Investigating the Conformational Flexibility of Calmodulin

By

Nicholas Paul Barton

A Thesis submitted for the Degree of Doctor of Philosophy

Department of Chemistry,

University of York,

September 2002

Abstract:

Calmodulin (CaM) is a ubiquitous 148 KDa highly acidic protein which responds to changes in intracellular Ca^{2+} concentration causing to activate its targets in a range of different systems. CaM interacts with a variety of different targets, which are highly diverse in nature, through high-affinity interactions, normally in a Ca^{2+}-dependent manner. The investigations in this thesis aim to study the extent to which this high-affinity, low-specificity is afforded by the conformational flexibility in the protein. This conformational flexibility is investigated initially through an analysis of the available structural information and subsequently through molecular simulations. The different behaviours of the C and N-term lobes of the protein are compared throughout in order to investigate the individual conformational characteristics of the protein segments. The flexibility of the CaM binding surfaces and the nature of the interaction between CaM and its targets are also investigated.

These investigations demonstrated considerable conformational flexibility in both lobes of CaM, particularly in the N-term lobe of the protein and the conformations of the lobes were shown to be dependent on Ca^{2+} and target association state. Observed conformational modes were shown to be well represented by the calculated normal modes of vibration. Molecular dynamics simulations of the system were shown to be highly sensitive to the initial conditions, which suggests that important considerations need to be addressed when performing simulations on such environmentally sensitive systems.

Acknowledgements:

The completion of this Thesis would not have been possible without the support assistance and consideration of the following people.

Firstly I would like to thank my supervisor, Dr. Leo Caves for the original idea for the project and his continuous assistance throughout. As a mind of useful information and a source of great intellect his input was invaluable to the planning, execution and completion of this work. Also for the procurement of funds from the BBSRC to provide for my financial needs.

I would also like to thank Dr. Igor Bronstein, without whom the unpublished experimental work on the CaM-Selectin interaction would not have been possible. Igor undertook the brave step of supervising a modeller in the wetlab and followed up with hours of patience and careful explanations. Also to Stephen Martin for assistance with fluorescence studies.

There are also many additional people at the YSBL who were essential to this project. Additional intellectual and technical input to the project came from many sources including Chandra Verma, Tom Oldfield, Graham Tyr(r)ell, Derek Smith, Ryan Smith, Barry Grant and Karim Elsawy. Computer support was provided by Steve Mumford, Stuart Priest and latterly Tim Kirk with the sometimes temperamental assistance of Silicon Graphics, Compaq and Roc computers. Emmotional support was frequently provided via the coffee room through numerous people including Philippa Seavers, Lisa Fisher (Wright?), Shirley Roberts, Joanne Ladds, Julia Walton, Helen Burrell, Jean Whittingham, Simon Grist, Sally Lewis, and others above not to mention much coffee and essential cakes and particularly to Derek for de-stressing afternoon tea-breaks!

I am also eternally grateful to family and friends who have helped me through the good times and particularly the bad, especially those from York Baptist Church including Matt & Ann, Anna & Chris, Chris & Sarah, Burchs, Herberts and many many more. To my

parents for laying the groundwork for my entire education. And to my daughter Rachel, whose arrival necessitated, and somewhat complicated the completion of this thesis.

And finally to my wife Chris, for everything else, if not everything.

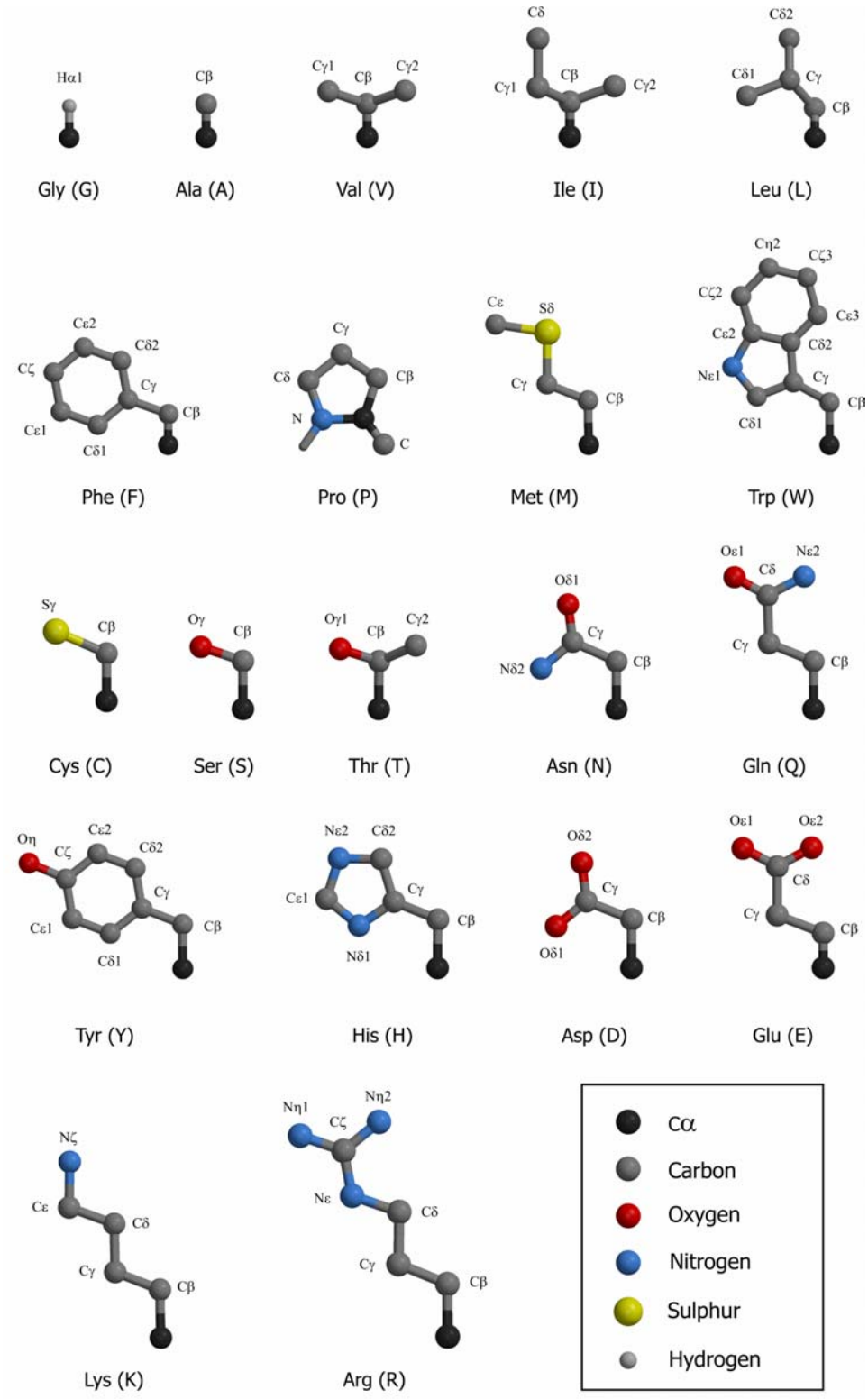
Forbid it Lord that I should boast.
Save in the death, of Christ my God!
All the vain things that charm me most-
I sacrifice them to his blood.

Table of Contents:

Table of Figures:

Table of Tables:

Amino Acid Composition & Atom Nomenclature



The twenty naturally occuring amino acid sidechains. Three letter and one letter amino acid codes are indicated along with standard atom names.

Chapter 1: Introduction

The conformation of the Ca^{2+} sensor protein, Calmodulin, is an *open and shut case* [Akke & Chazin 2001]. Calmodulin (CaM) is a highly acidic 148 amino acid alpha-helical protein which is highly conserved and ubiquitous in eukaryotic systems. CaM responds to changes in Ca^{2+} concentration through the high affinity binding of up to 4 Ca^{2+} ions (K_d 10^{-5} to 10^{-6} M) resulting in a conformational change. The Ca^{2+} activated protein is then able to activate a multitude of target proteins with high affinities (K_d 10^{-5} to 10^{-6} M) in a variety of different systems. [Crivici & Ikura 1995].

1.1 Ca^{2+} and Calmodulin

1.1.1 Ca^{2+} in Biological Systems

Calcium is the most important biological metal [Vogel 1994]. Of the four metals ions thought to be critical for life, Na^+ and K^+ primarily act as counter ions, Mg^{2+} is abundant in biological systems and is often involved in catalysis. Ca^{2+} acts as a biomineral in higher organisms however its most interesting role is as a secondary messenger in the regulation of cellular events throughout the body. In contrast to Mg^{2+} and other elements, Ca^{2+} concentration is carefully monitored with an extracellular free Ca^{2+} concentration of around 10^{-3} M and a cytoplasmic concentration of around 10^{-7} M. This considerable membranous Ca^{2+} gradient results in the potential for transient opening of calcium channels in the cell membrane to facilitate an intracellular increase in Ca^{2+} levels. Ca^{2+} activated ATPases and autoinhibitory mechanisms on the Ca^{2+} channels ensure that the intracellular concentration does not rise above 10^{-6} M. Excessive Ca^{2+} concentrations within the cell result in the activation of the Calpains, proteases which degrade intra-cellular proteins and are thought to play a role in cell death [Johnson 1990].

There are several properties of the Ca^{2+} cation which make it the signalling molecule of choice in many different processes [Williams 1985]. The higher charge density of the divalent ion enables it to form tighter complexes than monovalent ions such as Na^{+} and K^{+} and the large size of Ca^{2+} relative to Mg^{2+} makes the exchange of coordinating water molecules more achievable. The fact that Ca^{2+} can interact with uncharged carbonyl groups and does not require an octahedral coordination sphere make it possible for Ca^{2+} binding proteins to distinguish between Ca^{2+} and Mg^{2+} despite the much higher concentration of Mg^{2+} in the cell.

1.1.1.1 EF-Hand Proteins

Nearly all the proteins involved in Ca^{2+} signal transduction are members of the EF-hand family of proteins reviewed in detail by Kawasaki & Kretsinger [Kawasaki & Kretsinger 1994]. This Ca^{2+}-binding motif was first noted in the crystal structure of Carp Parvalbumin [Kretsinger & Nockolds 1973] in which consecutive helices E and F were connected by a loop containing several acidic residues. In addition to conserved acidic Ca^{2+} binding side chains in the loop, the 29-residue motif has subsequently been observed to include several conserved hydrophobic residues which help to form a hydrophobic core. This core is formed not only between the two helices but also facilitates the pairing of EF-hand domains, related by an approximate two-fold rotation, to form Ca^{2+}-binding lobes. Members of the 32 families of EF-hand proteins [Kawasaki & Kretsinger 1994] have between 2 and 8 EF-hand units normally arranged in pairs to form Ca^{2+} binding lobes. Not all EF-hands bind Ca^{2+} however; the regulatory light chain of Myosin (RLC) for example contains four EF-hands but only has one Ca^{2+} specific cationic binding site [Kwon *et al.* 1990].

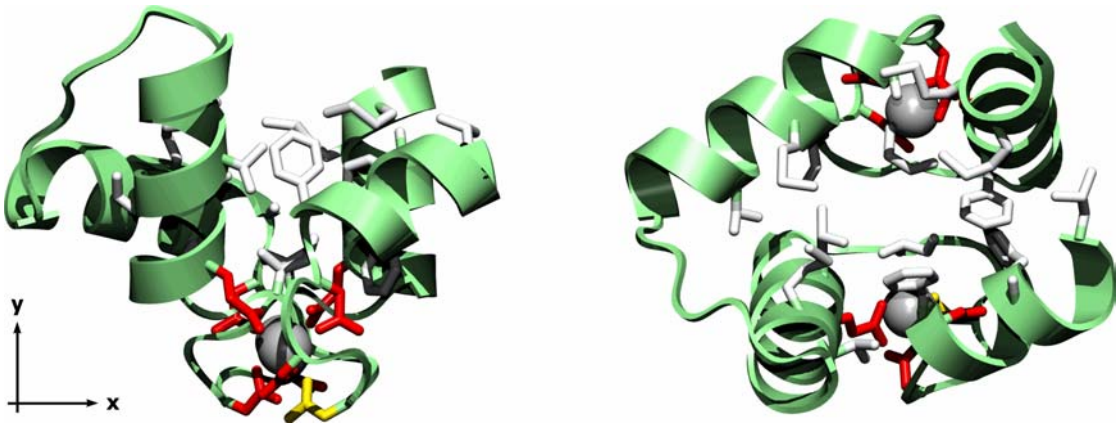


Figure 1.1: *Ribbon representation of the C-term lobe of CaM highlighting the location of conserved EF-hand residues with the liquorice representation. The conserved hydrophobic residues promoting inter-helical contacts are shown in white and the Ca^{2+}-binding residues in the EF-hands are shown in red and yellow for the acidic and polar residues respectively. The orthogonal representations show the side and top views of the lobe.*

1.1.1.2 Calmodulin and Calmodulin-like Proteins

Of the family of EF-hand calcium binding proteins, the most homologous to CaM are Troponin C (TnC) and the Regulatory (RLC) and Essential (ELC) Light Chains of Myosin. All four proteins share the same constituents with two lobes, each containing 2 EF-hands, connected by a peptide chain linker between EF-hands II and III. The light chains are an integral component of the Myosin molecular motor. Both light chains bind only one Ca^{2+} and are thought to perform a regulatory role in smooth muscle activity. In skeletal muscle however, the light chains are thought to perform a structural role and the regulation is carried out by a Troponin system. Ca^{2+} binding to the N-term lobe of Troponin C causes it to bind to Troponin I (TnI) which is responsible for inhibiting Troponin T, Tropomyosin and thus skeletal muscle activity. Consequently Ca^{2+} binding to the N-term lobe of TnC results in the activation of the myosin cross-bridge cycle [Lymn & Taylor 1971] and skeletal muscle activity.

CaM and TnC are similar in both structure and primary amino acid sequence despite having marked differences in function. In stark contrast to CaM, TnC has a single target, TnI, with which it interacts. Consequently, TnC and TnI are only found in skeletal and cardiac muscle cells and the majority of TnC is found already bound to TnI through the C-term lobe. CaM however is found throughout eukaryotic cells and due to its role in

countless systems is generally not bound to other proteins in the resting cell [Vogel 1994]. Moreover, a recent review suggested that despite its ubiquity, CaM is so much in demand in biological processes that competition for the protein is a limiting factor in its regulation of some systems [Persechini & Stemmer 2002].

It is likely that CaM and other EF-hands proteins can bind Mg^{2+} at resting cell concentrations of Ca^{2+}. A recent experiment by Allouche *et al.* attempted to characterise the differences between the Mg^{2+} and the Ca^{2+} states of Parvalbumin. The Mg^{2+} preference for an uncharged octahedral coordination sphere suggests that the acid groups in the EF-hand all interact with Mg^{2+} in a mono-dentate arrangement as observed in the Mg^{2+}-bound Parvalbumin structure [Allouche *et al.* 1999]. Computer simulation methods were used to demonstrate that such a change in coordination was also likely to occur for both CaM and Troponin C (TnC). Mg^{2+} binding to CaM however does not result in the conformational changes exhibited by Ca^{2+}-binding [Vogel 1994].

1.1.2 Function of CaM

The different roles of CaM in biochemical processes in the body are highly diverse. Amongst the many high profile roles of CaM are the effects of methionine oxidation, thought to play a critical role in protein aggregation and amyloid formation in ageing processes [Squier 2001], involvement in the formation of the Estrogen Receptor complex with DNA in the development of Breast Cancers [Biswas *et al.* 1998] and an essential role on the pathogenesis of Anthrax activity [Drum *et al.* 2002]. More generally, CaM targets can largely be separated into three different types [Crivici & Ikura 1995]:

- **Protein Kinases and Phosphatases**

There are a multitude of different CaM dependent protein kinases and phosphatases, the best characterised of which are skeletal and smooth muscle Myosin Light Chain Kinases (sk-MLCK & sm-MLCK) which regulate muscle contraction. Other enzymes include the CaM dependent protein Kinase II (CaMKII) and Calcineurin which is the major phosphatase in the brain.

- **Proteins Involved in Second-Messenger Generation**

CaM also plays a regulatory role in a variety of different signal transduction and intracellular second messenger pathways. CaM dependent enzymes include the Plasma membrane Ca^{2+}-ATPase and cyclic nucleotide phosphodiesterase.

In both of these types of system, CaM is though to interact with autoinhibitory domains in a Ca^{2+} dependent manner resulting in the exposure of an active site as illustrated in **Figure 1.2** below.

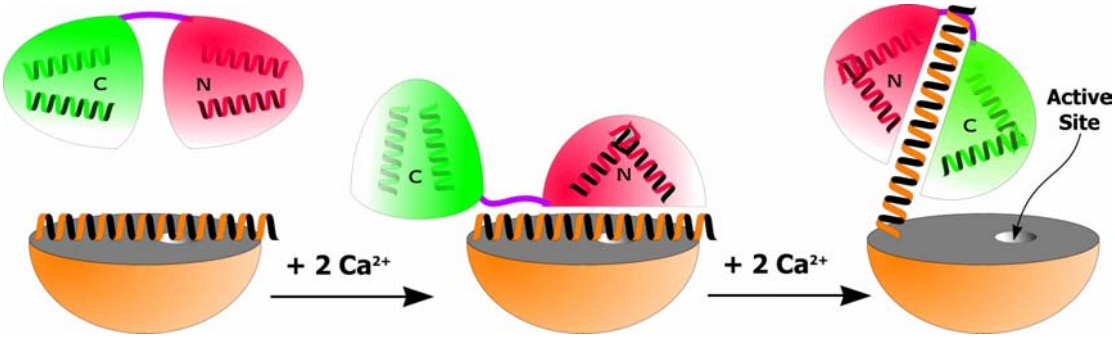


Figure 1.2: *Schematic diagram of suspected mode of action of CaM in the regulation of kinases and phosphatases and second messenger systems. Ca^{2+} binding to the N-term lobe of CaM causes it to interact with auto-inhibitory domains of its targets and subsequent Ca^{2+} binding to the C-term lobe of CaM causes a conformational change exposing an active site*

- **Proteins Involved in Regulation of Cytoskeletal Elements**

These are the least well characterised of CaM dependent systems. A Ca^{2+}-*independent* role for CaM binding to the regulatory domain of Myosin has been identified and CaM is also thought to participate in cell motility, growth and development.

1.2 Calmodulin Structure

The crystal structure of Calmodulin was first determined by Babu & co-workers in 1985 [Babu *et al.* 1985] to 3.0Å and subsequently improved to 2.2Å resolution using heavy atom derivatives and the Multiple Isomorphous Replacement technique. The resultant model suggested a dumbbell-shaped molecule which was some 65Å in length and

featured two lobes connected by a 7 turn α-helix. Both lobes were found to contain two EF-hand helix-loop-helix Ca^{2+}-binding domains connected with a short anti-parallel β-sheet. The EF-hand motif was found to be similar to that identified in Parvalbumin, the structure of which had previously been determined by Kretsinger et al. in 1973 [Kretsinger & Nockolds 1973].

Inspection of this first crystal structure of CaM identified the presence of large hydrophobic clefts on each of the two lobes which were proposed to be the sites of interaction between CaM and co-factors known to interact with the protein. This was finally confirmed by the solution of the first CaM-complex structure in 1992 by Means and co-workers [Meador *et al.* 1992]. Not only did this structure confirm the presence of extensive hydrophobic patches which acted as the target interaction sites for CaM but also the prevalence of Met residues on this surface, then termed *Methionine Puddles*.

Most recently the structure of *Paramecium Tetraurelia* CaM was determined to 1.0Å by Wilson and co-workers [Wilson & Brunger 2000]. To date there are some 40 or so different structures of CaM which have been determined, including NMR and X-ray Diffraction structures of the protein under different conditions and with various different co-factors. The relative merits of the experimental techniques used to determine these structures and their classification into different families are discussed in more detail in Chapter 2 and a full investigation of the structural ensemble is performed in Chapter 3 of this thesis.

1.2.1 The Anatomy of CaM

The different functional parts of CaM are highlighted in **Figure 1.3** below. The overall dumbbell shape of the protein is clearly visible on the left hand side of the figure with the two Ca^{2+} binding lobes at each end of the protein and the long linker helix in the middle shown in purple. The individual calcium binding lobes are reoriented and viewed on their own on the right hand side of the figure. The first EF-hand, EFI, of the protein is the helix-turn-helix motif formed by helices A and B and the turn in between them and is illustrated in red. The second EF-hand is between helices C and D and is coloured in pink in the right hand figure. These two helices are linked by a β-strand between helices B and C to and together will be referred to as the **N-term lobe** throughout this thesis.

The third EF-hand, EFIII, is formed between helices E and F and the fourth, EFIV, between helices G and H with a β-strand linker between helices F and G connecting the two Ca^{2+}-binding units. In the crystal, the 7-turn helix between the two lobes is a single alpha helix which incorporates both helices D and E and the linker helix from the figure below. For a better understanding of the constituents of the lobe, however, the structure is better understood by considering the central helix as three separate regions, according to the colouring in **Figure 1.3**.

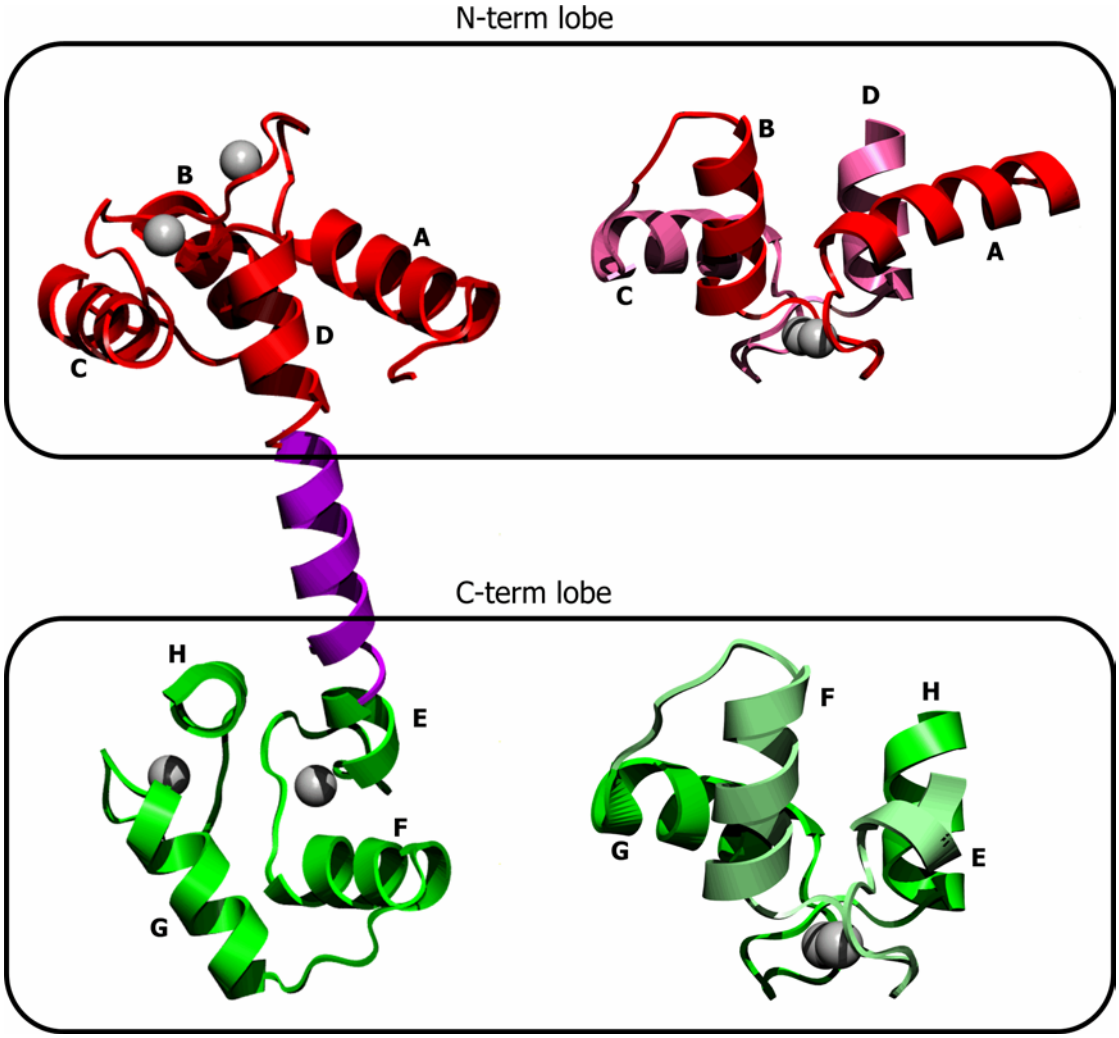


Figure 1.3: *Protein cartoon representation of the 1.0Å structure of CaM,* ***1exr****. Secondary structure elements are highlighted with a ribbon representation and the* Ca^{2+} *ions are shown as Van der Waals spheres. In the right hand elevation the* Ca^{2+} *ions are eclipsed showing the standard side view of the lobe used throughout this Thesis.*

1.2.1.1 EF-hands

The EF-hands are the structural and functional building blocks of CaM. The four Ca^{2+} binding loops coordinate the Ca^{2+} ions essential for CaM structure and function. **Figure 1.4** below illustrates the conformation of the C-terminal EFIII from the 1.0Å resolution structure **1exr** [Wilson & Brunger 2000] showing the residues involved. Here Glu104 interacts with the ion through a bi-dentate acid group whilst Asp93, Asp95, and Asn97 all coordinate through single oxygens. The backbone carbonyl group of Leu99 and a single water molecule (not shown) make up the final coordinating ligands.

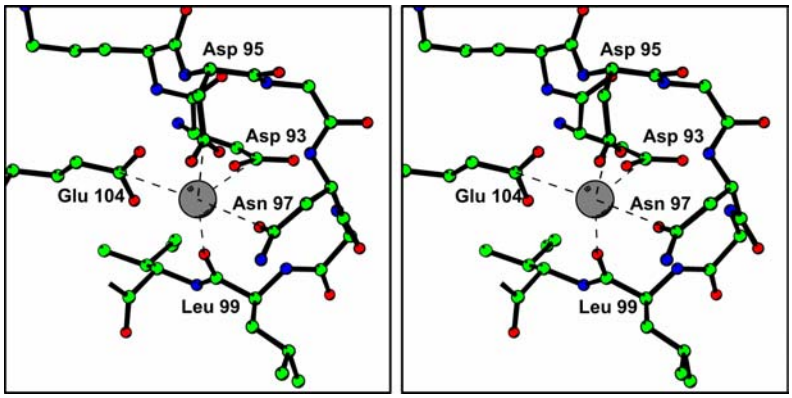


Figure 1.4: *WALLEYE stereo image of EFIII from the structure* ***1exr****. Figure created using Molscript [Esnouf 1999].*

Relative Ca^{2+} affinities of the EF-hands in CaM vary according to experimental conditions with some studies reporting independent and sequential binding to each of the 4 sites, some suggesting cooperativity within each lobe and others reporting a strong degree of cooperativity across all four sites [Van Eldik & Watterson 1998]. Under standard conditions however it seems most likely that Ca^{2+} coordination is cooperative *within* each lobe, with the C-term binding two equivalents of Ca^{2+} before the N-term. This cooperativity considerably heightens the sensitivity of Ca^{2+} binding which thus enhances the ability of the protein to switch over a narrow concentration gradient [Vogel 1994]. A recent study by Martin *et al.* confirmed that not only was Ca^{2+} binding to CaM dependent on factors such as ionic strength and pH, but the formation of a complex with a CaM binding domain significantly enhanced the Ca^{2+} affinity of the protein [Martin *et al.* 2000].

Further to structural comparisons of the four EF-hands in *Paramecium* CaM which indicated that EF-hands I and III were the most similar to one another and I and IV were the most different [Ban *et al.* 1994], Persechini *et al.* studied various CaM mutants in which the EF-hands of the different lobes were interchanged [Persechini *et al.* 1996]. This study investigated the effects on Ca^{2+} binding and target activation of interchanging EF-hands I and III and II and IV. A CaM lobe constructed from EF-hands I and IV was found to provide a structural mimic for Ca^{2+} binding of the wild type lobe constructed of EF-hands I and II. Conversely a lobe containing EF III and II was not able to reproduce the functionality of the wild-type lobe with EF-hands III and IV. Target activation was also lost in several of the mutants. These studies suggest that while equivalent EF-hands

(I & III and II & IV) may have similar Ca^{2+} binding loops, the Ca^{2+} binding of the lobe as a whole is dependent on the entire helix-loop helix motif rather than just the residues directly involved in the coordination of the Ca^{2+} ions.

The Ca^{2+} affinity of the EF-hands of CaM is also carefully tuned relative to other EF-hand proteins. Mutation of Leu39 which is situated at the top of helix B in **Figure 1.3** to a Phe as found in the Calbindin D_{9k} (thought to be involved in Ca^{2+} transport) was shown to dramatically increase the Ca^{2+} affinity of the lobe despite being removed from the Ca^{2+} binding loops [Ababou *et al.* 2001]. An interesting study by Siedlecka *et al.* demonstrated that various constructs using the sequence of the CaM binding loop from EFIII of rat CaM have been demonstrated to provide nucleation sites for helix formation C-terminal to the metal binding loop upon Lanthanide metal coordination [Siedlecka *et al.* 1999]. Other mutations normally sufficient to stabilise the closed conformation of the CaM lobe coupled to this L39F mutant were found to return the Ca^{2+} affinity of the lobe to wild-type levels. Such phenomena accentuate the importance of all the residues of the lobe in the formation and stabilisation of the *open* and *closed* conformations of the protein.

The conformational change observed in the lobes of CaM has been the subject of much investigation and is represented in **Figure 1.5** below. The first structures of apo (Ca^{2+}-free CaM), determined by NMR studies, demonstrated the significant differences in the organisation of the secondary structure units with respect to the crystal structures of Ca^{2+}-CaM [Zhang *et al.* 1995]. The primary feature of the conformational change is the angle between adjacent helices. In the closed, apo, conformation the helices are almost anti-parallel forming helical bundles in both of the Ca^{2+}-binding lobes. This is in marked contrast to the almost perpendicular arrangement of the helices seen in crystal structures of Ca^{2+}-CaM. It was also noted that the linker regions connecting the EF-hand units of each domain were highly flexible in the apo structures, as were the Ca^{2+}-binding loops [Zhang *et al.* 1995].

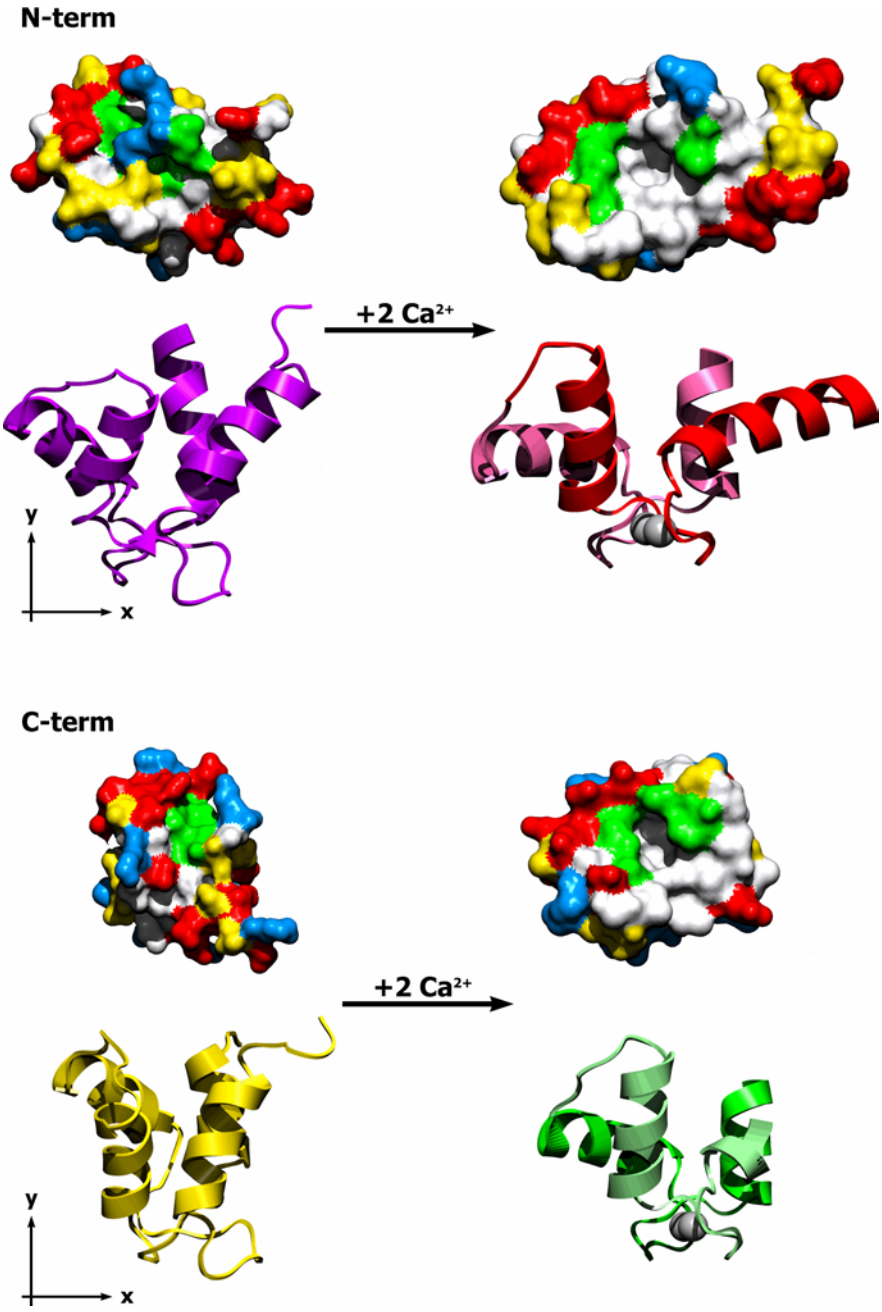


Figure 1.5: *The conformational change in the CaM lobes upon Ca^{2+} binding. The ribbon representations serve to illustrate the inter-helical angles in each of the EF-hands and the solvent accessible surface representations demonstrate the change in the binding surface characteristics. The binding surfaces are shown in a top view with a 90° rotation in the x-axis relative to the ribbon representation side view. Figures generated using the Ca^{2+}-free NMR structure of CaM* ***1cfc*** *and the 1.0Å crystal structure of Ca^{2+}-bound CaM* ***1exr****. The colouring of the molecular surface is according to* ***amino acid polarity*** *with the important Met residues highlighted in green: hydrophobic amino acids are coloured in white (Gly, Ala, Val, Ile, Leu, Phe, Pro, Trp), negatively charged sidechains in red (Asp, Glu), positively charged sidechains in blue (His, Lys, Arg) and polar sidechains in yellow (Cys, Ser, Thr, Asn, Gln, Tyr, His).*

The solvent accessible surface representation in **Figure 1.5** clearly highlights the differences in the binding surfaces of the two CaM lobes between the apo and Ca^{2+}-bound structures. The amino acid polarity colouring clearly highlights the hydrophobic binding surfaces of both lobes of the protein in the Ca^{2+}-bound state and the accessibility of the *Met puddles* highlighted in green and thought to be critical in CaM target interaction [Zhang & Yuan 1998]. Although parts of the hydrophobic patches remain visible in the apo structures, the large hydrophobic pockets are clearly not accessible to targets. The binding surfaces are much smaller in size and several polar and charged amino acids are situated on top of the lobes.

1.2.1.2 Flexible Tether

The tether linking the two lobes of CaM is one of the most interesting components of the protein. In CaM, the central linker helix has the ability to unwind and dis-order enabling the CaM lobes to come together to encompass CaM binding domains. Although the linker is generally accepted to be helical in Ca^{2+}-CaM crystal structures, certain anomalies have been observed such as extended hydrogen bond lengths in mammalian and *Drosophila* CaM not observed in *Paramecium* CaM [Ban *et al.* 1994]. In contrast to the X-ray analysis of CaM structures, NMR studies of the protein conformation suggest that the linker between the lobes is highly mobile in solution between residues 77 and 81 of the protein resulting in the independent tumbling of the two lobes [Barbato *et al.* 1992]. EPR studies of Ca^{2+}-CaM however identified a stabilisation of the linker upon Ca^{2+}-binding in the EF-hands [Qin & Squier 2001]. Structural studies of apo-CaM report considerable flexibility in the linker helix [Kuboniwa *et al.* 1995; Zhang *et al.* 1995].

The role of the central tether however is most apparent in structures of CaM complexes which commonly see the tether unwinding as the normally disparate lobes come together to encompass target peptides or ligands. **Figure 1.6** illustrates the helix of CaM in the Ca^{2+}-bound crystal structure **1cll** and the crystal structure of CaM in complex with the CaM binding domain of CaM dependent Kinase Kinase, **1iq5**. The central turn of the helix incorporating residues 78 to 80 of the protein is highlighted in both complexes.

This single 3 residue section is unwound in **1iq5** allowing a complete re-orientation of the CaM lobes in the complex. Other CaM complexes exhibit a greater degree of unwinding in the tether and an extensively delocalised region [Wall *et al.* 1997] which has been unresolved in some crystal structures.

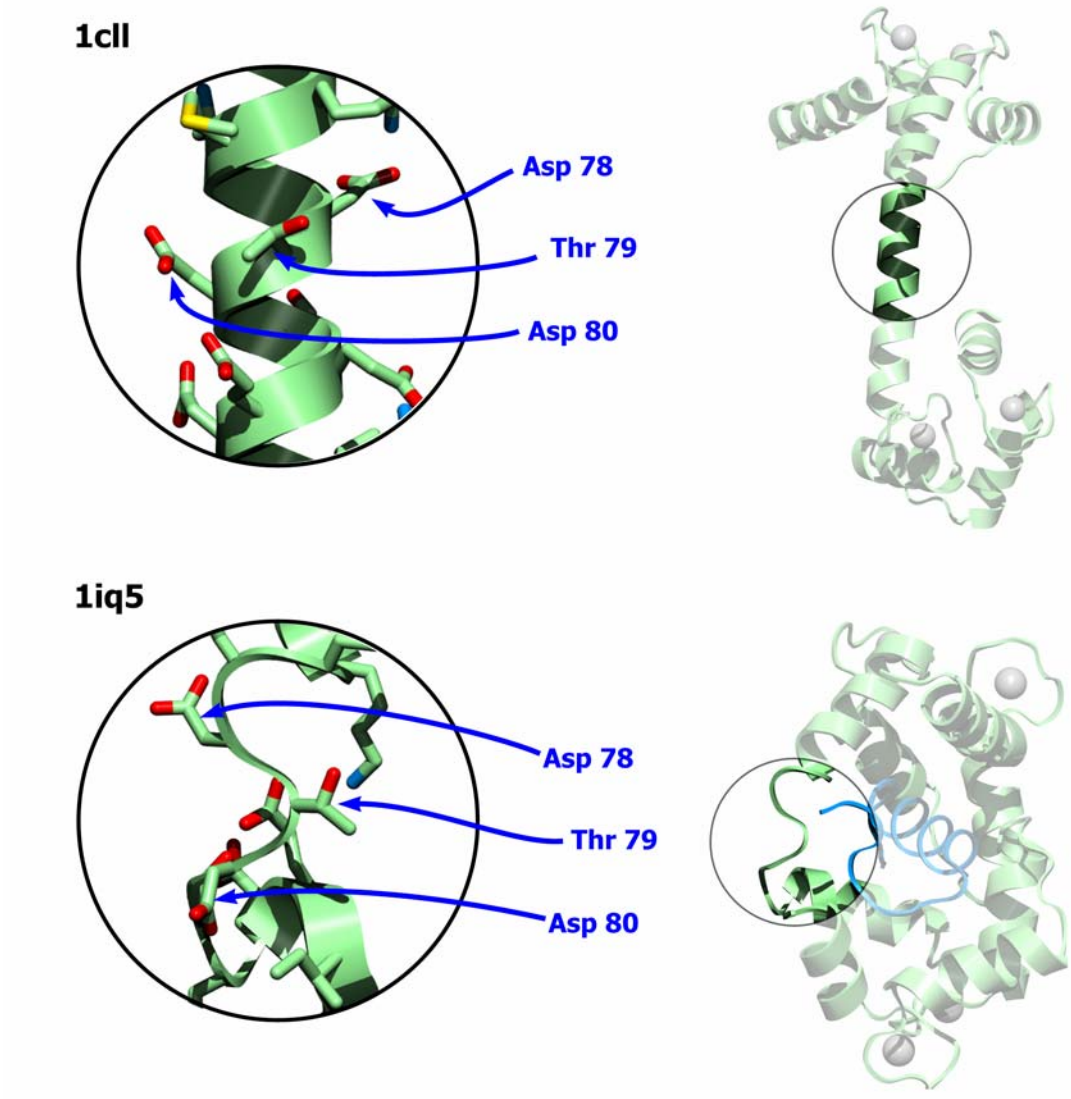


Figure 1.6: *Detailed representation of the central linker helix in the native Ca^{2+}-CaM structure* ***1cll*** *and the CaM-CaMKK complex structure* ***1iq5****. The three residue region central to the unwinding of the flexible tether are identified in both structures.*

Structural studies have also demonstrated that a single tri-cyclic TFP molecule is sufficient to cause a similar global rearrangement of the CaM lobes such that they both interact directly with the ligand [Cook *et al.* 1994].

Various mutation studies have been performed to investigate the importance of individual residues in the linker region of CaM. A deletion mutant which removed Glu84 from the linker suggested a compacted conformation due to bending of the shortened helix in the crystal structure despite presenting a classical dumbbell conformation, indistinguishable from the wild type protein in small angle scattering studies in solution [Kataoka *et al.* 1996]. The Ca^{2+}-free form of the mutant was also suggestive of a more extended conformation than the Ca^{2+}-bound form. It was concluded from these results that the conformation was highly flexible in solution and that the crystal structure had captured one particular possible conformation which is only transiently occupied in solution. The crystal structure of a double deletion mutant has also been determined by X-ray diffraction. The removal of residues 79 and 80 from chicken CaM resulted in a reorientation of the Ca^{2+}-binding lobes relative to one another (**Figure 1.7**) although the lobes themselves remained largely intact [Tabernero *et al.* 1997]. Surprisingly, the shortened linker remains helical in the structure. Functional studies of the mutant however revealed an inability to activate certain targets, thought to be due to the spatial rearrangement of the CaM lobes resulting in decreased accessibility to the hydrophobic binding surfaces.

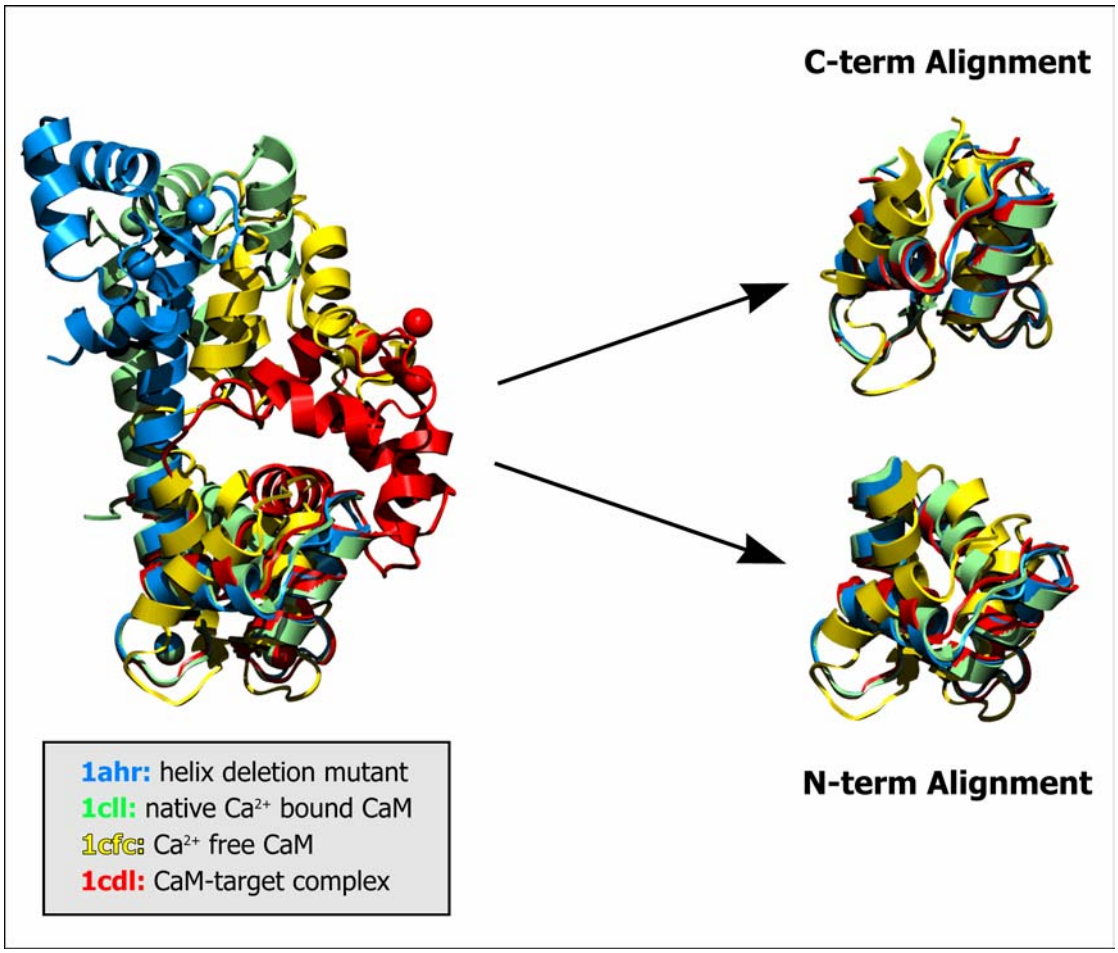


Figure 1.7: *Global superposition of four different CaM structures to highlight the flexibility of the central linker. 1ahr is a deletion mutant structure with two residues missing from the linker which changes the relative orientation of the two lobes but does not alter the linker integrity.* ***1cll*** *is a native Ca^{2+}-CaM structure which shows the normal conformation.* ***1cfc****, the Ca^{2+}-free structure shows slight unwinding of the central helix in solution.* ***1cdl*** *the CaM-target complex shows considerable unwinding to enable the two CaM lobes to encompass the target peptide. Superposition of the individual lobe fragments shown on the right hand side illustrates differences in these fragments*

Computer simulations of the CaM flexible tether alone and as part of intact Ca^{2+}-CaM have also demonstrated the flexibility of the helix and suggested this to be an intrinsic property of the amino acid sequence [Van Der Spoel *et al.* 1996; Wriggers *et al.* 1998].

1.2.1.3 CaM–target Interaction Surfaces

Since the determination of the first crystal structure of Ca^{2+}-CaM, the target interaction surfaces of the protein have been of particular interest [Babu *et al.* 1988]. The large

proportion of hydrophobic residues and the high concentration of methionine residues is a feature shared with the M-domain of the Signal Recognition particle [Walter *et al.* 2000], a protein also designed to interact with a broad range of amino acid motifs presented on an α-helical peptide.

The Methionine Puddles, as they are sometimes known [O'Neil & DeGrado 1990], are thought to be critical to the interaction between CaM and its targets. Mutation studies, however, have also shown that some of the Met residues are more important than others. Although mutation of Met residues to Gln was shown to decrease the affinity of CaM for the Plasma Membrane Ca-ATPase, a mutant in which all 9 Met residues of CaM had been converted to Gln was still able to activate the system [Yin *et al.* 1999]. Individual mutations demonstrated that the alterations to Met124 and 144 had the greatest effect and mutation of Met145 to Gln had no effect at all.

The effect of conformational change upon Ca^{2+}-binding is illustrated in **Figure 1.5** above demonstrating marked alterations to the character of the target binding surfaces. The hydrophobic binding pockets are not available in the Ca^{2+}-free conformation, however several hydrophobic residues are still clustered on the surfaces. Of the Met residues, only 124 and 144 are accessible to targets in the apo conformation [Zhang *et al.* 1995] however the Ca^{2+}-free structure of CaM represented in **Figure 1.5** also exhibits solvent accessible Met residues in the N-term [Kuboniwa *et al.* 1995].

1.2.2 Independent Character of the CaM Lobes

The overall topology of CaM reveals two separate domains connected by a single α-helix which is probably flexible in solution. The separation of the two Ca^{2+}-binding domains suggests some degree of independent behaviour confirmed by the independent tumbling of the lobes in NMR studies [Barbato *et al.* 1992]. Limited trypsinolysis of CaM has been shown to cleave the protein into its N and C-term lobes between residues 76 and 77 [Drabikowski *et al.* 1982]. This has proved to be very useful in the determination of the individual functional characteristics of the two lobes. Structural studies have demonstrated that the separate components largely retain the structure of the intact molecule [Finn *et al.* 1995; Olsson & Sjolin 2001]. Functional studies have been able to exploit the stability of the tryptic fragments to decouple the effects of the CaM lobes from one another and investigate their roles in target interaction [Medvedeva *et al.* 1996].

The functional and structural integrity of the independent tryptic fragments provides a rationale for the separate treatment of the CaM lobes. The illustration in **Figure 1.7** demonstrates how structural comparison is made difficult by the flexibility of the tether between the two lobes and how this can be largely overcome by the treatment of the N and C-term lobes of CaM as independent structures. This approach will be adopted throughout this thesis in an attempt to investigate the similarities and differences between the two Ca^{2+}-binding lobes of CaM.

1.2.3 Dynamics of CaM

Until recently the conformation of the CaM lobes was thought to be a relatively clear cut phenomenon [Akke & Chazin 2001]. Both CaM lobes exhibited an open conformation in the active Ca^{2+}-bound state and a closed conformation in the absence of Ca^{2+}. More recent studies however are starting to suggest that CaM conformation may well be far more adaptable in solution. NMR solution studies of CaM have consistently shown considerable conformational flexibility of the CaM lobes but the application of the newly emerging NMR structural technique of Residual Dipolar Couplings (RDC) has provided

a means of obtaining a single structure solution to NMR experimental data for the CaM lobes [Chou *et al.* 2001].

In these experiments, different initial conformations of CaM derived from a Ca^{2+}-CaM crystal structure, a Ca^{2+}-free NMR structure and a Parvalbumin derived homology model, all resulted in conformations which were more closed than the well conserved crystal structure conformations**Error! Reference source not found.**. This feature was observed in both lobes of the protein although the differences were more pronounced in the N-term lobe. The 1.0Å crystal structure of CaM [Wilson & Brunger 2000] was refined with anisotropic B-factors which indicated what was described as a *scissor-type motion* within the CaM lobes which is similar to an open-closed transition.

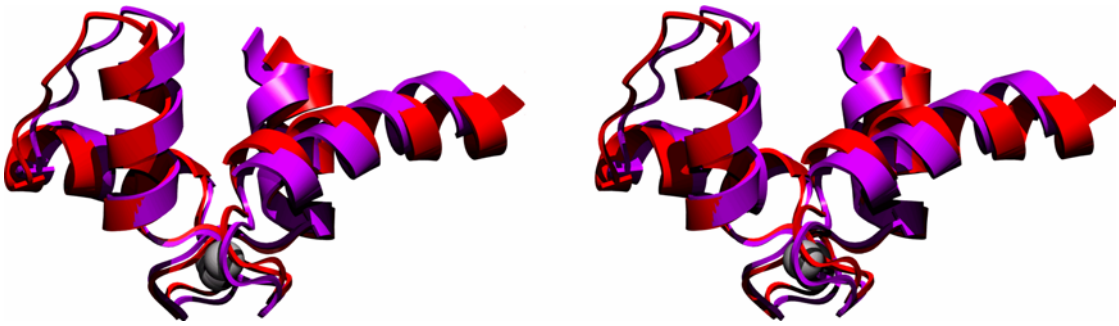


Figure 1.8: *WALLEYE Stereo Ribbon representation of the RDC NMR structure* ***1j7o*** *[Chou et al. 2001] in magenta and the crystal structure of the N-term lobe of CaM used as the model from which the NMR structure was determined shown in red.*

Molecular dynamics studies of holo-CaM [Wriggers *et al.* 1998] have suggested that there is more flexibility in the N-term lobe of the protein in solution. Simulations performed on the N-term lobe of the protein alone resulted in a trajectory which demonstrated transient closing of the N-term lobe towards a more apo-like conformation, a suggestion which was supported by small angle X-ray scattering experiments [Vigil *et al.* 2001]. Various studies of a mutant of the C-term lobe of CaM have also suggested some interesting dynamical properties. The E140Q mutant which disrupts the bi-dentate coordination in EFIV demonstrated exchange between open and closed conformations on the microsecond timescale [Evenas *et al.* 2001]. The authors also suggest such conformational exchange also happens in wild-type CaM albeit on a different timescale.

The experimental results of the studies of Chou *et al.* discount the possibility that the reduced degree of opening exhibited in their RDC structures is an artefact of conformational averaging [Chou *et al.* 2001]. It does however seem likely that the conformation of Calmodulin is far from an open and shut case and that the protein exhibits complex dynamical behaviour as a result of changes in environment and in particular Ca^{2+} and target availability.

1.3 Simulation Studies of Calmodulin

The small size of CaM makes it an ideal candidate for simulation studies of all types. Numerous different types of simulations have been performed on CaM and its homologues to investigate different aspects of the fascinating form and function of the protein, some of which are discussed below.

1.3.1 Molecular Dynamics Studies

One of the earliest computational studies of CaM was performed by Mehler *et al.* in 1991 [Mehler *et al.* 1991]. In this work the dynamics of CaM and TnC were investigated using CHARMM treating the solvent with a distance dependent dielectric with a few discrete water molecules positioned according to waters resolved in the crystal. Molecular dynamics studies of the order of hundreds of picoseconds revealed bending in the central helix which resulted in a more globular shape for the protein which had been suggested by NMR studies but was not observed in the crystal.

Seven years on and a subsequent experiment was performed on holo CaM, fully solvated to 44Å with counter ions to neutralise the induced charge of the highly acidic protein [Wriggers *et al.* 1998]. Here the dynamics were studied over a 3ns time frame although again the focus was on the dramatic bending effects observed in the central tether connecting the two lobes of the protein. The unwinding of the central tether resulted in a reorientation of the lobes such that the binding surfaces were facing one another, a change thought to be indicative of a preparative step towards target association. This

study also revealed some interesting differences between the behaviour of the two Ca^{2+}-binding lobes of CaM and an inhomogeneity of the water structure around the hydrophobic binding surfaces of the lobes.

A more focussed study has also been performed on the central tether of CaM alone in the absence of the two Ca^{2+} binding lobes [Van Der Spoel *et al.* 1996]. In this work the molecular dynamics results were supported with results from Normal Mode Analysis and Essential Dynamics to demonstrate that this bending of the central helix was an intrinsic property of the helix rather than an effect induced by the interaction of the two lobes with CaM targets. During the simulation, the peptide was found to unwind between residues 77 and 80 which enabled the two CaM binding lobes to come together in an orientation similar to that of a CaM-target complex [Meador *et al.* 1992]. This hinge bending motion has also been studied by Gerstein *et al.* and added to a database of macromolecular motions [Gerstein & Krebs 1998].

Another interesting observation obtained from the study of the molecular dynamics of CaM presented evidence for transient closing of the N-term lobe of Ca^{2+}-bound CaM [Vigil *et al.* 2001]. These observations were compared with small angle X-ray scattering data (SAXS) which were consistent with a model in which the N-term lobe of the protein fluctuated between open and closed states thus exposing and burying the hydrophobic binding surface of the protein.

1.3.2 Calcium binding and Molecular Mechanics

Critical to the use of Molecular Mechanics in the investigation of CaM is an appropriate treatment of the electrostatics of the system and particularly the parameterisation of the highly charged Ca^{2+}-binding sites and Ca^{2+} ions. Studies of the molecular dynamics of Calbindin D_{9k} which is another EF-hand Ca^{2+} binding protein, resulted in significant deviations from the classical calcium coordination observed in crystal structures of the protein [Marchand & Roux 1998]. In this case the lack of induced polarization effects were thought to have caused inaccuracies in the potential. It was found to be necessary to alter the parameters for the Ca^{2+} ions to reproduce the free energy of hydration based on the methods used in their simulations. The updated parameters refined here were

subsequently added to the standard CHARMM27 parameter set commonly used in protein studies and used in some of the calculations in this Thesis.

A free energy perturbation study has also been performed on Parvalbumin in order to examine the conformational rearrangement caused by switching between Ca^{2+} and Mg^{2+} in one of its EF-hands [Allouche *et al.* 1999]. Here careful attention was again paid to the treatment of the Ca^{2+} ions in the potential. The authors were again unsatisfied with the behaviour of the interactions and re-parameterisation was not successful with the studies of Parvalbumin [Sanejouand 2001].

1.3.3 Cam-Target Interaction Studies

A study by Afshar *et al.* had an entirely different basis to the largely dynamics based experiments of other investigators. Rigid body translations of a CaM-target peptide within a modelled CaM-target complex were evaluated by energy minimisation to investigate the ability of CaM to accommodate different orientations of the peptide [Afshar *et al.* 1994]. Decomposition of the resulting energy profiles for various translations and rotations of the peptide revealed considerable adaptability of the Van der Waals interactions in the system. Electrostatic interactions between the protein and peptide were not so adaptable and disruptions resulted in significant energy penalties that could not be accommodated by the flexibility in the system. The conclusions drawn from this analysis demonstrate that the conformational rearrangements required by the protein to modify the Van der Waals interactions with the target are relatively inexpensive. The plasticity of the CaM-binding surfaces can easily accommodate such changes through reorganisation of the hydrophobic side chains. The electrostatic interactions however are more specific and the reorganisation of the side chains required is more expensive and cannot readily be accomplished.

The study by Afshar *et al.* also focussed on a characterisation of the interaction between CaM and a target peptide in a model structure, this was largely achieved through the use of *Roadmaps* to illustrate the location of particular amino acid side chains on the protein and peptide surface. Roadmap [Chapman 1993] works by mapping the contact area of the protein side chains onto a cylinder inserted into the CaM binding tunnel. This

cylinder is then opened out to provide a projected topographical description of the location of those residues on the binding surface. This is then repeated for the peptide and the maps are superposed on top of one another to demonstrate the complementarities of the two substituents as shown in **Figure 1.9** below. A key feature of this representation is the complementarity of the four hydrophobic pockets of the protein surface with hydrophobic residues presented by the peptide.

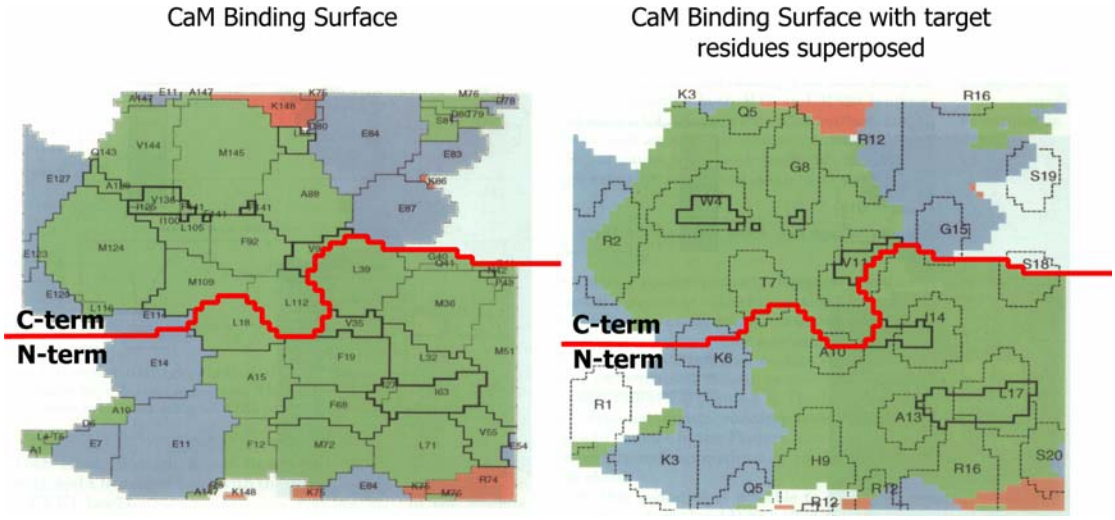


Figure 1.9: *Roadmap representation of the CaM binding surface. The left hand image illustrates the location of protein side chains presented by CaM on the binding surface, the hydrophobic clefts are highlighted by solid dark boxes. Hydrophobic residues are illustrated in green, negatively charged residues in blue and positively charged residues in brick red. The red line shows the separation of the N and C-term lobes. The right hand panel shows the colouring and hydrophobic pockets of the protein with projections of the side chain interactions of the target peptide superposed.*

1.4 Calmodulin Targets

Numerous unsuccessful attempts have been made to date to identify a universal motif for CaM targets which describe the full diversity of amino acid sequences with which CaM interacts. A review published in 1990 suggested that CaM binding domains could be described as Basic Amphiphillic Alpha helices, or BAA helices [O'Neil & DeGrado 1990]. This description is sufficient for a selection of Ca^{2+}-dependent CaM binding domains however a more comprehensive comparison of CaM-binding domains suggests that the motif is somewhat more complex.

Figure 1.10: *Cartoon illustration of the mode of action of CaM with Basic Amphiphillic alpha helices through both Ca^{2+} binding lobes of the protein. Reproduced from [O'Neil & DeGrado 1990]*

A comprehensive review of CaM binding domains by Rhoads and Friedberg compared a wide selection of amino acid sequences which had been identified as CaM-binding regions and three different classes of target [Rhoads & Friedberg 1997]. A Ca^{2+}-independent motif was the most common of the targets identified which was a modified version of the IQ motif found in unconventional Myosins [Cheney & Mooseker 1992]. Two separate Ca^{2+}-dependent motifs were identified; 1-8-14 and 1-5-10 in which the numbers represent the positions of conserved hydrophobic residues. These groups are

summarised in **Table 1.1** below. In these motifs φ represents the position of a hydrophobic residues such as Ile, Val or Leu, **B** a basic residue such as Lys or Arg and ***** any residue. Other residues are represented by their single letter codes.

Table 1.1: *CaM binding motif classes identified by Rhoads and Friedberg [Rhoads & Friedberg 1997].*

	Motif	Ca^{2+} Dependence	Frequency
IQ Motif	*****φQ***B****B**φ*****	Ca^{2+} independent	69
1-8-14 motif	**φ******φ*****φ**	Ca^{2+} dependent	38
1-5-10 motif	*****φ***φ****φ**	Ca^{2+} dependent	7

This collection of CaM binding sequences has more recently been updated and re assessed in the Calmodulin Target Database[1] [Yap *et al.* 2001]. Here, CaM targets are no longer specified as Basic amphiphilic helices, rather the motif is described as being *typically* hydrophobic and basic in nature, 15-30 residues in length and with a *propensity* to form an α-helix. This generalisation of CaM targets is associated with more general descriptions of different classes of CaM targets outlined in **Table 1.2** with the addition of a new undefined class of **Other** targets.

Table 1.2: *CaM binding motif classes identified in the CaM Target Database [Yap et al. 2001].*

	Motif	Ca^{2+} Dependence	Frequency
IQ class	*****φQ***B****B**φ*****	Ca^{2+}-independent	75
1-14 class	**BBBφ******φ****Bφ*****	Ca^{2+}-dependent	43
1-10 class	**BBBφ***φ****φ*******	Ca^{2+}-dependent	24
Other	**Undefined**	various	34

In addition to being a regularly updated database of known CaM targets, the resource provides various tools for the analysis of new targets. Amino acid sequences can be analysed for matches with the characteristic motifs from the above classes of CaM targets

[1] http://calcium.uhnres.utoronto.ca/ctdb

and various sub-classes. A search tool is also provided which can identify likely CaM binding regions in unknown protein sequences. This tool uses average hydrophobicity, average hydrophobic moment and average propensity for α-helix formation calculations to yield a score from 0 to 9. The authors state that this tools is 80% accurate in the identification of CaM binding domains although it can generate false positives [Yap *et al.* 2001].

Complementary to these analyses of known CaM-binding domains was a study performed on a Phage displayed library of peptides. The peptides bound by CaM commonly contained a Lys/Arg-Trp motif [Gao & Zhong 1999] which corresponds to the to **BBφ** motif identified in classes 1-10 and 1-14 in the CaM target database. Despite the diversity of CaM targets, these studies also demonstrated that the preferred targets of CaM were distinct from those of another EF-hand Ca^{2+}-binding protein, Calcineurin B which expressed a preference for negatively charged peptides rich in Phe.

An abundance of information is available on the way in which CaM interact with its targets which is far too extensive to be reviewed here. Several excellent reviews have been written describing some of the more interesting functions of CaM [O'Neil & DeGrado 1990; Crivici & Ikura 1995; Zhang & Yuan 1998] and Van Eldik has recently published a book on the protein and its role in signal transduction [Van Eldik & Watterson 1998]. The importance of CaM is clearly demonstrated through the activity of research in this field and although much is known, there is much more to be understood about the way in which such a small protein can interact with so many different targets with a high affinity.

1.5 Introduction to Thesis

The introduction above has outlined many of the interesting aspects of Calmodulin structure and function which make this protein so individual and fascinating. The intention of this thesis is to investigate some of these properties and better understand how the conformational flexibility inherent to CaM is able to enable it to interact with so

many different targets in living systems. The different components of the investigation are outlined in the layout of the Thesis described below.

Chapter 2: Tools and Techniques

This Chapter outlines the different methods used to investigate the various properties of CaM explored in this thesis. In addition to an outline of molecular modelling methods, the various analysis techniques used to extract information from the wealth of data generated from these investigations are described.

Chapter 3: Calmodulin Structural Variability

In this Chapter the available structural information on Calmodulin and some of its homologues is investigated. Principal Component Analysis is used to characterise the differences between the protein structures in different conformations in order to understand the observed distribution of conformations exhibited by the protein.

Chapter 4: Molecular Dynamics Force Field Comparisons

Here the effects of different force fields on the Molecular Dynamics of the Calmodulin lobes are investigated. Different parameterisations and solvent models are compared for multiple 1ns dynamics trajectories of the N and C-term lobes of the protein.

Chapter 5: Ca^{2+}-Dependent Molecular Dynamics of CaM

In this section, 10ns molecular dynamics trajectories are analysed to investigate the conformational flexibility of the CaM lobes. Simulations are performed for solvated N and C-term fragments and for a CaM-target complex. A method is also developed to investigate the effect of Ca^{2+} removal from the system resulting in a characterisation of the dependence of CaM on Ca^{2+} and target peptides for conformational stability.

Chapter 6: Entropic Analysis of the CaM lobes

Multiple static entropy calculations are performed on conformations generated from the Molecular Dynamics simulations of the CaM lobes. The entropy values are calculated from a Normal Mode Analysis to present information on the relative inherent flexibilities of the CaM lobes with a view to their different functional roles.

Chapter 7: CaM-Target interaction Analysis

In this Chapter a detailed inspection of the available CaM-target structures is performed in order to understand the effects of CaM conformational flexibility on the proteins ability to interact with different targets. In addition to a review of the global and local conformational flexibility of the system, the role of the protein side chains is investigated through an analysis of χ angle variations across the structural ensemble. CaM binding surface plasticity is also investigated through the use of molecular graphics techniques. The flexibility of the CaM binding surfaces described by NMA of the lobes is also discussed.

Chapter 8: Conclusions

In the final chapter the observations and conclusions of the thesis are collected and discussed. General conclusions are drawn from the results and opportunities for further studies are discussed.

Chapter 2: Tools & Techniques used in this Thesis

This chapter outlines the various tools used to analyse the conformational flexibility of Calmodulin investigated in this thesis, and the analysis techniques used to interpret the results. The first part of the chapter outlines the form and function of the molecular mechanics force field used to simulate the properties of CaM structures. The techniques employed using this force field are then introduced. This is followed by an exposition of the Principal Component Analysis technique used extensively throughout this thesis to identify and analyse trends in multivariate data, along with other analysis techniques used throughout. Finally the software and hardware used in these experiments are introduced.

Specific tools and techniques used only in a particular context are not discussed in this chapter; they will be described within the context of their usage to an appropriate degree of detail.

2.1 Molecular Mechanics Force Fields

Biomolecules are highly complex systems. Many attempts have been made to develop computational models which address this complexity and model the behaviour of these systems in a realistic manner. A recent article by Michael Levitt [Levitt 2001] described the origins of the first molecular force field, the Consistent Force Field (CFF), by Lifson *et al.* [Lifson 1981] in the late 1960's. This program was able to calculate the energy, forces and curvatures (second derivatives with respect to atomic displacements) for various molecules, culminating in the first energy minimisation of a protein in 1969 [Levitt & Lifson 1969]. From these early beginnings a plethora of molecular force fields now exist which are commonly used to describe the behaviour of biomolecules, recently reviewed by Kollman *et al.* [Wang *et al.* 2001], all of which use the same functional form developed for CFF and outlined below.

The development and parameterisation of the AMBER force field [Cornell *et al.* 1995] has been optimised primarily for the study of Nucleic Acid systems but can also be applied to proteins and organic molecules. The CHARMM force field was one of the first protein MM force fields and used an extended atom approach which only treated heavy atoms and polar hydrogens explicitly [Neria *et al.* 1996] developed by Brooks *et al.* in the early 1980's [Brooks *et al.* 1983]. More recently CHARMM has been re-parameterised as an all-atom potential [MacKerell *et al.* 1998]. The CVFF force field [Ewig *et al.* 1999] developed by BIOSYM (now Accelrys[2]) has a more complicated potential energy functional form than the form generally adopted by other force fields (see **Equation 2.1** below). GROMOS96 [Stocker & van Gunsteren 2000] is another well established force field developed by van Gunsteren *et al.* for the simulation of the molecular dynamics in the solution or crystalline state. The OPLS-AA force field [Kaminski & Jorgensen 1996] of Jorgensen *et al.* was developed for Monte Carlo and Molecular dynamics simulations of organic and molecular dynamics systems in solution. The MMFF force field was developed for the study of protein ligand complexes [Halgren 1996] through parameterisation for small molecules using Quantum Mechanical calculations.

2.2 The CHARMM Force field

The CHARMM Force field (Chemistry at HARvard Macromolecular Mechanics) is a program developed in 1983 [Brooks *et al.* 1983] to facilitate the manipulation of biomolecular macromolecules according to the following potential energy function:

[2] http://www.accelrys.com

2.2.1 Functional Form of the Potential

$$U(R) = \sum_{bonds} K_b (b - b_0)^2 + \sum_{angles} K_\theta (\theta - \theta_0)^2 + \sum_{UB} K_{UB} (S - S_0)^2$$

$$+ \sum_{torsions} K_\phi \left[1 + \cos(n\phi - \delta)\right] + \sum_{impropers} K_{imp} (\varphi - \varphi_0)^2 \qquad \textbf{Equation 2.1}$$

$$+ \sum_{VdW} \varepsilon_{ij} \left[\left(\frac{R_{\min_{ij}}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{\min_{ij}}}{r_{ij}} \right)^{6} \right] + \sum_{Elec} \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon_r r_{ij}}$$

2.2.1.1 Bonded Terms

The interaction between two covalently bonded atoms is described analytically by the well characterised Morse Function (**Equation 2.2**) for the electronic ground state of a diatomic molecule. This function characterises a minimum energy D_e at an equilibrium distance b_0 and which tends towards zero at infinite separation. Solution of the Schrödinger equation for the atomic nuclei resulted in a single adaptable parameter, α, which could be used to adapt the potential for a good approximation to the vibrational spectrum of all diatomics.

$$V = D_e \left[e^{-\alpha(b - b_0)} - 1 \right]^2 - D_e \qquad \textbf{Equation 2.2}$$

Due to the limited variability of chemical bonds, however, Lifson identified that biomolecular systems were unlikely to exhibit variations above 0.1Å. With such rigid systems the Morse potential could be accurately approximated through the use of a simpler quadratic harmonic function (**Figure 2.1** below) [Lifson 1981], although this was not sufficient for vibrational spectra.

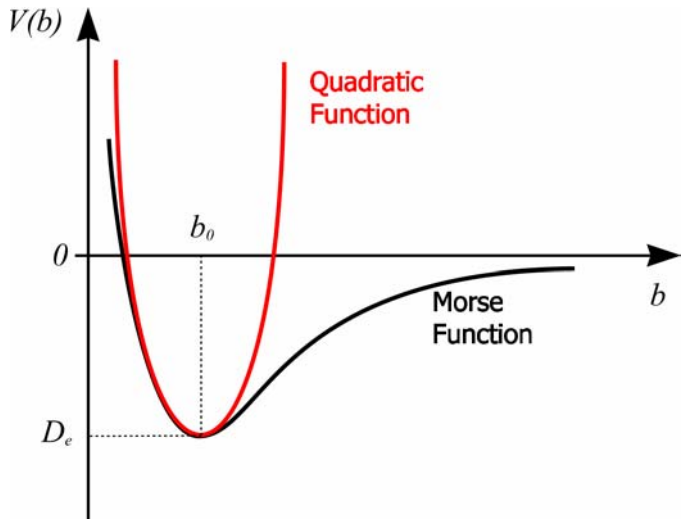


Figure 2.1: *Comparison of Morse and Quadratic representations of the potential energy variation with bond distance for a covalent interaction.*

In the absence of a suitable Quantum Mechanical relation, Lifson went on to describe deviations about the equilibrium bond angles also with a quadratic function, which when appropriately characterised turned out to be a suitable representation [Lifson 1981]. The penalty for deviation about the equilibrium out of plane angles (φ in **Figure 2.2** below) were also found to be accurately approximated using a quadratic function.

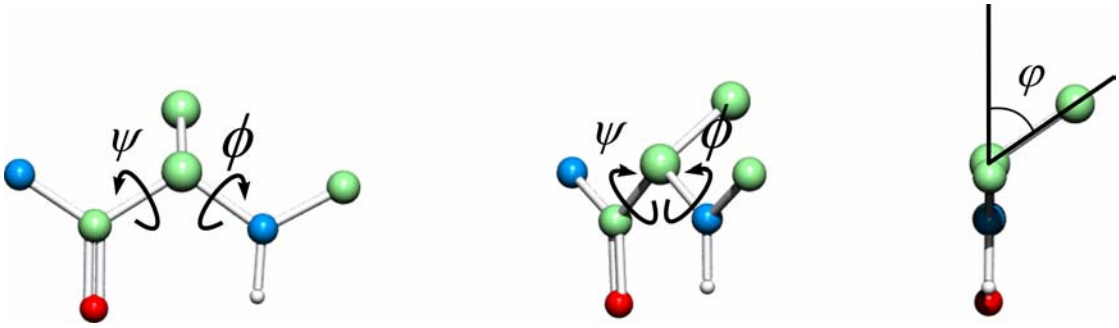


Figure 2.2: *Illustration of the assignment of the proper and improper torsions of the protein backbone in the CHARMM force field. The figure illustrates different view of an Ala residue highlighting the rotatable torsions ψ and ϕ and the out of plane, improper torsion angle φ.*

Torsional angles in the system, (ψ and ϕ in **Figure 2.2** below) were originally represented in the system with the empirical version of the Pitzer potential described in **Equation 2.1**. This trigonometric function represents the periodic nature of rotations around single C-C bonds. The multiplier, n, accounts for the periodicity e.g. in ethane which has

minima at 120° increments when the groups are staggered and n=3. Rotations about the peptide bond were not included in the original potential as the partial double bond character makes them highly resistant to rotation. More recently these rotations have been included and are treated with a quadratic function acting on the torsion peptide bond torsion angle ω.

The Urey-Bradley term has been extensively used in spectroscopic force fields but was introduced more recently into CHARMM, added to improve the accuracy of vibrational properties obtained with the potential. This quadratic term represents the distance dependent 1-3 interactions between an atom *i,* and the next-but-one atom, $i \pm 2$ although it is not used in this thesis.

2.2.1.2 Non-bonded Terms

The extensive intermolecular forces between atoms which are not directly involved in 1-2 bond length, angle, or Urey-Bradley terms are accounted for as non-bonded interactions which serve also to supplement the torsion angle terms. These are compiled for all pairs of atoms in the system which are not directly involved in the bonded terms and are divided into charged interactions which are most commonly treated using Coulombs Law and non-polar or Van der Waals interactions.

Attractive Van der Waals forces between non-polar atoms arise from induced dipole-dipole interactions, characterised by London et al. [London 1930] and vary according to $1/r^6$. At very small distances the repulsive forces between nuclear and electronic effects dominate and vary according to $1/r^N$ where N is an integer between 9 and 12. These effects are commonly modelled according to $1/r^{12}$ resulting in the Lennard-Jones (12-6) potential shown in **Equation 2.1** where R_{min} is the minimum energy separation between the atoms and ε_{ij} is the corresponding minimum energy.

Correct treatment of the electrostatic interactions between atoms in biochemical systems is crucial due to the long range effects of the interactions (proportional to 1/r) and correct parameterisation provides an incentive for extensive investigation (see section 2.2.1.4). Differences in the electro-negativity of atoms results in an uneven distribution of electron density and the formation of partial charges. Interactions between point charges

can be accounted for through the application of Coulombs Law (in **Equation 2.1**) to describe the interaction between point charges q_i and q_j at a separation r_{ij}.

2.2.1.3 Non-bonded interaction truncation

The pair-wise calculation of non-bonded interactions for biomolecular systems constitutes considerable computational requirements as they scale according to N^2. Consequently these interactions are often truncated at specified cutoff distances, although this results in a discontinuity in the functions at the cutoff distance. This discontinuity can be addressed by means of a cubic switching function (**Equation 2.3**) which smoothly reduces the potential to zero over the range r_{on} to r_{off}. Another treatment of the discontinuity involves shifting the potential to zero at the specified cutoff distance as described in **Equation 2.4** and **Figure 2.3**. These functions were introduced in the initial publication of the CHARMM program [Brooks *et al.* 1983].

$$Sw(r) = \left[\frac{(r_{off} - r_{ij})^2 \quad (r_{off} + 2r_{ij} - 3r_{on})}{(r_{off} - r_{on})^3} \right] \qquad \textit{Equation 2.3}$$

$$Sh(r) = \left[1 \; - \; \left(\frac{r_{ij}}{r_{off}} \right)^2 \right]^2 \qquad \textit{Equation 2.4}$$

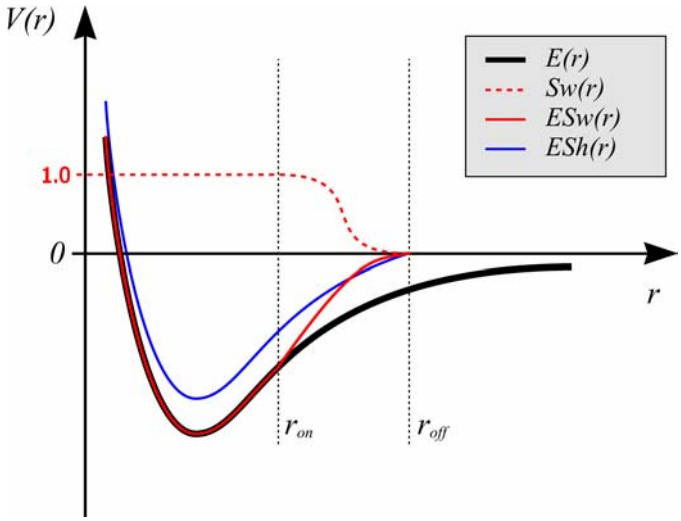


Figure 2.3: *Graphical representation of the switching and function and shifted potential often applied to the potential energy function when using non-bonded cutoffs. The non-bonded energy function is represented by E(r) and the switched and shifted potentials denoted by ESw(r) and ESh(r) respectively.*

Additional methods have also been derived for a more accurate treatment of the long range effects of electrostatics in highly charged systems such as CaM. In the *Extended Electrostatics approach*, the electric potential is separated into near and extended contributions, with the near contributions treated in a conventional atom based manner manner and the effects of the extended contributions treated with a different group based model [Stote 1991]. The most comprehensive of electrostatic treatments however is generally accepted to require the use of the *Ewald Summation* which uses a periodic function summing the electrostatic contributions of images of the unit cell under consideration. The *Particle Mesh Ewald Summation* is a hybrid mode and has been shown to be successful in the analysis of charged systems including CaM [Yang *et al.* 2001]. These methods are explained in detail in standard texts e.g. [Leach 1996].

2.2.1.4 Solvent treatments

As biomolecules do not function in a vacuum, it is common to incorporate some kind of treatment of the protein environment into simulations using these force fields. Different solvent models are discussed at length in standard texts such as Leach [Leach 1996] and Brooks *et al.* [Brooks *et al.* 1988]. The models most commonly used in these studies are outlined below.

The simplest treatment of solvent uses a dielectric screening constant to describe the effect of solvent molecules interfering with explicit inter-atomic charge interactions. This is normally treated with a distance dependent function such that $\varepsilon_r = r_{ij}$ (in Å). Where explicit solvent molecules are used the effective dielectric, ε_r, is set to unity.

