Identification of the active site of legumain links it to caspases, clostripain and gingipains in a new clan of cysteine endopeptidases

Jinq-May Chen, Neil D. Rawlings, Richard A.E. Stevens, Alan J. Barrett*

MRC Molecular Enzymology Laboratory, Babraham Institute, Babraham, Cambridge CB2 4AT, UK

Received 10 November 1998

Abstract We show by site-directed mutagenesis that the catalytic residues of mammalian legumain, a recently discovered lysosomal asparaginycysteine endopeptidase, form a catalytic dyad in the motif His-Gly-spacer-Ala-Cys. We note that the same motif is present in the caspases, aspartate-specific endopeptidases central to the process of apoptosis in animal cells, and also in the families of clostripain and gingipain which are arginyl/lysyl endopeptidases of pathogenic bacteria. We propose that the four families have similar protein folds, are evolutionarily related in clan CD, and have common characteristics including substrate specificities dominated by the interactions of the S1 subsite.

Key words: Legumain; Caspase; Clostripain; Gingipain; Peptidase clan

1. Introduction

Legumain (EC 3.4.22.34) is a cysteine endopeptidase that was until recently known only from plants [1,2] and *Schistosoma* [3]. In plants there is abundant evidence that legumain performs a protein-processing function, causing limited proteolysis of precursor proteins and protein splicing [2,4]. Following the discovery of legumain in mammalian cells, and the cloning and sequencing of legumain from human [5] and mouse [6], we have shown the enzyme to be predominantly lysosomal [6], but it has strict specificity for the hydrolysis of bonds on the carboxyl side of asparagine which is very different from that of any other lysosomal endopeptidase and adapts it particularly for limited proteolysis. Consistent with this, there is evidence for a key role of legumain in the processing of a bacterial antigen for the MHC class II system in the lysosomal/endosomal system of antigen presenting cells [7] and it also processes progelatinase A to the active enzyme (J.-M. Chen and A.J. Barrett, unpublished results).

The amino acid sequence of legumain shows it to belong to a distinct family of cysteine endopeptidases that has been termed peptidase family C13. (This and other identifiers for peptidase families and clans are as defined in the MEROPS database [8,9].) We here describe the identification of the catalytic residues of legumain, and show how they reveal relationships between family C13 and several other families of cysteine endopeptidases.

2. Materials and methods

2.1. Mutagenesis and plasmid construction of legumain mutants

Mouse legumain cDNA was removed from the plasmid construct pCMVmusleg [6] with *Eco*RI and *Xho*I and subcloned into the pCR-Script Amp^r SK(+) vector (Stratagene) at the same sites to produce a pCRmusleg construct. Mutagenesis was performed by PCR site-directed mutagenesis as described by Picard et al. [10] with modifications. Six mutagenic primers based on the nucleotide sequence of human legumain were designed to generate mutants H47A, C52S, H150A, H164A, C191S, and S195A. These primers with underlined nucleotides encoding the mutated residues were: 5′-GGTATAATTATAGGGCACAGGCAGACGC-3′ for H47A, 5′-AGGCAGACGCGTCCCATGCCTAC-3′ for C52S, 5′-CATTTACTTCACTGACGCTGGATCTACTGG-3′ for H150A, 5′-CTAATGATGATCTTGCTGTCAAGGACCTGA-3′ for H164A, 5′-CTACATTGAAGCCTCTGAGTCTGGGTCC-3′ for C191S, and 5′-GTGAGTCTGGGGCTATGATGAACC-3′ for S195A. Since the mutagenic primers were based on the human legumain sequence, some silent mutations were introduced into the mouse legumain constructs, and in mutant H150A the residue Ala-152 of mouse legumain was replaced by the Ser corresponding in the human sequence. Polymerase *Pfu* was used for all PCR reactions with a typical run of 95°C×1 min, 48°C×1 min, 72°C×2 min for 30 cycles. Introduction of the mutated codon was performed by PCR using a mutagenic primer, the downstream primer of mouse legumain (5′-CTGGTGTTGGTGTGGGACTTGACC-3′), and pCMVmusleg as the template to generate a megaprimer. The megaprimer and the legumain-specific upstream primer (5′-GACGCCCGGAATTCCCACGGTT-3′) were used for the second run of PCR with the same template to generate an 880-bp product containing the mutated nucleotide(s). This product was purified from the agarose gel by use of the Gene Clean Kit (BIO 101) and double digested with *Eco*RI and *Hin*cII. Wild-type pCRmusleg was similarly digested with *Eco*RI and *Hin*cII to remove the 880-bp wild type sequence, and the mutated PCR product was ligated into this vector to generate the mutated forms of pCRmusleg. Plasmid DNA of wild-type and mutant pCRmusleg was propagated in *E. coli* DH10B (Life Technologies) or XL10-gold (Stratagene) and purified by standard procedures [11]. The sequences of all mutated constructs were confirmed by sequencing. The full length cDNA of mutant legumain was then excised from the cloning vector pCRmusleg with *Eco*RI and *Xho*I and cloned into the mammalian expression vector pCMV-SPORT2 for transient expression.

