REGULATORY DOMAINS OF THE HUMAN CALPAIN FAMILY

by

RAVIKIRAN RAVULAPALLI

A thesis submitted to the Department of Biochemistry
In conformity with the requirements for
the degree of Doctor of Philosophy

Queen’s University
Kingston, Ontario, Canada
(February, 2009)

Copyright © Ravikiran Ravulapalli, 2009
Abstract

Calpains are intracellular enzymes that merge cysteine protease and calcium sensing activities together in one molecule. They respond to $\text{Ca}^{2+}$ signals and modify the activity of their targets by selective proteolysis. Calpains are involved in normal cellular processes like cell migration and apoptosis. The over-activation of calpain due to disturbances in $\text{Ca}^{2+}$ homeostasis or inactivation due to mutations, contribute to diseases like ischemic injury and muscular dystrophy.

The classical calpains 1 and 2 are heterodimeric enzymes containing a large (80 kDa) subunit and a small subunit (28 kDa). Dimerization occurs through the 5th EF-hand of penta-EF-hand (PEF) domains present in both large and small subunits. In this study, I have used structural genomics approaches to explore the PEF and C2-like regulatory domains of some of the other 12 human calpain isoforms. I have shown that recombinant PEF domain of skeletal muscle-specific calpain 3 exists as a stable homodimer when produced alone. Modelling studies suggest that there would be no barriers for dimerization of the full-length enzyme through the PEF domains which would place the protease cores at opposite ends of the dimer. Co-expression studies using small subunit were performed with PEF domains of calpains 1, 3, 8, 9, 11, 12 and 13. A differential tagging system was devised to differentiate heterodimers from
homodimers. The PEF domains of calpains 1, 3, 9 and 13 co-expressed with the small subunit, while the others failed to express. The PEF domains of calpains 1 and 9 formed heterodimers. Conversely, the PEF domain of calpain 3 formed a homodimer and that of calpain 13 predominantly formed a homodimer with a small amount of heterodimer. Homodimerization of calpains implies they are less-likely to be inhibited by the endogenous calpain inhibitor, calpastatin.

C2-like regulatory domains of calpains 5-13 were also studied. The structure of the distal C2-like domain of calpain 7 was solved. It is markedly different from canonical C2 domains and may not bind Ca$^{2+}$.
Acknowledgements

First and foremost, I would like to acknowledge my research supervisor, Dr. Peter L. Davies for having faith in me and providing me with the opportunity to carry out graduate studies in his wonderful lab. From the first telephone interview I had with him to the last day of my graduate research, Peter has been the same generous, caring and supporting individual. Peter’s enthusiasm and dedication to scientific research and his lab members is exemplary. His excellent guidance and motivation contributed immensely to my growth as a research scientist and as an individual. I consider it a great honor to have spent my graduate years under his guidance and I will be forever indebted to him for what he has given me both professionally and personally.

I wish to thank my loving wife Manoja Eswara who has been with me during these testing times. Being a Ph.D. student herself, she handles my temperament extremely well. Her steady support and motivation meant a lot to me over these years and I am grateful to her. Thanks also goes to my parents R.P.C. Rao and S.S.M. Lakshmi, for I would not be where I am today without their continuing love and support. My special thanks to my sister Sridevi and cousin brother’s Sridhar and Ravichandra, they bring joy to my life and I am fortunate to have them as part of my life. I would also like
to thank my aunt S.R. Rajeswari and uncle G. Tatarao, my grandparents S. Chittiah Chowdhary and S. Samrajyam for their support and good wishes.

I wish to acknowledge all the Davies lab members past and present for making my stay at Queen’s enjoyable and memorable: Dr. Dominic Cuerrier, Dr. Beatriz Garcia Diaz, Adam Middleton, Chris Garnham, Rachel Hanna, Jacqueline Kelly, Jack Gilbert, Dr. Andy Scotter, Dr. Laurie Graham, Nelson Lin, Sally Yu, Kyra Nabeta, Jin Qian, Dr. Yee-Foong Mok, Jordan Chou and Kristin Low. I am extremely lucky to have worked with such incredible people and I will never forget them. My special thanks reserved for life-line of Davies lab, Sherry Y. Gauthier. I can’t imagine these 4-5 years without Sherry. Sherry is not only an outstanding technician but also an outstanding person. During the past 4 years we became good friends. Her patience in answering my innumerable doubts is worth mentioning. I would also like to thank Dr. Robert Campbell for his guidance with PyMOL and other interesting crystallographic software and for successfully solving many “roblems”.

I wish to extend special acknowledgements to:

- Dr. Sirano Dhe-Paganon, our collaborator from Structural Genomics Consortium, Toronto. Sirano has been incredibly flexible and also accommodated me for six
months at SGC to gain technical expertise. Most of the work in this thesis would not have been possible without his guidance and support

- Advisory committee members Dr. Graham P. Côté and Dr. Wayne Snedden for giving timely suggestions and guiding my research
- Kim Munro of PFD facility for his technical assistance
- Dr. Zongchao Jia and Dr. Steven P. Smith labs for letting me use their lab equipment

I would also like to remember Late Shri. Nandamuri Taraka Ramarao, a legendary actor, unparalleled politician and an outstanding humanitarian. He is my idol and will continue to be a great inspiration throughout my life. Drawing inspiration from him, I would like to contribute my best to mankind through possible scientific, humanitarian and charity work.

Finally, I am extremely grateful to Queen’s University for funding me through Tuition Bursaries and Graduate Student Scholarships throughout my graduate study. I am also extremely grateful to the Canadian Institutes of Health Research and Heart and Stroke Foundation for providing funding for my research and the Canadian Foundation for Innovation for infrastructural support.
# Table of Contents

Abstract ................................................................................................................................. ii

Acknowledgements ............................................................................................................... iv

Table of Contents ................................................................................................................ vii

List of Figures ....................................................................................................................... xii

List of Tables ......................................................................................................................... xiv

Chapter 1 General Introduction .......................................................................................... 1

1.1 Cellular Ca^{2+} homeostasis and Ca^{2+} as intracellular messenger .......................... 1

1.1.1 Why Ca^{2+}? ............................................................................................................. 2

1.1.2 Elements of cellular Ca^{2+} signaling mechanism .................................................... 6

1.1.2.1 Ca^{2+} increase in cytoplasm .............................................................................. 6

1.1.2.2 Role of Ca^{2+} binding proteins ........................................................................... 9

1.1.2.3 Ca^{2+} export ......................................................................................................... 13

1.2 Cysteine proteases ......................................................................................................... 14

1.3 Calpains .......................................................................................................................... 15

1.3.1 Calpain nomenclature ............................................................................................. 17

1.3.2 Structure of calpains ............................................................................................... 19

1.3.3 Mechanism of calpain activation by Ca^{2+} ............................................................ 26

1.3.4 Calpastatin ................................................................................................................. 29
1.3.5 Other members of calpain family ......................................................... 33
  1.3.5.1 Calpain 3 ......................................................................................... 33
  1.3.5.2 Calpain 6 ......................................................................................... 34
  1.3.5.3 Calpain 8 ......................................................................................... 34
  1.3.5.4 Calpain 9 ......................................................................................... 35
1.3.6 Calpain in physiology and pathology ..................................................... 35
1.3.7 Research Objective ............................................................................ 36

Chapter 2 Homodimerization of calpain 3 penta-EF-hand domain .................. 38
  2.1 Abstract ............................................................................................... 39
  2.2 Introduction .......................................................................................... 39
  2.3 Materials and Methods ......................................................................... 42
    2.3.1 Cloning of p94DIV .......................................................................... 42
    2.3.2 Protein expression and purification .................................................. 43
    2.3.3 Size-exclusion chromatography ....................................................... 44
    2.3.4 CD spectroscopy ............................................................................. 45
    2.3.5 Intrinsic tryptophan fluorescence studies ........................................ 45
    2.3.6 Analytical ultracentrifugation .......................................................... 46
    2.3.7 Homology modelling of the p94DIV dimer ...................................... 47
    2.3.8 Graphic model of the m-calpain large-subunit dimer ....................... 48
2.4 Results ................................................................................................................................................. 49

2.4.1 Expression and purification of recombinant human p94 domain IV ........................................ 49

2.4.2 CD spectra and intrinsic tryptophan fluorescence studies ....................................................... 52

2.4.3 P94DIV is a stable dimer in solution ............................................................................................. 57

2.4.4 Homology modelling of p94DIV ................................................................................................. 62

2.5 Discussion ............................................................................................................................................. 67

Chapter 3 Distinguishing between calpain heterodimerization and homodimerization . 71

3.1 Abstract .................................................................................................................................................. 72

3.2 Introduction .......................................................................................................................................... 73

3.3 Materials and Methods ..................................................................................................................... 76

3.3.1 High throughput cloning .............................................................................................................. 76

3.3.2 Type III antifreeze protein-tagged human small subunit ............................................................. 79

3.3.3 Protein Expression and purification by Ni-NTA ........................................................................... 80

3.3.4 Ice Affinity Purification ............................................................................................................... 81

3.3.5 Modelling studies ......................................................................................................................... 82

3.3.6 Immunoblotting ............................................................................................................................ 83

3.4 Results .................................................................................................................................................. 83

3.4.1 Establishing the validity of the screening method using calpain 1 DIV ............................... 84

3.4.2 Calpain 3 DIV is a homodimer .................................................................................................... 89
3.4.3 Calpain 9 DIV forms a heterodimer with the small subunit ................. 90
3.4.4 Calpain 13 DIV .................................................................................. 93
3.5 Discussion .............................................................................................. 98

Chapter 4 Structural genomics of other regulatory domains ..................... 103
4.1 Introduction ............................................................................................ 104
4.2 Materials and Methods ......................................................................... 105
  4.2.1 High throughput cloning ................................................................ 105
  4.2.2 Expression and purification of recombinant proteins ................. 105
  4.2.3 Crystal Screening ............................................................................ 107
4.3 Results and Discussion ......................................................................... 107
  4.3.1 Expression of C2-like regulatory domains .................................. 107
  4.3.2 Calpain 7 distal C2-like domain structure .................................. 108
4.4 Structure of calpain 13 PEF domain .................................................... 115

Chapter 5 Characterization of human tissue-specific isoforms – Calpain 6 ..... 118
5.1 Introduction ............................................................................................ 119
5.2 Methodology ........................................................................................ 120
  5.2.1 Cloning, expression and purification of calpain 6 protease core (CC6) .... 120
  5.2.2 Mutagenesis of CC6 ........................................................................ 120
  5.2.3 Enzyme assays and intrinsic tryptophan fluorescence studies .......... 121
5.2.4 Sequence alignment .......................................................................................... 121
5.3 Results and Discussion ....................................................................................... 126
Chapter 6 ...................................................................................................................... 128
General Discussion ...................................................................................................... 128
6.1 Structural genomics approach to characterize calpains ...................................... 128
6.2 Understanding dimerization in calpain isoforms: implications for calpastatin
  inhibition ................................................................................................................. 131
6.3 Ca\(^{2+}\) binding properties of calpain isoforms ............................................... 137
  6.3.1 Ca\(^{2+}\) binding in PEF domains ................................................................. 137
  6.3.2 Ca\(^{2+}\) binding in C2-like domains ......................................................... 138
6.4 Summary and Conclusions ................................................................................ 139
Bibliography (or References) .................................................................................... 142
List of Figures

Figure 1.1. Coordination of Mg\(^{2+}\) and Ca\(^{2+}\) in Calbindin D\(_{9k}\).......................................................... 4

Figure 1.2. Control of cellular Ca\(^{2+}\)............................................................................................................. 7

Figure 1.3. EF-Hand and C2 domains- two widespread Ca\(^{2+}\) binding domains ......................... 11

Figure 1.4. Human Calpain System ............................................................................................................. 21

Figure 1.5. Three-dimensional structure of inactive m-calpain heterodimer ......................... 23

Figure 1.6. Realignment of active site cleft in mini-calpain......................................................... 27

Figure 1.7. Overview of calpastatin domain 4 (CAST4) bound to m-calpain (calpain 2) .. 31

Figure 2.1. Human calpain p94 domain IV and its purification ......................................................... 50

Figure 2.2. CD spectra and intrinsic tryptophan studies on p94DIV.............................................. 55

Figure 2.3. Size-exclusion (Sephadex G-75) chromatography standard curve ......................... 58

Figure 2.4. Sedimentation equilibrium distribution for p94IV ......................................................... 60

Figure 2.5. Homology model of p94DIV................................................................................................. 63

Figure 2.6. Graphic model of the large subunit dimer ................................................................. 65

Figure 3.1. Three possible scenarios from co-expression of recombinant PEF fusion proteins ................................................................................................................. 77

Figure 3.2. SDS-PAGE and immunoblot analysis of differentially tagged calpain 1 and 2 heterodimers ......................................................................................................................................... 87
Figure 3.3. SDS-PAGE and immunoblot analysis of DIV calpain 3 and 9 samples co-expressed with small subunit ................................................................. 91

Figure 3.4. Ice affinity purification of type III AFP-tagged small subunit and calpain 9 DIV .............................................................................................................. 94

Figure 3.5. SDS-PAGE analysis of calpain 13 DIV samples from Ni-NTA column .......... 96

Figure 4.1 Distal C2-like domain of calpain 7 (PDB ID: 2QFE) .................................. 111

Figure 4.2 Superimposition of the distal C2-like domain of calpain 7 on domain III of m-calpain ........................................................................................................... 113

Figure 4.3 High resolution crystal structure (1.8 Å) of calpain 13 PEF domain .......... 116

Figure 5.1 Multiple sequence alignment of calpain 1, 2 and 6 protease cores .......... 122

Figure 5.2 Inactive m-calpain (mC105S) proteolysis activity assay .......................... 124

Figure 6.1 Calpain PEF domains – Sequence alignment and dimerization structures ... 135
List of Tables

Table 1-1 Calpain Nomenclature ................................................................. 18
Table 2-1 Comparison of secondary structure content of p94DIV ....................... 53
Table 3-1 Screening results for PEF domains in calpains ..................................... 85
Table 4-1 Structural genomics of regulatory domains: Screening results ............. 109
List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAST</td>
<td>calpastatin</td>
</tr>
<tr>
<td>COP</td>
<td>coatamer protein</td>
</tr>
<tr>
<td>cADPR</td>
<td>cyclic ADP ribose</td>
</tr>
<tr>
<td>DIV</td>
<td>domain IV</td>
</tr>
<tr>
<td>DEAE</td>
<td>diethylamine ethyl</td>
</tr>
<tr>
<td>DMSO</td>
<td>dimethyl sulfoxide</td>
</tr>
<tr>
<td>DTT</td>
<td>dithiothreitol</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylene-diamine tetraacetic acid</td>
</tr>
<tr>
<td>ER</td>
<td>endoplasmic reticulum</td>
</tr>
<tr>
<td>IS</td>
<td>insertion sequence</td>
</tr>
<tr>
<td>IPTG</td>
<td>iso-propyl β-thiogalactoside</td>
</tr>
<tr>
<td>IWF</td>
<td>intrinsic tryptophan fluorescence</td>
</tr>
<tr>
<td>NAD</td>
<td>nicotinamide-adenine dinucleotide</td>
</tr>
<tr>
<td>Ni-NTA</td>
<td>Ni$^{2+}$-nitrilotriacetate–agarose</td>
</tr>
<tr>
<td>NMR</td>
<td>nuclear magnetic resonance</td>
</tr>
<tr>
<td>NS</td>
<td>N-terminal sequence;</td>
</tr>
<tr>
<td>PAGE</td>
<td>polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein Data Bank</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>PEF</td>
<td>penta-EF-hand</td>
</tr>
<tr>
<td>PLC</td>
<td>phospholipase C</td>
</tr>
<tr>
<td>PMSF</td>
<td>phenylmethyl sulfonyl fluoride</td>
</tr>
<tr>
<td>RMSD</td>
<td>root mean square deviation</td>
</tr>
<tr>
<td>ROC</td>
<td>receptor-operated channel</td>
</tr>
<tr>
<td>RTK</td>
<td>receptor tyrosine kinase</td>
</tr>
<tr>
<td>RYR</td>
<td>ryonodine receptor</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulfate</td>
</tr>
<tr>
<td>SER</td>
<td>sarco-endoplasmic reticulum</td>
</tr>
<tr>
<td>SGC</td>
<td>Structural Genomics Consortium</td>
</tr>
<tr>
<td>SOC</td>
<td>storage-operated channel</td>
</tr>
<tr>
<td>SR</td>
<td>sarcoplasmic reticulum</td>
</tr>
<tr>
<td>Tris</td>
<td>tris(hydroxymethyl) aminomethane</td>
</tr>
<tr>
<td>VOC</td>
<td>voltage-operated channel</td>
</tr>
</tbody>
</table>
Chapter 1

General Introduction

Calpains integrate two important and extensively studied processes; Ca\(^{2+}\) signaling and proteolysis, both of which are known to play a decisive role in a cell’s origin, survival and death [1-3]. The dependence of these proteases on Ca\(^{2+}\) makes them respond to Ca\(^{2+}\) signals, resulting in selective cleavage of specific downstream target proteins, which are often components from signaling pathways, thereby irreversibly modulating their function. Calpains contain a common papain-like Ca\(^{2+}\) dependent protease core which later evolved into a highly modular enzyme, presumably through fusion with other domains [4-6]. These additional domains are thought to add stability to the protease core, perform tissue-specific functions, and participate in the localization and regulation of the whole enzyme in response to Ca\(^{2+}\). My thesis is mainly focused on one such regulatory domain, the Ca\(^{2+}\)-binding penta EF-hand (PEF) domain, which is present in 9 calpain isoforms in humans.

1.1 Cellular Ca\(^{2+}\) homeostasis and Ca\(^{2+}\) as intracellular messenger

In the human body, most of the Ca\(^{2+}\) content is found in bone (∼99.5-99.9%) and the remaining amount (∼0.1-0.5 %) is distributed over the rest of the body mostly as
Ca^{2+} pools in extracellular spaces (ECV) and intracellular spaces (ICV) [7]. Ca^{2+} has long been known for its biomineralization role in stabilizing structures i.e. bone, shell and other hard tissues or for its utility in maintaining the integrity of cell-substratum and cell adhesions. More recently, the focus has been on its dynamic role in cell physiology, where the movement of Ca^{2+} from extracellular spaces into the cytoplasm and out again functions as a signal for many cellular processes including fertilization, muscle contraction, exocytosis/endocytosis, blood clotting, aspects of metabolism, enzyme activities and cell death [2, 8-12]. The Ca^{2+} level inside the cell in eukaryotes (~100 nM) is four orders of magnitude lower than that present outside. The control of Ca^{2+} concentration inside the cytoplasm is a decisive factor in the management of many essential cellular responses as high Ca^{2+} accumulation could lead to organelle damage, aggregation of nucleic acids and proteins, and precipitation of phosphates, causing cell damage and death. To avoid this risk, living organisms have evolved efficient Ca^{2+} regulatory mechanism involving Ca^{2+} influx/ export systems and Ca^{2+} buffering systems to maintain internal Ca^{2+} levels that are not harmful for the cell [13].

