Cloning and expression of mouse legumain, a lysosomal endopeptidase

Jinq-May CHEN[1], Pam M. DANDO, Richard A. E. STEVENS, Mara FORTUNATO and Alan J. BARRETT

MRC Peptidase Laboratory, The Babraham Institute, Babraham, Cambridge CB2 4AT, U.K.

Legumain, a recently discovered mammalian cysteine endopeptidase, was found in all mouse tissues examined, but was particularly abundant in kidney and placenta. The distribution in subcellular fractions of mouse and rat kidney showed a lysosomal localization, and activity was detectable only after the organelles were disrupted. Nevertheless, ratios of legumain activity to that of cathepsin B differed considerably between mouse tissues. cDNA encoding mouse legumain was cloned and sequenced, the deduced amino acid sequence proving to be 83% identical to that of the human protein [Chen, Dando, Rawlings, Brown, Young, Stevens, Hewitt, Watts and Barrett (1997) J. Biol. Chem. **272**, 8090–8098]. Recombinant mouse legumain was expressed in human embryonic kidney 293 cells by use of a vector containing a cytomegalovirus promoter. The recombinant enzyme was partially purified and found to be an asparagine-specific endopeptidase closely similar to naturally occurring pig kidney legumain.

INTRODUCTION

Legumain (EC 3.4.22.34) is a cysteine endopeptidase belonging to a family that was, until recently, known only from plants and a few invertebrate animals [1–4]. In plants, the enzyme is present in leguminous and other seeds. Legumain isolated from seeds such as moth bean (*Vigna aconitifolia*) or jack bean (*Canavalia ensiformis*) and from the blood fluke (*Schistosoma mansoni*) is specific for the hydrolysis of asparaginyl bonds [2,5,6]. Recently we discovered that legumain is also present in mammals. Human legumain was cloned and sequenced, and pig legumain was isolated from kidney cortical tissues and characterized [7]. Mammalian legumain was shown to be active in acidic conditions (pH 3–6), and was also specific for cleaving asparaginyl bonds. Its enzymic activity was inhibited by both chicken egg-white cystatin and human cystatin C, but not by compound E-64 [L-3-carboxy-2,3-*trans*-epoxypropionyl-leucylamido(4-guanidino)-butane], a potent inhibitor of many cysteine peptidases of the papain family, including most lysosomal cathepsins.

There is much to be learnt about the biochemistry and biological functions of mammalian legumain, and the mouse is a species that can offer suitable experimental systems. As a first step towards establishing such systems, we have now surveyed the distribution of legumain in the mouse at the tissue and subcellular levels, cloned and sequenced the cDNA for mouse legumain, and expressed and characterized the recombinant enzyme.

MATERIALS AND METHODS

Materials

Neurotensin was from Sigma, and Z-Ala-Ala-Asn-NHMec [where NHMec is 7-(4-methyl)coumarylamide and Z is benzyloxycarbonyl] was purchased from Dr. C. Graham Knight (Department of Biochemistry, University of Cambridge, Cambridge, U.K.). Cystatin from chicken egg white was prepared as described [8]. The His-tagged recombinant C-fragment of tetanus toxoid was kindly given by Dr. Colin Watts (Department of Biochemistry, University of Dundee, Dundee, Scotland, U.K.). Human embryonic kidney (HEK) 293 cells (no. CRL 1573) were from American Type Culture Collection (Rockville, MD, U.S.A.). All reagents for tissue culture were purchased from Life Technologies (Paisley, Scotland, U.K.).

Enzyme assays

The enzymic activities of cathepsin B and legumain were measured by continuous fluorimetric assay as previously described [7,9]. Briefly, the substrates were Z-Arg-Arg-NHMec for cathepsin B and Z-Ala-Ala-Asn-NHMec for legumain. The assays were performed at 30 °C in legumain assay buffer (0.1 M sodium citrate, pH 5.8, containing 1 mM EDTA, 1 mM dithiothreitol and 0.01% CHAPS). Rates of hydrolysis were determined in a Perkin–Elmer LS3B fluorimeter under the control of an IBM-compatible computer running the FLUSYS software [10]. The fluorescent product of hydrolysis was quantified at excitation and emission wavelengths of 360 nm and 460 nm respectively. One unit of activity was defined as that releasing 1 μmol of product/min under the standard conditions, and values for specific enzymic activity were calculated by use of protein concentration values determined in the Bradford assay [11].