2.2. Cell culture and transfection

Human embryonic kidney (HEK) 293 cells (ATCC no. CRL 1573) were maintained in minimal Eagle's medium supplemented with 10% fetal bovine serum and antibiotics. The calcium phosphate precipitation method was used to transfect HEK 293 cells with the wild-type and mutant legumain constructs as described [6].

2.3. Detection of recombinant legumain and enzymatic assays

HEK 293 cells transfected with legumain constructs and mock-transfected cells were harvested three days after transfection. Cells were disrupted by three cycles of freeze and thaw in the lysis buffer (0.1 M sodium citrate, pH 5.8, containing 1 mM EDTA and 1% *n*-octyl β-D-glucopyranoside). Cell lysates were collected after centrifugation at 18 000×*g* for 10 min. The recombinant protein of wild-type and mutant legumain expressed in the cell lysate were detected by SDS-PAGE immunoblot developed with a polyclonal antibody specific to pig legumain [6]. To measure the enzymatic activity of recombinant legumain, lysate (5 μl) was placed in the well of a 96-

*Corresponding author. Fax: (44) (1223) 496023.
E-mail: alan.barrett@bbsre.ac.uk

PII: S0014-5793(98)01574-9

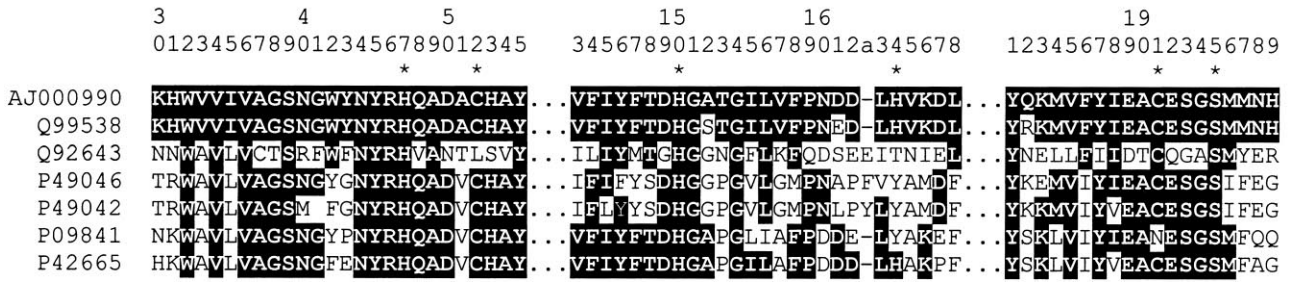

Fig. 1. Selection of candidate catalytic residues for mutagenesis. Only the parts of the sequences containing well-conserved residues of cysteine or histidine are shown. The sequences (numbered according to mouse preprolegumain [6]) are: AJ000990, mouse legumain; Q99538, human legumain; Q92643, human GPI8 protein; P49046, jack bean (*Canavalia ensiformis*) asparaginyl endopeptidase; P49042, castor bean (*Canavalia ensiformis*) vacuolar processing enzyme; P09841, *Schistosoma mansoni* haemoglobinase; and P42665, *Schistosoma japonicum* haemoglobinase.

well microtiter plate and the reaction was started by addition of 120 μl of benzyloxycarbonyl-Ala-Ala-Asn-7-(4-methyl)coumarylamide (Z-Ala-Ala-Asn-NHMec) solution in assay buffer (39.5 mM citric acid, 121 mM Na_2HPO_4, pH 5.8, containing 1 mM DTT, 1 mM EDTA and 0.1% CHAPS) to give 10 μM final concentration of the substrate. The plates were incubated at 25°C and readings of fluorescence (excitation 360 nm, emission 460 nm) were taken at 10 and 130 min in a PerSeptive Biosystems CytoFluor II fluorescence plate reader. The change between the two readings was the measure of activity. Specific legumain activity was obtained by subtracting the control value of mock-transfected cells.

