Regular paper

Cloning and characterization of the yak gene coding for calpastatin and *in silico* analysis of its putative product

Liping Zhang[1]★✉, Binyun Ma[2]★✉, Jianping Wu[1], Chunhong Fei[1], Lian Yang[1] and Hongling Wan[1]

[1]College of Animal Science and Technology, Gansu Agricultural University, Lanzhou, China; [2]School of Life Sciences, Lanzhou University, Lanzhou, China

The calcium-activated neutral proteases, μ- and m-calpain, along with their inhibitor, calpastatin, have been demonstrated to mediate a variety of Ca^{2+}-dependent processes including signal transduction, cell proliferation, cell cycle progression, differentiation, apoptosis, membrane fusion, platelet activation and skeletal muscle protein degradation. The cDNA coding for yak calpastatin was amplified and cloned by RT-PCR to investigate and characterize the nucleotide/amino-acid sequence and to predict structure and function of the calpastatin. The present study suggests that the yak calpastatin gene encodes a protein of 786 amino acids that shares 99% sequence identity with the amino-acid sequence of cattle calpastatin, and that the yak protein is composed of an N-terminal region (domains L and XL) and four repetitive homologous C-terminal domains (d1–d4), in which several prosite motifs are present including short peptide L54–64 (EVKPKEHTEPK in domain L) and GXXE/DXTIPPXYR (in subdomain B), where X is a variable amino acid. Our results suggest the existence of other functional sites including potential phosphorylation sites for protein kinase C, cAMP- and cGMP-dependent protein kinase, casein kinase II, as well as N-myristoylation and amidation sites that play an important role in molecular regulation of the calpain/calpastatin system. The regulation of the calpain/calpastatin system is determined by the interaction between dIV and dVI in calpains and subdomains A, B, and C in calpastatin.

INTRODUCTION

Calpains (CAPN, EC 3.4.22.17) belong to a family of calcium-dependent cytoplasmic cysteine proteinases that require Ca^{2+} for activity. Two types of CANP, μ-calpain (or CAPN I) and m-calpain (or CAPN II), which require approximately 50 and 300 μM Ca^{2+} for half-maximal activity *in vivo*, respectively, exist widely in mammalian tissues (Hirose *et al.*, 1999; Shiraha *et al.*, 2002; Goll *et al.*, 2003). Calpastatin (CAST) is a specific endogenous inhibitor of μ-calpain and m-calpain, which regulates calpain activity *in vivo* (Carragher *et al.*, 2002; Goll *et al.*, 2003). Whereas only one calpastatin gene has been identified, several forms of the gene product, resulting from differential mRNA splicing and posttranslational modifications such as phosphorylation and proteolysis, have

been detected (Croall & Demartino, 1991; Glading *et al.*, 2002; Weber *et al.*, 2004). Calpastatin is ubiquitously expressed and is translated as two main calpastatin isoforms, including a 110-kDa muscle tissue type and a 70-kDa erythrocyte type (Tullio *et al.*, 2000; Takano *et al.*, 2000). Typical calpastatin molecules have five structural domains: four repetitive domains (d1 to d4) and a unique N-terminal domain (L-domain). Calpain inhibitory activity has been detected in each repetitive domain, of which domain 1 possesses the strongest activity (Maki *et al.*, 1987; Emori *et al.*, 1988; Kawasaki *et al.*, 1989). The repetitive domains contain about 140 amino-acid residues each and show 20–30% sequence identity to each other, while the function of the N-terminal domain is not clear.

In the present study, we cloned the calpastatin gene from yak longissimus muscle, and investigated the structural and functional motifs of the yak calpastatin by bioinformatic methods.

MATERIALS AND METHODS

Animal tissue collection and total RNA isolation. Preharvest animal care and use was under control of local farmers in Tianzhu, Gansu production area, and was consistent with Gansu Agricultural University animal care and use requirements. Animals were harvested at a commercial facility that must comply with state regulations governing processing of meat animals. To clone the CAST genes, yak longissimus muscle tissues were collected from Tianzhu white yak (*Bos grunniens*) from Tianzhu, Gansu (China), within 10 min after slaughter. The samples were flash-frozen by immersion in liquid nitrogen, and stored at −70°C until thawed for RNA extraction.

Total RNA was extracted with Trizol reagent (Gibco BRL, Gaithersburg, MD, USA). Yak longissimus muscle tissue samples were homogenized in 1 ml of Trizol reagent per 50 to 100 mg of tissue using a glass homogenizer. Homogenized samples were incubated for 5 min at 15°C to 30°C to permit the complete dissociation of nucleoprotein complexes. Chloroform (0.2 ml/ml of Trizol) was added, and samples were shaken and incubated at 15°C to 30°C for an additional 2 to 3 min. Samples were

✉Liping Zhang: e-mail: zhangliping@gsau.edu.cn;
★Binyun Ma: e-mail: mabinyun@hotmail.com
★Both authors contributed equally to this work

Table 1. Primers for RT-PCR of CAST gene of yak

Primers sequences (5′ to 3′)	Size of interesting fragments (bp)	Temperature (°C)
P_{CAST}-F1:TGCGTCTCTCGGAACACAT P_{CAST}-R1:CTGATTCACTCAGGGGCTTG	1342	57
P_{CAST}-F2:ACCGGCCACAGATAAAGATG P_{CAST}-R2:AGTCACCATCTACCCGCTGA	1176	60.5

then centrifuged at $12\,000 \times g$ for 15 min at 2 °C to 8 °C. After centrifugation, the dissolved RNA was pipetted to a fresh tube, and the RNA was precipitated with isopropyl alcohol. Samples were then incubated at 15 °C to 30 °C for 10 min and centrifuged at $12\,000 \times g$ for 10 min at 2 °C to 8 °C. The supernatant was removed, and the RNA precipitate was washed with 75% ethanol. The RNA and ethanol were vortexed and centrifuged at $7\,500 \times g$ for 5 min at 2 °C to 8 °C. The RNA was then redissolved in 100% formamide (deionized) and stored at −80 °C.

PCR primers and RT-PCR. The oligonucleotide primers for reverse-transcription PCR (RT-PCR) were designed based on the mRNA sequences of CAST (Genbank Accession No AF159246) of the cattle (*Bos taurus*) published at the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/) (Table 1).

