Cloning, expression and characterisation of murine procathepsin E

Peter J. Tatnell, Wendy E. Lees, John Kay*

School of Molecular and Medical Biosciences, University of Wales, Cardiff, P.O. Box 911, Cardiff, CF1 3US Wales, UK

Received 18 February 1997; revised version received 26 March 1997

Abstract The cDNA encoding murine procathepsin E was isolated and sequenced and recombinant enzyme was produced in *Escherichia coli*. The activity of the purified recombinant mouse cathepsin E was characterised quantitatively using two synthetic peptide substrates and naturally occurring inhibitors. The majority of the recombinant enzyme was present as a homodimer (M_r ~80) in which the two monomers were linked by an intermolecular disulfide bond. By analogy to previous studies with human cathepsin E, this is most likely a consequence of the presence of a unique cysteine residue near the N-terminus of the mature proteinase. The availability of (i) recombinant murine enzyme in reasonable quantities and (ii) a full-length cDNA now enables structural investigations and attempts to generate 'knock-out' mice deficient in this important aspartic proteinase to be undertaken.

Key words: Murine procathepsin E, cloning; Murine procathepsin E, expression in *E. coli*; Recombinant cathepsin E; Characterization; Chromogenic substrate hydrolysis, inhibition

1. Introduction

Of the five aspartic proteinases currently documented to be produced in the human body, three (pepsin, gastricsin and renin) have well-defined physiological roles and the fourth, cathepsin D, is found ubiquitously in the lysosomes of most cells [1]. The fifth enzyme, cathepsin E, is readily distinguished from the others by its molecular architecture [2], its cytomorphological compartmentation [3] and its limited tissue and cellular distribution. The enzyme has been postulated to have roles in prohormone [4] and antigen processing [5] and in neurodegeneration, ischemia and ageing [6–8]. Definitive evidence in support of these putative roles has, however, remained elusive.

It was thus considered that one approach to establishing the physiological role(s) of cathepsin E unequivocally would be the generation of 'knock-out' mice lacking a functional *procathepsin E* gene. As a prelude to this extensive research programme, it was necessary to derive a cDNA encoding murine procathepsin E and to characterise the mouse enzyme.

*Corresponding author. Fax: (44) (1222)-874116.
E-mail: KayJ@Cardiff.ac.uk

Abbreviations: RT-PCR, reverse transcriptase polymerase chain reaction; RACE, rapid amplification of cDNA ends

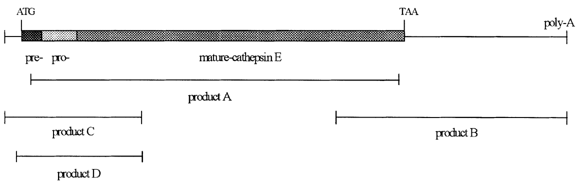
PII S0014-5793(97)00388-8

2. Materials and methods

2.1. Isolation of the cDNA sequence encoding murine procathepsin E

Specific amplification of the procathepsin E cDNA was achieved by a combination of RT-PCR and 5′- and 3′-RACE. The mRNA was isolated and purified from the spleen of BALB/c mice using the mRNA purification kit (Pharmacia Biotech Ltd., Milton Keynes, UK). First-strand cDNA synthesis was achieved by using reverse transcriptase with either random hexamers or a modified oligo(dT)$_{17}$ (see below). All oligonucleotides were purchased from Genosys Biotechnologies Inc., Cambridge, UK.

The initial PCR product was amplified from random-primed cDNA using degenerate primers (forward primer 1 = 5′-CTG CTG ITG CTC CTG GAI CTG G-3′ and reverse primer 2 = 5′-GGG AAC TGC IGG GGC CAI ICC CAC-3′ where I = inosine). The resultant (1167 bp) product A encompassed almost the full-length *procathepsin E* gene.



The 3′-RACE product (designated B) was amplified by utilising the strategy described previously [9], in which the first-strand cDNA was synthesised by using reverse transcriptase and the modified oligo(dT)$_{17}$ primer (CGG AGA TCT CCA ATG TGA TGG GAA TTC (T)$_{17}$). Specific PCR reactions were performed using the gene-specific forward primer 3 (5′-AT GGA ATG CAG TTC TGC GGC-3′) which was designed against authentic murine procathepsin E cDNA sequence and the reverse primer 4 corresponding to the oligo(dT)$_{17}$ primer used in the first-strand synthesis but devoid of the oligo(dT)$_{17}$ tail, i.e. (5′-CGG AGA TCT CCA ATG TGA TGG GAA TTC-3′). The resultant 760 bp product B (see above) contained a 152 bp overlap with product A and included the 3′-end flanking region and the poly-A sequence.