Mouse tissues and subcellular fractionation of kidney

BALB/c mice (3 weeks old) were obtained from the Babraham Institute small animal barrier unit. The mice were killed by cervical dislocation immediately before tissues were removed. Tissues were homogenized in 50 mM sodium citrate buffer, pH 5.8, containing 0.1 M NaCl, 1 mM EDTA and 2 mM 2-mercaptoethanol. Ten up-and-down strokes were used in a Braun Potter S homogenizer running at 1000 rev./min. After homogenization, three cycles of freezing and thawing were performed to disrupt organelles such as lysosomes. The homogenate was centrifuged at 12000 ***g*** for 10 min and the supernatant was collected for the quantification of legumain and

Abbreviations used: CMV, cytomegalovirus; E-64, L-3-carboxy-2,3-*trans*-epoxypropionyl-leucylamido(4-guanidino)butane; EST, expressed sequence tag; HEK, human embryonic kidney; NHMec, 7-(4-methyl)coumarylamide; Z, benzyloxycarbonyl.

[1] To whom correspondence should be addressed (e-mail jinq-may.chen@bbsrc.ac.uk).

The nucleotide sequence data reported have been submitted to the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under accession no. AJ000990.

cathepsin B activities, and for determination of protein concentration.

The procedure of Maunsbach [12] was adapted to prepare subcellular fractions from rat and mouse kidneys. The cortex of fresh kidney was dissected and homogenized in 8 parts (v/w) of 0.3 M sucrose in 0.1 M sodium citrate, pH 5.5, 1 mM EDTA and 2 mM β-mercaptoethanol. The homogenate was first centrifuged at 150 ***g*** for 10 min to sediment nuclei and unbroken cells. The pellet was designated the nuclear fraction. The supernatant was collected and centrifuged again at 9000 ***g*** for 3 min, excluding times for acceleration and deceleration. The pellet consisted of three layers of different colours: the dark brown bottom layer (lysosomal fraction), the yellow–brown middle layer (mitochondrial fraction) and the whitish top layer (brush border membranes). The three fractions were separated and collected by resuspending each in the citrate buffer containing 0.3 M sucrose. The supernatant from the 9000 ***g*** spin was further centrifuged at 100000 ***g*** for 1 h. The final supernatant was defined as the cytosolic fraction and the pellet was designated the microsomal fraction. Samples from these subcellular fractions were first assayed for legumain and cathepsin B activity in the citrate buffer containing 0.3 M sucrose, under which conditions most organelles remained intact. A portion (500 μl) of each sample was then frozen and thawed three times to disrupt membranes of organelles. The activities of legumain and cathepsin B in these fractions were measured in the assay buffer without sucrose.

Production of anti-legumain antibodies

Legumain purified from pig kidney [7] was used as an immunogen to raise polyclonal antisera in sheep. Briefly, 60 μg of purified legumain in 0.1 M citrate buffer, pH 5.8, and 1 mM EDTA was used for the initial injection, followed by two boosts 4 and 6 weeks later. For the initial injection, the antigen solution was mixed with an equal volume of Freund's complete adjuvant; incomplete adjuvant was used for the subsequent injections. The antisera were collected and the antibody against legumain was affinity-purified from a column of Sepharose beads conjugated with pig legumain by use of the procedure described previously [13].