The RT-PCR was carried out using an RNA PCR kit (AMV), Ver. 3.1 (TaKaRa, Shiga, Japan), for which avian myeloblastosis virus reverse transcription was used for the first-strand DNA synthesis, and *Taq* DNA polymerase was used for PCR in a single optimized RT-PCR buffer. First-strand cDNA synthesis was accomplished by RT-PCR at 30 °C for 10 min, 50 °C for 30 min, 99 °C for 5 min, and 5 °C for 5 min. The initial denaturation reaction was performed for 2 min at 94 °C for 1 cycle, and the amplification was carried out for 35 cycles, each comprising 30 s at 94 °C, 1.5 min at the temperatures specified in Table 1, and then 1.5 min at 72 °C, followed by the last elongation reaction at 72 °C for 10 min. Products were separated by electrophoresis on a 1.0% agarose gel. Fragments with the expected size were cut from the gel and purified using the PCR Preps DNA Purification System (Promega, Madison, WI, USA).

Cloning and sequencing of cDNA of the CAST genes. A plasmid library was constructed with T4 polymerase by ligating the amplified DNA fragments into the *Sph*I and *Sal*I sites of the cloning plasmid pGEM®-T Vector (Promega, Shanghai, China), following the manufacturer's procedures. Isolation of library DNA, transfection of competent *Escherichia coli* DH5α, and extraction of DNA from transfected cells were performed according to published methods (Sambrook & Russell, 2001). Recombinant plasmids containing relevant yak DNA fragments were sequenced on an ABI Prism 3100 genetic analyzer (Applied Biosystems, Foster City, CA, USA).

Sequence analysis and molecular characterization of CAST. The nucleotide and AA sequences of yak CAST were subjected to BLAST searching at the National Center for Biotechnology Information. Multiple alignment and comparisons of the nucleotide and amino acid sequences were then performed using BioEdit (Hall, 1999; Version 7.0.5.3) and DNAMAN (Version 5.2.2; Lynnon Biosoft). Protein signal peptides were analyzed using SignalP 3.0 software (Bendtsen *et al.*, 2004; http://www.cbs.dtu.dk/services/SignalP-3.0/). Secondary structures of CAST were predicted by DNAStar Protean (Chou & Fasman, 1978; Chou, 1990; Version

5.01), and the prosite motifs and functional regulating sites of CAST were predicted by PROSITE motif search of the ExPASy Server (http://www.expasy.org/prosite/).

RESULTS AND DISCUSSION

Cloning and molecular characterization of calpastatin

To study the molecular identity of calpastatin from the yak, gene-specific primers were used to amplify the entire cDNA encoding the CAST from yak longissimus muscle by RT-PCR as described in more detail in Materials and Methods (Table 1). The interesting fragments CAST-1 (about 1340 bp) and CAST-2 (about 1170 bp) were obtained by P_{CAST}-F1/P_{CAST}-R1 and P_{CAST}-F2/P_{CAST}-R2, respectively (Fig. 1).

Assembly of the CAST gene fragments generated a full-length CAST cDNA (2445 bp, Genbank Accession No EU009141). In the CAST cDNA complete open reading frame (2361 bp) was present encoding a protein of 786 amino acids 99% identical to that of the cattle CAST gene (Genbank Accession No NM_174003) (Fig. 2), and sharing 99% sequence identity with the amino-acid sequence of cattle calpastatin (Fig. 3).

Functional domains, motifs and sites of CAST

Calpastatin, an unstructured endogenous inhibitor of calpains that lack a well-defined 3D structure (Mucsi *et al.*, 2003), has been demonstrated previously by NMR and circular dichroism spectroscopy studies to be an intrinsically disordered molecule with local preformed transient structural elements which are important for calpain recognition (Uemori *et al.*, 1990; Kiss *et al.*, 2008a; 2008b; Toke *et al.*, 2009). The primary structure of calpastatin is composed of an N-terminal region (domains L and XL) and four repetitive homologous C-terminal domains (d1–d4) (Fig. 3). The function of the L-domain (N-terminal domain) is not clear, whereas the other four homologous domains are all capable of inhibiting a calpain molecule. In the calpastatin of yak, there are several prosite motifs and many functional sites, which may play important roles in a variety of Ca^{2+}-dependent processes

