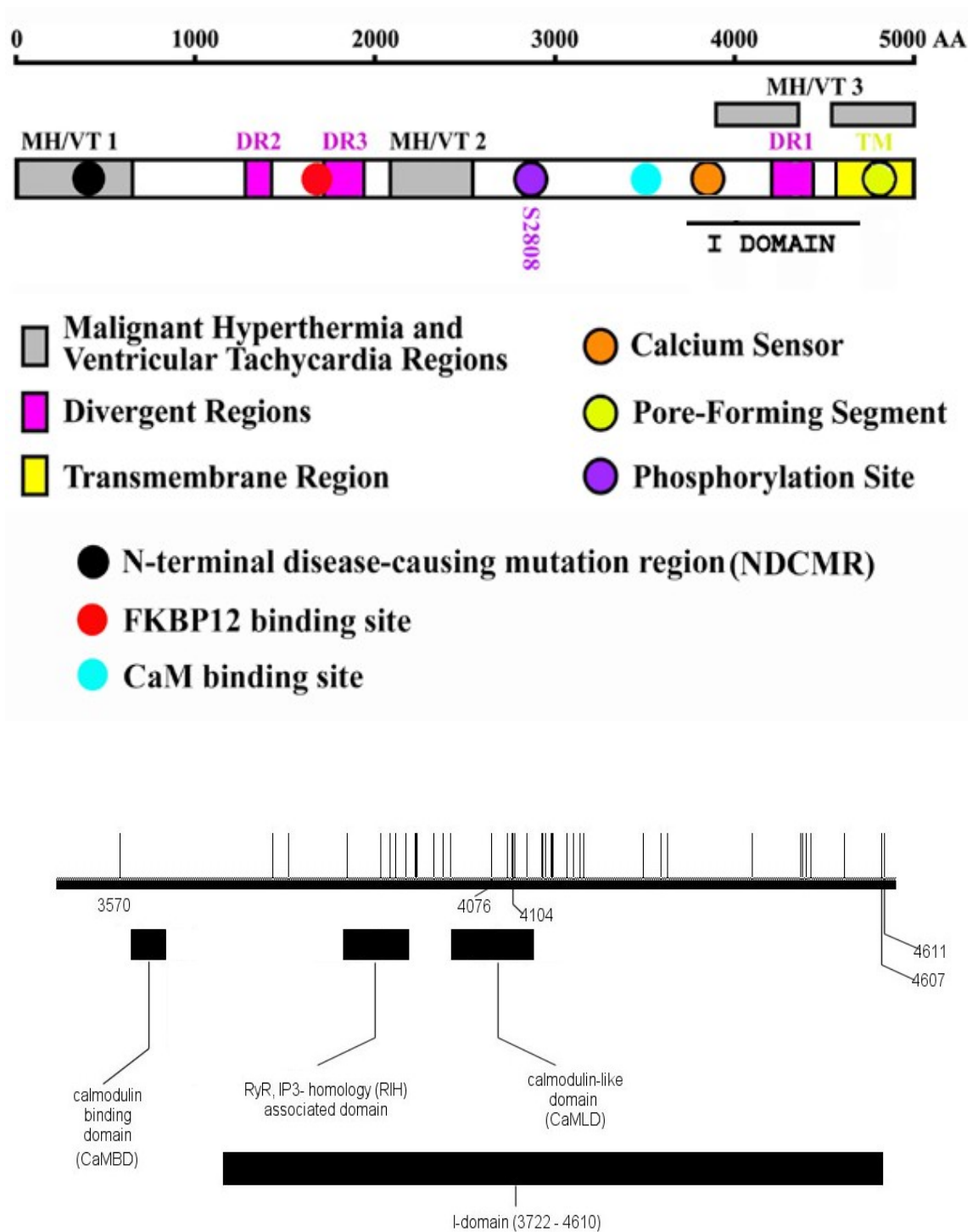
Figure 1.19   Key sequences and ligand binding sites of RyR

(Upper) Locations shown on linear map of RyR (D'Cruz, L.G., Yin, C.C., Williams, A.J., Lai 2009).

(Lower) Detailed section of RyR2 amino acid sequence displaying locations of CPVT mutations within this region and domains of interest to this thesis.

1.8.1.1 Calsequestrin, triadin and junctin

The role of calsequestrin in Ca^{2+} concentration sensitivity in ECC is mentioned earlier in this chapter. With this Ca^{2+} binding function, calsequestrin has been suggested as playing a role as part of a luminal Ca^{2+} sensor for RyR. Calsequestrin, which undergoes a conformational change with Ca^{2+} binding, forms a complex with triadin and junctin, which both have luminal and transmembrane regions. This complex appears to reduce the open probability of the RyR channel when Ca^{2+} concentration is low, however, this inhibition is reduces as the Ca^{2+} concentration increases (Györke et al. 2004).

1.8.1.2 Mutation clusters

The majority of identified disease causing mutations have been located in 3 cluster regions within RyR as shown in figure 1.14. Sequence analysis conducted on carrier families revealed the first set of known mutations in RyR2 (Priori et al. 2001). Equivalent clusters of mutations in the N and C terminal, and central regions are seen for RyR2 (CPVT and ARVD2) and RyR1 (MH and CCD). With specific mention to CPVT mutations in RyR2, the C terminal domain region is usually further split into the cytoplasmic I domain and transmembrane domain which also includes the C terminal region itself (Thomas et al. 2006).

1.8.1.3 Divergent regions

Within the RyR amino acid sequence, 3 areas of conspicuous variation were found which were labelled divergence regions D1, D2 and D3. These regions are, for RyR1 and RyR2 (in brackets), amino acids, D1 4254-4631 (4210-4562), D2 1342-1403 (1353-1397), and D3 1872-1923 (1852-1890) (Sorrentino & Volpe 1993). It is thought that D2 plays an important role in excitation-contraction coupling (ECC). In mutated mouse skeletal myocytes, in which the D2 region had been deleted, RyR was still able to function as a Ca^{2+} release channel but ECC activity was lost. Expression of wild type RyR1 restored ECC activity in these cells however, if RyR2 was expressed instead there was no return of ECC function (Yamazawa et al. 1997).

D3 has also been suggested to play some role in ECC. Through yeast two-hydrid experiments, D3 of RyR1 has shown to interact with a region of the L type Ca^{2+} channel. When D3 of RyR2 was substituted occasional, weak interactions were observed (Proenza et al. 2002). D1 appears to play a role in the channel's sensitivity to Ca^{2+} and caffeine concentrations. Deletion mutants, and chimeras of RyR 1 and 2 displayed altered responses to those of wild type channels in ryanodine binding experiments. Complete deletion of D1 did not remove caffeine or Ca^{2+} responses in the channel indicating that D1 does not contain binding sites for these molecules (Du et al. 2000). Tagging the divergent regions with green fluorescent protein, subsequent cryo-EM and comparison to the wild type during 3D reconstruction has suggested locations of these regions on the model of RyR.

1.8.1.4 FKBP 12/12.6

FK506 binding proteins (FKBP) are members of the immunophilin family which are binding targets of immunosuppressive drugs, in this case tacrolimus (initially designated FK506). They have peptidyl-prolyl isomerase activity catalysing the cis-trans isomerisation of peptide bonds. As regards to RyR, FKBPs were first identified when they were co-purified with RyR1. These FKBPs were named for their molecular masses, FKBP12 and FKBP12.6. One FKBP12 was shown to interact with one RyR1 subunit (Jayaraman et al. 1992), so that four FKBP molecules are associated with the functioning channel homotetramer. RyR1 associates with both FKBPs though RyR2 appears to preferentially bind FKBP12.6. However, a recent study indicates that there may be variation in FKBP type binding preference for RyR2 across species (Zissimopoulos et al. 2012a). A FKBP12 binding site is shown in figure 1.19, this site has not been consistently reported in the literature and other sites within the RyR sequence, in particular for RyR2, have been suggested (Zissimopoulos & Lai 2005a), (Zissimopoulos & Lai 2005b).

Both FKBPs have high affinity binding with RyR in either the open or closed state, however there is greater affinity for the closed channel (Jones et al. 2005). FKBP12.6 binding to RyR2 stabilises the closed conformation of the channel so inhibiting RyR2 activity. While there is some disparity in reported results, it is suggested that these may be due to other effects on the EC coupling machinery, a reduction in FKBP12.6 levels

leads to greater Ca^{2+} transients and more muscle contractions (Xiao et al. 1997). As expected, an increase in FKBP12.6 leads to reduction of channel activity shown by decreased calcium spark frequency, amplitude and frequency (Gómez et al. 2004).

1.8.1.5 Phosphorylation sites

A number of phosphorylation sites, their interactions with PKA and CaMKII and the associated changes in physiological function in RyR have been previously discussed in section 1.6 of this chapter.

1.8.1.6 Interacting (I) domain

This region has been studied in RyR2 and proposed to be involved in interdomain interaction and, as such, has been named the interacting (I) domain (George et al. 2004a). The I domain of RyR2 comprises of two distinct domains, the interaction domain located between residues 3722-4353 and the modulatory domain, containing two putative TM domains located at residues 4353-4610 (George et al. 2004a), (George et al. 2006). It should be noted that this region contains a high number of CPVT mutations. Selected CPVT mutations studied in the interacting (I) domain were seen to trigger increased signal variability in Ca^{2+} release and destabilise inter-domain interaction (George et al. 2006).

The interaction of the I domain may been through this region being able to behave like a hinge. It has been shown that the regulatory events at the cytoplasmic domain are not appropriately conducted to the pore domain following RyR2 mutations in the I domain (George et al. 2004a), (George et al. 2006). It is thought that a number of hydrophobic regions in the I domain are involved in the regulatory interactions that this section of RyR2 performs (George et al. 2006).

Homologous regions have been identified in RyR1 and RyR3. In addition, within the I domain, a series of amino acid repeat sequences that are present in IP3Rs and RyR have been identified. This region of repeats has been termed the RyR and IP_3R homology (RIH) associated domain, and has been proposed to possibly have protein binding or related function (Ponting 2000). Within the I domain a region containing 2 EF hand

motifs has been identified. This region has been termed the CaM like domain (CaMLD) and is discussed in later in this chapter.
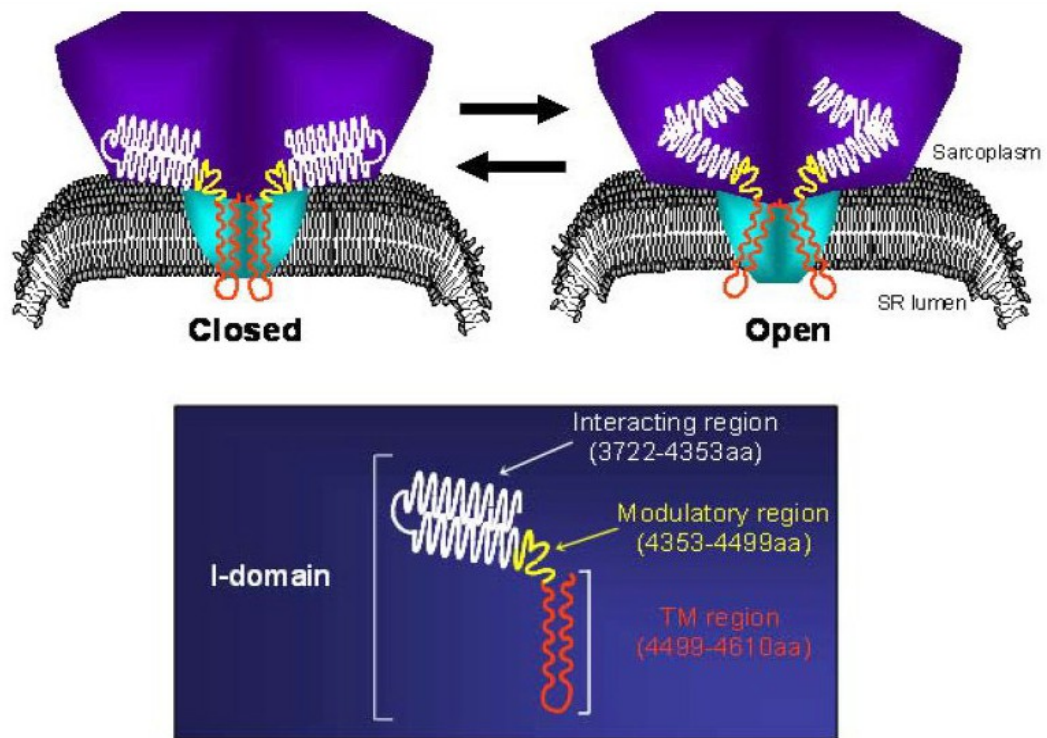
Figure 1.20   The RyR2 I Domain intrinsically controls RyR2 opening through conformational rearrangement (Jundi 2010).

Schematic diagram, showing that conformational re-ordering within the I domain is associated with RyR2 channel opening and closing, based on data obtained from fluorescence resonance energy transfer (FRET) studies of cytoplasmic (light grey) and transmembrane domain (dark grey) interactions. Functionally distinct sub-regions within the I domain are predicted from experimental data and computational analysis.

1.8.1.7 CaM Binding Site

A CaM binding site (CaMBD) has been identified in RyR1 between amino acids 3614-3643. An equivalent, highly conserved region was found in RyR2 between residues 3583-3603. Further details of this site and the potential significance of its flanking regions are discussed later in this chapter.

1.8.1.8 Transmembrane domains

The transmembrane domains are clearly the key region of the channel forming the pore within the SR membrane. A number of models exist predicting the number of transmembrane segments present in this domain. A site directed antibody approach appeared to support at least 4 segments of proposed models and considered the possibility of an even number up to a total of 10 segments. Additional data from this experiment indicated that the transmembrane domain was supported by a relatively short C terminal and longer N terminal cytoplasmic domains (Grunwald & Meissner 1995).

1.8.2 CaM and RyR regulation

CaM regulates the RyR channel both directly and indirectly. A CaM interaction of note is the activation of CaMKII. A range of experimental techniques reported in the literature indicates that CaMKII activity leads to a general increase in activity from Ca^{2+} signalling through to related downstream events such as ECC. This increased response is more Ca^{2+} sensitive and may occur through a response to increased Ca^{2+} levels (MacQuaide et al. 2007) but operates in a less well regulated fashion than the non-phosphorylated state. It should be noted that some of these effects are due to CaMKII having other targets in addition to Ser 2815. These include the L type Ca^{2+} channel, linked to increased Ca^{2+} signalling, and the phospholamban associated with SERCA, found to increase SERCA activity in preparation for the next excitation-contraction event (Niggli et al. 2013).

The interaction of CaM with the CaMBD of RyR is of particular interest to this thesis. CaM is known to inhibit the three RyR isoforms at Ca^{2+} concentrations greater than 1 μM. At submicromolar concentrations RyR1 and RyR3 are stimulated by Ca^{2+} (Tripathy et al. 1995). However, at these lower concentrations RyR2 was either unaffected (Fruen et al. 2000) or inhibited (Balshaw et al. 2001) by CaM.

CaM is known to bind with high affinity to α helical sections of protein. Several variations of CaM interaction motifs have been proposed from this structure and lead to the identification of CaM binding sites in a number of proteins that contain the "IQ

motif" (named after the first 2 amino acids in the sequence motif); seen in many proteins including RyR and myosin (Rhoads & Friedberg 1997).

Tryptic digests on RyR1 revealed that the arginine residues 3630 and 3637 were protected by apo and Ca-CaM binding (Moore et al. 1999). Both forms of CaM were shown to bind to a synthetic peptide matching amino acids 3614-3643. Deletion of the last 9 residues removed the ability to bind apo-CaM but not Ca-CaM (Rodney et al. 2001). Fluorescence resonance energy transfer (FRET) based experiments have subsequently agreed with the 3614-3643 location of a CaM binding site in RyR1 (Cornea et al. 2009). A highly conserved site was identified in RyR2, amino acids 3583-3603. Deletion of this region resulted in the prevention of CaM binding to RyR2 (Naohiro Yamaguchi et al. 2003).

Mutational studies conducted on the CaM binding site of RyR1 and RyR2 indicated that regulatory CaM binding is to the one highly conserved site (Naohiro Yamaguchi et al. 2003). Substitution of the non-conserved N and C terminal flanking regions to the CaM binding site in RyR1 and RyR2 resulted in the Ca^{2+} concentration response to CaM as expected for the opposite isoform. This result suggested that the Ca^{2+} concentration response to CaM by the RyR isoforms is as a result of the difference in sequence in these flanking regions. Further experimentation pointed to residues in the C terminal flanking region being responsible for the CaM inhibition of RyR2 (Yamaguchi et al. 2004).

S100A1, a member of the S100 protein family which contains two EF hand motifs, has been shown to bind to the same site on RyR1 as Ca-CaM. This S100A1 binding site in RyR consists of 12 residues identified in the highly conserved CaM binding site discussed above. In RyR1 this site is amino acids 3616-3627. An NMR structure of Ca^{2+} bound S100A1 interacting with a peptide of this 12 amino acid RyR1 binding site (PDB entry 2K2F) has been determined (Wright et al. 2008).

Comparison of this structure with Ca-CaM bound to a peptide of the RyR1 binding site (PDB entry 2BCX, (Maximciuc et al. 2006)) indicates 6 residues within the 12 amino acid peptide that have favourable interactions with S100A1 are involved with Ca-CaM binding. The location of these key binding residues within S100A1 binding site

contained within the CaM binding site of RyR supports the observation that Ca-CaM can still bind if the last 9 residues of the CaM binding site are deleted (Rodney et al. 2001).

Cryo-EM studies have suggested 2 different neighbouring sites for apo-CaM and Ca-CaM binding in RyR1. However, a 33 ± 5 Å distance between the centres of the 2 bound forms of CaM was observed (Samsó & Wagenknecht 2002). A number of possibilities were suggested by the authors to explain the distance between the 2 types of CaM that, the previously discussed data suggests, share one conserved binding site in RyR. Most of these suggestions involved the movement of CaM and/or the domains involved around the CaM binding site. Additional cryo-EM studies have been conducted which suggests that there are 2 overlapping binding sites for CaM, in RyR1, within the reported 30 amino acid binding site.

The involvement of Ca^{2+} appears to play a key role in the binding process. The conformational change triggered by Ca^{2+} binding to CaM appears to cause CaM to shift in its binding position and orientation. It has also been suggested that the binding of Ca^{2+} to RyR1 may well promote a conformational change in RyR that alters that the CaM binding site. A mutant form of CaM unable to bind Ca^{2+} was found to consistently bind to the apo-CaM position even in the presence of Ca^{2+}. As regards to RyR2, only one CaM binding location was observed by this set of cryo-EM experiments. Apo-CaM, binding to RyR2, appears to bind in the equivalent location of Ca-CaM in RyR1 (Huang et al. 2013).

This data would appear to agree with the observation that CaM binding to RyR2 inhibits the channel even at the lower Ca^{2+} concentrations that are non-inhibitory for the other RyR isoforms. It was suggested that this may form the basis of a structural regulation model in which there is an "inhibitory" and "activating" CaM binding location with the conserved binding site. The inhibitory location only appears to be functional in RyR2; it may well be that the flanking regions, particularly those downstream of the binding site, reported by Yamaguchi *et al (Naohiro Yamaguchi et al. 2003)*, (Yamaguchi et al. 2004)are somehow involved in this process.
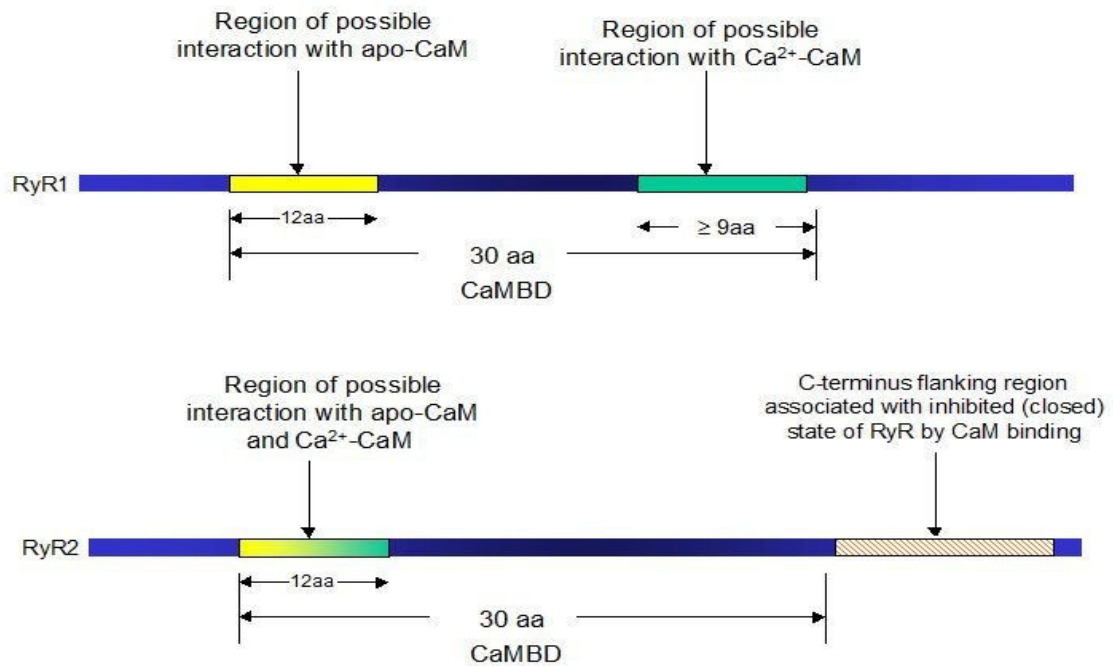
Figure 1.21   Proposed CaM binding locations within CaMBD for RyR1 and RyR2.

Schematic showing summary of literature for the proposed CaM lobe interaction sites within conserved CaMBD of RyR1 and RyR2. Where known, sizes of locations given in number of amino acids.

1.8.3 CaMLD

A region of rabbit RyR1, amino acids 4064-4210, was found to have CaM like structure containing 2 putative EF hands, binding Ca^{2+} and displaying a Ca^{2+} induced conformational change. Peptides of this region were found to interact with an identified CaM binding site, 3614-3643, within RyR1. Owing to this behaviour the region was termed the CaM-like domain (CaMLD) (Liangwen Xiong et al. 2006). Antibodies raised against CaMLD and the CaM binding site both significantly inhibited ryanodine binding to RyR1 suggesting interaction between the two domains is required for channel activation. Experiments conducted in which the CaM binding site was fluorescently labelled displayed a Ca^{2+} dependent interaction with CaMLD (Gangopadhyay & Ikemoto 2008). It was found in stimulated neonatal cardiomyocytes that CaM dissociates from RyR which lead to an increased rate of Ca^{2+} release events and eventually hypertrophy of the cells. When an anti CaM binding site antibody, so

preventing the interaction of CaMLD with the binding site, was included these events did not occur. It was suggested that the uncontrolled interaction between the 2 domains could lead to the disease state of the cardiomyocytes (Gangopadhyay & Ikemoto 2011). Due to the importance of this interaction, the focus of this thesis involves the study of this calcium binding domain within the I domain of RyR2. Structural and functional aspects of this domain are further examined in chapters 3 and 5 respectively.

1.9 EF hand motifs

In terms of RyR function the importance of the CaMLD/CaMBD association is highlighted by the fact that the CaMLD portion of the molecule contains a number of CPVT point mutations which are also prevalent in the sequence flanking this domain (see figure 1.19, lower panel). As previously stated, within CaMLD two EF hand motifs have been identified (Xiong, 2006). To further understand the potential mechanism of interaction between CaMBD and CaMLD, and the potential impact of CPVT point mutations within the CaMLD region requires the EF hand motif to be examined.

The presence of Ca^{2+} binding proteins (CaBPs) has allowed greater control in the propagation and site of effect of Ca^{2+} signalling (Case, 2007). It is probable that these CaBPs act as Ca^{2+} sensors and they transduce their effect via the EF hand motif. This motif enables the binding of Ca^{2+} to be converted into a functional output through the binding protein (Grabarek 2005).

The structure of a carp muscle CaBP protein, parvalbumin, was determined by x-ray crystallography. Within this structure a helix-loop-helix motif with Ca^{2+} bound in the loop was identified, a neighbouring repeat of this motif also containing Ca^{2+} was seen. As this motif could be visualised by a right hand with the thumb and index finger extended at a near right angle, the other fingers clenched (see Fig. 1.22), so the term EF hand was coined (Kretsinger & Nockolds 1973).
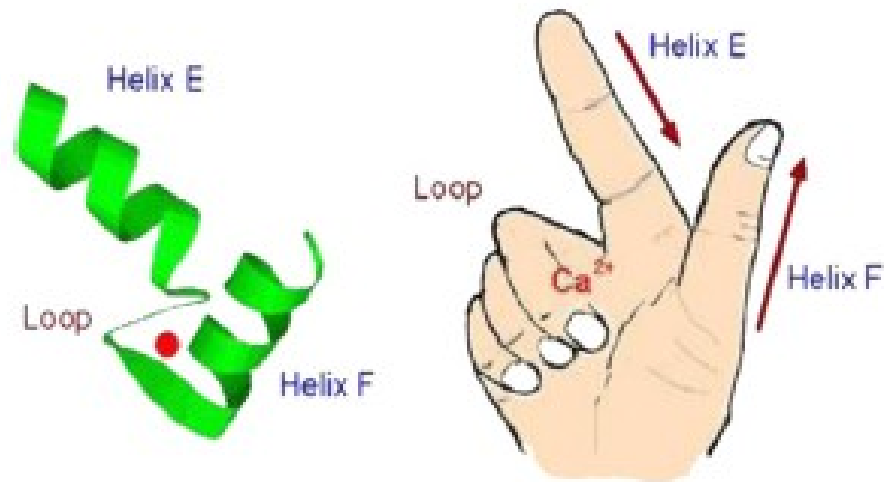
Figure 1.22   EF hand motif (Haiech et al. 2011).

The second of the pair of EF hand motifs is formed by helices E and F, and the loop between them. The index finger representing helix E points towards the N terminal of the protein. The clenched fingers describe the loop to which Ca^{2+} is bound. The thumb representing helix F points towards the protein C terminal. The first EF hand motif is formed by the helices C and D containing a loop. The two motifs are separated by a short linker.

Since the discovery of parvalbumin, thousands of distinct EF hand sequences have been reported. It is not clear how all these alterations between the sequences result in the variations in properties of the EF hands and the function of their associated CaBP. However, most known functional high affinity Ca^{2+} binding EF hands share a basic overall sequence which relies on the attraction of Ca^{2+} to a number of side chain oxygen molecules, commonly 6 or 7, present in the loop (Grabarek 2005), (Nelson et al. 2001).
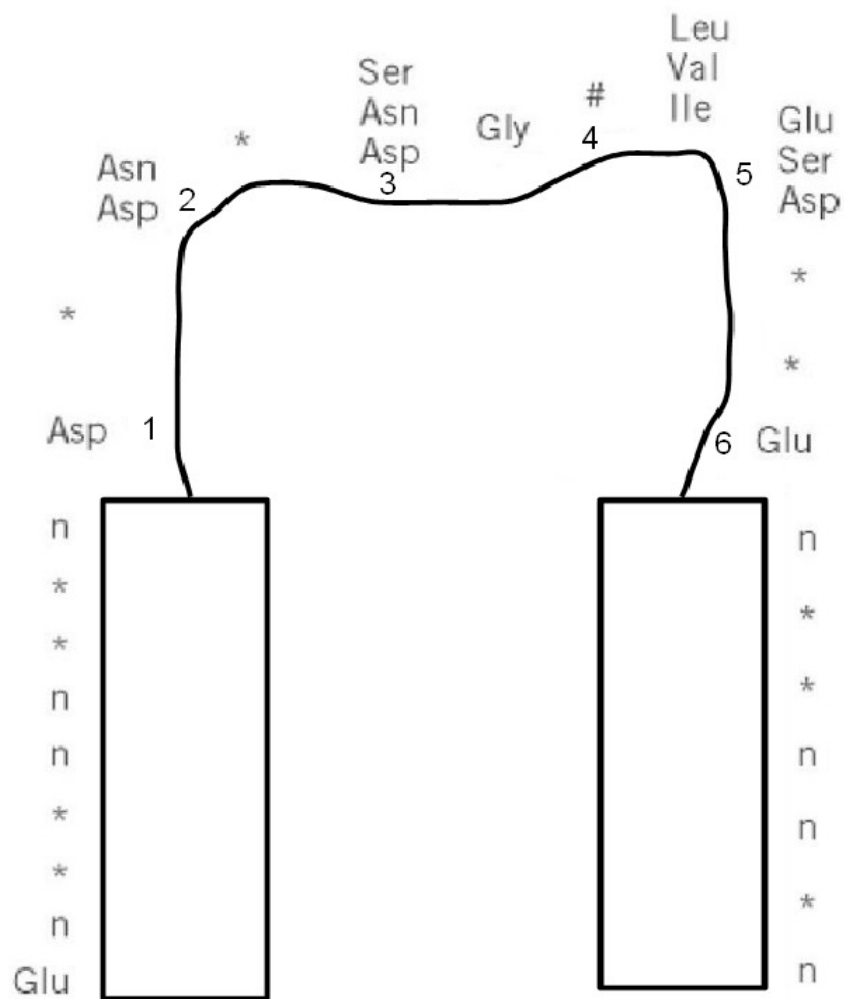
Figure 1.23   Canonical sequence of EF hand.

*Schematic image of helix-loop-helix motif displaying shared features of residues found in location of common EF hand. Nomenclature of canonical EF hand taken from parvalbumin where helix E is joined to helix F by a loop. Positions of hydrophobic residues in the helices are shown by n. The * positions are non-conserved and can be occupied by any amino acid. Commonly conserved amino acids, including variations, are shown. In the loop there are six conserved residues as indicated by the numbered residues. These positions are sometimes labelled X, Y, Z, -Y, -X and -Z, this system is based on the first three Ca^{2+} coordinating residues defining the axes of ligand geometry. Positions 1 (X), 2 (Y), 3 (Z) and 6 (-Z) are side chain oxygen ligand amino acids. The -Y position is indicated by a # symbol in the loop, this residue is a backbone carbonyl ligand. There are two other shown residues in the loop. One of these is a glycine which is not always conserved. The other is between positions 4 and 5, in earlier studied examples this was often an isoleucine residue.*

The canonical EF hand, as shown above (figure 1.23), consists of a twelve residue loop. The first coordinating ligand, position 1 (X), is commonly an aspartate and the last, position 6 (-Z), is predominantly a glutamate which contributes two oxygen atoms serving as a bidentate ligand. In addition with positions 1 (X), 2 (Y), 3 (Z) and 6 (-Z), the contributing two oxygen molecules from positions 4 (-Y) and 5 (-X), mean that bound Ca^{2+} is attached to 7 oxygen atoms in pentagonal bipyramid. 1 (X) and 5 (-X) act as the cap and base with the other positions forming the pentagon. The centrally positioned 4 (-Y) residue binds with the main carbonyl chain oxygen atom as opposed to a side chain oxygen. The immediate downstream residue is hydrophobic, prevailing valine, leucine or isoleucine as shown in the figure. The 4 (-Y) residue and its neighbour are part of a short β sheet. In paired EF hands, as seen for parvalbumin, these two residues form two hydrogen bonds with their equivalent residues in the matching β sheet so connecting the two loops (Strynadka, 1989). These two β sheets linked by hydrogen bonding are referred to as the EF hand β scaffold (Grabarek 2005).

There are examples of the 6 (-Z) position glutamate being substituted for another amino acid however there is usually compensation by insertion of additional residues in the N terminal of the loop. This restores the coordination geometry of the Ca^{2+} binding (Grabarek 2005). Mutational work has been conducted on this glutamate residue replacing it with an amino acid only able to provide 1 oxygen atom. The reduction of an oxygen is thought to reduce the stability of the orientation of the binding Ca^{2+} which is observed in a reduced affinity for Ca^{2+} (Maune et al. 1992). Mutations have been studied in the other residues of the loop in calbindin D9K. Glutamate and aspartate residues were replaced individually and in multiples by their amide equivalents glutamine and asparagine. Individual mutations in both ligating and non ligating residues in the loop caused only small structural differences though there was some accumulation for multiple mutations. Non ligating mutations were found to reduce Ca^{2+} affinity as well as those in coordinating positions (Linse et al. 1991).

Additional mutations in the hydrophobic residues, which have no direct contact with bound Ca^{2+}, in calbindin D9K have been studied. Mutations were selected to alter the side chain volume of hydrophobic residues. Structural differences were only observed when the mutated amino acid had grossly different structure to that of the wild type. The mutations were found to significantly decrease Ca^{2+} affinity and increase the rate of

release of bound Ca^{2+} (Kragelund et al. 1998).

There are no known monomeric CaBPs, a large number of known CaBPs contain paired EF hands, for example troponin C. Troponin T binding to troponin C provides a good example of how a minimal EF hand dimer connected by a helix binds to a partner protein, in this instance troponin I (see figure 1.24 below).
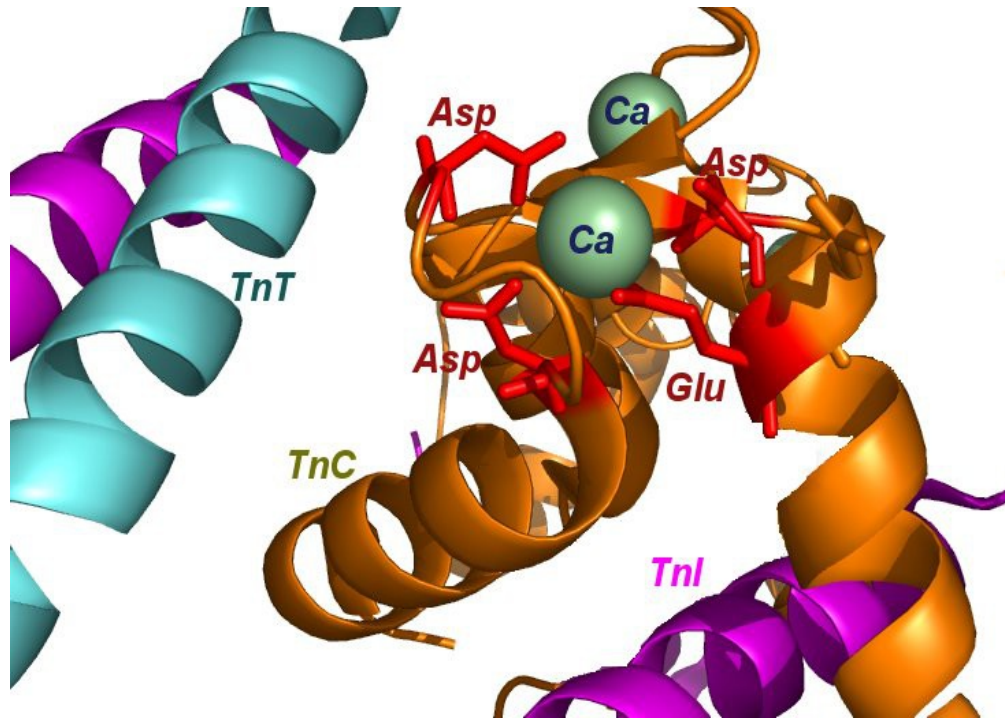
Figure 1.24   Cardiac muscle troponin complex showing interaction of EF-hand with partner proteins.

The cardiac muscle troponin structure showing the acidic residues within the loop region of the EF hand in Troponin C (ochre coloured) holding the calcium ion in position. The TnI molecule is an extended helix, clamped between the helices of the EF hand of troponin C. Troponin T is also an extended helix which swaps interactivity with troponin I in the organisation of the thin-filament. (Figure generated in pymol using coordinates from PDB: 1J1E).

In these structures there were two Ca^{2+} bound to the C terminal domain of the protein while the two EF hand motifs of the N terminal were unbound. It was observed that the two domains had very different conformations. The C terminal was in an open conformation which presented a large, solvent exposed hydrophobic pocket. The N

terminal had a closed conformation in which the domain helices were tightly packed against each other (Herzberg & James 1985). It was proposed that Ca^{2+} binding triggered the conformation from the closed to the open state. The exposed hydrophobic pocket of the open conformation acting as the target binding site (Herzberg et al. 1986). This transition model has been named the HMJ model, the name derived from the first letter of the authors surnames, Herzberg, Moult and James. Subsequent biochemical and structural data has supported this model. As structural data are required to fully confirm this type of conformational change it is unknown how extensive this model is amongst other EF hand CaBPs.

Other proteins, such as CaM, contain two pairs of EF hands. A few proteins contain odd numbers of EF hands, for example parvalbumin contains three, one of which functions in an ancillary role to the binding pair. As the EF hand motifs in the heterodimer have variations in their amino acid content, as shown in canonical EF hand motif (figure 1.23), this confers slightly different characteristics upon the functioning pair, most importantly Ca^{2+} binding affinity. It should be noted that under conditions of high concentration of magnesium ions (Mg^{2+}) and low Ca^{2+}, some EF hand motifs will bind Mg^{2+}. This binding impacts on the behaviour of the related CaBP though it is not known what component of the sequence of the binding loop confers Mg^{2+} binding. These variances allowed for greater control of Ca^{2+} signalling through the development of domains with different Ca^{2+} conformation responses that had specific binding targets. The N and C terminals of CaM are a well researched example of this difference in paired EF hand domains (Grabarek 2005), (Nelson et al. 2001).

The EF hand motif heterodimers are linked by the hydrogen bonds in the short section of β sheet and this linkage is further maintained by many hydrophobic contacts between the helices from the residues in the n positions (as shown in figure 1.23). The Ca^{2+} bound form receives extra strengthening from Ca^{2+} ligand interactions and hydrogen bonding between polar groups and closely surrounding water molecules (Strynadka & James 1989). The stability of the paired EF hand heterodimer has been compared to synthetic pairings of homodimers. Chemical denaturation studies combined with circular dichroism and NMR conducted on troponin C revealed that the heterodimer is more easily formed and has greater stability (Shaw et al. 1994). The EF hand scaffold structure has been found to occur without the support of the hydrophobic contacts.

However, the contacts between the helices greatly strengthen the overall structure of the EF hand pair (Wójcik et al. 1997).

There are two general functional divisions within the EF hand CaBPs. Firstly, signal modulators, including calbindin D9K and parvalbumin, which affect properties of the Ca^{2+} signal. These CaBPs do not undergo Ca^{2+} binding induced structural rearrangements or target interaction. This class is usually involved in calcium homoeostasis. The second type, of greater interest to this project, are calcium signal sensors. This class, which includes CaM and troponin C, undergo a calcium induced conformational changes that allows for binding to their target site (Nelson et al. 2001).

As stated previously, CaM is a well studied example of an EF hand containing CaBP. Crystal structures of CaM bound with calcium (Ca-CaM) and calcium free (apo-CaM) have been obtained showing that CaM follows the HMJ transition model (Finn et al. 1995), (Zhang et al. 1995). These structures combined with biochemical data display the differences between the N and C terminal paired EF hand containing lobes of CaM. The N lobe has been shown to dissociate Ca^{2+} more quickly than the C lobe. This has been reflected in the binding affinity with the C terminal having a higher affinity than the N terminal (Black et al. 2006). The Ca^{2+} dissociation rate and binding affinity can be affected by the target protein so further fine tuning the Ca^{2+} signal. Target proteins can display preferential binding for Ca-CaM or apo-CaM. Those selecting Ca-CaM cause a slower dissociation rate of Ca^{2+} from the bound CaM. However, proteins preferably binding apo-CaM tend to release CaM as the Ca^{2+} concentration increases. It is proposed that in cardiac myocytes that the low intracellular Ca^{2+} concentrations may not be sufficient to favour Ca-CaM over apo-CaM. As such apo-CaM may be bound near local sources of calcium, as Ca^{2+} concentration increases so CaM is released and goes onto bind Ca-CaM preferring sites (Saucerman & Bers 2012). Within cardiac myocytes one of the major proteins that CaM interacts with is RyR as discussed earlier in this chapter.

The use of mutational studies in examining the EF hand has been discussed. However, functional mutations related to cardiac disease have been recently observed within CaM including CPVT linked mutations (Nyegaard et al. 2012). Some of the studied mutations are associated with a disruption of the CaM inhibition of RyR2 which results in more

erratic Ca^{2+} release frequency and duration leading to impaired cardiac function. Cardiac hypertrophy and earlier than expected death has been observed in mouse models of these mutations (Arnáiz-Cot *et al*. 2013), (Yamaguchi et al. 2013). Additional mutations associated with long QT syndrome and cardiac arrest have been located within the conserved residues of the EF hand loop of CaM (Crotti, 2013). As these mutations are observed in CaM it is worth considering the locations of the known CPVT mutations with relation to the EF hand motifs of RyR2. The CPVT point mutation E4076K is contained in the loop of the 2nd EF hand motif of RyR2 (see figure 1.25 below) when compared to the canonical EF hand shown in figure 1.22. While this mutation is not in one the conserved positions of the loop it is not clear how it will effect the function of the EF hand. During the time of this thesis this is the only known CPVT mutation contained within the EF hand motifs of RyR2.

```
RyR2 EF2:    ETEFLLSCAETDENETLDYEEFVKRFHEP
Canonical:     En**nn**nD*N*?LD**En**nn*n
```

Figure 1.25   Sequence alignment of RyR2 2nd EF hand motif against canonical EF hand motif.

*RyR2 amino acid sequence shown as top line with closest matching canonical sequence underneath. Sequence within EF hand loop section is highlighted by box. CPVT point mutation E4076K location is shown in red in RyR2 sequence. As in figure 1.23, * indicates non conserved position, n indicates predicted hydrophobic residue within α helix either side of loop. Threonine residue present in RyR2 where ? symbol is used to represent expected Ser (or Asp or Asn) residue in canonical amino acid sequence.*

As illustrated above, while there is increasing knowledge about RyR and the role of mutations upon the channel's function there still much to be understood. Certain regions within RyR have undergone more scrutiny than others. The advancement in comprehending the role of these regions has increased as shown by the structures obtained for the N terminal region of RyR. The I domain has been studied and shown to be involved in domain interaction which is key for the function of the channel. Within the I domain, a CaM like domain which interacts with a CaM binding site has been identified. The interaction between these two domains is of particular interest to this project.

1.10 Aims of project

It is planned to focus upon the CaMLD which is proposed to interact with the CaM binding site. Figure 1.26 (see below) summarises what is currently known and indicates potential interactions between CaMLD and CaMBD, within the context of the proposed model of I domain regulation of channel gating. To test the role of CaMLD and CaMBD within this model (working hypothesis), it is intended to adopt a structure/function approach to investigate the relationship between the two domains and the potential effect of CPVT mutations within this region. Both regions will be made as recombinant proteins and used in various interaction studies. The recombinant proteins will also be characterised and used for crystallisation studies. Homology modelling of the two individual domains, and also both domains and their linking region will be undertaken to assess key residues essential to their interaction.
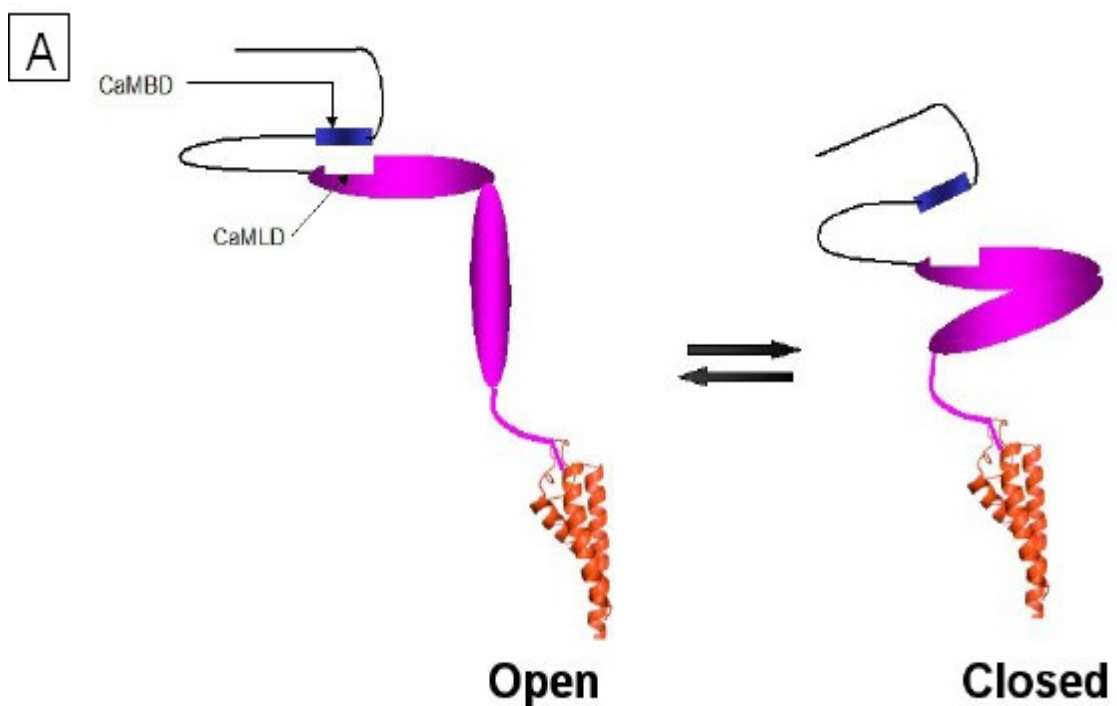
Figure 1.26a   Overview of proposed interactions between RyR2 central region and pore domain.

Magenta ovals and linker to pore represent I domain. Red pore section from KCSA pore x-ray determined structure. CaMBD and CaMLD are labelled.

(A) Basic overview of proposed interaction transitions of I domain, and CaMBD and CaMLD as RyR2 channel goes from open to closed state.
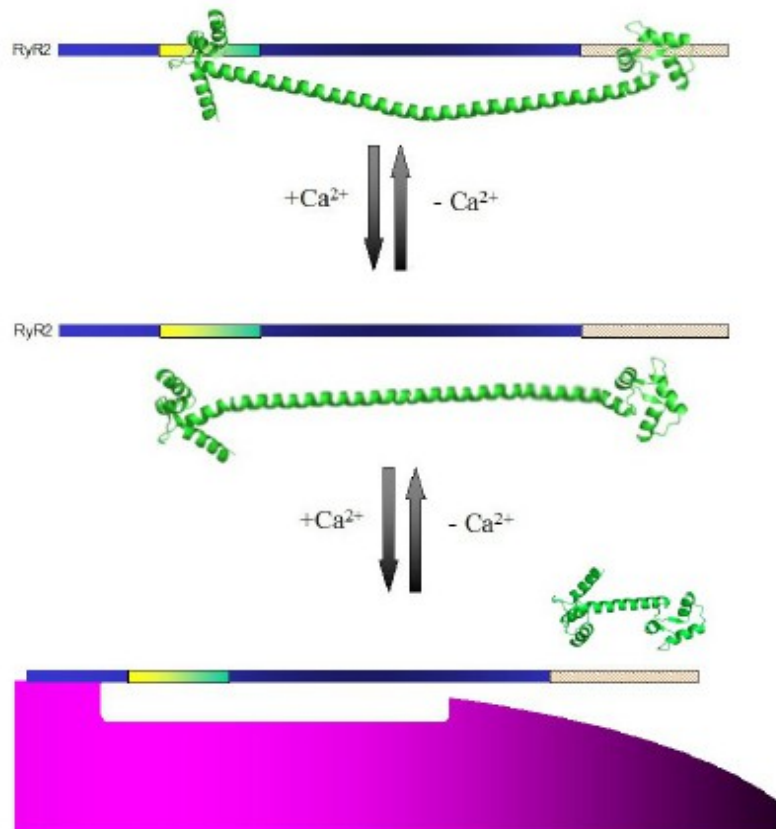
Figure 1.26b Overview of proposed interactions between RyR2 central region and pore domain.

Section of I domain (magenta) with indented region representing location of CaMLD. Within CaMBD (blue), CaM lobe interaction location (yellow-green) and C terminus flanking region (spotted orange) are shown.

(B) Focus on interaction between CaMBD and CaMLD from panel a. Basic proposed mechanism for disassociation of CaM from CaMBD with increasing Ca^{2+} concentration and transfer to binding of CaMLD to CaMBD. In this figure the linking region between the N and C terminal lobes of CaM is extended for graphical purposes to illustrate the possibility of the two lobes of CaM interacting with different locations of RyR2.

Chapter 2:

Materials and Methods

2.1 Materials

2.1.1 General Laboratory Reagents and Chemicals

All reagents and chemicals were of suitable or higher grade for the conducted experiments and were obtained from Sigma, Fisher or Calbiochem unless otherwise stated. Equipment for polyacrylamide gel electrophoresis (PAGE) was obtained from BioRad unless otherwise stated. All solutions were stored at room temperature, unless otherwise stated. All filter sterilization was through 0.2 μm filters (Sartorius).

2.1.2 General Biological Reagents

- CaCl$_2$, 1M stock: a 1 M solution was filter sterilised and stored at 4°C.
- DNA loading buffer, 5X stock: 50% v/v glycerol, 0.25% w/v orange G, 5x TAE.
- EGTA 1 M stock: Dissolved in H$_2$O at pH 8.0. Filter sterilised and stored at 4°C.
- IPTG, 1 M stock: prepared and filter sterilised as required.
- Molecular weight DNA markers: obtained from NEB and Sigma
- TAE, 50X stock: 2 M Tris, 2 M acetic acid, 50 mM EDTA

2.1.3 Protein Biochemistry Reagents

2.1.3.1 General reagents

- Ammonium persulfate: 10% w/v.
- Molecular weight protein markers: pre-stained Kaleidoscope (Broad Range) markers obtained from BioRad and pre-stained Colorplus (Broad Range) markers obtained from NEB.
- EDTA free protease inhibitor cocktail: obtained from Roche.
- Protein loading buffer, 5X stock: 0.2 M Tris, 10% w/v SDS, 40% v/v glycerol, 0.05% v/v 2-mercaptoethanol, 0.25% w/v bromophenol blue, pH adjusted to 6.8.

- Running buffer, 5X stock: 15 g/L Tris, 72 g/L glycine, 5 g/L SDS.
- Semi-dry transfer buffer: 48 mM Tris, 39 mM glycine, 0.0375% w/v SDS, 20% v/v methanol.

- TBS, 10X Stock: 0.2 M Tris, 1.37 M NaCl, pH adjusted to 7.5.

- TBS-T buffer: 1X TBS, 0.1% v/v Tween-20.

- Tris, 0.5 M: a 0.5 M solution was pH adjusted to 6.8.

- Tris, 1.5 M: a 1.5 M solution was pH adjusted to 8.8.

- Coomassie stain: 48% v/v MeOH, 42% v/v H_2O, 10% v/v glacial acetic acid, 0.2% w/v Coomassie brilliant blue R-250.

- Coomassie destain: 48% v/v MeOH, 42% v/v H_2O, 10% v/v glacial acetic acid.

2.1.3.2 Protein purification reagents and specific buffers

- Ni tag standard buffer: 93.2 mM Na_2HPO_4, 6.8 mM NaH_2PO_4, 300 mM NaCl, pH adjusted to 8.

- Ni tag resuspension buffer: 93.2 mM Na_2HPO_4, 6.8 mM NaH_2PO_4, 300 mM NaCl, 60 mM imidazole, 0.002% v/v 2-mercaptoethanol, pH adjusted to 8.

- Ni tag bead wash buffer: 93.2 mM Na_2HPO_4, 6.8 mM NaH_2PO_4, 300 mM NaCl, 70 mM imidazole, 0.002% v/v 2-mercaptoethanol, pH adjusted to 8.

- Ni tag bead elution buffer: 93.2 mM Na_2HPO_4, 6.8 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole, 0.002% v/v 2-mercaptoethanol, pH adjusted to 8.

- GST tag resuspension buffer: 57.7 mM Na_2HPO_4, 42.3 mM NaH_2PO_4, 300 mM NaCl, 2 mM DTT, pH adjusted to 7.

- GST tag bead wash buffer: 57.7 mM Na_2HPO_4, 42.3 mM NaH_2PO_4, 300 mM NaCl, 2 mM DTT, 5 mM glutathione reduced, pH adjusted to 7.

- GST tag bead elution buffer: 100 mM Tris, 300 mM NaCl, 2 mM DTT, 20 mM glutathione reduced, pH adjusted to 8.

- General assay buffer: 50 mM Tris, 300 mM NaCl, 2 mM DTT, pH adjusted to 8.

For all protein samples undergoing purification stages utilising a His tag, Ni sepharose (6 fast flow, GE healthcare) beads were used.

For all protein samples undergoing purification stages utilising a GST tag, glutathione sepharose 4B (GE healthcare) beads were used.

2.1.4 Bacterial Cell Culture Reagents

Growth media was obtained from Fisher, for some culture work ready made powdered LB media was purchased from Formedium. All glassware was washed in detergent-free water prior to use and in some instances also autoclaved (121 C, 15 min). Growth media were autoclaved under the same conditions prior to antibiotic addition. Procedures were conducted in a sterile fume hood or with the use of a Bunsen burner flame. All surfaces were cleaned with 70% v/v ethanol before and after use.

- LB broth: 10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl. Autoclaved. Medium cooled to <50°C before antibiotic addition. Stored at 4°C.
- LB-agar medium: LB broth, 15 g/L agar. Autoclaved. Medium cooled to <50°C before antibiotic addition. Stored at 4°C once poured.
- TB broth: 12 g/L tryptone, 24 g/L yeast extract, 4 ml/L glycerol made up to 900 ml with H_2O. Autoclaved. Medium cooled to <50°C before addition of 100 ml 0.17 M KH_2PO_4 and 0.72 M K_2HPO_4, and antibiotic.
- SOC medium: 20 g/L tryptone, 5 g/L yeast extract, 0.5 g/L NaCl, 0.18 g/L KCl, 0.95 g/L $MgCl_2$. Autoclaved. Medium was cooled to <50°C and 2% glucose w/v was added. Stored at -20°C.
- Ampicillin, 1000X stock: 100 mg/ml, dissolved in deionised H_2O, filter sterilised, stored at -20°C.
- Kanamycin, 1000X stock: 30 mg/ml, dissolved in deionised H_2O, filter sterilised, stored at -20°C.
- Chloramphenicol, 1000X stock: 40 mg/ml, dissolved in deionised H_2O, filter sterilised, stored at -20°C.

2.1.5 Oligonucleotides

Custom oligonucleotide primers were ordered from Sigma-Genosys and were obtained as lyophilised pellets. Pellets were resuspended in an appropriate volume of sterile deionised H_2O to give stock solutions of 100 μM, according to manufacturer's instructions. Working solutions of 20 μM or 3.2 μM, as appropriate, were prepared by dilution with H_2O and all primers stored at -20°C. The primers used in this study are shown in chapter 5 accompanying the relevant construct.

2.1.6 Plasmid Vectors

2.1.6.1 pET15b expression vector

A modified version of the pET15b vector (Novagen) was provided by Dr. M. Bochtler. This modified version of the commercial vector retains the lac promoter, the N-terminal His tag and allows for ampicillin selection. However the thrombin cleavage site has been removed meaning the His tag cannot be cleaved from the expressed protein sequence.
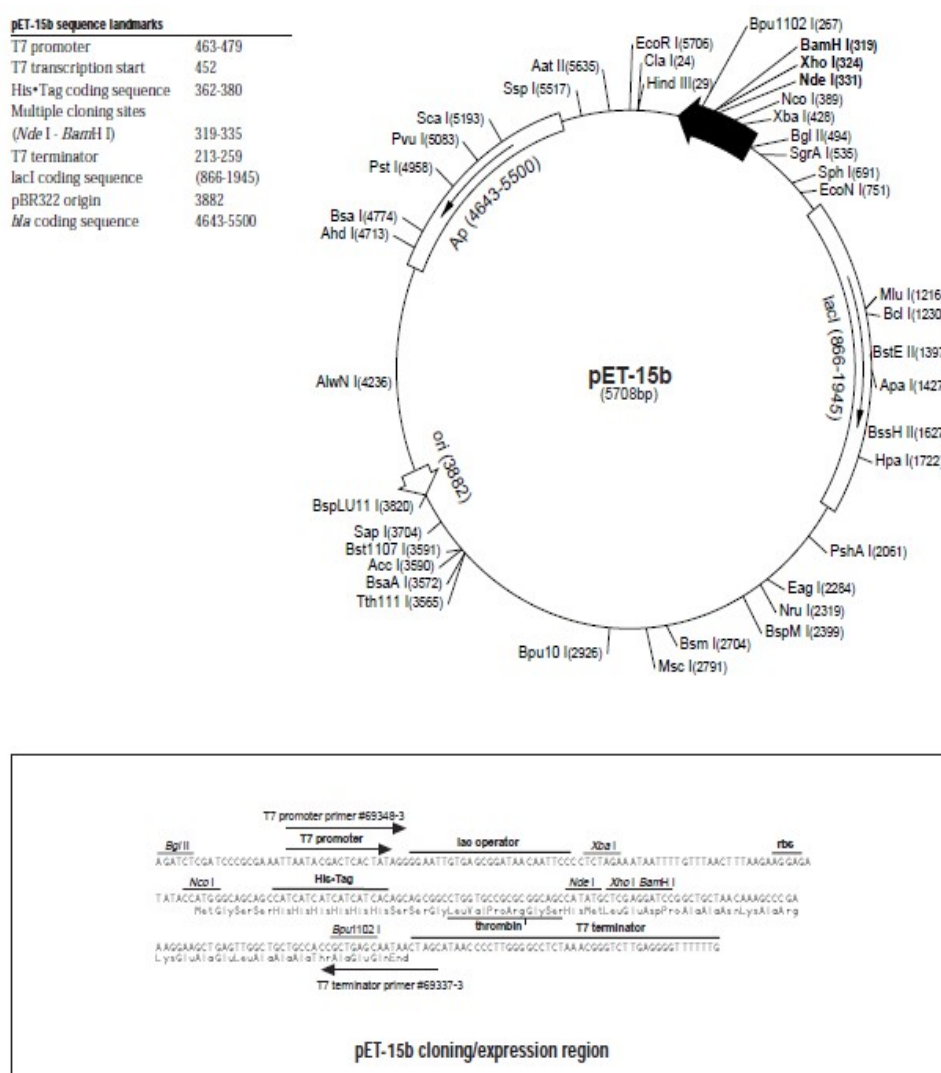
Figure 2.1.1   Vector map of pET15b (Novagen (Merck)).

*Note that modified version of vector used had thrombin cleavage site remov*ed.

2.1.6.2 pETMM Vectors

The pETMM series of expression vectors were generated at EMBL Heidelberg. The specific pETMM vectors used in this project were provided by Dr. M. Bochtler. The used vectors shared the features of allowing for kanamycin selection and a TEV cleavage site which allows for the removal of the N-terminal dual tag. The N-terminal tags consisted of a His tag and a highly soluble protein. Solubility tags used included thioredoxin (Trx) (pETMM20), maltose-binding protein (MBP) (pETMM41) and NusA (pETMM60), an *E. coli* protein associated with transcription termination. As the vectors are derived from pET vectors they share a common multiple cloning site (MCS) and the lac promoter.
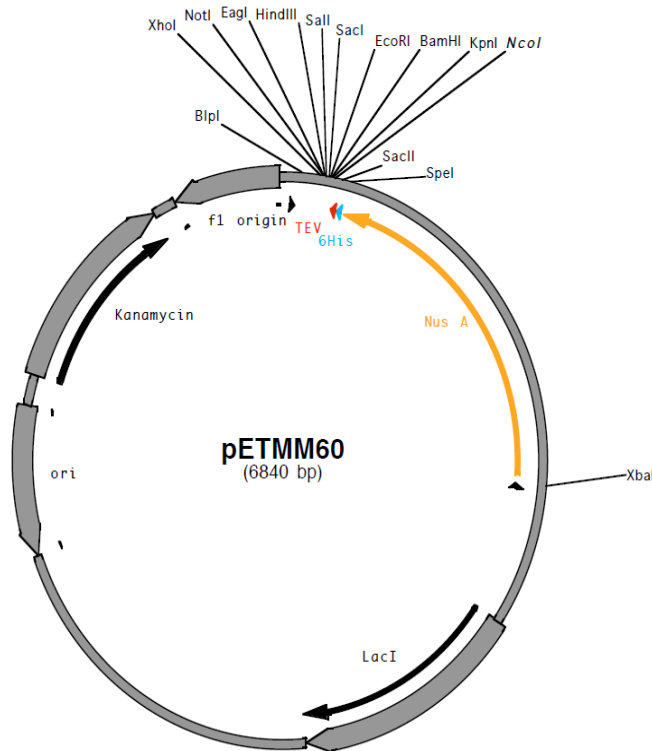
Figure 2.1.2   Vector map of pETMM 60 (NusA).

Only difference from other pETMM vectors used is the solubility tag.   Map from supplied information received with vector.

2.1.6.3 pGEX-6P-1 Vector

pGEX-6P-1 vector was obtained from GE Healthcare. The vector codes for an N-terminal glutathione-S-transferase (GST) tag, contains the lac promoter and allows for ampicillin selection.

Figure 2.1.3   Vector map of pGEX-6P-1 (GE Healthcare).

Expansion shows multiple cloning site of vector, demonstrating the frame of the GST coding sequence.

2.1.7 Antibodies

The following antibodies were used in this study:

- Ab-103T, rabbit polyclonal, raised against glutathione-S-transferase. Used at 1:10000 dilution for western blotting.
- Anti-NusA, mouse monoclonal. Obtained from Santa Cruz. Used at 1:20000 dilution for western blotting.
- anti-His (penta His), mouse polyclonal. Obtained from Qiagen. Used at 1:2000 dilution for western blotting.

- CaM (G-3), mouse monoclonal, raised against amino acids 1-149 human CaM. Obtained from Santa Cruz. Used at 1:2000 dilution for western blotting.

- Donkey anti-mouse IgG-HRP. Obtained from Santa Cruz. Used at 1:10000 dilution for western blotting.

- Donkey anti-rabbit IgG-HRP. Obtained from Santa Cruz. Used at 1:10000 dilution for western blotting.

2.1.8 Computer Software and Data Analysis

Stained protein gels and western blots were scanned at 600 dpi, when possible, using various image scanners and a densitometer (GS-700, BioRad), image processing was performed using Paint Shop Pro X (Corel). Standard spreadsheet packages present in office software (Microsoft and Libreoffice) were used to analyse data and generate graphs.

DNA and protein sequence analysis were performed using software available at the ExPASy (http://www.expasy.ch/tools/dna.html), and European Bioinformatics Institute (http://www.ebi.ac.uk/services) websites. In addition, protein secondary structure sequence prediction was conducted using the network protein sequence analysis web server based at the Pôle Bioinformatique Lyonnais (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_server.html). Open reading frame analysis of DNA sequence was performed using the Chromas Lite 2.01 (Technelysium).

In silico modelling and analyse software was used to generate the models shown in chapter 3 and 4. Models were generated using MODELLER (academic licence, Salilab, http://salilab.org/MODELLER/) was used with the Swift MODELLER GUI (Mathur, http://www.bitmesra.ac.in/swift-MODELLER/swift.htm). Model simulation was run in CHARMM (Karplus, http://www.charmm.org/), models were prepared in Insight II (Accelrys) prior to loading into CHARMM. Model analrysis was conducted using Procheck (EBI, http://www.ebi.ac.uk/thornton-srv/software/PROCHECK/) and VMD (Humphrey, http://www.ks.uiuc.edu/Research/vmd/). Gnuplot (http://www.gnuplot.info/) was used generate graphs from some of the data output from CHARMM. Specific information to the operation of these programmes is detailed in the relevant chapter.

2.1.9 Health and Safety

All reagents were handled and stored as recommended by manufacturer's safety sheets. All experiments were carried out in accordance with COSHH regulations and local college regulations.

All genetic manipulation was carried out following GMAG guidelines.

All bacterial cultures were disinfected with Haz-Tabs (Guest Medical) at manufacturer's recommended concentrations (1 tablet/L of water) prior to disposal.

2.2 Methods

Standard molecular and biochemical techniques were performed according to procedures described in *Molecular Cloning: A Laboratory Manual* (CSH) or *Short Protocols in Molecular Biology* (Wiley). DNA purification techniques and *in vitro* transcription/translation protocols were carried out according to users' manuals provided by system manufacturers (Promega, Qiagen and Pierce), as detailed below.