1.1.1 Why Ca^{2+}?

The evolutionary choice of Ca^{2+} as a universal and versatile intracellular messenger among other inorganic ions (e.g. H^{+}, Mg^{2+}, Zn^{2+}, Na^{+}, and K^{+}) is attributed to its
favourable chemical properties. Ca\(^{2+}\) possesses a couple of important prerequisites for being a good secondary messenger:

a) Flexible and rapid coordination

The ionic radius of Ca\(^{2+}\) (1.1-1.2 Å) is much larger than that of other divalent cations, such as Mg\(^{2+}\) and Zn\(^{2+}\). This gives Ca\(^{2+}\) coordination more flexibility making it an ideal ligand to bind carboxylate-rich sites on proteins with a wide-range of binding constants. The Ca\(^{2+}\) ion typically exhibits high coordination (6-8) numbers as opposed to the more stringent constraint of Mg\(^{2+}\) for six ligands arranged in an octahedral configuration, limiting its occurrence in proteins (Figure 1.1). Studies have also shown that kinetic rates of binding and dissociation of Ca\(^{2+}\) in comparison to Mg\(^{2+}\) are approximately 100-fold faster [14].

b) ability to reduce intracellular concentration levels

The principal force guiding the swiftness and efficiency of the Ca\(^{2+}\) signal is the 20,000-fold gradient maintained by cells between their intracellular (∼100 nM) and extracellular (>1 mM)[15] concentrations. In contrast, Mg\(^{2+}\) levels are constantly high inside the cell (∼10\(^{-3}\) M) and barely differ across the plasma membrane. The binding constants of most intracellular Ca\(^{2+}\) binding proteins inside the cells range from 10\(^{-3}\)–10\(^{-7}\) M; thus at physiological concentration, if Mg\(^{2+}\) were to be accommodated, it would always remain bound to these proteins while Ca\(^{2+}\) would be too low for occupancy. During Ca\(^{2+}\)
Figure 1.1. Coordination of Mg$^{2+}$ and Ca$^{2+}$ in Calbindin D$_{9k}$

(A): Octahedral oxygen coordination (red dotted lines) of Mg$^{2+}$ (yellow sphere) in the C-terminal EF-hand of calbindin D$_{9k}$. (PDB ID: 1IG5) (B): Coordination of Ca$^{2+}$ (orange sphere) in the C-terminal EF-hand of calbindin D$_{9k}$ (PDB ID: 4ICB). The essential side chains and backbone are shown as a stick model. The coordination number of Mg$^{2+}$ is lower (6 compared to 7 for Ca$^{2+}$). The bidentate ligand provided by Glu65 (position 12 of EF-Hand loop) in Ca$^{2+}$ bound form is too far from the Mg$^{2+}$ to be used; instead it ligates with an extra water molecule, one of the features that accounts for the lower affinity of Mg$^{2+}$ with the protein.
Figure 1.1

(A)

(B)
signaling, the intracellular Ca\textsuperscript{2+} concentration increases multi-fold, resulting in binding of Ca\textsuperscript{2+} to target proteins to modulate cellular processes. Soon after, the elevated Ca\textsuperscript{2+} is substantially reduced due to the capability of Ca\textsuperscript{2+} to form complexes. It associates with low-molecular weight metabolites, inorganic ions like phosphates, membrane (acidic) phospholipids and other Ca\textsuperscript{2+} transport proteins, making it a preferred on-off regulatory system [2].

1.1.2 Elements of cellular Ca\textsuperscript{2+} signaling mechanism

Ca\textsuperscript{2+} signaling in eukaryotes is triggered by various stimuli, resulting in Ca\textsuperscript{2+} entering the cytoplasm via Ca\textsuperscript{2+} channels and Ca\textsuperscript{2+} released from internal stores. This leads to activation of Ca\textsuperscript{2+}-sensitive processes by Ca\textsuperscript{2+} binding proteins, and finally Ca\textsuperscript{2+} export/removal from cytoplasm to turn off signaling (Figure 1.2) [16].

1.1.2.1 Ca\textsuperscript{2+} increase in cytoplasm

Intracellular Ca\textsuperscript{2+} levels are increased by using both internal and external sources of Ca\textsuperscript{2+}. External stimuli acting through various cell surface receptors like G-protein linked receptors and receptor tyrosine kinases (RTK) generate Ca\textsuperscript{2+}-mobilizing secondary messengers like inositol-1, 4, 5-trisphosphate (IP3), cyclic ADP ribose (cADPR) and
Figure 1.2. Control of cellular Ca^{2+}

Cytoplasmic Ca^{2+} levels are low (~100 nM) in resting cells compared to the extracellular space, which always has excess Ca^{2+} (> 1mM). During Ca^{2+} signaling intracellular Ca^{2+} levels elevate (~ 1-50µM) in response to various intracellular processes. External stimuli activate a variety of membrane receptors including G protein-coupled receptors and receptor tyrosine kinase (RTK) which trigger hydrolysis of phosphatidylinositol-4, 5-bisphosphate (PIP2) by a family of phospholipase C enzymes (PLCβ, PLCγ) to produce Inositol -1, 4, 5-trisphosphate (IP3) and diacylglycerol (DAG). IP3 and cyclic ADP ribose (cADPR) produced from nicotinamide-adenine dinucleotide (NAD) act as ligands for Ca^{2+} releasing receptors IP3R and ryanodine receptor (RYR) of sarco-endoplasmic reticulum (ER/SR) stimulating quick release of Ca^{2+} into cytoplasm. Other Ca^{2+} influx mechanisms included gated entry of Ca^{2+} through plasma membrane channels (VOC, ROC, SOC) as described in section 1.1.2. The intracellular Ca^{2+} pool is regulated by binding to Ca^{2+} sensitive proteins (like calmodulin, calpain, calbindin etc.) which influence many downstream targets; transport to organelles like mitochondria (through the uniporter) and endoplasmic reticulum (through sarco-endo plasmic reticulum Ca^{2+} ATPase, SERCA pumps). Finally, the signal is terminated by exporting elevated Ca^{2+} out of the cell via plasma membrane Ca^{2+} ATPase pumps (PMCA pumps), Na^{+}/Ca^{2+} exchangers (NCX) and the Na^{+}/Ca^{2+}-K^{+} exchangers (NCKX). Adapted and modified from Berridge et.al. [16].
nicotinic acid dinucleotide phosphate (NAADP). These exert their effect on Ca\(^{2+}\) channels such as inositol-1, 4, 5-trisphosphate receptor (IP3R) and ryanodine receptor (RYR) families to release Ca\(^{2+}\) from internal stores held within the endo(sarco)plasmic reticulum (ER/SR). These receptor families exhibit Ca\(^{2+}\)-induced Ca\(^{2+}\) -release (CICR) where Ca\(^{2+}\) elevates the stimulating effect of regulators. The most widely reported Ca\(^{2+}\) influx mechanism from extracellular space is the gated entry of Ca\(^{2+}\) by one of the three classes of plasma membrane Ca\(^{2+}\) channels: a) voltage-operated Ca\(^{2+}\) channels (VOC’s), activated by membrane depolarization and modulated by neurotransmitters, G proteins and diffusible messengers, b) receptor-operated Ca\(^{2+}\) channels (ROC’s), activated by binding of external ligand, c) store-operated channels (SOC’s), activated in response to depletion of internal Ca\(^{2+}\) stores [17].

1.1.2.2 Role of Ca\(^{2+}\) binding proteins

The Ca\(^{2+}\) released into the cytoplasm during the Ca\(^{2+}\) influx does not remain free. Instead, it becomes bound to a wide variety of Ca\(^{2+}\)-binding proteins thereby influencing numerous cellular processes. There are two main classes of Ca\(^{2+}\) binding proteins: a) Proteins that bind or interact with Ca\(^{2+}\) and whose main task is to regulate its concentration. These are also known as Ca\(^{2+}\)-buffering and Ca\(^{2+}\)-transporting proteins, important members of which include calsequestrin (in the endoplasmic reticulum; ER), parvalbumin (cytosol) and some intrinsic membrane proteins functioning as Ca\(^{2+}\)
channels (ATPases). b) Proteins that bind Ca\(^{2+}\) to translate its signal, also called Ca\(^{2+}\) sensors. While these proteins do contribute to the control of concentration, their main task is to transmit suitably processed Ca\(^{2+}\) signals to downstream targets. The classic Ca\(^{2+}\) sensors include calmodulin (CAM, involved in a multitude of processes) and troponin C (TnC, Ca\(^{2+}\) receptor in myofibrils) [1].

Proteins containing EF-hand motifs, where EF corresponds to a Ca\(^{2+}\) -coordinating helix-loop-helix sequence are the most common and best-characterized Ca\(^{2+}\) sensors (Figure 1.3a). The EF-hand motifs have distinct binding affinities for Ca\(^{2+}\) and have so far been found in nearly 600 proteins [18]. The most studied member of the family is calmodulin, a 17 kDa protein (Figure 1.3b) which, upon binding Ca\(^{2+}\), undergoes a variety of conformational changes making it active and ready to bind and modulate many structural proteins [19]. In addition, many Ca\(^{2+}\) binding proteins do not have EF-hand motifs but contain other domains capable of binding Ca\(^{2+}\) with lower affinity. [2] The best characterized of these are the C2 domains which occur in protein-kinase C (PKC), phopholipase A2 (cPLA2) and C (PLC) and synaptotagmins. A C2 domain contains an 8-stranded anti-parallel beta sandwich connected by variable loops containing negatively charged amino acids which are capable of binding 2-3 Ca\(^{2+}\) ions (Figure 1.3c)
**Figure 1.3. EF-Hand and C2 domains- two widespread Ca\(^{2+}\) binding domains**

(A) Single EF-hand motif from the N-terminal domain of calmodulin. The EF-hand motif is a 12-amino-acid loop structure capable of binding Ca\(^{2+}\) in a pentagonal bipyramidal fashion. Ca\(^{2+}\) (indicated by the orange sphere) is coordinated in this instance by seven oxygen atoms, three from the side-chain carboxyl groups of negatively charged amino acids (D20, D22, D24) and one each from a backbone carbonyl group (T26) and water molecule, hydrogen bonded to one of the side chains of the loop (T28). The remaining two coordinations come from a bidentate carboxylate ligand from the EF-loop’s twelfth residue (E31). The chelating residues of the loop are usually notated based on linear position or on the tertiary geometry imposed by their alignment on the axes of a pentagonal bipyramid: 1(\(X\)), 3(\(Y\)), 5(\(Z\)), 7(\(-Y\)), 9(\(-X\)), 12(\(-Z\)) [20, 21].

(B) The two independent domains of calmodulin connected by a flexible linker. Each domain contains a pair of EF hands capable of binding two Ca\(^{2+}\) ions (indicated by orange spheres) for a total of four per calmodulin molecule (PDB ID: 1CLL).

(C) C2 domain of protein kinase C \(\beta\) (PDB ID: 1A25). The scaffold is an antiparallel \(\beta\) sandwich (yellow) with Ca\(^{2+}\) ions (orange spheres) bound in variable loops (green). Alpha-helices are coloured red.
Figure 1.3

(A)

(B)

(C)
and in some cases phospholipids. C2 domains are known to associate with the membranes [22]. Both EF-hand and C2-like domains are integrated into the structure of most human calpains. In addition, calpains also contain two catalytic domains (protease core) capable of binding two Ca\(^{2+}\) ions, making calpains distinctive Ca\(^{2+}\) sensors, expertly integrating several Ca\(^{2+}\) binding motifs to translate Ca\(^{2+}\) signals into proteolysis of target proteins.

1.1.2.3 Ca\(^{2+}\) export

Ca\(^{2+}\) signaling is usually brief and, once finished, elevated Ca\(^{2+}\) is rapidly removed from the cytoplasm by energy-driven pumps and ion exchangers. The plasma membrane Ca\(^{2+}\) pumps (PMCA pumps) use energy derived from ATPases to pump Ca\(^{2+}\) out of the cell. These pumps complement the action of gradient exchangers such as Na\(^+\)/Ca\(^{2+}\) exchangers (NCX’s), which exchange one Ca\(^{2+}\) ion for three Na\(^+\) ions, and of Na\(^+\)/Ca\(^{2+}\)-K\(^+\) exchangers (NCKX), which co-transport one K\(^+\) ion with one Ca\(^{2+}\) ion in exchange for four Na\(^+\) ions. Inside the cell, endo (sarco) plasmic reticulum pumps (SERCA pumps), secretary pathway pumps (SPCA pumps, in Golgi membranes), and the mitochondrial Ca\(^{2+}\) uniporters all return Ca\(^{2+}\) to the internal stores. Ca\(^{2+}\) can also exit mitochondria through a Na\(^+\)/Ca\(^{2+}\) exchanger (MNCX) [2, 23].
1.2 Cysteine proteases

Proteases (also known as peptidases, proteinases or proteolytic enzymes) represent the class of enzymes that catalyze the cleavage of peptide bonds by hydrolysis. Proteases were originally classified into five groups based on the type of residue (nucleophile) that was directly involved in catalysis: a) serine proteases b) threonine proteases c) cysteine proteases d) aspartyl proteases e) metallo proteases [24]. This convention is still in use. However, the most widely used classification system currently followed is that adopted by the MEROPS database (http://www.merops.co.uk), which groups proteases based on similarities in structural organization of the active site, molecular structure, and evolutionary origin [25]. Accordingly, cysteine proteases (CPs), which include a complex set of enzymes containing the activated cysteine as the nucleophile, are further classified into clans (CA, CD, CE etc.) based on evolutionary relationship, and families (C1, C2, C3 etc) based on their homology and sequence similarity. Most prominent among CPs are clans CA (papains) and CD (legumains and caspases). Members belonging to clan CA (consisting of more than 25 families) share a structure and evolutionary history with papain and comprise the catalytic residues in the order cysteine; histidine; aspartine/aspartate [26]. The plant enzyme papain (family C1) and the related mammalian lysosomal cathepsins were the first studied members of the clan CA. They are the closest relatives to calpains,
which are also placed in clan CA, but as a distinct family, C2. Most of the enzymes of the papain family are synthesized as inactive precursors with a signal peptide and a multifunctional propeptide at their N terminus. The propeptide covers the active site and needs to be processed for the proenzyme to become activated and access substrates. [27] Conversely, calpains show a distinct mechanism of activation by Ca$^{2+}$ binding, which makes the enzyme active and competent for proteolytic processing. Other important members of CPs are the caspase (cysteine-dependent aspartate-specific proteases) family of proteases (Clan CD, family C14) that play key roles in several signal transduction pathways.

1.3 Calpains

Calpains were first described independently by two groups in the 1960s when proteolytic activities caused by a “Ca$^{2+}$-activated neutral protease” (CANP) were detected in rat brain [28] and skeletal muscle [29]. The enzyme was later isolated and purified to homogeneity from skeletal muscle [30, 31]. It is now called “calpain”, named for the calcium dependence of the papain-like enzyme [32]. The existence of two main isoforms μ-calpain and m-calpain (calpain 1 and calpain 2, respectively) differing primarily in Ca$^{2+}$ requirements was reported in the early 1980s [33, 34]. Both heterodimeric isoforms share an identical small subunit (30 kDa).
The nucleotide sequence coding for chicken calpain large subunit suggested for the first time the evolution of calpain by fusion of genes for a thiol protease (papain-like) and a Ca\(^{2+}\)-binding domain (calmodulin-like) [35]. This primary structure misled researchers into thinking that protease activity is regulated strictly by Ca\(^{2+}\) binding through calpain’s “calmodulin-like” domain. This view has changed with a) bioinformatic analyses of calpain genes from various organisms over last two decades, and b) structural studies.

The bioinformatics analyses have established that the only conserved feature among all calpains is the cysteine protease core [5, 36]. Indeed, the recent reviews by Sorimachi et al. [36] and Croall et al. [5] have revised the criterion used for classification of calpains. Homology as documented by > 20–25% sequence identity to the protease cores of the classical calpains (calpain 1 and 2) is now used as the decisive factor for designating a protein as a calpain. Using a similar approach, the analysis of genes available from various genome sequencing projects has revealed numerous calpain isoforms in vertebrates, invertebrates, plants and microorganisms [37-40]. So far, 14 calpain-like genes have been identified in humans, 4 in *Drosophila*, 12 in *Caenorhabditis elegans*, 2 in two species of fungi/yeast and 5 in the unicellular organism *Trypanosoma brucei*. Recently, a large family of genes coding for calpain homologs were discovered in
trypanosomes (*Trypanosoma brucei, T. cruzi* and species of *Leishmania*,) and ciliates (up to 26 members in *Tetrahymena*) [41, 42].

To reinforce the point about the primacy and commonality of the protease core, the “calmodulin-like” domain that was thought to be the key to calcium activation of calpain is missing in nearly half of the calpain isoforms in humans and from other organisms including protozoans, plants and fungi. This suggests that acquisition of the EF-hand domain through gene fusion is a relatively late event in calpain evolution.

(b) Structural studies by our lab revealed that the protease core of calpain 1 was enzymatically active as a recombinant protein in the absence of the other domains [4]. This led to the discovery of two new Ca$^{2+}$- binding sites within the protease core that act cooperatively to activate the enzyme [43]. X-ray crystallographic studies also showed that calpain’s “calmodulin-like” domain is in fact not at all like calmodulin but is instead a novel penta-EF-hand domain (PEF) that undergoes relatively small conformational changes on binding calcium [44-46].