The 437 bp 5′-RACE product (designated C) was amplified according to standard procedures [10,11] utilising reverse transcriptase and the procathepsin E gene-specific reverse primer 5 (5′-GAT TCC TGT CAG GCT CCC-3′). Homopolymeric tailing of the cDNA was achieved using terminal transferase and dCTP. PCR reactions were performed using a nested reverse primer 6 (5′-CGA TGG ATG GAA TAC TGG G-3′) and a forward oligo(dG)$_{15}$ primer. The resultant product C overlapped product A by 351 bp and extended the sequence information to include an additional 47 bp in the 5′-flanking region prior to the start of translation. Despite many attempts, this 5′-RACE product was obtained on only one occasion. Consequently, to confirm the cDNA sequence of this 5-′ region, a new primer was designed (forward primer 7 = 5′-GI GII AAG CTG CII IIC IIA C-3′) and used in conjunction with the nested reverse primer 6 to enable the amplification of a 401 bp PCR product (designated D).

All PCR amplifications were performed using 10–20 ng of reverse transcribed mRNA, 1 U of Amplitaq DNA polymerase (Perkin Elmer Corp., Roche Molecular Systems Inc., New Jersey), 1 mM $MgCl_2$, 0.5 mM of each primer and 200 mM dNTPs in a final volume of 50 ml. An initial denaturation step of 94°C for 3 min was typically followed by 40 cycles of 94°C for 1 min, 60°C for 1 min and 72°C for 1.5 min, after which a final elongation step of 72°C for 5 min was performed using a Perkin Elmer Cetus DNA thermal cycler 480 programmable thermoblock. Each PCR product was amplified on a minimum of three separate occasions and each independent product was cloned

```
                                                -18                                                              -1
                                                 M   K   P   L   L   V   L   L   L   L   L   L   L   D   L   A   Q   A
aagggagatcggagcagagtgagagagaagctaccccttggaccaca ATG AAG CCC CTC CTT GTG CTG CTG CTG CTG CTG CTC CTG GAT CTG GCT CAG GCC   101
 1P                                 10P                                     20P                                     30P
 Q   G   A   L   H   R   V   P   L   R   R   H   Q   S   L   R   K   K   L   R   A   Q   G   Q   L   S   E   F   W   R    30P
CAA GGT GCT CTG CAC AGA GTG CCC CTC AGA AGA CAT CAG TCC CTT CGG AAG AAA CTA CGG GCC CAA GGA CAG CTC TCA GAA TTC TGG AGG   191
 31P                     ↓      39P  ↓1          ∇               ↓      10                                      20
 S   H   N   L   D   M   T   R   L   S   E   S   C   N   V   Y   S   S   V   N   E   P   L   I   N   Y   L   D   M   E     21
TCT CAT AAC TTG GAC ATG ACC CGA CTC AGC GAG TCC TGT AAT GTG TAT TCG AGT GTC AAT GAA CCC CTC ATC AAC TAC CTG GAT ATG GAA   281
                                30               ♦                       ▼                                      50
 Y   F   G   T   I   S   I   G   T   P   P   Q   N   F   T   V   I   F   D   T   G   S   S   N   L   W   V   P   S   I     51
TAC TTT GGC ACC ATC TCC ATC GGC ACC CCG CCG CAG AAC TTC ACT GTC ATC TTT GAC ACC GGT TCA TCC AAC CTC TGG GTC CCT TCT ATC   371
                                60                                      70                                      80
 V   Y   C   T   S   P   A   C   K   A   H   P   V   F   H   P   S   Q   S   D   T   Y   T   E   V   G   N   H   F   S     81
GTG TAC TGC ACC AGC CCA GCA TGC AAG GCA CAC CCA GTA TTC CAT CCA TCG CAG TCC GAC ACA TAC ACG GAG GTA GGG AAT CAT TTC TCC   461
                                90                                      100                                     110
 Q   Y   G   T   G   S   L   T   G   I   I   G   A   D   Q   V   S   V   E   G   L   T   V   D   G   Q   Q   F   G   E    111
CAG TAT GGT ACC GGG AGC CTG ACA GGA ATC ATT GGA GCT GAT CAA GTC TCT GTG GAA GGG TTG ACT GTG GAT GGC CAG CAG TTT GGA GAA   551
                                120                                     130                                     140
 S   V   K   E   P   G   Q   T   F   V   N   A   E   F   D   G   I   L   G   L   G   Y   P   S   L   A   A   G   G   V    141
AGT GTC AAG GAG CCA GGC CAG ACC TTT GTG AAT GCA GAG TTT GAT GGG ATT CTG GGT CTG GGA TAC CCC TCA TTG GCT GCT GGA GGA GTG   641