2.2.1 Nucleic Acid Biochemistry Techniques

2.2.1.1 PCR amplification of DNA

PCR reactions were carried out according to reagent manufacturers' instructions, using a GeneAmp 9700 (Applied Biosystems) PCR machine. *Taq* DNA polymerase (Promega) was used for diagnostic reactions only, whilst Phusion polymerase (NEB) was used for generation of high fidelity PCR products for cloning purposes. Standard reaction mixtures and thermal cycling conditions, including annealing temperature and extension times, as recommended for the polymerase used were followed.

Following analysis of PCR products by agarose gel electrophoresis, PCR products were purified by excision from the agarose gel, using the QIAEX II Agarose Gel Extraction Kit (Qiagen).

2.2.1.2 Agarose gel electrophoresis

DNA fragments were analysed by agarose gel electrophoresis and compared to DNA

standard size markers. Agarose (Euorgentec), usually 1% w/v agarose, was added to TAE buffer (1X) and heated at high power in a microwave oven until the agarose dissolved. The solution was allowed to cool to <50°C and ethidium bromide was added to a final concentration of 0.2 μg/ml. The solution was cast in a gel tray, assembled according to manufacturer's instructions (BioRad), and a comb inserted to mould the wells. Once the gel had set, DNA samples in 1X DNA-loading buffer were loaded, along with standard DNA molecular weight markers (NEB). Electrophoresis was then performed in a gel tank containing fresh 1x TAE at a constant voltage (70-120V). The gel was visualised on a UV transluminator and the image acquired using Quantity One software (BioRad).

2.2.1.3 Cloning of DNA fragments

Plasmid DNA was digested with 100 U of predetermined endonuclease enzymes (NEB), typically 100 μg DNA in 100 μl reactions for 2 hr at 37°C, in the appropriate buffer as instructed by the enzyme manufacturer. Primers were designed to allow compatible double digests within the same vector system. The digestion mixture was heat denatured at 80°C for 20 minutes prior to treatment with alkaline phosphatase (~100 units CIP, NEB) for 1 hr at 37°C. The digested product was purified using the QIAquick PCR purification kit (Qiagen) with the DNA eluted eluted in a minimal volume (<50 μl) dH$_2$O. Purified PCR products were digested following the same protocol except the heat denaturing and phosphatase treatment steps were omitted.

Ligations were carried out using a molar ratio of 3:1 or even 5:1 of insert to vector. The ligation was performed using T4 DNA ligase (NEB) in supplied 1X buffer and incubated at 4°C overnight. The ligation mixture was subsequently used for transformation into chemically competent *E. coli* bacteria as described below.

2.2.1.4 Bacterial cell culture

Bacteria were cultured under aseptic conditions at 37°C in LB medium, either in suspension with rotation at 180 rpm (Innova 4300 shaker incubator, New Brunswick) or on solid support medium on LB-agar plates (plate incubator, Heraeus). Where bacteria were transformed with plasmid DNA, the growth medium was supplemented with the

appropriate antibiotic.

For every positive clone, a glycerol stock of the host bacteria was prepared for long-term storage; 1 ml of a saturated overnight culture wad pelleted by centrifugation at 10,000g for 5 min and resuspended in 1 ml of LB:glycerol 1:1, containing the appropriate antibiotic. The glycerol stock was then snap frozen in liquid nitrogen and stored at -80°C. Bacteria were revived from stock by streaking a sample of the frozen stock on an LB-agar plate (containing the appropriate antibiotic) to produce individual colonies.

The cell density of growing cultures, where required, was determined by measuring the light absorbance of the suspension at a wavelength of 600 nm.

2.2.1.5 Preparation of competent bacteria

Chemically competent TOP10, BL21 (DE3) and Rosetta (DE3) cells were made in the laboratory using the $CaCl_2$ method. Aliquots (120 μl) were flash frozen using liquid nitrogen and stored at −80 C for up to six months.

Rosetta 2 (DE3) (Novagen/Merck) were purchased for some stages of the project involving mutant constructs.

2.2.1.6 Transformation of competent bacteria

5 μl of ligation reaction mixture was added to chemically competent *E. coli* as soon as they started to thaw. The cells were then incubated on ice for 30 minutes and then subjected to 'heat-shock' by being placed at 42°C in a water bath for 45 sec and then returned to ice for a further 3-5 minutes. Cells were then suspended in 800 μl SOC medium and incubated at 37°C for one hour with shaking at 225 rpm. Unequal volumes of the cell culture (commonly the equivalent of 100 μl and 800 μl) were plated on LB-agar containing the appropriate antibiotic and incubated overnight at 37°C. Culture plates were then stored for a maximum of 4 weeks at 4°C.

2.2.1.7 Analysis of positive clones by PCR screening

From the transformation culture plates a number of colonies were screened for the

presence of the recombinant plasmid. Colonies of interest were picked from the plate and transferred to a PCR reaction containing primers specific to the subcloned insert. Following cycling, the PCR reactions were analysed by agarose electrophoresis as above. The selected colonies were also used to innoculate 10 ml of LB medium containing any appropriate antibiotic which was grown overnight at 37°C with shaking at 225 rpm. Plasmid DNA was purified from any culture that the parent colony that had a given a positive result from PCR screening. Purification was performed using the Wizard SV Minipreps plasmid purification kit (Promega) following the manufacturer's recommended protocol except DNA was eluted with 30 µl dH$_2$O per kit spin column. Some of the colonies were further tested by restriction digest following the protocol described above.

The presence and correct orientation of the subcloned insert was confirmed by DNA sequencing of the plasmid DNA. Sequencing samples were prepared using the BigDye Terminator v3.1 cycle sequencing kit (Applied Biosystems) following the manufacturer's instructions. The prepared sample was sent for sequence reading and the returned chromatograms compared against predicted sequence. Samples of plasmid DNA from sequencing confirmed colonies were stored at -20°C.

2.2.1.8 Quantification of DNA

DNA concentration was determined by spectrophotometric quantification of a 1:100 dilution sample (in duplicate) by measuring the absorbance at 260 nm (A$_{260}$) in a quartz cuvette using a Perkin-Elmer MBA2000 spectrophotometer. The concentration was calculated using the equation: 1unit A$_{260}$ = 50   g/ml of double stranded DNA.

2.2.1.9 Point mutation of recombinant DNA constructs

Mutation of recombinant DNA constructs was performed using the QuikChange II XL Site-Directed Mutagenesis Kit (Stratagene) according to manufacturer's instructions. Complementary primers incorporating the desired mutation were designed for use with the kit for which manufacturer's instructions were followed.
Post the thermal cycling step, 1 µl of DpnI was added to digest the methylated template DNA and the reaction incubated for 1 hr at 37°C. 5 µl of the reaction was then transformed into the supplied XL10-Gold competent cells. Recombinant colonies were

screened by PCR and restriction digest, and the presence of the mutation confirmed by DNA sequencing.

2.2.2 Protein Biochemistry Techniques

2.2.2.1 SDS-polyacrylamide gel electrophoresis

Protein sample quality was assessed by comparison to standard molecular weight (MW) markers by running SDS-polyacrylamide gel electrophoresis (SDS-PAGE). Gels were made, assembled and run according the instructions in the Mini-Protean tetra cell manual (Bio-Rad). The acrylamide/Bis percentage of the separating gels were determined by the MW of the proteins to be assessed. The separating gels were covered with water until set which was removed using blotting paper prior to pouring the stacking gel. Wells were flushed with water then filled with running buffer prior to sample loading. Samples were heated treated at 80°C for 10 minutes in 1X loading buffer. Electrophoresis was conducted at constant voltage (usually 150V) until the dye front had travelled far enough to allow visualisation of the protein samples. Gels were visualised by staining with Coomassie stain followed by several washes of destain. Gels were stored in water prior to scanning.

2.2.2.2 Transfer of proteins to membranes

Proteins from SDS-PAGE gels were transferred to a polyvinylidene difluoride (PVDF) (Immobilon-P, Millipore) membrane using a semi-dry transfer apparatus (BioRad). The SDS-PAGE gel and membrane were pre-equilibrated in semi-dry transfer buffer for 20 minutes at room temperature (PVDF membranes were pre-soaked in methanol for 1 minute). The transfer apparatus was assembled according to the manufacturer's instructions, with the membrane between the gel and the anode. The proteins were transferred by electrophoresis at a constant voltage (15 V) for 45-60 minutes, dependent on protein size, at room temperature.

2.2.2.3 Western blot analysis

After protein transfer, the membrane was incubated in TBS-T buffer containing 5% (w/v) skimmed milk powder (Marvel) (TBS-T/milk) overnight at 4°C with gentle mixing. Primary antibodies were added in the appropriate dilution in TBS-T/milk for 60

minutes at room temperature with gentle mixing. The membrane was then washed three times (10 minutes/wash) with TBS-T/milk. The secondary antibody was added in an appropriate dilution in TBS-T/milk and incubated at room temperature for 60 minutes with gentle mixing. The membrane was then washed three times (10 minutes/wash) with TBS-T. The membrane was then sprayed with a single step chemiluminescent substrate, ECL reagent (Calbiochem/Merck). The reagent was allowed to work for 3-5 minutes prior to exposure to film (Kodak) for various periods of time as required for visualisation of immunoreactive bands. The exposure and subsequent film development (Compact X4 developer, Xograph) was conducted in a darkroom.

2.2.2.4 Determination of protein concentration

Protein concentration was mainly determined by application of the Beer-Lambert law. Absorbance readings were taken at 280 nm using a MBA 2000 (Perkin Elmer) spectrophotometer with a 1 cm light path. The theoretical molar extinction coefficients used in the calculation were generated by amino acid sequence analysis by the ProtParam tool (ExPASy site tool).

For some protein samples protein concentration was determined using the BCA assay kit (Pierce) according to manufacturer's instructions. The assay was performed in a 96-well plate and the absorbance read at 560 nm in a Multiscan EX (Labsystems) plate reader using the Genesis (Labsystems) software program. Sample measurements were taken in duplicate for three appropriate dilutions. These were then compared to a standard curve produced by measurement of various known concentrations of BSA (62.5 µg/ml – 2 mg/ml) using the same method.

2.2.2.5 General expression and affinity purification of recombinant proteins

E. coli Rosetta (DE3) bacteria were transformed with appropriate recombinant tag-fusion vector as described above. A 25 ml culture of LB containing the appropriate antibiotic for the selected vector and chloramphenicol, to select for the Rosetta, was inoculated with one colony of transformed bacteria and grown at 37°C overnight in a shaking incubator. A 500 ml culture of TB containing both the vector selection antibiotic and chloramphenicol was inoculated with 5 ml of the overnight culture. This culture was then grown at 37°C and 180 rpm in a shaking incubator until the optical density

(OD), measured by absorbance at 600 nm reached 0.7-0.8 at which point the incubator was set to cool to 16°C. The OD and temperature were then closely monitored, the OD commonly reached 1.0 by the time the temperature had fallen to below 20°C. At this stage expression was induced by addition of 100 µg/ml (final concentration) IPTG, and the culture medium incubated at 16°C for a further 12-14 hrs.

The culture medium was then centrifuged at 7000g (JLA-8.1000, Beckman) for 15 min at 4°C to pellet the bacteria. The pellet was transferred to universal tubes and stored at -80°C until usage. Stored pellet was commonly resuspended in 20 ml/L (relative to original culture volume), occasionally larger volumes were required, of the respective buffer for the tag system to be used for purification. EDTA free protease inhibitor cocktail tablets (Roche), 1 per 50 ml, were added to the resuspended pellet mixture which was then left to mix by rolling for 1-2 hrs at 4°C. A pressure cell homogeniser (Stansted) was pre-chilled to 4°C and flushed through with water followed by the appropriate resuspension buffer. The pellet mixture was processed through the homogeniser a minimum of two times in 50 ml batches with processed sample being collected on ice. Lysed pellet was centrifuged at 50,000g (JA-25.50, Beckman) for 30 min 4°C. The supernatant was mixed with the appropriate affinity beads, which had been pre-washed in resuspension buffer, for the tag system being used. Beads were allowed to mix by rolling at 4°C for a minimum of two hours though more usually for 14 hours. Post this incubation phase, the mixed slurry was poured into small (less than 2 cm diameter) columns (Bio-Rad) allowing the beads to pack under gravity. Once all the liquid from the slurry had flowed through the column then the beads were washed with 2 column volumes of appropriate resuspension buffer. Beads were then further washed with 10-20 column volumes of bead specific wash buffer as required. Protein elution from the column was monitored using the basic protein assay, stated above, with no further wash buffer being used once no further blue colour was observed in the assay. 1 ml fractions were eluted using the relevant elution buffer, once again these fractions were monitored using the basic protein assay. Fractions were analysed for purity by SDS-PAGE.

It should be noted that for the pETMM60 CaM protein that the Ni tag buffer systems were modified so that the harvested pellet was lysed in 20 mM imidazole, the incubated beads washed with 40 mM imidazole and the CaM fusion protein eluted with 100 mM

imidazole. These modifications were also applied for the production of the NusA-6His tag control protein.

If no further purification steps were required for the protein sample prior to downstream application use, predominantly functional assays, then affinity purified protein samples underwent dialysis. Desired protein fractions were pooled and loaded into SnakeSkin dialysis tubing (Pierce/Thermo Scientific), either 10K or 3K MWCO depending on protein size. Dialysis tubing was placed in an excess volume (min x500 volumes) of the appropriate downstream assay starting buffer, usually the general assay buffer, which was left at 4°C overnight with gentle stirring. If required then the dialysis tubing was transferred to a similar excess of buffer and left for an additional 3-4 hr period under the same conditions. Dialysed protein samples were concentrated using Vivaspin (Sartorius) centrifugal concentrators (PES membrane, relevant MWCO to protein sample) until suitable concentration for the downstream application was achieved.

2.2.2.6 Further stages of purification of recombinant proteins

The affinity purified protein samples underwent further purification stages prior to certain experiments described in chapters 5 and 6.

2.2.2.6.1 Tag cleavage from fusion proteins

Fusion tags were cleaved from recombinant protein samples by incubation with the appropriate protease, TEV for pETMM and 3C for pGEX based constructs. Commonly the cleavage was combined with the post affinity purification dialysis step with the protease added to the respective dialysis tubing usually in the concentration range 1:20 to the fusion protein. The protease and cleaved fusion tag was removed by an incubation, 30 minutes at 4°C with rolling, with an equivalent amount of relevant affinity bead as used in the purification step for that batch of protein. The slurry was passed through a gravity drip column and the follow through, containing the cleaved protein, was retained. Recombinant 3C and TEV were produced in the laboratory using the procedures described above except that the Ni tag standard buffer was used to lyse the TEV harvested pellet and the affinity purification wash step was omitted for both proteases.

2.2.2.6.2 Gel filtration

A sephacryl HR-300 column (GE healthcare) was pre-washed with the general assay buffer using an ÄKTA FPLC system running Unicorn software (GE healthcare). Affinity purified protein was concentrated, usually 1-5 mg/ml, and a sample of less than 10 ml was injected onto the column. A constant flow rate of <0.5 ml was maintained, elution was monitored by UV absorbance and fractions were collected during the course of the run. Elution peaks of interest were analysed by SDS-PAGE.

2.2.3 Functional Assay Techniques

2.2.3.1 Pull down assays

Protein and antibody concentrations were tested to allow sufficient detection of potential interactions by western blot analysis. A slight excess of protein, commonly 20-30 µg per protein per assay, was subsequently used in the pull down assays. To allow detection by western blot the tagged fusion proteins were used except for CaM which had its tag removed. Some experiments were conducted with commercially obtained CaM-agarose beads (Stratagene). An excess of affinity beads, min vol 30 µl, were used to immobilise the bait protein. The beads had been previously washed in general assay buffer containing the desired Ca^{2+} for the experiment. The pull down assay Ca^{2+} and EGTA buffers were made by addition of the relevant stock solution to general assay buffer, the Ca^{2+} buffers were not chelated. The potential binding partner protein was added with pull down assay buffer, to give a total volume of x10 the bead volume used, and allowed to incubate with mixing for a minimum of 20 minutes. Post incubation the beads were centrifuged <100 g to remove liquid from the beads. This was followed by 3 further washes of x3-4 volume of beads using the specific pull down assay buffer. Post the final centrifuge step the beads were resuspended in x2 volume protein loading buffer prior to SDS-PAGE and western blot analysis.

For a limited number of the interaction experiments glutaraldehyde cross linking was used. The protocol as above was followed except 0.25% v/v glutaraldehyde solution (Sigma) was added during the incubation phase. Post the incubation stage 20% v/v hydrazine solution (Sigma) was added to prevent further cross linking.

2.2.3.2 Circular Dichroism

Purified protein samples were dialysed into a buffer system suitable for circular dichroism (CD), commonly 20 mM Tris, 300 mM NaF, pH adjusted to 8 using H_2SO_4 rather than HCl. Post dialysis the samples were concentrated, if possible into the range 1-2 mg/ml, and then checked for UV absorbance in the wavelength range 190-250 nm if the sample was to be tested for secondary structure features. CD data was collected in the far (190-250 nm) and near (250-350 nm) UV regions using a DSM20 Monochromator (Olis) using a range of sample cell lengths, dependent on the protein sample quality, on the B23 beamline at the Diamond Light Source synchrotron. Data were analysed using Globalworks software (OLIS). Additional data was collected using a Chirascan CD spectrometer (Applied Photophysics), these data were directly entered into a spreadsheet package for analysis. On both systems data collection rates (scan speed, nm/sec) were optimised to provide best signal quality against noise within a practical time frame. Multiple data sets were collected per experiment and experiments were conducted across several batches of purified protein.

Ca^{2+} sensitivity and potential interaction experiments were mainly conducted in the 250-350 nm range monitoring changes in the 3ry structure of aromatic residues of the tested protein samples. Although the protein samples being tested had different relative contributing signal strengths; data was recorded using 1:1 concentrations for interaction studies. Experimental Ca^{2+} concentrations were achieved by addition of stock $CaCl_2$ and EGTA solutions. Volumes were pre-calculated to minimize dilution effects on protein signal strength with H_2O used as an equivalent volume when no stock solutions were added. Variations in protein concentrations between purification batches were also taken into account to allow comparison across the data collected.

2.2.3.3 Mass Spectrometry

Gel plugs (1.5 mm diameter) were manually excised, placed in a 96-well plate and peptides recovered following trypsin digestion using a slightly modified version of the (Shevchenko et al. 1996) method. Sequencing grade modified trypsin (Promega UK Ltd) was used at 6.25 ng/µl in 25mM $NH4HCO_3$ and incubated at 37ºC for 3 hours. Finally the dried peptides were resuspended in 50% (v/v) acetonitrile in 0.1% (v/v) trifluoroacetic acid (TFA; 5 µl) for mass spectrometry (MS) analysis and an aliquot

corresponding to 10% of the material (0.5 µl) was spotted onto a 384 well MS plate. The samples were allowed to dry and the overlaid with α-cyano-4-hydroxycinnamic acid (CHCA, Sigma, Dorset, UK; 0.5 µl prepared by mixing 5mg matrix with 1ml of 50% (v/v) acetonitrile in 0.1% (v/v) TFA).

Mass spectrometry was performed using a MALDI TOF/TOF mass spectrometer (Applied Biosystems 4800 MALDI TOF/TOF Analyzer; Foster City, CA, USA) with a 200 Hz solid state laser operating at a wavelength of 355nm (Medzihradszky et al. 2000), (Bienvenut et al. 2002), (Brennan et al. 2009). MALDI mass spectra and subsequent MS/MS spectra of the 8 most abundant MALDI peaks were obtained following routine calibration. Common trypsin autolysis peaks and matrix ion signals and precursors within 300 resolution of each other were excluded from the selection and the peaks were analysed with the strongest peak first. For positive-ion reflector mode spectra 800 laser shots were averaged (mass range 700-4000 Da; focus mass 2000). In MS/MS positive ion mode 4000 spectra were averaged with 1 kV collision energy (collision gas was air at a pressure of 1.6 x 10-6 Torr) and default calibration.

Combined PMF and MS/MS queries were performed using the MASCOT Database search engine v2.1 (Matrix Science Ltd, London, UK) (Perkins et al. 1999) embedded into Global Proteome Server (GPS) Explorer software v3.6 (Applied Biosystems) on the Swiss-Prot database. Searches were restricted to proteins of human and *E.coli* origin, with digestion-profiles limited to tryptic fragmentation. The tolerances were set for peptide identification searches at 50 ppm for MS and 0.3 Da for MS/MS. Cysteine modification by iodoacetamide was employed as a fixed modification with methionine oxidation as a variable modification. Search results were evaluated by manual inspection and conclusive identification confirmed if there was high quality tandem MS (good y-ion) data for ≥2 peptides (E value $p < 0.05$ for each peptide; overall $p < 0.0025$) or one peptide (only if E value was $p < 0.0001$).

Chapter 3:

Computational modelling of human RyR2

3.1 Introduction

3.1.1 The role of the C terminus of RyR

As discussed in chapter 1 there are a number of key regions within RyR related to function. The carboxy-terminal, ~ 1000 amino acids of RyR, is a very important region for the functioning of the protein as a channel. It contains the pore that is responsible for the exit of calcium ions from the sarcoplasmic reticulum (SR) (Bhat et al. 1997), (Xu et al. 2000), a number of calcium interaction sites (Du et al. 2001) and the site for ryanodine binding (Callaway et al. 1994), (Witcher et al. 1994). This region is also critical for the tetrameric organisation and assembly of the intact RyR channel (Gao et al. 1997), (Stewart et al. 2003). As described in Chapter 1, the disruption of the amino-acid sequence in key regions of the RyR contributes to the pathology observed in malignant hyperthermia (or malignant hyperpyrexia) and in stress induced polymorphic ventricular tachycardia (George et al. 2003), (Marx et al. 2000b), (McCarthy et al. 2000), (Thomas et al. 2007). It is believed that the activation of the RyR channel depends on the rotation of the C-terminal transmembrane assembly and the important changes to the conformation at the N-terminal cytoplasmic domains (el-Hayek et al. 1995), (Orlova et al. 1996b), (Sharma et al. 2000). The organisation of this region according to the current literature is shown below.

Human RyR2 : amino acids 3579 - 4967

"Calmodulin binding" domain (CamBD)

"Calmodulin like" domain (CamLD)

Transmembrane regions 4-6 of RyR family

"RyR, IP3 homology" (RIH) associated domain

```
PQRSKKAVWHKLLSKQRKRAVVACFRMAPLYNLPRHRAVNLFLQGYEKSWIETEEHYFEDKLIEDLAKPGAEPPEE   3654
DEGTKRVDPLHQLILLFSRTALTEKCKLEEDFLYMAYADIMAKSCHDEEDDDGEEEVKSFEEKEMEKQKLLYQQAR   3730
LHDRGAAEMVLQTISASKGETGPMVAATLKLGIAILNGGNSTVQQKMLDYLKEKKDVGFFQSLAGLMQSCSVLDLN   3806
AFERQNKAEGLGMVTEEGSGEKVLQDDEFTCDLFRFLQLLCEGHNSDFQNYLRTQTGNNTTVNIIISTVDYLLRVQ   
ESISDFYWYYSGKDVIDEQGQRNFSKAIQVAKQVFNTLTEYIQGPCTGNQQSLAHSRLWDAVVGFLHVFAHMQMKL   3958
SQDSSQIELLKELMDLQKDMVVMLLSMLEGNVVNGTIGKQMVDMLVESSNNVEMILKFFDMFLKLKDLTSSDTFKE   4034
YDPDGKGVISKRDFHKAMESHKHYTQSETEFLLSCAETDENETLDYEEFVKRFHEPAKDIGFNVAVLLTNLSEHMP   4110
NDTRLQTFLELAESVLNYFQPFLGRIEIMGSAKRIERVYFEISESSRTQWEKPQVKESKRQFIFDVVNEGGEKEKM   4186
ELFVNFCEDTIFEMQLAAQISESDLNERSANKEESEKERPEEQGPRMAFFSILTVRSALFALRYNILTLMRMLSLK   4262
SLKKQMKKVKKMTVKDMVTAFFSSYWSIFMTLLHFVASVFRGFFRIICSLLLGGSLVEGAKKIKVAELLANMPDPT   4338
QDEVRGDGEEGERKPLEAALPSEDLTDLKELTEESDLLSDIFGLDLKREGGQYKLIPHNPNAGLSDLMSNPVPMPE   4414
VQEKFQEQKAKEEEKEEKEETKSEPEKAEGEDGEKEEKAKEDKGKQKLRQLHTHRYGEPEVPESAFWKKIIAYQQK   4490
LLNYFARNFYNMRMLALFVAFAINFILLFYKVSTSSVVEGKELPTRSSSENAKVTSLDSSSHRIIAVHYVLEESSG   4566
YMEPTLRILAILHTVISFFCIIGYYCLKVPLVIFKREKEVARKLEFDGLYITEQPSEDDIKGQWDRLVINTQSFPN   4642
NYWDKFVKRKVMDKYGEFYGRDRISELLGMDKAALDFSDAREKKKPKKDSSLSAVLNSIDVKYQMWKLGVVFTDNS   4718
FLYLAWYMTMSVLGHYNNFFFAAHLLDIAMGFKTLRTILSSVTHNGKQLVLTVGLLAVVVYLYTVVAFNFFRKFYN   4794
KSEDGDTPDMKCDDMLTCYMFHMYVGVRAGGGIGDEIEDPAGDEYEIYRIIFDITFFFFVIVILLAIIQGLIIDAF   4870
GELRDQQEQVKEDMETKCFICGIGNDYFDTVPHGFETHTLQEHNLANYLFFLMYLINKDETEHTGQESYVWKMYQE   4946
RCWEFFPAGDCFRKQYEDQLN                                                         4967
```

Pore outer helix

Pore helix · Selectivity filter · Pore inner helix

Pore inner helix

Figure 3.1.1   Proposed organisation of functional domains at the C-terminus of human RyR2.
Regions of interest in this section of RyR2 are highlighted in the amino acid sequence.

Conflicting results from colleagues have demonstrated that residues 3900-4450 in RyR2 is a span of hydrophobic residues that facilitate the transduction of modulatory events from the cytoplasmic domain to the pore region at the carboxy-terminus (George et al. 2004b). Experiments using GFP-tagged RyR2 fragments corresponding to the 4TM and 10TM assemblies of RyR (aa 4485-4967 and aa 3722-4967, respectively) have indicated intra-RyR interaction between these regions and a dsRed-tagged cytoplasmic (N-terminal) RyR2 fragment, detected by FRET technology (George et al. 2004b). This is in contrast to the reported transmembrane regions that lie in this region as reported by other laboratories (Bhat et al. 1997), (Brandt et al. 1992), (Wagenknecht & Radermacher 1995).

Nevertheless, this region constitutes important domains that denote the functionality of the protein as a calcium channel. Thus, knowledge of the 3D structure of the various domains and its interaction is important to understand how these various regions communicate with each other to function as a channel. The experimental structural determination of this region is under way in the laboratory and chapter 5 details the role of this project in that operation. However this process is fraught with experimental and

technical difficulties. Since there are multiple domains within this region as shown in figure 3.1.1, it is possible to construct a theoretical model of this region, partly by homology modelling. Comparative modelling relies on the assumption that protein domains of sufficiently matching sequence homology will share equivalent structural similarity. If suitable amino acid sequence alignment can be made between a potential 3D template and a protein of interest then a homology model can be generated. The cut off point for the shared similarity between two protein sequences is generally accepted to be a 30% match. Ideally 40% or greater homology yields a more accurate modelled structure in which 90% of main chain atoms are modelled with an RMSD (root mean square distance) error of about 1 Angstrom. 30% homology corresponds to a RMSD error of 3.5 Angstrom for about 80% of main chain backbone atoms. Loops and side-chains mainly contribute to the structural difference which is amplified by insertions and deletion regions as the sequence similarity diminishes. Under 30% homology usually severely limits the detection of homologue structures as alignment becomes further reduced so drastically increasing insertion and deletion regions between the compared sequences (Sánchez & Sali 1997), (Xiang 2006).

As such it is important to achieve at least 30% sequence homology between the experimentally determined 3D template sequence and the protein sequence of yet undetermined structure (Hill et al. 2002), (Park et al. 1998). In the case of modelling a multi-domain protein thus, it is possible to use multiple templates structures (Carpenter et al. 2007), (Fraternali & Pastore 1999). When there are intervening regions between identified domains which cannot be modelled directly using a 3D-template, the secondary structure of these regions can be predicted using well-known computational methods (eg. Chou-Fassman algorithms, SOPMA) (Chou & Fasman 1974), (Geourjon & Deléage 1995). These can be folded as helices or sheets using modelling-software methods (Insight Software, Accelrys) or within MODELLER itself (Reddy & Kaznessis 2007). Subsequent molecular dynamics runs coupled with energy minimisation methods can be carried out to obtain a theoretical 3D fold of the domain. The secondary structure prediction of the amino acid residues 3579-4085 in RyR2 using SOPMA (Geourjon & Deléage 1995) is shown below.

```
              10        20        30        40        50        60        70
               |         |         |         |         |         |         |
     PQRSKKAVWHKLLSKQRKRAVVACFRMAPLYNL PRHRAVNLFLQGYEKSWIETEEHYFEDKLIEDLAKPG
     ccchhhhhhhhhhhhhhhhhhhhccccccccccccchheehhhhhheehhcchhhhhhhhhhhhhhcc
     AEPPEEDEGTKRVDPLHQLILLFSRTALTEKCKLEEDFLYMAYADIMAKSCHDEEDDDGEEEVKSFEEKE
     cchhhhhhhccccccchhhhhhhhhhhhhhhhhhhcccchhhhhhhhhhhhhhhhhhhhhhhccchhhhhhhhhhh
     MEKQKLLYQQARLHDRGAAEMVLQTISASKGETGPMVAATLKLGIAILNGGNSTVQQKMLDYLKEKKDVG
     hhhhhhhhhhhhhhhhhhhhhhhhhccccccchhhhhhhhheeeeettccchhhhhhhhhhhhttthh
     FFQSLAGLMQSCSVLDLNAFERQNKAEGLGMVTEEGSGEKVLQDDEFTCDLFRFLQLLCEGHNSDFQNYL
     hhhhhhhhhhhhhhhheehhhhhhhhhhhtteeeeccccccccccchhhhhhhhhhhhhhhttccchhhhh
     RTQTGNNTTVNIIISTVDYLLRVQESISDFYWYYSGKDVIDEQGQRNFSKAIQVAKQVFNTLTEYIQGPC
     hcccccccceeeeehhhhhhhhhhhhhhhhhhhheeettcceacttccchhhhhhhhhhhhhhhhhhhecccc
     TGNQQSLAHSRLWDAVVGFLHVFAHMQMKLSQDSSQIELLKELMDLQKDMVVMLLSMLEGNVVNGTIGKQ
     ccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhhhttcccccchhhh
     MVDMLVESSNNVEMILKFFDMFLKLKDLTSSD TFKEYDPDGKGVISKRDFHKAMESHKHYTQSETEFLLS
     hhhhhhhccchhhhhhhhhhhhhhhhhhccchhhheecttttceechhhhhhhhhhhttccchhhhhheeh
     CAETDENETLDYEEFVK ◄── calmodulin-like domain
     hccctttcheehhhhhh

Sequence length :    507

SOPMA :
    Alpha helix      (Hh) :    341 is   67.26%
    3₁₀ helix        (Gg) :      0 is    0.00%
    Pi helix         (Ii) :      0 is    0.00%
    Beta bridge      (Bb) :      0 is    0.00%
    Extended strand  (Ee) :     35 is    6.90%
    Beta turn        (Tt) :     24 is    4.73%
    Bend region      (Ss) :      0 is    0.00%
    Random coil      (Cc) :    107 is   21.10%
    Ambigous states  (?)  :      0 is    0.00%
    Other states          :      0 is    0.00%
```

Figure 3.1.2   Secondary structure prediction of the amino acid residues 3579 – 4085 of human RyR2 using SOPMA

The boxed residues are the calmodulin binding domain and calmodulin like domain, respectively. The regions outside (between) these two boxed domains will be modelled using the above secondary-structure predicted results.

3.1.2 CaM Binding Domain and CaM Like Domain

Studying elements of the role of calmodulin (CaM) in relation to RyR aids with the understanding of the CaM binding domain (CaMBD) and CaM like domain (CaMLD) in the process of modelling these domains. As discussed in chapter 1, CaM is universally recognised as a major calcium sensor and regulator of cell-cycle events through its role in a diverse group of cellular proteins. Due to this important role, there is a need to define the region of interaction between CaM and its cellular targets, and the action of CaM on target protein function.

CaM binds with high affinity to a short α helical region found in many proteins. The interaction of CaM to its target protein sequence mainly consists of a single amino acid that acts as a hydrophobic anchor point and this residue characteristically reaches into the N-lobe core, thus the CaM binding protein or peptide (CaMBP) typically is encased between both lobes of CaM (Ikura, Barbato, et al. 1992), (Ikura, Clore, et al. 1992). In many CaM-CaMBP complexes, the CaM molecule almost engulfs the entire binding region, reducing the solvent accessible area of the binding motif to less than 20%. This typical interaction is illustrated schematically below.
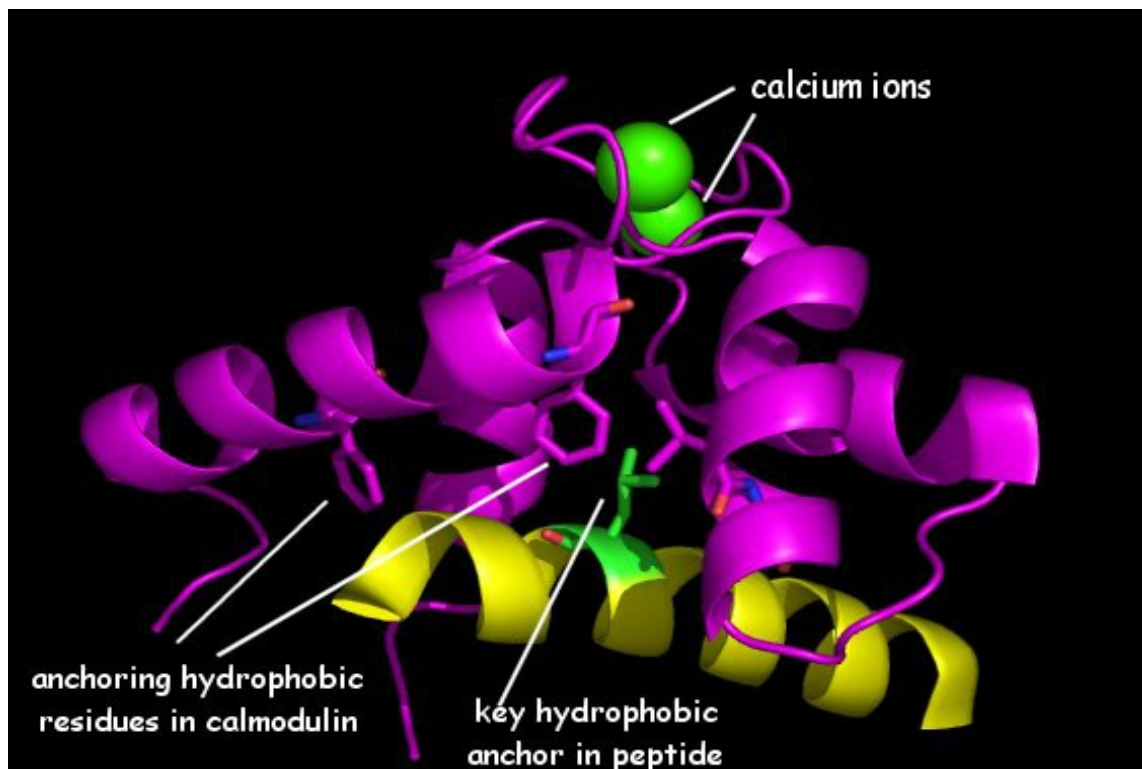
Figure 3.1.3   Typical interaction between calmodulin (CaM) and a calmodulin binding peptide or protein (CaMBP).

A key hydrophobic residue in the CaMBP acts as anchor, reaching into the core of CaM. A few hydrophobic residues (in this example from PDB entry 1G4Y) also serve to clamp onto the binding motif.

The CaMBP consensus sequence is conserved through evolution (see figure 3.1.4) and the presence of such a sequence in a previously uncharacterised protein would serve to identify it as one that would interact with CaM to varying degrees.

Figure 3.1.4 Schematic representation of CaM with a calmodulin binding peptide (CaMBP) from PDB entry 2BCX.

(a) A side-on view of CaM with the binding region shown as a helix in magenta.

(b) View showing the C-terminal end of the CaMBP proximal to the reader.

(c) View showing the N-terminus of CaMBP proximal to the reader, both these views demonstrate the interaction of side chains of CaMBP with the "clamp-like" cavity generated by CaM.

(d) the conservation of residues of the CaMBP of the ryanodine receptor across the animal kingdom, minor residue changes in organisms imply that those residues if mutated within the human isoform may be tolerated with little implication to function.

The CaMBP region in RyR1 and equivalent CaMBD in RyR2, residues 3583-3603, (N. Yamaguchi et al. 2003) have been identified. It is the CaMBP from RyR1 that is shown in figure 3.1.4. Previous studies on skeletal muscle RyR (RyR1) have shown an important segment resembling the structure of CaM within aa 4064-4210 of the rabbit sequence, resembling the structure of CaM containing possible Ca^{2+} binding sites. This region has also been shown to interact with CaMBP in RyR1 and has been termed the CaM like domain (CaMLD) (L. Xiong et al. 2006).
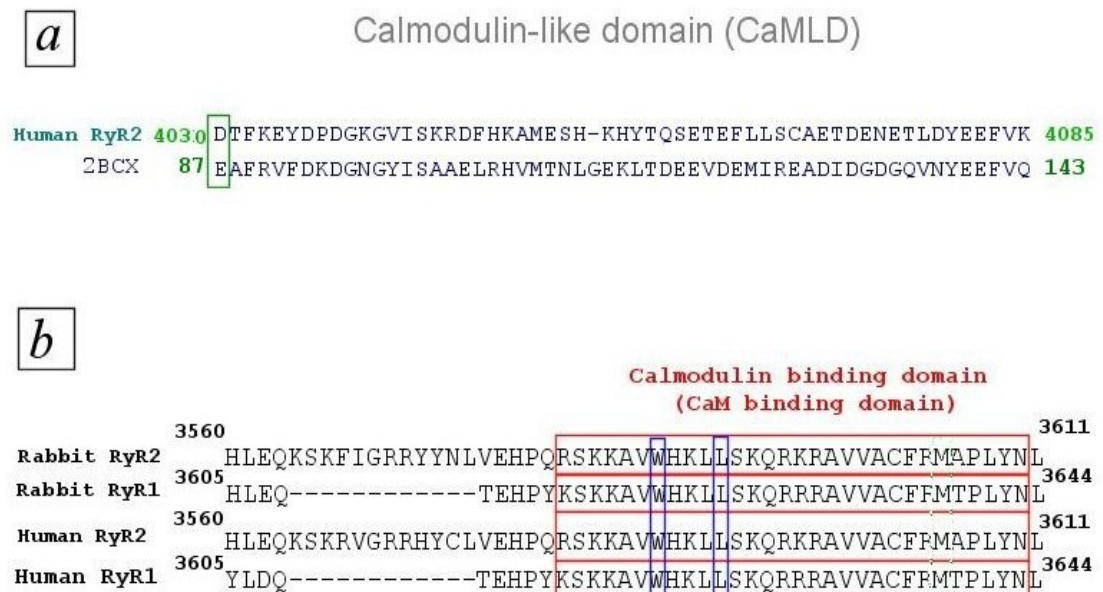
Figure 3.1.5 **Residues in the ryanodine receptor having a calmodulin like domain (CaMLD) and a corresponding calmodulin binding domain (CaMBD).**

(a) Sequence homology between a known EF hand domain of the experimentally determined structure – 2BCX.

(b) shows sequence alignment between rabbit RyR1 and RyR2 and comparisons with corresponding sequences from human RyR1 and RyR2 further illustrating the conservation of residues in this region. The region boxed in red shows the sequence that is implicated in binding to CaM. The conserved residues boxed in blue when mutated in previous studies have shown dramatic reduction of Ca^{2+} binding properties.

It should be noted that the EF hand domain shown in (a) figure 3.1.5 is from CaM as shown in figure 3.1.4 reflecting the CaM "like" structure of CaMLD. A major component of this similarity is a result of the 2 EF hand motifs associated with Ca^{2+}

binding contained with CaMLD. The two lobes of CaM contain 2 EF hands each which lead to comparison of CaMLD to a lobe of CaM both in structure and proposed function (Liangwen Xiong et al. 2006). Although there are amino acid variations within different types of EF hand, there is a high conservation of residues critical to the Ca^{2+} binding function of the domain within EF hand families (Grabarek 2005).

```
CaM       ADQLTEEQIAEFKEAFSLFDKDGDGTITTKELGTVMRSLGQNPTEAELQD
CaMLD     ----------TSSDTFKEYDPDGKGVISKRDFHKAMESH-KHYTQSETEF
                    .::*. :* **.*.*:.::: ..*.*  :: *::* :

CaM       MINEVDADGNGTIDFPEFLTMMARKMKDTDSEEEIREAFRVFDKDGNGYI
CaMLD     LLSCAETDENETLDYEEFVKRFHEPAKDIGFN------------------
                    ::. .::* * *:*: **:. : .   ** . :

CaM       SAAELRHVMTNLGEKLTDEEVDEMIREADIDGDGQVNYEEFVQMMTAK
CaMLD     ----VAVLLTNLSEHMPNDT----------------------RLQ
                    :  ::***.*::.::                     :
```

Figure 3.1.6   Amino acid sequence alignment of CaM with CaMLD.

The region of match between the two sequences relates to the area of the 2 EF hand motifs within CaMLD. The 2nd matching region is located a few residues downstream of the 2nd EF hand motif.

The immediate surrounding regions, including portions of α helix, to the EF hands have been identified in the "clamp like" binding of CaM to CaMBP (Kawasaki & Kretsinger 2012).

Figure 3.1.7   Predicted secondary structure of CaM.

Boxed section is the 2nd matching region of CaM in figure 3.1.6. This boxed section overlaps 2 predicted sections of α helix which are predicted to play a role in binding to CaMBP.

There are a range of disease mutation sites located throughout RyR including the CPVT mutations with CaMLD as discussed in chapter 1. Studying the nature of the interaction

between CaMLD and CaMBD will aid with the understanding of the pathophysiology of CPVT and, in the case of RyR1, malignant hyperthermia. Intuitively, it is envisaged that such minute changes in interaction between two domains may be brought about by small changes to the overall structure.

3.2 Results

To aid the further study of the interaction between the CaMBD and CaMLD domains, a primary aim of the earlier stages of the *in-silico* work was to produce a model of the overall region of RyR2 containing the two domains. As this process would consist of several stages so initially models of the two separate domains were made prior to commencing the construction of the larger planned model.

The following sections detail the process of homology modelling of human RyR2 CaMBD and CaMLD, and the subsequent computations that were performed to refine the models.

3.2.1 *In silico* modelling of CaMLD and CaMBD regions of RyR2

To generate the models of the domains the homology modelling software MODELLER (A Sali & Blundell 1993) was used, all models were generated with MODELLER v9.10. A front end GUI (graphical user interface), swift MODELLER v2.0 (Mathur et al. 2011), was used to run MODELLER in automated mode so that sequence alignment and model generation were controlled by the software. Discrete Optimised Protein Energy (DOPE) scores are used in MODELLER to evaluate how the model compares with the template structure used in homology modelling. Briefly, this method estimates the positions of non-interacting atoms in an artificial sphere. The radius of the sphere is calculated based on the extents of the native structure of the entire template molecule. Thus, the difference between the calculated artificial sphere encasing the new model and the volume of the sphere containing the entire template molecule is the basis of the DOPE score (Shen & Sali 2006). The lowest DOPE scores in a collection of models is desirable. Another method of assessment is the GA341 scores, these always range from 0.0 (worst) to 1.0 (native like). Molecular PDF energy (molpdf) scores are also used for assessment of models, these report the restraint energies in the model structure. Lower

energies indicate less strain on the structure so lower molpdf scores are desirable.

CaMLD shares the high sequence homology with other members of other EF hand families as described in chapter 1. In terms of the identification of a template for CaMLD, none of the more closely matched EF hand structures were present in the Protein Data Bank (PDB). Structures of other members of other less closely sequence related EF hand families have been solved. Initially trials of modelling CaMLD focussed on using the highest percentage sequence homology match template possible. However it was not possible to gain matches to the secondary structure predicted regions of α helix, associated with CaMBD binding, with these templates. Small regions of structures were identified that could be used as donors for these few α helical regions. While multiple template modelling can be performed the use of these very small donor sections was impractical. While CaM does not have such a close match to CaMLD it still falls above the required 30% homology threshold and contains the predicted required secondary structure elements as discussed in the introduction. There are a variety of of available CaM structures, including the CaM component of 2BCX which was selected as the template for modelling CaMLD.

Twenty models of CaMLD were generated by the process stated above using the CaM component of 2BCX. Models were evaluated by DOPE scores.

CaMLD model no.	molpdf	DOPE	GA341
1	436.78894	-6067.64014	0.56606
2	482.53687	-6239.92725	0.44066
3	474.44733	-5858.98779	0.63272
4	473.14145	-6116.59668	0.26535
5	447.55273	-5871.44971	0.40209
6	360.47827	-5984.33252	0.53040
7	445.15585	-6048.05908	0.51714
8	394.35928	-5980.94629	0.49276
9	532.34747	-5981.55469	0.22836
10	446.59985	-5960.28857	0.48357
11	413.00021	-6165.89990	0.66968
12	398.90665	-5993.46191	0.36087
13	*450.62045*	*-5849.70117*	*0.65303*
14	370.50214	-6035.63623	0.51170
15	484.21292	-5692.58838	0.57380
16	372.79172	-6271.97900	0.39410
17	471.26294	-5904.84521	0.65400
18	438.81503	-6157.76025	0.53179
19	449.60641	-5992.08203	0.55847
20	373.29279	-5954.35742	0.29167

Table 3.2.1   Assessment scores of generated CaMBD models.

Model 13 is highlighted as having the lowest DOPE score and one of the highest GA341 scores.
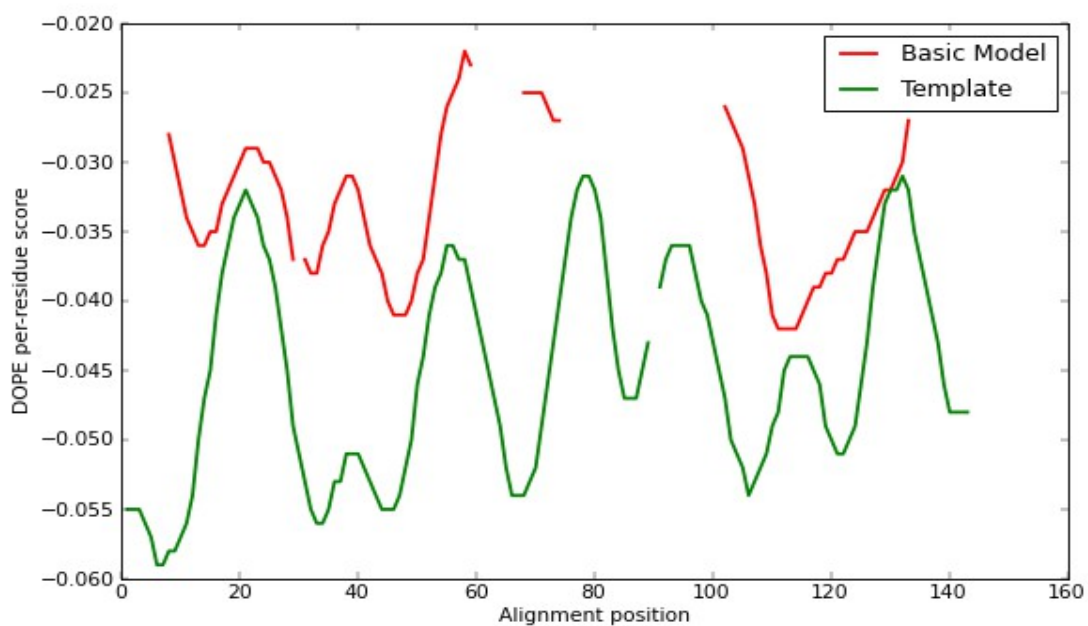
Figure 3.2.1   DOPE scores per residue of CaMLD model 13.

The CaM template is shown by the green trace and the CaMLD model shown in red. Usually better models are produced when there is template structure covering the entirety of the model sequence as in this case. As expected the DOPE score per residue of the template is lower than the model, the closer the model scores are to the template indicates lesser deviation from the template structure in the model.
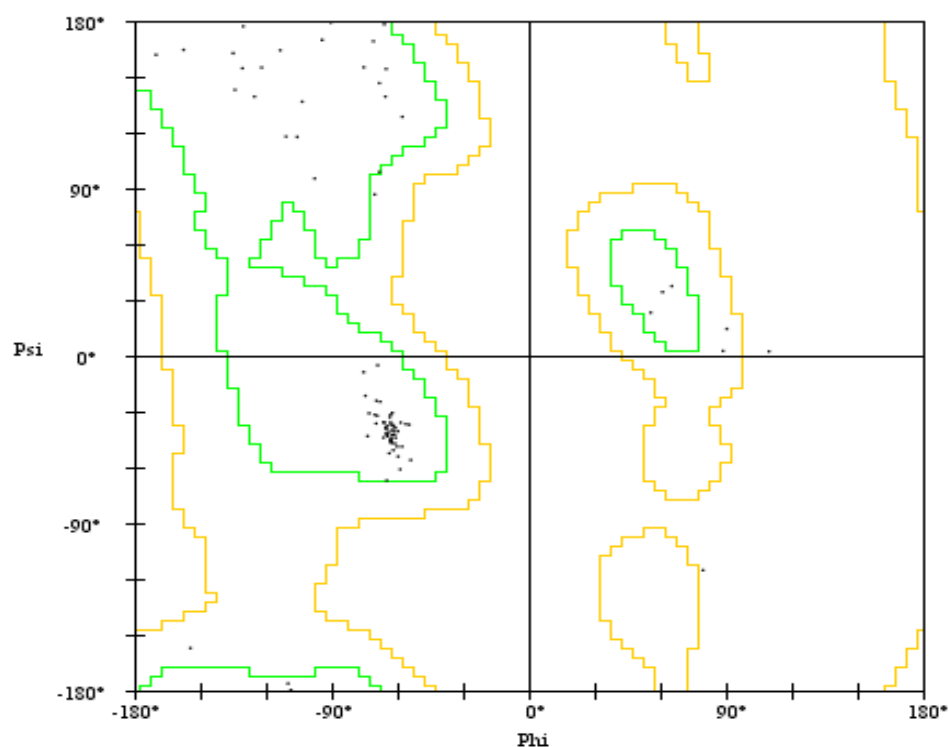
Figure 3.2.2   Ramachandran plot of CaMLD model 13.

Most favourable areas are shown in green, next most favourable in orange. The better the model the more residues (black spots) that are expected within the most favourable areas. Poor models or those requiring further improvement would have a number of residues outside of the favourable regions. For this model only one such residue is observed.

The Ramachandran plot provides a visual measure of the limits of the movement of the bond angles that are present in the backbone of the protein molecule. The Phi (Φ) and Psi (Ψ) angles describe these torsion angles between the atoms.
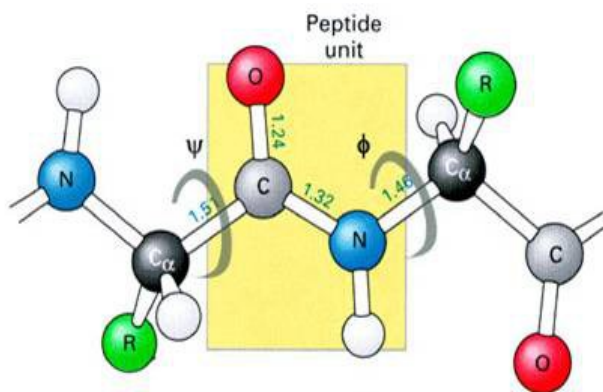
Figure 3.2.3   Phi and psi torsion angles within protein backbone.

Analysis of known structures provides a reference of tolerances of the torsion angles. The favourable areas include phi and psi angles that are observed within this reference set. Residues with bonds falling outside of these areas can indicate an "unlikely" angle for a valid protein structure. The Ramachandran plot provides a tool for the location of these residues so that the model structure can be analysed. Issues such as kinks in the structure and overlapping bonds can be seen though in some cases an outlying residue is not considered a problem if the structure appears viable.
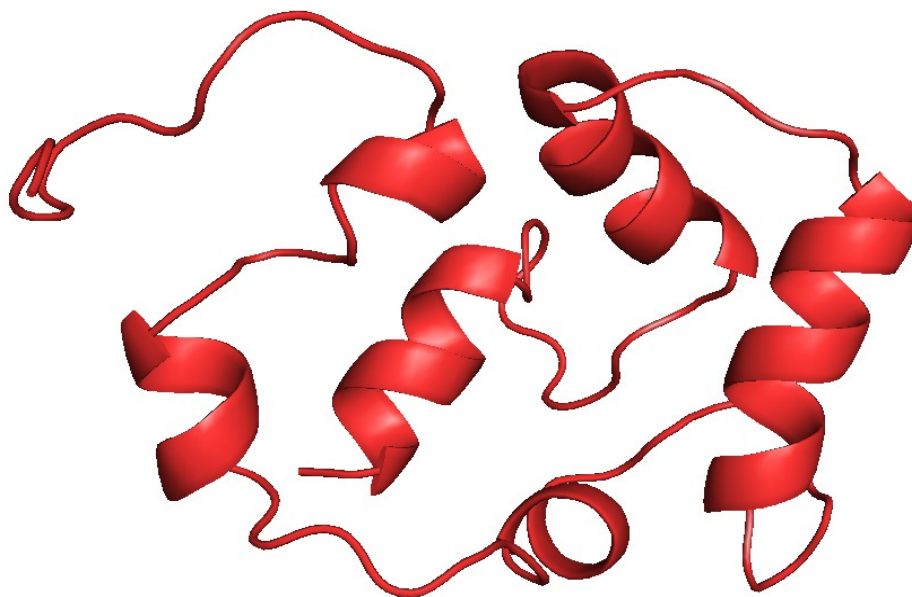
Figure 3.2.4   Structure of CaMLD model 13.

While not having as much α helix content as CaM (see figure 3.1.4) there are clear areas of conserved secondary structure within this model. The overall structure seems viable with no obvious structural deviations.

As already shown the PDB entry 2BCX contains a section of RyR1 CaMBP which has very close homology to the equivalent region of interest in RyR2, so was used a template for modelling CaMBD. The process as described above for the CaMLD model was followed. As the CaMBP structure is a simpler, tube of magenta α helix (figure 3.1.4), so an even closer model to the template with no associated problems was produced.
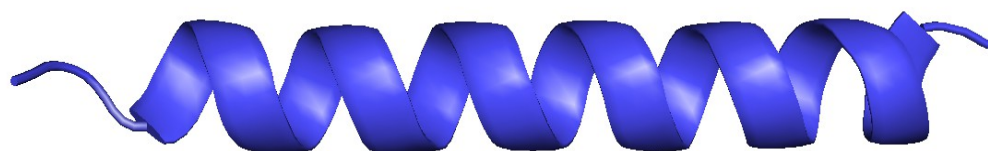
Figure 3.2.5   Computer-g**enerated model of RyR2 CaMBD.**

The CaMBD region folds naturally as a helical segment as shown here.

3.2.2 *In silico* modelling of RyR2 between residues 3578-4085

The CaMLD and CaMBD regions were folded by homology modelling in MODELLER. However, the regions between these two were subsequently modelled using information that had been rationally deduced from the secondary structure bioinformatic analysis.

```
>P1;CMBD
sequence:CMBD:1 :A: 507 :A:::0.00:0.00
PQRSKKAVVHKLLSKQRKRAVVACFRMAPLYNLPRHRAVNLFLQGYEKSWIETEEHYFEDKLIEDLAKPGAEPPE
EDEGTKRVDPLHQLILLFSRTALTEKCKLEEDFLYMAYADIMAKSCHDEEDDDGEEEVKSFEEKEMEKQKLLYQQ
ARLHDRGAAEMVLQTISASKGETGPMVAATLKLGIAILNGGNSTVQQKMLDYLKEKKDVGFFQSLAGLMQSCSVL
DLNAFERQNKAEGLGMVTEEGSGEKVLQDDEFTCDLFRFLQLLCEGHNSDFQNYLRTQTGNNTTVNIIISTVDYL
LRVQESISDFYWYYSGKDVIDEQGQRNFSKAIQVAKQVFNTLTEYIQGPCTGNQQSLAHSRLWDAVVGFLHVFAH
MQMKLSQDSSQIELLKELMDLQKDMVVMLLSMLEGNVVNGTIGKQMVDMLVESSNNVEMILKFFDMFLKLKDLTS
SDTFKEYDPDGKGVISKRDFHKAMESH-KHYTQSETEFLLSCAETDENETLDYEEFVK
*

>P1;2BCXb
structureX:2BCXb:  3614 :B: 3640 :B:undefined:undefined:-1.00:-1
--KSKKAVVHKLLSKQRRRAVVACFRMTP------------------------------------------------
----------------------------------------------------------------------------
----------------------------------------------------------------------------
----------------------------------------------------------------------------
----------------------------------------------------------------------------
----------------------------------------------------------------------------
-------------------------------------------------------------
*

>P1;2BCXa
structureX:2BCXa:    87 :A: 143 :A:undefined:undefined:-1.00:-1
----------------------------------------------------------------------------
----------------------------------------------------------------------------
----------------------------------------------------------------------------
----------------------------------------------------------------------------
----------------------------------------------------------------------------
----------------------------------------------------------------------------
-EAFRVFDKDGNGYISAAELRHVMTNLGEKLTDEEVDEMIREADIDGDGQVNYEEFVQ
*
```

Figure 3.2.6   Sequence alignment between human RyR2 (labelled here as CMBD) and the experimentally determined template sequence of 2BCX.

The sequence alignment is presented in the "PIR" format. For simplicity of coding in the scripts, the actual amino acid residue numbers of RyR2 are not used (ie: proline 3578 is represented as proline 1; lysine 4085 is represented as lysine 507). The asterisks at the end of each line denote the "end" of the sequence.

To ensure that the CaMBD segment is positioned "beneath" the CaMLD segment, several force restraints; rsr (real space residuals: where in xray crystallography, rsr is a measure of the agreement between an electron-density map calculated directly from a structural model and one calculated from experimental data) were applied as follows.

Figure 3.2.7   Distances between CaMLD and CaMBD.

To position the CaMLD of RyR above CaMBD during the modelling process, rsr-distance restraints were applied.

(a) Firstly, the distances (in Angstroms) between 2 residues in CaMLD and 3 residues in CaMBD were measured using the measuring tool in Pymol.

(b) These distances were compared to the equivalent residues in the RyR2 segment, and the script file for MODELLER is shown here incorporating these distance restraints.

As evident from the alignment in figure 3.2.6, the middle portion of the sequence bears no resemblance to 2BCX, there was less than 30% homology to other structures deposited in the PDB. Thus this region was analysed using SOPMA, as shown earlier in figure 3.1.2. Helices were modelled *ab initio* within MODELLER using the following

script segment.

```
#         Residues 43 through 49 should be an alpha helix:
          rsr.add(secondary_structure.alpha(self.residue_range('43:',
'49:')))
#         Residues 56 through 68 should be an alpha helix:
          rsr.add(secondary_structure.alpha(self.residue_range('56:',
'68:')))
#         Residues 73 through 79 should be an alpha helix:
          rsr.add(secondary_structure.alpha(self.residue_range('73:',
'79:')))
#         Residues 86 through 104 should be an alpha helix:
          rsr.add(secondary_structure.alpha(self.residue_range('86:',
'104:')))
#         Residues 108 through 127 should be an alpha helix:
          rsr.add(secondary_structure.alpha(self.residue_range('108:',
'127:')))
#         Residues 131 through 167 should be an alpha helix:
          rsr.add(secondary_structure.alpha(self.residue_range('131:',
'167:')))
#         Residues 175 through 183 should be an alpha helix:
          rsr.add(secondary_structure.alpha(self.residue_range('175:',
'183:')))
#         Residues 194 through 205 should be an alpha helix:
          rsr.add(secondary_structure.alpha(self.residue_range('194:',
'205:')))
#         Residues 209 through 225 should be an alpha helix:
          rsr.add(secondary_structure.alpha(self.residue_range('209:',
'225:')))
#         Residues 227 through 236 should be an alpha helix:
          rsr.add(secondary_structure.alpha(self.residue_range('227:',
'236:')))
#         Residues 276 through 281 should be an alpha helix:
          rsr.add(secondary_structure.alpha(self.residue_range('276:',
'281:')))
#         Residues 295 through 311 should be an alpha helix:
          rsr.add(secondary_structure.alpha(self.residue_range('295:',
'311:')))
#         Residues 327 through 345 should be an alpha helix:
          rsr.add(secondary_structure.alpha(self.residue_range('327:',
'345:')))
#         Residues 354 through 380 should be an alpha helix:
          rsr.add(secondary_structure.alpha(self.residue_range('354:',
'380:')))
#         Residues 386 through 408 should be an alpha helix:
          rsr.add(secondary_structure.alpha(self.residue_range('386:',
'408:')))
#         Residues 417 through 428 should be an alpha helix:
          rsr.add(secondary_structure.alpha(self.residue_range('417:',
'428:')))
#         Residues 433 through 448 should be an alpha helix:
          rsr.add(secondary_structure.alpha(self.residue_range('433:',
'448:')))
```

Figure 3.2.8   Scripting of helices.

Segment from the script that ensures specific residues predicted by SOPMA (secondary structure algorithm) as shown in figure 3.1.2 were modelled as helices within MODELLER. Without this segment of scripting, the regions between CaMBD and CaMLD (that bear < 30% sequence homology to any PBD data) would be modelled as one long extended piece of amino acid residues.