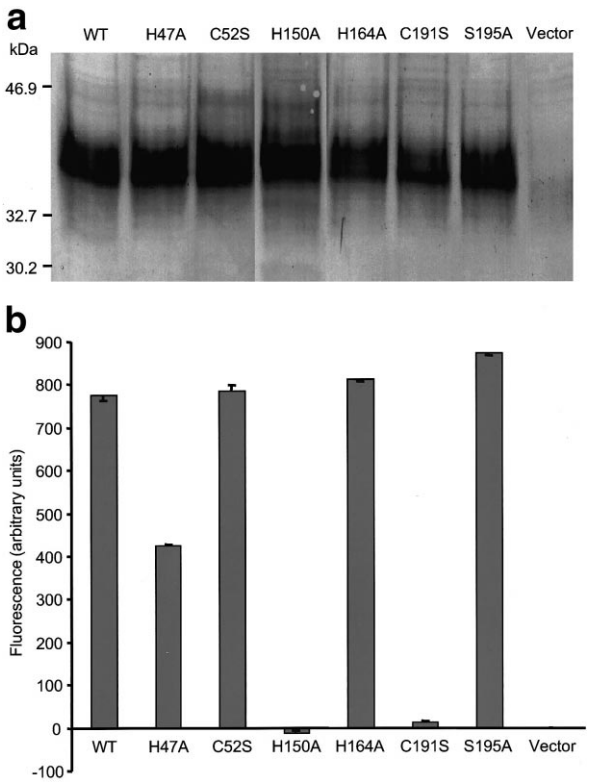

Fig. 2. Expression and activity of wild-type and mutant forms of legumain. a: Lysates (5 μl except H150A 20 μl) of HEK 293 cells transfected with legumain constructs in the pCMV-SPORT2 vector, or mock-transfected cells, were run in SDS-PAGE, and transblotted for the detection of legumain as described in Section 2. b: Legumain activity of cell lysates as in a, assayed with Z-Ala-Ala-Asn-NHMec as described in Section 2. The blank value of activity in the mock-transfected cells that has been subtracted is 622 fluorescence units. Bars show S.D. ($n = 3$).

3. Results and discussion

3.1. Catalytic dyad of legumain

The catalytic residues in all known cysteine endopeptidases are a dyad of cysteine and histidine, which may occur in either order, cysteine–histidine or histidine–cysteine, in the amino acid sequence. Accordingly, the residues His-47, Cys-52, His-150 and Cys-191 were selected as targets for mutagenesis by inspection of an alignment of the sequences for family C13, part of which is shown in Fig. 1. Site directed mutagenesis was done as described in Section 2 to generate mutant constructs H47A, C52S, H150A and C191S. Two additional constructs, H164A and S195A, were made as controls. Wild-type and mutant proteins were transiently expressed in HEK 293 cells under the control of the CMV promoter. Immunoblotting analysis showed that recombinant proteins from all constructs were expressed and processed to mature forms similar in molecular size (35 kDa) to the wild-type protein (Fig. 2a). The H150A mutant was expressed about four-fold less efficiently than the other forms as judged by inspection of the immunoblots.

Assays of legumain activity in extracts of cells expressing the mutated forms of the enzyme revealed that mutants H150A and C191S were completely inactive, whereas H47A and C52S were 54% and 101%, respectively, as active as the wild type. The two control mutants, H164A and S195A, were also as active as the wild type (Fig. 2b). In separate experiments (not shown) the C191S mutant of human legumain was made and again was found to be expressed normally but to be inactive. We conclude that His-150 and Cys-191 are essential for the catalytic activity of legumain and can be assumed to represent the catalytic dyad of the enzyme. The legumain of *Schistosoma mansoni* apparently lacks Cys-191 (Fig. 1), and we assume that the published sequence is that of an inactive variant of the enzyme. This would be consistent with the failure of Götz and Klinkert [12] to detect activity of the *Schistosoma mansoni* enzyme expressed in insect cells, but Cys-191 is present in the *S. japonicum* sequence. The other generally conserved cysteine residue, Cys-52, is absent from the human GPI8 protein that catalyses the attachment of glycosylphosphatidylinositol anchors to proteins [13,14].