                                150                                     160                                     170
 T   P   V   F   D   N   M   M   A   Q   N   L   V   A   L   P   M   F   S   V   Y   L   S   S   D   P   Q   G   G   S    171
ACC CCA GTG TTT GAC AAC ATG ATG GCC CAG AAC CTT GTG GCT CTG CCT ATG TTT TCT GTC TAC TTG AGC AGT GAC CCT CAA GGT GGC TCA   731
                                180                                     190                                     200
 G   S   E   L   T   F   G   G   Y   D   P   S   H   F   S   G   S   L   N   W   I   P   V   T   K   Q   A   Y   W   Q    201
GGC AGT GAG CTG ACT TTC GGA GGC TAT GAC CCC TCT CAT TTC TCT GGG AGC CTC AAC TGG ATT CCA GTC ACC AAG CAA GCC TAT TGG CAG   821
                                210                                     220                  ▼                  230
 I   A   L   D   G   I   Q   V   G   D   T   V   M   F   C   S   E   G   C   Q   A   I   V   D   T   G   T   S   L   I    231
ATT GCC TTG GAT GGA ATC CAG GTG GGA GAC ACT GTG ATG TTC TGC TCC GAA GGC TGT CAG GCC ATA GTG GAC ACA GGG ACC TCT CTC ATC   911
                                240                                     250                                     260
 T   G   P   P   D   K   I   K   H   L   Q   E   A   I   G   A   T   P   I   D   G   E   Y   A   V   D   C   A   T   L    261
ACT GGC CCC CCC GAC AAG ATC AAA CAT CTT CAA GAG GCC ATT GGG GCC ACA CCC ATT GAT GGA GAA TAT GCA GTG GAT TGT GCC ACT CTC  1001
                 ♦              270                                  ♦  280                                     290
 D   T   M   P   N   V   T   F   L   I   N   E   V   S   Y   T   L   N   P   T   D   Y   I   L   P   D   L   V   D   G    291
GAC ACG ATG CCA AAC GTT ACC TTC CTC ATC AAC GAG GTT TCA TAT ACC CTC AAC CCA ACT GAC TAC ATC CTG CCG GAC TTG GTG GAT GGA  1091
                                300                                     310                                     320
 M   Q   F   C   G   S   G   F   Q   G   L   D   I   P   P   P   A   G   P   L   W   I   L   G   D   V   F   I   R   Q    321
ATG CAG TTC TGC GGC AGT GGC TTT CAA GGA CTT GAC ATT CCA CCT CCA GCT GGG CCC CTC TGG ATC CTG GGG GAT GTC TTC ATC CGA CAG  1181
                                330                                     340
 F   Y   S   V   F   D   R   G   N   N   Q   V   G   L   A   P   A   V   P   *                                            340
TTC TAC TCA GTC TTT GAC CGT GGA AAT AAC CAA GTG GGA TTG GCC CCC GCA GTT CCC TAA agagggatgtatgcctacatatggatgcctgataccat  1280

ttaacctgttagatacctttgtaactatcaaagccgtcatttcccatggggtgtagccaccccagagtattcagaccaatcaaagcataagagtgcaccccactcactgcaaacacaca  1399
cacacacacaccacctctaccatcaccacgatggaagaagttctgtctatagtcttcactgcttattgttgactttctattatggaaatctctaaacatgtacacagtagacgtgatgg  1518
caagataaatacccacacacctctgcctcaggtcacaacccatccatgtgtggaccagactctctatcttccatccctctggttccacgcctctagatttggaagcagattctaagcac  1637
caggtcattttatctaatgtctaacatccttacaaatcagaatttaaatgcctcaccctctcataaatgtggatctgtttttttacagttggtttatttgtatcaggattaaaaccagat  1756
ccataactggacaaaaaaacccataactgatttgattttaaatatc(a)n                                                                        1802
```

Fig. 1. The nucleotide and deduced amino acid sequence of the cDNA encoding murine procathepsin E. The residues in the putative signal peptide (numbers −18 to −1) are shown in italics. The Gln residue (indicated by 1P) is the putative first amino acid of the pro-part. Residues 1–340 represent the sequence of mature cathepsin E. The positions of the two active-site aspartic acid residues, the three potential N-linked glycosylation sites and the Cys residue involved in the formation of the interchain disulfide bridge are indicated with ▼, ♦ and ▽, respectively. The vertical arrows show the cleavage sites where the recombinant precursor undergoes auto-activation.

using either the Promega pGEM-T PCR product cloning system (Promega Ltd., Southampton, UK) or the Pharmacia Sureclone Ligation Kit. Elucidation of the nucleotide sequences was achieved by dideoxy sequencing of both DNA strands using the Thermo Sequenase Kit (Amersham International, Amersham, UK) with primers labelled with fluorescent dyes and a Hybaid Licor Automated DNA Sequencer.