The use of explicit solvent brings great computational expense and introduces additional problems such as where and how should the solvent be placed and how are the boundaries of the solvated particle to be maintained over the course of the simulations. One established technique used to overcome these issues is the use of *Periodic Boundary Conditions*. This involves the construction of a unit cell (usually a cube) which is surrounded by an appropriate number of images of itself (26 in the case of a cube) and the forces exerted by the atoms in the surrounding image boxes are incorporated into the calculations. More commonly, the particle of interest is solvated in a droplet of water to sphere of a particular radius or to a minimum shell depth [Grubmueller 1996]. Boundary conditions may be imposed on the water droplet to retain the shape of the solvated particle. A more economical solution to the problem of solvation can be provided by the process of *Active Site Solvation* where only the region of interest in a system is solvated explicitly. Atoms not contained within this core are often harmonically constrained or fixed.

A more detailed description of different solvent treatments is included in Chapter 4 where the effects of different solvent models on simulations of CaM fragments are investigated.

2.2.2 Parameterisation of the Potential

The empirical nature of molecular mechanics force fields makes the parameterisation critical for the correct treatment of biomolecular systems. The optimisation of parameters for CFF was performed by least squares fitting to obtain the best match between theoretical and experimental data for the vibrations and enthalpies of small molecules and biopolymers [Lifson 1981]. Much the same approach is used today for the derivation of Van der Waals parameters from liquid or solid state small molecular

systems [Wang *et al.* 2001]. Electrostatic parameter optimisation however is intrinsically linked to the solvent treatment and is increasingly performed by the derivation of partial charges to optimally reproduce the quantum mechanical electrostatic potential. Critical to this analysis is the interplay between the so-called holy trinity of solute-solute, solute-solvent and solvent-solvent interactions. CHARMM uses a modified version of the established TIP3 water model of Jorgensen *et al.* [Jorgensen *et al.* 1983] which has three interaction sites per molecule, situated at the three atom centres of a rigid water molecule.

In the original CFF, it was found that no specific hydrogen bond term was necessary for the potential and that the interaction were adequately treated by the polar interactions of charged groups [Lifson 1981]. The early implementations of CHARMM used an entirely extended atom approach in which hydrogen atoms were incorporated into the adjoining heavier nuclei, resulting in the need for additional hydrogen bonding terms. This model was later succeeded by the polar hydrogen potential, PARAM19 [Neria *et al.* 1996] which included explicit hydrogen atoms on N and O. This has now also been largely superseded by the PARAM27 all atom potential, the full parameterisation of which is outlined in detail in the publication by MacKerell *et al.* [MacKerell *et al.* 1998].

2.3 Molecular Mechanics Applications

The CHARMM potential energy function described above can be used to calculate the energy of a system in a given conformation on the basis of the Cartesian coordinates of the system. This could be used to generate a 3N (where N is the number of atoms in the system) dimensional potential energy surface for the system describing the potential energy for any conformation. Full knowledge of this potential energy surface can be used to recover the thermodynamic and kinetic properties of a system. The techniques used in this thesis include *Energy minimisation*, *Molecular Dynamics* and *Normal Mode Analysis* as described below.

2.3.1 Energy Minimisation

One of the first applications of CFF was the minimisation of myoglobin and lysozyme [Levitt & Lifson 1969], a process then known as *Energy Refinement*. The multi-dimensional character of these biomolecular energy landscapes is such that the single lowest energy conformation of any system will rarely be discovered. Rather, energy minimisation techniques are used to identify physiologically relevant and realistically accessible *local* energy minima using various different algorithms. Various such techniques exist and are discussed in some detail in standard texts [e.g. Leach 1996]

In the simplest minimisation technique of *Steepest Descents* (SD), a line search is performed along the vector with the steepest gradient and the energy is calculated at several points and the minimum point calculated by interpolation. In the *Conjugate Gradients* (CG) method, a linear combination of the vector corresponding to the steepest descent gradient of the current iteration, *i*, and the previous iteration, *i-1*, results in more accurate minimisation steps and better convergence as the minimum is approached. The more computationally intensive technique of *Adopted Basis Newton Raphson* (ABNR) minimisation uses the second derivatives from an expanded approximation to the potential energy function to find a minimum directly.

Energy minimisation strategies are normally developed which incorporate a combination of these techniques to exploit their characteristics for different phases of the minimisation. The minimisation protocol commonly used in this thesis is outlined below and a schematic of the process is illustrated in **Figure 2.4**. Initially a harmonic constraint is applied to all the heavy atoms in the system which is initially set to 10 kcal mol^{-1} $Å^{-2}$. The system is then subjected to 100 steps of SD minimisation to remove the worst atom clashes followed by 400 steps of ABNR minimisation. The harmonic restraints are then reduced by 0.5 kcal mol^{-1} $Å^{-2}$ and the SD and ABNR minimisations repeated. Once all the harmonic restraints have been removed the system is subjected a much more extended ABNR minimisation routine with 80 cycles of 400 steps of ABNR minimisation with no constraints on the system. The final stage of the process is to request 3000 steps of minimisation to a gradient tolerance of 10^{-5} kcal mol^{-1} $Å^{-1}$. The final tolerance gradient is normally achieved within the 3000 steps in which case the minimisation request is terminated.

Where this minimisation protocol is performed prior to NMA calculations, the structure is minimised more intensively to a gradient of 10^{-8} kcal mol^{-1} $Å^{-1}$ in order to ensure that the matrix of second derivatives is a positive definite and yields no negative eigen values (Section 2.3.3).

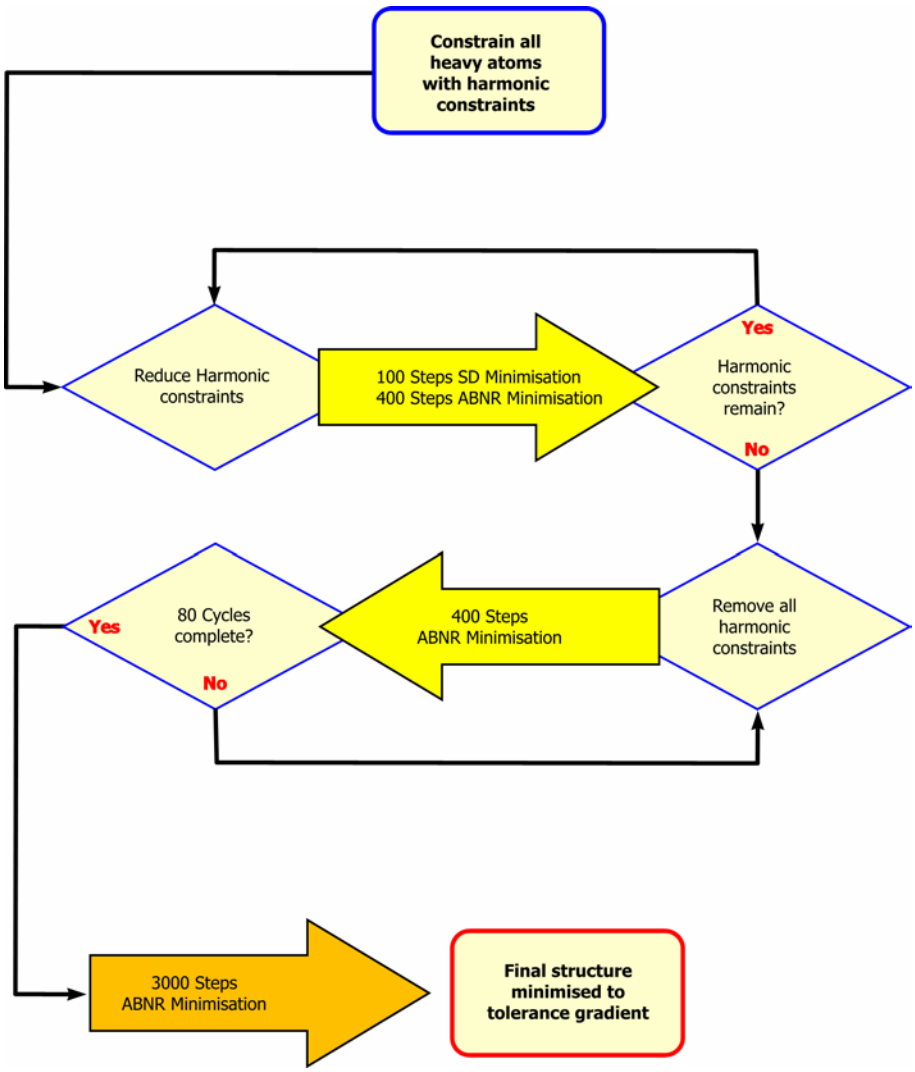


Figure 2.4: *Schematic flow diagram of the minimisation protocol used for all minimisations in these experiments.*

2.3.2 Molecular Dynamics Simulations (MD)

Molecular dynamics is a technique for the investigation of time dependent behaviour through the integration of Newton's equations of motion. MD trajectories are

determined by solving the differential equations resulting from the second law ($F=ma$) outlined in **Equation 2.5** below. Here the acceleration is described as the second derivative of the coordinates with respect to time which is equal to the Force along a particular coordinate F_{xi} acting on a mass m_i.

$$\frac{d^2x_i}{dt^2} = \frac{F_{x_i}}{m_i} \qquad \textit{Equation 2.5}$$

This differential equation can be solved by one of several different integration approximation methods. Finite Difference methods separate the integration into small time steps and use the Taylor series to approximate the physical properties of the system. Predictor-corrector methods of integration offer an alternative technique using a three stage process. Positions, velocities and accelerations are approximated with a Taylor expansion, forces calculated at the new positions, and the accelerations are then corrected using the differences between predicted and calculated values. Heat can be transferred to the system simply by scaling the velocities at each time step, however a smoother alternative method was developed by Berendsen *et al.* using an external temperature bath [Berendsen *et al.* 1984]. Using this method the velocities are scaled such that the rate of change of temperature is proportional to the difference in temperature between the temperature bath and the system. Details of these techniques are discussed at great length in standard texts [e.g. Leach 1996].

The MD simulations in this thesis were carried out at constant temperature and integrated using a finite-difference method with an integration step of 1fs. After completion of the minimisation protocol on the system, the protein fragments were gradually heated to the temperature of interest by performing 10ps MD simulations at 25K temperature increments from 0 to 300K. This was done using a tight temperature coupling of 0.01ps used to ensure that the energy was fully transferred to the system [Berendsen *et al.* 1984]. Judgements regarding the time required for the system to equilibrate under different conditions were incorporated into the analysis of the trajectories.

2.3.3 Normal Mode Analysis (NMA)

Normal mode analysis is a technique which can be used to compute the vibrational states of a molecule based on the harmonic approximation. As outlined in section 2.2.1 above, the potential energy minima of molecular systems can locally be characterised by a simple harmonic approximation **Figure 2.1**. This approximation is only accurate for small deformations about the minima but has been shown to be useful in the identification of functionally relevant motions [Van Der Spoel *et al.* 1996].

Any molecule has a set of 3N-6 internal vibrational modes (plus 3 rotational modes and 3 translational). This information can be calculated from the mass weighted matrix of second derivatives of the potential energy function to produce a the *Hessian* matrix of force constants for the system. Diagonalisation of this matrix of force constants results in a set of 3N orthogonal eigen vectors describing the shape of the normal modes with corresponding eigen values which detail the energy (or frequency) of the vibrations. The first 6 modes generated from the system have eigen values which are approximately zero as they correspond to overall translational and rotational degrees of freedom in the system. The lowest frequency modes commonly describe the most functionally relevant motions in the system. The frequencies of the modes can also be used to calculate the thermodynamic properties of a system such as ΔG and ΔS through the harmonic oscillator partition function which is described in detail elsewhere [McQuarrie 1973; Leach 1996].

- **Reduced Basis Normal Mode Analysis**

This technique allows for the normal modes of vibration to be calculated for a subsection of a protein leaving the rest of the system fixed. The technique can be used to constrain particular degrees of freedom thus leaving them out of the basis for the calculation. In this way the deformations described by the resultant normal modes include information for the unconstrained parts of the system whilst incorporating the effects of the fixed regions into the calculations. This technique is not widely used although it has had some exposure [Fischer *et al.* 2001]. The technique is used in Chapter 6 to study the thermodynamics of the separate CaM lobes in a CaM-target complex.

2.4 Catalogues of Protein Structure

Since the first CaM structure was determined crystallographically in 1988, numerous structures have been determined by both NMR and X-ray crystallography under a variety of different conditions. Structures have also been determined for some of the homologous proteins from the wider family of EF-hand proteins. This provides a plethora of structural information which may help to explain how CaM is able to adapt its structure under different conditions and act as an environmental sensor. This structural information is held in the Brookhaven Protein Database [Sussman *et al.* 1998], currently hosted by the Research Collaboratory for Structural Biology (www.rcsb.org) [Berman *et al.* 2000]. This database is freely available over the Internet and a number of websites have tools for searching the information contained within it.

A keyword search for *Calmodulin* using the RCSB PDB browser (www.rcsb.org) yields a total of 49 hits, which, in addition to real CaM structures with several false positive hits: Examples include **1phk**; a Phosphorylase Kinase structure which contains a CaM binding domain and Troponin C through identification of the structure as a known homologue. Removal of these false positives generated a list of 40 different CaM structures in various different target association states and Ca^{2+} states.

In order to supplement the list of CaM structures with any appropriate structural homologues, various different structural classification databases were interrogated. The characteristics of the different databases are outlined below.

2.4.1 Structural Classification Databases

In the light of the increasing size of the structural database, several attempts have been made to classify protein structures using different approaches to help understand structure function relationships and taxonomy. Each system has a different methodology for developing rules about what constitutes a relationship and consequently each has advantages and disadvantages dependent on the problem being addressed. A comparative study of the three most widely used tools SCOP [Murzin *et al.* 1995], FSSP [Holm & Sander 1996] and CATH [Orengo *et al.* 1997] undertaken by Hadley and Jones in 1999 [Hadley & Jones 1999] concluded that none of the databases was superior and that each had advantages for specific tasks. These three databases were investigated to supplement the list of CaM-like structures obtained from the keyword search of the PDB with some CaM homologues.

2.4.1.1 SCOP Database *(Structural Classification of Proteins)*

Developed in 1995 [Murzin *et al.* 1995], the SCOP database was probably the first database to attempt to organise the rapidly increasing number of structures publicly available in the PDB according to a structural classification. This is a manual and hierachic organisation of protein domains into four main classes; all-α, all-β, α/β and α+β as defined by Levitt and Chothia [Levitt & Chothia 1976] (an additional six exceptional classes have subsequently been defined) before classification into different folds, superfamilies and families. Classification at the fold level identifies favourable topologies while members of the same superfamily share common structural and functional features and members of the same family exhibit similar sequence or highly homologous structures. **Figure 2.5** below provides some indication of the scale of this classification task for the first release of SCOP based on some 3179 protein domains relative to the latest release (release 1.55 1/3/2001) which classified some 31,474 protein domains.

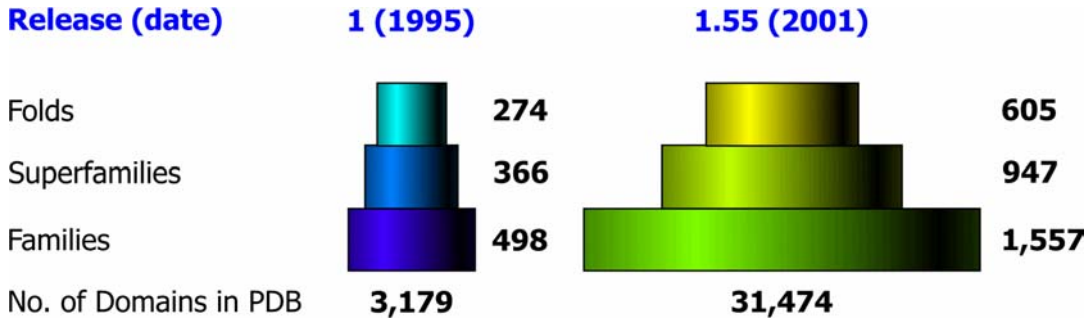


Figure 2.5: *SCOP classification pyramid of protein structures according to the original 1995 release (left) and the latest March 2001 release 1.55.*

2.4.1.2 CATH Database (Class Architecture Topology Homology)

CATH was developed by Thornton and co-workers in 1997 [Orengo *et al.* 1997]. Rather than trying to classify all the available structural information, the CATH database is complied using only one representative structure from each sequence family. High resolution (3.0 Å or better) crystal and NMR structures are extracted from the PDB and classified into sequence families with 35% or more identity from which the highest resolution structure is selected. This is then subjected to various protocols which identify and separate the various domains in the structure. Domains are then allocated, automatically where possible, into one of four Classes (mainly α, mainly β, α β and no secondary structure) before being organised into fold groups (Topologies) and then homologous superfamilies (Homologies). Architecture is defined manually by visual inspection. The CATH Database is commonly represented using the CATH wheel shown below.

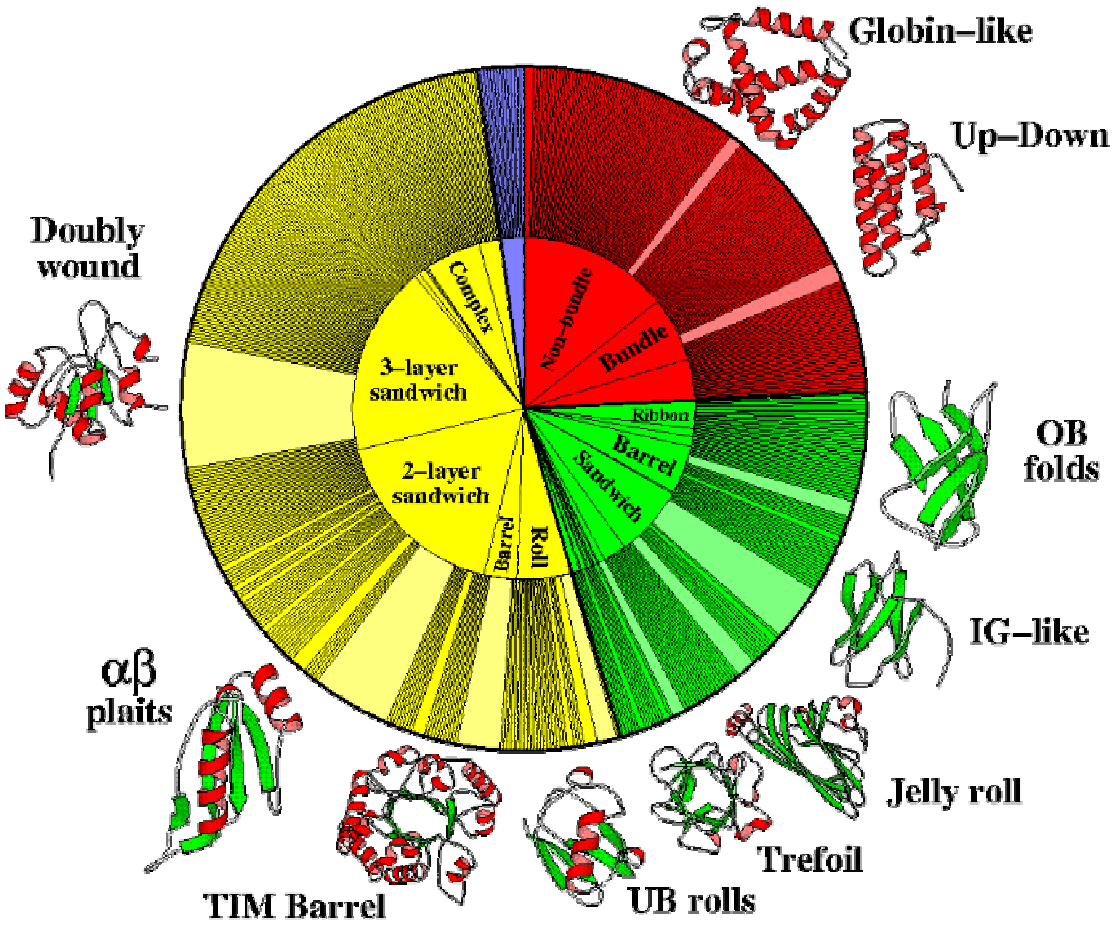


Figure 2.6: *CATH wheel illustrating the proportion of the different Classes, Architectures and Topologies found in the CATH database. The sizes of the α β and α/β regions (re, green and yellow) indicating the H-level families belonging to them.*

2.4.1.3 FSSP *(Fold classification based on Structure-Structures alignment of Proteins)*

FSSP is a fully automated tool developed by Holm and Sander of the EMBL-EBI Structural Genomics Group at Hinxton [Holm & Sander 1996]. This tool aims to associate protein structures automatically without an assignment into different classes or families. Structures with 25% or more sequence identity are removed and the remainder separated into domains to give a set of representative structures. A pairwise comparison of these structures then yields a Z-score entered into a results file for the target structure. The accompanying distance matrix alignment tool, Dali [Holm & Sander 1998], performs the same kind of pairwise comparison within a set of homologous structures and thus is able to identify all the CaM-like structures through their sequence homology.

2.4.1.4 CAMINE

Another fully automated tool used to identify structural homologues of a target molecule has been developed by Tom Oldfield [Oldfield 2001]. CAMINE reduces the structural representation of a target protein to a fragment has-table of virtual angles and torsions between the α-carbon atoms. This can then be used as a characteristic profile of the target structure which can be matched against similar profiles for a multitude of other structures in a highly rapid fashion. By generating a sequence window of a pre-defined length and moving it along a protein structure to generate this profile, a whole structure is searched for matches against the target fragment. Once structural matches have been found, sequence similarities can also be highlighted to provide further evidence for a match.

2.4.2 Protein Structure Solution by NMR and X-ray Crystallography

Nearly all protein structures are determined by either Single Crystal X-ray Diffraction or NMR Spectrometry. Whilst both of these techniques provide invaluable information about the structure of proteins there are considerable differences in the nature of the information that they provide. When attempting to directly compare the results of these two techniques it is therefore important to understand some of these differences.

The single protein crystal used for X-ray diffraction contains a tightly packed lattice of protein molecules only 50% aqueous by volume [Creighton 1993]. A consequence of the crystallisation is that numerous contacts will be made between protein units and other entities such as solvents or additives used to help promote crystallisation, most of which would not occur under natural conditions. Many recent crystal structures have been determined at low temperatures of -150^{o}C [Brändén & Tooze 1999], which helps improve resolution and preserve the lifetime of the crystal in the X-ray beam. Such low temperatures, however, also represent non-physiological conditions for the protein environment.

Solution structure determination is now possible by NMR for proteins up to 25kDa in size [Brändén & Tooze 1999]. The protein concentration used for such studies however has to be very high, of the order of 1-2mM and the pH generally has to be below 6.0 to

minimise mainchain NH exchange which also results in non-physiological conditions. Members of NMR ensembles are generated by solutions to the distance constraints, often generated by refinement of randomly generated conformations which can differ by 1-2 Å [Lee & Kollman 2001]. Such discrepancies are often the consequence of considerably fewer observables per atom than are obtained for by X-ray Crystallography. A recent review by Lee & Kollman [Lee & Kollman 2001] suggested that considerable inaccuracies are often obtained from inappropriate modelling of solvent during model refinement.

At 17 kDa, CaM is well within the size limit for NMR structure determination and the flexibility of the protein can cause problems in crystal structure determination [Cook *et al.* 1994] making it an excellent subject for NMR Structure determination. Both techniques have been exploited to investigate different properties of CaM with crystallography providing high-resolution data on crystalline CaM [Wilson & Brunger 2000] and NMR providing valuable information on CaM solution dynamics [Malmendal *et al.* 1999]. X-ray structures however clearly provide more precise data with most good structures providing atomic positions to within 0.1-0.2 Å [Brändén & Tooze 1999] in contrast to NMR structures consisting of ensembles of conformations covering several Å around an average structure.

2.5 Data Analysis Techniques

Much of the work in this thesis is based on the comparison of protein structures derived from various origins including experimentally determined protein structures and conformations resulting from simulation studies. The techniques outlined below detail many of the different techniques and tools used to perform these comparisons throughout these experiments. Some of these techniques are available within the CHARMM program although most of the data analysis was performed externally using the R Statistical package [Gentleman & Ihaka 2000]. Plots were generated with the postscript plotting functions of R and then edited with the graphics package Xara X [Various 2001].

2.5.1 RMSD Comparison

The Root Mean Square Difference (RMSD) is the most commonly quoted statistic for comparisons of protein structures and other information. The square root of the mean of the squared differences between the coordinates of equivalent atoms in the two or more structures reports a value in Å according to the relation outlined below. The coordinates for N atoms in the structures are represented by x_i and y_i over which the comparisons are made.

$$RMSD = \sqrt{\frac{1}{N}\sum (x_i - y_i)^2} \qquad \textbf{Equation 2.6}$$

2.5.2 Structural Superposition

Structural superpositioning of ensembles of structures is commonly performed to minimise the RMSD between them and put them in a common reference frame. This process is most commonly performed on the basis of the RMSD for the more invariant parts of the structure such as the Cα atoms or the protein backbone and in certain cases the superposition is calculated on a rigid section of the structure rather than an entire protein chain. A rotation matrix is calculated for the best fit of the invariant regions and then applied to the entire structure to obtain the superposition. The choice of this reference can be critical to the interpretation of the resultant fit [Karplus & Ichiye 1996]. Throughout this thesis the established technique of least squares fitting [Kabsch 1976] is used to optimise the fit of structures through the minimisation of the RMSD between them.

2.5.3 Principal Components Analysis

Principal components analysis (PCA) is used extensively to identify underlying trends in complex multidimensional data. It is a widely used multivariate statistical technique discussed extensively in standard texts [Chatfield & Collins 1980] which has been applied

to a multitude of different applications from the characterisation of Beef [Destefanis *et al.* 2000] to the interpretation of NMR spectra [Howe 2001]. Application to the interpretation of molecular dynamics is becoming increasingly common [Amadei *et al.* 1993] and more recently its application to the interpretation of NMR structural ensembles has been realised [Howe 2001].

Protein structure comparison is a multidimensional problem with 3 degrees of freedom per atom to be considered. Whilst value judgements can be made about differences between structures purely by visual inspection with the use of molecular graphics, PCA provides a means to not only quantify the differences between protein structures, but also to characterise these differences with orthogonal descriptors. The process used to perform this analysis and examples of the types of information recovered are outlined below.

Central to the analyses in this thesis are the differences in the conformation of CaM across a structural ensemble using the Cα atoms alone. The Cartesian coordinates for the α carbon atoms of the appropriate protein fragments (after having been superposed onto a common reference frame) were extracted and compiled into a single data matrix, **X**, with p rows and m columns were p=3N (and N is the number of residues or Cα-atoms) and m is the number of structures. A p dimensional square covariance matrix of the data, **S(X)**, is then constructed describing the variation for each of the independent variables across the m structures relative to a mean structure, $\bar{x}$, according to the relation described in **Equation 2.7** below. Diagonalisation (eigen analysis) of this matrix yields a set of 3N orthogonal vectors, each in 3N dimensions, with an associated eigen value, which together describe all the variation in the original data matrix. These are the *Principal Components* (PCs) of the distribution.

$$S(X) = \frac{1}{m-1} X_d' X_d = \frac{1}{m-1} \sum_{i=1}^{m} (x_i - \bar{x})'(x_i - \bar{x}) \qquad \textbf{Equation 2.7}$$

An eigen spectrum indicates the magnitude of the variance described by each of the PCs. Throughout this thesis, these are scaled to indicate the proportion (%) of the total variance (the trace of **S(X)**)described by each PC. This is sometimes known as a *Scree Diagram* and is illustrated in the fourth panel of **Figure 2.7** below. It is not uncommon in

these studies for the first few PCs to describe a large proportion of the data as demonstrated in **Figure 2.7**, where the first three PCs describe over 80% of the variance and only 19 PCs are required to describe 99% of the total variance.

The vector descriptions of the PCs themselves can be described by linear projections of the average structure along the eigenvectors as illustrated in **Figure 2.8**. Although this representation does not necessarily have a physical significance, it is analogous to the projection of a structure along a vibrational normal mode and is useful in the characterisation of the structural variation described by a particular PC.

A means for the identification of the contributions of different structure to the PCs is provided by the *Conformational Subspace Projections* shown in **Figure 2.7**. After subtraction of the average structure from the initial distribution, the residuals are multiplied by the vector component of the PCs as described in **Equation 2.8**. If the differences between a structure and the average correspond to the structural variation described by a particular PC, then the resultant displacement will be significant from the ensemble. In such cases it can be said that a particular PC accurately describes the way in which a structure differs from the average of the structural ensemble. In PC1 of **Figure 2.7**, the data point (structure) labelled **1dmo** can be seen to be considerably displaced from the origin, thus the displacements described by PC1 provide a good description of the way in which 1dmo differs from the average of the structural ensemble. In these plots the axes are scaled by the square root of the number of atoms to yield values which are comparable to the RMSD values commonly quoted for differences between protein structures.

$$Q_i = a_i \bullet (x_i - \bar{x}) \qquad \textbf{Equation 2.8}$$

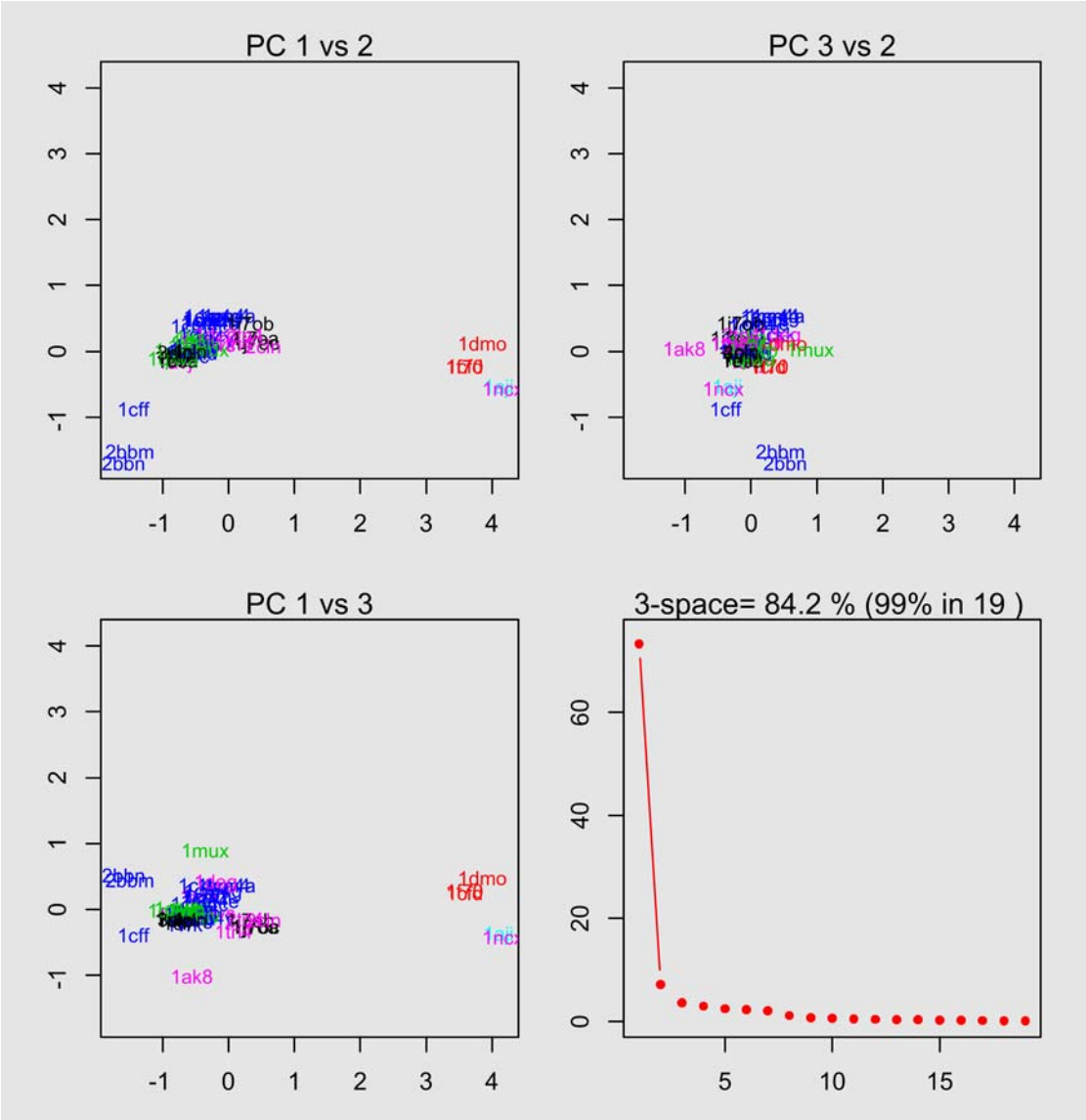


Figure 2.7: *Conformational subspace projections for the first 3 PCs of a distribution of experimentally determined structures for the N-term lobe of CaM. The fourth panel presents the eigen spectrum for the distribution.*

Figure 2.7 above illustrates the results of a PCA for a set of experimentally determined structures for the N-term lobe of CaM (discussed fully in Chapter 3). In this analysis the xyz coordinates of the 69 Cα atoms from the structures were compiled in a data matrix. The covariance matrix was then calculated and diagonalised. The Scree diagram illustrated in the fourth panel of **Figure 2.7** shows that over 70% of the variance in the observed distribution can be described by a single vector representing *concerted* variations in Cartesian coordinates for all the atoms. Projections of the original structures onto the subspace described by the three most significant PCs allows the distribution of the data to be visualised with proximities in the projected subspace reflecting the (full

dimensional) RMSD to an extent proportional to the variance spanned by the respective PC's. In the case of the projection onto the subspace of PC 1 & 2 it is clear that many of the structures are clustered at the origin and do not contribute significantly to either of these PCs. The cluster of several structures at 4Å on PC 1 however highlights the fact that these structures differ from the main cluster in a manner which is well described by PC 1, illustrated in **Figure 2.8**. In the same way, the difference between **2bbm** and **2bbn** and the other structures is characterised by PC 2.

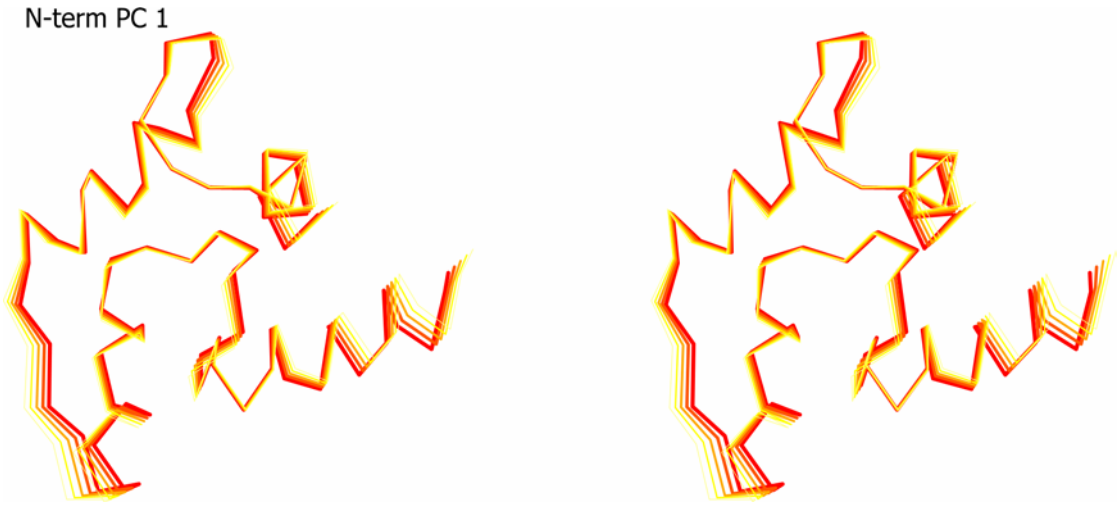


Figure 2.8: *Wall-eye stereo plot of the projection of the average structure from the distribution along PC1 for the N-term.*

The character of the PCs can be visualised by the projection of the average structure in Cartesian space as shown in **Figure 2.8**. These projections provide an insight into the nature of the structural variability observed across the structural ensemble which is comparable with the normal modes of vibration discussed in section 2.3.3 above.

2.5.4 Essential Dynamics

The application of Principal components Analysis to molecular dynamics trajectories was popularised by Amadei *et al.* [Amadei *et al.* 1993]. In this study it was noted that after translation and rotational degrees of freedom had been removed, the diagonalisation of the atomic positional fluctuations could be used to describe the collective motions of the system in only a few degrees of freedom. It was found that all other degrees of freedom

in the system represented Gaussian fluctuations orthogonal to this *Essential Subspace* and the technique was consequently termed called Essential Dynamics (**ED**). A study by Hayward *et al.* [Hayward *et al.* 1995] compared the subspaces described by the (harmonic) computed Normal Modes and the anharmonic principal components describing the essential subspace of an MD simulation of BPTI. This study concluded that for a large subspace, the free energy surface described by the two sets of orthogonal degrees of freedom was largely similar after a scaling factor had been applied. These findings were confirmed in a subsequent study using only the Cα atoms to define the essential subspace for a 200ps MD simulation of SH3 which were then compared with a NMA of the same system [Van Aalten *et al.* 1997]. The functional relevance of such studies was also highlighted in a recent review by Kitao and Go [Kitao & Go 1999].

2.5.5 Cluster Analysis

Hierarchical cluster analysis is another technique which is commonly used to analyse complicated multivariate datasets. Data are compared on the basis of some kind of similarity index such as RMSD in the case of protein structures studied here. Cluster analysis attempts to organise the data such that at the highest level, all elements of the data set are members of a single cluster whilst at the other extreme, each element represents a single unique cluster. Between the two extremes elements are organised together into subgroups which show a degree of similarity but are separate from other elements or subgroups which are different. The hierarchical organisation of the data is best represented in a *Dendogram* as illustrated in **Figure 2.9** below.

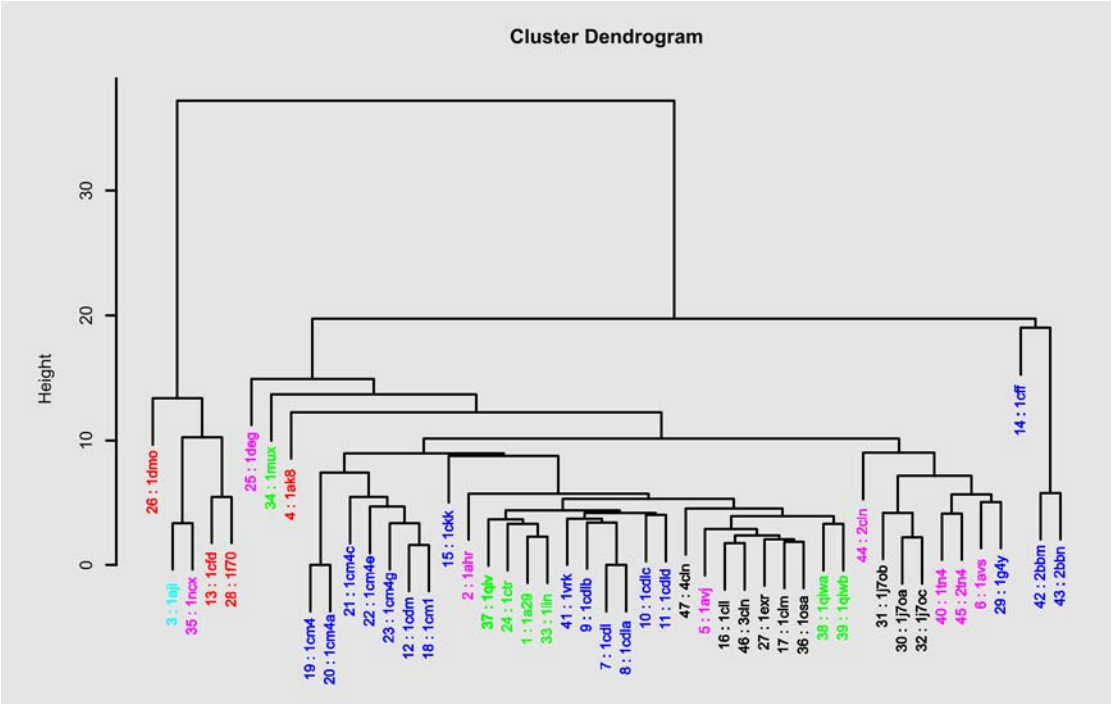


Figure 2.9: *Hierarchical cluster dendogram of the CaM N-term structures. Clustering performed by the average linkage method on pairwise Euclidean distances between superposed coordinates.*

Cluster analysis performed in this thesis is according to the *hierarchical agglomerative* method using a technique called *average linkage*, although many other methods are also available. This is a bottom up method in which all elements of the dataset are initially considered as individuals and the clusters are built towards a single cluster with increasing dissimilarity. In the example of the data above, elements 19 & 20 and 7 & 8 are the first to be clustered together at Euclidean distance close to zero. The average distance between the elements of the new clusters is then compared with all other elements in the dataset as new clusters are formed. In the example above, the separate cluster on the left hand side including structures 26, 3, 35, 13 and 28 which is nearly 40 units away from the main cluster. This separation can also be seen in **Figure 2.7** as the members of this cluster are around 4Å away from the main group of structures in the conformational subspace projections onto PC 1.

In S*ingle Linkage* methods it is the shortest distance between any member of a cluster and any other element which is compared to generate new cluster. Conversely in *Complete Linkage* methods it is the furthest member of the clusters which are compared to ensure complete incorporation within the nearest cluster. There are no fixed rules as to which

method results in the best representation of the data structure. Single linkage clustering however is commonly thought to result in elongated clusters [Leach 1996]. Average linkage methods are often chosen on the basis of the central position between complete and single linkage methods and are thought to best represent the Euclidean structure of the data.

2.5.6 Amino Acid Sequence Comparison

The primary sequence alignments performed in this thesis were all carried out by visual inspection or using the amino acid sequence alignment package *ClustalX* [Thompson *et al.* 1997]. CaM sequence alignments were relatively simple to perform due to the high degree of conservation across different species, providing an indication of the importance of each of the 148 amino acids in the protein [Yang 2001]. The short length and high diversity seen across CaM target peptides however present enormous challenges for sequence alignment techniques.

Various different scoring matrices have been developed to compare non-identical equivalent pairs of amino acids on the basis of the minimum number of genetic substitutions required for the match [Fitch 1966] or chemical similarity [McLachlan 1972]. More recently developed scoring matrices are based on the concept of accepted point mutations; the frequencies of observed mutations across a set of protein sequences of known homology are used to generate a scoring matrix indicating the likelihood of a particular (position independent) mutations being accepted. The details of the alignment output are outlined below **Figure 2.10**.

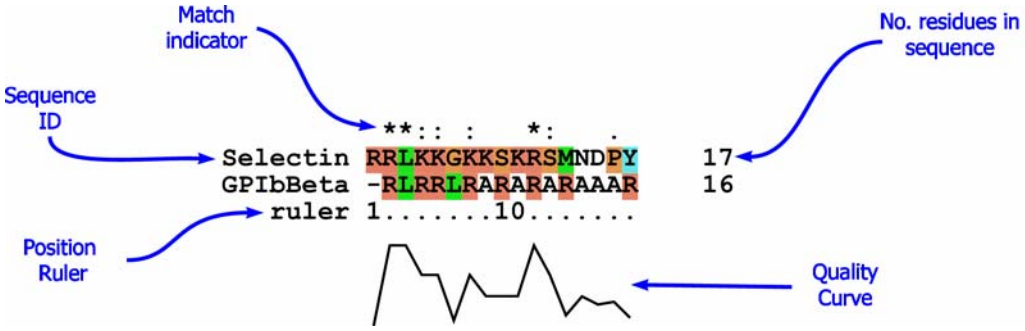


Figure 2.10: *Exemplary annotated sequence aligment from* ClustalX. *Features of this plot detailed below.*

- **Match Indicator:**

These indicators are used to highlight matches across the amino acid sequences compared. The Asterisk (*****) denotes a perfect match across all sequences, the colon (**:**) denotes a match of very similar residues and the full stop (**.**) denotes a match of slightly similar residues.

- **Residue Ruler:**

The position ruler can be used for reference and numbered positions correspond to the last character of the number, i.e. position 10 is marked by the 0 and not the 1.

- **Quality Curve:**

The Quality curve also provides a simple reference for identifying regions of good and poor alignment on the basis of a *conservation score.*

Chapter 3: Calmodulin Structural Variability

3.1 Introduction

Understanding the link between protein structures and their function is central to the understanding of living systems. To this end, huge efforts are being invested in the elucidation of protein structures to enlarge the protein structure database. Whilst it was once the case that a single structure was used to understand a protein's role, more recently the trend has been to determine multiple structures of a system under different conditions.

Changes in Ca^{2+} concentration result in a structural rearrangement of Calmodulin which changes the proteins affinity for a particular target. Interaction with the target also results in structural alterations to the protein. It is known that CaM has a high affinity for a multitude of highly diverse targets which have to be accommodated by adaptations in the protein. Whilst the role of the highly flexible CaM linker in facilitating the global rearrangement and reorganisation of the lobes relative to one another and CaM targets is well characterised [Van Der Spoel *et al.* 1996], the local adaptations of conformation within the lobes are particularly of interest [Chou *et al.* 2001].

Calcium (Ca^{2+}) binding in the EF hands results in significant alterations to the inter-helical angles as shown in **Figure 3.1** and consequently the CaM target binding surfaces. This 7-fold coordination of the Ca^{2+} ion in turn stabilises each EF-hand unit, both constraining the inter-helical angles and forming the β strands which hydrogen bond to each other holding the two EF-hands together. This results in a Ca^{2+}-bound form of the protein which is more *structured* than the Ca^{2+} free form [Kuboniwa *et al.* 1995].

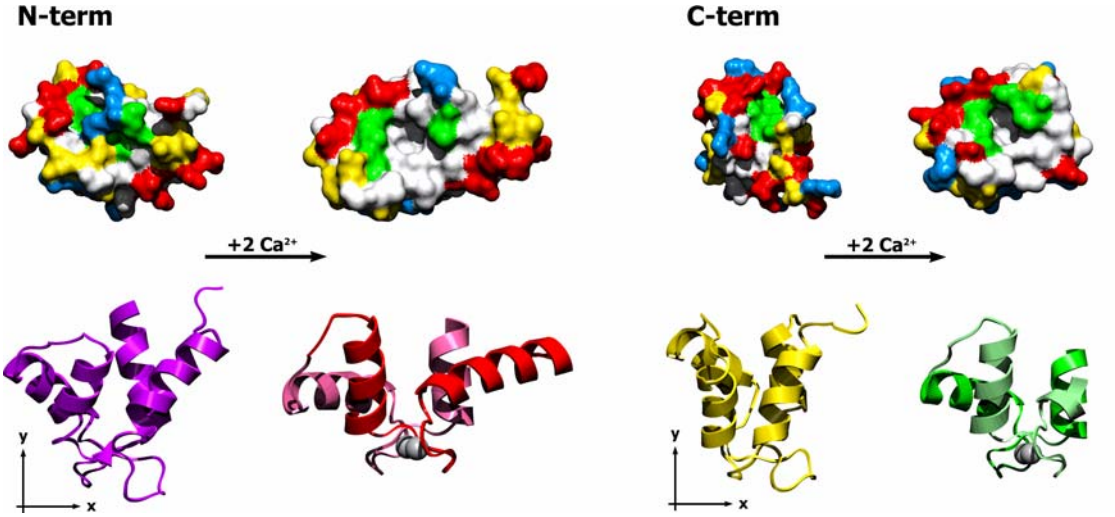


Figure 3.1:*Figure illustrating the conformational change taking place on Ca^{2+} binding The upper figures illustrate a plan view of the CaM binding surfaces and the lower figures show ribbon representations of apo CaM in* ***1cfc*** *and the Ca^{2+}-bound conformations of* ***1exr****.*

In this investigation we will explore the wealth of structural information available for CaM and some of its structural homologues. A structural superposition of the various CaM-like fragments will highlight differences in Cα structure which are then characterised via a Principal Component Analysis. This technique is used to decompose the multivariate nature of the observed conformational flexibility and re-describe the variability in terms of its most significant statistical characteristics. A Normal Mode Analysis is also employed to explore the nature of vibrations about the local energy minima occupied by the CaM lobes and how they relate to the observed conformational variability. These studies help to characterise the conformational flexibility available to the CaM lobes which confer functionality as a signal transducer.

3.1.1 Reference Frame Selection

The correct selection of an appropriate reference frame for protein superposition can be crucial to the interpretation of results under certain circumstances [Karplus & Ichiye 1996]. In this analysis, an attempt was made to describe a reference frame appropriate to the dataset and the conformational flexibility being investigated using the AvgCore software of Gerstein and Altman [Gerstein & Altman 1995]. The core structure

determined by this analysis was compared to the all Cα atom reference frame more commonly used. A detailed description of the technique is provided in section 3.2.4.

3.2 Materials & Methods

3.2.1 Structure Selection

CaM structures were identified from the PDB [Berman *et al.* 2000] by means of a keyword search[3] for *Calmodulin*. False positives, including mainly CaM-binding proteins, were not included in the analysis and the list of CaM structures was supplemented with appropriate homologues identified using FSSP as outlined in the results section 3.3.1. The Calpain structures **1alv** and **1alw** were included initially as these are also members of the EF-hand calcium-binding proteins. (A complete list of structures used in the initial analysis is also included below in **Table 3.6**).

3.2.2 PDB Pre-processing

The Protein Data Bank files held in the PDB are primarily designed to contain the atomic coordinates of the protein structures that have been determined. As experimental techniques are developed, more information is becoming available as part of the structure determination procedure and consequently demanded by the user. Much of this information has to be incorporated into the PDB file resulting in a highly complex file [Community 1996]. Consequently, it is often the case that a considerable amount of pre-processing is necessary to remove much of the extraneous information from the files.

After the Cα coordinates have been extracted from the PDB file, there are various problems that need to be addressed as a consequence of the different experimental techniques used for each structure elucidation, the main points of which are covered below.

[3]A keyword search tool is available using the RCSB interface to the PDB, available at www.rcsb.org

3.2.2.1 Disordered Atoms

Averaging of the many copies of the asymmetric unit in the crystal means that normally, only one conformation is identified in a crystal with high B-values obtained for the more flexible regions of the protein. For the highest resolution structures however, it is quite common to see alternate conformations for some amino acid sidechains or even whole sections of protein. In this case multiple sets of coordinates are entered into the PDB file for the different locations with a corresponding occupancy value. The identifiers are a letter (A, B, C, …) adjacent to the atoms number and conventionally the location with the highest occupancy is found first and entered as location A. In those structures where such disorder had been identified, location A was selected in preference to any of the other locations.

3.2.2.2 Handling Multiple Conformers

In addition to multi-model NMR files, the CaM dataset includes several structures where multiple models have been refined by crystallography. These fall into two categories outlined below:

- **1cm4 – A multiple Conformer Refinement:**

This is a rare type of experiment in which the special techniques of Bragg and Diffusive X-ray scattering were applied to provide extra information about the motions of CaM in the crystal [Wall *et al.* 1997]. In addition to the multiple conformer structure however, a single representative structure was released corresponding to the same experimental data which was used for this analysis. Since these separate structures represent distinct experimental observations they were treated as such in the analysis and split into the separate fragments **1cm1**, **1cm4a**, **1cm4c**, **1cm4e** & **1cm4g**.

- **Multiple Copies in the Asymmetric unit:**

The asymmetric unit is the smallest building block of a protein crystal. Symmetry operations yield multiple copies of the asymmetric unit throughout the unit cell and the packing of the unit cell makes up the crystal lattice. The asymmetric unit often only contains one copy of the protein in which case only one structure is released in the PDB. In some crystals however, multiple copies of the protein are found in the asymmetric

unit published in the PDB file. This was the case for **1cdl** [Meador *et al.* 1992] and **1qiw** [Harmat *et al.* 2000] and these were again included as separate structures.

3.2.2.3 NMR Multi-model Files

NMR files were separated into the individual models published from the experiment and the representative model suggested by the NMRCore/NMRClust package (see section 3.2.3) was used for the comparative analysis. Where published, a minimised average structure was also included in the analysis for comparison.

3.2.3 NMR Representative Structure Selection

NMR Structure solutions are published as a structural ensemble which represents model solutions to the distance constraints generated by the experiment (discussed in Chapter 2). Some experimentalists publish a minimised average structure in addition to the ensemble to be used as a representative structure for the ensemble but this has not been experimentally observed. Where a single representative structure is required, as in this case, it is therefore necessary to find a structure which can be reliably used to summarise the results of the experiment. Lee and Kollman [Lee & Kollman 2001] simply used the first model presented in the ensemble although they provide no rationale for this decision. Sutcliffe *et al.* [Kelley *et al.* 1996; Kelley *et al.* 1997] produced a method for selecting a representative structure from the ensemble which was thought to be preferable.

- **NMR Core & NMR Clust:**

The first program, *NMRCore* uses a sorted list of the variances in individual dihedral angles across the ensemble to define the core. This is followed by the automatic clustering of the variances in pairwise inter-atom distances across the ensemble by the program *NMRClust*. Example results are shown below.

Table 3.1: *Results of NMRCore analysis for the NMR ensemble 1cmg. The largest local structural domain (LSD) contains 34% of the protein residues. 10 domains are required to cover the entire structure*

Domain No.	Residues	% core	% core (cum.)
1	{ 85-89 91-93 99-100 134-142 }	33.9%	33.9%
2	{ 101-107 }	12.5%	46.4%
3	{ 122-127 }	10.7%	57.1%
4	{ 94-98 }	8.9%	66.1%
5	{ 118-121 }	7.1%	73.2%
6	{ 143-146 }	7.2%	80.4%
7	{ 82-84 }	5.3%	85.7%
8	{ 108-110 }	5.4%	91.1%
9	{ 128-130 }	5.4%	96.4%
10	{ 116-117 }	3.6%	100.0%

Table 3.2: *Results from the program NMRClust indicating the clustering of structures in the NMR ensemble and showing the representative model from each cluster.*

Cluster No.	Representative	Cluster Members
1	Model 13	{1 2 6 10 13 17 }
2	Model 7	{7 15 16 19 }
3	Model 12	{8 9 12 }
4	Model 5	{5 14 18 }

These results were obtained using the web interface for the program available at http://neon.chem.le.ac.uk/olderado/ where the user enters a PDB accession code or can upload a PDB format multi-model file. This technique was used to obtain representative models for all the NMR structures used in this thesis.

3.2.4 Core Structure Determination

The *AvgCore* software of Gerstein and Altman [Gerstein & Altman 1995] was used to determine a conserved core region of the CaM dataset which could be used to provide a consistent reference frame for the superposition of the structure prior to PCA.

Provided with an un-gapped alignment of sequences of identical length and the corresponding PDB files, the program calculates error ellipsoids at each atomic position, and then calculates the respective volumes. The atom corresponding to the largest error ellipsoid is then discounted and the structures superposed without that atom. This process repeats until all the atoms have been removed and the program reports a *throw-out* order for the atoms with corresponding volumes for the atom removed at that iteration. Various options are then offered to the user in order to decide what should constitute the core:

Table 3.3: *Featured Methods for core determination in AvgCore. Each of the four methods provides a way of deciding what cutoff to use to separate the core and non-core regions of the structure on the basis of error ellipsoid size.*

Method	Description
1	The final 75% of the atoms thrown out are defined as the core.
2	The core cutoff is defined as the first iteration for which the volume of the ellipsoid being thrown out of the core is less than a constant threshold, currently 0.25Å.
3	The core, for a structure with n atoms, is defined as (a) the last 'floor(n/2)' atoms to be thrown out, plus (b) any other atoms with throwout volumes within 2.0 standard deviations of the mean of the volumes of the atoms in (a) at iteration 'floor(n/2)'.
4	The core is determined by looking at the variance of the volumes of the non-core atoms fitted to each other at every iteration. The first peak of this variance, beyond the first 20% of the iterations, is reported as the core cutoff.

Method 2 was used for this investigation using a threshold appropriate to the data reported in the results section 3.3.4.1.

3.2.5 Normal Mode Analysis

A NMA analysis of the 1.0Å resolution Ca^{2+}-bound CaM crystal structure **1exr** was performed on an all atom representation of the two lobes using the CHARMM 27a1 program [Brooks *et al.* 1983]. Polar hydrogens were added in QUANTA [Accelrys 2001] and parameterised with the PARAM19 and TOPH19 parameter sets [Neria *et al.* 1996]. For the analysis of the N-term, residues 2-74 were used and for the C-term residues 75-147. A non-bonded cutoff of 9Å was used with a switching function applied to Van der

Waals interactions over a 5-8 Å distance. Solvent screening was implicitly modelled via the shifted distance dependent dielectric function [Brooks *et al.* 1988].

The minimisation protocol used is outlined in Chapter 2 to a final tolerance of 1 x 10^{-8} kcal mol^{-1} $Å^{-1}$. The Vibran module of CHARMM was used to generate 3N (N= number of atoms) vibrational modes via the diagonalisation of the mass-weighted matrix of second derivatives of the potential energy. The frequencies of these vibrations were then used to calculate the associated entropic components as discussed in Chapter 6. The normal mode displacement vectors were scaled to appropriate magnitudes for comparison with the principal components of variation in the observed structural dataset.

3.2.6 Principal Components Analysis (PCA)

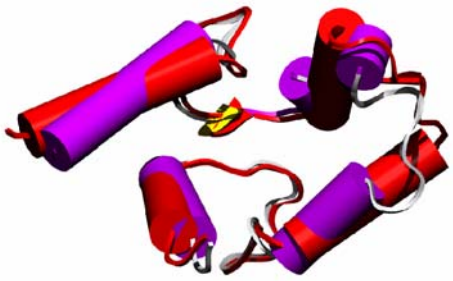
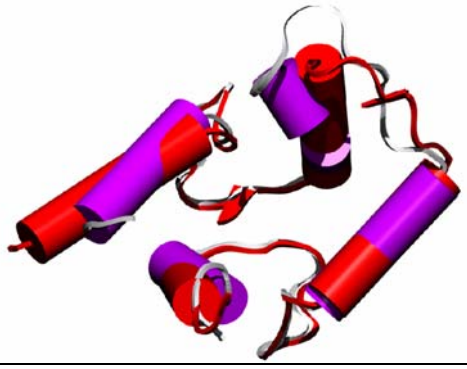
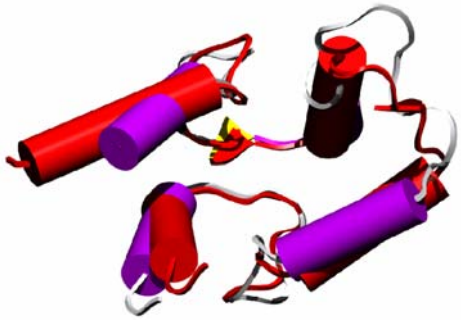
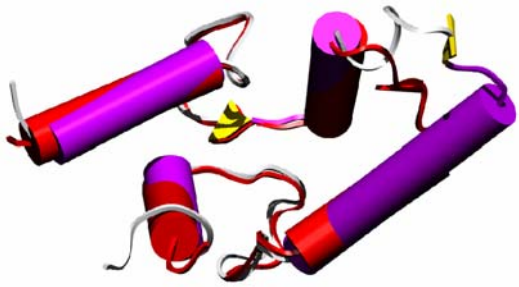
The multivariate analysis was performed on the Cartesian coordinates of different structural fragments using the R statistical package [Gentleman & Ihaka 2000]. Plots were generated initially using R and formatted externally. Cα coordinates were read into data matrices in R and then superposed to a common structure. An eigen analysis of the covariance matrices of the structures gave the principal components of the distributions as outlined in Chapter 2.

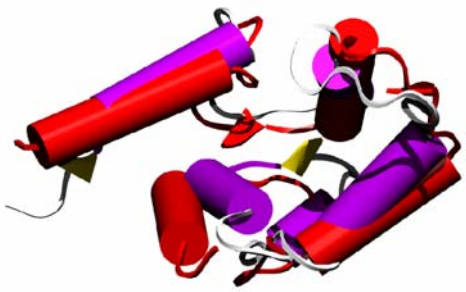
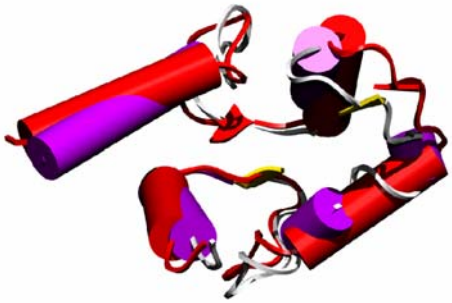
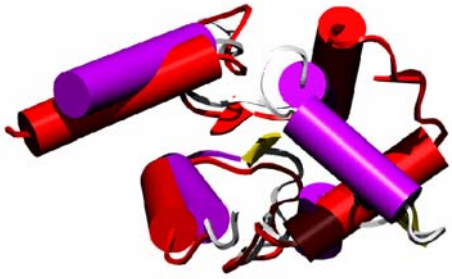
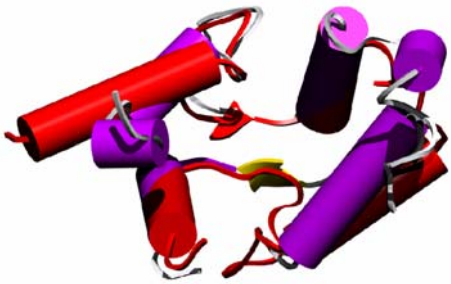
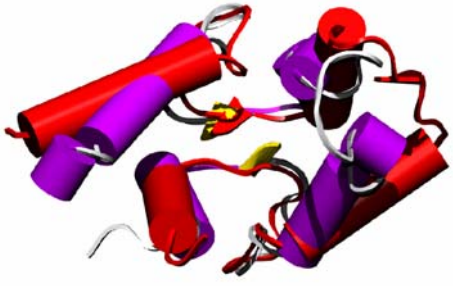
3.3 Results & Analysis

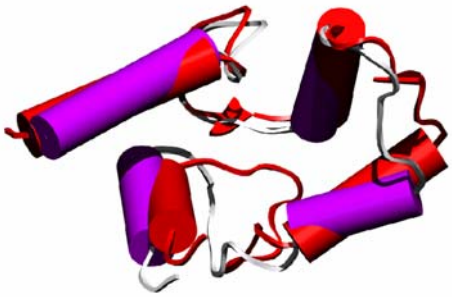
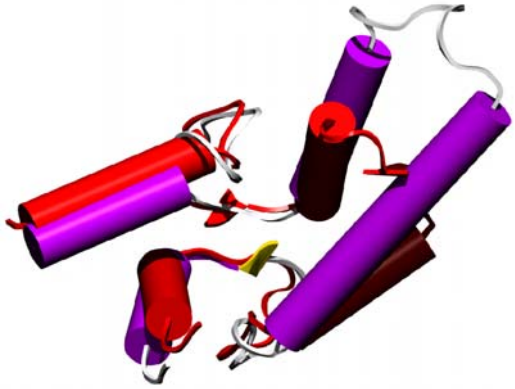
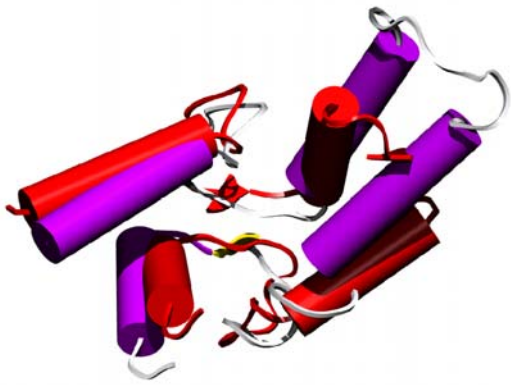
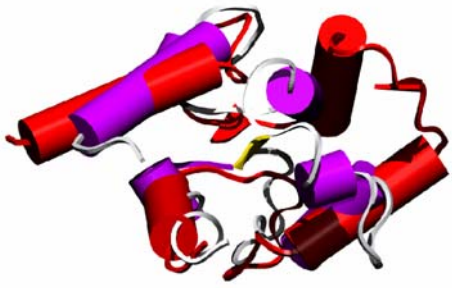
3.3.1 Structural Comparison by Inspection

A review of the Structural classification databases outlined in Chapter 2 suggested FSSP [Holm & Sander 1996] to be the best tool for the identification of Calmodulin Structural homologues on the basis of its continuous update and comprehensive listings. The domain identifier corresponding to the lobes of the 1.0Å CaM structure, **1exr** [Wilson & Brunger 2000], ***DC_3_241_1***, was used to identify the structures of the same fold. The Z-scores for the matches in the full results table, included in the appendix, range from 21 for the identity through to 2. Visual inspection of these structures reveals a high degree of variability across the set of structural partners. The general fold of the CaM lobe is conserved across the dataset and all structures exhibit the double helix-turn-helix EF-hand motif. The linker between these units however is more variable in both length and conformation as can be seen in the structures shown below (**Table 3.4**). From these observations, thirteen of the apparent structural homologues were removed immediately due to obvious differences in structure. Fragments of Troponin C were the only structures found to be sufficiently similar to CaM structures to be included in the PCA.

Table 3.4: *Images and comments in this table indicate the reasons for removal of some CaM homologues from the PCA. In all cases the reference structure (**1exr**) is coloured in red and the target structure is coloured according to secondary structure; α-helices coloured in purple, β-sheet coloured in yellow and loops etc in white. Helix nomenclature corresponds to that outlined in Figure 1.3 in Chapter 1. The linker (right) joins the two EF-hands together.*

PDB Code.	***Superposed Image***	***Comments***
1aui		Both helices B and C show considerable difference to those of CaM. Helix B is *broken* half way along and helix C is considerably shorter than the corresponding helix in CaM
1a75		Here the linker between helices B and C is extended relative to the CaM structure resulting in a distorted helix B.
1c7v		The linker region between helices B and C is considerably different in this structure and helices B and C are different lengths to those in the reference structure.
1dgu		Helix C in this structure is much longer that that in CaM.

1djx		An interesting anomaly in this structure is the extra helical section observed in the Ca^{2+} binding loop found between helices C and IV.
1mxl		In this structure, helix C is unwound into two separate short helices. The conformation of the first EF-hand is also considerably different to that of CaM.
1psr		Here the linker between helices B and C has formed a helix which is much longer than the normal linker in CaM.
1rec		In this structure, helices B and C are considerably longer than normal and the linker between them includes a single turn helix.
1sra		The linker between helices B and C in this protein fragment is quite short and again includes a very short helix unit.

1wdc		This fragment only has one EF-hand structure capable of binding Ca^{2+}.
2sas		Once again this structure has considerably elongated helices B and C
2scp		This structure has two elongated helices, B and C, with a short π -helix (5 residues per turn) at the bottom of helix C.
4icb		This structure exhibits an extended linker between helices B and C relative to CaM which again includes a very short helix.

3.3.2 Amino acid Sequence Comparison

In order to perform this analysis it was necessary to have continuous data for the protein fragments. All CaM structures determined by X-ray crystallography have been unsuccessful in resolving the electron density at the extreme ends of CaM sequence. The highest resolution structure, **1exr**, only has coordinate information from residue 2 and up to residue 147, older structures determined to lower resolution have even more missing density at the termini. There are also problems resulting from missing coordinates in the central helix of the protein, partly where it has been too flexible to resolve and partly where the protein sequence had been truncated in order to study one or other of the two lobes independently. In order to acquire the continuous information across the full ensemble therefore, the structures were truncated to the regions for which structural coordinates were available in all structures. The resulting N-terminal fragment included 69 protein residues and the C-term fragment 62.

Sequence alignments of the protein fragments for the N and C-terminal lobes are illustrated in **Figure 3.2** and **Figure 3.3** below. These alignments illustrate the high level of homology between CaM sequences from different sources. The CaM sequences shown here include examples of Human, Bovine, Drosophila, Rat and other CaMs which are all highly similar. *Paramecium tetraurelia* CaM in **1clm**, **1osa** and **1exr** show minor differences to other species as does the synthetic CaM construct in **1vrk**. Even the sequence from the Troponin C structures are very similar to CaM. The Calpain sequence however in structures **1alv** and 1alw however shows marked differences to CaM, hence the positioning at the bottom of the C-term alignment (**Figure 3.3**).

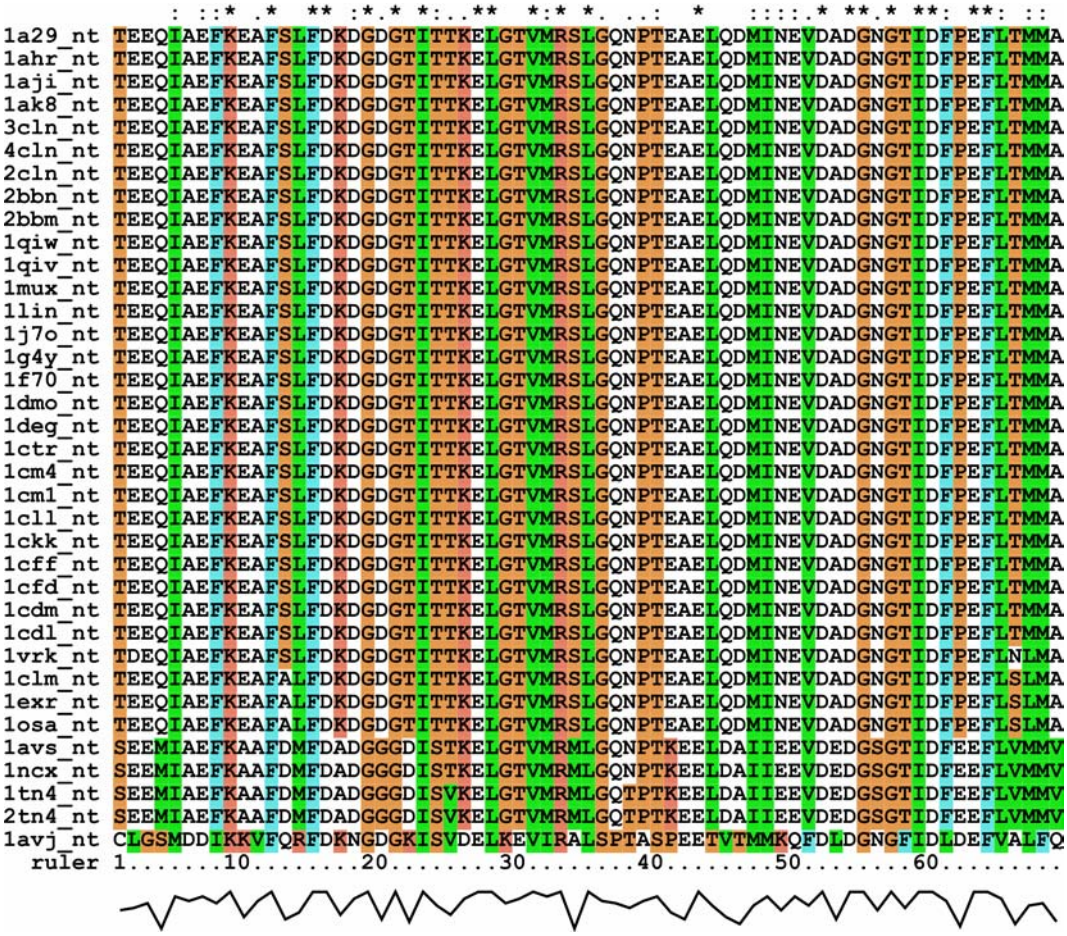


Figure 3.2: *ClustalX Sequence alignment of the N-terminal fragments used in the initial analysis.*

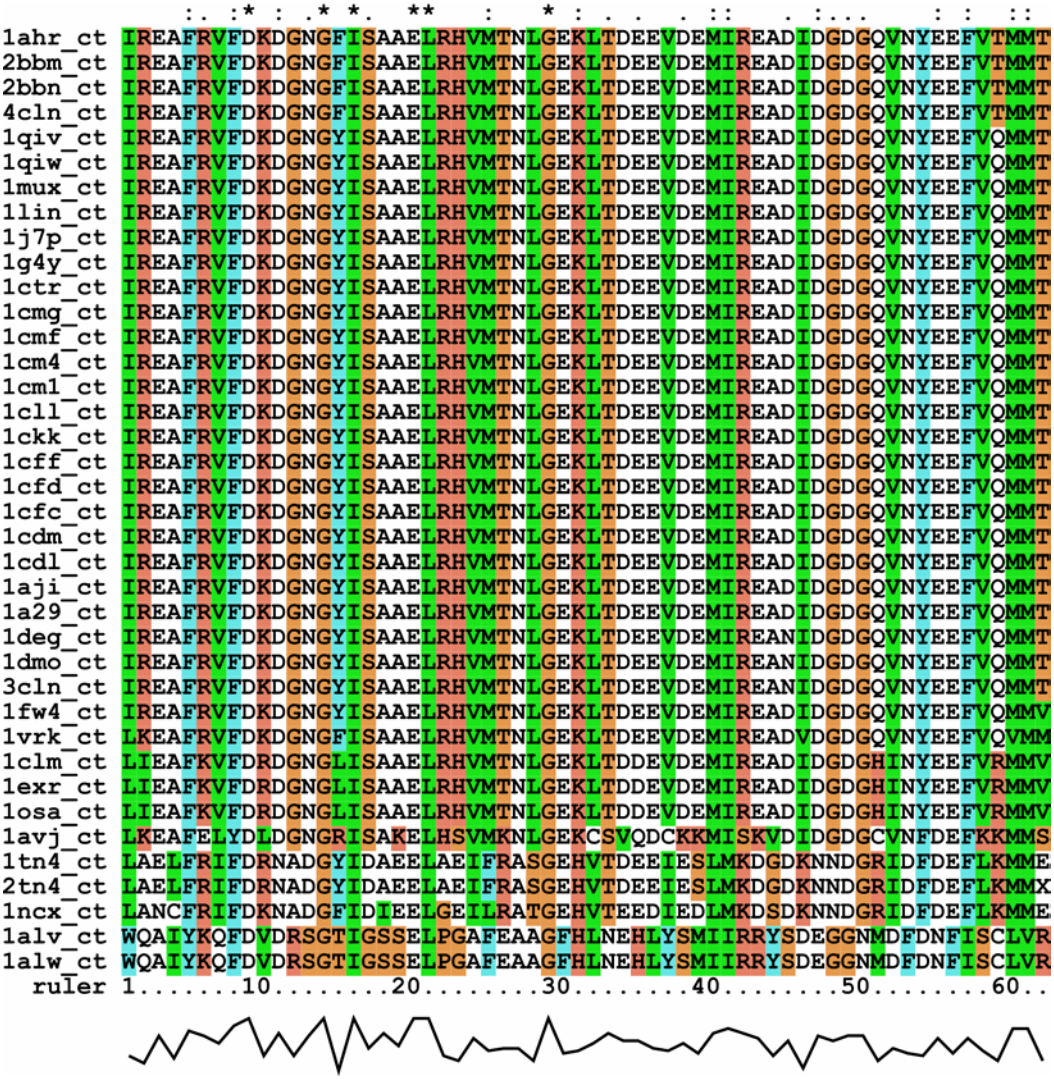


Figure 3.3: *ClustalX Sequence alignment of the C-terminal fragments used in the initial analysis.*

3.3.3 Principal Components Analysis

Details of the full structural ensemble of structures used in the initial analysis are included in **Table 3.6** in which the structures are classified according to the categories of **Table 3.5**. This classification by colour helps to highlight the differences between the structures presented in the results plots.

Table 3.5: *Categorisation and colour key for subspace projection plots from the PCA.*

Category	*Description*
Native : (Black)	Ca^{2+} bound CaM with no ligands or peptides associated to the binding surfaces.
Apo: (Red)	Ca^{2+} free CaM lobes.
Ligand: (Green)	Structures with a small molecule associated with the CaM lobe.
Target: (Blue)	These structures have a full target peptide interacting with the CaM lobe.
Model: (Cyan)	Structures derived from modelling rather than observed experimental data.
Special: (Magenta)	Structures of particular note. These include structures with critical mutations and non-CaM structures.

The figures below summarise the results of the Principal Components Analysis for the full structural data set of CaM and its homologues in **Figure 3.4** for the N-term lobe and **Figure 3.5** for the C-term. Data points are labelled by PDB code identifiers and coloured according to the category (**Table 3.5**). The eigen value spectrum beneath the conformational subspace projections indicate the extent to which the principal components describe the variation in the data.

Table 3.6: *The full CaM Structural Dataset of experimentally determined*

Code	Descriptor	Experimental Technique	Classification
1a29	Calmodulin	XRD	Ligand
1ahr	Calmodulin Deletion Mutant (T79, D80)	XRD	Special
1aji	Calmodulin & Myosin Seq.	Model	Model
1ak8	Calmodulin (N-term)	NMR (23 models)	Apo
1alv	Calpain	XRD	Special
1alw	Calpain	XRD	Special
1avj	Calmodulin-Type Tch2 Protein	Model	Special
1avs	Troponin C (res. 1-90)	XRD	Special
1cdl	Calmodulin & sm-MLCK	XRD	Target
1cdla	Calmodulin & sm-MLCK Chain A	XRD	Target
1cdlb	Calmodulin & sm-MLCK Chain B	XRD	Target
1cdlc	Calmodulin & sm-MLCK Chain C	XRD	Target
1cdld	Calmodulin & sm-MLCK Chain D	XRD	Target
1cdm	Calmodulin & CaMKII	XRD	Target
1cfc	Calmodulin	NMR (25 models)	Apo
1cfd	Calmodulin	NMR (Avg)	Apo
1cff	Calmodulin	NMR (26 models)	Target
1ckk	Calmodulin	NMR (30 models)	Target
1cll	Calmodulin	XRD	Native
1clm	Calmodulin	XRD	Native
1cm1	Calmodulin & CaMKIIα	XRD	Target
1cm4a	Calmodulin & CaMKIIα Model A	XRD	Target
1cm4c	Calmodulin & CaMKIIα Model B	XRD	Target
1cm4e	Calmodulin & CaMKIIα Model C	XRD	Target
1cm4g	Calmodulin & CaMKIIα Model D	XRD	Target
1cmf	Calmodulin (C-term)	NMR (20 models)	Apo
1cmg	Calmodulin (C-term)	NMR (20 models)	Native
1ctr	Calmodulin & TFP	XRD	Ligand
1deg	Calmodulin Mutant (Del E84)	XRD	Special
1dmo	Calmodulin	NMR (30 models)	Apo
1exr	Calmodulin	XRD	Native
1f70	Calmodulin (N-term res 1-76)	NMR (RDC)	Apo
1fw4	Calmodulin (C-term res 78-148)	XRD	Native
1g4y	Calmodulin & Rsk2	XRD	Target
1j7oa	Calmodulin (N-term) Model A	NMR (RDC)	Native
1j7ob	Calmodulin (N-term) Model B	NMR (RDC)	Native
1j7oc	Calmodulin (N-term) Model C	NMR (RDC)	Native
1j7pa	Calmodulin (C-term) Model A	NMR (RDC)	Native
1j7pb	Calmodulin (C-term) Model B	NMR (RDC)	Native
1j7pc	Calmodulin (C-term) Model C	NMR (RDC)	Native
1lin	Calmodulin	XRD	Ligand
1mux	Calmodulin	NMR	Ligand
1ncx	Troponin C	XRD	Special
1osa	Calmodulin	XRD	Native
1qiv	Calmodulin	XRD	Ligand
1qiwa	Calmodulin Chain A	XRD	Ligand
1qiwa	Calmodulin Chain B	XRD	Ligand
1tn4	Troponin C	XRD	Special
1vrk	Calmodulin & RS20	XRD	Target
2bbm	Calmodulin & sk-MLCK	NMR (Avg)	Target
2bbn	Calmodulin & sk-MLCK	NMR (21 models)	Target
2cln	Trimethyl-calmodulin & TFP	Model	Special
2tn4	Troponin C Mutant (C98L)	XRD	Special
3cln	Calmodulin	XRD	Native
4cln	Calmodulin	XRD	Native

- **N-Term Results:**

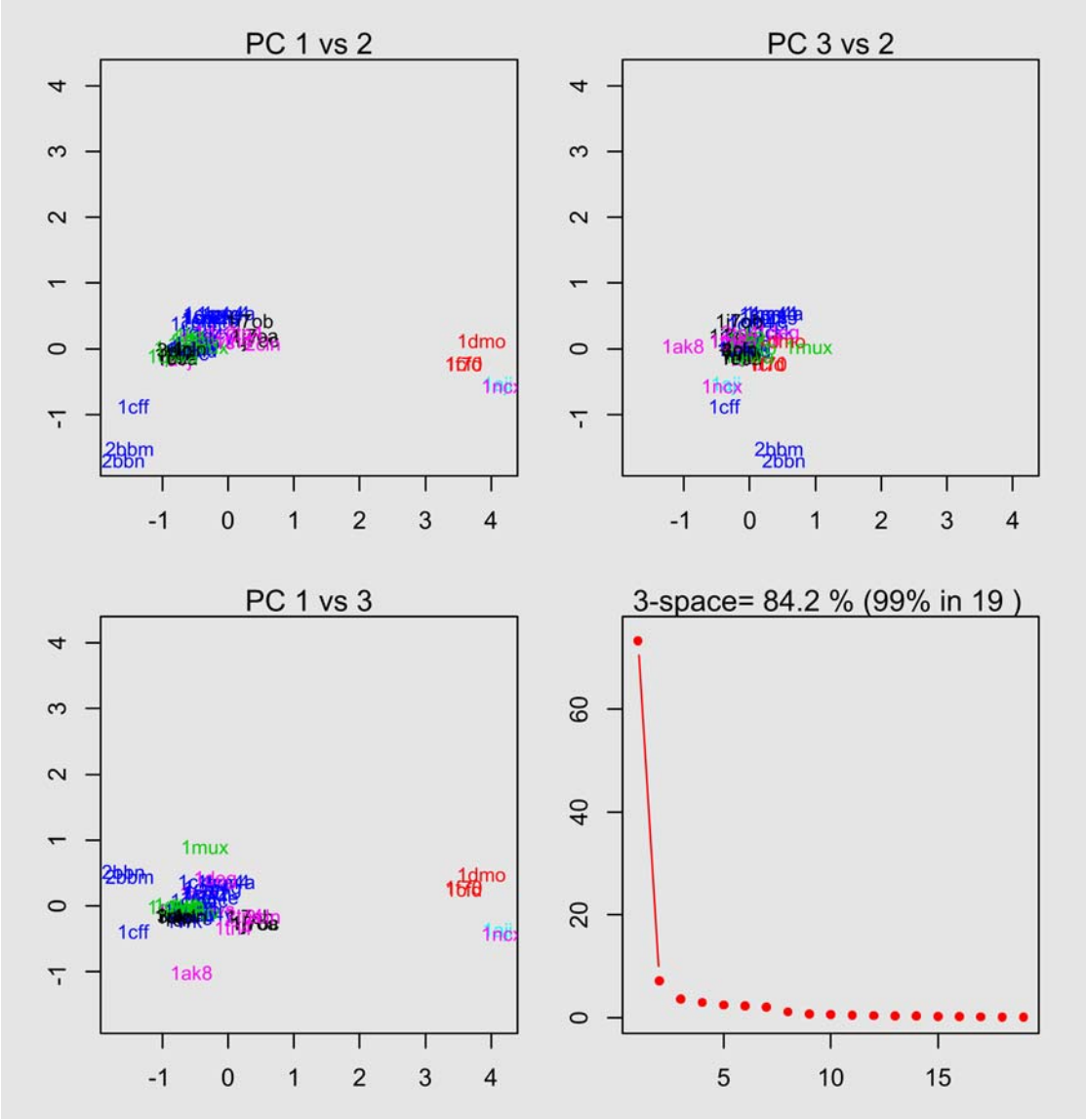


Figure 3.4: *Conformational subspace projections for the full structural dataset. Structures are projected onto the first three principal components of the data for the N-term lobe. The Eigen spectrum also shows the dimensionality of the data.*

The eigen value spectrum for the N-term indicates that the first principal component (PC 1) describes 75 % of the variation across the while structural data set. This property is illustrated in the subspace projection in as much as the data points are primarily distributed parallel to the x-axis (PC 1). The large cluster of structures around the origin indicates that most of the structures do not project well onto these first two PC's as the distribution is heavily skewed by the outliers shown. The projection of the structures onto PC's 2 and 3 shows how these PC's describe very little of the variation as all the structures are tightly clustered.