Primary culture of rat kidney cells and immunolabelling

A kidney was removed from a 12-week old rat and the cortical tissue was dissected and minced under sterile conditions. The minced tissue was stirred at 30 °C in a solution containing 0.05 % trypsin and 0.02 % EDTA to dissociate cells. The solution became turbid after about 20 min. Tissue aggregates were removed by sedimentation and discarded, and dissociated cells were collected by centrifugation at 400 ***g*** for 5 min. The trypsinized cells were plated on cover slips coated with rat-tail collagen to facilitate their attachment [14]. The cells were cultured in sodium pyruvate-free, high-glucose Dulbecco's modified Eagle's medium supplemented with 10 % (v/v) fetal bovine serum. At 4 days after plating, cells were fixed in 3 % (w/v) paraformaldehyde in PBS (137 mM NaCl, 2.7 mM KCl, 4.3 mM Na_2HPO_4, 1.47 mM KH_2PO_4, pH 7.1) and permeabilized in 0.4 % Triton X-100 in PBS. Cells were incubated with the affinity-purified sheep anti-legumain antibody, followed by FITC-conjugated anti-(sheep immunoglobin) secondary antibody (Sigma). After immunostaining, cells were examined using a Zeiss Axiophot microscope equipped with an epifluorescence light source and a filter set for fluorescein, and photographed using a Zeiss camera system.

SDS/PAGE and immunoblotting

Electrophoresis was in the buffer system of Laemmli [15], and transfer of proteins to nitrocellulose was as described by Towbin et al. [16]. The blot was first incubated in blocking solution containing 5 % horse serum in PBS for 2 h, and then probed with the affinity-purified sheep anti-legumain antibody (10 μg/ml IgG). The incubation with primary antibody was at 20 °C for 2 h, followed by a PBS wash. Alkaline-phosphatase-conjugated secondary antibody (Sigma) was applied, and the colour reaction was developed using 0.005 % 5-bromo-4-chloro-3-indolyl phosphate as the substrate in 100 mM Tris/HCl, pH 9.3, containing 0.005 % Nitro Blue Tetrazolium, 50 mM $MgCl_2$ and 100 mM NaCl [17].

Database search and sequence analysis of mouse legumain clones

To identify mouse clones homologous to human legumain, we used nucleotide and amino acid sequences of both human (EMBL accession no. Y09862) and pig [7] legumain cDNAs as query sequences to search for homologous expressed sequenced tags (ESTs) with the BlastN and TBlastN programs [18]. The sequences of the mouse EST clones identified as potential legumain homologues were analysed and assembled into a full-length legumain sequence using the Fragment Assembly system of the GCG package [19]. By use of the Translate and BestFit programs, the amino acid sequence deduced from the mouse legumain cDNA assembled from the EST clones was aligned and compared with the known sequence of human legumain.

Five mouse EST clones with overlapping sequences (W14837, W36795, W83121, AA000961 and AA107459) were obtained from Research Genetics, Inc. (Huntsville, AL, U.S.A.). The insert in each clone was first analysed by PCR using vector-specific primers and legumain-gene-specific primers, and by restriction endonuclease digestion. Legumain-specific primers were generated by the GCG Prime program based on the sequences of mouse legumain EST clones. Two pairs of primers, Mcf/Mcr (5′-CCAAGGAGTCATCTTATGCC-3′/5′-TGTCTCTCAGCAGTTTCCCC-3′) and Mef/Mer (5′-AGCACTGGGTGGTGATTGTG-3′/5′-ACATCTGTGCCGTTAGGTCG-3′), were selected and used in this study. cDNA from each clone was propagated and purified by standard procedures [20]. Nucleotide sequences of the entire inserts of clones AA107459 (from a mouse embryo 8.5 days *post coitus*) and W83121 (from a 19-week-old mouse) were determined by Taq DyeDeoxy terminator cycle sequencing reactions in conjunction with an Applied Biosystems Model 373 DNA Sequencer. The primers for cycle sequencing included T3 and T7 primers for clone W83121 in pT7T3D vector, and pUC/M13f and pUC/M13r primers for clone AA107459 in pCMV-SPORT2 vector (where CMV is cytomegalovirus), as well as the legumain-specific primers.