3.2. An active site motif recognised in legumain is present in other families

The amino acid sequence of the segment of the polypeptide chain containing the catalytic residues identified in legumain

Fig. 3. Alignment of the catalytic sites of cysteine endopeptidases of families of legumain (C13), caspase-1 (C14), clostripain (C11) and gingipain R (C25). a: The sequence segments containing the known or putative catalytic residues (magenta and yellow) in each of the four families were aligned manually by the introduction of gap residues '.'. It can be seen that two or three residues N-terminal to each of the catalytic residues there is a block of four predominantly hydrophobic residues (green). Other residues that tend to be conserved between the families are printed in white on black. Asterisks are placed over the His and Cys residues in each family for which they have been identified experimentally as catalytic, and in addition the His residue in family C25 was proposed to be catalytic by Pavloff et al. [19] on the basis of its conservation in the family. Key to sequences: AL000990, mouse legumain; Q99538, human legumain; P49046, jack bean (*Canavalia ensiformis*) asparaginyl endopeptidase; P42665, *Schistosoma japonicum* haemoglobinase; P49048, *Caenorhabditis elegans* hypothetical protein T05E11.6; Q92643, human GPI8 protein; P42574, human caspase 3; P55210, human caspase 7; P55212, human caspase 6; O01382, *Drosophila melanogaster* caspase; P29466, human caspase 1; P42573, *Caenorhabditis elegans* CED3 protein; P09870, *Clostridium histolyticum* α-clostripain; Q51816, *Porphyromonas gingivalis* gingipain R. b: The corresponding secondary structures are those determined crystallographically for human caspase-1 (family C14; PDB 1IBC) or predicted for the C13, C11 and C25 families by the PHD program [20]. The β-strand residues are marked 'b', the α-helix residues 'a' and the turn residues 't'. Both catalytic residues in family C14 are preceded by β-strands (strands S3 and S4, respectively) and with one exception these strands (containing the blocks of hydrophobic residues noted in a above) are predicted for the other families. The secondary structure elements shown here for the caspase family can be seen in the three-dimensional representation of Fig. 4.

was compared with sequences of cysteine peptidases in other families. As shown in Fig. 3, it was found that a motif **His**-Gly-spacer-Ala-**Cys** could be recognised not only in the legumain family but also in the families of caspase-1 (C14), clostripain (C11) and gingipain R (C25). The caspases are mammalian cytosolic endopeptidases that play key roles in apoptosis, whereas clostripain and gingipain are cysteine endopeptidases from the pathogenic bacteria *Clostridium histolyticum* and *Porphyromonas gingivalis*, respectively [15,16]. The cysteine residues in this motif have already been identified as the catalytic cysteines of all three of these other families [17–19] and the histidine has been identified as catalytic in C14 [18] and suggested tentatively for C25 [19]. Closely preceding each of the catalytic residues in all four families is a block of four hydrophobic amino acids (marked in blue in Fig. 3a). The corresponding parts of the secondary structures predicted for mouse legumain, clostripain, and gingipain R by