Inspection of the plot for PC 1 vs. PC 2 identifies the members of the small cluster at 4Å on PC 1 as **1aji**, **1dmo**, **1cfc** and **1ncx**. **1aji** is a modelled structure [Houdosse *et al.* 1996] and consequently worth special consideration. **1dmo** and **1cfc** are both NMR structures of Apo (Ca^{2+} free) CaM and thus can be expected to present an alternate conformation to classical Ca^{2+} bound CaM. **1ncx** is a structure of Troponin C binding Cd^{2+} and again some structural difference is perhaps to be expected.

Outliers on PC 2 are **2bbm**, **2bbn** and **1cff** which are all NMR structures of CaM in complex with target peptides. This result adds weight to concerns NMR and crystal structure comparison at high levels of detail (note the discussion in section 2.4.2 regarding the different types of information provided from NMR and X-Ray diffraction studies of protein structure).

- **C-Term Results:**

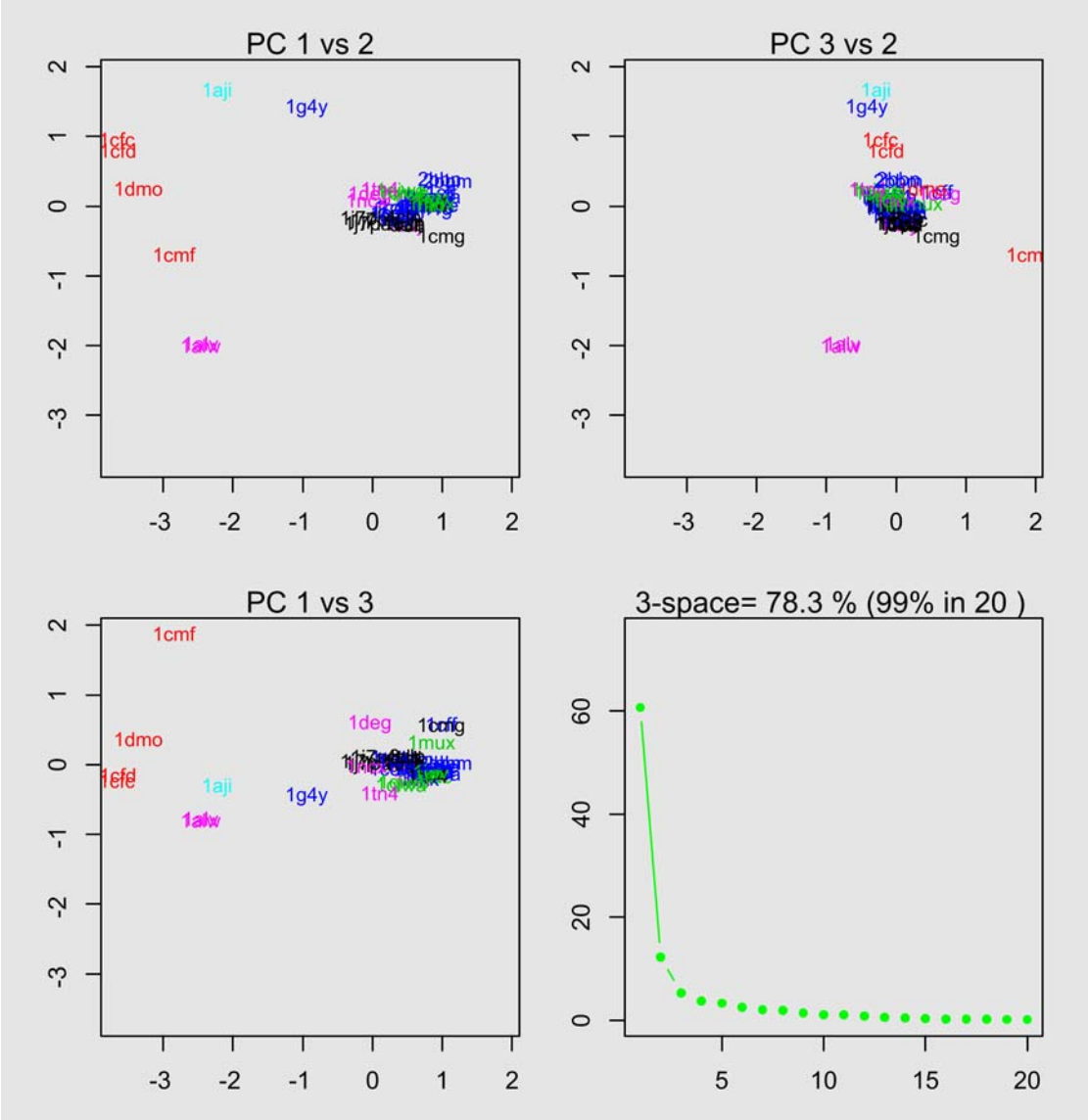


Figure 3.5: *Conformational subspace projections for the full structural dataset. Structures are projected onto the first three principal components of the data for the C-term lobe. The Eigen spectrum also shows the dimensionality of the data.*

The eigen value spectrum for the C-term shows that a second and third PC is more significant in the description of structural distribution here. Again this is illustrated in the subspace projections as the structures are spread more out along the PC's. Once again the majority of the structures are clustered around the origin and the few outliers cause much of the variation in the ensemble.

The outliers can be identified as the NMR apo-Ca^{2+} structures **1cfc**, **1cfd**, **1dmo** and **1cmf**. Also weighting the distribution are the modelled structure **1aji** and the 2.0Å crystal structure of Calpain, **1alw**. The structure **1g4y** has no Ca^{2+} in the C-term lobe and consequently lies separate from the main cluster of CaM lobe structures for the C-term, however it cannot be distinguished from the main cluster of N-term structures highlighting the independent nature of the two CaM lobes.

- **Overview:**

It is clear from these results that the subtleties of conformational variation cannot be seen in this analysis as the most accurately determined structures are tightly clustered together around the origin of the plots. Many of the structures differing significantly from the main cluster of structures can be identified as being structures determined by NMR experiments, modelling experiments or non-CaM structures. The removal of such structures from the analysis will help to identify the more detailed trends in CaM conformational flexibility.

3.3.3.1 High Resolution CaM Dataset

The final data set was obtained by repeating the analysis with a reduced set of structures which excluded those found to considerably skew the analysis. All the NMR structures were removed as were the modelled structures. The structure **1deg** was also removed as the structure had only been resolved to 2.9 Å and included coordinates for only the alpha-carbon atoms. All none CaM structures were also removed from the analysis at this stage although it is interesting to note however, that the Troponin C structure **1avs** in particular showed a very similar N-term structure to CaM. The C-term fragment of the CaM complex structure **1g4y** also was removed although the N-term was included in the analysis.

- **N-Term Results:**

The conformational subspace plots shown below highlight trends in structural variablity for the N-term lobe of CaM to a very high resolution. Note that the total range of this data is now only 1.9Å.

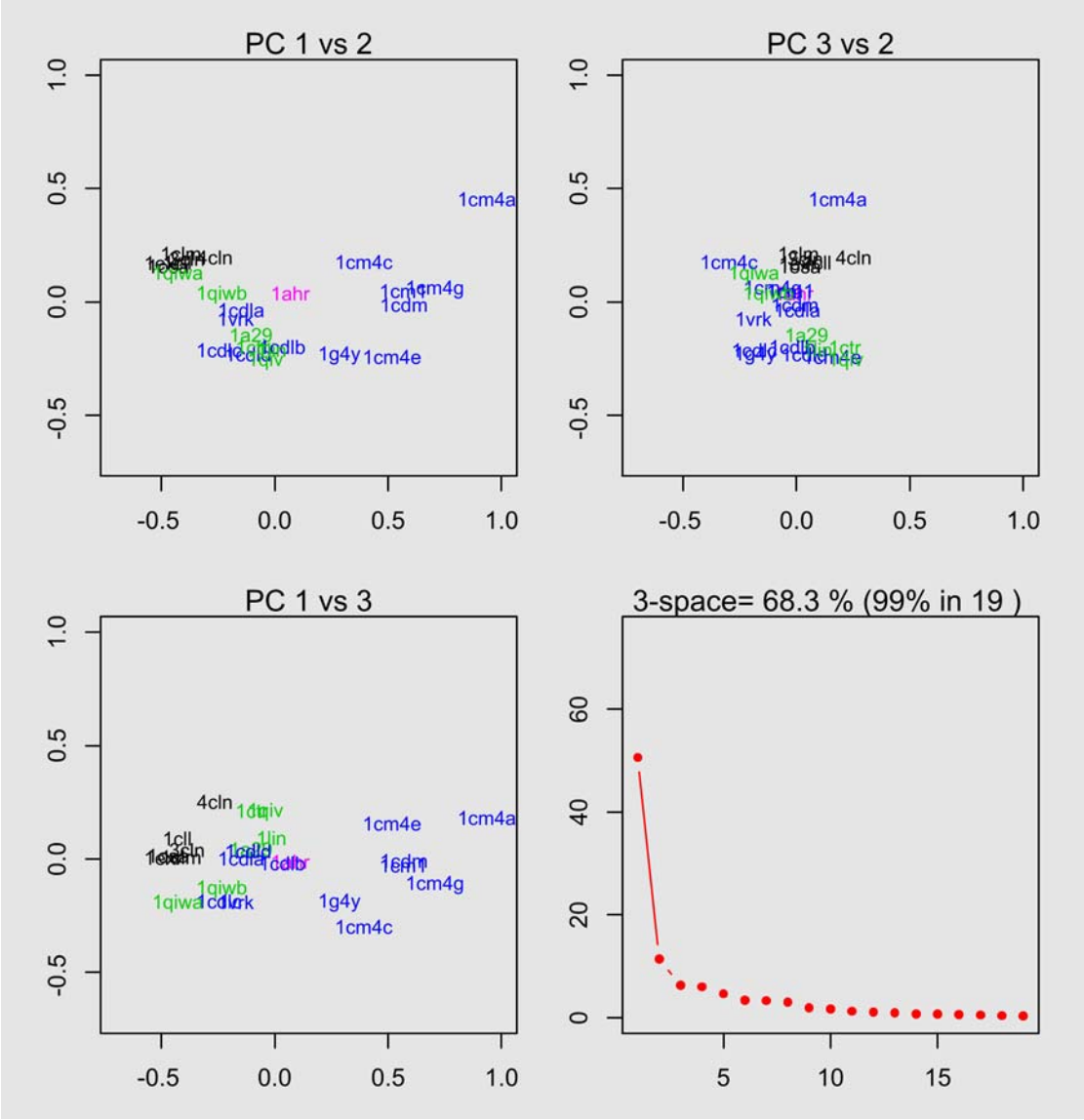


Figure 3.6: *Conformational subspace projections for the high resolution structures of CaM onto the first three PC's. The eigen spectrum indicates the extent to which the variation is described by the first few PC's.*

The variation of this data is again largely one-dimensional with PC 1 describing over 50% of the observed flexibility in structure with PC's 2 and 3 describing 11% and 6% of the data respectively. The native Ca^{2+}-bound structures (black) are tightly clustered together

(**4cln** to a lesser extent) whilst target peptide bound structures (blue) lie further along PC 1 in a positive direction.

This result suggests that a conformational change takes place in the CaM N-term lobe when it interacts with a target, this is partly observed for small molecule inhibitors but is variable in magnitude dependent on the nature of the target. Structures for the A, B, C and D chains of **1cdl** lie in a similar position along PC 1 as do chains A, C, E, and G of **1cm4**. The inhibitor structures (green) are also projected slightly further along PC 1 and PC 2 than the native structures but not as far as most of the peptide bound structures. The ligand-bound crystal structure **1qiw** [Harmat *et al.* 2000] shows little structural deformation with respect to the native structures.

Projection of the average structure from this analysis along the eigen vectors generated in this analysis can help in the understanding of the conformational change described by these PC's. Such projections can also be animated in molecular visualisation packages such as Molviewer [Hartshorn 1996] which aids the analysis.

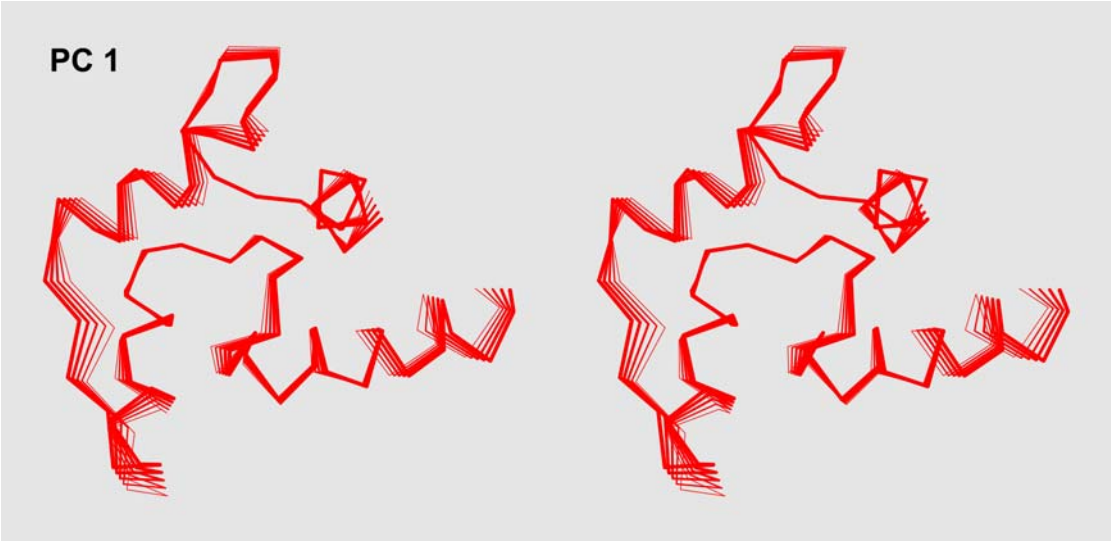


Figure 3.7: *(Cross-eye stereoplot) Projection of the average structure along PC 1 for the N-term lobe.*

This displacement of the average structure along PC 1 is mostly afforded by helices A and D (see the anatomy of CaM in Figure 1.3 of Chapter 1) of the lobe. The Ca^{2+}-binding loops of the lobes however are well retained across the ensemble. Much of the flexibility in the helices appears to come from compaction of the helices along their principal axis. There is also a degree of motion perpendicular to the axes of the helices

which enables the helices to pivot about the Ca^{2+}-binding loops in the same was as the conventional CaM open-closed conformational transition [Nelson & Chazin 1998].

Hierarchical cluster analysis can also be used to analyse the similarities and differences between the structures as shown in **Figure 3.8**. The distinction between target-bound and native structures is again clear as seen in PCA projections. The structure **1ahr** however appears to be quite similar to other structure in the conformational subspace projections but here it is separated from all the other clusters. Re-inspection of the results in **Figure 3.6** reveals that **1ahr** stays close to the origin in PCs 1, 2 and 3 suggesting that these are not good descriptions of the differences between **1ahr** and other conformations. This analysis suggests that **1ahr** is not the same as any of the other types of conformation, native, inhibitor-bound or target-bound.

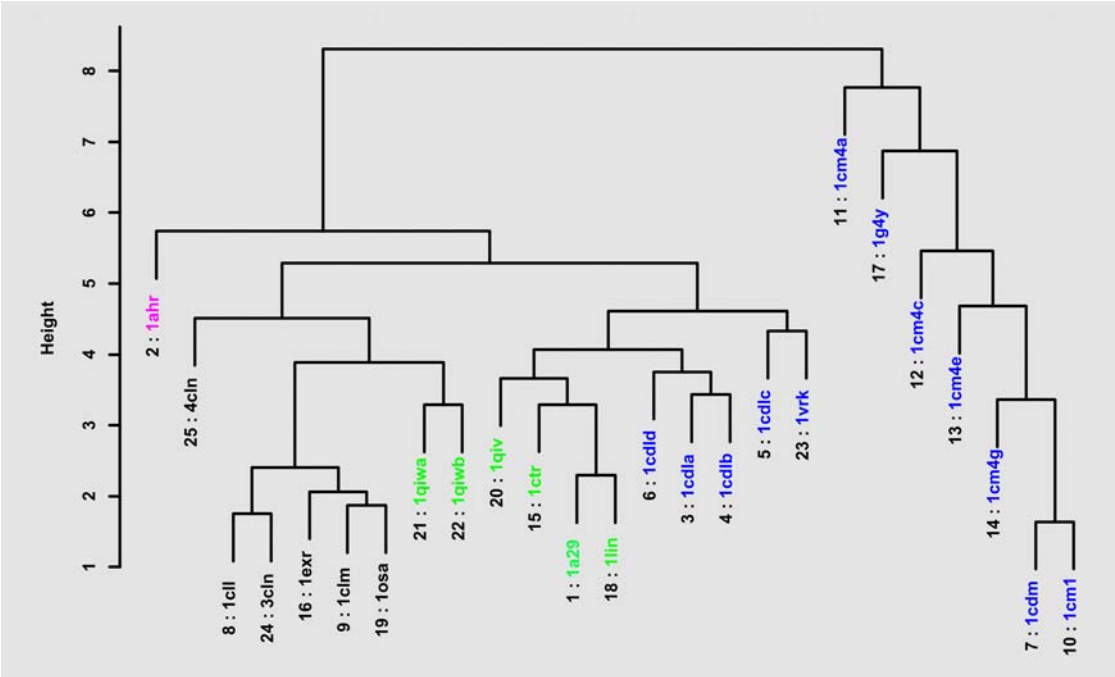


Figure 3.8: *Hierarchical cluster diagram of observed structures. Relations are constructed by average linkage clustering of the distances between the transposed coordinates. Structure identifiers coloured according to the structure classification Table 3.5*

- **C-Term Results:**

Conformational subspace projections and the eigen-spectrum for the analysis of the C-term structures are shown below (**Figure 3.9**). The eigen-spectrum for these structures is flatter than for the N-term and PC 1 only describes 37% of the total variation compared to over 50% for the N-term. PC's 2 and 3 are also more significant than for the N-term describing 16 and 10% of the total variation respectively compared to 11 and 6% for the N-term. The distribution of structures projected onto these PC's is more compact covering slightly less than 1Å in total along PC 1.

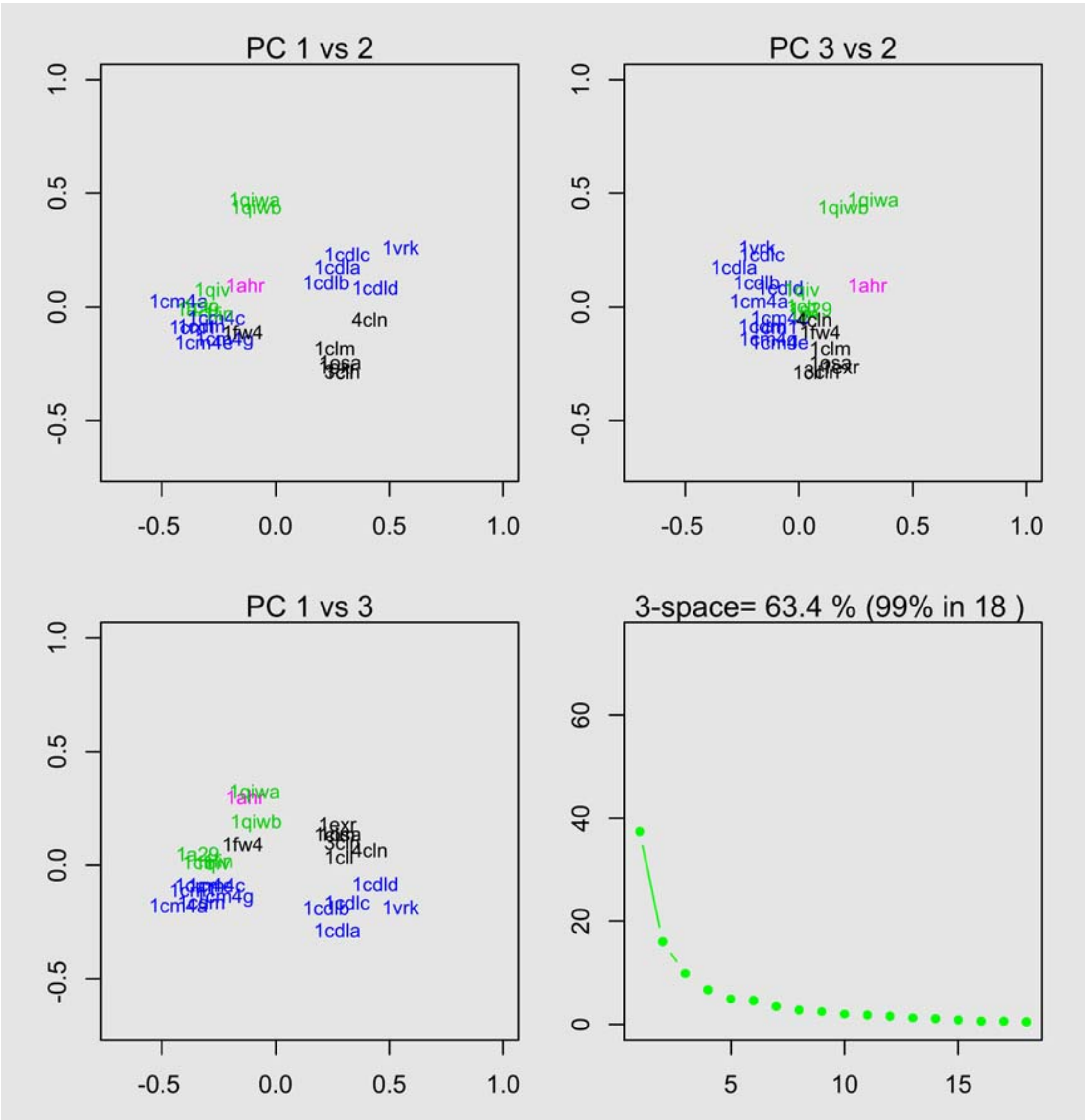


__Figure 3.9:__ Conformational subspace projections for the high resolution CaM C-term structures onto the first three PC's. The eigen spectrum shows that 18 PC's are required to describe 99% of the variation in the data.

The projection of these structures onto PC 2 reveals some separation into bound and unbound conformations although this is less well defined than in the N-term. Clustering of observed conformations however is more pronounced in this case with four main clusters observed for the projection onto PC's 1 and 2. One cluster includes most of the native Ca^{2+}-bound CaM structures. A second cluster incorporates the four conformers of **1cdl** and **1vrk** which are all target-bound structures with a similar target peptide. The third cluster includes a mixture of target-bound structures (**1cm1** and **1cm4**), the native C-term only structure **1fw4**, most of the inhibitor structures and also the linker mutant **1ahr**. The fourth cluster contains only the two conformers of the ligand-bound structure, **1qiw**.

Displacement of the average structure along PC 1 shows similar trends to those of the N-term. Most of the flexibility is observed in helix H at the C-term end of the protein which involves displacements perpendicular to the axis of the helix. The Ca^{2+}-binding-regions of the lobe are more constrained while the linker between the two EF-hands again shows some conformational freedom. PC 2 describes more motion in the flexible loop regions of the structure either side of the EF-hands and the linker region rather than in the secondary structure units.

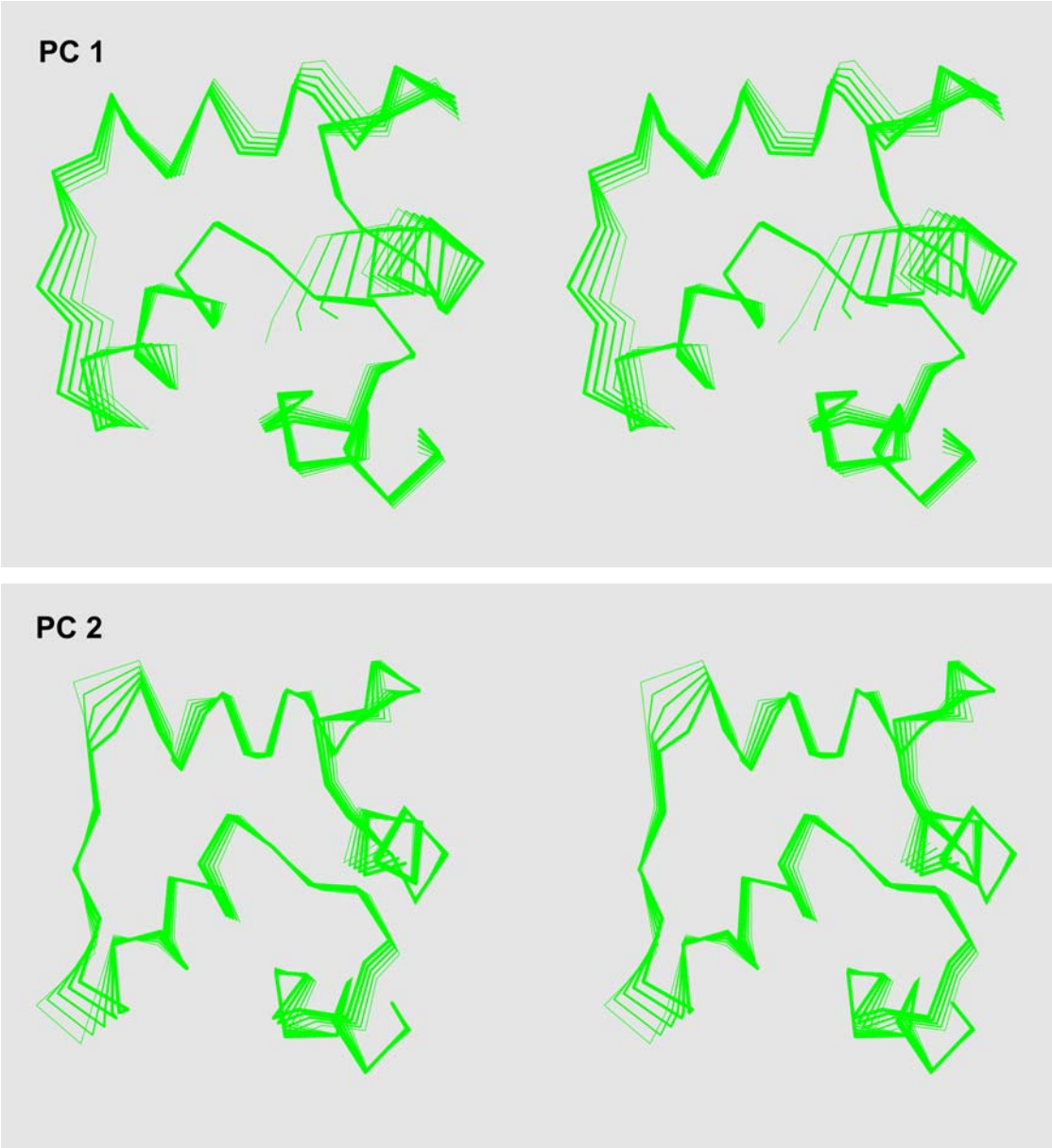


Figure 3.10: *(Cross-eye stereoplot) Dsiplacement of the average structure along PC 1 and PC 2 for the C-term lobe.*

The hierarchical cluster analysis of the superposed C-term structures shows a clearer separation of the different categories of structure than any of the individual PC's. The native structures cluster tightly together on the right hand side of the plot but share a branch with the **1cdl** and **1vrk** structures. The native C-term only structure **1fw4**, however, is associated more with the **1cdm** and **1cm4** structures which are found to be similar to the inhibitor structures. The inhibitor structure **1qiw** lies separate from the other inhibitor structures adopting a separate conformation.

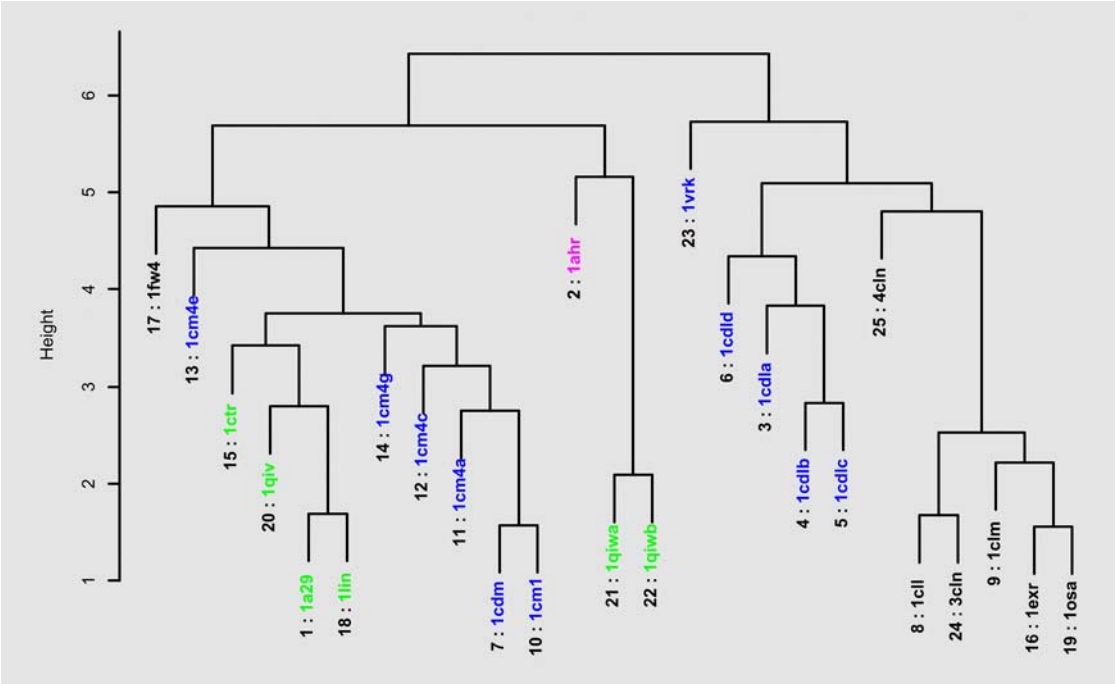


Figure 3.11: *Hierarchical cluster dendogram of the CaM C-term structures. Clustering performed by the average linkage method on pairwise Euclidean distances superposed coordinates.*

- **N and C-term Together:**

By truncating the N-term data set appropriately, it is possible to perform the analysis on both N and C term structures together in order to compare the relative flexibility and conformations of the different lobes. The N-term fragments were further shortened to a region of only 62 residues corresponding to the same fragments of the C-term lobe.

The eigen-spectrum for the whole structural ensemble is much flatter than the individual lobe analyses with 34 PC's required to describe 99% of the variation in comparison with 16 for the C-term and 11 for the N-term. Projections of the structures onto this combined N and C-term subspace shows a much more even spread of data points across the distribution. Interestingly, there is no overall distinction between C- and N-term lobe structures. The different structures still form clusters of the same lobe and same target association state but the cluster of Ca^{2+}-bound N-CaM is closer to Ca^{2+}-bound C-CaM than to any other N-CaM clusters. The same feature is observed in the hierarchical cluster analysis shown in **Figure 3.13**.

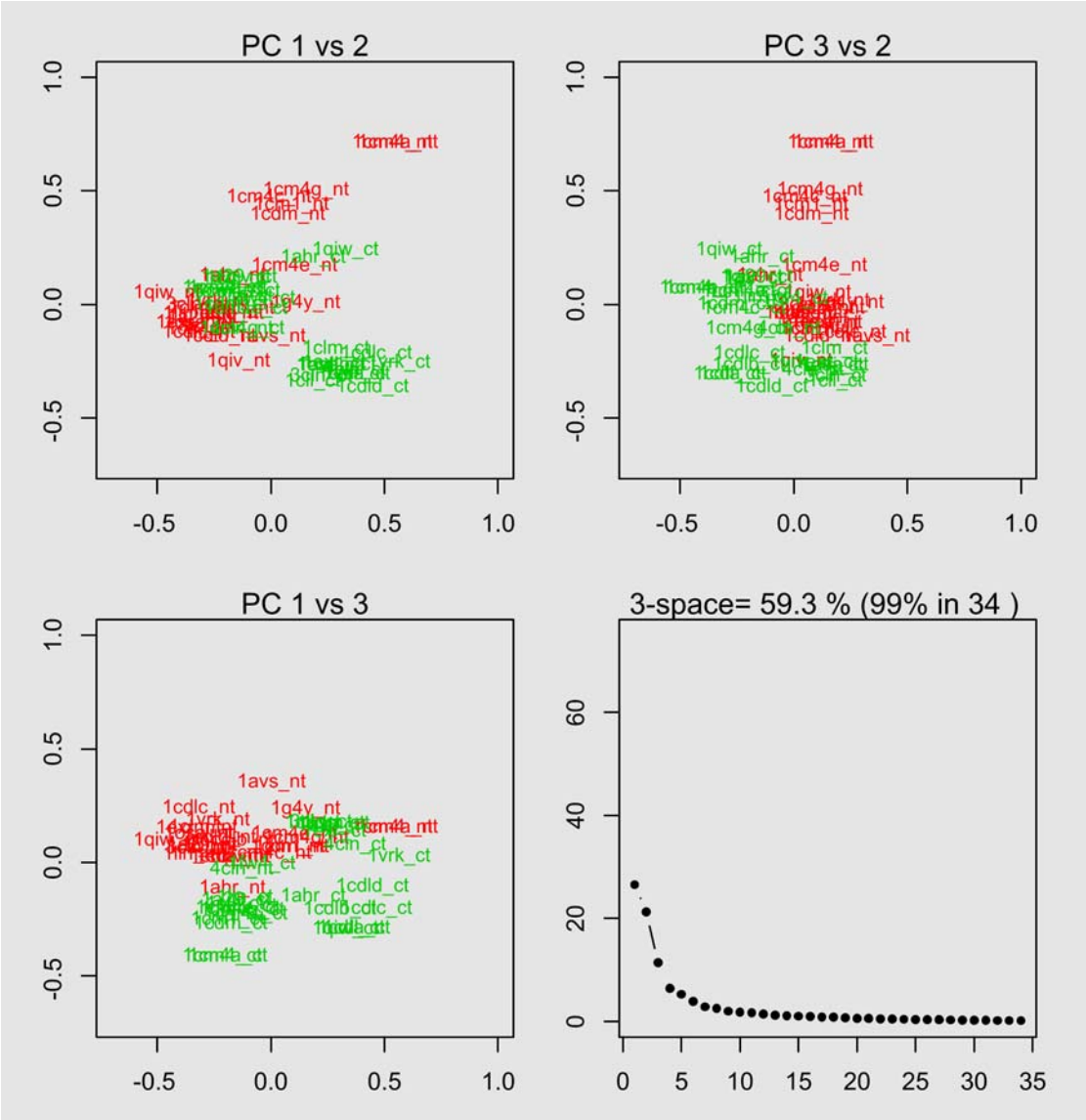


Figure 3.12: *Conformational subspace projections for the N and C lobes together with the corresponding eigen spectrum. N-term structures highlighted in* ***red*** *and C-term structures in* ***green****. The eigen spectrum indicates that PC's 1,2 and 3 are all significant in the description of the data.*

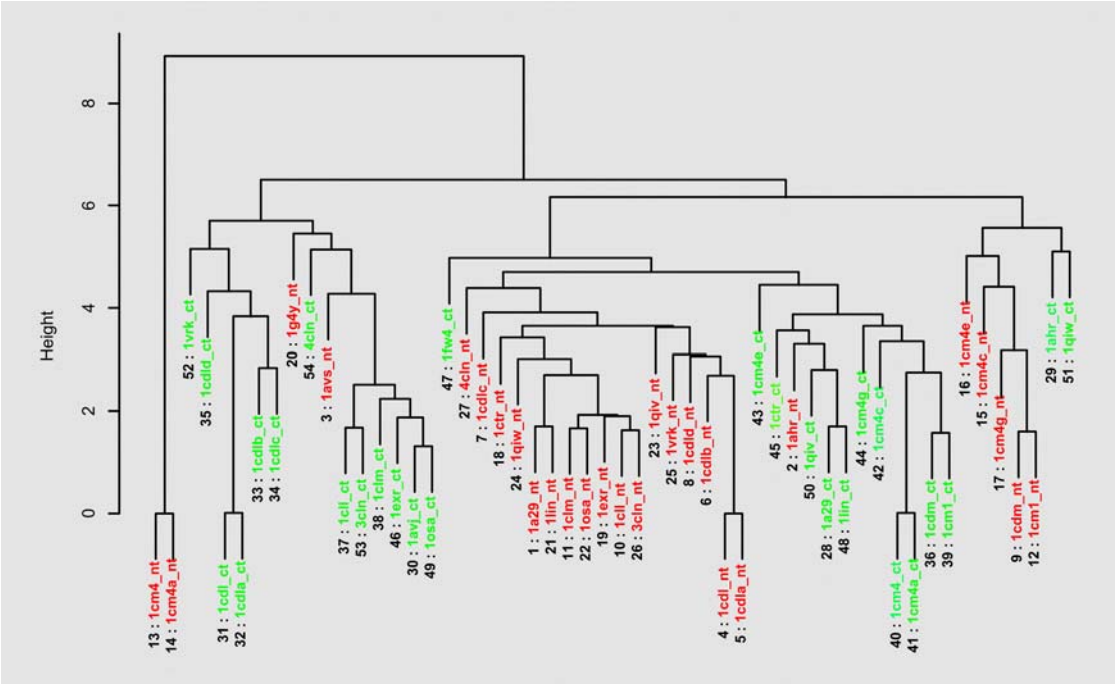


Figure 3.13: *Hierarchical clustering dendogram of the N and C-term lobes together. N-term structures highlighted in **red** and C-term structures in **green**.*

3.3.4 Normal Mode Analysis (NMA)

The minimised structure was superposed onto the same reference frame as the other structures in the PCA data-set and the corresponding transformation matrix applied to the Cα components of the normal mode vectors calculated in the simulation. The transformed vectors were then normalised and multiplied by the transposed matrix of principal components to obtain correlation coefficients. Projections of the ten lowest frequency normal mode vectors centred about the minimised structures of **1exr** are shown in the **Figure 3.14** and **Figure 3.15** below. The normal mode displacement vectors are projected onto the conformational subspace of the PCA of the N and C-term observed structures.

Although the crystal structure of **1exr** is a member of the tight black cluster of Ca^{2+}-CaM structures, minimisation alters the structure such that it no longer projects onto the same place on these PC's. These conformational subspace projections clearly show that the minimisation protocol applied to the crystal structure for **1exr** has caused considerable alterations to the original structures but this is necessary in order to explore the character

of the energy surface. The correlation values shown in the fourth plot are generated from a matrix multiplication of the *normalised* (but no longer orthogonal) modes after they have been transposed onto the new reference frame. Because of this, direct comparison of the magnitudes with the subspace projections is not possible. Mode 3 for example shows a better correlation with PC 6 than it does with either 1, 2 or 3 which means the angle between PC 6 and mode 3 is less than the others shown. However, since higher mode numbers represent higher frequencies, the corresponding amplitude of vibration will be smaller at any given temperature. It is primarily for this reason that only the first three modes have been projected onto the conformational subspace of the principal components: Mode 3 is already quite difficult to identify on the plots despite having correlation values comparable with modes 1 and 2.

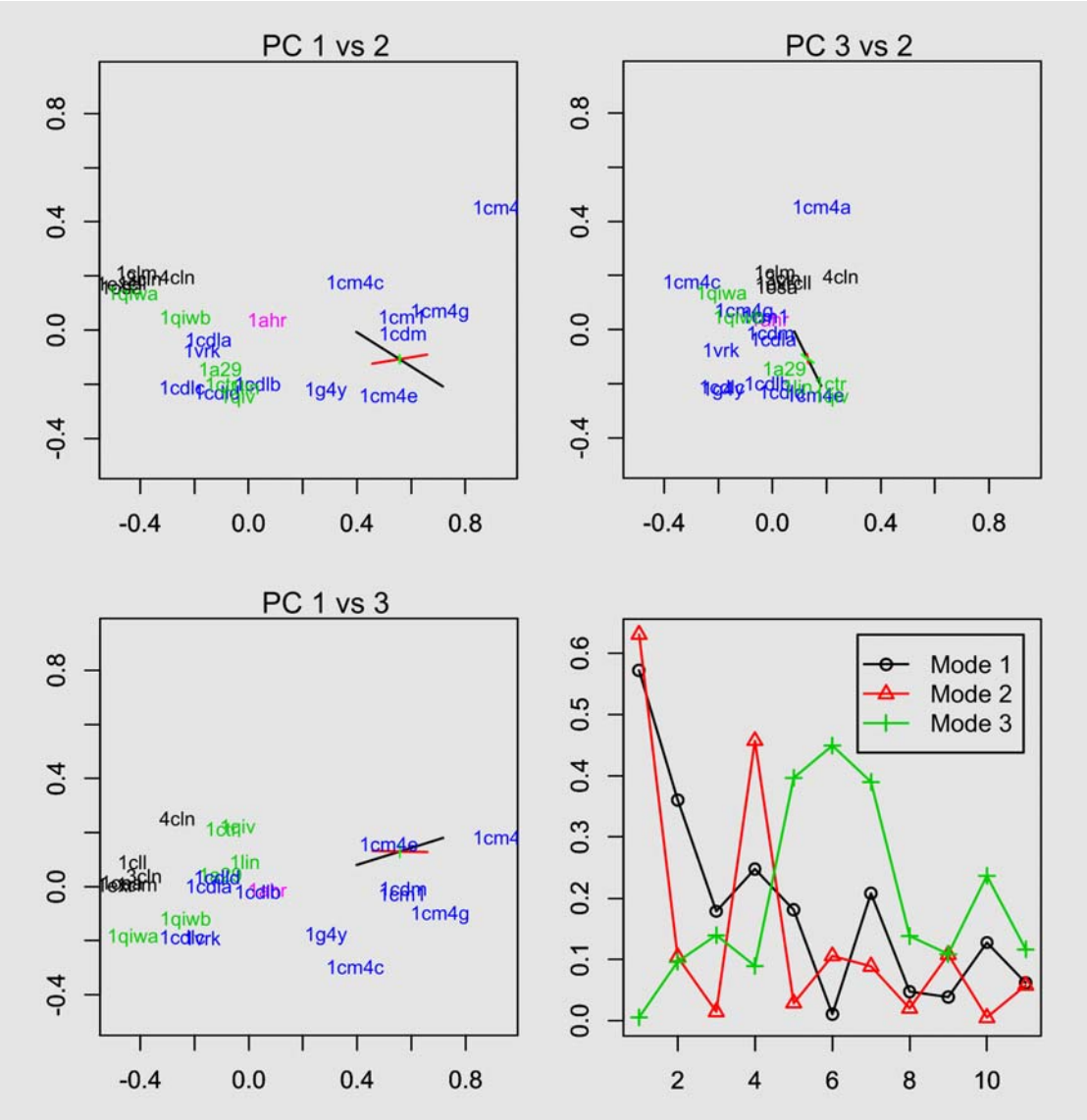


Figure 3.14: *Conformational subspace projections for the observed N-term structures and the normal modes 1,2 and 3 (black, red and green lines). Correlation coefficients for the first three normal modes with the first 11 PC's*

The conformational subspace projections illustrate that the first normal mode of the structure accurately described the differences in conformation between the native and target/inhibitor-bound structures of the distribution for the N-term. The correlation plots shows that both mode 1 and mode 2 are relatively similar in direction to PC1, however the first mode also shares significantly more character of PC2 than mode 2. Mode 3 on the other hand does not describe a conformational displacement which is apparent in the subspace projections. The correlation plots suggests that mode 3 shares significant character with PCs 5 to 7 although this describes little of the observed variation in the dataset as illustrated in **Figure 3.6**.

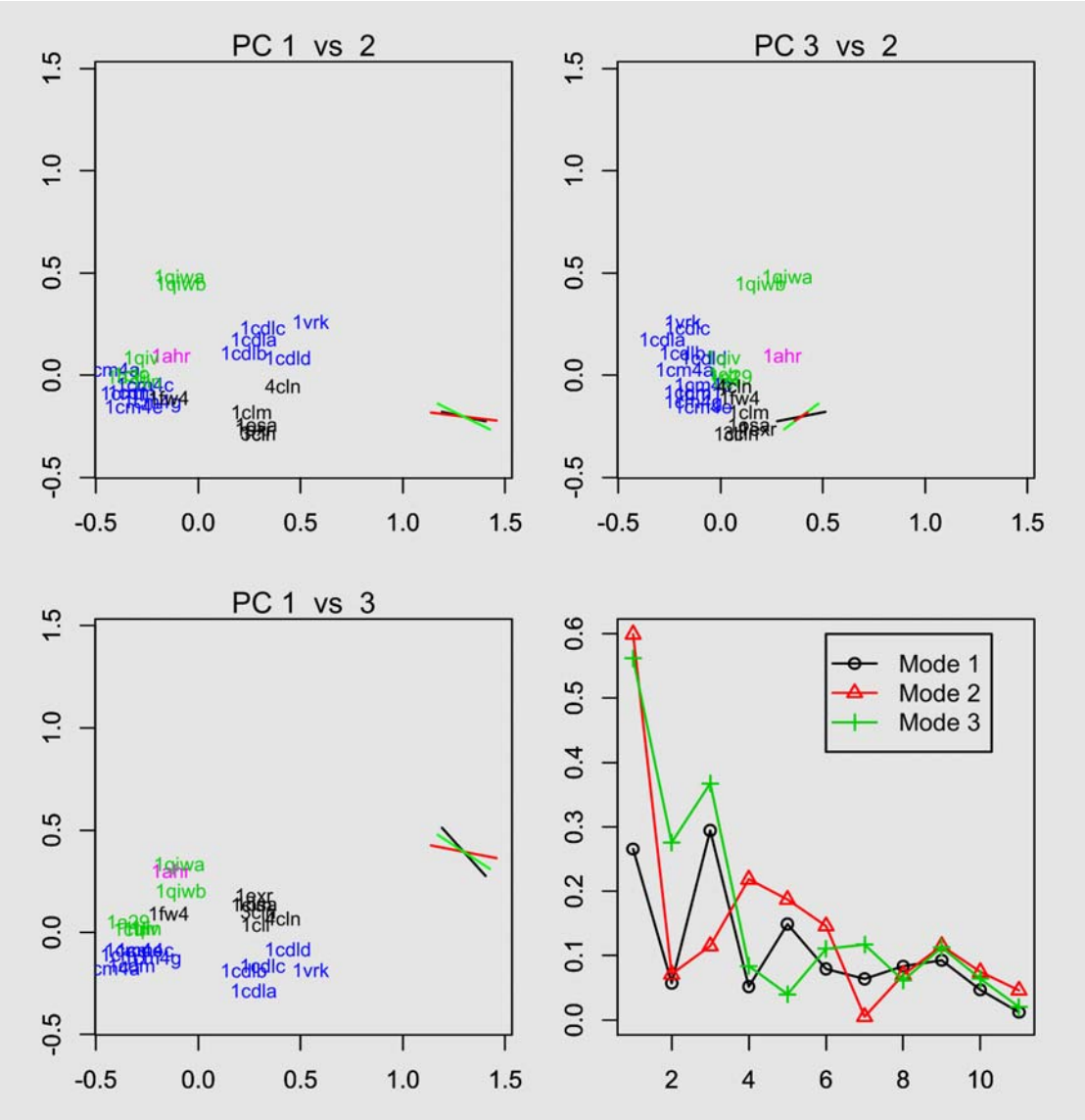


Figure 3.15: *Conformational subspace projections for the observed C-term structures and the normal modes 1,2 and 3 (black, red and green lines). Correlation coefficients for the first three normal modes with the first 11 PC's*

The comparison of the PCs and modes 1 to 3 in the correlation plot suggest that the lowest frequency modes and the first few PCs both describe the same subspace. In this analysis, mode 1 correlates reasonably well with both PC1 and PC3 but not with PC2. Modes 2 and 3 both correlate well with PC1 and mode 3 also correlates reasonably well with PCs 2 and 3. There is no clear separation between the native and target-bound states as in the N-term analysis and consequently no particular mode can be said to describe the same conformational flexibility.

Of the modes 1 to 3, mode 3 shows the greatest similarity to PC2 along which the native structures are separated from the **1cdl** / **1vrk** cluster of structures. However the poor representation in the subspace projections suggests that this distribution may be better described by normal modes higher up in the spectrum. As outlined in the eigen spectrum for the analysis of the C-term, the vectors describing the degrees of freedom in the structure are less specific than in the N-term and the flexibility appears to be more limited and less specific.

3.3.4.1 Structural Core Superpositioning

The Corefind program [Gerstein & Altman 1995] was used to determine the conserved regions of the structures used in the analysis of the N-term lobe structures. A cutoff value of 0.02Å produced a structural core including 29 of the 69 Cα atoms, the location of which are shown in the figure below. The PCA was performed in the usual way with only the core atoms used to perform the superposition. Calculation of the covariance matrix and the subsequent eigen analysis were then performed in the usual way.

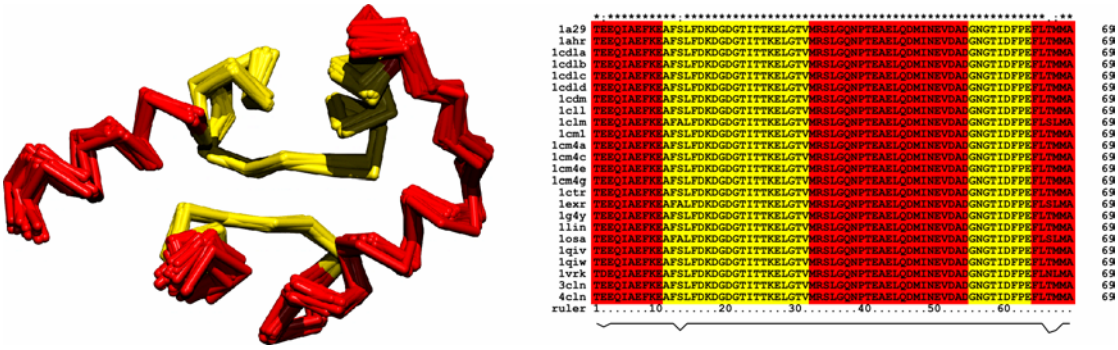


Figure 3.16: *Liquorice representation of N-term crystal CaM structures coloured to highlight regions of core structure. The same regions are also highlighted in the amino acid sequence alignment.*

The results of the PCA after the structures have been superposed on the core atoms only (**Figure 3.17**), are very similar to those corresponding to the all atom superposition. Projections of the structures onto the PC's indicates the same differentiation between structures of different classes seen for the all atom superposition. Since these structures are very similar to one another, it seems that there is little to be gained through the use of this reference frame, when it has been argued that an all atom frame of reference is most

appropriate in the context of the computation of concerted motions in proteins [Karplus & Ichiye 1996].

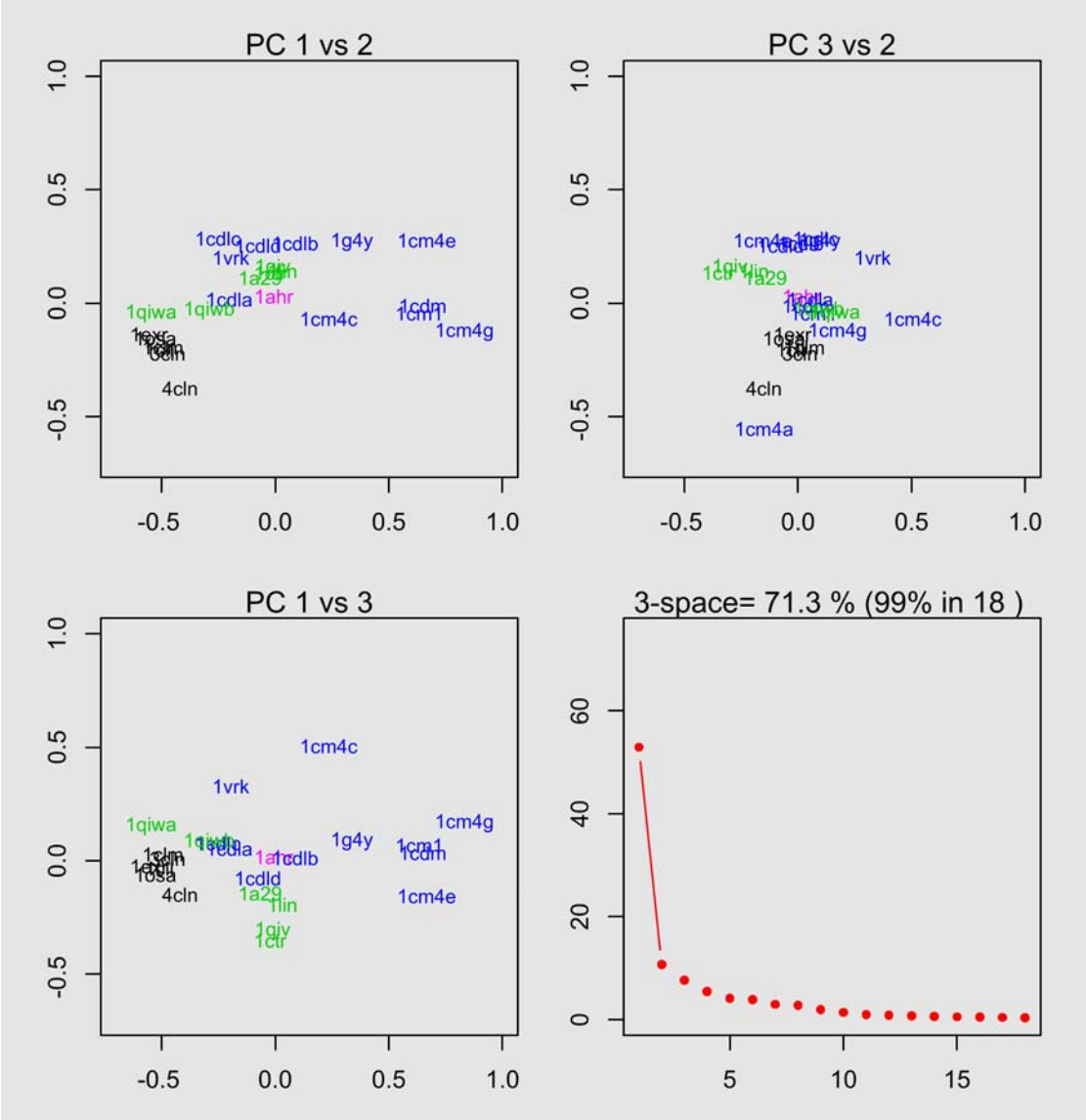


Figure 3.17: *Conformational subspace projections of PC's 1 to 3 and the associated eigen-spectrum for the N-term structures superposed in the core atoms only.*

3.4 Conclusions

The Principal Components Analysis of these structures has highlighted subtle conformational changes that take place as a result of the interaction between either the C or N-term lobe of Calmodulin and different targets. Comparison of the different crystal structures for the CaM lobes in the unbound state show near identical conformations for proteins from different sources. The introduction of ligands and co-factors however has been shown to induce alternate conformations. The distinction between protein complexes involving small molecule inhibitors and target peptide fragments, suggests that whilst certain conformational adjustments are required by the inhibitors, the target peptides are able to induce more dramatic changes in the overall conformations of both the N and C-terminal lobes of CaM.

Comparison of the two different CaM domains indicate that these conformational changes are more apparent in the N term than for the C term lobe which suggests a higher degree of flexibility, potentially linked to the lower affinity of the N term lobe for Ca^{2+} [Barth *et al.* 1998]. Plots showing the contributions of the residues to the structural variation are shown below for the first 5 PC's of each lobe. Whilst the general trend is for the higher contributions to come from the more flexible loop and terminal regions, it is interesting to note that different regions contribute to different PC's as demonstrated for the C-term in **Figure 3.10**.

Comparison of the PCA results with the orthogonal vector space representation generated by the NMA results highlighted some interesting observations. The high correlation between the first few PC's and the lowest frequency normal modes suggest that the conformational sampling achieved through the simulations is similar to the observed variability in the structural ensemble. Projections of the normal mode displacement vectors onto the principal component subspace seems to indicate that the process of lobe opening identified as a consequence of interaction of CaM with its targets is also proposed by the simulations of the un-bound protein. Thermodynamic calculations based on these normal mode calculations provide further insight on the individual characteristics of the N and C-term lobes of the protein and are investigated in detail in Chapter 6.

Chapter 4: Molecular Dynamics Force Field Comparisons

4.1 Introduction

Proteins are complex systems [Caves *et al.* 1998]. Whilst protein structures from NMR and XRD experiments can provide valuable insights into the structure function relationships of proteins, it is important to understand that they are also dynamical systems. Consequently the biological activity of such systems can only be understood through the study of the structure, function and *dynamics* of proteins [Brooks *et al.* 1988].

Modern technological developments have resulted in experimental techniques to monitor biochemical processes on increasingly shorter timescales [Weiss 1999]. The latest spectroscopic techniques are able to monitor some processes taking place on a fs timescale. The most interesting biochemical techniques such as conformational exchanges and ligand binding events however more commonly take place on a μs timescale. Biomolecular modellers have at the same time been able to extend towards a μs timescale [Duan & Kollman 1998]. Developments in methodologies and enhancements in computer performance have made it possible to routinely perform all-atom molecular mechanics simulations on solvated systems over several nanoseconds. The increasing accessibility of physiologically relevant timescales for protein simulations maximises their potential for the prediction and characterisation of biophysical phenomenon.

Molecular dynamics simulations traverse the potential energy surface and sample the conformational space accessible to a protein system under specific conditions. Optimising the conditions for such studies is essential for the fidelity of the simulation and can only be achieved by experimentation with the different parameters and environmental representations which are available. There have been countless developments in molecular mechanics potentials and numerous different models have been constructed to simulate protein environments using the minimum of computational resources. In this study, some of these different techniques will be addressed in order to

determine the optimum approach for the simulation solution molecular dynamics for CaM on a nanosecond timescale using available resources.

4.1.1 Previous Studies of CaM dynamics

As far back as 1991, Mehler and co-workers were experimenting with molecular dynamics studies of CaM and Troponin C in order to ascertain whether CaM exhibited a more globular fold in solution than the extended conformation represented in X-ray diffraction experiments [Mehler *et al.* 1991]. Their simulations indicated that use of a fixed dielectric for the solvent yielded unsatisfactory results. The incorporation of waters with positions determined in the crystal structure, and a distance dependent dielectric (detailed in Chapter 2) however resulted in more realistic simulations. Studies using their optimised conditions suggested a compaction of the dumbbell like structure of holo-CaM facilitated through flexibility in central helix.

A study by Yasri *et al.* [Yasri *et al.* 1996] used an auto-correlation analysis of inter-atomic distances during molecular dynamics simulations to rationalise the choice of appropriate simulation parameters for CaM. Their experiments with implicit solvent models suggested that the best results were obtained though the use of long heating periods (in excess of 10ps) and a strong coupling to the temperature bath (0.1 ps).

One of the key challenges of this study is ensuring that the potential used can correctly accommodate the electrostatic effects of the Ca^{2+} ions on the system. Molecular mechanics potentials are highly optimised for either protein or DNA systems and whilst they can adequately accommodate commonly occurring metal ions such as Ca^{2+} and Fe^{3+} ions, it is necessary to pay careful attention to the behaviour of metal containing systems. A study by Allouche and co-workers in 1999 investigated the effects of *Alchemically* changing the Mg^{2+} ions in Parvalbumin to Ca^{2+} in a Free Energy Perturbation technique [Allouche *et al.* 1999]. In these studies the bi-dentate glutamate sidechains were found to convert to mono-dentate interactions yielding an octahedral Ca^{2+} coordination which is generally unfavoured for the ion. Subsequent work yielded accurate values for interaction energies between the ions and water clusters but attempts to incorporate these values

into a protein potential for Parvalbumin generated unsatisfactory coordination behaviour in the Ca^{2+} binding sites [Sanejouand 2001].

In 1998, Roux & Marchand investigated the molecular dynamics of Calbindin D_{9K} under different Ca^{2+}-states (apo, singly and doubly loaded) [Marchand & Roux 1998]. Using the CHARMM forcefield and a fully solvated extended electrostatic model they used the all atom PARAM22 [MacKerell *et al.* 1998] potential but used specially derived many body non-bonded parameters for the Ca^{2+} atoms. The electrostatic parameters were optimised to reproduce the experimental free energy of hydration of Ca^{2+} (-381 kcal mol^{-1}) using free energy perturbation experiments, initially with 25 but later 50 and 100 explicit water molecules. The resulting parameters in the empirical potential were then compared favourably to corresponding results for *ab initio* calculations of the interaction energy between Ca^{2+} and a single water molecule. They attributed previous problems with Ca^{2+}-containing systems in such simulations to use of the polar hydrogen potential and cutoff distances which were too short to include the interactions between adjacent Ca^{2+} (some 13Å apart in CaM) sites in EF-hand proteins.

These previous studies emphasise the importance of the identification of an appropriate model for the system, particularly with respect to the treatment of the Ca^{2+} ions in the system. The high charge density and coordination requirements of Ca^{2+} ions present considerable challenges for molecular mechanics potentials and require careful attention to detail in order to obtain meaningful results.

4.1.2 Protein Solvent Models

Appropriate treatment of protein environments in molecular mechanics simulations invariably presents many problems. Increasing the detail of solvent models inherently increases the complexity of the calculations and thus the tractability of the system. The use of an insufficiently detailed solvent model however will generate inaccurate results as indicated in previous studies of CaM [Mehler *et al.* 1991]. The simplest representation of protein environments is provided by dielectric functions which can be fixed or more commonly distance dependent functions.

Various techniques have been developed to introduce *implicit* solvent into molecular mechanics potentials resulting in a more complex model which is computationally more expensive but still cheaper then using explicit solvent. The Effective Energy Function (EEF1) developed by Lazaridis *et al.* is one such technique which has been implemented into CHARMM 27 [Lazaridis & Karplus 1999]. This model is based on the assumption that the Hamiltonian for the potential energy function of the system can be separated into solvent-solvent, solvent-solute and solute-solute energy terms. Thus, "*the solvation free energy of a protein molecule is a sum of group contributions which are determined from values for small molecule compounds"*. According to this assumption, a solvent exclusion model can be used to calculate the solvation free energy of the system which can then be combined with the CHARMM polar hydrogen energy function [Neria *et al.* 1996]. Initial results were found to be highly comparable for those performed using explicit solvent molecules although it was noted that polar groups contained within α-helices of proteins were found to exhibit anomalous behaviour.

Only through the incorporation of explicit solvent molecules in a simulation can the effects of solvent correctly be accounted for, although this results in a considerable increase in the computational requirements and puts further constraints on the simulation. The CHARMM forcefield uses the simple TIP3 water model [Jorgensen, 1983]. This uses a rigid water molecule with a total of 3 electrostatic interaction sites with the partial positive charges on the H atoms balanced by a negative charge on the O.

Various different techniques can be used to solvate the protein fragment under investigation. In the case of periodic boundary simulations [Leach 1996], the protein fragment is usually immersed in an appropriately sized box of pre-equilibrated water molecules. Waters found within the space occupied by the protein fragment are removed. The IMAGE function of CHARMM [Community 2001] can be used to generate the repeating unit of the system in all directions such that a water molecule migrating out of one unit is simultaneously incorporated into the calculations through an adjacent unit. The *SOLVATE* package [Grubmueller 1996] can be used to generate a solvent shell around a protein fragment either by generating a shell of a given radius or a minimum thickness around the solute. This process also facilitates the incorporation of counter ions to neutralise the charge on a protein and include salt ions to a physiological concentration.

The most comprehensive simulations currently performed use the Periodic Boundary Conditions in conjunction with explicit counter ions and the use of the Ewald summation for the long range electrostatics. In this treatment, the interactions between any particle in the unit cell under consideration and all the other particles in that cell and all the other images of the unit cell in an infinite array are considered. Conversion of this summation into a series results in a function which converges rapidly towards some cutoff [Leach 1996]. This approach is highly intensive and can only be used for small systems and for relatively short simulations without the use of extensive computational resources.

4.1.3 Considerations in this study

This study will focus on various different solvent treatments and parameter sets in order to uncover the most appropriate experimental conditions for the simulation of CaM molecular dynamics.

- **Solvent treatment**

Vacuum simulations of the protein fragments will be compared with results using the EEF1 implicit solvation model [Lazaridis & Karplus 1999], and explicit solvent simulations of three different types: Firstly all waters molecules included in the crystal structure will be considered. These will then be supplemented using the program *SOLVATE* [Grubmueller 1996] to a minimum thickness of 5.0Å around the protein fragments. Thirdly a solvent shell of minimum thickness 5.0Å generated entirely using *Solvate* will be used.

- **Parameterisation**

In the light of the work by Roux & Marchand described above, the results of simulations using the extended atom parameter set PARAM19, were compared with results using the all atom PARAM27 potential. Non-bonded cutoff distances were ensured to be sufficient to incorporate any interaction between the adjacent Ca^{2+} sites.

- **Initial Conditions**

Recent MD studies of Crambin [Caves *et al.* 1998] indicated that multiple short simulations provided better sampling of the potential energy surface than single extended trajectories. With so many different variables to be explored, there is no possibility of performing numerous trajectories on a meaningful timescale for this work, however each experiment will be performed using two different initial velocity assignments as test of reproducibility. A full summary of the different simulations performed in this study is shown in **Table 4.1** below.

4.2 Materials & Methods

4.2.1 Structure Selection & Preparation

Simulations were performed on the 1.0Å crystal structure **1exr** [Wilson & Brunger 2000]. The highly flexible tether connecting the two CaM domains was removed from the system resulting in two fragments corresponding to E2-A73 and L85-S147 for the N and C-term lobes of the protein respectively. For residues with multiple conformations as determined experimentally, the conformation corresponding to dis-order locator **A** was used in all cases. The N-termini of the two fragments were patched with N-Acetyl groups to remove the induced charges from truncation of the sequences shown in **Figure 4.1** below. The C-termini of the fragments were patched with an acetyl (ACET) patch. Coordinates for the patches were built from standard internal coordinates and hydrogens were added using 5 iterations of the HBUILD function. In the case of studies with the PARAM19 parameter set, only polar hydrogens were added, and for the PARAM27 simulations all hydrogen atoms were considered explicitly.

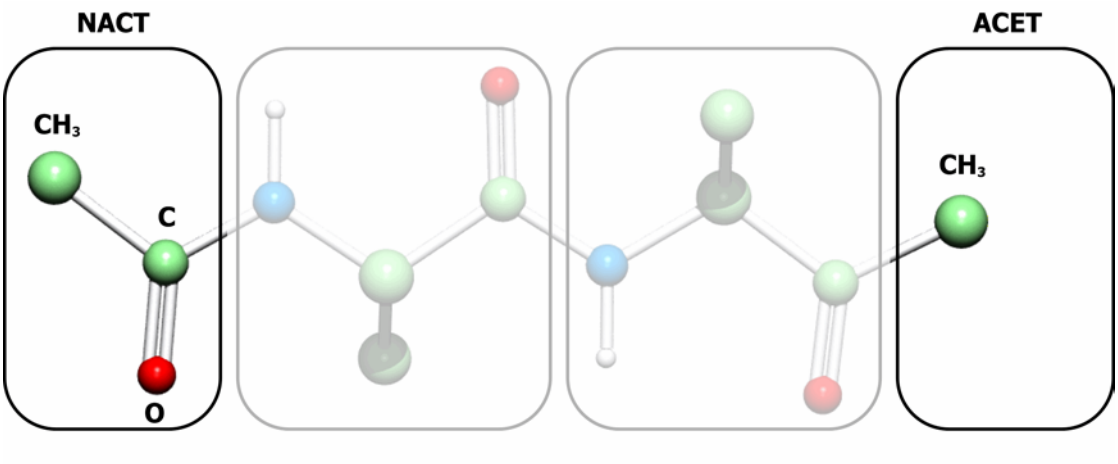


Figure 4.1: *Diagram showing the atoms added as part of the patches on the termini of the protein fragments used in the simulations. The pale boxes in the centre represent standard amino acid repeating units with the NACT patch shown at the N-terminus of the peptide chain and the ACET at the C-terminus. Non-polar hydrogens are not shown here.*

4.2.2 Protein Solvation

Protein fragments were solvated using the *SOLVATE* package of Grubmüller *et al.* [Grubmueller 1996]: The protein fragment to be solvated was passed to *SOLVATE* which generated a shell of TIP3 [Jorgensen *et al.* 1983] waters to a minimum shell depth of 5.0Å around the fragment. Inspection of the resulting distribution of water molecules indicated that the key solvent molecule used to co-ordinate the Ca^{2+} ions had been appropriately placed by the algorithm.

4.2.3 Experimental Details

4.2.3.1 Non-bonded parameters

The non-bonded parameters were specific for the different types of simulation and the details for each experiment are outlined below.

- ***In Vacuo* simulations:**

A distance dependent dielectic ($\varepsilon_r = r_{ij}$) was used for these simulations with the corresponding parameters recommended in the CHARMM documentation [Community

2001]. The electrostatic potential was shifted to a zero cutoff at 9.0Å and Van der Waals interaction were smoothed to zero at the cutoff with the use of a switching function between 8.0 and 9.0Å.

- **Effective Energy Function (EEF1) simulations:**

The non-bonded parameters recommended for this potential differ from the vacuum set in the use of the group electrostatic treatment [Community 2001].

- **Explicitly solvated simulations:**

In the case of simulations using explicit solvent, a constant dielectric (ε=1.0) is used. In accordance with the recommendations in the documentation [Community 2001], the electrostatic forces were shifted to zero at 13.0Å and the Van der Waals interactions were treated with a switching function between 12.0 and 13.0Å for a smooth cutoff. This is a sufficiently large distance to incorporate any interaction between the adjacent Ca^{2+}-binding sites in the CaM fragments in accordance with the recommendations of Marchand *et al.* [Marchand & Roux 1998].

In addition to the incorporation of the non-polar hydrogen atoms, the use of the PARAM27 parameter set included the updated Ca^{2+} parameters also calculated and validated by Marchand *et al.* [Marchand & Roux 1998].

Table 4.1: *Summary of experimental details for the simulations performed. Six different types of simulation were performed on both the N and C-term lobes of CaM. Each simulation was repeated using a different assignment of initial velocities. CPU time is recorded in hours according to the internal CHARMM log.*

Identifier	*Description*	*Parameter set*	*Dielectric*	*No. atoms*	*CPU Time*
ntvac & ntvac1	*In vacuo* simulations	TOPH19 & PARAM19	Radial ($\epsilon=r_{ij}$)	667	4.41
ctvac & ctvac1				607	6.92
nteef & nteef1	*In vacuo* simulations using the EEF1 implicit solvation potential.	Special TOPH19 & PARAM19	Radial (Group)	668	7.88
cteef & cteef1				608	6.21
ntxtal & ntxtal1	Protein surrounded by crystal waters	TOPH19 & PARAM19	Fixed (ϵ=1.0)	905	19.48
ctxtal & ctxtal1				890	17.97
ntxtalsolv & ntxtalsolv1	Protein and crystal waters solvated to 5Å	TOPH19 & PARAM19	Fixed (ϵ=1.0)	5573	168.71
ctxtalsolv & ctxtalsolv1				5585	171.16
ntsolv & ntsolv1	Protein fragment solvated to 5Å	TOPH19 & PARAM19	Fixed (ϵ=1.0)	4694	203.54
ctsolv & ctsolv1				4214	176.87
ntpar27 & ntpar271	Protein fragment solvated to 5Å	TOPH27 & PARAM27	Fixed (ϵ=1.0)	5516	261.21
ctpar27 & ctpar271				4595	231.94

4.2.3.2 Minimisation

The standard minimisation protocol outlined in Chapter 2 was applied to remove any bad interactions in the system generated by the building of patches and the hydrogen atoms and to remove any structural defects caused by crystal packing or other effects. The system was minimised to a final tolerance gradient of 10^{-5} kcal mol^{-1} $Å^{-1}$.

4.2.3.3 MD Heating Phase

The structures were then heated to 300K by performing 10ps of dynamics on the system at 25K increments. The use of a tight temperature coupling to the water bath [Berendsen *et al.* 1984] was applied at this stage to ensure that the energy was fully transferred to the system.

4.2.3.4 MD Production Phase

The molecular dynamics of the different systems were simulated over a 1ns period using steps of 1fs and coordinates were saved every ps. A relatively tight temperature coupling constant of 0.5 was used for all the simulations as recommended by Yasri et al. [Yasri *et al.* 1996].

4.2.4 Computational Details

All these simulations were carried out on Compaq Alpha machines. Repeat simulations were performed using identical protocols but using different random number seeds for the assignment of the initial velocities.

The vacuum simulations using a radial dielectric were the fastest simulations requiring around 5 hours of CPU time to perform the minimisation and 1ns of molecular dynamics. The potential incorporating an implicit solvation model (**EEF1**) resulted in a doubling of the CPU time. The extra atoms included in the solvated systems resulted in a considerable increase in the computational requirements which was also affected by the much longer cutoffs used with the continuous dielectric function. The additional expense incurred by the use of the PARAM27 all atom potential was relatively limited due to the high proportion of solvent in the system which already include polar hydrogens.

4.2.5 Analysis Techniques

The different techniques used to analyse the trajectories are outlined in Chapter 2. Molecular visualisation of the trajectories was performed using the QUANTA package [Accelrys 2001] and data manipulation and analysis was performed using the R statistical package [Gentleman & Ihaka 2000].

4.3 Results & Analysis

The aim of these experiments was to determine the most appropriate conditions for molecular dynamics simulations of CaM. On this basis, the results and analysis presented

here will concentrate on general characteristics of the simulations that have been performed rather than a detailed characterisation of the trajectories. The criterion used to assess these characteristics are as follows:

- **RMSD**

Root mean square deviation information relative to a reference structure is a conventional means to assess the overall behaviour of a simulation. In this analysis the crystal structure used as the template for all these simulations will provide the central frame of reference about which the comparisons will be made. In order to minimise the computational requirements of the analysis, only Cα coordinates will be considered which provide sufficient detail to monitor significant conformational fluctuations.

- **RGYR**

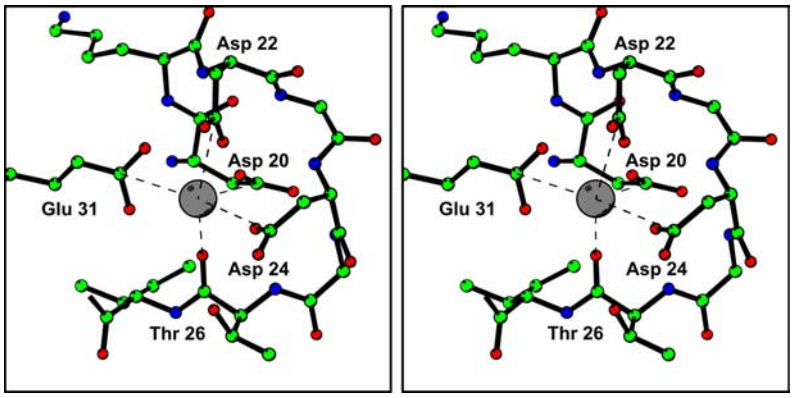
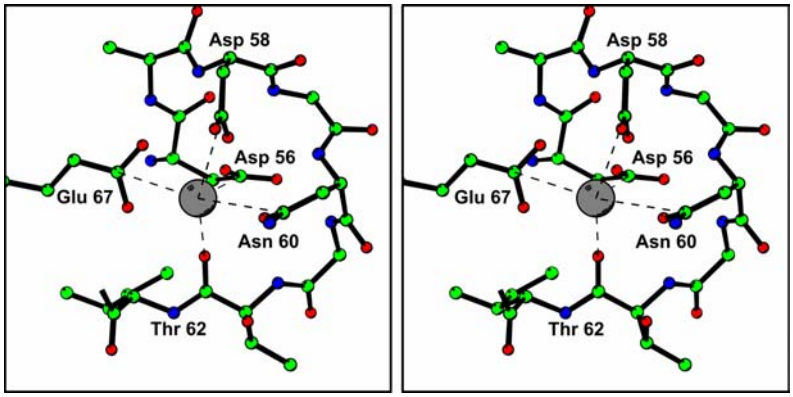
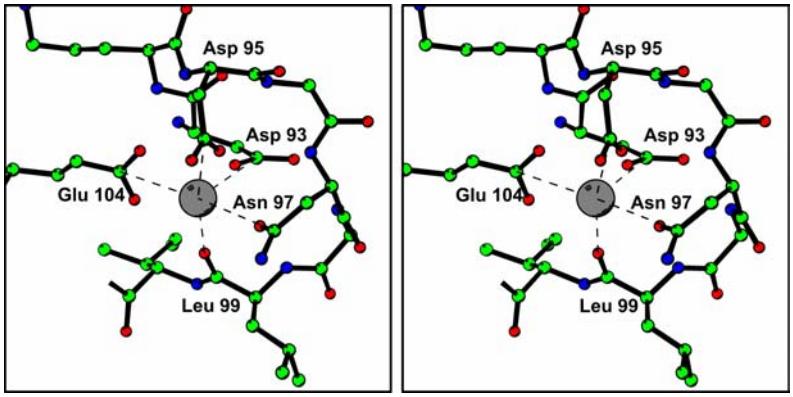
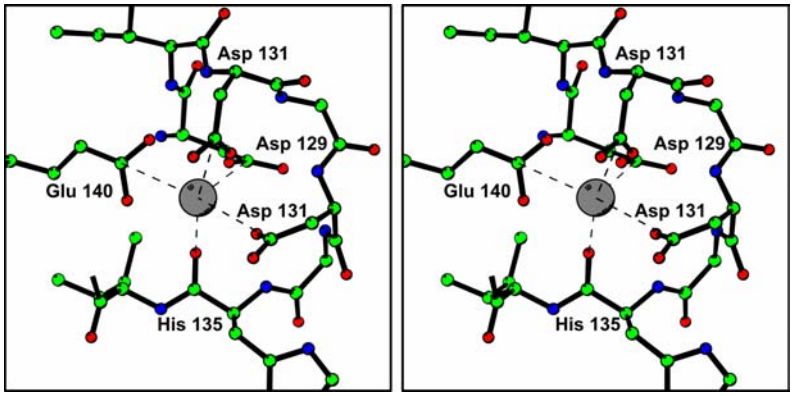
The Radius of Gyration (RGYR) of the protein fragments yields slightly more detail about the significance of conformational rearrangements occurring in the system. In the specific case of CaM two different conformations have been identified as the open conformation observed for Ca^{2+}-bound CaM as in the structure **1exr** and the Ca^{2+} free conformation seen in **1dmo**. There are associated marked differences in the Cα RGYR values of these structures with **1exr** reporting values of 11.9 and 10.6Å for the N and C-term lobes respectively and the Ca^{2+}-free structure, **1dmo,** 11.1 and 10.4Å. The N-term in particular suggests marked differences for the two conformations and this parameter will provide useful information on the extent to which the protein is adapting over the course of the simulation. The calculated RGYR values determined for the conformations explored in the simulations are compared with the mean and standard deviation of the RGYR values calculated for the set of crystal structures used in the analysis in Chapter 3 for the N and C-term structures respectively.