Expression of recombinant mouse legumain

A full-length mouse legumain cDNA ligated into the mammalian expression vector pCMV-SPORT2 (Life Technologies), termed pCMVmusleg, was used to transfect HEK 293 cells [21] for recombinant protein expression. Transfection was performed with the Calcium Phosphate Transfection Kit (Invitrogen). Cells were plated at approx. 1×10^6 cells per 100 mm dish 1 day before transfection, and incubated with a calcium phosphate DNA precipitate containing 20 μg of pCMVmusleg plasmid and 10 μg of pAdVAntage vector (Promega). The pAdVAntage vector carries the adenoviral VAI RNA gene required for efficient translation in adenovirus-infected cells [22]. The transfection

medium was removed after 16 h and cells were subjected to shock with 15% glycerol in Hepes-buffered saline (25 mM Hepes, 140 mM NaCl, 1.5 mM Na_2HPO_4, pH 7.1) for 1 min. The transfected cells were then rinsed once with PBS, and cultured in minimal Eagle's medium supplemented with 10% (v/v) heat-inactivated fetal bovine serum until harvest.

Preparation of HEK 293 cell lysates and partial purification of recombinant mouse legumain

Cells were harvested and disrupted with the lysis buffer (0.1 M sodium citrate, pH 5.8, containing 1 mM EDTA and 1% n-octyl β-D-glucopyranoside). After three cycles of freezing (in solid CO_2) and thawing (30 °C), the cell lysate was centrifuged at 20000 ***g*** for 5 min. The supernatant was collected and stored at 4 °C. To purify the recombinant mouse legumain, the cell lysate from pCMVmsleg-transfected HEK 293 cells (ten 100 mm dishes) was further centrifuged at 100000 ***g*** for 1 h. The supernatant was diluted to 10 mM sodium citrate. The sample was then applied to a Mono S column (Pharmacia HR 5/5) equilibrated with 50 mM sodium citrate buffer, pH 5.5, containing 1 mM EDTA and 0.1% CHAPS. A step gradient of NaCl (0.15 M, 0.4 M and 1 M) in the same buffer was applied to wash the column and to elute bound proteins. The active recombinant mouse legumain was eluted from the column at 0.4 M NaCl.

HPLC analysis

A Waters BioDiscovery® HPLC system incorporating photodiodide array detection was used to analyse peptide cleavage by legumain. The product mixtures were run on a Waters Deltapak C_{18} 300 Å (30 nm), 5 μm-pore-size, column (2 mm × 150 mm) with a linear gradient of 5–55% (v/v) acetonitrile containing 0.01% trifluoroacetic acid at a flow rate of 0.2 ml/min. Each run covered a period of 1 h with monitoring at 214 nm.

RESULTS

Distribution of legumain and its subcellular localization in mouse tissues

We first examined the expression of legumain in tissues of a 3-week-old mouse. Protein samples (30 μg) from various tissue homogenates were analysed by SDS/PAGE and immunoblotting. An immunoreactive band of 35 kDa was detected with the anti-legumain antibody in all tissues, and was most intense in kidney and placenta homogenates (Figure 1A). Compared with the legumain purified from pig kidney (34 kDa; lane 10), mouse legumain in tissue homogenates (lanes 1–9) showed a slightly higher molecular mass. The pig legumain has been shown previously to be glycosylated [7], and the potential glycosylation sites are conserved in mouse legumain (see below). The difference in size between mouse and pig legumains could be due to different degrees of glycosylation. In lane 10, in addition to the 34 kDa band, a lower-molecular-mass immunoreactive band (indicated by the arrow) was also observed. This band has been shown by N-terminal micro-sequencing to be a degraded fragment of pig legumain.

Fluorimetric assays showed that the specific activity of legumain was highest in kidney and placenta (Figure 1B), consistent with the results from immunoblot analysis.