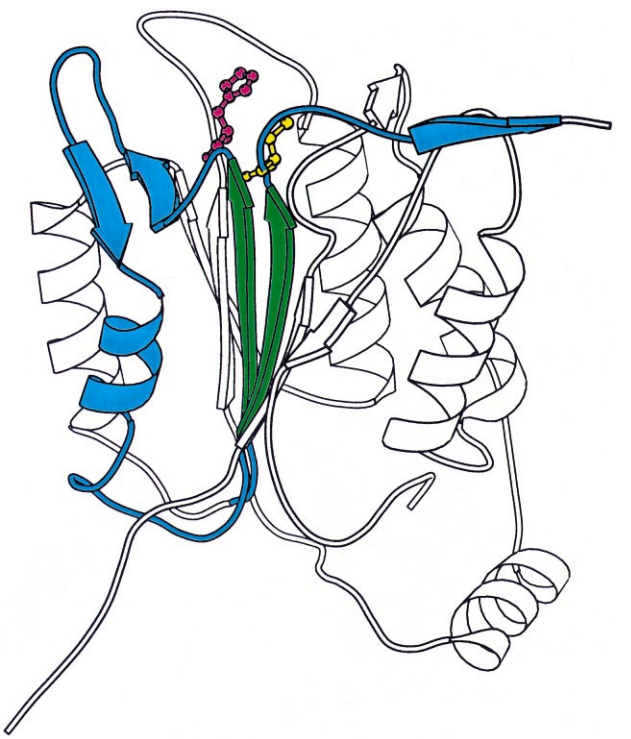

Fig. 4. Tertiary structure of part of the caspase-1 molecule (peptidase family C14) drawn from PDB record 1IBC by use of the MOLSCRIPT program [26]. Shown in green are the two β-strands that support the catalytic residues, histidine (pink) and cysteine (yellow), and the remainder of the segment included in the alignment of Fig. 3 is coloured in cyan.

the PHD program [20] are shown in Fig. 3b, and are generally in good agreement with the structure known for caspase-1 [18]. The longest spacer sequence exists in clostripain, but it is notable that the maturation of clostripain normally involves the autolytic excision of the nonapeptide 156-159E, although this is not required for activity [21]. As shown in Fig. 4, the catalytic residues of caspase-1 occur at the ends of two parallel β-strands. These are two central strands in a set of six that form a β-sheet at the core of the molecule, and the blocks of hydrophobic residues seen in the alignment are contained in the strands. This arrangement is expected to be mirrored in the other families.

3.3. The active site motif is unique

We next asked whether the sequence motif that contains the catalytic dyad in each of these four families of cysteine endopeptidases occurs in other proteins too. First, Motif 1 (Table 1) was formulated to include the blocks of hydrophobic residues preceding each of the catalytic residues as highlighted in Fig. 3, and the two non-hydrophobic residues immediately preceding the catalytic cysteine, and this was used to search the protein sequence databases. All 37 sequences retrieved proved to be known members of peptidase families C11, C13, C14 or C25, but many peptidases of these families were not recognised by Motif 1, so the more relaxed Motif 2 was constructed. This allowed a greater or lesser number of residues spacing the catalytic His and Cys, any hydrophobic residue at any position in the blocks preceding each member of the catalytic dyad, and restrictions on residue 189. With Motif 2, 66 hits were retrieved, and these included all but seven of the known members of the four families together with just three apparently unrelated false positive sequences all for hypothetical proteins (Table 1).

It is evident from these results that Motif 2 is very effective in screening protein sequences for the peptidases of just four families. Sequences belonging to all other known families of cysteine peptidases were quite unrecognised, notably including those from the papain (C1) and the calpain (C2) families, and those of many viruses.

Table 1
Database searches for proteins containing sequence motifs based on the active site sequence identified in legumain

Motif 1:	[AFILV]-[CILV]-[AFINVY]-[FILMY]-[ALMST]-X-H-G-X(2-4,47)-[ILPV]-[FILV]-[AFIY]-[AFIV]-X(1,2)-A-C
Motif 2:	[AFILMVWY]-[ACFILMVWY]-[ACFILMNVWY]-[AFILMVWY]-[AFILMSTVWY]-X-H-G-X(20,60)-[AFILMPVWY]-[AFILMVWY](3)-[EDQN]-X(0,1)-[ACST]-C

Sequences retrieved

C11	P09870
C13	O24325, O24326, O46047, P42665, P49018, P49042, P49043, P49044, P49045, P49046, P49047, P49048, Q39044, Q39119, Q92643, Q99538
C14	O01382, O02002, O02229, O02433, O08736, O08738, O14676, O14822, O35397, O35669, O42284, O55194, P29452, P29466, P29594, P42573, P42574, P42575, P43527, P45436, P49662, P51878, P55210, P55211, P55212, P55213, P55214, P55215, P55865, P55866, P55867, P70343, P70677, P89116, P97864, Q14790, Q60431, Q92851, Q98943
C25	O33441, P28784, P95493, Q51816, Q51838, Q51839, Q51844

Sequences not in families C11, C13, C14 or C25
O22161, O28117, Q53839

Sequences not retrieved although recognised members of families C11, C13, C14 or C25