From the SOPMA analysis in figure 3.1.2, the region between the CaMBD and CaMLD of human RyR2 contains 6 beta-turns. A beta-turn is a type of non-regular secondary structure in proteins causing a change in direction of the polypeptide chain. These beta-turns are important in nature since it renders most proteins a globular shape instead of being linear molecules.

beta turn type 1

beta turn type 2

Turn Type	Dihedral angles ($^\circ$)			
	ϕ_{i+1}	ψ_{i+1}	ϕ_{i+2}	ψ_{i+2}
I	-60	-30	-90	0
I'	60	30	90	0
II	-60	120	80	0
II'	60	-120	-80	0
IV	-61	10	-53	17
VIa1	-60	120	-90	0
VIa2	-120	120	-60	0
VIb	-135	135	-75	160
VIII	-60	-30	-120	120
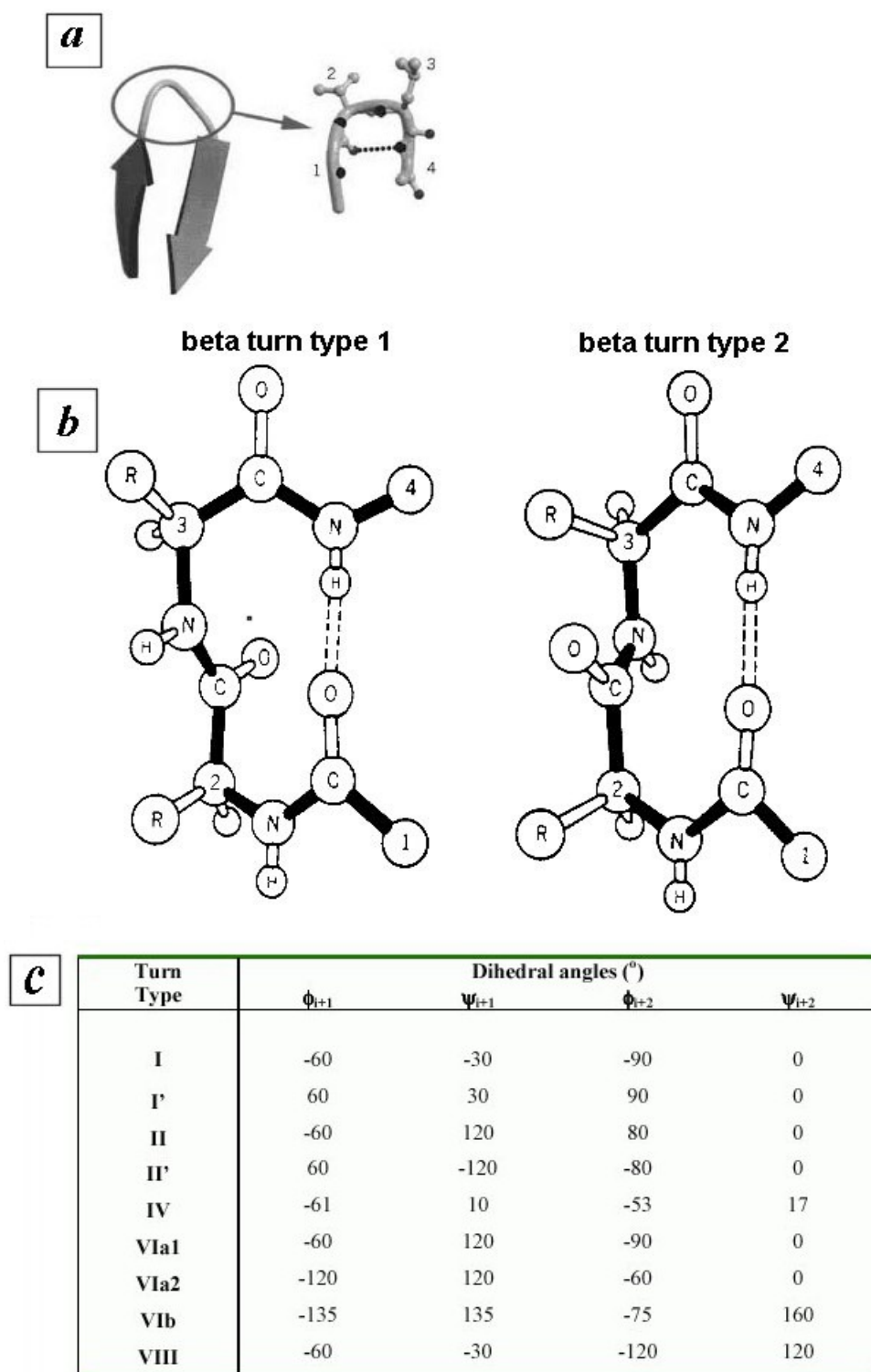
Figure 3.2.9   Beta turns.

In a beta turn, a hydrogen bond is formed between the backbone carbonyl oxygen of one residue (i) and the backbone amide NH of the residue three positions further along the chain (i + 3).

There is another type of structural turn, called a "g-turn", which has a (i) to (i + 2) hydrogen bond. Two thirds of all beta turns that occur in nature are either type I or type II beta-turns. The difference between these turns are, as such, shown below.

| a | Type 1 beta turns - residue preference : |
Position (i) : Asn, Asp, or Ser
Position (i + 1) : Asp, Ser, Thr or Pro
Position (i + 2) : Asp, Ser, Asn, or Arg
Position (i + 3) : Gly, Trp, or Met

Type 2 beta turns - residue preference :

Position (i + 1) : Pro
Position (i + 2) : Asn or Gly
Position (i + 3) : Gln or Arg

b

Beta turns predicted in RyR2 between CaMBD and CaMLD

```
1.   NGGN   : 189 - 193
2.   KKDV   : 206 - 209
3.   EGLG   : 237 - 240
4.   EGHN   : 270 - 273
5.   SGKD   : 316 - 319
6.   EQGQ   : 322 - 325
```

C

43 46   51   56
KPKMIGGIGGFIKVRQY

Figure 3.2.10 Beta-turn residue preference and predicted beta-turns in RyR2.

(a) The various residue preferences in beta-turns that occur in nature are shown. However, these are not definitive and exceptions to these residues do occur. The SOPMA algorithm incorporates various parameters to identify uncommon residues that are involved.

(b) The predicted beta-turns in the RyR2 segment between CaMBD and CaMLD with the residue numbers given (note the residue numbers refer to the simplified residue numbers corresponding to those in figure 3.2.6).

(c) As beta-hairpin turns are ubiquitous in nature, an example of this was taken from PDB entry 1HPV {Kim, 1995 #25}. The sequence between residue 49-52 (Gly-Ile-Gly-Gly) was chosen as the template for modelling beta-hairpins in the RyR2 sequence.

Following the above analysis of the sequence between the CaMBD and CaMLD domains, models were generated using the MODELLER software. For this process a script was written to run directly in MODELLER as opposed to using an automated

front end GUI. This script was initially run without optimisation which generated a number of clashes and erroneous structures as shown below.


```
a.starting_model= 1                    # index of the first model
a.ending_model  = 50                      # index of the last model
                        # (determines how many models to calculate)

# very thorough VTFM optimisation
a.library_schedule = autosched.slow
a.max_var_iterations = 1000

# Thorough MD optimisation:
a.md_level = refine.slow

# Repeat the whole cycle 2 times and
#  don't stop unless obj.fun >1E6
a.repeat_optimization = 2
a.max_molpdf = 1e6

a.make()                        # do homology modeling
```

Figure 3.2.11 Preliminary optimisation of structures.

(a) When structures were generated within MODELLER without energy minimisation, erroneous structures were obtained. Examples of these show where loops traverse within helices and loops occur within loops, both instances are not possible in nature. Such examples also highlight the need for visualisation and examination of structures by eye.

(b) To circumvent and prevent erroneous structures occurring, optimisation using the Variable Target Function Method (VTFM) that incorporates the conjugate gradient (CG) method was used. The script for this is shown here. The molecular dynamics (MD) optimisation step incorporates a simulated annealing (SA) protocol for optimisation.

As with the modelling of the CaMBD and CaMLD regions, DOPE scores are used to evaluate the model quality. DOPE score analysis requires comparison to known structures, since a large proportion of this model does not have any sequence identity to any known experimental structure so DOPE score is not entirely relevant as an

assessment tool for the models generated. Despite this issue the DOPE scores can be used as guide towards indication of better quality models. As previously, lower DOPE and low molpdf scores, and GA341 scores closer to 1 are desirable. The evaluation of a selection of the better models generated is shown below.

model no.	molpdf	DOPE	GA341
2	2135.01514	-33378.89844	1.0
7	2170.64136	-32306.58594	1.0
10	4140.45801	-31879.91211	1.0
12	2341.98706	-34170.73828	1.0
14	3968.89355	-32610.49609	1.0
19	2063.74683	-33129.16016	1.0
24	3582.28345	-32514.34961	1.0
25	2179.01733	-34727.08594	1.0
30	2051.01611	-34238.48438	1.0
31	2517.95605	-33879.74609	1.0
39	4442.13916	-32013.15820	1.0
41	1321.98376	-33801.53516	1.0
42	1445.89270	-34180.05078	1.0
43	1173.77771	-34627.82812	1.0
46	2194.13086	-32852.93750	1.0
47	3564.65674	-32716.01562	1.0
48	4012.54736	-31355.48828	1.0
49	1196.78613	-33920.44531	1.0
50	2861.83032	-35177.90234	1.0

Table 3.2.2   Assessment scores for selected models.

The various computer-generated models are shown here with the MOLPDF and DOPE scores.   Models that have a combination of the lowest molpdf values and the lowest (more negative) DOPE scores were desirable
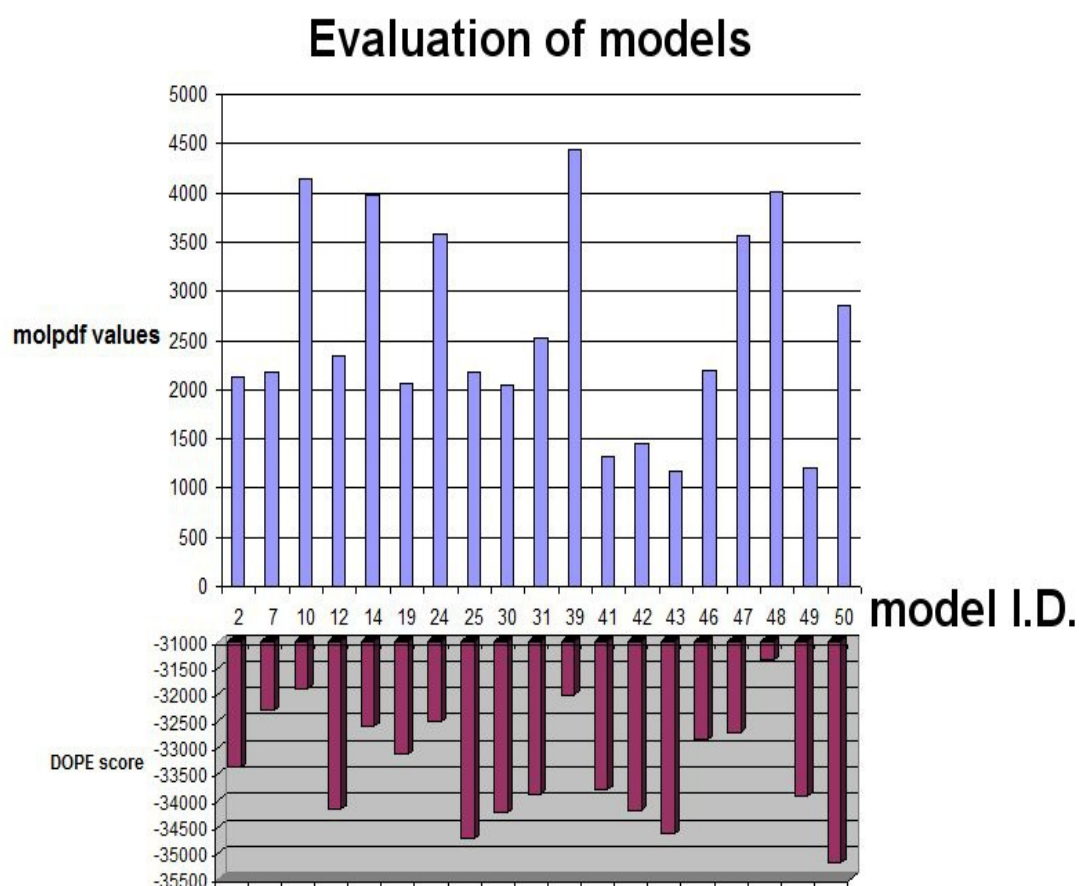
Figure 3.2.12 Evaluation of models by molpdf and DOPE scores.

Models that were successfully generated and had molpdf values less than $1x10^6$ are shown above. Models that have a combination of the lowest molpdf values and the lowest (more negative) DOPE scores were desirable.

However, following optimisation by conjugate gradients (CG) minimisation and MD-SA, the models were evaluated by Ramachandran analysis for any deviations of the phi-psi torsion angles.

Ramachnadran plots were generated for the 4 models, selected from the table below, which had no residues in the disallowed regions, so considered potentially better models.

model no.	Number of residues in region			
	Most favoured	Additional allowed	Generously allowed	Disallowed
2	*449*	*16*	*3*	*0*
7	436	16	9	7
10	*438*	*25*	*5*	*0*
12	451	10	4	3
14	425	27	6	6
19	454	10	1	3
24	448	15	4	1
25	*442*	*24*	*2*	*0*
30	441	20	6	1
31	440	19	8	1
39	431	31	5	1
41	*449*	*16*	*3*	*0*
42	449	16	2	1
43	449	18	0	1
46	449	15	3	1
47	422	34	9	3
48	430	31	6	1
49	456	8	2	2
50	443	18	4	3

Table 3.2.3   List of residues by Ramachandran analysis region.

Models with no residues in disallowed regions are highlighted in the table.
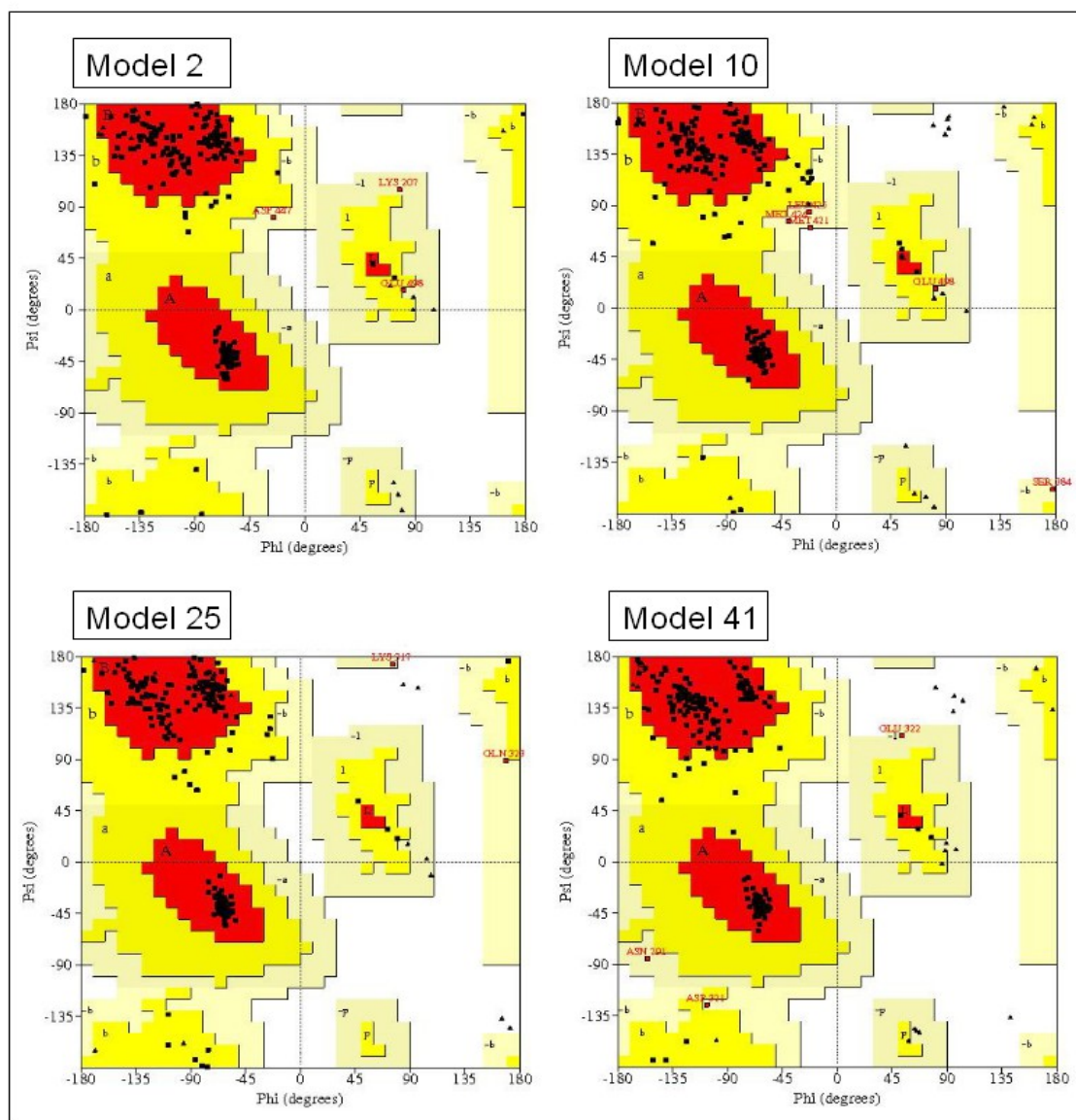
Figure 3.2.13 Ramachandran plots of selected models.

Models with no residues in disallowed regions are shown.

Although no residues were present in the disallowed regions for these model indicating reasonable models, further energy minimisation steps were run on the selected models to see if it was possible to increase the quality of the model. Steepest descent minimisation was used to try to improve models 2, 10, 25 and 41. Each model was run through 1000 steps of steepest descent and then evaluated by Ramachandran analysis.
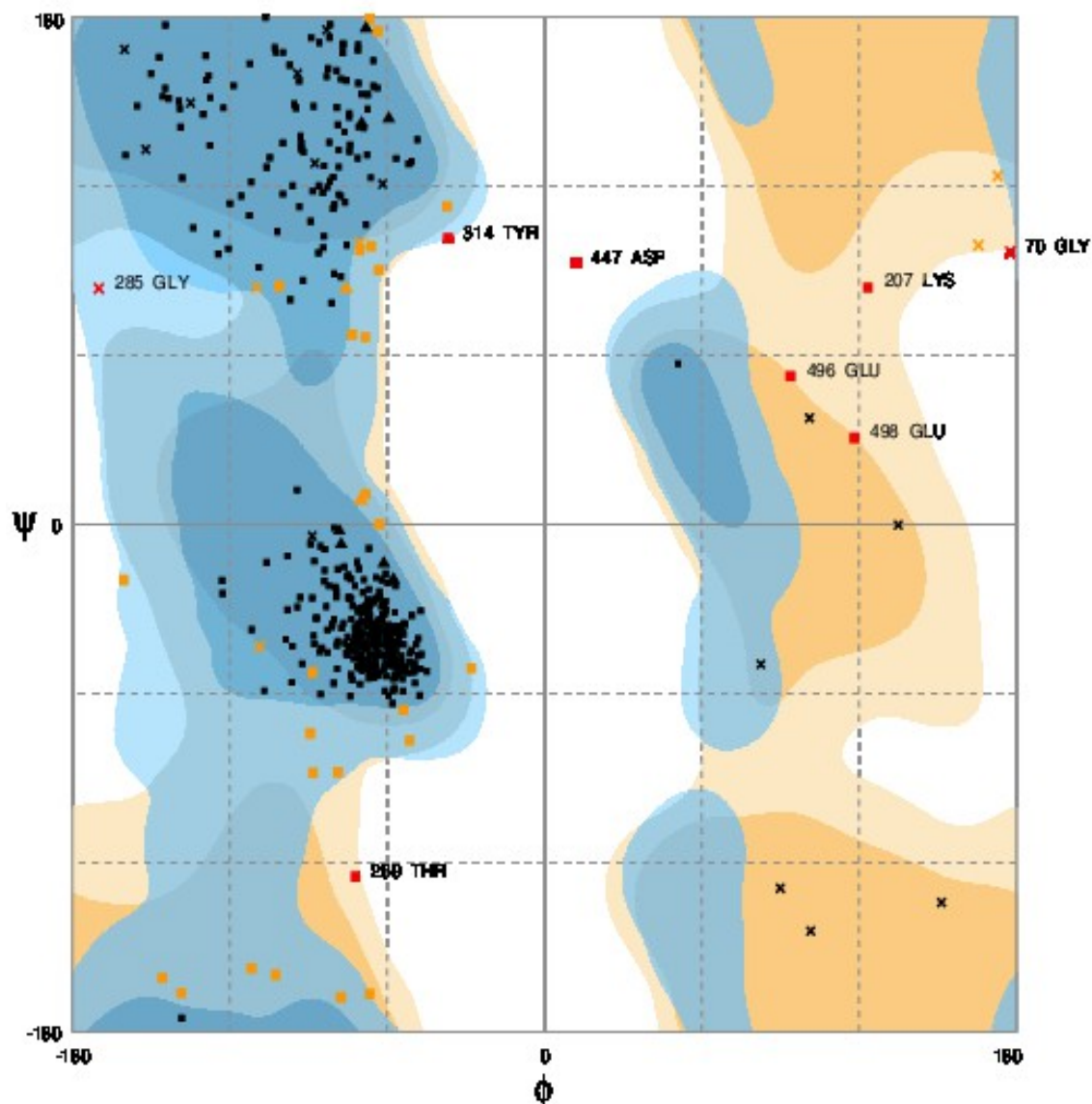
Figure 3.2.14 Ramachandran plots of model 2 after 1000 steps of steepest descent minimisation.

Residues in disallowed regions are shown in red.

Prior to the 1000 steps of steepest descent energy minimisation model 2 had no residues in the disallowed region, post this step there were 8. A similar increase in residues in disallowed regions was observed for the other selected models after the steepest descent minimisation. This indicates that the quality of the models has decreased. As the steepest descent minimisation is a relatively rapid means of decreasing the overall energy state of the model this may be a too rapidly drastic technique for these models. The loop region of these models contains a number of residues which may be prone to unfavourable conformations when exposed to steepest descent minimisation. While the

overall energy state of the model may appear to be decreasing bonds within these amino acids have been forced into unrealistic positions so actually worsening the model.

Conjugate gradient minimisation unlike steepest descent allows for an a potential increase in energy within the model per step while trying to reach an overall lower energy state, usually after many steps. This approach potentially allows more freedom of movement of the residues which may have the possibility of reducing the number of unfavourable residue positions. To reduce computing time model 2 underwent 5000 steps of conjugated gradient minimisation and the resulting model was tested by Ramachandran analysis.

Figure 3.2.15 Ramachandran plot of model 2 after 5000 steps of conjugate gradient minimisation.

Residues in disallowed regions are shown in red.

As with the steepest descent approach a clear increase in the number of resides in disallowed regions can be observed. 24 residues were in disallowed regions post the conjugate gradient minimisation.

3.3 Discussion

The aim of this chapter was to simulate regions of the C terminal of human RyR2 following two main stages. Firstly, the production of suitable quality models of CaMLD and CaMBD regions. If this stage was successful, secondly, model the entire region between and including the two domains and then attempt to improve the quality of this model.

The first step of stage one was the selection of existing known structures to serve as templates to the models. As discussed the percentage matching homology between the amino acid sequence to be modelled and those of published structures has to be >30% to be able to produce a model of acceptable accuracy. In the case of CaMBD there are structures of CaM binding proteins in the PDB. For this project the CaM binding protein region of RyR1, equivalent to CaMBD in RyR2, in PDB entry 2BCX served as an excellent donor template. The CaMLD region of RyR2 proved to be more of a challenge to find a suitable matching structure. As CaMLD is meant to be, to some extent, a CaM mimic in terms of binding function to CaMBD, so it was decided to use the CaM structure in the PDB entry 2BCX. CaM meets the 30% homology requirement and also includes all the secondary structure elements assumed to be required for CaMLD binding to CaMBD. The use of these templates allowed models to be generated that met the requirements of stage one.

Stage 2, the modelling of the region 3578-4085 proved to contain a significant number of technical challenges. Primary of these was the strategy required to model the region between the two known domains as there was no available template for this region. If this region has been modelled completely *ab initio* without defining sections within the area then it is likely that a long string of undefined structure with the two defined domains at either terminal would have been produced. The first step to correct this was to set distance restriction limits on how far away CaMBD and CaMLD could be from

each other. As 2BCX, the donor template for the domains, consists of CaM interacting with a CaMBP so the distances between selected residues were measured and these distances applied to the equivalent residues in the RyR2 sequence. The second step was to analyse the predicted secondary structure of the region between the domains. Areas of α helix and beta turns were identified and the amino acid coordinates of these areas logged. Scripts were written to tell the MODELLER software of these logged areas so they would be produced in the models of the region.

It was initially thought that the two steps described above would be sufficient for the generation of reasonable quality models of the region. However examination of the produced models revealed a number of problems, as seen in panel A in figure 3.2.11, including physically impossible structural elements. To address these issues short energy minimisation and molecular dynamics steps were added to the modelling script. The use of these steps was aimed to generate models in which were in a more favourable energy state. While a wide range of possibilities were expected with this approach including models with potentially even more flawed structures it was hoped to produce some models which had reduced overall energy states while also having viable structures. Initially 50 models were generated using the revised script. As expected a number of these models had worse structures which was reflected in high molpdf scores indicating that they were less energetically probable as viable models. The models which appeared to be potential solutions were analysed by Ramachandran plot, of these 4 had no residues in disallowed regions.

It was decided to attempt to increase the quality of these models by using further energy minimisation steps. Energy minimisation steps can be conducted using a solvated molecule. Water shells or a water box is placed around the molecule simulating an environment closer physiological conditions than a vacuum. Boundaries can be placed to limit the size of the solvent area that is simulated around the molecule though solvent-solvent interactions are modelled within these boundaries. Even with increasing available computing power these calculations can be very taxing on the simulating system and severely slow down the minimisation stages of refinement. As the models produced by this project it would have meant that a relatively very large volume of water would have had to been simulated to solvate the model. As such it was decided to conduct the minimisation stages in a vacuum to reduce processing time.

At first steepest descent was trialled by use of 1000 steps. Steepest descent applies a simpler algorithm than other minimisation techniques. This algorithm moves the coordinates of the model towards energetically lower states and allows for the most rapid convergence to an achievable minimum energy state. This technique tends converge to an overall minimum for the whole molecule with little tolerance for localised structure. This can force the model into unrealistic localised structures which although appearing to reduce the overall energy state of the molecule in that step actually adversely limit the progress towards the potential minimum. It is likely that this is the outcome that was observed with the 4 models trialled with this method as their overall structures were shown to worsen. Steepest descent is commonly used in limited steps to improve new models. While a rapid step, on reflection this may not have been a good choice as an approach as the models had already undergone limited minimisation steps during the modelling process.

Conjugate gradient minimisation was subsequently used on model 2. This method is iterative taking into account the previous steps as well as progression towards an energy minimum in determination of the next step. As a result this technique tends to have better convergence characteristics towards an energy minimum state than steepest descent. However this method can become limited in attempts to move coordinates with no real overall improvement of the energy state over several steps and still enforce worse structures upon the model. Some protein conformations are more prone to suffer these problems and it is likely that the modelled region is one of those structures. As a result it was decided to not trial any of the other models with this technique. Improvements to the basic conjugate gradient method are included in the CHARMM software. Mostly these rely on closer step based monitoring of the structure and can place boundaries of acceptable movement for regions within the molecule which can be updated in a series of blocks of steps rather than step by step potential variation. However owing to the overall loop shape of the modelled region of RyR2 it is unlikely that these alterations to the method would result in an improved structure.

Although there are other minimisation algorithms available within the CHARMM software as the two methods trialled have both led to an overall decreasing of the quality of the model it was not viewed as being worthwhile to pursue further minimisation techniques. The improved modelling process already contains some

limited minimisation, and also molecular dynamics, steps so further minimisation does not appear to yield any further benefits. Selected models generated from this modelling process already have no residues present in disallowed regions when examined by Ramachandran analysis. This meets the basic criteria of the second stage of the chapter and the overall aim which was to produce a model suitable to undergo molecular dynamics to enable further analysis of the function of CaMBD and CaMLD regions of RyR2.

Chapter 4:

Refinement of the computational model

of human RyR2

4.1 Introduction

The region of human RyR2 between amino acid residues 3579 – 4085 is an important region that has a putative regulatory role influencing the function of the overall protein as a calcium channel. Thus, studying how the structure of this region contributes in this regulatory role would aid in the understanding of the function of the channel and the impact of disease related mutations within this area.

The tertiary folding of secondary structure elements into domains is a key stage in attaining the functional conformation of a protein. The tertiary structure of a domain often facilitates the function of that domain. The loop model generated in chapter 3 was processed to a state in which it contained relevant secondary elements and the overall structure could not be further energy minimised. However to understand the regulatory role of this region the structure needs to be folded into a three dimensional state. This chapter explores the *in silico* theoretical folding of the protein.

4.1.1 The problem of protein folding

There are numerous ways that a polypeptide chain can possibly fold but there are certain steric limitations where if two atoms were too close together, they would face large opposing forces preventing that conformation from being stable in nature. Normally three conformations within polypeptide chains persist in nature that connects one amino acid to two others (one at the N-terminus and the other at the C-terminus). These are the extended conformation, cis-configuration and the trans-configuration as shown in the figure below.
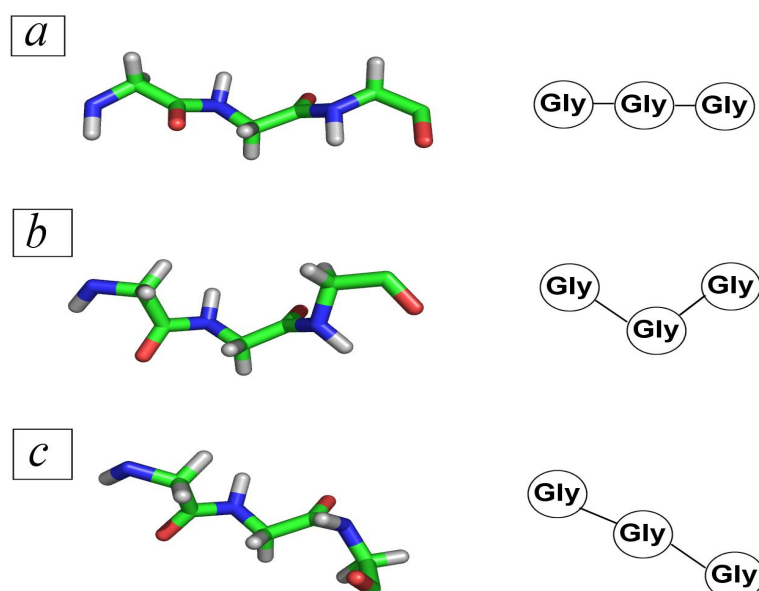
Figure 4.1.1   A triple Glycine repeat (Gly-Gly-Gly) in three conformations that are sterically stable.

a) extended conformation

b) cis configuration

c) trans configuration

Assuming these 3 states are the only ones allowed in nature, and bearing in mind that the first and the last amino acid only connect with one other amino acid, so in a polypeptide chain of "n", where "n" is the number of amino acids in the chain, n-2 amino acids are able to exist in any of the 3 conformations as shown in figure 4.1.1.

Thus, for the length of the section of RyR2 protein that has been modelled, between amino acids 3579-4085, giving a polypeptide length of 506 residues, there are (506-2) or 504 residues that can exist in 3^{504} or 2.945 x 10^{240} different conformations.

If protein folding were a completely random process in nature, then each protein being synthesised in the cell would undergo random perturbations in conformation. For example a 506 amino acid residue would undergo 2.945 x 10^{240} conformational changes until it arrived at the final "desired one", this would logically take a very long time. Hypothetically, if a million conformations could be tried in a second, a polypeptide of this length would take:

2.945 x 10^{240} /10^6 = 3.4 x 10^{229} days

However, it is know that in nature, from experiments on the folding of large proteins such as cytochrome C that proteins fold in millisecond time-scales (Yang & Smith 1997). The rationalisation of the protein folding problem is not a new concept, the apparent incredulity at the speed at which proteins fold in nature unlike the predicted theoretical time it would take if it were a random process was first discussed by Cyrus Levinthal in 1969 (Levinthal 1969). This apparent disagreement of the theoretical rate of folding with that observed in nature is known as "Levinthal's paradox" (Zwanzig et al. 1992), (Ben-Naim n.d.).

It could be logically predicted that the end conformation after so many attempts would differ each time if the folding process were truly random. Of course, each polypeptide chain would not have to undergo every change and the criteria to stop trying different conformations would include a random stop command. This does not happen in nature. Each polypeptide chain folds to a destined shape, thus there has to be a set of predefined rules that guides the folding process.

Early research into folding proposed that the code for tertiary folding must be somehow included in the amino acid sequence of the protein itself. Nobel laureate, Christian Anfisen had conducted a historic experiment to test this theory (Anfinsen 1973).

The Observation:

Native (100% active) → 1. Reduce 2. 8 M urea → Denatured (inactive) → 1. Remove urea 2. Oxidize → Native (>90% active)

The Control:

Native → 1. Reduce 2. 8 M urea → Denatured → 1. Oxidize 2. Remove urea → "Scrambled" (1-2% active)
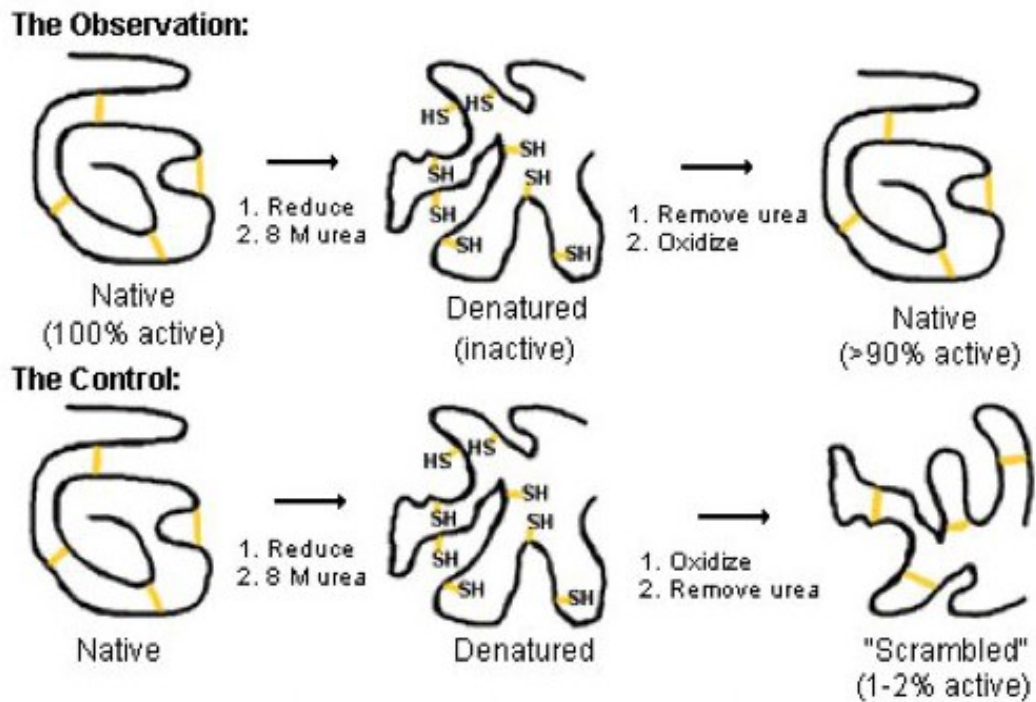
Figure 4.1.2   The Anfisen experiment.

Ribonuclease (124 aa) contains four disulphide bridges. When ribonuclease was treated with 8M urea + β-mercaptoethanol, the protein linearised and lost enzymatic activity. When this linearised, denatured protein was allowed to oxidise in the presence of urea (as shown in the control experiment above), the sulphydryl (SH) groups randomly pair, forming four disulphide bridges. Despite the formation of the S-S bonds, enzymatic activity was still extremely weak or lost. However, when the 8M urea was completely dialysed out, and the S-S bridges were then allowed to form, enzymatic activity was restored.

The Anfinsen experiment proposed support for what became known as the "thermodynamic hypothesis". This hypothesis states that the native folded state of a protein in nature is the one where the Gibb's energy of the whole system is the lowest. The Anfinsen experiment supported the concept that the information which determines the way a protein folds is contained within the amino acid sequence. The unknown nature of this information which influences the tertiary folding of a polypeptide chain is the essence of the "protein folding problem".

Further research has been conducted on this problem. Martin Karplus and workers proposed the "diffusion-collision-adhesion model" to address the protein folding

problem (Karplus & Weaver 1994). In this model a polypeptide chain was computationally simulated in which micro-domains of the polypeptide were encouraged to move diffusively and repeatedly collide. Another model is the "hydrophobic collapse" model (Dill 1985). In this model, the various hydrophobic stretches of amino acids gradually eliminate water from their vicinity bringing about a collapse of the structure, thereby two previously distant hydrophobic regions gradually migrate towards each other. Many other proposed models of protein folding exist but it is likely that many of these models are in fact working together in nature to enable the complex folding process.

4.1.2 Simulating protein folding

In terms of structure, very little data was available pertaining to the protein sequence between amino acid residues 3579 – 4085 of human RyR2. As discussed in Chapter 3, this region is known to contain a domain (CaMBD) that interacts with calmodulin (CaM), subsequent mutation studies of these regions have demonstrated loss of interaction with calcium (N. Yamaguchi et al. 2003). Additional studies have implicated another domain (CaMLD) with similar structure and function to CaM (L. Xiong et al. 2006), (Gangopadhyay & Ikemoto 2006), (Zhu et al. 2004). As little information is known about the intervening region between the two domains, so this region was subjected to a secondary structure prediction using the SOPMA algorithm (Geourjon & Deleage 1995). Afterwards the region was folded in MODELLER (A. Sali & Blundell 1993) using the secondary structure prediction as demonstrated in figure 3.2.8.

In this chapter, the tertiary fold of all the section in this region were encouraged to evolve using molecular dynamics software within CHARMM (Brooks et al. 2009). The PDB model files generated in MODELLER (A. Sali & Blundell 1993) were imported into INSIGHT 2000 (Accelrys) in order to convert the coordinates to the CHARMM format and to generate the protein structure files (PSF). Based on the findings of chapter 3, the models were then subjected to closely monitored energy minimisation using improved algorithms. The minimised structures were then subjected to 1000 steps of "heating" from 240 K to 800 K. The final structure from this heating step was then fed into an "equilibration" cycle where the temperature was gradually reduced from 800 K

down to 310 K. The temperature values quoted are not to be literally interpreted as a protein heated to 800 K (527°C) would obviously have denatured well before this temperature was reached. Temperature in molecular dynamics is a reference to a statistical quantity.

To fully appreciate the contribution of temperature in molecular dynamics (MD), the following discussion is necessary.

Newton's first law of motion is stated as "objects at rest remain at rest and a moving object remains in motion with the same speed and direction unless an unbalanced force acts upon the object either changing its direction or altering the speed". This is represented as follows:

$$F = ma$$

Where F = force, m = mass of the object and a = acceleration.

Acceleration can also be represented as dv/dt where v = velocity and t = time.

An adiabatic system is one where thermodynamic energy or heat cannot enter or escape the system. Observing energy changes in such a system is fortuitous, since one confounding variable (namely heat) is eliminated, therefore any change in energy would have to be due to either changes in intra-molecular forces or changes in kinetic energy. Within an adiabatic system, it is possible to observe changes in potential and kinetic energies with no change in the total energy. Within such an adiabatic system, Boyle's law for perfect gasses applies, this is summarised as below:

$$PV = nRT \qquad or \qquad PV = nkT$$

Where P = pressure, V = volume of the system, n = number of moles of molecules/atoms in the system, R = gas constant, k = Boltzmann's constant and T = temperature.

From the equipartition theorem, for a system in thermal equilibrium (such as in a closed adiabatic system) all matter in such a system has the following average degrees of

freedom (degrees of freedom meaning independent ways in which matter can move):

kT / 2

Where k = Boltzmann constant and T = temperature.

Since there are 3 coordinate axes specifying three directions of possible motion (hence 3 directions for the degrees of freedom), the total kinetic energy of the system can be thus represented as:

$E_{kinetic} = 3 \times kT/2$

But kinetic energy can also be represented as follows:

$E_{kinetic} = \frac{1}{2} mv^2$

Thus the previous two equations can be represented as follows:

$3kT/2 = \frac{1}{2} mv^2$

and therefore the relationship of "temperature" and "velocity" in the previous equation can be thus represented as follows:

$T \approx v^2$

So "increasing the temperature" in a molecular dynamics simulation is a computational way of increasing the velocity of atoms/molecules in a computer model. Obviously, the amount of kinetic energy provided to the system should not result in abnormal fluctuations in the bond-lengths or bond-angles.

```
! Script to read parameter, psf and crd file
!
OPEN READ UNIT 21 CARD NAME "/xtal/CHARMMPARAM/top_all27_prot_na.rtf"
READ RTF UNIT 21 CARD
CLOSE UNIT 21
OPEN READ UNIT 21 CARD NAME "/xtal/CHARMMPARAM/charmm27_support.rtf"
READ RTF UNIT 21 CARD APPEND
CLOSE UNIT 21
OPEN READ UNIT 20 CARD NAME "/xtal/CHARMMPARAM/par_all27_prot_na.prm"
READ PARA UNIT 20 CARD
CLOSE UNIT 20
```

Figure 4.1.3   Example of section MD script referring CHARMM software to use set of parameters.

In an MD system the bond-lengths, angles etc. are kept to within empirically determined parameters (shown in italics and underlined) by the code above in the molecular dynamics script.

Otherwise, the molecules can also be restrained within empirical parameters using the SHAKE algorithm (16). Thus in summary, "movement" of molecules in a computer system can be simulated at different kinetic energy levels using the relationship of kinetic energy to heat energy. This way, there can be an audit trail of various fluctuations of energy within the system.

4.2 Results

4.2.1 Computer simulated folding of RyR2 aa 3579-4085

The final configuration of the models following *in silico* folding by MODELLER, as described in Chapter 3, is shown below.
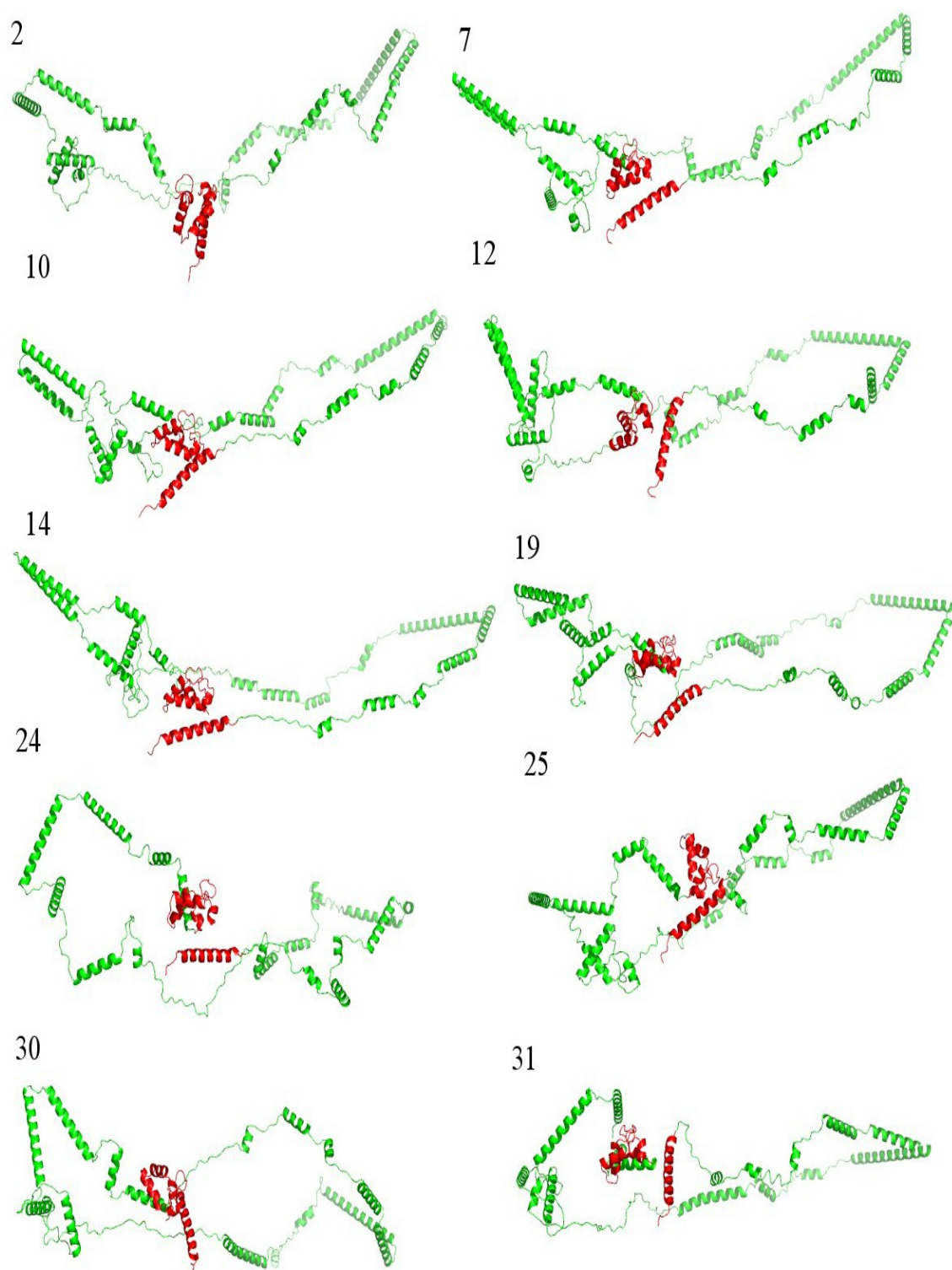
Figure 4.2.1a The initial models produced from homology modelling using MODELLER as described in chapter 3.

The CAMLD and the CAMBD domains are both highlighted in red.

39

41

42

43

46

47

48

49

50
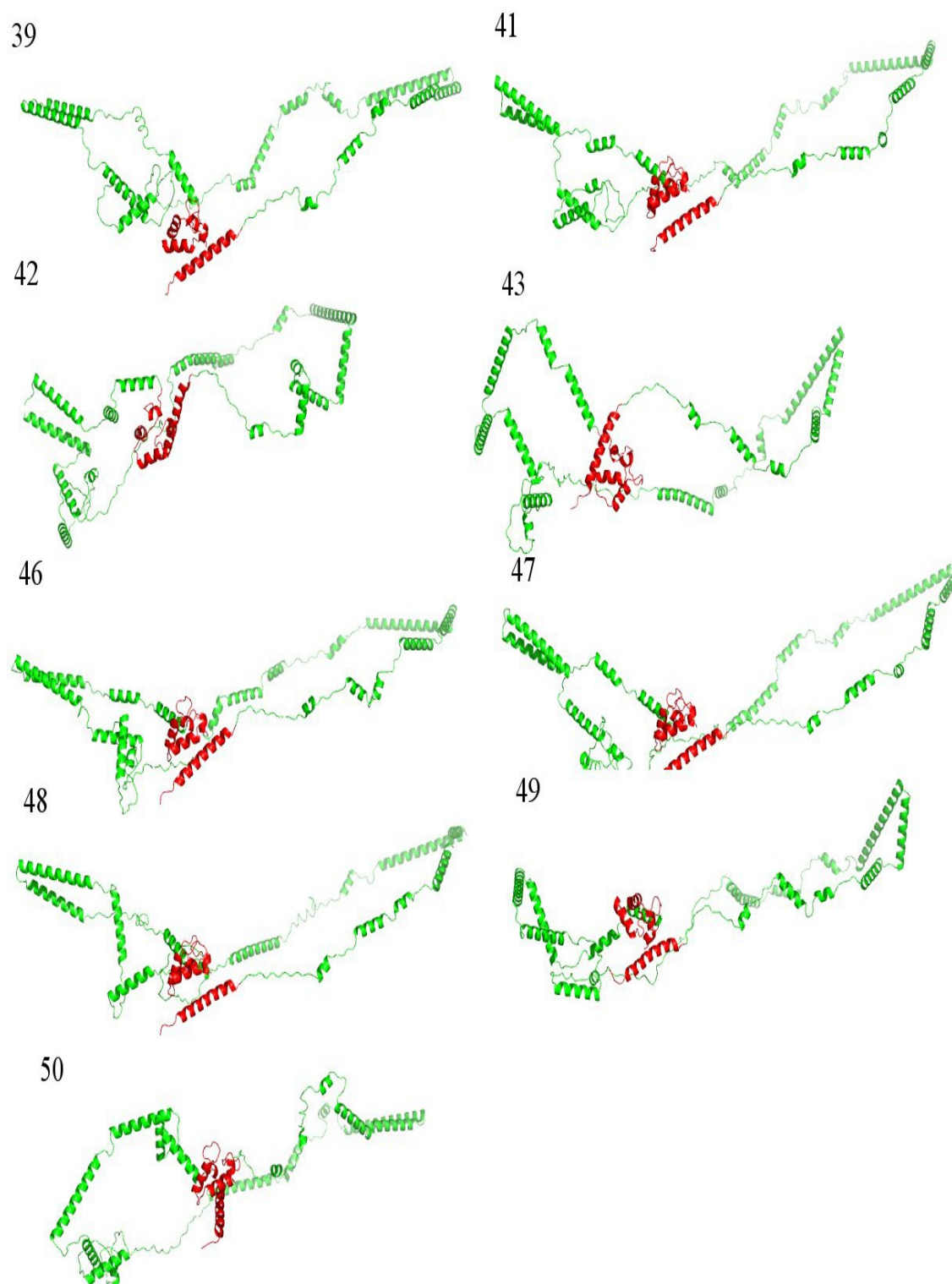
Figure 4.2.1b The initial models produced from homology modelling using MODELLER as described in chapter 3.

The CAMLD and the CAMBD domains are both highlighted in red.

Various conformations are present between the 2 domains since the loop regions (regions of high entropy and disorder) may adopt any conformation and thus influence the overall shape of the molecule. As discussed in chapter 3, large portions of the RyR2 region amino acids 3579-4085 had to folded using *de novo* methods. Minimal energy minimisation was carried out within MODELLER itself (using the conjugate gradient method). These models were used as a starting point for subsequent simulation.

Each of the models shown in figure 4.2.1 were subjected to 1000 steps of steepest descent and 1000 steps of conjugate gradient energy minimisation using CHARMM. Each energy minimisation run was scripted to produce a detailed log file. Figure 4.2.2 shows the gradual minimisation of overall energies using both methodologies on the various models that were generated.
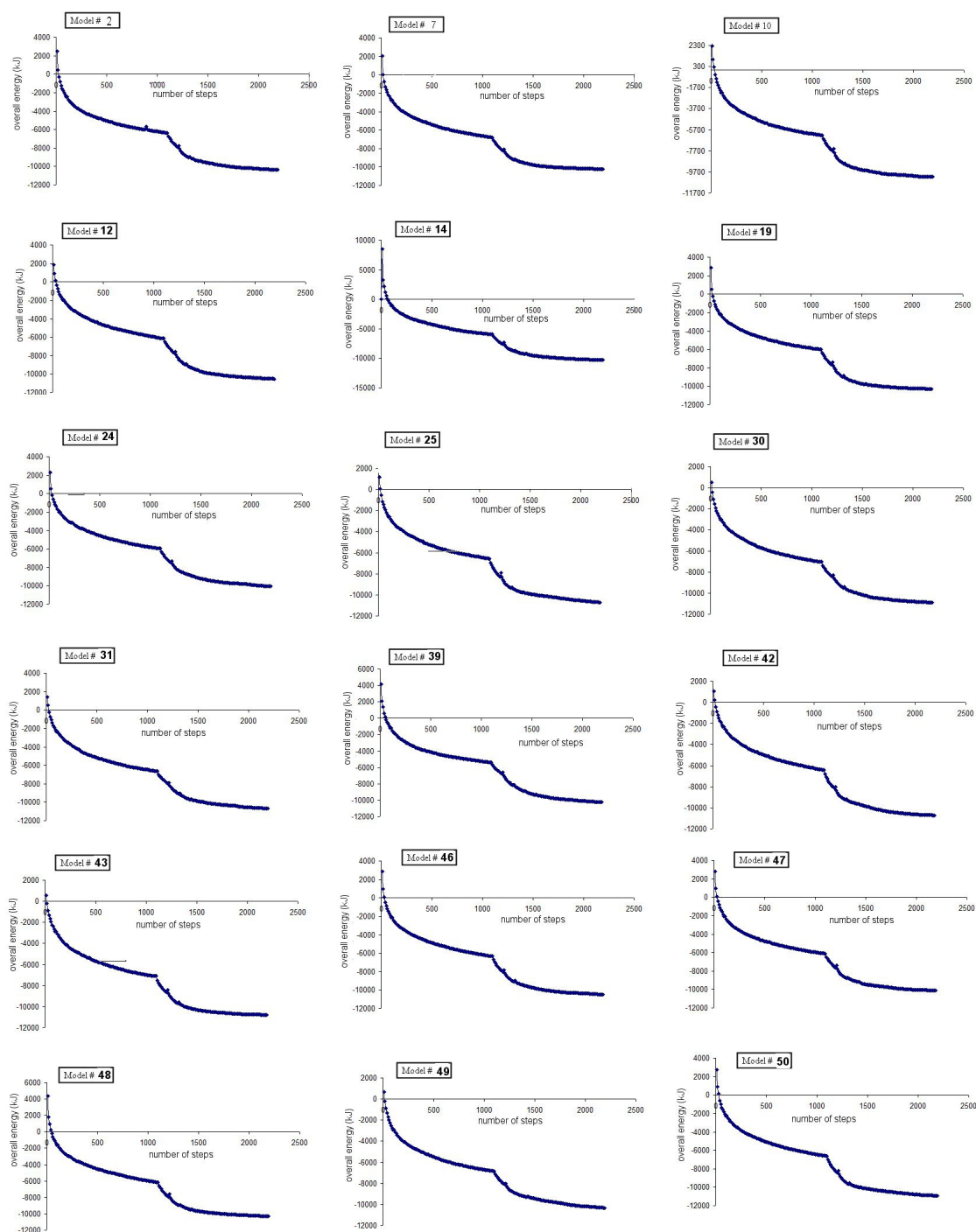
Figure 4.2.2 Energy minimisation of the various models generated using MODELLER.

The first concave curve portion signifies the energy minimisation using the steepest descent protocol. The second concave curve shows the minimisation carried out using the conjugate gradients protocol. The data was extracted from the log files using the following script: awk '/MINI>/ {print $2, $3;}' name_of_logfile.log > energy_output.txt., the graphs were plotted using gnuplot.

The composite figures in figure 4.2.2 show that both methodologies of minimisation were necessary to achieve a final structure that had all the bond angles, bond lengths and erroneous distances between atoms corrected according to empirically determined values contained in the CHARMM parameter files. The first hyperbolic curve portion signified the minimisation carried out using the steepest descent protocol, this method quickly corrects any gross problems with the structure but does not approach minima easily. Subsequent minimisation, indicated by the second hyperbolic curve portion, using the conjugate gradients method, further corrected any problems and reduced the overall energy to that one of a more stable structure.

Between each minimisation step, there would be changes to the overall structure and this can be monitored by evaluating the root-mean square deviation (RMSD) of the positions of the C-alpha carbon atom chain between the starting model and that of the final refined structure. If there were no difference between the two compared structures, the RMSD value would be zero. Lower values of RMSD indicate a lesser degree of deviation of the geometry of the compared structure against the starting model. Figures 4.2.3-5 show the RMSD deviations of the various models following *ab initio* folding by MODELLER, then refinement using steepest descent minimisation followed by conjugate gradients minimisation. Also shown in these figures are the structures of the domains CAMLD and CAMBD.
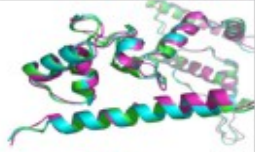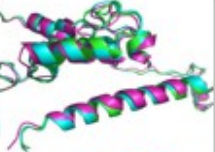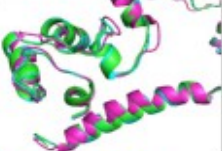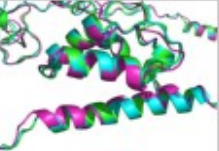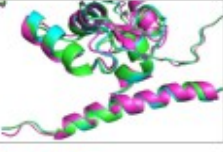
Starting model #	RMSD after SD refinement	RMSD after CG refinement	Alignment of model, SD and CG refined models
2	0.43	1.346	
7	0.445	1.456	
10	0.456	1.413	
12	0.451	1.458	
14	0.468	1.458	
19	0.421	1.436	

Figure 4.2.3   RMSD values following energy minimisation.

The difference between the original model and the refined model after steepest descents (SD) and conjugate gradient (CG) minimisation are given above. A value of 0 indicates that there is no difference between the compared structure and the original model. The larger the value, the greater the difference between the two compared structures.

Cyan = original model, Green = SD refined, Magenta = CG refined.

Starting model #	RMSD after SD refinement	RMSD after CG refinement	Alignment of model, SD and CG refined models
24	0.426	1.399	
25	0.432	1.405	
30	0.435	1.426	
31	0.458	1.478	
39	0.456	1.607	
42	0.444	1.330	

Figure 4.2.4   RMSD values following energy minimisation.

The difference between the original model and the refined model after steepest descents (SD) and conjugate gradient (CG) minimisation are given above. A value of 0 indicates that there is no difference between the compared structure and the original model. The larger the value, the greater the difference between the two compared structures.

Cyan = original model, Green = SD refined, Magenta = CG refined.

Starting model #	RMSD after SD refinement	RMSD after CG refinement	Alignment of model, SD and CG refined models
43	0.432	1.343	
46	0.468	1.440	
47	0.457	1.451	
48	0.474	1.502	
49	0.408	1.450	
50	0.443	1.438	

Figure 4.2.5   RMSD values following energy minimisation.

The difference between the original model and the refined model after steepest descents (SD) and conjugate gradient (CG) minimisation are given above. A value of 0 indicates that there is no difference between the compared structure and the original model. The larger the value, the greater the difference between the two compared structures.

Cyan = original model, Green = SD refined, Magenta = CG refined.

Ramachandran analysis of the models post these stages of minimisation was conducted.
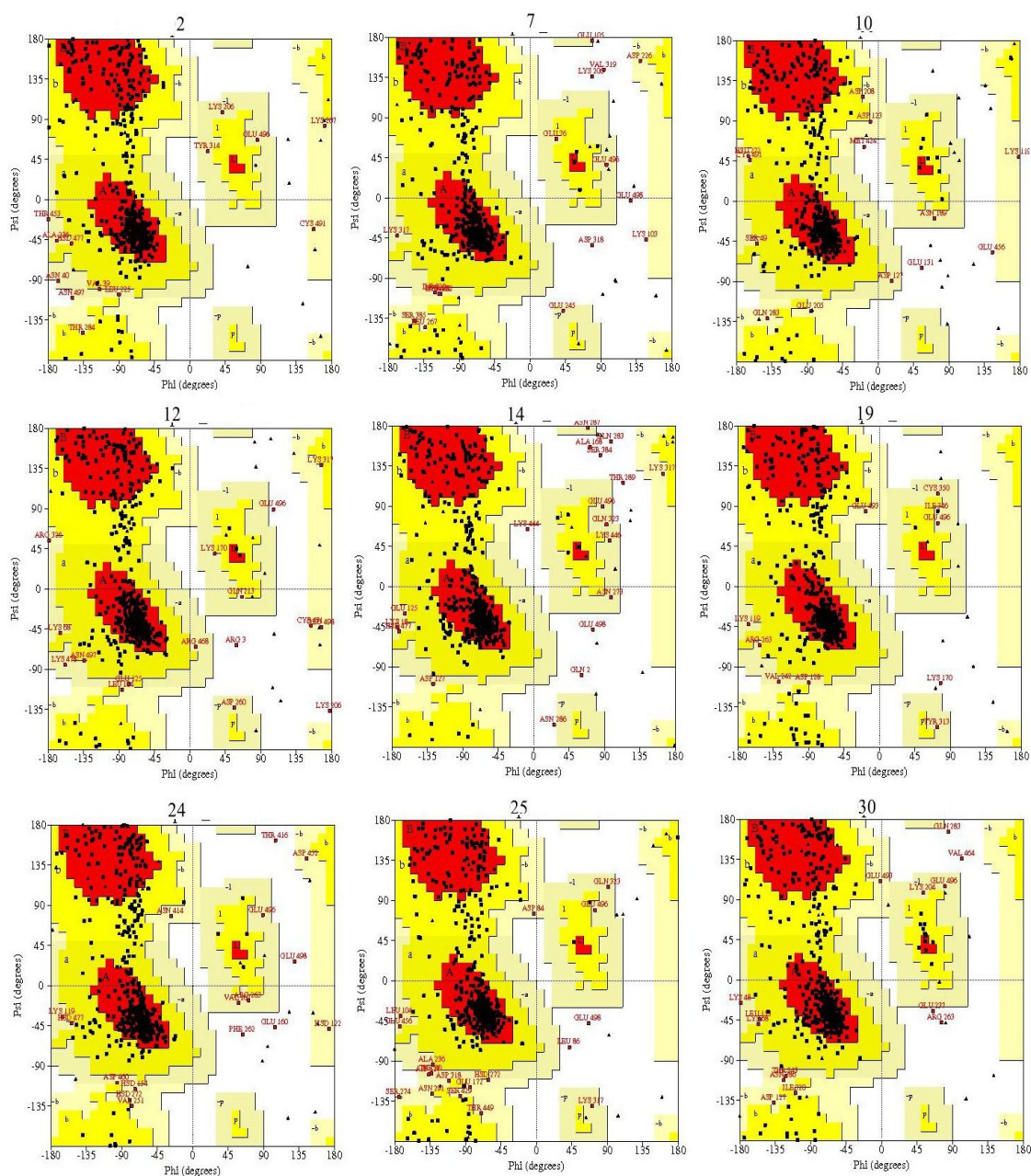
Figure 4.2.6a Composite of Ramachandran plots of models post-refinement (SD and CG).
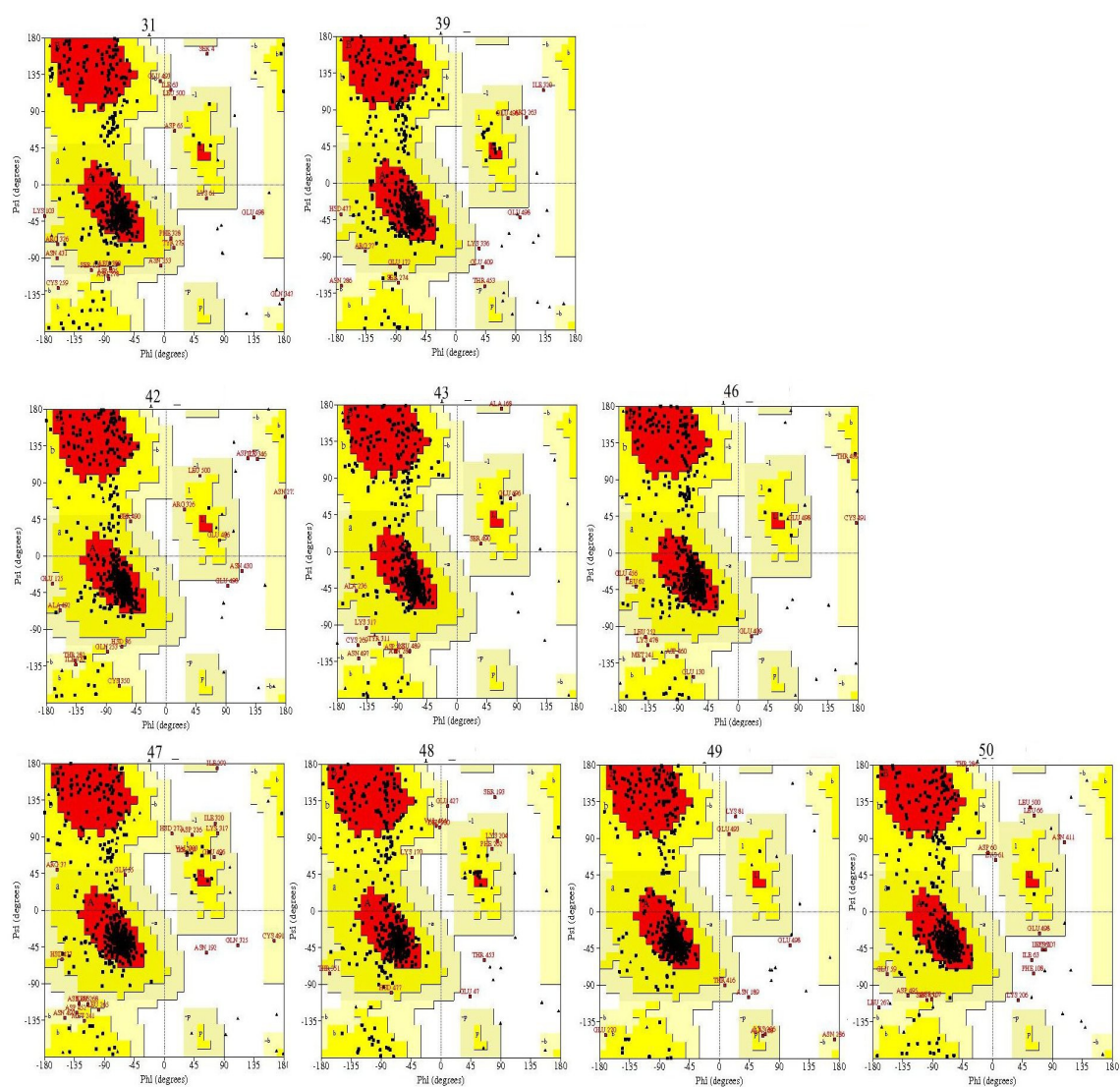
Figure 4.2.6b Composite of Ramachandran plots of models post-refinement (SD and CG).