To determine the subcellular localization of legumain within tissues, the procedure of Maunsbach [12] was used to prepare subcellular fractions from rat kidney, including brush border, mitochondria, lysosomes, microsomes and cytosol. Samples (6 μg of total protein) from these fractions were analysed by immunoblotting with the anti-legumain antibody. The immunoreactive band of 35 kDa, presumably mature legumain, was found to be most intense in the lysosomal fraction (Figure 2, lane 5). Similar results were obtained with mouse kidney (not shown).

When the primary culture of rat kidney cells from the renal cortex was labelled with the anti-legumain antibody, the immuno-

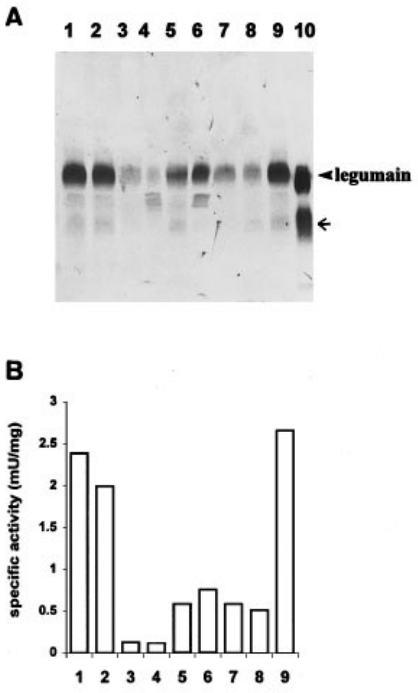

Figure 1 Distribution of legumain in mouse tissues

Legumain in tissue extracts from a 3-week-old mouse was detected by (**A**) immunoblotting for protein expression and (**B**) the fluorimetric assay to determine the specific activity of legumain by hydrolysis of Z-Ala-Ala-Asn-NHMec. Tissue samples were: 1, placenta; 2, kidney (sample 1); 3, heart; 4, lung; 5, liver; 6, spleen; 7, testis; 8, thymus; 9, kidney (sample 2). Lane 10 of (**A**) contains legumain purified from pig kidney serving as the control. The lower-molecular-mass immunoreactive band indicated by the arrow is known to be a degradation fragment of legumain.

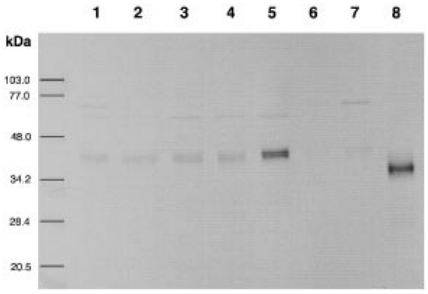

Figure 2 Immunoblot analysis of rat kidney subcellular fractions

Subcellular fractions of rat kidney cortex were prepared using the method of Maunsbach [12]. Portions (6 μg) of total protein from the tissue homogenate (lane 1), nuclear fraction (lane 2), brush border (lane 3), mitochondria (lane 4), lysosomes (lane 5), cytosol (lane 6) and microsomes (lane 7) were run in electrophoresis and transferred to a PVDF membrane. The blot was probed with the anti-legumain antibody. The 35 kDa immunoreactive band of legumain is seen to be most intense in the lysosomal fraction (lane 5). Purified legumain from pig kidney (about 34 kDa; lane 8) served as a control. Two faint bands between the size standards of 48 and 77 kDa appeared to be the result of non-specific reactions.

