Crystal structure of calpain reveals the structural basis for Ca^{2+} -dependent protease activity and a novel mode of enzyme activation

Christopher M.Hosfield, John S.Elce,
Peter L.Davies and Zongchao Jia¹

Department of Biochemistry, Queen's University and The Protein Engineering Network of Centres of Excellence, Kingston, Ontario, Canada K7L 3N6

¹Corresponding author
e-mail: jia@crystal.biochem.queensu.ca

The combination of thiol protease activity and calmodulin-like EF-hands is a feature unique to the calpains. The regulatory mechanisms governing calpain activity are complex, and the nature of the Ca^{2+} -induced switch between inactive and active forms has remained elusive in the absence of structural information. We describe here the 2.6 Å crystal structure of m-calpain in the Ca^{2+} -free form, which illustrates the structural basis for the inactivity of calpain in the absence of Ca^{2+} . It also reveals an unusual thiol protease fold, which is associated with Ca^{2+} -binding domains through heterodimerization and a C_2 -like β -sandwich domain. Strikingly, the structure shows that the catalytic triad is not assembled, indicating that Ca^{2+} -binding must induce conformational changes that re-orient the protease domains to form a functional active site. The α -helical N-terminal anchor of the catalytic subunit does not occupy the active site but inhibits its assembly and regulates Ca^{2+} -sensitivity through association with the regulatory subunit. This Ca^{2+} -dependent activation mechanism is clearly distinct from those of classical proteases.

Keywords: calcium/calpain/cysteine proteases/protease activation/protease structure

Introduction

Calpains are the only known mammalian enzymes that combine protease activity with a dependence on Ca^{2+} -binding to EF-hands in one molecule. The μ - and m-calpains are cytosolic cysteine proteases that are ubiquitously expressed and differ in their sensitivity to Ca^{2+} . They consist of an isoform-specific catalytic ~80 kDa subunit and a common regulatory ~28 kDa subunit. While the exact physiological roles of calpains remain to be elucidated, their functional characteristics and wide distribution suggest that they have important cellular roles. The important functions to which calpains are commonly assumed to contribute include signal transduction, apoptosis, cell cycle regulation and cytoskeletal reorganization (Molinari and Carafoli, 1997; Sorimachi *et al.*, 1997; Carafoli and Molinari, 1998; Ono *et al.*, 1998). Calpains have been called biomodulators because their physiological activity involves cleavage of substrates at inter-domain boundaries, serving to modulate the function

of these substrates, rather than simply digesting them (Suzuki and Sorimachi, 1998). Several substrates have been identified both *in vitro* and *in vivo*, including p53, protein kinase C, spectrin, Ca^{2+} -ATPase, talin and fibronectin (Kubbutat and Vouden, 1997; Carafoli and Molinari, 1998; Dourdin *et al.*, 1999; Harada *et al.*, 1999). In recent years, many tissue-specific calpain isoforms have been discovered, in most cases only as cDNA sequences, and not yet as proteins (Sorimachi *et al.*, 1997; Matena *et al.*, 1998; Franz *et al.*, 1999). Among these isoforms, calpain-3, or p94, in particular, has drawn significant attention since the discovery that defects in the p94 gene causing loss of p94 activity lead to the development of limb-girdle muscular dystrophy 2A (Richard *et al.*, 1995; Fougousse *et al.*, 1998).

In addition to its physiological functions, calpain has been implicated in several pathological conditions, including cerebral and cardiac ischemia, cataract formation, and Alzheimer's disease (Wang and Yuen, 1997). These conditions, typified by altered Ca^{2+} -homeostasis resulting from some sort of cellular stress, appear to result in disrupted regulation and excessive activation of calpain. Consequently, calpain is a potentially important therapeutic target, since the development of specific inhibitors could be valuable in the treatment of these diseases.

With the exception of calpain, all known cellular proteolytic enzymes are synthesized as zymogens, or inactive precursors, to prevent inappropriate or premature degradation of substrates (Wang and Yuen, 1997; Khan and James, 1998). Zymogen conversion to the active enzyme can occur by several mechanisms, depending on the protease family, and can be initiated by various factors such as change in pH. In trypsin-like enzymes, activation involves very limited cleavages, which permit critical adjustments in the active site. The activation of profactor D, which is otherwise a normal serine protease, follows a unique pathway, involving re-orientation of residues to form a self-inhibited mature enzyme (Jing *et al.*, 1999). In all known cysteine proteases, activation involves autolytic removal of a pro-peptide to generate the mature protease, and the inactivity of the zymogen is due to the occupation of the active site by the pro-peptide, providing a steric block for other substrates (Khan and James, 1998). The active sites of almost all zymogens and their related active proteases are otherwise virtually indistinguishable, although one example has recently been described in which a significant degree of active site assembly contributes to activation. In proplasmepsin II, the pro-segment does not block the 'immature' active site (Bernstein *et al.*, 1999), and removal of the pro-segment promotes active site formation. The cellular mechanisms for regulation of calpain activity are far more complex. While calpain appears to have some zymogen-like characteristics, it clearly does not function as a typical zymogen and thus is generally not referred to in this manner.