2

Plot statistics

Residues in most favoured regions [A,B,L]	368	78.6%
Residues in additional allowed regions [a,b,l,p]	87	18.6%
Residues in generously allowed regions [~a,~b,~l,~p]	13	2.8%
Residues in disallowed regions	0	0.0%
Number of non-glycine and non-proline residues	468	100.0%
Number of end-residues (excl. Gly and Pro)	1	
Number of glycine residues (shown as triangles)	28	
Number of proline residues	10	
Total number of residues	507	

7

Plot statistics

Residues in most favoured regions [A,B,L]	361	77.1%
Residues in additional allowed regions [a,b,l,p]	91	19.4%
Residues in generously allowed regions [~a,~b,~l,~p]	11	2.4%
Residues in disallowed regions	5	1.1%
Number of non-glycine and non-proline residues	468	100.0%
Number of end-residues (excl. Gly and Pro)	1	
Number of glycine residues (shown as triangles)	28	
Number of proline residues	10	
Total number of residues	507	

10

Plot statistics

Residues in most favoured regions [A,B,L]	356	76.1%
Residues in additional allowed regions [a,b,l,p]	99	21.2%
Residues in generously allowed regions [~a,~b,~l,~p]	10	2.1%
Residues in disallowed regions	3	0.6%
Number of non-glycine and non-proline residues	468	100.0%
Number of end-residues (excl. Gly and Pro)	1	
Number of glycine residues (shown as triangles)	28	
Number of proline residues	10	
Total number of residues	507	

12

Plot statistics

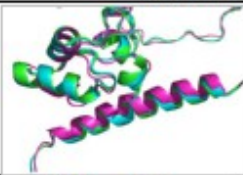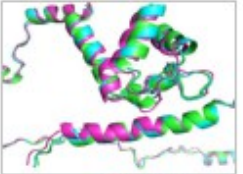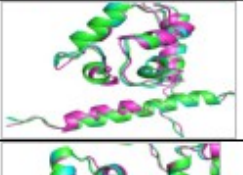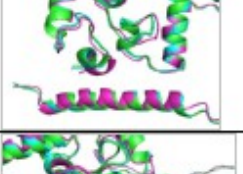|		
Residues in most favoured regions [A,B,L]	361	77.1%
Residues in additional allowed regions [a,b,l,p]	91	19.4%
Residues in generously allowed regions [~a,~b,~l,~p]	13	2.8%
Residues in disallowed regions	3	0.6%
Number of non-glycine and non-proline residues	468	100.0%
Number of end-residues (excl. Gly and Pro)	1	
Number of glycine residues (shown as triangles)	28	
Number of proline residues	10	
Total number of residues	507	

14

Plot statistics

Residues in most favoured regions [A,B,L]	353	75.4%
Residues in additional allowed regions [a,b,l,p]	97	20.7%
Residues in generously allowed regions [~a,~b,~l,~p]	10	2.1%
Residues in disallowed regions	8	1.7%
Number of non-glycine and non-proline residues	468	100.0%
Number of end-residues (excl. Gly and Pro)	1	
Number of glycine residues (shown as triangles)	28	
Number of proline residues	10	
Total number of residues	507	

19

Plot statistics

Residues in most favoured regions [A,B,L]	359	76.7%
Residues in additional allowed regions [a,b,l,p]	99	21.2%
Residues in generously allowed regions [~a,~b,~l,~p]	9	1.9%
Residues in disallowed regions	1	0.2%
Number of non-glycine and non-proline residues	468	100.0%
Number of end-residues (excl. Gly and Pro)	1	
Number of glycine residues (shown as triangles)	28	
Number of proline residues	10	
Total number of residues	507	

24

Plot statistics

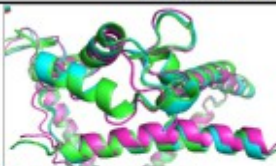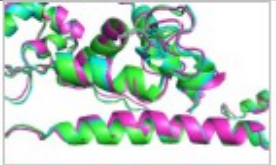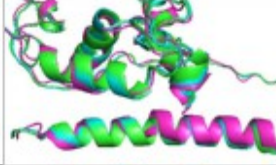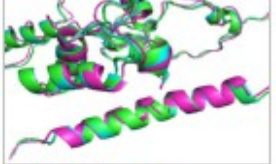|		
Residues in most favoured regions [A,B,L]	354	75.6%
Residues in additional allowed regions [a,b,l,p]	98	20.9%
Residues in generously allowed regions [~a,~b,~l,~p]	12	2.6%
Residues in disallowed regions	4	0.9%
Number of non-glycine and non-proline residues	468	100.0%
Number of end-residues (excl. Gly and Pro)	1	
Number of glycine residues (shown as triangles)	28	
Number of proline residues	10	
Total number of residues	507	

25

Plot statistics

Residues in most favoured regions [A,B,L]	360	76.9%
Residues in additional allowed regions [a,b,l,p]	90	19.2%
Residues in generously allowed regions [~a,~b,~l,~p]	16	3.4%
Residues in disallowed regions	2	0.4%
Number of non-glycine and non-proline residues	468	100.0%
Number of end-residues (excl. Gly and Pro)	1	
Number of glycine residues (shown as triangles)	28	
Number of proline residues	10	
Total number of residues	507	

30

Plot statistics

Residues in most favoured regions [A,B,L]	354	75.6%
Residues in additional allowed regions [a,b,l,p]	100	21.4%
Residues in generously allowed regions [~a,~b,~l,~p]	10	2.1%
Residues in disallowed regions	4	0.9%
Number of non-glycine and non-proline residues	468	100.0%
Number of end-residues (excl. Gly and Pro)	1	
Number of glycine residues (shown as triangles)	28	
Number of proline residues	10	
Total number of residues	507	

Figure 4.2.7a Ramachandran statistics for models post-refinement.

The various values following the Ramachandran analysis of the structures following energy minimisation of SD followed by CG methods. The ramachandran plots were calculated using PROCHECK.

31

Plot statistics

Residues in most favoured regions [A,B,L]	351	75.0%
Residues in additional allowed regions [a,b,l,p]	98	20.9%
Residues in generously allowed regions [~a,~b,~l,~p]	16	3.4%
Residues in disallowed regions	3	0.6%
Number of non-glycine and non-proline residues	468	100.0%
Number of end-residues (excl. Gly and Pro)	1	
Number of glycine residues (shown as triangles)	28	
Number of proline residues	10	
Total number of residues	507	

39

Plot statistics

Residues in most favoured regions [A,B,L]	335	71.6%
Residues in additional allowed regions [a,b,l,p]	121	25.9%
Residues in generously allowed regions [~a,~b,~l,~p]	7	1.5%
Residues in disallowed regions	5	1.1%
Number of non-glycine and non-proline residues	468	100.0%
Number of end-residues (excl. Gly and Pro)	1	
Number of glycine residues (shown as triangles)	28	
Number of proline residues	10	
Total number of residues	507	

42

Plot statistics

Residues in most favoured regions [A,B,L]	375	80.1%
Residues in additional allowed regions [a,b,l,p]	77	16.5%
Residues in generously allowed regions [~a,~b,~l,~p]	12	2.6%
Residues in disallowed regions	4	0.9%
Number of non-glycine and non-proline residues	468	100.0%
Number of end-residues (excl. Gly and Pro)	1	
Number of glycine residues (shown as triangles)	28	
Number of proline residues	10	
Total number of residues	507	

43

Plot statistics

Residues in most favoured regions [A,B,L]	371	79.3%
Residues in additional allowed regions [a,b,l,p]	86	18.4%
Residues in generously allowed regions [~a,~b,~l,~p]	11	2.4%
Residues in disallowed regions	0	0.0%
Number of non-glycine and non-proline residues	468	100.0%
Number of end-residues (excl. Gly and Pro)	1	
Number of glycine residues (shown as triangles)	28	
Number of proline residues	10	
Total number of residues	507	

46

Plot statistics

Residues in most favoured regions [A,B,L]	358	76.5%
Residues in additional allowed regions [a,b,l,p]	99	21.2%
Residues in generously allowed regions [~a,~b,~l,~p]	10	2.1%
Residues in disallowed regions	1	0.2%
Number of non-glycine and non-proline residues	468	100.0%
Number of end-residues (excl. Gly and Pro)	1	
Number of glycine residues (shown as triangles)	28	
Number of proline residues	10	
Total number of residues	507	

47

Plot statistics

Residues in most favoured regions [A,B,L]	335	71.6%
Residues in additional allowed regions [a,b,l,p]	113	24.1%
Residues in generously allowed regions [~a,~b,~l,~p]	17	3.6%
Residues in disallowed regions	3	0.6%
Number of non-glycine and non-proline residues	468	100.0%
Number of end-residues (excl. Gly and Pro)	1	
Number of glycine residues (shown as triangles)	28	
Number of proline residues	10	
Total number of residues	507	

48

Plot statistics

Residues in most favoured regions [A,B,L]	339	72.4%
Residues in additional allowed regions [a,b,l,p]	118	25.2%
Residues in generously allowed regions [~a,~b,~l,~p]	7	1.5%
Residues in disallowed regions	4	0.9%
Number of non-glycine and non-proline residues	468	100.0%
Number of end-residues (excl. Gly and Pro)	1	
Number of glycine residues (shown as triangles)	28	
Number of proline residues	10	
Total number of residues	507	

49

Plot statistics

Residues in most favoured regions [A,B,L]	393	84.0%
Residues in additional allowed regions [a,b,l,p]	66	14.1%
Residues in generously allowed regions [~a,~b,~l,~p]	6	1.3%
Residues in disallowed regions	3	0.6%
Number of non-glycine and non-proline residues	468	100.0%
Number of end-residues (excl. Gly and Pro)	1	
Number of glycine residues (shown as triangles)	28	
Number of proline residues	10	
Total number of residues	507	

50

Plot statistics

Residues in most favoured regions [A,B,L]	368	78.6%
Residues in additional allowed regions [a,b,l,p]	83	17.7%
Residues in generously allowed regions [~a,~b,~l,~p]	8	1.7%
Residues in disallowed regions	9	1.9%
Number of non-glycine and non-proline residues	468	100.0%
Number of end-residues (excl. Gly and Pro)	1	
Number of glycine residues (shown as triangles)	28	
Number of proline residues	10	
Total number of residues	507	

Figure 4.2.7b Ramachandran statistics for models post-refinement.

The various values following the Ramachandran analysis of the structures following energy minimisation of SD followed by CG methods. The ramachandran plots were calculated using PROCHECK.

From evaluation of the results in figure 4.2.7, model 49 showed the highest percentage of residues in most favoured regions compared to the others. Model 2 did not have any residues in the disallowed regions post-refinement. Thus, it was decided to pursue molecular dynamics analysis on these two models.

4.2.2 Molecular Dynamics

4.2.2.1 Analysis of Molecular Dynamics files

Each molecular dynamics run using CHARMM generates an "*.ene" file, which contains all the energy output data.

An example content from an "*.ene" file is as follows.

STEP	TIME	TOTE	TOTK	ENER
EP-K	TEMP	BOND	ANGL	DIHE
MPR	VDW	ELEC	HBON	HARM
DMC	RGY			
1610	1.6100	-2745.0474	5163.5804	-7908.6278
-13072.2082	320.9724	2064.4737	3078.3075	2160.0468
233.8503	-924.3428	-14934.3152	0.0000	0.0000
0.0000	0.0000			
1620	1.6200	-2669.6006	5108.0907	-7777.6914
-12885.7821	317.5231	2172.5351	3090.1985	2179.4730
239.2218	-918.8821	-14955.6961	0.0000	0.0000
0.0000	0.0000			
1630	1.6300	-2601.1064	5196.0770	-7797.1834
-12993.2604	322.9924	2012.9682	3193.3932	2160.6832
245.6919	-937.4140	-14907.3020	0.0000	0.0000
0.0000	0.0000			
1640	1.6400	-2534.1869	5272.6931	-7806.8800
-13079.5731	327.7549	2048.3893	138.3341	2152.4378
259.3095	-939.0818	-14892.7498	0.0000	0.0000
0.0000	0.0000			
1650	1.6500	-2464.9783	5284.8835	-7749.8618
-13034.7453	328.5127	2017.4168	3210.9169	2166.2347
239.4728	-929.3629	-14883.1323	0.0000	0.0000
0.0000	0.0000			

Figure 4.2.8   Example of the contents of an "*.ene" file following a CHARMM simulation.

The objective was to plot the time-steps or the time (in picoseconds) against the energies. To preserve the original file, in case of any error in data processing, the original file is copied and renamed. Subsequently, the new file is edited and the lines coloured in red (as shown above) are deleted. In their place, the following three lines (in green) are substituted in the file as shown below.

```
replacement_dummy_line_1
replacement_dummy_line_2
replacement_dummy_line_3

1610            1.6100          -2745.0474      5163.5804       -7908.6278
-13072.2082     320.9724        2064.4737       3078.3075       2160.0468
233.8503        -924.3428       -14934.3152     0.0000          0.0000
0.0000          0.0000

1620            1.6200          -2669.6006      5108.0907       -7777.6914
-12885.7821     317.5231        2172.5351       3090.1985       2179.4730
239.2218        -918.8821       -14955.6961     0.0000          0.0000
0.0000          0.0000
```

Figure 4.2.9   Example of the "*.ene" file after editing in preparation for data extraction.

The starting few header lines are deleted in the original file and replaced with three "dummy" lines, which will serve as null-strings in subsequent data extraction steps.

The newly edited file is then saved under a different filename, such as "new_ene_file.ene".

In order to extract the kinetic energy values from this huge data file, without having to manually scan through the entire file, the following command was typed at the Linux prompt.

awk '{print $1, $4;}' new_ene_file.ene > 2column_energyfile_steps_vs_kineticE.txt

This command grabs the 1st and the 4th column from the file, and then places it in a new file (2column_energyfile_steps_vs_kineticE.txt) containing two columns of data.

However, as the data needed is contained in every 4th line of these columns, the other lines contain data that pertains to other parameters. Thus, in order to grab every 4th line and output this into a new file, the following command was invoked at the Linux prompt.

awk '0 == NR % 4' 2column_energyfile_steps_vs_kineticE.txt > every4th_line.txt

This newly generated file ("every4th_line.txt) contains data that can be plotted in a standard x,y coordinate system, using any graphical program. Gnuplot was used to generate the graphs for this section of the project.

4.2.2.2 Molecular Dynamics – Heating Step

The heating step of the molecular dynamics run served to "mobilise" the molecule, so relieving any residues that are tethered to one another by virtue of van der walls forces, protein-protein interaction or ionic charges. The heating step in a molecular dynamics script is typically written as follows:

```
OPEN WRIT UNIT 31 CARD NAME "cmbd_heat.rst"
OPEN WRIT UNIT 32 FILE NAME "cmbd_heat.dcd"
OPEN WRIT UNIT 33 FILE NAME "cmbd_heat.vel"
OPEN WRIT UNIT 34 CARD NAME "cmbd_heat.ene"

DYNAmics VERL STRT -
  TIME 0.001000 NSTEP 1000 -
  FIRSTT 240.000000 FINALT 800.000000 TEMINC 5.000000 -
  ISEED 314159 -
  IHTFRQ 10 IEQFRQ 0 IASORS 1 IASVEL 1 ICHECW 1 -
  INBFRQ -1 CUTNB 180.000000 -
  CTONNB 11.000000 CTOFNB 140.000000 VSWITCH SWITCH -
  CDIE EPS 1.000000 -
  IHBFRQ 0 IMGFRQ -1 -
  ISVFRQ 100 NSAVC 10 NSAVV 10 NPRINT 10 -
  IUNREA -1 IUNWRI 31 IUNCRD 32 IUNVEL 33 KUNIT 34

OPEN WRIT UNIT 47 CARD NAME "cmbd_heat_1000.crd"
WRIT COOR UNIT 47 CARD
CLOSE UNIT 47

OPEN WRIT UNIT 48 CARD NAME "cmbd_heat_1000.psf"
WRIT PSF UNIT 48 CARD
```

Figure 4.2.10 Example script of a CHARMM MD heating step.

In the above script, it can be seen that the time-step-size is fairly small, (0.001). Larger step sizes would allow the simulation to run much faster as there would be less sampling steps. However, larger step sizes also mean that for the "in-between" steps, there would be a certain amount of data that is not recorded and therefore lost. There is also a danger that simulations would crash with larger step sizes, this is primarily when there is a loss of recording of data. If there is a need to reference a particular point during the simulation which points to a time step not present in the data file, then the simulation would crash.

In addition, the heating step is taken from 240 K to 800 K. As discussed previously in section 4.1.2, these values are related to kinetic energy and allows a means to provide the appropriate amount of translational movement to a molecule in computation.
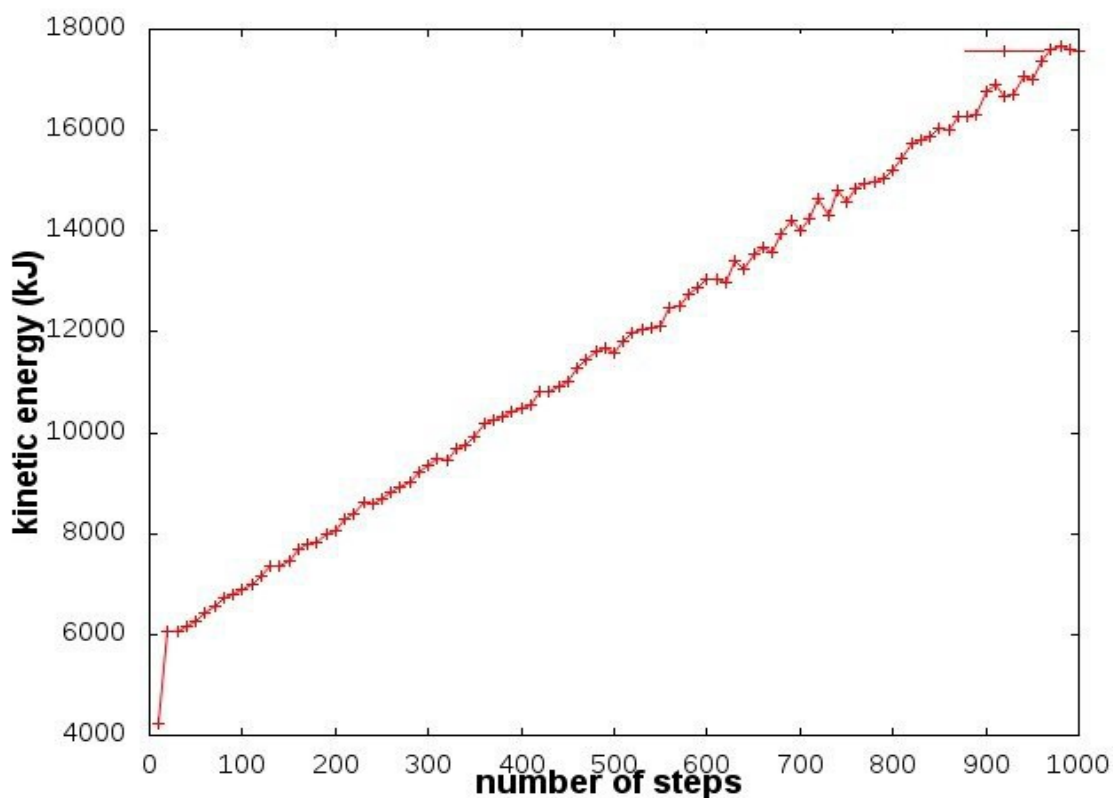
Figure 4.2.11 The heating step during CHARMM MD analysis of model 2.

The linear line shows a gradual rise in kinetic energy with the number of steps.

Since there is no sudden dip in the line, the overall starting structure must have been adequately refined (minimised) to relieve any erroneous strain that may have prevented parts of the molecule from mobilising appropriately.
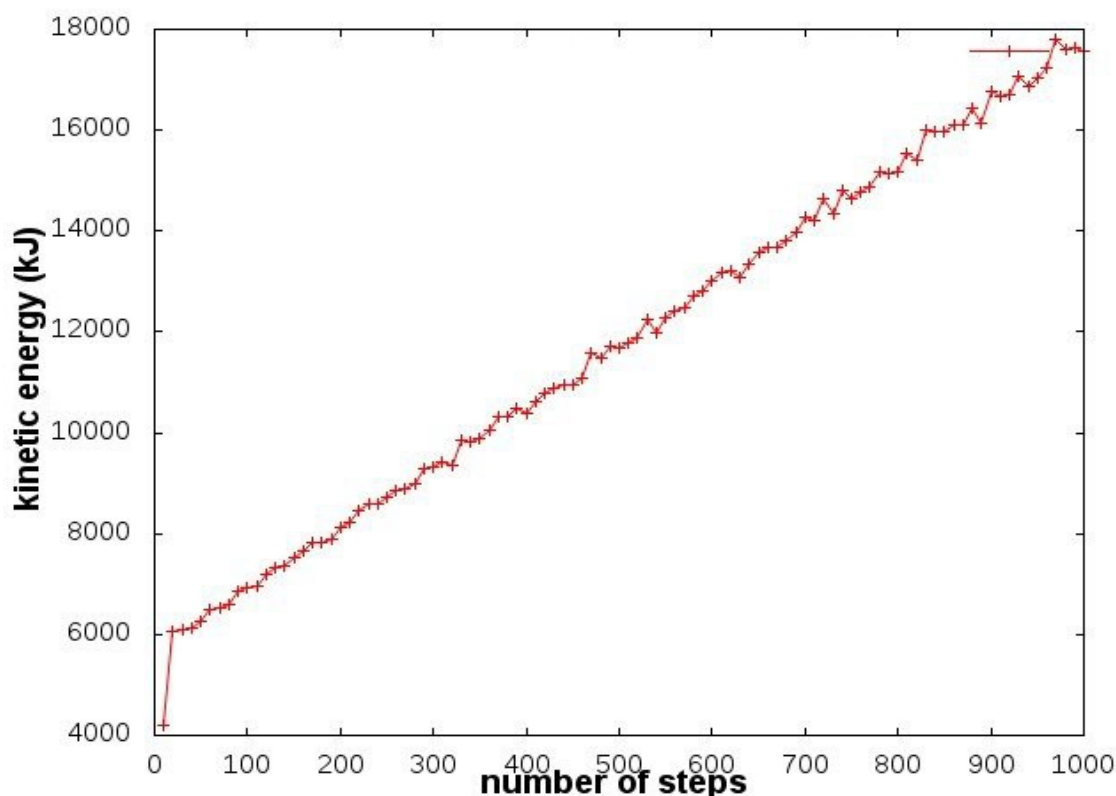
Figure 4.2.12 The heating step during CHARMM MD analysis of model 49.

The linear line shows a gradual rise in kinetic energy with the number of steps.

Both models, 2 and 49, show a stable and gradual heating step. It is difficult to assess what the maximal heating step should be for molecules. This is usually performed on a trial by error method, gradually increasing the maximal heating step in a simulation empirically until a point is reached where the molecule then totally unfolds or disintegrates, or the system crashes. Thus, a slightly lower temperature from that point would be normally used. In this instance, however, a consensus temperature commonly used was applied whilst carrying out the heating step.

4.2.2.3 Molecular Dynamics – Equilibration Step

The subsequent process in the molecular dynamics simulation is the equilibration step. In this process, the molecule is gradually cooled down to a stable temperature, at which it is know that the molecule would be stable in a liquid state. This is usually selected at normal physiological body temperature 310 K (37°C) although any temperature above O K (-273°C) would suffice.

```
! Dynamics equilibration stage
!

OPEN READ UNIT 30 CARD NAME "cmbd_heat.rst"
OPEN WRIT UNIT 31 CARD NAME "cmbd_equil.rst"
OPEN WRIT UNIT 32 FILE NAME "cmbd_equil.dcd"
OPEN WRIT UNIT 33 FILE NAME "cmbd_equil.vel"
OPEN WRIT UNIT 34 CARD NAME "cmbd_equil.ene"

DYNAmics VERL REST -
  TIME 0.001000 NSTEP 30000 ISEED 314159 -
  FIRSTT 800.000000 FINALT 310.000000 -
  TWINDH 10.000000 TWINDL -10.000000 -
  IHTFRQ 0 IEQFRQ 5 IASORS 0 ISCVEL 0 ICHECW 1 -
  INBFRQ -1 CUTNB 180.000000 -
  CTONNB 11.000000 CTOFNB 140.000000 VSWITCH SWITCH -
  CDIE EPS 1.000000 -
  IHBFRQ 0 IMGFRQ -1 -
  ISVFRQ 100 NSAVC 10 NSAVV 10 NPRINT 10 -
  IUNREA 30 IUNWRI 31 IUNCRD 32 IUNVEL 33 KUNIT 34

OPEN WRIT UNIT 47 CARD NAME "cmbd_equil_30000.crd"
WRIT COOR UNIT 47 CARD
CLOSE UNIT 47

OPEN WRIT UNIT 48 CARD NAME "cmbd_equil_30000.psf"
WRIT PSF UNIT 48 CARD
CLOSE UNIT 48

OPEN WRIT UNIT 49 CARD NAME "cmbd_equil_30000.pdb"
WRIT COOR PDB UNIT 49
CLOSE UNIT 49
```

Figure 4.2.13 Example of the CHARMM molecular dynamics equilibration step.

In this script, the molecule is slowly cooled from the final temperature of the previous heating step (800 K) down to 310 K. The idea is to continue the equilibration until the kinetic energy fluctuations of the molecule are minimal, whereby the molecule is at its most stable conformation.

The kinetic energy at each step of this stage of the equilibration is recorded allowing for this data to be plotted as seen in figures 4.2.14 (model 2) and 4.2.15 (model 49). A protein structure file (.psf) and a trajectory file (.dcd) of the changes of spatial positioning of each amino acid are also generated. When these two files are both loaded into suitable software, VMD (Humphrey et al. 1996) was used, the movement of the model as it undergoes conformational changes during equilibration can be viewed as a film. This allows for clear observation of movement within the structure including the reduction in movement, as the kinetic energy decreases, as a stable conformation is

adopted. Figures 4.2.16-17 (model 2) and 4.2.18-20 (model 49) show images captured at points selected based on the kinetic energy from the graphs below.
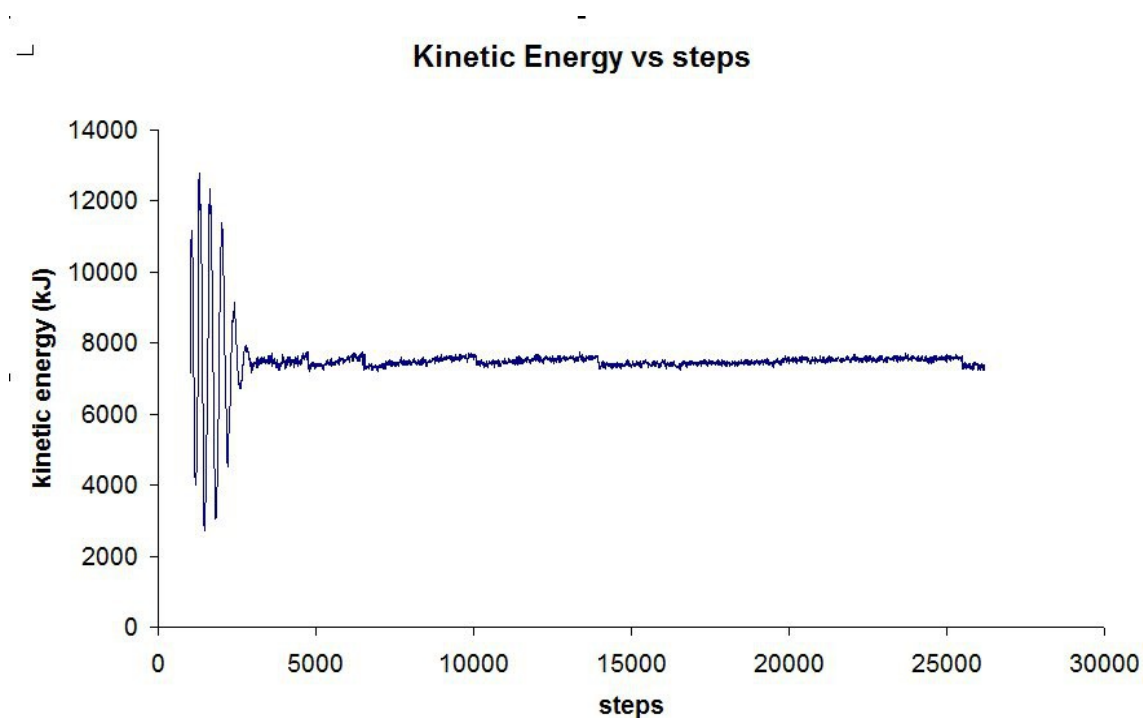
Figure 4.2.14 The CHARMM equilibration step of model 2.

Wide fluctuations in kinetic energy are seen below 4500 steps, above which, the system starts to equilibrate, albeit minor fluctuations. These fluctuations continue to diminish approximately at 15000 steps.

Kinetic Energy vs. # of steps
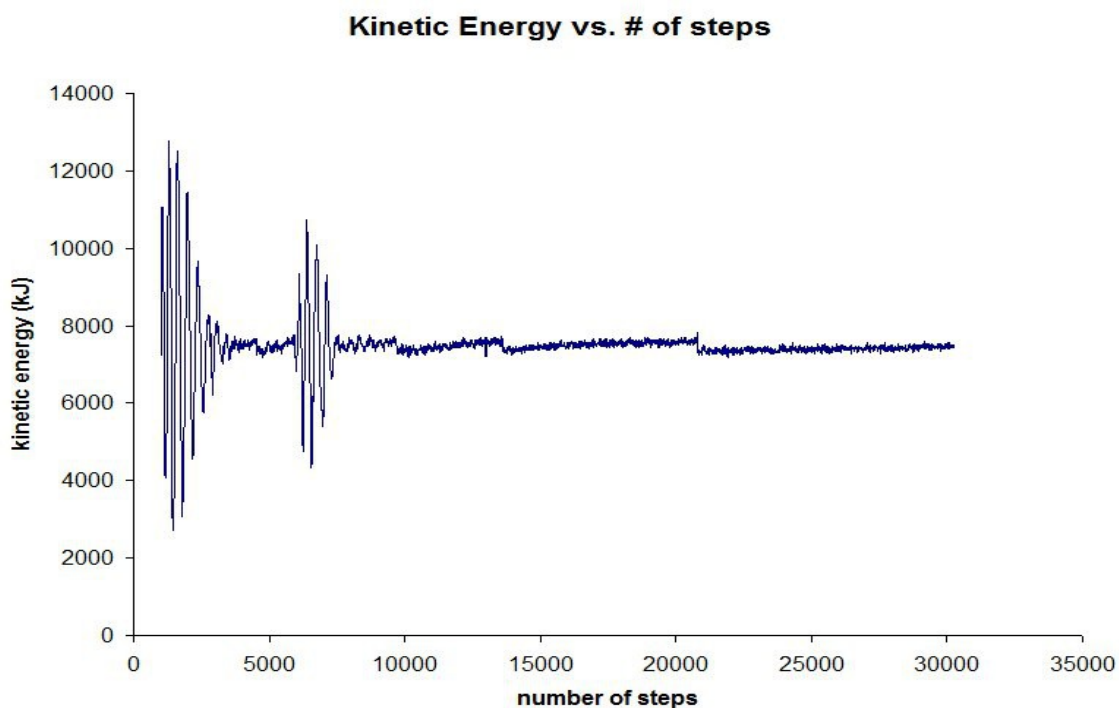
Figure 4.2.15 The CHARMM equilibration step of model 49.

Two large groups of wide fluctuations in kinetic energy are seen below 7000 steps, above which, the system starts to equilibrate, albeit minor fluctuations.

The following images are aligned, for both models, so that CaMLD in close proximity to CaMBD is shown in the lower centre of the images. Also images at the different step stages are taken from the same viewpoint allowing for easier observation of the change in conformation.
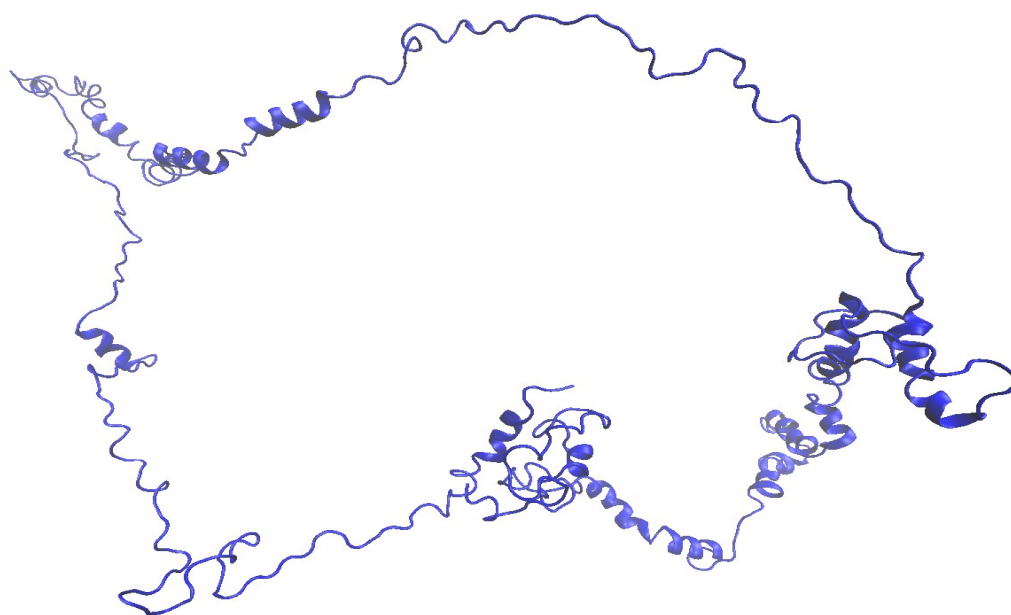
Figure 4.2.16 Structure of model 2 during equilibration at high kinetic energy fluctuation point (step 2500).
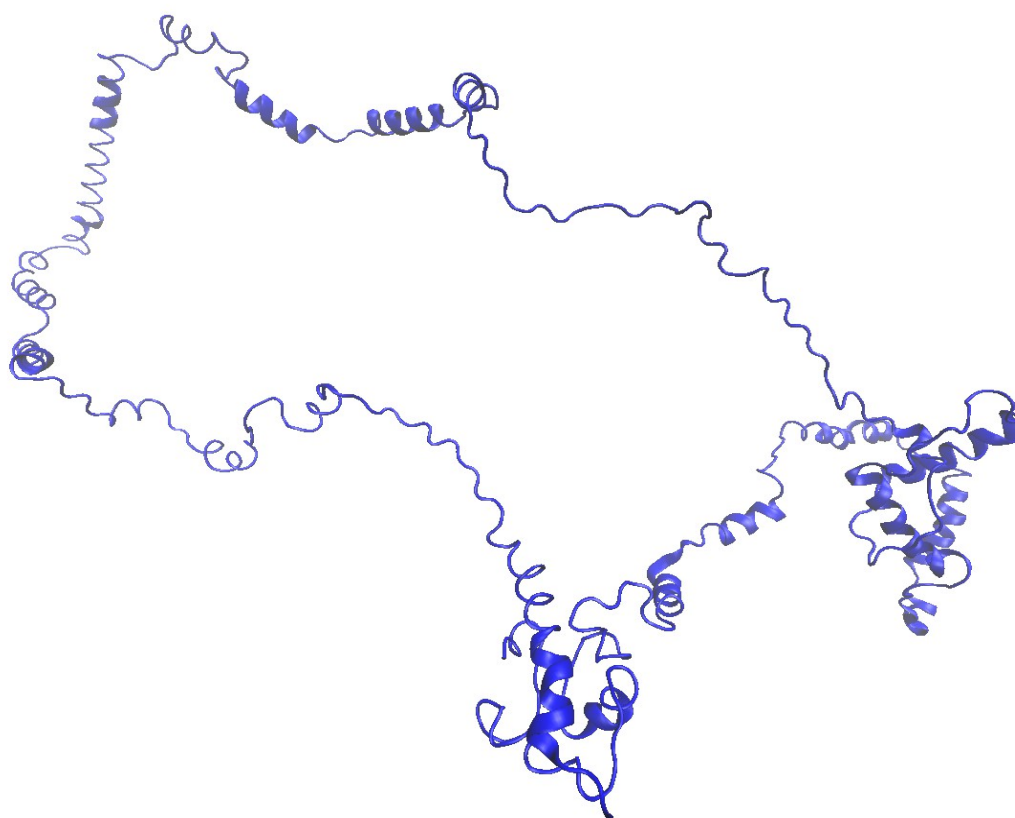
Figure 4.2.17 Structure of model 2 at end of equilibration steps.

144
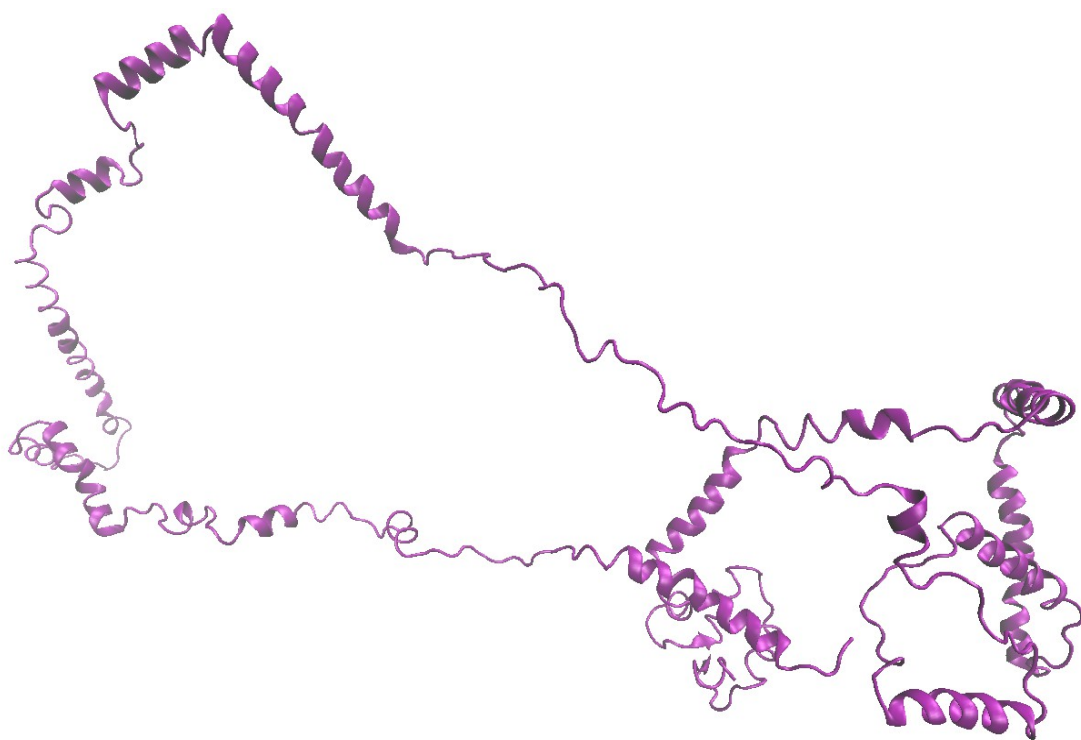
Figure 4.2.18 Structure of model 49 during equilibration at first high kinetic energy fluctuation point (step 2000).
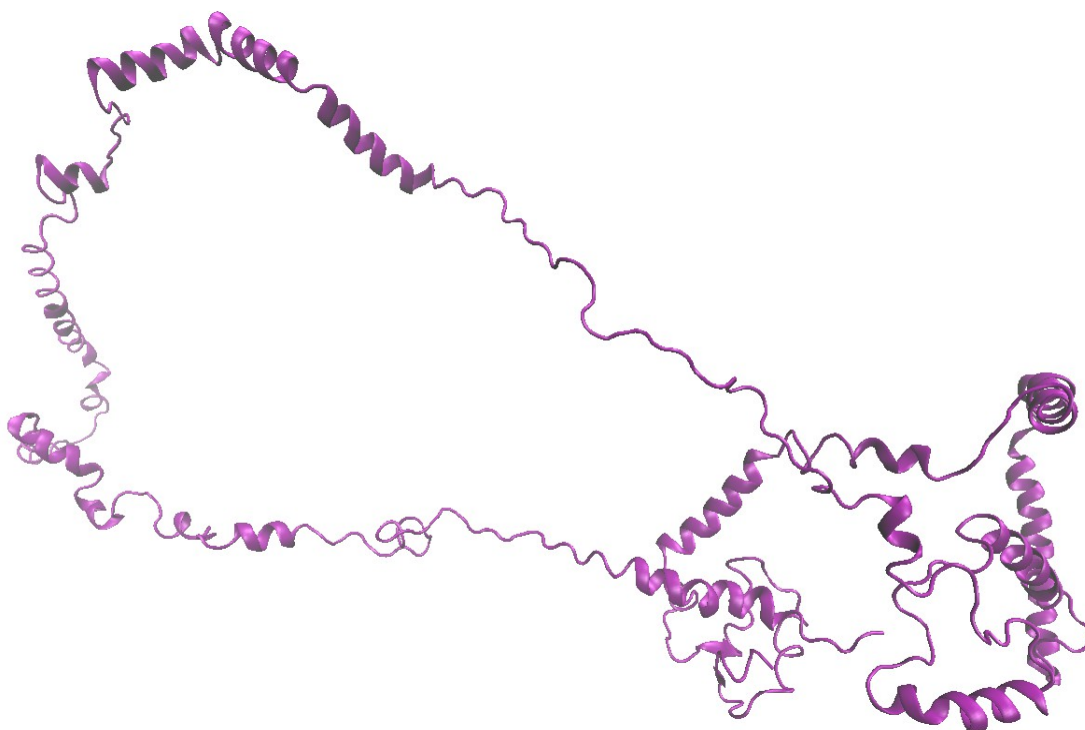
Figure 4.2.19 Structure of model 49 during equilibration at second high kinetic energy fluctuation point (step 6500).

145
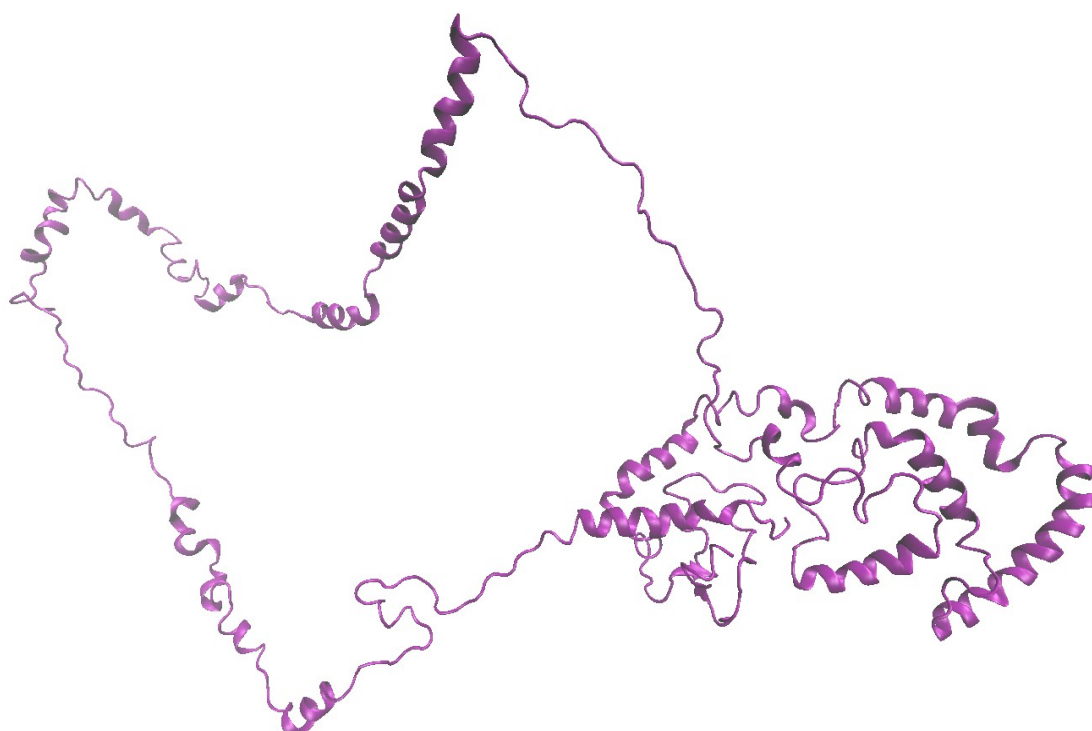
Figure 4.2.20 Structure of model 49 at end of equilibration steps.

In the above structure images the movement of the regions towards the stable conformation adopted at the end of the equilibration stage can be observed. This is clearer for model 49 were two snapshots have been taken during the two zones of kinetic energy fluctuation allowing for clearer comparison with the stable conformation. For both models CaMLD stays in close proximity with CaMBD, so not greatly differing from the input structure for these parts of the models. The majority of movement is observed in the loop between the two RyR2 domains.

4.3 Discussion

The aim of this chapter was to improve the models generated at the end of chapter 3 so they could offer greater potential as a tool for the study of human RyR2 between amino acid residues 3579-4085. Within this area of RyR2, two domains with known structures had been identified, CaMBD and CaMLD. The PDB entry 2BCX served as a template for both domains as it contained the structure of two equivalent proteins interacting. At the end of chapter 3 models had been generated containing CaMBD and CaMLD which

had been positioned relative to each other based on distances obtained from 2BCX.

However, the remainder of the loop region between these two domains only had a basic overall structure including predicted secondary structure elements. As discussed in the introduction of this chapter, the tertiary structure of a domain has significant influence on its function. Therefore, to improve the quality of the models it was decided to simulate the folding of the model structure into a three dimensional state.

To enable the *in silico* theoretical folding of the protein a number of processes had to be performed. An additional round of energy minimisation was conducted on the models from chapter 3. This was to confirm that these models were in the lowest achievable energy state prior to the simulated folding. Ramachandran analysis post energy minimisation revealed that model 2 had no residues present in disallowed regions and that model 49 had the highest percentage of residues within the most favoured regions. Thus, these two models were selected for further improvement.

As discussed in the introduction there are various theories postulated about how the amino acid translates into a tertiary folded protein structure. Within current computing limitations the best simulation of a folding process is through a molecular dynamics approach. This is achieved by a theoretical heating and then cooling phase. The relationship between temperature and velocity (kinetic energy) within molecular dynamics has been discussed. Basically the process involves increasing the energy available (heating) to the molecule which allows movement within the structure. The energy level is gradually reduced (cooling) from this high energy state so that the structure adopts an energetically stable conformation. As mentioned, this process does not contain the complexity or finesse of nature and, as a result, the final conformation obtained represents a probable likelihood of the structure of the modelled region.

Analysing the equilibration phase revealed that the majority of movement within the structures to achieve the stable conformation occurred in the loop region between the two identified domains. The CaMLD and CaMBD domains remained closely together as in the input model structure. These two domains were already folded as based on their respective template structures. The two domains were placed at distances observed in the structure of PDB entry 2BCX. Thus, as the folding and positioning of these two

domains was based on known structure, i.e. the result of protein that had undergone folding in nature, these parts of the model structure were expected to be least affected by the simulated folding. However, the region between the two domains was modelled *ab intio* as, unlike the two domains, no template was available. Thus, this region contained less structural information so was likely to be most influenced by the simulated folding process.

The difference in final structures obtained for the two models reflects the limitations of the process. The conformation at the end of the equilibration phase is a product of the molecular dynamics process upon the input model structure. As stated previously, this process is a best possible simulation, and as the input models varied, it is to be expected that variation in the final conformations would be observed. However, the process of emulating some of the aspects of tertiary structure through simulated folding of the models has increased their structural quality. As a result, the products of this process are better tools with which to study this region of RyR2.

It is assumed that some of the disease mutations within the human RyR2 region 3579-4082 would disrupt the structure so impacting on the predicted regulatory role of this region within the calcium channel. It is proposed that the disruption to structure would be reflected in the energy states observed in relevant simulated models. As shown by the generation of the two models of wild type RyR2, the tools exist to simulate this region and monitor the energy state of the structure during this process. Further information about the structural impact of mutations in this region can be gained by the computational simulation and manipulation of models containing studied mutations.

Chapter 5:

Cloning and expression of the recombinant RyR-fragments and Calmodulin.

5.1 Introduction

As discussed in chapter 1 the interaction between various domains within the ryanodine receptor (RyR) are known to play a crucial role in the regulation of the channel acting as a calcium release store. As further elaborated in chapter 3 the calmodulin like domain (CaMLD) and CaM binding domain (CaMBD) within RyR are indicated to be involved in one of these key interactions. The binding of these two domains along with the association of CaM is believed to play a role in the response of the channel to increasing calcium concentration leading to channel opening. The presence of some of the known catecholaminergic polymorphic ventricular tachycardia (CPVT) mutations within CaMLD is postulated to disrupt the mechanism of the interaction of the two RyR domains and hence affect channel opening leading to arrhythmia.

As the binding of these two domains and the involvement of CaM with this interaction was targeted for this study so constructs of these regions were required. Chapters 3 and 4 discuss the computer based modelling of these regions. While chapter 4 describes simulating a much larger region of RyR linking the two domains of interest for the purposes of generating recombinant protein a similar approach to the smaller modelling of the domains as followed in chapter 3 was adopted. A number of the initial studies of these regions, as described in the literature, were conducted in RyR1. This study identified equivalent regions in RyR2 for the basis of construct design, subsequent additions to the literature regarding the CaMLD region in RyR2 have followed a similar approach. The specifics of the construct domain boundaries are discussed further in the relevant sections below.

The RIH associated domain was identified as a series of amino acid repeat sequences present in IP3Rs and RyR that it was proposed may have possible protein binding or related function (Ponting 2000). Studies conducted by colleagues indicate that residues 3722-4353 of the I domain of RyR2 are sufficient for interdomain interaction (George et al. 2004a). It is possible that the RIH associated domain may be involved with this process. While this may separate to CaMLD function, the RIH associated domain may play a role in supporting the binding of CaMLD with CaMBD. It is because of this possibility that the RIH associated domain has been identified as a region for potential

further study.

The aim of this section of the project was to produce the required recombinant proteins, including CaMLD constructs containing known CPVT mutations, required for further experimentation to study the domain interactions. To enable this to occur so structurally stable soluble recombinant protein has to be produced; the construct design phase is a key stage in this process. A variety of factors as discussed in chapter 1 have to be considered. With consideration to the regions of RyR being studied in this project the identification of stable "pieces" that can be independently lifted from the whole molecule is required. Commonly complete domains and active regions are targeted, in this case potential sections have already been identified in the literature. However to enable the production of recombinant protein to the quantities and quality required for structural biology studies then further work is usually needed to determine suitable domain boundaries, and a number constructs and expression variables need to be tested.

5.1.1 Design of CaMLD constructs

As stated above RyR2 based CaMLD constructs have been reported in the literature. This study used the RyR1 based Xiong, 2006 (Liangwen Xiong et al. 2006) paper as a basis for identifying domain boundaries for RyR2 CaMLD constructs. The basis of most CaMLD constructs are designed around the two predicted EF hand motif regions contained within the earlier section of the interacting (I) domain.
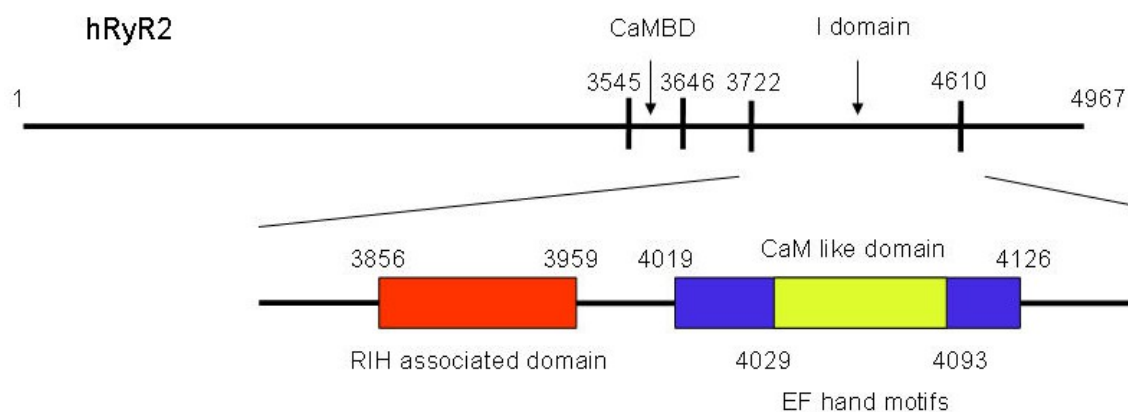
Figure 5.1.1   Map of predicted regions in earlier section of I domain including CaMLD.

Amino acid coordinates are given for the relevant regions. RIH associated domain determined from Ponting, 2000 (Ponting 2000). CaMLD coordinates in RyR2 derived from Xiong, 2006 (Liangwen Xiong et al. 2006). Two EF hand motifs contained within CaMLD from Pfam predictive software. Location of CaMBD relative to CaMLD within RyR2 also shown.

As stated the initial coordinates for the domain boundaries for CaMLD constructs were derived from the Xiong, 2006 paper (Liangwen Xiong et al. 2006). This paper reported interaction between a CaM binding motif and Ca^{2+}-binding domain, 4064-4210, in RyR1. The sequence homology of the relevant amino acid sequences between RyR1 and 2 were compared.

```
1  ---------------ELLKELLDLQKDMVVMLLSLLEGNVVNGMIARQMV
2  FAHMQMKLSQDSSQIELLKELMDLQKDMVVMLLSMLEGNVVNGTIGKQMV
                   ******:*************:********* *.:***

1  DMLVESSSNVEMILKFFDMFLKLKDIVGSEAFQDYVTDPRGLISKKDFQK
2  DMLVESSNNVEMILKFFDMFLKLKDLTSSDTFKEYDPDGKGVISKRDFHK
   *******.*******************:..*::*::* .* :*:***:**:*

1  AMDSQKQFSGPEIQFLLSCSEADENEMINCEEFANRFQEPARDIGFNVAV
2  AMESHKHYTQSETEFLLSCAETDENETLDYEEFVKRFHEPAKDIGFNVAV
   **:*:*::: .* :******:*:**** :: ***.:**:***:********

1  LLTNLSEHVPHDPRL--------------------------------H
2  LLTNLSEHMPNDTRLQTFLELAESVLNYFQPFLGRIEIMGSAKRIERVYF
   ********:*:*.**                                  .
```

Figure 5.1.2 Amino acid homology sequence comparison between reported Ca²⁺-binding domain in RyR1 and equivalent region in RyR2.

Sequence 1 is 4064-4210 section of RyR1. Sequence 2 is 3950-4150 region of RyR2

As can be seen in figure 5.1.2 the potential matching region of RyR2 to the reported RyR1 sequence is slightly longer than the given coordinates in figure 5.1.1. This is because these residue numbers were determined by further analysis of the sequence in the surrounding region of RyR2. Initially the exon boundaries of the nucleotide sequence were mapped to the amino acid sequence to try to identify complete coding sequences within RyR2.

```
          1         2         3         4         5         6
1         0         0         0         0         0         0
          GEEEVKSFEEKEMEKQKLLYQQARLHDRGAAEMVLQTISASKGETGPMVAATLKLGIAIL  3766

          NGGNSTVQQKMLDYLKEKKDVGFFQSLAGLMQSCSVLDLNAFERQNKAEGLGMVTEEGSG  3826
          EKVLQDDEFTCDLFRFLQLLCEGHNSDFQNYLRTQTGNNTTVNIIISTVDYLLRVQESIS  3886
RIH       DFYWYYSGKDVIDEQGQRNFSKAIQVAKQVFNTLTEYIQGPCTGNQQSLAHSRLWDAVVG  3946
domain
          FLHVFAHMQMKLSQDSSQIELLKELMDLQKDMVVMLLSMLEGNVVNGTIGKQMVDMLVES  4006
                                        EF-hand
          SNNVEMILKFFDMFLKLKDLTSSDTFKEYDPDGKGVISKRDFHKAMESHKHYTQSETEFL  4066
CaMLD          EF-hand
domain    LSCAETDENETLDYEEFVKRFHEPAKDIGFNVAVLLTNLSEHMPNDTRLQTFLELAESVL  4186
          NYFQPFLGRIEIMGSAKRIERVYFEISESSRTQWEKPQVKESKRQFIFDVVNEGGEKEKM  4246
          ELFVNFCEDTIFEMQLAAQISESDLNERSANKEESEKERPEEQGPRMAFFSILTVRSALF  4306
          ALRYNILTLMRMLSLKKSLKKQMKKVKKMTVKDMVTAFFSSYWSIFMTLLHFVASVFRGFF  4366
          RIICSLLLGGSLVEGAKKIKVAELLANMPDPTQDEVRGDGEEGERKPLEAALPSEDLTDL  4426
          KELTEESDLLSDIFGLDLKREGGQYKLIPHNPNAGLSDLMSNPVPMPEVQEKFQEQKAKE  4486
          EEKEEKEETKSEPEKAEGEDGEKE                                       4510
```

Figure 5.1.3 Exon boundaries mapped onto amino acid sequence of earlier section of I domain of RyR2

Amino acid sequence number of RyR2 is shown to right of sequence. The exon boundaries are illustrated by a vertical separator (|). The Regions mapped in figure 5.1.1 including coordinates: 3826-3959 (RIH associated domain – boxed in red); 4019-4126 (CaMLD -boxed in blue) and the EF hand motifs lie between 4027-4055 and 4062-4090 (highlighted in ochre). Selected CPVT mutations shown in bold.

Based on the analysis of the exon boundaries 4 potential construct start locations were identified, 3 prior to the start of the RIH associated domain and 1 before CaMLD. However using the end point of the exon encoding CaMLD would result in an impractically long recombinant protein with increased technical difficulties for expression and downstream experiments. To determine potential C termini of potential constructs the predicted secondary structure of the I domain was examined.

```
I domain      AMESHKHYTQSETEFLLSCAETDENETLDYEEFVKRFHEPAKDIGFNVAVLLTNLSEHMPNDTRLQTFLE
DSC           hhhhhhccccchhhhhhhhcccccccccchhhhhhhhhcccchhcccceeeeeccccccccccccchhhhhhh
MLRC          hhhhhccccchhhheeehcccccccccchhhhhhhhhhhhhchchcccchhheehhhhhcccccchhhhhhhh
PHD           hhhhhhccchhhhhhhhhhhhhcccccccchhhhhhhhhhcchhhhcchhhhhhhhhhhhhcccchhhhhhhh
Sec.Cons.     hhhhhhccchhhhhhhhhhhcccccccccchhhhhhhhhcchhhccchhheehhhhhcccccchhhhhhhh

I domain      LAESVLNYFQPFLGRIEIMGSAKRIERVYFEISESSRTQWEKPQVKESKRQFIFDVVNEGGEKEKMELFV
DSC           hhhhhhhhhcccccceeeeccchhhhhhhhheeeccccchhhhhhhccccchhhhhhheeeecccccccccceeeee
MLRC          hhhhhhhhhhhhhhchhhhhccchhhhheeeehcccccccccccgccchhhheeeeeeecccccchhhhhhhh
PHD           hhhhhhhhhhhhhccceeeeeccchhhhhhhhchhhhhhhhhccgchhhhhhhhhhhhhhcchhhhhhhhh
Sec.Cons.     hhhhhhhhhhhhhcceeee?ccchhhhhhheeehccchhhhhccgcccchhhhhheeee?cccchhhhhhh

I domain      NFCEDTIFEMQLAAQISESDLNERSANKEESEKERPEEQGPRMAFFSILTVRSALFALRYNILTLMRMLS
DSC           ccccchhhhhhhhhhhhccgchhhhhhhhhhhhhhhccccccccccccccceeeehhhhhhhhhhhhcchhhhhhhhhh
MLRC          hhchhhhhhhhhhhhhhccgchhhhccchhhhhhhhhhhhcccchhhhhhhhhhhhhhhhhhhhhhhhhhhh
PHD           hhhhhhhhhhhhhhhhhhccgccccccccccccccccchhhccccccchhhhhhhhhhhhhhhhhhhhhhhhh
Sec.Cons.     hhchhhhhhhhhhhhhhccgchhhhccchhhhhcchhhccccccc?hhhhhhhhhhhhhhhhhhhhhhhhh
```

Figure 5.1.4   Predicted secondary structure of I domain region downstream of CaMLD.

Top line is amino acid sequence of RyR2 including and immediately downstream of CaMLD, end section of CaMLD sequence highlighted in turquoise. Lowest line is consensus sequence of predicted secondary structure generated from 3 algorithms. Location of potential C termini residue selections highlighted in green.

A number of secondary structure prediction algorithms were used to generate a consensus sequence so increasing the likelihood that the selected amino acid had the predicted secondary structure feature. As discussed above candidate residues were searched for in regions of random coil that were clearly downstream of helix regions that may be associated with the Ca^{2+} interaction region. The role of supporting helical regions around EF hands to aid binding to target sites is discussed further in chapter 1. The first identified C terminus site, amino acid 4163, also met the additional criterion of being a proline residue. A second site, 4228, was also selected to test if better quality recombinant protein required a longer sequence.

Based on the coordinates derived from the analysis of the exon boundaries and predicted secondary structure so 8 potential constructs were identified. Within these; 2 were identified as potentially preferred constructs. Firstly, the region encompassing the matching Ca^{2+}-binding domain of RyR1, construct 7 below, this construct was later referred to as the Ca^{2+} interacting region (CIR). Secondly, the construct containing the

RIH associated domain in addition to the Ca^{2+}-binding domain, construct 5 below, later referred to as CaMLD.
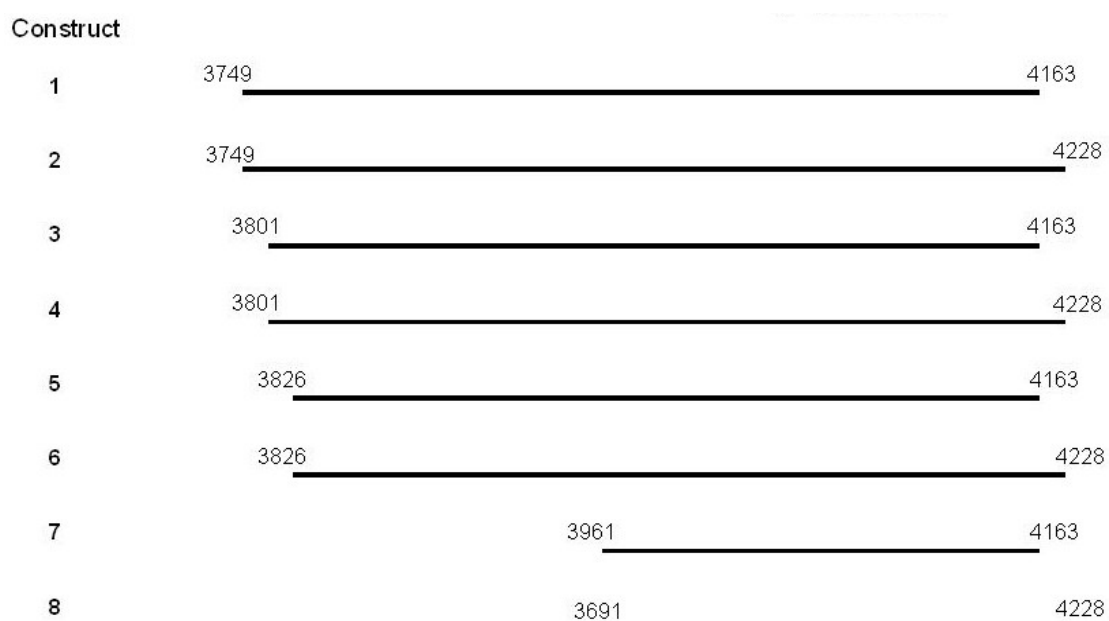
Figure 5.1.5   Coordinates of 8 potentially identified constructs within the earlier section of the I domain.

The locations of known CPVT mutations within these 8 potential constructs was identified with a view to generating mutant forms of the constructs.

```
L3778F
C3800F
G3946S    -in RIH associated domain
F4020L    -in Ca²⁺ interaction region
E4076K    -in 2nd EF hand motif
N4097S    -in Ca²⁺ interaction region
N4104K/I  -in Ca²⁺ interaction region
H4108N/Q  -in Ca²⁺ interaction region
S4124T
E4146K
T4158P
Q4201R
```

Figure 5.1.6   Known CPVT mutations contained within region of 8 potential constructs.