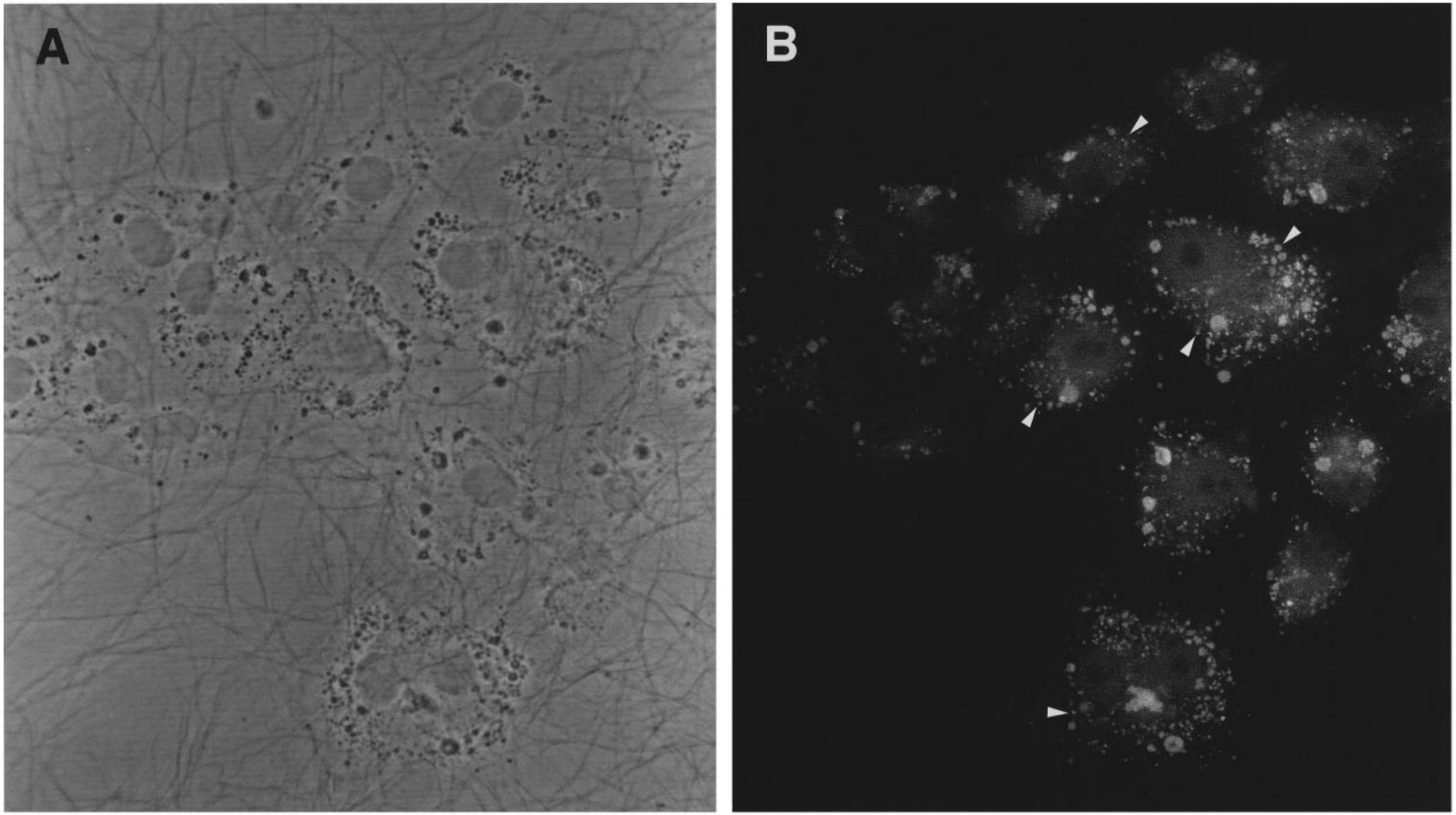

Figure 3 Immunofluorescence of legumain in the primary culture of kidney cells

Rat kidney cells were prepared from the cortex and cultured as described in the Materials and methods section. Cells on rat-tail collagen substratum seen in phase contrast (**A**) were stained with the anti-legumain antibody and observed under epifluorescence illumination with the fluorescein filter set (**B**). Both images show large vacuoles and lysosome-like organelles (some indicated by arrowheads in **B**) which were labelled positively for legumain. A control preparation treated only with the secondary antibody (not shown) revealed some yellow autofluorescence only of the large vacuoles.