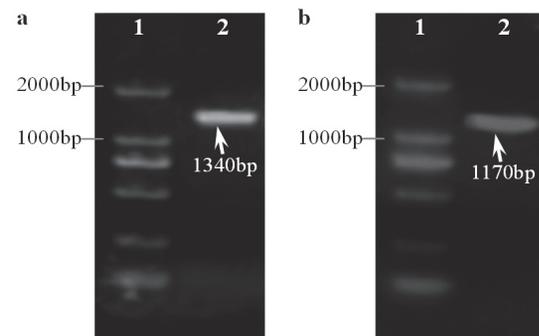**Figure 1. Amplification of CAST gene from yak longissimus muscle total RNA.**
(a) PCR product with primers P_{CAST}-F1 and P_{CAST}-R1, lane 1 shows DNA Marker DL 2000 (TaKaRa, Shiga, Japan), lane 2 shows amplified fragment of 1340 bp (CAST-1); **(b)** PCR product with primers P_{CAST}-F2 and P_{CAST}-R2, lane 1 shows DNA Marker DL 2000, lane 2 shows amplified fragment of 1170 bp (CAST-2).

```
-30                                                          tgcgtctctcggaacacatccatcgtcgcc

   1  ATGGCATTTGCAAGCTGGTGGTACAAGACGCATGTCAGTGGAAAAACCAGTGGATCGCCTTCCAAGTCAGGAGAAAAGAAAGGATCAGAT
   1   M  A  F  A  S  W  W  Y  K  T  H  V  S  G  K  T  S  G  S  P  S  K  S  G  E  K  K  G  S  D

  91  GAGAAAAAAGCAACAAGCCTTGGGAGCAGTCAGCTCTCCAGAACTCAAGCTGGTGAAAAAGCCCCGGTCCCCAAGGTAACTACTTCCTCT
  31   E  K  K  A  T  S  L  G  S  S  Q  L  S  R  T  Q  A  G  E  K  A  P  V  P  K  V  T  T  S  S

 181  GCGTCAGCCAGCAAGTCTTCCAGTATGAATCCCACAGAAGCCAAGGCTATTCCAGGCAGCAAACAGCTGGAAGGACCGCATTCTCCTAAC
  61   A  S  A  S  K  S  S  S  M  N  P  T  E  A  K  A  I  P  G  S  K  Q  L  E  G  P  H  S  P  N

 271  AAGAAAAGACACAAAAAACAGGCTGTAAAAACAGAACCTGAGAAGAAGCCACAATCATCTAAGCCATCTGTGGTTCATGAGAAAAAAACC
  91   K  K  R  H  K  K  Q  A  V  K  T  E  P  E  K  K  P  Q  S  S  K  P  S  V  V  H  E  K  K  T

 361  CAAGAAGTAAAGCCAAAGGAACACACAGAGCCAAAAAGCCTACCCAAGCACTCATCAGATACAGGAGGCAAGCATGCTCCTAAGGAAAAA
 121   Q  E  V  K  P  K  E  H  T  E  P  K  S  L  P  K  H  S  S  D  T  G  S  K  H  A  P  K  E  K

 451  GCCGTTTCCAAATCAAGCGAGCAGCCACCATCAGAGAAATCAACAAAACCAAAGACCAAGTCACAGGACAAGATCTCCGGTGGTGGAAAG
 151   A  V  S  K  S  S  E  Q  P  P  S  E  K  S  T  K  P  K  T  K  S  Q  D  K  I  S  G  G  G  K

 541  AGCGCTGTTCCTGCTGCTGCTGCTGCAGCATCTGCCGAACCAGCTGACAAGAATAAAGAAAATAAATTGTTAACATCGGCCGTACCAGCT
 181   S  A  V  P  A  A  A  A  A  S  A  E  P  A  D  K  N  K  E  N  K  L  L  T  S  A  V  P  A

 631  GAATCCAAACCAAGTAAACCGTCTGGAAAGTCAGACATGGACACTGCTCTGGATGACTTAATAGACACTTTAGGAGAACCTGAAGAGATG
 211   E  S  K  P  S  K  P  S  G  K  S  D  M  D  T  A  L  D  D  L  I  D  T  L  G  E  P  E  E  M

 721  AAAGAAGATAACACAATATATACCGGACCGGAAGTGTCGGATCCAATGAGTTCTACCTACATAGAGGAACTGGGTAAAAGAGAATCCACA
 241   K  E  D  N  T  I  Y  T  G  P  E  V  S  D  P  M  S  S  T  Y  I  E  E  L  G  K  R  E  S  T

 811  CTTCCTCCAAAATATAAGGAACTTCTGAATAAAGAAGAAGGGATCGCAGGGCCTCCTCCAGACTCCTTGAAACCCCTGGGGCCCAATGAT
 271   L  P  P  K  Y  K  E  L  L  N  K  E  E  G  I  A  G  P  P  P  D  S  L  K  P  L  G  P  N  D

 901  GCCATCGATGCCTTGTCATCCGACTTCACCTGCAGTTCCCTACAGCTGATGCAAAGAAAACTGAGAAAGAGAAGTCTACAGAAGAGGCT
 301   A  I  D  A  L  S  S  D  F  T  C  S  S  P  T  A  D  A  K  K  T  E  K  E  K  S  T  E  E  A

 991  TTAAAAGCTCAGTCAGCTGGGGTGATCAGAAGTGCTGCTCCACCCCAAGAGAAAAAAAGGAAAGTGGAAAAGGATGCCATGACTGAGCAC
 331   L  K  A  Q  S  A  G  V  I  R  S  A  A  P  P  Q  E  K  K  R  K  V  E  K  D  A  M  T  E  H

1081  GCCCTGGAGGCCCTGTCAGCCTCCCTGGGCACCCGGAAGCCGGAGCCGGAGCTCGACCCCAGCTCCATTAAGGAGGTCGATGAGGCAAAA
 361   A  L  E  A  L  S  A  S  L  G  T  R  K  P  E  P  E  L  D  P  S  S  I  K  E  V  D  E  A  K

1171  GCCAAAGAAGAGAAAGTAAAGAAAATGTGGTGAAGATGAGGAAACAGTCCCATCGGAGTACAGATTAAAACCGGCCACAGATAAAGATGGA