2.2. Expression of procathepsin E in E. coli

To facilitate the expression of recombinant murine procathepsin E in *Escherichia coli*, the DNA sequence encoding the propart and mature enzyme regions was amplified by PCR using the forward primer 8 = 5′-G GAA TTC CAT ATG CAA GGT GCT CTG CAC AGA-3′ and the reverse primer 9 = 5′-CCC AGA TCT TTA GGG AAC TGC GGG GGG-3′. These primers contain *Nde*I and *Bgl*II restriction sites, respectively, to facilitate subsequent subcloning into the T7-RNA polymerase expression vector pET16b (AMS Biotech Ltd., Witney, Oxon, UK). DNA amplification was performed using 100 ng of cDNA encoding murine procathepsin E, 1 U of Vent polymerase (New England Biolabs, Beverly, MA), 2 mM $MgSO_4$, 0.5 mM of each primer and 200 mM dNTPs in a final volume of 50 ml. An initial denaturation step of 94°C for 3 min was followed by 10 cycles of 94°C for 1 min, 63°C for 1 min and 72°C for 1 min, after which a final elongation step of 72°C for 5 min was performed. The resultant 1162 bp product was cloned into pUC18 using the Sureclone Ligation Kit and, after authentication of the nucleotide sequence, the fragment was excised and subcloned into the *Nde*I/*Bam*HI sites of pET16b. Expression was carried out in *E. coli* BL21(DE3) pLysS as described previously [12,13].

The analysis of recombinant protein samples by SDS-PAGE and Western blotting, enzyme purification, N-terminal sequence analysis and determination of kinetic parameters for chromogenic substrate hydrolysis and inhibitor binding were carried out as described previously [13,14].

3. Results and discussion

3.1. Analysis of DNA sequence encoding murine procathepsin E

The nucleotide sequence of the cDNA-encoding murine procathepsin E (see Fig. 1) was determined by analysing both strands of the amplified DNA products generated as described in Section 2. It has been deposited in the EMBL, Genbank and DDBJ databases under the accession number X97399. The sequence consisted of 47 bp of 5′-untranslated region, a 1194 bp open-reading frame and 561 bp of 3′-flanking region prior to the poly(A) tail. The open-reading frame predicts a 397-amino acid polypeptide (calculated M_r $\sim$42 937) consisting of three regions; a signal peptide (18 residues), a propart region of 39 residues and the mature enzyme (340 residues). The hallmark which distinguishes cathepsin E from all other aspartic proteinases was readily evident in the murine sequence. This is the unique Cys^{4} residue (Fig. 1) which forms an intermolecular disulfide bond resulting in cathepsin E existing as a homodimer of $\sim$80 kDa [15].

The coding sequence of murine procathepsin E and the deduced amino acid sequence were compared to those known for procathepsin E from other species. The identities (cDNA and amino acid, respectively) were 82% and 79% with human [16], 81% and 79% with guinea pig [17], 80% and 79% with rabbit [18] and 93% and 91% with rat [19] procathepsin E.