The importance of calpain as a Ca^{2+} -dependent biomodulator is illustrated by the extensive array of mechanisms designed to regulate calpain activity *in vivo*. In addition to the absolute requirement for Ca^{2+} to produce a functional protease, calpain activity is carefully controlled by several other mechanisms. It is known that the concentrations of Ca^{2+} required for activity *in vitro* (~10–50 μM in μ -calpain and ~300–500 μM in m -calpain) are considerably higher than physiological Ca^{2+} concentrations, which are generally <1 μM (Goll *et al.*, 1992). This suggests a requirement for endogenous activators, which have indeed recently been reported (Melloni *et al.*, 1998). In contrast, the calpain inhibitor calpastatin found in all cells acts in an opposing manner and specifically blocks calpain activity. Since many preferred calpain substrates are associated with the cytoskeletal matrix, Ca^{2+} -dependent translocation to the cell membrane is often suggested as a mechanism to allow escape from inhibition by calpastatin, while also utilizing cellular localization as one factor affecting substrate specificity (Mellgren, 1987; Molinari and Carafoli, 1997). A further regulatory factor is autolytic cleavage, by which a short N-terminal peptide is cleaved from the large subunits of μ - and m -calpain very rapidly after addition of Ca^{2+} . The details are still not entirely clear, but removal of the prosegment is apparently not a strict requirement for activity (Molinari *et al.*, 1994), nor is its removal alone sufficient to activate the protease (Elce *et al.*, 1997a), as is the case with all known zymogens (Khan and James, 1998). It is well established that cleavage after residue nine (and later possibly after residue 19) in the catalytic subunit of m -calpain greatly increases the *in vitro* sensitivity of the enzyme towards Ca^{2+} (Goll *et al.*, 1992). Autolytic cleavage of the N-terminal 86 residues that constitute D-V of the regulatory subunit is yet another process accompanying activation, although it does not by itself affect the Ca^{2+} requirement, and its significance is not well understood (Elce *et al.*, 1997b). Finally, an interesting regulatory mechanism involving dissociation of the regulatory subunit has been proposed (Yoshizawa *et al.*, 1995a,b; Suzuki and Sorimachi, 1998). In this model, the calpain subunits dissociate on exposure to Ca^{2+} , thereby liberating a catalytic subunit with increased Ca^{2+} -sensitivity and a capacity for membrane association. There remains some uncertainty about this model, since it was not supported by various studies involving immunoprecipitation and affinity chromatography of natural and mutant calpains in the presence of Ca^{2+} (Zhang and Mellgren, 1996; Dutt *et al.*, 1998). However, the question should be considered as remaining open, because of the great technical difficulties arising from the characteristic autolysis and aggregation of calpain in the presence of Ca^{2+} . Clearly, the mechanisms governing activation and inhibition of calpain *in vivo* are complex, and further experiments will be required for a full understanding of calpain regulation.

To investigate these mechanisms, including that of direct Ca^{2+} -dependent activation, and to assist in the development of therapeutic drugs, we have determined the structure of m -calpain at 2.6 Å resolution in the absence of Ca^{2+} , using the method of multi-wavelength anomalous dispersion (MAD).

Results and discussion

Overall structure

The molecule is an elongated, multi-domain assembly, with dimensions of ~100 × 60 × 50 Å, which is shown both as a stereo view of the α -carbon backbone (Figure 1A) and as a ribbon diagram (Figure 1B). The large subunit consists of four distinct domains (D-I to D-IV) in addition to an α -helical N-terminal anchor of 19 residues. This new domain classification (Figure 1C), derived from the crystal structure, differs in some details from that previously defined on the basis of amino acid sequence (Ohno *et al.*, 1984). As in other cysteine proteases, the catalytic triad residues (Cys105 in D-I, His262 and Asn286 in D-II) are located at the interface between D-I and D-II (Berti and Storer, 1995; Groves *et al.*, 1998). D-I (residues 20–210) has a novel fold consisting of a central helix flanked on three faces by a cluster of α -helices, and two anti-parallel β -sheets. Apart from the α -helix containing Cys105, which is slightly shorter than the corresponding helix in members of the papain family, D-I is otherwise entirely unrelated to the corresponding domain in the typical thiol protease, since it is larger and shows no overall structural similarity (Kamphuis *et al.*, 1984). The sequence and structural differences shown by D-I when compared with other thiol proteases may contribute to the proteolytic specificity characteristic of calpain, which usually cleaves its substrates at a limited number of sites between domains to form fragments with modified function, rather than causing complete degradation (Sorimachi *et al.*, 1997). The fold of D-II is similar to that of the corresponding domain in other cysteine proteases (Kamphuis *et al.*, 1984; Groves *et al.*, 1998) and contains two three-stranded anti-parallel β -sheets. These strands orient His262 and Asn286 of the catalytic triad towards the interface of D-I and D-II. In addition, a three-turn α -helix in D-II makes several contacts with D-III, which is an eight-stranded anti-parallel β -sandwich that shares structural characteristics with the C₂-domain (Sutton *et al.*, 1995; Rizo and Sudhof, 1998) (see below). D-III, which makes contacts with each domain in the enzyme, leads into an ~15 residue extended linker that connects D-III to the Ca^{2+} -binding D-IV. The linker lacks secondary structure, with the exception of three residues (516–518) that form a short anti-parallel β -sheet with three residues (636–638) from D-IV. D-IV and D-VI, the calmodulin-like domains of each subunit, are ~50% identical to each other in sequence and are predominantly α -helical, each containing five EF-hand motifs. Their structures are very similar (r.m.s.d. of 1.7 Å on main chain atoms) and have pseudo-2-fold symmetry. Heterodimerization of the catalytic and regulatory subunits occurs primarily through hydrophobic interactions in the C-terminal regions of D-IV and D-VI, as predicted from the known crystal structure of D-VI (which is a homodimer) (Blanchard *et al.*, 1997; Lin *et al.*, 1997). In the homodimer of D-VI, it has been shown that EF-hands 1–4 bind Ca^{2+} while EF-hand 5 does not bind Ca^{2+} (Blanchard *et al.*, 1997; Lin *et al.*, 1997). The structure of D-VI in the heterodimer described here is virtually identical to its structure in the homodimer (r.m.s.d. of 1.29 Å on main chain atoms).

In the absence of Ca^{2+} , m -calpain also contains several

Fig. 1. (A) Stereodiagram of the C α trace for m-calpain. Residues labeled with an A represent catalytic subunit residues, while those with a B represent regulatory subunit residues. (B) Ribbon diagram of m-calpain. The active site cleft exists at the interface of D-I and D-II. Dotted lines represent highly disordered regions that were omitted. (C) Schematic diagram illustrating the domain organization of m-calpain. The 80 kDa subunit 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). The color scheme is the same as in (A) and (B). Catalytic triad residues are indicated in red, and approximate domain boundaries are indicated by residue number. Figures 1A and B, 2, 3A and 4 were prepared with the programs MOLSCRIPT (Kraulis, 1991) and RASTER3D (Merritt and Bacon, 1997).

disordered regions, which is a characteristic shared by other inactive proteases. Disordered regions in proteases generally play an important role in maintaining the enzyme in an inactive conformation, and undergo significant conformational changes and rigidification during the activation process (Jing *et al.*, 1999).