Location of mutation when present within identified domain of early I domain is given.

Within these mutations, 4 were selected of particular interest for further investigation. G3496S owing to its position within the RIH associated domain. E4076K because of its presence within the predicted EF hand motifs and also because it is a charge swap mutation with the loss of the glutamic acid in this location. The potential importance of the change of a negative to positive charge within an EF hand is discussed in further length in chapter 1. N4104K is located in a surrounding α helix adjacent to the EF hand motif believed to be important in supporting binding to CaMBD. Also this mutation results in the gain of a positive charge from neutral with the change to a lysine residue. Although H4108Q is not included in the predicted α helical regions downstream of the EF hand motifs it may have some overlap as seen in the modelling conducted in chapter 3. Also this mutation results in the removal of a positive charge introducing a neutral glutamine.

5.1.2 Design of CaMBD constructs

There are PDB entries containing CaMBD structures, 2bcx is cited in chapter 3 of this thesis, these involve a short peptide with an α helical tube like structure. For the purposes of this project it was decided to make a longer recombinant protein based on human RyR2. Constructs were designed around the N and C terminal flanking regions of RyR1, linked with CaM concentration response in RyR, and CBD2 as identified by Yamaguchi et al (Yamaguchi, 2004). The CBD2 is a CaM binding site of RyR, this region has high homology between the RyR isoforms and is located within the region marked CaMBD in figure 5.1.1. Initially 4 constructs were designed all of which contained the CaM binding site of RyR2 in addition to the following specified regions. The construct labelled "full" included both flanking regions, "N1" the N terminal and "C1" the C terminal. The construct initially designated "short" and later referred to as CaMBD was designed to include the CaM binding site and short sections of the N and C terminal flanking regions. The domain boundaries of the construct, which determined the amount of content of the flanking regions, were selected with reference to predicted 2rt structure using the same criteria as used in the design of the CaMLD constructs.

```
UNK_274470  TEDEVRDIIRSNIHLQGKLEDPAIRWQMALYKDLPNRTDDTSDPEKTVERVLDIANVLFHLEQKSKRVGR
DSC         cchhhhhhhhcceeecccccccchhhhhhhhhhccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhh
19
PHD         chhhhhhhhhhhhhhcccccccccchhhhhhhhhhhccccccccccchhhhhhhhhhhhhhhhhhhhhhccccccc
Sec.Cons.   cchhhhhhhhcceeecccccccchhhhhhhhhhcccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhh

                  3580      3590      3600      3610      3620      3630      3640
                   |         |         |         |         |         |         |
UNK_274470  RHYCLVEHPQRSKKAVWHKLLSKQRKRAVVACFRMAPLYNLPRHRAVNLFLQGYEKSWIETEEHYFEDKL
DSC         heeeecccccccchhhhhhhhhhhhccccceeecccccccccccccccchhhhhhcccccccccchhhhhhhhh
19
PHD         ccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccchhhhhhhhhhhhhhhhhhhhcccchhhhh
Sec.Cons.   heeeecccccccchhhhhhhhhhhhcccceeecccccccccccccccchhhhhhcccccccccchhhhhhhhh

                  3650      3660      3670      3680      3690      3700      3710
                   |         |         |         |         |         |         |
UNK_274470  IEDLAKPGAEPPEEDEGTKRVDPLHQLILLFSRTALTEKCKLEEDFLYMAYADIMAKSCHDEEDDDGEEE
DSC         hhhhhccccccccccccccccccchhhhhhhhhhhhhhhhccccccchhhhhhhhhhhhhhhhhccccccccccchhh
19
PHD         hhhhhcccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhchhhhhhhhhhhhhhhccccccccccccc
Sec.Cons.   hhhhhccccccccccccccccccchhhhhhhhhhhhhhhhccchhhhhhhhhhhhhhhhhccccccccccchhh
```

Figure 5.1.7   Predicted secondary structure of CaM binding site and flanking regions of RyR2.

Top line is amino acid sequence of region of RyR2 containing CaM binding site. Section highlighted in turquoise is equivalent region to CBD2 and both flanking regions identified in RyR1. Lowest line is consensus sequence of predicted secondary structure generated from 3 algorithms.

Based on location of proline residues relative to areas of conserved predicted helical sections the following amino acids were identified as domain boundaries for construct design: E3545 and K3646.

This design process resulted in 4 potential constructs with the following domain boundaries in RyR2:

1) Full (includes both flanking regions)   3536-3691
2) Short (based on predicted secondary structure)   3545-3646
3) N1 (CBD2 + N term flanking region)   3536-3646
4) C1 (CBD2 + C term flanking region)   3545-3691

As discussed in chapter 1 there have a number of studies examining the role of mutations in the regions flanking CaMBD particularly in reference to the difference in calcium sensitivity between the RyR isoforms (Yamaguchi, 2004, 2005, 2009). However, at the time of this project there are no published CPVT mutations within the region of CaMBD being examined.

5.2 Results

5.2.1 Production of CaMLD constructs

5.2.1.1 Cloning and expression in pET15bmod vector

Initially all 8 potential constructs were selected for cloning and expression trials prior to further selection. pET15b was chosen as a cloning vehicle because it contained a very short linker sequence downstream of the hexa-histidine (6His) encoding sequence within the plasmid. This would mean that there would be that fewer non-specific amino acids are incorporated into the final expressed recombinant protein. The pET15b plasmid is shown in figure 2.1.1.

Using the complimentary DNA (cDNA) of the human Ryanodine receptor isoform 2 (RyR2) as a template, the polymerase chain reaction (PCR) was used to generate a number of different synthetic amplimers that were then ligated into pET15bmod (a modified derivative of pET15b (+)). These were transformed into TOP10 cells. The primers used to generate the amplimers are shown below.

Construct ID	Primer sequence	Amplimer size

1	EcoRI 5'- CC<u>GAATTC</u>GGTGAAACTGGACC -3' (1F) XhoI 5'- CCT<u>CTCGAG</u>CTACTCTTCCGGCC -3' (1R)	1.25 kb
2	EcoRI 5'- CC<u>GAATTC</u>GGTGAAACTGGACC -3' (1F) XhoI 5'- CTC<u>CTCGAG</u>CTAGGGCTTCTCCC -3' (2R)	1.44 kb
3	EcoRI 5'- CG<u>GAATTC</u>AGTGTCCTTGACCTAAATGC -3' (2F) XhoI 5'- CCT<u>CTCGAG</u>CTACTCTTCCGGCC -3' (1R)	1.09 kb
4	EcoRI 5'- CG<u>GAATTC</u>AGTGTCCTTGACCTAAATGC -3' (2F) XhoI 5'- CTC<u>CTCGAG</u>CTAGGGCTTCTCCC -3' (2R)	1.28 kb
5	EcoRI 5'- CG<u>GAATTC</u>GGAGAAAAGGTTCTGC -3' (3F) XhoI 5'- CCT<u>CTCGAG</u>CTACTCTTCCGGCC -3' (1R)	1.01 kb
6	EcoRI 5'- CG<u>GAATTC</u>GGAGAAAAGGTTCTGC -3' (3F) XhoI 5'- CTC<u>CTCGAG</u>CTAGGGCTTCTCCC -3' (2R)	1.21 kb
7	EcoRI 5'- CG<u>GAATTC</u>GATTCCAGTCAAATTGAGC -3' (4F) XhoI 5'- CCT<u>CTCGAG</u>CTACTCTTCCGGCC -3' (1R)	0.61 kb
8	EcoRI 5'- CG<u>GAATTC</u>GATTCCAGTCAAATTGAGC -3' 4F) XhoI 5'- CTC<u>CTCGAG</u>CTAGGGCTTCTCCC -3' (2R)	0.80 kb

Table 5.2.1.1 Primer sequence and amplimer sizes of CaMLD pET15b constructs.
All forward primers contained the EcoRI restriction site whereas the reverse primers contained the XhoI restriction site. The restriction sites are underlined in the primers.
The amplimers were double digested using these enzymes prior to ligation into pET15bmod plasmid that had been digested similarly earlier.
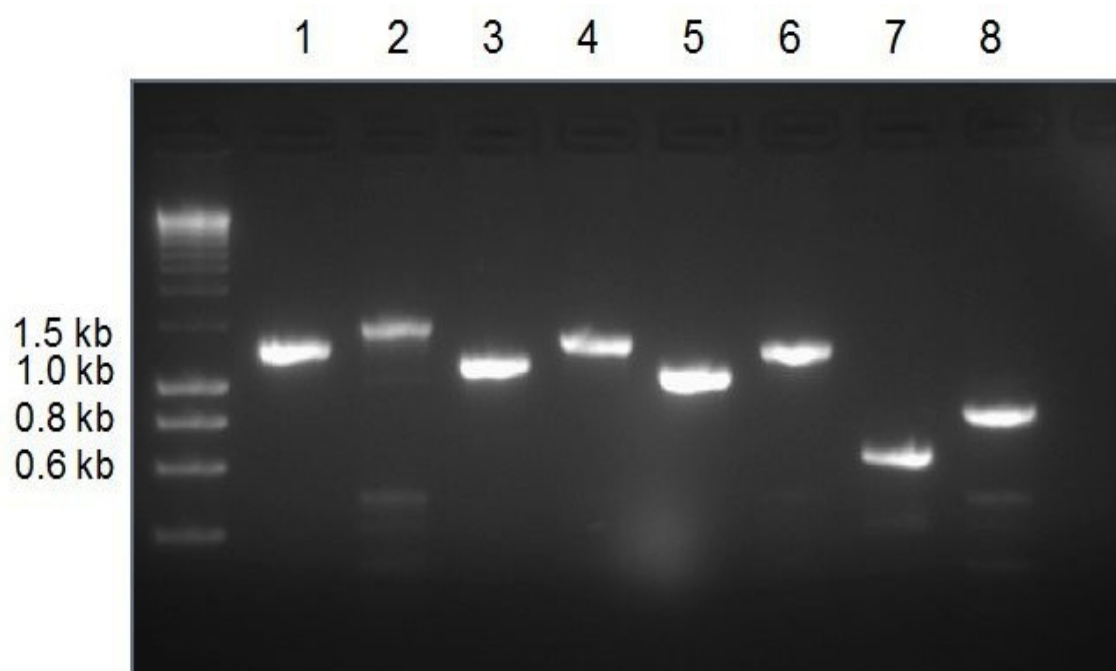
Figure 5.2.1.1 Agarose gel of PCR products of 8 initial potential constructs.

Size of expected products by construct number.

1) 1.25 kb
2) 1.44 kb
3) 1.09 kb
4) 1.28 kb
5) 1.01 kb
6) 1.21 kb
7) 0.61 kb
8) 0.80 kb

PCR products of the expected size were generated for all 8 potential constructs. As detailed in chapter 2 the digested PCR products were ligated into pET15bmod vector. Colony PCR (cPCR) was conducted on a range colonies from antibiotic selection agar plates.
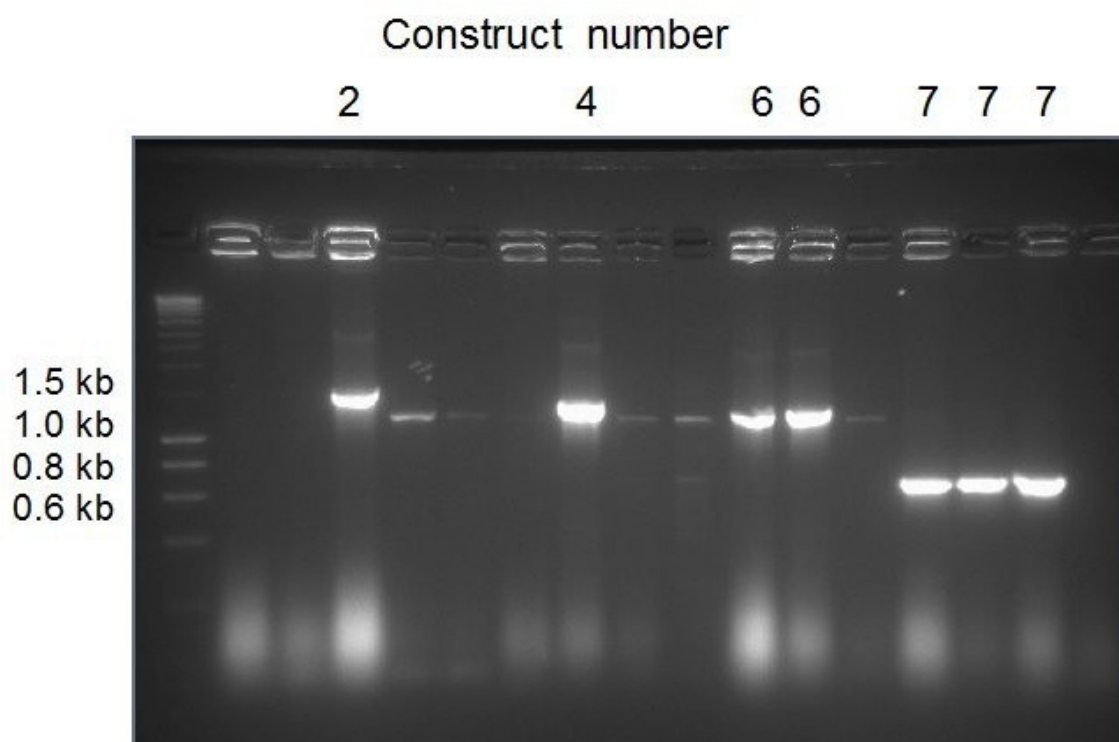
Figure 5.2.1.2 Agarose gel of cPCR of selected pET15bmod constructs.

Bands of expected size observed for colonies of constructs 2 (1.44kb), 4 (1.28kb), 6 (1.21kb) and 7 (0.61kb).

Initially only positive cPCR results were observed for colonies of constructs 2, 4, 6 and 7. While not all potential constructs had given positive results the majority of the designed domain boundaries could be tested. Construct 7 was the only positive one to feature the residue 4163 C terminus, however as identified by the analysis conducted during the construct design phase construct 7 was targeted as a key construct to test. It was decided to run protein expression trials with these constructs.
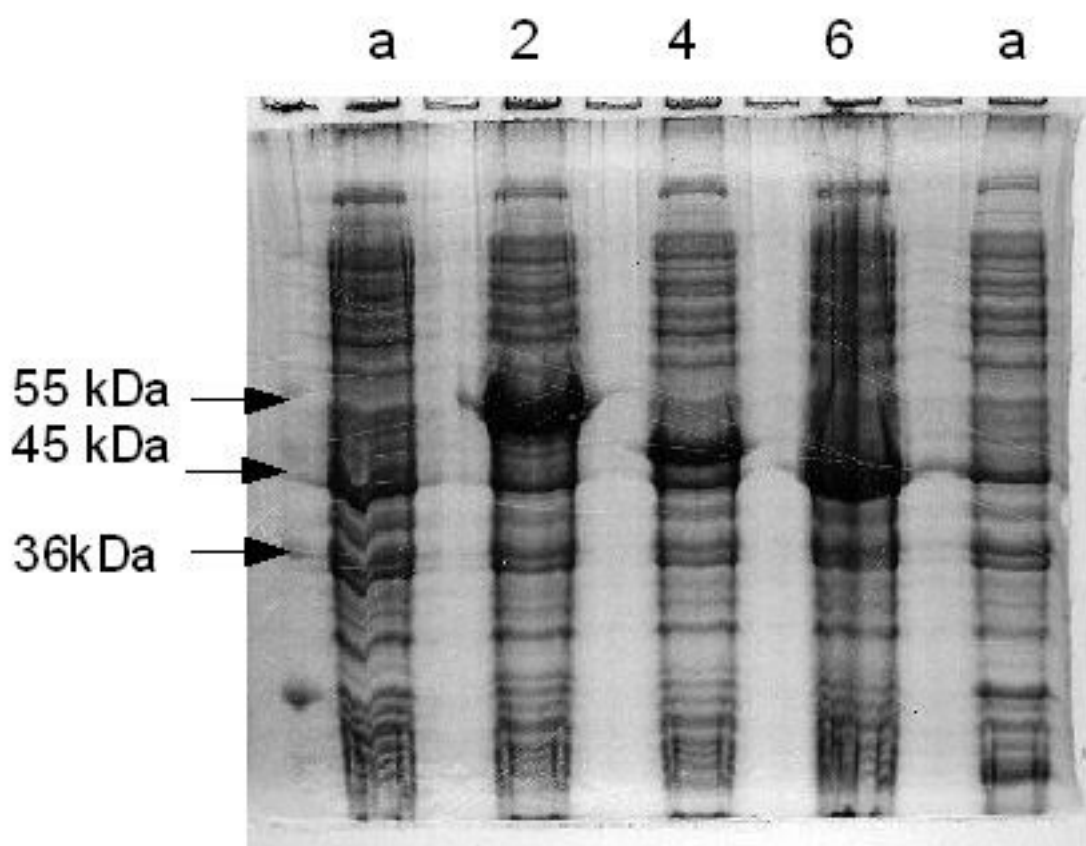
Figure 5.2.1.3 Expression of constructs in pET15bmod vector in Rosetta (DE3).

The 6His constructs 2, 4 and 6 were lysed, and the entire protein lysate was subjected to SDS-PAGE alongside protein expressed from native pET15bmod (without cloned inserts). This figure shows that the constructs labelled 2, 4 and 6 gave strong protein bands in the expected sizes. Lanes labelled (a) were the lysates from the cells transformed with native pet15b, in comparison, these control lanes do not show strong protein bands as in the lanes labelled 2,4 and 6.

Expected size of constructs including tag from pET15bmod vector:

2) 56.0 kDa

4) 50.5 kDa

6) 47.9 kDa

Expression bands of the expected size were observed for constructs 2,4 and 6. Although not shown similar expression levels were observed for construct 7 in later trials. To observe the soluble protein content the lysate was centrifuged then applied to Ni-chelate chromatography. The amount of recombinant protein that remained in the pellet following lysis was compared to the amount of recombinant protein that was

successfully eluted from Ni-chelate chromatography. The eluate fraction represented the soluble portion of the protein preparation whereas those that remained in the pellet following lysis represented the insoluble fraction of the protein preparation. By this quick screening method, it was possible to determine which constructs had higher probability to deliver good soluble protein that would be useful and amenable to downstream purification processes.

Upon analysis of the above samples it was noted that no soluble expression was obtained with the protein occurring in inclusion bodies. A number of variables within the growth and induction phase were trialled. These included alterations to the timings, temperatures and IPTG concentration stated in chapter 2. A range of additives associated with reported increase in soluble expression including alcohol and glycerol were included with the growth media. Additional Ca^{2+} in the growth media was also trialled in case the presence of the EF hand motifs in the expressed constructs were disrupting the Ca^{2+} homeostasis of the expressing cells. The addition of glycerol and Ca^{2+}, and also variations within the stated lysis buffers were tested. None of these tested variables made an observable difference in the quantity of soluble protein.

The only tested variable that made any observable difference to the amount of soluble protein was the addition of detergent to the lysis buffer. As an initial trial NP40 lysis buffer (Pierce) was added into the existing lysis buffer. The growth, induction and Ni affinity purification protocols as detailed in chapter 2 were followed with the exception that the basic Ni affinity protocol was used which omits the 60mM imidazole lysis and 70mM imidazole wash steps. As a result few of the contaminant bands, as observed in the crude lysate preparation in figure 5.2.1.3, were removed from the sample. To try to improve the purity of the sample, ion exchange chromatography (Q sepharose with NaCl gradient) followed by gel filtration, as per chapter 2 protocol, was used. Samples of the selected pooled fractions from the NP40 containing preparation of construct 4 are shown in the figures below.
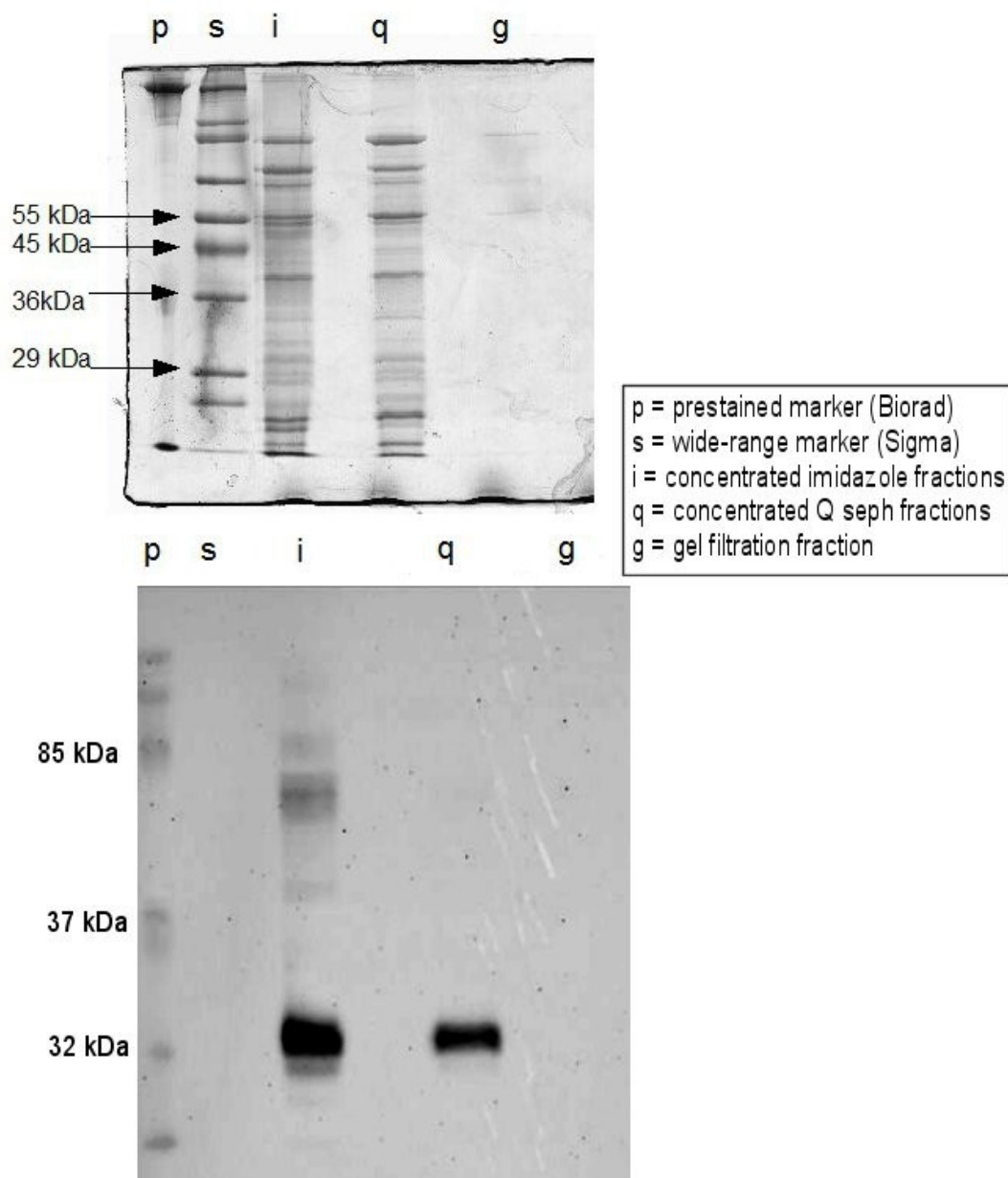
Figure 5.2.1.4   NP40 treated purification of construct 4/pET15bmod/Rosetta (DE3)

(Upper): Coomassie stain of equivalent SDS-PAGE to Western blot of NP40 purification of construct 4/pET15bmod/Rosetta (DE3).

(Lower): Western blot of NP40 preparation of construct 4/pET15bmod/Rosetta (DE3). Construct 4 expected size 50.5 kDa.

Whilst the majority of protein remained in the insoluble fraction during the NP40 preparation sufficient material was obtained to track a band of the expected size by SDS-PAGE through the purification stages (upper panel). The Western blot showed several bands for the imidazole fractions though the strongest signal is observed for the band above the 32 kDa marker (lower panel). This was the only band seen after ion exchange chromatography but is not clearly observable after gel filtration. It should be noted that unlike the other eluted peaks the gel filtration fractions were not concentrated.

Although only shown for construct 4, the above experiments were repeated for the other constructs including 7. As observed, the other constructs displayed no significant increase in soluble protein content.

5.2.1.2 Cloning and expression in pETMM vectors

In order to address the solubility issues observed with the expression of the recombinant protein when using a hexa-histidine tag a range of other tags were selected to address the solubility issue. The pETMM series of vectors contains the same multiple cloning site (MCS) but with different tags that will be expressed at the N terminus of the cloned insert. This allows the same set of primers to be used to generate insert suitable for the chosen vectors in the range. The dual tag vectors pETMM 60 (NusA-6His), see figure 2.1.2, and pETMM 20 (Trx-6His) were initially selected owing to the reported increased solubility of their associated tags.

To further increase the speed of testing for the expression of soluble protein; 2 of the initial 8 constructs were selected to be inserted into the pETMM vector series. These were constructs 5 (CaMLD) and 7 (CIR). Despite failing to show clear soluble expression with the 6His tag these constructs were selected, as of all the designed constructs, their domain boundaries most closely matched the predicted internal domains within the region of the I domain being studied.

Using the complimentary DNA (cDNA) of the human Ryanodine receptor isoform 2 (RyR2) as a template, the polymerase chain reaction (PCR) was used to generate a number of different synthetic amplimers that were then ligated into either pETMM20 or

pETMM60. These were transformed into TOP10 cells. The primers used to generate the amplimers are shown in the table below.

Construct ID	Primer sequence	Amplimer size
CaMLD	EcoRI 5'-GAGC<u>GAATTC</u>GGGAGAAAAGGTTCTGCAGG-3' (CaMLD_F) SalI 5'- CGTC<u>GTCGAC</u>TCAGGGCTTCTCCCACTGGGTTCG -3' (PET_R1)	1.01 kb
CIR	EcoRI 5'- GAGC<u>GAATTC</u>GGATTCCAGTCAAATTGAGCT-3' (CIR_F) SalI 5'- CGTC<u>GTCGAC</u>TCAGGGCTTCTCCCACTGGGTTCG-3' (PET_R1)	0.61 kb

Table 5.2.1.2 Primer sequence and amplimer sizes of CaMLD pETMM constructs.
All forward primers contained the EcoRI restriction site whereas the reverse primers contained the SalI restriction site. The restriction sites are underlined in the primers.

The amplimers were then double digested with these enzymes and ligated into either pETMM20 or pETMM60 plasmid vectors that had been digested similarity earlier. The regions that were being amplified did not host these restriction sites prior to them being engineered in.
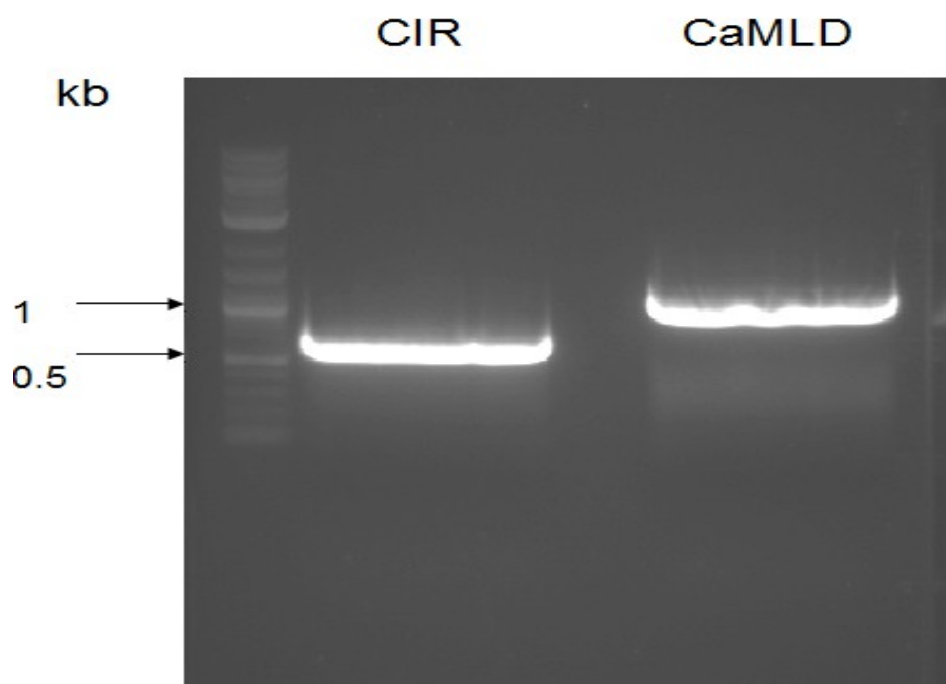
Figure 5.2.1.5 Amplimers of CaMLD and CIR.

Bands of the expected size (CaMLD 1.01 kb and CIR 0.61 kb) were observed on an agarose gel from samples run from the PCR used to generate the amplimers.

PCR products of the expected size were generated for all 8 potential constructs. As detailed in chapter 2 the digested PCR products were ligated into pET15bmod vector. Colony PCR (cPCR) was conducted on a range colonies from antibiotic selection agar plates. These PCR products were digested and ligated into the pETMM 20 and 60 vectors. Restriction digests were conducted on plasmid DNA prepared from 2 selected colonies per vector post ligation transformation.
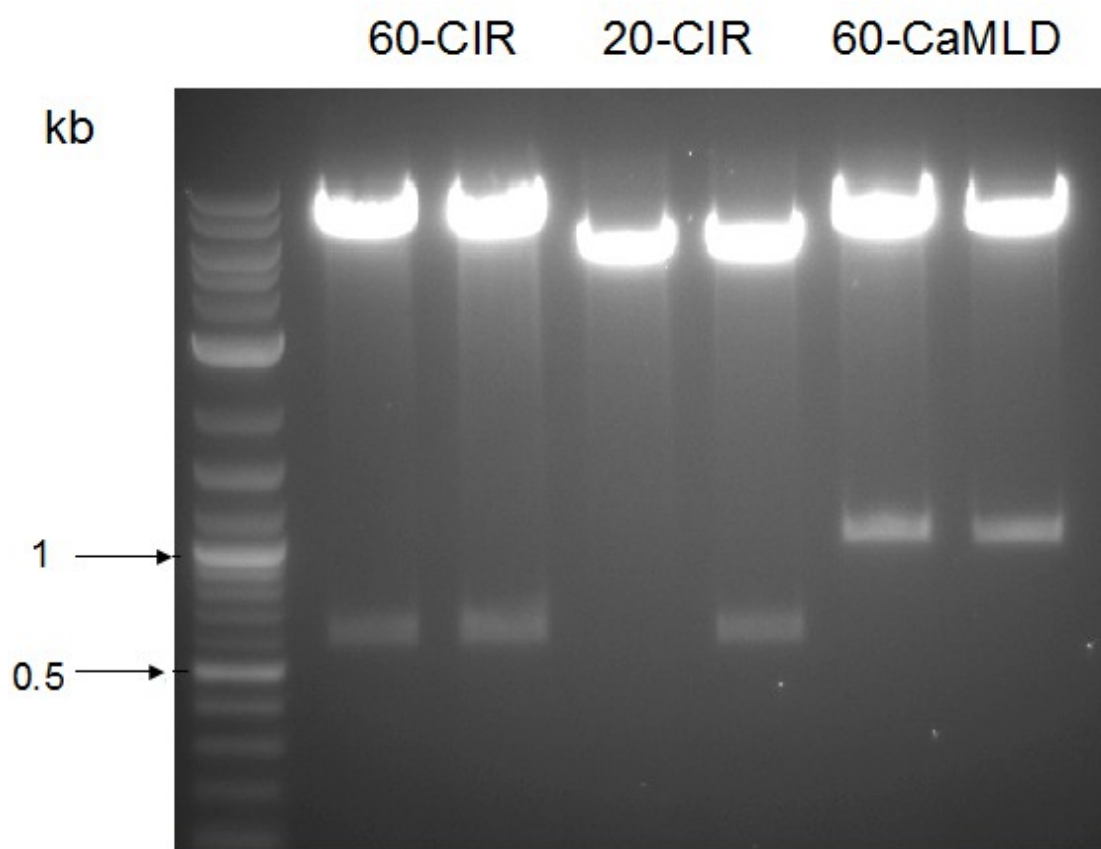
Figure 5.2.1.6 Restriction digests of CaMLD and CIR constructs in pETMM 20 and 60.

Bands of the expected size were observed for the tested colonies (CIR 0.61kb, CaMLD (1.01kb) except for one of the CIR pETMM 20 colonies. No colonies were obtained for CaMLD in pETMM20 so this construct was not tested at this time.

The positive colonies from the restriction digest were transformed into TOP10 cells and from these plasmid DNA prepared. This DNA was sequenced and the presence of the desired inserts confirmed. The successful plasmid constructs were then transformed into BL21(DE3) and Rosetta(DE3) *E.coli* cells for protein expression trials. The growth and induction conditions as described in chapter 2 were followed. Cells were lysed and the expression levels examined by SDS-PAGE.
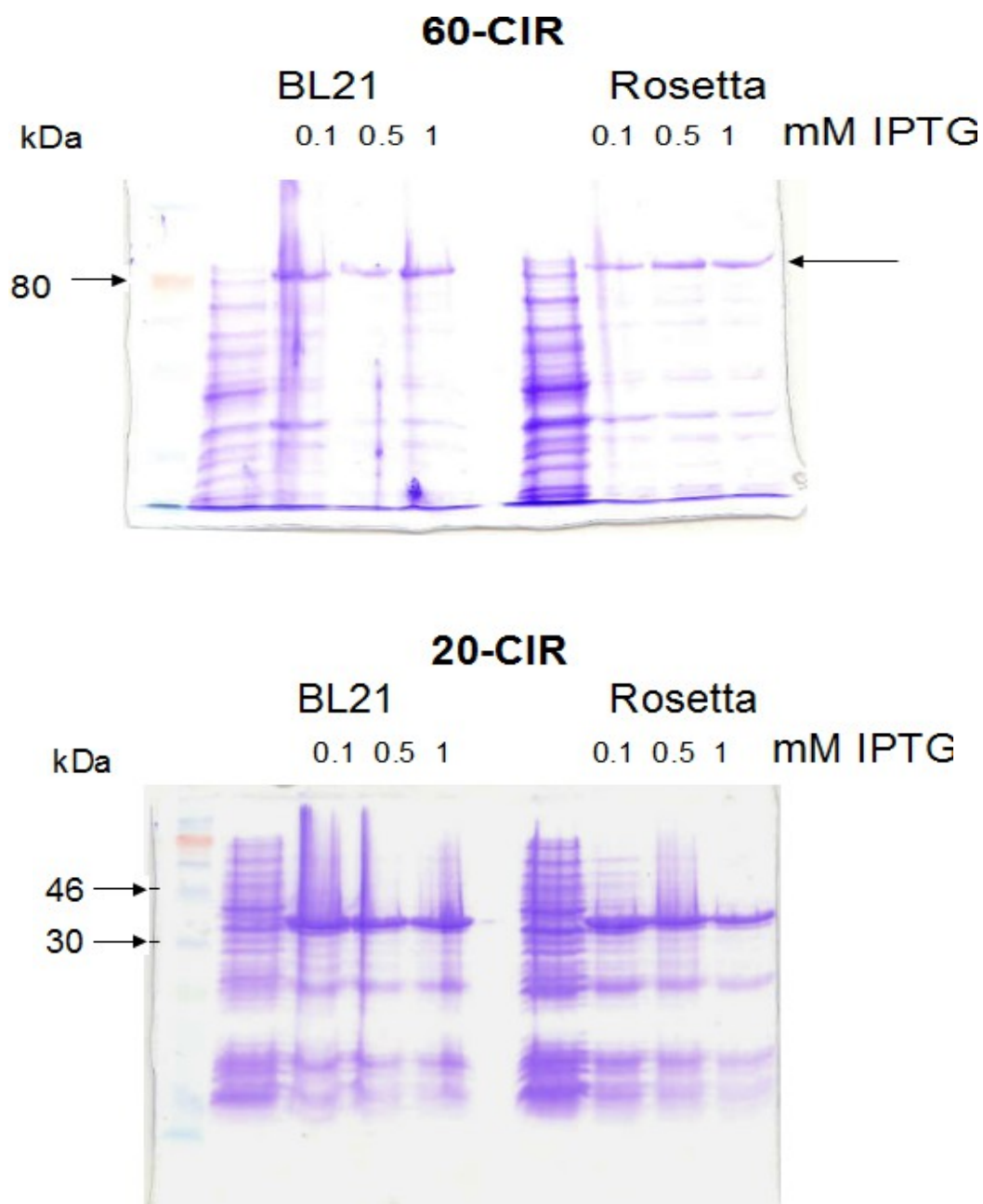
Figure 5.2.1.7a   SDS-PAGE analysis of expression of CIR and CaMLD constructs in pETMM 20 and 60 vectors.

IPTG concentrations were varied in testing for the construct and vector combinations in both BL21 (DE3) and Rosetta (DE3) cells. Cell pellets were lysed and analysed by SDS-PAGE alongside protein expressed from the native non-recombinant vector (shown on left of gels). This figure shows that the CIR construct in both vectors gave strong protein bands in the expected sizes. However weak expression was observed for the 60-CaMLD construct. No apparent differences were observed between the expression cell

60-CaMLD

BL21   Rosetta

kDa   0.1 0.5 1   0.1 0.5 1   mM IPTG

80→

Figure 5.2.1.7b   SDS-PAGE analysis of expression of CIR and CaMLD constructs in pETMM 20 and 60 vectors.

IPTG concentrations were varied in testing for the construct and vector combinations in both BL21 (DE3) and Rosetta (DE3) cells. Cell pellets were lysed and analysed by SDS-PAGE alongside protein expressed from the native non-recombinant vector (shown on left of gels). This figure shows that the CIR construct in both vectors gave strong protein bands in the expected sizes. However weak expression was observed for the 60-CaMLD construct. No apparent differences were observed between the expression cell line used. (Expected band sizes : 60-CIR -80 kDa; 20-CIR -42 kDa; 60-CaMLD 99 kDa)
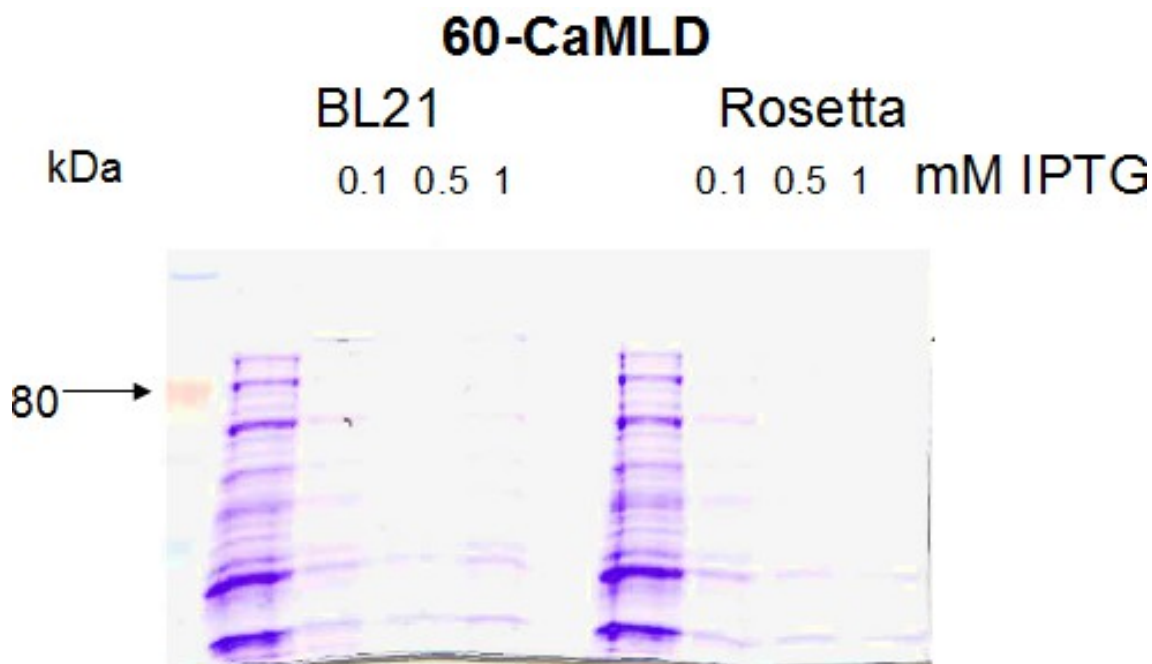
Despite the low level of expression seen for the 60-CaMLD construct it was decided to include samples from this construct when testing for soluble expression. As for the pET15bmod constructs the Ni-chelate chromatography approach was adopted.
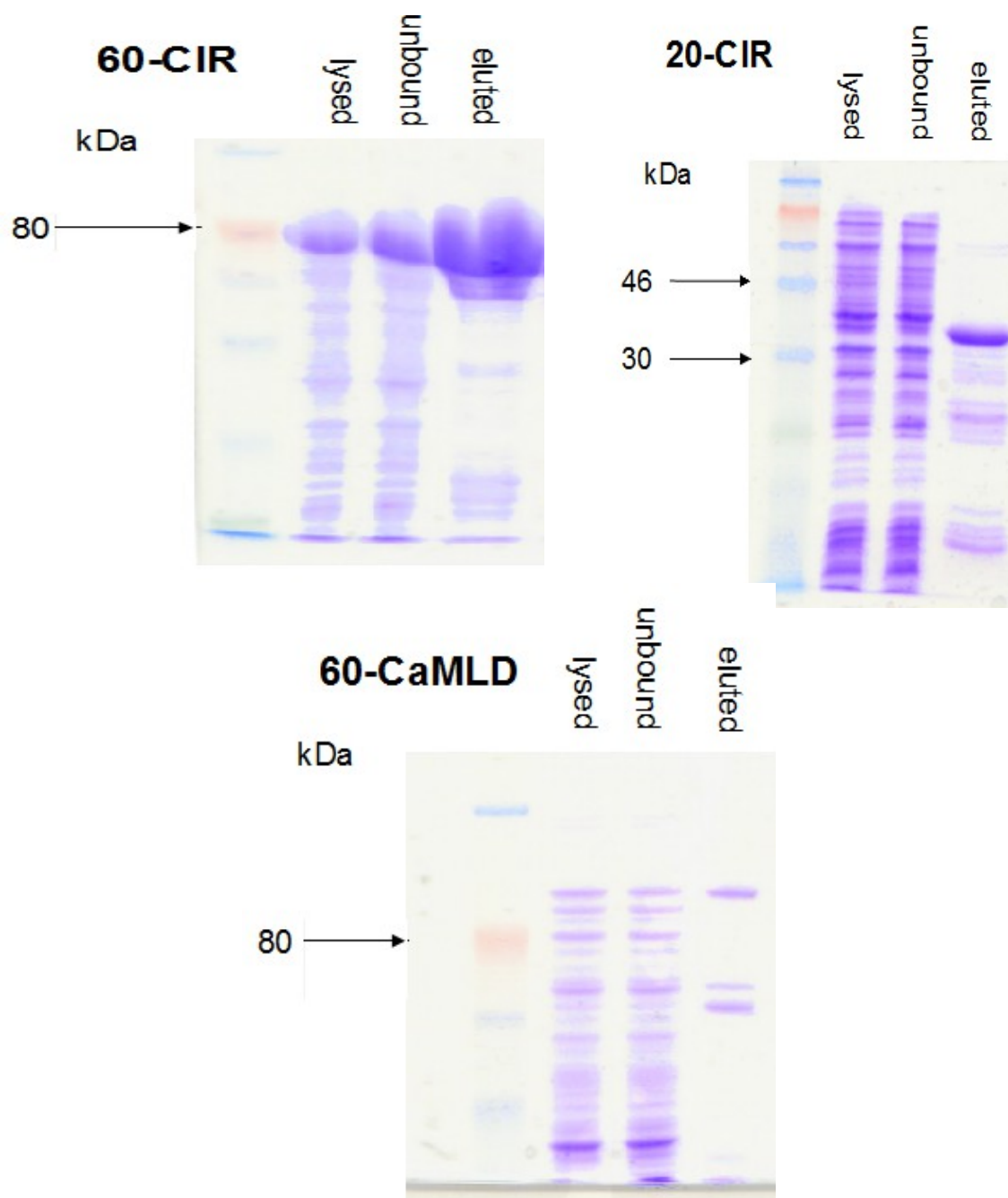
Figure 5.2.1.8 Ni-chelate chromatography purification of CIR and CaMLD constructs in pETMM20 and 60 vectors.

Protein bands of the expected size were observed for the all the constructs tested. The lysed band lane shows the lysed cell pellet material that was loaded onto Ni-chelate affinity beads. The unbound lane is a sample of the material that was washed off the beads. The eluted lane is the material eluted from the beads representing the soluble partially purified content obtained from the cell pellet. Expected size of construct bands including tag from pETMM vectors: 60-CIR -80 kDa; 20-CIR -42 kDa;

60-CaMLD -99 kDa

As discussed in the construct design section of this chapter the CaMLD construct had been identified as the construct offering most possibility for further research. Despite the lower yield relative to the other constructs, the apparent lesser number of contaminants offered more potential for higher purity samples downstream. It has to be acknowledged that usually higher yielding proteins are selected owing to the expected loss at each stage of handling. Also the contaminants for 60-CaMLD appear to be at a similar concentration to the desired protein which can often mean separating a meaningful amount of wanted material can be more difficult.

Later further colonies of 60-CaMLD were tested for expression levels, from these, colonies in Rosetta (DE3) cells that had similar expression levels to those seen for the CIR constructs were identified. Subsequently these colonies gave matching yields from Ni-chelate chromatography. Despite concerns that an increase in total protein might result in an increase in the number of contaminants, only the 2 major contaminant bands observed in figure 5.2.1.8 still remained though were present at obviously higher levels. Once the initial apparent issue of low protein yields had been solved so 60-CaMLD expressed in Rosetta (DE3) was selected for further work in this project.

To aid the identification of the bands observed post Ni-chelate chromatography for 60-CaMLD so mass spectrometry was conducted. Enzymatically digested samples from the bands were run on an ABI 4800 MALDI TOF/TOF Analyzer and the generated peptide fingerprint maps analysed.
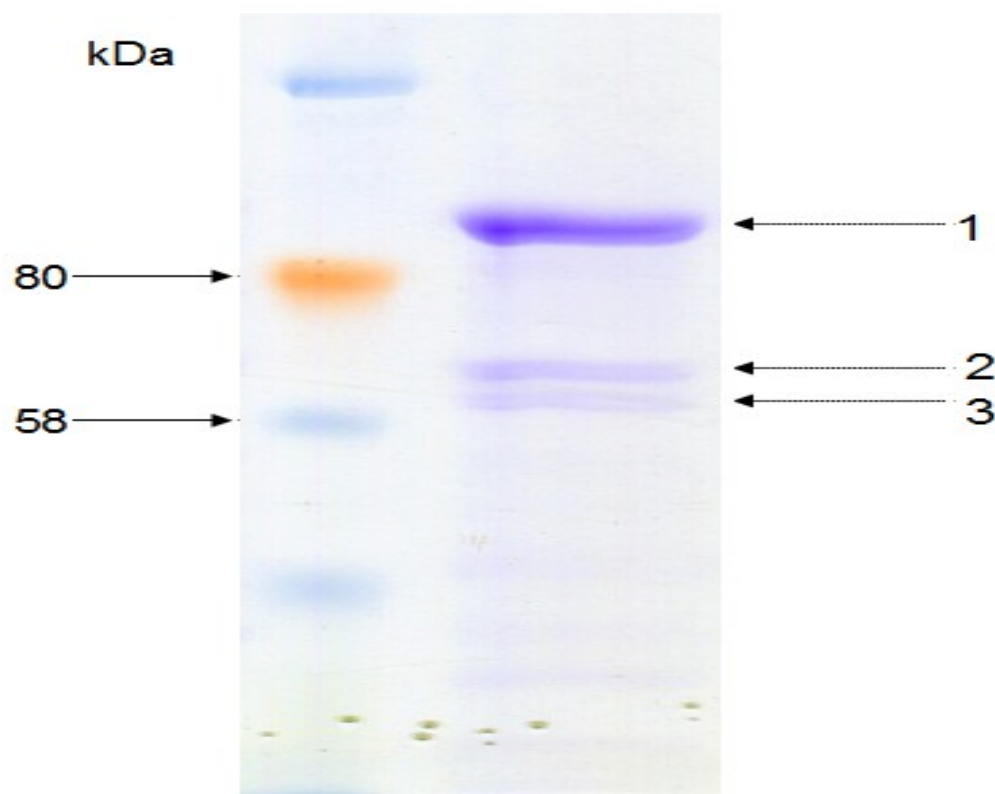
Figure 5.2.1.9 Mass spectrometry analysed SDS-PAGE samples from Ni-chelate chromatography purified 60-CaMLD.

This figure illustrates the increased protein yield from 60-CaMLD as described above. Samples from the numbered bands were analysed by mass spectrometry. Band 2 was the only sample identified to a high probability confidence level.

Band 1: fragments of NusA and RyR2

Band 2: bi-functional polymyxin resistance protein ArnA (75 kDa E.coli protein)

Band 3: insufficient ions produced to give clear result

The mass-spectrometry data of band-1 correlated best with the known fingerprint tryptic digest of the 60-CaMLD protein, hence confirming that the correct protein was being purified. Band 2, the ArnA protein, is an *E.coli* protein associated with antibiotic resistance and lipid metabolism. It does not have any obvious connection to RyR2 and is expected to be present to some extent in all the expressed construct protein samples. Band 3 appears to be selected during purification with the 60-CaMLD protein. Although not identified by mass spectrometry it is proposed that it may be part of the fusion tag from the vector, the band appears close to the expected size of the NusA-6His tag.

Purification optimisation trials were conducted to remove bands 2 and 3, seen in figure 5.2.1.9, from 60-CaMLD samples. Following manufacturer recommendations initially increasing concentrations of imidazole used as a series of wash steps was tested.
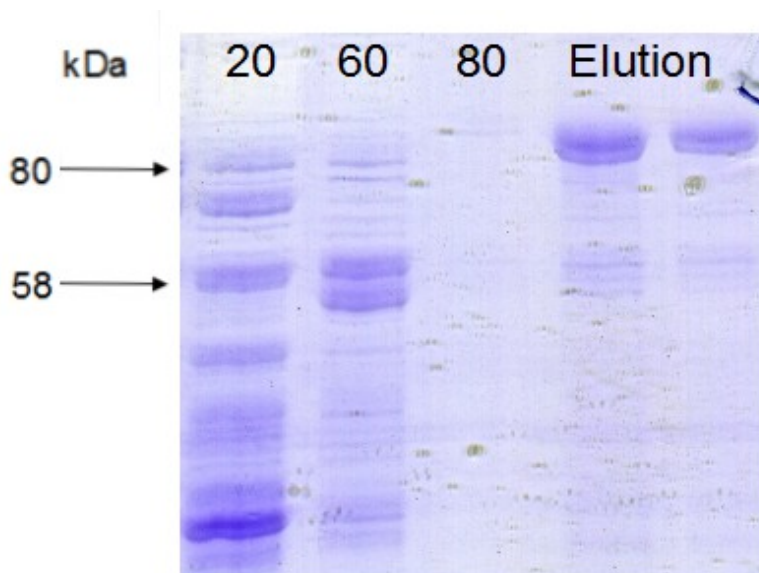

Figure 5.2.1.10 SDS-PAGE of increasing imidazole concentration wash steps during Ni-chelate chromatography on 60-CaMLD protein sample.

The numbered lanes indicate the concentration (mM) of imidazole wash used. The elution lanes represent 2 samples taken during the standard 250mM imidazole elution step. The strong band at the top of the elution lanes is the 60-CaMLD protein. While there is a small amount of the contaminant bands 1 and 2 present, the amount is much reduced compared to the previous Ni-chelate purification trials shown. The 20mM imidazole wash step does remove some of the contaminant material but the 60mM step is much more effective. The increase to 80mM does not bring any obvious increase in sample purity while removing 60-CaMLD protein from the beads.

Prior to to these increasing imidazole concentration wash steps being conducted, cell pellet had been lysed in 20mM imidazole and washed in 40mM imidazole. As a result of this trial pellet was lysed in 60mM and washed with 70mM imidazole prior to elution with 250mM imidazole as previously.
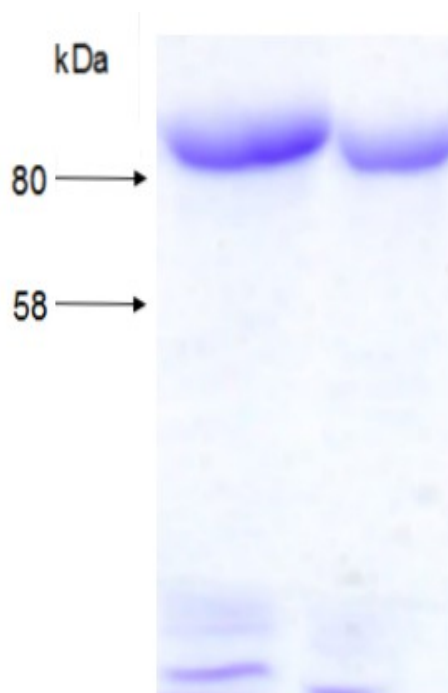
Figure 5.2.1.11 SDS-PAGE of Ni-chelate chromatography purified 60-CaMLD protein samples using 60mM imidazole lysis and 70mM wash.

Samples of differing concentrations of two purification preparations of 60-CaMLD. While a few minor lower weight contaminants remain, the 2 major contaminant bands have been removed using the increased imidazole concentration modification to the Ni affinity purification protocol. For all further experiments requiring tagged 60-CaMLD, protein of this purification quality was used.

As part of the purification process N terminal dual tag was cleaved from the fusion protein. The pETMM vectors contain a TEV cleavage site downstream of the tag before the expressed insert enabling the removal of the tag.
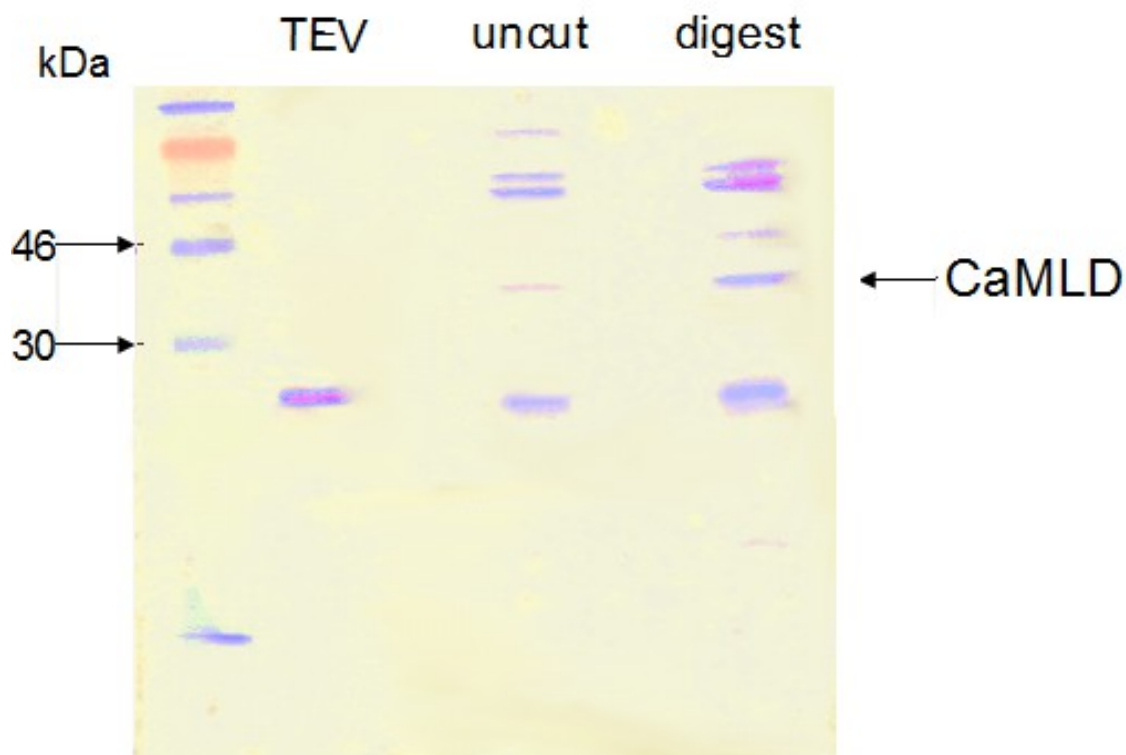
Figure 5.2.1.12 Trial of TEV protease cleavage of fusion tag from 60-CaMLD protein.

SDS-PAGE in figure shows purified in house produced TEV that was used for tag cleavage digests. A low concentration sample of 60-CaMLD protein, uncut lane, was incubated with TEV for 2 hours and a concentrated sample from the digest mixture loaded onto the gel, digest lane. A band, at the size of cleaved CaMLD (39 kDa), was present in the digest lane indicating cleavage of the tag from the fusion protein.

A 2nd pass through Ni affinity beads was used to remove the cleaved tag and TEV protease, TEV used was expressed with 6His tag allowing for purification and subsequent removal from sample post digest.
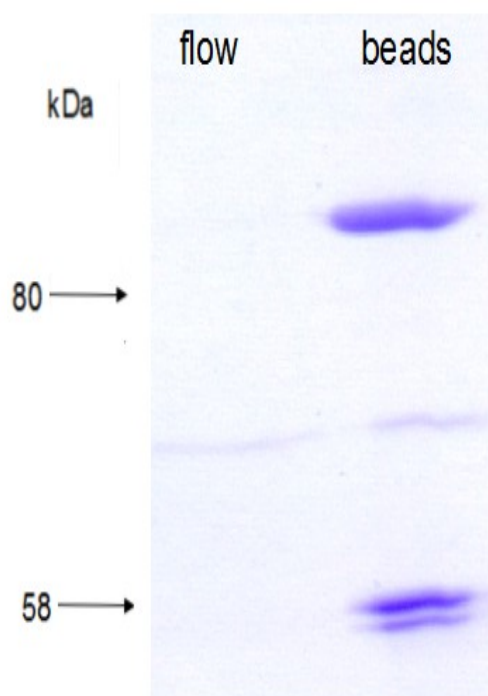
Figure 5.2.1.13 CaMLD protein samples after 2nd pass through Ni affinity beads post TEV protease tag cleavage.

The beads lane is a sample of the Ni beads after the flow through was collected. This lane contains CaMLD (middle band), also present are cleaved NusA tag (top band) and TEV protease (lowest band). The lane marked flow indicates material that was the collected flow through from Ni beads, this sample only contains the cleaved CaMLD protein.

Initial tag cleavage trials were conducted on material that had yet to undergo the increased imidazole wash steps hence the presence of the contaminant bands in the samples shown. A higher yield of CaMLD protein was obtained from later samples that had been purified using the increased imidazole concentration steps. Also in these preparations the dialysis step, to reduce the imidazole concentration, was combined with the TEV cleavage step. Although TEV had lower activity in the dialysis buffer than TEV specific buffers, sufficient activity was obtained when TEV was used at higher amounts, 1:10 - 1:20 ratio range to CaMLD, that >90% tag cleavage was observed. This level of TEV activity was observed when the protease was added to the dialysis samples and left overnight at 4°C, as the standard protocol involved overnight dialysis this allowed two purification steps to be conducted in the same time frame. The excess TEV

was removed during the 2nd pass with Ni beads. The initial incubation time with the beads had been 1 hour however this was reduced to 30 minutes which improved recovery of the CaMLD protein from the beads while the TEV and cleaved NusA tag remained behind.

Cleaved protein was combined and stored at 4°C. Examination of the pooled sample after overnight storage revealed a large amount of aggregation. SDS-PAGE analysis showed that the aggregate was composed of the cleaved protein, losses of ~80% of the soluble material were observed. Increased NaCl concentration in the dialysis/cleavage buffer was found to increase the storage time. As a result all buffers from lysis onwards contained 300mM NaCl, as listed in the purification protocol in chapter 2. Storage of smaller aliquots, rather than one pooled sample, and keeping concentrations below 2mg/ml also helped to reduce losses due to aggregation. All these measures combined provided a window of less than 48hrs with which to work with the cleaved protein. Uncleaved protein could be stored for 5 days at 4°C prior to usage.

Once soluble purified CaMLD protein could be produced then work to produce the mutant forms of CaMLD was conducted. As discussed in the chapter introduction 4 CPVT mutations had been identified for further study, G3496S, E4076K, N4104K and H4108Q. Mutation of recombinant wild type (wt) CaMLD DNA was performed using the QuikChange II XL Site-Directed Mutagenesis Kit (Stratagene) to generate 4 clones each containing one of the single mutations listed. The mutated plasmids were used to transform TOP10 cells and fresh mutated plasmid DNA produced. Restriction digests were conducted to check for the presence of inserts of the correct size after mutagenesis.
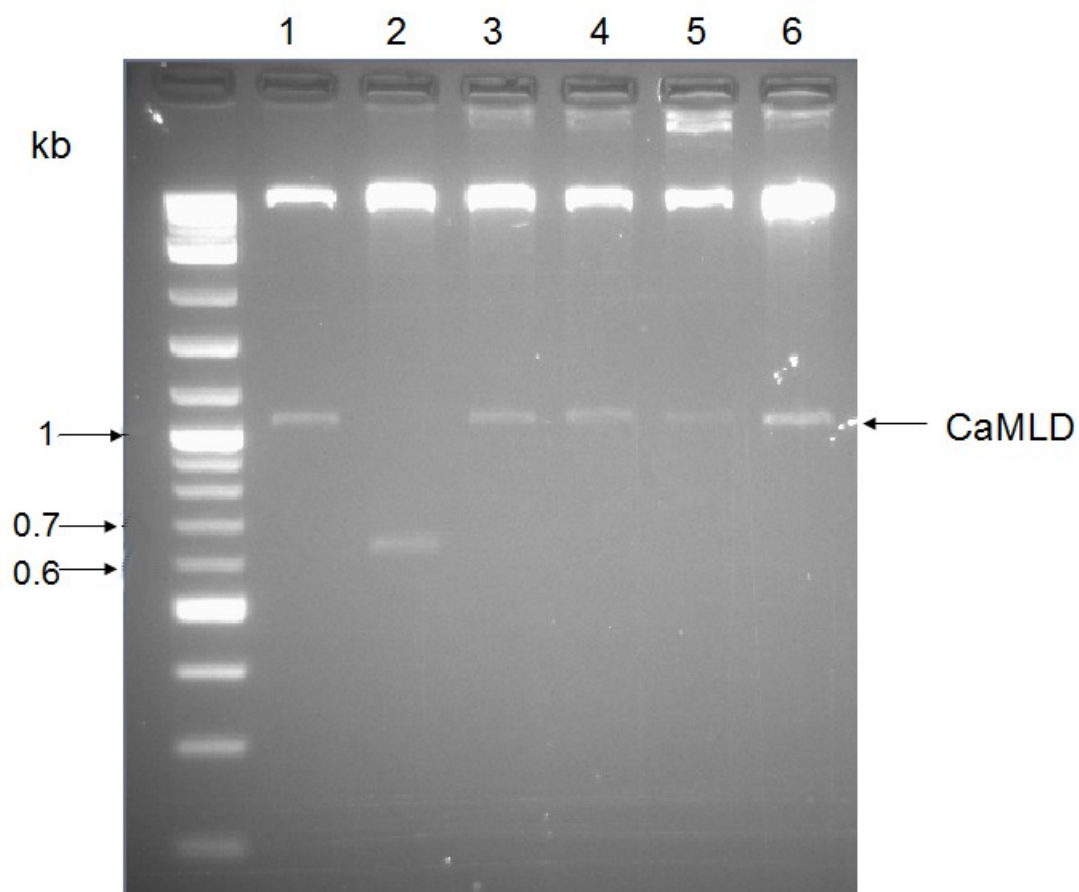
Figure 5.2.1.14 Restriction digests of CaMLD DNA post mutagenesis.

Lanes 1 and 2 contain wt CaMLD that was checked prior to mutagenesis. Lane 2 contains a band of the incorrect size so was rejected for further use. The other lanes contain DNA post mutagenesis which is of the expected size.