**1.3.1 Calpain nomenclature**

Human calpain genes are numbered 1-15 in order of their discovery. In addition, most of the calpain isoforms are also referred to in the literature by their original or common names (Table 1.1). The first studied members of the calpain family, calpains 1
## Table 1-1 Calpain Nomenclature

<table>
<thead>
<tr>
<th>Calpain protein</th>
<th>Gene Name</th>
<th>Common Names</th>
<th>Tissue Distribution pattern</th>
<th>Expression (Highest)</th>
<th>Genbank accession</th>
<th>Chromosome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calpain 1</td>
<td>cspn1</td>
<td>μ-calpain</td>
<td>Ubiquitous</td>
<td>Placenta, Esophagus, Trachea</td>
<td>NM_005186</td>
<td>11q13</td>
</tr>
<tr>
<td>Calpain 2</td>
<td>cspn2</td>
<td>μ-calpain</td>
<td>Ubiquitous</td>
<td>Kidney, Lung, Trachea</td>
<td>NM_001748</td>
<td>1q41</td>
</tr>
<tr>
<td>Calpain 3</td>
<td>cspn3</td>
<td>nCl-1</td>
<td>Tissue-specific</td>
<td>Skeletal muscle (Testis, retina)</td>
<td>NM_000070</td>
<td>15q15.1</td>
</tr>
<tr>
<td>Calpain 5</td>
<td>cspn5</td>
<td>nCl-3</td>
<td>Ubiquitous</td>
<td>Brain, Kidney, Liver</td>
<td>NM_004055</td>
<td>11q14</td>
</tr>
<tr>
<td>Calpain 6</td>
<td>cspn6</td>
<td>CAEN6</td>
<td>Tissue-specific</td>
<td>Placenta</td>
<td>NM_014269</td>
<td>xq23</td>
</tr>
<tr>
<td>Calpain 7</td>
<td>cspn7</td>
<td>CAEN7</td>
<td>Ubiquitous</td>
<td>NA</td>
<td>NM_014296</td>
<td>3p24-p25</td>
</tr>
<tr>
<td>Calpain 8</td>
<td>cspn8</td>
<td>CAEN8</td>
<td>Tissue-specific</td>
<td>Stomach, Digestive tract, Brain</td>
<td>AA_043095</td>
<td>1q41</td>
</tr>
<tr>
<td>Calpain 9</td>
<td>cspn9</td>
<td>CAEN9</td>
<td>Tissue-specific</td>
<td>Digestive tract, heart, Stomach</td>
<td>NM_006615</td>
<td>1q42</td>
</tr>
<tr>
<td>Calpain 10</td>
<td>cspn10</td>
<td>CAPNL0</td>
<td>Ubiquitous</td>
<td>Heart</td>
<td>NM_021251</td>
<td>2q37.3</td>
</tr>
<tr>
<td>Calpain 11</td>
<td>cspn11</td>
<td>CAPNL1</td>
<td>Tissue-specific</td>
<td>Testis</td>
<td>NM_007058</td>
<td>6p12</td>
</tr>
<tr>
<td>Calpain 12</td>
<td>cspn12</td>
<td>CAPNL2</td>
<td>Tissue-specific</td>
<td>Hair follicle</td>
<td>NM_144691</td>
<td>19q13.2</td>
</tr>
<tr>
<td>Calpain 13</td>
<td>cspn13</td>
<td>CAPNL3</td>
<td>Tissue-specific</td>
<td>Lung, Testis</td>
<td>NM_144575</td>
<td>2p22.2-p22.3</td>
</tr>
<tr>
<td>Calpain 14</td>
<td>cspn14</td>
<td>CAPNL4</td>
<td>NA</td>
<td>NA</td>
<td>NR_017746</td>
<td>2p22.2-p22.3</td>
</tr>
<tr>
<td>Calpain 15</td>
<td>cspn15</td>
<td>CAPNL5</td>
<td>NA</td>
<td>NA</td>
<td>NM_005632</td>
<td>16p13.3</td>
</tr>
<tr>
<td>Calpain small subunit 1</td>
<td>cspn1</td>
<td>CAPN4</td>
<td>Ubiquitous</td>
<td>Heart, Pancreas, Kidney</td>
<td>NM_001743</td>
<td>19q13.1</td>
</tr>
<tr>
<td>Calpain small subunit 2</td>
<td>cspn2</td>
<td>CPN2</td>
<td>Tissue-specific</td>
<td>膀胱,食道,前列腺</td>
<td>AC026002</td>
<td>16</td>
</tr>
</tbody>
</table>
and 2 are both ubiquitously expressed whereas some isoforms are tissue specific. The small subunit lacks the protease core domain and is a member of the PEF domain family. It has been misleadingly annotated earlier as CAPN4. It is now designated as CPNS after its gene name representing calpain small subunit [5]. Human genome sequencing also revealed another small subunit variant CPNS2 whose function is not determined.

1.3.2 Structure of calpains

Calpains are complex multi-domain proteins with diverse amino-terminal or carboxy-terminal modules flanking a conserved protease core. Phylogenetic analyses suggest they evolved by a process of fusion, duplication and divergence of multiple distinct genes. Based on domain composition, calpains are grouped into a) classical calpains (also referred to as ‘typical’ or ‘conventional’ calpains) and b) non-classical calpains (also referred to as ‘atypical’ or ‘non-conventional’ calpains). The classical calpains include members of the calpain family that have a penta-EF-hand domain (homologous to domain IV of calpains 1 and 2). The atypical calpains include all other members of the family containing DI-DII but lacking the PEF domain. These might be ancestral calpains as they are predominantly found in lower organisms.

The first and best-studied isoforms are calpains 1 and 2. Of the 14 human calpain isoforms, 8 share a domain structure similar to calpains 1 and 2 as illustrated in Figure
1.4. Other calpain isoforms contain specialized modules like C2-like domains, PBH (PalB homologous domains), SOH (small optic lobe homologous) domains, T domains (tra-3-specific, now called C2 domains) and Zn–finger domains. These additional domains might provide clues to the possible physiological function of the calpains bearing them. Some calpain isoforms have been found lacking one or more of the critical catalytic amino-acid residues, suggesting unique functions unrelated to proteolysis (eg. calpain 6 in mammals and Dm_CalpC in Drosophila). Additional diversity in calpains is sometimes achieved through the presence of insertion sequences as observed in calpain 3 or the formation of splice variants, including some that lack domains outside the protease core.

The classical calpains 1 and 2 are heterodimers of two distinct subunits: an 80-kDa catalytic subunit and a small 30-kDa regulatory subunit (CPNS1) (Figure 1.5.). The large subunit contains four characteristic domains (DI-DIV) preceded by an N-terminal anchor helix (previously referred to as a prosegment) which directly interacts with DVI of CPNS1 and helps provide structural stability between the two subunits in the calcium-free form. The N-terminal sequence is unique and varies in length, it undergoes autolysis and consequent dissociation from CPNS1 following calpain activation. In calpain 1 part of it acts as a mitochondrial targeting sequence [47] guiding calpain 1 localization to the mitochondrial intermembrane space. Some members of the calpain
Figure 1.4. Human Calpain System

Bar diagram showing the linear organization (DI-DIV) of human calpains. The first two domains DI and DII form the protease core and contain catalytic residues cysteine (C), histidine (H), aspartic (N). Domain III is a C2-like domain. Domain IV is a pentahelical domain which dimerizes with homologous small subunit DVI. Calpastatin contains an N-terminal domain DL followed by four repeating domains DI-DIV each of which contain 3 conserved subdomains A, B, C.

Abbreviations used are: NS, N-terminal insertion sequence; IS1 and IS2, Insertion sequences 1 and 2; T, TRA3-specific domain; A^, Anchor sequence, MIT, microtubule interacting and trafficking molecule domain; PBH, PalB homologous domain, Zn, Zinc-finger motif, SOH, SOL homologous domain; G-rich, glycine rich. The length of the calpain constructs (amino acids) is indicated on the right.
Figure 1.4

Calpains 1, 2, 8, 9, 11, 12, 13, 14

Calpain small subunit

Calpain 3

Calpain 5

Calpain 6

Calpain 7

Calpain 10

Calpain 15

Figure 1.4

Calpains 1, 2, 8, 9, 11, 12, 13, 14

Calpain small subunit

Calpain 3

Calpain 5

Calpain 6

Calpain 7

Calpain 10

Calpain 15

Figure 1.4

Calpains 1, 2, 8, 9, 11, 12, 13, 14

Calpain small subunit

Calpain 3

Calpain 5

Calpain 6

Calpain 7

Calpain 10

Calpain 15
Figure 1.5. Three-dimensional structure of inactive m-calpain heterodimer

Ribbon representation of m-calpain (1KFU) structure in the absence of Ca$^{2+}$. The active site cleft exists at the interface of DI and DII. A bar diagram representing the domain organization of m-calpain is also enclosed. The large subunit of m-calpain (80 kDa) is composed of a 19 residue anchor (red), protease domains I and II (blue and cyan, respectively), domain III (green), an ~15 residue transducer (magenta) and domain IV (yellow). The regulatory subunit contains only domain VI (orange). These figures were adapted and modified from Hosfield et al. [48, 49].
family have even more diverse sequences at their N-termini including Zn-fingers (calpain 15), MIT domains (calpain 7) and trans-membrane sequences (Eg. phytocalpains). Domains I and II (also referred to as Ila and IIb)[6] form the catalytic core of the protease. Together they contain a catalytic triad (Cys-His-Asn) similar to the one found in papains and cathepsins. The core domains are larger than those in papain and have additional elements of secondary structure, including the loops that bind the two calcium ions [4, 48, 49]. The active site cleft resides at the interface of these two domains.

Domain III shares no sequence homology with other proteins but structurally resembles C2 domains found in various proteins including protein kinase C and phospholipase C. DIII stabilizes the protease core and is thought to bind Ca\textsuperscript{2+} and phospholipids. Interestingly, no Ca\textsuperscript{2+} was bound in a recent Ca\textsuperscript{2+} bound crystal structure of calpain inhibited by calpastatin [50]. A 16-residue linker joins DIII to DIV. Domain IV is a Ca\textsuperscript{2+}-binding domain containing five EF-hand motifs (PEF domain), the fifth of which does not bind Ca\textsuperscript{2+} but is involved in the dimerization with the small subunit (CPNS1), through the interaction with the equivalent EF-hand motif in domain VI (DVI). The small subunit is composed of two domains (DV and DVI). DV contains clusters of glycine residues and is suggested to direct interaction of calpain with the plasma membrane.
[51-53]. It is unresolved in the crystal structure of human calpain [49]. DVI is homologous to DIV of the large subunit.

1.3.3 Mechanism of calpain activation by Ca\(^{2+}\)

Crystallographic structural information obtained on calpains in recent years reveals their complex regulation of activity involving the interaction of multiple domains. The Ca\(^{2+}\) free heterodimeric structure (apoform) shows a circular arrangement of domains which imposes constraints on the protease core to be held in the open, inactive conformation [48, 49]. A wide open cleft disrupts the catalytic triad, keeping the enzyme in an inactive state. The crystal structure of recombinant mini-calpain 1 (DI-DII) revealed the conformational changes brought about by binding of Ca\(^{2+}\) to two novel Ca\(^{2+}\) binding sites in the protease core [4]. Binding of Ca\(^{2+}\) aligns the active site cleft and converts the core into active enzyme (Figure 1.6). In addition to the core, two other domains, DIV-DVI, also bind Ca\(^{2+}\). This was established for the latter by structural analysis of the recombinant small subunit DVI homodimer in presence and absence of Ca\(^{2+}\) [45, 46]. By inference the same sites bind Ca\(^{2+}\) in DIV as established by EF-hand mutant studies on both domains [54]. Although these domains were originally described
Figure 1.6. Realignment of active site cleft in mini-calpain.

Ribbon representation of the protease core domains (DI-DII) of calpain in absence (top, 1KFU) and presence (bottom, 1KFX) of Ca$^{2+}$. Catalytic triad residues are coloured in green, and the active site tryptophan in magenta. Orange-coloured spheres indicate the positions of the two Ca$^{2+}$ atoms. In the absence of Ca$^{2+}$, the catalytic cysteine and histidine are separated by a distance of 10Å, resulting in an inactive catalytic triad. Upon binding Ca$^{2+}$ movement of domains I and II brings the catalytic triad to the correct spacing for catalysis. These figures were adapted and modified from Strobl et al. [4, 43, 55].
Figure 1.6
as calmodulin-like, the X-ray crystal structures of DVI revealed that they do not undergo major conformational changes upon binding Ca\(^{2+}\).

The complex mechanism of calpain activation can be divided into two major parts. One involves the release of constraints imposed on the protease core by the circular arrangement of the domains. This includes displacement of the N-terminal anchor helix upon Ca\(^{2+}\) binding to DVI. Other events associated with this step are exposure of hydrophobic surface patches (that are calpastatin binding sites) in DIV-DVI following the Ca\(^{2+}\)-induced subtle conformational changes in DIV and DVI. These changes shift the two PEF domains and DIII closer towards the protease core [50, 56]. There are also suggestions of subunit dissociation occurring, which would of course release all constraints on the protease core [57-60]. The second part involves alignment of the active site cleft caused by the cooperative binding of two Ca\(^{2+}\) to DI-DII [61]. It is not clear if these two parts are sequential or simultaneous.

1.3.4 Calpastatin

Calpastatin is the endogenous protein inhibitor of calpain. It is encoded by a single gene that is ubiquitously expressed. It is the only known inhibitor that is completely specific for calpain and it inhibits calpain in a reversible manner. When Ca\(^{2+}\) levels are high the interaction with calpain is very tight, but calpastatin is released when
Ca\(^{2+}\) levels fall. Calpastatin contains five domains, four homologous domains (CAST1–4) each containing approximately 140 amino acids and capable of inhibiting heterodimeric calpains 1 and 2, and a unique N-terminal domain, termed domain L, with no inhibitory activity [62-64]. Calpastatin is capable of simultaneously binding and inhibiting four molecules of calpain [65]. It is intrinsically unstructured and exhibits a complex interaction with calpain. Each inhibitory domain of calpastatin has three subdomains, A, B, and C [66]. Subdomains A and C interact with the two PEF domains of calpain (A binds to DIV and C binds to DVI). Subdomain B binds and occupies the catalytic cleft to inhibit the activity. Neither subdomain A nor C have inhibitory activity without subdomain B, but they increase the overall affinity of the interaction [67-69]. The recent X-ray crystallographic structure of Ca\(^{2+}\)-activated calpain 2 heterodimer bound by one of the four inhibitory domains of calpastatin (CAST4) revealed the novel mechanism of inhibition(Figure 1.7a)[50]. Subdomains C and A bind to the exposed hydrophobic patches on the small subunit (DVI) and DIV as previously described [56]. Subdomain B binds to the activated enzyme on either side of the active site but loops out around active site cysteine thus escaping cleavage in a novel manner (Figure 1.7b). The structure also reveals for the first time the Ca\(^{2+}\)-bound structure of calpain, which is more compact than the apoform. The compactness of the overall structure is caused by
Figure 1.7. Overview of calpastatin domain 4 (CAST4) bound to m-calpain (calpain 2)

(a) Ribbon representation of CAST4 (purple) structure bound to the inactive C105S mutant of m-calpain. CAST4 is unstructured in the absence of calpain; it forms three alpha-helices when in complex with the enzyme. Helices of subdomains A and C bound DIV (yellow) and DVI (orange) respectively, and the helix in subdomain B, associates with the protease core DI-II (blue and cyan, respectively). DIII is coloured green.

(b) Stereo representation of subdomain B of calpastatin (purple) looping away from the mutant active-site serine (yellow) of calpain (shown uncoloured in space-filling mode)[50].
Figure 1.7

(A)

(B)
shifting of domains DIII, DIV and DVI) towards the protease core and facilitates binding of CAST4 over the surface of calpain to make contact with each domain of the enzyme.

1.3.5 Other members of calpain family

1.3.5.1 Calpain 3

Calpain 3 (p94 or CAPN3) was the first and by far the best studied tissue-specific isoforms [70]. It is predominantly expressed in skeletal muscle, where its mRNA is more abundant than those of the classical isoforms (calpains 1 and 2) [71]. The domain composition of calpain 3 is similar to that of the large subunit of the classical calpains, but contains three unique sequences: a 47-residue-long N-terminal extended sequence (NS) which replaces the anchor helix, insertion sequence 1 (IS1, 48 residues) present in domain II acts as an internal propeptide [72], and insertion sequence 2 (IS2, 77 residues) located between domains III and IV, which is thought to interact with titin [73]. CAPN3 is highly unstable during extraction or recombinant expression and undergoes rapid autolysis. The lack of stability of p94 has been attributed to the presence of its unique insertion sequences IS1 and IS2 [74]. Lp82 is a lens-specific variant of p94 in rodents that lacks NS, IS1 and IS2 and has been shown to be enzymatically active and relatively stable [75]. Other CAPN3 splice variants lacking one of these insertion sequences are also
reported [76-78]. CAPN3 plays a crucial role in muscle physiology and has been implicated in cellular processes including myofibrillogenesis and sarcomere remodeling [79, 80]. Inactivating mutations in CAPN3 cause limb girdle muscular dystrophy type 2A (LGMD-2A) [81].

1.3.5.2 Calpain 6

Human Calpain 6 is predominantly expressed in placenta. This calpain isoform is unique because its active site residue cysteine is mutated to a lysine, which presumably makes it catalytically inactive [82]. However, its closest homologue, calpain 5, has a catalytic cysteine and is clearly active in *C. elegans* where it functions in sex determination [83]. Calpain 6 also differs in its domain composition from the large subunit of Calpains 1 and 2 in that the PEF domain IV is replaced by a tra-3-like domain common to Calpain 5 and first described in *C. elegans*. Recent reports on calpain 6 describe its role in microtubule stabilization and tumorigenesis [84, 85].

1.3.5.3 Calpain 8

Human Calpain 8 (nCL-2) is a stomach-specific calpain [86]. The domain structure of Calpain 8 is identical to that of m-calpain. However, it has a splice variant nCL-2’ that lacks domains III and IV. nCL-2 was shown to proteolyze beta-COP, a subunit of the COPI coatomer complex, in the Golgi upon stimulation with Ca²⁺, resulting in the dissociation
of beta-COP from the Golgi and suggesting a novel function for nCL-2 in intracellular membrane trafficking [87].

1.3.5.4 Calpain 9

Human calpain 9 (nCL-4) is mostly expressed in the digestive track. It is a functional protease containing a similar domain composition to the classical isoforms and is suggested to play a tumor-suppressing role [86, 88]. Down-regulation of calpain 9 has been implicated in gastric cancer, hypertensive heart and kidney disease [89, 90].

1.3.6 Calpain in physiology and pathology

The physiological role of calpains remains unclear. They have been shown to be involved in processes such as cleavage of cytoskeletal/membrane attachments that aid in cell motility (e.g. Integrin) [91], proteolytic modification of proteins in signal transduction pathways (e.g. PKC) [92], degradation of enzymes controlling cell cycle progression (e.g. ezrin) [93], regulation of gene expression (c-Jun, p53) [94], substrate degradation in some apoptotic pathways (e.g. procaspase 12) [95], and long-term potentiation of neurons [96]. Most of these roles have been credited to classical calpains. Other tissue-specific calpains are involved in diverse roles as described in previous sections. Also, the discovery of calpain family members (e.g. calpain 6) lacking
catalytic residues suggest that calpains may have functions that do not require proteolytic activity [84].

Pathological conditions have been associated with disturbances of the calpain system. Hyper-activation of calpains as a consequence of uncontrolled influx of Ca\(^{2+}\) into the cell results in excessive proteolysis and consequent tissue damage. Some of the complications associated with altered Ca\(^{2+}\) homeostasis include ischemic injury and Alzheimer’s disease [97, 98]. Two well-known genetic disorders have been linked to tissue-specific calpains. Mutations in the calpain 3 (p94) gene are responsible for limb-girdle muscular dystrophy type 2A [81], and Calpain 10 has been identified as a susceptibility gene for type II diabetes mellitus [99]. Recently, both Calpain 9 and 6 have been linked to tumorigenesis [85, 90]. The role of the calpains in various pathologies has been the subject of several recent reviews [100, 101].