C11	None
C13	O24539, P09841
C14	P42576
C25	O07442, P72194, Q51817, Q51818

The two motifs were formulated according to the conventions of PROSITE [27] and used to search the SWISSPROT and TREMBL databases (10 August 1998; 223 599 sequences) by use of the SCANPROSITE program at the ExPASy server [27]. Motif 1 retrieved the 37 sequences underlined, all of which are recognised members of families C11, C13, C14 or C25. Motif 2 retrieved 66 sequences of which 63 are recognised members of these families, but three are not. Seven sequences of known members of families C11, C13, C14 or C25 were not retrieved by either motif. These were O24539, *Vicia narbonensis* legumain (with Cys-147); P09841, *Schistosoma mansoni* haemoglobinase (with Cys191Asn replacement); P42576, isoform of the positive P42575; O07442, P72194 and Q51817, Q51818, aberrant forms of gingipain. Five sequence fragments that did not overlap the consensus sequence were ignored.

3.4. Four families in clan CD

We believe that these findings represent strong evidence that the four families of cysteine peptidases C11, C13, C14 and C25 are evolutionarily related, albeit very distantly, and that they retain topological similarities in the catalytic sites. Testable predictions from this proposal are that the same histidine residue (His-150 in Fig. 3) will be found experimentally to be essential for catalysis in clostripain and gingipain R, and that when three-dimensional structures are determined for representatives of families C11, C13 and C25 they will show similarities to those of the caspases.

Family C14 has already been assigned to clan CD in the classification of the MEROPS database [9] on the grounds of its distinctive protein fold, and has been the only family in this clan, but we now propose that families C11, C13 and C25 should be added to clan CD. One of the purposes of assembling clans of peptidases is the expectation that the evolutionary and structural similarities between the peptidases in related families will tend to be reflected in similar enzymological properties and sometimes even functions. The enzymes of clan CD do indeed have striking similarities in catalytic activity. All are very specific endopeptidases, with a strict requirement for the side chain of the P1 amino acid residue that provides the carbonyl group in the scissile peptide bond: arginine for clostripain (family C11), asparagine for legumain (C13), aspartate for the caspases (C14) and arginine or lysine for gingipain R or gingipain K, respectively, in family C25. This type of specificity is not characteristic of the more familiar but quite unrelated cysteine peptidases that are homologues of papain in clan CA, which generally show broad specificity with greatest selectivity in subsite S2 [22,23]. There are also similarities in inhibition characteristics. All of the peptidases in the families now assigned to clan CD are resistant to irreversible inhibition by compound E-64 [L-3-carboxy-2,3-*trans*-epoxypropionyl-leucylamido-(4-guanidino)butane], which has often been thought of as a general reagent for cysteine endopeptidases. (Reversible inhibition of gingipain R and clostripain is attributable to the fact that E-64 is an arginine analogue [24,25].) However, the cysteine endopeptidases of clan CD are susceptible to peptide aldehyde and peptidylchloromethane inhibitors, and it now appears that it may be possible to take advantage of successes in the structure-based design of inhibitors for the caspases to design inhibitors for enzymes in the other families, modifying specificity simply by

matching the different P1 requirements in a similar active site geometry.

3.5. Evolution of clan CD

The families of clan CD are widely distributed. Clostripain, the only known member of family C11, occurs in Gram-positive bacteria, and the gingipains of family C25 in Gram-negative bacteria. In contrast, members of family C13 are known from plants, fungi and animals, and the caspases of family C14 are known only from animals. The idea that these four families are derived from a common ancestor raises the question of the route by which the caspases and the legumains reached mammals following the origin of the clan presumably in the prebiotic genote. The eukaryote sequences are very different from those in bacteria, and thus give no indication of horizontal transfer from bacteria to eukaryotes. None of the four families is yet reported to be represented in archaea or protista, but it seems that representatives of the clan must have existed there at some time, and we suggest they may yet be found in the modern members of these groups.

Acknowledgements: This work was supported by the Medical Research Council (UK). We thank Drs Pam Dando and Harald Thidemann Johansen for valuable comments.