The alignment is shown in Fig. 2. From this, a number of distinguishing features are apparent.

```
       -18     -10         1P       10P       20P       30P      +1  ▽     10        20
Mouse   MKPLLVLLLLLLL-DLAQAQGALHRVPLRRHQSLRKKLRAQGQLSEFWRSHNLDMTRLSESCNVYSSVNEPLINYLDMEY
Rat         F        L       V                       D         IEF      DKGI
Rabbit    T --P      L  G    T D      QP                KA KV  VQYT T TMEQ A
Human     T --     V LE GE   S         P  K     RS      K      IQFT   SMDQ AK
G.pig     TF--     V LE G  P        S RE            T L K Q  N DQ---- STIQ A

               30  ♦     ▼         50        60        70        80        90        100
Mouse   FGTISIGTPPQNFTVIFDTGSSNLWVPSVYCTSPACKAHPVFHPSQSDTYTEVGNHFSIQYGTGSLTGIIGADQVSVEGL
Rat        V   S S                                     S  M
Rabbit         S           V                QM  Q R    N  S   TP   A                 Q
Human          S                             T SR Q    S  SQP QS          S
G.pig          S                            QT       L S  R    S

               110       120       130       140       150       160       170       180
Mouse   TVDGQQFGESVKEPGQTFVNAEFDGILGLGYPSLAAGGVTPVFDNMMAQNLVALPMFSVYLSSDPQGGSGSELTFGGYDP
Rat       E                                V
Rabbit    V                                                 S       M  N E             S
Human     V        T       D               V                D       M  N E  A    I     H
G.pig     V        Q   K   H                                        M  N -

               190       200       210       220  ▼    230       240       250       260
Mouse   SHFSGSLNWIPVTKQAYWQIALDGIQVGDTVMFCSEGCQAIVDTGTSLITGPPDKIKHLQEAIGATPIDGEYAVDCATLD
Rat                    G                                     K   Q         M           N
Rabbit           V     G       E    GSP   P                 S   IQ  A      M      E EN N
Human            V             N    G                       S    Q  N    A V      E  N N
G.pig            V                   S                       G   Q    L   YV EG S Q  N N

           ♦   270      ♦280       290       300       310       320       330       340
Mouse   TMPNVTFLINEVSYTLNPTDYILPDLVDGMQFCGSGFQGLDIPPPAGPLWILGDVFIRQFYSVFDRGNNQVGLAPAVP.
Rat     M         G     S  A                      Q               K
Rabbit  I  D   V  G P   SA A T   F                Q                        S R
Human   V  D   T  G P   S  A T L F       S        H                          R
G.pig   M LD   I  G P      A T L F     V ST  E  E Q                  A       R
```

Fig. 2. Alignment of the deduced amino acid sequences of procathepsin E from mouse, rat, rabbit, human and guinea pig (G. pig). The sequence for the murine procathepsin E is given in full. At any given position where no residue is shown, sequence identity with murine procathepsin E exists. (−) indicates the absence of a residue. The positions of the active-site aspartic acid residues, potential N-linked glycosylation sites and the Cys residue involved in the formation of the interchain disulfide bridge are indicated by ▼, ♦ and ▽, respectively.

(i) In addition to the N-linked oligosaccharide attachment site which is present in the enzyme from all of the species (∼Asn^{34}–X–Thr^{36}∼, see Fig. 2), murine cathepsin E contains two additional potential N-glycosylation sites located towards the C-terminus (Asn^{266} and Asn^{279}), i.e. murine cathepsin E possesses three sites in total. The glycosylation site at Asn^{266} is present in rat cathepsin E but not in the guinea pig sequence; whereas the 'motif' at Asn^{279} is present in the guinea pig sequence but not in rat cathepsin E. Thus, rat and guinea pig enzymes each have two N-glycosylation sites while cathepsin E of human and rabbit origin has only the one ∼Asn^{34}–X–Thr^{36}∼ 'motif' (Fig. 2). The extent and type of N-glycosylation at the additional Asn–X–Thr 'motifs' remains to be elucidated [15].

(ii) Whereas the unique Cys residue is conserved (as described above), the context of the sequence in which it is located is the most highly variable of the entire procathepsin E molecule (Fig. 2).

(iii) Modelling of the sequence of the murine cathepsin E on the basis of its homology to human cathepsin E [2] suggests that, of those residues predicted to make contact with features of a substrate or inhibitor bound in the active site [2,20], only one residue differs significantly between the mouse and the human enzymes. In mouse cathepsin E, the residues which contribute to the structural feature of aspartic proteinases known as the polyproline loop [21] are ∼Pro^{305}–Pro^{306}–Pro^{307}∼ (Fig. 2); in human cathepsin E, this sequence is ∼His–Pro–Pro∼ while in the guinea pig, rat and rabbit enzymes, the sequence is ∼Gln–Pro–Pro∼. In order to investigate whether such changes might influence the nature and activity of the cathepsin E molecule significantly, the DNA encoding the pro- and mature regions of murine procathepsin

Table 1
Kinetic constants for the interactions of chromogenic substrates and inhibitors with recombinant mouse and human cathepsin E

Substrate or inhibitor	Mouse	Human	Mouse	Human	Mouse	Human	Mouse	Human
	K_i (nM)		K_m (mM)		k_{cat} (s^{-1})		k_{cat}/K_m ($mM^{-1}s^{-1}$)	
Substrate 1	–	–	50	65	130	70	2600	1070
Substrate 2	–	–	60	160	110	125	1830	780
Isovaleryl-pepstatin	< 0.1	< 0.1	–	–	–	–	–	–
Ascaris protein	1.0	3.0	–	–	–	–	–	–

Substrate 1 = Pro–Pro–Thr–Ile–Phe*Nph–Arg–Leu and Substrate 2 = Lys–Pro–Ile–Glu–Phe*Nph–Arg–Leu. (Nph = nitrophenylalanine).
Substrate 2 was used for the evaluation of the inhibitors. The estimated precision of all values obtained was in the range 10–15%.