Structural basis for calpain inactivity in the absence of Ca²⁺

A fundamental question in calpain regulation is how this enzyme is maintained in an inactive conformation prior to Ca²⁺-binding. Our structure reveals a regulatory mechanism that is highly unusual in proteases and is

unprecedented within the cysteine protease family: while the α -helix containing Cys105 in D-I, and the fold of D-II, are each similar to those of other thiol proteases, the active site is not assembled. Numerous crystal structures of cysteine proteases have clearly shown that the catalytic Cys and His residues form an ion-pair in the active site. Specifically, the inter-atomic distance between the S atom of Cys and the N δ atom of His is ~ 3.7 Å (Kamphuis *et al.*, 1984). In this orientation, the His-N δ is at an appropriate distance to coordinate the hydrogen atom bonded to the Cys-S, significantly decreasing the pK_a of the sulfur, increasing its negative charge, and rendering it nucleophilic. Our structure reveals that in calpain, the catalytic Cys105-S in D-I is ~ 10.5 Å away from His262-N δ and therefore is too remote to form a competent catalytic triad with its counterparts His262 and Asn286 in D-II (Figure 2A). Accordingly, the Ca^{2+} -induced conformational change must reduce this distance to ~ 3.7 Å in order to assemble the triad and form an active protease. In the absence of a Ca^{2+} -bound structure, the details of the conformational change remain speculative, but it is clear that the catalytic triad must be assembled into the required geometry for activity. In a simple molecular model, such a conformational change is readily achieved by small rotations of D-I and D-II relative to each other ($\sim 5^\circ$), together with a small translation of ~ 1 – 2 Å. These observations provide the first example of a cysteine protease catalytic triad that undergoes a substantial conformational rearrangement during enzyme activation. This finding suggests an attractive strategy for the design of inhibitory drugs specific for calpain, as it may prove possible to synthesize inhibitors that recognize the unique pre-active site, thereby preventing calpain from attaining an active conformation.

Structural features that contribute to inactivity: anchor-regulatory subunit interactions and the transducer

It is apparent from the structure that D-I and D-II, which house the catalytic triad residues, are held apart and have their movement restricted by both N- and C-terminal constraints. The N-terminal anchor (residues 2–19 of the catalytic subunit) is an α -helix that tethers D-I to D-VI in the regulatory subunit (Figure 3A), and plays a key role in the activation process. This helix makes several contacts with a hydrophobic pocket in D-VI, and has a strong helical dipole–dipole interaction with a helix in D-VI (Figure 3B and C). Most interestingly, this anchor is autolyzed upon Ca^{2+} activation, implying that exposure of the N-terminal segment is required prior to cleavage. Activation may result in loss of the contacts between the anchor and D-VI through Ca^{2+} -induced conformational changes, with subsequent release of the anchor from the regulatory subunit. Alternatively, the anchor could be released through dissociation of the two subunits upon Ca^{2+} -binding (Suzuki and Sorimachi, 1998). Release of the anchor would increase the conformational freedom of D-I, and allow it to move towards D-II. Autolysis of the exposed anchor at Ala9–Lys10 and subsequently at Gly19–Ser20 is not strictly required for activity, yet in practice occurs rapidly during activation (Molinari *et al.*, 1994; Elce *et al.*, 1997a). The crystal structure shows that these cleavage sites are too far away from the active site (~ 40 Å)

to be autolyzed by an intramolecular mechanism, so that the cleavage must be intermolecular. In natural calpains, D-V of the regulatory subunit is also autolyzed, possibly in several steps. The details of regulatory subunit autolysis cannot be determined from this structure, but the final step must again be intermolecular, since this cleavage site, which is close to the N-terminus of the recombinant regulatory subunit used here, is also ~ 40 Å from the active site. While autolyzed calpain has a much lower Ca^{2+} -requirement, indicating that an inhibitory activity associated with the anchor has been lost, it is important to note that the enzyme is still absolutely dependent on Ca^{2+} . Calpain is therefore similar to a ‘zymogen’ to the extent that its anchor is analogous to a pro-segment, but differs from classical zymogens in that removal of the anchor is not by itself sufficient for protease activation.

The fact that calpain remains Ca^{2+} -dependent following subunit autolysis indicates that additional conformational changes are required to assemble the active site. Assuming that D-I is free to move following release of the anchor, it follows that movement of D-II is also required to form the competent catalytic triad. Based on the structure, an appealing mechanism is suggested by the linker of ~ 15 residues that directly connects the Ca^{2+} -binding D-IV to D-III. It appears that this extended segment is acting as an intramolecular signal transducer, communicating a conformational change resulting from Ca^{2+} -binding in D-IV, through D-III, to the protease domains. D-III is able to mediate these conformational changes because it makes several contacts with each domain in the protein, including a significant interaction at the interface with the protease D-II. At this interface, there are several electrostatic interactions involving Lys226, Lys230 and Lys234, which reside on an amphipathic helix in D-II, and several acidic residues including Asp395, Asp398, Glu399 and Glu504 on loops in D-III. These strong interactions suggest that both D-III and D-II may move in unison in assembling the active site, but the nature of these movements is presently unknown.

This Ca^{2+} -dependent activation mechanism is further supported by crystal structure analyses of the isolated regulatory subunit (Blanchard *et al.*, 1997; Lin *et al.*, 1997). Comparison of the crystal structures of the homodimeric D-VI in Ca^{2+} -bound and Ca^{2+} -free states showed that the N-terminal helix of D-VI undergoes a small but critical conformational change upon Ca^{2+} -binding (Figure 2B). Considering the high degree of structural similarity between D-IV and D-VI, it is probable that the corresponding helix in D-IV undergoes a similar conformational change upon Ca^{2+} -binding. This mobile D-IV helix is directly connected to the 15-residue linker or transducer, which in the absence of Ca^{2+} may prevent movement of D-III and D-II. On addition of Ca^{2+} , conformational changes in D-IV would be transmitted through the transducer to permit movement of D-III and D-II towards D-I, generating a functional catalytic triad (Figure 2C), the fundamental requirement for activation.