- **Ca^{2+}-binding behaviour**

The key to the function CaM is its Ca^{2+}-binding response and as such it is appropriate that an assessment of these experiments should incorporate some characterisation of the Ca^{2+}-binding behaviour. This will be performed by monitoring the distances between reference atoms in the Ca^{2+}-coordination protein sidechains in the four EF-hands of the protein. **Table 4.2** below includes stereo figures of the four Ca^{2+} binding sites in **1exr** and shows the distances between the reference atoms and the Ca^{2+} ion and reports the

distances observed in the crystal structure. These will be monitored for the different simulations and the resulting statistics reported.

Table 4.2: *Ca^{2+} coordination characteristics of the four Ca^{2+} sites in 1exr*

Ca^{2+} Coordination Geometry	*Distances*
	**EF I** **1:** Cγ Asp20-Ca **3.405Å** **2:** Cγ Asp22-Ca **3.525Å** **3:** Cγ Asp24-Ca **3.337Å** **4:** O Thr26 -Ca **2.355Å** **5:** Cδ Glu31-Ca **2.820Å**
	**EF II** **6:** Cγ Asp56-Ca **3.430Å** **7:** Cγ Asp58-Ca **3.421Å** **8:** Cγ Asn60-Ca **3.322Å** **9:** O Thr62 -Ca **2.407Å** **10:** Cδ Glu67 -Ca **2.883Å**
	**EF III** **1:** Cγ Asp93-Ca **3.455Å** **2:** Cγ Asp95-Ca **3.211Å** **3:** Cγ Asn97-Ca **3.293Å** **4:** O Leu99 -Ca **2.354Å** **5:** Cδ Glu104-Ca **2.871Å**
	**EF IV** **6:** Cγ Asp129-Ca **3.479Å** **7:** Cγ Asp131-Ca **3.288Å** **8:** Cγ Asp133-Ca **3.415Å** **9:** O His135 -Ca **2.284Å** **10:** Cδ Glu140-Ca **2.839Å**

4.3.1 RMSD Analysis

- **N-term simulations:**

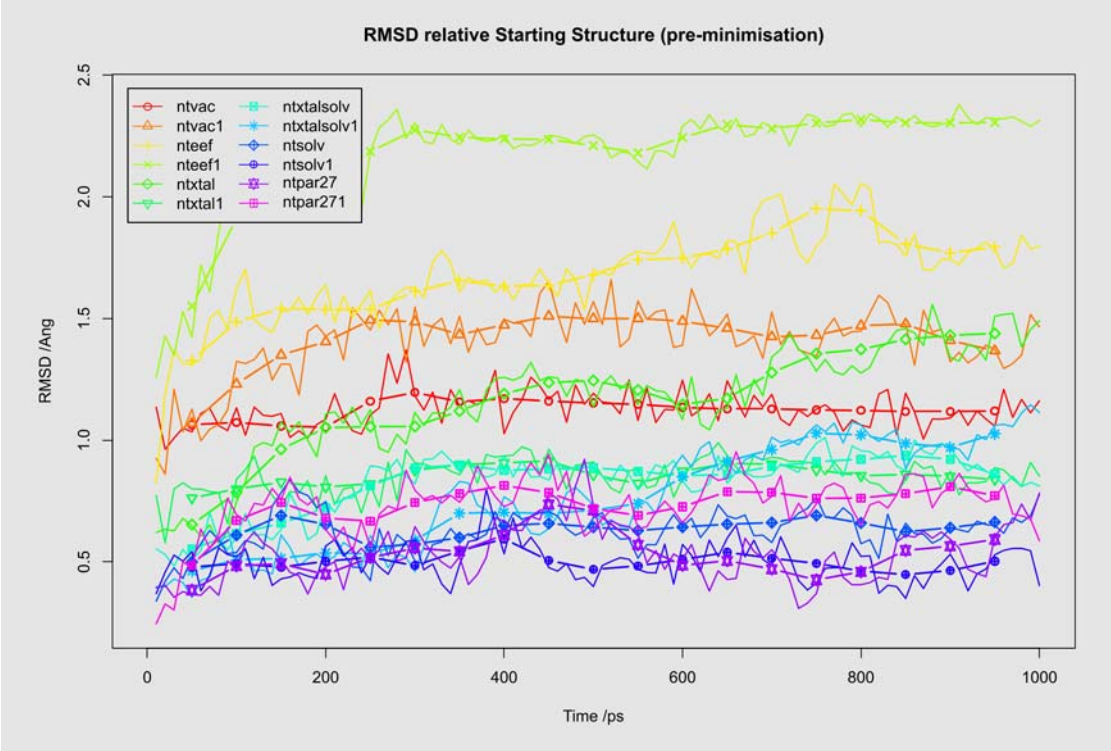


Figure 4.2: *RMSD time series for the N-term CaM simulations. Continuous lines indicate instantaneous values at 10ps increments, connected points indicate average values over subsequent 100ps periods.*

This summary description of the 12 different simulations clearly highlights the different characteristics of the trajectories calculated (**Figure 4.2**). The legend lists the simulations in order of increasing complexity (or increasing number of degrees of freedom) and thus computational demand. From the beginning of the production phase, the simulations performed using the EEF1 potential show marked deviation from the crystal structure, and all the other simulations. The second EEF1 trajectory, **nteef1**, appears to exhibit two different stable conformations however in the period 300-600 and 600-1000 ps, for which actual fluctuations about the conformation are limited. **nteef** however seems much less stable and the floating average values show no real continuity at any point in the simulation.

The vacuum simulations **ntvac** and **ntvac1** show less erratic behaviour although deviation from the crystal structure after minimisation and heating is still significant. **ntvac** in particular appears to establish a conformation with an RMSD of 1.1 Å after some 300ps of dynamics which is largely retained for the remainder of the trajectory. The perceived stability of these simulations however is in marked contrast to the inconsistent behaviour observed using the EEF1 potential.

The addition of the waters from the crystal structure to the system helps to stabilise the system although the two simulations appears to be split between the more variable solvent-free trajectories and those performed with a fully solvated water shell. The first trajectory, **ntxtal**, does not seem to converge towards any particular conformation over the time period and continues to diverge beyond an RMSD of 1.5Å from the crystal structure while the second, **ntxtal1**, appears to converge rapidly within the first 200ps to 0.8Å.

Solvating the system to the full 5.0Å shows a significant effect on the simulations, particularly during the minimisation and equilibration phases of the simulations. All 6 of the fully solvated experiments are almost within 0.5Å of the crystal structure after completion of the heating dynamics. Both **ntxtalsolv** and **ntxtalsolv1** however exhibit trajectories which take them away from the original conformation over the 1ns production phase of the dynamics. **ntxtalsolv** equilibrates during the first 300ps and then seems to stabilise at 0.75Å, while **ntxtalsolv1** exhibits conformational changes later on in the trajectory which are most apparent between 500 and 700ps and apparently continue beyond the end of the 1ns time period.

The two remaining types of experiment show little in the way of difference between them. The **ntpar271** trajectory appears to equilibrate rapidly over the first 200ps although the variation in floating averages suggest that this probably not a single stable conformation. The **ntpar27** trajectory on the other hand is relatively stable over the first 400ps and after a deviation around 500ps, comes back to within 0.5Å of the crystal structure for much of the final 400ps. The PARAM19 trajectory, **ntsolv**, shows only very slight fluctuations in the floating average values which are almost constant under 0.5Å after 400ps of the production phase. The second trajectory, **ntsolv1**, however largely mirrors the behaviour of **ntpar27** with a stable phase at the beginning of the simulation,

a deformation around 400ps and then a return towards the crystal conformation which is conserved throughout the second half of the trajectory.

- **C-term simulations:**

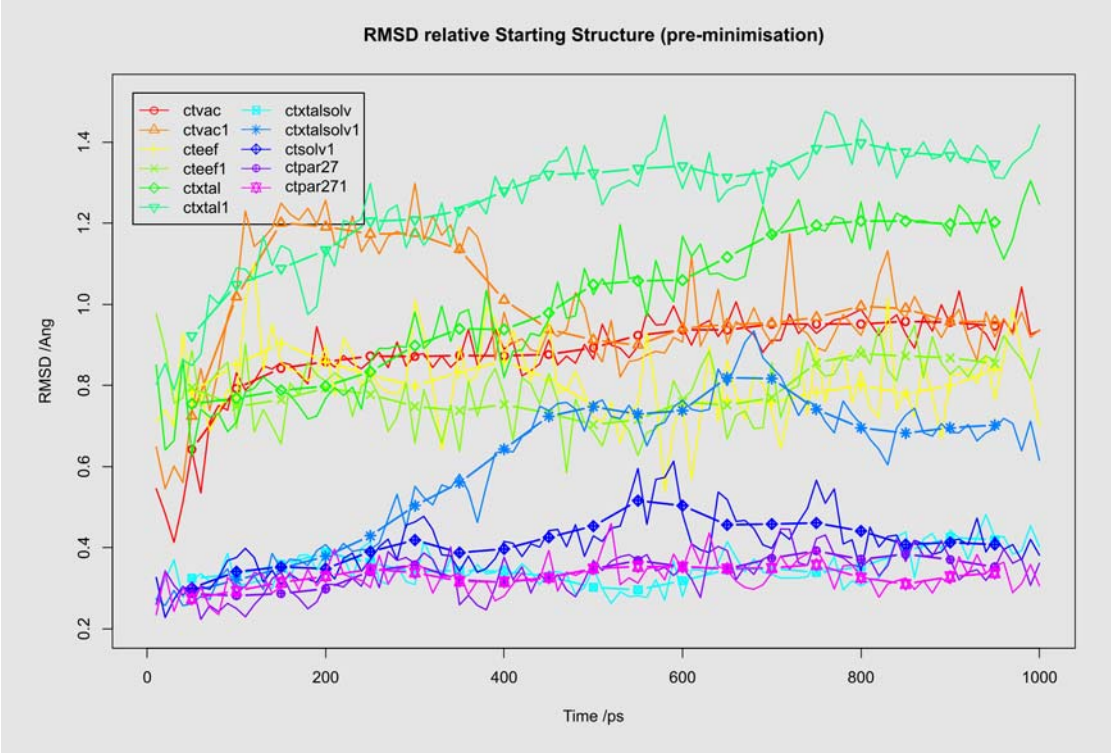


Figure 4.3: *RMSD time series for the C-term CaM simulations. Continuous lines indicate instantaneous values at 10ps increments, connected points indicate average values over subsequent 100ps periods.*

These RMSD time series in **Figure 4.3** show a clear distinction between those with and without a fully solvated 5.0Å shell of explicit water. In these experiments it is the trajectories solvated with crystal waters alone which show the most significant deviations from the original crystal structure, **ctxtal** and **ctxtal1**. Neither of these trajectories seem to discover any particularly stable conformations and the floating average values for RMSD relative to the crystal structure continue increasing throughout the 1ns simulation time.

Both the vacuum simulations show similar values after the heating dynamics phase although the second trajectory, **ctvac1**, exhibits significant conformational changes over the first 150ps of the production phase. **ctvac** equilibrates over the same timescale to a

different conformation 0.9Å from the crystal structure which appears to be largely stable throughout the rest of the trajectory. Interestingly, 500ps into the production phase of the simulation, **ctvac1** seems to return towards the crystal structure and fluctuates around 0.9Å away from the crystal structure, a value similar to **ctvac**.

These results of the C-term simulations seem to indicate the EEF1 potential is rather better behaved than suggested by the simulations of the N-term lobe with the conformations explored by **cteef** and **cteef1** both staying within 1.0Å of the crystal structure. This represents a considerable improvement in comparison to the 1.5 - 2.0Å differences observed in the N-term simulations. Once again there is a clear split between those structures that have been notably altered during the minimisation and heating phases and those which have not, and clearly simulations using the EEF1 potential fall in the former category.

All 5 of the simulations with a full 5.0Å solvation shell deviate to the same extent after minimisation and heating, and throughout the first 200ps of equilibration during the production phase. As with the simulations of the N-term lobe, there is a clear split between the two trajectories in which the solvation shell was built around the crystal waters. **ctxtalsolv1** gradually progresses up to 0.8Å away from the crystal structure over the first half of the trajectory while **ctxtalsolv** remains within 0.4Å over the full 1ns.

The results of the first fully solvated PARAM19 simulation, **ctsolv**, generated corrupted binary trajectory files rendering the results unreadable and attempts to correct the error were unsuccessful. The second PARAM19 simulation, **ctsolv1**, showed relatively conservative behaviour staying within 0.5Å throughout the full simulation. Variation in the floating averages however suggests that the trajectory did not find any single conformation which was retained for a prolonged period.

The PARAM27 simulations of the C-term lobe, **ctpar27** and **ctpar271** show the most conservation throughout the time period. Not only are the floating averages for the RMSD within 0.4Å of the crystal structure over the full nanosecond for both simulations, but fluctuations about those averages are rather limited.

4.3.2 RGYR Analysis

- **N-term simulations:**

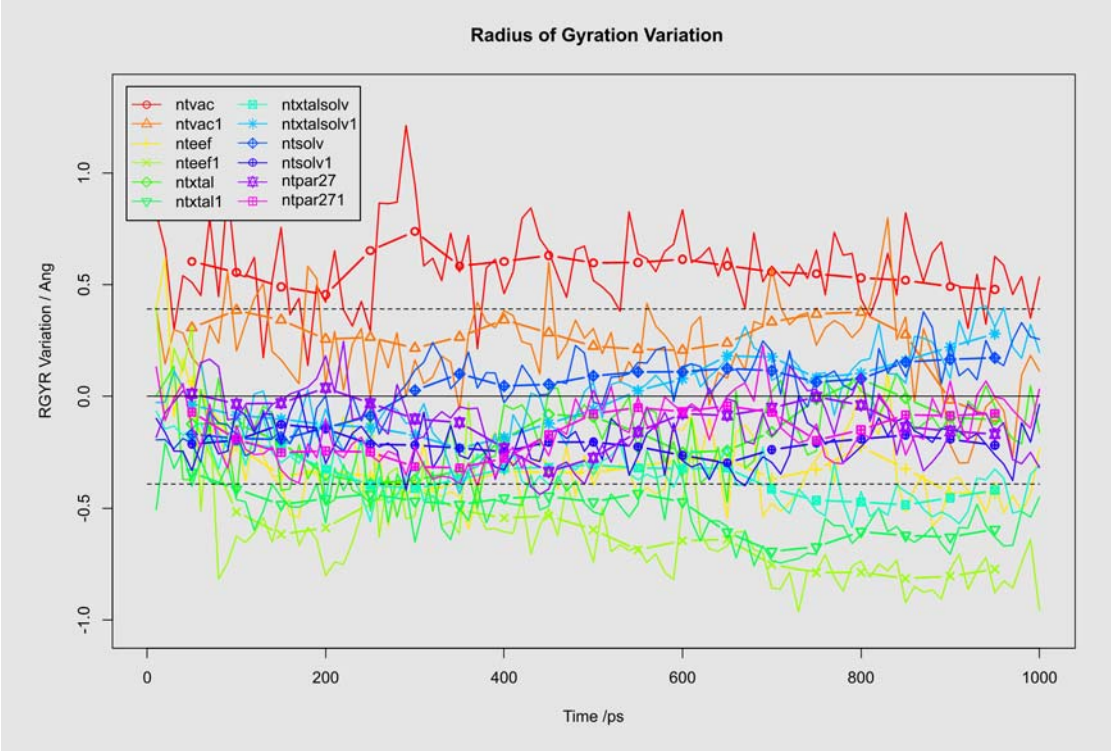


Figure 4.4: *RGYR variation for the 12 different simulations of the N-term lobe of CaM. Deviations about the mean value of the observed dataset are presented with dotted lines marking the upper and lower limit of 5 standard deviations of the observed variation.*

Generally the RGYR variation analysis of the N-term lobe trajectories shown in **Figure 4.4** shows similar trends to the RMSD analysis. Once again there is a clear distinction between the first 6 simulations using minimal solvent representations which show large variations in RGYR and the second 6 fully solvated systems which show more conserved RGYR values over the time period. This analysis does however make slightly clearer distinctions in the nature of the *inaccuracies* observed in the first 6 simulations.

The results clearly illustrate that the vacuum simulations in **ntvac** and **ntvac1** result in larger values of RGYR relative to the crystal structures used to provide a reference frame. The simulations performed using the EEF1 implicit solvent potential however result in lower estimates for RGYR relative to the crystal structures. This suggests that the

vacuum simulations tend towards a more open conformation for the system while the EEF1 and crystal water solutions are more closed than the reference structures. This kind of representation of the results does show that there are no exchanges between significantly different conformations on this timescale. Any significant unfolding events would also be detailed in this analysis by significant and sustained increases in RGYR values, indicative of the formation of extended structures.

The fully solvated simulations largely demonstrate general reductions in RGYR over the time period with values in between the mean and 5 standard deviations of the observed distribution. Only the trajectories **ntsolv** and **ntxtalsolv1** exhibit floating average values of RGYR persistently above the mean value.

- **C-term simulations:**

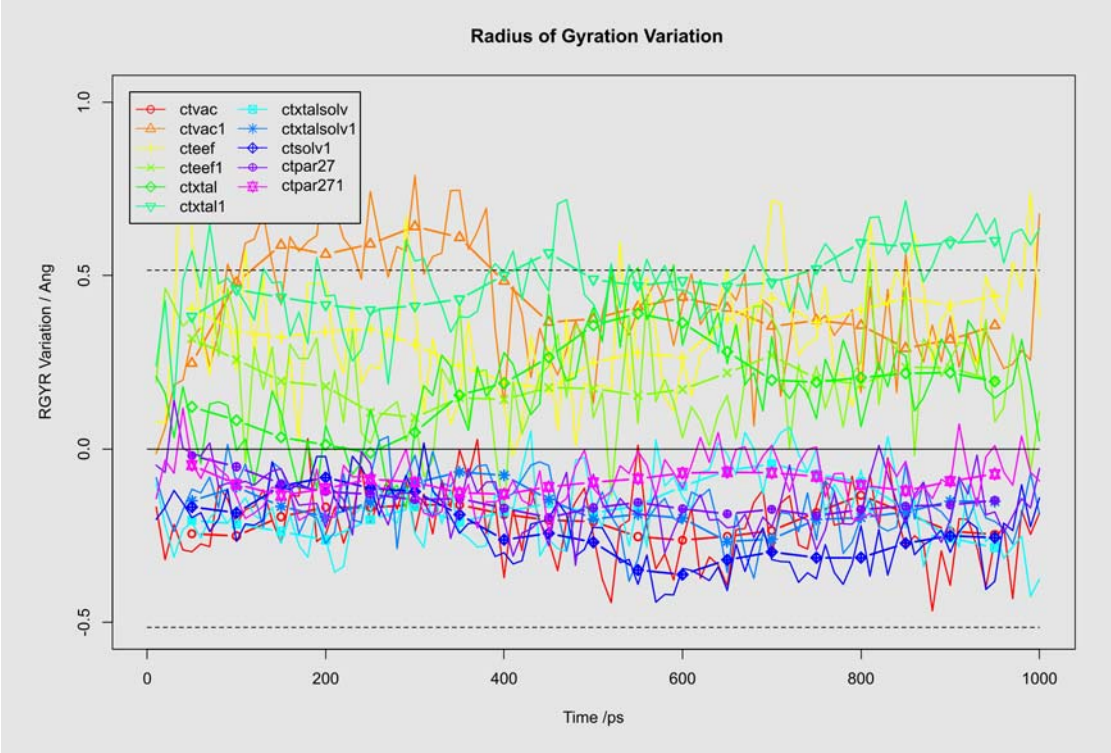


Figure 4.5: *RGYR variation for the 12 different simulations of the C-term lobe of CaM. Deviations about the mean value of the observed dataset are presented with dotted lines marking the upper and lower limit of 5 standard deviations of the observed variation.*

The variation in RGYR illustrated in **Figure 4.5** for the simulations of the C-term lobe of CaM show an even clearer distinction than the results of the N-term. All 5 of the fully solvated simulations are almost entirely contained between the mean and 5 standard deviations below the mean of the observed distribution. With the exception of the vacuum simulation **ctvac**, the experiments with less sophisticated representations exhibit considerable fluctuations in RGYR well above the mean value of the crystal structures.

Interestingly the limits corresponding to 5 multiples of the standard deviation of the observed RGYR values was around 0.4Å for the N-term but is around 0.5Å in the C-term. Although there is a greater variation in RGYR for the observed crystal structures of the C-term lobe, there is considerably less variation in RGYR from the conformations generated by the simulations in the C-term than the N-term.

4.3.3 Calcium Coordination Analysis

The variation in Ca^{2+}-coordination distances for the N and C-term simulations are shown in **Figure 4.6** and **Figure 4.7** below. The details of the reference distances and observed values in the crystal structure **1exr** are shown in **Table 4.2** above, with identities of the reference positions 1 to 10 for the N and C-term lobe shown on the abscissa in the results. The scale on the ordinate represents the mean and standard deviation of the differences between the structures generated by the dynamics and the reference values of the crystal structure in Å. The calculated values for the mean and standard deviation values are also included in a tabular form in the Appendix.

- **N-term simulations:**

In general the simulations show good preservation of the reference Ca^{2+} coordination distances with some notable exceptions. The **nteef** trajectory clearly mishandles the Ca^{2+}-binding in the first EF hand (EF I) also shows relatively poor behaviour through ligands 6 and 10 of the second EF hand (EF II). The second EEF1 trajectory shows better performance although the mean values over the course of the simulation are not particularly close to zero in the first EF hand and ligand 10 is not very accurate in the EFII.

The mean values obtained in the vacuum simulations are not very consistent with those observed in the crystal structure although the standard deviations are very low indicating tight control by the protein. The incorporation of crystal waters into the system results in larger standard deviations throughout, indicating greater fluctuations in the distances. Similar variation is observed in the fully solvated simulations, **ntxtalsolv**, **ntxtalsolv1**, **ntsolv** and **ntsolv1**, all of which exhibit slightly inaccurate mean values and greater standard deviations than the PARAM27 simulations. A time series analysis of the results Ca^{2+}-binding behaviour in **ntpar271** shows that the cause of the large fluctuation observed for ligand 8 in EFII is a temporary phenomenon rather than sustained variation. The Asn60 side chain temporarily adopts a short lived alternate conformation, although this single event is sufficient to affect the statistics significantly.

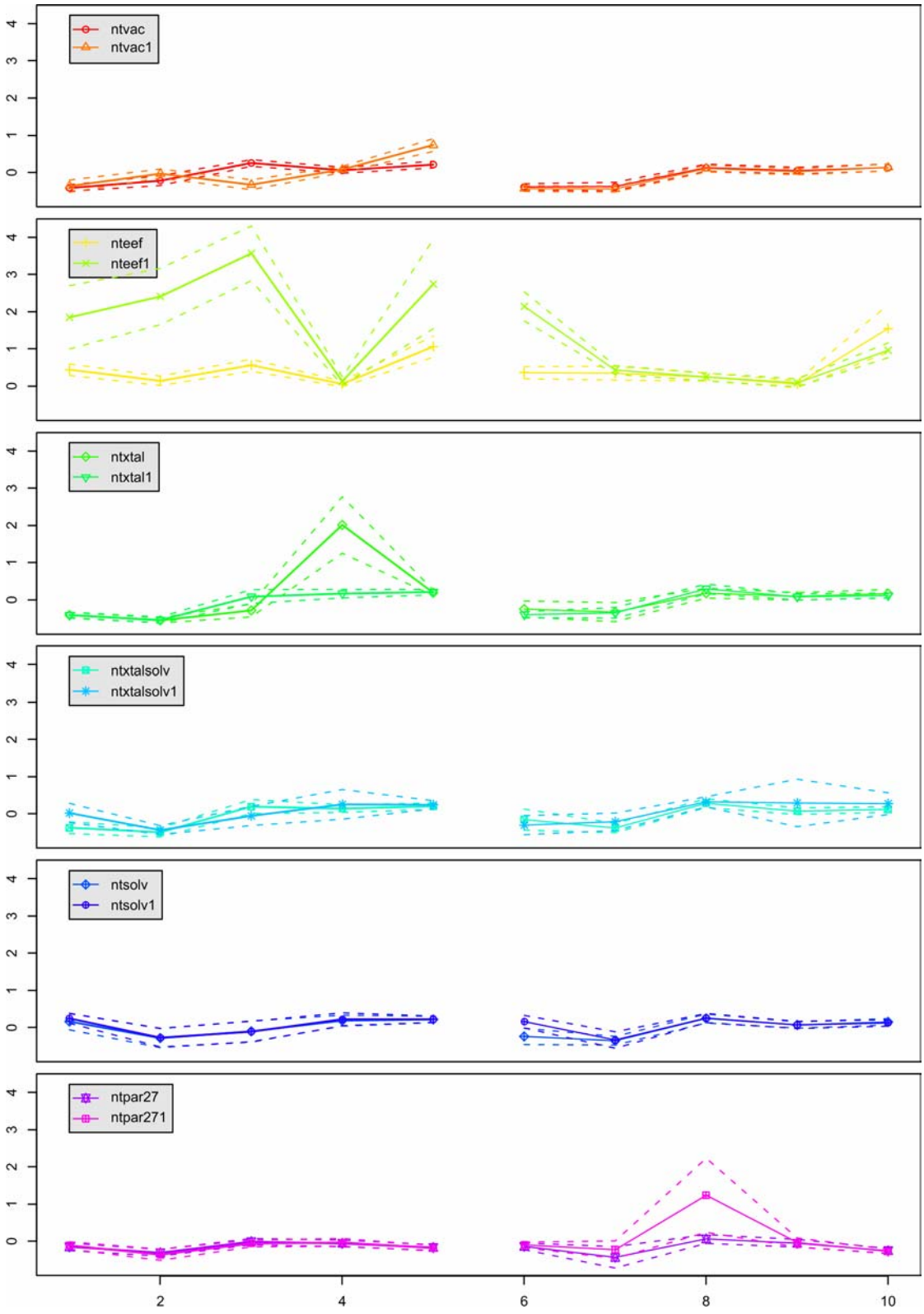


Figure 4.6: *Ca^{2+}-binding statistics for the 1ns MD simulations of the N-term lobe of CaM. The abscissa correspond to the reference distances 1 to 10 detailed in* ***Table 4.2****. Points and solid lines indicate mean deviations from the reference values over the simulation in Å and standard deviation values are indicated by the dotted lines.*

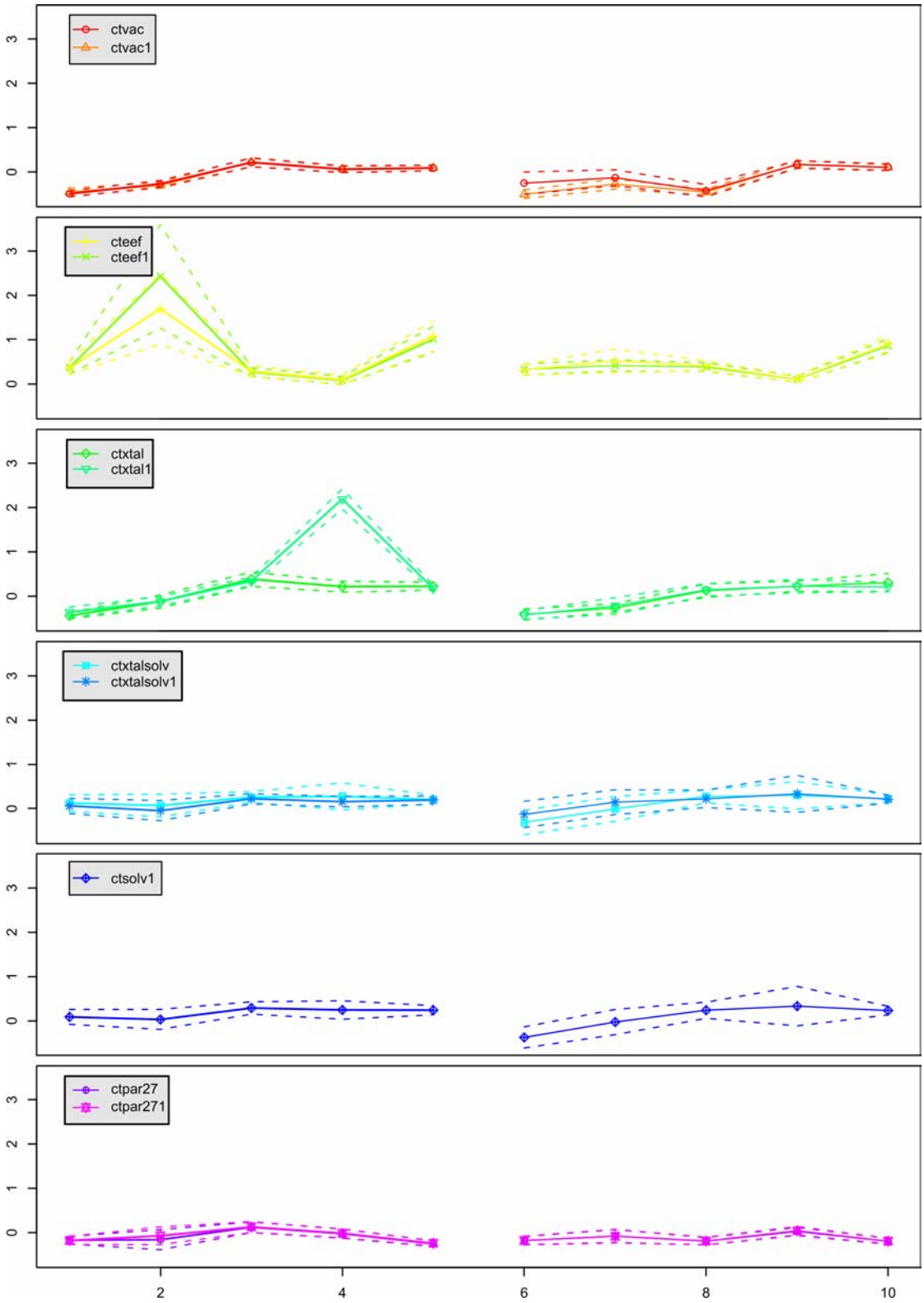


Figure 4.7: *Ca^{2+}-binding statistics for the 1ns MD simulations of the C-term lobe of CaM. The abscissa correspond to the reference distances 1 to 10 detailed in* ***Table 4.2****. Points and solid lines indicate mean deviations from the reference values over the simulation in Å and standard deviation values are indicated by the dotted lines.*

- **C-term simulations:**

Results of the C-term simulations shown in **Figure 4.7** show similar trends to those of the N-term results. Again simulations using the EEF1 potential seem to result in variable and inaccurate Ca^{2+} coordination distances throughout. Vacuum simulations suggest relatively little variation in the Ca^{2+}-binding distances although the mean values do not correspond well to the observed values in the crystal structure. Incorporation of solvent into the system generally results in more accurate Ca^{2+}-binding distances which are more variable over the course of the simulation.

Simulations using the all-atom PARAM27 potential once again demonstrate both the most accurate and least variable Ca^{2+}-coordination behaviour observed in these simulations.

4.4 Discussion & Conclusions

The application of the 6 different techniques to study the molecular dynamics of the two lobes of CaM has generated 23 different trajectories and a total of 23,000 different conformations for the protein. The interpretation of this wealth of information presents a considerable challenge in order for the relative merits of these different techniques to be adequately compared. If one of the more simplistic models is found to be sufficient in its treatment of CaM then opportunities abound to perform multiple long molecular dynamics experiments to further explore the conformational space accessible to the protein. If however a more sophisticated model is required to accurately accommodate the high charge associated with the protein and the special requirements of the Ca^{2+} ions, then the tractability of extended simulations will be more limited.

It is within this context that the relative performance of the different techniques employed here need to be assessed on the basis of the three different analysis techniques used to characterise the behaviour of the MD experiments performed on the CaM lobes.

4.4.1 RMSD Analysis

RMSD is the most commonly used comparison between any two structures in the sphere of structural biology. As such the technique provides a readily accessible and easily understood framework for the comparison of the different conformations generated by the molecular dynamics simulations under investigation here. In the context of these studies, a large and fluctuating RMSD over the course of a particular trajectory would indicate a simulation in which the conformation of the protein was poorly maintained. As the crystal structure used to generate these trajectories was observed at low temperature (100K) and in a crystal lattice, the result of minimisation and subsequent heating to a physiological temperature of 300K will inherently result in some changes in conformation. Of greater importance however is the extent of flexibility in the latter stages of the simulation in terms of the character of the sampling of conformational space.

The RMSD analysis of the 23 different trajectories studied here highlighted clear differences between the first three methods and the second three for both simulations: Experiments in which there was no complete solvation shell encompassing the protein fragment (**Class I**), suggested considerably greater conformational flexibility in the protein than those with a complete solvent shell of minimum depth 5.0Å (**Class II**). This distinction was most apparent, particularly in the case of the simulations of the C-term lobe, at the beginning of the production phase of the molecular dynamics. This would suggest that the solvent plays a critical role in the minimisation process and perhaps particularly in the absorption of heat. An alternative explanation could be that despite the use of a tight temperature coupling constant during the heating phase of the dynamics, the transfer of heat is more rapid for the water molecules which absorb much of the kinetic energy of the system effectively damping the protein fragment and limiting its potential for flexibility.

Within these fully solvated **Class II** simulations there are obvious differences of particular significance. It is perhaps noteworthy that the **xtalsolv** simulations seem to exhibit split behaviour in these simulations particularly in the case of **ctxtalsolv1** which exhibits greater conformational than any of the other **Class II** simulations. On the basis

of this analysis alone however it is almost impossible to uncover differences between the relative performance of PARAM19 and PARAM27.

4.4.2 RGYR Analysis

The use of Radius of Gyration information in this context provides an interesting supplement to the more general RMSD data. The RGYR measure can be considered a vector property which provides information about the *character* of changes in the system in contrast to RMSD which purely returns information on the *magnitude* of any changes. In particular considerable increases in RGYR can normally be attributed to the formation of a more extended and less well-folded conformation normally associated with poor structures.

In the context of this work, the RGYR analysis once again highlighted differences between **Class I** and **Class II** trajectories. These differences were particularly apparent in the C-term simulations, which, with the exception of **ctvac,** indicated that the **Class I** simulations resulted in conformations which were considerably more extended than the crystal structures used for comparison. In the case of the N-term, the **Class I** simulation were split between more extended and more globular conformations relative to the **Class II** simulations and the reference crystal structures.

This measure is particularly appropriate to the considerations of CaM conformation in the light of recent work in which RGYR variation suggests a compaction of the CaM lobes in solution relative to **1exr** and other crystal structures [Chou *et al.* 2001; Vigil *et al.* 2001]. In the light of these suggestions, the trend of the **Class II** simulations towards a more globular conformation than observed in the reference crystal structures is therefore particularly appropriate.

4.4.3 Calcium Coordination Analysis

Given the central role of Ca^{2+}-binding in CaM function, it seems appropriate that a more specific criterion should be introduced for the comparison of these experiments. The

analysis technique presented here does not demonstrate a rigorous characterisation of the Ca^{2+}-binding behaviour over the course of these simulations. It does however present a simple and accessible measure of the way in which the Ca^{2+} ions are accommodated by the protein and as such provides useful information on the behaviour of the system during simulations.

Only at this level of detail do differences begin to emerge *within* the **Class I** and **Class II** simulations. The vacuum simulations present precise Ca^{2+}-binding through well-maintained distances throughout the simulations. These distances do not however compare particularly accurately with those observed in the crystal structure. The addition of explicit solvent molecules to the system seems to present a distraction to the system and allows for more flexibility in the side-chains coordinating the Ca^{2+} ions. As a water molecule is the 6^{th} naturally occurring Ca^{2+} ligand in each of the EF hands, it could be that the completion of the coordination geometry allows for greater flexibility in the system.

It does however seem clear that the use of the PARAM27 potential provides not only more precise Ca^{2+} coordination, demonstrated through more consistent distances, but also greater accuracy in the average values with respect to the crystal geometry.

4.4.4 General Conclusions

The assessment of the simulations performed here has suggested several underlying conclusions.

- **The use of the EEF1 implicit solvent potential is inappropriate for use with the MD of CaM fragments.**

- **The use of explicit solvent in these simulations results in more robust simulations less susceptible to large conformational changes on a short timescale, particularly during MD heating phases.**

- **Subtle differences can be identified between MD trajectories which have been solvated in different ways.**

- **The increased computational requirements of explicit solvent models are worthwhile and that the use of the PARAM27 all atom potential can be justified on the basis of better Ca^{2+}-binding behaviour.**

Chapter 5: Ca^{2+}-Dependent Molecular Dynamics of Calmodulin

5.1 Introduction

The conformational flexibility of Calmodulin is undoubtedly of profound importance to the function of the protein as a molecular switch. The central paradigm for the function of CaM in solution is based on a conformational change that takes place upon Ca^{2+}-binding to the EF-hands of the protein resulting in exposure of the hydrophobic CaM target binding surfaces. This activated Ca^{2+}-bound state is then able to bind to diverse targets initiating different processes in living systems. Such behaviour can only be accommodated through an inherently flexible system which is able to adopt numerous different conformations under different conditions. The studies detailed in this chapter use the technique of molecular dynamics (MD) in an attempt to better understand the range of conformations accessible to the CaM system and the role of Ca^{2+}-activation.

5.1.1 Solution Structure of Ca^{2+}-CaM

The detailed comparison of the ensemble of experimentally determined structures of CaM has already been investigated in Chapter 3. This work concentrated on the crystal structures of Ca^{2+}-CaM and those of the protein in complex with different inhibitors and fragments of CaM target domains. A principal component analysis of these structures for both the C and N-term lobes of CaM illustrated that the Ca^{2+}-bound conformation was relatively well conserved across the observed structural ensemble. It was also shown that the interaction between the protein and different co-factors induced conformational changes in both lobes of the protein. At the time of this investigation, there were no solution (NMR) structures for Ca^{2+}-CaM in the absence of any co-factors such as small molecule inhibitors or CaM-target mimics.

The developing method of Residual Dipolar Coupling (RDC) [Chou *et al.* 2000] has recently been used to determine the structure of the individual CaM lobes in a liquid

crystal matrix [Chou *et al.* 2001]. These experiments were carried out in a dilute aqueous suspension of oriented particles at a temperature of 32°C and it was demonstrated that the liquid crystal had no effect on the structure of the CaM domains. Refined solutions to the RDC data were obtained using three different initial structures; apo CaM (**1f70** & **1f71**) [Chou *et al.* 2000], a Parvalbumin derived homology model and the 1.0Å crystal structure of Ca^{2+}-CaM (**1exr**) [Wilson & Brunger 2000] used in these simulations, all of which resulted in highly similar final models.

Comparisons of these NMR solution structures with crystallographic structures showed a similar result for the C-term lobe of CaM although the structures derived for the N-term lobe exhibited a considerably more closed conformation than that observed in any of the crystal structures. It was suggested that the more open conformation of the N-term in crystal structures was the result of crystal contacts. Incorporation of the RDC NMR structures into the PCA highlights these conformational differences and shows these structures to be substantially different from other crystallographically determined structures for the N-term. The conformational subspace projections of the PCA for the N and C-term structures respectively are shown in **Figure 5.1** and **Figure 5.2** below. A more detailed presentation of the results from this analysis is shown in the Appendix D in conjunction with illustrations of the shapes of the first three PC's from the distribution.

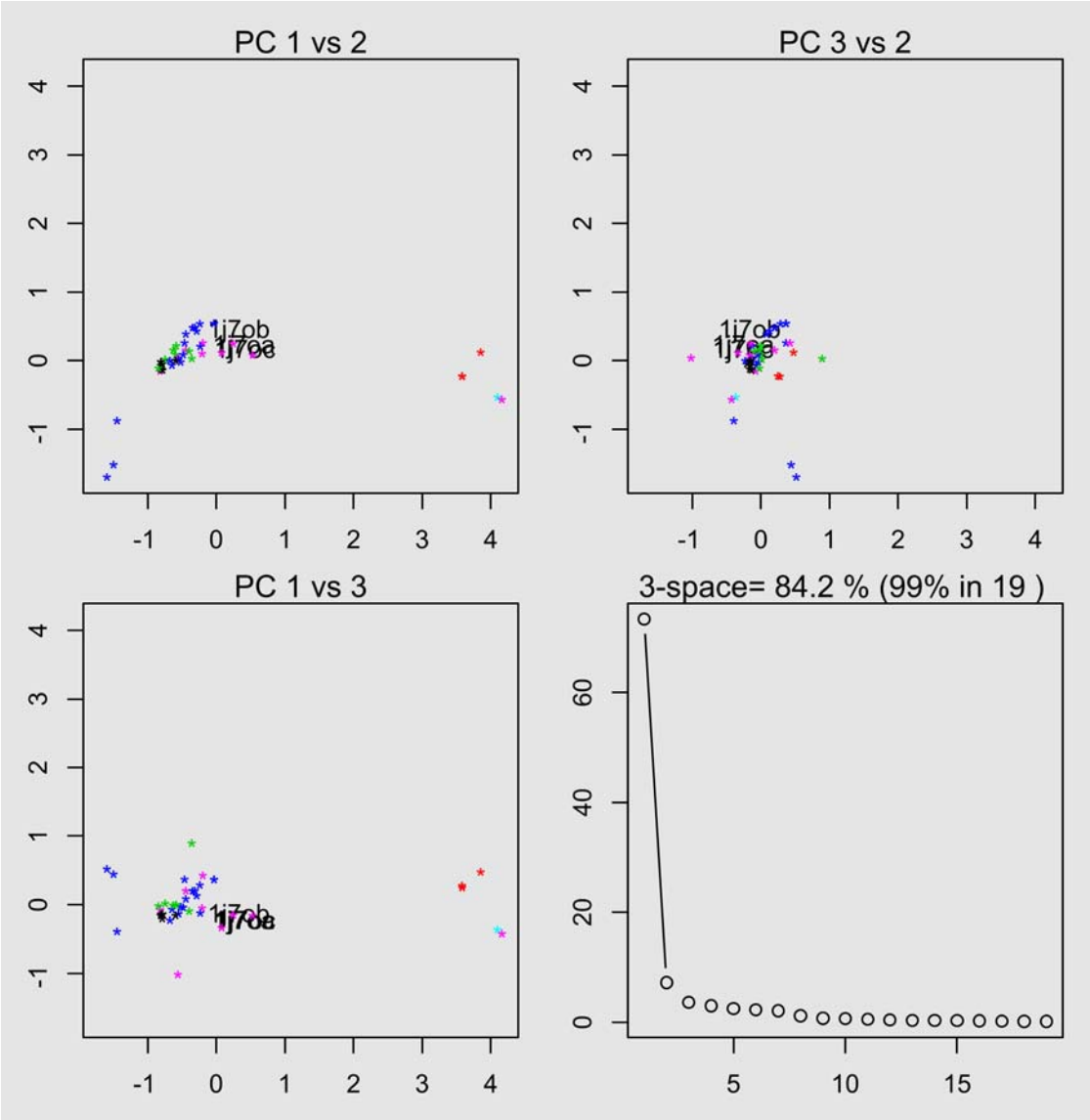


Figure 5.1: *Conformational subspace projections for PC's 1 to 3 of the N-term observed structures. The PCA was performed on the structures used in the full analysis described in an Appendix with the addition of the RDC structures of. The* Chou et al *original structures are marked with asterisks and coloured according to their classification (see Chapter 3). The new RDC structures are indicated with PDB codes and chain identifiers for the three different experimental solutions.*

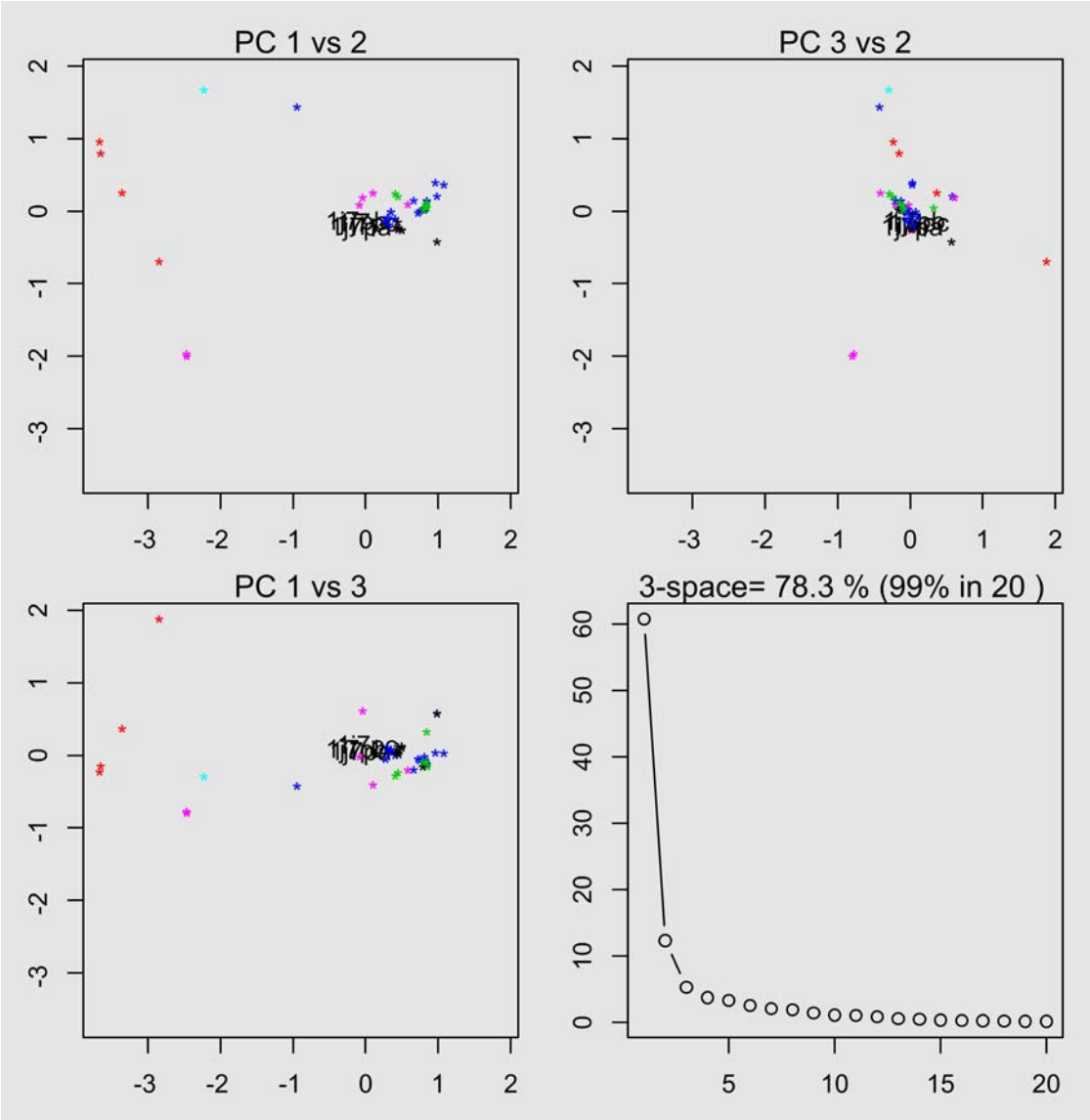


Figure 5.2: *Conformational subspace projections for PC's 1 to 3 of the C-term observed structures. The PCA was performed on the structures used in the full analysis described in Appendix with the addition of the RDC structures of Chou et al. The original structures are marked with asterisks and coloured according to their classification (see Chapter 3). The new RDC structures are indicated with PDB codes and chain identifiers for the three different experimental solutions.*

The conformational subspace projections onto the first two PC's of the structural dataset including these new structures clearly illustrates the similarity of the RDC structures derived from significantly different starting structures. The C-term structures are slightly distinct from the other experimentally observed Ca^{2+}-CaM structure shown in black and show little similarity to any of the other classes of inhibitor or target-bound structures (green and blue respectively). The analysis of the N-term structures shows more

significant structural differences between the RDC structures and the tight cluster of Ca^{2+}-CaM crystal structures. The three RDC structures are quite similar to each other and also to the TnC structures **1tn4**, **2tn4** and the modelled structure **2cln**.

From these results it is clear that the structures determined by this relatively new NMR technique using residual dipolar couplings show interesting differences to other experimentally determined structures of Ca^{2+}-CaM. The comparisons made by the authors of these new structures are based on differences in the angles between the helices of the EF-hands. On the basis of such comparisons, the RDC structures for the N-term are seen to be *semi-open*; roughly half way between open and closed. When compared using the PCA technique illustrated above however, these new structures show much greater similarity to the open conformation observed in other Ca^{2+}-bound structures. This suggests that the PC's describe other more global structural differences between open and closed conformations in addition to the differences inter-helical angles and thus provide a better basis for comparison of CaM conformation.

5.1.2 Conformational Exchange in Solution

A view is emerging in the field of CaM structural research which suggests that the C-term lobe of CaM has the potential to exchange between different conformations in solution [Malmendal *et al.* 1999]. The possibility that such an exchange also exists in the N-term lobe adds a level of considerable complexity to the understanding of CaM function. The classical view suggests that CaM exists either in an inactive, closed, Ca^{2+}-free conformation or in the activated, open Ca^{2+}-loaded state. Recent work discussed below however suggests that one or possibly both apo-CaM lobes exist in dynamic equilibrium between open and closed conformations with a Ca^{2+}-dependent equilibrium constant.

In 1997 Evenäs *et al.* Published the results of NMR studies of the dynamics of an E140Q mutant of the C-term tryptic fragment (TR2C) of CaM [Evenas *et al.* 1997]. This mutant removes the crucial bi-dentate glutamate residue resulting in a significant reduction of the Ca^{2+} affinity of the fourth EF-hand of the protein. The doubly loaded $(Ca^{2+})_2$-state of the fragment was seen to exist in a dynamic equilibrium between the open and closed

conformation of which only 65% was in the classical, loaded, open conformation. Subsequent work [Evenas *et al.* 2001] confirmed such a conformational exchange and suggested a mean rate of 21 μs with a similar exchange occurring in the wild-type protein. The exchange was described as a complex process involving all residues in the protein fragment.

The NMR RDC structural investigations of Chou *et al.* [Chou *et al.* 2001] rule out any such dynamic exchange. A review accompanying this article [Akke & Chazin 2001] points out that the anisotropic refinement of the low temperature (100K) 1.0Å crystal structure of CaM [Wilson & Brunger 2000] resolves B-factors corresponding to a scissor-type fluctuation of each EF-hand which would correlate with an open-closed transition.

A recent NMR study by Mercier *et al.* looked at the dynamical behaviour of the C-term lobe of Ca^{2+}-loaded TnC in complex with the binding domain of TnI [Mercier *et al.* 2001]. This work clearly demonstrated that interaction of the protein domain with a target sequence considerably reduced the conformational space accessible to the fragment.

5.1.3 Molecular Dynamics Simulations of CaM

The size and flexibility of the CaM system represent a challenging yet exciting proposition for investigation with molecular dynamics simulations. Early studies concentrated on the flexibility of the central helix linking the two Ca^{2+}-binding domains and its ability to allow them to come together and envelope target molecules. Van der Spoel and co-workers [Van Der Spoel *et al.* 1996] used various molecular mechanics techniques to probe the flexibility of the linker helix in the absence of the Ca^{2+}-binding lobes. These studies discovered that the bending facilitated by residues 77-80 of the linker was an intrinsic property of the amino acid sequence and not caused by any essential interactions of the two CaM lobes.

A more extensive dynamics study was undertaken by Wriggers *et al.* in which they studied the molecular dynamics of intact Ca^{2+}-CaM in a sphere of water with additional counter

ions to neutralise the charges associated with the acidic protein [Wriggers *et al.* 1998]. Again this work concentrated on the effects observed in the flexibility of the linker helix although some interesting observations were made regarding the lobes themselves. Comparisons of the flexibility of the two lobes revealed greater flexibility in the N-term than the C-term. The re-organisation of the central helix enabled a re-orientation of the relative positions of the two lobes such that the two binding surfaces were facing one another in an orientation similar to that found in the CaM-target complex structures. A conformational change was also observed 1.3ns into the 3ns trajectory which resulted in a more closed conformation for the lobe. The details of this conformational change are not explored in detail however the authors do correlate the observed differences in lobe flexibility with the relative affinities of the lobes for CaM targets. It was also noted on the basis of these conformational rearrangements that it is necessary to run such simulations for as long a time as possible as the more significant conformational deformations require considerable time periods to propagate throughout the system.

A molecular dynamics study of Yamaotsu *et al.* studied the crystal structure **1lin** of CaM in complex with multiple copies of the inhibitor ligand TFP [Yamaotsu *et al.* 2001]. The focus of this study was the interaction between CaM and the TFP ligands in an attempt to determine the character and location of all the potential binding sites on the CaM lobe surfaces. Again however some interesting observations were noted regarding the relative flexibility of the two CaM lobes. In these simulations the N-term lobe of the structure became disordered over the course of the simulation and the Ca^{2+}-binding loops of the C-term lobe lost their EF-hand conformations towards the end of the 2ns simulation. These results lead the authors to conclude that the CaM lobes were able to alternate between ordered and disordered conformations in solution.

The most relevant piece of work to this study was published only after the initial phase of these studies had been completed and was carried out by Vigil *et al.* [Vigil *et al.* 2001]. This group performed MD experiments on the N-term lobe of CaM in both the presence and absence of Ca^{2+}. In these studies the open conformation of the N-term lobe determined experimentally [Chattophadhyaya *et al.* 1992] was seen to undergo a conformational change from the open towards a closed conformation, after 1ns of simulation. Small angle X-ray scattering data was also used to confirm the presence of a more closed structure than suggested by X-ray crystallography. The closed conformation

observed in the simulation however was found not to be the same as the NMR structure for the Ca^{2+}-free structure of CaM. The conformation derived from the simulation included some solvent exposed hydrophobic amino acids which are not seen in the NMR structures for the protein fragment. These data were used to propose a model in which the Ca^{2+}-loaded N-term lobe of CaM exists in dynamic equilibrium between open and closed conformations under physiological conditions. Once again it is suggested that crystal packing forces act to stabilise or even induce the open conformation seen in the crystal. Interestingly this work remains uncited elsewhere in the literature suggesting perhaps that this proposed mode of action for CaM has not been generally accepted at this time.

5.1.4 Conclusions from previous work

The literature reviewed above is representative of a considerable amount of work in the investigation of the dynamical properties and conformational flexibility of Calmodulin in solution. All the experimental structural evidence suggests that this is a highly flexible system which is able to explore a variety of different conformations in solution. The different structures of these conformations appear not to be limited to closed and open states and the Ca^{2+} dependency of interchange between different conformation also remains undetermined.

The molecular dynamics studies presented in this chapter seek to further characterise the conformational space accessible to the C and N-term lobes of CaM and the role that Ca^{2+}-binding has to play in their structure and dynamics.

5.2 Materials & Methods

5.2.1 Molecular Dynamics Simulations

Molecular Dynamics simulations for the separate lobes of CaM were performed on the 1.0Å crystal structure **1exr** [Wilson & Brunger 2000]. Fragments corresponding to the regions E2-A73 and L85-S147 were used for the N and C-term lobes respectively. For residues with multiple conformations as determined experimentally, the conformation corresponding to dis-order locator A was used throughout the structures. The simulations of the CaM-CaMKIIα complex was performed on the first model (chains **A** and **B**) of the 2.0Å crystal structure **1cm4**. The protein fragment resolved in the experiment included residues L4-T146 and the peptide corresponding to a fragment of the CaM binding domain of CaMKIIα (residues F293-T310).

The N-termini of both protein and peptide fragments were patched with extra groups to remove the induced charges from truncation of the sequence. Coordinates for the patches were built using the ICBUILD function in CHARMM and hydrogens were subsequently added using 5 iterations of the HBUILD function. Polar hydrogens corresponding to the PARAM19 parameter set or a full hydrogen representation from the all atom PARAM27 parameter set were used for the simulations according to the table below.

5.2.1.1 Protein Solvation

Protein fragments were solvated using the *SOLVATE* package of Grubmüller *et al.* [Grubmueller 1996]. The protein fragment to be solvated was passed to *SOLVATE* which generated a shell of waters with a minimum solvent shell depth of 5.0Å around the fragment.

5.2.1.2 Non-bonded Parameters

A fixed dielectric with a constant (ε_r=1.0) was used for these simulation. Electrostatic forces were shifted to zero at 13.0Å and the Van der Waals potential was treated with a switching function between 12.0 and 13.0Å for a smooth cutoff. The non-bonded cutoff

is sufficiently large to incorporate interactions between the adjacent Ca^{2+} sites in the CaM fragments which are less than 12Å in the crystal structure. This is in accordance with the recommendations of Marchand *et al.* [Marchand & Roux 1998].

5.2.1.3 Minimisation & Heating

The minimisation protocol outlined in the introduction was applied to remove any bad interaction generated by the building of patches and the hydrogen atoms and to remove any poor Van der Waals contacts. The final gradient tolerance was 10^{-5} kcal mol^{-1} $Å^{-1}$. The structures were then heated to 300K by performing 10ps of dynamics on the system at 25K increments. The use of a tight temperature coupling to the water bath was applied at this stage to ensure that the kinetic energy was distributed evenly throughout the system.

5.2.1.4 Molecular Dynamics Production Phase

The molecular dynamics of the different systems were calculated over a 10ns period using steps of 1fs with coordinates saved every ps. A tight temperature coupling constant of 0.5ps was used for the production phase.

Two different types of simulation were performed, the details of which are summarised in **Table 5.1** below. Simulations were performed using both the polar hydrogen PARAM19 potential [Neria *et al.* 1996] and the all atom PARAM27 potential [MacKerell *et al.* 1998] of CHARMM for the CaM fragments. Only the PARAM19 potential was used for the MD simulation of the CaM-target complex. All the simulations were run for a production phase of 10ns, in the case of **ntext**, the 10ns was performed using the coordinates and velocities generated by the previous 10ns of simulation, recorded in **ntsolv**. The simulation **ctroc** used exactly the same initial coordinates, potential, random number seed and protocol as **ctsolv** but was performed on a different computing platform. The particulars of the Ca^{2+}-bound and Ca^{2+}-expulsion experiments are outlined below:

- **Ca^{2+}-bound simulations**

The Ca^{2+} ions are included in these simulations and initially positioned according to their positions in the crystallographic structures **1exr** for the lobe fragments and **1cm4** for the protein-target complex. Note that for the simulations of the N-term lobe, positions for 3 different Ca^{2+} ions were determined, all of which were included in the simulation.

- **Ca^{2+} Expulsion simulations**

This study was designed specifically to investigate the structural role of the Ca^{2+} ions in CaM. In these simulations (denoted with an **apo** in the descriptor code) the full crystal structure with added hydrogens and modified chain termini is minimised and gently heated to 300K as described above. Prior to the production phase of the simulation however, all the Ca^{2+} ions are removed from the system and the simulation then continued in the normal way. This instantaneous removal forces the system to adapt to accommodate the loss of the Ca^{2+} and from the induced response it is hoped to establish the structural role of the ions. After any initial response, the simulation should continue to explore conformations characteristic of the Ca^{2+}-free system.

Table 5.1: *Summary of the 10ns MD simulations performed in this section.*

Code	*Description*			*No. of atoms*
ntsolv	N-term	Ca^{2+}-expulsion	PARAM19	4694
ntapo	N-term	Ca^{2+}-expulsion	PARAM19	4691
ntext	N-term	Ca^{2+}-bound	PARAM19	4694
ntpar27	N-term	Ca^{2+}-bound	PARAM27	5117
ntp27apo	N-term	Ca^{2+}-expulsion	PARAM27	5114
ctsolv	C-term	Ca^{2+}-bound	PARAM19	4214
ctapo	C-term	Ca^{2+}-expulsion	PARAM19	4212
ctroc	C-term	Ca^{2+}-bound	PARAM19	4214
ctpar27	C-term	Ca^{2+}-bound	PARAM27	4596
ctp27apo	C-term	Ca^{2+}-expulsion	PARAM27	4594
ntcomp **ctcomp**	CaM-CaMKIIα Complex	Ca^{2+}-bound	PARAM19	9188

5.2.2 Analysis Techniques

The details of the different techniques used to analyse the trajectories are outlined in Chapter 2. Principal component analysis was performed on the Cα coordinates of the 10,000 1ps frames from the trajectories as outlined in the technique of *Essential Dynamics* developed by Amadei *et al.* [Amadei *et al.* 1993].

A standard reference frame was also calculated using the observed structural dataset of CaM structures. These structures were superposed onto a common reference frame and the principal components were calculated. The structural dataset used for this analysis and projections of the average structure along these PC's are provided in the Appendix. Where appropriate this observed PC space was used for comparison of the different conformations generated from the MD simulations.

Molecular visualisation of the trajectories was performed using the QUANTA package [Accelrys 2001] and data manipulation and analysis was performed using the R statistical language [Gentleman & Ihaka 2000].

5.3 Results & Analysis

5.3.1 N-term Simulations Overview

The characteristics of the different simulations performed on the N-term lobe of CaM are outlined in **Table 5.1** above. The general trends observed during these molecular dynamics trajectories are best illustrated in plots indicating the time dependence of the RMS difference with respect to the initial structure prior to minimisation. The plots for the six simulations performed on the N-term lobe of CaM are shown below (**Figure 5.3**). This plot clearly shows a significant conformational change occurring 4ns into the **ntsolv** trajectory. This new conformation appears to be largely retained throughout the rest of the simulation and also throughout the subsequent 10ns of dynamics performed in **ntext**. Structural transitions are less prominent in the other simulations. The **ntpar27**

simulation proves to be the least excursive system with the conformation restricted to within 1.0Å of the crystal structure throughout the simulation.

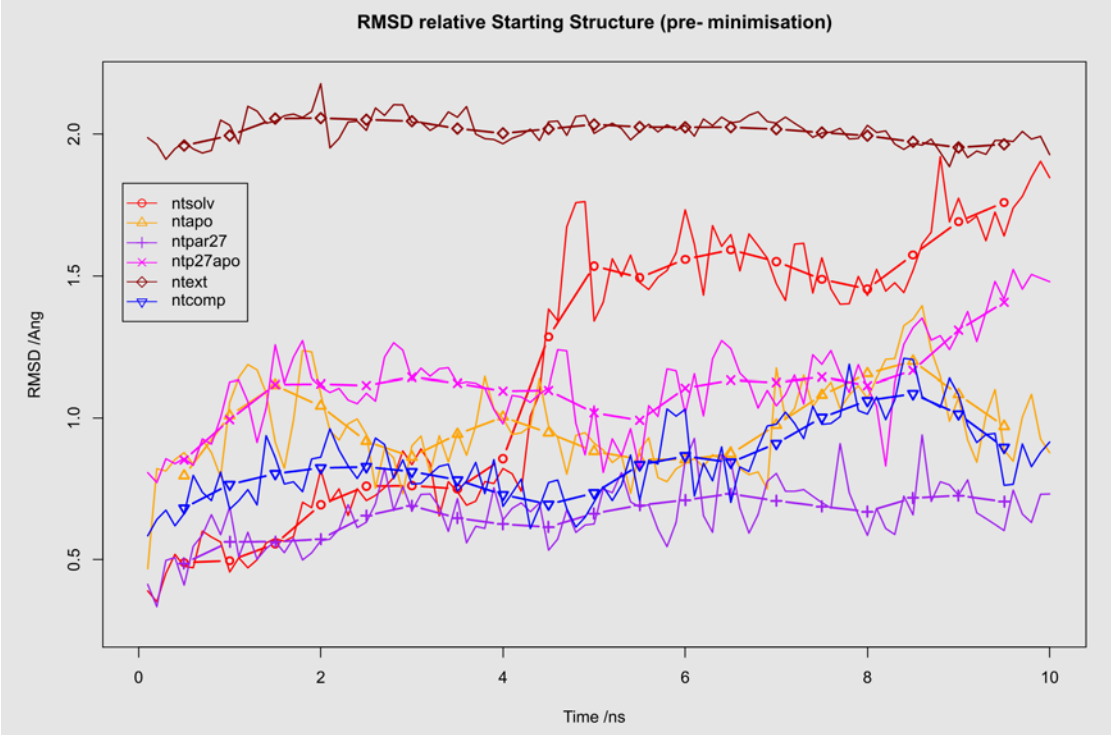


Figure 5.3: *Time dependence of RMSD relative to* ***1exr*** *for the different 10ns MD trajectories of the N-term lobes of CaM. Codes in the legend correspond to those outlined in* ***Table 5.1***

5.3.2 N-term PARAM19 Simulations

In **Figure 5.4** below, frames from the **ntsolv** trajectory are projected onto the conformational subspace described by combinations of the first three PC's, which describe almost 82% of the variation of the trajectory. Superposed onto this subspace are projections of the experimentally determined CaM observed structures detailed in Chapter 3.

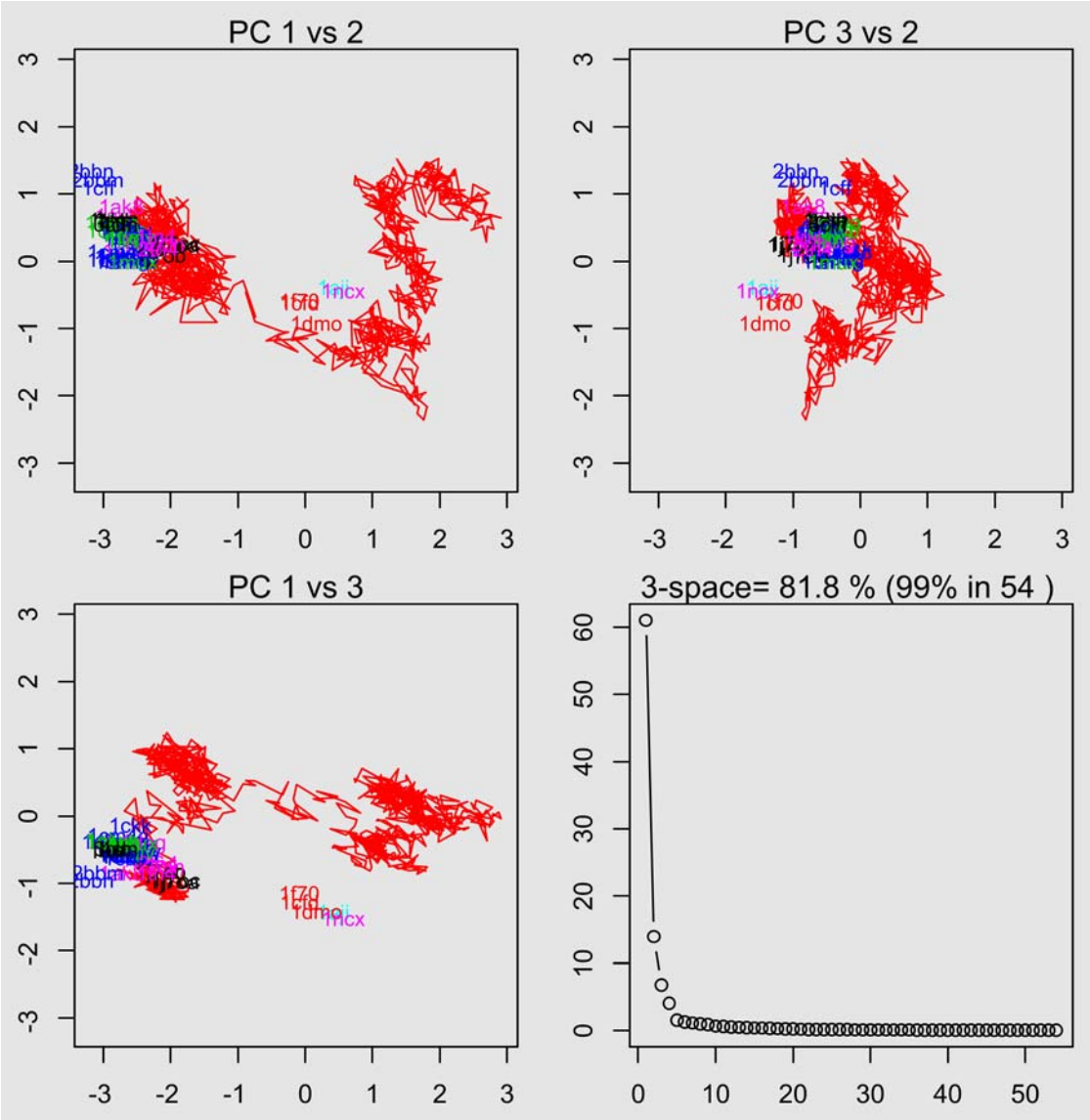


Figure 5.4: *Conformational subspace projection of the 10ns Ca^{2+}-CaM MD trajectory for the N-term onto the first three principal components of the trajectory. The trajectory* ***ntsolv*** *is shown in red.*

From these plots it is apparent that the *open* conformation observed by most of the experimentally determined structures of the N-term of Calmodulin is explored by the MD trajectory over the first few ns of the simulation. The conformational change observed 4ns into the trajectory however, results in a displacement towards the *closed* conformation associated with Ca^{2+} free CaM. **Figure 5.5** below shows a structural superposition of the coordinates of the structure at 4 and 5ns into the **ntsolv** trajectory. The structures are compared with the Ca^{2+}-bound structure **1exr** used to generate the trajectory and the apo (closed) structure **1dmo**. From this superposition it can be seen that the structure prior to the most significant conformational change is already

somewhat different to the crystal structure of the protein. The structure at 5ns however does seem to be very similar to the **1dmo** structure representative of the Ca^{2+}-free conformation of the lobe, particularly in terms of inter-helical angles. The differences shown in the conformational subspace projections are largely due to axial displacements along the helices A and D[4] as described in the PC projections in the appendix.

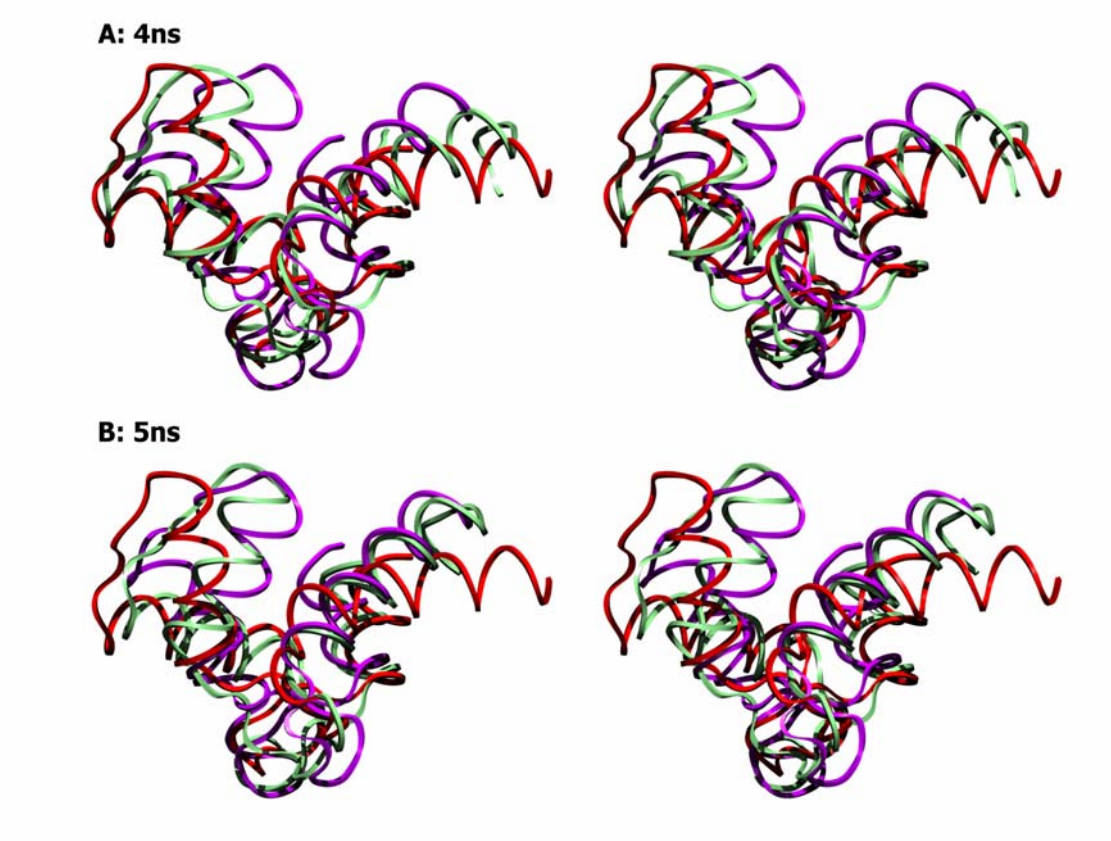


Figure 5.5: *Stereo structural superpositions of the CaM lobes before and after the conformational rearrangement at 4ns in* ***ntsolv****. The open Ca^{2+} structure of* ***1exr*** *is shown in red and the apo structure* ***1dmo*** *is shown in magenta. The structure generated by the simulation is shown in green at 4ns in figure* ***A*** *and 5ns in figure* ***B****.*

It should however be noted that the final conformation observed in this simulation is not the closest to the apo structure **1dmo**. The final conformation in this simulation (which is retained in **ntext**) has an RMSD of 1.4Å relative to **1dmo** although at its closest points in conformational intermediates at ca. 5ns (shown above), the trajectory comes to within

[4] This nomenclature is defined in the Anatomy of CaM section of the introduction

1.1Å of **1dmo**[5]. Rather than an exploration of the conformations described from the experimental results as the open (Ca^{2+}-bound) and closed (Ca^{2+}-free), this simulation clearly samples new conformations not currently observed by experimental structure determination. The subsequent 10ns of dynamics performed in the simulation **ntext**, show little deviation from the final conformation adopted in this simulation. This can be seen in **Figure 5.6** below.

The Cartesian coordinates corresponding to these trajectories can also be projected onto the subspace described by the principal components of the observed structures calculated previously, the results of which are illustrated in the appendix. This provides a different view of the data and can help to identify different trends between the trajectories and experimentally observed structures. **Figure 5.6** below shows projections of the 10ns MD trajectories **ntsolv**, **ntext** and **ntapo** onto the subspace described by combinations of the top three PC's from the observed structural distribution. In this subspace the **ntsolv** trajectory shows the same overall form as observed in **Figure 5.4** above. The final conformation generated in **ntsolv** and **ntext** however appears to be more similar to the apo structures **1cfc** and **1dmo**. Projection of the Ca^{2+} expulsion run **ntapo** onto this space highlights the limited effect of deleting the Ca^{2+} ions from the structure. The variance plots illustrate the extent to which the observed PC's describe the variance of the individual dynamics trajectories. From these plots it is clear that although PC1 of the observed distribution is a good descriptor of the variance in **ntsolv** describing some 27% of the total. The variance of the trajectory **ntapo** is also shown to be well described by the first few PC's of the observed structures although none of the PC's describe more than 6% of the total variance in the **ntext** trajectory.

[5] These pairwise RMSD values are based on a superposition of all structures to an average structure over the full set of observed structures for the N-Term.

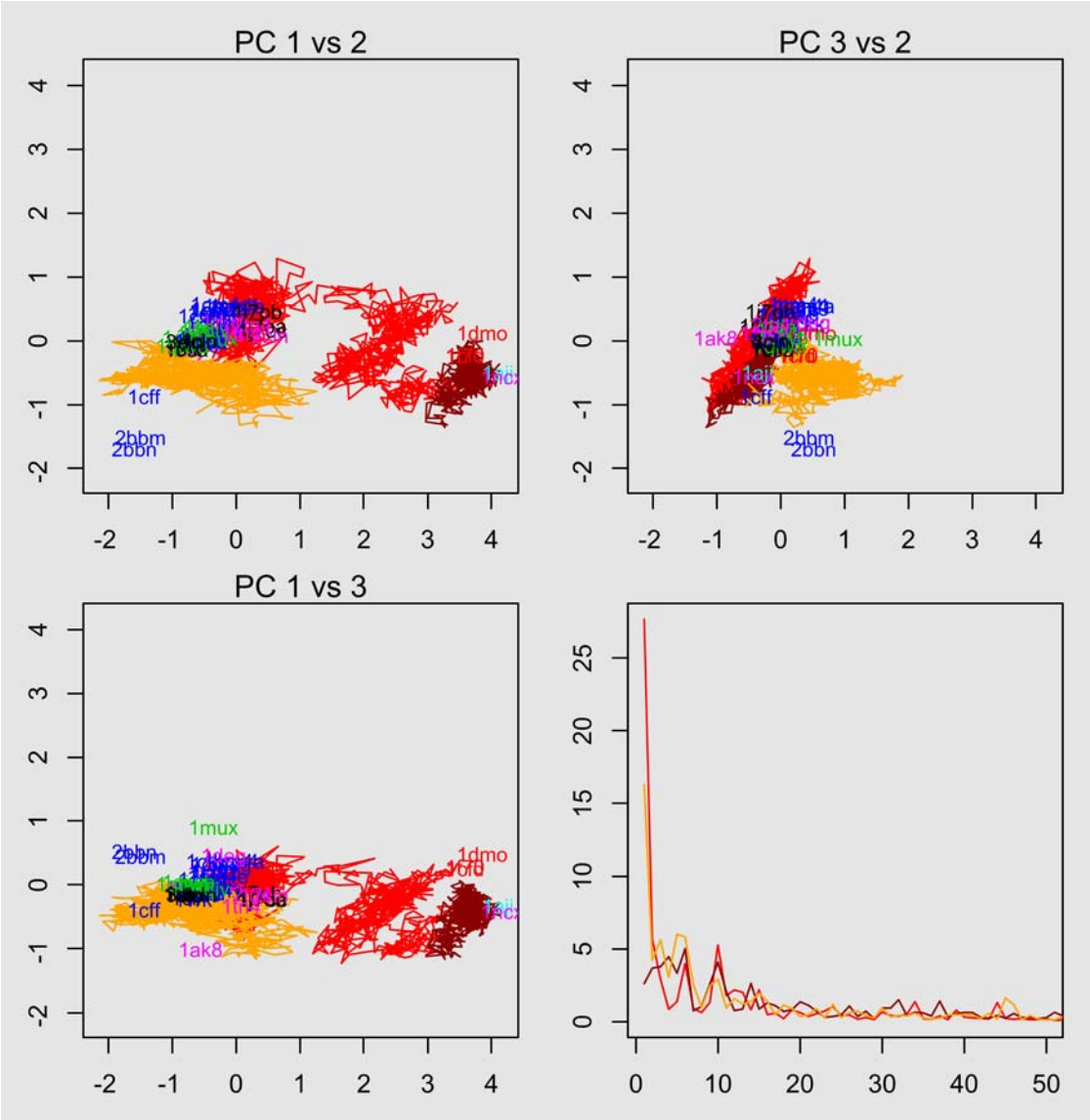


Figure 5.6: *Conformational subspace projection of MD trajectories for the N-term onto the first three principal components of the observed structures. The trajectory* ***ntsolv*** *is shown in red, the extended 10ns* ***ntext*** *is shown in burgundy and the* Ca^{2+}*-free trajectory* ***ntapo*** *is shown in orange. The fourth plot of the panel shows the percentage of total variance for a given trajectory described by the top 50 PC's for the observed structures.*

The expulsion of Ca^{2+} from the protein fragments would be expected to cause the system to adopt a more closed apo-like conformation. From these results this is clearly not the case. Over the course of this 10ns simulation, there appears to be no propensity for the trajectory to progress towards a closed Ca^{2+}-free like conformation. Experiments were also performed in which the Ca^{2+} ions were expelled at different points along the **ntsolv** trajectory in to investigate if these resulted in any significant alterations to the different

conformations. In the 100ps subsequent to deletion of the Ca^{2+} ions, there was no significant effect on the trajectory (data not shown).

- **Ca^{2+}-binding behaviour**

The Ca^{2+}-binding behaviour of the system was investigated by monitoring the distances involving key atoms in the Ca^{2+} coordinating sidechains of the protein over the course of these simulations. The identity of these atoms and their reference values in the crystal structure **1exr** are shown in the previous chapter (Chapter 4) where the technique was introduced. Comparison of differences in these distances relative to the initial crystal structure with the overall RMSD from the initial structure can be used to show correlations between the local Ca^{2+}-binding conformation and global conformation of the lobe.