Lane 3: G3496S

Lane 4: E4076K
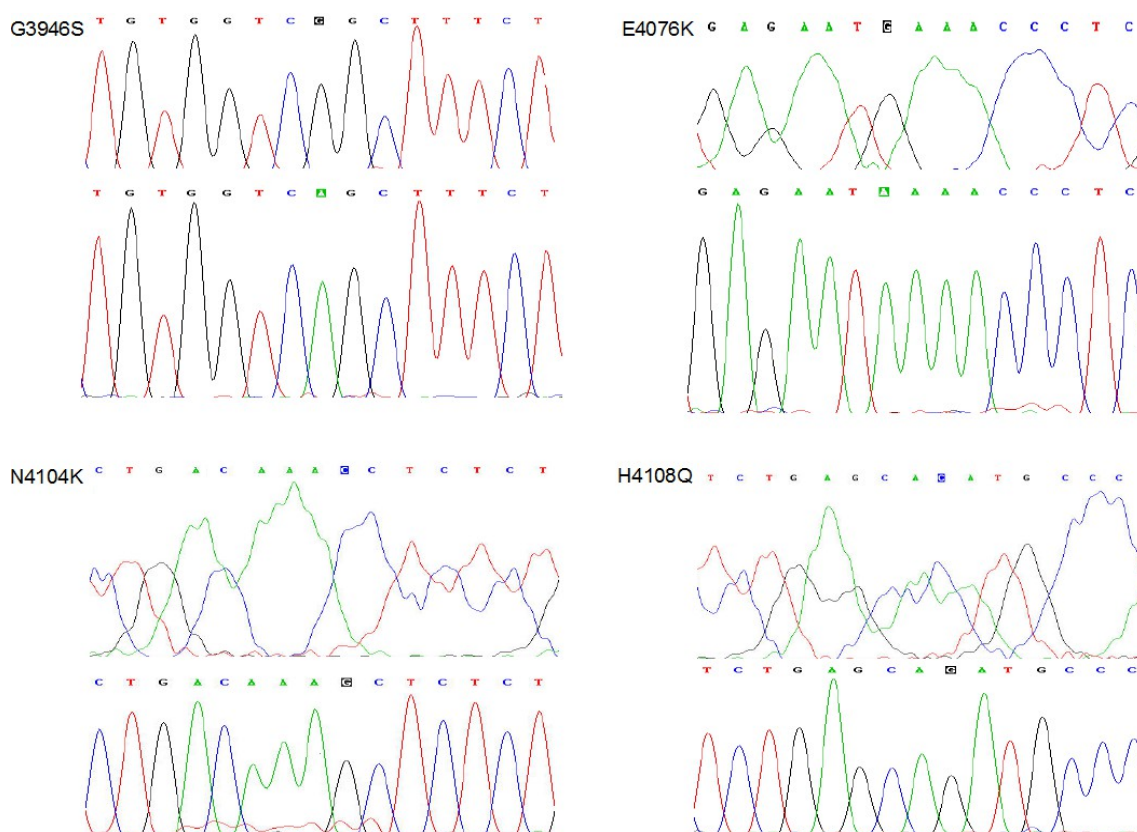
Lane 5: N4104K

Lane 6: H4108Q

Figure 5.2.1.15 Sequencing chromatograms of mutated forms of CaMLD.

Figure is split into 4 sections, one per mutation. In each section there are 2 panels showing the region of DNA in which the mutation listed occurs. The upper panel contains the chromatogram from the wt CaMLD and the lower displays the mutant sequence. In the DNA sequence shown above each chromatogram the location of the single nucleotide mutation required to generate the mutant form is highlighted. For all 4 mutant forms created the correct mutation is found.

Once the correct mutations has been confirmed by sequencing then the mutant plasmid DNA was used to transform Rosetta (DE3) cells. Expression and purification was conducted as for the wt CaMLD following the protocols listed in chapter 2. A batch of the wt and mutants forms of CaMLD was prepared. Post tag cleavage these samples underwent gel filtration to give very pure samples. A western blot probing with an anti-NusA antibody was conducted to check the detection of the wt and mutant forms of CaMLD.
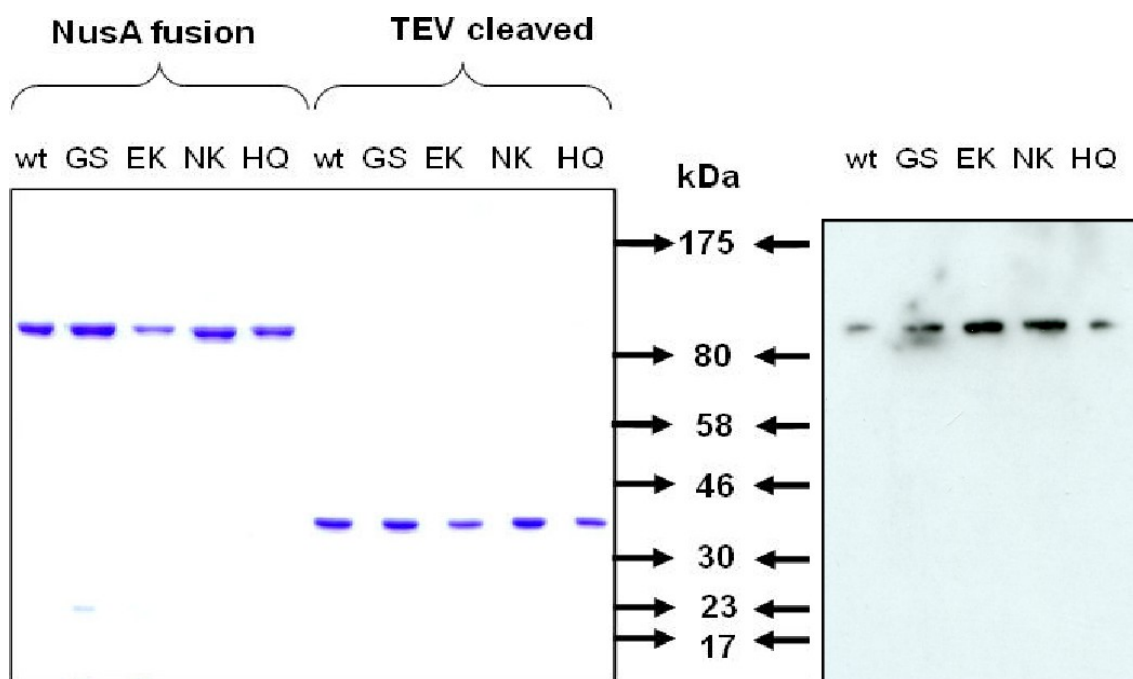
Figure 5.2.1.16 SDS-PAGE and western blot of wt and mutant forms of CaMLD.

Forms of CaMLD are referred to by amino acid single letter code of mutation. Left hand side of SDS-PAGE shows Ni affinity purified material which was used for the western blot. As no antibody was available to the RyR2 CaMLD region the western blot was probed using an anti-NusA antibody. Only a single band is observed for each sample in the western blot which indicated the use of the anti-NusA antibody as a detection tool. The right hand side of the SDS-PAGE displays CaMLD samples that had gone through gel filtration after the fusion tag removal process. As can be seen very pure samples were produced by these final steps however only a very small of protein could be practically produced this way.

5.2.2 Production of CaMBD constructs

5.2.2.1 Cloning and expression in pGEX-6P-1 vector

Initially all 4 potential constructs were selected for cloning and expression trials prior to further selection. pGEX-6P-1 was chosen as a cloning vehicle because it contained an N terminal GST tag, mainly this system was selected as this tag was not present in any of the pETMM vectors used so avoiding potential downstream problems when working with tagged CaMLD proteins. The presence of the tag would also allow for affinity

purification of the expressed protein. The pGEX-6P-1 plasmid is shown in figure 2.1.3. Using the complimentary DNA (cDNA) of the human Ryanodine receptor isoform 2 (RyR2) as a template, the polymerase chain reaction (PCR) was used to generate a number of different synthetic amplimers that were then ligated into pGEX-6p-1. These were transformed into TOP10 cells. The primers used to generate the amplimers are shown below.

Construct ID	Primer sequence	Amplimer size
Full	EcoRI 5'- GGAC<u>GAATTC</u>AACAGGACTGATGATACCTCAG -3' (1F) XhoI 5'- GCCA<u>CTCGAG</u>TCAATAGGCCATATATAAAAAATC -3' (1R)	0.46 kb
Short (CaMBD)	EcoRI 5'- CCTC<u>GAATTC</u>GAGAAGACGGTAGAAAGAG -3' (2F) XhoI 5'- GGTC<u>CTCGAG</u>TCATTTTGCTAAATCTTCTATCAG -3' (2R)	0.3 kb
N1	EcoRI 5'- GGAC<u>GAATTC</u>AACAGGACTGATGATACCTCAG -3' (1F) XhoI 5'- GGTC<u>CTCGAG</u>TCATTTTGCTAAATCTTCTATCAG -3' (2R)	0.33 kb
C1	EcoRI 5'- CCTC<u>GAATTC</u>GAGAAGACGGTAGAAAGAG -3' (2F) XhoI 5'- GCCA<u>CTCGAG</u>TCAATAGGCCATATATAAAAAATC -3' (1R)	0.44 kb

Table 5.2.2.1 Primer sequence and amplimer sizes of CaM binding site pGEX-6p-1 constructs.

All forward primers contained the EcoRI restriction site whereas the reverse primers contained the XhoI restriction site. The restriction sites are underlined in the primers. The amplimers were double digested using these enzymes prior to ligation into pGEX-6P-1 plasmid that had been digested similarly earlier.

Restriction digests were conducted on plasmid DNA prepared from selected transformed colonies grown after ligation.
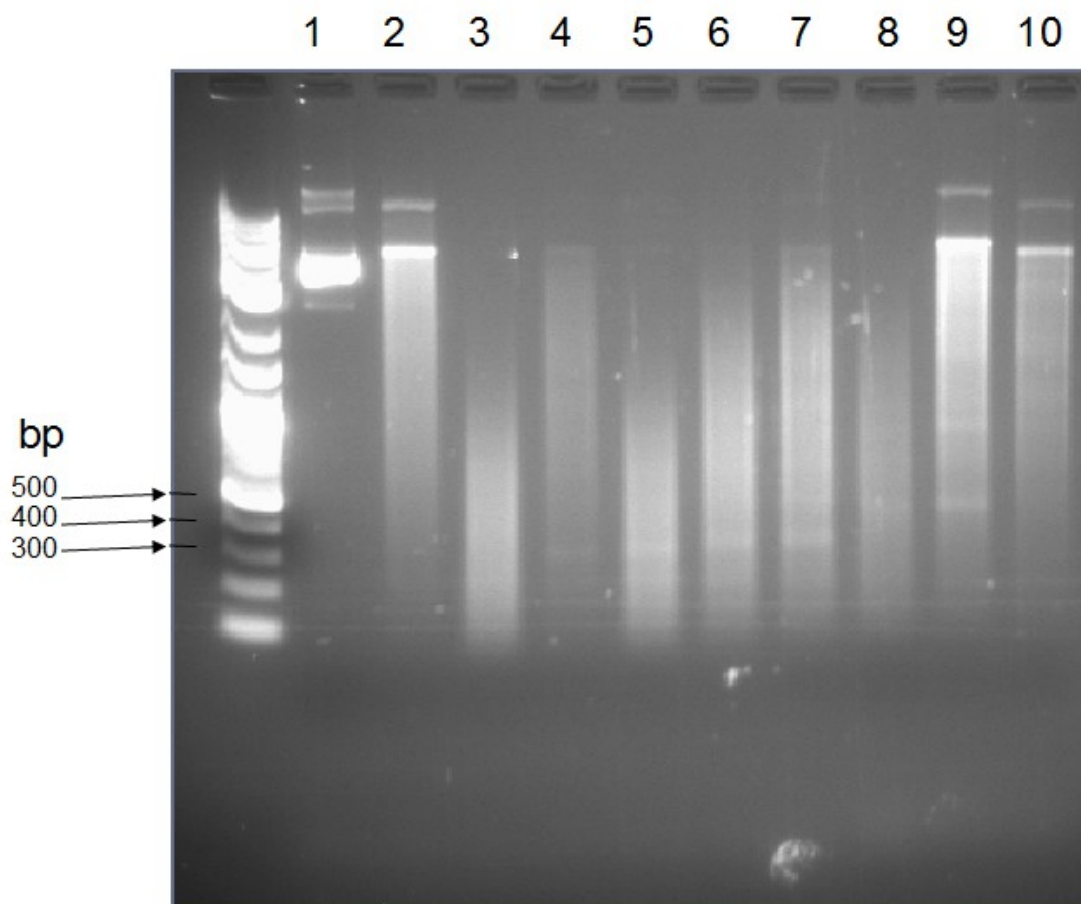
Figure 5.2.2.1 Restriction digests of CaM binding site constructs in pGEX-6P-1.

2 colonies were selected from each potential construct and double digested.

Lane 1: pGEX 6P-1 (vector reference)

Lanes 2 + 3: Full, 460 bp

Lanes 4 + 5: Short, 300 bp

Lanes 6 + 7: N1, 330 bp

Lanes 8 + 9: C1, 440 bp

Lane 10: negative control colony not containing insert

Bands of the expected size were observed for the tested colonies of the short, N1 and C1 constructs. No insert was observed for the full construct colonies.

The presence of the short (CaMBD) and N1 inserts were later confirmed by sequencing.

Expression trials were conducted for the short, N1 and C1 constructs using Rosetta (DE3) cells.
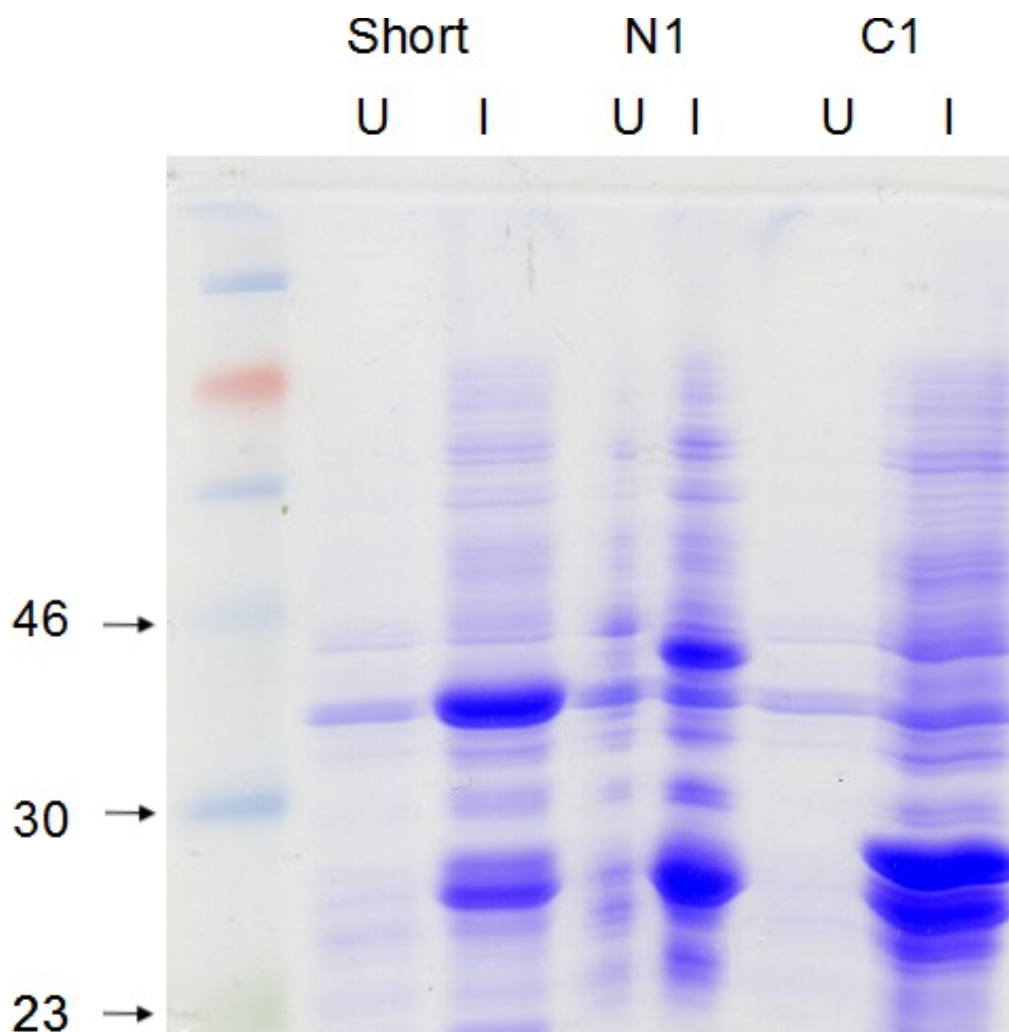
Figure 5.2.2.2 Expression trial of CaM binding constructs in pGEX-6P-1 vector in Rosetta (DE3).

Lanes are headed with appropriate construct, U indicates uninduced sample and I induced sample.

There is a clear increase in an induced band of the expected size for the short and N1 constructs. A much weaker equivalent band is seen for C1 suggesting much lower expression of this construct. The GST tag is 26kDa, it is thought that the presence of the large band at this size in all the induced lanes is GST.

Expected size of constructs including tag from pGEX-6P-1 vector:

GST-short   38 kDa

GST-N1   39 kDa

GST-C1   47.9 kDa

A western blot was conducted to analyse the expression from these samples. The blot was probed using an anti-GST antibody as this would confirm the presence of the tagged protein and possibly also the identity of the lower size band.
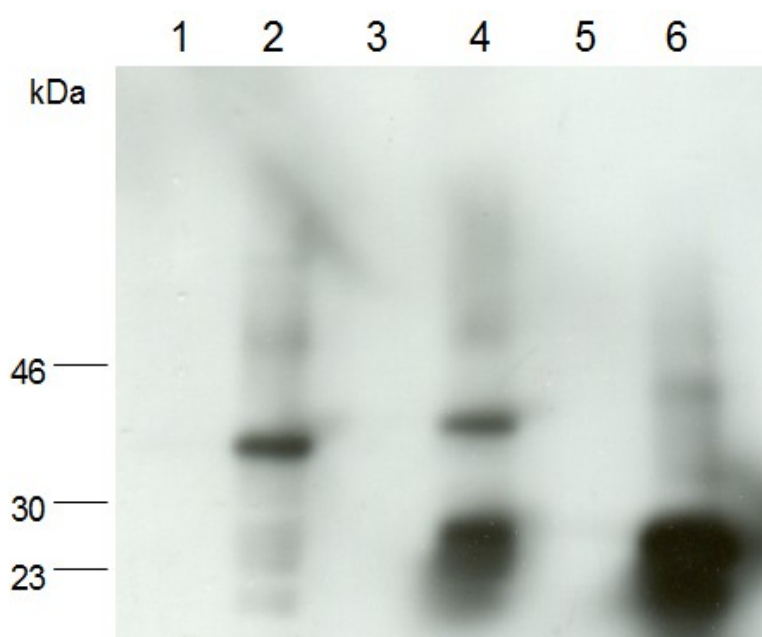
Figure 5.2.2.3 Western blot of expression trial of CaM binding site constructs probed with anti-GST antibody.

Bands of expected size are observed for all the constructs but as seen in the SDS-PAGE there is very small amount of the C1 construct. Very heavy bands are observed at the size expected of the GST tag for the N1 and C1 constructs, there is a much weaker band at this size for the short construct.

Based on the comparison of the SDS-PAGE and western blot there appeared to be slightly more short (CaMBD) protein with much less GST present. As such the CaMBD construct appeared to be a better selection for continuation into purification and downstream applications.

5.2.2.2 Purification of CaMBD recombinant protein

Post cell lysis, GST-CaMBD was first affinity purified following the manufacturer's recommended protocol buffer system for the glutathione sepharose 4B beads (GE healthcare).
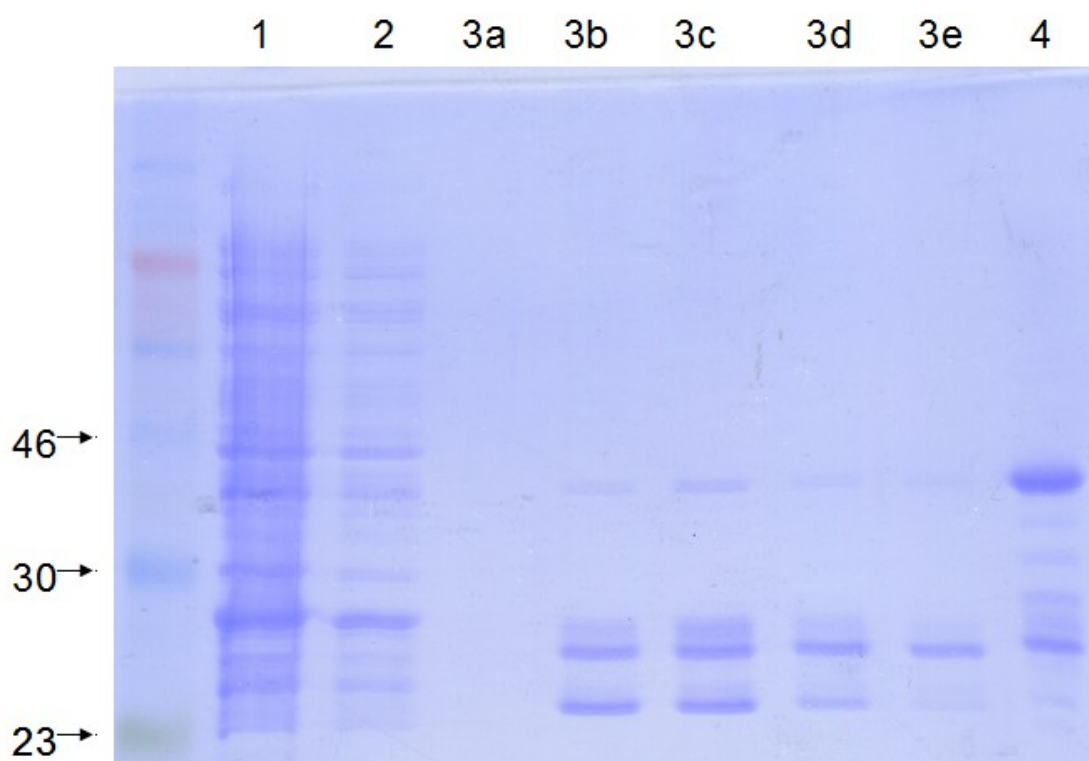
Figure 5.2.2.4 Initial affinity purification of GST-CaMBD following bead manufacturer protocol buffer system.

Expected size of GST-CaMBD, 38kDa.

Lane 1: Unbound material from beads

Lane 2: Binding buffer wash

Lanes 3a-e: 0.5 column volume elution fractions

Lane 4: Bead sample post elution step

Although some GST-CaMBD is eluted from the beads the vast majority remains on the beads. However some of the GST tag contaminant material is removed by the elution step.

The manufacturer recommended protocol uses a 10mM reduced glutathione in TRIS at pH 8 elution step, this glutathione concentration was found to be insufficient to elute the GST-CaMBD. Through a series of trials, modifications were made to the recommended protocol. This lead to the buffer system described in chapter 2. It was found that a 5mM reduced glutathione wash step in phosphate buffer at pH 7 was sufficient to remove a large quantity of the contaminant GST tag material without eluting the GST-CaMBD protein. A higher concentration of reduced glutathione could have been used in this buffer without eluting the GST-CaMBD but this was found not to be necessary. The

transfer to TRIS buffer at pH 8 was required to elute the GST-CaMBD, 20mM reduced glutathione was found to be sufficient for this purpose.
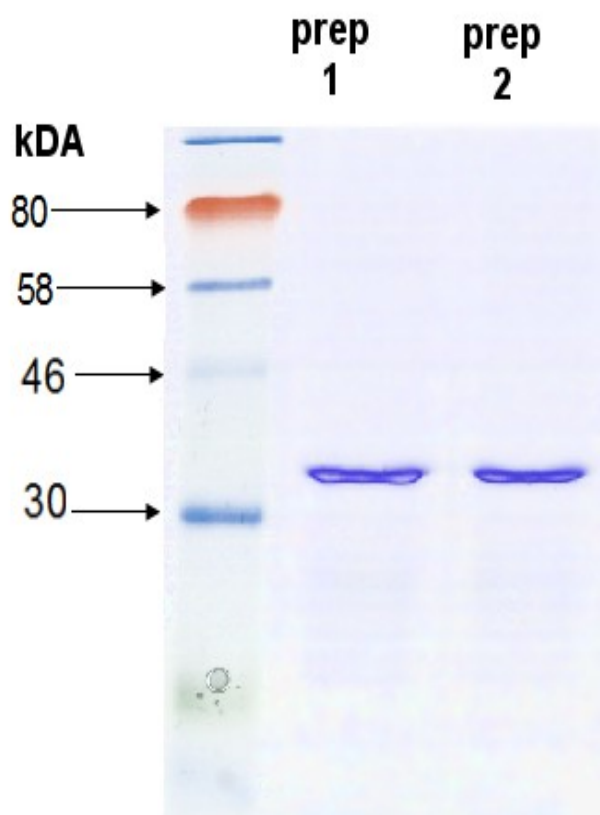
Figure 5.2.2.5 Affinity purified GST-CaMBD following buffer system listed in chapter 2.

Samples from 2 preparations of GST-CaMBD using buffer system described above. Using this protocol high purity GST tagged CaMBD protein could be prepared. This material could be stored at 4°C for 1 week prior to further usage.

The pGEX-6P-1 vector contains a 3C protease site allowing for the cleavage of the GST tag. 3C protease was found to have sufficient activity when used in combination with the dialysis allowing a similar protocol to that used for the CaMLD constructs to be adopted. The 2nd pass through glutathione sepharose beads was not as successful for CaMBD as the equivalent step for the CaMLD constructs. It proved to be not possible to practically remove a large amount of the cleaved GST tag which remained in the CaMBD sample. Similar aggregation problems to cleaved CaMLD constructs were observed for cleaved CaMBD. As with CaMLD, cleaved CaMBD had a storage life of <48 hrs.

5.2.3 Cloning and expression of CaM construct

A human CaM clone was obtained in the plasmid pAED4 (gift from Z. Grabarek). This plasmid was transformed into Rosetta (DE3) cells and expression trials conducted. The growth and early stages of purification were followed as described by Tan (Tan, 1996). Although CaM protein was produced the reported levels could not be matched. Owing to the successful use of the pETMM vectors with the CaMLD constructs, CaM was sub-cloned into the pETMM60 vector. The 60-CaM plasmid was transformed into Rosetta (DE3) and expression trials performed.
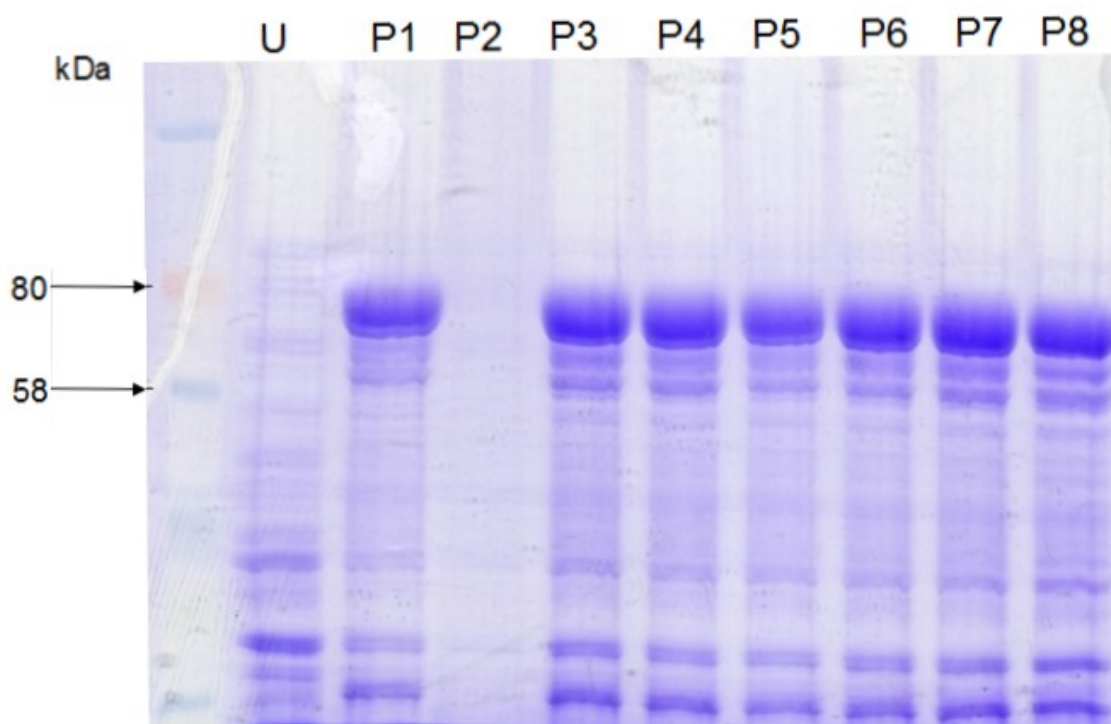

Figure 5.2.3.1 Expression trial of 60-CaM.

60-CaM expected size 76kDa

Lane labelled U is uninduced pellet sample. Lanes P1-P8 are samples from induced cell pellets, induced expression can be observed for all these lanes except P2. Pellets P1, P3-P8 were stored for in preparation for purification.

60-CaM was purified following the CaM specific protocols listed in chapter 2.
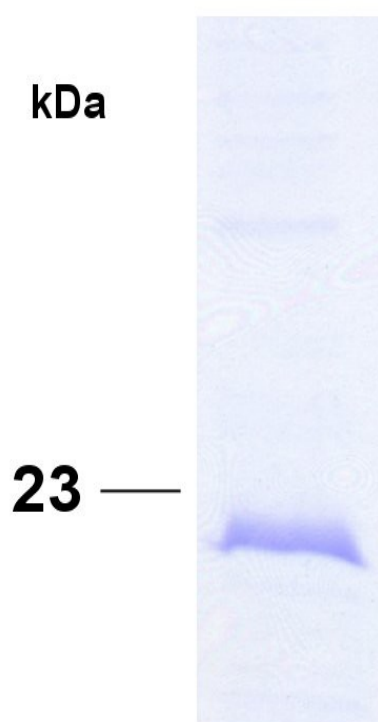
Figure 5.2.3.2 CaM protein sample post TEV cleavage and 2ⁿᵈ pass through Ni beads.

The SDS-PAGE illustrates the purity to which CaM protein samples could be prepared. This material could be stored at 4°C for 1 week prior to further usage.

5.3 Crystallography Trials

Crystallisation trials of were conducted for CaMBD. The protein used in this trial was concentrated to 8mg/ml after a gel filtration purification step following affinity purification. Owing to the small amount of protein that could be produced in this concentration range only 4 crystallisation screens could be trialled.

200 nl protein with 200 nl screen condition sitting drops were set up in 96 well Intelli-Plate (Art Robbins) using an Art Robbins Phoenix crystallisation robot (Alpha Biotech UK). The PACT (pH, anion and cation crystal trial testing pH within PEG/ion), JCSG (PEG and salt conditions devised for Joint Centre for Structural Genomics), and CSS (clear strategy screen) I and II screens (all Molecular Dimensions) were used. Plates were observed on a regular basis for the formation of aggregates and any potential crystals.
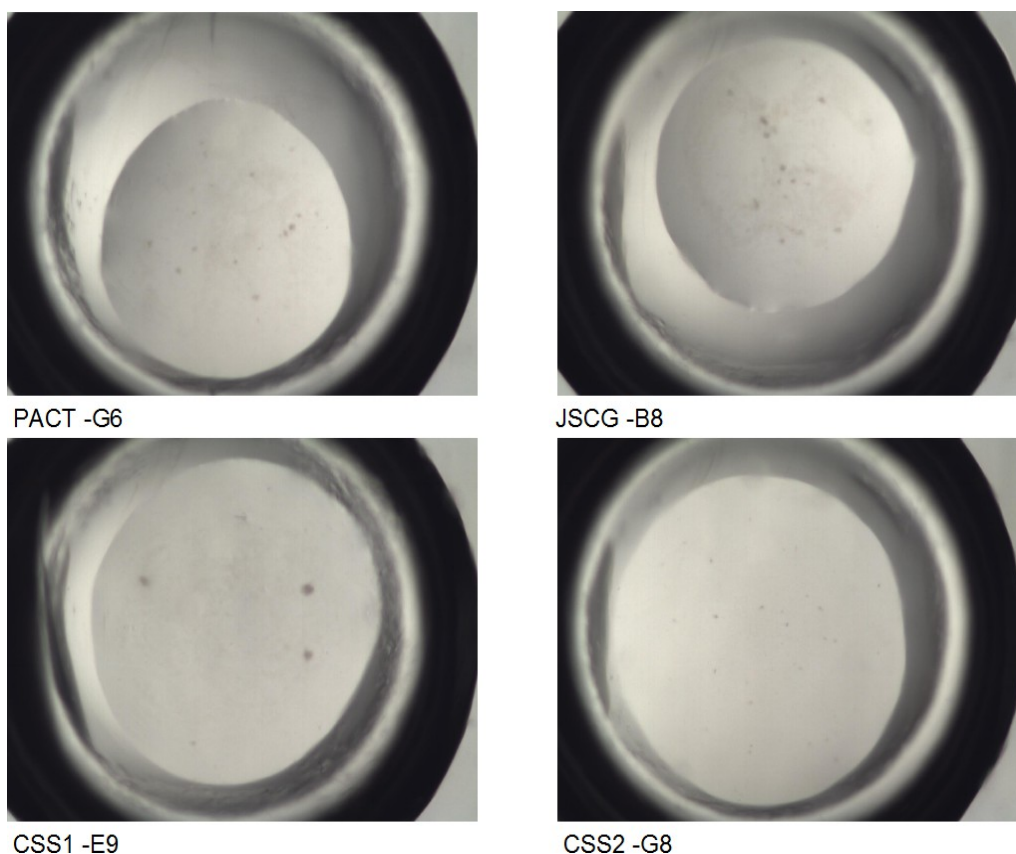
Figure 5.3.1   Selection of trialled screen conditions from CaMBD crystallisation trial.

A few conditions in both screens tested had silt like residues develop in the plate wells. The presence of silt like residues may be a result of protein and/or salt aggregation. As the screens tested are sparse matrix screens so a wide range of possible conditions are tested within one screen. As such wells with silt like residues are considered for further trials using finer screening conditions based around the original condition. No obvious crystal formation was observed in the screens trialled.

5.4 Discussion

As discussed in the introduction to this chapter the aim of this section of the project was generate the protein tools required to examine the binding between the CaMLD and CaMBD regions of RyR2, and also the involvement of CaM within this interaction. Previous constructs of the regions of the RyR domains have been described in the literature. This project designed constructs in the equivalent areas of RyR2 based around the reported regions in RyR1.

In the case of the CaMLD region the identification of a CaM like domain (Xiong, 2006) was used to locate an equivalent region in RyR2. This strategy has subsequently been reported in the literature where an equivalent region in RyR2 (4020-4166) was described (Gangopadhyay, 2011). The equivalent construct designed in this project had amino acid coordinates 3961-4163. While having a similar C terminal, more residues are included at the N terminal as exon boundaries were considered beyond a straight match to the RyR1 reported region. The selected CaMLD construct of this project also included the RIH associated domain which has been suggested may play a role in domain interaction in RyR (Ponting 2000).

pET15bmod was selected as an initial expression vector for the CaMLD based constructs as only an N terminal 6His tag with a short linker was expressed with the recombinant protein. This strategy was adopted with a view towards crystallography work so that as little as possible extra amino acids were included with the protein of interest while still allowing for an affinity purification step. In addition short (<5 residue) linker regions between co-expressed tags have been identified as increasing the likelihood of crystal formation as the potential movement of the tag is assumed to be limited relative to the protein of interest so reducing disorder within the forming crystal structure (Smyth et al. 2003). Unfortunately there were major solubility issues with the expressed protein using this vector. The addition of detergent to the lysis buffer did increase the amount of soluble protein this approach was not followed for 2 major reasons. Firstly, as detergent would need to be present in all downstream buffers to maintain solubility this would almost certainly hinder downstream experiments and would also severely limit the number of crystallisation screens that could be used. Secondly, while the loss of protein is expected with every additional handling step when working with the sample, the yield obtained by the time the protein was sufficiently pure was so low that it would have been impractical to continue with this approach.

Based on the solubility issues observed with the CaMLD constructs it was decided that co-expressing a solubility tag may offer a better solution than additives to the lysis buffer. The pETMM series of vectors offered a dual tag system consisting of a range of proven solubility tags with a 6His tag for affinity purification which could be removed using the included protease site located between the dual tag and the insert. This series

of vectors also shared the same MCS allowing a range of solubility tags to be trialled using the same cloning strategy (Dümmler et al. 2005). As it was unknown if any of the solubility tags would be beneficial; only 2 constructs were used to increase the speed of the trials. As discussed in the chapter introduction these constructs were CIR and CaMLD as it was felt that these offered most further experimental options. The CaMLD construct when joined with the NusA tag proved to be the most successful combination tested. With optimisation to the purification protocol it proved possible to produced high purity CaMLD protein though obviously serious stability issues remained once the NusA tag had been removed. From the observations in this chapter it appears this region of RyR2 is unstable. This instability may well be increased further when trying to make larger pieces of the I domain beyond the strictly identified CaM like domain of RyR1. Even production of this piece resulted in expression predominantly in inclusion bodies and had to utilise a refolding step (Liangwen Xiong et al. 2006). Through conversations with colleagues working on the I domain section 3722-4353 (George et al. 2004a) this region was found to be particularly unstable and difficult to produce even in small quantities. As such this CaMLD construct may be one of the first soluble expressed proteins in this region produced above small scale quantities.

Although the CaM binding domain of RyR2 has been identified (Naohiro Yamaguchi et al. 2003) it was decided to attempt incorporate sections of the N and C terminal flanking regions associated with CaM concentration response in the different isoforms of RyR. These regions were described in relation to CaMBD in RyR1 (Yamaguchi et al. 2004), and it was from these coordinates that an equivalent region was identified in RyR2. The inclusion of the N and C terminal flanking regions would allow potential testing of CaM and Ca^{2+} concentration responses. However the primary aim of the CaM binding domain constructs was to produce stable protein that contained the α helical tube structure as identified in PDB entries such as 2bcx that would allow for CaM and, presumably, CaMLD binding to this region.

Based on assessment of the SDS-PAGE (figure 5.2.2.2) and accompanying western blot (figure 5.2.2.3) the short (CaMBD) and N1 constructs both appeared to be expressed at similar levels. However the CaMBD construct had a much weaker signal at the band size linked to GST indicating a much lower concentration of GST in this sample compared to the N1 construct. As observed in both images the C1 construct is present at

low levels so was rejected for further use. Owing to the large amount of GST material seen for the N1 and C1 constructs it is considered that these constructs may be less stable than the CaMBD construct. If these constructs were less stable and degrading during expression then large amounts of GST compared to the construct protein would be observed. The CaMBD construct was selected owing to this apparent greater stability. It would appear that trying to express the CaM binding site with either flanking region attached reduces the stability of the produced protein. It is difficult to compare against the literature as whole RyR was expressed in HEK cells in the papers used during construct design (Naohiro Yamaguchi et al. 2003), (Yamaguchi et al. 2004). Most other experiments involving the CaM binding domain reported in the literature commonly use a synthesised peptide of shorter length than the CaMBD construct.

CaM is a well characterised protein and can be easily commercially obtained. For the potential scale of protein required by this project it was decided to recombinantly produce the protein. Following the accompanying protocol supplied with the donor plasmid as described by Tan (Tan et al. 1996) could not be repeated to generate sufficient CaM. As optimisation conditions for use of the pETMM60 vector had already been established so this allowed for larger scale production of CaM. The tag was successfully removed from the fusion protein and high purity samples generated. This approach allowed production of relatively pure CaM protein above the levels required for the project.

Crystallography trials were conducted for GST-CaMBD however they provided inconclusive results. The presence of silt like residues in some of the wells indicated that some conditions may be worth pursuing in further trials. As expected with crystallography based work a much greater number of trials involving a larger range of crystallisation screens is required for progression in this area. A significantly large amount of protein would have been required to enable these trials to be conducted. This would represent a major investment in time and resources with the possibility of little advancement of the project. These concerns would be amplified when considering CaMLD owing to the greater instability of the cleaved purified protein that would be required. While there are examples citing successful crystallography projects in which crystals have been grown with a solubility tag still attached to the protein, these are rare and usually involve a short linker between the tag and expressed protein (Smyth et al.

2003). Also these examples cite use of MBP, Trx and GST tags, use of NusA is commonly limited to that of purely a solubility tag which is removed as part of the purification process. This approach was adopted in the trial with CaMBD, it was hoped that the presence of the GST tag would not prove detrimental to the crystallisation process while being beneficial in aiding protein stability and allowing a purer sample to be used. Although a sufficiently pure sample of CaMLD with the NusA tag removed can be produced there were serious concerns about the stability of the cleaved protein. While not trialled the initial expectation was that the CaMLD protein would rapidly aggregate out of solution prior to sufficient time for crystal formation.

Further time could have been spent within this project with the aim to produce more stable and higher quality protein. However, as discussed, for a number of reasons this would be very difficult and may not be practically achievable. A compromise between optimised protein quality and sample suitability for further work had to be reached to allow the project to continue. As the aim of this section of the project was to generate protein to investigate domain interaction it was felt that this point had been reached with the work described in this chapter.

Chapter 6:

Functional Studies of the Calmodulin Like Domain of Human RyR2

6.1 Introduction

The primary aim of this chapter was to examine the interaction between the RyR2 domains of CaMBD and CaMLD. As previously discussed there are examples of this interaction in the literature and the PDB contains structures of CaM bound to CaMBD. This project aims to further understand the interaction between CaMLD, CaMBD and CaM, and gain insight into any potential mechanism leading between the transition of binding of CaM and CaMLD to CaMBD. The role of calcium within this interaction mechanism is believed to play a key role and, as such, is a studied factor. In addition, selected known CPVT mutations within CaMLD have been studied to see if further knowledge can be gained at a molecular level towards why these mutations lead to arrhythmia.

Two main types of analytical technique were employed to conduct this investigation. Firstly the interaction between the domains was examined using pull down assays. This technique has been successfully used by colleagues investigating other potential binding and interaction sites within RyR (Zissimopoulos et al. 2006), (Zissimopoulos et al. 2012b). This established technique offered a means of testing the interaction between the recombinant protein constructs of RyR2 used in this project.

The second technique employed was circular dichroism (CD). Initially this technique was approached to see if folded recombinant protein had been produced. Further examination in the far UV region (170-260nm) would allow assessment of the secondary structure of the produced protein. If sufficient data was collected then the specific percentage components of secondary structure could be assigned (Kelly et al. 2005), (Greenfield 2006). A repeat of the change observed in secondary structure composition by the addition of Ca^{2+} to wt CaMLD (Liangwen Xiong et al. 2006) could be conducted and extended to examination of the CPVT mutant forms of CaMLD.

The rearrangement of secondary structure elements within CaMLD would expected to be accompanied by a conformational change within the protein which may not necessarily observed in the far UV region. The near UV region (250-350nm) is commonly used to look at changes in 3ry structure through comparison of spectra

recorded under different conditions. This allows analysis of the impact of variations in reagent concentration and the presence of binding partners by examination of divergence in the obtained spectra. The spectra are generated from a combination of the signals given by the aromatic amino acid content of the studied protein. Tryptophan is the main contributor to the signal with a peak close to 290nm. Tyrosine residues provide a smaller addition peaking between 275-282nm. Phenylalanine has a weak contribution between 250-270nm. Disulfide bonds may give very weak contributions throughout the entire wavelength range. The differences observed in spectra under varying conditions result from the changing of conditions in the local environment of the aromatic residues altering the strength of signal given by the affected amino acid. This means that alterations in 3ry structure which do not affect aromatic residues are unlikely to be observed in the near UV. The signal strength in the near UV region is weaker than the far UV requiring more concentrated protein samples to compensate for the drop in signal. However the near UV can be more sensitive for detecting subtle differences in conformation that would not always be seen by far UV (Kelly et al. 2005).

As stated the aromatic acid composition of the studied protein plays a key role in the spectra obtained in the near UV region. Trytophan and tyrosine are the main contributors to the signal though the absence of tryptophan residues usually indicates that a very weak signal will be obtained from a protein. As such the amino acid composition of the studied proteins by number of residues and overall percentage component has to be considered.

CaMLD:

Trp (W)	3	0.9%
Tyr (Y)	11	3.3%

CaMBD:

Trp (W)	2	2.0%
Tyr (Y)	4	3.9%

CaM:

Trp (W)	0	0%
Tyr (Y)	2	1.3%

Based on the analysis of amino acid content, both CaMLD and CaMBD are potential good candidates for examination using near UV however CaM may provide a weak signal.

6.2 Results

6.2.1 Chemical denaturation

Chemical denaturation studies were conducted on wt and mutant CaMLD protein that had been affinity purified, and the NusA tag cleaved and removed. The unfolding state of protein was monitored by fluorescence spectroscopy while the sample was denatured by addition of guanidine hydrochloride (GndHCl).
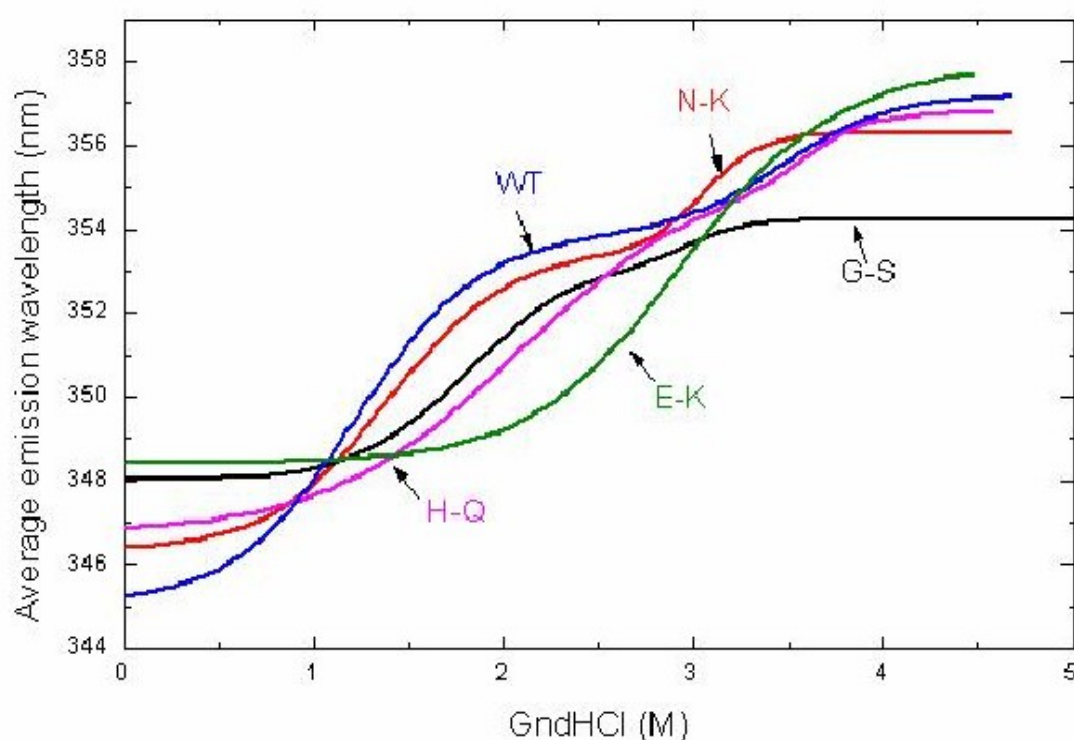

Figure 6.2.1.1 Weighted average emission wavelength versus the concentration of guanidine hydrochloride (GndHCl), plots for wt and mutant CaMLD proteins.

	ΔG^{N-I} (cal/mol)	[GndHCl] 50%	ΔG^{I-D} (cal/mol)	[GndHCl] 50%	ΔG^{N-D} (cal/mol)	
WT	-2214	1.21	-7207	3.46		3 state
HQ	-2416	2.07	-13511	3.63		3 state
NK	-2597	1.38	-12200	3.05		3 state
GS	-3695	1.80	-11960	2.99		3 state
EK					-2874	2 state

Table 6.2.1.1 Free energy (ΔG) of unfolding states of wt and mutant CaMLD proteins.

The information from the denaturation was analysed and a 2 or 3 state denaturation process accordingly fitted to the data. The shape of the plots (figure 6.2.1.1) is indicative of this 2 or 3 state denaturation process. The ΔG (calories/mole) values required for the transfer from native (N) to intermediate (I) and denatured (D) states are shown in the above table. The three mutations, G3946S, N4104K and H4108Q, increased the wt stability with respect to an intermediate state. On the other hand the E4076K mutation severely affected the stability profile of the CaMLD construct and decreased its stability with respect to the fully unfolded state.

6.2.2 Pull Down Assays

The majority of protein used in the pull down assays was affinity purified samples which had the tags attached. Limited trials with cleaved tag material using SDS-PAGE for detection were conducted. It was found that smaller amounts of protein could be used with better detection rates when using the more sensitive western blot technique for detection of potential interaction. For the western blots to work, with the exception of CaM, it was only possible to probe against the co-expressed tags of the CaMLD and CaMBD constructs. Using the tagged protein reduced some of the issues surrounding the stability of the protein samples though the western blot technique was sensitive enough to detect degradation in some of the samples. Pull down assays were performed

following the protocol described in chapter 2.

6.2.2.1 CaM agarose bead pull down assay

The first set of pull down assays performed examined in there as was any potential interaction between CaM and CaMLD, and, if so, did the CPVT mutations affect this interaction. For the purposes of this experiment CaM agarose beads were used.
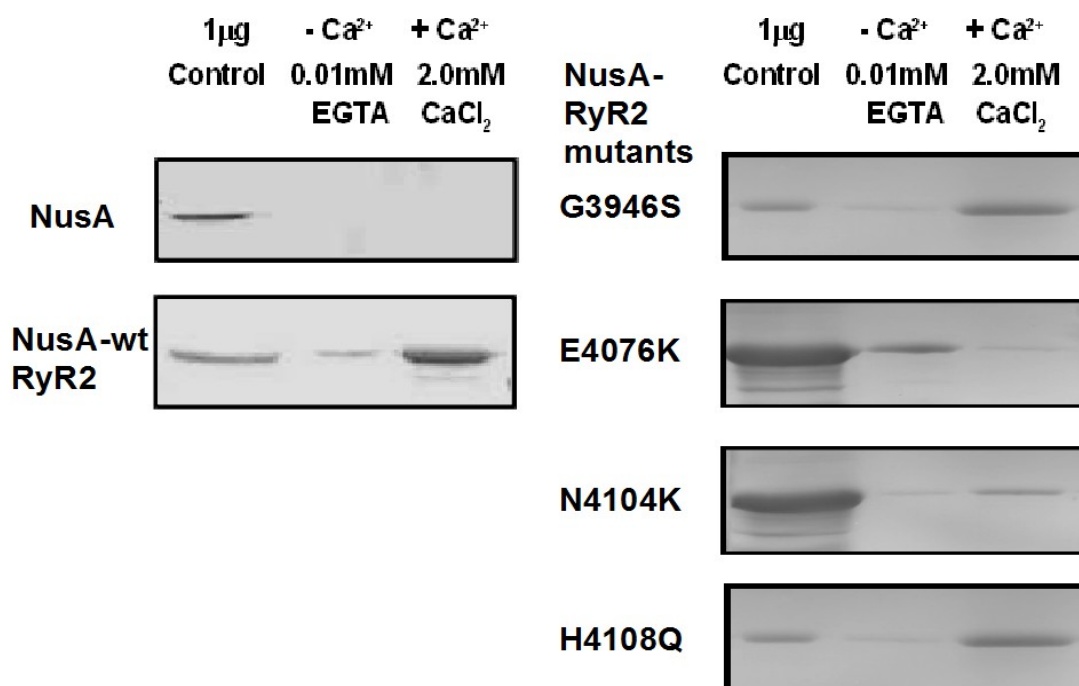
Figure 6.2.2.1 Western blots of CaM agarose bead pull down assay with wt and mutant CaMLD.

Western blot probed with anti-NusA antibody.

Control sample present in left lane to show detection of sample. Without (middle) and with (right) Ca^{2+} pull down assay samples.

wt RyR2 CaMLD construct and its mutants except E4076K displayed CaM binding in Ca^{2+} dependent manner. NusA fusion tag, which served as a negative control for this experiment, did not display CaM binding.

The binding of the CaMBD construct with CaM agarose beads was also tested, weak Ca^{2+} independent interaction was observed. Within this experiment the binding of wt CaMLD was also re-tested, CaMLD displayed the same interaction profile as seen in the

above figure showing stronger interaction than CaMBD.

6.2.2.2 Pull down assays using CaMLD, CaMBD & CaM constructs

As recombinant protein could be produced for the potential interactions to be tested so the CaM agarose beads were no further used. This decision was also because bovine CaM was used for the beads whereas all human versions of RyR2 had been recombinantly made allowing for the examination of human-human versions of the proteins.

Initially the recombinant purified CaM protein was tested for any potential unwanted interactions with either set of the affinity beads that could be used to immobilise the bait protein.
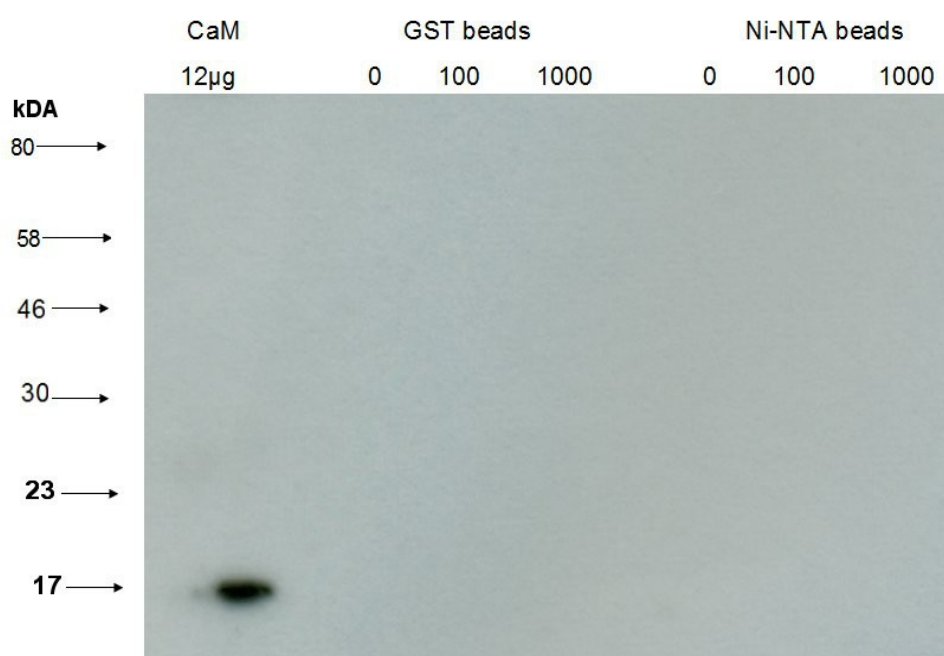
Figure 6.2.2.2 Western blot of check of CaM interaction with affinity beads.

Membrane probed with anti-CaM antibody. Left hand contains detection control sample of CaM protein. CaM protein incubated with listed affinity bead in assay conditions containing shown concentration (μM) of calcium.

As the CaM protein did not display any interaction with either set of potential affinity beads the decision to use tagged protein material in the assays initially appeared to offer a range options regarding towards bait and capture protein variations. However control condition testing revealed a number of unwanted interactions between the affinity

media, used to immobilise the bait protein, and potential capture sample.
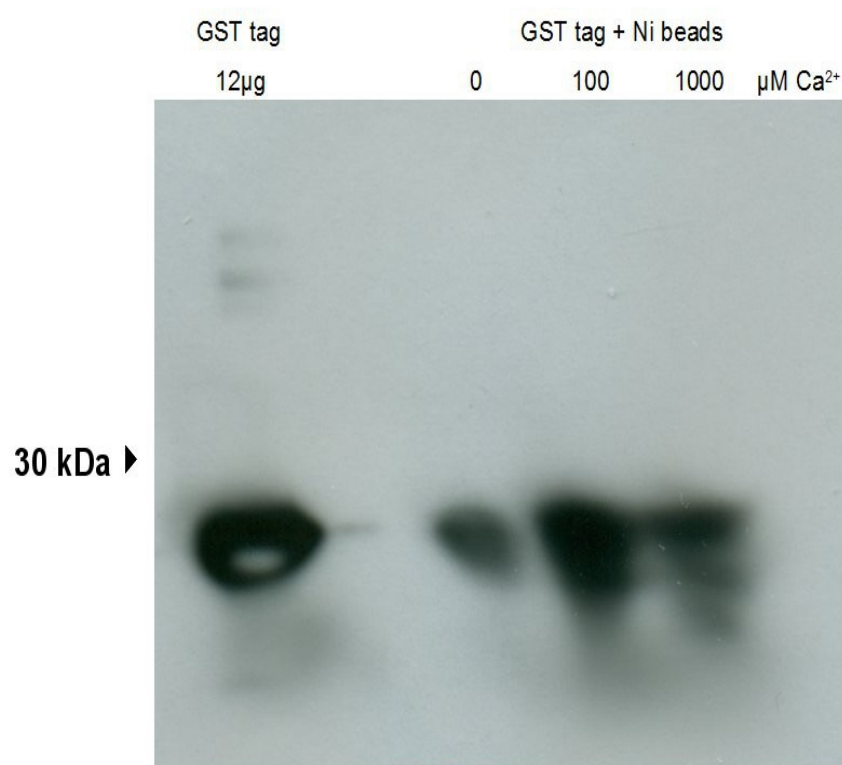
Figure 6.2.2.3 Western blot of GST tag incubated with Ni affinity beads under increasing Ca^{2+} concentrations.

Membrane probed with anti-GST. Left hand lanes contains detection control sample of GST tag. Signal detected for presence of GST across Ca^{2+} range.

The apparent binding of the GST tag to Ni affinity beads meant that Ni beads could not be used to immobilise CaMLD protein to test binding with CaMBD. Glutathione sepharose beads were then tested with CaMLD, this was with the aim of immobilising CaMBD protein.
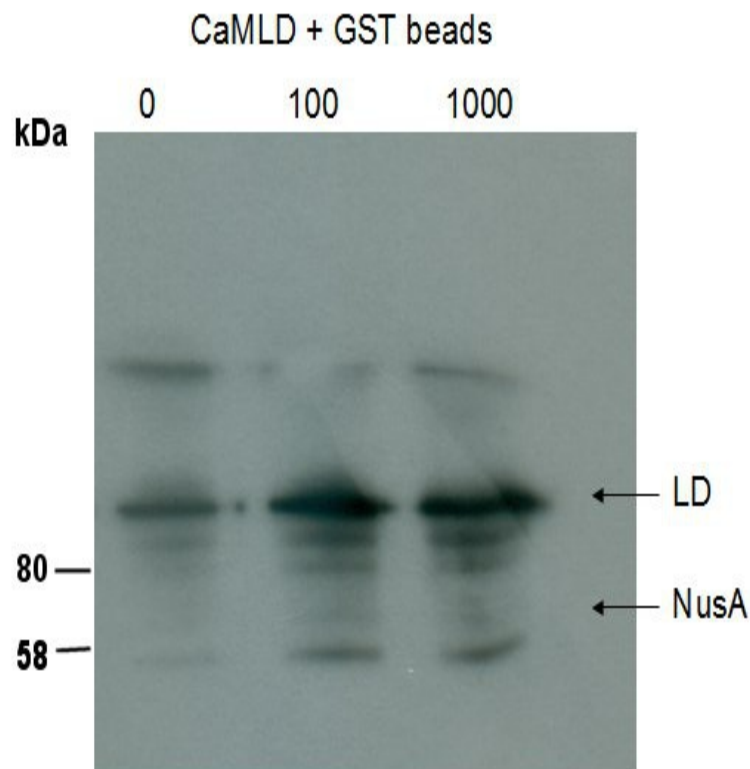
Figure 6.2.2.4 Western blot of CaMLD incubated with glutathione sepharose affinity beads under increasing Ca^{2+} concentrations.

Identity of bands based on comparison with size markers on companion gel stained with Coomassie R250.

As CaMLD appeared to bind with glutathione sepharose affinity beads this severely limited the range of experiments that could be conducted using the pull down assay technique. To bypass the issues related the affinity beads cross linking experiments were run. The interaction reaction was run in the presence of a cross linking agent, glutaraldehyde. The cross linking reaction was stopped prior to introducing affinity beads to pull down one of the reaction components. The detection of cross linked products was conducted by western blot analysis as previously.

As the interaction of the CaM binding site with CaM has been reported in the literature so the binding of CaMBD with CaM was selected to test the cross linking technique.
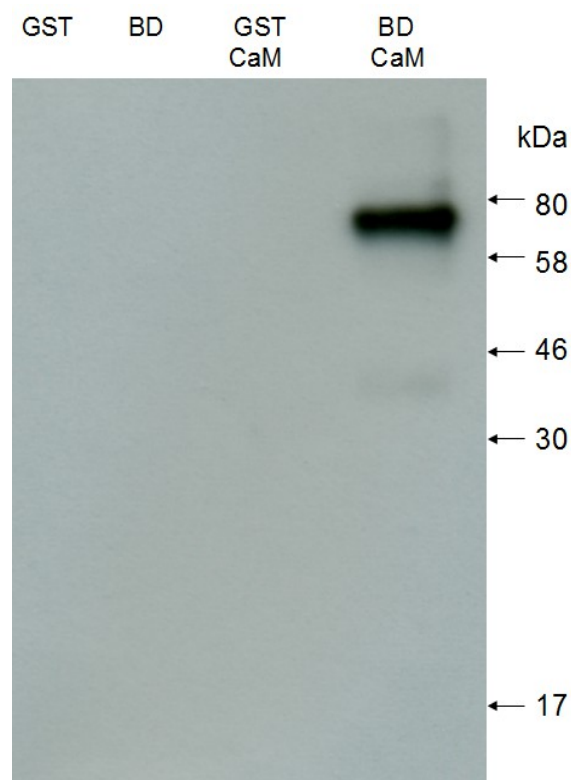
Figure 6.2.2.5 Western blot of cross linking trial of CaMBD with CaM.

Membrane probed with anti-CaM antibody. GST-CaMBD incubated with CaM reaction sample in right hand lane. GST and BD included as negative control samples in their respective lanes. GST tag incubated with CaM as further negative control.

Expected sizes of proteins:

GST   26kDa

BD   40kDa

CaM   17kDa

BD + CaM   57kDa

Bands were expected for CaM (17kDa) in the GST tag with CaM lane, and BD + CaM (57kDa) and possibly also CaM in the CaMBD with CaM lane. As the expected unbound CaM signal from the GST tag incubated with CaM reaction was not detected so the reaction mixtures were repeated and timed samples collected.
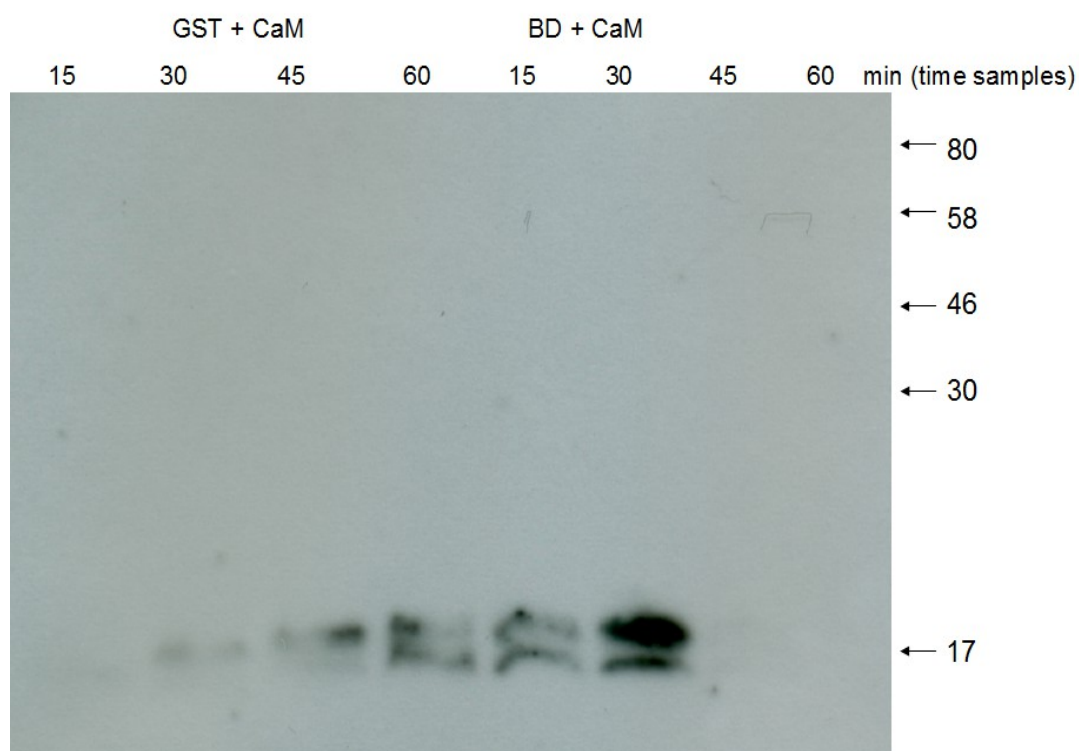
Figure 6.2.2.6 Western blot of repeat cross linking of CaMBD with CaM.