Comparison with other cysteine proteases

The protease module of calpain is both structurally and functionally distinct from those of most known cysteine proteases. Primary sequence comparisons show that there is little amino acid identity except in the immediate vicinity

Fig. 2. (A) Structural basis for the inactivity of calpain. Stereodigram of the active site of apo- Ca^{2+} -calpain (colors as in Figure 1) superimposed with the active site of papain (red, PDB accession code 9PAP) (Kamphuis *et al.*, 1984). His262 and Asn286 of calpain are arranged in an orientation similar to that in papain. Inactivity of calpain in the absence of Ca^{2+} is due to the catalytic Cys105 being displaced by ~ 7 Å too far from His262 [the distance between His- N_δ and Cys-S is ~ 10.5 Å in calpain and is ~ 3.7 Å in active cysteine proteases (Kamphuis *et al.*, 1984)]. (B) Comparison of the D-IV–D-VI heterodimer with the crystal structures of the Ca^{2+} -free and Ca^{2+} -bound D-VI homodimer. D-VI and D-IV (colors as in Figure 1) are structurally similar to the Ca^{2+} -free form of the homodimer (blue, PDB accession code 1AJ5) (Blanchard *et al.*, 1997). Ca^{2+} -binding results in a conformational change in the N-terminal helices, as indicated by red arrows (PDB accession code 1DVI) (Blanchard *et al.*, 1997). (C) Proposed activation mechanism. Ca^{2+} -binding causes conformational changes in D-IV, which may allow the transducer to release constraints on other domains, and affords increased flexibility to D-II. This may occur by change in the interactions between basic residues (blue) in D-II and acidic residues (red) in D-III, or by movement of D-II and D-III together. Concomitantly, Ca^{2+} -binding to D-VI causes the release of the anchor, yielding a more flexible D-I. Release of the conformational restraints imposed on D-I and D-II would permit formation of the active site at the interface of D-I and D-II through rotations of D-I and D-II relative to each other. Interestingly, Trp 288 (gold) may act as a wedge that also prevents D-I from associating with D-II.

of the active site (Berti and Storer, 1995), suggesting that divergent evolution has played a role in modifying features of the substrate binding site, other than the actual catalytic triad, that contribute to specificity. The crystal structure has revealed several significant differences from those of

papain or the cathepsins. One notable difference is clearly the greater size and lack of similarity of D-I when compared with the corresponding domain in typical cysteine proteases. This domain (D-I, ~ 20 kDa) is nearly as large as the entire papain molecule and is similar to papain

Fig. 3. Calpain has a unique N-terminal anchor. (A) The helical anchor (residues 2–16 are shown) makes contacts only with D-VI (colors as in Figure 1). (B) View down the helical axis highlights interactions between the residues in the anchor (magenta type) and D-VI (black type), represented as an electrostatic GRASP surface (Nicholls *et al.*, 1991) (red, acidic; blue, basic). (C) Side view of (B) illustrates the depth of the hydrophobic pocket in D-VI, which interacts with hydrophobic residues Ala2, Gly3, Ile4, Ala5, Leu8 and Ala9 of the anchor. This anchor inhibits active site assembly by associating with the regulatory subunit, thus restricting flexibility of protease D-I. The anchor also acts as a co-chaperone in concert with D-VI, ensuring proper folding of the catalytic subunit. (B) and (C) were created with the program GRASP (Nicholls *et al.*, 1991).

only in regions defining the active site and substrate-binding cleft. D-II, while possessing a fold similar to that of papain, is also larger and more flexible, at least in the inactive conformation. There are ~50 residues in D-II that are disordered or not visible in the electron density maps. Flexibility has been observed previously in zymogens, and is thought to contribute to their inactivity (Jing *et al.*, 1999), although this mechanism is not prevalent in the cysteine protease family (McGrath, 1999). In calpain, flexibility may result from the fact that despite having 17 cysteine residues, calpain contains no disulfide bonds, stabilizing factors present in most cysteine proteases.

The most remarkable difference between calpain and the typical cysteine proteases is the structure and regulatory role of the N-terminal regulatory anchor, as mentioned

earlier. In the four known crystal structures of cysteine protease zymogens (procaricain, procathepsin L, procathepsin B and procathepsin K) (Coulombe *et al.*, 1996; Cygler *et al.*, 1996; Groves *et al.*, 1996; Sivaraman *et al.*, 1999), it has been demonstrated that the function of the pro-segment is to provide a steric block of a preformed active site. Specifically, the pro-segment binds to the active site cleft in a reverse orientation compared with native substrates, providing an effective mode of inhibition. Thus, there is an absolute requirement for removal of the pro-segment to provide access for substrates to the active site.

In order to draw comparisons between the active site of calpain and those in other cysteine proteases, we have attempted to model the active form of calpain using other

Fig. 4. Domain III shares similar characteristics with a C_2 domain. A typical C_2 domain exists as an anti-parallel β -sandwich with several acidic residues at one end that form a binding cradle for Ca^{2+} . The first C_2 domain from synaptotagmin (cyan, PDB accession code 1RSY) (Sutton *et al.*, 1995) and D-III (green) have approximately the same overall dimensions, though slightly differing topologies. Numerous acidic residues (red) result in a highly negative potential, which is partially stabilized by adjacent basic residues (blue).

cysteine protease structures as a guide. Since the active site conformation in known thiol protease structures is highly conserved, we are confident that the modeled structure accurately represents the active site of Ca^{2+} -activated calpain. The only major steric barrier to assembling the active conformation was the aromatic side chain of the conserved Trp288 of D-II (Figure 2C), which clashes with D-I upon formation of the active site. In cysteine proteases, the side chain of this conserved Trp residue has a weak interaction with the His residue of the catalytic triad, and helps to maintain the His orientation. Mutation of Trp288 to Tyr reduced the activity of calpain to 5% of the wild-type value, a result that was consistent with the role of this Trp residue in other cysteine proteases (Arthur *et al.*, 1995). This observation strongly supports our modeled active site and is in agreement with the observed spatial geometry of the catalytic triad and Trp288. In the absence of Ca^{2+} , it appears that Pro287 (which is a Ser in other cysteine proteases) forces Trp288 into a non-papain-like conformation, so that Trp288 may act as a 'wedge', which would help to prevent the cleft between D-I and D-II from closing. In the presence of Ca^{2+} , Trp288 must rotate to the conformation normally seen in other thiol proteases in order to allow active site formation. Finally, Gln99, which is expected to contribute to the oxyanion hole of calpain (Arthur and Elce, 1996), is not in a mature orientation, and the side chain of Gln99 must also rotate into the appropriate conformation upon activation.