 391   A  K  E  E  K  V  K  K  C  G  E  D  E  E  T  V  P  S  E  Y  R  L  K  P  A  T  D  K  D  G

1261  AAACCACTCTTGCCAGAGGCTGAAGAAAAACCCAAGCCCTGAGTGAATCAGAACTCATCGATGAACTCTCAGAAGATTTTGACCAGTCT
 421   K  P  L  L  P  E  A  E  E  K  P  K  P  L  S  E  S  E  L  I  D  E  L  S  E  D  F  D  Q  S

1351  AAGTGTAAAGAAAAACAATCTAAGCCAACTGAAAAAACAGAGGCATCCCCGGCTGCTGCCCCCGTGCCCGTGGCAGAGGACGTGCCTCGG
 451   K  C  K  E  K  Q  S  K  P  T  E  K  T  E  A  S  P  A  A  A  P  V  P  V  A  E  D  V  P  R

1441  ACCTCTATGTGTTCTGTGCAGTCGGCTCCGCCCACAGCAGCTCCAGCGAAGGGCATGGTGCCAGACGATGCTGTCGAAGCCTTGGCTGGA
 481   T  S  M  C  S  V  Q  S  A  P  P  T  A  A  P  A  K  G  M  V  P  D  D  A  V  E  A  L  A  G

1531  AGCCTGGGCAAAAAGGAAGCAGATCCAGAAGACGGGAAAGCCTGTGGAGGATAAAGTCAAGGAGAAAGCCAAAGAAGAAGGGATCGTGAGAAA
 511   S  L  G  K  K  E  A  D  P  E  D  G  K  P  V  E  D  K  V  K  E  K  A  K  E  E  D  R  E  K

1621  CTTGGTGAAAAAGAAGAAAGCGATTCCTCCTGATTACAGATTAGAAGAAGCCAAGGATAAAGACGGAAAAACCACTGCTGCCAAAAGAGGTC
 541   L  G  E  K  E  E  T  I  P  P  D  Y  R  L  E  E  A  K  D  K  D  G  K  P  L  L  P  K  E  V

1711  AAGGAACCGCTCCCACCCTTGAGTGAAGACGTCCTCCTCGATGCTCTGTCCAAGGACTTCACTGTCCCCTCAGACACATCATCGCCTCAA
 571   K  E  P  L  P  P  L  S  E  D  V  L  L  D  A  L  S  K  D  F  T  V  P  S  D  T  S  S  P  Q

1801  TTTGAAGATGCTAAACTCTCAGCTGTCGTCTCTGAAGTGGTTTCCCAAACCCCAGCTCCAACCACCCAGGCAGCCGGTCCACCCCCCAGC
 601   F  E  D  A  K  L  S  A  V  V  S  E  V  V  S  Q  T  P  A  P  T  Q  A  A  G  P  P  P  S

1891  ACTGCGCAGCGTGACAACAAAGAAACTTGACGATGCCCTGGATCAACTTTCTGACAGTCTCGGGCAAAGACGCCTGATCCAGATGAGGAT
 631   T  A  Q  R  D  N  K  E  L  D  D  A  L  D  Q  L  S  D  S  L  G  Q  R  Q  P  D  P  D  E  D

1981  AAACCCGTAGAGGATAAAGTCAAGGAAAAAGCCAAAGCTGAACACAGAGACAAGCTGGGAGAAAGAGATGACACCATCCCACCTAAATAC
 661   K  P  V  E  D  K  V  E  K  A  K  A  E  H  R  D  K  L  G  E  R  D  D  T  I  P  P  K  Y

2071  CAACATCTTTTGGATGACAACAAGGAGGGCACACCCGGGAAGCCAAAGGCATCAGAGAAGCCCAAGGCATCAGAGAAACCTGCAGGTGCC
 691   Q  H  L  L  D  D  N  K  E  G  T  P  G  K  P  K  A  S  E  K  P  K  A  S  E  K  P  A  G  A

2161  CAGGACCCCATTGATGCCCTCTCAGGGGACTTTGACAGCTGTCCCTCGACTACAGAAACCTCGACAGACACCACCAAAGGACAAAGACAAG
 721   Q  D  P  I  D  A  L  S  G  D  F  D  S  C  P  S  T  T  E  T  S  T  D  T  P  K  D  K  D  K

2251  AAGCCTGCTTCCAGTGCCGAAGCACCTAGGAATGGCGGGAAAGCAAAGGATTGCACAAAGGCAAAGGAGGAAACTTCCAAGCCAAAAGCT
 751   K  P  A  S  S  A  E  A  P  R  N  G  G  K  A  K  D  C  T  K  A  K  E  E  T  S  K  P  K  A

2341  GATGGGAAAAAGTACAAGTTAAagttcacactatttggtatctgcatataaaatcttcagcgggtagatggtgact
 781   D  G  K  S  T  S end
```
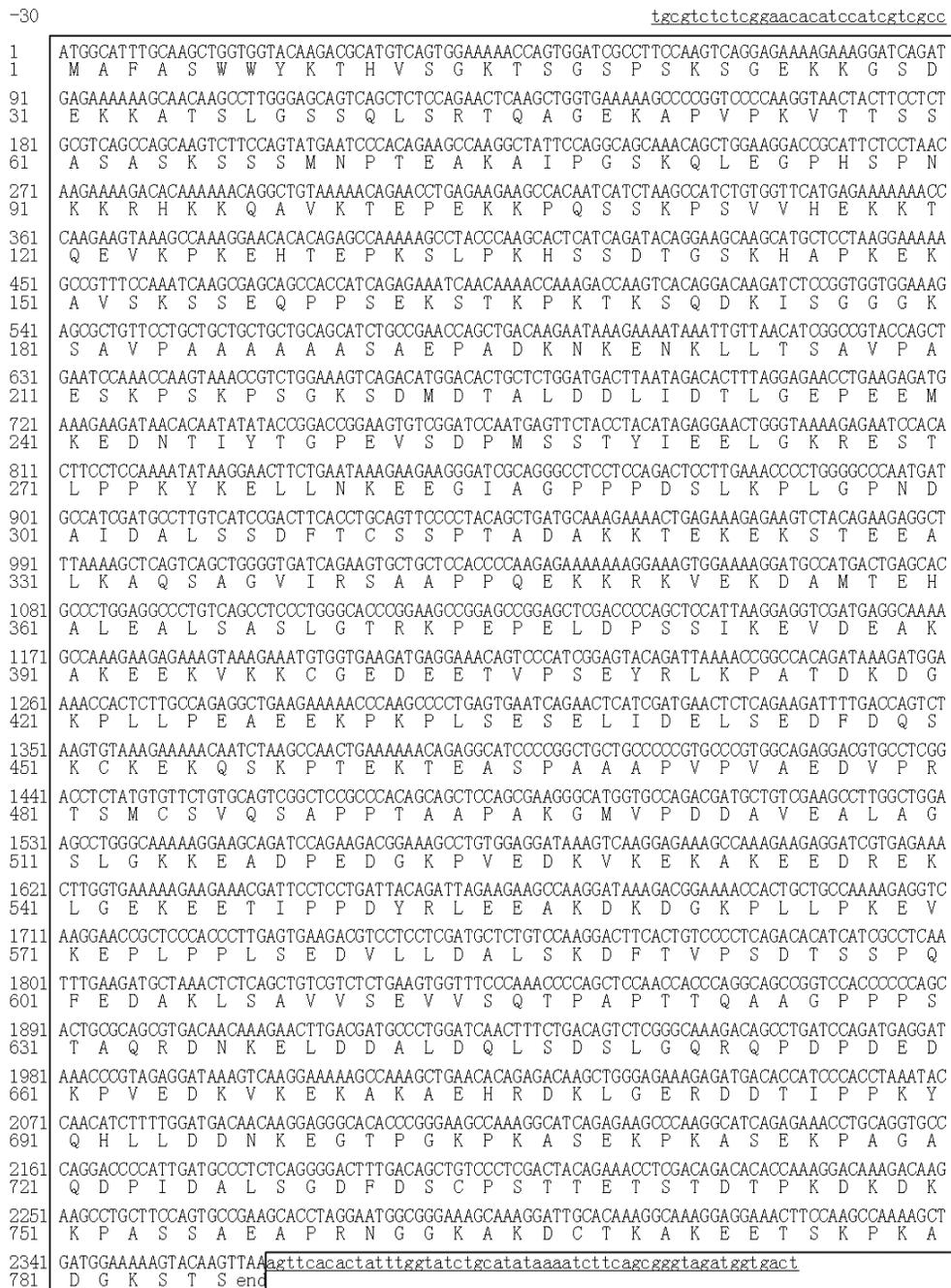
Figure 2. Full-length nucleotide sequence of CAST cDNA from yak (2445 bp) and deduced amino-acid sequence (786 aa). Positions of nucleotides and amino acids are indicated on the left, and sequences are numbered relative to the translation initiation site. The open reading frame (ORF) and encoded protein are boxed. The 3′-untranslated region (3′-UTR) and 5′-untranslated region (5′-UTR) are underlined. There is no signal peptide or membrane-anchoring domain. Potential prosite motifs and functional sites are shown in Fig. 3.