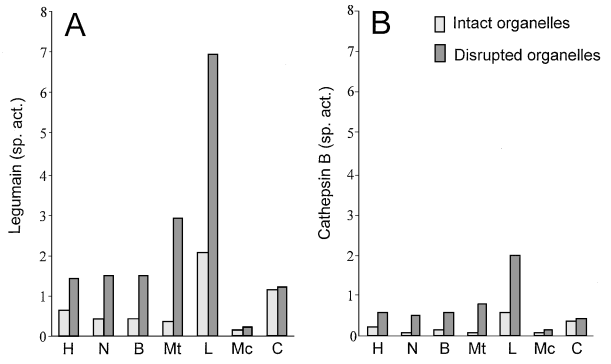

Figure 4 Dependence of legumain activity on disruption of organelles

Subcellular fractionation of rat renal cortex and measurement of the specific activities (sp. act.; m-units/mg of protein) of legumain and cathepsin B were as described in the Materials and methods section. Fractions were: tissue homogenate (H), nuclear fraction (N), brush border (B), mitochondria (Mt), lysosomes (L), microsomes (Mc) and cytosol (C). Specific activities of legumain (**A**) and cathepsin B (**B**) were found to be greatest in the lysosomal fraction whether organelles remained intact or were disrupted. In the lysosomal fraction (as well as in several other fractions that may have been contaminated with lysosomes), activity increased markedly following disruption of membranes by freeze/thawing. In contrast, there was no increase in the activity of the cytosolic fraction, which is expected to contain a fraction of lysosomal enzymes incidentally liberated by disruption of lysosomes even in the presence of sucrose.

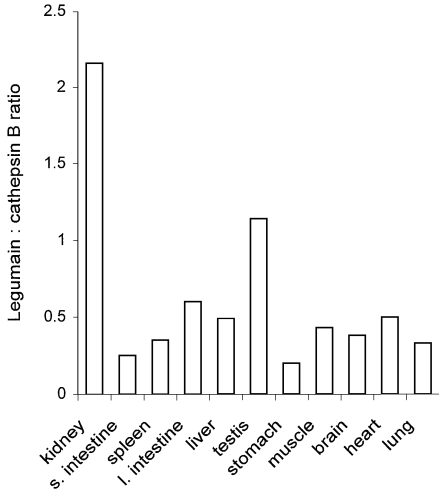

Figure 5 Comparative distributions of legumain and cathepsin B

The specific activities of legumain and cathepsin B in various mouse tissue homogenates were determined as described in the Materials and methods section. Values on the *y*-axis represent ratios of the specific activity (m-units/mg of protein) of legumain to that of cathepsin B.

fluorescence image showed a punctate pattern (Figure 3), also suggesting a lysosomal localization of legumain.

It is known that lysosomal enzymes are inactive in test-tube assays while lysosomes remain intact, because the lysosomal membrane prevents contact between the enzyme and the exogenous substrate [23]. This property of legumain and cathepsin B (a known lysosomal enzyme) is demonstrated in Figure 4. The specific activity of legumain (Figure 4A), like that of cathepsin B (Figure 4B), was highest in the lysosomal fraction. Furthermore, both legumain and cathepsin B activities were increased about 3.5-fold when membranes of organelles were disrupted by freezing and thawing. The results show a latency of legumain, typical of lysosomal enzymes. A small percentage of the activity of each enzyme was present as a soluble form in the cytosol, for which there was no increase in activity following freeze/thawing.

Having established that legumain is a lysosomal enzyme, we asked whether the amount of legumain parallels that of cathepsin B. The specific activities of legumain and cathepsin B were determined for a range of tissues and the ratios between the two were calculated. The results (Figure 5) revealed that the specific activity of legumain was twice that of cathepsin B in kidney, but in other tissues the legumain activity was similar to (testis) or less than that of cathepsin B. We concluded that mouse kidney lysosomes are particularly rich in legumain.