Despite the similarity of the catalytic triad in the active conformation to that expected from papain, the nature of the substrate specificity of calpain differs from that of conventional papain-like proteases. Calpain shows a weak preference for Leu in the P2 position, but does not otherwise possess a strong sequence specificity for substrate cleavage, and cleaves predominantly at exposed

peptide regions between domains in its substrates (Suzuki and Sorimachi, 1998). Comparison of the active site and substrate-binding cleft of conventional cysteine proteases and their inhibitor-bound complexes with the modeled active conformation of calpain has illustrated some features that may contribute to the observed differences in substrate specificity. The protease residues that interact with the substrate backbone are partially conserved, including a Gly-Gly repeat commonly found in cysteine proteases (residues 197 and 198 in calpain, residues 65 and 66 in papain) that stabilizes S2-P2 backbone interactions through hydrogen bonding (McGrath, 1999). In calpain, however, interactions with substrate side chains may be significantly reduced in specificity because most of the subsite-forming side chains in calpain are considerably smaller than those in papain or the cathepsins (McGrath, 1999). For example, well-defined S2-S3 substrate-binding residues in papain include Asp158, Tyr67, Pro68 and Val133, while the corresponding residues in calpain are Gly261, Ala199, Thr200 and Gly 239, all significantly smaller. The substrate-binding cleft in calpain therefore appears to be wider and less strict in its binding specificity than in most other thiol proteases.

C2-like domain III and a putative role in membrane targeting

The function of D-III has remained unclear, partly owing to its lack of sequence homology with any known protein. The C_2 -like fold revealed by this crystal structure, however, suggests a role for this domain in membrane-association processes. D-III makes several contacts to the protease domains, and possesses an anti-parallel β -sheet sandwich structure (Figure 4) with similar overall dimensions to that of the C_2 -domain. In typical C_2 -domains, several loops contribute acidic residues to form a binding cradle at one end of the β -sandwich, which is often surrounded

Table I. Crystallographic analysis

Crystal	λ (Å)	$d_{\min}$ (Å)	Observed reflections	Unique reflections	Completeness (%)	$I/\sigma(I)$	R_{merge}^c (%)
SeMet $P2_1$ ^a							
low energy remote	1.069	2.6	80 764	28 841	92.7 (66.1)	25.3	2.7 (15.6)
anomalous peak	0.9795	2.6	155 288	57 860	93.4 (88.7)	21.5	3.1 (21.9)
inflection point	0.9799	2.6	136 982	54 126	88.3 (81.1)	19.7	3.2 (26.0)
high energy remote	0.9252	2.6	153 100	57 581	93.2 (89.0)	21.6	3.2 (21.5)
SeMet $P1$ ^b							
anomalous peak	0.9795	2.6	128 734	74 656	93.7 (92.1)	16.9	2.7 (37.4)
inflection point	0.9799	2.6	67 174	43 860	57.2 (55.1)	18.7	2.0 (23.4)
Refinement statistics							
native $P1$ (25 to 2.6 Å)							
unique/free reflections	34 794/3454						
R -factor ^d / R -free ^e (%)	22.3/29.3						
protein/water atoms	6428/361						
r.m.s.d. bond lengths (Å)	0.0095						
r.m.s.d. bond angles (°)	1.452						

Unit cell dimensions: ^a $P2_1$: $a = 51.85$, $b = 156.7$, $c = 64.43$ Å, $\beta = 95.32^\circ$; ^b $P1$: $a = 65.17$, $b = 79.91$, $c = 81.59$ Å, $\alpha = 108.5^\circ$, $\beta = 103.37^\circ$, $\gamma = 112.95^\circ$.
^c $R_{\text{merge}} = \sum |I(k) - \langle I \rangle| / \sum I(k)$, where $I(k)$ and $\langle I \rangle$ represent the diffraction intensity values of the individual measurements and the corresponding mean values. The summation is over all measurements.
^d R -factor = $\sum ||F_o| - |F_c|| / \sum |F_o|$
^e $R_{\text{free}} = \sum ||F_{o(\text{free})}| - |F_{c(\text{free})}|| / \sum |F_{o(\text{free})}|$
 F_o is the observed structure factor, F_c is the calculated structure factor based on the model.
Values given in parentheses refer to reflections in the outer resolution shell, 2.7–2.6 Å.
All MAD data were collected in inverse beam mode, with the exception of the low energy remote of the $P2_1$ form, which served as the ‘native’ for phasing.
No sigma cutoff was applied to the data. 10% of reflections were excluded from refinement for calculation of R -free.
r.m.s.d., root mean square deviation.

by basic residues (Sutton *et al.*, 1995; Rizo and Sudhof, 1998). In addition to possessing a fold similar to a Ca^{2+} -dependent C_2 -domain, although with a different topology, D-III also has a large number of acidic residues in the loop region, flanked by basic residues. Further, primary sequence analysis indicates that a C_2 -domain exists in the C-terminal region of Tra-3, a calpain homologue found in *Caenorhabditis elegans* (Barnes and Hodgkin, 1996). Many C_2 -domains have been described as Ca^{2+} -dependent lipid-binding domains (Sutton *et al.*, 1995; Rizo and Sudhof, 1998). Since calpain activation *in vivo* is commonly believed to depend in part on Ca^{2+} -dependent interactions with the membrane (Mellgren, 1987), D-III may target the active enzyme to the membrane where it would have access to physiologically relevant substrates associated with the cytoskeletal matrix. It is also possible that Ca^{2+} -induced conformational changes in D-III could enhance its affinity for the cell membrane. Although overall similarity of the fold is observed between calpain D-III and Ca^{2+} -dependent C_2 -domains, whether D-III of calpain actually binds to the membrane needs to be explored.