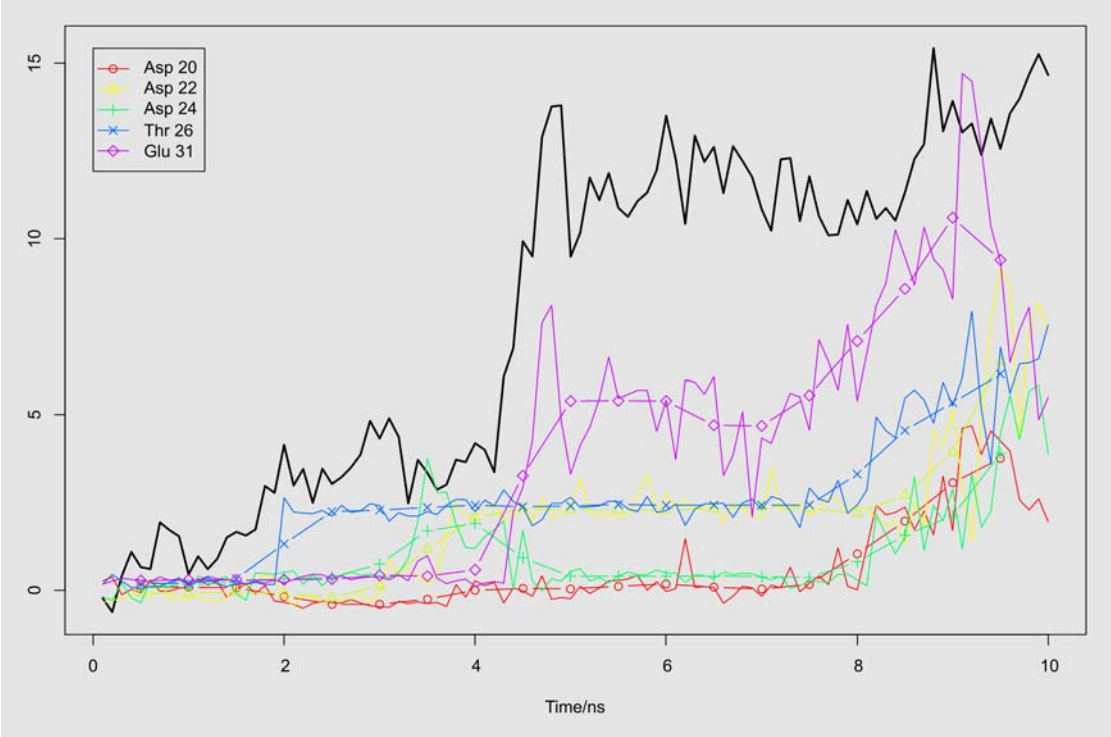


Figure 5.7: *Deviation in Ca^{2+}-binding distances (Å) for the first EF-hand during the simulation* ***ntsolv****. Actual differences relative to reference values are shown in solid lines with 500ps floating averages plotted by points. The overall Cα RMSD for the trajectory is shown in black.*

This analysis shows that the first ligand to alter its coordination is the backbone oxygen of residue Thr26. This change appears to occur as a result of the modest conformational

change taking place approximately 1.5ns into the simulation. The new arrangement of the group is some 2.5Å further away than in the reference structure and is conserved for the next 6ns of simulation suggesting an alternate mode of interaction with the Ca^{2+} or another part of the protein. Careful inspection of the changes taking place around 4ns into the simulation suggests that the loss of Ca^{2+} through the essential bi-dentate Glu31 occurs as a result of the conformational change in the protein. The progressive effects of this conformational change result in the complete loss of Ca^{2+} coordination in this site after 7.5ns. The Ca^{2+}-binding in this site is never regained in the subsequent 10ns simulation **ntext**.

Similar analysis of the second EF-hand of the N-term fragment shows that Asn60 is affected by the change in conformation 1.5ns into the trajectory as the Cγ-Ca^{2+} distance increases by some 6Å. This deformation is conserved for the rest of the simulation. Other Ca^{2+} coordination geometries are well maintained throughout the simulation.

- **CaM-complex MD simulation**

The 10ns simulation of the CaM-CaMKIIα complex was designed to examine to what extent the flexibility of the complex was limited by the presence of a target peptide in comparison to the unliganded structures. **Figure 5.8** shows projections of the conformations explored during this simulation onto the subspace described by PC's 1 to 3 of the observed distribution. These plots clearly show that the complex has a far more limited range of conformations accessible than the unliganded state of the protein. The conformations appear to be relatively restricted to those which have been determined experimentally for Ca^{2+}-CaM in the presence of various co-factors. The relatively flat variance plot also indicates that there are no significant motions in the system which correspond to other PC's of the observed distribution.

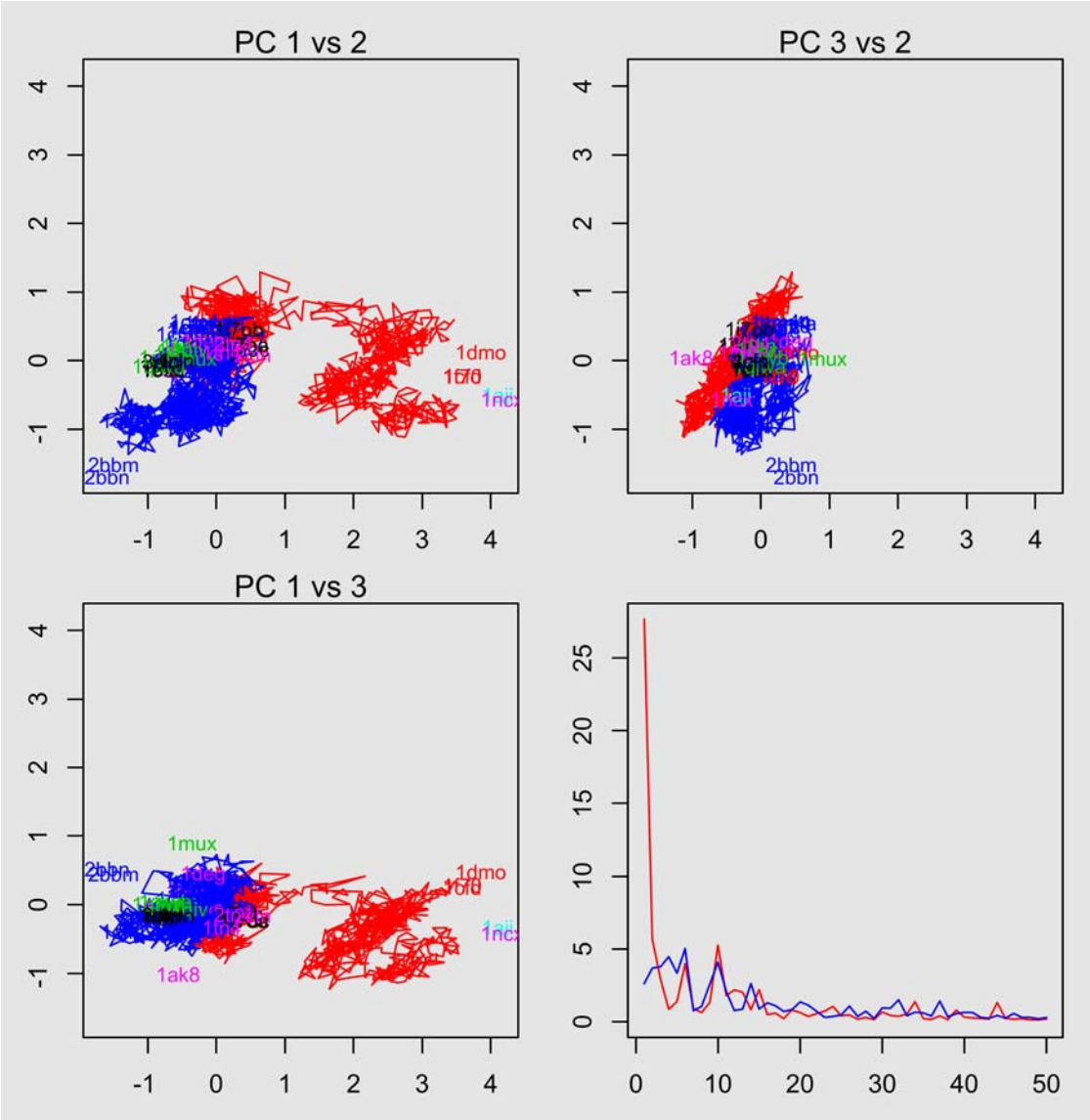


Figure 5.8: *Conformational subspace projection of MD trajectories for the N-term onto the first three principal components of the observed structures. The trajectory* ***ntsolv*** *is shown in red, the N-term fragment of the simulation of the complex* ***ntcomp*** *is shown in blue. The fourth plot of the panel shows the percentage of total variance for a given trajectory described by the top 50 PC's for the observed structures.*

5.3.3 N-term PARAM27 Simulations

The investigations outlined in the previous chapter suggested that the all atom PARAM27 CHARMM potential performs better at modelling the Ca^{2+} coordination in CaM. The RMSD plots shown in **Figure 5.3** suggest that the degree of conformational

flexibility in the **ntpar27** and **ntp27apo** simulations is significantly smaller in comparison to the **ntsolv** trajectory.

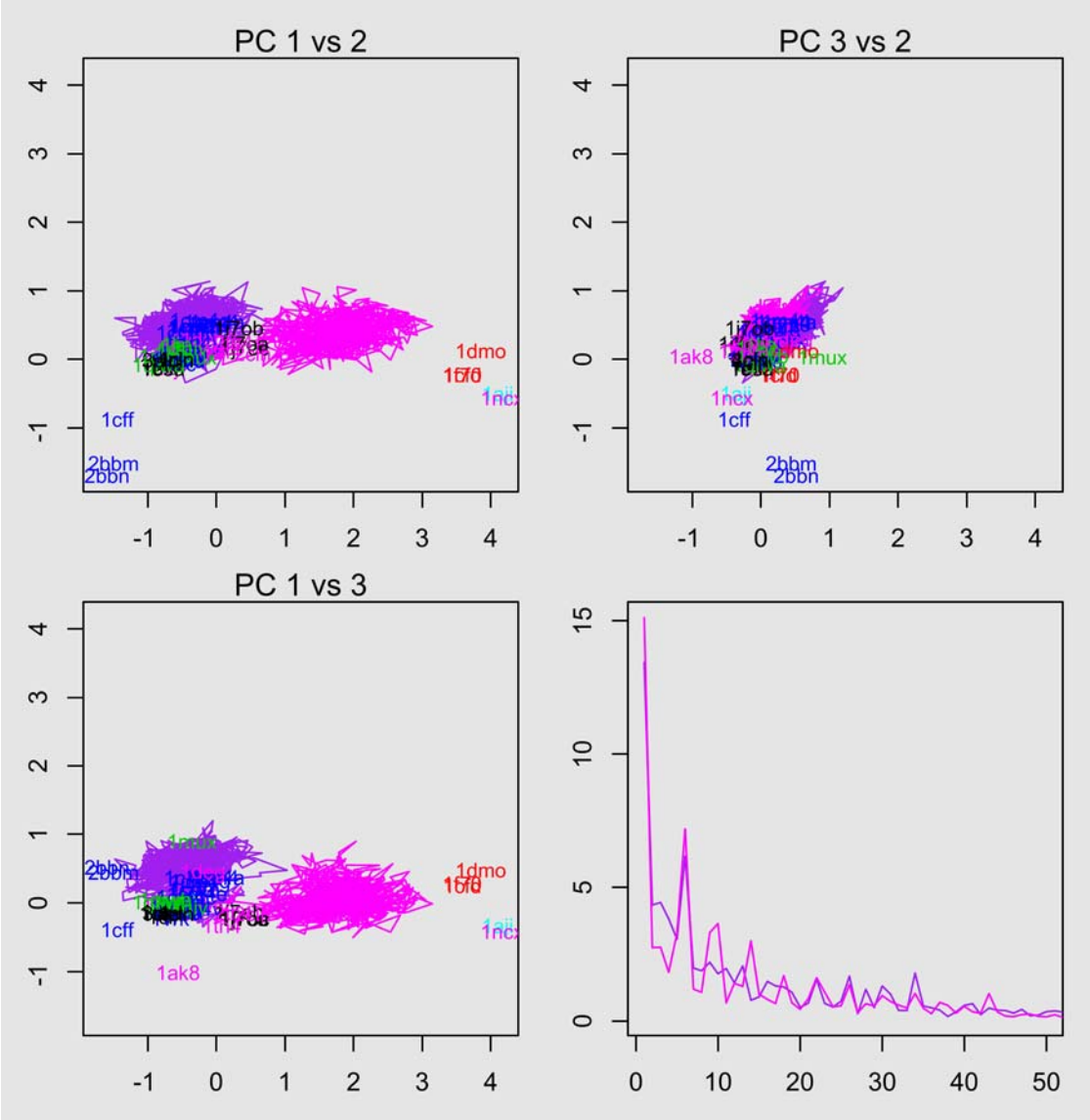


Figure 5.9: *Conformational subspace projection of MD trajectories for the N-term onto the first three principal components of the observed structures. The trajectory* ***ntpar27*** *is shown in purple and the* Ca^{2+}*-free trajectory* ***ntp27apo*** *is shown in magenta. The fourth plot of the panel shows the percentage of total variance for a given trajectory described by the top 50 PC's for the observed structures.*

These simulations show an entirely different behaviour to the PARAM19 results above. PCA of the full 10ns of the **ntpar27** simulation does not identify more than one distinct conformation unlike the **ntsolv** simulations above that are shown to explore multiple conformations over the course of the trajectory. This observation suggests there is a

stable conformation for the system which is similar to the experimentally determined structures for Ca^{2+}-CaM. These simulations also show a marked effect upon removal of the Ca^{2+} ions from the system. Loss of these ions immediately induces a conformational change resulting in a structure which shows greater similarity to that of Ca^{2+}-free CaM than the Ca^{2+}-ligated structure from the view in **Figure 5.9** above. The RMSD from the Ca^{2+}-free structure **1dmo** is variable throughout this simulation but decreases from 1.4Å at the beginning of the trajectory to a final conformation only 0.9Å away from the observed apo conformation.

- **Ca^{2+}-binding behaviour**

The analysis of the Ca^{2+}-binding distances over the course of the **ntpar27** trajectory shows the Ca^{2+}coordinating ligands to be well maintained. **Figure 5.10** below shows some minor variations in the Asp58 residue coordination throughout the first 6ns of the trajectory. Restoration of this interaction however seems to cause a significant effect on the Asn60 coordination which is considerably affected. These changes in coordination do not correspond to any significant changes in the global conformation indicated by the RMSD progression along the trajectory. Distances for the other sidechains, including all the Ca^{2+} coordinating residues in the first EF-hand, are well maintained throughout the simulation.

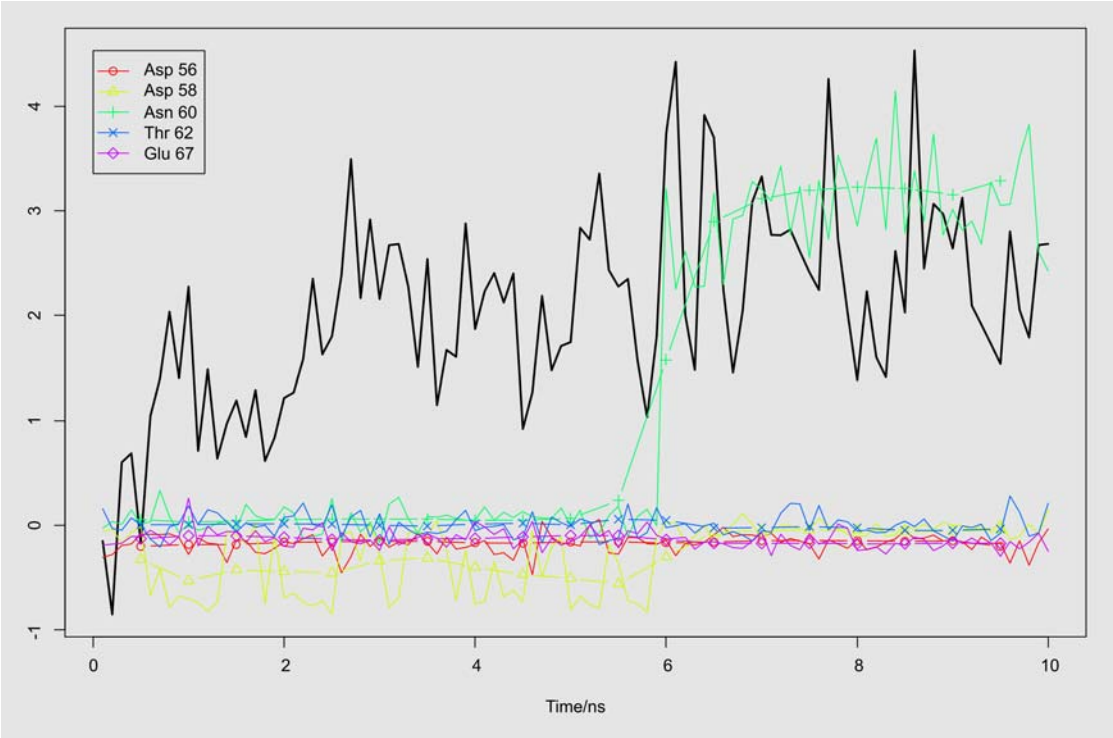


Figure 5.10: *Deviation in Ca^{2+}-binding distances in Å for the second EF-hand during the simulation* ***ntpar27****. Actual differences relative to reference values are shown in solid lines with 500ps floating averages plotted by points. The overall RMSD from the crystal structure for the trajectory is shown in black.*

5.3.4 C-Term Simulations Overview

The details of the different 10ns MD simulations performed on the high resolution X-ray structure **1exr** are shown above (**Table 5.1**). The time dependence of the RMS deviation from the crystal structure for the different simulations are shown in (**Figure 5.11**). Conformational changes are seen only a few ns into the **ctsolv**, **ctroc** and **ctapo** trajectories. The **ctpar27** and **ctcomp** simulations however are relatively stable throughout the observed time period. The scale of the changes observed in these simulations is less than the 2.0Å observed for the N-term, but remain significant. PCA was used to help comprehend the characteristics of the conformational changes in the simulations and the Ca^{2+} coordination analysis was used to help understand the Ca^{2+}-dependency of the system.

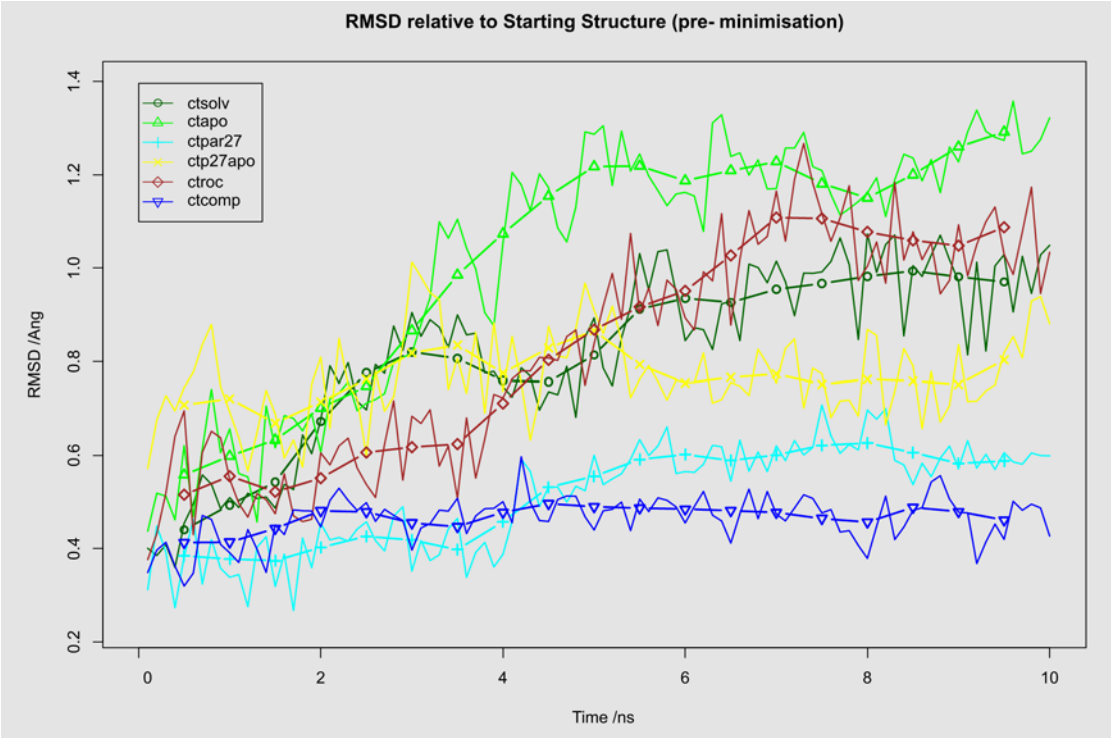


Figure 5.11: *Time dependence of RMSD relative to* ***1exr*** *for the different 10ns MD trajectories of the C-term lobes of CaM. Codes in the legend correspond to those outlined in* ***Table 5.1****.* ***NB*** *the abscissa scale is not the same as that used for the N-term lobe in* ***Figure 5.3****.*

5.3.5 C-term PARAM19 Simulations

Principal component analysis of the **ctsolv** trajectory suggested that several different conformations were explored over the course of the 10ns simulation. The conformational subspace projections below show both the trajectory and the experimentally observed C-term structures projected onto the first three PC's of the MD trajectory.

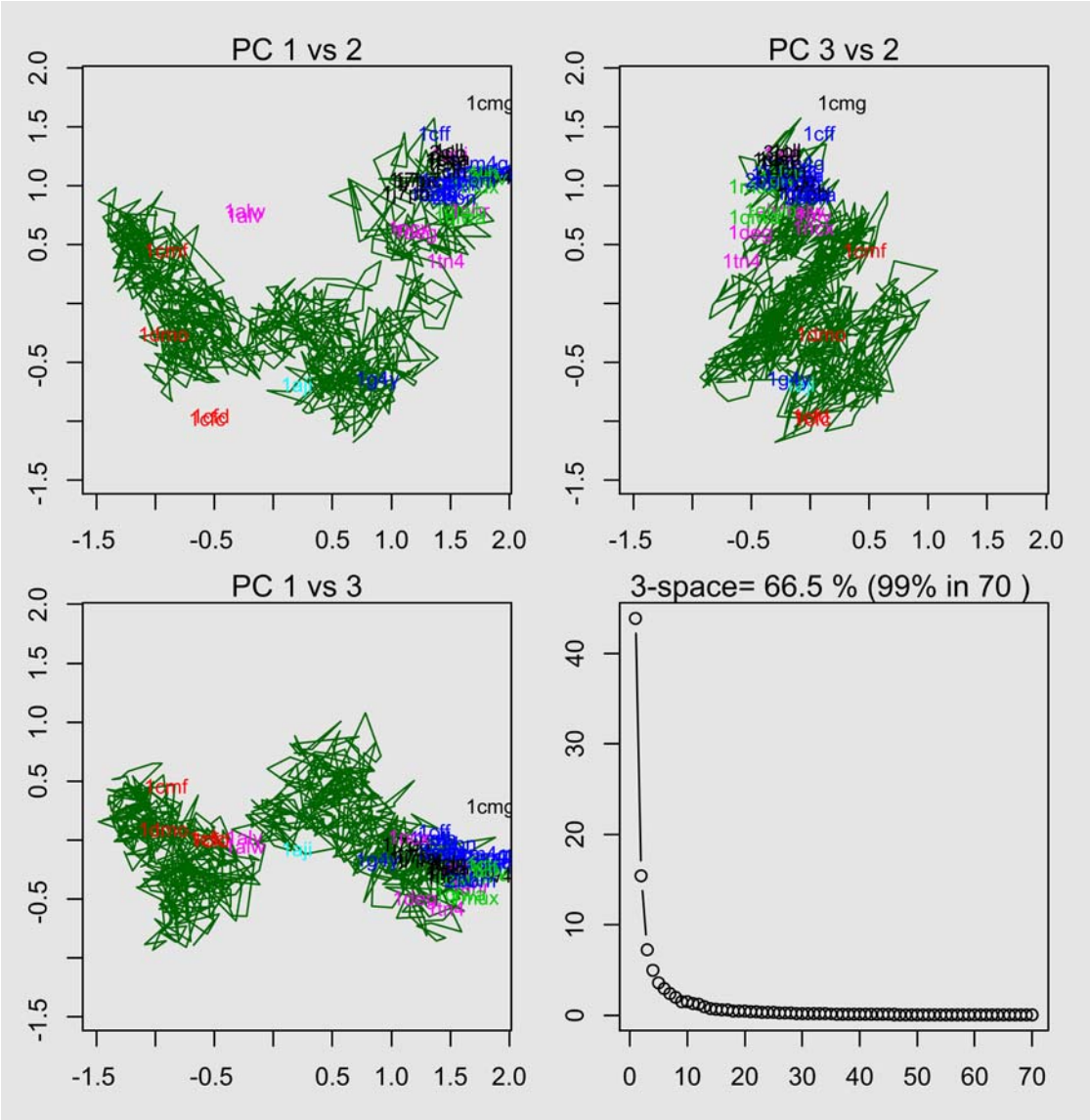


Figure 5.12: *Conformational subspace projection of the 10ns Ca^{2+}-CaM MD trajectory for the C-term onto the first three principal components of the trajectory. The trajectory* ***ctsolv*** *is shown in dark green.*

The trajectory is seen to explore at least three distinct conformational regions over this time period. The projection of the experimentally observed structures for CaM indicates that this initial conformation is similar to that observed for Ca^{2+}-CaM under various different target association states. A second conformation is explored during the 3-5ns time frame before the final conformation is adopted some 6 ns into the simulation. This final conformation again seems to correspond well to the observed structures for Ca^{2+}-free CaM when viewed in this subspace. Projection of the **ctsolv** and **ctapo** onto the principal components of the distribution of observed structures is shown in **Figure 5.13** below.

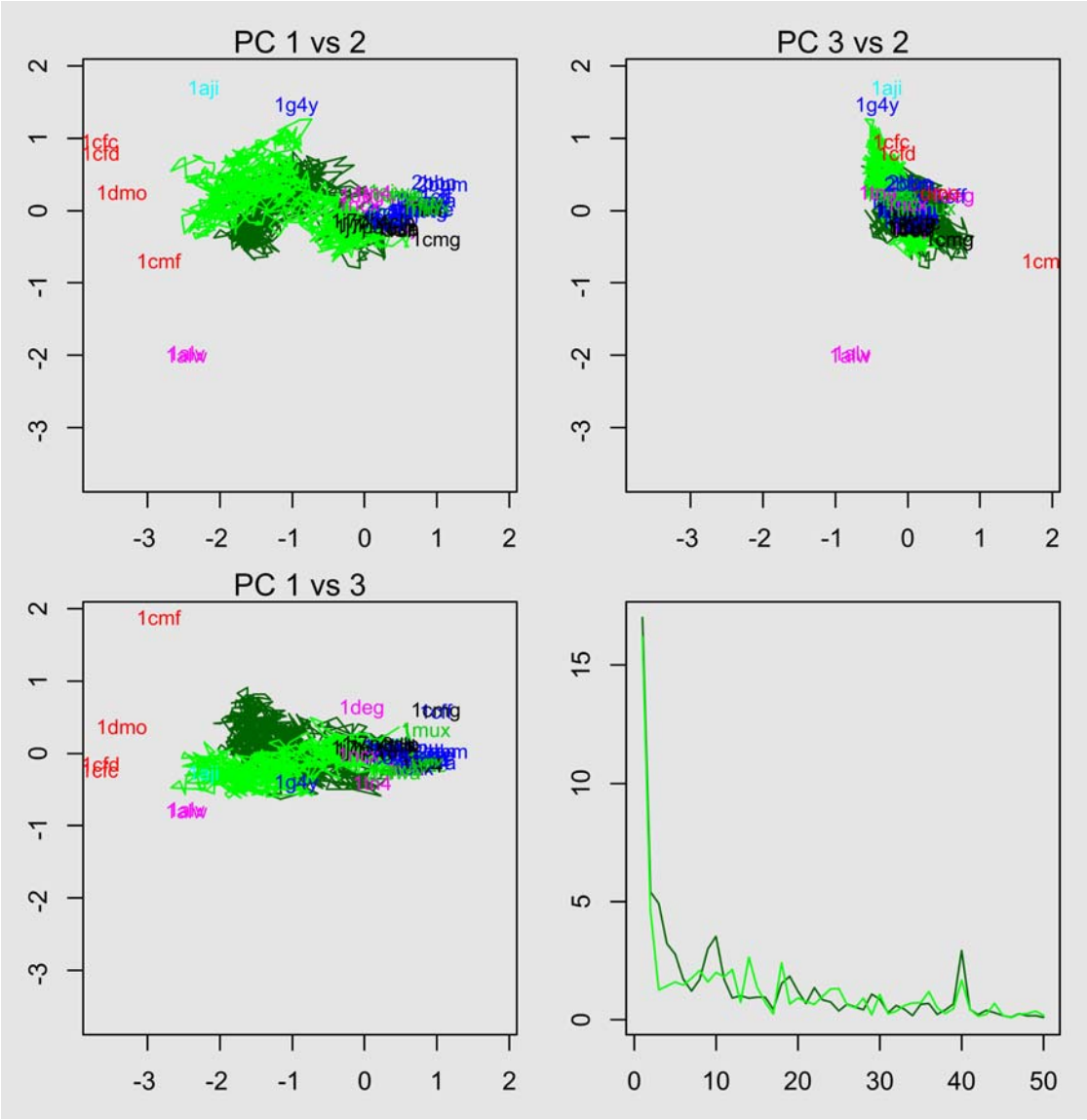


Figure 5.13: *Conformational subspace projection of MD trajectories for the C-term onto the first three principal components of the observed structures. The trajectory* ***ctsolv*** *is shown in dark green and the* Ca^{2+}*-free trajectory* ***ctapo*** *is shown in light green. The fourth plot of the panel shows the percentage of total variance for a given trajectory described by the top 50 PC's for the observed structures.*

The variance plot shows that PC1 of the observed distribution also describes 15% of the variance in both the apo and Ca^{2+}-bound trajectories and more than any other PC of the distribution. However, from this view it can be seen that neither the **ctsolv** nor the **ctapo** trajectory sample the sub-space defined by the observed Ca^{2+}-free structures. It is however clear from the different views of these trajectories that under these conditions,

the protein has a tendency to adopt a more closed apo-like conformation than has been observed in the crystal.

The repetition of this experiment on a different machine showed some even more extra-ordinary results. Performing the same simulation on different computer platforms using the same parameters, potential energy function and random number seed for the assignment of initial velocities would be expected to generate a slightly different trajectory due to differences in the assignment of initial velocities. The results obtained from this simulation however show changes in conformation in direct opposition to those obtained previously. From the projection of the trajectory and the experimentally observed structures onto the first three PC's of the dynamics trajectory **Figure 5.17**, the trajectory is seen to progress *away* from the apo conformation. Inspection of the different lobe conformations observed in this simulation confirm that the majority of frames do indeed exhibit a more open conformation for the protein than the crystal structure **1exr** used to initiate the simulation. This demonstrates a whole new region of conformational flexibility accessible to the protein which was unexpected. There are no experimentally determined structures in this region of conformational space.

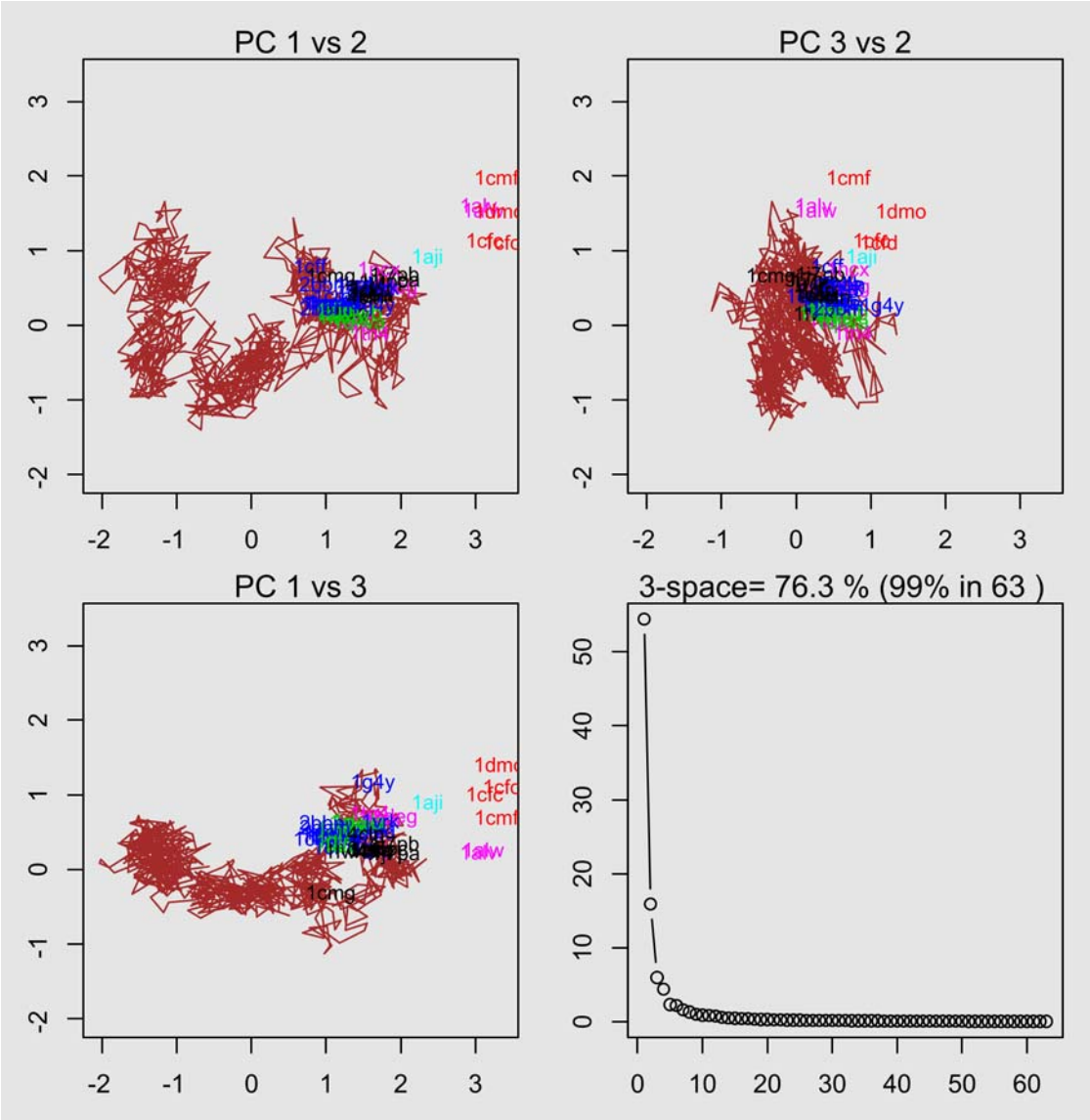


Figure 5.14: *Conformational subspace projection of the 10ns Ca^{2+}-CaM MD trajectory for the C-term onto the first three principal components of the trajectory. The trajectory* ***ctroc*** *is shown in brown.*

The final frame of the **ctroc** simulation is superposed onto structures for the Ca^{2+} bound structure **1exr** and the Ca^{2+}-free NMR structure **1dmo** is shown in **Figure 5.15** below. From this superposition it can clearly be seen that the main difference between the Ca^{2+} conformation of **1exr** and the structure generated by the simulation is the orientation of the C-term helix (front left hand side in this view). The rest of the lobe conformation is relatively similar to the experimental Ca^{2+}-bound conformation.

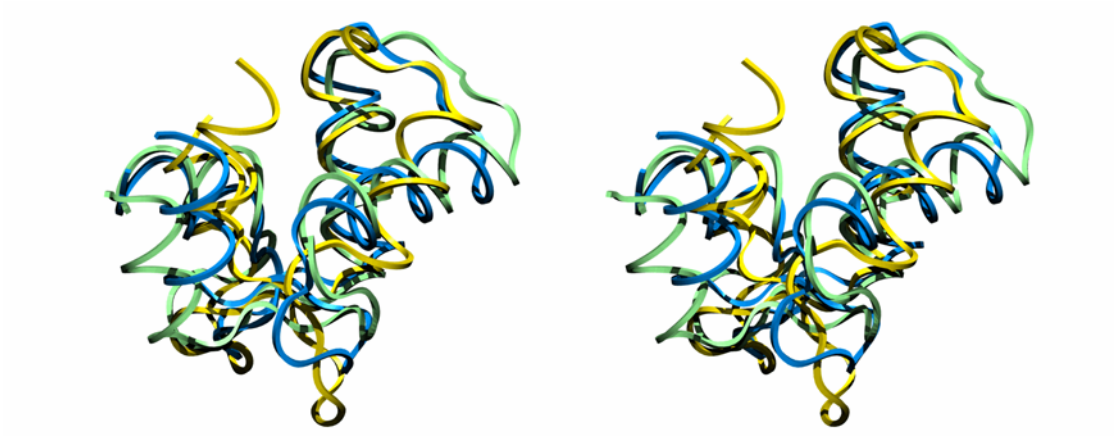


Figure 5.15: *Stereo structural superpositions of the CaM lobes at the end of the* ***ctroc*** *simulation. The open Ca^{2+}-bound structure of* ***1exr*** *is shown in blue and the apo structure* ***1dmo*** *is shown in yellow. The structure generated by the simulation is shown in green.*

- **Ca^{2+}-binding behaviour**

This analysis of the Ca^{2+}-binding behaviour over the course of the 10ns C-term simulation **ctsolv** shows that the conformational changes observed for this simulation have no direct connection with significant fluctuations in the Ca^{2+}-binding distances which are generally well maintained throughout this simulation. The loss of coordination by the backbone carbonyl of Leu99 occurs after ca. 6.5ns of simulation but is recovered less than 1ns later.

Analysis of the **ctroc** simulation indicates more erratic behaviour in the Ca^{2+}-binding sites of CaM, as shown in **Figure 5.16** below. Asn97 is clearly seen to fluctuate considerably throughout the course of the simulation as the reference atom moves up to 6Å further away from the Ca^{2+} than in the original crystal structure. This behaviour does not appear to be directly linked to the conformational changes exhibited by the system over the course of the simulation. Ca^{2+} coordination by Asp95 is also variable over the course of the simulation. Although the observed distances are close to the reference value for much of the time series, minor fluctuations are observed throughout the initial 5.5ns with a prolonged re-organisation of the group between 5.5 and 9ns. This looks to be a possible consequence of the conformational change taking place after 4ns. The distance between the reference atom on Asp93 and the Ca^{2+} also shows a gradual decrease resulting from the conformational changes in the protein fragment.

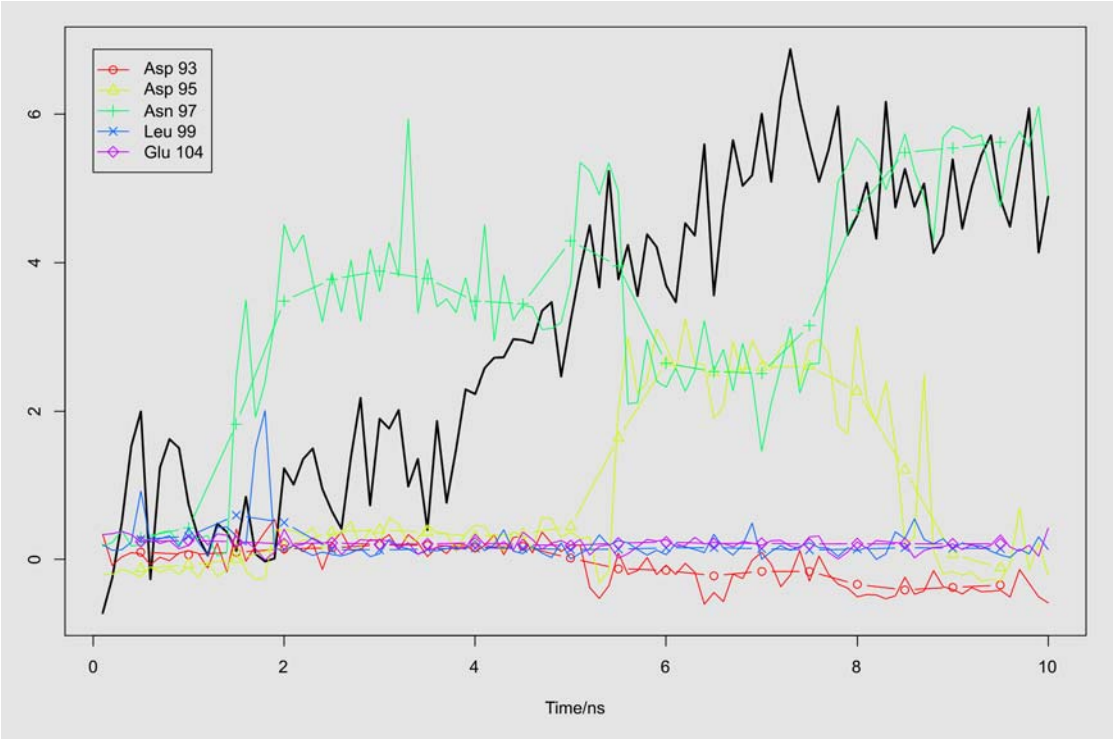


Figure 5.16: *Deviation in Ca^{2+}-binding distances in Å for the first EF III during the simulation* ***ctroc****. Actual differences relative to reference values are shown in solid lines with 500ps floating averages plotted by points. The overall RMSD from the crystal structure for the trajectory is shown in black.*

5.3.6 C-term PARAM27 Simulations

Simulations of the C-term fragments of the crystal structure **1exr** were performed in the presence and absence of Ca^{2+} using the all-atom PARAM27 potential with interesting results. Projection of the structures generated from the **ctpar27** run onto the PC's for the experimentally observed structures suggest a very tight distribution of conformations about the Ca^{2+}-bound conformation for CaM. PCA of the trajectory itself however highlighted two distinct conformations for the protein as illustrated below (**Figure 5.17**). For the first few nanoseconds the protein explores only the conformation observed in all the Ca^{2+}-bound structures of CaM. The fragment then undergoes a rapid transition to a different conformation that is retained throughout the subsequent 7ns of simulation. The scale of this conformational change is relatively small with only a 0.2Å increase in RMSD relative to the initial crystal structure.

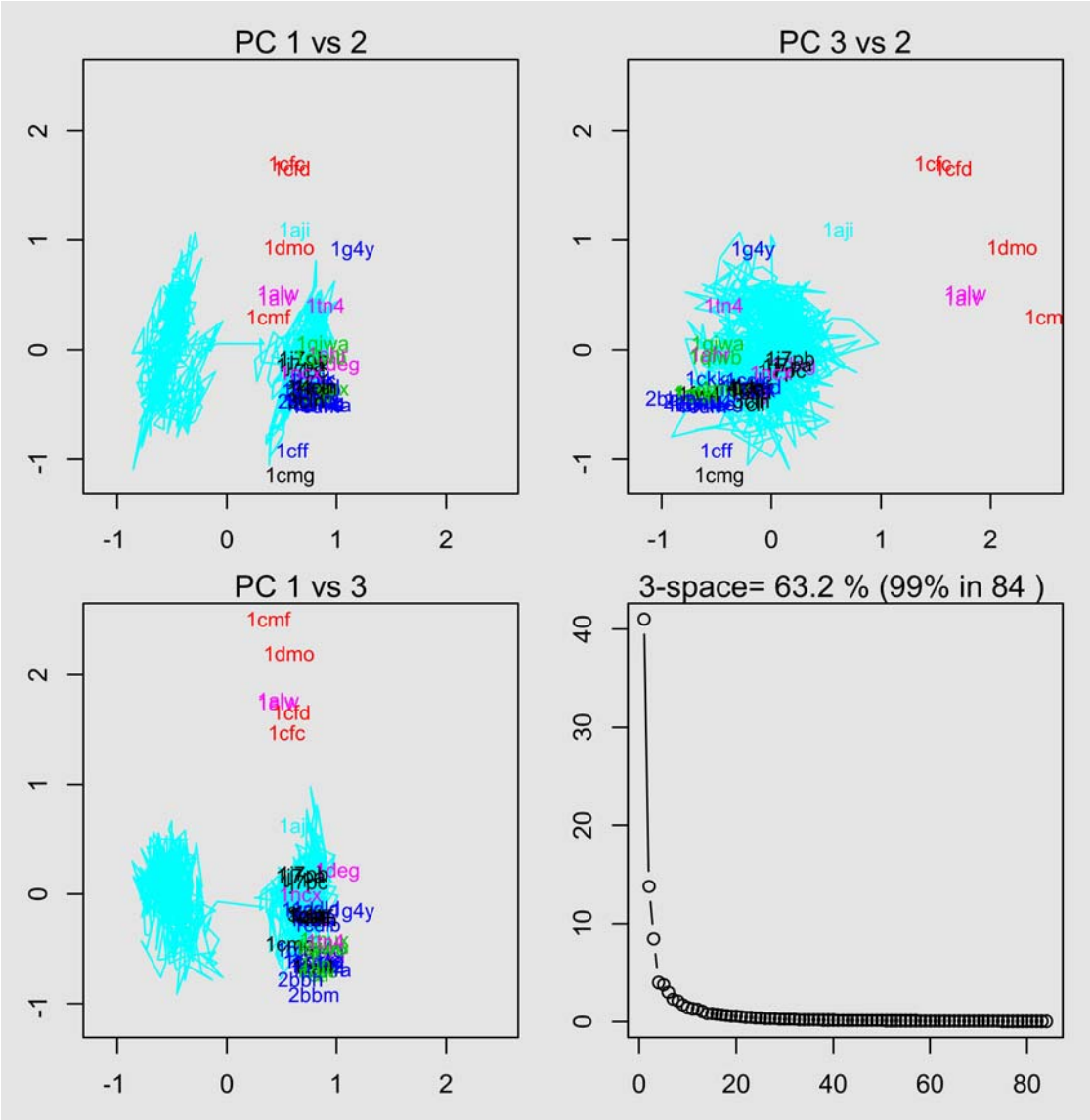


Figure 5.17: *Conformational subspace projection of the 10ns Ca^{2+}-CaM MD trajectory for the C-term,* ***ctpar27,*** *onto the first three principal components of the trajectory. Experimentally observed structures are also projected onto the same subspace.*

The projection of the trajectory onto the PC's of the observed structures indicate that the conformational change is not related to any of the other recognised modes of flexibility which have already been identified. Analysis of the Ca^{2+}-binding distances for the simulation show that there are no Ca^{2+} coordination changes that induce the change in conformation although the Asn97 group alters in orientation approximately 1ns afterwards.

Comparison of the Ca^{2+}-bound and apo trajectories when projected onto the PC's for the observed structures indicates that the Ca^{2+} removal allows for considerably more

flexibility in the system. Close inspection of the RMSD time series indicates a rapid response to the Ca^{2+} deletion during the first 200ps of the simulation. There are several minor fluctuations throughout the trajectory although the RMSD is relatively stable at around 0.7Å away from the crystal structure. PCA of the trajectory itself suggests the exploration of several different conformers over the 10ns but these do not correspond to any of the observed structures.

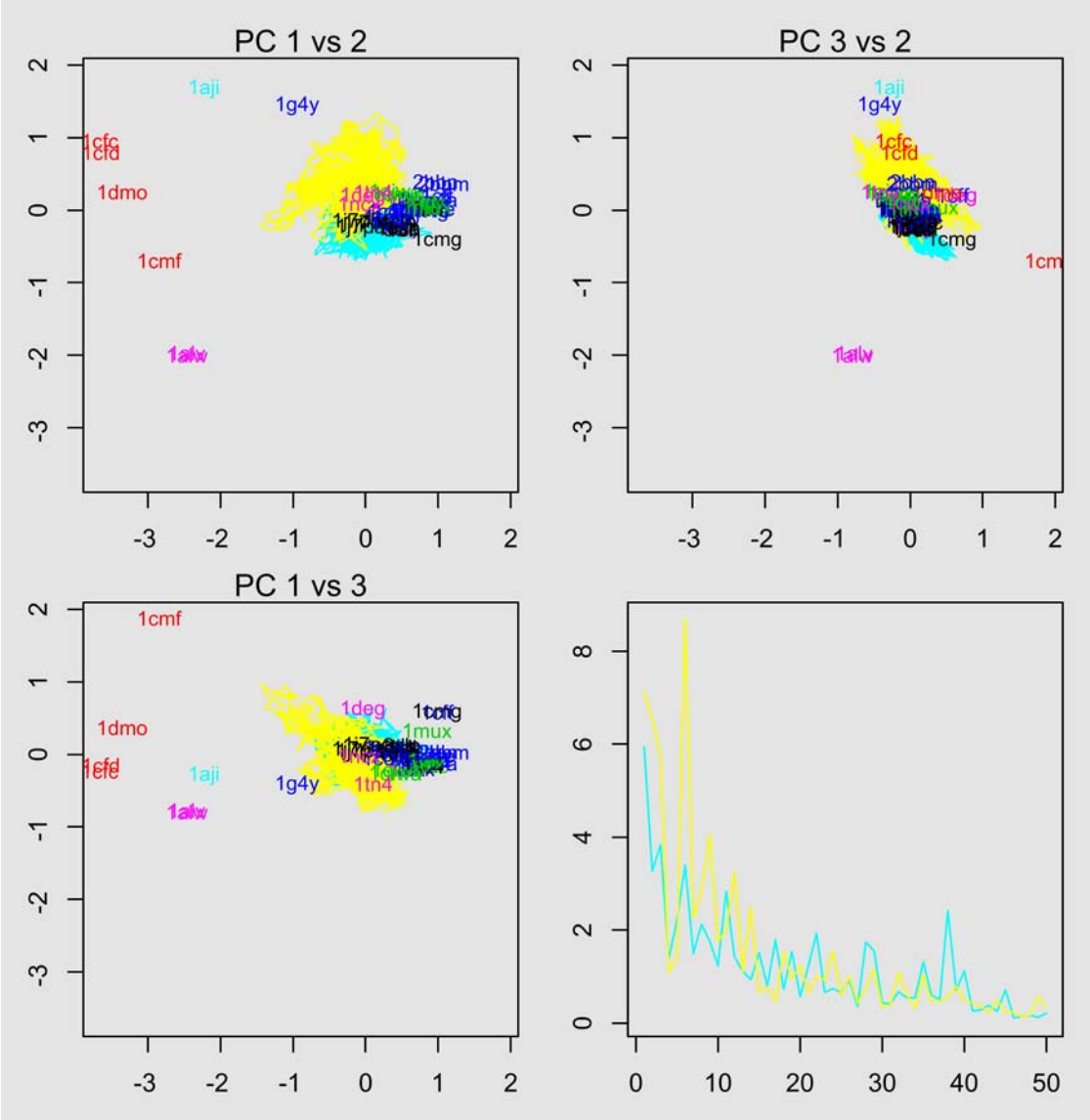


Figure 5.18: *Conformational subspace projection of MD trajectories for the C-term onto the first three principal components of the observed structures. The trajectory* ***ctpar27*** *is shown in cyan and the* Ca^{2+}*-free trajectory* ***ctp27apo*** *is shown in yellow. The fourth plot of the panel shows the percentage of total variance for a given trajectory described by the top 50 PC's for the observed structures.*

5.4 Discussion of Results

5.4.1 CaM flexibility

The picture that emerges from these simulations is one of considerable conformational flexibility in CaM. The range of different conformations adopted by both lobes of CaM in the 11 different simulations is unprecedented. These investigations represent a significant exploration of the CaM conformational space although the diverse characteristics of these simulations present significant difficulties for the identification of characteristic modes of flexibility in the CaM lobes.

In addition to the RMSD time series plots shown in the previous section, RMS atomic fluctuation analysis can be used to compare relative flexibility within the structure over the course of the simulation. After allowing a 2ns equilibration period for each simulation, **Figure 5.19** shows the averaged fluctuations for the different data sets over the time period 2 to 10ns relative to the mean structure of the trajectory over this timeframe. This analysis provides a different description of the variation over the course of the trajectory by detailing the variation inherent to the trajectory rather than relative to the initial conformation.

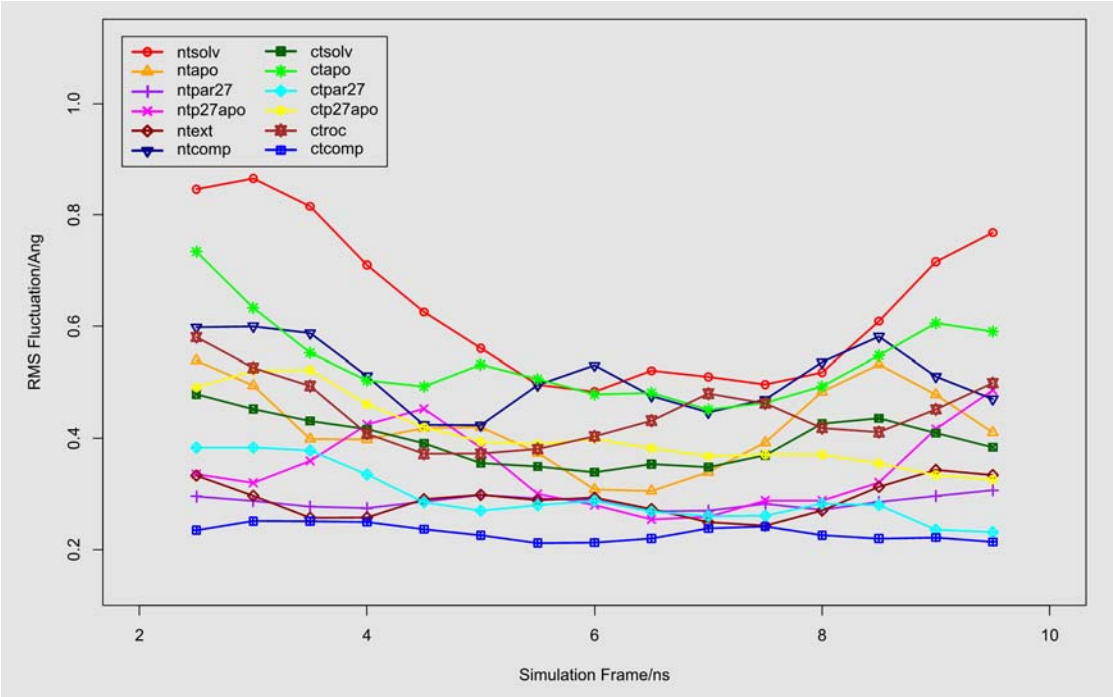


Figure 5.19: *RMS Fluctuations for the simulations over the 2-10ns time period. Values are calculated over 100ps time frames.*

There is no clear distinction in these plots between N and C-term simulations, however the most variable trajectory is for the N-term (**ntsolv**) and the most stable system is the C-term (**ctcomp**). With the exception of the simulation of the CaM-target complex and the extended N-term simulation (**ntext**), the PARAM27 simulations demonstrate lower magnitude fluctuations than those using PARAM19. Simulations of Ca^{2+}-free CaM are also shown to be more variable than those of Ca^{2+}-bound CaM when using the PARAM27 potential. This supports the principal of a structural role for Ca^{2+} in CaM structure.

Perhaps the most surprising results here are the trajectories for **ntpar27** and **ntp27apo** both of which are highly stable according to the RMSF plots. The N-term is thought to provide much of the flexibility in the system which suggests that it should be able to show the kind of flexibility seen at least in the **ntcomp** simulation. The simulations performed here, however, illustrate a highly stable system which retains a single conformation throughout the simulation which deviates only 0.3Å from an average structure over the whole 10ns of simulation.

Comparison of **ntcomp** and **ctcomp** indicate the type behaviour expected of the classical view of CaM target interaction. Here the C-term of the protein forms a tight and relatively inflexible conformation with the protein which corresponds to the high affinity interaction commonly observed between the protein and CaM binding domains. Similarly the N-term lobe of CaM, which is known to interact more weakly with the targets of the protein is shown to exhibit considerable conformational flexibility during the simulation.

5.4.2 Modes of Flexibility

Having compared the general characteristics of all the different simulations, it is also of interest to attempt to identify different conformations explored in the various different simulation. To do this, a PCA was performed on conformers generated at frame intervals of 10ps from all of the different experiments together. The projections of these trajectories onto the space described by the first 3 PC's are shown below in **Figure 5.22**. The eigen spectrum and variance plots shown in **Figure 5.20** clearly show that not all trajectories contribute equally to the dimensions of the principal conformational subspace sampled over all these simulations.

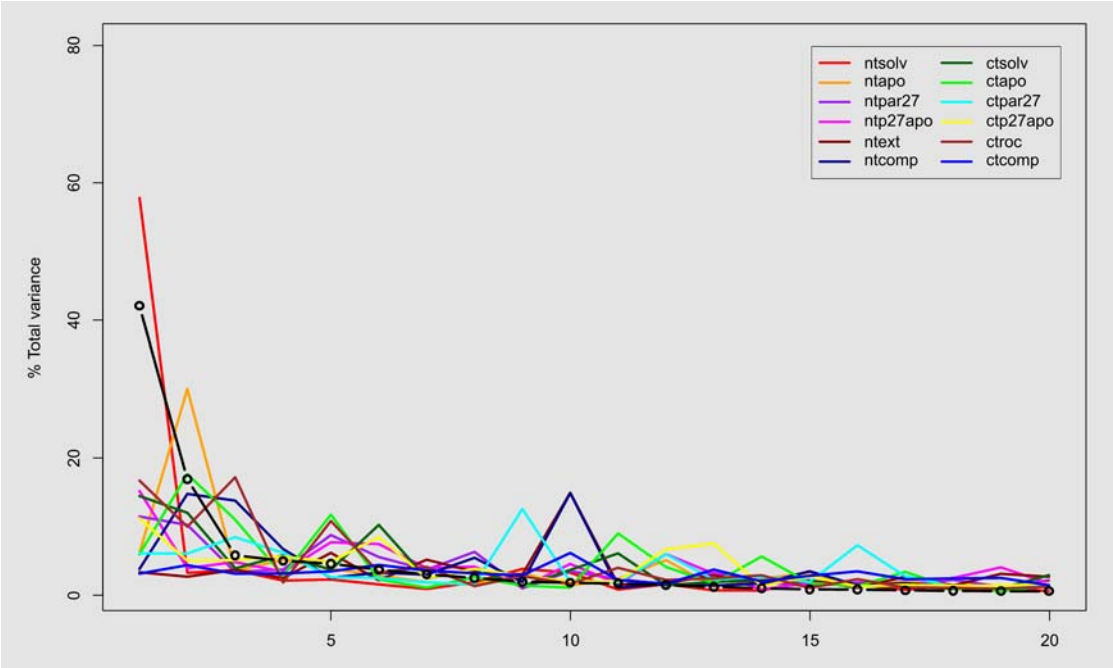


Figure 5.20: *Variance spectra for the first 20 PC's of the distribution. The over all eigen spectrum is shown in black with circles, percentage variances for each of the individual simulations are shown according to the colours in the legend.*

With only 64% of the total variance described by the first 3 PC's, this eigen spectrum is somewhat flatter than those of the observed distributions indicating a larger number of modes of flexibility than have been observed experimentally. Nearly 60% of the significant variance in the **ntsolv** trajectory, however, is described by PC1 although this PC only describes 40% of the total variance across all the trajectories indicating a significant bias. No other trajectory is so well described by any of the other PC's with 30% of the variance in **ntapo** described by PC2 and just under 20% of **ctroc** described by PC3.

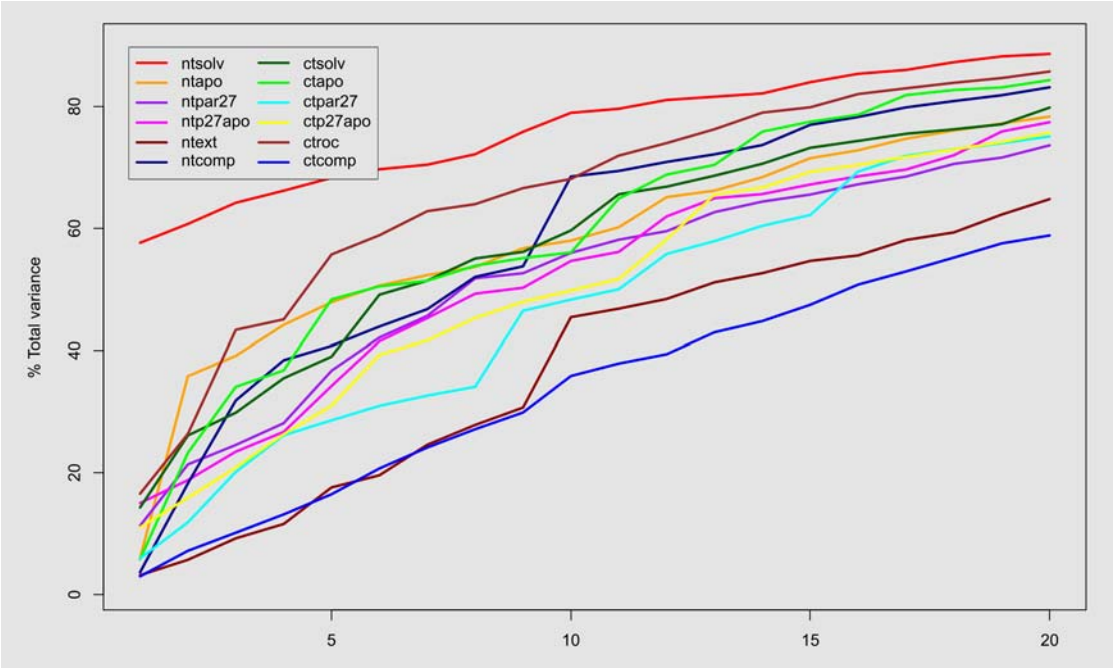


Figure 5.21: *Cumulative variances for each of the simulations for the first 20 PC's of the distribution of conformers generated by all 11 simulations.*

The cumulative variance shows a better separation of the different trajectories in **Figure 5.21**. From this plot it might be expected to see certain PC's to be more characteristic of variance in the N-term and one for the C-term but this is not the case. At this level of detail the two lobes of CaM do not appear to present different modes of flexibility. It would have been intriguing to have identified different orthogonal modes of flexibility for the two CaM lobes and would go some way towards the identification of different functional roles for the domains. Since this is not the case, it must be concluded that the conformational space accessible to both lobes of the protein is effectively shared with both parts of the protein able to access structurally similar conformations.

The conformational subspace projections of these trajectories support these observations. The sampling of the conformational space appears to be more extensive for the N-term of the protein, particularly in the **ntsolv** trajectory, which explores the most diverse range of conformations of the simulations performed here.

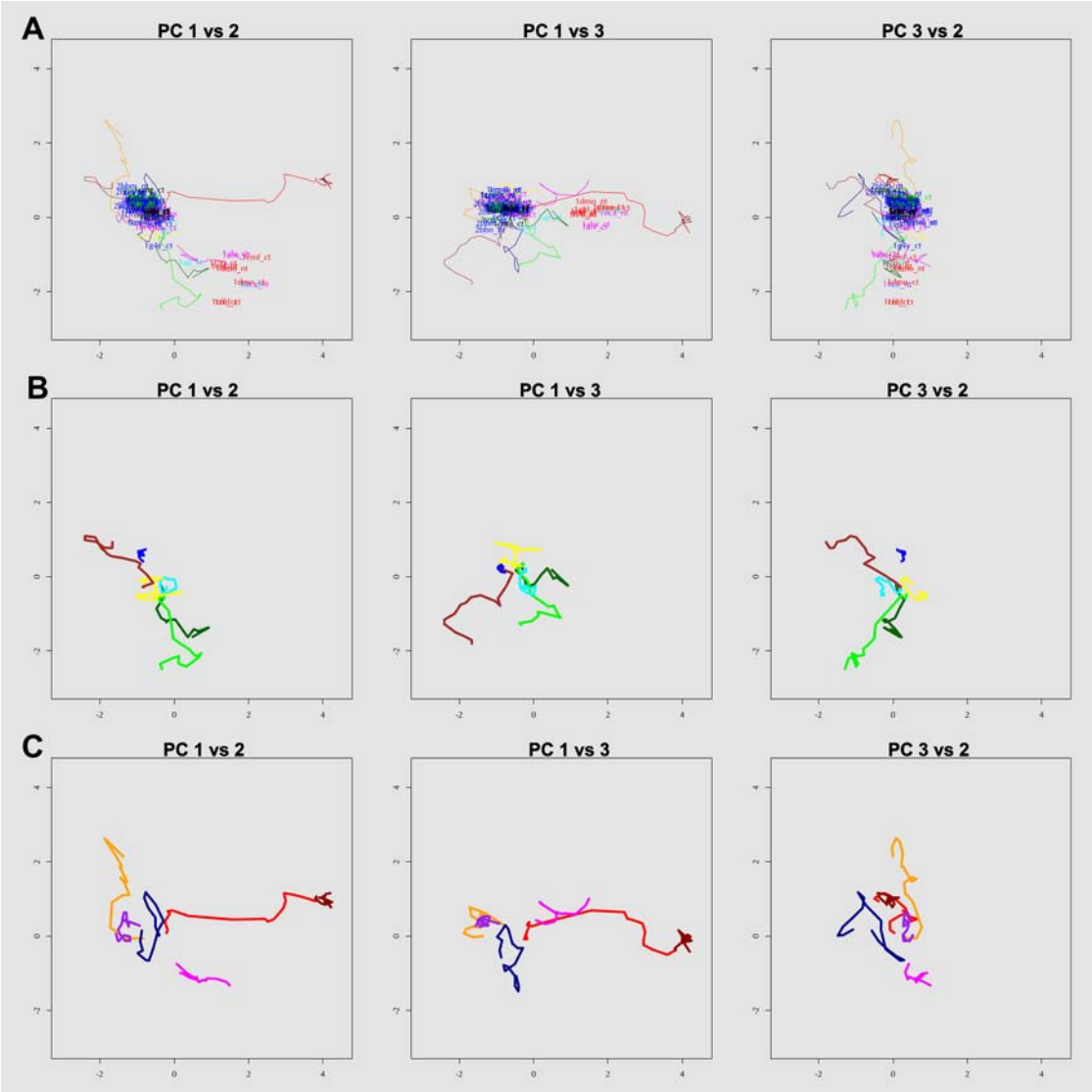


Figure 5.22: *Conformational subspace projections of the observed structures and frames from all the 10ns trajectories onto to the first 3 PC's of the distribution. Panel* ***A*** *shows projections for the observed structures and all 11 trajectories, Panel* ***B*** *shows the C-term simulations and Panel* ***C*** *the N-term simulations.*

5.4.3 Structural role of Calcium in CaM

The collection of experimentally determined structures of CaM clearly indicate that the protein has a conformation which is dependent on the Ca^{2+}-binding state of the protein. Although no crystal structures of Ca^{2+}-free (apo) CaM have been determined, the evidence strongly suggests that the apo conformation is far less open than the Ca^{2+}-bound form of the protein. The experimental studies of Evenäs *et al.* [Evenas *et al.* 1999]

and the work of Garcia *et al.* [Vigil *et al.* 2001] however have indicated the possibility of the CaM lobes existing in dynamic equilibrium between the open and closed conformations with rapid exchange between them.

Both the **ctsolv** and **ntsolv** trajectories support the idea of a dynamical exchange between open and closed conformations for the CaM lobes in solution conditions with the **ntsolv** trajectory demonstrating that the Ca^{2+} can dissociate from one of the EF-hands. Removal of calcium from these systems should therefore result in a response from the system and a conversion to something resembling the Ca^{2+}-free conformation. This is clearly not observed in these investigations. The response to Ca^{2+} deletion is more apparent in the PARAM27 simulations and both the Ca^{2+}-free trajectories **ntp27apo** and **ctp27apo** do tend towards a more closed and apo like conformation than the Ca^{2+}-bound counterparts. The conservation of conformation observed in the **ntext** trajectory, suggests that the conformational change seen in **ntsolv** is non-reversible. It is however possible that the reverse transition towards a Ca^{2+}-bound open state requires the re-association of Ca^{2+} and would therefore be a diffusion-controlled process on a longer timescale than can currently be explored by conventional MD simulation.

The use of the all-atom PARAM27 potential results in what appear to be more stable trajectories for both the N and C-term lobes of the protein. A possible explanation for this observation is that the improved force field delays the conformational changes seen with the PARAM19 simulations. Propagation of these trajectories on a longer timescale may reveal similar conformations to those already seen with the less sophisticated potentials.

It has been suggested by Martin *et al.* [Martin *et al.* 1996] that the interaction of CaM with its targets can increase the Ca^{2+} affinity of the system. This finding correlates with the observation in the **ntsolv** trajectory that the conformational change takes place prior to the dissociation of Ca^{2+}. The simulations of the CaM-target complex indicate the structure is more conserved under these conditions which helps to preserve the organisation of Ca^{2+}-binding sidechains in the EF-hands.

5.5 General Conclusions

It is clear from these simulations that the range of conformations accessible to the two lobes of CaM is by no means limited to a simple open and closed conformation for each lobe. The magnitude and complexity of the conformational changes in this system are pushing the boundaries of current MD simulation techniques. A better understanding of the true structure of CaM in solution and its potential for exchange between different conformations requires considerably more data than has been generated in this study. The analysis of 11 different 10ns Molecular Dynamics trajectories performed here has revealed a high degree of conformational flexibility which is potentially the key to the ability of the protein to accommodate such a diverse range of targets.

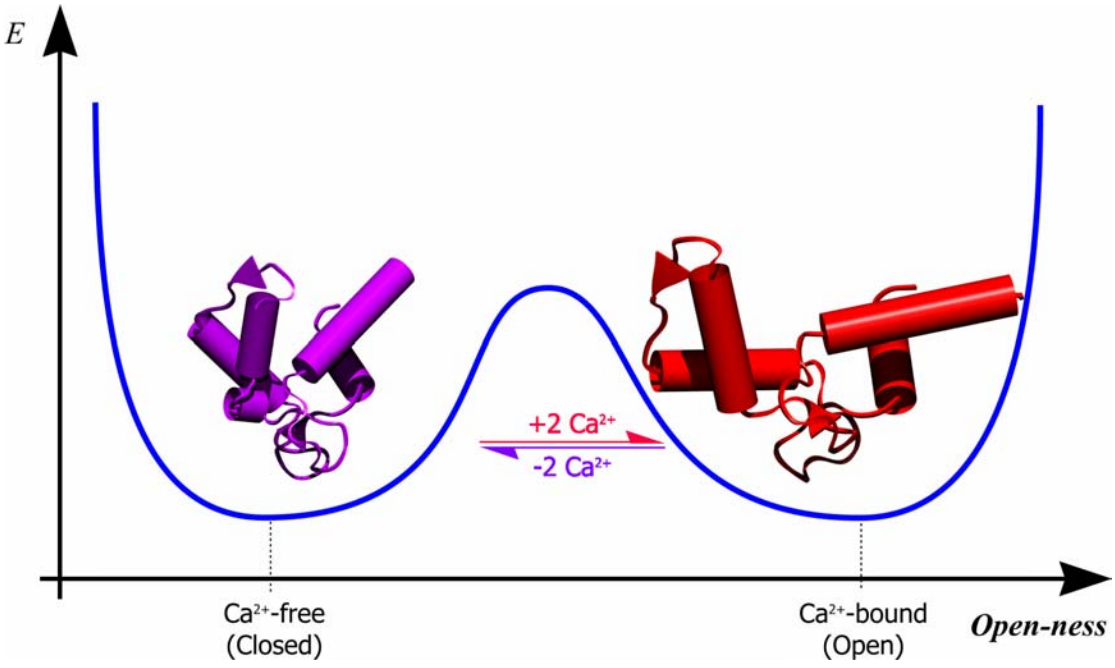


Figure 5.23: *Potential energy landscape schematic illustrating the classical view of two different CaM lobe conformations, closed and open.*

The range of conformers generated over the course of these simulations is suggestive of an expansive conformational landscape for both lobes of the protein which is accessible under physiological conditions. The X-ray and NMR structures currently available for CaM in the presence and absence of Ca^{2+} and various different co-factors has largely resulted in the identification of two main conformations, open (Ca^{2+}-bound) and closed (Ca^{2+}-free). These two states, and the possibility of inter-conversion between them were

largely thought to describe the structure of the protein under different conditions as shown in **Figure 5.23**. The studies conducted here, however, suggest that the range of potential conformations for the system is much greater as suggested in the schematic of **Figure 5.24** below.

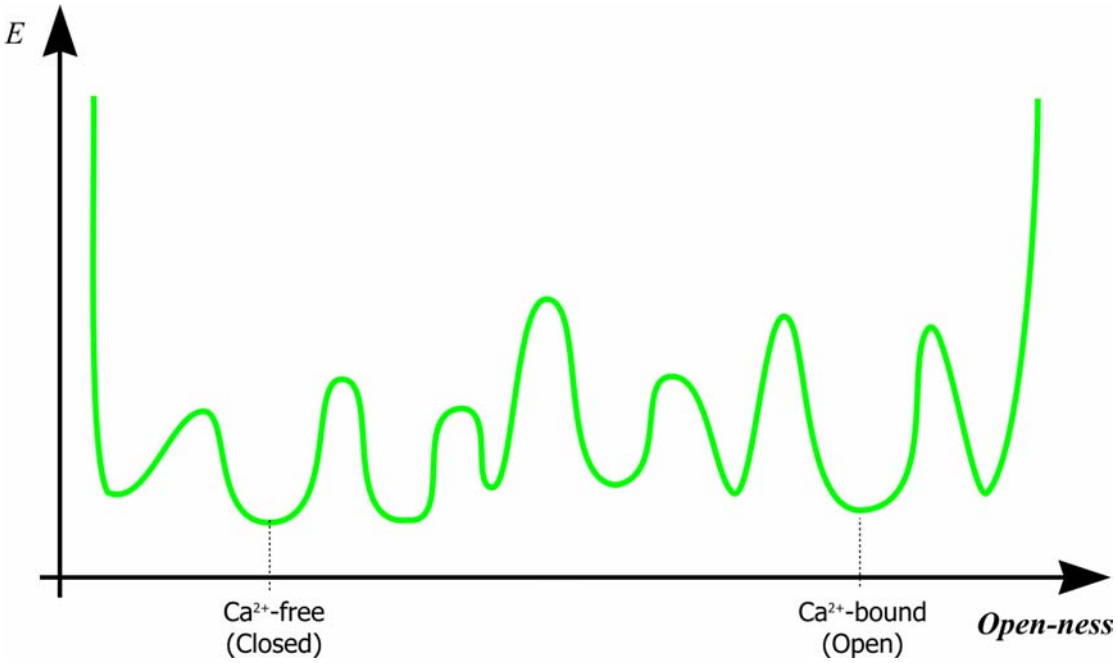


Figure 5.24: *Potential energy landscape schematic illustrating the multi-minima situation proposed from the results of these simulations. The range of barrier heights shown in this illustration is representative of the sensitivity of the protein to inter-conversion between different conformations on the Ca^{2+} concentrations and other physiological and environmental factors on the system.*

The sensitivity to initial conditions observed in these investigations also suggests considerable adaptability in the system. The simulations presented here show CaM to have the ability to either exchange between different conformers or retain the original conformation. This suggests that the energy barrier to inter-conversion between different conformers is critically poised and highly tuneable to the environment in which the protein is operating. The functional implications of such a system are clear; not only is the protein sensitive to critical changes in Ca^{2+}-concentration in its environment but the presence or absence of other effectors such as co-factors, temperature and pH could result in the subtle adaptation of the energy barriers between significantly different conformations.

It has not been possible to use these results to conclusively answer the question regarding the true conformation of CaM in solution. Rather, the results presented here suggest that future experimental structure determinations of the protein will yield an even greater range of conformations for the system which are largely dependent on the experimental conditions. The relatively conserved structures observed by X-ray crystallography and studied in detail previously (see Chapter 3) showed target dependent variations around a conserved conformation which could be supported by the contacts resulting from the formation of a crystal lattice. The NMR structures representative of structural ensembles show more variation in conformation although there are noticeable distinctions between Ca^{2+}-bound and Ca^{2+}-free conformations. The recently published NMR RDC structures however show relatively subtle differences to the crystal structure conformations. The authors strongly believe that the experimental conditions used to align the protein fragments do not significantly influence the conformation observed. The additional conformations generated from the MD simulations support the suggestion that X-ray crystallography presents a somewhat constrained view of CaM conformation. The variability observed across all the simulations however is not helpful in the identification of a *true* conformation (or range of conformations) for the protein fragments in solution.

These simulations have also yielded some interesting results regarding the Ca^{2+}-dependency of CaM conformation. Whilst the Ca^{2+} coordination through the bi-dentate Glu residues appears to be closely coupled to the conformational change in **ntsolv**, the loss of coordination after the conformational change event suggests that this is consequential rather than causative. If this is a true representation of the situation then alternative causes for the significant effects on the conformation of the protein fragment must be sought. Interestingly, there is no loss of Ca^{2+} coordination in the second EF-hand of the N-term lobe over this period or indeed during the subsequent 10ns of simulation in **ntext**. Such dramatic effects on Ca^{2+} coordination are not observed in any of the other simulations, which in turn suggests that the computational model is treating the Ca^{2+} ions appropriately.

Protein dynamics simulations are commonly directed towards the understanding of interactions between proteins and small molecules and seek to characterise small changes in protein sidechain orientation and key interactions. The enormous range of conformational variation observed in these experiments has shown that this system

exhibits too much flexibility to characterise fully at this time (even before the consideration of the role of the central tether in CaM function). The findings of Caves *et al.* [Caves *et al.* 1998] which suggested that many different trajectories for an accurate characterisation of the dynamics of a system, are entirely in tune with these findings. In the case of CaM however it is clear that the conformational exchange processes of particular interest require tens if not hundreds of nano-seconds of simulation to further characterise.

Chapter 6: Entropic Analyses of CaM Lobes[6]

6.1 Introduction

The spatial separation of the CaM lobes in the target free state is a likely indicator of the design of a two component system. Although a large degree of cooperativity in Ca^{2+}-binding is observed *within* each CaM domain, the differences in Ca^{2+} affinities *between* the two domains result in largely independent activation of targets at different Ca^{2+} concentrations [Bayley *et al.* 1996]. Despite the high sequence homology observed for the two domains of CaM it seems clear that rather than being ambidextrous, the N and C domains of CaM have different characteristics and thus different roles.

CaM is a highly flexible system. NMR studies have indicated that Ca^{2+} -free CaM exhibits considerably more conformational sampling than the Ca^{2+}-bound form of the protein [Zhang & Yuan 1998] and that the backbone flexibility of the system is further reduced by the presence of a target peptide. Experimental structure determination of CaM is complicated by the dynamic properties of the system which contribute to conformational averaging (in the case of NMR) and non-resolvable regions (in the crystallographic electron density). This is particularly apparent at the N-terminus of the protein and in the central linker. Such flexibility is characteristic of signalling proteins which generally make low-affinity, low specificity interactions, in contrast to catalytic systems which are more conserved and better ordered in crystal structures [Dunker *et al.* 2002]. Even in an atomic resolution (d_{min}=1.0Å) low temperature crystal structure of CaM [Wilson & Brunger 2000], significant disorder was observed throughout the protein and the anisotropic B-factors revealed scissor-type opening/closing motions within each of the EF-hands.

Comparison of the experimentally determined structures and simulated trajectories in the previous chapter suggest greater conformational flexibility for the N-domain of the protein compared to the C-domain. This result is of interest in the context of experimental observations that the C-domain of CaM has a greater affinity for Ca^{2+} and

[6] This chapter has been revised from a paper currently awaiting publication in J. Phys Chem.

some CaM targets while the N-domain is less specific in its choice of target motif [Barth *et al.* 1998]. In this chapter this flexibility is quantified via estimation of the configurational entropy of the independent CaM lobes.

6.2 Materials & Methods

Simulations were carried out for the separate lobes of CaM using the 1.0Å crystal structure **1exr** [Wilson & Brunger 2000]. Fragments corresponding to the regions Asp2-Ala73 and Leu85-Ser147 were used for the N and C-term lobes respectively. For residues with multiple conformations, the first listed (or "A" conformer) was used. Additionally, computations on the CaM-CaMKIIα complex were carried out on the first model (chains A and B) of a 2.0Å crystal structure **1cm4**, [Wall *et al.* 1997]. The structure resolved in the experiment included residues Leu4-Thr146 of CaM and the peptide corresponding to a fragment of the CaM binding domain of CaMKIIα (residues Phe293-Thr310). The details of the MD simulations are outlined in the previous chapter.

The selected frames of the trajectories were reoriented relative to the initial minimised structure by least squares superposition [Kabsch 1976] to remove any global rotations/translations. Structures were then selected such that the RMS difference between consecutive selected structures was at least 0.13Å. The threshold was adapted in a manner analogous to that of Elber & Karplus [Elber & Karplus 1987] for each simulation to select approximately 100 structures from the 10ns trajectory to represent the entire conformational space explored. The trajectory names and summarised details are listed in **Table 6.1** below.