1.3.7 Research Objective

Much of the research on calpains over the last two decades has been concentrated on the wild-type calpain 1 and 2 (m and μ) isoforms. The functional role(s) of other ubiquitous calpain isoforms (like 7 and 9) and tissue-specific isoforms (like 3 and 6) and their response to increased Ca\(^{2+}\) levels has not yet been extensively studied. The key
objective of my thesis research is to increase our understanding of these other human calpain family members. I will employ both conventional biochemical techniques as well as high-throughput structure-based approaches to study individual domains of these isoforms. Determining the structure of these calpain components and their mode of association will increase understanding of the function and mode of regulation of the whole enzymes and their role in health and disease.
Chapter 2
Homodimerization of calpain 3 penta-EF-hand domain

Preface:

This chapter was published in the Biochemical Journal:


Ravikiran Ravulapalli was responsible for cloning, protein expression, purification and all biochemical experiments in the investigation of recombinant calpain 3 domain IV function. The experiments were performed under supervision and guidance of Dr. Beatriz Garcia Diaz and Dr. Peter L. Davies. Modelling studies and their analysis were performed by Dr. Robert L. Campbell. Kim Munro from the Protein Function Discovery (PFD) facility helped with CD spectroscopy and analytical ultracentrifugation. The manuscript was written by Ravikiran Ravulapalli with guidance from Dr. Beatriz Garcia Diaz and Dr. Peter L. Davies and editorial input from Dr. Robert L. Campbell.


2.1 Abstract

Calpains 1 and 2 are heterodimeric proteases in which large (relative molecular mass \( M_r \) 80000) and small (\( M_r \) 28000) subunits are linked through their respective PEF (penta-EF-hand) domains. The skeletal muscle-specific calpain 3 is believed not to form a heterodimer with the small subunit but might homodimerize through its PEF domain. Size-exclusion chromatography and analytical ultracentrifugation of the recombinant PEF domain of calpain 3 show that it forms a stable homodimer that does not dissociate on dilution. Molecular modelling suggests that there would be no barriers to the dimerization of the whole enzyme through the PEF domains. This orientation would place the catalytic centres at opposite ends of the dimer.

2.2 Introduction

Calpains form a widely distributed superfamily of \( \text{Ca}^{2+} \)-dependent cytosolic cysteine proteases. Numerous calpain isoforms have been identified in vertebrates [6, 70, 101], invertebrates [6, 37-39, 101], plants [40, 102] and microorganisms [6, 39, 101]. These enzymes are supposed to play a role in many intracellular processes linked to calcium signaling, including cell motility, apoptosis, cell differentiation and cell-cycle regulation [3, 103, 104]. They function by modulating the biological activities of their substrates through limited proteolysis [91]. Abnormal changes in calpain activity, including those
due to defects in calcium homoeostasis or mutations in calpain genes, contribute to pathologies such as ischemic injury, Alzheimer’s disease and other neurological disorders, muscular dystrophy, cancer and Type II diabetes [100, 101].

The two abundant ubiquitous mammalian calpains 1 and 2 (μ- and m- respectively) are the best-characterized members of the family. They are heterodimers of a large (relative molecular mass $M_r$ 80000) and a small ($M_r$ 28000) subunit. The $M_r$ 80000 catalytic subunit has four structural domains (I–IV) [48, 49]. The first two domains (I and II) make up the papain-like catalytic core common to all calpains, whereas domains III and IV are the C2-like and PEF (penta-EF-hand) domains respectively [6]. The small subunit contains two domains (V and VI). Its PEF domain (VI) makes extensive contact with the homologous PEF DIV (domain IV) of the large subunit through pairing of the fifth EF hand [44]. Both isoforms are catalytically inactive in the absence of Ca$^{2+}$. They undergo a series of conformational changes on Ca$^{2+}$ binding to several domains [4, 105, 106]. In addition to these widely distributed μ- and m-isoforms, a dozen more distinctive isoforms have been identified in humans. Although the proteolytic core domains (I and II) are conserved in all isoforms of the calpain family, the other domains can be quite varied. Some homologues seem to lack the small subunit (calpains 3, 5, 6, 7 and 10), suggesting that they may not form heterodimers.
Calpain 3 (p94), the first identified ‘tissue-specific’ isoform, is produced predominantly in skeletal muscles [70]. p94 is implicated in myofibrillogenesis and sarcomere remodelling. Mutations in p94 are genetically linked to limb girdle muscular dystrophy type 2A, signifying a crucial role for p94 in muscle physiology [79, 81]. It has a domain organization similar to that of the μ- and m-calpain large subunits, and its amino acid sequence is 54 and 51% identical with their respective large subunits [70]. p94 is $M_r$ 14000 larger than the μ- and m-calpain large subunits. This is due to the presence of three additional unique sequences; an N-terminal extension sequence of 47 amino acids (referred to as NS) that replaces the anchor helix of m-calpain, and two insertion sequences of 48 and 77 amino acids (referred to as IS1 and IS2 respectively). IS1 is located in domain II as an insert that flanks the active site cleft and acts as an internal propeptide [72]; IS2 is located between domains III and IV and binds titin [73].

Attempts to isolate p94 from muscles or to produce it using recombinant systems have been hampered by its rapid autolysis [74, 107, 108]. Owing to this instability, the physiological role of p94 and the structure–function aspects of its different domains and additional unique sequences remain obscure. When inactive [C129S (Cys129→Ser)] rat p94 was expressed in COS cells, it was stable, and during size-exclusion chromatography, it eluted with an $M_r$ $\sim$180000, as though it were a dimer [109]. However, there have been no definitive studies performed on the issue of the
oligomeric state of p94 in vitro or in vivo. Also, we note that PEF proteins do not naturally occur as monomers [110], suggesting that the PEF domain DIV could be the logical site for dimerization. In the present study, we describe experiments that establish the intrinsic capability of the p94 PEF domain to homodimerize, and we show by molecular modelling that there are no obvious barriers to dimerization of the whole enzyme through this interaction.

2.3 Materials and Methods

2.3.1 Cloning of p94DIV

PCR primers (5'-TATACATATGGGCGCTCGATCAGGAAAGTG-3' and 5'-CACCTCGAGGGC-ATACATGGTGAGCTGCAGCCACTC-3') were designed to generate the cDNA fragment encoding the DIV region of p94 by PCR amplification of reverse transcripts (reverse transcriptase–PCR) of total RNA from human skeletal muscles (Stratagene, La Jolla, CA, U.S.A.) using an RT-Thermoscript kit (Invitrogen, Carlsbad, CA, U.S.A.) and Expand high-fidelity DNA polymerase (Roche, Indianapolis, IN, U.S.A.). The amplified product was cloned into the pCR 2.1-TOPO vector (Invitrogen). DNA sequences of positive clones were verified. The verified cDNA sequence was ligated into the Ndel and Xhol sites of the pET24a expression vector (Novagen, Madison, WI, U.S.A.)
to incorporate an N-terminal Met codon and a C-terminal histidine tag into the p94DIV construct (Figure 2.1A).

2.3.2 Protein expression and purification

The plasmid encoding p94DIV was transfected into *Escherichia coli* BL21 (DE3) (Novagen) grown in LB (Luria–Bertani) medium under kanamycin selection to an absorbance $A_{600}$ of 0.8–1.0 at 37°C. Protein expression was induced with 0.4 mM isopropyl β-D-thiogalactoside for 3 h. Cells harvested from the culture by centrifugation were resuspended in lysis buffer (25 mM Tris/HCl, pH 7.6, 5 mM EDTA, 5%, v/v, glycerol, 10 mM 2-mercaptoethanol and 0.1 mM PMSF) and lysed by sonication. The lysate was centrifuged at 27000 $g$ for 45 min and the supernatant was applied to a DEAE-Sepharose column. Protein was eluted by a 0–0.75 M NaCl salt gradient in lysis buffer, and fractions containing p94DIV were identified by the presence of a prominent $M_r$ 22000 band on SDS/PAGE. These fractions were pooled and applied to a 10 ml NTA (Ni$^{2+}$-nitrilotriacetate–agarose) column (Qiagen, Chatsworth, CA, U.S.A.). The column was washed with N-buffer (50 mM Tris/HCl, pH 7.6, 100 mM NaCl, 5 mM imidazole and 0.01% sodium azide) and eluted with N-buffer and 250 mM imidazole. The fractions containing p94DIV were further purified by gel-permeation chromatography.
2.3.3 Size-exclusion chromatography

A Sephadex G-75 column (2.6 cm × 100 cm; Amersham Biosciences) was equilibrated overnight with 25 mM Tris/HCl (pH 7.6), 150 mM NaCl, 5 mM EDTA, 2% glycerol and 0.05% sodium azide at a flow rate of 0.5 ml/min. The column was then calibrated with the following protein standards: Blue Dextran (\(M_r 2000000\)), BSA (\(M_r 66000\)), ovalbumin (\(M_r 44000\)), carbonic anhydrase (\(M_r 29000\)), myoglobin (\(M_r 17000\)) and cytochrome c (\(M_r 12400\)), with inorganic phosphate to mark the total volume. The protein sample containing p94DIV was loaded on to the column and eluted at 0.5 ml/min. All chromatographic runs were performed at 4°C. Peak volumes are standardized to a \(K_D\) (diffusion constant) value \[111\] given by \(K_D = (V_e - V_o)/(V_t - V_o)\), where \(V_e\) is the elution volume of the centre of a protein peak, \(V_o\) the void volume (determined by elution of Blue Dextran) and \(V_t\) the total elution volume (determined by elution of inorganic phosphate). The absorbance of each fraction was measured at 280 nm, and those in which Blue Dextran and myoglobin eluted were measured at 256 and 415 nm respectively. The presence of phosphate in the eluate was detected by a colorimetric assay \[112\].

Purified p94DIV was concentrated using a Centricon \(M_r 10000\) concentrator (Millipore, Bedford, MA, U.S.A.) to a final protein concentration of 25 mg/ml in 10 mM
Hepes (pH 7.5) and 10 mM DTT (dithiothreitol), then divided into 400 µl aliquots and flash-frozen in liquid nitrogen for storage.

2.3.4 CD spectroscopy

CD spectra were recorded using an Olis RSM 1000 CD spectrophotometer (Bogart, GA, U.S.A.). The CD spectra spanning wavelengths 260–180 nm were collected using a 0.1 mm path length quartz cuvette at 4°C. p94DIV was dialysed into 5 mM Tris-HCl (pH 7.6), 0.1 mM DTT and 0.2 mM EDTA. The protein concentration was 2.2 mg/ml. Five scans were recorded, averaged and corrected for buffer background signal. Deconvolution of the data was performed using CDNN software [113]. Results are expressed in terms of mean residue ellipticity [θ] (deg·cm²·dmol⁻¹).

2.3.5 Intrinsic tryptophan fluorescence studies

Intrinsic tryptophan fluorescence measurements were performed using a PerkinElmer LS50B fluorescence spectrophotometer equipped with a stirrer-adapted 4 ml cuvette (Hellma, Concord, Ontario, Canada) at a constant temperature of 24°C. Excitation and emission wavelengths were set at 280 and 340 nm respectively. The reaction buffer was 50 mM Tris-HCl (pH 7.6) containing 100 mM NaCl, 1 mM EDTA and 1 mM DTT. p94DIV (2
\( \mu M \) was added to the assay buffer and intrinsic fluorescence was recorded for 10 min. At this point, 50 mM CaCl\(_2\) dissolved in the assay buffer was pumped continuously into the cuvette at 4 \( \mu l/min \) using an injector pump (Harvard Apparatus pump 22). Similar experiments were performed under the same conditions but using MgCl\(_2\) instead of CaCl\(_2\). In a control experiment, intrinsic tryptophan fluorescence changes were monitored during CaCl\(_2\) addition to purified rat \( \mu l-II \) (1 \( \mu M \)) [4] by the method described above for p94DIV.

### 2.3.6 Analytical ultracentrifugation

Ultracentrifugation experiments were performed on p94DIV at 20\( ^\circ \)C in a Beckman (Palo Alto, CA, U.S.A.) Optima XL-I analytical ultracentrifuge equipped with automated scanning absorption optics (at 280 nm) by the method of Laue and Stafford [114]. Protein solutions were exhaustively dialysed against 20 mM Pipes (pH 6.5), containing 150 mM NaCl, 0.1 mM DTT and either 1 mM EDTA or 2 mM CaCl\(_2\). The rotor used was a Beckman An60-Ti. Sedimentation velocity runs were performed at two different speeds, 30000 and 42000 rev./min, for 7 and 4 h duration respectively and a final protein concentration of 0.25 mg/ml both in the absence and presence of Ca\(^{2+}\) (2 mM). The results were analysed by SEDFIT software [115]. Sedimentation equilibrium experiments were performed at three different speeds, 12000, 16000 and 20000 rev./min, and at six
different protein concentrations (0.05, 0.1, 0.2, 0.3, 0.4 and 0.5 mg/ml) both in the presence and absence of Ca^{2+} (2 mM). Each speed was maintained until there was no significant difference in $r^2/2$ versus absorbance. After 24 h of equilibration at each speed, a scan was taken at each hour for the next 5 h (total five scans). The results were analysed using Origin 4.1 software (Beckman).

2.3.7 Homology modelling of the p94DIV dimer

Modelling of the p94DIV dimer was performed using PyMOL, MODELLER and GROMACS [116-119]. A total of 100 models of p94DIV were created with the program MODELLER by mapping DIV (Glu^{649}-Ala^{821}) on to the crystal structure of the rat calpain small subunit DVI homodimer (Glu^{1}-Ser^{173}) [46] to verify that there was sufficient sequence identity to form a dimer. The structures were examined with PyMOL to detect clashes at the interface, and the complementarity of the models was measured with the program sc from the CCP4 program suite [120]. The model with the highest complementarity was subjected to energy minimization for 10000 steps or until convergence (tolerance, 1000 kJ · mol^{-1} · nm^{-1}) and molecular dynamics (3 ns or 30000 steps, at 300 K, with a constant temperature bath) using GROMACS. Coordinates for the p94DIV dimer have been deposited in the Protein Data Bank (PDB ID: 1Y9V). A second model was generated to examine the possibility of heterodimerization of p94DIV with DVI. In this case, the
sequence of p94DIV was mapped on to one subunit of the rat calpain DIV homodimer, whereas the sequence of the human DVI was mapped on to the other. As a control, the process of energy minimization and molecular dynamics was repeated for the crystal structure of the rat DVI homodimer (PDB ID: 1DVI) and for the DIV–DVI heterodimer (PDB ID: 1KFU) from the Ca\(^{2+}\)-free human calpain crystal structure. The trajectories from the molecular dynamics simulations were analysed for energetic stability and complementarity between the subunits.

2.3.8 Graphic model of the m-calpain large-subunit dimer

A homodimer of two m-calpain large subunits was generated by superimposing DIV of one m-calpain heterodimer on to the small subunit (DVI not shown) of a second m-calpain heterodimer (Figure 2.6). This computer-graphics model of a full-length m-calpain large-subunit dimer associated through their PEF domains was constructed using PyMOL. The structure was visually examined for structural clashes but no other manipulations were performed.
2.4 Results

2.4.1 Expression and purification of recombinant human p94 domain IV

The DNA sequence encoding p94DIV is shown in Figure 2.1(A) above the amino acid sequence. The predicted protein has 189 residues, including the His6 C-terminal tag, with a theoretical pl of 5.92 and a calculated $M_r$ of 22181. The amino acid sequence is approx. 52% identical with DIV of μ-calpain and approx. 47% identical with DIV of m-calpain. p94DIV was produced as the major soluble product in *E. coli* supernatant (Figure 2.1B, lane 2) and was purified using a combination of three column separations, DEAE-Sepharose, Ni-NTA and Sephadex G-75. The pooled fractions from the DEAE-Sepharose column contained several higher $M_r$ contaminating *E. coli* proteins, which were largely removed in the subsequent Ni-NTA affinity chromatography step. The final purification product from Sephadex G-75 was estimated to be more than 95% pure, as indicated by the SDS/PAGE (Figure 2.1B, lane 6). Soluble p94DIV was stable throughout the purification and concentration steps. A final yield of approx. 250 mg of protein was obtained from a 4 litre culture.
Figure 2.1. Human calpain p94 domain IV and its purification

(A) The cDNA sequence of domain IV of human calpain p94 is shown above the encoded amino acid sequence. (B) SDS/PAGE of fractions obtained during purification. Lane 1 shows the molecular-mass standards indicated at the side of the gel. Lanes 2 and 3 were loaded with an aliquot of *E. coli* supernatant and pellet respectively. Lanes 4–6 represent pooled fractions from purifications steps on DEAE-Sephrose, Ni-NTA and Sephadex G-75 columns respectively. The amount of protein in lane 5 is twice that loaded in lane 6.
Figure 2.1

A

\[ \text{NdelI} \]

\[
\begin{array}{c}
\text{1} \\
\text{101}
\end{array}
\]

ATGGCCATCTCTGATCAAGAAATGAGAAACAGCAAAATCTCCTGGAAACATTTTCAACAG

\[
\begin{array}{c}
\text{21}
\end{array}
\]

IAAGDDDMHEICABDELKHKVLYIV

\[
\begin{array}{c}
\text{121}
\end{array}
\]

GTGCAACGACCAAAAGGAACTGACAGAAGCAGACTGCTCTCTC

\[
\begin{array}{c}
\text{41}
\end{array}
\]

VHKKHDKDLKINGFEGIFLESCHRSM

\[
\begin{array}{c}
\text{101}
\end{array}
\]

ATGGCCATCTCTGATCAAGATGCTGAGAAGCAGACAGACTCTGCAAGAGATCCACACAGCTC

\[
\begin{array}{c}
\text{61}
\end{array}
\]

IALNDTDGSGSKLHIQEFHHL

\[
\begin{array}{c}
\text{241}
\end{array}
\]

TGGAACAGATAAGGCTGGTCAAGAATTTTCGAACAGCAAATGACAAGACAGACTCCAGC

\[
\begin{array}{c}
\text{81}
\end{array}
\]

WHKIRAWQKIFRHYDTQSG

\[
\begin{array}{c}
\text{301}
\end{array}
\]

ACCTCAAGCAGAGTGGATGCAAGATGCAAGACGCTCAGAGATACCCCTCAACGAC

\[
\begin{array}{c}
\text{101}
\end{array}
\]

TINSYEHNANVNDAGFHHLN

\[
\begin{array}{c}
\text{361}
\end{array}
\]

CAAGCTCAAGATGATCTCTGATCAAGCTGGCTGACAGACAGACAGACCCTCTTAC

\[
\begin{array}{c}
\text{121}
\end{array}
\]

QLYDIRITMRYADHRHMNIDFD

\[
\begin{array}{c}
\text{421}
\end{array}
\]

AGTTCTCATCTCTCTGGTATGGCAAGAGCAATGGTACAGCTCCTCTAACATGCTAC

\[
\begin{array}{c}
\text{141}
\end{array}
\]

SFCICCFVRLLEGFMFRARFHAED

\[
\begin{array}{c}
\text{481}
\end{array}
\]

AAGGAAGATGGATGATGATGCTATCATCAAGAGCTGCTCTCTGAGAGAACCTCACTGCTATAT

\[
\begin{array}{c}
\text{161}
\end{array}
\]

KDGDSGIIKLMHVLWQLLTMY

\[
\begin{array}{c}
\text{541}
\end{array}
\]

GGCTCAGACACCAACACCACACACACAC

\[
\begin{array}{c}
\text{181}
\end{array}
\]

ALEEHHHHHH

B

\[
\begin{array}{c}
\text{kDa} \\
66.4 \\
55.6 \\
42.7 \\
36.5 \\
26.6 \\
22k-> \\
20.0 \\
14.3
\end{array}
\]

\[
\begin{array}{cccccccc}
1 & 2 & 3 & 4 & 5 & 6 \\
\end{array}
\]
2.4.2 CD spectra and intrinsic tryptophan fluorescence studies

To assess that recombinant p94DIV is indeed properly folded; we determined its secondary structure by CD (Figure 2.2A). The CD spectra of p94DIV are those of a predominantly α-helical protein. Deconvolution of the CD spectra showed a high α-helical content (∼59%) and a very low percentage of random coil. A comparison of the secondary structure profile of p94DIV with that of the homologous rat small subunit determined from its crystal structure (PDB ID: 1DVI) shows that the two proteins share high structural similarity (Table 2.1). There are only a few percentage points difference in the abundance of each secondary structure type. The same is true for the model of p94DIV. These results strongly suggest that p94DIV was produced as a well-folded protein in E. coli.