such as signal transduction, cell proliferation, cell cycle progression, differentiation, apoptosis, membrane fusion, platelet activation and skeletal muscle protein degradation (Fig. 3 and Table 2). Calpastatin protein signal peptides were analyzed using SignalP 3.0 software. Results from the protein signal peptide analysis suggest that calpastatin is a nonsecretory cytoplasmic protein. Using the Protean program of DNAStar, potential secondary structures of CAST were predicted, and the results indicated that there are 33 α-helices and eight β-sheets in the sequence of CAST (Fig. 3).

Molecular basis for interaction of calpain-calpastatin system

Based on a previous research report, we predict that the N-terminal region domain L of the yak calpastatin that contains a short peptide L54–64 (EVKPKEHT-EPK) may be responsible for Ca^{2+} channel repriming function (Hao *et al.*, 2000; Minobe *et al.*, 2006) (Fig. 3). The homologous inhibitory domains contain three short conserved segments (subdomains A, B, and C) of about 20 amino acids each that are primarily responsible for
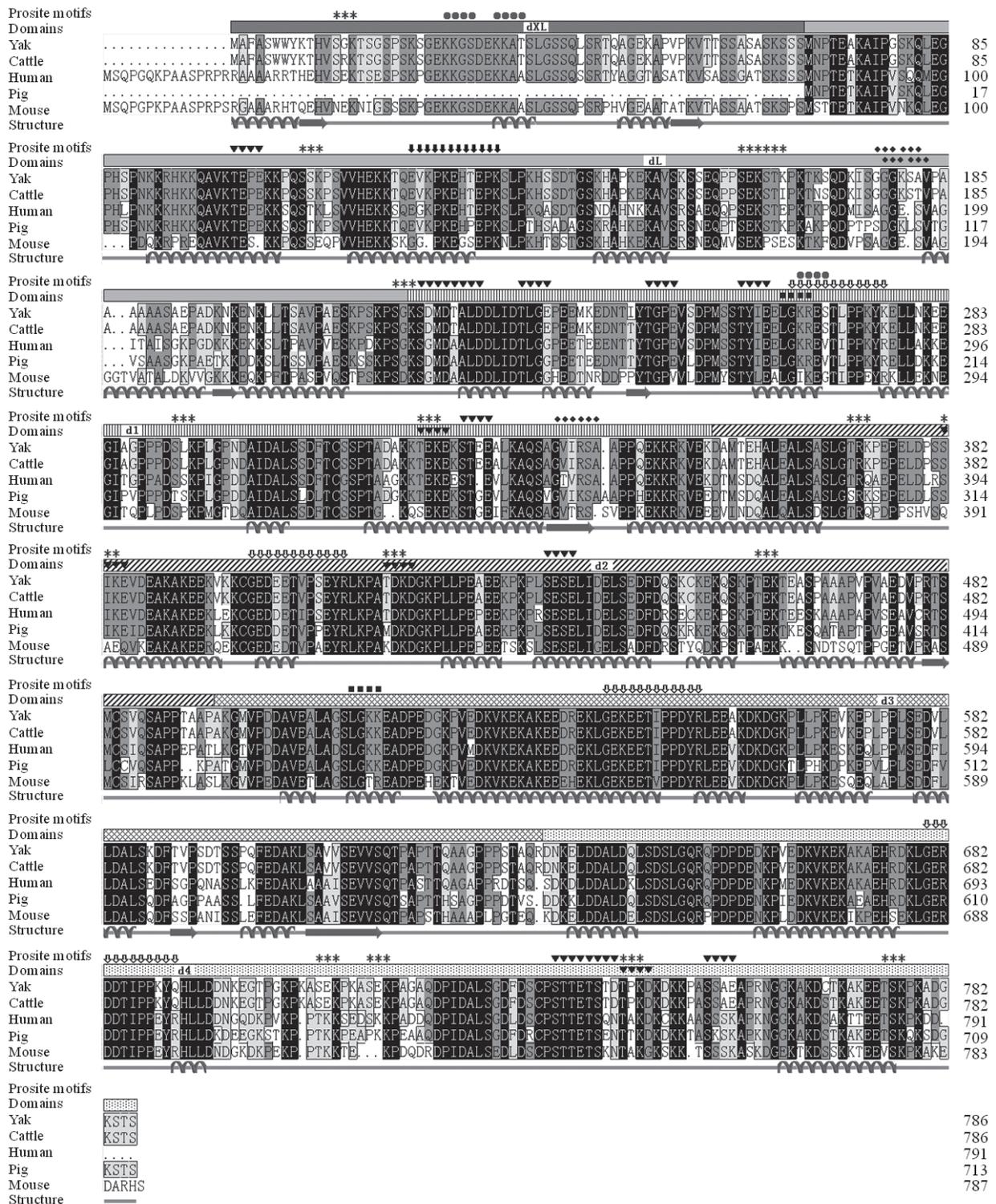
Figure 3. Multiple alignment and analysis of amino-acid sequences of mammalian calpastatins.
Secondary structural elements of the core region of yak calpastatin were predicted by DNAStar Protean, and amino-acid sequence align-ment of calpastatin of the yak, cattle, human, pig, and mouse was performed by DNAMAN software (prediction identity =81.69%). Do-mains of yak calpastatin are depicted by different symbols: dXL (1 to 68), dL (69 to 219), d1 (220 to 354), d2 (355 to 495), d3 (496 to 634), d4 (635 to 786). Broad arrows and wavy lines under the sequences denote potential β-sheets and α-helices, respectively. Possible prosite motifs and phosphorylation sites were obtained with PROSITE motif search of the ExPASy Server (Table 2): protein kinase C phosphoryla-tion sites are marked by stars, cAMP- and cGMP-dependent protein kinase phosphorylation sites are marked by closed circles, casein kinase II phosphorylation sites are marked by triangles, N-myristoylation sites are marked by diamonds, amidation sites are marked by squares. The L54–64 short peptide in domain L is indicated by closed arrows, and the highly conserved repetitive 12-aa peptides in each of subdomains B are indicated by open arrows above the sequences. Letters in blue-black blocks indicate 100% homology between these sequences, and those in gray blocks indicate ≥60% homology.