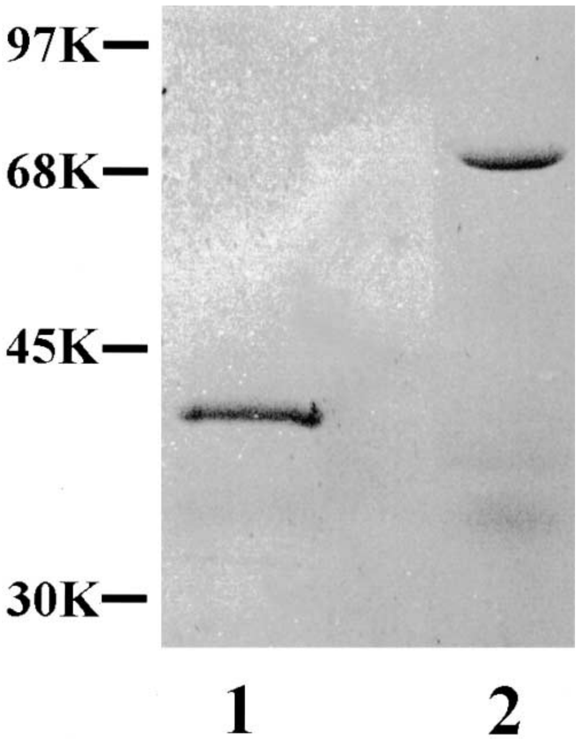


Fig. 3. SDS-PAGE of samples containing recombinant mouse cathepsin E. Samples (containing ≈1 mg of protein) were analysed under reducing conditions in the presence of 50 mM β-mercaptoethanol (lane 1) or in the absence of reducing agent (lane 2). Staining was with Coomassie Blue and molecular mass markers (kDa) migrated as indicated.

E was sub-cloned into pET16b to facilitate the production of recombinant enzyme.

3.2. Expression and characterisation of murine procathepsin E in E. coli

The resultant recombinant plasmid was transformed into *E. coli* and induced with 1 mM IPTG. After 3 h, the cells were harvested, lysed and the zymogen was extracted with urea and re-folded by rapid dilution at pH 9.5 [12,13]. After concentration, the pH of the suspension was adjusted to pH 3.1 and the solution was incubated for 1 h at 37°C in order to generate mature cathepsin E by autoactivation of the recombinant precursor. Following centrifugation, the supernatant was neutralised, centrifuged again and aliquots were taken for further analyses.

Analysis by SDS-PAGE under reducing conditions (Fig. 3) revealed a single homogeneous protein band of approximately ∼40 kDa consistent with the molecular mass predicted from the deduced amino acid sequence of the mature form of murine cathepsin E. In contrast, when the electrophoresis was performed under non-reducing conditions, the predominant band observed migrated at ∼80 kDa (Fig. 3), indicating that the recombinant enzyme was largely present in the form of a disulfide-linked homodimer. A very faint band migrating slightly faster under the non-reducing conditions (Fig. 3, lane 2) than the single band obtained in the presence of β-mercaptoethanol (Fig. 3, lane 1) confimed that only a very small proportion of the protein existed as the monomer. Samples of the purified murine enzyme were subjected to Edman degradation but a single unique N-terminal sequence was not observed; rather, three distinct overlapping sequences were elucidated. All were exactly coincident with the deduced amino acid sequence for the mature murine cathepsin E as predicted by the cDNA sequence. These N-terminal isoforms were present at a ratio of 2:1:0.1, respectively.

Sequence 1	T-R-L-S-E-S-C-N-V-Y-S-S-V-N-E-P-L-I-N-Y-L∼
Sequence 2	S-E-S-C-N-V-Y-S-S-V-N-E-P-L-I-N-Y-L∼
Sequence 3	S-V-N-E-P-L-I-N-Y-L∼

The predominant N-termini (sequences 1 and 2) are positioned identically to the sites located in the human procathepsin E sequence at which it has been shown previously [13] that autoactivation of the zymogen takes place. This leads to the N-terminal microheterogeneity of the resultant mature enzyme. It is evident that, in both cases, these activation sites are situated immediately upstream of the unique cysteine residue of cathepsin E so that the intermolecular disulfide bond is positioned within the mature enzyme (and not in the propart region).