The anchor-regulatory subunit interaction has a key role in enzyme assembly
In addition to its role in activation and in modulating Ca^{2+} -affinity, the observation that the N-terminal anchor associates exclusively with the regulatory subunit implies that their functions must be inter-related. It has been demonstrated previously that the regulatory subunit has chaperone-like effects on the refolding of the denatured catalytic subunit (Yoshizawa *et al.*, 1995b). Chaperonins

such as GroE are also able to assist in the folding of the denatured catalytic subunit in the presence of ATP, though not as effectively as the regulatory subunit (Yoshizawa *et al.*, 1995b). It is well documented that classical pro-segments often promote folding of pro-enzymes (Baker *et al.*, 1993), and we suggest that the N-terminal anchor has a chaperone-like activity. In accordance with this suggestion is the finding that expression of various N-terminal large subunit-truncated constructs yielded calpains that were stable heterodimers but lacked enzymatic activity (Elce *et al.*, 1997a), presumably due to misfolding. Taken together with the fact that the crystal structure shows a direct interaction between these structural elements, it seems likely that the anchor and the regulatory subunit act as co-chaperones that promote productive folding of the catalytic subunit (Figure 3).

Modeling of μ -calpain
The amino acid sequences of the μ - and m-calpain catalytic subunits are ~60% identical, but the two enzymes have very different Ca^{2+} requirements. Given this sequence homology, we were able to model the μ -calpain structure without much difficulty, but preliminary inspection has revealed no obvious differences that might explain the difference in Ca^{2+} requirement (unpublished data).

Structure-based mechanism of calpain regulation
The fusion of calmodulin-like and papain-like activities suggests a unique mechanism for Ca^{2+} -dependent regulation of calpain, whose significance *in vivo* is indicated by the numerous pathological conditions resulting from excessive or inappropriate calpain activation. On the

basis of the crystal structure, we propose the following mechanism of calpain regulation from the level of enzyme assembly and Ca^{2+} -induced activation, to autolytic processing and cellular localization. As indicated from the structure, the regulatory subunit will facilitate the formation of a properly folded catalytic subunit through interactions with the N-terminal anchor, as well as by the D-IV–D-VI interactions. The assembled heterodimer remains inactive in the absence of Ca^{2+} as a result of a network of conformational restraints that effectively hold the active site apart. Binding of Ca^{2+} to EF-hands induces conformational changes in D-IV and D-VI, and these changes have two immediate results: the release of the catalytic subunit N-terminal anchor from the regulatory subunit, permitting movement of D-I; and movement of the transducer attached to D-IV, permitting movement of D-II and D-III. Consequently, D-I and D-II are able to move together to form a competent active site. Following formation of the active site, intermolecular autolysis of the N-terminal anchor activates calpain by permitting activity at lower Ca^{2+} -concentrations. Subunit dissociation, if it occurs, will affect these events because of the close association of the N-terminal anchor of the catalytic subunit with the regulatory subunit. It has been proposed that subunit dissociation is reversible, as long as catalytic subunit autolysis has not occurred. In this case, calpain activation could be either reversible (prior to catalytic subunit autolysis), or irreversible (following catalytic subunit autolysis), depending on cellular conditions. Finally, the C_2 -like D-III may be responsible for promoting binding of active calpain to the membrane in response to Ca^{2+} , thereby relieving inhibition from calpastatin and promoting digestion of physiologically relevant substrates. The crystal structure of calpain has provided many new insights into these issues and should lay the groundwork for extensive biochemical studies that may allow us to understand more clearly the regulatory mechanisms governing this intriguing enzyme.

Materials and methods

Protein expression, purification and crystallization have been described previously (Elce *et al.*, 1995; Hosfield *et al.*, 1999). Here, m-calpain refers to a recombinant rat (80 + 21 kDa) m-calpain that has a C-terminal histidine-tag and a mutation of the active site Cys105 to Ser in the large subunit. The small subunit was expressed as a truncated 21 kDa form, from which the natural D-V has been omitted. D-V is ~90 residues in length, and contains ~30% glycine residues (Sorimachi *et al.*, 1996). This domain is partially degraded when expressed in *Escherichia coli* and purified, giving rise to heterogeneous preparations of calpain (Elce *et al.*, 1997b), which was expected to hinder crystallization. Attempts to crystallize calpain in the presence of Ca^{2+} were unsuccessful as all attempts led to massive aggregation. Alternatively, we attempted to soak native crystals in solutions containing Ca^{2+} ; however, addition of Ca^{2+} immediately shattered the crystals. Data were also collected and the structure was determined for enzymatically active calpain containing the natural Cys105. The data were at a lower resolution but the structure was fully consistent with the model of the Ser105 mutant.

Selenomethionine-containing crystals of the Cys105Ser form in the absence of Ca^{2+} were grown in $P1$ and $P2_1$ forms, and MAD data were collected for both space groups using an ADSC Quantum IV-CCD at the Stanford Synchrotron Radiation Laboratory (SSRL) beamline 1–5. Diffraction data were processed with the HKL program suite (Otwinowski and Minor, 1997) (Table I) and the CCP4 program suite (Collaborative Computational Project, No. 4, 1994), and 17 of 19 selenium positions were determined by direct methods with the DREAR/SnB package (Weeks and Miller, 1999). Heavy-atom refinement and phasing with

SHARP (De la Fortelle and Bricogne, 1997) gave interpretable electron density maps (figure of merit 0.63 and 0.38 for $P2_1$ and $P1$ crystals, respectively) that were substantially improved upon solvent flattening with SOLOMON (Abrahams and Leslie, 1996). The model was traced with XFIT (McRee, 1992) independently in both space groups to reduce the possibility of tracing mistakes in disordered regions and was subjected to iterative cycles of manual fitting and maximum-likelihood refinement as implemented in the CNS package (Brünger *et al.*, 1998). The final model (refined against data collected previously at the Cornell High Energy Synchrotron Source on the $P1$ native crystal form) (Hosfield *et al.*, 1999) is well defined with the exception of the following disordered regions that were omitted: 245–260, 273–278, 292–321, 437–459 and 565–566 (indicated by dotted lines in Figure 1B). The quality of the model was assessed with PROCHECK (Laskowski *et al.*, 1993), and displays good stereochemistry with 83.2% of residues in the most favorable regions and no residues in the disallowed regions of the Ramachandran plot.