Table 6.1: *Summary of the 10ns MD simulations used for the NMA calculations.*

Code	Description			No. of atoms
ntsolv	N-term	Ca^{2+}-expulsion	PARAM19	4694
ntapo	N-term	Ca^{2+}-expulsion	PARAM19	4691
ctsolv	C-term	Ca^{2+}-bound	PARAM19	4214
ctapo	C-term	Ca^{2+}-expulsion	PARAM19	4212
ntcomp ctcomp	CaM-CaMKIIα complex	Ca^{2+}-bound	PARAM19	9188

An eigen analysis of the covariance matrix of the conformers selected for the NMA studies was performed to obtain the principal components of the distribution of conformers generated for the N and C domains separately using the R statistical package [Gentleman & Ihaka 2000].

Explicit solvent atoms were then removed from the selected structures and they were subjected to extensive energy minimisation (convergence criterion was norm of gradient vector less than $1x10^{-8}$ kcal mol^{-1} $Å^{-1}$) with a distance dependant ($\varepsilon_r = r_{ij}$ in Å) dielectric constant. The VIBRAN module [Brooks *et al.* 1995] of CHARMM was used to diagonalise the full Hessian, and the normal modes of vibration were computed. The thermodynamic properties (within the classical harmonic oscillator formalism [McQuarrie 1973]) were evaluated over all the (3N-6) non rotational-translation modes. In the case of the simulation of the CaM-Target complex, modes were calculated for the fragments corresponding to the N and C domains while the rest of the system was kept fixed using the Reduced Basis functionality of CHARMM [Fischer *et al.* 2001].

6.3 Results and Discussion

The RMSD plots for the conformations extracted from the different trajectories confirm that the sampling described by the selected frames is similar to that of the original trajectories. The range of RMSD values is similar to that of the full trajectory detailed in

Chapter 5 and the distribution along the time axis suggests that the full 10ns time frame has also been largely explored.

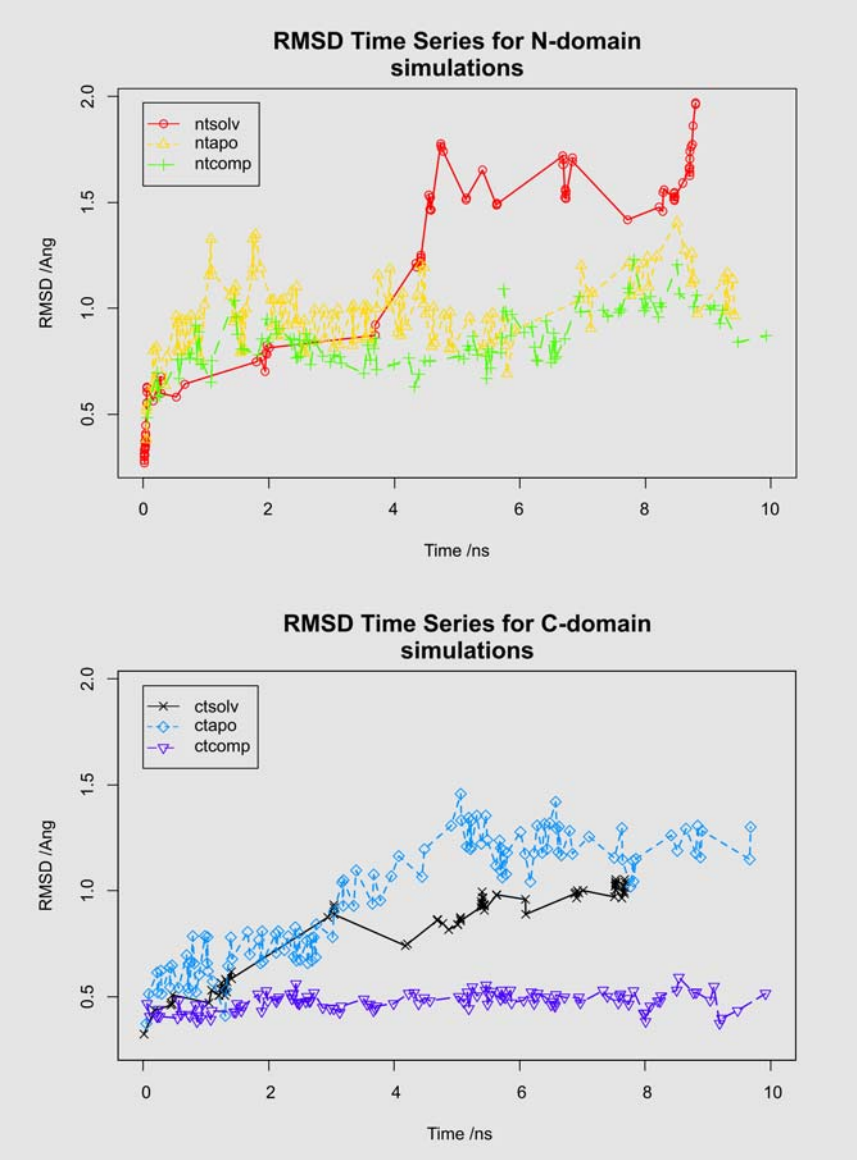


Figure 6.1: *RMSD time series for the selected frames from the MD trajectories of the 10ns simulations of the N (top) and C (bottom) -term lobes of CaM.*

The simulation of the CaM-target complex exhibits marked differences between the behaviour of the N and C-domains of the protein. The N-domain of the protein explores a variety of different conformations over the 10ns simulation ranging within RMSDs of 0.5 Å to1.3 Å to the crystal structure **1exr**. The conformation of the C-term lobe in the simulation of the complex however is highly conserved throughout the 10ns of

simulation, remaining close to that found for the domain in the absence of a target peptide.

These simulations indicate marked differences in the behaviour of the two Ca^{2+}-binding domains of CaM in solution dynamics experiments. Ours and other studies have shown however that MD simulations of proteins can be very sensitive to initial conditions (of conformation and momentum) reflecting high-dimensional and multi-minimum nature of the potential energy surface [Caves *et al.* 1998]. This character may only be amplified in the case of signal transduction proteins, whose architecture is primed for conformational switching [Dunker *et al.* 2002]. Only by comparison of tens of simulated trajectories on this timescale could realistic conclusions be drawn regarding the full-extent of the conformational space accessible to the protein fragments. Such conformational sampling is computationally intractable at this time.

Within this context, we further examined the relative flexibility of the CaM domains using the technique of Normal Mode Analysis (NMA). This technique was used to compute the vibrational properties of multiple energy minima generated from conformations explored in the molecular dynamics simulations. Conformations were selected to ensure that the conformational space explored by the simulations has been fully characterised. The sampling here is sufficiently extensive (ca. 100 different conformations per trajectory) to ameliorate the problems of sensitivity to initial conditions for NMA studies of proteins [van Vlijmen & Karplus 1999].

The distributions of (classical harmonic) configurational entropies calculated from the ca. 100 different conformations selected from each of the 5 different MD trajectories are summarised in **Table 6.2** and illustrated graphically in **Figure 6.2.** The total mean entropic contributions to the vibrational free energy at 300K were 245.9 (**ntsolv**), 247.9 (**ntapo**), 234.5 (**ntcomp**), 220.5 (**ctsolv**), 221.2 (**ctapo**) and 209 (**ctcomp**) kcal mol^{-1}. Values summarised in **Table 6.2** below are per degree of freedom to account for the different number of atoms in each lobe.

Table 6.2: *Summary of mean entropy values from the distributions of NMA calculations on the conformations extracted from the MD trajectories. Standard deviations for the distributions are shown in brackets in units of kcal mol^{-1} per degree of freedom.*

Simulation	N-domain Entropy	C-domain Entropy
Ca^{2+}-bound **ctsolv ntsolv**	0.368 (0.00290)	0.363 (0.00277)
Ca^{2+} expulsion **ctapo ntapo**	0.373 (0.00236)	0.365 (0.00320)
Ca^{2+}-CaM-target complex **ctcomp ntcomp**	0.364 (0.00356)	0.346 (0.00270)

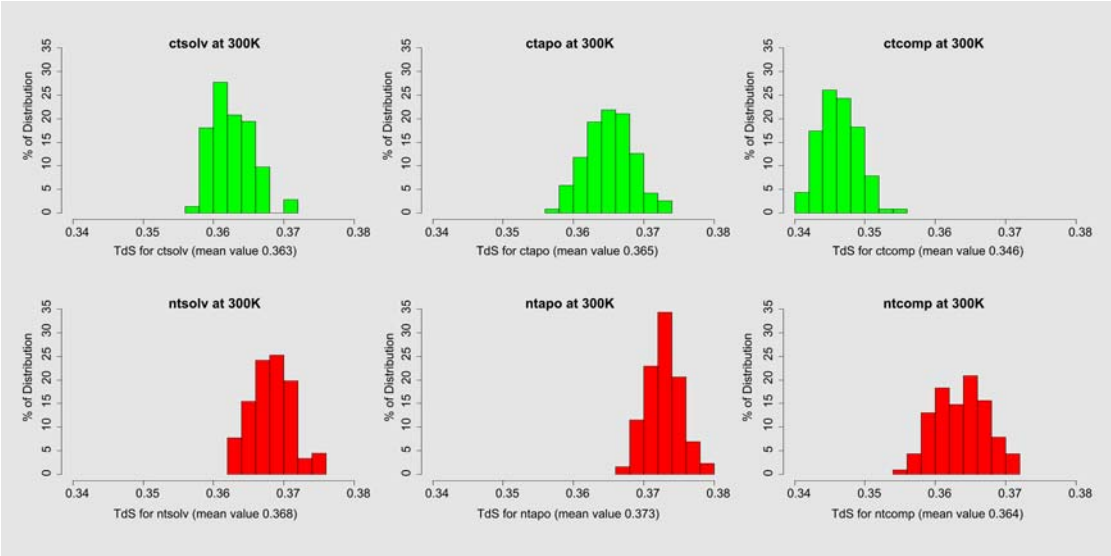


Figure 6.2: *Histograms illustrating the results of the configurational entropy calculations from the conformations extracted from the MD trajectories. Calculations for the C-term are shown on the top row in green and the N-term results are at the bottom in red.*

In each experiment, the mean value of the distribution of configurational entropy values is larger for the N-term than it is for the C-term **Table 6.2** the statistical significance of these differences was established using the standard t-test. This provides clear evidence for a different character of the two Ca^{2+}-binding domains of CaM. Of particular interest is the observation that the difference is most evident in the simulations of the CaM-target complex (**Figure 6.2**). Both the RMSD time-series (**Figure 6.1**) and the configurational

entropy values calculated in these experiments confirm that the N-domain is more inherently flexible than the C-domain of the protein.

Examination of the density of states in the 6 systems [van Vlijmen & Karplus 1999] shows that while the overall distribution is very similar (**Figure 6.3** inset), the differences between the systems are largely confined to the low frequency motions (<= 60 cm^{-1}) (**Figure 6.3**). Indeed, the decrease in entropy for a 1 cm^{-1} increase in frequency is rapid and goes from -12% from 1 cm^{-1} to –1% at 35cm^{-1}. This also suggests that the major differences in flexibility result from collective displacements delocalised over the whole of each lobe rather than any specific region.

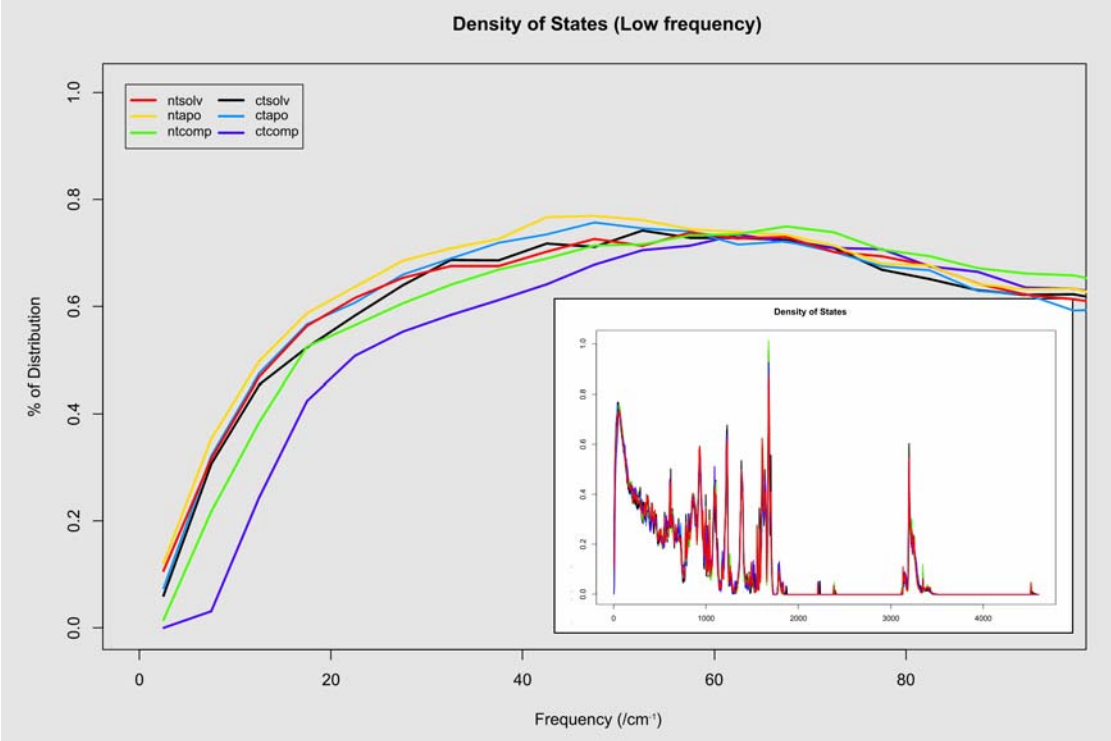


Figure 6.3: *Density of states for the lowest energy frequencies of vibration (main plot) and the full vibrational spectrum (inset) generated for the different trajectories. These are generated from the distribution of vibrational frequencies resulting from the normal mode analyses performed on multiple conformations taken from the different MD trajectories.*

6.4 Conclusions

These differences in flexibility of the N and C lobes of CaM can be considered in the context of different properties of the two lobes. The more conserved C-lobe of the protein has a higher affinity for Ca^{2+} and a higher affinity for the majority of CaM targets [Bayley *et al.* 1996].

In this work we see a distinct difference in the inherent flexibility of the N- and C-term lobes of CaM. Additionally, we see how the presence of a target peptide results in a more conserved and less flexible conformation for the C-term lobe of the protein while the N-term lobe retains much of the flexibility observed in the absence of the target peptide.

The identification of different functional roles for the domains [Barth *et al.* 1998], combined with the observation of a variety of different target binding modes [Zhang & Yuan 1998] (with the domains in a variety of relative spatial arrangements) helps to explain the diversity in primary sequence observed for the plethora of CaM target sequences and suggests that target recognition is performed on a per-domain basis with separate motifs for the N and C domains of the protein [Afshar *et al.* 1997]. The non-contiguous nature of some targets has been further exposed in recent crystal structures [Schumacher *et al.* 2001; Drum *et al.* 2002]. The identification of different functional roles of the two CaM domains, coupled with their relative inherent flexibility and potential for global re-orientation afforded by the central tether are essential elements for CaM function.

Chapter 7: Calmodulin-Target Interaction

7.1 Introduction

The characterisation and analysis of the mode of interaction between Calmodulin and its targets has been the subject of countless investigations. In this chapter we will attempt to draw on our improved understanding of CaM conformational flexibility obtained through the previous investigations of this thesis and the interpretation of the structural information available for different CaM-target systems.

The ability of CaM to interact with such a diverse range of different target systems is achieved through its conformational flexibility. This conformational flexibility is conferred on three different levels:

- **Global Conformational Flexibility:** This describes the rearrangement of the CaM lobes with respect to each other and CaM targets
- **Local Conformational Flexibility:** This describes the extent to which the conformation of the lobes themselves adapt to accommodate different targets, particularly in the context of the *open*, *closed* and *semi-open* lobe conformations representative of different Ca^{2+}-binding states.
- **Binding Surface Plasticity:** The extent to which the side chains of CaM are able to alter the binding surface of the protein in the context of the backbone conformation of the protein.

7.1.1 Global Conformational Flexibility

The flexibility of the central tether in CaM which connects the two Ca^{2+}-binding domains has long been recognised and well characterised. Computational investigations have been used to confirm that this tether is inherently flexible [Van Der Spoel *et al.* 1996] in the absence of the Ca^{2+}-binding domains themselves. Not only does this flexibility enable the

hydrophobic patches on the two lobes, which are some 40Å apart in the crystal structure **1exr**, to come within 10Å of one another to accommodate target binding domains, but the separation of these domains is highly adaptable depending on the nature of that target as illustrated in **Figure 7.2** below. The differences between **1cdl** and **1cm1** are quite subtle at this degree of detail in that the main difference is the separation of the bulky hydrophobic residues characteristic of classical CaM-binding domains. In **1cdl** the Trp which interacts with the C-term lobe of CaM is separated by 12 residues from the Leu residues which interacts with the N-term lobe. In **1cm1** however, a Leu sidechain, which locates in the hydrophobic pocket of the C-term lobe, is separated by 8 residues from the Leu interacting with a similar region on the N-term lobe. The structures in the figure are superposed on the C-term lobe of the protein which highlights the reorganisation of the N-term lobe to move further along the CaM target.

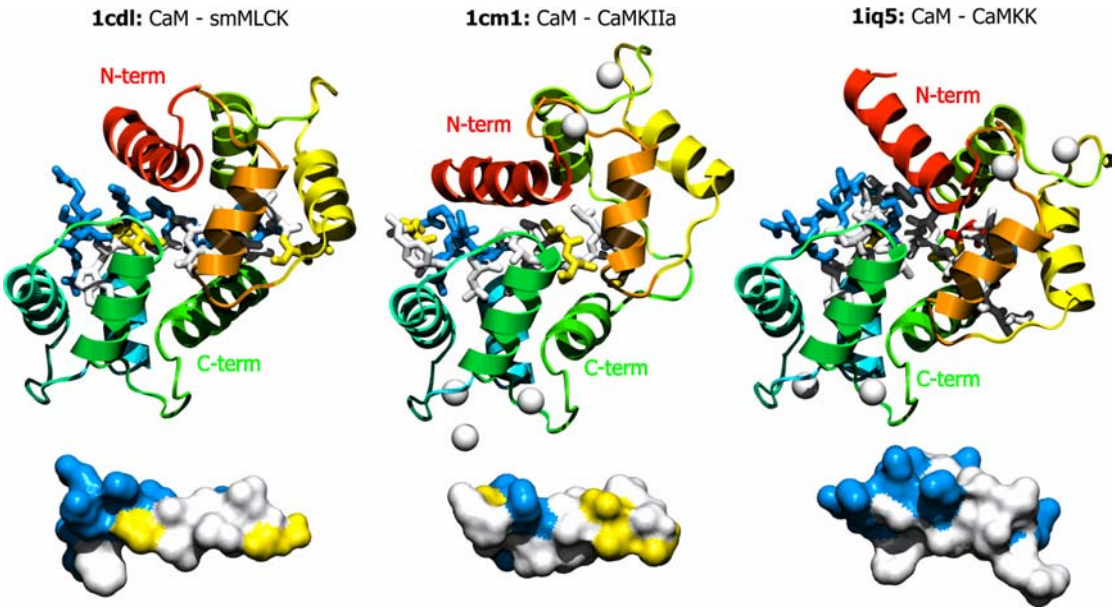


Figure 7.1: *Figure illustrating the global organisation of* ***classical*** *CaM-target interactions. Calmodulin is represented with a ribbon protein cartoon to show the organisation of the secondary structure elements in the complex. CaM targets are illustrated with a liquorice representation in the complexes and with a solvent accessible surface underneath in the same orientation. CaM colouring is by sequence; from red (N-term) through to Cyan (C-term). Target sequence colouring is by amino acid polarity, hydrophobic residues are coloured in white, positively charged residues in blue, negative residues in red and polar residues in yellow.*

A further degree of global rearrangement is required in the case of the interaction of CaM and the CaM binding domain of CaM Kinase Kinase which was initially determined by NMR [Osawa *et al.* 1999] and later by X-ray diffraction [Kurokawa *et al.* 2001]. In this complex the target peptide incorporates a short region of β-sheet in addition to the helical region normally seen in these complexes. This is associated with a further displacement of the N-term lobe of CaM enabling it to wrap further around the target peptide to correctly associate the Phe on the target sequence with the hydrophobic pocket on the N-term lobe of the protein.

Such global conformational rearrangements observed in *classical* CaM-target interactions as illustrated in **Figure 7.1** show a relatively high degree of conservation in comparison to the *non-classical* systems in **Figure 7.2**. The structure **1aji** is a homology model developed to detail the mode of interaction between CaM and IQ-motifs in the neck regions of non-conventional myosins [Houdosse *et al.* 1996]. CaM plays a reverse Ca^{2+}-dependent role in this system in which an increase in Ca^{2+} concentration results in a conformational change in the system and release or complete reorganisation of the mode of interaction in the system. The CaM lobes themselves adopt alternative conformations in this system largely due to the absence of Ca^{2+}, although this structure again demonstrates the relative independence of the lobes facilitated by the flexible tether.

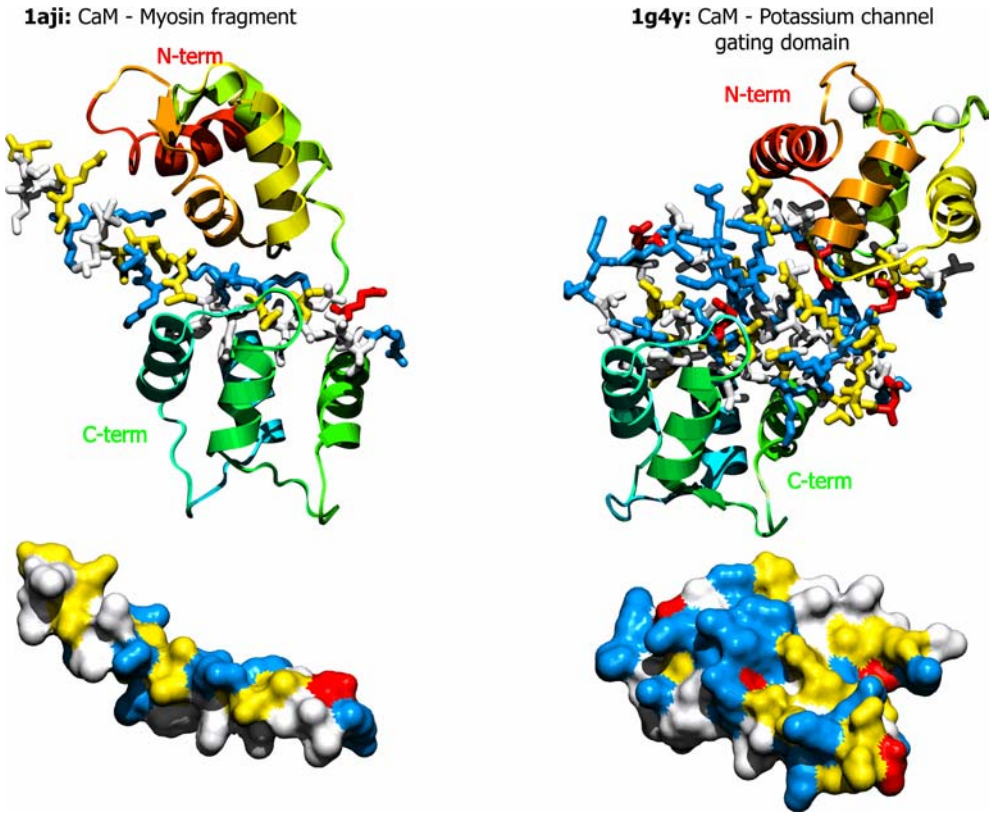


Figure 7.2: *Figure illustrating the global organisation of* ***non-classical*** *CaM-target interactions. Calmodulin is represented with a ribbon protein cartoon to show the organisation of the secondary structure elements in the complex. CaM targets are illustrated with a liquorice representation in the complexes and with a solvent accessible surface underneath in the same orientation. CaM colouring is by sequence; from red (N-term) through to Cyan (C-term). Target sequence colouring is by amino acid polarity, hydrophobic residues are coloured in white, positively charged residues in blue, negative residues in red and polar residues in yellow.*

This propensity for diversity in the global arrangement of the CaM lobes with respect to their targets and the high diversity observed in the primary sequence of CaM targets was thought to be suggestive that rather than a single target motif, CaM is able to recognise two separate binding regions, one for each lobe of the protein [Afshar *et al.* 1997]. This possibility was essentially confirmed with the crystal structure determination of CaM in complex with the gating domain from small conductance potassium channel, **1g4y** [Schumacher *et al.* 2001]. In this structure, two copies of Ca^{2+}-CaM each bind to two separate chains from the gating domain resulting in the formation of a tetramer. The observation that CaM binds to two separate molecules simultaneously confirms that there are separate target motifs for the N- and C- lobes of the protein. The representation in **Figure 7.2** above shows the extensive target domain enclosed between

the two CaM lobes. However it can be seen from this illustration that this global arrangement of the lobes can only be accommodated through the significant flexibility conferred by the tether. The character of the CaM binding regions in this structure can be seen to be markedly different to others in the illustration through the surfaced representations in **Figure 7.1**. The interaction areas on the target fragment however are still clearly hydrophobic despite the absence of the locator pins characteristic of other systems.

The flexibility of the tether has not been investigated in depth in this thesis as it has been the focus of several previous modelling studies [Van Der Spoel *et al.* 1996; Wriggers *et al.* 1998] and numerous experimental studies: A CaM mutant in which residues critical for flexibility in the helix were deleted, Thr-79 and Asp-80, was shown to exhibit a relative re-orientation of the CaM lobes such that they were facing each other in the crystal structure (**1ahr** [Tabernero *et al.* 1997]). Thus inhibition of phosphodiesterase by this mutant was also shown to be affected. Another interesting experiment demonstrated that the transplantation of the Asp-Thr-Asp region of the CaM linker into Troponin C enabled it to fully activate phosphodiesterase *in vitro* [Gulati *et al.* 1993].

7.1.2 Local Conformational Flexibility

Another feature of CaM which is of particular interest in this context is the ability of the two separate lobes to respond to changes in Ca^{2+} concentration resulting in several different local conformations, often described as *open*, *closed* and occasionally, *semi-open* [Nelson & Chazin 1998]. The principal component analysis performed in Chapter 3 of this thesis facilitated the identification of these different conformations observed in structures representative of different physiological conditions. The molecular dynamics studies investigated the stability and dynamics of these conformations in the simulations in Chapter 5.

The central paradigm for CaM function suggests that an increase in local Ca^{2+} concentration results in the inter-conversion of apo-CaM in a *closed* conformation to Ca^{2+}-CaM which exhibits an *open* conformation in both lobes and the exposure of the

hydrophobic binding surfaces which bind to the majority of CaM targets. Significant exceptions to this paradigm have also been identified. The IQ-motif class of CaM targets was the largest group identified in the CaM target review of Rhoads & Friedberg [Rhoads & Friedberg 1997] and the CaM Target Database [7][Yap *et al.* 2001]. The notion of an IQ motif originates in the binding motif of the myosin light chains which are CaM homologues. As these are mostly Ca^{2+} independent CaM binding regions it expected that CaM will bind in a *semi-open* or *closed* conformation adopted by the Myosin light chains rather than the *open* conformation normally exhibited in CaM-target complexes. As outlined above, the PDB structure **1aji** [Houdosse *et al.* 1996] is a model structure based on a *semi-open* conformation for the C-term lobe of CaM and a *closed* conformation for the N-term lobe. These observations are clearly seen in the conformational subspace projections of the observed structural distribution (**Figure 7.3**). This model has also been suggested to explain the reverse Ca^{2+}-dependent interaction between CaM and Neuromodulin [Zhang & Yuan 1998].

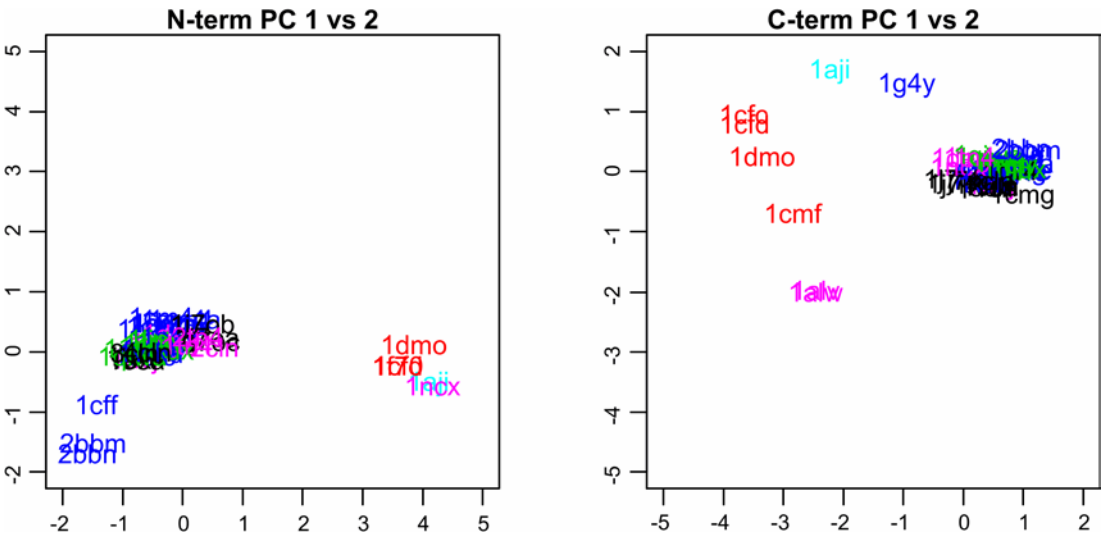


Figure 7.3: *Conformational subspace projections for the CaM N and C-term lobe structures. A full description of the analysis is included in the Appendix. The structures are represented by their PDB codes and coloured according to target association state.*

Other clear outliers in the distribution of experimentally determined structures are also noteworthy. The crystallographically determined complex of CaM with the gating domain from small conductance potassium channel (**1g4y** [Schumacher *et al.* 2001]) not only

[7] http://calcium.uhnres.utoronto.ca/ctdb/

demonstrates a novel global conformation but the non-classical motif induces a different local conformation in the N-term lobe of CaM. Careful inspection of this complex reveals that the presence of a Leu in the centre of the C-term lobe of CaM which has forced a new hydrophobic pocket in the CaM binding surface. This has in turn removed the classical C-term hydrophobic binding pocket to such an extent that the Trp in the target peptide is forced to interact with the CaM surface rather than being buried in a hydrophobic pocket. This seems to be coupled to a slight closing of the CaM domain resulting in a novel local conformation for the C-term not observed for any of the other CaM structures. In this context it is interesting that Ca^{2+}-binding to the N-term lobe is essential for CaM function in this system while interaction through the C-term lobe of CaM is Ca^{2+} independent [Schumacher *et al.* 2001].

The fragment from the gating domain from small conductance potassium channel and the IQ motif exhibited in **1aji**, have a relatively similar amino acid sequence at the C-term binding site. The Ile, Gln and Phe in **1aji** are in very close spatial proximity to Leu, Arg, and Phe in **1g4y** after superposition of the C-term lobes of CaM. A homology model for the unconventional myosin complex built using **1g4y** as a template would inevitably result in a very similar conformation for the C-term lobe of CaM.

The MD simulations performed on the CaM fragments suggested that the N-term of CaM exhibits greater flexibility than the C-term lobe of the protein. The magnitude of conformational changes observed were generally larger in the N-term and the range of conformations observed in the PCA of the two fragments also suggested that there was greater flexibility in the N-term lobe. Comparison of the different ligand bound structures presented through the results of the PCA indicates that the lobe conformation is relatively well conserved for most structures. Notable exceptions include the NMR structures **2bbn** and **2bbm** [Ikura *et al.* 1992] and the structure of CaM in complex with a fragment from the Ca^{2+} pump **1cff** [Elshorst *et al.* 1999]. One possible reason for the anomalous conformation N-term lobe in **2bbn** is that this structure has a large Phe group interacting with the hydrophobic pocket on the protein. There are however several inhibitor structures which also have aromatic ring groups in this pocket without causing a similar effect on the lobe conformation. The target peptide from skMLCK is very similar to that of smMLCK in **1cdl** and the sequence of the two fragments can easily be aligned.

The absence of an obvious cause for the differences in conformation of the protein fragments must therefore be attributed to the different experimental conditions and techniques. The solution NMR structure, **1cff** [Elshorst *et al.* 1999], shows the target peptide interacting only with C-term lobe of the CaM. There are no suggestions as to why this structure should adopt a different conformation in the N-term lobe compared to other crystal and solution structures for the protein.

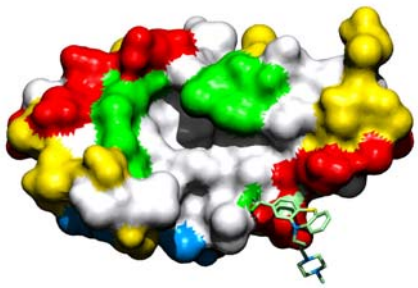
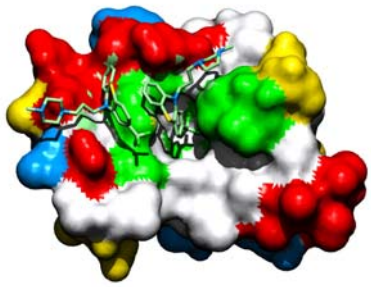
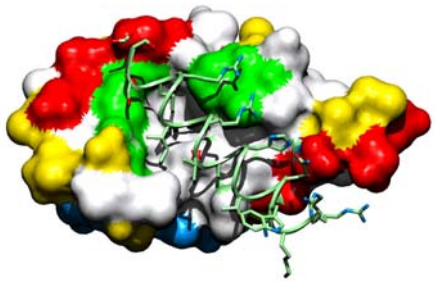
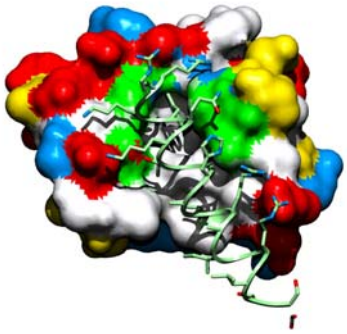
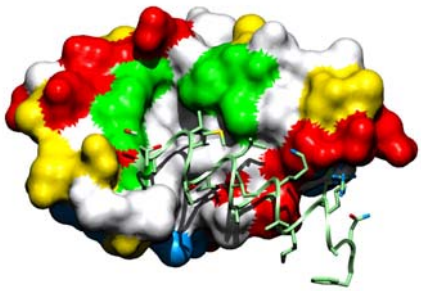
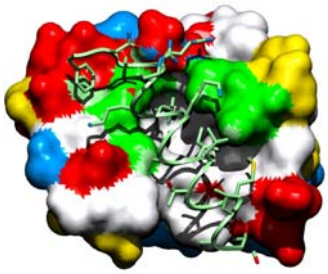
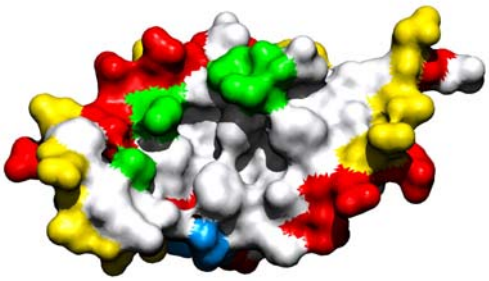
7.1.3 Binding Surface Plasticity

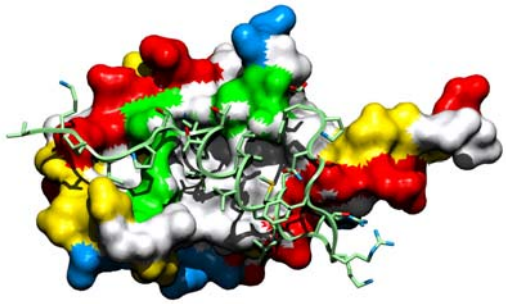
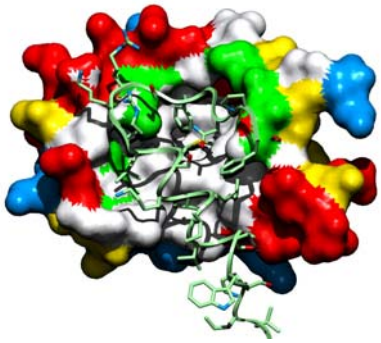
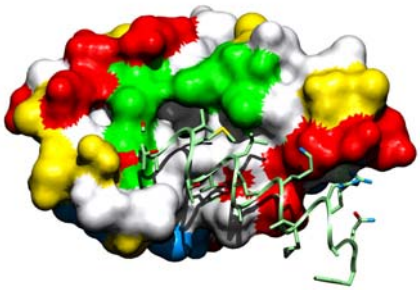
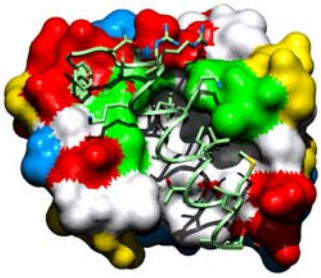
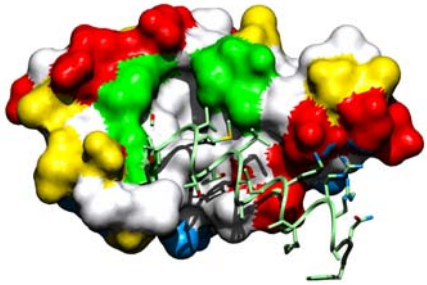
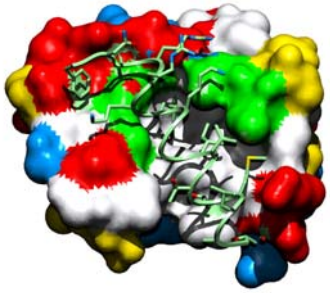
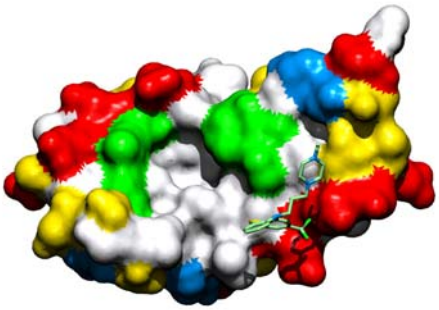
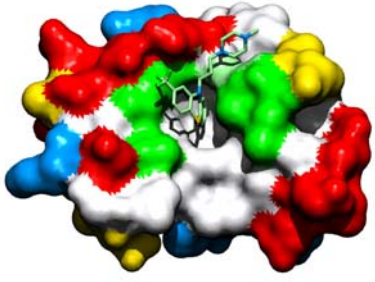
The PCA performed on the crystal structures presented earlier in this thesis suggested that subtle adaptations in the conformations of the CaM lobes take place on target association. These adaptations could be well characterised by individual principal components for the N and C-term lobes which detail the conformational changes observed. The scale of these conformational changes however is very small and Cα RMS differences between the structures are of the order of only tenths of an Å. Similarly, the conformational variability of the amino acid sidechains interacting at the surface of the CaM binding seems to be limited. Superposition of the N-term lobes from all the CaM-target crystal complexes show that the only binding site residues to undergo significant conformational rearrangement are the Met residues. These residues are highly flexible with an unbranched chain of four heavy atoms including a sulphur atom making them highly polarisable [Zhang & Yuan 1998]. The plasticity of the CaM domains is often referred to as the key to the CaM's ability to interact with a wide variety of target sequences. This was recently challenged through the determination of the RDC NMR structures of Chou *et al.* [Chou *et al.* 2001] which indicated that the χ_1 angles for the Met residues were fixed across various different protein structures.

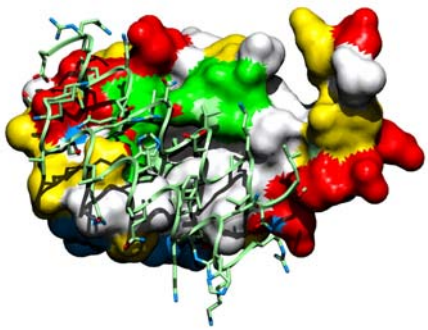
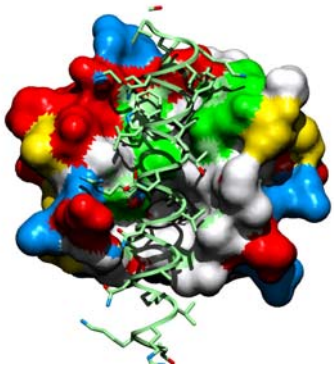
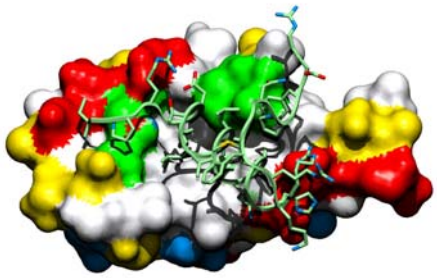
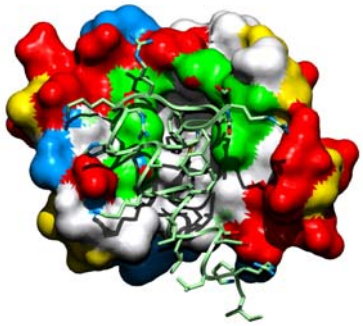
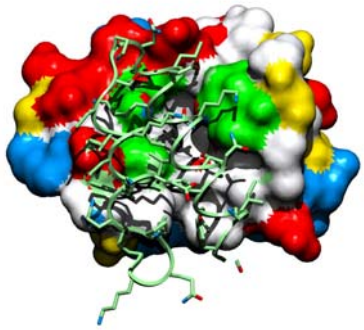
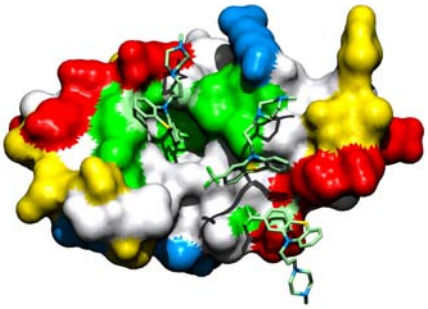
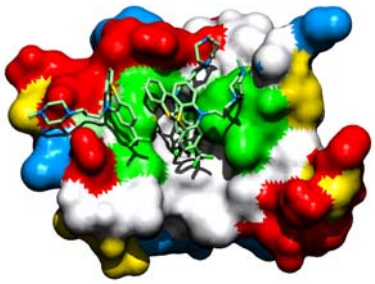
The understanding of the CaM binding surface plasticity is a complex problem which we have attempted to address in various different ways. The full selection of CaM target complexes (including fragments of CaM binding domains and inhibitor bound structures) has been interrogated in detail with the use of Molecular Graphics [Accelrys 2001]. Different techniques of graphical representation and molecular surface generation have been used to inspect the adaptability exhibited by the CaM structure in the

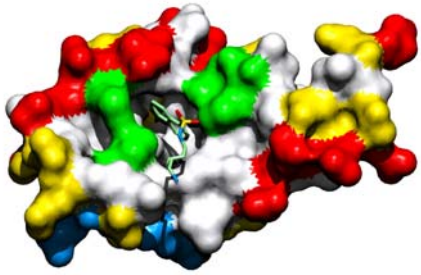
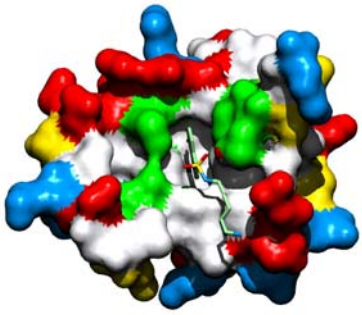
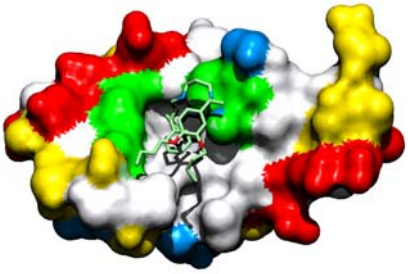
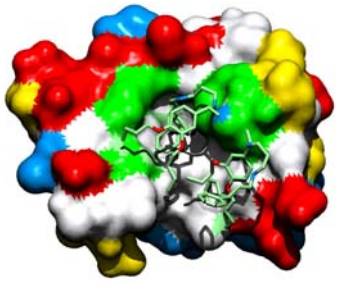
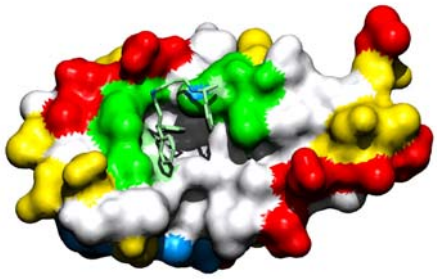
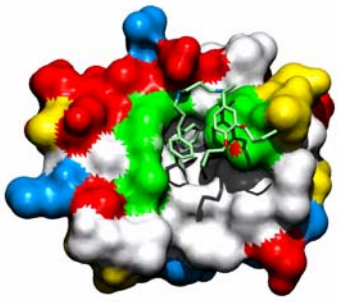
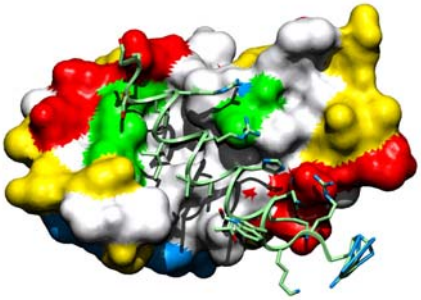
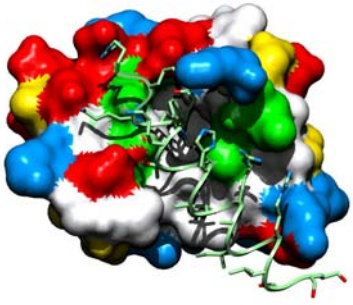
accommodation of different ligands and peptides across the ensemble of experimental structures.

In the figure below (**Figure 7.4**), rendered representations of the solvent accessible molecular surface presented by CaM (in the absence of co-factors) in response to different targets are coloured to highlight the character of the binding surfaces of the N and C-term lobes. These representations clearly show the hydrophobic binding surface of the CaM lobes surrounded by charged amino acid side chains at the edges. The importance of the Met residues in the ligand binding is also clearly evident from these representations. The protein fragments are aligned to a common reference frame in order to provide comparable orientations for the different structures.

1a29	
	
1cdl	
	
1cdm	
	
1cff	
	

1ckk	
	
1cm1	
	
1cm4	
	
1ctr	
	

1g4y	
	
1iq5	
	
	1k90
	
1lin	
	

1mux	
	
1qiv	
	
1qiw	
	
1vrk	
	

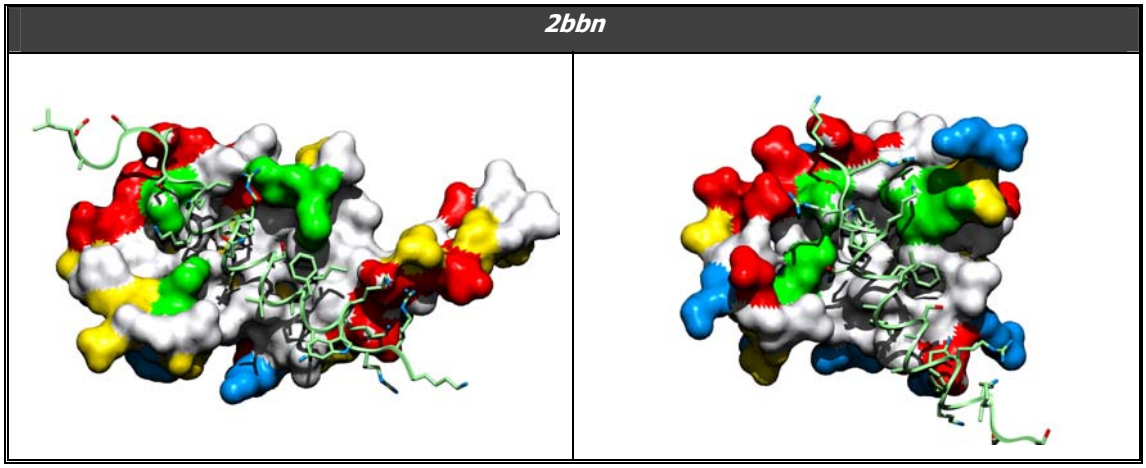


Figure 7.4: *Surfaced representation of the CaM lobes. N-term lobes are on the left hand side of each panel and C-term lobes on the right. Target peptides and ligands are represented with ribbon cartoons for the peptide backbones and liquorice models for the amino acid sidechains.*

These representations of the CaM binding surface show the overall shape to be largely conserved for the different CaM targets. The size and shape of the hydrophobic pockets on both the N and the C-term lobes changes noticeably across the ensemble. It is important to note that the most significant differences are observed in the NMR structures **1cff**, **1ckk**, **1mux** and **2bbn**. As discussed previously, the direct comparison of experimentally determined structures from different techniques should be approached with some caution (see discussion in Chapters 2 & 3) although useful information can be obtained from both types of structure. Comparisons of side chain orientations are particularly susceptible to problems in this area.

7.2 Calmodulin Side Chain Flexibility Analysis

Whilst the role of global and local conformational flexibility is important in the interactions of CaM with different target peptides and ligands, most of the binding energy comes from key interactions between the protein sidechains and the chemical groups on the target. In fact one of the key features of CaM target interaction is that there are no interaction between the protein and the backbone of the target peptides. The images of the different interfaces shown above (**Figure 7.4**) suggest that the plasticity of the CaM binding surfaces is important to these interactions although the role

of individual amino acids cannot be detailed through such an approach. Here the different rotameric states of all the CaM amino acid side chains are studied in order to identify the alternate conformations adopted to accommodate the range of different targets.

The torsion angles Chi 1, 2 and 3 (χ_1, χ_2 and χ_3 in the **Figure 7.5**) represent the most significant degrees of freedom for side chain orientation. Alterations to these internal degrees of freedom across different protein structures are largely determined by the immediate environment of the protein side chain namely non-bonded and H-bonded interaction with other groups. Chi torsion angle data were compiled from the CaM structural dataset and analysed through various different graphical representations of the data.

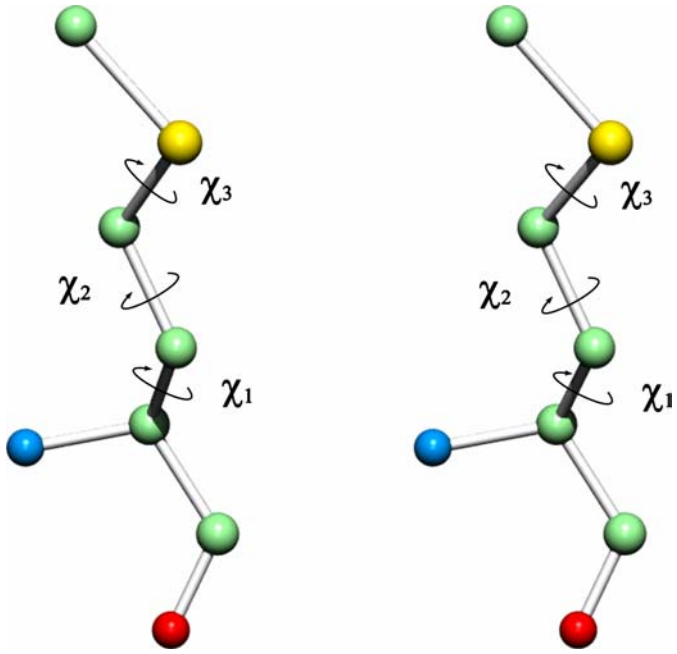


Figure 7.5: *WALLEYE stereo image of a single Met residue in a ball and stick representation labelled with the rotameric torsion angles χ_1, χ_2 and χ_3.*

7.2.1 Data processing

CaM structures were obtained from the PDB and solvent and dis-order information were removed. The first model from multiple-model structures was used and the most representative NMR models were selected according to the *OLDERADO* results [Kelley

et al. 1997]. Structures were then passed through *PROCLEAN* [Laskowski *et al.* 1993] to restore standard IUPAC nomenclature. *SQUID* [Oldfield 1992] was then used to generate the χ_1, χ_2 and χ_3 angle data. This data was then analysed using the R package [Gentleman & Ihaka 2000]. The first stage of the processing involved adjusting the data around the mean of each distribution such that no value was more than 180^0 from the mean value. This is an important consideration when dealing with the inherent periodicity in data from internal coordinates [Reijmers *et al.* 2001]. The rotational symmetry of particular residues was then incorporated for the residues Asp, Glu and Phe for which χ values separated by 180^0 are structurally equivalent. Appropriate values were offset by 180^0 to avoid obtaining an artificially bi-modal distribution.

Information regarding the rotameric angle variability was compared with solvent accessibility and protein-ligand contact information compiled using the functions in the Protein Design module of QUANTA [Accelrys 2001]. These functions are based on the small molecule probe algorithm [Lee & Richards 1971] designed to analyse the potential for water permeation through protein structures. This analysis was performed on the most representative model of the NMR structure 2bbn [Ikura *et al.* 1992] using a probe of radius 2.0Å. Despite the uncertainty regarding the precision of structural information from NMR experiments, this model was selected on the basis that it was the most reliable model to have a target peptide interacting with the entire CaM binding surfaces on both lobes. Many of the other CaM-target structures have target fragments which only interact with part of the binding surfaces on one or both of the CaM binding lobes (see **Figure 7.4** above).

Boxplots are used extensively in the description of this data as they provide a condensed description of the distribution of the data [Tukey 1977]. This representation reports information on the *range*, *upper* and *lower quartiles* and *median* of the statistical data in a simple reduced representation as described in the figure below. After ranking of the data values, the *Median* value represents the middle value of the distribution, the *upper quartile*, the middle value of the top half of the data and the *lower quartile* the middle value of the lower half of the distribution.

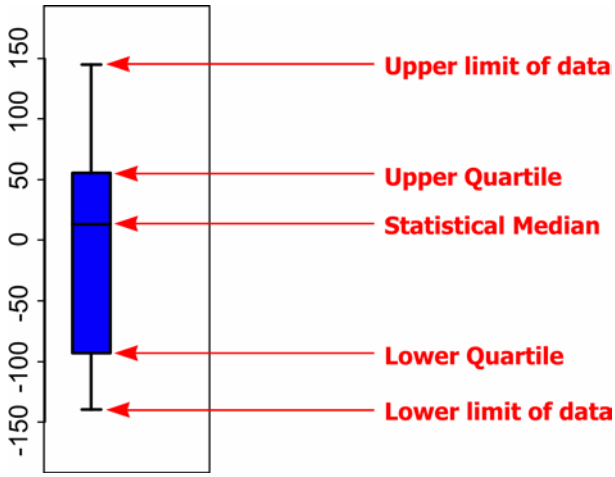


Figure 7.6: *Description of the statistical data distribution as represented by the boxplots used in this analysis.*

7.2.2 Results & Interpretation

The data compiled from this analysis are presented in **Figure 7.7** and **Figure 7.8** below for the N and C-term lobes of CaM respectively. In these figures the data are aligned with the amino acid sequence of the corresponding CaM lobe, included at the bottom of each section. The colour scheme for the all the data for each residue is determined by the calculation of target contact areas in the lowest plot according to a blue to red spectrum. In this way, the residues most important in target interactions are highlighted throughout in red with those not involved in binding in blue. The bar plots in the lower panel are aligned with the sequence ruler and the box plots of the Chi angle data in the upper panel.

7.2.2.1 Target Contact Data:

As outlined in section 7.2.1 above, the contact data for the side chains of the protein was obtained using an algorithm involving a 2.0Å molecular probe to determine the accessibility of different parts of the **2bbn** system. This data was also used to construct a colour spectrum which was applied to all the other results in which the tightest contacts between CaM and the sk-MLCK peptide are coloured in red and the residues making no contacts with the peptide are coloured in blue.

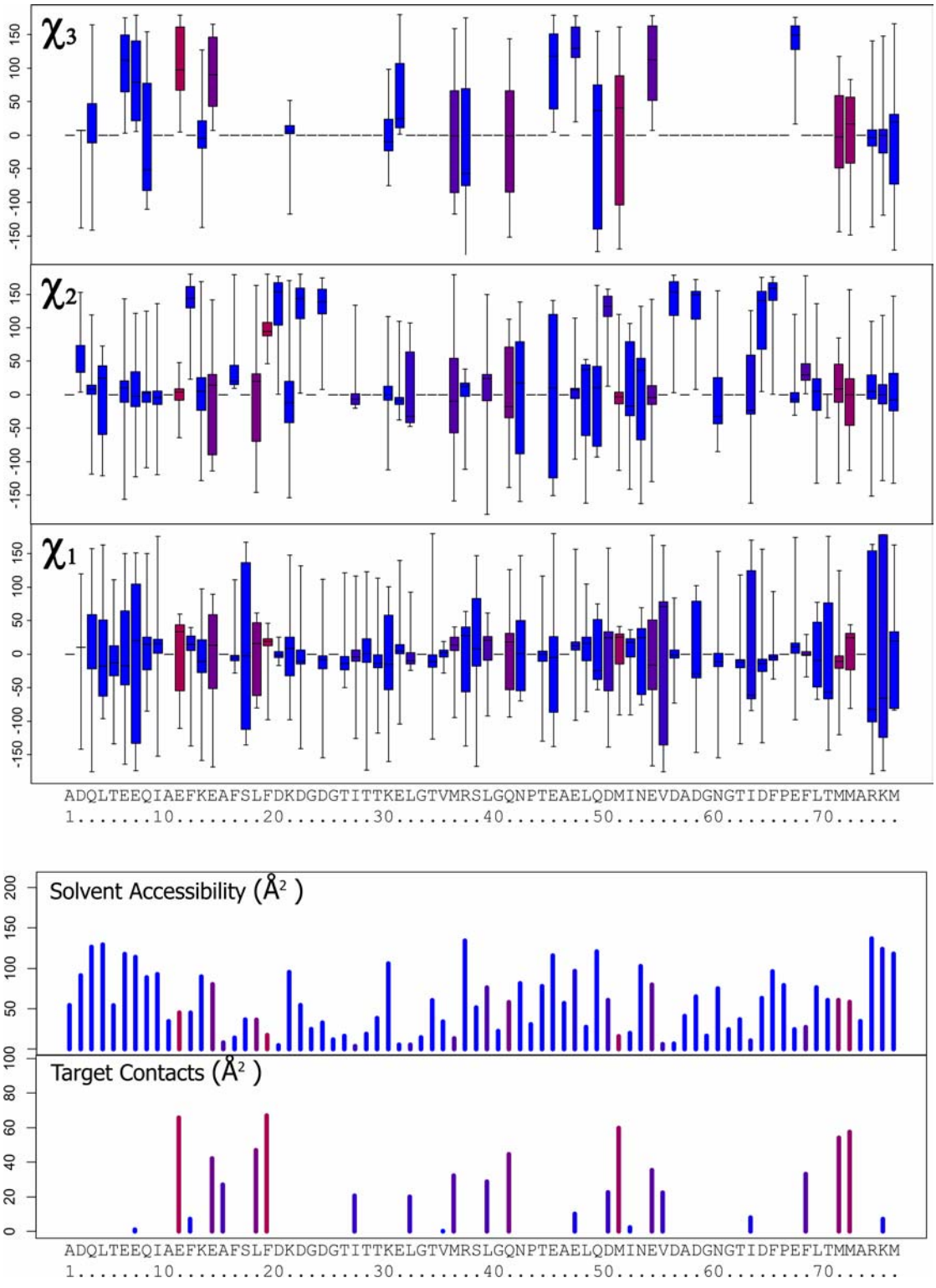


Figure 7.7: *Side chain rotameric variability data for the N-term lobe of CaM. The top section illustrates boxplots for the distribution of values for χ_1, χ_2 and χ_3 data which are aligned with the amino acid sequence at the bottom of the plot. The lower section of the plot includes solvent accessibility and target contact data again aligned with a sequence ruler. The data is coloured throughout according to a blue to red spectrum generated from the target contact data for the full structure.*

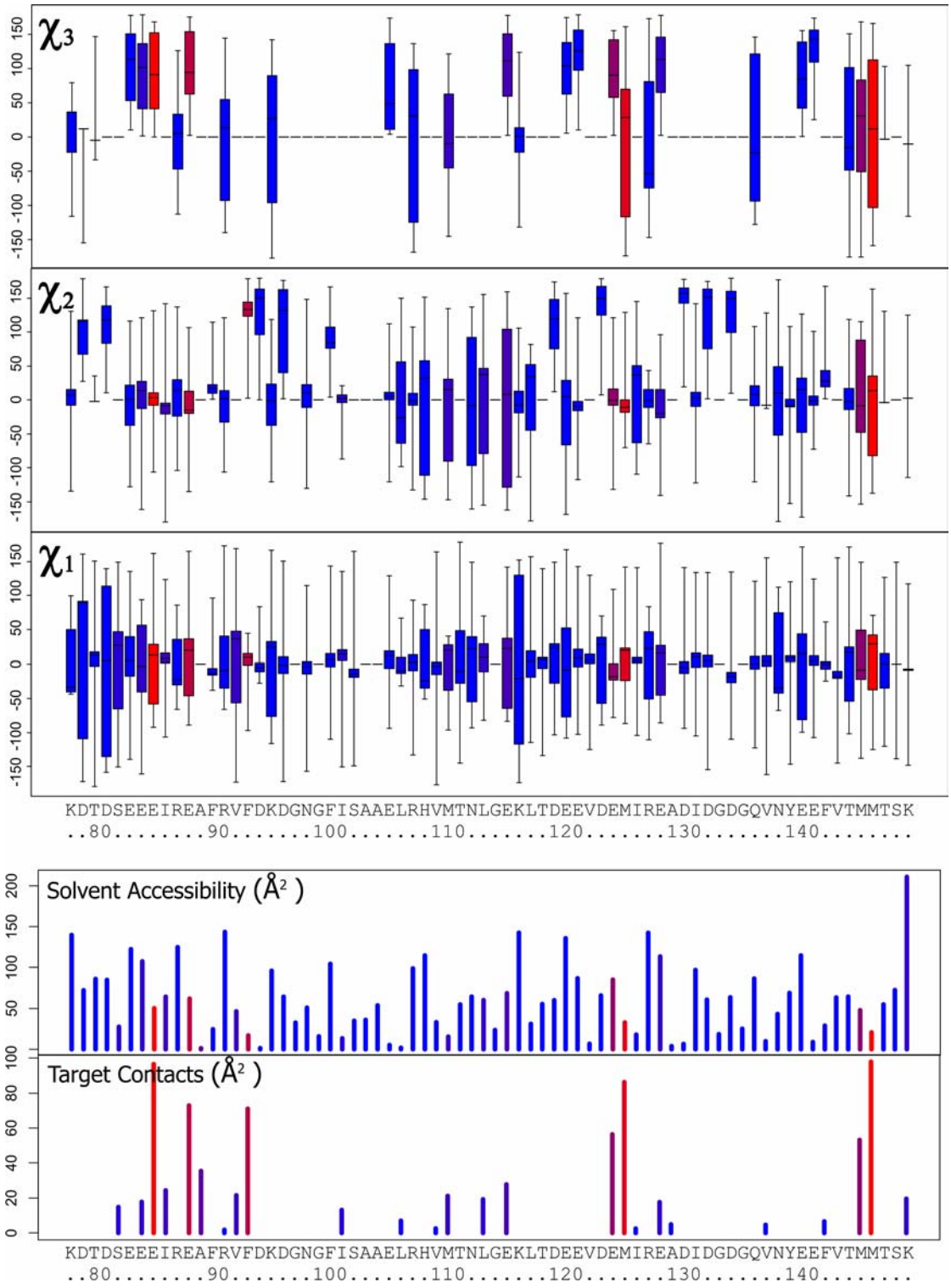


Figure 7.8: *Side chain rotameric variability data for the C-term lobe of CaM. The top section illustrates boxplots for the distribution of values for χ_1, χ_2 and χ_3 data which are aligned with the amino acid sequence at the bottom of the plot. The lower section of the plot includes solvent accessibility and target contact data again aligned with a sequence ruler. The data is coloured throughout according to a blue to red spectrum generated from the target contact data for the full structure.*

From this colouring, it is apparent that more close contacts appear to be formed between the C-term lobe of the protein and the peptide. This observation is in accordance with the experimental work of Bayley *et al.* on this system [Bayley *et al.* 1996]. These studies established that the strongest component of the interaction was between the C-term lobe of CaM and the N-term end of the target peptide.

The results of this study also suggest that the Met residues at the C-terminus of both lobes of the protein make significant VdW contacts with the peptide. Glu11, Glu84 and Glu87 also establish important electrostatic interactions with basic residues on the target peptide. Phe19 and Phe92 also make extensive contacts in this structure through the N and C-term lobes of the protein respectively. Phe19 is positioned on the top of the N-term lobe binding surface and Phe92 forms an end on interaction with the Trp side chain of the peptide in the C-term lobe hydrophobic pocket.

7.2.2.2 Solvent Accessibility

Unlike the target contact data which is reported for side chains only, the solvent accessibility information is summed over the protein backbone and side chain for each residue. Initial results were analysed to the side chains only but these were found to show artificially low values, particularly for the terminal residues which are highly solvent exposed. Normally a probe radius of 1.4Å would be used for this analysis to mimic the size of a water molecule. This was increased to 2.0Å to generate smaller values, particularly for amino acids which are in the hydrophobic core of the protein and are in reality completely inaccessible to water molecules.

The residues which had been identified as being important in the interactions between the protein and the target are generally less accessible to the solvent whereas other residues in the system are more accessible to the solvent. Not surprisingly, many of the residues with high solvent accessibility are polar or charged.

7.2.2.3 X_1 Rotamers

The box plots summarising the distribution of χ_1 values show that these are generally the most constrained of the three torsions in the side chains although there are specific exceptions. Ser17 in the N-term lobe exhibits considerable flexibility although this side

chain points straight out into the solvent suggesting fewer constraining steric interactions. Val55 also exhibits considerable flexibility and forms an important part of the N-term lobe hydrophobic pocket. Closer inspection of the structure indicates that this residue is closely associated with Ile63 which is also highly flexible in χ_1 and that the two side chains are orientated towards one another in EFII just above the Ca^{2+} binding site.

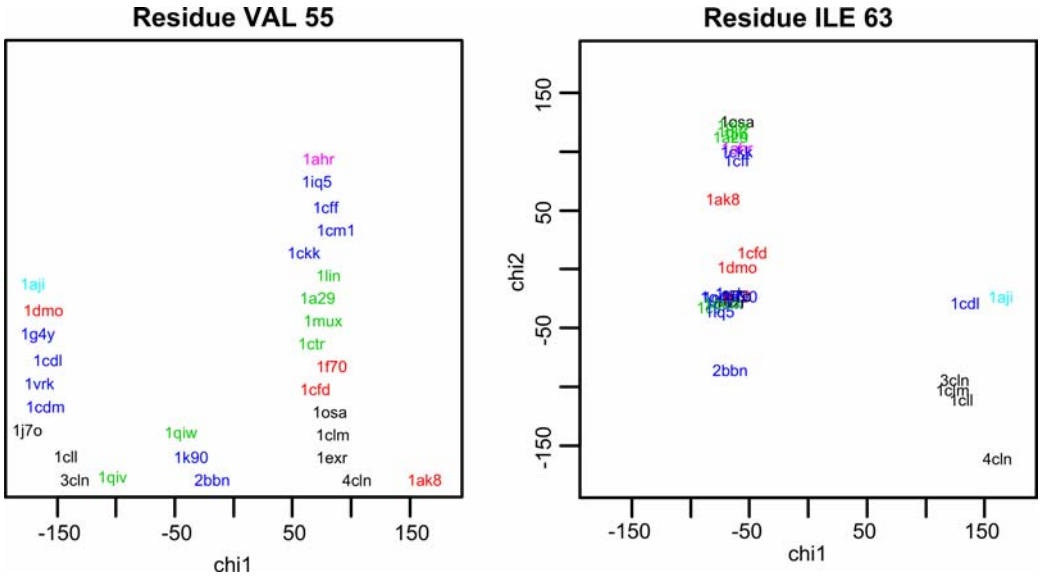


Figure 7.9: *Scatter plots indicating the distribution of χ_1 and χ_2 values for Residues Val55 and Ile63. In the absence of χ_2 values for Val, the distribution is scattered sequentially along the ordinate. Positions are identified by PDB accession codes and coloured according to structural classification (See Appendix)*

The actual distribution of χ_1 and χ_2 values for the different structures is presented in the scatterplots in **Figure 7.9** above. From this representation it is clear that there are several preferred orientations for the different sidechains and in particular for Val55. Here the preferred χ_1 angle is around 60° with an alternate conformation of –150°. The rotamer studies of Dunbrack *et al.* studied the conformations of amino acid side chains in different secondary structure conformations identified that 71% of Ile residues in β-sheet regions had a χ_1 value of -60° [Dunbrack & Karplus 1994]. The reasons for these differences however are not particularly clear, most of the native (Ca^{2+}-bound target free) structures are found at 60° although the target bound and inhibitor structures exhibit χ_1 angles over the full range. In the case of Ile63 the situation is similarly complicated with a range of different orientations but with no definite pattern. Several of the native structures have χ_1 and χ_2 values of 150° and -75° respectively although there are others

including the 1.0Å structure **1exr** with χ_1 and χ_2 of -60° and -30 ° respectively. Once again the target bound structures exhibit a range of different values. Thus it is difficult to conclude whether these residues play a direct role in the specification of the N-term CaM target binding surface. The location of these residues at the hinge point of the N-term lobe suggests that they may play a significant role in the degree to which the N-term lobe opens up in its accommodation of different CaM binding domains (as demonstrated in Chapter 3). The significant variation in the χ_1 angles of residues 74-76 can be attributed to their location in the middle of the highly flexible linker helix and the large solvent accessibility of these residues.

The high flexibility in χ_1 angles of Asp78 and Asp80 is intrinsically associated with the role of these residues in the hinge point of the flexible tether. A deletion mutant constructed without residues 79 and 80 was found incapable of activating various CaM targets due to the inability of the CaM lobes to come together and encompass the target peptide [Tabernero *et al.* 1997]. The only other residue with significant χ_1 flexibility is Lys115 which is again a long flexible side chain exposed to solvent.

7.2.2.4 X_2 Rotamers

When looking at the variability of χ_2 across the different side chains, the flexibility of residues interacting with CaM targets begins to become apparent. Glu14 and Leu18 exhibit the greatest variability in the first half of the N-term lobe and both are involved in target interaction. Similarly Met36 and Gln41 are both quite variable in χ_2 and involved in target interaction in **2bbn**. Gln42 and Glu45 are also variable in χ_2 although neither is directly involved in target interaction. Gln42 is largely directed towards solvent interactions while Glu45 has an interesting role. In one orientation the side chain of Glu45 points outwards into the solvent as in the inhibitor bound structure **1a29.** In all the **native** structures[8] however, the side chain forms an intra-molecular hydrogen bond with Thr29 in the adjacent helix. Although this cannot be correlated with an observed conformational change (see **Figure 7.10** below), the intra-molecular connection may be important in the stabilisation of the open conformation of the N-term lobe of CaM in the absence of target peptides or other co-factors. Glu45, Thr26 and Thr29 are conserved across the sequences of the structures in the ensemble.

[8] See Chapter 3 or the Appendix for the classification of the CaM target association states.

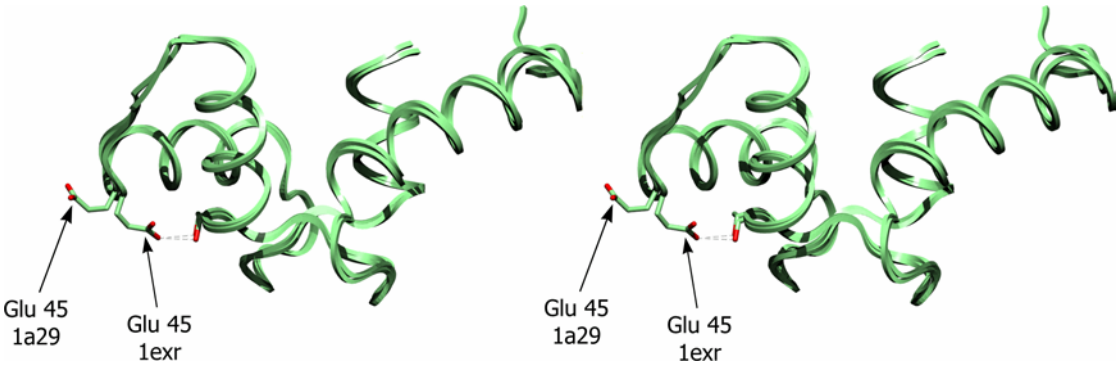


Figure 7.10: *WALLEYE stereo protein cartoon highlighting the alternate conformations of Glu45 in the native structure* ***1exr*** *and the inhibitor bound structure* ***1a29****. The resultant intra-molecular Hydrogen bond is found only in target free and ligand free structures.*

The χ_2 angle of Asp95 exhibits considerable variability and inspection of the protein structure suggests that this side chain is able to make a range of intra-molecular interactions with adjacent side chains or indeed solvent molecules. There is no apparent conserved interaction which can be identified as structurally essential to any of the different target association states. Variability in χ_2 for Leu105 appears to be similar to that of Val55 and Ile63 discussed above. This side chain forms a similar pairing with Val121 on an adjacent helix. The scatter plot for this residue suggests that the conformation is relatively well conserved for the native structures (with the exception of **4cln**) but the presence of target fragments and inhibitors can induce various different rotamers in χ_2.

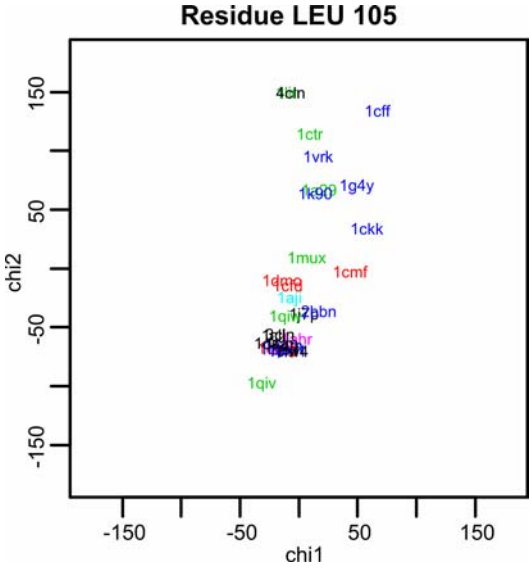


Figure 7.11: *Scatter plot indicating the distribution of χ_1 and χ_2 values for Residue Leu105. Positions are identified by PDB accession codes and coloured according to structural classification .*

Residues Asn111, Leu112 and Glu114 are all located on the edge of the CaM binding site. These side chains are carefully positioned for subtle changes in conformation to accommodate different target groups. Glu114 was identified in the target contact analysis of 2bbn to be involved in interaction with the target peptide and the adaptability of this residue to form interactions with either target groups or the solvent is likely to be another important component of the CaM system. The scatter plot (**Figure 7.12**) again suggests that the side chain orientation for this residue is relatively well maintained for the native structures but exhibits various different conformations in the presence of different co-factors.

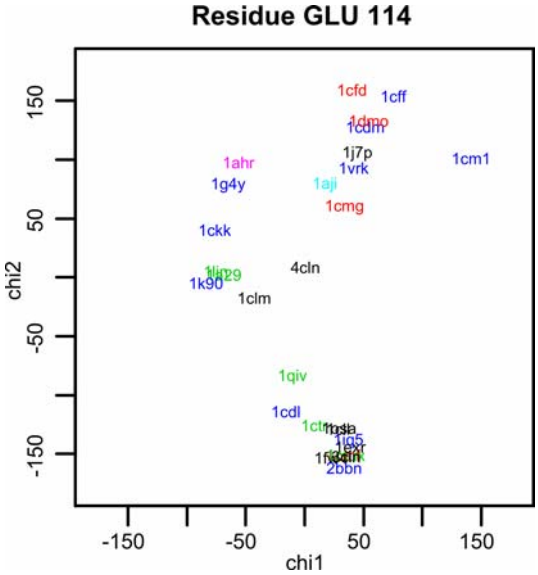


Figure 7.12: *Scatter plot indicating the distribution of χ_1 and χ_2 values for Residue Leu105. Positions are identified by PDB accession codes and coloured according to structural classification .*

7.2.2.5 X_3 Rotamers

As the χ_3 angle requires a γ and δ atom in the amino acid side chain, it is only going to be the longer and inherently more flexible amino acids that are considered. Here the distinction between flexible residues such as Gln49 which interact with the solvent and Met51 which is directly involved in target interactions is quite clear. It is also in this degree of freedom in which the Met residues are at their most flexible. Consequently the variation in binding site residue χ_3 values is dominated by the effects of the Met side chains. Glu11 and Glu54 however also play a vital role, positioned at opposite sides of the CaM binding surface. Clear contributors to the characteristics of the CaM binding surface on the C-term lobe are residues Glu84, Glu87 and Glu123, again at opposite sides of the binding surface and are seen to have a broad range of χ_3 values.

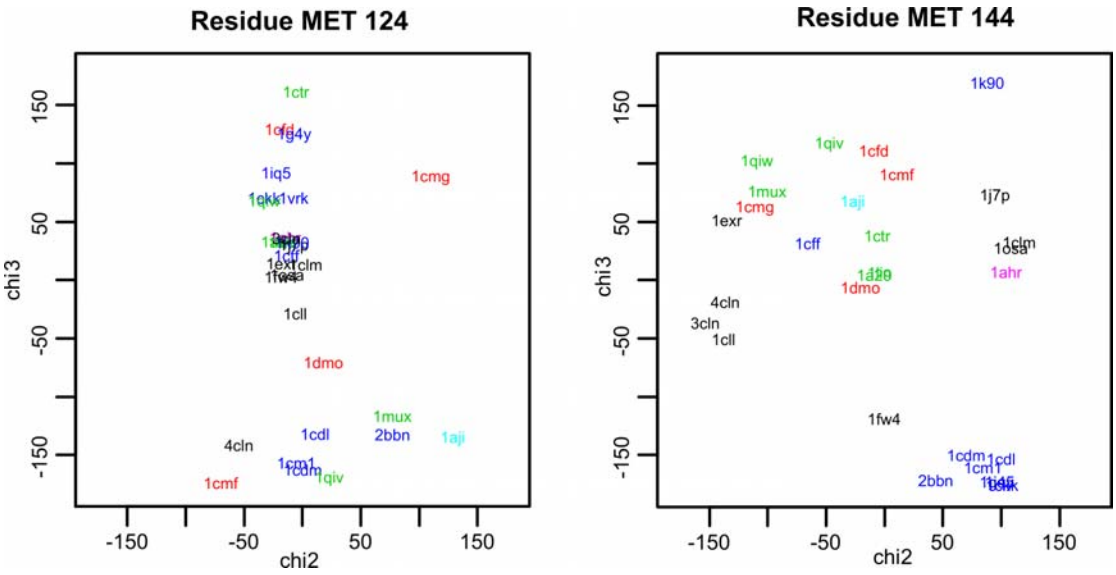


Figure 7.13: *Scatter plots indicating the distribution of χ_2 and χ_3 values for Residues Met124 and Met144. Positions are identified by PDB accession codes and coloured according to structural classification.*

In the above figure, two very different types of Met side chain behaviour are described. The conformation of Met124 appears to be well conserved for the native CaM structures with the exception of **4cln**, and the variation in χ_2 angle is also quite limited. The χ_3 data for Met124 is highly variable across the ensemble of different target and inhibitor bound structures. In contrast Met 144 is highly flexible in both χ_2 and χ_3 angles, particularly across the native structures. Several target-bound structures exhibit a similar combination of χ_2 and χ_3 however there is significant variation in χ_1 across these structures and in the backbone so the overall orientation is not conserved.

7.2.3 Conclusions

This study addressed the variability of the side chain conformations in CaM across the structural ensemble. The analysis was also performed in the absence of NMR structures to analyse the distributions of the data without the influence of the less precise information provided by the NMR experiment. Scatter plots for χ_1, χ_2 and χ_2, χ_3 for all 148 residue for the full structural ensemble and crystal structures alone were generated and analysed but were not included in this thesis due to the volume of information.

This analysis has aided in the identification of certain key residues in the protein which form interesting intra-molecular interactions under certain conditions which may have significant structural implications. Mutation experiments to probe the particular roles of residues highlighted above such as Glu45 and the Val55 Ile63 pairings may provide interesting results on the role of these interactions. The importance of the acidic residues arranged around the edges of the hydrophobic binding surfaces of both CaM lobes has also been highlighted. Once again, this provides good motivation for the mutation studies to investigate the significance of these residues in target peptide binding. The high level of sequence homology between CaMs of different origins also highlights the importance of the majority of side chains in CaM function.