Two ∼40 kDa molecules are thus connected to generate the ∼80 kDa moiety observed by SDS-PAGE under non-reducing conditions. The importance of this disulfide bond for enzymic stability and correct intracellular trafficking of cathepsin E has been described previously [12,22]. The third, minor sequence detected indicated that a small proportion of the zymogen molecules (<5%) had undergone processing downstream from the cysteine residue. In this case, the resultant mature cathepsin E could not be dimeric and, indeed, SDS-PAGE did show a minor component (the lowest band in lane 2, Fig. 3) with a slightly lower molecular mass under non-reducing conditions. The position of this minor activation site is similar to one of the processing sites observed upon activation of the Cys^4Ala mutant form of human cathepsin E [12]. Perhaps it should be emphasised that mature cathepsin E, irrespective of the species of origin, is longer at its N-terminus by 8 or 11 residues (as a result of the observed N-terminal microheterogeneity) by comparison with other aspartic proteinases of fungal or vertebrate species (such as pepsin or cathepsin D). One consequence of this is that the intermolecular disulfide bond is located in this N-terminal 'extension' to the archetypal aspartic proteinase molecule; the corollary is that the propart released upon autoactivation of procathepsin E is necessarily shorter than those generated from other aspartic proteinase precursors by an equivalent number of residues. In this regard, it is noteworthy that, in the propart region of many aspartic proteinase precursors, there is a lysine residue in a ∼Lys–Tyr∼ 'motif' [23] which is believed to play a key role in the folding and 3-dimensional structure of the zymogens [24]. In the propart region of mouse (and other) procathepsin E sequences, this ∼Lys–Tyr∼ 'motif' is not only absent but taking its place are the residues which contribute the predominant autoactivation cleavage sites. In this context, then, the N-terminus of the minor component observed upon autoactivation of murine procathepsin E is located almost exactly coincident with the N-terminal residues of other aspartic proteinases such as pepsin, gastricsin and cathepsin D [2,15].

The concentration of purified recombinant murine cathepsin E was determined by active site titration using isovaleryl-pepstatin as described in Section 2. On this basis, it was estimated that the yield of active enzyme was ∼5 mg/l of

E. coli culture. This is considerably higher (≈10- and 3-fold, respectively) than those obtained previously for recombinant wild-type and mutant (Cys^4Ala) human enzymes [12,13] and was sufficient to enable crystallisation trials to begin with a view to the structural determination of murine cathepsin E.

3.3. Kinetic parameters for the interaction of substrates and inhibitors with recombinant murine cathepsin E

The ability of the purified recombinant murine enzyme to hydrolyse two chromogenic peptide substrates was compared with that of human cathepsin E. The kinetic parameters obtained are listed in Table 1. For each substrate, the specificity constant derived for the murine enzyme was comparable (∼2.3-fold higher) to the values derived for the recombinant human enzyme and closely similar to those reported previously for naturally occurring cathepsin E isolated from human and rat tissues [2,13,14]. Sub-nanomolar K_i values were determined for the interaction of isovaleryl-pepstatin with the recombinant mouse and human enzymes (Table 1). A further distinguishing feature of cathepsin E is its susceptibility to inhibition by a protein (M_r ∼17) from the parasitic worm, *Ascaris lumbricodes* [25]. A K_i value of 1 nM was determined for the interaction of the recombinant murine enzyme (Table 1) with this protein inhibitor; this value is at most only 2–3-fold different from the value determined for inhibition of human cathepsin E.

In terms of its interactions with selected substrates and inhibitors, murine cathepsin E differs only marginally from its human counterpart, and this may be related to differences in the sequences of the two proteins. Whether the small changes observed are in themselves significant may be discerned from a more detailed investigation using mutagenesis together with a much larger series of synthetic substrates with systematic variation of individual residues [2]. In its other characteristics, murine cathepsin E would appear to be virtually indistinguishable from its human counterpart and would thus serve as a model to facilitate investigations into the human enzyme. To this end, the murine cDNA has been used as a probe to isolate a genomic clone spanning the entire murine *procathepsin E* gene locus (W. Roth, C. Peters, J. Kay and P. Tatnell, unpublished). This should facilitate attempts to generate a 'knock-out' mouse lacking the gene for this important aspartic proteinase.