Acknowledgements

Special thanks go to D.Hosfield for excellent discussions and support. We thank C.Hegadorn and Q.Ye for skilled technical assistance, T.Moldoveanu for insightful discussions, and H.Bellamy and other staff at SSRL for technical support at beamline 1–5. This work was funded by grants from MRC, PENCE and Warner-Lambert Canada. C.M.H. is a recipient of studentships from NSERC and MRC. Atomic coordinates for m-calpain have been deposited with the RCSB, accession code 1DF0.

References

- Abrahams, J.P. and Leslie, A.G.W. (1996) Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr. D*, **52**, 30–42.
- Arthur, J.S.C. and Elce, J.S. (1996) Interaction of aspartic acid-104 and proline-287 with the active site of m-calpain. *Biochem. J.*, **319**, 535–541.
- Arthur, J.S.C., Gauthier, S. and Elce, J.S. (1995) Active site residues in m-calpain: identification by site-directed mutagenesis. *FEBS Lett.*, **368**, 397–400.
- Baker, D., Shiau, A.K. and Agard, D.A. (1993) The role of pro regions in protein folding. *Curr. Opin. Cell Biol.*, **5**, 966–970.
- Barnes, T.M. and Hodgkin, J. (1996) The tra-3 sex determination gene of *Caenorhabditis elegans* encodes a member of the calpain regulatory protease family. *EMBO J.*, **15**, 4477–4484.
- Bernstein, N.K., Cherney, M.M., Loetscher, H., Ridley, R.G. and James, M.N.G. (1999) Crystal structure of the novel aspartic proteinase zymogen proplasmepsin II from *Plasmodium falciparum*. *Nature Struct. Biol.*, **6**, 32–37.
- Berti, P.J. and Storer, A.C. (1995) Alignment/phylogeny of the papain superfamily of cysteine proteases. *J. Mol. Biol.*, **246**, 273–283.
- Blanchard, H., Grochulski, P., Li, Y., Arthur, J.S.C., Davies, P.L., Elce, J.S. and Cygler, M. (1997) Structure of a calpain Ca^{2+} -binding domain reveals a novel EF-hand and Ca^{2+} -induced conformational changes. *Nature Struct. Biol.*, **4**, 532–538.
- Brünger, A.T. *et al.* (1998) Crystallography and NMR system: a new software system for macromolecular structure determination. *Acta Crystallogr. D*, **54**, 905–921.
- Carafoli, E. and Molinari, M. (1998) Calpain: a protease in search of a function. *Biochem. Biophys. Res. Commun.*, **247**, 193–203.
- Collaborative Computational Project No. 4. (1994) The CCP4 suite. *Acta Crystallogr. D*, **50**, 760–763.
- Coulombe, R., Grochulski, P., Sivaraman, J., Menard, R., Mort, J.S. and Cygler, M. (1996) Structure of human procathepsin L reveals the molecular basis of inhibition by the prosegment. *EMBO J.*, **15**, 5492–5503.
- Cygler, M., Sivaraman, J., Grochulski, P., Coulombe, R., Storer, A.C. and Mort, J.S. (1996) Structure of rat procathepsin B: model for inhibition of cysteine protease activity by the proregion. *Structure*, **4**, 405–416.
- De la Fortelle, E. and Bricogne, G. (1997) Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.*, **276**, 472–494.
- Dourdin, N., Balcerzak, D., Brustis, J.J., Poussard, S., Cottin, P. and Ducastaing, A. (1999) Potential m-calpain substrates during myoblast fusion. *Exp. Cell Res.*, **246**, 433–442.