More generally the flexibility of the Met residues has been illustrated here, particularly in the χ_3 values. It seems clear that ability of these side chains to accommodate a multitude of different conformations is indeed exploited by CaM albeit to different extents. The residues Met71 and Met72 in the N-term lobe and Met144 and Met145 in the C-term lobe exhibited a greater variety of conformations than others in the system. The positioning of these residues immediately adjacent to the hydrophobic pockets of each lobe provides the opportunity to adapt to the characteristics of different targets.

7.3 Structural Analysis of CaM Target-Interactions

Previous studies have highlighted the difficulties in the identification of a characteristic motif for CaM binding [Rhoads & Friedberg 1997; Yap *et al.* 2001]. The introduction to this chapter discusses the role of conformational flexibility in the interaction at a global and local level and section 7.2 above investigates the role of the side chains. In this section the investigation centres on the particular functional groups presented by the target in their interaction with the CaM binding surfaces.

Rather than an investigation of the primary amino acid sequences of different CaM targets which has been performed elsewhere [Rhoads & Friedberg 1997; Yap *et al.* 2001], this investigation examines the interface between target peptides and small molecule inhibitors with the CaM binding surfaces. As such, this represents a combined sequence structure approach to the understanding of the CaM-target interface. Structural superposition of the various target peptide and inhibitor bound structures on the N or C-term lobe of the 1.0Å crystal structure **1exr** [Wilson & Brunger 2000] provides a common reference frame in which to compare the different structures.

7.3.1 Mapping the CaM –target interaction surfaces

Given the flexibility afforded by the range of side chain conformations outlined in the above analysis, it is clear that CaM-target interaction is not characterised by specific interactions between key residues on the protein and the target peptide. Here the description of the binding site presented by the CaM lobes is taken to be to a solvent accessible surface representation provided by the molecular graphics program QUANTA [Accelrys 2001]. This representation is calculated with the use of a 1.4Å molecular probe to search accessible space using a standard algorithm [Lee & Richards 1971] and coloured according to amino acid polarity. Once this solvent accessible surface has been calculated for the protein or peptide fragments, the surface is coloured according to the amino acid polarity of the nearest atom to the surface. This colouring has been used throughout this thesis; acidic amino acids are coloured red, basic residues are coloured blue, polar-

uncharged amino acids are coloured yellow and hydrophobic residues are coloured white. Non-standard amino acids and small molecules are coloured identified in pale green and the Met residues of the protein are highlighted on the binding surfaces with bright green colouring.

These surfaced representations are accompanied by ClustalX sequence alignments for the CaM binding domain peptides. The alignments here are coloured using the amino acid polarity colour scheme described above. The sequences are aligned according to the best match at the hydrophobic locator pin site for the N and C-term lobes respectively resulting from the structural superposition of the CaM lobes. In order to perform this alignment it was necessary to invert the amino acid sequence for the CaM binding domains of **1ckk**, **1iq5** and **1g4y** in order for the all the sequences to be in the same orientation with the left hand side of the sequence being that which interacts with the C-term lobe of CaM and vice versa.

After superposition onto the structure of **1exr**, the CaM lobes are surfaced, coloured and rendered. The ligands for each complex are then rotated through 180° in the y-axis so that both the ligand and the receptor interfaces are visible side by side for each structure. A grid of cross-wires is used in **Figure 7.14** and **Figure 7.16** to provide reference points for the location of the hydrophobic pocket and locator pin in a classical inhibitor structure. Comparisons with other structures can then be made to see how the position of the pocket changes between structures and how the ligands move around relative to the surface.

- **N-term Alignment**

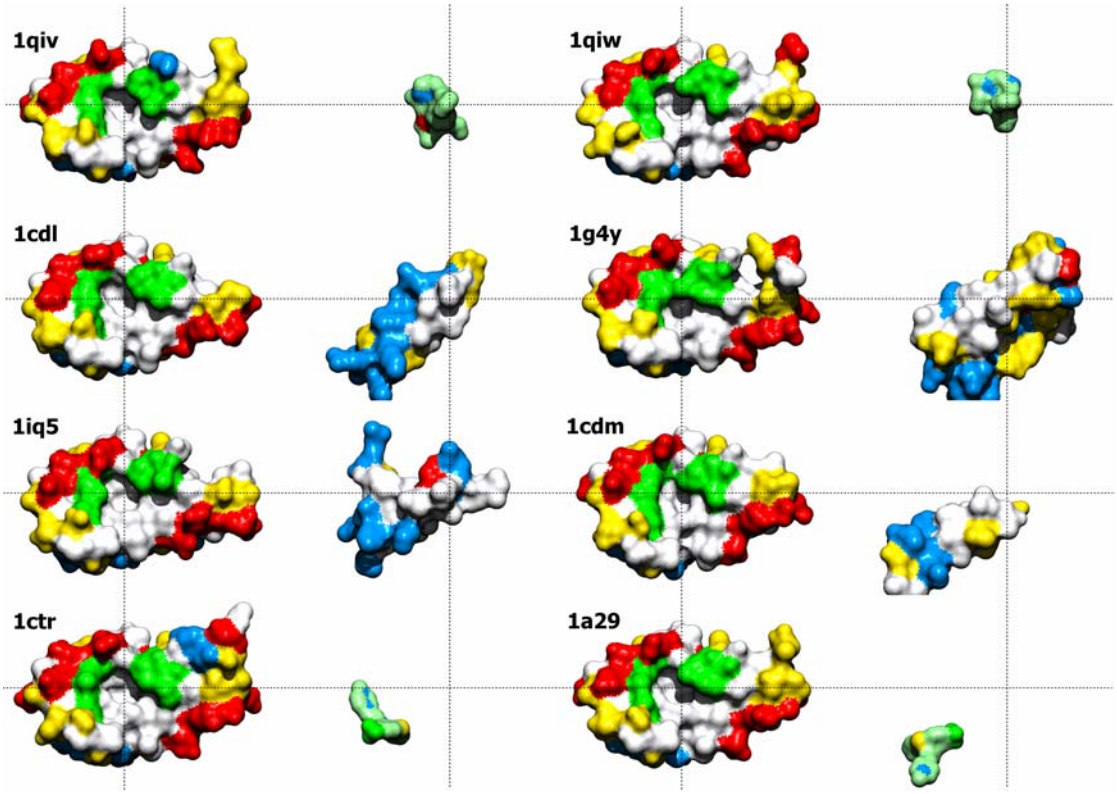


Figure 7.14: *Molecular surface complementarity of the N-term lobes of CaM complex structures. The ligand representation is rotated through 180° in the y-axis relative to the protein representation. Cross-wires are used to define a common reference frame based on the hydrophobic pockets and locator pins in the inhibitor structure* ***1qiv*** *and* ***1qiw****.*

Figure 7.14 above illustrates some of the changes in the CaM binding surface of the N-term lobe as it accommodates the range of different ligands. The inhibitor structures **1qiv** and **1qiw** reveal the large hydrophobic binding pocket which accommodates the aromatic systems of the inhibitor ligands. The Met residues to the right of the pocket (Met71 & Met72) generate a ledge which is a slightly different shape in the two structures although the Met residue lining to the left of the pocket (Met36 & Met51) is more conserved between the two structures. The CaM-smMLCK structure, **1cdl**, also represents a classical interaction. Again the well formed hydrophobic pocket is evident on the receptor with a locator pin presented by a Leu side chain on the target peptide at the conjunction of the cross-wires.

The CaM-camKK structure, **1iq5**, represents a departure from the classical target in that the peptide conformation includes a β-strand section. In this representation however there are no distinct differences from the classical conformations aside from the slightly more extended nature of the target peptide. The locator pin is very evident and the basic residue sticking up on the ligand wraps around the met shelf to the right of the pocket on the receptor. In the structure **1g4y** and **1cdm** the locator pin on the target peptide has moved further down the CaM binding surface and is not as prominent as in the more classical structures. A re-organisation of the met residues on the CaM binding surface shifts the hydrophobic pocket further down the receptor resulting in a much shallower pocket. A 180° rotation about χ_1 of Ile63 also contributes the shallowing of the pocket in **1cdm**. The inhibitor structure **1ctr** and **1a29** illustrate that small molecule ligands do not need to interact in the hydrophobic pocket in order to associate with CaM. In both of these structures the hydrophobic pocket remains well defined although it is not used by the ligands.

In addition to the interactions through the hydrophobic pocket, this subset of the structure also shows regions of conserved hydrophobic character around the surface of the pocket. In the structures **1cdl** and **1iq5** in particular there are extensive regions of basic residues below and to the left of the hydrophobic pin, there are largely accommodated by the acidic regions below and to the right of the pocket on the receptor. These acidic regions on the N-term binding surface are relatively conserved in these representations, however the analysis of the side chain rotamers above (section 7.2) did reveal the ability for considerable flexibility in these side chains.

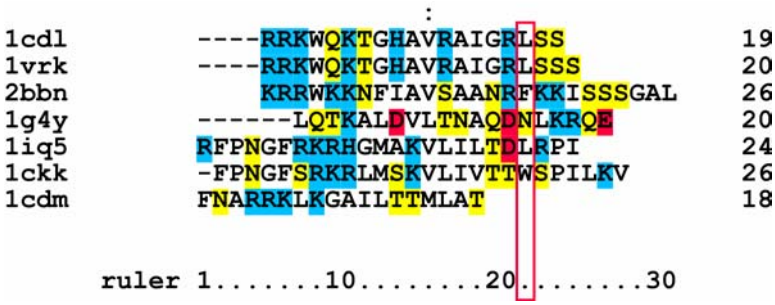


Figure 7.15: *Primary sequence alignment on based on the structural superposition for CaM binding domains interacting with the N-term lobe of CaM. Colouring corresponds to the amino acid polarity. The position of the hydrophobic locator is indicated with the red box.*

The sequence alignment in **Figure 7.15** is based on the structural superpostion performed on the CaM lobes. After the CaM receptors have been superposed, the sequence alignment is constructed on the basis of the position of the Cα-atoms of the peptide fragments at the position of the hydrophobic locator pin. The apparent poor quality of the alignment serves to demonstrate the extent to which CaM can accept different residue types with only limited adaptations to the receptor surface. It should however be noted that due to the helical nature, less than half of the amino acid side chains will interact with the CaM binding surface as some will interact with the C-term lobe, and some will be exposed to the solvent. The red box on the alignment indicates the position of the hydrophobic locator pin on classical CaM targets and the presence of a hydrophobic patch 4-5 residues (1 turn of an α-helix) up stream of the locator pin. The *presence* of the basic residues at the N-term end of the fragments is relatively well conserved however the *position* is not.

It has been noted that CaM targets do not interact with CaM through the peptide backbone [Meador *et al.* 1992]. The helical conformation of CaM targets serves to largely satisfy the Hydrogen bonding capacity of the peptide backbone. This also enables the helical peptides to bind in either direction and make adjustments to the position of the helix on the binding surfaces through the re-orientation of side chains with significant conformational expense.

- **C-term Alignment**

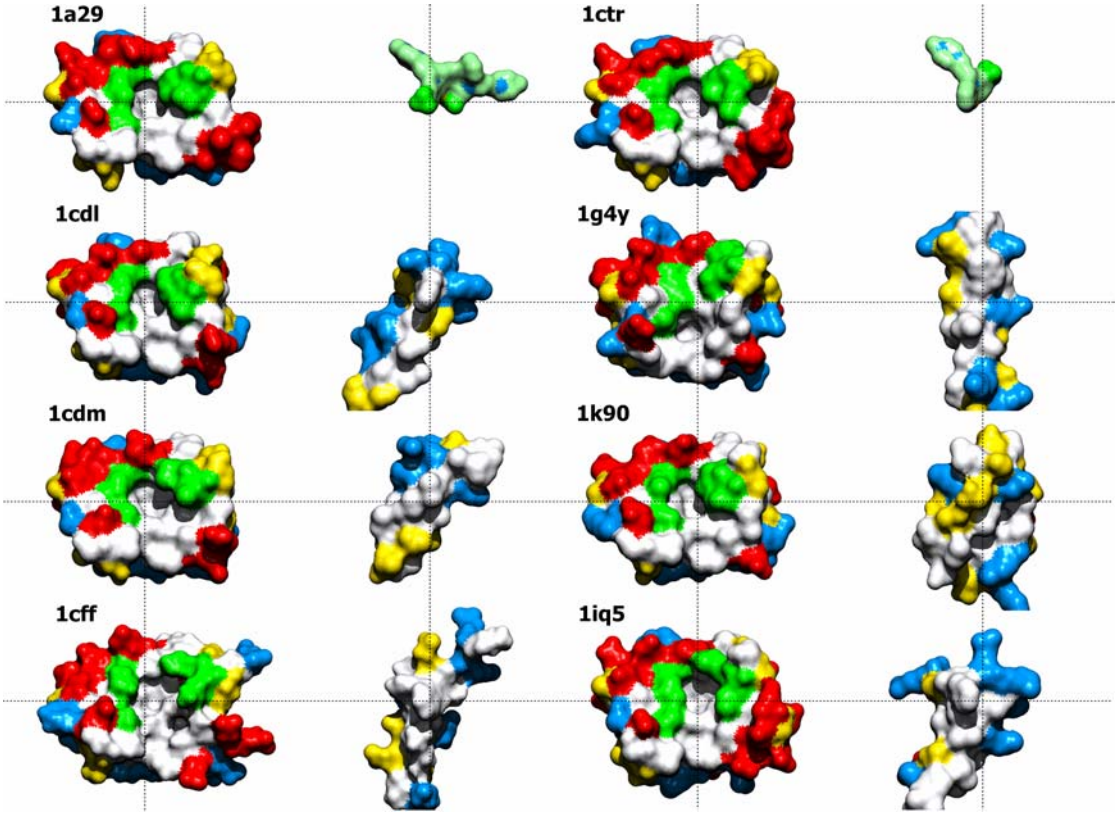


Figure 7.16: *Molecular surface complementarity of the C-term lobes of CaM complex structures. The ligand representation is rotated through 180° in the y-axis relative to the protein representation. Cross-wires are used to define a common reference frame based on the hydrophobic pockets and locator pins in the inhibitor structure* ***1a29*** *and* ***1ctr****.*

The cross-wires on the molecular surface comparison for the C-term lobes of CaM in **Figure 7.16** are aligned on the locator pockets of the inhibitor structures **1a29** and **1ctr** and their corresponding groups on the ligands. Along with the various inhibitor structures, the classical structures **1cdl** and **1cdm** exhibit a large, well-conserved hydrophobic binding pocket surrounded by Met residues. Indeed all the currently available CaM structures featuring a small molecule inhibitor or target peptide present chemical groups which interact with the pocket on the C-term lobe, all except **1g4y**.

The PCA performed in Chapter 3 of this thesis, indicated that the conformation of the C-term lobe of CaM was considerably different in **1g4y** than other crystal structures. The detailed descriptions of the CaM and protein surface in this representation also help to

characterise the structural individuality of this complex. The Trp residue which is optimally position in the target peptide to interact with the pocket on the C-term lobe lies in the plane of the surface and the binding pocket on the receptor is effectively closed by a rearrangement of the Met side chains. A smaller hydrophobic pocket is however formed further down the receptor surface with which a Leu residue interacts from further along the target sequence **Figure 7.17**. Another interesting observation regarding this structure is in the orientation of the target peptide. In the other structures, the peptide sequence is roughly orientated at 45° to the cross-wires. The axis of the peptide interacting with the C-term lobe of CaM in **1g4y** however is parallel to the vertical cross-wire.

The complex illustrated in the structure **1k90** also presents some interesting features. Rather than the standard hydrophobic residue, this target sequence has a Gln residue which is located in the hydrophobic binding pocket of CaM. This representation also suggests that there is little reorganisation of the receptor surface required to accommodate this difference. The Met residues lining the left hand side of the pocket exhibit slightly different conformations to other structures although these differences are not significant. In addition to being a reverse binding sequence, the Adenylate Cyclase CaM binding domain also includes two anti-parallel sequences although only one helix interacts directly with CaM. There is no significant conformational re-arrangement required for this interaction according to PCA of the Cα (data not shown).

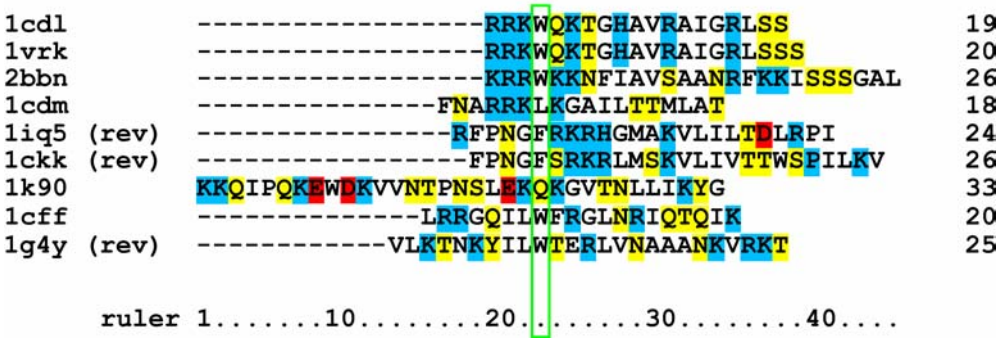


Figure 7.17: *Primary sequence alignment on based on the structural superposition for CaM binding domains interacting with the N-term lobe of CaM. Colouring corresponds to the amino acid polarity used in* ***Figure 7.16*** *above. The position of the hydrophobic locator is indicated with the green box.*

Above the hydrophobic pocket, the presence of a positively charged residue is quite well conserved as illustrated at position 19 in **Figure 7.17**. Most sequences have an extended hydrophobic area below the pocket around position 26 and then a polar residue beyond that around position 30. The structure based sequence alignment illustrated in **Figure 7.17** demonstrates much greater homology across the targets than the N-term alignment. The basic residues upstream of the hydrophobic locator pin are well conserved as are the presence of the hydrophobic and polar regions further downstream although the exact positions are not conserved. This leads to the hypothesis that the motif for the C-term CaM binding surface is more specific than the N-term.

- **General Comments:**

It is generally accepted that interactions of the C-term lobe of CaM with CaM-binding domains are generally of higher affinity than those made through the N-term lobe [Bayley *et al.* 1996]. It is therefore particularly interesting that interactions through the hydrophobic pocket of the C-term lobe are far more consistent across the structural data set than in the N-term. The primary sequence alignments in **Figure 7.15** and **Figure 7.17** must be interpreted with care in this context as adjacent amino acids in the sequence may have entirely different orientations in real space. These alignments do however suggest that there is a better conservation of a basic region to interact with the C-term lobe than can be identified for the N-term lobe. In all the single helical structures (both **1k90** and **1g4y** exhibit two α-helices interacting with CaM) there are basic residues immediately around the key hydrophobic residue for the C-term pocket. These residues point away from the C-term lobe of CaM and interact with either the N-term lobe or the solvent.

Another key observation from this analysis is that there are chemical groups interacting with the hydrophobic pocket of the C-term lobe of CaM in all the ligand and target bound structures. This is clearly not the case for the N-term. In the structures **1k90** and **1cff** and several of the inhibitor structures there are no groups at all interacting with the N-term lobe of CaM. The structures of the complex between CaM and CaMKII illustrate that although the N-term lobe is involved in the formation of the complex, the interaction does not necessarily involve the hydrophobic binding pocket. The alignment of target peptides and ligands on the C-term forms a much tighter ensemble of chemical

groups with directly correlates with suggestions that the C-term forms relatively stronger interactions with CaM targets [Bayley *et al.* 1996].

7.3.2 Identification of Key Chemical Groups in CaM-Target interactions

The technique of MCSS (Multiple Copy Simultaneous Searching) was developed in 1991 by Karplus & Miranker [Miranker & Karplus 1991] to characterise the active sites of proteins through the identification of high affinity small molecule binding sites. Multiple copies of functionally relevant small molecules are distributed throughout a region of particular interest in a protein structure. Energy minimisation is used to identify suitable binding sites around the active site of the protein and clustering techniques are used to "prune" these binding sites resulting in the identification of several ligand binding hotspots and orientations.

This technique was used to find ligand binding hotspots for the Trp side chain on the binding surfaces of the N- and C-term lobes of the 1.0Å structure of Ca^{2+}-CaM, **1exr** [Wilson & Brunger 2000]. The results for the 20 lowest energy arrangements of the Trp side chain groups from this experiment are summarised in **Figure 7.18** below.

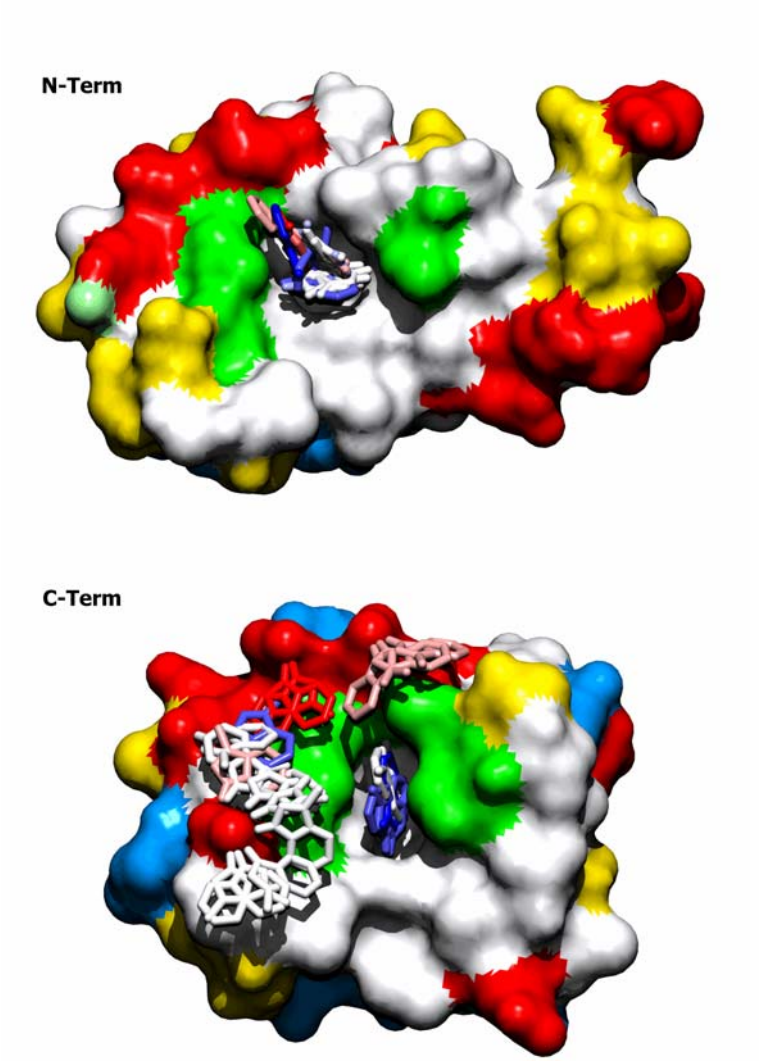


Figure 7.18: *Surfaced representation of the N and C-term lobes of CaM coloured by amino acid polarity. Results of the MCSS calculation illustrated with Liquorice representation of the Trp side chains coloured by energy with the lowest energy conformation shown in red and the highest in blue.*

In the N-term calculations all the lowest energy positions for the Trp group were identified in the hydrophobic pocket with different orientations, the lowest energy of which forms an H-bond with a backbone carbonyl group. Interestingly, none of the lowest energy conformations are buried right down in the bottom of the pocket suggesting that the buried conformation more common in the C-term pocket is not as energetically favourable in the N-term. The C-term calculations show the most favourable interaction sites to be on the surface of the protein with interactions between the NH on the Trp and acid groups around the surface. Only three of the 20 lowest energy arrangements include interaction in the C-term binding pocket. This could be

attributable to the fact that the receptor surface is fixed during these calculations and could not be sufficiently large to properly accommodate the Trp group in the structure **1exr**. If either of the hydrophobic binding pockets are not large enough to accommodate the ligand, there is no opportunity for the protein to adapt to accommodate it.

In order to compare these proposed ligand-binding hotspots with experimental data, the ligand/target bound structural ensemble was used to identify patterns in the arrangement of amino acid side chains on the binding surfaces of the protein. Based on the assumption that those side chains located close to the binding surface of the CaM lobes are involved in key interactions, conservation of particular groups on specific parts of the surface can be used to highlight those key interactions.

- **N-term ligands**

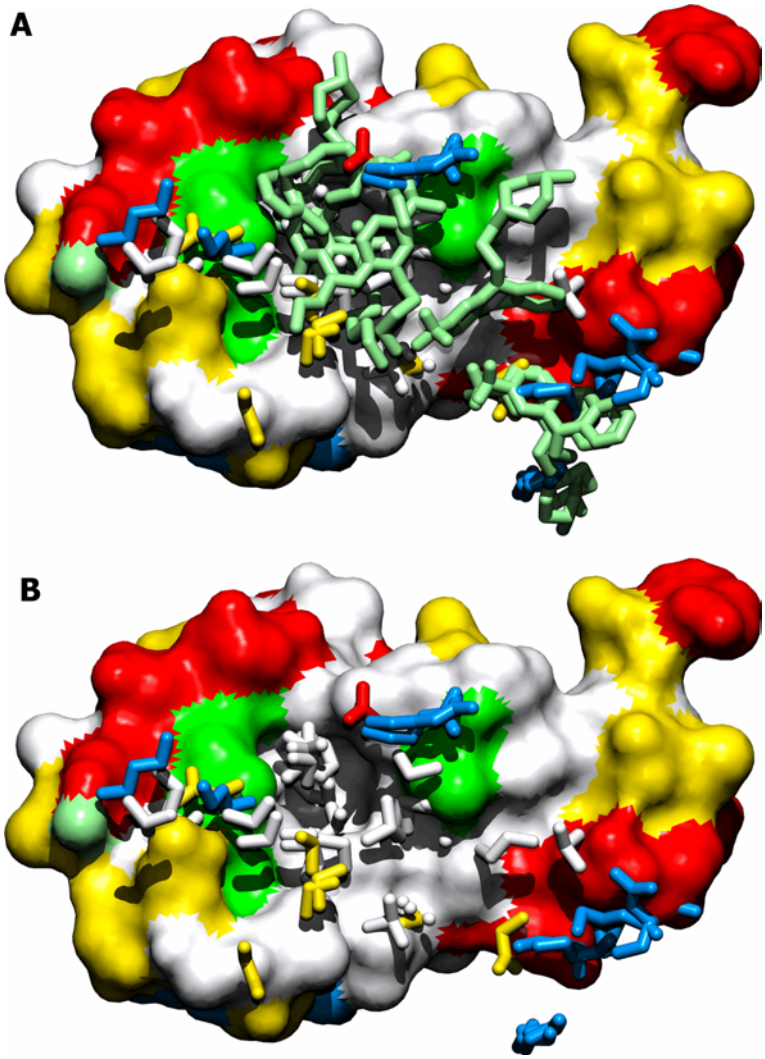


Figure 7.19: *Surfaced representation of the N-term lobe of CaM showing the location of chemical groups within 3.2Å of the protein surface. Panel A illustrates groups from the ensemble of all small molecule and peptide bound structures and Panel B illustrates side chains from the target bound structures only.*

Chemical groups associating with the N-term lobe of CaM are dominated by the large inhibitor groups which are found scattered across the surface from different structures. These inhibitors can be seen to interact with the hydrophobic pocket and other parts of the hydrophobic surface. Removal of the inhibitors from the distribution provides a clearer representation of the amino acid side chain groups interacting with the CaM binding surface from the structure of CaM and CaM-binding domain peptide analogues. Once again a group of hydrophobic residues are clearly seen interacting with the

hydrophobic pockets although the presence of other specific interactions is less clear. There is evidence of hydrophobic side chain groups interacting with the extensive hydrophobic surface although these are not particularly well conserved in any region. The cluster of basic residues towards the bottom left hand corner of the image correspond to the interactions identified in the ligand summary analysis above (Section 7.3.1). In this representation the single acidic group presented in the structure **1g4y** sits rather uncomfortably associated to a hydrophobic group on the CaM surface. This and other apparent mismatches demonstrate the ability of CaM to alter the surface to accommodate different groups across the binding surface.

Comparison of this representation with the MCSS results above confirms that most significant feature the Ca^{2+}-CaM binding surface is the hydrophobic pocket which attracts hydrophobic side chains. The other hydrophobic areas of the binding site are not specific in the accommodation of particular target side chains. Similar experiments using charged ligand probes such as Lys side chains would undoubtedly identify potential interaction sites similar to those represented here by the actual interaction sites for the charged side chains of the different CaM targets.

- **C-term Ligands**

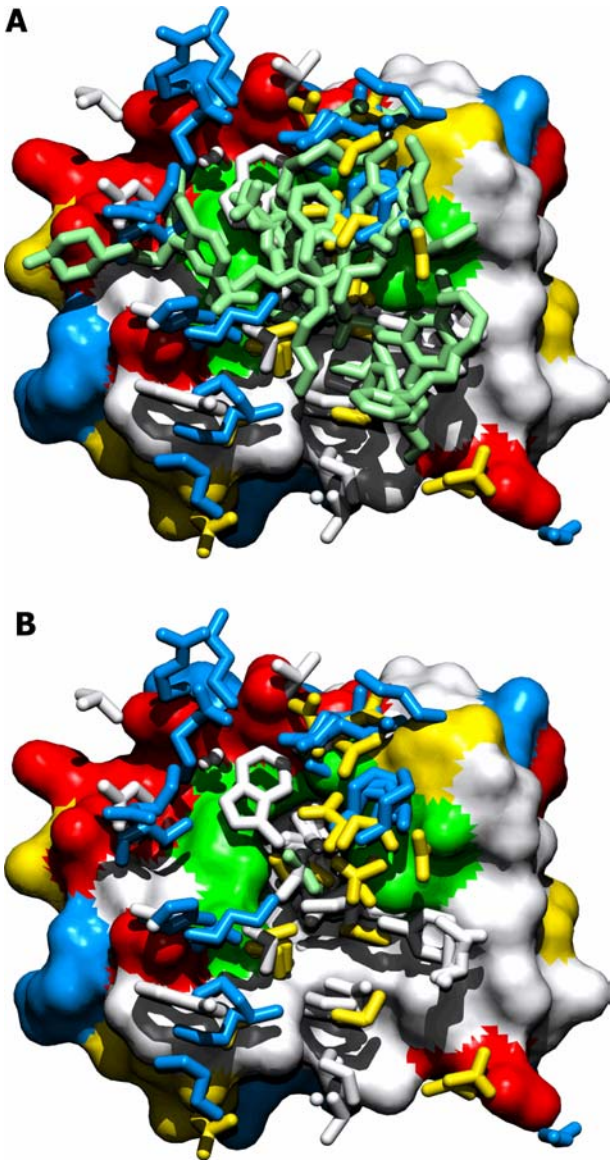


Figure 7.20: *Surfaced representation of the C-term lobe of CaM showing the location of chemical groups within 3.2Å of the protein surface. Panel A illustrates groups from the ensemble of all small molecule and peptide bound structures and Panel B illustrates target bound structures only.*

Once again the interaction with the C-term lobe binding surface are dominated by the ligands from the inhibitor structures. Removal of these ligands reveals a plethora of different interactions between CaM and the various target peptide side chains. Only the interactions with the hydrophobic pockets are clearly apparent and well conserved across the structural ensemble. Other interactions appear to be less specific. This representation suggests that the interactions between the C-term lobe of the protein and the targets are

far more numerous, an observation consolidated by the fact that two structures (**1k90** and **1cff**) reveal a target peptide which interacts with the C-term lobe alone. This representation suggests that the acidic side chains distributed around the edges of the C-term binding surfaces are well used to accommodate the basic side chains presented in the different targets. It is also possible that these residues play an important role in controlling the initial stages of target association due to the long range effects of electrostatic interactions prior to the formation of the much shorter range Van der Waal hydrophobic interactions.

The presence of a Trp residue interacting with the surface rather than the pocket of the C-term domain is particularly interesting in the light of the results of the MCSS calculation which suggests that the most favourable Trp interaction sites are on the exposed surface of the lobe. This side chain, presented in the structure **1g4y**, does not however make the hydrogen bond suggested by MCSS which stabilises this conformation. The Glu120 residue involved in the lowest energy complexes from the MCSS could re-arrange to make a hydrogen bond with the Trp NH group on the target peptide although this is not observed in the crystal structure.

7.3.2.1 Amino acid sequence alignments

Having identified the residues which are likely to be involved in interaction with the CaM binding surface in each of the target bound complexes, the position of these amino acids in the sequence alignments of target peptides was investigated. The alignment based on the position of the C-term hydrophobic locator pin, illustrated in **Figure 7.17** above was annotated with the residue contact information from the ligand binding hotspots experiments above. **Figure 7.21** and **Figure 7.22** below show sequence alignments of the different target peptides with sequences for **1iq5**, **1ckk** and **1g4y** reversed to preserve the overall orientation. Note that two separate sequences must be considered for **1g4y** as the different lobes of CaM interact with different peptide sequences.

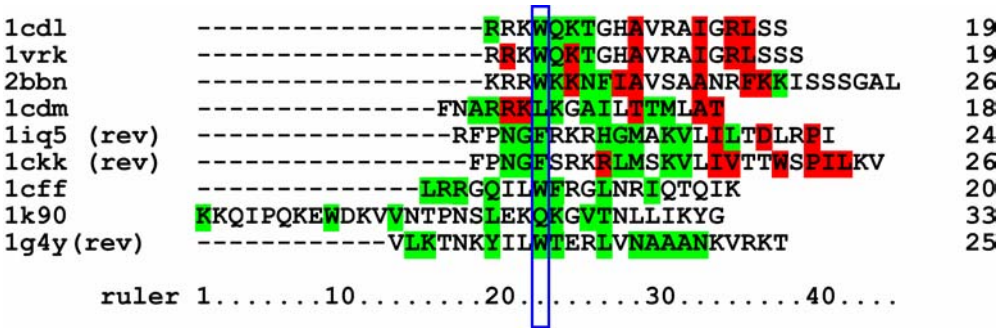


Figure 7.21: *Primary sequence alignment of CaM target peptides based on the structural superposition for CaM binding domains interacting with the C-term lobe of CaM. Colouring corresponds to the residue contact information obtained from the experiments above. Residues interacting with the C-term lobe are highlighted in green and those interacting with the N-term lobe in red.*

In this representation the organisation of residues interacting with the C-term on the left and N-term residues on the right is quite evident. The distribution of C-term interacting residues on the peptide fragments is conserved for the hydrophobic binding pocket although there is no other single position where in which a residue from each structure interacts with the C-term lobe of CaM. Position 19 is relatively conserved with the Pro residues in **1ckk** and **1iq5** being too short to touch the protein surface, although the basic residues in **1vrk** and **2bbn** are also not identified as contact residues. The helical nature of the targets is most evident in **1cdm**, with periodic contacts with the C-term and the N-term; starting at position 18 there are two interactions with the C-term, two with the N-term, two with the C-term, one absence, etc. It is interesting to note however that there is a five residue stretch in **1g4y** which all interact with the C-term lobe.

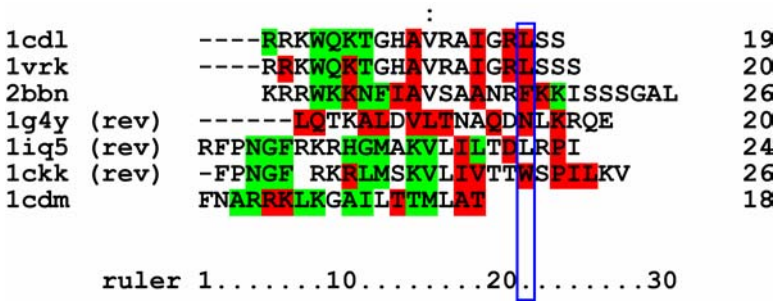


Figure 7.22: *Primary sequence alignment of CaM target peptides based on the structural superposition for CaM binding domains interacting with the N-term lobe of CaM. Colouring corresponds to the residue contact information obtained from the experiments above. Residues interacting with the C-term lobe are highlighted in green and those interacting with the N-term lobe in red.*

The alignment of the N-term hydrophobic pins shown in **Figure 7.22** again retains the general organisation of most of the N-term contacts on the right hand side and the C-term contacts on the left. The position of the hydrophobic locator pin, indicated in blue, shows that not all sequences interact with the pocket, the CaMKII sequence in **1cdm** stops short of the pocket and it seems that the Leu side chain in **1iq5** is too far away from the reference structure to register an interaction. Again there are no other positions in the sequence where contacts are conserved across all the targets. This is almost achieved at position 18 although the Leu residue in **1iq5** is in closer proximity to the C-term lobe than the N-term.

If the residues interacting with the respective lobes of CaM identify the most important binding sites on the protein lobe surfaces then an alignments of these contacting residues should result in the identification of a true motif for each of the CaM lobes. The sequence alignments presented in **Figure 7.23** below were performed with *ClustalX* using default alignment parameters.

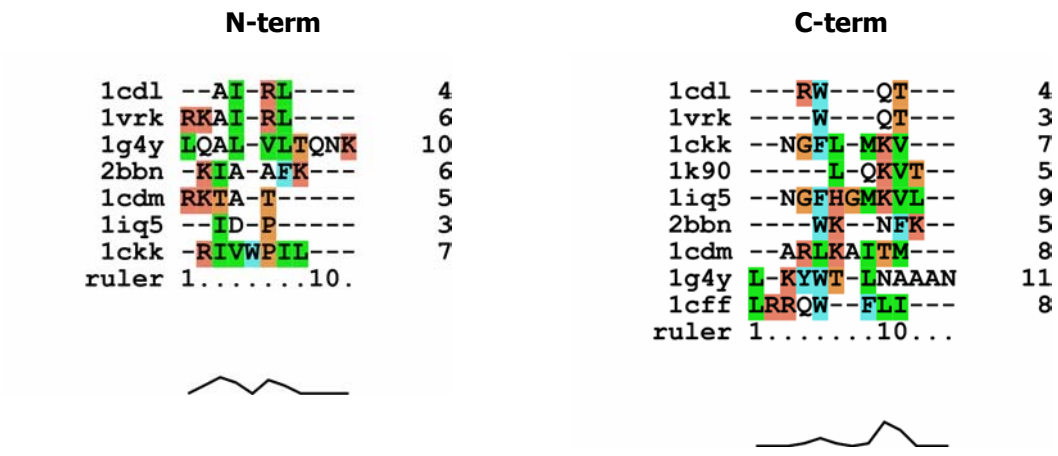


Figure 7.23: *Primary Sequence alignments of the residues contacting the N and C-term lobes of* ***1exr****.*

The alignment of the N-term residues identifies a conserved hydrophobic residue at position 3 of the alignment and a largely hydrophobic alignment at position 4 although this incorporates the Asp from **1iq5**. This alignment clearly does not correctly identify the hydrophobic locator pin residues which are spread throughout the distribution. The alignment of the C-term residues is slightly more successful with the correct alignment of the hydrophobic locators pins at position 5 in all except **1k90** which should have the Gln

at this position. Every single other position however has a least one mismatch on the basis of residue polarity (basic, acidic, polar or hydrophobic). Even with this restricted amount of information the identification of a conserved target motif for either of the CaM domains has been largely unsuccessful.

7.3.3 Normal Mode Trajectories

The PCA performed in Chapter 3 of this thesis outlined conformation changes which were found to correspond to different target association states of the CaM lobes. Projection of the minimised structure along the first normal mode of the N-term analysis and second mode of the C-term analysis were found to highlight similar types of conformational flexibility to those exhibited in the observed crystal structures between native and target bound states. Here the modes are scaled to describe motions at an artificially high temperature of 1000K to accentuate the deformations described by the modes. Structures for these projections are surfaced, rendered and coloured according to amino acid polarity to highlight the conformational changes occurring at the CaM binding surfaces in these normal mode descriptions.

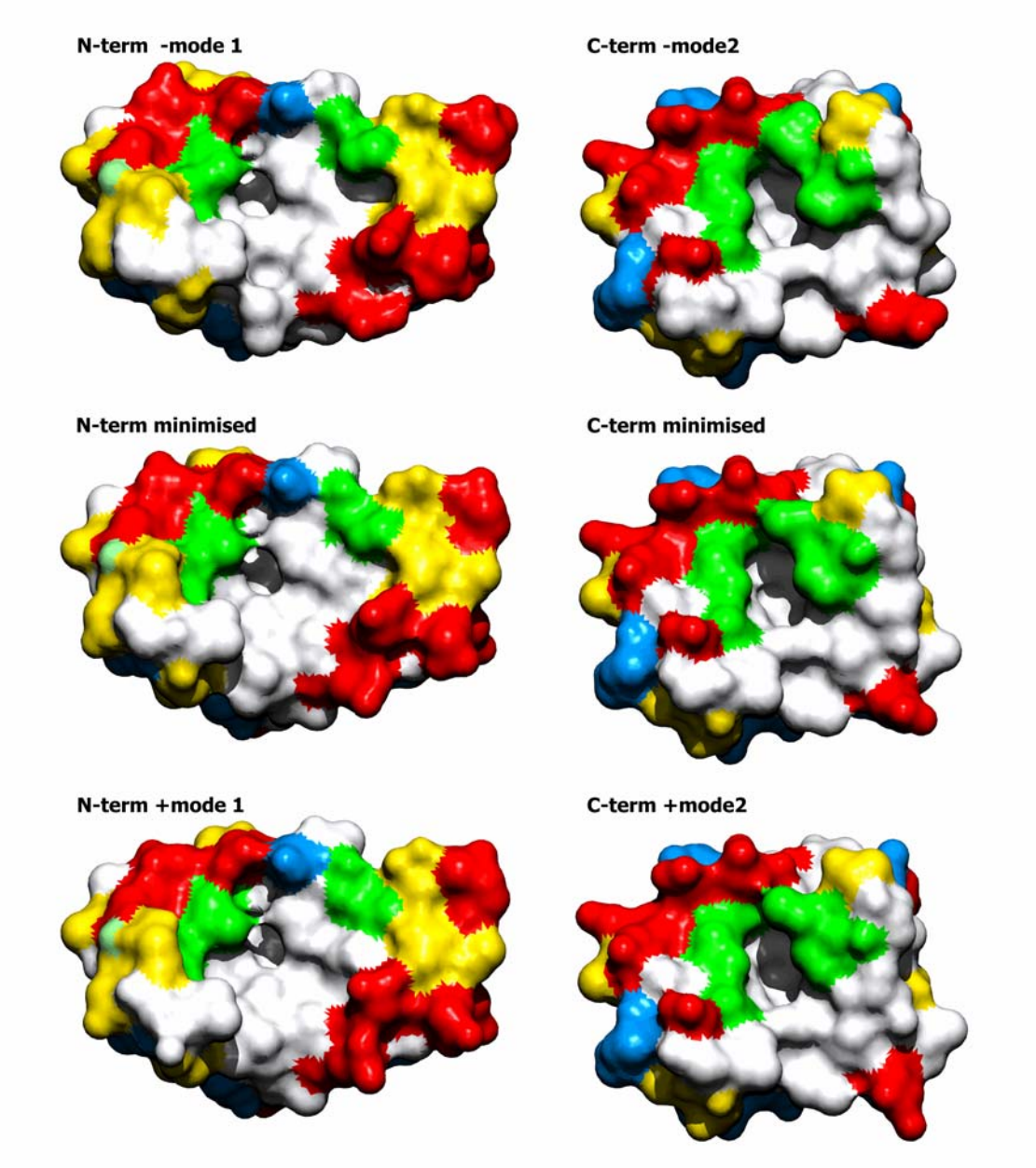


Figure 7.24: *Projections of the minimised structure from the Normal mode analysis of the 1.0Å structure of CaM,* ***1exr****, along mode 1 of the N-term and mode 2 of the C-term. Solvent accessible surfaces for the structures are coloured according to amino acid polarity with met residues highlighted in green.*

It is clear from the representation in **Figure 7.24** that the deformations on the surface of CaM are limited even at an artificially high temperature of 1000K. Some of the characteristics however can be determined from this representation. The most mobile parts of the structure in both lobes are the solvent accessible residues situated around the edges of the binding sites. The acid group at the bottom right hand side of the C-term

lobe is seen to be highly mobile. The other region of flexibility in this lobe is the Met flap covering the right hand side of the hydrophobic pocket, although the pocket itself does not seem to change shape or size significantly. The same can be seen in the N-term lobe, although the process of minimisation has resulted in significant changes to the hydrophobic binding cleft on this lobe. The most mobile parts of this structure are again in the acidic patch on the bottom right of the lobe and the Met residue above and to the right of the hydrophobic binding pocket. In the top structure, the Met side chain protrudes significantly into the binding surface however in the bottom image this is almost flat.

At this level of detail it is impossible to make valuable comparisons between the behaviours of the N and C-term lobes of the protein. One of the more important observations however is that the minimisation of the N-term CaM fragment prior to NMA clearly results in a closing of the hydrophobic pocket. The PCA performed in these structures indicated that the minimised structure of **1exr** was significantly displaced along a PC which described the *closing* of the CaM domain towards a more apo (Ca^{2+}-free like) conformation.

7.4 Discussion & Conclusions

It is clear that CaM manages to interact with a wide range of highly diverse target sequences using several different types of conformational flexibility. The global conformational flexibility afforded by the central tether allows relative re-orientations of the CaM domains to interact with separate CaM binding motifs on a single target molecule with significant spatial separations. The local conformational flexibility of the CaM lobes themselves, which is probably associated with the Ca^{2+} state of the lobe, facilitates significantly different characteristics on the CaM binding surfaces. This flexibility is reinforced by the plasticity of the CaM binding surfaces themselves afforded by met and other flexible side chains which interact with the various targets. These generally accepted findings have been reinforced by the observations detailed in this chapter.

Detailed analysis of the side chain flexibility through the variation in chi angles highlights the anticipated flexibility in Met residues in particular. Comparison of this flexibility with the solvent accessibility and contacts made in the structure **2bbn** serve to highlight those residues which are flexible across the ensemble of protein structures due to target interaction involvement rather than solvent interactions. Interestingly, this analysis suggested that the Met residues are not universally flexible. Residues 71 and 72 on the N-term and 144 and 145 on the C-term lobe, which form a flap over the hydrophobic binding pockets in the respective lobes, are more flexible across the structural ensemble that the other Met residues on the CaM binding surfaces. Detailed comparison of the available structures of different complexes highlighted the way in which this flexibility was utilised to change the accessibility of the pocket. In the extreme case of the structure **1g4y** however, Met124 plays an interesting role by completely blocking off the hydrophobic binding cavity in the C-term lobe as illustrated in **Figure 7.16**. This exceptional behaviour suggests the structures determined so far have demonstrates only a limited degree of the plasticity accessible in the CaM-target interface. This structure also demonstrates that it is possible for the CaM lobe to adopt a *semi-open* conformation and yet still exhibit a fully exposed hydrophobic binding surface. The analysis presented here serves also to identify some key residues which could be used to target mutation studies for a fuller characterisation of the binding surfaces.

The studies performed in this analysis have helped to demonstrate the independent characteristics of the CaM lobes. The publication of structures in which only the C-term lobe of CaM interacts with target peptides (**1k90**, **1cff**, and several inhibitor structures) has added weight to the suggestion that CaM may indeed be attached to its targets at resting Ca^{2+}-levels prior to activation [Martin *et al.* 2000]. The detailed inspection of the available CaM complex structures also indicates that more interactions occur between the C-term lobe of CaM and targets than the N-term lobe. This observation correlates with suggestions that the C-term lobe of the protein forms the most thermodynamically significant part of the interaction [Bayley *et al.* 1996].

Attempts to identify a characteristic target motif for either of the CaM lobes on the basis of the structures currently available have highlighted the complexity and particularly the diversity of CaM-target interactions. A lobe-based analysis of the full plethora of CaM targets however may serve to identify separate putative motifs for the N and C-term

lobes. After the identification of a residue that interacts with one of the CaM lobes, a framework of residues predicted to interact with the N and C-lobes respectively could be devised to treat the two surfaces separately [Afshar *et al.* 1997].

The *Holy Grail* of studies of CaM target interactions is probably a full characterisation of a consensus motif for a CaM binding domain for each of the CaM lobes. This is however immediately complicated by the different conformational states (*open*, *semi-open* and *closed*) each of which may have a different motif. The analysis performed here on the range of available CaM complex structures has made little impression on this goal. Several key features of the interaction have been highlighted and these considerations of CaM will only be confirmed or refuted as more structural information becomes available.

Chapter 8: Conclusions & Further Work

8.1 Calmodulin Structural Variability

8.1.1 Aims & Objectives

In Chapter 3 of this thesis, the available structural information on Calmodulin (CaM) and some of its structural homologues was analysed using Principal Components Analysis (PCA) to identify the underlying trends of variation in CaM structure. The conformations of the N and C-term lobes of the protein under different target association states and at different extents of Ca^{2+}-saturation were compared in order to understand the key features of flexibility in CaM conformation.

8.1.2 Overview of Results & Conclusions

Comparison of CaM structure with proteins of the same superfamily with the same **domain identifier** (what does this mean?) highlighted the similarities and differences between CaM and its homologues. The general fold of a CaM domain, featuring a pair of helix-turn-helix motifs can be identified in various different systems. The length of the helices in this motif however, and also the nature of the linker between the two EF-hands are relatively unique features of CaM. Only the Troponin C structures features EF-hand motifs with helices of the same length and a similar linker to CaM.

The PCA of the CaM structural dataset highlighted the differences between the Ca^{2+}-bound and the Ca^{2+}-free conformations of the protein. Projection of structures from the same lobe of the protein onto the subspace described by the principal components (PC's) of the structural ensemble revealed that the conformations of CaM in the different Ca^{2+}-states could clearly be distinguished.

Comparison of the crystal structures of CaM identified more detailed conformational differences. With the more structurally diverse conformations of the protein removed

from the analysis it is possible to identify differences in conformation directly attributable to the target association state of the protein. Conformational changes within a lobe of the protein corresponding to the different Ca^{2+}-states are of the order of 5.0Å RMSD whereas the conformational changes resulting from the association of CaM with a target peptide or small molecule inhibitor closer to 1.0Å for the Cα atoms. At this level of detail it is also possible to identify minor conformational differences between the conformations of the protein with *different* ligands. The PCA of the crystal structures clearly demonstrated that complexes with target peptides exhibited greater conformational change in the lobes than the small molecule inhibitors and that different ligands effected slightly different conformations. Crystal structures of CaM from a range of different species show remarkably similar conformations for the protein. This supports the suggestion that conformational variation in CaM is due to different environmental conditions of the protein rather than experimental differences.

One of the more surprising observations of this analysis was the similarity between the two lobes of the protein. The conformation of the Ca^{2+}-bound form of the N-term lobe of the protein was very similar to that of the C-term, and differences between target association states were more significant than between the different lobes of the protein in the same target association state. Much of the analysis in this thesis is suggestive of differing roles for the two lobes of CaM; it is evident that the two lobes have different affinities for both Ca^{2+} and many target peptides [Barth *et al.* 1998]. Differences in the conformational flexibility of the two lobes with respect to one another are however more apparent --- with the N-term lobe of the protein exhibiting more flexibility than the C-term across the ensemble of structures.

The comparison of the results from the (computed) normal mode analysis (NMA) of the protein lobes and the observed variability in the structural ensemble showed interesting similarities. The high correlation of the lowest frequency modes with the most significant PCs did suggest a significant overlap in conformational space described by the two techniques. From this result it can be inferred that the conformational variation of the N and C-term lobes of CaM under the influence of a variety of factors (Ca^{2+} load, target association, etc) lies principally along intrinstically low energy deformation modes. Further, a significant proportion of the conformational variability can be described by relatively few degrees of freedom, representing global collective atomic displacements.

8.1.3 Potential Extensions of the Investigation

This analysis has proved useful in the identification and characterisation of the variation in CaM conformation from the available structural data. As additional structural information becomes available, it would be useful to repeat these investigations with the incorporation of the new data. Structures of CaM in complex with new peptides and inhibitors will undoubtedly serve to expand the boundaries of the conformational space described for the protein. Some structures will add additional weight to the conformations already described for the protein while others will suggest new conformations which were previously unidentified. As different conformations become better represented it should become possible to extend the analysis from considerations of the Cα atoms alone to identify the detailed local effects of individual residues in different target proteins on the conformation of the CaM. This may in turn help to explain how CaM is able to accommodate so many different target peptides and bind them with such a high affinity.

8.2 Molecular Dynamics Force Field Comparisons

8.2.1 Aims & Objectives

Macromolecular simulations are based on theoretical principles and rational approximations to observed physical behaviour, as outlined in the second chapter of this thesis. In order to have confidence in the results of these simulations, it is therefore necessary to assess the models and methods themselves before attempting detailed interpretation of the results obtained from them. The parameterisation of the CHARMM force field has been the subject of a great deal of effort in order to ensure that calculations using the potential energy function are as accurate as possible [Neria *et al.* 1996; MacKerell *et al.* 1998]. There are however many different ways to set up simulations to investigate the conformational flexibility of CaM and it is therefore necessary to invest a certain amount of effort to ensure that the appropriate approach is

adopted in these calculations. Despite considerable advances in computer technology and affordability, the simpler the experiment that can be used to probe the molecular dynamics of CaM, the greater the conformational sampling that can be obtained. It is therefore of interest to adopt the most reasonable experiment which can be used to generate the most reliable and extensive conformational sampling for this system.

Analysis of the results from the different types of simulation focussed on some overall properties of the system in order to get a feel for the reliability of the results of the different types of experiment rather than the detail of the results being obtained. The use of the standard measure of the deviation (RMSD) from the intitial (observed) structure was supplemented by the radius of gyration (RGYR) data and an analysis of the Ca^{2+}-coordination analysis to provide additional information important to the features of this system.

8.2.2 Overview of the Results & Conclusions

The overall conclusion that the greater the complexity of the model, the more reliable the results was not immediately apparent and only attained after a full consideration of all the results. The RMSD and RGYR data clearly demonstrated the importance of a fully solvated protein in these investigations with the experiments separated into two separate classes. **Class II** experiments in which the protein was simulated in a water droplet of minimum depth 5.0Å demonstrated more consistent behaviour throughout the simulations, although the reasons for this are far from certain. **Class I** simulations on the other hand, and in particular the use of the EEF1 implicit solvent model [Lazaridis & Karplus 1999], demonstrated more significant conformational variations. This was thought to suggest that the protein interactions were not being appropriately represented in these **Class I** experiments.

The simple monitoring of the Ca^{2+}-binding distances in the protein over the course of the MD simulations revealed interesting properties of the different types of experiment. It was perhaps unsurprising to note that the use of the all atom PARAM27 force field resulted in Ca^{2+}-coordination geometries which were closer to those observed in the

crystal and more closely regulated over the course of the simulations than those of the trajectories using the extended atom PARAM19 force field. It was however quite unexpected to discover that the solvent-free simulations resulted in Ca^{2+}-coordination geometries which were highly maintained throughout the trajectories. These distances were not as close to those of the initial crystal structure as those obtained from the PARAM27 potential, but whether this was the cause or effect of the more variable protein conformations in the solvent-free simulations is unclear.

It was therefore concluded from these investigations that more sophisticated representations were necessary for the realistic simulation of the CaM fragments. The consequence of this conclusion is that the conformational sampling accessible in the subsequent MD simulations of CaM would be limited to a shorter timescale with fewer independent runs than would be ideal.

8.2.3 Potential Extensions of the Investigation

MD methods are continually being developed and new protocols are being refined for simulations of biomolecules. As the power of computers inevitably advances the scope for more sophisticated and complex models broadens. Such models will undoubtedly result in different trajectories and possibly provide more accurate information on the conformational flexibility of CaM as they become available. One obvious improvement which could ideally be applied to this system is the use of the *Particle-Mesh Ewald* Summation [Leach 1996]. In this method the calculation of the long-range electrostatic interactions is efficiently performed. This would be particularly helpful for a more detailed characterisation of the behaviour of the high charge density in the Ca^{2+}-binding sites. Such simulations have already been performed on CaM [Yang *et al.* 2001] although the accessible timescales are somewhat restrictive. Other techniques which could be applied to this system include the use of hybrid Quantum Mechanics / Molecular Mechanics (QM/MM) methods. This method involves a full calculation of the electrostatic structure of an area of interest such as the Ca^{2+}-binding sites coupled to a standard molecular mechanics treatment of the residual protein fragments. Of greatest interest would be very long (microsecond and beyond) dynamics simulations in order to more fully probe the various conformational states and transitions of CaM.

8.3 The Ca^{2+}-dependent Molecular Dynamics of CaM

8.3.1 Aims & Objectives

Chapter 5 of this thesis presents a set of 10ns MD trajectories performed on the separate lobes of CaM to investigate the conformational dynamics of the system. In accordance with the observations from the previous chapter, simulations were performed on fully solvated systems using both the PARAM19 and PARAM27 potentials. An additional simulation was also performed on a CaM-target complex to investigate the effects on the system. A protocol was also developed to investigate the effect of removing the Ca^{2+}-ions form the system and understand the structural implications.

8.3.2 Overview of the Results & Conclusions

Initial results from the simulations were highly encouraging and suggested that both lobes of the protein had the ability to inter-convert between open and closed conformations on a ns timescale as suggested in the PARAM19 trajectories **ntsolv** and **ctsolv**. These observations were directly in-line with suggestions of previous work by Garcia *et al.* and Evenäs *et al.* which had demonstrated similar behaviour in the N and C-term lobes of the protein respectively [Evenas *et al.* 2001; Vigil *et al.* 2001].

As more simulations on the system were performed however the interpretation of the results became less clear. Deletion of the Ca^{2+}-ions from the system was expected to cause the protein to adopt the closed conformation seen in all experimental Ca^{2+}-free structures and the use of the PARAM27 potential was expected to demonstrate similar trends to those of the PARAM19 experiments. The situation was then further complicated by the observation that performing the same experiment on two different machines produced entirely different trajectories highlighting the extreme sensitivity of the simulations to the initial conditions.

The comparison of the results from all these simulations clearly demonstrates that the conformational flexibility of CaM is a highly complex issue and that the conformational space accessible to the protein is very extensive. This flexibility is at the core of the function of the protein and in particular its ability to accommodate a wide variety of target peptide sequences with relatively high affinities.

Some of the more specific observations of these experiments are none the less very interesting and worthy of further investigation. The presence of a target peptide interacting with both lobes of the protein had appeared to accentuate the differences between the two lobes of the protein. In the MD simulation of this complex the flexibility of the C-term and N-term lobe was considerably more limited than in the simulation of the isolated lobes. The N-term lobe of the protein did however demonstrate considerably greater conformational flexibility than the C-term.

It is also important to understand the sequence of events regarding the conformational changes taking place in the **ntsolv** and **ctsolv** trajectories. Despite the complete loss of Ca^{2+}-coordination on the **ntsolv** trajectory, this was shown to be a *consequence* of the conformational change taking place in the protein. If the Ca^{2+} coordination geometry had altered prior to the conformational change then this could have been an indication that the potential was not behaving correctly and was unable to consider the electrostatic effects of the high charge density in the Ca^{2+}-binding sites.

8.3.3 Potential Extensions of the Investigation

These studies of the conformational flexibility of the CaM lobes have produced some very interesting results; however it is clear that the situation requires much more investigation for a better understanding of the system. Inconsistencies in the findings of the experiments are entirely in accordance with suggestions by Caves *et al.* that multiple trajectories are required to fully characterise the conformational space accessible to a protein using MD [Caves *et al.* 1998]. With the observed conformational changes taking place after several ns in the case of the PARAM19 simulations, it is highly unlikely that ten 1ns simulations would yield the same sampling as a single 10ns trajectory.

Consequently it seems that 10ns represents a minimum timeframe for the simulation of this system and that multiple long trajectories are the key to exploring this system.

Given the sensitivity to initial conditions within a particular physical representation, it is difficult to assess the effect of the different potentials to the observed results. Significantly more conformational sampling would help to address this issue. These extensions to the investigation will become readily achievable as computational resources become ever more powerful and available.

8.4 Entropic Analysis of the CaM lobes

8.4.1 Aims & Objectives

Chapter 6 of this thesis investigated the relative flexibilities of the CaM lobes directly through the application of NMA to a selection of conformations generated from the MD simulations of the previous chapter. The classical harmonic vibrational entropies of ca. 100 different conformations extracted from the PARAM19 MD trajectories were calculated to provide a quantitative measure for the flexibility of the lobes. The aim of this study was to investigate the different functional roles of the N- and C-term lobes of CaM and thus estimate the configurational entropy of the two lobes in isolation to see if there were significant differences. Through comparison of the data with that obtained from simulations of the Ca^{2+}-free structures and a CaM-target complex, the effects of these co-factors on the flexibility of the system could also be investigated.

8.4.2 Overview of the Results & Conclusions

Each conformation extracted from the MD trajectory was treated independently and energy minimised to a relatively strict gradient tolerance of 1×10^{-8} kcal mol^{-1} $Å^{-1}$. Comparison of the configurational entropy values calculated from the NMA demonstrated that the mean values for the N-term lobe were significantly higher than those for the C-term lobe in the presence and absence of Ca^{2+} and in the presence and

absence of a target peptide. Indeed the largest differences between the two lobes were seen between the calculations of the N and C-term conformations extracted from the MD simulation of the CaM-target complex. Conformations extracted from the MD trajectories **ntsolv** and **ctsolv** were the most similar to one another with considerable overlap between the distributions and a difference of only 0.005 kcal mol^{-1} ($T\Delta S$ at 300K) between the mean values.

These differences are again suggestive of different functional roles of the N and C-term lobes of the protein. The C-term lobe of the protein has the more rigid conformation, it also has the higher affinity for Ca^{2+} and target peptides. The more flexible character of the N-term lobe of the protein however allows for a greater range in specificity in its interaction with target peptides and perhaps also with Ca^{2+}.

8.4.3 Potential Extensions of the Investigation

The data presented here uses only conformations from the PARAM19 simulations as there is no MD trajectory for an all-atom PARAM27 simulation of the CaM-target complex at this time. Completion of this simulation would enable the same comparison to be made from a different set of conformations in order to more fully evaluate the significance of the results presented here. Other techniques (such as quasiharmonic analysis) have also been developed for the calculation of the entropic properties of a system from simulation data. Most of these techniques however take the trajectory data as is, thus it is necessary to use a fully representative trajectory. As observed in the data from the previous chapter, the MD trajectories generated for CaM are highly dependent on the initial conditions of conformation and momentum and consequently the entropic data calculated from these trajectories would demonstrate the same sensitivity. The feature of the approach adopted here is that each conformation is treated independently and each individual entropy estimate is complete, within the approximations of the method. The multiple entropy estimates computed allow for a statistical comparison of results. It is however acknowledged that the distributions of the data will also be affected by the conformational sampling of the MD trajectories.

8.5 Calmodulin-Target Interactions

8.5.1 Aims & Objectives

At the core of this thesis is an attempt to gain further understanding of the way in which CaM is able to interact with a highly diverse range of molecular targets with a relatively high affinity. In order to understand this distinctive characteristic, the ability of the two CaM lobes to alter their conformation has been studied in some detail in the previous chapters of this thesis. In Chapter 7 however the focus changes from the overall conformations of the lobes, to focus on the binding surfaces which interact with CaM targets and inhibitors. The structures of CaM in complex with the CaM binding domains of different target enzymes and a selection of small molecule inhibitors are used to investigate the plasticity of the CaM binding surfaces.

8.5.2 Overview of the Results & Conclusions

Analysis of the available structures clearly demonstrates the way in which the flexible tether between the two CaM lobes allows them to accommodate a range of different target fragments, mostly α-helical in nature. Two of the most recently determined structures have aptly demonstrated this flexibility showing CaM interacting with a twin helix target through both lobes of the protein as in **1g4y** and through the C-term lobe of the protein alone in **1k90**. Another interesting feature of the structure **1g4y** is that the C-term of the CaM adopts a conformation which resembles the semi-open conformation rather than the fully open conformation seen in other CaM-target complexes.

Analysis of the CaM binding surfaces presented in the different target peptide and inhibitor structures shows suggested that the major properties of a hydrophobic binding pocket, methionine puddles and charged perimeters of the binding surfaces were conserved between both lobes of the protein and in most structures. **1g4y** showed a reorganisation of the hydrophobic binding pocket in the C-term lobe which was otherwise conserved throughout the experimentally determined structures. Interactions

with this pocket were also shown to be the most conserved feature of target interaction throughout the structural ensemble with other features showing far more variability.

A relatively novel approach presented in this investigation was the analysis of the sidechain torsion angles of the amino acid side chains across the CaM structural ensemble. This was used to highlight the residues of the protein which were most variable in conformation, in the context of their role in the protein as target interacting residues or their degree of solvent exposure. From this analysis it was possible to identify flexibility in key residues which are probably involved in conformational integrity (can you elaborate on this?) such as Glu45 and other residues which are involved in target interaction such as Glu114.

Analysis of the amino acid side chains of the target peptides and also the small molecules interacting with the CaM binding surfaces did not indicate any key interactions which were conserved across the ensemble of complexes apart from the presence of a hydrophobic group in the large pocket on the C-term lobe of CaM. The plasticity of the CaM binding surface provides the potential for a plethora of different interactions which means that there are none of the key, specific and conserved interactions associated with some other protein-ligand systems.

Attempts to determine a consensus target motif from the available structures were largely unsuccessful in this analysis. The only common feature of these structures remains the presence of a hydrophobic side chain interacting with the C-term lobe of the protein. There is however a relatively invariant presence of basic residues interacting with the perimeters of both lobes of the protein. It was thought that consideration of residues interacting with the N and C-term lobes of the protein separately may help in the identification of separate motifs for the two lobes of the protein, however attempts at this were also inconclusive.

These studies have clearly demonstrated the complexity of CaM-target interactions. Calmodulin has been shown to demonstrate conformational flexibility on several different levels which enables the protein to interact with diverse targets with relatively high affinities through a multi-faceted approach.

8.5.3 Potential Extensions of the Investigation

An alternative approach to the description of the CaM binding surfaces has the potential to be very useful in these comparative studies. The program *ROADMAP* [Chapman 1993] has been used previously to approach this problem with some success [Afshar *et al.* 1994] however this software was unavailable for these investigations. The benefit of this approach is the reduction in dimensionality of the problem with the simplification of a highly flexible 3-D surface into two dimensions which can be used to more clearly annotate the role of individual side chains. When compared with the 2-D representations of the target peptides, the discussion regarding the complementarity of the protein and target *may* have been more easily achieved. Alternative methods for a reduction in the dimensionality and thus complexity of the problem would also be beneficial to these studies.

It would be of interest to perform *in silico* site-directed mutagenesis study for this system in which the effect of mutations in target peptides could be investigated. The individual residues of an appropriate high affinity target peptide sequence would individually be altered within the framework of an appropriate force field in order to asses the energy penalty for the mutation. The combinatorics of this approach would require careful handling. Such an approach would however provide a key insight into the interactions which are important in recognition of targets for this system

8.6 Final Conclusions

These simulations have shown that CaM is a system which intrinsically has a great capacity for conformational changes, both globally through the flexible linker between the 2 lobes, but also locally (and the focus of this thesis) through flexibility of the individual lobes.

Multivariate analysis was used to allow for discrimination and classification of CaM conformations from the ensemble of observed structures. It was revealed that there are

characteristic conformational changes which can be associated with different states of the protein, eg. Ca^{2+} load, target association etc. This approach can be extended as the database of CaM structures increases. It also serves as an important frame of reference for simulation studies of CaM conformational dynamics.

The principal observed variations in CaM conformation under a variety of conditions can be well described by relatively few degrees of freedom, which correspond to intrinsically low energy deformation modes representing global collective atomic displacements.

The problems associated with sensitivity to initial conditions in biomolecular MD simulations were highlighted in this study. Considerable differences in trajectories were found, even within the same physical representation. This effect may well be amplified by the intrinsic characteristics of the CaM architecture, which serve its role of an environmentally sensitive, adaptive target recognition system.

Classical harmonic configurational entropy estimates derived from NMA on biomolecules can be more robustly estimated by considering a sample of multiple conformations derived from MD simulation. The results from calculations for the CaM lobes suggest that the N-lobe is significantly more flexible than the C-lobe, which correlates with observed functional differences of the lobes. These differences in configurational entropy have been shown to be conserved in the presence and absence of Ca^{2+} and target peptide fragments, however the differences are particularly apparent in the presence of both Ca^{2+} and a peptide.

An extensive and in depth analysis of the CaM binding surfaces in the available database of CaM structures has demonstrated that there is significant plasticity observed across the ensemble of structures. Analysis of the sidechain conformations has revealed important roles for some of the Met residues but also for residues surrounding the binding surfaces of CaM.

A more comprehensive simulation study of conformational variability in CaM requires considerably more extensive conformational sampling, such as might be obtained from multiple very long molecular dynamics simulations. These studies will be facilitated by the ever increasing power of computers.

References:

Ababou, A., R. A. Shenvi*, et al.* (2001). *"Long-range effects on calcium binding and conformational change in the n-domain of calmodulin."* Biochemistry **40**(42): 12719-26.

Accelrys (2001). *QUANTA 2001*. San Diego, Accelrys Inc.

Afshar, M., L. S. Caves*, et al.* (1994). *"Investigating the high affinity and low sequence specificity of calmodulin binding to its targets."* J Mol Biol **244**(5): 554-71.

Afshar, M., L. S. D. Caves*, et al.* (1997). *A unified structural model of target recognition in calmodulin and myosin light chains*. J Muscle Res Cell Motil, (unpublished).

Akke, M. and W. J. Chazin (2001). *"An open and shut case."* Nat Struct Biol **8**(11): 910-2.

Allouche, D., J. Parello*, et al.* (1999). *"Ca2+/Mg2+ exchange in parvalbumin and other EF-hand proteins. A theoretical study."* J Mol Biol **285**(2): 857-73.

Amadei, A., A. B. M. Linssen*, et al.* (1993). *"Essential Dynamics of Proteins."* Proteins **17**(4): 412.

Babu, Y. S., C. E. Bugg*, et al.* (1988). *"Structure of Calmodulin refined at 2.2 A resolution."* J. Mol. Biol. **204**: 191-204.

Babu, Y. S., J. S. Sack*, et al.* (1985). *"Three-dimensional structure of calmodulin."* Nature **315**(6014): 37-40.

Ban, C., B. Ramakrishnan*, et al.* (1994). *"Structure of the recombinant paramecium tetraurelia Calmodulin at 1.68A resolution."* Acta Cryst D **50**: 50-63.

Barbato, G., M. Ikura*, et al.* (1992). *"Backbone dynamics of calmodulin studied by 15N relaxation using inverse detected two-dimensional NMR spectroscopy: the central helix is flexible."* Biochemistry **31**(23): 5269-78.

Barth, A., S. R. Martin*, et al.* (1998). *"Specificity and symmetry in the interaction of calmodulin domains with the skeletal muscle myosin light chain kinase target sequence."* J Biol Chem **273**(4): 2174-83.

Bayley, P. M., W. A. Findlay*, et al.* (1996). *"Target Recognition by Calmodulin: Dissecting the kinetics and affinity of interaction using short peptide synthesis."* Protein Science **5**: 1215 - 1228.

Berendsen, H., J. Postma*, et al.* (1984). *"Molecular dynamics with coupling to an external bath."* J. Chem. Phys. **81**(8): 3684-3690.

Berman, H. M., J. Westbrook*, et al.* (2000). *"The Protein Data Bank."* Nucleic Acids Res **28**(1): 235-42.

Biswas, D. K., P. V. Reddy*, et al.* (1998). *"Calmodulin is essential for estrogen receptor interaction with its motif and activation of responsive promoter."* J Biol Chem **273**(50): 33817-24.

Brändén, C.-I. and J. Tooze (1999). *Introduction to protein structure*. New York, Garland Pub.

Brooks, B. R., R. B. Bruccoleri*, et al.* (1983). *"CHARMM: A program for macromolecular energy minimisation and dynamics calculations."* J. Comp. Chem. **4**: 187-217.

Brooks, B. R., D. Janezic*, et al.* (1995). *"Harmonic Analysis of Large systems. I Methodology."* J. Comp. Chem. **16**(12): 1522-1542.

Brooks, C. L., M. Karplus*, et al.* (1988). *Proteins : a theoretical perspective of dynamics, structure, and thermodynamics*. New York, J. Wiley.

Caves, L. S., J. D. Evanseck*, et al.* (1998). *"Locally accessible conformations of proteins: multiple molecular dynamics simulations of crambin."* Protein Sci **7**(3): 649-66.

Chapman, M. S. (1993). *"Mapping the surface properties of macromolecules."* Protein Sci **2**(3): 459-69.
Chatfield, C. and A. J. Collins (1980). *Introduction to multivariate analysis*. London ; New York, Chapman and Hall :.
Chattophadhyaya, R., W. E. Meador*, et al.* (1992). *"Calmodulin structure refined at 1.7A resolution."* J. Mol. Biol. **228**: 1177-1192.
Cheney, R. E. and M. S. Mooseker (1992). *"Unconventional myosins."* Curr Opin Cell Biol **4**(1): 27-35.
Chou, J. J., S. Li*, et al.* (2001). *"Solution structure of Ca(2+)-calmodulin reveals flexible hand-like properties of its domains."* Nat Struct Biol **8**(11): 990-7.
Chou, J. J., L. Shipeng*, et al.* (2000). *"Study of conformational rearrangement and refinement of structural homology models by the use of heteronuclear dipolar couplings."* J. Biolmol. NMR **18**: 217-227.
Community, C. D. (2001). *CHARMM27 Documentation Index*.
Community, P. D. (1996). *PDB Format Description version 2.2*, RCSB. **2001**.
Cook, W. J., L. J. Walter*, et al.* (1994). *"Drug binding by Calmodulin: Crystal structure of a Calmodulin-TFP Complex."* Biochemistry **33**: 15259-15265.
Cornell, W. D., P. Cieplak*, et al.* (1995). *"A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules."* Journal- American Chemical Society **117**(19): 5179.
Creighton, T. E. (1993). *Proteins : structures and molecular properties*. New York, W.H. Freeman.
Crivici, A. and M. Ikura (1995). *"Molecular and structural basis of target recognition by calmodulin."* Annu Rev Biophys Biomol Struct **24**: 85-116.
Destefanis, G., M. T. Barge*, et al.* (2000). *"The use of principal component analysis (PCA) to characterize beef."* Meat Science **56**(3): 255-259.
Drabikowski, W., H. Brzeska*, et al.* (1982). *"Tryptic fragments of calmodulin. Ca2+- and Mg2+-induced conformational changes."* J Biol Chem **257**(19): 11584-90.
Drum, C. L., S. Z. Yan*, et al.* (2002). *"Structural basis for the activation of anthrax adenylyl cyclase exotoxin by calmodulin."* Nature **415**(6870): 396-402.
Duan, Y. and P. A. Kollman (1998). *"Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution."* Science **282**(5389): 740-4.
Dunbrack, R. L., Jr. and M. Karplus (1994). *"Conformational analysis of the backbone-dependent rotamer preferences of protein sidechains."* Nat Struct Biol **1**(5): 334-40.
Dunker, K. A., C. J. Brown*, et al.* (2002). *"Intrinsic disorder and protein function."* Biochemistry **(Review)**.
Elber, R. and M. Karplus (1987). *"Multiple conformational states of proteins: a molecular dynamics analysis of myoglobin."* Science **235**(4786): 318-21.
Elshorst, B., M. Hennig*, et al.* (1999). *"NMR solution structure of a complex of calmodulin with a binding peptide of the Ca2+ pump."* Biochemistry **38**(38): 12320-32.
Esnouf, R. M. (1999). *"Further additions to MolScript version 1.4, including reading and contouring of electron-density maps."* Acta Crystallogr D Biol Crystallogr **55**(Pt 4): 938-40.
Evenas, J., A. Malmendal*, et al.* (2001). *"Dynamics of the Transition between Open and Closed Conformations in a Calmodulin C-Terminal Domain Mutant."* Structure **9**: 185-195.
Evenas, J., E. Thulin*, et al.* (1997). *"NMR Studies of the E140 mutant of the carboxy terminal domain of Calmodulin reveal global conformational exchange in the Ca saturated state."* Biochemistry **36**: 3448-3457.

Ewig, C. S., T. S. Thacher*, et al.* (1999). *"Derivation of Class II Force Fields. 7. Nonbonded Force Field Parameters for Organic Compounds."* Journal of Physical Chemistry B **103**(33): 6998-7014.

Finn, B. E., J. Evenas*, et al.* (1995). *"Calcium induced structural changes in domain autonomy in calmodulin."* Nat. Struc. Biol. **2**(9): 777-783.

Fischer, S., J. C. Smith*, et al.* (2001). *"Dissecting the vibrational entropy change on protein/ligand binding: Burial of a water molecule in Bovine Pancreatic Trypsin Inhibitor."* J. Phys. Chem. B.

Fitch, W. M. (1966). *"An improved method of testing for evolutionary homology."* J Mol Biol **16**(1): 9-16.

Gao, Z. H. and G. Zhong (1999). *"Calcineurin B- and calmodulin-binding preferences identified with phage-displayed peptide libraries."* Gene **228**(1-2): 51-9.

Gentleman, R. and R. Ihaka (2000). *R statistical language*. Aukland.

Gerstein, M. and R. B. Altman (1995). *"Average Core Structures and Variability Measures for Protein Families: Application to the Immunoglobulins."* J. Mol. Biol. **251**: 161-175.

Gerstein, M. and W. Krebs (1998). *"A database of macromolecular motions."* Nucleic Acids Res **26**(18): 4280-90.

Grubmueller, H. (1996). *Solvate 1.0*, Grubmueller, H.

Gulati, J., A. Babu*, et al.* (1993). *"Identification of the regions conferring calmodulin-like properties to troponin C."* J Biol Chem **268**(16): 11685-90.

Hadley, C. and D. T. Jones (1999). *"A systematic comparison of protein structure classifications: SCOP, CATH and FSSP."* Structure **7**: 1099-1112.

Halgren, T. A. (1996). *"Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94."* Journal of Computational Chemistry **17**(5/6): 490-519.

Harmat, V., Z. Bocskei*, et al.* (2000). *"A new potent antagonist with arylalkylamine structure: Crystallographic and Functional studies."* J. Mol. Biol. **297**: 747-755.

Hartshorn, M. (1996). *Molviewer*. York.

Hayward, S., A. Kitao*, et al.* (1995). *"Harmonicity and anharmonicity in protein dynamics: a normal mode analysis and principal component analysis."* Proteins **23**(2): 177-86.

Holm, L. and C. Sander (1996). *"Mapping the Protein Universe."* Science **273**: 595-602.

Holm, L. and C. Sander (1998). *"Dictionary of recurrent domains in protein structures."* Proteins **33**(1): 88-96.

Houdosse, A., M. Silver*, et al.* (1996). *"A model of Ca free calmodulin binding to unconventional myosins reveals how calmodulin acts as a regulatory switch."* Structure **4**(12): 1475-1490.

Howe, P. W. (2001). *"Principal components analysis of protein structure ensembles calculated using NMR data."* J Biomol NMR **20**(1): 61-70.

Ikura, M., G. M. Clore*, et al.* (1992). *"Solution structure of a Calmodulin-target peptide complex by multidimensional NMR."* Science **256**: 632-638.

Johnson, P. (1990). *"Calpains (intracellular calcium-activated cysteine proteinases): structure-activity relationships and involvement in normal and abnormal cellular metabolism."* Int J Biochem **22**(8): 811-22.

Jorgensen, W. L., J. Chandrasekhar*, et al.* (1983). *"Comparison of simple potential functions for simulating liquid water."* J. Chem. Phys. **79**: 926 -935.

Kabsch, W. (1976). *"A solution for the best rotation to relate two sets of vectors."* Acta. Cryst. A. **32**: 922-923.

Kahn, J., B. Walcheck*, et al.* (1998). *"Calmodulin regulates L-selectin adhesion molecule expression and function through a protease-dependent mechanism."* Cell **92**(6): 809-18.

Kaminski, G. and W. L. Jorgensen (1996). *"Performance of the AMBER94, MMFF94, and OPLS-AA Force Fields for Modeling Organic Liquids."* Journal of Physical Chemistry **100**(46): 18010-18013.

Karplus, M. and T. Ichiye (1996). *"Comment on a "Fluctuation and cross correlation analysis of protein motors observed in nanosecond molecular dynamics simulations.""* J. Mol. Biol. **263**: 120-122.

Kataoka, M., A. Persechini*, et al.* (1996). *"The linker of Calmodulin lacking Glu 84 is elongated in solution although it is bent in the crystal."* Proteins **25**: 335-341.

Kawasaki, H. and R. H. Kretsinger (1994). *"Calcium-binding proteins. 1: EF-hands."* Protein Profile **1**(4): 343-517.

Kelley, L. A., S. P. Gardner*, et al.* (1996). *"An Automated Approach For Clustering An Ensemble of NMR Derived Protein Structures Into Conformationally-Related Subfamilies."* Prot. Eng. **9**: 1063-1065.