Acknowledgements: This work was supported by a Wellcome Trust Grant. N-terminal sequence analysis was performed within the Protein Chemistry Core Facility, University of Florida, Gainesville, FL, USA. We are most grateful to Drs. B.M. Dunn (University of Florida), D. Bur (Hoffman La Roche, Basel, Switzerland), and Y. Peng Loh (National Institute of Health, Bethesda, MD) for their valuable contributions to this research.

References

[1] P. Saftig, M. Hetman, W. Schmahl, K. Weber, L. Heine, H. Mossmann, A. Köster, B. Hess, M. Evers, K. von Figura, C. Peters, EMBO J. 14 (1995) 3599–3608.
[2] C. Rao-Naik, K. Guruprasad, B. Batley, S. Rapundalo, J. Hill, T. Blundell, J. Kay, B.M. Dunn, Proteins: Struct. Funct. Genet. 22 (1995) 168–181.
[3] T. Kageyama, Meth. Enzymol. 248 (1995) 120–136.
[4] W.E. Lees, S. Kalinka, J.D. Meech, S.J. Capper, N.D. Cook, J. Kay, Biochem. J. 273 (1990) 99–102.
[5] K. Bennet, T. Levine, J.S. Ellis, R.J. Peanasky, I.M. Samloff, J. Kay, B.M. Chain, Eur. J Immunol. 22 (1992) 1519–1524.
[6] H.G. Bernstein, B. Wiederanders, Brain Res. 667 (1994) 287–290.
[7] H. Nakanishi, K. Tominaga, T. Amano, I. Hirotsu, T. Inoue, K. Yamamoto, Exp. Neurol. 126 (1994) 119–128.
[8] H. Nakanishi, T. Tsukuba, T. Kondou, T. Tanaka, K. Yamamoto, Exp. Neurol. 121 (1993) 215–223.
[9] P. Kille, P.E. Stephens, J. Kay, Biochim. Biophys. Acta 1089 (1991) 407–410.
[10] M.A. Frohman, M.K. Dush, G.R. Martin, Proc. Natl. Acad. Sci. 85 (1988) 8998–9002.
[11] R.L. Dorit, O. Ohara, W. Gilbert, Meth. Enzymol. 218 (1993) 36–47.
[12] S.D. Fowler, J. Kay, B.M. Dunn, P.J. Tatnell, FEBS Lett. 366 (1995) 72–74.
[13] J. Hill, D.S. Montgomery, J. Kay, FEBS Lett. 326 (1993) 101–104.
[14] R.A. Jupp, A.D. Richards, J. Kay, B.M. Dunn, J.B. Wyckoff, I.M. Samloff, K. Yamamoto, Biochem. J. 254 (1988) 895–898.
[15] Kay, J. and Tatnell, P.J. (1997) In Handbook of proteolytic enzymes (Barrett, A.J., Rawlings, N.D. and Woessner, J.F., eds.), in press, Academic Press, London.
[16] T. Azuma, G. Pals, T.K. Mohandas, J.M. Couvreur, R.T. Taggart, J. Biol. Chem. 264 (1989) 16748–16753.
[17] T. Kageyama, M. Ichinose, S. Tsukada, K. Miki, K. Kurokawa, O. Koiwai, M. Tanji, E. Yakabe, S.B.P. Athauda, K. Takahashi, J. Biol. Chem. 267 (1992) 16450–16459.
[18] T. Kageyama, Eur. J. Biochem. 216 (1993) 717–728.
[19] K. Okamoto, Y. Hao, Y. Misumi, Y. Ikehara, K. Yamamoto, Arch. Biochem. Biophys. 322 (1995) 103–111.
[20] M. Tanji, E. Yakabe, T. Kageyama, K. Takahashi, J. Biochem. 120 (1996) 647–656.
[21] B.L. Sibanda, T. Blundell, P.M. Hobart, M. Fogliano, J.S. Bindra, B.W. Dominy, J.M. Chirgwin, FEBS Lett. 174 (1984) 102–111.
[22] E.M. Finley, S. Kornfeld, J. Biol. Chem. 269 (1994) 31259–31266.
[23] G. Koelsch, M. Mares, P. Metcalf, M. Fusek, FEBS Lett. 343 (1994) 6–10.
[24] M.N.G. James, A.R. Sielecki, Nature 319 (1986) 33–38.
[25] M.R. Martzen, B.A. McMullen, N.E. Smith, K. Fujikawa, R.J. Peanasky, Biochemistry 29 (1990) 7366–7372.