To study the effect of Ca^{2+} on the conformation of p94DIV, we monitored its intrinsic tryptophan fluorescence in the presence and absence of this ion. The p94DIV monomer has three tryptophan residues and five putative EF-hand Ca^{2+}-binding sites. p94DIV displayed a significant shift in fluorescence intensity on the addition of Ca^{2+} (Figure 2.2B), which is indicative of a conformational change in the protein caused by binding Ca^{2+}. Under the same conditions, Mg^{2+} did not produce a change in intrinsic tryptophan fluorescence. For a positive control, we used the proteolytic core of μ-
Table 2-1 Comparison of secondary structure content of p94DIV, based on CD spectra or modelling, with that of the rat Mr 21000 small-subunit dimer obtained from its crystallographic structure (PDB ID: 1DVI)
### Table 2-1

<table>
<thead>
<tr>
<th></th>
<th>Rat small subunit (%)</th>
<th>P94DIV (%)</th>
<th>P94DIV model (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Helix</td>
<td>63</td>
<td>59</td>
<td>64</td>
</tr>
<tr>
<td>Random coil</td>
<td>17</td>
<td>16</td>
<td>17</td>
</tr>
<tr>
<td>β-Strand</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>β-Turn</td>
<td>14</td>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td>Total</td>
<td>99</td>
<td>93</td>
<td>100</td>
</tr>
</tbody>
</table>
Figure 2.2. CD spectra and intrinsic tryptophan studies on p94DIV

(A) Plot of the molar ellipticity $\theta$ as a function of wavelength. Spectra of p94DIV were collected over the far- and near-UV range (180–260 nm). (B) $\text{Ca}^{2+}$-induced conformational change in p94DIV (black) and $\mu$-II (light grey). Intrinsic fluorescence intensity was monitored at 340 nm ($I_{340}$) by exciting p94DIV and $\mu$-II at 280 nm while continuously adding CaCl$_2$. Under similar conditions, no change in protein conformation was observed when MgCl$_2$ was added to p94DIV (dark grey).
Figure 2.2

A

$[\theta] \times 10^{-5}$ (degree cm$^2$ dmol$^{-1}$)

wavelength (nm)

B

$[\theta]_{340\text{nm}}$

time (sec)

$\text{Ca}^{2+} / \text{Mg}^{2+}$
2.4.3 P94DIV is a stable dimer in solution

The fraction obtained after Ni-NTA purification was chromatographed on Sephadex G-75, both as a purification step and to determine its $M_r$. A standard curve was generated by plotting the $K_D$ values for proteins of known molecular masses against the logarithm of their molecular masses (Figure 2.3). Recombinant p94DIV eluted immediately after ovalbumin as a single macromolecular species with an $M_r \sim 40000\pm 5000$, estimated by interpolation of its $K_D$ value on to the standard curve. This is twice the expected molecular mass ($M_r$ 22181) of p94DIV. These results suggest that p94DIV forms a homodimer in solution. To confirm this and obtain quantitative information on the strength of dimerization, ultracentrifugation analysis was performed.

Sedimentation velocity analysis of recombinant p94DIV showed that the sample solution was homogeneous both in the presence and absence of 2 mM CaCl$_2$. A value of $s_{20, w}^0 \sim 3.41$ was obtained, which in this case corresponds to a $M_r$ of approx. 44000, assuming the protein to be globular in shape. From sedimentation equilibrium studies, a $M_r$ of 43580 was obtained, which is consistent with the expected theoretical value for the p94DIV dimer (44360). The dependence of $M_r$ on the ultracentrifuge cell-loading concentration $c$ is shown in Figure 2.4 for the entire range of protein concentrations used (approx. 0.05–0.6 mg/ml). $M_r$ is constant as a function of $c$, indicating that the sample behaves as an ideal single component. No evidence for dissociation of the
Figure 2.3. Size-exclusion (Sephadex G-75) chromatography standard curve

$K_D$ values obtained from peak volumes of different protein standards are plotted against $\log M_r$. The points a–e correspond to the protein standards used: (a) BSA ($M_r$ 66000), (b) ovalbumin ($M_r$ 44000), (c) carbonic anhydrase ($M_r$ 29000), (d) myoglobin ($M_r$ 17000) and (e) cytochrome c ($M_r$ 12400). Point x represents the intersection of the $K_D$ value of p94IV obtained from its elution volume with the standard curve giving a $\log M_r$ value. The $M_r$ was then calculated to be approx. 40000 with an error value of $\pm$ 5000.
Figure 2.3
The $M_r$ is plotted against the cell loading concentration $c$ (mg/ml). The graph shows that the $M_r$ of p94DIV is constant as a function of concentration ($c$), indicating an ideal single component in sample. Graphical data are consistent with the expected molecular mass range for the p94DIV domain dimer (~43500) throughout the concentration range in the absence of $\text{Ca}^{2+}$. Similar values were obtained in the presence of 2 mM $\text{Ca}^{2+}$.
Figure 2.4
homodimer (at concentrations of 0.05, 0.1 and 0.2 mg/ml) was observed either in the presence or absence of Ca2+, even at the lowest concentrations tested, suggesting strong dimer interaction.

2.4.4 Homology modelling of p94DIV

Mapping of the sequence of p94DIV on to the rat calpain DVI homodimer showed that the dimer interface is well conserved (Figure 2.5). Energy minimization and molecular dynamics did not show any evidence that the p94DVI homodimer is unstable. Calculation of the average complementarity from every 100th structure in the dynamics trajectories for the human m-calpain DIV–DVI heterodimer, the rat DVI homodimer and the p94DIV homodimer models gave values of 0.76, 0.74 and 0.73 respectively. Average complementarity of the p94DIV homodimer lies within the minimum and maximum values calculated for the trajectory for the human m-calpain DIV–DVI heterodimer (0.70–0.80 respectively). This suggests that the p94DIV is capable of dimerization. When a model of the p94DIV-DVI heterodimer was generated, there were no obvious steric clashes or charge repulsions that would prevent heterodimerization and the average complementarity value 0.72 was comparable with the others.
Figure 2.5. Homology model of p94DIV

(A) Molecular surface of the ‘A’ chain at the dimer interface between the two monomers of the rat DVI dimer. Residues that are identical between rat DVI and p94DIV are coloured white, while those that differ are coloured black. (B) Alpha-carbon trace of the superimposed structures of rat DVI dimer (black) and the p94DIV dimer (light grey) created with the modeller program. The view in this Figure is rotated 90° about the vertical axis from that in (A). The ‘A’ chain of each dimer is drawn in thick tubes, while the ‘B’ chain is drawn in thin tubes. The spheres represent Ca$^{2+}$ ions.
Figure 2.5

A.

B.
Figure 2.6. Graphic model of the large subunit dimer

(A) Schematic representation of the large subunit domain structures of m-calpain and p94. The domain organization of the two proteins is similar, but p94 has three additional unique sequences; an N-terminal (NS) extension sequence and two insertion sequences (IS1 and IS2). (B) Homodimer of two m-calpain large subunits generated by superimposing domain VI of one m-calpain heterodimer on to domain IV of the other (domain VI not shown). They associate with each other through the PEF domain IV. The locations of IS1 and IS2 and the NS sequence in p94 (coloured purple and cyan respectively) are indicated by thick arrows. The inset to (B) shows a side view (rotated 90° about the vertical axis).
Figure 2.6
2.5 Discussion

The premise for our study of p94DIV dimerization stems from the circumstantial evidence that p94 might associate with itself rather than with the small subunit ($M_r$ 28000) found in calpains 1 and 2. This small subunit has not been seen in preparations of p94 from skeletal muscles [109], even though muscle cells do produce the small-subunit-containing calpains [71]. Also, yeast two-hybrid assays on p94 (both inactive and active) failed to show binding of the small subunit to p94 [73, 121]. Evidence for homodimerization comes mainly from reports that an active-site mutant of p94 (C129S) elutes as a $M_r$ 180000 protein during size-exclusion chromatography [109]. This is consistent with homodimer formation, although an asymmetrical p94 monomer could also produce the same result. In a comparable system, the homologous large subunits of chicken calpain ($M_r$ 80000) have been reported to form homodimers when expressed alone in SF9 cells, but form heterodimers when co-expressed with the $M_r$ 28000 small subunit [109, 122]. Most PEF proteins are known to form either heterodimers or homodimers. This has been established by biochemical analyses, co-immunoprecipitation and yeast two-hybrid assays on PEF proteins such as calpain, ALG-2, grancalcin, sorcin and peflin [110]. This observation that PEF domains are typically paired as heterodimers [48, 49, 101, 123] or homodimers [124-126] argues for p94
homodimerization and points to the PEF domain as being the probable homodimerization interface.

p94DIV was massively produced in E. coli and gave yields that were comparable with those obtained with the homodimer of the rat calpain small subunit [126]. It accumulated in the soluble fraction and behaved during column chromatography as a single entity with a sharp elution profile, consistent with it being a stable, well-folded protein. This was confirmed by analytical ultracentrifugation and CD spectroscopy studies. On Ca\(^{2+}\) binding, it underwent a conformational change that was detected by an increase in intrinsic tryptophan fluorescence. Evidence for p94DIV being a homodimer comes from both size-exclusion chromatography and analytical ultracentrifugation studies. The $M_r$ values of the p94DIV domain deduced from size-exclusion chromatography (40000±5000) and analytical ultracentrifugation studies (43500) are consistent with a calculated theoretical dimer molecular mass of 44362. The value determined by size-exclusion chromatography is also deemed to be reliable because crystallography of calpain PEF domain homo- and heterodimers shows them to be roughly globular in shape (47 Å × 48 Å × 54 Å; 1 Å=10^{-10} m) both with and without Ca2+ present [45, 46]. The dimer molecular mass calculated from analytical ultracentrifugation ($M_r$ 43500) is constant over a 10-fold concentration range, suggesting that the dimer is very stable.
To investigate further the implications of our biophysical studies described above, we performed some modelling studies. Initially, we mapped the p94DIV sequence on to the rat small subunit (DIV) and observed that the dimer interface is well conserved. We generated a homology model of p94DIV and performed energy minimization and molecular dynamics studies. These showed that p94DIV is capable of forming a homodimer through the same interface as the rat DVI homodimer and the human m-calpain heterodimer (Figure 2.5).

Given the known propensity of PEF domain proteins to form dimers, we explored the possibility that full-length p94 could dimerize through the pairing of its PEF domain. In the absence of a p94 structure, we made a computer-graphics model of an m-calpain large-subunit homodimer (Figure 2.6). The m-calpain large subunit has a domain composition similar to that of p94 but lacks the three unique additional sequences (NS1, IS1 and IS2). The regions of the human m-calpain that are replaced by these additional sequences (NS, IS1 and IS2) in p94 are highlighted in Figure 2.6. Since these additional sequences have no known structural homologues, we were not able to model their structures with confidence. Nevertheless, we can draw some conclusions from this mock-up. The IS1 sequence is remote from the dimerization contact and is unlikely to be affected by dimer formation. The NS sequence of p94 replaces the anchor helix of m-calpain. Although it appears to be in close proximity to the DIV homodimer, its sequence
is rich in proline and glycine residues and shows no homology to the anchor helix of m-calpain. Therefore we predict that it will be intrinsically disordered and will not occupy the same space as the m-calpain anchor helix. Thus it is unlikely to interfere with dimer association. The IS2 sequence is present in the m-calpain ‘transducer’ region, which lies away from the dimer interface. Although we have no structural information about IS2, there is nothing in our model that suggests that it should interfere with homodimerization of p94. Also, since IS2 is the binding site for titin, we assume that it would project outwards to favour titin association. In our model, the catalytic cores are orientated in opposite directions (at the two extreme ends), and they are unlikely to autoproteolyse each other. Dimerization removes the fifth EF-hand of DIV as a potential binding site for the interactions of p94 with other proteins. However, dimerization means that all other possible contact sites are duplicated, which could result in stronger binding or simultaneous binding to two ligands (e.g. binding sites on titin) [73].

Homodimerization in this manner might apply to all of the single-chain calpains that have a PEF domain, such as the Drosophila calpains A–C. Furthermore, it should not be assumed that other PEF-containing mammalian calpains (e.g. 8, 11, 12 and 13) will necessarily form a heterodimer with the small subunit in preference to a homodimer, until this is proven by experimentation as with calpains 1, 2 and 9 [127].
Chapter 3

Distinguishing between calpain heterodimerization and homodimerization

Preface:

This chapter was published in FEBS Journal:

Ravikiran Ravulapalli, Robert L. Campbell, Sherry Y. Gauthier, Sirano Dhe-Paganon, Peter L. Davies (2009)

Distinguishing between calpain heterodimerization and homodimerization. FEBS J 276, 973 - 982

Ravikiran Ravulapalli was responsible for the high-throughput cloning, protein expression, purification and all biochemical experiments in the investigation of the dimerization of calpain isoforms. The experiments were performed under supervision and guidance of Dr. Peter L. Davies (Queen’s University) and Dr. Sirano Dhe-Paganon (University of Toronto). Sherry Y. Gauthier was responsible for the cloning and expression of type III antifreeze protein and the small subunit fusion construct, ice affinity experiments and immunoblotting. Modelling studies and analysis were performed by Dr. Robert L. Campbell. The manuscript was written by Ravikiran Ravulapalli with guidance from Dr. Peter L. Davies and editorial input from Sherry Y. Gauthier and Dr. Robert L. Campbell.
3.1 Abstract

The two main mammalian calpains, 1 and 2, are heterodimers of a large 80 kDa and a small 28 kDa subunit that together bind multiple calcium ions during enzyme activation. The main contact between the two subunits of these intracellular cysteine proteases is through a pairing of the 5th EF-hand of their C-terminal penta-EF-hand (PEF) domains. From modelling studies and observation of crystal structures it is not obvious why these calpains form heterodimers with the small subunit rather than homodimers of the large subunit as suggested for calpain 3 (p94). Therefore, we have used a differential tagging system to determine which of the other PEF domain-containing calpains form heterodimers and which form homodimers. His6-tagged PEF domains of calpains 1, 3, 9 and 13 were co-expressed with the PEF domain of the small subunit that had been tagged with an antifreeze protein. As predicted, the PEF domain of calpain 1 heterodimerized and that of calpain 3 formed a homodimer. The PEF domain of digestive tract-specific calpain 9 heterodimerized with the small subunit and that of calpain 13, prevalent in lung and testis, was mainly found as a homodimer with a small amount of heterodimer. These results indicate whether recombinant production of a particular calpain requires co-expression of the small subunit, and whether this calpain is likely to be active in a small subunit knock-out mouse. Furthermore, as the endogenous inhibitor calpastatin binds to PEF domains on the large and small subunit, it
is less likely that the homodimeric calpains 3 and 13 with two active sites will bind or be silenced by calpastatin.

3.2 Introduction

Calpains are a family of intracellular cysteine proteases. They are Ca$^{2+}$-dependent and function by modulating the biological activities of target proteins through selective cleavage[101]. Genome sequencing projects have revealed numerous calpain isoforms in vertebrates, invertebrates, plants, microorganisms and recently in kinetoplastid parasites [37-41, 101, 128]. In the human genome, fourteen different calpain isoforms have been identified to date. Several calpain isoforms are ubiquitously expressed, whereas many demonstrate tissue-specific expression patterns [71]. Although their precise functions are poorly understood, calpains are implicated in many intracellular processes linked to calcium signaling such as cell motility, apoptosis, and cell cycle regulation, as well as cell-type specific functions like cell fusion in myoblasts and long-term potentiation in neurons [91, 93, 96, 129]. A number of pathologic conditions (ischemic injury, Alzheimer’s disease, limb-girdle muscular dystrophy 2A, type II diabetes mellitus, gastric cancer etc.) have been associated with disturbances of the calpain
system [81, 90, 97, 98, 130, 131]. Therefore, elucidating the specific role of calpains in these pathologies may facilitate treatment of these diseases.

The ubiquitous and well-characterized members of the family, calpains 1 and 2 (µ- and m-, respectively) are heterodimers, containing a large 80 kDa subunit (domains I-IV) and a small 28 kDa subunit (domains V and VI) [48, 49, 132]. Both enzymes share a papain-like protease core (domains I-II) characterized by the presence of the catalytic triad residues cysteine, histidine and asparagine. Domains III and IV are the C2-like and PEF (penta-EF-hand) domains, respectively. The PEF domain (IV) of the large subunit pairs with the homologous PEF domain VI of the small subunit through EF-hand 5, thus forming a heterodimer. In the absence of Ca\(^{2+}\) both isoforms are catalytically inactive and upon binding Ca\(^{2+}\) the heterodimer undergoes multiple structural changes to form the active calpain enzyme. Structural events such as, autoproteolysis, subunit dissociation, intra-/inter- domain rearrangement and phospholipid binding are suggested to be involved in this complex regulation of activation [4, 105, 106, 133].

Five of the human calpains (calpains 5, 6, 7, 10 and 15) have significantly different domain compositions from those of the conventional calpain large subunit, suggestive of distinct functions [84, 133-136]. In particular, they lack a PEF domain with which to dimerize and are presumed to be monomers. The other members of the
calpain family (calpain 3, 8, 9, 11, 12 and 13) do have a PEF domain (DIV). Considering their similarity in domain arrangement to the classic calpains 1 and 2, these isoforms have the potential to form heterodimers with the small subunit. However, recent biophysical studies on recombinant PEF domain of calpain 3 (C3DIV) showed that it forms a very stable homodimer [137]. Molecular modelling demonstrated that this interaction could be the basis for homodimerization of the whole enzyme. A 180 kDa protein was formed by recombinant expression of inactive calpain 3 in the absence of small subunit, which is consistent with homodimerization [109]. The situation with native calpain 3 (p94) is unclear because the enzyme is unstable and rapidly autoproteolysed during purification, but the small subunit does not seem to co-purify with the 94 kDa large subunit. Thus it cannot be assumed that the presence of a C-terminal PEF domain in other calpain isoforms will lead to heterodimerization with the small subunit. One of the reasons why it is important to establish which calpains form heterodimers is that calpastatin, the natural inhibitor of calpains 1 and 2, [66] binds to sites on the PEF domains of both the large and small subunits [50, 56, 138]. In the presence of Ca\(^{2+}\), subdomains A and C of calpastatin tightly associate with PEF domains DIV of the large subunit and DVI of the small subunit, respectively. This binding ensures a high local concentration of subdomain B that binds and blocks the active site cleft of the enzyme. In the absence of the small subunit, calpastatin would lose one of its
binding sites and might not associate tightly enough with large subunit to inhibit it. More to the point, a homodimer of the large subunit would have two active sites at opposite ends of the molecule, and these certainly could not both be inhibited by one calpastatin. In this context, we sought to examine all known PEF domains from human calpain isoforms including calpain 3 to establish if they exist as heterodimers or homodimers.