References

[1] Kembhavi, A.A., Buttle, D.J., Knight, C.G. and Barrett, A.J. (1993) Arch. Biochem. Biophys. 303, 208–213.
[2] Hara-Nishimura, I., Kinoshita, T., Hiraiwa, N. and Nishimura, M. (1998) J. Plant Physiol. 152, 668–674.
[3] Dalton, J.P. and Brindley, P.J. (1998) in: Handbook of Proteolytic Enzymes (Barrett, A.J., Rawlings, N.D. and Woessner, J.F., Eds.) pp. 749–754, Academic Press, London.
[4] Min, W. and Jones, D.H. (1994) Nat. Struct. Biol. 1, 502–504.
[5] Chen, J.-M., Dando, P.M., Rawlings, N.D., Brown, M.A., Young, N.E., Stevens, R.A., Hewitt, E., Watts, C. and Barrett, A.J. (1997) J. Biol. Chem. 272, 8090–8098.
[6] Chen, J.-M., Dando, P.M., Stevens, R.A.E., Fortunato, M. and Barrett, A.J. (1998) Biochem. J. 335, 111–117.
[7] Manoury, B., Hewitt, E.W., Morrice, N., Dando, P.M., Barrett, A.J. and Watts, C. (1998) Nature, in press.
[8] Rawlings, N.D. and Barrett, A.J. (1998) Nucleic Acids Res., in press.
[9] Rawlings, N.D. and Barrett, A.J. *MEROPS* – the peptidase database. Release 3.03. http://www.bi.bbsrc.ac.uk/Merops/Merops.htm
[10] Picard, V., Ersdal-Badju, E., Lu, A. and Bock, S.C. (1994) Nucleic Acids Res. 22, 2587–2591.
[11] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
[12] Götz, B. and Klinkert, M.-Q. (1993) Biochem. J. 290, 801–806.
[13] Yu, J., Nagarajan, S., Knez, J.J., Udenfriend, S., Chen, R. and Medof, M.E. (1997) Proc. Natl. Acad. Sci. USA 94, 12580–12585.
[14] Benghezal, M., Benachour, A., Rusconi, S., Aebi, M. and Conzelmann, A. (1996) EMBO J. 15, 6575–6583.
[15] Porter, W.H., Cunningham, L.W. and Mitchell, W.M. (1971) J. Biol. Chem. 246, 7675–7675.
[16] Potempa, J. and Travis, J. (1998) in: Handbook of Proteolytic Enzymes (Barrett, A.J., Rawlings, N.D. and Woessner, J.F., Eds.) pp. 762–768, Academic Press, London.
[17] Gilles, A.-M., De Wolf, A. and Keil, B. (1983) Eur. J. Biochem. 130, 473–479.
[18] Walker, N.P.C., Talanian, R.V., Brady, K.D., Dang, L.C., Bump, N.J., Ferenz, C.R., Franklin, S., Ghayur, T., Hackett, M.C., Hammill, L.D., Herzog, L., Hugunin, M., Houy, W., Mankovich, J.A., McGuiness, L., Orlewicz, E., Paskind, M., Pratt, C.A., Reis, P., Summani, A., Terranova, M., Welch, J.P., Xiong, L. and Möller, A. (1994) Cell 78, 343–352.
[19] Pavloff, N., Pemberton, P.A., Potempa, J., Chen, W.C.A., Pike, R.N., Prochazka, V., Kiefer, M.C., Travis, J. and Barr, P.J. (1997) J. Biol. Chem. 272, 1595–1600.
[20] Rost, B. (1996) Methods Enzymol. 266, 525–539.
[21] Witte, V., Wolf, N. and Dargatz, H. (1996) Curr. Microbiol. 33, 281–286.
[22] Brömme, D., Bonneau, P.R., Lachance, P. and Storer, A.C. (1994) J. Biol. Chem. 269, 30238–30242.
[23] Storer, A.C. and Ménard, R. (1996) Perspect. Drug Discov. Design 6, 33–46.
[24] Kembhavi, A.A., Buttle, D.J., Rauber, P. and Barrett, A.J. (1991) FEBS Lett. 283, 277–280.
[25] Rangarajan, M., Smith, S.J.M., U, S. and Curtis, M.A. (1997) Biochem. J. 323, 701–709.
[26] Kraulis, P.J. (1991) J. Appl. Crystallogr. 24, 946–950.
[27] Bairoch, A. and Bucher, P. (1994) Nucleic Acids Res. 22, 3583–3589.