Membrane probed with anti-CaM antibody. GST tag with CaM incubation run as negative control to GST-CaMBD with CaM test condition. Timed samples were taken from both incubation reaction mixtures.

An increasing strength CaM signal is detected for the GST tag with CaM reaction. A similar signal is observed for the CaMBD with CaM incubation but only for the first 2 samples taken.

Repeats of the cross linking experiments were conducted with further inconclusive results.

6.2.3 Circular Dichroism

Some early CD data was collected using cleaved tagged material that had been further purified by gel filtration however the majority of data was collected using tagged protein. All samples collected in the far UV range underwent dialysis into CD suitable buffer prior to data collection. Most samples required concentration so that sufficient detectable signal was generated from the sample. As a precautionary measure all samples were briefly centrifuged prior to data collection to remove any aggregates from

solution. While samples were monitored during data collection owing to the time taken to collect the spectra there was degradation of samples during this process which adversely affected the quality of the recorded spectra. Baseline data was recorded for the buffer used for each data set which was subsequently subtracted from protein spectra during analysis. It should be noted that data was collected on 2 CD instruments. Data sets collected on the monochromator system on beamline 23 could be processed in the accompanying Globalworks software (OLIS) which allowed for more complex manipulation of the raw data during analysis. Data collected on the Chirascan spectrometer (Applied Photophysics) had to processed in spreadsheet software owing to the data output format.
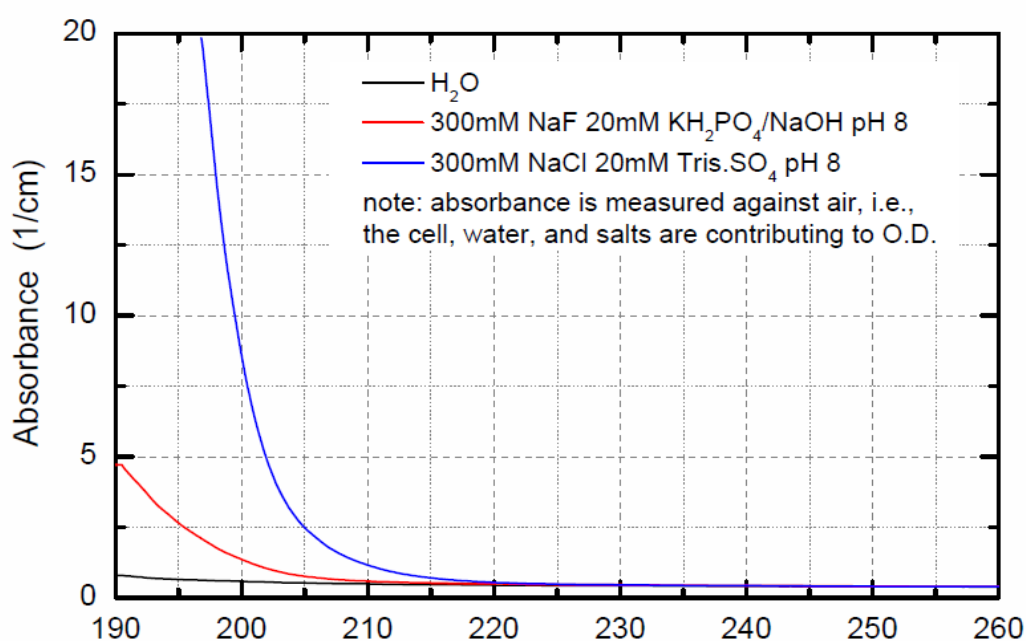
Figure 6.2.3.1 Comparison of UV absorbance of purification buffer against buffer system designed for CD data collection.

Black trace for H_2O. Blue trace for buffer system used for latter stage of purification. Red trace for CD buffer system. CD buffer has much lower absorbance than purification buffer below 210nm, this difference is greatly amplified as the wavelength shortens. However CD buffer still has noticeably higher absorbance than H_2O reference.

6.2.3.1 Far UV region (170-260nm)

Some data in this region was obtained for early batches of CaMLD construct protein. The data were of poor quality and it was not possible to collect lower wavelength data however the protein signal appeared to be influenced by more than one secondary structure element. The further work conducted on CD buffer and protein sample quality optimisation allowed use of this technique for further investigation of the RyR2 domains. Despite these optimisation steps, although the spectrometer systems were capable of collecting data within the specified range, it was not possible to record viable data in the shorter wavelengths owing to protein sample and buffer constraints.

The spectra of cleaved protein samples, so there was no tag influence on the data collected, were recorded for CaMBD, and wt and EK CaMLD.

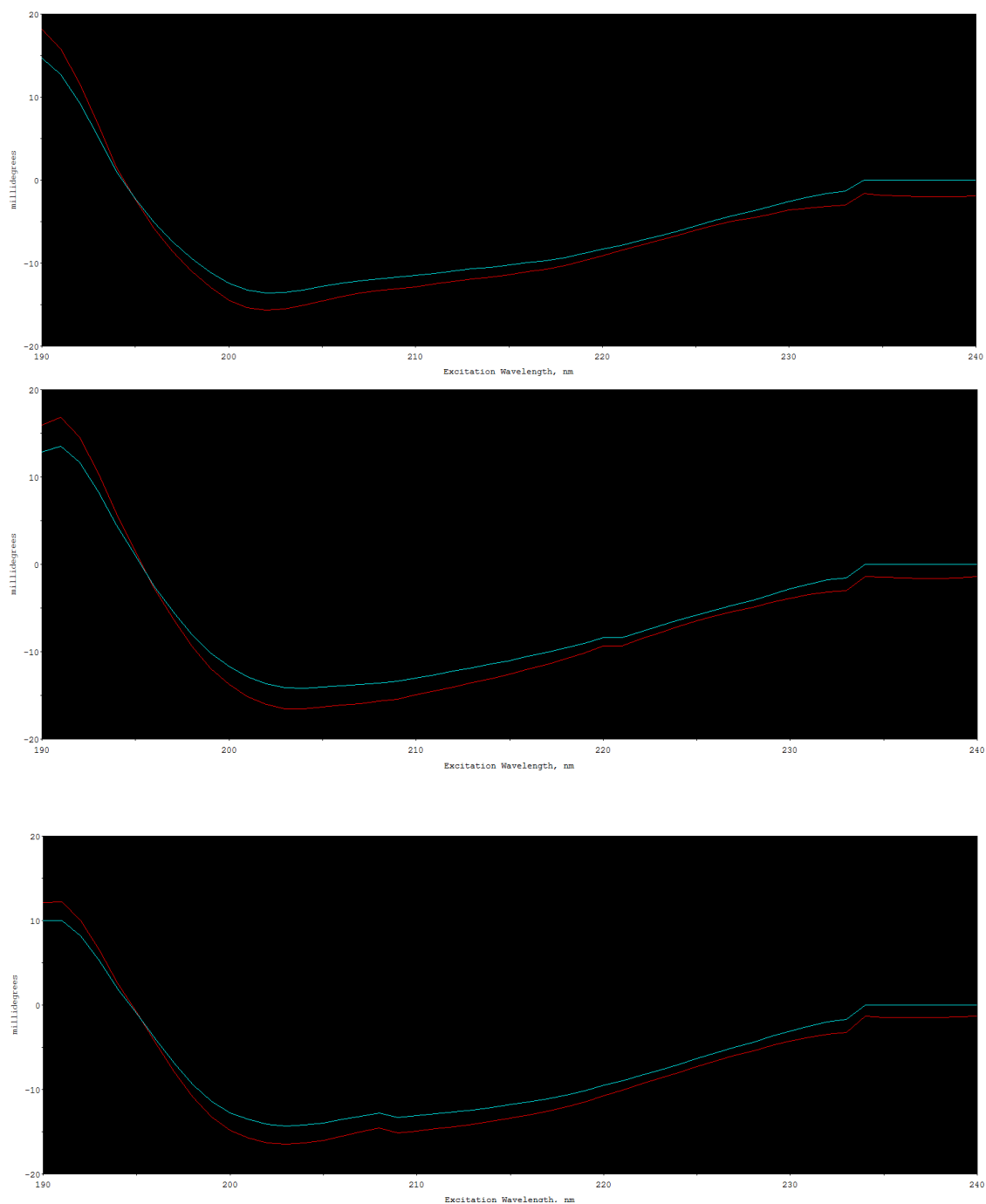**Figure 6.2.3.2 Far UV range data collected for CaMLD (top), E4076K (middle) and CaMBD (lower).**

secondary structure fit analysis fitted with CDSSTR algorithm using the SP43 reference set.

Red spectrum are averaged experimental data with buffer baseline subtracted.

Turquoise spectrum are fitted data for secondary structure prediction showing deviation required for fit against experimental data.

	CaMLD		E4076K		CaMBD	
	Predicted	Experimental	Predicted	Experimental	Predicted	Experimental
α helix	42.9%	52%	42.3%	46%	56.9%	45%
β sheet	19.5%	22%	19.5%	18%	12.8%	18%
other	37.6%	26%	38.2%	36%	30.3%	37%

Table 6.2.3.1 Predicted and experimentally derived secondary structure for CaMLD, E4076K and CaMBD

Experimentally derived secondary structure obtained from fit analysis generated in above figure.

Other category in table includes turns and random coils, category outputs differ between predictive software and CD spectral analysis software.

Although there is a difference between the experimental spectra and the fitted data this is mainly in the signal strength (given in millidegrees). The key component of the spectra, the shape of the curve which is defined by the secondary structure elements of the sample, is closely mirrored in the fitted structure. The E4076K CaMLD mutant protein has a close match between the predicted secondary structure and the experimentally derived values. CaMBD and wt CaMLD show variance between these two sets of values.

As the chamber holding the sample in the monochromator could be temperature regulated so it was possible to conduct thermal denaturation studies of the protein samples. The temperature at which the protein lost its secondary structure characteristics was considered to be the point at which the sample had denatured. Samples were heated to 90°C in 5°C increments from 10°C and returned to 20°C to observe if they was a return to a folded state with cooling. Tag cleaved CaMBD was initially tested.
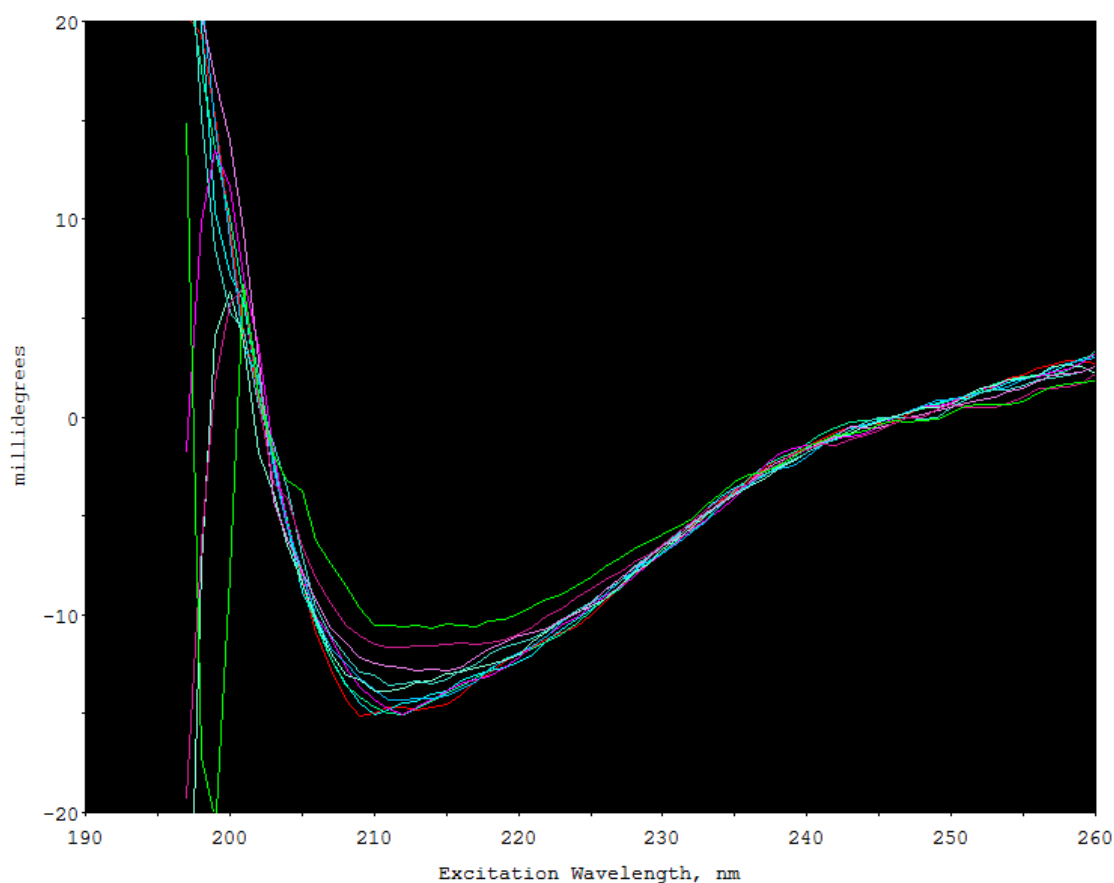
Figure 6.2.3.3 Thermal denaturation spectra series for CaMBD.

Spectra from 10°C to 55°C shown. Change in spectra observed for 45°C (light pink), 50°C (dark pink) and 55°C (light green). These spectra are 3 weakest (highest points) at 210nm and noticeably dip down much earlier at 200nm.

The change in the shape of the spectra for CaMBD between 45°C and 55°C indicates a decrease in secondary structure content as the spectra becomes flatter when compared to lower temperature spectra. For the purposes of comparison the thermal denaturation series for CaMBD that had been incubated with E4076K CaMLD was recorded.

212
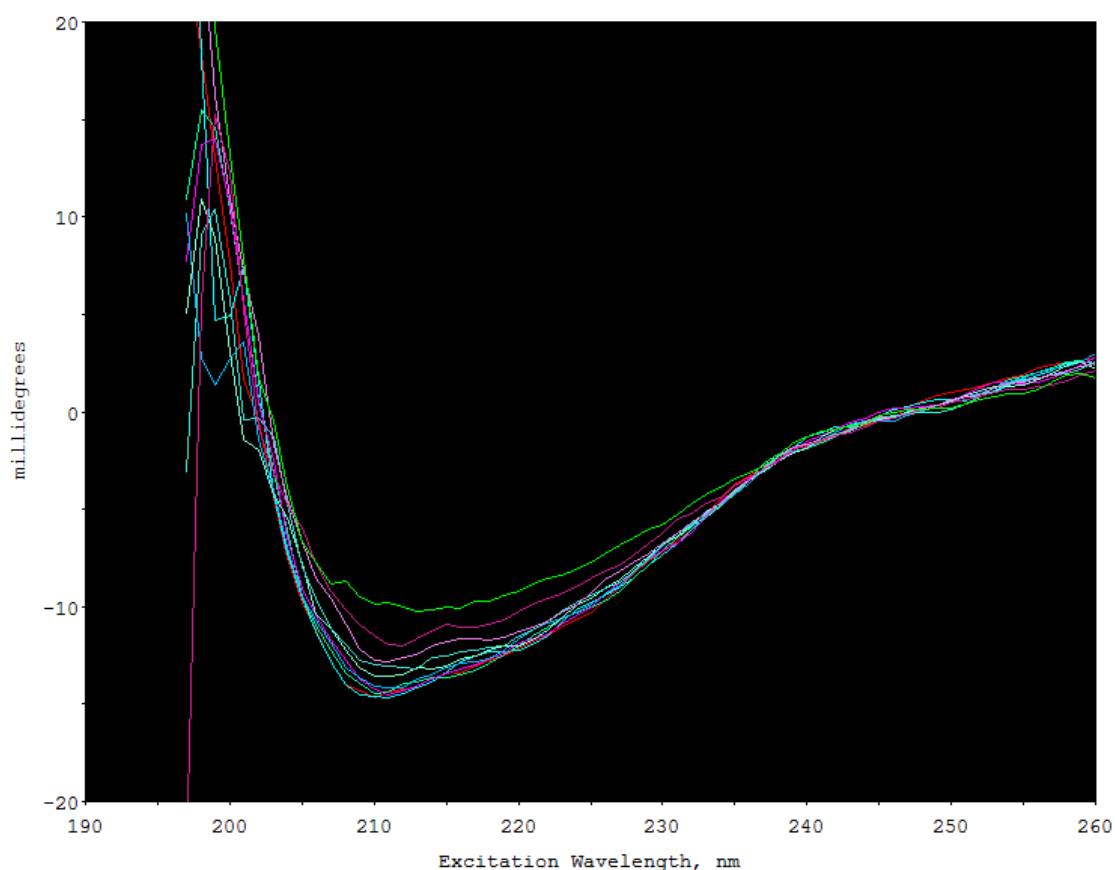
Figure 6.2.3.4 Thermal denaturation spectra for CaMBD incubated with E4076K CaMLD.

Spectra from 10°C to 55°C shown. Change in spectra observed for 45°C (light pink), 50°C (dark pink) and 55°C (light green). These spectra are 3 weakest (highest points) at 210nm and noticeably dip down much earlier at 200nm.

Observed spectra are very similar to just CaMBD thermal denaturation spectra, temperature at which spectra change shape are matching.

Ca^{2+} induced spectral change, predominantly a change in α helix content, was reported for wt CaMLD in the literature. This experiment was repeated with the recombinant proteins produced for this project including CaMBD and two of the mutant forms of CaMLD. 1mM Ca^{2+} was added in the relevant samples as per the original trial and dilution factors were taken into account for the spectra recorded without added Ca^{2+}. For all figures shown below red spectrum is averaged data for protein sample and turquoise spectrum is averaged data for Ca^{2+} added sample. Variance in spectra below 200nm is likely to be attributed to protein sample quality.
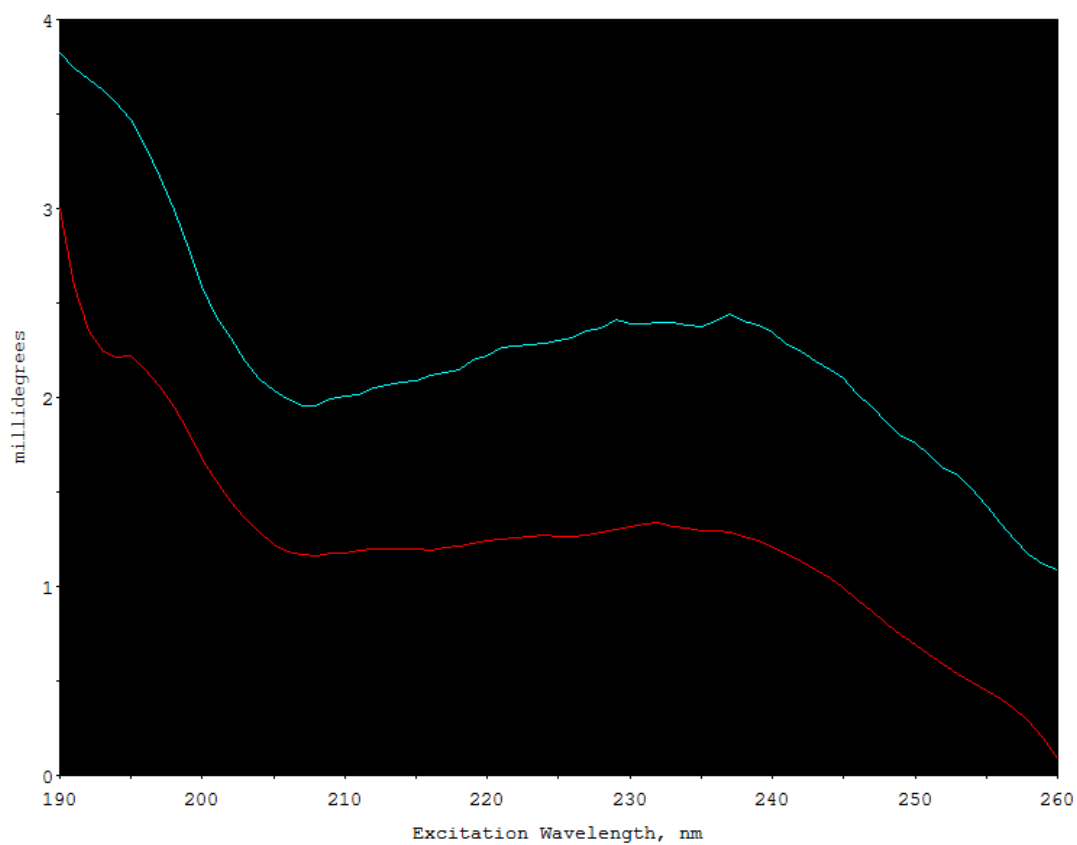
Figure 6.2.3.5 Spectra of wt CaMLD and wt CaMLD with calcium.
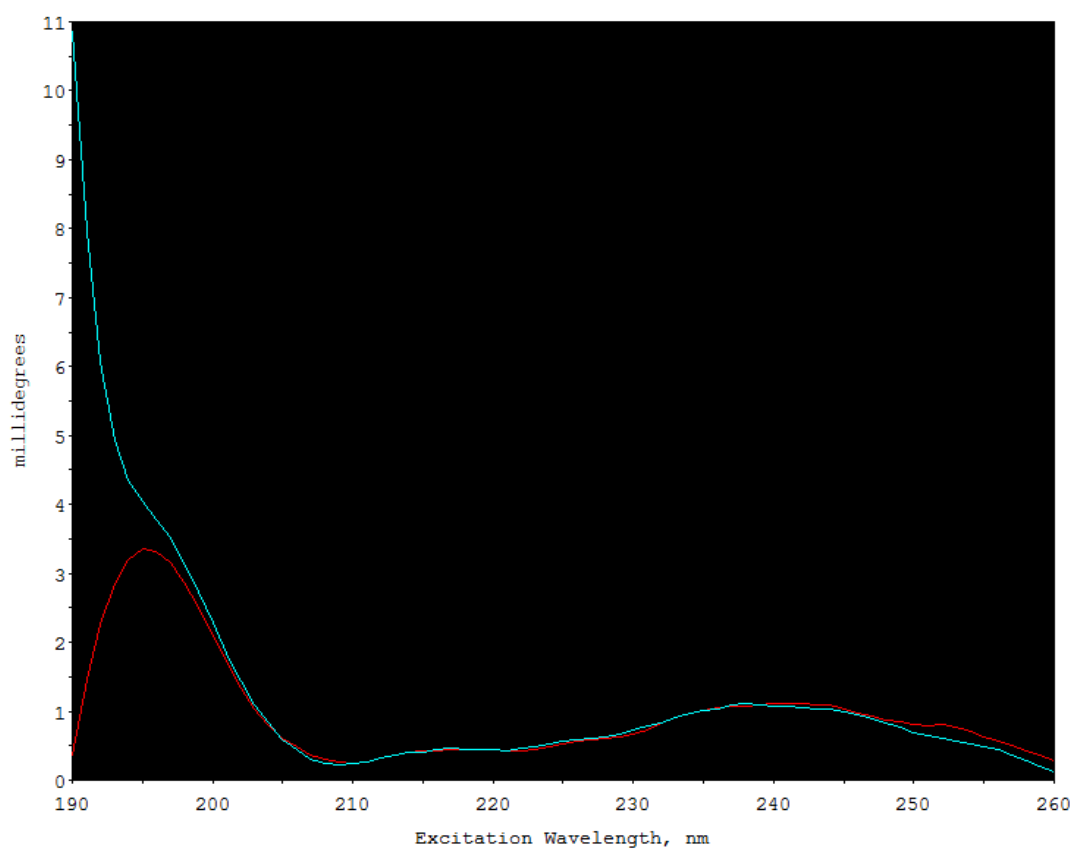
Figure 6.2.3.6 Spectra of E4076K and E4076K with calcium.
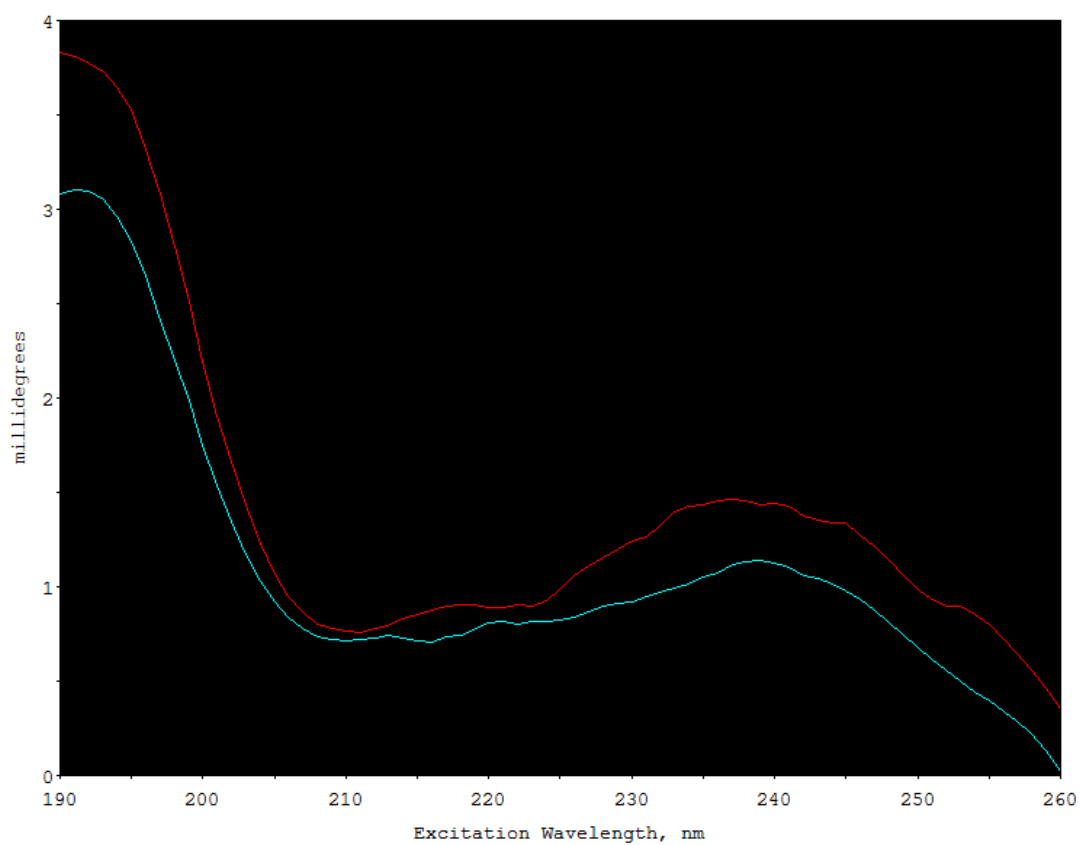
Figure 6.2.3.7 Spectra of N4104K and N4104K with Ca^{2+}.
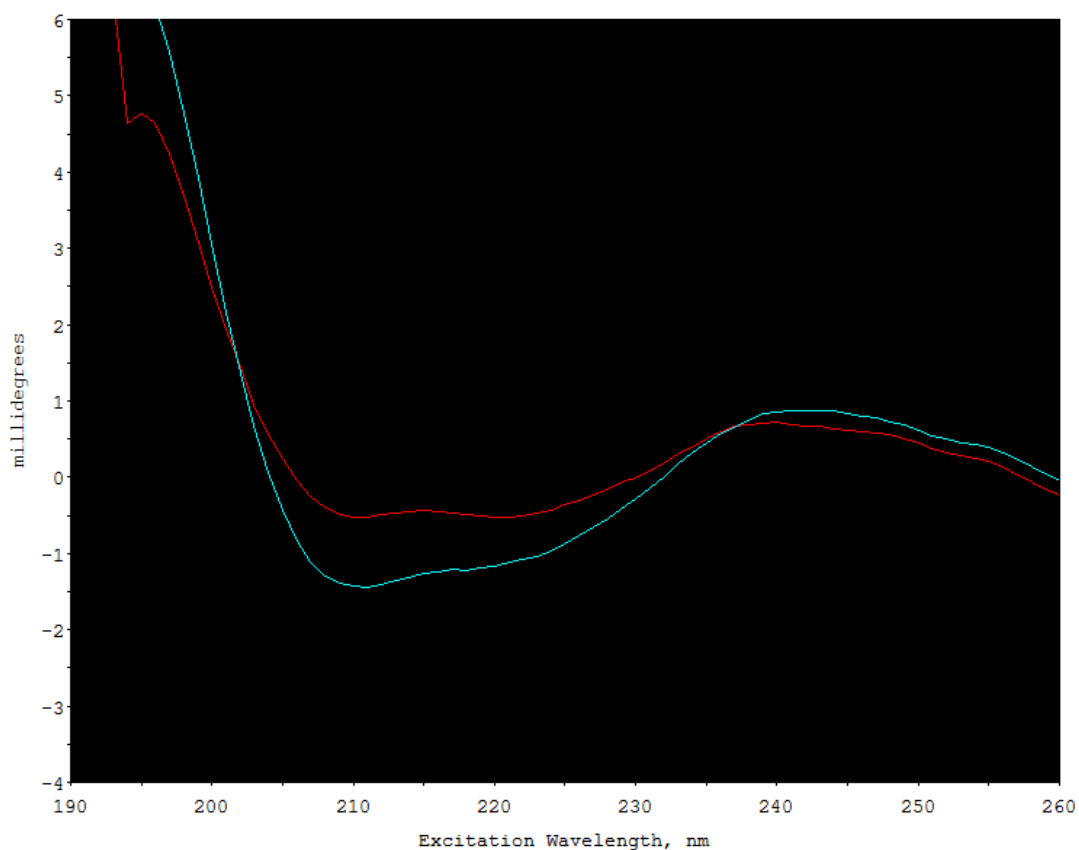
Figure 6.2.3.8 Spectra of CaMBD and CaMBD with calcium.

For the above spectra the wavelength region 208-220 associated with α helix is being examined. For most of the samples there are appears to be very little variation in the shape of the spectra in this region with or without added calcium.

6.2.3.2 Near UV region (250-350nm)

Initially spectra of individual proteins recorded in the near UV region to allow for later comparison.
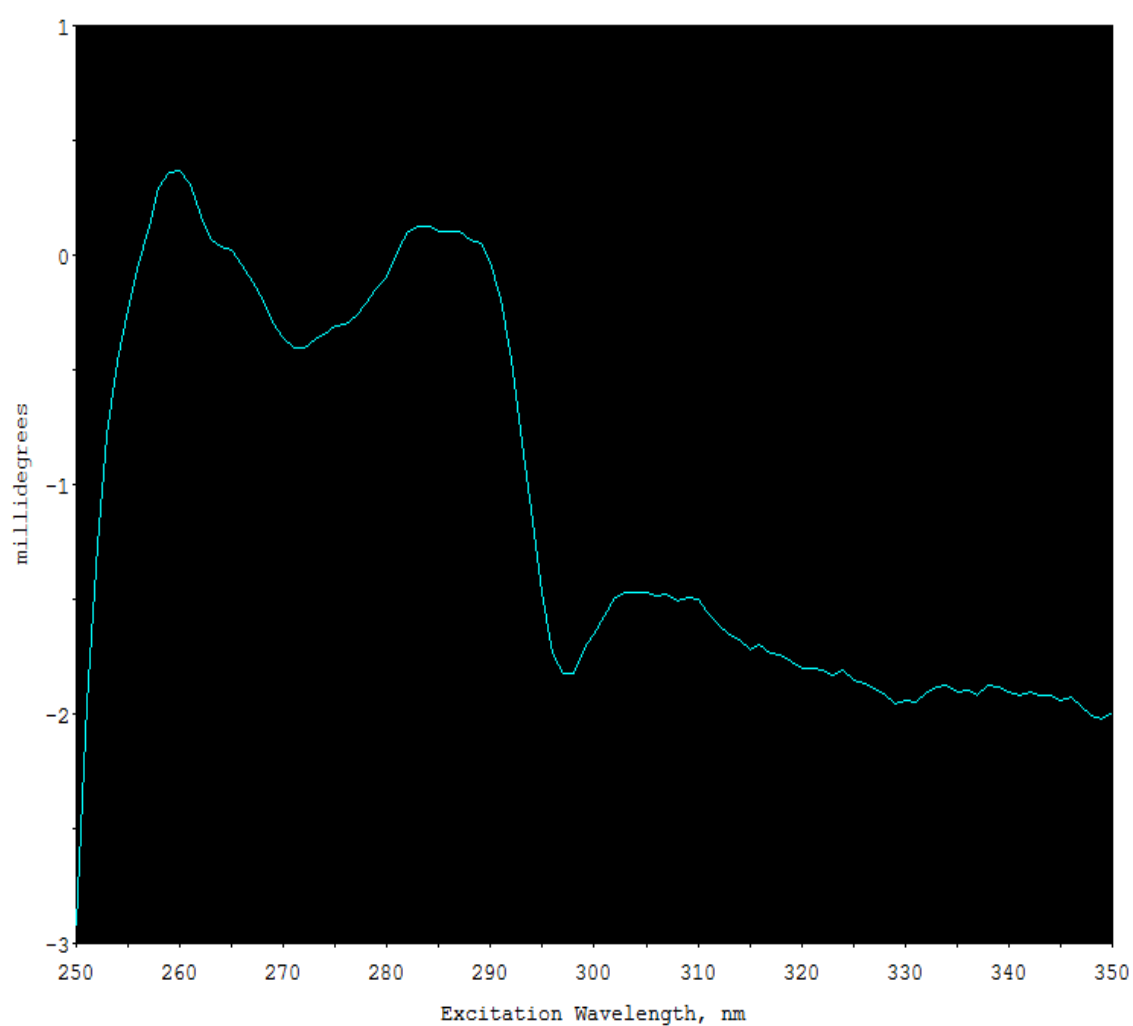
Figure 6.2.3.9 Near UV region spectra of CaMBD.
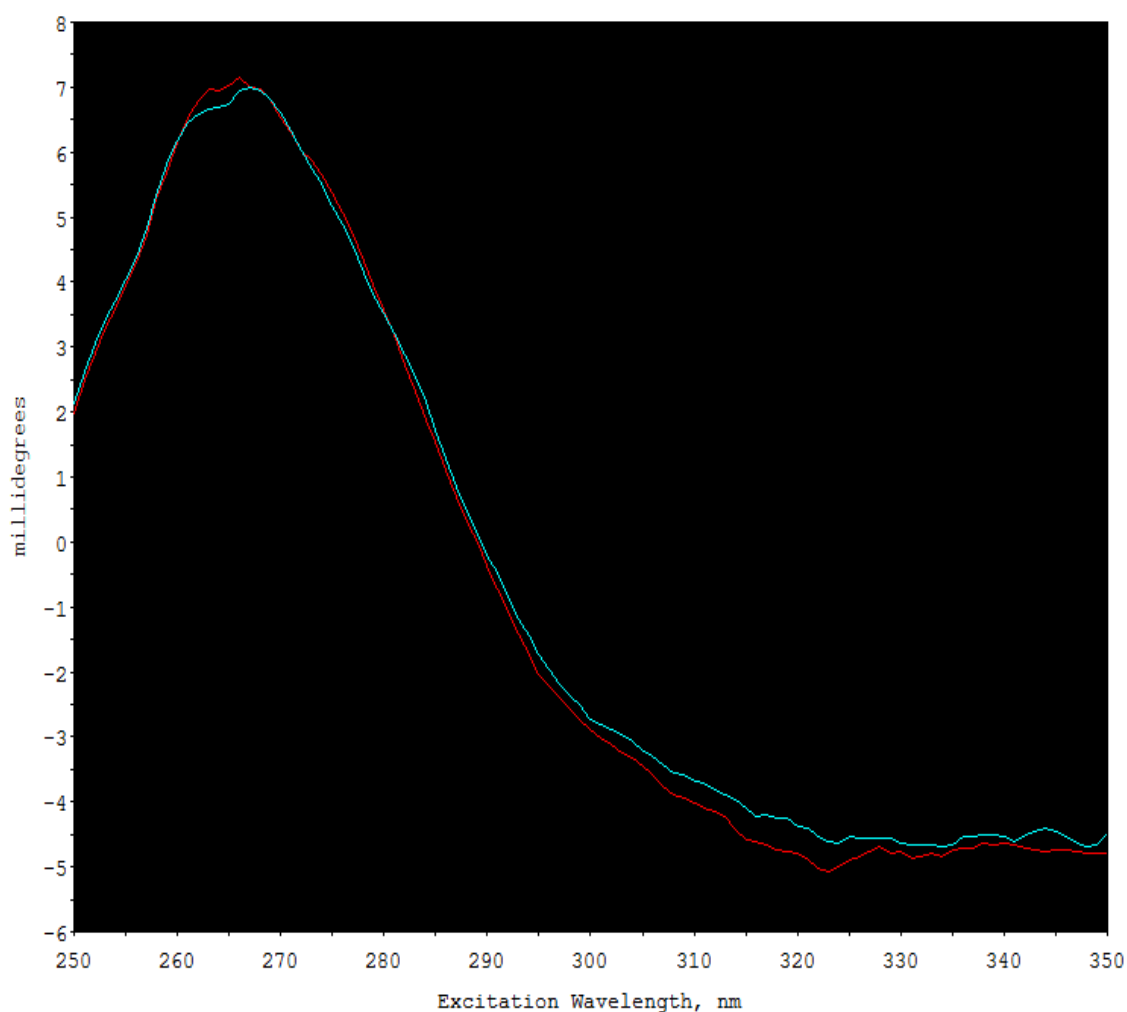
Figure 6.2.3.10 Near UV region spectra of wt CaMLD and wt CaMLD with calcium.

Red spectrum is averaged data for wt CaMLD.

Turquoise spectrum is averaged data for wt CaMLD Ca^{2+} (1mM) added sample.

Change in spectral shape observed in 260-270nm range between wt CaMLD with or without 1mM Ca^{2+} samples. The sample with Ca^{2+} present shows dip in this region compared to sample without Ca^{2+} added. Less clear spectra were recorded for the CaMLD mutants N4104K and E4076K, any potential dip is lost in signal noise.

Owing to the clear difference between the spectra for CaMBD and the CaMLD proteins it was decided to monitor the interaction between the 2 domains in the near UV region.
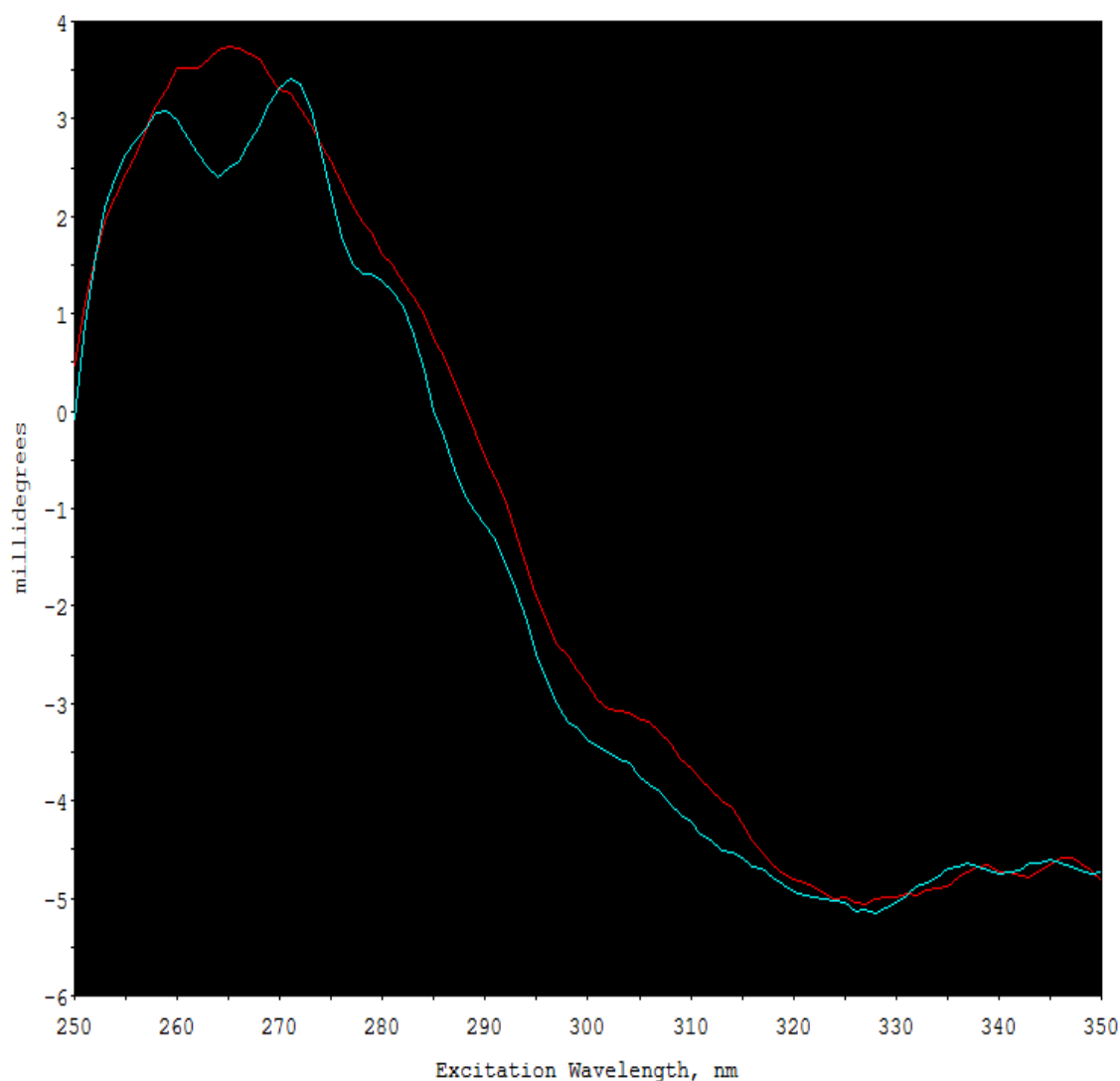
Figure 6.2.3.11 Near UV range spectra of wt CaMLD incubated with CaMBD with and without presence of calcium.

Red spectrum is averaged data for wt CaMLD and CaMBD incubated without Ca^{2+}.

Turquoise spectrum is averaged data for wt CaMLD and CaMBD incubated with 200µM Ca^{2+}.

Mixture of wt CaMLD with CaMBD has spectrum resembling that of just CaMLD. Presence of Ca^{2+} in this incubation triggers change to spectrum observed for CaMBD.
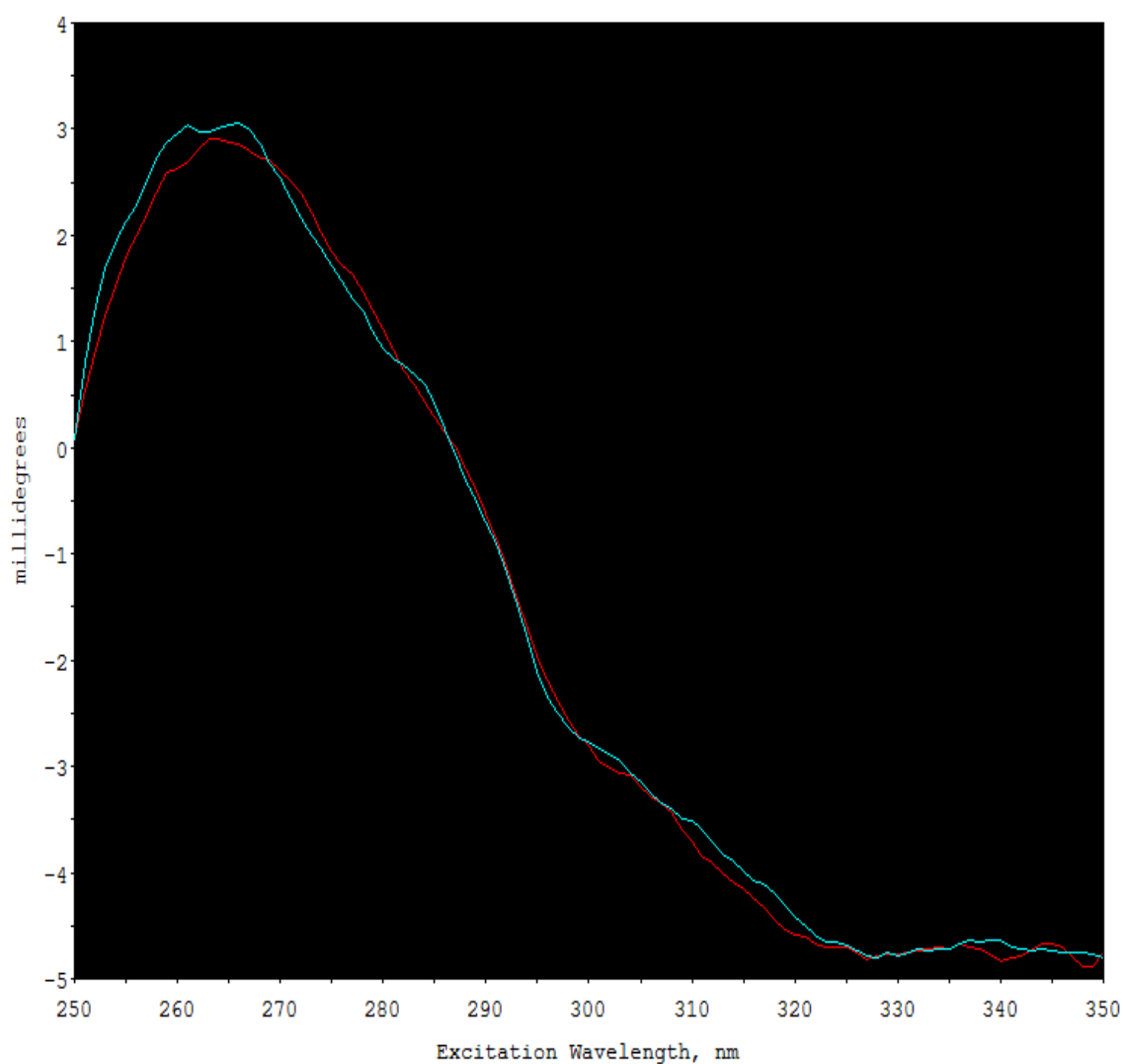
Figure 6.2.3.12 Near UV range spectra of E4076K CaMLD incubated with CaMBD with and without presence of calcium.

Red spectrum is averaged data for E4076K CaMLD and CaMBD incubated without Ca²⁺.

Turquoise spectrum is averaged data for E4076K CaMLD and CaMBD incubated with 200μM Ca²⁺.

Difference between spectra is not as pronounced as for wt CaMLD however change to more CaMBD like spectra, with addition of Ca^{2+}, is observed.

NK LD + BD

NK LD + BD with 200uM Ca
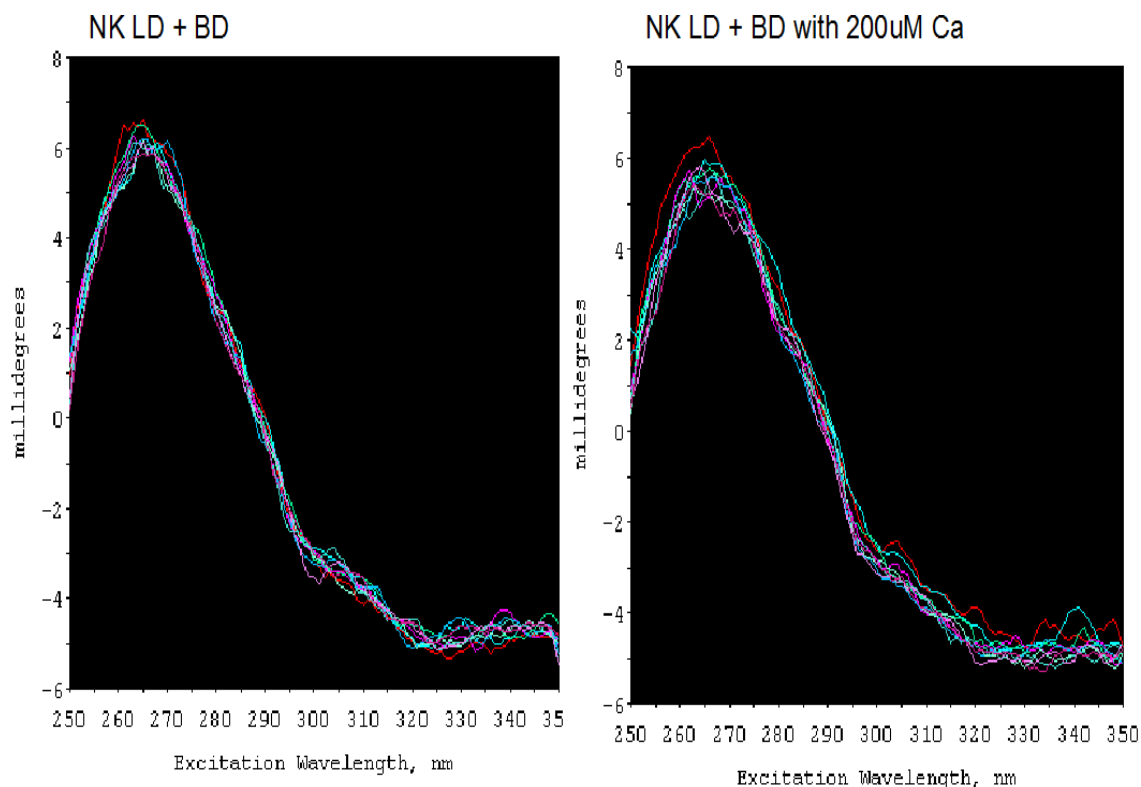
Figure 6.2.3.13 Near UV range spectra of N4108K CaMLD incubated with CaMBD with and without presence of calcium.

Non-averaged data is shown for NK mutant. A few of the spectra recorded for the incubation containing Ca^{2+} show the spectral shape more associated with CaMBD however when the data is averaged the influence of these spectra is minimal.

The role of Ca^{2+} concentration was further explored with relation to the interaction of CaMLD and CaMBD. It should be noted that at this point that the system used to collect the CD data was changed, this system allowed data to be collected over a wider range of wavelengths however had reduced signal sensitivity as a result. A repeat of the wt CaMLD with CaMBD incubation was conducted with new protein material as a check of recordable data.
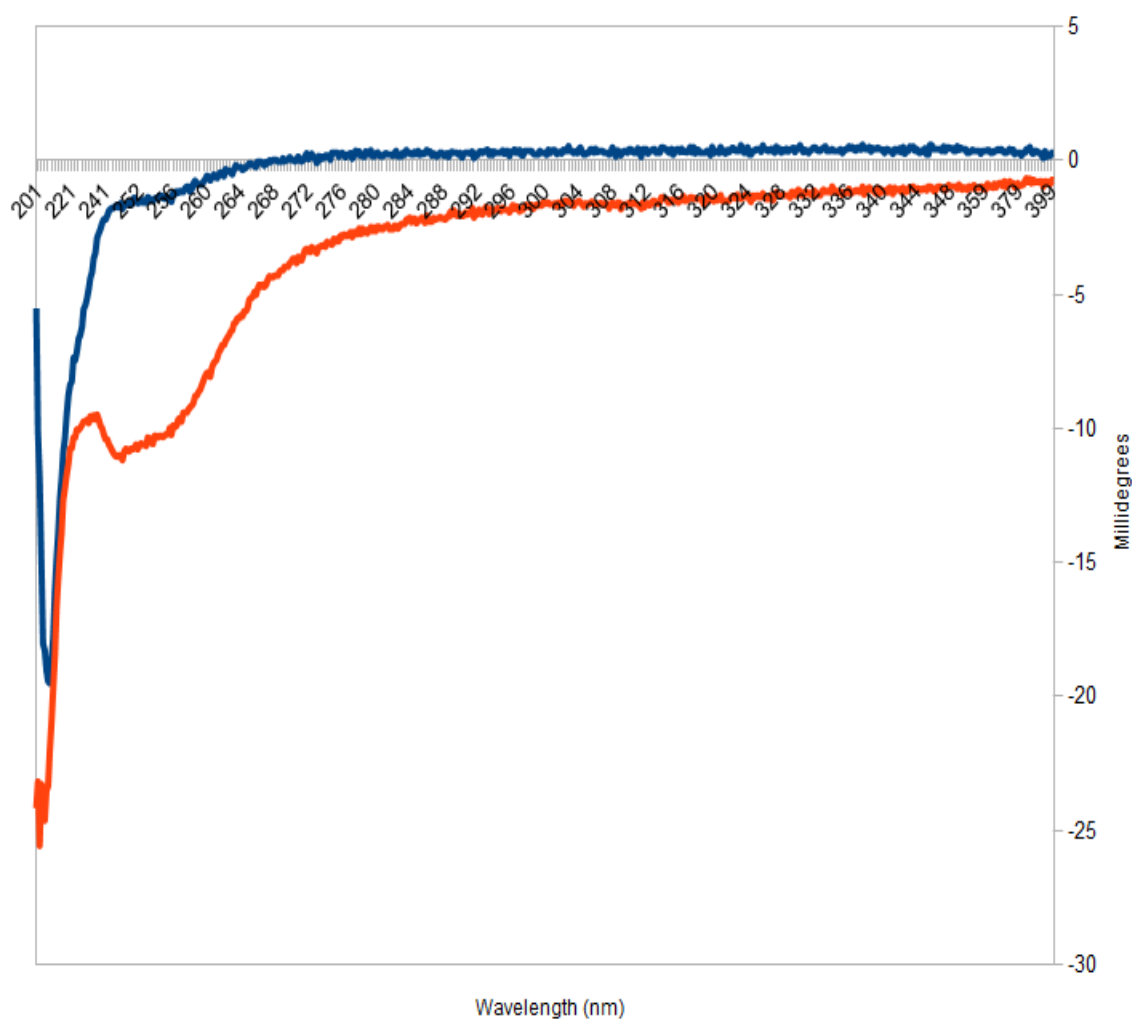
Figure 6.2.3.14 Repeat of near UV range spectra of wt CaMLD incubated with CaMBD with and without presence of calcium.

Blue trace is wt CaMLD with CaMBD.

Red trace is a repeat of the same incubation but with 200μM Ca²⁺ added.

Although the spectra are slightly different in appearance to the previously recorded data, similar characteristics of spectral shape change can be observed. The distinctive 2ⁿᵈ peak, though shifted in wavelength, associated with the CaMBD spectrum is observed when Ca^{2+} is added to the incubation.

A range of Ca^{2+} concentrations were tested to investigate the required concentration to trigger the swap between spectral shapes observed in the above figure.
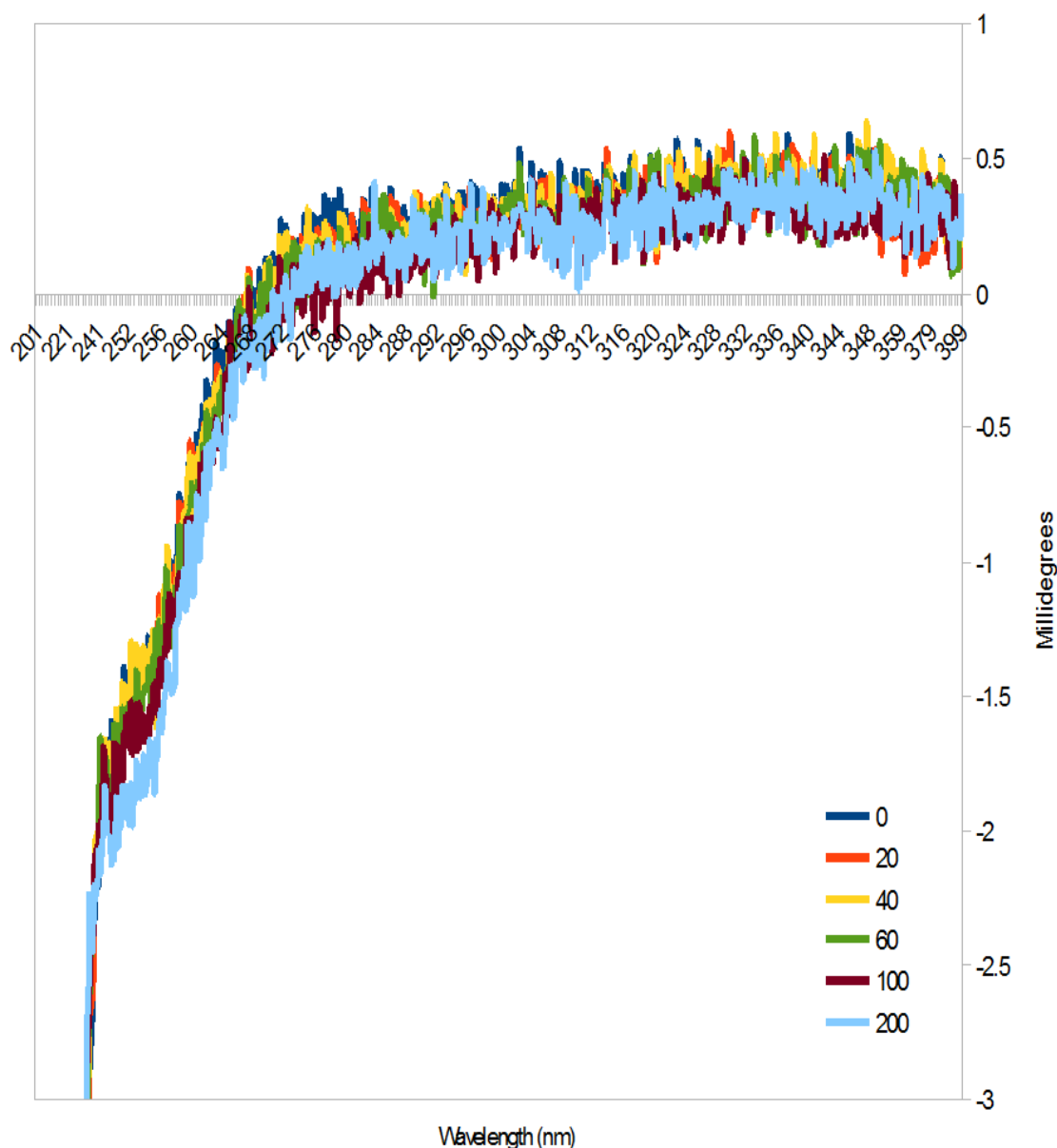
Figure 6.2.3.15 Spectra of CaMLD with CaMBD incubated with range of calcium concentrations.

Ca^{2+} concentrations (μM) shown in key in lower right corner.

Ca^{2+} concentration increase was achieved by serial addition of stock Ca^{2+} solution to same sample.

There is no clear shift in spectra shape as observed with the incubation with a single Ca^{2+} concentration. There is a slight shift down in the spectra in the 250-270nm range. A dilution effect would be expected to shift the spectra up towards the x axis, with the suggestion of a 2nd peak for the 200μM Ca^{2+} trace. A smaller partial shift may have occurred for the 100μM Ca^{2+} trace but this is not clear.

Repeats of this experiment were conducted using individual samples for each Ca^{2+} concentration tested so that no dilution effects were involved. As previously the swap between spectral types was observed for 200μM, and higher, Ca^{2+} concentrations. The presence of 100μM Ca^{2+} did not produce a conclusive result, in different experiments a range of spectral shapes between those associated with CaMLD and CaMBD were observed. Ca^{2+} concentrations below 100μM did not result in a spectral shape change, the trace remained the same as a CaMLD sample.

These experiments were conducted using the CaMLD mutants E4076K and N4108K and as previously it was difficult to obtain consistent results. As seen in earlier data 200μM Ca^{2+} seemed to start a trigger towards a change in spectra shape. However use of higher Ca^{2+} concentrations did not consistently result in the level of difference seen for wt CaMLD.

As these experiments were conducted using tagged protein, the tags were separately tested against the presence and absence of Ca^{2+}, and in control trials in which one of the RyR2 proteins was replaced with just its associated tag. In all these cases the tags were not observed to influence the recorded spectrum in any significant manner.
The interaction of CaM with CaMBD was also tested.
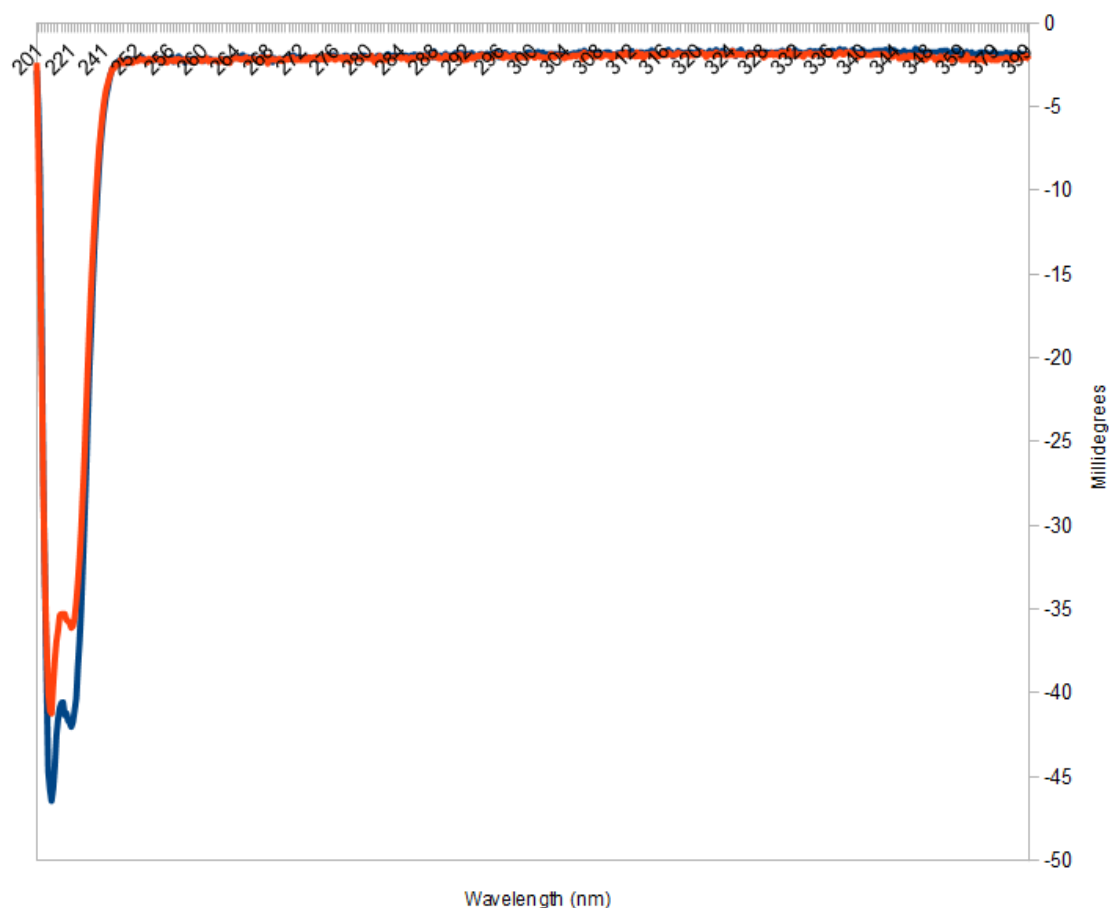
Figure 6.2.3.16 Near UV region spectra of CaM incubated with CaMBD in presence and absence of calcium.