In order to screen these PEF domains, a co-expression system with differential tags on the recombinant proteins was established. The human small subunit lacking the glycine-rich domain (21 kDa) was tagged with type III antifreeze protein (7 kDa) \([139]\) in its place on the N terminus, while the recombinant domain IV’s of other calpain isoforms had a His\(_6\)-tag on the N terminus (Figure 3.1.). This approach gave us the opportunity to exploit two distinct purification methods, ice affinity purification \([140]\) and Ni-NTA chromatography to characterize these recombinant proteins.

3.3 Materials and Methods

3.3.1 High throughput cloning

The cDNA fragments encoding the DIV regions of calpains 1, 9, 11, 12 and 13 were obtained by PCR amplification of full-length cDNA templates of human calpain 1, 9, 11,
Figure 3.1. Three possible scenarios from co-expression of recombinant PEF fusion proteins

(A) Homodimer model of His$_6$-tagged PEF domain. (B) Homodimer model of type III antifreeze protein-tagged PEF domain (calpain small subunit DVI). (C) Heterodimer model of fusion protein containing type III antifreeze protein (blue) tagged small subunit (cyan) forming a dimer with His$_6$-tagged (light brown) PEF domain (orange). Rat small subunit (1AJ5) was used for modelling. All structures were drawn with PyMOL [141].
12 and 13 obtained from the Mammalian Gene Collection, using Expand high-fidelity DNA polymerase (Roche, Indianapolis, IN, U.S.A.). Human calpain 8 DIV region (C8DIV) was obtained by PCR amplification of reverse transcripts (reverse transcriptase–PCR) of total RNA from human stomach (Stratagene, La Jolla, CA, U.S.A.) using an RT-Thermoscript kit (Invitrogen, Carlsbad, CA, U.S.A.) and Expand high-fidelity DNA polymerase. Human calpain 3 DIV was obtained as previously described. [30] Multiple constructs were designed for each of these domains. The amplified fragments encoding domain IV regions of calpains 1, 8, 9, 11, 12 and 13 were inserted using the infusion ligation independent cloning system (BD Biosciences) into a modified pET28-LIC expression vector (EMD-Novagen) using a 96-well format high-throughput approach[142], downstream of the nucleotide sequence encoding MGSSHHHHHHSSGLVPRLS. This 20-amino-acid sequence contains a hexahistidine tag (His$_6$-tag) and a thrombin cleavage site.

### 3.3.2 Type III antifreeze protein-tagged human small subunit

The cDNA fragment encoding DVI of the human small subunit was obtained by PCR amplification of reverse transcripts (reverse transcriptase–PCR) of total RNA from human stomach (Stratagene, La Jolla, CA, U.S.A.) using an RT-Thermoscript kit
(Invitrogen, Carlsbad, CA, U.S.A.) and Expand high-fidelity DNA polymerase (Roche, Indianapolis, IN, U.S.A.). The amplified product was cloned into the modified pET vector (pAC-pET) as previously described [143]. The type III antifreeze protein (AFP) sequence was previously prepared by gene synthesis [144]. It was cloned into the pAC-pET vector 5' of the truncated 21 kDa subunit sequence. At the protein level the two domains are joined by a linker of three alanine residues.

### 3.3.3 Protein Expression and purification by Ni-NTA

The pET28-LIC vectors encoding the DIV regions were transformed along with the pAC-pET plasmid containing the small subunit fusion construct into *Escherichia coli* BL21 (DE3) cells (Novagen) by electroporation. The transformed cells were grown in 1L of LB (Luria-Bertani) medium under kanamycin and ampicillin selection. The medium was grown to an OD_{600} of 0.8–1.0 at 37 °C. Protein expression was induced at 16 °C using 0.4 mM isopropyl b-D-thiogalactoside for 16 h. The cells were collected by centrifugation, re-suspended in lysis buffer (25 mM Tris-HCl, pH 7.6, 5 mM EDTA, 5%, v/v, glycerol, 10 mM 2-mercaptoethanol and 0.1 mM PMSF) and lysed by sonication. The resulting lysate was clarified by centrifugation at 27000 x g for 45 min. The supernatant obtained was incubated with 5 ml Ni-NTA (Ni^{2+}-nitrilotriacetate–agarose) resin (Qiagen, Chatsworth, 80
CA, U.S.A.) for 30 min at 4 °C with constant stirring. The Ni-NTA resin was later transferred into a column and washed with N-buffer (50 mM Tris-HCl, pH 7.6, 100 mM NaCl, 5 mM imidazole and 0.01% sodium azide). His$_6$-tagged proteins were eluted with the lysis buffer containing 250 mM imidazole. The samples collected were later analyzed by SDS-PAGE. The inactive recombinant rat calpain 2 large subunit (C105S-m-80 kDa) [143] was also co-expressed with the AFP-tagged small subunit and purified as described there.

3.3.4 Ice Affinity Purification

Ice affinity purification [140] was explored as a way of isolating and identifying products containing the type III AFP fusion. In this method the AFP fusion protein was adsorbed from solution (50 ml) into growing polycrystalline ice frozen onto a cooled brass cold finger. The growth of the ice was controlled by circulating cold ethylene glycol solution through the hollow cold finger. After a thin layer of ice (~1 mm) was initially formed on the cold finger it was immersed in the AFP-containing solution pre-chilled to 1 °C in an insulated beaker. The solution was gently mixed using a stir bar and the temperature of the cold finger was gradually reduced at a linear rate (-0.5 to -2.5 °C over 36 h) using a temperature programmable water bath (Neslab) until ~1/2 to 2/3 of the volume was
incorporated into the ice hemisphere. The ice hemisphere was then removed from the liquid and allowed to melt for ~10 min to remove any protein that was nonspecifically bound to surface of the ice. The ice hemisphere was melted to release the AFP. Samples (2 µl and 5 µl) from both melted ice (ice fraction) and leftover liquid (liquid fraction) were analyzed by SDS-PAGE [145].

### 3.3.5 Modelling studies

Shape complementarity of various dimer structures and models was calculated using the program Sc from the CCP4 program suite [120]. Crystal structures of the rat small subunit homodimer (PDB ID: 1DVI) and of the human calpain 2 heterodimer (PDB ID: 1KFU) were used as references. Homology models of the C3DIV homodimer, the heterodimer of C3DIV with the small subunit and of a calpain 2 DIV homodimer were generated using the program Modeller 9v3[117]. The best of 100 models were then used in an energy minimization and molecular dynamics protocol using the program GROMACS 3.3 [146]. The protein was solvated; energy minimized using the steepest descents protocol and subjected to position restrained molecular dynamics to relax the solvent. This was followed by a 2ns molecular dynamics simulation. Structures were extracted from the trajectory every 20ps and the surface complementarity at the dimer
interface was calculated with the program Sc from the CCP4 program suite [120]. The average Sc value from these 100 structures is reported. For comparison, the same molecular dynamics protocol was used on the crystal structures of the rat small subunit homodimer (PDB ID: 1DVI) and of the human calpain 2 heterodimer (PDB ID: 1KFU).

3.3.6 Immunoblotting

Immunoblotting was done using 10% Tris–Tricine SDS–PAGE gels transferred onto PVDF membranes. Polyclonal antibodies against the His-tag and against type III AFP were raised in rabbits. The secondary antibody was anti-rabbit IgG conjugated to horse radish peroxidase (Promega) which was detected by ECL chemiluminescence (Perkin-Elmer).

3.4 Results

Multiple constructs representing the domain IV region of human calpain isoforms 1, 3, 8, 9, 11, 12 and 13 were designed in an effort to improve the likelihood of expressing these recombinant isoforms. Recombinant calpains 1, 3, 9 and 13 domain IV constructs produced high yields when expressed alone or when co-expressed with the human small
subunit (Table 3.1). Constructs of calpain 8, 11 and 12 failed to express. Further trials to stabilize their expression by co-expression with human small subunit did not influence the yield.

3.4.1 Establishing the validity of the screening method using calpain 1 DIV

To test the functionality of the N-terminally AFP-tagged small subunit in forming a natural heterodimer [49, 132] we co-expressed it with the inactive rat calpain 2 (C105S-m-80 kDa) large subunit and with human C1DIV. The rat large subunit was chosen for this purpose because the human ortholog is poorly expressed in E. coli and the residues involved in heterodimer formation are highly conserved in the two mammals. As expected, the calpain 2 large subunit as well as the isolated domain C1DIV (21 kDa) both formed heterodimers with recombinant type III AFP-tagged human small subunit (28 kDa). This was established by Ni-NTA column purification where both the co-expressed constructs were detected in the imidazole eluted fractions (Figure 3.2A, lane 4 and Figure 3.2C, lane 3). In lane 4 (Figure 3.2A), the relative staining of large (80 kDa) and small (21 kDa) subunits is consistent with their 1:1 stoichiometry. When an immunoblot of this gel was probed with anti-AFP antibody the only protein band detected was the 28kDa AFP-tagged small subunit (Figure 3.2B, lane 1). Similarly, when a duplicate
Table 3-1 Screening results for PEF domains in calpains

Table 1 shows screening results of domain IV constructs from calpains 1, 3, 8, 9, 11, 12, 13. Column 1 shows calpains used for screening, column 2 shows the number of constructs designed and cloned. Column 3 shows the number of constructs expressed. Column 4 indicates the yields of constructs when expressed alone in the absence of human small subunit. Column 5 indicates the results from co-expression of the domain IV construct with human small subunit. Column 6 shows the results obtained by biophysical analysis of these constructs. * indicates data not available. ^ indicates predominant form found as homodimer. + + +, + +, + indicates very high, high and low expression respectively.
<table>
<thead>
<tr>
<th>Construct</th>
<th>Cloned</th>
<th>Expressed</th>
<th>Yield</th>
<th>Co-expression yield</th>
<th>Dimerization</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1DIV</td>
<td>1</td>
<td>1</td>
<td>+++</td>
<td>+++</td>
<td>Heterodimer</td>
</tr>
<tr>
<td>C3DIV</td>
<td>5</td>
<td>5</td>
<td>+++</td>
<td>+++</td>
<td>Homodimer</td>
</tr>
<tr>
<td>C8DIV</td>
<td>5</td>
<td>Nil</td>
<td>No Expression</td>
<td>No Expression</td>
<td>*</td>
</tr>
<tr>
<td>C9DIV</td>
<td>14</td>
<td>11</td>
<td>+++</td>
<td>+++</td>
<td>Heterodimer</td>
</tr>
<tr>
<td>C11DIV</td>
<td>11</td>
<td>Nil</td>
<td>No Expression</td>
<td>No Expression</td>
<td>*</td>
</tr>
<tr>
<td>C12DIV</td>
<td>6</td>
<td>Nil</td>
<td>No Expression</td>
<td>No Expression</td>
<td>*</td>
</tr>
<tr>
<td>C13DIV</td>
<td>12</td>
<td>12</td>
<td>+++</td>
<td>+++</td>
<td>Homodimer</td>
</tr>
</tbody>
</table>
Figure 3.2. SDS-PAGE and immunoblot analysis of differentially tagged calpain 1 and 2 heterodimers

(A) Lane 1 shows the molecular-mass standards indicated at the side of the gel. Lanes 2, 3 and 4 contain flow-through, wash and eluate samples, respectively, from the Ni-NTA column chromatography of 80 kDa subunit (C105S-m-80 kDa) (triangle) co-expressed with 28 kDa AFP-tagged small subunit (dot). (B) Lanes 1 and 2 are immunoblots of lane 4 from Figure 3.2A probed with anti-AFP-antibody and anti-His-tag antibody, respectively. (C) Lanes 1, 2 and 3 represent flow-through, wash and eluate samples from the Ni-NTA column chromatography of C1DIV (square) co-expressed with 28 kDa AFP-tagged small subunit (dot). Both co-expressed constructs are predominantly detected in fractions eluted with imidazole.
Figure 3.2

(A) Protein Marker, Flow-through, Wash, Elute

(B) Anti-AFP, Anti-His

(C) Flow-through, Wash, Elute

- 28 kDa subunit
- 80 kDa subunit
- C1DIV (21.8 kDa)
immunoblot was probed with anti-His-tag antibody the only protein band detected was the 80kDa His-tagged large subunit (Figure 3.2B, lane 2).

One immediate value of the type III AFP-tagged small subunit construct is the increase in its molecular mass from 21 kDa to 28 kDa that readily distinguishes it from DIV constructs. Thus, in lane 3 of Figure 3.2C, the upper 28 kDa band of the small subunit is well separated from the lower, more abundant His6-tagged C1DIV. Although the presence of AFP-tagged small subunit in the affinity purified His6-tagged C1DIV shows that the two different PEF domains form heterodimers, the relative staining of these two bands suggests that C1DIV is present in excess.

3.4.2 Calpain 3 DIV is a homodimer

Calpain 3 domain IV (C3DIV) is suggested to favour homodimerization even though small subunit-containing calpains are produced in muscle cells. Earlier studies on recombinant C3DIV when expressed in isolation formed a homodimer [137]. In further support of the argument we show below that His6-tagged recombinant C3DIV co-expressed with type III AFP-tagged human small subunit (28 kDa) exclusively forms a homodimer. Upon purification by Ni-NTA chromatography the 28 kDa subunit was not detected in the
imidazole eluted fraction along with C3DIV (Figure 3.3A, lane 4). The 28 kDa subunit was present in the fractions that did not bind to the Ni-NTA column (Figure 3.3A, lane 2). Indeed, it was the most abundant protein in the flow-through fraction from that column.

3.4.3 Calpain 9 DIV forms a heterodimer with the small subunit

The recombinant calpain 9 domain IV (C9DIV) construct has 200 amino acids including its His\textsubscript{6} N-terminal tag. It has a theoretical pI of 5.71 and a calculated molecular mass of 23130 Da. The amino acid sequence is 43% identical with DIV of calpain 1 and 40% identical with the small subunit (28 kDa). When C9DIV was co-expressed with the 28 kDa small subunit fusion protein it formed a heterodimer. Both subunits were detected in the imidazole eluted fraction (Figure 3.3B, lane 3). Their stoichiometry was close to 1:1. To confirm the identity of the two subunits the gel was immunoblotted and probed with the two antibodies used in Figure 3.2B. The anti-AFP antibody detected the upper band as 28kDa AFP-tagged small subunit (Figure 3.3C, lane 1). Similarly, the anti-His-tag antibody reacted with the N-terminally His\textsubscript{6}-tagged C9DIV (Figure 3.3C, lane 2).
Figure 3.3. SDS-PAGE and immunoblot analysis of DIV calpain 3 and 9 samples co-expressed with small subunit

(A) Lane 1 shows the molecular-mass standards indicated at the side of the gel. Lanes 2, 3 and 4 contain flow-through, wash and eluate samples, respectively, from the Ni-NTA column chromatography of His-tagged C3DIV (triangle) co-expressed with 28 kDa AFP-tagged small subunit (dot). Only the C3DIV domain is detected in the eluant. (B) Lanes 1-3 represent flow-through, eluate and wash samples from the Ni-NTA column of His-tagged C9DIV (square) co-expressed with 28 kDa AFP-tagged small subunit (dot). Both the human small subunit and C9DIV are present in the eluant. (C) Lanes 1 and 2 are immunoblots of lane 3 from Figure 3.3B probed with anti-AFP-antibody and anti-His-tag antibody, respectively.
Figure 3.3

(A) Protein Marker, Flow-through, Wash, Eluate

(B) Flow-through, Wash, Eluate

(C) Anti-AFP, Anti-his

- 28 kDa subunit ▲ C3DIV (22 kDa) ■ C9DIV (23 kDa)
In the converse approach using ice affinity purification His$_6$-tagged C9DIV was included into ice because of its heterodimerization with the AFP-tagged small subunit (Figure 3.4, lane 2). Here the amount of the His$_6$-tagged C1DIV in the ice fraction is slightly lower than the expected 1:1 stoichiometry with the small subunit as seen in the liquid fraction (lane 3). This seems to be due to a small amount of subunit dissociation that occurs as the ice grows over and pushes past the adsorbed AFP-tagged subunit. The shearing forces of the ice are apparently sufficient to disrupt quaternary structure in a portion of the dimers but do not break covalent bonds between the AFP moiety and a fusion partner [140]. A similar partial dissociation of subunits was seen during ice affinity purification of full length $\mu$-calpain heterodimerized to the AFP-tagged subunit (results not shown). The control experiment in this series showed that His$_6$-tagged C9DIV when expressed alone was not included into ice but remained in the liquid fraction (Figure 3.4, lanes 4 and 5, respectively).

3.4.4 Calpain 13 DIV

The recombinant calpain 13 domain IV (C13DIV) construct contains 174 amino acids including the His$_6$ N-terminal tag. It has a theoretical pI of 6.75 and a calculated molecular mass of 19901 Da. Unlike other calpain PEF domains it has low sequence
Figure 3.4. Ice affinity purification of type III AFP-tagged small subunit and calpain 9 DIV

Lane 1 shows the molecular-mass standards indicated at the side of the gel. Lanes 2 and 3 contain equal volumes of the ice and liquid fractions obtained from the distribution of co-expressed 28 kDa AFP-tagged small subunit (dot) with His-tagged C9DIV (square). Lanes 4 and 5 contain equal volumes of the ice and liquid fractions obtained from the distribution of His-tagged C9DIV (square) in the absence of 28 kDa AFP-tagged small subunit.
Figure 3.4
Figure 3.5. SDS-PAGE analysis of calpain 13 DIV samples from Ni-NTA column

Lane 1 shows the molecular-mass standards indicated at the side of the gel. Lanes 2, 3 and 4 contain the flow-through, wash and eluate fractions from the column, respectively. The 28 kDa subunit and C13DIV proteins are marked by dot and square symbols, respectively.
Figure 3.5
identity with DIV of calpain 1 (28%) and the small subunit (29%). When the recombinant C13DIV construct was co-expressed with type III AFP-tagged human small subunit (28 kDa) C13DIV was predominantly seen in the eluant. The 28 kDa small subunit was mainly observed in the flow-through, although a faint band was seen in the wash and eluant (Figure 3.5). Based on these SDS-PAGE results a small amount of heterodimer is produced but C13DIV is predominantly a homodimer.