- Dutt, P., Arthur, J.S.C., Croall, D.E. and Elce, J.S. (1998) m-Calpain subunits remain associated in the presence of calcium. *FEBS Lett.*, **436**, 367–371.
- Elce, J.S., Hegadorn, C., Gauthier, S., Vince, J.W. and Davies, P.L. (1995) Recombinant calpain II: improved expression systems and production of a C105A active-site mutant for crystallography. *Protein Eng.*, **8**, 843–848.
- Elce, J.S., Hegadorn, C. and Arthur, J.S.C. (1997a) Autolysis, Ca^{2+} requirement and heterodimer stability in m-calpain. *J. Biol. Chem.*, **272**, 11268–11275.
- Elce, J.S., Davies, P.L., Hegadorn, C., Maurice, D.H. and Arthur, J.S.C. (1997b) The effects of truncations of the small subunit on m-calpain activity and heterodimer formation. *Biochem. J.*, **326**, 31–38.
- Fougerousse, F., Durand, M., Suel, L., Pourquie, O., Delezoide, A.L., Romero, N.B., Abitbol, M. and Beckmann, J.S. (1998) Expression of genes (CAPN3, SGCA, SGCB and TTN) involved in progressive muscular dystrophies during early human development. *Genomics*, **48**, 145–156.
- Franz, T., Vingron, M., Boehm, T. and Dear, T.N. (1999) Capn7: a highly divergent vertebrate calpain with a novel C-terminal domain. *Mamm. Genome*, **10**, 318–321.
- Goll, D.E., Thompson, V.F., Taylor, R.G. and Zalewska, T. (1992) Is calpain activity regulated by membranes and autolysis or by calcium and calpastatin? *BioEssays*, **14**, 549–556.
- Groves, M.R., Taylor, M.A., Scott, M., Cummings, N.J., Pickersgill, R.W. and Jenkins, J.A. (1996) The prosequence of procaricain forms an α -helical domain that prevents access to the substrate-binding cleft. *Structure*, **4**, 1193–1203.
- Groves, M.R., Coulombe, R., Jenkins, J. and Cygler, M. (1998) Structural basis for specificity of papain-like cysteine protease proregions toward their cognate enzymes. *Proteins*, **32**, 504–514.
- Harada, K., Maekawa, T., Abu Shama, K.M., Yamashita, T. and Yoshida, K. (1999) Translocation and down-regulation of protein kinase C- α , - β and - γ isoforms during ischemia-reperfusion in rat brain. *J. Neurochem.*, **72**, 2556–2564.
- Hosfield, C.M., Ye, Q., Arthur, J.S.C., Hegadorn, C., Croall, D.E., Elce, J.S. and Jia, Z. (1999) Crystallization and X-ray crystallographic analysis of m-calpain: a Ca^{2+} -dependent protease. *Acta Crystallogr. D*, **55**, 1484–1486.
- Jing, H., Macon, K.J., Moore, D., DeLucas, L.J., Volanakis, J.E. and Narayana, S.V.L. (1999) Structural basis of profactor D activation: from a highly flexible zymogen to a novel self-inhibited serine protease, complement factor D. *EMBO J.*, **18**, 804–814.
- Kamphuis, I.G., Kalk, K.H., Swarte, M.B.A. and Drenth, J. (1984) Structure of papain refined at 1.65 Å resolution. *J. Mol. Biol.*, **179**, 233–256.
- Khan, A.R. and James, M.N.G. (1998) Molecular mechanisms for the conversion of zymogens to active proteolytic enzymes. *Protein Sci.*, **7**, 815–836.
- Kraulis, P.J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.
- Kubbutat, M.H. and Vousden, K.H. (1997) Proteolytic cleavage of human p53 by calpain: a potential regulator of protein stability. *Mol. Cell. Biol.*, **17**, 460–468.
- Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993) PROCHECK—a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 283–291.
- Lin, G.D. *et al.* (1997) Crystal structure of calcium bound domain VI of calpain at 1.9 Å resolution and its role in enzyme assembly, regulation and inhibitor binding. *Nature Struct. Biol.*, **4**, 538–547.
- Matena, K., Boehm, T. and Dear, T.N. (1998) Genomic organisation of mouse Capn5 and Capn6 genes confirms that they are a distinct calpain subfamily. *Genomics*, **48**, 117–120.
- McGrath, M.E. (1999) The lysosomal cysteine proteases. *Annu. Rev. Biophys. Biomol. Struct.*, **28**, 181–204.
- McRee, D.E. (1992) A visual protein crystallographic software system for X11/XView. *J. Mol. Graph.*, **10**, 44–46.
- Mellgren, R.L. (1987) Calcium-dependent proteases: an enzyme system active at cellular membranes? *FASEB J.*, **1**, 110–115.
- Melloni, E., Michetti, M., Salamino, F. and Pontremoli, S. (1998) Molecular and functional properties of a calpain activator protein specific for μ -isoforms. *J. Biol. Chem.*, **273**, 12827–12831.
- Merritt, E.A. and Bacon, D.J. (1997) RASTER3D: photorealistic molecular graphics. *Methods Enzymol.*, **277**, 505–524.
- Molinari, M. and Carafoli, E. (1997) Calpain: a cytosolic proteinase active at the membranes. *J. Membr. Biol.*, **156**, 1–8.
- Molinari, M., Anagli, J. and Carafoli, E. (1994) Ca^{2+} -activated neutral protease is active in the erythrocyte membrane in its nonautolyzed 80-kDa form. *J. Biol. Chem.*, **269**, 27992–27995.
- Nicholls, A., Sharp, K. and Honig, B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins*, **11**, 281–296.
- Ohno, S., Emori, Y., Imajoh, S., Kawasaki, H., Kisaragi, M. and Suzuki, K. (1984) Evolutionary origin of a calcium-dependent protease by fusion of genes for a thiol protease and a calcium-binding protein? *Nature*, **312**, 566–570.
- Ono, Y., Sorimachi, H. and Suzuki, K. (1998) Structure and physiology of calpain, an enigmatic protease. *Biochem. Biophys. Res. Commun.*, **245**, 289–294.
- Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.
- Richard, I. *et al.* (1995) Mutations in the proteolytic enzyme calpain 3 cause limb-girdle muscular dystrophy type 2A. *Cell*, **81**, 27–40.
- Rizo, J. and Sudhof, T.C. (1998) C_2 -domains, structure and function of a universal Ca^{2+} -binding domain. *J. Biol. Chem.*, **273**, 15879–15882.
- Sivaraman, J., Lalumiere, M., Menard, R. and Cygler, M. (1999) Crystal structure of wild-type human procathepsin K. *Protein Sci.*, **8**, 283–290.
- Sorimachi, H., Amano, S., Ishiura, S. and Suzuki, K. (1996) Primary sequences of rat μ -calpain large and small subunits are, respectively, moderately and highly similar to those of human. *Biochim. Biophys. Acta*, **1309**, 37–41.
- Sorimachi, H., Ishiura, S. and Suzuki, K. (1997) Structure and physiological function of calpains. *Biochem. J.*, **328**, 721–732.
- Sutton, R.B., Davletov, B.A., Berghuis, A.M., Sudhof, T.C. and Sprang, S.R. (1995) Structure of the first C_2 domain of synaptotagmin I: a novel Ca^{2+} /phospholipid-binding fold. *Cell*, **80**, 929–938.
- Suzuki, K. and Sorimachi, H. (1998) A novel aspect of calpain activation. *FEBS Lett.*, **433**, 1–4.
- Wang, K.K.W. and Yuen, P.-W. (1997) Development and therapeutic potential of calpain inhibitors. *Adv. Pharmacol.*, **37**, 117–152.
- Weeks, C.M. and Miller, R. (1999) Optimizing shake-and-bake for proteins. *Acta Crystallogr. D*, **55**, 492–500.
- Yoshizawa, T., Sorimachi, H., Tomioka, S., Ishiura, S. and Suzuki, K. (1995a) Calpain dissociates into subunits in the presence of calcium ions. *Biochem. Biophys. Res. Commun.*, **208**, 376–383.
- Yoshizawa, T., Sorimachi, H., Tomioka, S., Ishiura, S. and Suzuki, K. (1995b) A catalytic subunit of calpain possesses full proteolytic activity. *FEBS Lett.*, **358**, 101–103.
- Zhang, W. and Mellgren, R.L. (1996) Calpain subunits remain associated during catalysis. *Biochem. Biophys. Res. Commun.*, **227**, 890–896.

Received September 20, 1999; revised and accepted October 27, 1999