Blue trace is CaM with CaMBD.

Red trace is a repeat of the same incubation but with 200μM Ca²⁺ added.

No change in spectral shape observed between the two conditions.

For this experiment spectrum without Ca²⁺ was recorded. Ca²⁺ was then added to this sample and allowed to incubate before recording with Ca²⁺ spectrum. Addition of Ca²⁺ solution to sample will have diluted protein concentration which is reflected in reduction in protein signal seen as shift of spectrum closer to x axis.

6.3 Discussion

The aim of this chapter was to examine the binding of the studied RyR2 domains, further the understanding of the involvement of CaM within this interaction and also how Ca^{2+} concentrations affected the binding in these instances. In addition, this chapter also shows how the selected CPVT mutations in CaMLD would disrupt these processes.

Initially chemical denaturation was used to examine the structural stability differences between wt and mutant CaMLD. As the proteins were denatured the ΔG values are negative though with differences between the wt and mutant forms. This technique revealed that, apart from the E4076K mutation, the mutant forms of CaMLD had similar structural integrity to the wt with all of them having an intermediate, partially unfolded, state between folded and fully denatured states. The three mutations, G3946S, N4104K and H4108Q, increased the wt stability with respect to an intermediate state, though this may not increase the overall stability of the protein. The lower ΔG values seen for the mutants for the transfer from the intermediate to unfolded state, compared to the wt, means that less energy would be released from this transfer so implying greater stability in the intermediate state than the wt. The E4076K mutant does not share the intermediate state rather going directly to an unfolded state. This mutation severely affects the stability profile of the CaMLD region. Interestingly, the absence of the intermediate state indicates that the E4076K mutant is less likely to destabilise than the wt and other mutants, though under conditions favouring denaturation the EK mutant is expected to be much less stable.

The pull down assays initially offered an established method of analysing the interaction between the RyR domains and CaM. The use of western blots for detection over just SDS-PAGE provided greater sensitivity within the technique though this caused some issues with clarity of data as protein degradation products were also seen. The unspecific interactions with the affinity beads limited the potential range of experiments that could be conducted. Attempts to saturate the beads binding sites were unsuccessful and doubts remained about the validity of the results produced. Despite these issues, data of interest was generated using this technique.

The use of CaM agarose beads with the CaMLD constructs appeared to show predominantly Ca^{2+} dependent interaction between CaMLD and CaM. This interaction is not reported in the literature with only CaMBD binding to either CaM or CaMLD. To initially explain this potential interaction it was proposed that there may be some partial interaction between CaM and CaMLD during the transfer on binding with CaMBD. Based on CaM binding mechanisms described in the literature and the absence of a probable CaM binding site in the structure of CaMLD this is unlikely. Experiments conducted under the same conditions showed weaker binding between CaMBD and CaM than were observed for CaMLD with CaM. It is more probable that there was a flaw with this pull down assay than a genuine interaction between CaM and CaMLD. Subsequent repeats using the recombinantly produced human CaM that did not display any interaction between CaM and CaMLD reinforce this idea.

The use of cross linking reagents addressed some of the above problems described for the pull down assays. Based on cross linking operating on that any genuine interacting proteins would be close enough for long enough to cross link whereas random contact between non-interacting protein in solution would not produce significant detectable amounts of x-linked product. The resultant cross linked product would also be of greater mass so removing it from the regions of the blot affected by degradation products. Glutaraldehyde, a bi-functional cross linking reagent that reacts with lysine residues on the exterior of proteins, was used for these experiments. The use of glutaraldehyde can results in the formation of large aggregates. This may explain the large size band seen for the CaM with CaMBD incubation (figure 6.2.2.5) but not the absence of the expected CaM signal for GST with CaM incubation. There were also unexplained CaM signals produced from the timed sample incubations (figure 6.2.2.6). A constant strength signal was expected for the unbound CaM in the incubation with the GST tag as opposed to an increasing signal with time. For the CaM with CaMBD incubation an increase in CaM-CaMBD product signal was expected with a decrease in unbound CaM. It is possible that some of the issues with cross linking could have been resolved. The use of other cross linking reagents was being trialled as the initial data from CD experiments was obtained. As CD appeared to have greater potential for experimentation, the pull-down assays were not pursued.

For work in the far UV region the buffer composition used played a major factor. Owing to the stability issues found with the RyR2 domain proteins so limitations were placed on buffer components that could be used. The CD buffer system used in this project served as a direct replacement for purification buffer in an attempt to maintain protein stability while removing as many of the far UV absorbing components as possible. Chloride has a strong absorbance below 200nm so was replaced with fluoride. Imidazole has a strong absorbance around 210nm which made thorough dialysis and/or the use of gel filtration extremely important in sample preparation. Obviously this approach did not produce an ideal CD buffer, data collection was limited to >190nm (figure 6.2.3.1) even when there was a high protein concentration giving good signal strength. However as the main data wanted from the CD experiments was for the α helix regions (>190nm, mainly 200-220nm) then this was viable solution.

Initially CD in the far UV region was used as a tool to assess folding of the recombinantly produced proteins. As can be seen in figure 6.2.3.2, spectra for folded protein was obtained for all the tested samples. The accompanying table reveals potential differences in the structure between wt and E4076K CaMLD. The secondary structure predictions show wt CaMLD and EK as very similar in expected structure however this not observed in the experimentally derived values. The EK values are close to the predicted but wt has higher α helical content with a matching reduction in other components, β sheet content is similar to predicted. It is possible that wt CaMLD has more defined secondary structure than under predicted conditions. The presence of the EK mutation may disrupt this structure so resulting in more random coil (included in other) in the structure. The experimental values for CaMBD show less α helix than predicted, and increased β sheet and other content. It is not clear why this occurs. It is possible, that as with wt LD, experimental conditions differ from those of predicted resulting in variance between the two sets of values. Also the algorithm used to fit the secondary structure values to the experimental spectra may cause differences in values. As data could not collected below 190nm this limited the choice of algorithms that could be selected and increased the deviation between the fitted and recorded data (figure 6.2.3.2). The CDSSTR program (Sreerama & Woody 2000) combined with SP43 (reference set containing 43 soluble protein examples) was considered to be the most accurate option available with the data obtained.

The thermal denaturation experiments were trialled as companion data to the chemical denaturation work. Owing to the equilibration time delay required at each new temperature point the programme to run thermal denaturation was several hours long. This limited the amount of data that could be collected within allocated B23 beamtime sessions. As no denaturation data was held for CaMBD so this protein was tested first. The flattening of BD spectra (figure 6.2.3.3) indicates loss of secondary structure, this indicates that the protein is denaturing, probably structure is lost somewhere between 50-55°C. There is no real observable difference between the just CaMBD and complex with E4076K CaMLD spectra. Admittedly further data, in particular with wt CaMLD, would need to be recorded to allow full comparison. Owing to the time taken to collect data and the initial unclear data it was decided not to further pursue the thermal denaturation experiments.

One of the experiments conducted in the far UV region was to repeat the spectra recorded for CaMLD with and without Ca^{2+} reported in the literature (Liangwen Xiong et al. 2006). This data is shown using ellipticity rather than millidegrees, ellipticity is the millidegree values corrected for protein concentration, however this difference in units does not change the shape of the spectra which is the examined component when assessing condition variables. The experimental data from this project reveals little difference between the presence or absence of 1mM Ca^{2+} for wt and mutant CaMLD. Possibly a slight difference can be seen for wt and NK protein though the EK sample appears unchanged (figures 6.2.3.5-8). Based on the expectation of the affect of the mutations, EK is thought to be less Ca^{2+} sensitive though 1mM Ca^{2+} should be well above any triggering threshold. The literature example represents somewhere in the region of a 5% change in α helix content, while this represents an important rearrangement of the structure of the protein it is not a huge difference in terms of CD signal and maybe easily lost due to signal noise in a non perfect sample. An illustration of the impact of signal variation is shown by the CaMBD sample appearing to show a difference between conditions. This is not expected as CaMBD does not contain a known Ca^{2+} binding site.

The quality of data collected could have possibly been improved by using more concentrated protein samples. The increase in signal strength would have also allowed for a shorter path length cell to be used potentially enabling data collection at shorter

wavelengths despite the buffer imposed limitations. These experiments were conducted using cleaved tag protein samples, the problems associated with handling this material and achieving higher concentrations are discussed in chapter 5.

As the expected conformational change with CaMLD was difficult to observe within the far UV region is was decided to collect readings in the near UV region as this offered a great chance of detection. As discussed in the chapter introduction the RyR2 domain proteins appeared to be good candidates for work in this region. Commonly when working in this region the relative contributions to the spectra based on their percentage aromatic acid composition is determined and the concentrations in the CD sample altered to provide matching signal strength of the proteins involved. It was calculated that CaMLD gives a stronger signal than CaMBD. Experiments were conducted using these corrected concentrations alongside samples using 1:1 concentrations as would be expected if there was direct binding between the two proteins. No obvious difference was observed so all further experiments were conducted using 1:1 concentrations. The only issue from this approach was that the signal contribution from CaMBD might be diminished compared to CaMLD but the observed changes in spectrum shape indicated that this was not a problem.

The repeat of the addition of Ca^{2+} to wt CaMLD (figure 6.2.3.10) revealed a difference in spectra indicating a conformational change in the CaMLD structure triggered by addition of Ca^{2+}. This result was less clear for the N4104K and E4076K mutants. The wavelength of the spectra where variation was observed is attributed to phenylalanine residues within the protein. There are several Phe residues located in CaMLD; 2 upstream of 1st EF hand motif, 2 per EF hand motif and 1 downstream of 2nd EF hand motif. It is plausible that there could be a Ca^{2+} induced conformational change which alters the environment around the Phe residues, this is almost certainly the case for the amino acids in the EF hand motifs, which resulted in the observed spectral change.

It should be noted that 200µM Ca^{2+} was used instead of 1mM for this experiment and all subsequent work involving the triggering of a conformational change in CaMLD. The CaMLD region in RyR1 was reported to have an apparent Ca^{2+} binding affinity of 60 (±12) µM. It is assumed that the conformational triggering threshold concentration is somewhere in this region. The switch to 200µM Ca^{2+} in experimental conditions was an

attempt to use a value closer to this threshold and set a lower top concentration for experiments attempting to discover this threshold concentration.

Based on the successful observation of a Ca^{2+} induced conformational change in CaMLD, the interaction between CaMLD and CaMBD was investigated using near UV region CD. For the wt CaMLD a clear change in the shape of the spectra is observed (figure 6.2.3.11). This indicates that a clear conformational change is occurring which suggests binding with CaMBD. As previously found 200μM Ca^{2+} is sufficient trigger conformational change, as this structural rearrangement in CaMLD is considered to be required prior to CaMBD binding so it is not surprising that this concentration is sufficient. As the change in spectra is different to that observed for just the addition of Ca^{2+} so it is likely that the difference is a result of the presence of CaMBD. As the change in spectrum shape transfers between that associated for CaMLD to CaMBD it is even more suggestive that CaMBD binding is the cause. What is not so clear is why the spectra should so clearly differentiate between these two component associated spectra. This difference suggests that in the absence of Ca^{2+} CaMLD is the dominant protein signal. The addition of Ca^{2+} triggers the conformational change in CaMLD which is assumed to adopt the "wrap round" binding of CaMBD, as observed for CaM in the PDB structures, so making CaMBD, or a CaMBD like, signal the dominant observed component of the spectrum. This potential CaMBD binding result is not so clear for the NK and EK mutants but appears to occur to some extent suggesting possibly weaker or a lower percentage of the CaMLD population binding with CaMBD, as thought to be seen in the NK spectra (figure 6.2.3.13), than compared to wt.

With the change of CD equipment so the CaMLD binding to CaMBD experiment was repeated. As expected a very similar result is observed with the shift to the CaMBD associated spectra with the addition of Ca^{2+}. Although the 2nd peak of the CaMBD spectrum appears to be shifted closer to 250nm (figure 6.2.3.14), appears closer to 260nm for previous data sets, this is an artefact produced by the change in x axis scaling resulting from the change in frequency of data points collected.

To determine the CaMLD conformational trigger Ca^{2+} concentration threshold a series of Ca^{2+} concentrations were tested (figure 6.2.3.15). The data from this experiment is unclear, it appears that 200μM Ca^{2+} is required, though 100μM may be sufficient. The

use of serial additions of Ca^{2+} may well skew the results in this experiment. Repeats using individual Ca^{2+} concentrations seem to confirm these observations, 100µM Ca^{2+} may be close to the threshold for this recombinant wt CaMLD under these experimental conditions. A 200µM Ca^{2+} concentration is required to start to observe a similar triggering of conformational change for the EK and NK mutants. As no Ca^{2+} concentrations were tested between 100 and 200µM a closer threshold value cannot be assigned to these mutations beyond stating that the assumed required concentration >100µM. This difference in required Ca^{2+} concentrations suggests towards lesser Ca^{2+} sensitivity for E4076K and N4104K mutants.

The experiment including the incubation of CaMBD with CaM in the presence and absence of Ca^{2+} was conducted as a check against previously observed results using the CaMBD construct protein. The only difference seen in the spectra (figure 6.2.3.16) is as a result of a dilution factor of the addition of Ca^{2+} to the sample to record the with Ca^{2+} spectrum. Based on the mechanism of CaM binding so any observable conformational change would be expected in CaM rather than CaMBD. Previously tested CaM samples, although generating spectra, had shown no discernible difference in the absence or presence of Ca^{2+}, as discussed in the chapter introduction this was expected. Owing to this lack of signal change CaM was not further used within CD experiments.

The data obtained in this chapter indicates that the recombinant CaMLD and CaMBD proteins bind. This process may first require a Ca^{2+} induced conformational change in CaMLD. The interaction between domains occurs for the mutant forms E4076K and N4104K though may be affected by the possible lesser Ca^{2+} sensitivity of these mutations. In addition the E4076K mutation severely weakens the stability profile of the CaMLD protein, this may be reflected in the decrease of α helix and β sheet content seen in the experimentally derived secondary structure.

Chapter 7:
General Discussion

7.1 General Discussion

The aim of this thesis was to examine the interaction between the CaMBD and CaMLD of RyR2 and the impact that certain CPVT point mutations within CaMLD have upon this interaction. A number of computational, structural and functional techniques, as discussed in previous chapters, were applied to this study. The data collected from these methods is now discussed with regards to the studied domain interaction within RyR2 and how certain CPVT mutations may disrupt this mechanism of interaction.

The amino acids 3545-4163 of the human RyR2 contains the CaMBD and the CaMLD. Work carried out previously in this lab using FRET technology have shown that the cytoplasmic domain of the ryanodine receptor communicated with the inner pore region (membrane –embedded channel pore segment) via an interacting domain (I-domain) consisting of amino acids 3722-4610 (George et al. 2004a). The majority of residues of this region are expected to traverse through the inner regions of the channel, thus apart from the CaMLD and CaMBD regions, they are not expected to be solvent exposed. Recently, discussion and the presence of this region in modulating the interaction between the cytoplasmic domain and that of the inner pore region has also been suggested by the Van Petegem group as shown in the figure below.
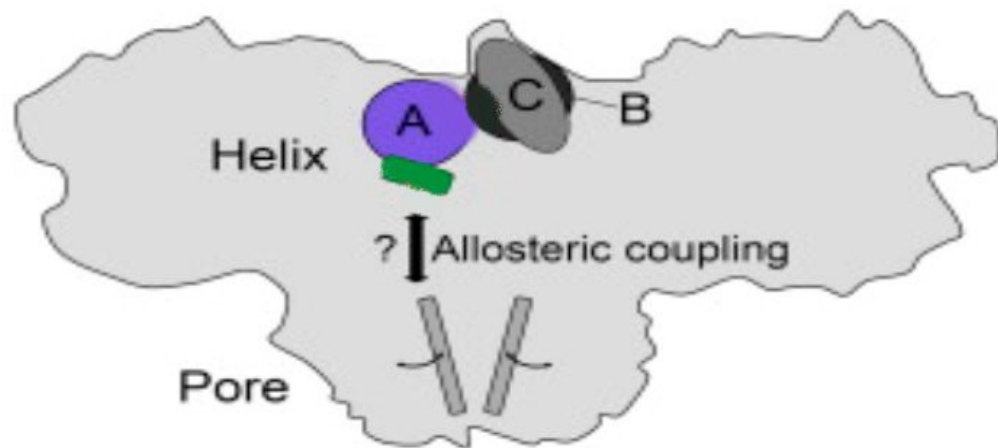
Figure 7.1   Hypothetical schematic model of cross-domain interaction in the RyR.

Model showing the possible interaction between the N-terminal region of RyR and the transmembrane pore region. Regions labelled A, B and C are the interacting domains within the N-terminal segment of the RyR. The segment shaded in green represents the region in RyR encoded by exon 3, seen in severe CPVT phenotypes where the entire exon is deleted. Figure adapted from (Lobo et al. 2011).

The current accepted model of allosteric interaction between the cytoplasmic domain and the pore region is that multiple modulatory interactions exist between these two important domains. It is thought that the normal gating of the RyR as a channel occurs via allosteric regulation between the mutation hotspot region in the N-terminal domain and the pore region, where the N-terminal region plays an important role in slowing down the opening of the pore. Domain A (as shown in figure 7.1) interacts with the pore via a helical segment within it (Lobo et al. 2011).

The region encoded by this area of interest which incorporates the CaMLD and CaMBD region contains a high proportion of amino acids that are highly hydrophobic in nature, as examined in the computational modelling work in chapter 3. This is to be expected of a set of amino acid residues that lie buried within a structure, largely hydrophilic amino acids are usually found at the solvent exposed surfaces of a protein molecule. As such, it is extremely difficult to generate recombinant versions of such highly hydrophobic proteins since the purification methods rely on the use of water based buffers and separation media reliant on aqueous conditions.

Experiments with CaMBD and CaMLD required the production of recombinant protein. The design of both constructs relied heavily on the identification of secondary structure elements believed to play important roles in the interaction of the RyR2 domains. The constructs designed for CaMBD were based around a section of protein that had well-established structures. CaMLD posed more problems. Traditionally when trying to produce stable protein then whole proteins or clearly identified domains within the protein of interest are targeted. There are published reports of a CaMLD area within RyR1 and RyR2 that interact (Liangwen Xiong et al. 2006) (Gangopadhyay & Ikemoto 2008) (Gangopadhyay & Ikemoto 2011) which are all based on the initial report in RyR1 by Xiong and colleagues (Liangwen Xiong et al. 2006). While CaMLD appears to offer an identifiable domain, it is not entirely clear from a structural viewpoint where the domain begins and ends. When compared to a lobe of CaM then portions of predicted α helix that are proposed to play a role in binding to CaMBD can be identified. It is probably around these residues, close to the published coordinates, that CaMLD starts and finishes.

Various recombinant constructs of CaMBD and CaMLD were tested as explained in chapter 5, however, only few were marginally soluble. The expression of such hydrophobic constructs is also not conducive to the growth and propagation of *E.coli*, thus protein recovery from *E.coli* is also a major factor to be considered. Furthermore, to protect itself from the recombinant protein, *E.coli* frequently sequester the protein as soon as it is translated into inclusion bodies, enveloping it with various heat shock proteins so that it does not interfere with the cytoplasmic milieu of the bacteria. Thus, when bacteria are first lysed and the broken cell membrane portions are centrifuged as a pellet for removal, the inclusion body sequestered protein also aggregates with the pellet fraction. Isolating the inclusion body sequestered proteins is possible but the method involves denaturing the tight interaction between the recombinant protein from the inclusion body and heat shock proteins using strong denaturants such as urea or detergents as detailed in chapter 5. The downside of this method is that the tertiary folding of a protein is often lost upon these treatments. While refolding protocols can be applied there is no guarantee that the functional folded protein will be obtained and, as such, the recovery of the protein of interest at the end is very poor at best.

To combat these issues various affinity tags were also used to encourage the solubility of the recombinant protein. The NusA protein is one of a complex of proteins that is used by bacteriophage lambda to prevent the early termination of an RNA transcriptional step. Thus, with the use of these "anti-termination" proteins, RNA transcription can be coaxed to continue for longer than intended. The NusA protein is also a highly soluble protein molecule, which is also used as a recombinant protein expression modifier. Proteins that poorly expressed were highly hydrophobic and not amenable to standard purification methods usually were "workable" again. These proteins showing better expression profiles, enhanced purification and yield from recovery steps when expressed as a fusion partner with the 495 amino acid NusA tag at the N-terminus. However, upon cleavage of the solubility tag such as NusA, these proteins frequently started to form insoluble aggregates and dropped out of solution. Examples of this phenomena are related in chapter 5.

Other solubility tags tried and used for improving expression of recombinant proteins were the Trx tag (thioredoxin tags), MBP tags (maltose binding protein tags) and the GST tag (glutathione S-transferase tag). However, as shown in chapter 5, none of these tags were as effective at attempting to gain the large milligram scale production of recombinant proteins that is necessary for the production of crystals for determination of structure by x-ray crystallography.

There is some evidence in the literature that disulphide bonds are possible and present within the RyR molecule (Favero et al. 1995), however it is not certain if the region encoded by aa 3545-4163 contains any disulphide bridges in the intact whole RyR molecule *in vivo*. The cytoplasm of *E.coli* is a very strong reducing environment and recombinant proteins that are normally stabilised by disulphide bridges would not be correctly folded as these disulphide bonds would be broken within *E.coli* using enzymes such as thioredoxin and glutaredoxin. Such proteins would then be randomly folded, form aggregates within the cytoplasm of *E.coli* and be sequestered into inclusion bodies. If disulphide bonds are crucial for the maintenance of protein structure, then there are *E.coli* strains, such as the commercially available Origami strains, that have oxidising properties in the cytoplasm which allow the maintenance of disulphide linkages. Since it is unknown if disulphide bonds are crucial for the folding of this region of the RyR, so no steps were taken to preserve those S-S bonds.

Certain additives can be used to increase the expression of highly hydrophobic proteins and to encourage their presence within the aqueous milieu of the cell extract following lysis. These additives, namely the osmolytes such as betaine have been used but to little effect in this thesis. Various growth conditions, low temperatures for induction of protein from the plasmid, different formulations of media and other techniques have been tried but to little effect as described in chapter 5. Due to this, the production of large quantities of recombinant protein were not feasibly possible, and thus the initial aims of producing adequate amounts of protein for the purpose of obtaining protein crystals to try to solve the 3D structure of the regions of interest, was not met.

Since it was not possible within the time frame of this research study to persevere with the ultimate goal of crystallisation, it was decided to obtain some preliminary answers by theoretical and homology modelling. The computational based work produced models of the studied region. These models contained both domains of interest and the predicted secondary structure elements in the proposed loop region between the two domains. The theoretical modelling was carried out using the homology modelling software, MODELLER (A Sali & Blundell 1993). Within MODELLER, it was also possible to force the secondary folding of regions of the protein based on the data obtained from secondary structure prediction software. Thus, after the CaMBD-CaMLD regions were folded using a homologous template, MODELLER was used to force the folding of stretches of protein with unknown structure by *ab-initio* folding methods. Following this, the model was refined to incorporate bond lengths and angles that were within the prescribed values that have been empirically established.

Having finished the theoretical modelling using first principles (*ab initio* modelling), the model was coaxed to fold further by the interaction of proximal molecules. In molecular simulations, usually the non-bond values of molecules are specified to approximately 12-15 Å. In reality, two atoms that are not bonded to each other would still exert at attractive force on each other that tails off to infinity, the force of attraction may be minuscule at infinite distance but is still present. In a computer simulation however, due to the enormous amounts of computation required should the non-bond values be specified to infinity, it is common to specify them to approximately 12-15 Å. In the conducted simulations, to coax the molecule to fold in 3D-space, the non-bond

value was set to the largest dimension between two atoms placed furthest from each other. In the time taken for the simulation to run, the structure was observed to become more compact, however, for a simulated folding to occur *in-silico,* this would be expected to take a much longer time and require much greater computational processing power. It is interesting that protein folding *in-vivo* takes place in an extremely short period of time, as described in chapter 4. The structure presented in chapter 4 therefore is best taken as a structure that is still in the process of folding.

Although various energy parameters, bond-distances and angles were analysed exhaustively in chapter 4, one of the most remarkable observations is that despite the high mobility of the molecule during the simulations, the CaMBD and CaMLD domains did not migrate very far away from each other. This shows that the non-bond interacting forces (such as Van der Waals forces, ionic-interaction, etc) are sufficient to keep the two domains proximal to each other. Thus, there is a natural propensity for these two domains to interact with each other.

The functional assays that were conducted displayed interaction between the recombinant protein of CaMBD and CaMLD constructs. As the interaction was also observed when the solubility tags had not been removed this addressed some of the issues raised above. This suggested that functional structure of RyR2 domain recombinant proteins is not affected by the presence of the tags either at the level of steric hindrance or through unwanted interaction. The tags were shown not to be responsible for potential false positive binding in control experiments.

A calcium induced conformational change was observed in CaMLD which seemed to be required before interaction with CaMBD. While a specific threshold concentration to trigger the conformational change, for these construct proteins, was not not found, data suggested in the range 100-200 μM calcium. This value is higher than the 60 ±12μM calcium binding with CaMLD concentration reported from fluorescence binding assays by Xiong (Liangwen Xiong et al. 2006). While this concentration is above the sub μM calcium concentrations observed for RyR2 activation it is below the mM inhibitory levels (Lamb, 2000). As CaM is expected to unbind from CaMBD with the increasing calcium levels during RyR2 channel activation, it is plausible that the conformational change required for CaMLD binding to CaMBD could be triggered in the range of

100μM calcium. The same set of experiments indicated that the CPVT point mutations E4076K and N4104K may decrease calcium sensitivity as CaMLD construct protein containing these mutations appeared to require higher calcium concentrations than the wt for binding with CaMBD.

The requirement for a calcium induced conformational change prior to binding further strengthens the comparison between CaMLD and CaM. CaMLD is seen to share very similar structure and functional response to a lobe of CaM. However, beyond showing CaM interaction with the recombinant CaMBD protein this project did not address the potential competition between CaM and CaMLD for binding to CaMBD. Competition assays between CaM and CaMLD under varying calcium concentrations were attempted but with no clear data produced. Part of this was due to the low CaM signal given in near UV CD experiments as discussed in chapter 6. As such, it is not possible to discuss the conditions required for CaM to dissociate from CaMBD so that CaMLD will bind using results from this project. It is known that this transition in binding occurs as part of the channel function, the impact of calcium concentrations are discussed in the literature (Balshaw et al. 2001) (Naohiro Yamaguchi et al. 2003).

The mutation, E4076K was found to significantly weaken the structural integrity of the CaMLD domain. The other mutations studied, G3946S, N4104K and H4108Q had a similar denaturation profile to that of the wt sharing an intermediate stage before becoming unfolded. However, these three mutations were less likely to transfer to this intermediate state than the wt. This would appear to link to work conducted by colleagues which showed that mutations in the I domain caused variation in calcium release that were not observed for central domain mutations. N4104K was one of these studied mutations and was implicated in causing conformational instability in the functional channel (George et al. 2006). If the ability of the protein to undergo a calcium induced conformational change is reflected in the ease of a transfer from native to intermediate state, i.e. internal rearrangement, then the data obtained in this project would start to explain the structural and functional impact of the CPVT mutation N4104K. The wt version would respond to a lower calcium concentration and then have a lower energy requirement to undergo a conformational change as part of the channel opening in response to increasing calcium levels.

E4076K was identified at the start of the project as a CPVT point mutation of particular interest in RyR2. Its position within the EF hand loop, although not in a conserved position, lead to speculation that this mutation would disrupt the function of CaMLD though it was not clear to what extent this mutation would disrupt function. The data from this thesis, as discussed above, indicated a structural weakness and a reduced calcium sensitivity compared to the wt recombinant protein when this mutation was present. However it would appear, that from the studies conducted in this thesis, that even with this mutation present CaMLD is still able to function in its expected CaMBD interaction role. This is consistent with the understanding of CPVT in that the condition does not present until the heart is in a "stressed" state as described in relation to the phosphorylation state of RyR in chapter 1. As such, this thesis has provided further understanding of the possible mechanism by which certain CPVT point mutations within CaMLD disrupt the interaction with CaMBD and so lead to the manifestation of CPVT symptoms.

7.2 Future directions

One of the main concerns with the functional experiments conducted in this project was the quality of protein. High throughput studies for crystallisation work are increasingly generating a large number of potential constructs which test a large range of domain boundaries encompassing the region of interest. A range of vectors and expression systems are also trialled. This approach, because of its required scale, may not be fully applicable within the scale of a project such as this but smaller versions can be attempted and external laboratories offer this as a service. However, for reasons as discussed above, even this approach may not find a noticeably improved solution to the issue of CaMLD based construct protein stability. As discussed in chapter 5, a compromise for the production of usable protein has to be reached at some point.

The use of other techniques in functional assays would allow further investigation into the interaction between the two RyR2 domains and CaM. A number of techniques were considered though it was considered that isothermal titration calorimetry (ITC) or surface plasmon resonance (SPR) would offer the most possibilities. The use of SPR had been considered for the purposes of competition assays. This technique would allow

CaM or CaMLD to be flowed over CaMBD, bound to a chip, under changing Ca^{2+} concentrations to monitor binding affinities. Once these had been established then competition between CaM and CaMLD under varying conditions could be examined. There were some concerns over the methods to attach the protein to the chip. Most of the chemical binding options as the acidic conditions required would denature the proteins. One of the tagged based systems would probably provide the solution, the presence of tags did not appear to affect the interaction between CaMLD and CaMBD though this could be test more fully be examination of binding affinities under control condition experiments using SPR. It may also be possible to examine if the RIH associated domain plays any role in binding for CaMLD with this technique by the use of the range of CaMLD constructs generated for this project.

Structural biology was planned as a significant content of the original thesis. Prior to commencement cryo-EM structures of RyR with predictions of the locations of domains had been published. During the course of the project a number of papers containing crystal x-ray derived structures were published. Notable amongst these were the work of the Van Petegem group examining the N terminal of RyR and the structural implications of mutations within this region (Lobo & Van Petegem 2009), (Tung et al. 2010). These papers illustrate the additional level of understanding that can be gained when the physical impact of a mutation can be seen upon the structure. It was hoped that further progress to equivalent result in the domains studied in this project could be made but as discussed in chapter 5 this did not prove possible. This is clearly an area for further exploration beyond this project. Gaining a structure of CaMLD bound to CaMBD would further help to see any similarities with CaM binding and help to clearly identify key residues involved in CaMLD binding. Comparison of structures involving mutant forms of CaMLD would greatly aid in the understanding of CPVT mutations in this region. While likely to be resource and time heavy this approach would offer potentially great benefits.

One of the techniques that can be used to identify more likely crystallisation buffer conditions is the thermal shift assay. The strength of a fluorescent signal is monitored as the protein sample is heated in a variety of buffer conditions. As the protein unfolds so the signal is quenched, the more stable the protein is in a certain condition then the higher the temperature as which it denatures. This technique was trialled, using sypro

orange (Sigma) in a qPCR system, for the cleaved tag proteins of CaMLD and CaMBD in attempt to identify better storage buffers. Unfortunately owing to the existing stability issues of CaMLD it did not prove possible to identify an improved buffer system. With regards to crystallography work, if a protein sample is more stable under a certain condition then there is more probability that it may crystallise in that condition than one in which it degrades though this is not guaranteed. However it should be noted that minor degradation, particularly the loss of N and/or C terminal residues, may result in a more stable form of protein. This is same theory behind limited tryptic digests of potential crystallisation target proteins as the aim is to remove away "flappy" amino acids on ends of protein. This refers back to the ideas regarding domain boundaries and may offer an additional layer of identification.

Part of the process to identify further mutations to study could be addressed by development of the model system shown in chapter 4. This approach could potentially offer a quicker way of identifying the potential impact of mutations on the structure so identifying more likely targets prior to the commencement of laboratory based studies. Single or multiple mutation models could be produced as required and the impact on model structure and energy levels examined. The limitations of this approach are the capabilities of the software, which are steadily increasing as additional experimental data is accrued, the users interpretation and choice of route to direct the software, and the processing power and time required. The latter concern is becoming less of an issue though as increased computing power becomes available then more complex, and potentially realistic, simulations may be performed.

Bibliography

Amador, F.J. et al., 2009. Crystal structure of type I ryanodine receptor amino-terminal beta-trefoil domain reveals a disease-associated mutation "hot spot" loop. *Proceedings of the National Academy of Sciences of the United States of America*, 106(27), pp.11040–11044.

Anfinsen, C.B., 1973. Principles that govern the folding of protein chains. *Science*, 181(4096), pp.223–30.

Anon, 1981. Timolol-induced reduction in mortality and reinfarction in patients surviving acute myocardial infarction. *N Engl J Med*, 304(14), pp.801–7.

Arnáiz-Cot, J.J. et al., 2013. Cardiac calcium signaling pathologies associated with defective calmodulin regulation of type 2 ryanodine receptor. *The Journal of physiology*.

Balshaw, D.M. et al., 2001. Calmodulin binding and inhibition of cardiac muscle calcium release channel (ryanodine receptor). *The Journal of biological chemistry*, 276(23), pp.20144–20153.

Berridge, M.J., Bootman, M.D. & Lipp, P., 1998. Calcium--a life and death signal. *Nature*, 395(6703), pp.645–648.

Bers, D.M., 2002. Cardiac excitation-contraction coupling. *Nature*, 415(6868), pp.198–205.

Bhat, M.B. et al., 1997. Functional calcium release channel formed by the carboxyl-terminal portion of ryanodine receptor. *Biophysical journal*, 73(3), pp.1329–1336.

Bienvenut, W.V. et al., 2002. Matrix-assisted laser desorption/ionization-tandem mass spectrometry with high resolution and sensitivity for identification and characterization of proteins. *Proteomics*, 2(7), pp.868–876.

Black, D.J., Leonard, J. & Persechini, A., 2006. Biphasic Ca2+-dependent switching in a calmodulin-IQ domain complex. *Biochemistry*, 45(22), pp.6987–6995.

Blayney, L.M. & Lai, F.A., 2009. Ryanodine receptor-mediated arrhythmias and sudden cardiac death. *Pharmacology & therapeutics*, 123(2), pp.151–177.

Bootman, M.D. et al., 2001. Calcium signalling--an overview. *Seminars in cell & developmental biology*, 12(1), pp.3–10.

Brandt, N.R. et al., 1992. Mapping of the calpain proteolysis products of the junctional foot protein of the skeletal muscle triad junction. *The Journal of membrane biology*, 127(1), pp.35–47.

Braunwald, E., 2007. *Braunwald's Heart Disease : A textbook of Cardiovascular*

Disease 8th ed., Philadelphia: Saunders Elsevier.

Brennan, P. et al., 2009. Quantitative nuclear proteomics reveals new phenotypes altered in lymphoblastoid cells. *PROTEOMICS – Clinical Applications*, 3(3), pp.359–369.

Brooks, B.R. et al., 2009. CHARMM: the biomolecular simulation program. *J Comput Chem*, 30(10), pp.1545–614.

Callaway, C. et al., 1994. Localization of the high and low affinity [3H]ryanodine binding sites on the skeletal muscle Ca2+ release channel. *The Journal of biological chemistry*, 269(22), pp.15876–15884.

Campbell, K.P., Franzini-Armstrong, C. & Shamoo, A.E., 1980. Further characterization of light and heavy sarcoplasmic reticulum vesicles. Identification of the "sarcoplasmic reticulum feet" associated with heavy sarcoplasmic reticulum vesicles. *Biochimica et biophysica acta*, 602(1), pp.97–116.

Capes, E.M., Loaiza, R. & Valdivia, H.H., 2011. Ryanodine receptors. *Skeletal muscle*, 1(1), p.18.

Carpenter, T., Khalid, S. & Sansom, M.S.P., 2007. A multidomain outer membrane protein from Pasteurella multocida: modelling and simulation studies of PmOmpA. *Biochimica et biophysica acta*, 1768(11), pp.2831–2840.

Case, R.M. et al., 2007. Evolution of calcium homeostasis: from birth of the first cell to an omnipresent signalling system. *Cell calcium*, 42(4-5), pp.345–350.

Chen, S.R. et al., 1998. Molecular identification of the ryanodine receptor Ca2+ sensor. *The Journal of biological chemistry*, 273(24), pp.14675–14678.

Chou, P.Y. & Fasman, G.D., 1974. Prediction of protein conformation. *Biochemistry*, 13(2), pp.222–245.

Cornea, R.L. et al., 2009. FRET-based mapping of calmodulin bound to the RyR1 Ca2+ release channel. *Proceedings of the National Academy of Sciences of the United States of America*, 106(15), pp.6128–6133.

D'Cruz, L.G., Yin, C.C., Williams, A.J., Lai, F.A., 2009. Insights into the Three-Dimensional Organisation of the Ryanodine Receptor. In K. Mikoshiba, ed. *Handbook of Neurochemistry and Molecular Neurobiology*. Springer, pp. 463–488.

Davison, F.D., D'Cruz, L.G. & McKenna, W.J., 2000. Molecular motors in the heart. *Essays Biochem*, 35, pp.145–58.

Dill, K.A., 1985. Theory for the folding and stability of globular proteins. *Biochemistry*, 24(6), pp.1501–9.

Du, G.G. et al., 2001. Functional characterization of mutants in the predicted pore region of the rabbit cardiac muscle Ca(2+) release channel (ryanodine receptor isoform 2). *The Journal of biological chemistry*, 276(34), pp.31760–31771.

Du, G.G., Khanna, V.K. & MacLennan, D.H., 2000. Mutation of divergent region 1 alters caffeine and Ca(2+) sensitivity of the skeletal muscle Ca(2+) release channel (ryanodine receptor). *The Journal of biological chemistry*, 275(16), pp.11778–11783.

Dümmler, A., Lawrence, A.-M. & de Marco, A., 2005. Simplified screening for the detection of soluble fusion constructs expressed in E. coli using a modular set of vectors. *Microbial cell factories*, 4, p.34.

Endo, M., 2011. Reiji Natori, Setsuro Ebashi, and excitation-contraction coupling. *Progress in biophysics and molecular biology*, 105(3), pp.129–133.

Favero, T.G., Zable, A.C. & Abramson, J.J., 1995. Hydrogen peroxide stimulates the Ca2+ release channel from skeletal muscle sarcoplasmic reticulum. *The Journal of biological chemistry*, 270(43), pp.25557–25563.

Finn, B.E. et al., 1995. Calcium-induced structural changes and domain autonomy in calmodulin. *Nature structural biology*, 2(9), pp.777–783.

Franzini-Armstrong, C., 1970. STUDIES OF THE TRIAD : I. Structure of the Junction in Frog Twitch Fibers. *The Journal of cell biology*, 47(2), pp.488–499.

Fraternali, F. & Pastore, A., 1999. Modularity and homology: modelling of the type II module family from titin. *Journal of molecular biology*, 290(2), pp.581–593.

Fruen, B.R. et al., 2000. Differential Ca(2+) sensitivity of skeletal and cardiac muscle ryanodine receptors in the presence of calmodulin. *American journal of physiology. Cell physiology*, 279(3), pp.C724–733.

Gangopadhyay, J.P. & Ikemoto, N., 2011. Aberrant interaction of calmodulin with the ryanodine receptor develops hypertrophy in the neonatal cardiomyocyte. *The Biochemical journal*, 438(2), pp.379–387.

Gangopadhyay, J.P. & Ikemoto, N., 2008. Interaction of the Lys(3614)-Asn(3643) calmodulin-binding domain with the Cys(4114)-Asn(4142) region of the type 1 ryanodine receptor is involved in the mechanism of Ca2+/agonist-induced channel activation. *The Biochemical journal*, 411(2), pp.415–423.

Gangopadhyay, J.P. & Ikemoto, N., 2006. Role of the Met3534-Ala4271 region of the ryanodine receptor in the regulation of Ca2+ release induced by calmodulin binding domain peptide. *Biophys J*, 90(6), pp.2015–26.

Gao, L. et al., 1997. Evidence for a role of C-terminal amino acid residues in

skeletal muscle Ca2+ release channel (ryanodine receptor) function. *FEBS letters*, 412(1), pp.223–226.

George, C.H. et al., 2006. Arrhythmogenic mutation-linked defects in ryanodine receptor autoregulation reveal a novel mechanism of Ca2+ release channel dysfunction. *Circulation research*, 98(1), pp.88–97.

George, C.H. et al., 2004a. Ryanodine receptor regulation by intramolecular interaction between cytoplasmic and transmembrane domains. *Molecular biology of the cell*, 15(6), pp.2627–2638.

George, C.H. et al., 2004b. Ryanodine receptor regulation by intramolecular interaction between cytoplasmic and transmembrane domains. *Molecular biology of the cell*, 15(6), pp.2627–2638.

George, C.H., Higgs, G.V. & Lai, F.A., 2003. Ryanodine receptor mutations associated with stress-induced ventricular tachycardia mediate increased calcium release in stimulated cardiomyocytes. *Circulation research*, 93(6), pp.531–540.

Geourjon, C. & Deleage, G., 1995. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput Appl Biosci*, 11(6), pp.681–4.

Geourjon, C. & Deléage, G., 1995. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Computer applications in the biosciences: CABIOS*, 11(6), pp.681–684.

Gillespie, D., Chen, H. & Fill, M., 2012. Is ryanodine receptor a calcium or magnesium channel? Roles of K+ and Mg2+ during Ca2+ release. *Cell calcium*, 51(6), pp.427–433.

Gómez, A.M. et al., 2004. FKBP12.6 overexpression decreases Ca2+ spark amplitude but enhances [Ca2+]i transient in rat cardiac myocytes. *American journal of physiology. Heart and circulatory physiology*, 287(5), pp.H1987–1993.

Grabarek, Z., 2005. Structure of a trapped intermediate of calmodulin: calcium regulation of EF-hand proteins from a new perspective. *Journal of molecular biology*, 346(5), pp.1351–1366.

Greenfield, N.J., 2006. Using circular dichroism spectra to estimate protein secondary structure. *Nature protocols*, 1(6), pp.2876–2890.

Grunwald, R. & Meissner, G., 1995. Lumenal sites and C terminus accessibility of the skeletal muscle calcium release channel (ryanodine receptor). *The Journal of biological chemistry*, 270(19), pp.11338–11347.

Györke, I. et al., 2004. The role of calsequestrin, triadin, and junctin in conferring cardiac ryanodine receptor responsiveness to luminal calcium. *Biophysical journal*, 86(4), pp.2121–2128.

Haiech, J. et al., 2011. Revisiting intracellular calcium signaling semantics. *Biochimie*, 93(12), pp.2029–2037.

Hakamata, Y. et al., 1992. Primary structure and distribution of a novel ryanodine receptor/calcium release channel from rabbit brain. *FEBS letters*, 312(2-3), pp.229–235.

Hayashi, T. et al., 2009. Three-dimensional electron microscopy reveals new details of membrane systems for Ca2+ signaling in the heart. *J Cell Sci*, 122(Pt 7), pp.1005–13.

el-Hayek, R., Yano, M. & Ikemoto, N., 1995. A conformational change in the junctional foot protein is involved in the regulation of Ca2+ release from sarcoplasmic reticulum. Studies on polylysine-induced Ca2+ release. *The Journal of biological chemistry*, 270(26), pp.15634–15638.

Herzberg, O. & James, M.N., 1985. Structure of the calcium regulatory muscle protein troponin-C at 2.8 A resolution. *Nature*, 313(6004), pp.653–659.

Herzberg, O., Moult, J. & James, M.N., 1986. A model for the Ca2+-induced conformational transition of troponin C. A trigger for muscle contraction. *The Journal of biological chemistry*, 261(6), pp.2638–2644.

Hill, E.E., Morea, V. & Chothia, C., 2002. Sequence conservation in families whose members have little or no sequence similarity: the four-helical cytokines and cytochromes. *Journal of molecular biology*, 322(1), pp.205–233.

Huang, X. et al., 2013. Two potential calmodulin binding sequences in the ryanodine receptor contribute to a mobile, intra-subunit calmodulin binding domain. *Journal of cell science*.

Humphrey, W., Dalke, A. & Schulten, K., 1996. VMD: visual molecular dynamics. *Journal of molecular graphics*, 14(1), pp.33–38, 27–28.

Ikura, M., Clore, G.M., et al., 1992. Solution structure of a calmodulin-target peptide complex by multidimensional NMR. *Science (New York, N.Y.)*, 256(5057), pp.632–638.

Ikura, M., Barbato, G., et al., 1992. Solution structure of calmodulin and its complex with a myosin light chain kinase fragment. *Cell calcium*, 13(6-7), pp.391–400.

Iwasaki, Y.K. et al., Atrial fibrillation pathophysiology: implications for management. *Circulation*, 124(20), pp.2264–74.

Jayaraman, T. et al., 1992. FK506 binding protein associated with the calcium release channel (ryanodine receptor). *The Journal of biological chemistry*, 267(14), pp.9474–9477.

Jones, J.-L. et al., 2005. Ryanodine receptor binding to FKBP12 is modulated by channel activation state. *Journal of cell science*, 118(Pt 20), pp.4613–

4619.

Jundi, H., 2010. *The I-domain of the Ryanodine receptor*. Cardiff University.

Karplus, M. & Weaver, D.L., 1994. Protein folding dynamics: the diffusion-collision model and experimental data. *Protein Sci*, 3(4), pp.650–68.

Kawasaki, H. & Kretsinger, R.H., 2012. Analysis of the movements of helices in EF-hands. *Proteins*, 80(11), pp.2592–2600.

Kelly, S.M., Jess, T.J. & Price, N.C., 2005. How to study proteins by circular dichroism. *Biochimica et biophysica acta*, 1751(2), pp.119–139.

Kimlicka, L. et al., 2013. Disease mutations in the ryanodine receptor N-terminal region couple to a mobile intersubunit interface. *Nature communications*, 4, p.1506.

Kragelund, B.B. et al., 1998. Hydrophobic core substitutions in calbindin D9k: effects on Ca2+ binding and dissociation. *Biochemistry*, 37(25), pp.8926–8937.

Kretsinger, R.H. & Nockolds, C.E., 1973. Carp muscle calcium-binding protein. II. Structure determination and general description. *The Journal of biological chemistry*, 248(9), pp.3313–3326.

Lai, F.A. et al., 1988. Evidence for a Ca2+ channel within the ryanodine receptor complex from cardiac sarcoplasmic reticulum. *Biochemical and biophysical research communications*, 151(1), pp.441–449.

Lai, F.A. et al., 1989. The ryanodine receptor-Ca2+ release channel complex of skeletal muscle sarcoplasmic reticulum. Evidence for a cooperatively coupled, negatively charged homotetramer. *The Journal of biological chemistry*, 264(28), pp.16776–16785.

Levinthal, C., 1969. *How to fold graciously.*, Monticello, Illinois: University of Illinois Press.

Linse, S. et al., 1991. Electrostatic contributions to the binding of Ca2+ in calbindin D9k. *Biochemistry*, 30(1), pp.154–162.

Liu, Z. et al., 2005. Localization of a disease-associated mutation site in the three-dimensional structure of the cardiac muscle ryanodine receptor. *The Journal of biological chemistry*, 280(45), pp.37941–37947.

Liu, Z. et al., 2002. Three-dimensional reconstruction of the recombinant type 2 ryanodine receptor and localization of its divergent region 1. *The Journal of biological chemistry*, 277(48), pp.46712–46719.

Lobo, P.A. et al., 2011. The deletion of exon 3 in the cardiac ryanodine receptor is rescued by β strand switching. *Structure (London, England: 1993)*, 19(6), pp.790–798.

Lobo, P.A. & Van Petegem, F., 2009. Crystal structures of the N-terminal domains of cardiac and skeletal muscle ryanodine receptors: insights into disease mutations. *Structure (London, England: 1993)*, 17(11), pp.1505–1514.

Ludtke, S.J. et al., 2005. The pore structure of the closed RyR1 channel. *Structure (London, England: 1993)*, 13(8), pp.1203–1211.

MacKrill, J.J., 1999. Protein-protein interactions in intracellular Ca2+-release channel function. *The Biochemical journal*, 337 (Pt 3), pp.345–361.

MacQuaide, N., Dempster, J. & Smith, G.L., 2007. Measurement and modeling of Ca2+ waves in isolated rabbit ventricular cardiomyocytes. *Biophysical journal*, 93(7), pp.2581–2595.

Marks, A.R. et al., 1989. Molecular cloning and characterization of the ryanodine receptor/junctional channel complex cDNA from skeletal muscle sarcoplasmic reticulum. *Proceedings of the National Academy of Sciences of the United States of America*, 86(22), pp.8683–8687.

Marx, S.O. et al., 2000a. PKA phosphorylation dissociates FKBP12.6 from the calcium release channel (ryanodine receptor): defective regulation in failing hearts. *Cell*, 101(4), pp.365–376.

Marx, S.O. et al., 2000b. PKA phosphorylation dissociates FKBP12.6 from the calcium release channel (ryanodine receptor): defective regulation in failing hearts. *Cell*, 101(4), pp.365–376.

Mathur, A., Shankaracharya & Vidyarthi, A.S., 2011. SWIFT MODELLER: a Java based GUI for molecular modeling. *Journal of molecular modeling*, 17(10), pp.2601–2607.

Maune, J.F., Klee, C.B. & Beckingham, K., 1992. Ca2+ binding and conformational change in two series of point mutations to the individual Ca(2+)-binding sites of calmodulin. *The Journal of biological chemistry*, 267(8), pp.5286–5295.

Maximciuc, A.A. et al., 2006. Complex of calmodulin with a ryanodine receptor target reveals a novel, flexible binding mode. *Structure (London, England: 1993)*, 14(10), pp.1547–1556.

McCarthy, T.V., Quane, K.A. & Lynch, P.J., 2000. Ryanodine receptor mutations in malignant hyperthermia and central core disease. *Human mutation*, 15(5), pp.410–417.

Medzihradszky, K.F. et al., 2000. The characteristics of peptide collision-induced dissociation using a high-performance MALDI-TOF/TOF tandem mass spectrometer. *Analytical chemistry*, 72(3), pp.552–558.

Meissner, G., Rousseau, E. & Lai, F.A., 1989. Structural and functional correlation of the trypsin-digested Ca2+ release channel of skeletal muscle sarcoplasmic reticulum. *The Journal of biological chemistry*,

264(3), pp.1715–1722.

Mohamed, U., Napolitano, C. & Priori, S.G., 2007. Molecular and electrophysiological bases of catecholaminergic polymorphic ventricular tachycardia. *Journal of cardiovascular electrophysiology*, 18(7), pp.791–797.

Moore, C.P. et al., 1999. Apocalmodulin and Ca2+ calmodulin bind to the same region on the skeletal muscle Ca2+ release channel. *Biochemistry*, 38(26), pp.8532–8537.

Nagasaki, K. & Fleischer, S., 1988. Ryanodine sensitivity of the calcium release channel of sarcoplasmic reticulum. *Cell calcium*, 9(1), pp.1–7.

Ben-Naim, A., Levinthal's question revisited, and answered. *J Biomol Struct Dyn*, 30(1), pp.113–24.

Nakai, J. et al., 1990. Primary structure and functional expression from cDNA of the cardiac ryanodine receptor/calcium release channel. *FEBS letters*, 271(1-2), pp.169–177.

Nelson, M.R., Chagot, B. & Chazin, W.J., 2001. EF-hand Calcium-binding Proteins. In *eLS*. John Wiley & Sons, Ltd. Available at: http://onlinelibrary.wiley.com/doi/10.1002/9780470015902.a0003056.pub 2/abstract [Accessed August 21, 2013].

Niggli, E. et al., 2013. Posttranslational modifications of cardiac ryanodine receptors: Ca(2+) signaling and EC-coupling. *Biochimica et biophysica acta*, 1833(4), pp.866–875.

Norman, M.W. & McKenna, W.J., 1999. Arrhythmogenic right ventricular cardiomyopathy: perspectives on disease. *Zeitschrift für Kardiologie*, 88(8), pp.550–554.

Nyegaard, M. et al., 2012. Mutations in calmodulin cause ventricular tachycardia and sudden cardiac death. *American journal of human genetics*, 91(4), pp.703–712.

Orlova, E.V. et al., 1996a. Two structural configurations of the skeletal muscle calcium release channel. *Nature structural biology*, 3(6), pp.547–552.

Orlova, E.V. et al., 1996b. Two structural configurations of the skeletal muscle calcium release channel. *Nature structural biology*, 3(6), pp.547–552.

Otsu, K. et al., 1990. Molecular cloning of cDNA encoding the Ca2+ release channel (ryanodine receptor) of rabbit cardiac muscle sarcoplasmic reticulum. *The Journal of biological chemistry*, 265(23), pp.13472–13483.

Park, J. et al., 1998. Sequence comparisons using multiple sequences detect three times as many remote homologues as pairwise methods. *Journal of molecular biology*, 284(4), pp.1201–1210.

Perkins, D.N. et al., 1999. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, 20(18), pp.3551–3567.

Ponting, C.P., 2000. Novel repeats in ryanodine and IP3 receptors and protein O-mannosyltransferases. *Trends in Biochemical Sciences*, 25(2), pp.48–50.

Priori, S.G. et al., 2001. Mutations in the cardiac ryanodine receptor gene (hRyR2) underlie catecholaminergic polymorphic ventricular tachycardia. *Circulation*, 103(2), pp.196–200.

Proenza, C. et al., 2002. Identification of a region of RyR1 that participates in allosteric coupling with the alpha(1S) (Ca(V)1.1) II-III loop. *The Journal of biological chemistry*, 277(8), pp.6530–6535.

Radermacher, M. et al., 1994. Cryo-electron microscopy and three-dimensional reconstruction of the calcium release channel/ryanodine receptor from skeletal muscle. *The Journal of cell biology*, 127(2), pp.411–423.

Reddy, B.V.B. & Kaznessis, Y.N., 2007. Use of secondary structural information and C alpha-C alpha distance restraints to model protein structures with MODELLER. *Journal of biosciences*, 32(5), pp.929–936.

Rhoads, A.R. & Friedberg, F., 1997. Sequence motifs for calmodulin recognition. *FASEB journal: official publication of the Federation of American Societies for Experimental Biology*, 11(5), pp.331–340.

Rodney, G.G. et al., 2001. Calcium binding to calmodulin leads to an N-terminal shift in its binding site on the ryanodine Receptor. *The Journal of biological chemistry*, 276(3), pp.2069–2074.

Sali, A. & Blundell, T.L., 1993. Comparative protein modelling by satisfaction of spatial restraints. *Journal of molecular biology*, 234(3), pp.779–815.

Sali, A. & Blundell, T.L., 1993. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol*, 234(3), pp.779–815.

Samsó, M. & Wagenknecht, T., 2002. Apocalmodulin and Ca2+-calmodulin bind to neighboring locations on the ryanodine receptor. *The Journal of biological chemistry*, 277(2), pp.1349–1353.

Samsó, M., Wagenknecht, T. & Allen, P.D., 2005. Internal structure and visualization of transmembrane domains of the RyR1 calcium release channel by cryo-EM. *Nature structural & molecular biology*, 12(6), pp.539–544.

Sánchez, R. & Sali, A., 1997. Evaluation of comparative protein structure modeling by MODELLER-3. *Proteins*, Suppl 1, pp.50–58.

Saucerman, J.J. & Bers, D.M., 2012. Calmodulin binding proteins provide domains of local Ca2+ signaling in cardiac myocytes. *Journal of*

molecular and cellular cardiology, 52(2), pp.312–316.

Serysheva, I.I. et al., 1995. Electron cryomicroscopy and angular reconstitution used to visualize the skeletal muscle calcium release channel. *Nature structural biology*, 2(1), pp.18–24.

Serysheva, I.I. et al., 1999. Structure of the skeletal muscle calcium release channel activated with Ca2+ and AMP-PCP. *Biophysical journal*, 77(4), pp.1936–1944.

Sharma, M.R. et al., 1998. Cryoelectron microscopy and image analysis of the cardiac ryanodine receptor. *The Journal of biological chemistry*, 273(29), pp.18429–18434.

Sharma, M.R. et al., 2000. Three-dimensional structure of ryanodine receptor isoform three in two conformational states as visualized by cryo-electron microscopy. *The Journal of biological chemistry*, 275(13), pp.9485–9491.

Sharma, P. et al., 2012. Structural determination of the phosphorylation domain of the ryanodine receptor. *The FEBS journal*, 279(20), pp.3952–3964.

Shaw, G.S. et al., 1994. Relative stabilities of synthetic peptide homo- and heterodimeric troponin-C domains. *Protein science: a publication of the Protein Society*, 3(7), pp.1010–1019.

Shen, M.-Y. & Sali, A., 2006. Statistical potential for assessment and prediction of protein structures. *Protein science: a publication of the Protein Society*, 15(11), pp.2507–2524.

Shevchenko, A. et al., 1996. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Analytical chemistry*, 68(5), pp.850–858.

Smyth, D.R. et al., 2003. Crystal structures of fusion proteins with large-affinity tags. *Protein science: a publication of the Protein Society*, 12(7), pp.1313–1322.

Sorrentino, V. & Volpe, P., 1993. Ryanodine receptors: how many, where and why? *Trends in pharmacological sciences*, 14(3), pp.98–103.

Sreerama, N. & Woody, R.W., 2000. Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Analytical biochemistry*, 287(2), pp.252–260.

Stewart, R., Zissimopoulos, S. & Lai, F.A., 2003. Oligomerization of the cardiac ryanodine receptor C-terminal tail. *The Biochemical journal*, 376(Pt 3), pp.795–799.

Strynadka, N.C. & James, M.N., 1989. Crystal structures of the helix-loop-helix calcium-binding proteins. *Annual review of biochemistry*, 58, pp.951–998.

Takeshima, H. et al., 1989. Primary structure and expression from

complementary DNA of skeletal muscle ryanodine receptor. *Nature*, 339(6224), pp.439–445.

Tan, R.Y., Mabuchi, Y. & Grabarek, Z., 1996. Blocking the Ca2+-induced conformational transitions in calmodulin with disulfide bonds. *The Journal of biological chemistry*, 271(13), pp.7479–7483.

Thomas, N.L. et al., 2007. Ryanodine receptor mutations in arrhythmias: advances in understanding the mechanisms of channel dysfunction. *Biochemical Society transactions*, 35(Pt 5), pp.946–951.

Thomas, N.L., George, C.H. & Lai, F.A., 2006. Role of ryanodine receptor mutations in cardiac pathology: more questions than answers? *Biochemical Society transactions*, 34(Pt 5), pp.913–918.

Tripathy, A. et al., 1995. Calmodulin activation and inhibition of skeletal muscle Ca2+ release channel (ryanodine receptor). *Biophysical journal*, 69(1), pp.106–119.

Tung, C.-C. et al., 2010. The amino-terminal disease hotspot of ryanodine receptors forms a cytoplasmic vestibule. *Nature*, 468(7323), pp.585–588.

Valdivia, H.H., 2012. Ryanodine receptor phosphorylation and heart failure: phasing out S2808 and "criminalizing" S2814. *Circulation research*, 110(11), pp.1398–1402.

Wagenknecht, T. & Radermacher, M., 1995. Three-dimensional architecture of the skeletal muscle ryanodine receptor. *FEBS letters*, 369(1), pp.43–46.

Wang, R. et al., 2007. Localization of an NH(2)-terminal disease-causing mutation hot spot to the "clamp" region in the three-dimensional structure of the cardiac ryanodine receptor. *The Journal of biological chemistry*, 282(24), pp.17785–17793.

Witcher, D.R. et al., 1994. Photoaffinity labeling of the ryanodine receptor/Ca2+ release channel with an azido derivative of ryanodine. *The Journal of biological chemistry*, 269(18), pp.13076–13079.

Wójcik, J. et al., 1997. Isolated calcium-binding loops of EF-hand proteins can dimerize to form a native-like structure. *Biochemistry*, 36(4), pp.680–687.

Wright, N.T. et al., 2008. S100A1 and calmodulin compete for the same binding site on ryanodine receptor. *The Journal of biological chemistry*, 283(39), pp.26676–26683.

Xiang, Z., 2006. Advances in homology protein structure modeling. *Current protein & peptide science*, 7(3), pp.217–227.

Xiao, R.P. et al., 1997. The immunophilin FK506-binding protein modulates Ca2+ release channel closure in rat heart. *The Journal of physiology*, 500 (Pt 2), pp.343–354.

Xiong, L. et al., 2006. A Ca2+-binding domain in RyR1 that interacts with the calmodulin binding site and modulates channel activity. *Biophysical journal*, 90(1), pp.173–182.

Xiong, L. et al., 2006. A Ca2+-binding domain in RyR1 that interacts with the calmodulin binding site and modulates channel activity. *Biophys J*, 90(1), pp.173–82.

Xu, X. et al., 2000. Molecular cloning of cDNA encoding a drosophila ryanodine receptor and functional studies of the carboxyl-terminal calcium release channel. *Biophysical journal*, 78(3), pp.1270–1281.

Yamaguchi, N. et al., 2013. Cardiac hypertrophy associated with impaired regulation of cardiac ryanodine receptor by calmodulin and S100A1. *American journal of physiology. Heart and circulatory physiology*, 305(1), pp.H86–94.

Yamaguchi, N. et al., 2004. Different regions in skeletal and cardiac muscle ryanodine receptors are involved in transducing the functional effects of calmodulin. *The Journal of biological chemistry*, 279(35), pp.36433–36439.

Yamaguchi, N. et al., 2003. Molecular basis of calmodulin binding to cardiac muscle Ca(2+) release channel (ryanodine receptor). *The Journal of biological chemistry*, 278(26), pp.23480–23486.

Yamaguchi, N. et al., 2003. Molecular basis of calmodulin binding to cardiac muscle Ca(2+) release channel (ryanodine receptor). *J Biol Chem*, 278(26), pp.23480–6.

Yamazawa, T. et al., 1997. A region of the ryanodine receptor critical for excitation-contraction coupling in skeletal muscle. *The Journal of biological chemistry*, 272(13), pp.8161–8164.

Yang, H. & Smith, D.L., 1997. Kinetics of cytochrome c folding examined by hydrogen exchange and mass spectrometry. *Biochemistry*, 36(48), pp.14992–9.

Yuchi, Z., Lau, K. & Van Petegem, F., 2012. Disease mutations in the ryanodine receptor central region: crystal structures of a phosphorylation hot spot domain. *Structure (London, England: 1993)*, 20(7), pp.1201–1211.

Yusuf, S. et al., 1985. Beta blockade during and after myocardial infarction: an overview of the randomized trials. *Prog Cardiovasc Dis*, 27(5), pp.335–71.

Zhang, M., Tanaka, T. & Ikura, M., 1995. Calcium-induced conformational transition revealed by the solution structure of apo calmodulin. *Nature structural biology*, 2(9), pp.758–767.

Zhu, X. et al., 2004. The calmodulin binding region of the skeletal ryanodine receptor acts as a self-modulatory domain. *Cell Calcium*, 35(2), pp.165–

77.

Zissimopoulos, S. et al., 2012a. Disparities in the association of the ryanodine receptor and the FK506-binding proteins in mammalian heart. *Journal of cell science*, 125(Pt 7), pp.1759–1769.

Zissimopoulos, S. et al., 2012b. Disparities in the association of the ryanodine receptor and the FK506-binding proteins in mammalian heart. *Journal of cell science*, 125(Pt 7), pp.1759–1769.

Zissimopoulos, S. et al., 2006. Ryanodine receptor interaction with the SNARE-associated protein snapin. *Journal of cell science*, 119(Pt 11), pp.2386–2397.

Zissimopoulos, S. & Lai, F.A., 2005a. Central domain of the human cardiac muscle ryanodine receptor does not mediate interaction with FKBP12.6. *Cell biochemistry and biophysics*, 43(2), pp.203–219.

Zissimopoulos, S. & Lai, F.A., 2005b. Interaction of FKBP12.6 with the cardiac ryanodine receptor C-terminal domain. *The Journal of biological chemistry*, 280(7), pp.5475–5485.

Zwanzig, R., Szabo, A. & Bagchi, B., 1992. Levinthal's paradox. *Proc Natl Acad Sci U S A*, 89(1), pp.20–2.