3.5 Discussion

The penta-EF-hand (PEF) domain was first described in calpain[44-46] and has since been found in other proteins such as ALG-2, grancalcin, sorcin and peflin [110, 147]. It is characterized by having a fifth EF-hand available to pair with that of another PEF domain to form hetero- or homodimers. More than half of the human calpain isoforms (1, 2, 3, 8, 9, 11, 12 and 13) have a penta-EF-hand domain. Of these, the ubiquitous well-studied calpains 1 and 2 are known to form heterodimers with the small subunit PEF domains. However, previous investigations on calpain 3 suggest that PEF domain-containing calpain isoforms need not necessarily form a heterodimer like calpains 1 and 2. In this study we set out to determine what kind of dimers the different calpain isoforms make.
Modelling studies using shape complementarity as a tool to measure the likelihood of forming a hetero- or homodimer were done using calpain 2, the previously generated model of C3DIV [137] and the small subunit structures as a guide. In addition, models were generated for artificial structures of the C3DIV-small subunit heterodimer and of the calpain 2 DIV homodimer. Shape complementarity values differed only slightly between the different dimers. In order of best through worst, the complementarity values were C3DIV homodimer (0.751), small subunit homodimer (0.751), C3DIV-small subunit heterodimer (0.734), C2DIV-small subunit heterodimer (0.734) and C2DIV homodimer (0.715). These values are not significantly different from each other and therefore do not appear to provide a method for distinguishing correct from incorrect dimers. Comparison of the buried surface areas for the various complexes also shows little variation with the calpain 2 DIV homodimer displaying the smallest surface area (average value of 1182 Å²) compared to the others (average values ranging from 1311 Å² to 1391 Å²). Since tight packing of residues involved in the dimerization interfaces might not be the only factor influencing dimer formation we used experimentation to distinguish which isoforms form hetero or homodimers.

The recombinant small subunit DVI has a molecular mass of 21264 Da and forms a homodimer when expressed alone [124]. Its molecular mass is close in value to those of isolated calpain PEF domains (domain IV) making it hard to distinguish whether they
formed homodimers or heterodimers when co-expressed. In order to overcome this uncertainty we devised a differential tag approach whereby all the calpain PEF domains contain a His$_6$ N-terminal tag and the small subunit has a N-terminal type III AFP tag (7 kDa) allowing us to distinguish these two domains by size. Like the rat small subunit, the recombinant 28 kDa human small subunit fusion protein formed a homodimer when expressed alone (results not shown).

Calpains 1, 3, 9 and 13 PEF domains were successfully cloned and co-expressed as soluble recombinant products. However, numerous trials to express calpains 8, 11, 12 PEF domain constructs in *E. coli* were unsuccessful and thus the dimerization potential of these PEF domains could not be analyzed. Since the wild-type calpains 1 and 2 are both known to form heterodimers, we used calpain 2 large subunit and calpain 1 domain IV, as controls in our experiments. Even in the absence of its adjacent domains C1DIV formed a heterodimer with the small subunit, rather than a homodimer. It should be noted that this construct lacks the N-terminal anchor peptide that, based on the structure of calpain 2 [48, 49], should make additional heterodimerization contacts between the large and small subunits.

Recombinant C3DIV was previously shown to form a homodimer when expressed alone [30]. In this study it was co-expressed with small subunit (28 kDa) but still formed a
homodimer, further supporting the argument that calpain 3 is a natural homodimer. Calpain 9 has been previously suggested to form a heterodimer when co-expressed with small subunit in the baculovirus-expression system [127]. Co-expression of recombinant proteins C9DIV and small subunit fusion product (28 kDa) led these proteins to associate as a heterodimer in agreement with these previous studies. As with calpain 1, the absence of the other domains in the large subunit did not alter the propensity of the C9DIV to heterodimerize. When expressed alone C9DIV formed an oligomer unlike other PEF domains (results not shown). Calpain 13 is a tissue-specific calpain expressed predominantly in testis and lung. Its physiological role is not well understood and its dimerization state is unknown [71]. Calpain 13 domain IV appeared as a predominant homodimer when co-expressed with small subunit fusion protein (28 kDa), although there were small amounts of heterodimer present in the eluate from the Ni-NTA column.

Most of the PEF domains in calpain isoforms share a high degree of sequence identity; however it is not clear as to why they prefer one form of dimerization over the other. Further analysis of these constructs by determining their structure through crystallography may help us gain more insight into the preference for homodimerization vs. heterodimerization. Meanwhile, based on these results we predict that calpain 9 can be bound and silenced by calpastatin. Silencing of calpains 3 and 13 would require the
simultaneous binding of two calpastatin inhibitory domains. Although this is a theoretical possibility, especially since calpastatin has four inhibitory domains and is an intrinsically unstructured protein, the absence of a small subunit in these two calpains would deprive calpastatin of one of its three calpain-binding sequences. The loss of this binding site would significantly weaken the overall binding interaction.
Chapter 4

Structural genomics of other regulatory domains

Preface:

This chapter will form the basis for manuscripts on a) Characterization of C2-like regulatory domains in calpains and b) Penta-EF-hand domain structures of tissue-specific calpains.

Ravikiran Ravulapalli was responsible for high-throughput cloning, protein expression and purification of all the constructs detailed in this chapter. Crystallization screening and optimization of distal C2-Like domain of human calpain-7 and domain IV of calpain 13 were performed by Dominic Cuerrier (University of Toronto) and K. Ng (University of Toronto), respectively. The crystallographic structure for both the domain constructs was solved by John R. Walker (University of Toronto). The experiments were performed under supervision and guidance of Dr. Peter L. Davies (Queen’s University) and Dr. Sirano Dhe-Paganon (University of Toronto). The chapter was written by Ravikiran Ravulapalli with guidance from Dr. Peter L. Davies.
4.1 Introduction

Calpains are family of cysteine proteases that are found in mammals and many other organisms [37-40]. In mammals, calpains form a large gene family with currently 14 members identified in humans. Calpain family member’s exhibit great sequence divergence and varied domain structure but have homologous catalytic domains. Given the large number of calpain targets and complexity involved in characterizing each isoform using conventional methods, a structural genomics-based approach present an opportunity for high-throughput analysis of these targets. In addition, multiple constructs are designed in this approach, for each target of interest, which improves the probability of expressing and crystallizing a representative construct. The structural information about a protein obtained from solving its crystal structure may help in predicting its potential function. However, targeted research on a protein of interest using a combination of biophysical and biochemical approaches will be crucial for identifying its specific role in the cell.

The following chapter details an exploration of the calpain regulatory domains using a structural genomics approach in collaboration with Dr. Sirano Dhe-Paganon’s group at the Structural Genomics Consortium (SGC) in Toronto. The Consortium has the infrastructure and technologies necessary for rapid, parallel structure determination.
4.2 Materials and Methods

4.2.1 High throughput cloning

BD In-Fusion™ Dry-Down PCR Cloning Kits were used as a simple, convenient way to clone PCR products [148]. These kits allow the cloning of any PCR fragment into any vector without using restriction enzymes, ligase, or blunt-end polishing. The procedure involved a simple 30-minute bench-top incubation of the PCR product with the linearized vector. The method relied on BD In-Fusion Enzyme, a protein that efficiently and precisely fuses PCR-generated sequences to linearized vectors. SGC custom-made cloning vector pET 28-MHL was used for this purpose. All the targets were amplified and cloned using a 96-well format high-throughput approach. Using the following methodology multiple constructs representing different regulatory domains were cloned.

4.2.2 Expression and purification of recombinant proteins

*E. coli* was used for expression of target recombinant proteins. Plasmids that encode particular genes of interest were used to transform competent bacterial cells, typically BL21. Culture growth was done in a high-throughput manner using a simple bubbling
system developed by Dr. Dhe-Paganon. Regular air was pumped through a simple frit system that simultaneously exposes bacteria to oxygen and creates a mixing action. A manifold allowed aeration of up to 48 bottles. Temperature was controlled by a circulating water bath. The system is called LEX and is an innovative approach to high-throughput bacterial growth. Once the desired bacterial density was reached, growth was inhibited in order to redirect most of the bacterial energy towards protein production. This was accomplished by dropping the temperature to 16°C, adding IPTG, and allowing protein expression to continue overnight. Cells were then harvested by centrifugation and bacterial pellets stored at -80°C.

Cell lysis was accomplished by sonication (Virtis408912, Virsonic) on ice and the resulting lysate was clarified by centrifugation. Recombinant protein obtained was purified by a combination of Ni-NTA Resin / Talon Resin and size-exclusion chromatography techniques. Column chromatography was accomplished using the AKTA Express system (GE Healthcare). This system combined serial and parallel purification of proteins allowing it to purify 4 recombinant proteins per module at a time (with the option to have up to 12 modules as a unit). Fractions were collected using a 96-well fraction collector integrated for each module.
4.2.3 Crystal Screening

Reagent and protein solutions were dispersed automatically into 96-well sitting drop crystal plates by a liquid handling robot and a “mosquito” robot (http://www.ttplabtech.com), respectively. The liquid handling robot was used to transfer the reagents, typically 100 micro-liters, to crystal plates as mother liquor from a 96-deep-well reservoir plate. Using the “mosquito” robot, protein was dispensed in volumes ranging from 100 nano-liters to 1.2 micro-liters and mixed with mother liquor in pre-determined ratios. SGC-designed custom screens (SGC-1 Screen) and Red-Wings screen (courtesy of Alexei Savchenko's group at Structural Proteomics in Toronto) were used for screening.

4.3 Results and Discussion

4.3.1 Expression of C2-like regulatory domains

Using the structural genomics approach a concerted effort was made to express almost all of the putative C2-like domains from the different calpain isoforms. Table 4.1 shows a summary of the results.
Human calpain 5 and 6 isoforms both have a C2-like domain III and a tra3-like domain IV (Figure 1.4). However, none of the C2-like and tra3-like domain constructs could be expressed.

b) Human calpain 7 has a PBH-like domain III and C2-like domain IV. It also has a 217-amino-acid N-terminal extension sequence considered to contain two Zn-fingers (Figure 1.4). Both the calpain 7 PBH (C7PBH) domain and the C2-like domain IV (C7DIV) were expressed as recombinant soluble products. Also, some of the N-terminal extension sequence constructs produced soluble products (data not included in Table 4.1).

c) Calpain 10 has two C2-like domains (Figure 1.4). The distal C2-like domain was successfully expressed as a recombinant soluble product.

d) Calpain 9 and 13 each have a C2-like domain III. Calpain 9 C2-like domain constructs did not yield expression products despite numerous efforts, whereas, a couple of human calpain 13 C2-like targets were soluble when expressed.

4.3.2 Calpain 7 distal C2-like domain structure

Crystals of the calpain 7 distal C2-like domain were grown at room temperature using the hanging drop method by mixing equal volumes of 1.4 M sodium phosphate (pH 7.0) and 13 mg/mL protein in binding buffer (500 mM NaCl, 50 mM Tris-HCl pH 8.0).
Table 4-1 Structural genomics of regulatory domains: Screening results.

The number of constructs cloned and expressed are indicated in the last two columns. * Indicates no expression was obtained.
The crystals were cryo-protected by treating with a 50% mixture of mineral oil and paratone.

A high resolution structure (1.45 Å, PDB ID: 2QFE) was determined from data collected at the Cornell High Energy Synchrotron Source (New York, USA). Like other C2-like domains it forms an 8-stranded beta-sandwich connected by loops; however, the topology is distinct from canonical C2 domains like that of synaptotagamin I (Figure 4.1). The structure also shows some variation from previously determined calpain C2-like domains. When superimposed with the C2-like domain III of calpain 2 (PDB ID: 1KFU), the RMSD value was 7.1Å (Figure 4.2). There is a notable difference in the loop connecting the beta4 and beta5 strands. This loop lacks helix alpha2 and forms a novel beta-strand with beta4 in calpain 7. The distal C2-like domain of calpain 7 also lacks the acidic and basic loops that form extensive contacts between domain III and the proteolytic core in calpain 2 suggesting that the distal C2-like domain of calpain 7 performs a role separate from that of structural support for the proteolytic core. Moreover, despite the presence of Ca$^{2+}$ in the crystallization buffer, the structure did not bind Ca$^{2+}$ suggesting that this domain does not directly bind Ca$^{2+}$. This structure serves to further emphasize the structural divergence of calpain C2-like domains from canonical C2 domains.
Figure 4.1. Distal C2-like domain of calpain 7 (PDB ID: 2QFE)

Ribbon diagram representation of the high resolution crystal structure of the C2-like domain of calpain 7. Beta-strands are colored yellow, loops green and helices red.
Figure 4.2. Superimposition of the distal C2-like domain of calpain 7 on domain III of m-calpain

The distal C2-like domain of calpain 7 and domain III of m-calpain are superimposed in this ribbon diagram representation. The C2-like domain of calpain 7 is colored purple (PDB ID: 2QFE). The C2-like domain of m-calpain (PDB ID: 1KFU) is colored green. Some of the structural deviations are highlighted in the figure. The RMSD value was calculated to be ~7.1 Å.
Figure 4.2

- basic loop absent
- α2 helix absent
- acidic loop absent
- novel β-strand (between β4 & β5)
- complete structural divergence in β4-β5 loop
4.4 Structure of calpain 13 PEF domain

Crystallization trials of recombinant calpain 13 domain IV (C13DIV) were carried out using the hanging drop vapor diffusion method. Diffracting crystals leading to the structure grew when the protein drop was equilibrated against a reservoir solution (1:1 volume ratio) of a final concentration of 25% PEG4K, 0.2M ammonium sulfate, 0.1M sodium acetate (pH 4.6), 10 mg/mL protein solution, in the absence of Ca$^{2+}$. The crystal was cryoprotected with 25% glycerol.

A high resolution structure (1.80 Å) [Figure 4.3a] was determined from data collected at the Advanced Photon Source (Chicago, USA). Two molecules were present in the unit cell and they formed a homodimer by crystallographic symmetry. C13DIV can be superimposed on the analogous domain of calpain 2 (PDB ID: 1KFU), with an RMSD of 2.1Å. It also aligns well with the common regulatory small subunit whose structure has been determined both with (PDB ID: 1AJ5) and without (PDB ID: 1DVI) Ca$^{2+}$ to give RMSD values of 1.5Å and 1.1 Å, respectively. C13DIV was purified in the presence of Ca$^{2+}$, but no additional Ca$^{2+}$ was added during crystallization. A single Ca$^{2+}$ atom is seen between α2 and α3. It binds in pentagonal bipyramidal coordination (Figure 4.3b). Unfortunately, EF-hand 1 was truncated during efforts to produce high levels of C13DIV. While C13DIV exhibits good structural similarity to the rat small subunit structure, many critical residues responsible for Ca$^{2+}$ coordination are not conserved in other EF hands.
(EF3-EF5), so it is not clear at this point if these EF-hands do not bind Ca\(^{2+}\) or will need high calcium concentrations to bind Ca\(^{2+}\).

**Figure 4.3 High resolution crystal structure (1.8 Å) of calpain 13 PEF domain.**

(a) Ribbon representation of calpain 13 domain IV (C13DIV) modeled as a dimer. Ca\(^{2+}\) is coloured orange. Note that C13DIV forms a dimer by crystallographic symmetry.

(b) Residues involved in pentagonal bipyramidal Ca\(^{2+}\) coordination at EF hand 2 of C13DIV. Five ligands (X, Y, Z, W, -Y) provide five oxygens for coordination and the 12th residue (-Z) is coordinated in a bidentate fashion. Although the second acidic residue (position Y) is not conserved in calpain 13, it was still able to bind Ca\(^{2+}\).
Chapter 5

Characterization of human tissue-specific isoforms – Calpain 6

Preface:

This chapter will form the basis for manuscript on the characterization of human tissue-specific isoform – Calpain 6.

Ravikiran Ravulapalli was responsible for cloning, protein expression, purification and all biochemical experiments in the investigation of recombinant calpain 6 protease core (CC6) function. The experiments were performed under supervision and guidance of Dr. Peter L. Davies (Queen’s University). The chapter was written by Ravikiran Ravulapalli with guidance from Dr. Peter L. Davies.
5.1 Introduction

Calpain 6 is a unique member of the human calpain family that is thought to be a non-proteolytic enzyme due to the absence of the catalytic triad residue, cysteine, which is replaced by lysine [82]. The calpain 6 expression pattern is different from those of other calpain family members. During embryogenesis it is expressed in multiple tissues including skeletal muscle, heart, lung, kidney and placenta, but is downregulated and restricted to the placenta after birth. Calpain 6 also lacks the EF-hand-containing domain found at the C terminus of the majority of calpains. It contains instead a C-terminal tra-3-like domain, also present in calpain 5 and first described in C. elegans [83]. Of interest to calpain researchers is what role calpain 6 plays in the cell physiology as an inactive enzyme. It has been implicated recently in processes related to microtubule stabilization, cytoskeletal organization and tumorigenesis [84, 85]. In our study on calpain 6 we have tested the protease core of calpain 6 for proteolytic activity and explored the possibility of restoring its activity by mutating the lysine residue back to cysteine.
5.2 Methodology

5.2.1 Cloning, expression and purification of calpain 6 protease core (CC6)

The cDNA encoding calpain 6 protease core was amplified using RT-PCR of total RNA from human placenta (Stratagene) using a RT-Thermoscript kit (Invitrogen) and Expand high fidelity DNA polymerase (Roche). The amplified construct was cloned into a pET 24a vector. The recombinant calpain protease core 6 (CC6) was expressed in *E. coli*. CC6 was purified by a three-step purification process using DEAE-Sepharose, Ni/nitriloacetic acid-agarose and size-exclusion (Sephadex G75) chromatography.

5.2.2 Mutagenesis of CC6

Primers (containing diagnostic restriction sites) were designed which could generate amino acid replacements within the CC6 protease core sequence. Two single amino acid replacement mutants, K81C and L286W, were made using the Kunkel mutagenesis method [149]. The K81C mutant was designed to see if restoring the active site residue would confer activity. L286W was designed to introduce a diagnostic tryptophan to be used in intrinsic tryptophan fluorescence (IWF) to track Ca^{2+}-induced conformational changes in CC6, just as the equivalent tryptophan (Trp298) residue
proved useful in similar studies on m- and µ-calpain isoforms.[4] Both mutants (CC6K81C and CC6L286W) were expressed in E. coli. They were purified following the same protocol used for CC6.

5.2.3 Enzyme assays and intrinsic tryptophan fluorescence studies

Autoproteolytic activity was measured by incubating 1mg/ml solution of inactive m-calpain (mC105S) as the substrate with 1mg/ml solution of either CC6 or its active site mutant construct CC6 K81C in buffer containing HEPES-NaOH (pH 7.6), 10 mM DTT, 100 mM Ca^{2+} at room temperature. Aliquots were removed at different time points (0-6h), and the reaction was stopped by the addition of 3X SDS sample buffer (187.5 mM Tris-HCl (pH 6.8), 6% SDS, 30% glycerol, and 0.3% bromphenol blue). The progress of autolysis was analyzed by 10% SDS-PAGE and Coomassie Blue staining. Intrinsic tryptophan fluorescence studies were performed as described in chapter 1.