Table 2. Analysis of prosite motifs of yak calpastatin with ExPASy Server[1]

Prosite motifs and sites[2]	Sequences and residues
L54–64 short peptide	E^{122}VKPKEHTEPK
12-AA short peptide	G^{265}KRESTLPPKYK, G^{400}EDEETVPSEYR, G^{542}EKEETIPPDYR, G^{680}ERDDTIPPKYQ
LYS_RICH	K^{91}krhkkqavktepekkpqsskpsvvhekktqevkpkehtepkslpkhssdtgskhapkekavsksseqppsekstkpktksqdK
PKC_PHOSPHO_SITE	S^{13}gK, S^{109}sK, S^{161}eK, S^{164}tK, S^{218}gK, S^{292}lK, T^{321}eK, T^{371}rK, S^{382}iK, T^{416}dK, T^{460}eK, S^{708}eK, S^{714}eK, T^{744}pK, T^{775}sK
CAMP_PHOSPHO_SITE	K^{26}KgS, K^{32}KaT,K^{266}ReS
CK2_PHOSPHO_SITE	T^{101}epE, S^{221}dmD, T^{225}alD, T^{233}lgE, T^{248}gpE, T^{259}yiE, T^{321}ekE, S^{326}teE, S^{382}ikE, T^{416}dkD, S^{435}esE, S^{736}ttE, T^{740}stD, T^{744}pkD, S^{754}saE
MYRISTYL	G^{177}GgkSA, G^{178}GksAV, G^{337}VirSA
AMIDATION	I^{264}GKR, I^{512}GKK

[1]http://www.expasy.org/prosite/. [2]LYS_RICH, lysine-rich region; PKC_PHOSPHO_SITE, protein kinase C phosphorylation site; CAMP_PHOSPHO_SITE, cAMP- and cGMP-dependent protein kinase phosphorylation site; CK2_PHOSPHO_SITE, casein kinase II phosphorylation site; MYRISTYL, N-myristoylation site; AMIDATION, amidation site.

Figure 4. Alignment and analysis of yak calpastatin inhibitory domains and their subdomains A, B, and C.
The four domains of yak calpastatin inhibitory to calpain are shown aligned, which subdomains A (20 aa), B (27 aa), and C (20 aa) indicated by broad line above their sequences. Identity or strong similarity of residues extending to four (black) or three (gray) domains are marked by shading (Tompa et al., 2002). Analysis by DNAMAN indicated that the identity of subdomains A, B, and C in domains 1, 2, 3, and 4 are 61.25%, 62.96%, and 59.52%, respectively. The highly conserved repetitive 12-aa peptides in each of subdomains B are indicated by arrows (Fig. 5).
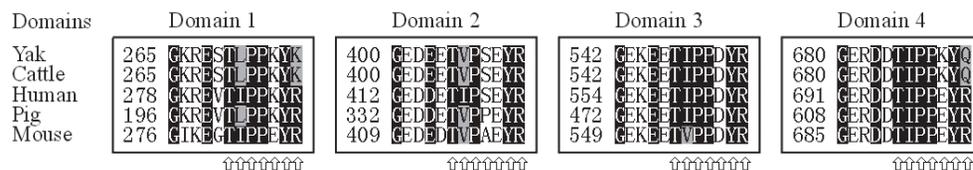
Figure 5. Alignment of the highly conserved repetitive 12-aa peptides of calpastatin domains 1, 2, 3, and 4 from yak, cattle, human, pig, and mouse.
Each of these repeating domains has inhibitory activity with functional sequence TIPPXYR indicated by open arrows. Residues of each peptide are numbered on the left.

calpain inhibition (Fig. 4). Subdomain B, a 27-residue peptide, is responsible for inhibition by binding to the active site of the enzyme, whereas subdomains A and C only potentiate this inhibitory effect by anchoring the inhibitor to the calmodulin-like domains of the large and small subunits of CAPN, respectively, in a strictly Ca^{2+}-dependent manner (Tompa et al., 2002; Mucsi et al., 2003; Melloni et al., 2006) (Fig. 4). Hence the yak calpastatin has potentially dual function, namely inhibition of calpain (domains 1–4) and the regulation of the Ca^{2+} channel (domain L) (Minobe et al., 2006). Subdomain B areas of the d1, d2, d3 and d4 domain contain a sequence Gly-X-X-Glu/Asp-X-Thr-Ile-Pro-Pro-X-Tyr-Arg (GXXE/DXTIPPXYR, where X is a variable amino acid) that appears to be important for protease inhibition (Croall & Demartino, 1991) (Fig. 5).