Kelley, L. A., S. P. Gardner*, et al.* (1997). *"An Automated Approach For Defining Core Atoms And Domains in an Ensemble of NMR-Derived Protein Structures."* Prot. Eng. **10**: 737-741.

Kitao, A. and N. Go (1999). *"Investigating protein dynamics in collective coordinate space."* Curr Opin Struct Biol **9**(2): 164-9.

Kretsinger, R. H. and C. E. Nockolds (1973). *"Carp muscle calcium-binding protein. II. Structure determination and general description."* J Biol Chem **248**(9): 3313-26.

Kuboniwa, H., N. Tjandra*, et al.* (1995). *"Solution Structure of Ca Free Calmodulin."* NSB **2**(9): 768-776.

Kurokawa, H., M. Osawa*, et al.* (2001). *"Target induced Conformational Adaptation of Calmodulin Revealed by the Crystal Structure of a Complex with Nematode Ca2+/Calmodulin-dependent Kinase Kinase Peptide."* J. Mol. Biol. **312**(1): 59-68.

Kwon, H., E. B. Goodwin*, et al.* (1990). *"Isolation of the regulatory domain of scallop myosin: role of the essential light chain in calcium binding."* Proc Natl Acad Sci U S A **87**(12): 4771-5.

Laskowski, R. A., M. W. MacArthur*, et al.* (1993). *"PROCHECK: a program to check the stereochemical quality of protein structures."* Journal of Applied Crystallography **26**(2): 283.

Lazaridis, T. and M. Karplus (1999). *"Effective energy function for proteins in solution."* Proteins **35**(2): 133-52.

Leach, A. R. (1996). *Molecular modelling : principles and applications*. Harlow, England, Longman.

Lee, B. and F. M. Richards (1971). *"The interpretation of protein structures: estimation of static accessibility."* J Mol Biol **55**(3): 379-400.

Lee, M. R. and P. A. Kollman (2001). *"Free-Energy Calculations Highlight Differences in Accuracy between X-Ray and NMR Structures and Add Value to Protein Structure Prediction."* Structure (Camb) **9**(10): 905-16.

Levitt, M. (2001). *"The birth of computational structural biology."* Nat Struct Biol **8**(5): 392-3.

Levitt, M. and C. Chothia (1976). *"Structural patterns in globular proteins."* Nature **261**(5561): 552-8.

Levitt, M. and S. Lifson (1969). *"Refinement of protein conformations using a macromolecular energy minimization procedure."* J Mol Biol **46**(2): 269-79.

Lifson, S. (1981). *Potential Energy Functions for Structural Molecular Biology*. Maratea, S. Italy, NATO Advanced Study Institute / FEBS Advanced Course No. 78.

London, F. (1930). Z. Physik. Chem. B **11**: 222.

Lymn, R. W. and E. W. Taylor (1971). *"Mechanism of adenosine triphosphate hydrolysis by actomyosin."* Biochemistry **10**(25): 4617-24.

MacKerell, A. D., D. Bashford*, et al.* (1998). *"All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins."* Journal of Physical Chemistry B **102**(18): 3586-3616.

Malmendal, A., J. Evenas*, et al.* (1999). *"Structural Dynamics in the C-terminal domain of Calmodulin at low Ca levels."* J. Mol. Biol. **293**: 883-899.

Malmendal, A., J. Evenas*, et al.* (1998). *"When size is important. Accommodation of magnesium in a calcium binding regulatory domain."* J Biol Chem **273**(44): 28994-9001.

Marchand, S. and B. Roux (1998). *"Molecular Dynamics study of Calbindin D9k in the apo and singly and doubly Ca loaded states."* Proteins **33**: 265-284.

Martin, S. R., P. M. Bayley*, et al.* (1996). *"Spectroscopic characterization of a high-affinity calmodulin-target peptide hybrid molecule."* Biochemistry **35**(11): 3508-17.

Martin, S. R., L. Masino*, et al.* (2000). *"Enhancement by Mg2+ of domain specificity in Ca2+ dependent interactions of calmodulin with target sequences."* Protein Science **9**: 2477-2488.

McLachlan, A. D. (1972). *"Repeating sequences and gene duplication in proteins."* J Mol Biol **64**(2): 417-37.

McQuarrie, D. A. (1973). *Statistical thermodynamics*. New York,, Harper & Row.

Meador, W. E., A. R. Means*, et al.* (1992). *"Target Enzyme recognition by Calmodulin: 2.4A structure of calmodulin-peptide complex."* Science **257**: 1251-1254.

Medvedeva, M. V., E. A. Kolobova*, et al.* (1996). *"Interaction of proteolytic fragments of Calmodulin with caldesmon and calponin."* Biochemical Journal **315**: 1021-1026.

Mehler, E. L., J. L. Pascual-Ahuir*, et al.* (1991). *"Structural dynamics of calmodulin and troponin C."* Protein Eng **4**(6): 625-37.

Mercier, P., L. Spyracopoulos*, et al.* (2001). *"Structure, dynamics, and thermodynamics of the structural domain of troponin C in complex with the regulatory peptide 1-40 of troponin I."* Biochemistry **40**(34): 10063-77.

Miranker, A. and M. Karplus (1991). *"Functionality maps of binding sites: a multiple copy simultaneous search method."* Proteins **11**(1): 29-34.

Murzin, A. G., S. E. Brenner*, et al.* (1995). *"SCOP: A structural classification of Proteins Database for the investigation of Sequences and Structures."* J. Mol. Biol. **247**: 536-540.

Nelson, M. R. and W. J. Chazin (1998). *"An interaction based analysis of Calcium induced conformational changes in Ca sensor proteins."* Protein Sci **7**(2): 270-82.

Neria, E., S. Fischer*, et al.* (1996). *"Simulation of activation free energies in molecular systems."* J. Chem. Phys. **105**(5): 1902-1921.

Oldfield, T. J. (1992). *"SQUID: a program for the analysis and display of data from crystallography and molecular dynamics."* J Mol Graph **10**(4): 247-52.

Oldfield, T. J. (2001). *CAMINE*.

Olsson, L. L. and L. Sjolin (2001). *"Structure of Escherichia coli fragment TR2C from calmodulin to 1.7 A resolution."* Acta Crystallogr D Biol Crystallogr **57**(Pt 5): 664-9.

O'Neil, K. T. and W. F. DeGrado (1990). *"How calmodulin binds its targets: sequence independent recognition of amphiphilic alpha-helices."* Trends Biochem Sci **15**(2): 59-64.

Orengo, C. A., A. D. Michie*, et al.* (1997). *"CATH- a hierarchic classification of protein domain structures."* Structure **5**: 1093-1108.

Osawa, M., H. Tokumitsu*, et al.* (1999). *"A novel calmodulin target recognition revealed by its NMR Structure in complex with a peptide derived from Ca2+ Calmodulin dependent protein kinase kinase."* Nat. Struc. Biol. **6**: 819.

Persechini, A. and P. M. Stemmer (2002). *"Calmodulin is a limiting factor in the cell."* Trends Cardiovasc Med **12**(1): 32-7.

Persechini, A., P. M. Stemmer*, et al.* (1996). *"Localization of unique functional determinants in the calmodulin lobes to individual EF hands."* J Biol Chem **271**(50): 32217-25.

Qin, Z. and T. C. Squier (2001). *"Calcium-dependent stabilization of the central sequence between Met(76) and Ser(81) in vertebrate calmodulin."* Biophys J **81**(5): 2908-18.

Reijmers, T. H., R. Wehrens*, et al.* (2001). *"Circular effects in representations of an RNA nucleotides data set in relation with principal components analysis."* Chemometrics and Intelligent Laboratory Systems **56**(2): 61-71.

Rhoads, A. R. and F. Friedberg (1997). *"Sequence motifs for calmodulin recognition."* Faseb J **11**(5): 331-40.

Sanejouand, Y.-H. (2001). *Ca2+ handling potential energy function*. Personal Communication.

Schumacher, M. A., A. F. Rivard*, et al.* (2001). *"Structure of the gating domain of a Ca2+-activated K+ channel complexed with Ca2+/calmodulin."* Nature **410**(6832): 1120-4.

Siedlecka, M., G. Goch*, et al.* (1999). *"alpha helix nucleation by a calcium binding peptide loop."* PNAS **96**: 903-908.

Squier, T. C. (2001). *"Oxidative stress and protein aggregation during biological aging."* Exp Gerontol **36**(9): 1539-50.

Stocker, U. and W. F. van Gunsteren (2000). *"Molecular dynamics simulation of hen egg white lysozyme: a test of the GROMOS96 force field against nuclear magnetic resonance data."* Proteins **40**(1): 145-53.

Stote, R. H., States, D.J., Karplus, M, (1991). *"J. Chimie Physique."*

Sussman, J. L., D. Lin*, et al.* (1998). *"Protein Data Bank (PDB): database of three-dimensional structural information of biological macromolecules."* Acta Crystallogr D Biol Crystallogr **54**(1 (Pt 6)): 1078-84.

Tabernero, L., D. A. Taylor*, et al.* (1997). *"The structure implications of a calmodulin mutant with a deletion in the central helix: implications for molecular recognition and protein binding."* Structure **5**(5): 613-622.

Thompson, J. D., T. J. Gibson*, et al.* (1997). *"The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools."* Nucleic Acids Res **25**(24): 4876-82.

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, Mass., Addison-Wesley Pub. Co.

Van Aalten, D. M. F., B. L. De Groot*, et al.* (1997). *"A Comparison of Techniques for Calculating Protein Essential Dynamics."* J. Comp. Chem. **18**(2): 169-181.

Van Der Spoel, D., B. L. De Groot*, et al.* (1996). *"Bending of the Calmodulin Central Helix: A theoretical Study."* Protein Science **5**: 2044-2053.

Van Eldik, L. J. and D. M. Watterson (1998). *Calmodulin and signal transduction*. San Diego, Academic Press.

van Vlijmen, H. W. T. and M. Karplus (1999). *"Analysis of calculated normal modes of a set of native and partially unfolded proteins."* J. Phys. Chem. B **103**(15): 3009-3021.

Various (2001). *Xara X*. Hemel Hempstead, UK, Xara Ltd.

Vigil, D., S. C. Gallagher*, et al.* (2001). *"Functional Dynamics of the Hydrophobic Cleft in the N-Domian of Calmodulin."* Biophys. J. **80**: 2082-2092.

Vogel, H. J. (1994). *"The Merck Frosst Award Lecture 1994. Calmodulin: a versatile calcium mediator protein."* Biochem Cell Biol **72**(9-10): 357-76.

Wall, M. E., J. B. Clarage*, et al.* (1997). *"Motions of calmodulin characterised using both Bragg and Diffuse X-ray scattering."* Structure **5**(12): 1599-1612.

Walter, P., R. Keenan*, et al.* (2000). *"Perspectives: structural biology. SRP--where the RNA and membrane worlds meet."* Science **287**(5456): 1212-3.

Wang, W., O. Donini*, et al.* (2001). *"Biomolecular simulations: recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions."* Annu Rev Biophys Biomol Struct **30**: 211-43.

Weiss, S. (1999). *"Fluorescence spectroscopy of single biomolecules."* Science **283**(5408): 1676-83.

Williams, R. J. (1985). *"16th Sir Hans Krebs lecture. The symbiosis of metal and protein functions."* Eur J Biochem **150**(2): 231-48.

Wilson, M. A. and A. T. Brunger (2000). *"The 1.0 A crystal structure of Ca2+ Calmodulin: An analysis of disorder and implications for functionally relevant plasticity."* J. Mol. Biol. **301**: 1237.

Wriggers, W., E. Mehler*, et al.* (1998). *"Structure and Dynamics of Calmodulin in Solution."* Biophys. J. **74**: 1622-1639.

Yamaotsu, N., M. Suga*, et al.* (2001). *"Molecular Dynamics Simulation of the Calmodulin-Trifluoperazine complex in Aqueous Solution."* Biopolymers **58**: 410-421.

Yang, C., G. S. Jas*, et al.* (2001). *"Structure and dynamics of calcium-activated calmodulin in solution."* J Biomol Struct Dyn **19**(2): 247-71.

Yang, K. (2001). *"Prokaryotic Calmodulins: Recent Developments and Evolutionary Implications."* J. Mol. Microbiol. Biotchnol. **3**(3): 457-459.

Yap, K. L., J. Kim*, et al.* (2001). *"Calmodulin Target Database."* J. Struc. Func. Geno. **1**: 8-14.

Yasri, A., L. Chiche*, et al.* (1996). *"Rational choice of Molecular Dynamics simulation parameters through use of three-dimensional auto-correlation method: application to calmodulin flexibility study."* Prot. Eng. **9**(11): 959-976.

Yin, D., H. Sun*, et al.* (1999). *"Nonessential role for methionines in the productive association between calmodulin and the plasma membrane Ca-ATPase."* Biochemistry **38**(41): 13654-60.

Zhang, M., T. Tanaka*, et al.* (1995). *"Calcium induced conformational transition revealed by the solution structure of apo Calmodulin."* Nat. Struc. Biol. **2**(9): 758-767.

Zhang, M. and T. Yuan (1998). *"Molecular mechanisms of calmodulin's functional versatility."* Biochem Cell Biol **76**(2-3): 313-23.

Appendix A: Abbreviations

CaM	Calmodulin
ED	Essential Dynamics
MARCKS	Myristoylated Alanine-rich C Kinase Substrate
MD	Molecular Dynamics
MRP	MARCKS Related Protein
N	Number of Atoms
NMA	Normal Mode Analysis
NMR	Nuclear Magnetic Resonance
PCA	Principal Components Analysis
PCs	Principal Components
PDB	Protein DataBank hosted by the RCSB at www.rcsb.org
RCSB	Research Collaboration for Structural Biology
RDC	Residual Dipolar Coupling
RGYR	Radius of Gyration
RMSD	Root Mean Square Deviation
SAXS	Small-Angle X-ray Scattering
Sel	Wild type peptide corresponding to the cytoplasmic domain of L-Selectin
TnC	Troponin C
VdW	Van der Waals
W-Sel	Tryptophan labelled derivative peptide from L-Selectin
XRD	X-ray Diffraction

Appendix B: FSSP Results

The following table presents the results of the FSSP classification of CaM family of structures. The domain identifier corresponding to CaM is DC_3_241_1 which includes all CaM structures contained in the database and homologous protein structures such as Troponin C. Other domain identifiers are reported in the results however these were not studied in the structural comparison analysis performed in Chapter 3.

A discussion of FSSP and other structural classification databases is included in the Intrduction.

Table A: *Complete results of the FSSP Classification of the CaM family of proteins.*

PDB ID	Z	RMSD	lali	%ide	Ifir	Jlas	Lseq2	D.C. number		Descriptor
1cll	21.1	0	144	100	1	144	144	DC_3_241_1	DC_3_241_1	Calmodulin (vertebrate)
3cln	19.5	0.3	143	99	2	143	143	DC_3_241_1	DC_3_241_1	Calmodulin
1clm	19.2	0.7	144	89	1	144	144	DC_3_241_1	DC_3_241_1	Calmodulin (paramecium tetraurelia) (wild type)
1exrA	19.1	0.7	144	89	1	146	146	DC_3_241_1	DC_3_241_1	Calmodulin
1osa	19.1	0.6	144	89	1	147	148	DC_3_241_1	DC_3_241_1	Calmodulin
4cln	18.2	0.8	144	98	1	147	148	DC_3_241_1	DC_3_241_1	Calmodulin
1qiwB	14.4	5.3	84	93	2	132	143	DC_3_241_1	DC_3_241_1	Calmodulin
1lin	14.2	7.3	84	83	2	138	146	DC_3_241_1	DC_3_241_1	Calmodulin
1qivA	14.1	6.6	87	85	1	138	144	DC_3_241_1	DC_3_241_1	Calmodulin
1cdlA	14	7.3	81	87	2	136	142	DC_3_241_1	DC_3_241_1	Calmodulin complexed with Calmodulin-binding peptide from smooth muscle myosin light chain kinase
1a29	14	5.5	76	92	2	136	144	DC_3_241_1	DC_3_241_1	Calmodulin biological_unit
1cdlD	13.9	7.5	81	87	2	136	141	DC_3_241_1	DC_3_241_1	Calmodulin complexed with Calmodulin-binding peptide from smooth muscle myosin light chain kinase
1cm1A	13.7	3.8	76	95	2	137	143	DC_3_241_1		Calmodulin Calmodulin-dependent protein kinase ii-alpha fragment
1qiwA	13.7	7.9	88	96	1	139	145	DC_3_241_1	DC_3_241_1	Calmodulin
1vrkA	13.6	8.2	91	91	2	141	148	DC_3_241_1	DC_3_241_1	Calmodulin Mutant biological_unit rs20
1ctr	13.5	0.9	67	100	78	141	141	DC_3_241_1		Calmodulin complexed with trifluoperazine (1:1 complex)
1cdmA	13.4	4.6	72	96	2	126	133	DC_3_241_1		Calmodulin complexed with calmodulin-binding domain of calmodulin-dependent protein kinase ii
1ahr	13.2	5.8	76	92	2	138	146	DC_3_241_1	DC_3_241_1	calmodulin Mutant
1avsB	13.2	0.7	70	45	75	79	82	DC_3_241_1		troponin c fragment
1avsA	13.1	0.6	70	45	75	78	81	DC_3_241_1		troponin c fragment
1ncx	12.6	1.2	67	36	78	160	162	DC_3_241_1	DC_3_241_6	troponin c
1ckkA	12.6	3.2	83	92	1	140	148	DC_3_241_1	DC_3_241_1	calmodulin (cam) biological_unit rat ca2+CALMODULIN DEPENDENT PROTEIN KINASE fragment (camkk)

1cm4A	12.5	7	81	95	1	132	143	DC_3_241_1	DC_3_241_1	calmodulin calmodulin-dependent protein kinase ii-alpha fragment
1ak8	11.2	1.4	70	100	2	75	76	DC_3_241_1		calmodulin fragment (calmodulin cerium tr1c-domain, residues 1 - 75)
1auiB	11.1	1.3	61	41	82	154	165	DC_3_241_1	DC_3_241_1	serineTHREONINE PHOSPHATASE 2B (calcineurin)
1cmg	10.7	1.9	70	46	2	73	73	DC_3_241_1		calmodulin (vertebrate) (calmodulin calcium tr2c-domain)
1wdcB	10.6	10.5	130	28	2	137	142	DC_3_241_1	DC_3_241_2	scallop myosin fragment
2bbmA	10.3	1.6	67	96	78	147	148	DC_3_241_1		Calmodulin (calcium-bound) complexed with rabbit skeletal myosin light chain kinase (calmodulin-binding domain) (NMR, minimized average structure)
1mxlC	10	2	76	65	2	89	89	DC_3_241_1		troponin c fragment (cntnc) troponin i fragment (cntni peptide)
1cffA	9.4	8.7	80	89	37	147	148	DC_3_241_1	DC_3_241_1	calmodulin (cam) calcium pump fragment (c20w)
1a75A	9.3	1.5	67	35	78	106	106	DC_3_241_1		parvalbumin
1dguA	9.1	13	116	22	2	171	183	DC_3_241_1	DC_3_241_1	calcium-saturated cib
1c7vA	8.6	2.9	66	46	79	67	68	DC_3_241_1		calcium vector protein fragment (cavp)
1djxB	8.4	2.1	66	15	78	124	561	DC_3_241_1	DC_3_241_5	phosphoinositide-specific phospholipase c, isozyme delta1 (plc-d1) Mutant biological_unit
4icb	7.6	3.3	65	30	77	72	76	DC_3_241_1		Calbindin d9k (minor a form)
1rec	7	9.5	87	21	8	165	185	DC_3_241_1	DC_3_241_8	Recoverin (calcium sensor in vision)
2scpA	6.7	2.8	76	17	1	163	174	DC_3_241_1		Sarcoplasmic calcium-binding protein
1alvA	6.7	1.9	66	20	77	75	173	DC_3_241_1	DC_3_241_4	calpain fragment (s-camld)
1psrA	6.5	3.5	69	26	1	86	100	DC_3_241_1		psoriasin (s100a7)
2sas	6.1	5.6	85	20	2	155	185	DC_3_241_1	DC_3_241_1	Sarcoplasmic calcium-binding protein (iso type ii)
2cblA	6	2.8	64	11	79	201	305	DC_3_241_3	DC_4_34_1	proto-oncogene cbl fragment zap-70 fragment
1sra	5.5	2.8	63	18	78	140	151	DC_3_241_1	DC_7_251_1	sparc fragment (bm-40, osteonectin)
1wdcC	4.7	3.6	66	44	78	149	152	DC_3_241_2	DC_3_241_6	scallop myosin fragment
1f70A	4.4	4.3	66	75	3	71	76	DC_3_241_1		calmodulin fragment
1cmf	3.9	3.2	61	74	78	67	73	DC_3_241_1		calmodulin (vertebrate) (calmodulin apo tr2c-domain)
1eh2	3.3	3.1	54	19	79	92	95	DC_3_241_7		eps15 fragment (eh2, epidermal growth factor receptor substrate 15)
1dmo	3.3	7.5	56	91	2	110	148	DC_3_241_1		calmodulin

1dkvA	3.2	3.3	77	17	1	688	699	DC_3_241_8	DC_7_349_1	m-calpain fragment (calcium-activated neutral proteinase) Mutant calpain fragment (calcium-activated neutral proteinase)
1b4aA	2.9	3.1	45	16	86	65	146	DC_3_161_5	DC_6_12_2	arginine repressor biological_unit
1utg	2.7	3.3	52	14	24	70	70	DC_6_153_1		Uteroglobin (oxidized)
1thg	2.7	2.8	49	5	81	425	544	DC_3_190_1	DC_6_157_1	Lipase triacylglycerol hydrolase
1ecl	2.6	3.6	59	5	9	497	552	DC_3_161_7	DC_3_173_5	escherichia coli topoisomerase i (escherichia coli omega protein) Mutant engineered
1qsaA	2.5	3.5	52	2	5	542	618	DC_3_229_1	DC_3_251_1	soluble lytic transglycosylase slt70 fragment
1ytfD	2.5	3.8	40	8	2	48	100	DC_6_81_1		yeast tata-box binding protein fragment (ytbp) yeast transcription factor iia (ytfiia, toa1TOA2) cyc1 tata-box DNA
1cii	2.4	4.5	63	7	1	573	602	DC_3_48_3	DC_3_227_1	colicin ia biological_unit
1bkrA	2.4	4	60	10	3	108	108	DC_3_245_1		spectrin beta chain fragment (calponin homology (ch) domain) biological_unit
2fokA	2.4	5.2	42	13	21	370	558	DC_3_161_3	DC_3_161_6	foki restriction endonuclease fragment (r.Foki) biological_unit
1hstA	2.3	3.3	45	12	89	73	74	DC_3_161_1		Histone h5 (globular domain)
1r2aA	2.3	1.8	35	10	1	45	46	DC_3_125_2		camp-dependent protein kinase type ii regulatory subunit fragment (riia(1-44))
1bh9A	2.3	2.8	35	24	97	38	45	DC_3_153_2		tafii18 fragment (transcription initiation factor tfiid 18 kd subunit) tafii28 fragment (transcription initiation factor tfiid 28 kd subunit)

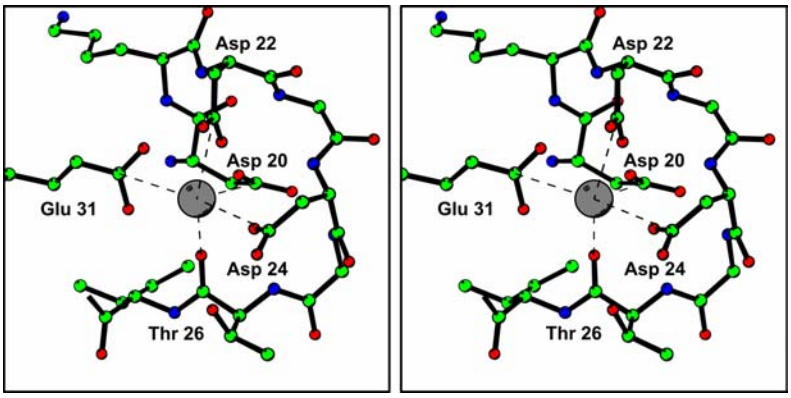
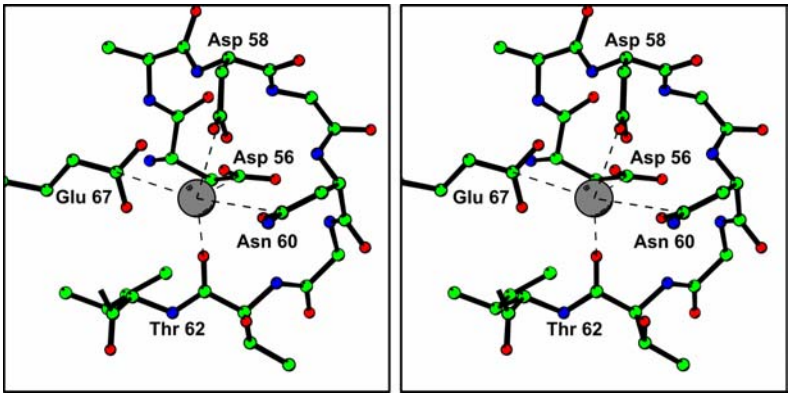
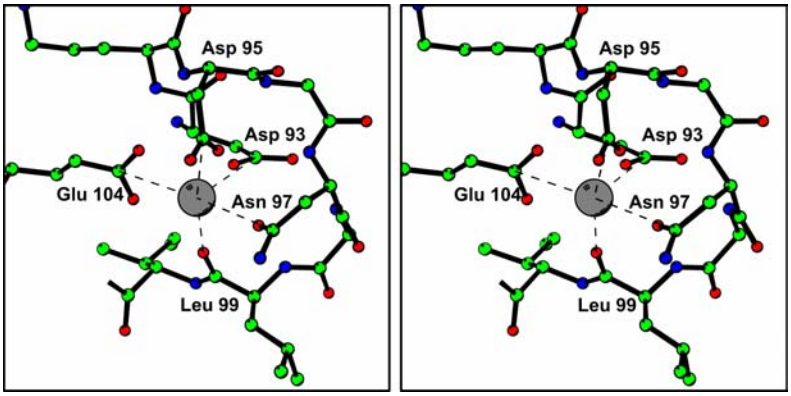
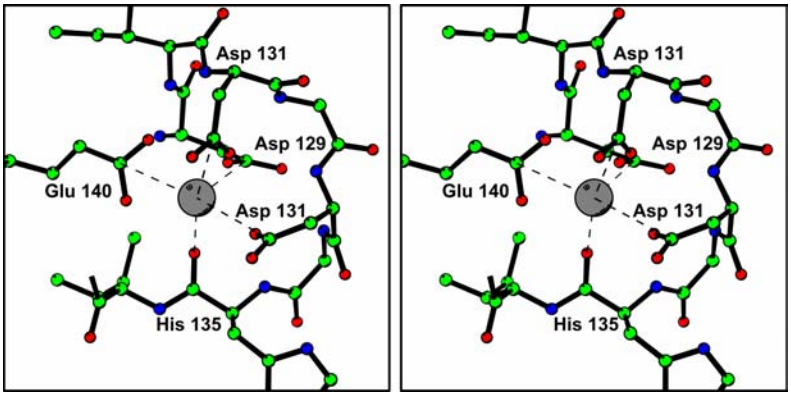
Appendix C: Ca^{2+} Binding Statistics from the 1ns Trajectories

This appendix presents the results of the Calcium coordination analysis performed as part of the Molecular Dynamics force field analysis in Chapter 4.

Table B contains stereo images of the four Ca^{2+}-binding sites of **1exr,** the crystal structure used to produce the different MD simulations. Reference atoms and corresponding Ca^{2+} binding distances are also included for the crystal structure.

Table C and **Table D** show the mean and standard deviation for each of these reference distances during the 1ns molecular dynamics runs for the N and C-term lobes of CaM respectively.

Table B: *Ca^{2+} coordination characteristics of the four Ca^{2+} sites in 1exr*

Ca^{2+} Coordination Geometry	*Distances*
	**EF I** Cγ Asp20-Ca **3.405Å** Cγ Asp22-Ca **3.525Å** Cγ Asp24-Ca **3.337Å** O Thr26 -Ca **2.355Å** Cδ Glu20-Ca **2.820Å**
	**EF II** Cγ Asp56-Ca **3.430Å** Cγ Asp58-Ca **3.421Å** Cγ Asn60-Ca **3.322Å** O Thr62 -Ca **2.407Å** Cδ Glu67 -Ca **2.883Å**
	**EF III** Cγ Asp93-Ca **3.455Å** Cγ Asp95-Ca **3.211Å** Cγ Asn97-Ca **3.293Å** O Leu99 -Ca **2.354Å** Cδ Glu104-Ca **2.871Å**
	**EF IV** Cγ Asp128-Ca **3.479Å** Cγ Asp131-Ca **3.288Å** Cγ Asp133-Ca **3.415Å** O His135 -Ca **2.284Å** Cδ Glu140-Ca **2.839Å**

Simulation	*Difference in Mean distance relative to observed structure 1exr (SD values)*									
	1: Asp 20	*2: Asp 22*	*3: Asp 24*	*4: Thr 26*	*5: Glu 31*	*6: Asp 56*	*7: Asp 58*	*8: Asn 60*	*9: Thr 62*	*10: Glu 67*
Observed	**3.405**	**3.525**	**3.337**	**2.355**	**2.820**	**3.430**	**3.421**	**3.322**	**2.407**	**2.883**
ntvac	-0.420 (0.088)	-0.222 (0.131)	0.252 (0.090)	0.059 (0.081)	0.207 (0.089)	-0.392 (0.085)	-0.381 (0.114)	0.125 (0.095)	0.048 (0.088)	0.128 (0.094)
ntvac1	-0.367 (0.164)	-0.030 (0.121)	-0.328 (0.133)	0.082 (0.083)	0.739 (0.172)	-0.432 (0.077)	-0.442 (0.081)	0.107 (0.091)	0.021 (0.078)	0.136 (0.094)
nteef	0.440 (0.154)	0.139 (0.129)	0.559 (0.160)	0.054 (0.075)	1.058 (0.283)	0.356 (0.162)	0.343 (0.184)	0.241 (0.108)	0.048 (0.091)	1.546 (0.639)
nteef1	1.840 (0.851)	2.401 (0.756)	3.562 (0.736)	0.123 (0.110)	2.736 (1.202)	2.135 (0.392)	0.424 (0.117)	0.238 (0.101)	0.078 (0.108)	0.954 (0.197)
ntxtal	-0.412 (0.068)	-0.554 (0.072)	-0.290 (0.172)	2.011 (0.758)	0.186 (0.074)	-0.249 (0.215)	-0.332 (0.256)	0.177 (0.132)	0.093 (0.099)	0.163 (0.115)
ntxtal1	-0.415 (0.081)	-0.546 (0.079)	0.084 (0.187)	0.163 (0.116)	0.206 (0.076)	-0.405 (0.084)	-0.352 (0.143)	0.292 (0.130)	0.077 (0.085)	0.112 (0.067)
ntxtalsolv	-0.372 (0.159)	-0.497 (0.122)	0.196 (0.187)	0.141 (0.103)	0.205 (0.085)	-0.155 (0.281)	-0.380 (0.124)	0.286 (0.118)	0.071 (0.092)	0.119 (0.073)
ntxtalsolv1	0.026 (0.258)	-0.432 (0.118)	-0.051 (0.257)	0.260 (0.400)	0.248 (0.110)	-0.299 (0.257)	-0.209 (0.234)	0.327 (0.138)	0.297 (0.639)	0.279 (0.291)
ntsolv	0.163 (0.222)	-0.281 (0.258)	-0.100 (0.276)	0.186 (0.147)	0.219 (0.085)	-0.237 (0.218)	-0.353 (0.116)	0.245 (0.122)	0.070 (0.085)	0.154 (0.087)
ntsolv1	0.255 (0.132)	-0.265 (0.249)	-0.102 (0.281)	0.231 (0.172)	0.232 (0.085)	0.166 (0.170)	-0.326 (0.220)	0.264 (0.124)	0.080 (0.096)	0.131 (0.085)
ntpar27	-0.151 (0.100)	-0.311 (0.090)	-0.012 (0.085)	-0.059 (0.089)	-0.172 (0.078)	-0.139 (0.091)	-0.433 (0.288)	0.061 (0.118)	-0.055 (0.105)	-0.260 (0.064)
ntpar271	-0.116 (0.097)	-0.364 (0.151)	-0.062 (0.091)	-0.035 (0.104)	-0.188 (0.073)	-0.101 (0.076)	-0.230 (0.234)	1.234 (0.997)	-0.037 (0.122)	-0.277 (0.059)

Table C: *Table of mean distances between reference atoms on Ca^{2+} coordinating ligands and the Ca^{2+} ions for the simulations of the N-term. Values in brackets indicate the standard deviation for the distribution over the simulation. Reference values come from the crystal structure used for the simulations, prior to minimisation and dynamics.*

Simulation	*Difference in Mean distance relative to observed structure 1exr (SD values)*									
	1: Asp 93	*2: Asp 95*	*3: Asn 97*	*4: Leu 99*	*5: Glu 104*	*6: Asp 128*	*7: Asp 131*	*8: Asn 133*	*9: His 135*	*10: Glu 140*
Observed	**3.455**	**3.211**	**3.293**	**2.354**	**2.871**	**3.479**	**3.288**	**3.415**	**2.284**	**2.839**
ctvac	-0.494 (0.067)	-0.269 (0.075)	0.215 (0.099)	0.060 (0.076)	0.089 (0.064)	-0.253 (0.253)	-0.130 (0.177)	-0.422 (0.138)	0.168 (0.086)	0.107 (0.077)
ctvac1	-0.465 (0.089)	-0.302 (0.069)	0.216 (0.104)	0.063 (0.076)	0.087 (0.062)	-0.508 (0.097)	-0.279 (0.116)	-0.459 (0.066)	0.171 (0.085)	0.096 (0.060)
cteef	0.360 (0.145)	1.697 (0.795)	0.289 (0.119)	0.096 (0.105)	1.085 (0.338)	0.321 (0.124)	0.524 (0.262)	0.413 (0.110)	0.109 (0.070)	0.913 (0.183)
cteef1	0.368 (0.141)	2.430 (1.169)	0.268 (0.100)	0.079 (0.088)	1.010 (0.292)	0.330 (0.124)	0.417 (0.129)	0.382 (0.096)	0.106 (0.071)	0.862 (0.156)
ctxtal	-0.442 (0.071)	-0.121 (0.147)	0.383 (0.161)	0.215 (0.124)	0.220 (0.089)	-0.416 (0.126)	-0.267 (0.095)	0.123 (0.152)	0.225 (0.117)	0.306 (0.201)
ctxtal1	-0.370 (0.117)	-0.124 (0.106)	0.330 (0.095)	2.188 (0.226)	0.167 (0.080)	-0.431 (0.098)	-0.228 (0.188)	0.137 (0.141)	0.220 (0.150)	0.200 (0.108)
ctxtalsolv	0.125 (0.187)	0.070 (0.257)	0.260 (0.122)	0.277 (0.304)	0.201 (0.098)	-0.311 (0.275)	-0.007 (0.280)	0.276 (0.147)	0.301 (0.308)	0.221 (0.088)
ctxtalsolv1	0.061 (0.171)	-0.046 (0.228)	0.222 (0.116)	0.157 (0.117)	0.194 (0.094)	-0.132 (0.298)	0.146 (0.281)	0.221 (0.197)	0.333 (0.419)	0.214 (0.088)
ctsolv1	0.087 (0.167)	0.029 (0.226)	0.288 (0.139)	0.243 (0.209)	0.239 (0.103)	-0.377 (0.240)	-0.030 (0.284)	0.237 (0.183)	0.330 (0.445)	0.229 (0.098)
ctpar27	-0.172 (0.085)	-0.161 (0.226)	0.120 (0.118)	-0.022 (0.103)	-0.252 (0.063)	-0.181 (0.092)	-0.079 (0.146)	-0.196 (0.084)	0.026 (0.092)	-0.197 (0.066)
ctpar271	-0.186 (0.092)	-0.081 (0.200)	0.118 (0.123)	-0.026 (0.100)	-0.252 (0.061)	-0.187 (0.100)	-0.088 (0.153)	-0.198 (0.085)	0.034 (0.096)	-0.203 (0.062)

Table D: *Table of mean distances between reference atoms on Ca^{2+} coordinating ligands and the Ca^{2+} ions for the simulations of the C-term. Values in brackets indicate the standard deviation for the distribution over the simulation. Reference values come from the crystal structure used for the simulations, prior to minimisation and dynamics.*

Appendix D: PCA of Experimentally Observed Structures

This appendix contains the full results of the Principal Component analysis of observed structures used as a reference frame for the analysis of the Molecular Dynamics trajectories studied in Chapter 5. The structures used in this analysis are detailed in **Table E** below. The analysis was performed according to the method described in Chapter 3. Analysis on the N and C-term lobes was performed separately.

Eigen spectra for the PCA are shown in **Figure A** and **Figure B**. Conformational subspace projections of the observed structures onto combinations of the first three PC's of the N and C-term lobe distributions are illustrated in **Figures C**, **E** and **G** for the N-term and **Figures I**, **K** and **M** for the C-term. The experimentally observed structures are represented on by the codes detailed in **Table E** and coloured according to the classification also detailed in **Table E**.

Wall-eye stereo schematics are also included to indicate the shapes of the structural variation corresponding to the PC's in **Figures D**, **E** and **G** for the N-term and **Figures J**, **L** and **N** for the C-term. These representations were constructed by projecting the average structures from the distributions along increments of PC's 1 to three from the PCA.

Table E: *Descriptions of the full dataset of experimentally determined structures used in the PCA. Codes are PDB codes supplemented where necessary.*

Code	Descriptor	Experimental Technique	Classification
1a29	Calmodulin	XRD	Ligand
1ahr	Calmodulin Deletion Mutant (T79, D80)	XRD	Special
1aji	Calmodulin & Myosin Seq.	Model	Model
1ak8	Calmodulin (N-term)	NMR (23)	Apo
1alv	Calpain	XRD	Special
1alw	Calpain	XRD	Special
1avj	Calmodulin-Type Tch2 Protein	Model	Special
1cdl	Calmodulin & sm-MLCK	XRD	Target
1cdla	Calmodulin & sm-MLCK Chain A	XRD	Target
1cdlb	Calmodulin & sm-MLCK Chain B	XRD	Target
1cdlc	Calmodulin & sm-MLCK Chain C	XRD	Target
1cdld	Calmodulin & sm-MLCK Chain D	XRD	Target
1cdm	Calmodulin & CaMKII	XRD	Target
1cfc	Calmodulin	NMR (25)	Apo
1cfd	Calmodulin	NMR (Avg)	Apo
1cff	Calmodulin	NMR (26)	Target
1ckk	Calmodulin	NMR (30)	Target
1cll	Calmodulin	XRD	Native
1clm	Calmodulin	XRD	Native
1cm1	Calmodulin & CaMKIIα	XRD	Target
1cm4a	Calmodulin & CaMKIIα Model A	XRD	Target
1cm4c	Calmodulin & CaMKIIα Model B	XRD	Target
1cm4e	Calmodulin & CaMKIIα Model C	XRD	Target
1cm4g	Calmodulin & CaMKIIα Model D	XRD	Target
1cmf	Calmodulin (C-term)	NMR (20)	Apo
1cmg	Calmodulin (C-term)	NMR (20)	Native
1ctr	Calmodulin & TFP	XRD	Ligand
1deg	Calmodulin Mutant (Del E84)	XRD	Special
1dmo	Calmodulin	NMR (30)	Apo
1exr	Calmodulin	XRD	Native
1f70	Calmodulin (N-term res 1-76)	NMR (RDC)	Apo
1fw4	Calmodulin (C-term res 78-148)	XRD	Native
1g4y	Calmodulin & Rsk2	XRD	Target
1j7oa	Calmodulin (N-term) Model A	NMR (RDC)	Native
1j7ob	Calmodulin (N-term) Model B	NMR (RDC)	Native
1j7oc	Calmodulin (N-term) Model C	NMR (RDC)	Native
1j7pa	Calmodulin (C-term) Model A	NMR (RDC)	Native
1j7pb	Calmodulin (C-term) Model B	NMR (RDC)	Native
1j7pc	Calmodulin (C-term) Model C	NMR (RDC)	Native
1lin	Calmodulin	XRD	Ligand
1mux	Calmodulin	NMR	Ligand
1ncx	Troponin C	XRD	Special
1osa	Calmodulin	XRD	Native
1qiv	Calmodulin	XRD	Ligand
1qiwa	Calmodulin Chain A	XRD	Ligand
1qiwa	Calmodulin Chain B	XRD	Ligand
1tn4	Troponin C	XRD	Special
1vrk	Calmodulin & RS20	XRD	Target
2bbm	Calmodulin & sk-MLCK	NMR (Avg)	Target
2bbn	Calmodulin & sk-MLCK	NMR (21)	Target
2cln	Trimethyl-calmodulin & TFP	Model	Special
3cln	Calmodulin	XRD	Native
4cln	Calmodulin	XRD	Native

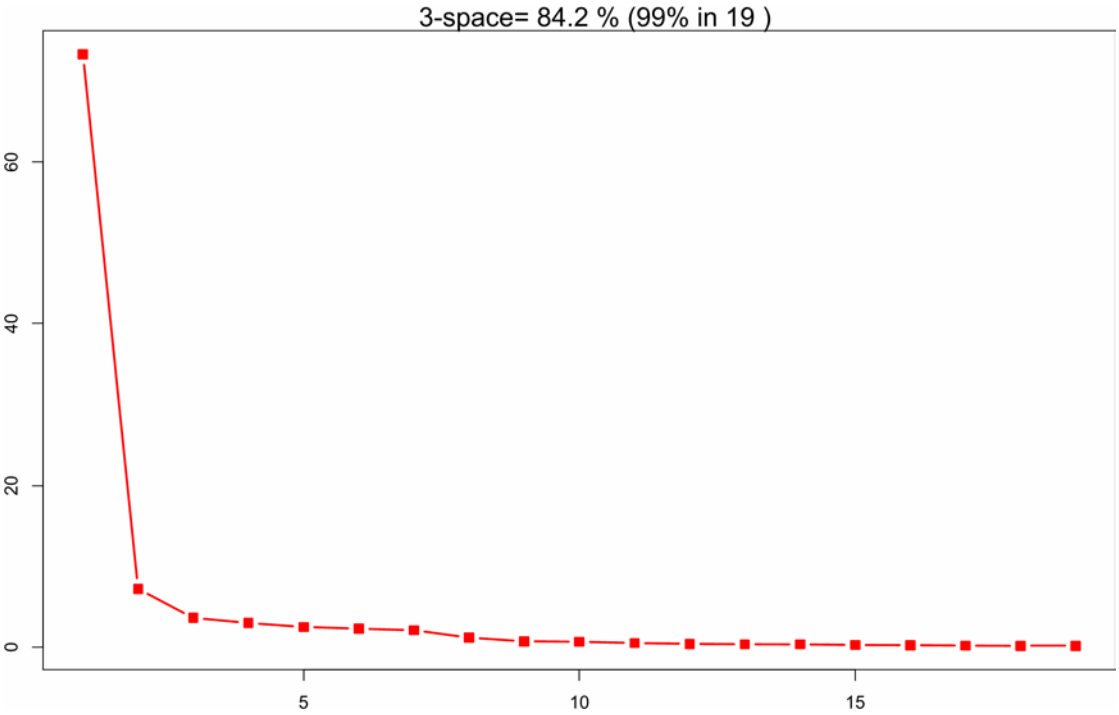


Figure A: *Eigen spectrum for the PCA of the observed structures for the N-term*

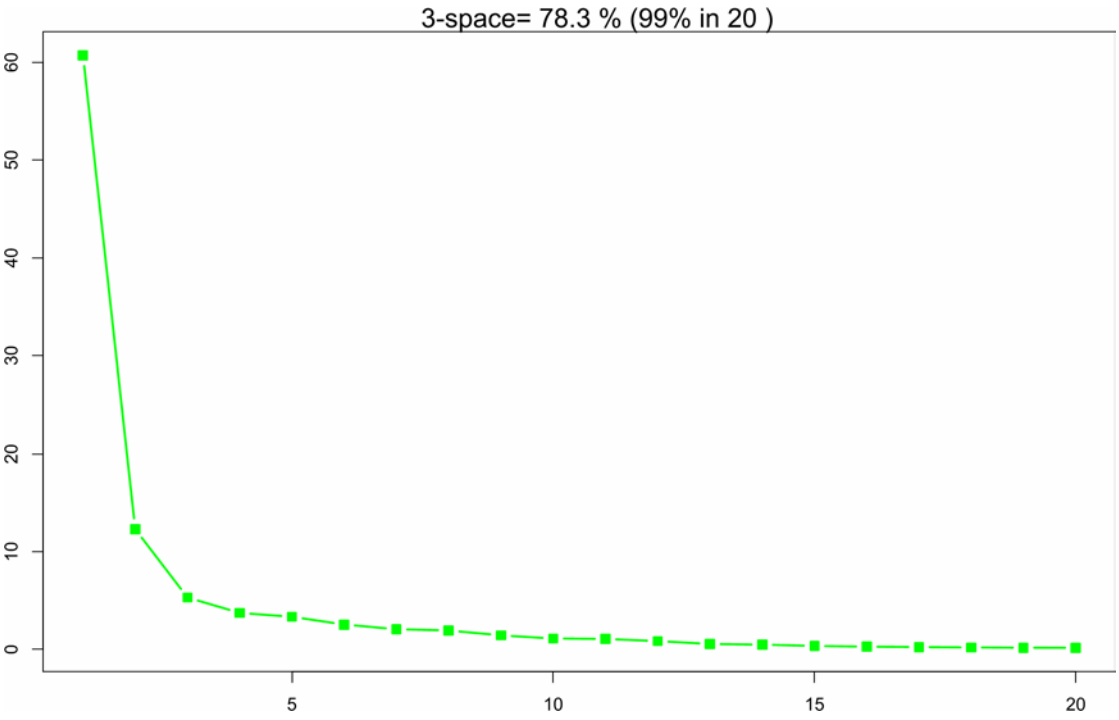


Figure B: *Eigen spectrum for the PCA of the observed structures for the C-term*

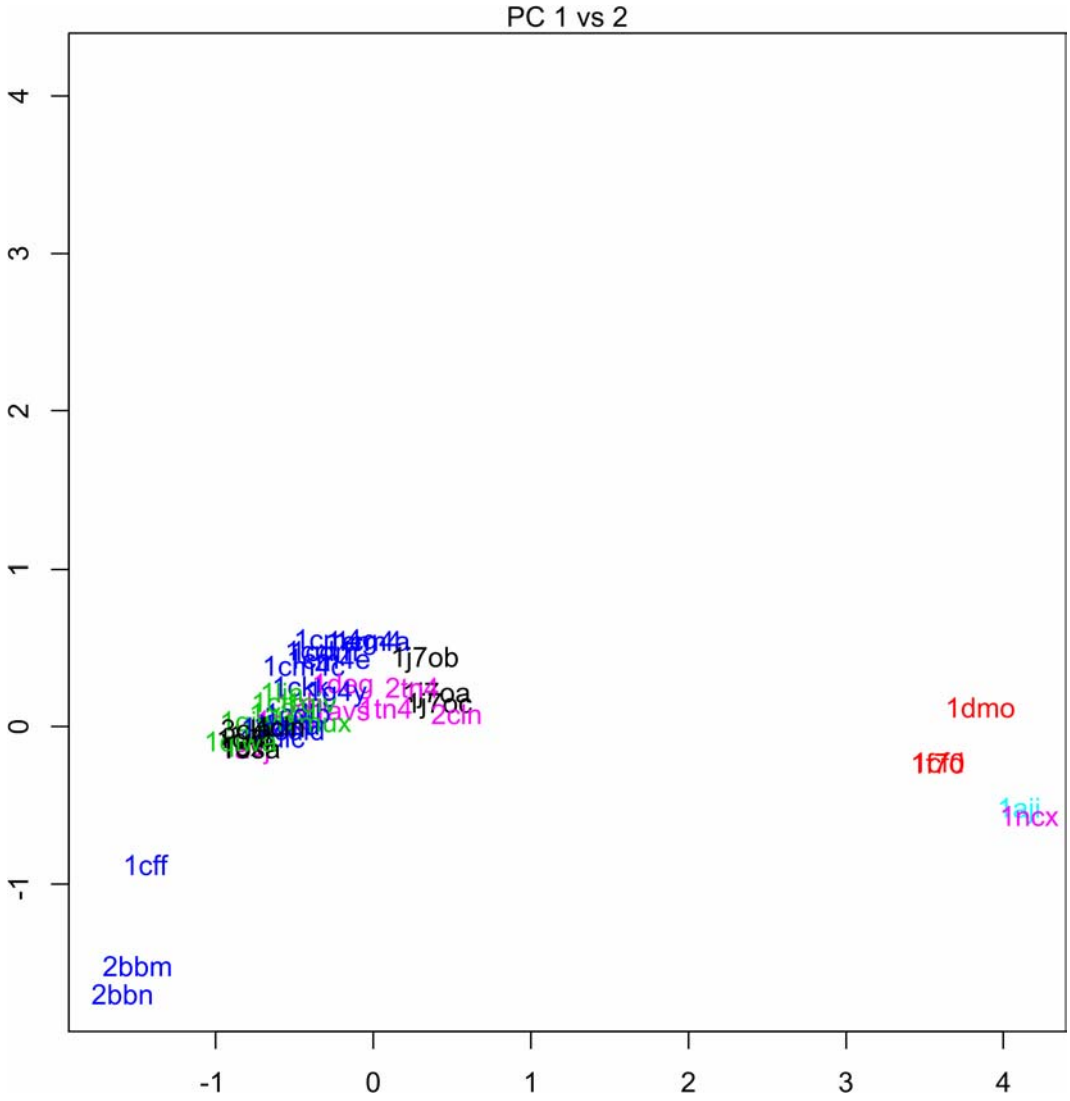


Figure C: *Conformational Subspace projection for the observed structures onto PC's 1 and 2 of the observed distribution for the N-term.*

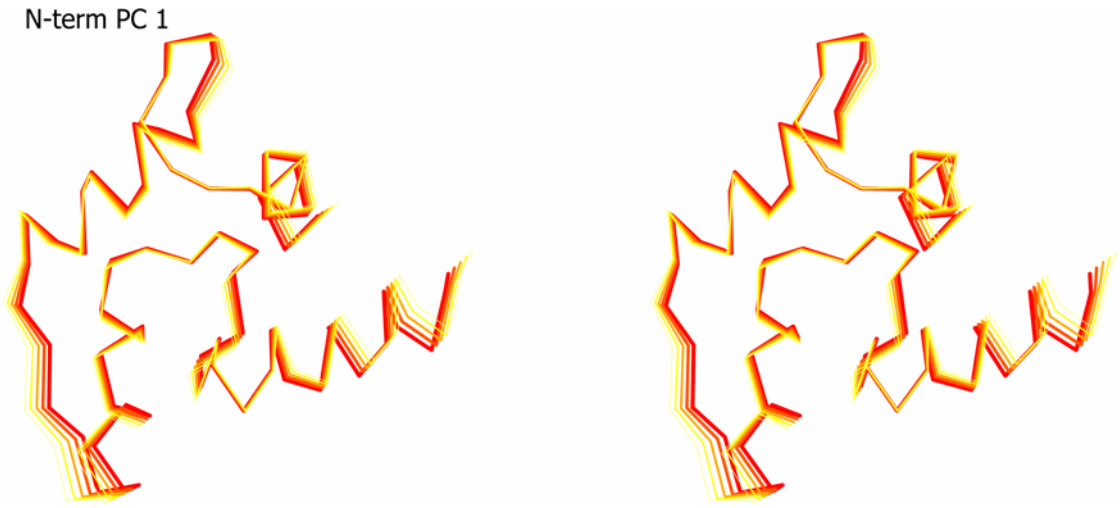


Figure D: *Wall-eye stereo plot of the projection of the average structure from the distribution along PC1 for the N-term.*

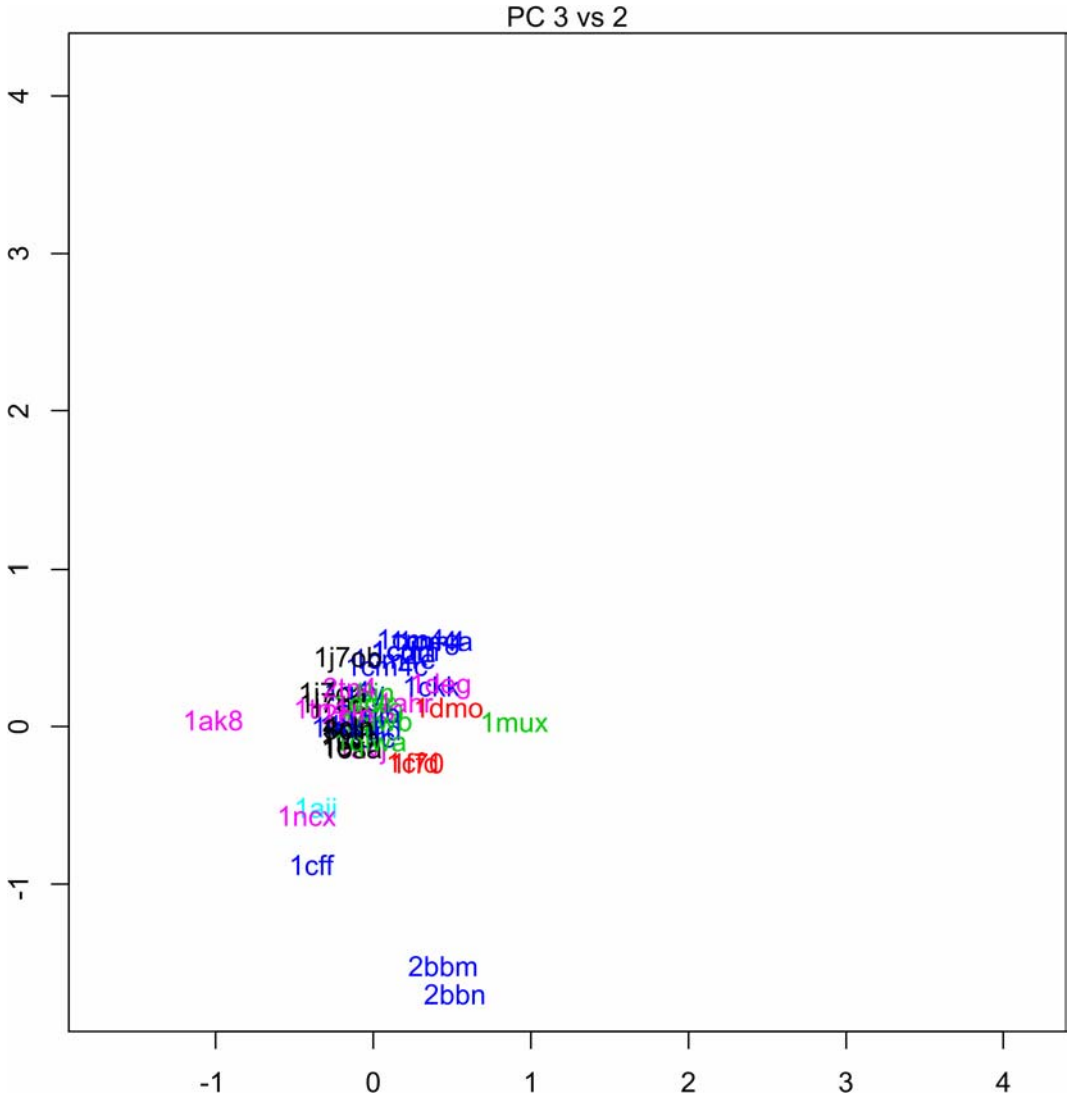


Figure E: *Conformational Subspace projection for the observed structures onto PC's 3 and 2 of the observed distribution for the N-term.*

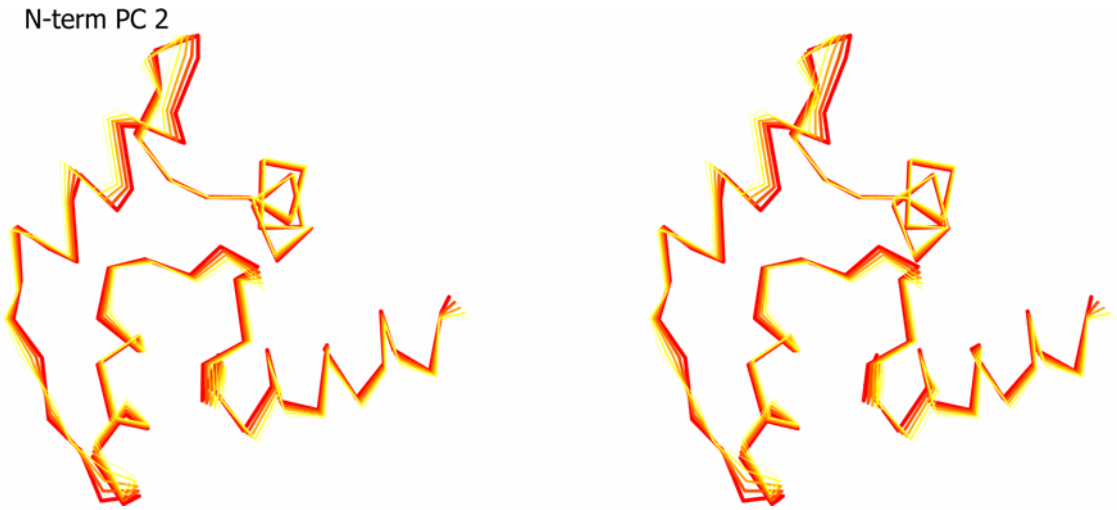


Figure F: *Wall-eye stereo plot of the projection of the average structure from the distribution along PC2 for the N-term.*

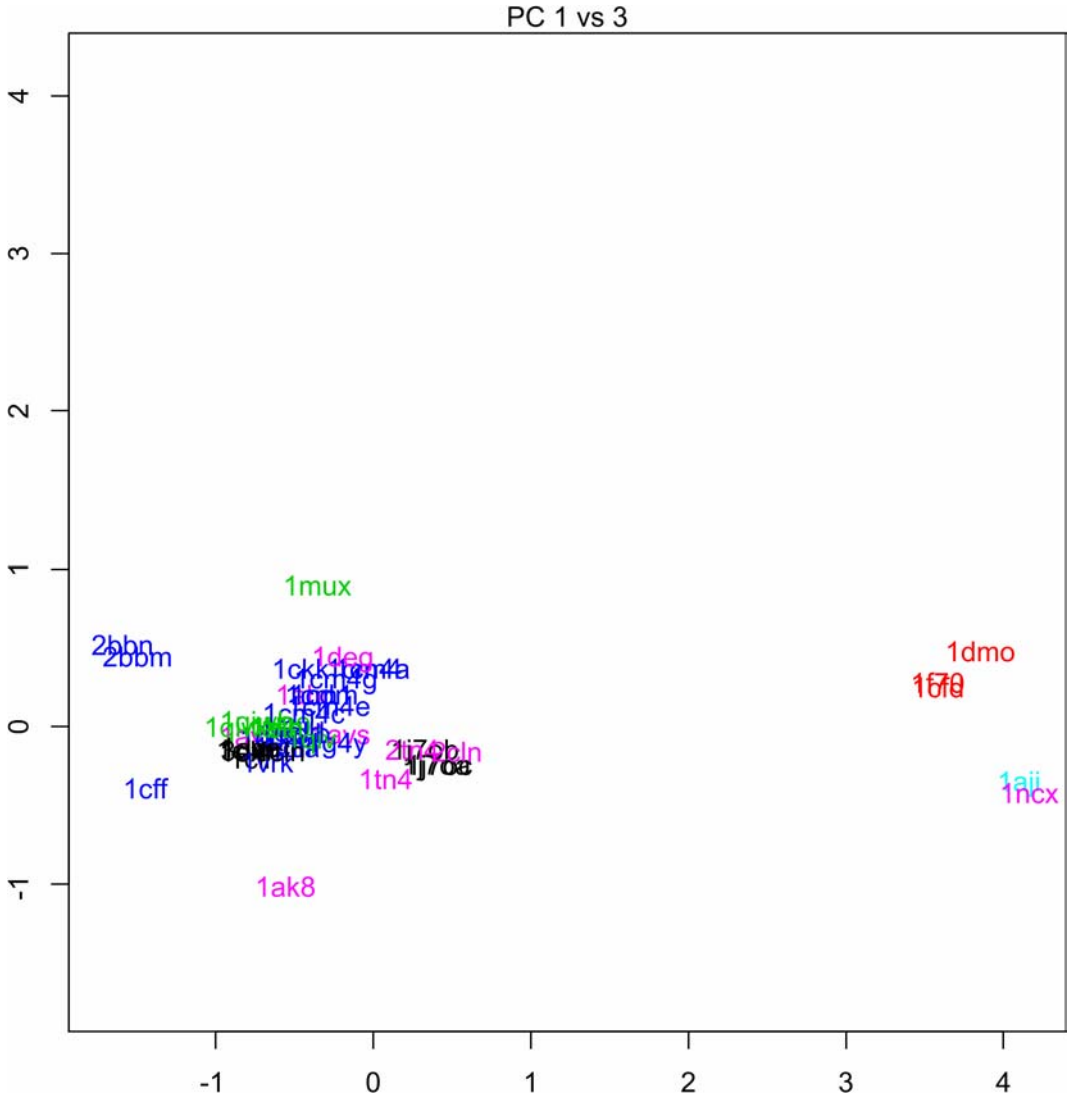


Figure G: *Conformational Subspace projection for the observed structures onto PC's 1 and 3 of the observed distribution for the N-term.*

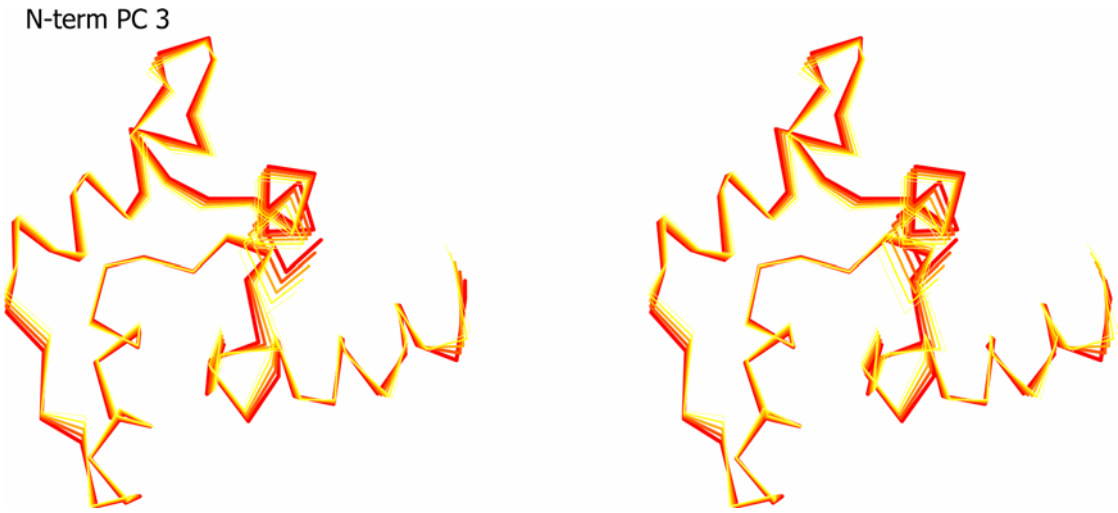


Figure H: *Wall-eye stereo plot of the projection of the average structure from the distribution along PC3 for the N-term.*

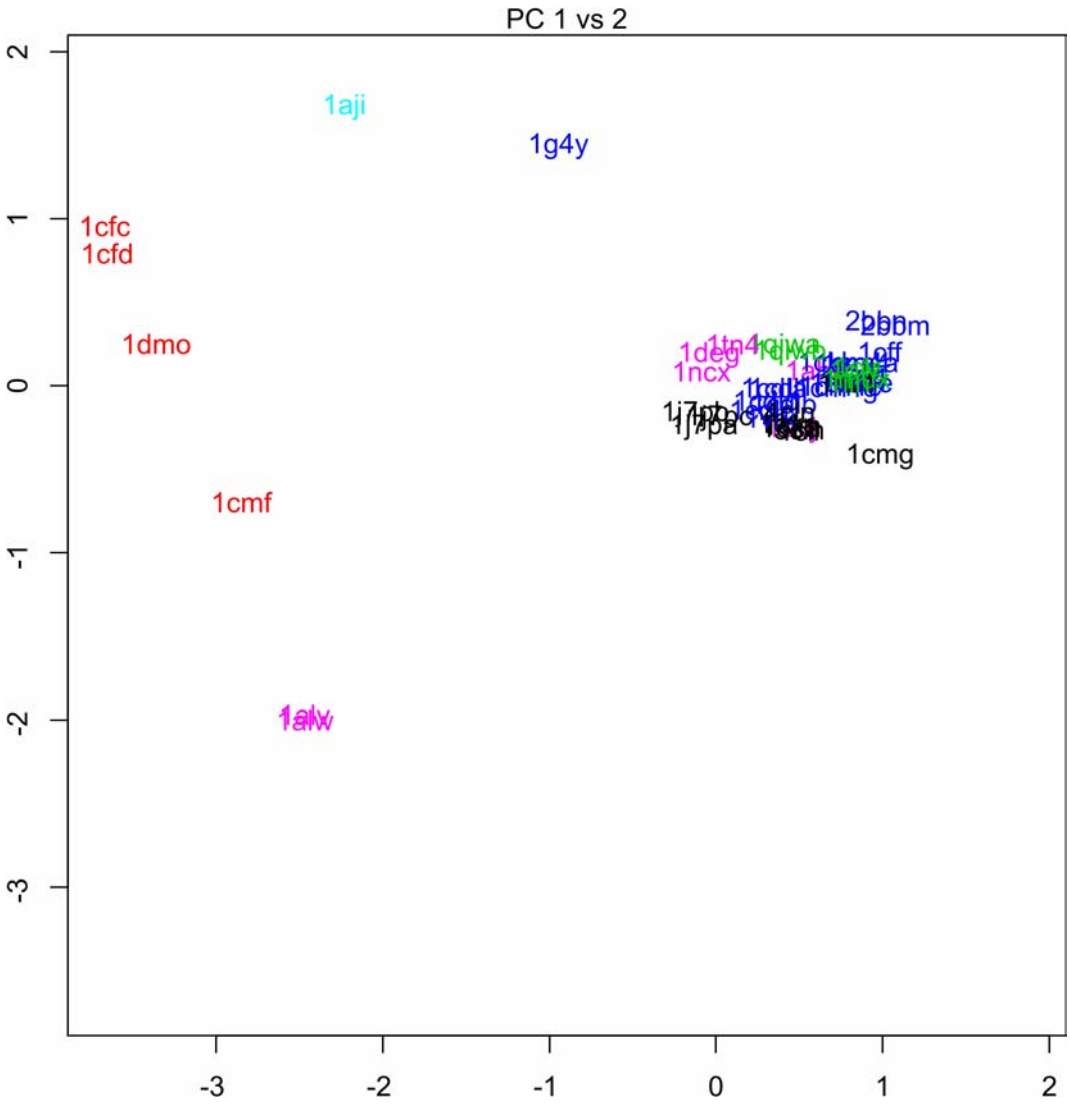


Figure I: *Conformational Subspace projection for the observed structures onto PC's 1 and 2 of the observed distribution for the C-term.*

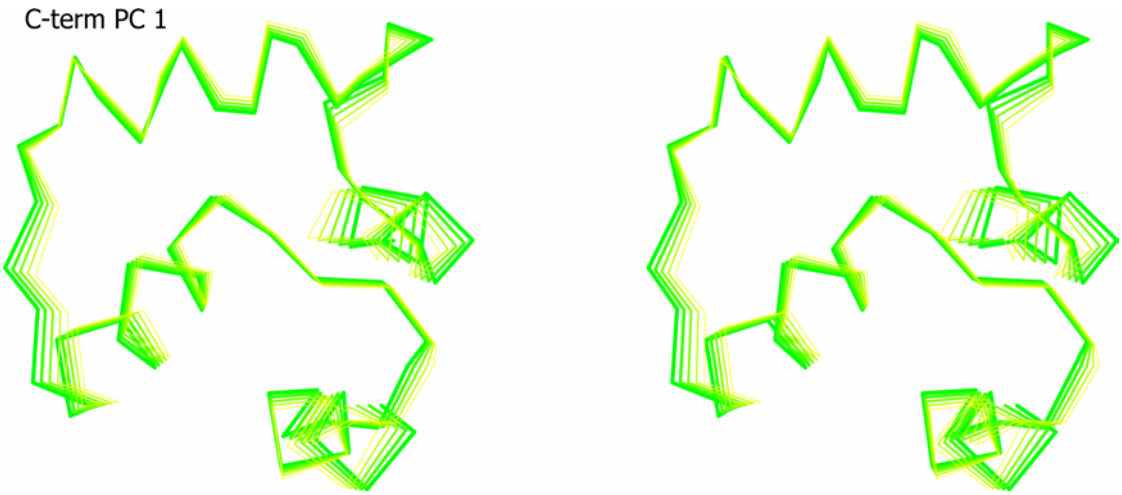


Figure J: *Wall-eye stereo plot of the projection of the average structure from the distribution along PC1 for the C-term.*

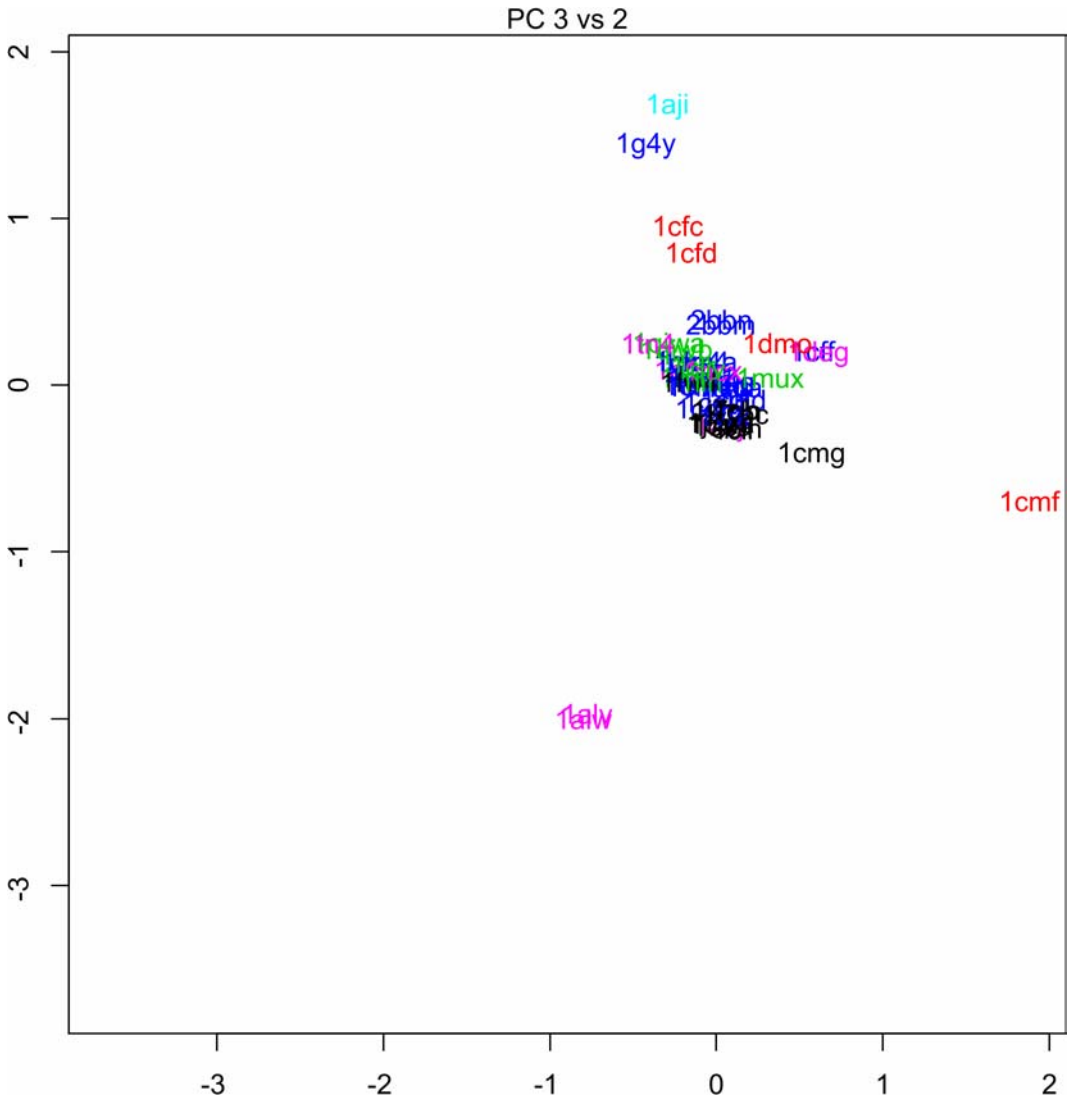


Figure K: *Conformational Subspace projection for the observed structures onto PC's 3 and 2 of the observed distribution for the C-term.*

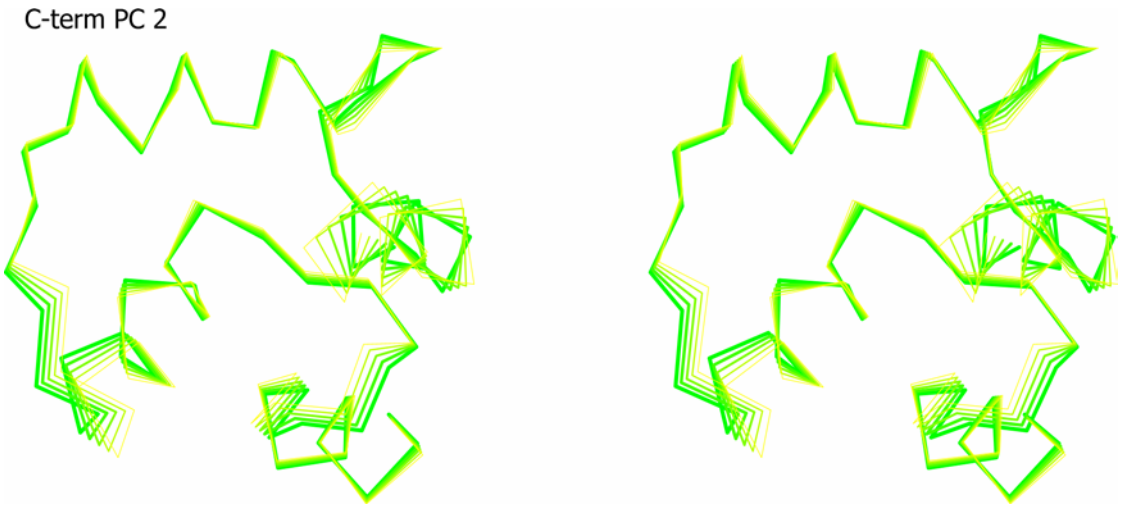


Figure L: *Wall-eye stereo plot of the projection of the average structure from the distribution along PC2 for the C-term.*

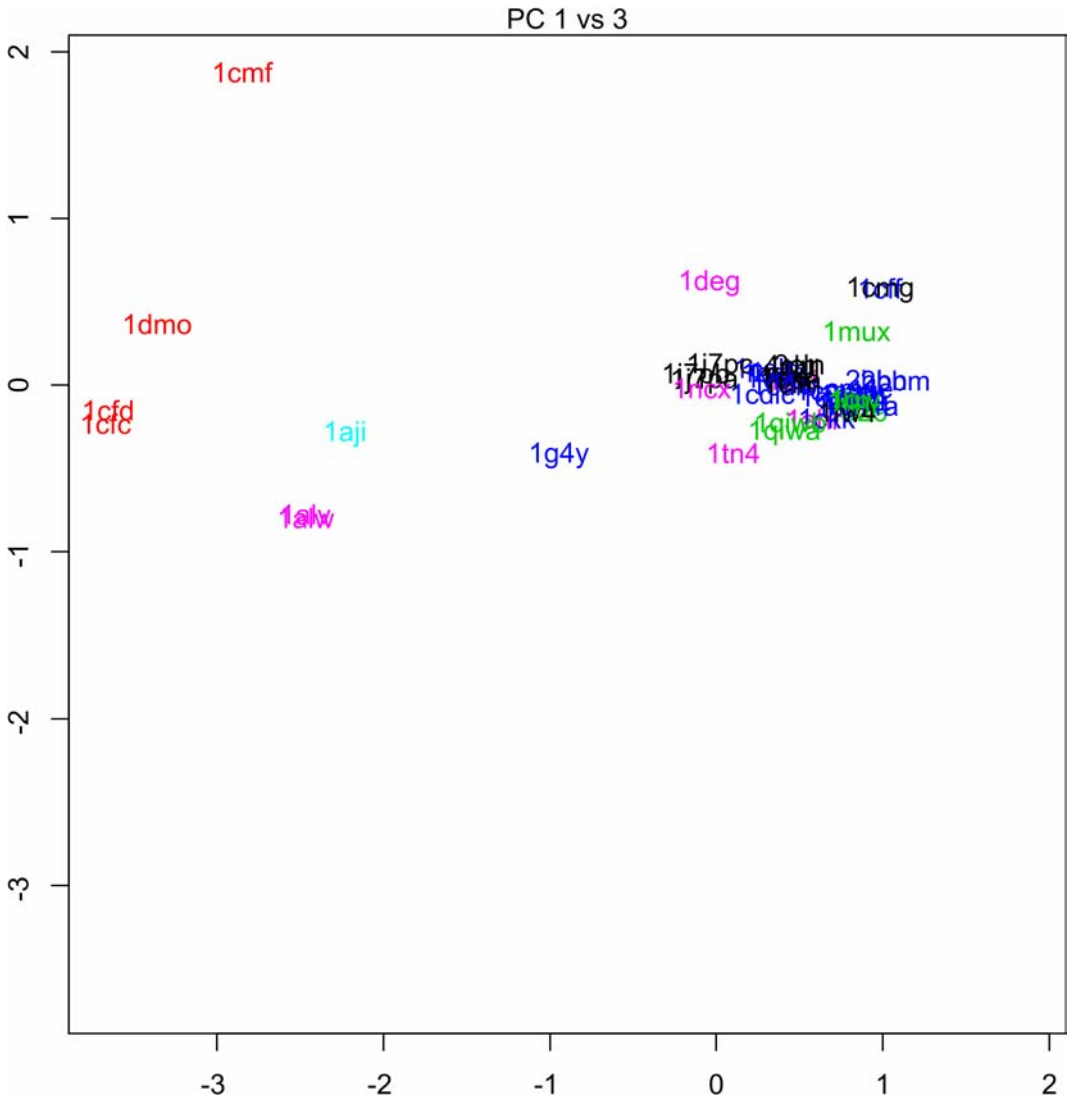


Figure M: *Conformational Subspace projection for the observed structures onto PC's 1 and 3 of the observed distribution for the C-term.*

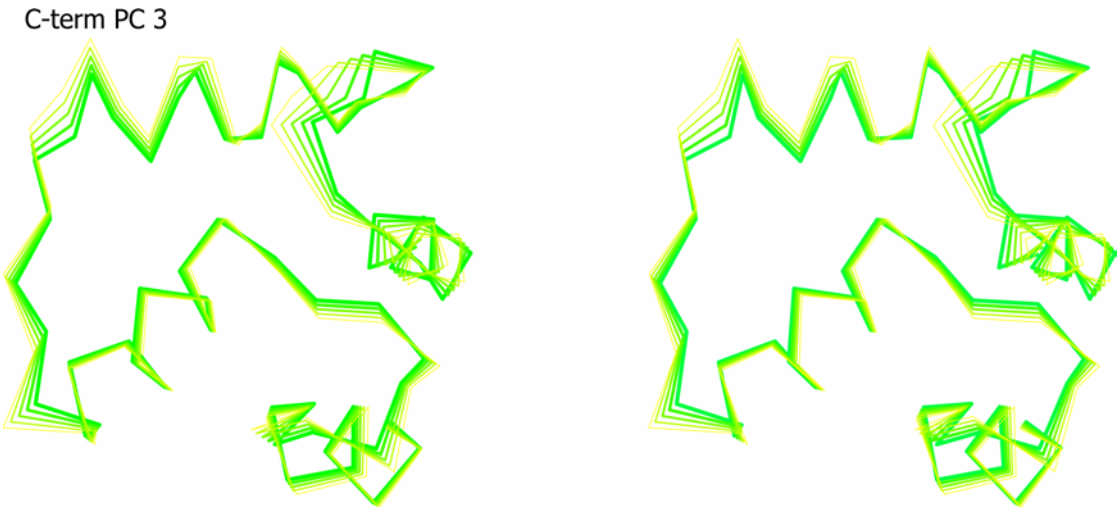


Figure N: *Wall-eye stereo plot of the projection of the average structure from the distribution along PC3 for the C-term.*