5.2.4 Sequence alignment

Multiple sequence alignment studies were performed using program CLC sequence viewer (CLC Bio). Calpain 6 protease core was aligned with the classic isoforms calpain 1 and 2 as detailed in figure 5.1
Figure 5.1. Multiple sequence alignment of calpain 1, 2 and 6 protease cores

The amino acid sequences corresponding to calpains 1, 2 and 6 were aligned using CLC sequence viewer (CLC Bio). Based on their increasing identity score residues are colored blue to red, with blue being the least identity score and red being the highest identity score. The residues involved in Ca$^{2+}$ coordination in calpain 1 and 2 are indicated by arrows. Green arrows correspond to residues involved in coordinating the first Ca$^{2+}$ in the protease core; Blue arrows correspond to residues involved in coordinating the second Ca$^{2+}$ ion. Catalytic triad residues are indicated by a star.
Figure 5.1
Figure 5.2. Inactive m-calpain (mC105S) proteolysis activity assay

This assay monitored the digestion of the 80 kDa large subunit of mC105S by μ I-II (control lane 1, indicated by *) and by (a) CC6 (lanes 2-6) (b) CC6K81C (lanes 2-6) into its cleavage products. Each digestion was done at 24°C in the presence of 100 mM Ca^{2+}. The enzyme to substrate molar ratio was 1:1. The time of digestion is listed in hours above each lane (0, 1 h, 3 h, and 6 h).
5.3 Results and Discussion

Recombinant calpain 6 protease core (CC6) was produced as a soluble product and it remained soluble throughout the purification processes and eluted as a single entity with a sharp elution profile during column chromatography, suggesting it to be a well-folded protein. The Mr value of the CC6 domain deduced from size-exclusion chromatography was approximately 42 kDa (theoretical mass = 41,011 Daltons).

Given the propensity of calpain protease cores (μ, m and p94 minus IS1) to hydrolyze inactive recombinant m-calpain heterodimer (mC105S) in the presence of Ca\(^{2+}\), we performed similar proteolysis studies using CC6 (Figure 5.2a). After 1 h in presence of Ca\(^{2+}\), three digestion products were produced with digestion from mC105S by the control protein μI-II, while the recombinant protein CC6 was unable to hydrolyze the inactive m-calpain heterodimer even after 6h, suggesting it to be completely inactive as expected. Mutagenesis studies were performed to bring back or introduce activity by replacing the lysine with cysteine (CC6K81C mutant). Autolysis studies performed on this mutant (CC6K81C) met with a similar outcome like the wild-type (Figure 5.2b), suggesting that more residues in the protease core need to be mutated to make it active.

Similarly another residue (L286) was mutated to generate the CC6L286W mutant. The tryptophan residue in classical calpains 1 and 2 (W298 in calpain 1 and
W288 in calpain 2) forms a wedge between the two domains in the protease core (DI-DII) in the absence of Ca\(^{2+}\). Upon binding Ca\(^{2+}\) this residue contributes most of the overall change in fluorescence caused by Ca\(^{2+}\) that can be monitored by intrinsic fluorescence studies [4]. When the protease cores of calpains 1, 2 and 3 are titrated with Ca\(^{2+}\) in the resultant realignment of their two domains is accompanied by a large increase in IWF. Similar trials were performed on CC6 and CC6L286W, but no increase or change in IWF was observed on titrating Ca\(^{2+}\) (data not shown). Multiple sequence alignment of calpain 6 against the classical isoforms 1 and 2 shows that some critical residues of domain I involved in site 1 Ca\(^{2+}\) coordination and the active site cysteine have been replaced in calpain 6. (Fig 5.1) Eg. Side chain residue D106 of calpains 1 and 2 which offers two coordinating atoms for Ca\(^{2+}\) binding is replaced by Asparagine (Q). It is not clear at this point if calpain 6 is capable of binding Ca\(^{2+}\) to bring about realignment of active site cleft as occurs in the classical isoforms. Further active site mutation studies need to be performed in concert with structure determination to understand the functional role of calpain 6.
Chapter 6

General Discussion

6.1 Structural genomics approach to characterize calpains

Human genome sequencing increased the number of identified calpain isoforms to 14. Most of these calpain isoforms have not yet been characterized and their physiological role remains unknown. These enzymes are of great interest owing to their involvement in several disease states. The large number of calpain targets (including whole enzymes, unique domains and domain combinations) are too numerous to be characterized by conventional methods. But structural genomics approaches presented an opportunity for high-throughput analysis of these proteins and attracted the attention of the Structural Genomics Consortium (SGC) at Toronto. Dr. Sirano Dhe-Paganon’s group at SGC tried to express full-length versions of all known calpain isoforms as recombinant soluble products but were unsuccessful. It is worth noting that, our group at Queen’s University were fortunate to have chosen rat calpain 2 (m-calpain) as an initial target for calpain research because the recombinant enzyme is expressed in *E. coli* as a soluble protein in high yield. In contrast, the closely related calpain isoforms including human and rat calpain 1 (μ-calpain) and human calpain 2 are hard to express as full-length, stable recombinant products. Dr. Dhe-Paganon’s group later started to look at the
protease cores (minicalpains) of all the calpain isoforms, a similar approach to the one our laboratory pioneered for calpain 1, which led to the discovery of two novel calcium binding sites in the protease core of calpain 1[4] and the first inhibitor-bound calpain structures. We learnt about Dr. Dhe-Paganon’s group’s efforts in this area from the deposition of the calpain 9 protease core structure in the PDB. Subsequently, we collaborated with their group to screen other regulatory domains of calpains, which were not explored earlier, and to further characterize some of the protease cores.

Multiple constructs were designed for each target of interest, to improve the probability of expressing and crystallizing a representative domain. We cloned nearly 205 constructs representing different individual domains or combinations corresponding to the regulatory regions of all the calpain isoforms as described in chapter 3. Using the *E. coli* expression system we were able to successfully produce some PEF domain constructs (calpains 3, 9 and 13), C2-like domain constructs (calpains 7 and 10) and other regulatory domain constructs as recombinant soluble products. Conversely, in spite of our best efforts, we were unable to express the majority of these domains (eg. 36 constructs comprising calpain 5 and 6 regulatory domains). Even though the putative domain structures are well-defined for these modules and they share a high degree of sequence similarity, we were unable to predict which specific N- and C-terminal boundaries would be most compatible for expression of soluble protein. One of the
limiting factors for the PEF domains is that the C-terminal region of these proteins cannot be truncated as it is involved in dimerization. For example, with calpain 12 domain IV (C12DIV) we were unable to express any of the eight N-terminally truncated constructs, but a few constructs with C-terminal truncations gave reasonable expression. However, the C-terminal truncated versions were not used for dimerization studies as loss of C-terminal residues might have influenced the dimerization status of the enzyme.

Most of the expression results shown in this thesis are from “first generation” constructs (constructs of a particular target domain or protein that were initially designed based on its known domain boundaries). Given the knowledge obtained from expressing these constructs, it is feasible to generate new constructs that could improve their expression. In addition, there is the option to try a different host system, (for example, a baculovirus vector in insect cells) for testing the expression of both full-length enzymes and individual domains. For the constructs which were already expressed as soluble products in low or reasonable quantities or which produced insoluble aggregates, we would make 2nd generation constructs to improve the recombinant protein production. In addition, we could express these domains as fusion proteins with some solubility enhancing fusion tags like a His$_6$-maltose-binding protein tag [150].
The structures of calpain 13 PEF domain (chapter 4) and C2-like distal domain of calpain 7 (chapter 4) have been solved. The calpain 13 PEF domain has a high degree of structural similarity to the small subunit (cpns1). In contrast, the distal C2-like domain of calpain 7 showed greater deviation from the C2-like domain III of calpain 2 suggestive of unique functions for these domains in different tissue-specific calpain isoforms (calpain 5, 6, 7 and 10). Other soluble recombinant proteins are undergoing crystallization trials.

6.2 Understanding dimerization in calpain isoforms: implications for calpastatin inhibition

Calpain 3, which is predominantly expressed in skeletal muscle, has a similar domain composition to that of the large subunit of the calpains 1 and 2. Like the classical calpains, it was thought to form a heterodimer. However, there were reports in the late 90's that the small subunit was not present in preparations of calpain 3 from skeletal muscles [109]. In addition, yeast two-hybrid assays on both inactive and active forms of calpain 3 failed to show binding of the small subunit to the large subunit [73]. So, there was ambiguity about the dimerization status of calpain 3 and no definitive studies were done to tackle this issue.
Owing to the problems of instability and rapid autolysis often associated with calpain 3 full-length enzyme, we have expressed the isolated domain IV construct of calpain 3 (C3DIV) in *E. coli*. The recombinant construct was soluble and stable throughout the purification processes. We have established through size-exclusion chromatography and analytical ultracentrifugation that it exists as a homodimer in solution as detailed in chapter 2. Modelling studies also suggest that there would be no barriers to formation of a homodimer via the EF-hand 5 of the PEF domain in the full-length enzyme. The homodimerization potential of calpain 3 prompted us to look at the dimerization status of other PEF domain-containing calpain isoforms. The PEF domains of all the calpain isoforms have a high degree of similarity and it was not clear either by sequence analysis or modelling studies whether the domains would homodimerize or heterodimerize. As detailed in chapter 3, we resorted to experimentation and co-expressed DIV of all PEF domain-containing calpains with the small subunit. Domain IV’s of calpain 1, 3, 9, 13 successfully co-expressed with the small subunit. While calpain 1 and 9 formed heterodimers, calpain 3 formed a homodimer and calpain 13 existed predominantly as a homodimer. The homodimerization of a calpain would mean that all its domains are duplicated, which would give the enzyme more surface area to access substrates, and the capability to react with more binding partners or ligands at the same time. An active enzyme like calpain 3 could potentially hydrolyze target substrates more
rapidly and effectively because it would have two protease cores. Also, calpain 3 dimerization would give it the ability to bind a large protein like titin more tightly. The IS2 region present between domain III and domain IV is thought to be a flexible structure that points outward from the enzyme and is considered to be the binding site for titin. Duplication of IS2 would provide two opportunities to bind titin and increase the strength of binding. Binding to titin protects calpain 3 from autolysis. Mutations in the IS2 region that results in loss of anchorage are clearly deleterious for muscle and have pathological consequences [151] signifying the importance of this interaction. Inactivating mutations of calpain 3 have been linked to limb girdle muscular dystrophy 2A (LGMD2A), emphasizing the importance of having a functioning protease core.

Understanding and predicting dimerization in calpains from linear amino acid sequence is not yet feasible. Given the high sequence similarity and conservation among PEF domains it is still not clear to why calpains prefer one dimer form over the other. EF hand 5 is used as the dimerization interface and the interactions between the two PEF domain monomers are principally hydrophobic [101]. The hydrophobic residues involved are indicated in figure 6.1A. Interestingly, the same residues with only a few conservative substitutions are also present in corresponding positions in domain IV from other calpains (eg. Calpains 1, 2, 3 and 9), suggesting their significance. Despite the
conservation of these residues, calpain 3 prefers homodimerization unlike classical isoforms 1 and 2, which prefer heterodimerization. A comparison of crystal structures of the small subunit homodimer (1DVI), calpain 2 domain IV with the small subunit heterodimer (1KFU), and calpain 3 domain IV homodimer model (1Y9V) (Figure 6.1B) together with space complementarity and other modelling studies reported in chapter 3 fail to explain the dimerization preferences of the calpains.

Homodimerization of calpains has important implications for the role of the endogenous calpain inhibitor, calpastatin, which might not inhibit calpain dimers as efficiently as it does the classical calpain heterodimers. Sub-domains C and A of each calpastatin inhibitory domain (CAST1-4) bind small subunit and domain IV, respectively. This binding ensures a high local concentration of sub-domain B which inhibits the calpain active site. Absence of the small subunit means that calpastatin would lose one of its anchorage points on calpain 3 and, therefore, bind less effectively and be less likely to inhibit its activity. However, as mentioned in chapter 3 it is theoretically possible to inhibit one calpain 3 by using two calpastatin inhibitory domains. In this scenario, only sub-domains A and B would be bound the enzyme [65]. It is not clear if this theoretical deployment of calpastatin would bind tightly enough to its target to inhibit the enzyme.
Figure 6.1. Calpain PEF domains – Sequence alignment and dimerization structures

A) A multiple sequence alignment of penta-EF-hand domains of calpains 1, 2, 3, 9 and the 28 kDa small subunit is shown below. Amino acid residues are colored blue to red based on increasing identity score. The blue color corresponds to residues with lowest identity and red corresponds to residues with highest identity. Amino acids identified in the dimerization of domain VI (small subunit) crystal structures of rat, porcine and human are indicated by green arrows. CLC sequence viewer (CLC BIO) was used to generate the alignment.

B) PEF domain dimer structures of (i) domain IV of m-calpain with the small subunit [1KFU] (ii) homodimer of rat small subunit [1DVI] and (iii) homodimer model of calpain 3 domain IV. Dimerization occurs through EF hand 5.
6.3 Ca\(^{2+}\) binding properties of calpain isoforms

As detailed earlier, calpains are remarkable for their Ca\(^{2+}\) binding properties. The protease core is capable of binding two Ca\(^{2+}\) ions. This is true of all the protease cores characterized to date (calpain 1, 2, 8, and 9) [4, 152, 153]. Beyond the protease core, the number of Ca\(^{2+}\) ions bound to other regulatory domains varies among calpain isoforms. The regulatory domains are potentially the first sensors of Ca\(^{2+}\) in calpains, which then remove those restrictions on the protease core that prevent alignment of the active site cleft.

6.3.1 Ca\(^{2+}\) binding in PEF domains

EF-hand domains are helix-loop-helix regions capable of coordinating Ca\(^{2+}\). Theoretically, each EF-hand is capable of binding Ca\(^{2+}\). In calpains 1 and 2 both PEF domains (DIV and small subunit) are capable of binding up to 4 Ca\(^{2+}\) (EF1-EF4). EF5, does not bind Ca\(^{2+}\), and instead interacts with the corresponding EF-hand in a second molecule resulting in dimer formation. The small subunit crystal structure showed that EF1-EF3 bind Ca\(^{2+}\) at 1mM where as EF4 required nearly 5mM Ca\(^{2+}\) [46]. Sequence analysis of PEF domains in different calpain isoforms shows that most of them have conserved EF hands resembling the small subunit except calpain 12 and 13. EF3 and EF4
of calpain 12 and 13 do not contain a consensus EF-hand motif but might bind Ca$^{2+}$ in an atypical manner. The crystal structure of calpain 13 domain IV (chapter 4) showed binding of 1 Ca$^{2+}$ in EF2 even with Ca$^{2+}$ absent from the crystallization conditions. The EF1 was truncated for expression and crystallization purposes and it has yet to be determined whether other EF-hands (EF3 and EF4) in C13DIV bind Ca$^{2+}$. Using Isothermal calorimetry (ITC) and mass spectrometry we attempted to determine the Ca$^{2+}$ binding stoichiometry of the small subunit and calpain 1, 2, 3 and 9 PEF domains (results not shown). But interpreting the data was too complex because of the dimerization. In parallel with crystallography, it should be possible to probe the calcium-binding properties of these calpain PEF domains using intrinsic tryptophan fluorescence and $^1$H-$^1$N NMR spectroscopy.

### 6.3.2 Ca$^{2+}$ binding in C2-like domains

C2-like domains in calpains differ in sequence, but share structural homology with other C2-like domains or the canonical C2 domains described in homologs of PKC. The classical C2 domains exhibit Ca$^{2+}$ binding and phospholipid binding. So it was assumed that all C2 domains contain Ca$^{2+}$-binding motifs. But with the increasing number of C2-like domain structures available in the structural database it is now clear that Ca$^{2+}$ and phospholipid binding is not obligatory, and C2-like domains can function as diverse protein modules.
Domain III of the classical calpains was thought to bind Ca\(^{2+}\) [105]. But the recent structure of calpastatin (CAST4) bound to calpain in the presence of Ca\(^{2+}\) showed no Ca\(^{2+}\) coordinated in the domain III region[50]. Also the distal C2-like domain of calpain 7 did not bind Ca\(^{2+}\) although the crystallization conditions contained Ca\(^{2+}\). These results point to deviations in C2-like domains from the classical model suggesting unique functional roles. Now that we are able to express isolated domain constructs of C2-like domains, they can be also tested for Ca\(^{2+}\) binding and phospholipids binding using 1H-\(^{15}\)N NMR spectroscopy.

6.4 Summary and Conclusions

1. A recombinant penta-EF-hand domain construct of calpain 3 (C3DIV) formed a stable homodimer when produced alone. Modelling studies showed full-length calpain 3 can dimerize via EF5 pairing between the PEF domains without any obstacles. Dimerization would place the two protease cores of the enzyme at opposite ends of the dimer.

2. A structural genomics approach using high-throughput cloning and batch expression in \textit{E. coli} was used to examine regulatory domains of human calpain isoforms. We had success in producing regulatory domains corresponding to the PEF domains of calpains 1, 3, 9 and 13 and the C2-like domains of calpains 7, 10 and 13.
3. A differential-tagging system using a 6-His-tag and type III antifreeze protein was developed for screening calpains for dimerization. The PEF domains of calpain 1, 3, 8, 9, 11, 12, and 13 containing an N-terminal His-tag were co-expressed with the small subunit AFP-fusion protein (28 kDa). Calpain 1, 3, 9 and 13 co-expressed as recombinant products and others failed.

4. The dimerization study using Ni-NTA chromatography showed the PEF domains of calpain 1(C1DIV) and 9(C9DIV) formed heterodimers. Conversely, the PEF domain of calpain 3 formed a homodimer. Calpain 13 was predominantly found as a homodimer with a small amount of heterodimer.

5. The 28-kDa small subunit fusion protein containing type III antifreeze protein (7 kDa) at its N-terminus was a homodimer when produced alone and was purified by ice affinity purification (results not shown). However, when the co-expressed product containing fusion protein and PEF domain of calpain 9 (C9DIV) was purified by ice-affinity, overgrowth by the ice resulted in partial dissociation of the heterodimer.

6. Recombinant calpain 6 protease core (CC6) was enzymatically inactive, as predicted from the literature. Mutating the inactivating lysine residue at active site back to cysteine (K81C) did not restore its activity (chapter 5).

7. The structure of the distal C2-like domain of calpain 7 has been solved by X-ray crystallography. It shows striking deviation from canonical C2 domains. It does not bind Ca$^{2+}$ even though Ca$^{2+}$ was present in the crystallization condition. The
structure suggests C2-like domains in calpain isoforms might not bind Ca\(^{2+}\) and might, therefore, possess diverse functions (chapter 4).

8. The structure of calpain 13 PEF domain (C13DIV) has been solved in the absence of Ca\(^{2+}\). However, one Ca\(^{2+}\) was seen bound to EF2 and exhibits pentagonal bipyramidal coordination (chapter 4). C13DIV formed a homodimer by crystallographic symmetry and showed a high degree of structural identity with the small subunit structure (1AJ5)
Bibliography (or References)


Guroff, G. (1964) A Neutral, Calcium-Activated Proteinase from the Soluble Fraction of Rat Brain. J Biol Chem. 239, 149-155


36 Sorimachi, H. (2000) [Structure and function of calpain and its homologues]. Seikagaku. 72, 1297-1315
of calcium bound domain VI of calpain at 1.9 A resolution and its role in enzyme assembly, regulation, and inhibitor binding. Nat Struct Biol. 4, 539-547


151


152