The calpains and their specific competitive inhibitor calpastatin form a calpain–calpastatin system (Goll et al., 2003). The activity of calpains is tightly controlled by the endogenous inhibitor calpastatin in the presence of calcium (Wendt et al., 2004; Hanna et al., 2007), but the mechanism of inhibition by calpastatin and the basis for its absolute specificity, to date, have remained speculative (Betts et al., 2003; Todd et al., 2003; Pfizer et al., 2008). μ-Calpain (or CAPN I) and m-calpain (or CAPN II) are 100-kDa heterodimers with homologous large subunits comprising four domains (DI–DIV) and a common small subunit with two domains (DV, DVI). Calpastatin binds to the calpains at three sites on the calpain molecule: subdomain A of calpastatin to domain IV in calpain, subdomain C of calpastatin to domain VI in calpain, and subdomain B of calpastatin to an area near the active site (domains IIa or IIb or both) of calpain (Fig. 6). However, it was not clear how subdomain B of the unstructured protein inhibits calpains without being cleaved itself and how calcium induces changes facilitated the binding of calpastatin to calpain (Hanna
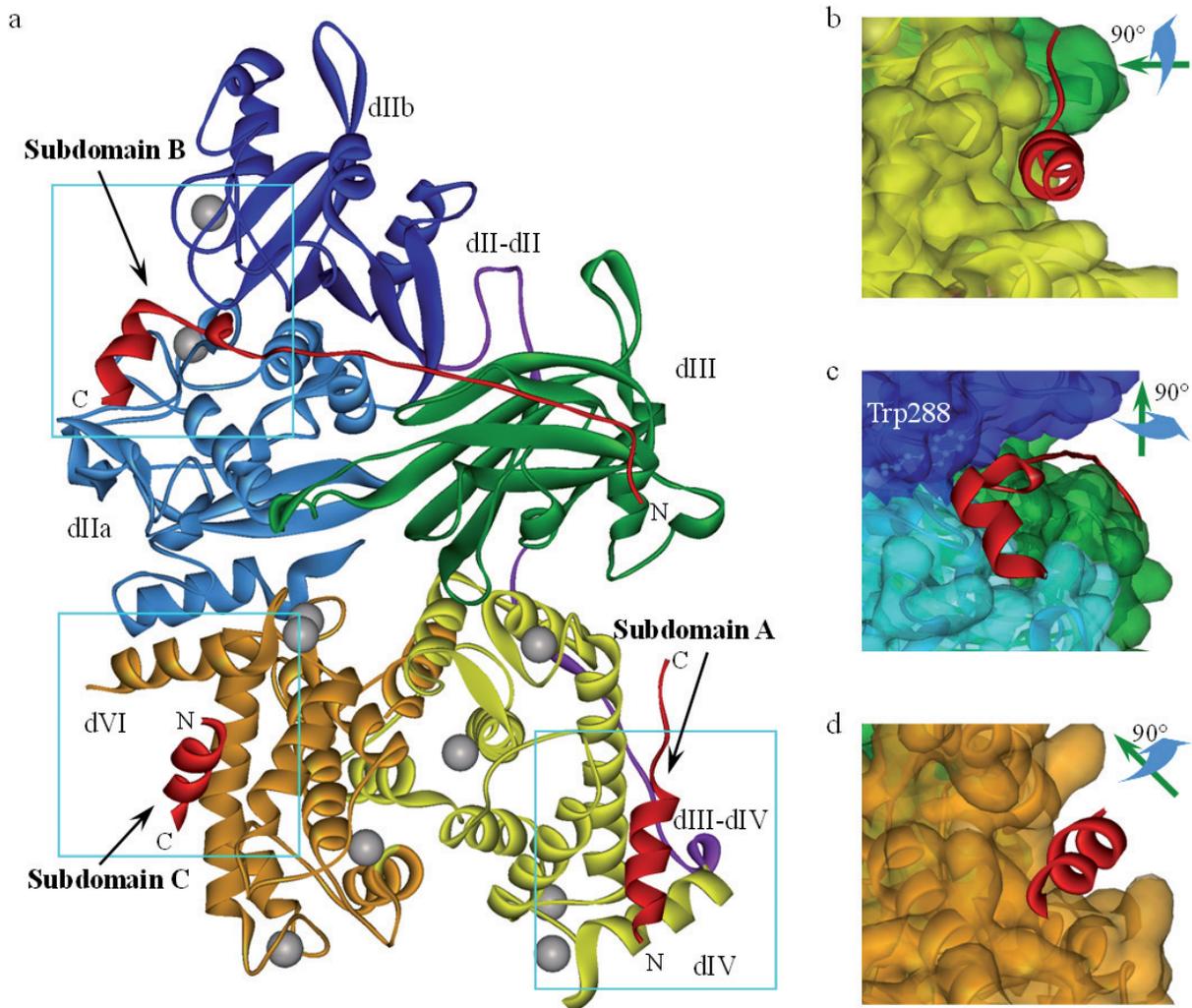
Figure 6. Structural diagram of calpastatin domain 4 (CAST4) bound to m-calpain (CAPN II).
The diagram is designed as indicated by Hanna and coworkers in their structure of m-calpain in complex with domain 4 of calpastatin (Hanna *et al.*, 2008). **(a)** Schematic representation of the domain structure of the large and small subunits of m-calpain in complex with domain 4 of calpastatin. CAST4 is unstructured in the absence of calpain, and forms three α-helices (red) when in complex with the enzyme. The helices in subdomains A and C are in contact with dIV (yellow) and dVI (orange), and the helix in subdomain B is in contact with the protease core dIIa (light blue) and dIIb (blue), and they are shown in solid ribbon representation. The ten Ca²⁺ ions in dIIa, dIIb, dIV, and dVI are shown as spheres (gray). Boxes represent approximate area where subdomains A, B, and C of CAST4 bind to domains of m-calpain. **(b–d)** Representation of binding area of subdomains A, B, C of CAST4 with the domains of calpain, obtained from Fig. 6a by a 90° rotation around an axis shown in Fig. 6b, 6c, and 6d and shown in solid ribbon with solid surface. The helices of subdomains A, B, and C and N-terminal side of subdomain B are buried in hydrophobic groove patch in calpain. The important residue Trp288 in the calcium-dependent activation of the protease core is shown in Fig. 6c (Todd *et al.*, 2003).

et al., 2008; Moldoveanu *et al.*, 2008). When bound to calpain, each subdomain A and subdomain C forms an amphipathic α-helix that binds to an exposed hydrophobic groove patch on DIV (on the large subunit) and DVI (on the small subunit), respectively (Todd *et al.*, 2003; Hanna *et al.*, 2008; Moldoveanu *et al.*, 2008). Subdomain B of calpastatin inhibits calpain by binding to the activated enzyme on the side of the active site cleft that requires interaction between calpain DIV and DVI and calpastatin subdomains A and C, respectively (Fig. 6b–d). The N-terminal side of subdomain B forms hydrophobic and electrostatic interactions with a shallow groove in DIII that becomes aligned with the catalytic cleft (Fig. 6c).

Each of the domains 1, 2, 3, and 4 of calpastatin can inhibit the proteolytic activity of either μ- or m-calpain, whereas the L domain of calpastatin alone has no inhibitory activity. Theoretically, one calpastatin molecule can inhibit four calpains (Emori *et al.*, 1988; Maki *et al.*, 1988). The ability of the individual domains to inhibit the calpains is as follows: d1 > d4 > d3 > d2, from the most to the least effective (Emori *et al.*, 1988; Kawasaki *et al.*, 1989). In fact, the conserved 12-amino-acid sequence (GXXE/DXTIPPXYR) in the C-terminal part of subdomains B areas is essential for the inhibitory activity (Figs. 5 and 6). It is reported that the maximum effective inhibition of the calpains by calpastatin requires all calpastatin subdomains, A, B, and C to bind simultaneously to calpain (Goll *et al.*, 2003). Previous research (Hanna *et al.*, 2008) indicated that 1) a shift caused by Ca²⁺ in the EF hands exposes hydrophobic areas and provides enough room for the binding of subdomains A and C to domains of the calpain; 2) a large structure change, caused by calcium binding to the protease core

to align the active site, Trp 288 in dIIb of calpain, forces the subdomain B to occupy and bind to the cleft; 3) the rearrangement of the domains relative to one another allows the simultaneous interaction of calpastatin with both dIII and the protease core, which increases the area of interaction, and thus the overall affinity of calpastatin for calpain.

Acknowledgements

REFERENCES

Bendtsen JD, Nielsen H, Heijne G, Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* **340:** 783–795.

Betts R, Weinsheimer S, Blouse GE, Anagli J (2003) Structural determinants of the calpain inhibitory activity of calpastatin peptide B27-WT. *J Biol Chem* **278:** 7800–7809.

Carragher NO, Westhoff MA, Riley D, Potter DA, Dutt P, Elce JS, Greer PA, Frame MC (2002) v-Src-induced modulation of the calpain-calpastatin proteolytic system regulates transformation. *Mol Cell Biol* **22:** 257–269.

Chou PY (1990) Prediction of protein structural classes from amino acid composition. In *Prediction of protein structure and the principles of protein conformation*. Plenum Press, New York, NY.

Chou PY, Fasman GD (1978) Prediction of the secondary structure of proteins from their amino acid sequence. *Adv Enzymol* **47:** 45–148.

Croall DE, Demartino GN (1991) Calcium-activated neutral protease (calpain) system: structure, function, and regulation. *Physiol Rev* **71:** 813–847.

Emori Y, Kawasaki H, Imajoh S, Minami Y, Suzuki K (1988) All four repeating domains of the endogenous inhibitor for calcium-dependent protease independently retain inhibitory activity. *J Biol Chem* **263:** 2364–2370.

Glading A, Lauffenburger DA, Wells A (2002) Cutting to the chase: calpain proteases in cell motility. *Trends Cell Biol* **12:** 46–54.

Goll DE, Thompson VF, Li H, Wei W, Cong J (2003) The calpain system. *Physiol Rev* **83:** 731–801.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* **41:** 95–98.

Hanna RA, Robert LC, Peter LD (2008) Calcium-bound structure of calpain and its mechanism of inhibition by calpastatin. *Nature* **456:** 409–412.

Hanna RA, Garcia-Diaz BE, Davies PL (2007) Calpastatin simultaneously binds four calpains with different kinetic constants. *FEBS Lett* **581:** 2894–2898.

Hao LY, Kameyama A, Kuroki S, Takano J, Takano E, Maki M, Kameyama M (2000) Calpastatin domain L is involved in the regulation of L-type Ca^{2+} channels in guinea pig cardiac myocytes. *Biochem Biophys Res Commun* **279:** 756–761.

Hirose K, Kadowaki S, Tanabe M, Takeshima H, Iino M (1999) Spatiotemporal dynamics of inositol 1,4,5-trisphosphate that underlies complex Ca^{2+} mobilization patterns. *Science* **284:** 1527–1530.

Kawasaki H, Emori Y, Imajoh-ohmi S, Minami Y, Suzuki K (1989) Identification and characterization of inhibitory sequences in four repeating domains of the endogenous inhibitor for calcium dependent protease. *J Biochem* **106:** 274–281.

Kiss R, Bozoky Z, Kovacs D, Rona G, Friedrich P, Dvortsak P, Weisemann R, Tompa P, Perczel A (2008a) Calpain-induced tripartite binding of intrinsically disordered calpastatin to its cognate enzyme, calpain. *FEBS Lett* **582:** 2149–2154.

Kiss R, Kovacs D, Tompa P, Perczel A (2008b) Local structural preferences of calpastatin, the intrinsically unstructured protein inhibitor of calpain. *Biochemistry* **47:** 6936–6945.

Maki M, Takano E, Mori H, Sato A, Murachi T, Hatanaka M (1987) All four internally repetitive domains of pig calpastatin possess inhibitory activities against calpains I and II. *FEBS Lett* **223:** 174–180.

Maki M, Takano E, Osawa T, Ooi T, Murachi T, Hatanaka M (1988) Analysis of structure-function relationship of pig calpastatin by expression of mutated cDNAs in *Escherichia coli*. *J Biol Chem* **263:** 10254–10261.

Melloni E, Averna M, Stifanese R, Tullio RD, Defranchi E, Salamino F, Pontremoli S (2006) Association of calpastatin with inactive calpain. *J Biol Chem* **281:** 24945–24954.

Minobe E, Hao LY, Saud ZA, Xu JJ, Kameyama A, Maki M, Jewell KK, Parr T, Bardsley RG, Kameyama M (2006) A region of calpastatin domain L that represses cardiac L-type Ca^{2+} channels. *Biochem Biophys Res Commun* **348:** 288–294.

Moldoveanu T, Gehring K, Green DR (2008) Concerted multi-pronged attack by calpastatin to occlude the catalytic cleft of heterodimeric calpains. *Nature* **456:** 404–408.

Mucsi Z, Hudecz F, Hollósi M, Tompa P, Friedrich P (2003) Binding-induced folding transitions in calpastatin subdomains A and C. *Protein Sci* **12:** 2327–2336.

Pfizer J, Assfalg-Machleidt I, Machleidt W, Chaschke N (2008) Inhibition of human mu-calpain by conformationally constrained calpastatin peptides. *Biol Chem* **389:** 83–90.

Sambrook J, Russell D (2001) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Press, Cold Spring Harbor, NY.

Shiraha H, Glading A, Chou J, Jia Z, Wells A (2002) Activation of m-calpain (calpain II) by epidermal growth factor is limited by protein kinase A phosphorylation of m-calpain. *Mol Cell Biol* **22:** 2716–2727.

Takano J, Watanabe M, Hitomi K, Maki M (2000) Four types of calpastatin isoforms with distinct amino-terminal sequences are specified by alternative first exons and differentially expressed in mouse tissues. *J Biochem (Tokyo)* **128:** 83–92.

Todd B, Moore D, Deivanayagam CC, Lin GD, Chattopadhyay D, Maki M, Wang KK, Narayana SV (2003) A structural model for the inhibition of calpain by calpastatin: crystal structures of the native domain VI of calpain and its complexes with calpastatin peptide and a small molecule inhibitor. *J Mol Biol* **328:** 131–146.

Toke O, Banoczi Z, Tarkanyi G, Friedrich P, Hudecz F (2009) Folding transitions in calpain activator peptides studied by solution NMR spectroscopy. *J Pept Sci* **15:** 404–410.

Tompa P, Mucsi Z, Orosz G, Friedrich P (2002) Calpastatin subdomains A and C are activators of calpain. *J Biol Chem* **277:** 9022–9026.

Tullio RD, Averna M, Salamino F, Pontremoli S, Melloni E (2000) Differential degradation of calpastatin by μ- and m-calpain in Ca^{2+}-enriched human neuroblastoma LAN-5 cells. *FEBS Lett* **475:** 17–21.

Uemori T, Shimojo T, Asada K, Asano T, Kimizuka F, Kato I, Maki M, Hatanaka M, Murachi T, Hanzawa H, Arata Y (1990) Characterization of a functional domain of human calpastatin. *Biochem Biophys Res Commun* **166:** 1485–1493.

Weber H, Jonas L, Hühns S, Schuff-Werner P (2004) Dysregulation of the calpain-calpastatin system plays a role in the development of cerulein-induced acute pancreatitis in the rat. *Am J Physiol Gastrointest Liver Physiol* **286:** G932–G941.

Wendt A, Thompson VF, Goll DE (2004) Interaction of calpastatin with calpain: a review. *Biol Chem* **385:** 465–472.