DNA sequencing and structural analysis of mouse legumain

Mouse EST clones AA107459 and W83121 were identified as being homologous to human and pig legumain by use of the BlastN and TBlastN programs. Restriction enzyme digest and PCR analysis revealed a 1.1 kb insert in clone W83123 and a 1.9 kb insert in clone AA107459. This suggested that clone AA107459 might carry a full-length mouse legumain coding sequence. The clone was sequenced, giving the cDNA sequence and deduced amino acid sequence shown in Figure 6. When the sequence was aligned with that of human legumain [7], it became apparent that the 1889 nucleotides include the complete coding region together with 5′ and 3′ non-coding regions. The open reading frame of 1305 nucleotides is predicted to encode 435 amino acid residues with a calculated mass of 49373 Da. The sequence includes a putative leader peptide, four potential N-linked glycosylation sites, and an RGD motif. Using the Bestfit program, 83% identity was observed between the mouse and human sequences at the amino acid level, consistent with the two gene products being orthologues, i.e. species variants of the same protein.

Expression of recombinant legumain

Recombinant mouse legumain was expressed in HEK 293 cells driven by the CMV promoter. The lysate prepared from the pCMVmusleg-transfected cells hydrolysed the fluorogenic substrate of legumain, Z-Ala-Ala-Asn-NHMec. The enzymic activity was inhibited by 50 nM egg-white cystatin, but was unaffected by 10 μM compound E-64, exactly as had been found previously for the naturally occurring enzyme from pig kidney [7]. No activity was detected in cells transfected with the pCMV-SPORT2 vector only or with both the pCMV-SPORT2 and pAdVAntage vectors. The immunoblot (Figure 7) showed that the recombinant mouse legumain was expressed primarily as the mature enzyme of 35 kDa (lane 1); no antigen was detected in cells transfected only with the control vector plasmid (lane 2). A minor band of approx. 48 kDa was also immunoreactive with the antibody (lane 1). This band is tentatively identified as pro-legumain, since the calculated molecular mass of unglycosylated mouse pro-legumain is 47634 Da. As seen previously, mouse legumain of recombinant origin (lane 1) or from kidney (lane 3) showed a slightly higher molecular mass than legumain purified from pig kidney (lane 4).

Biochemical characterization of recombinant mouse legumain

Recombinant mouse legumain was maximally active close to pH 6.0, and was irreversibly inactivated at pH 7. These characteristics are identical with those of naturally occurring pig legumain [7]. The effects of a range of potential inhibitors on recombinant legumain were also similar to those on pig kidney legumain. Thus the general inhibitors of thiol-dependent enzymes, i.e. iodoacetate, iodoacetamide, *N*-ethylmaleimide and

```
GCTCTGAGTCTGCGCGACGCCCGGAATTCCCACGGTTCTGCAGTCACCGCGGCGATCACC   60
CGCCCAGTCTTCTGTAGCGGACACGGGGTGCAGAATGACCTGGAGAGTGGCTGTGCTTCT  120
                                  M  T  W  R  V  A  V  L  L      9
CAGCCTGGTGCTGGGTGCTGGTGCCGTTCCCGTCGGTGTGGACGATCCCGAGGATGGAGG  180
 S  L  V  L  G  A  G  A  V  P  V  G  V  D  D  P  E  D  G  G     29

CAAGCACTGGGTGGTGATTGTGGCGGGCTCCAATGGCTGGTATAATTACCGACACCAGGC  240
 K  H  W  V  V  I  V  A  G  S  N  G  W  Y  N  Y  R  H  Q  A     49

AGACGCATGCCACGCCTACCAGATCATCCACCGGAACGGGATTCCTGACGAGCAGATCAT  300
 D  A  C  H  A  Y  Q  I  I  H  R  N  G  I  P  D  E  Q  I  I     69

AGTGATGATGTATGACGACATTGCCAACTCTGAAGAAAACCCTACCCCAGGTGTTGTGAT  360
 V  M  M  Y  D  D  I  A  N  S  E  E  N  P  T  P  G  V  V  I     89

CAACCGACCTAACGGCACAGATGTATACAAGGGAGTCCTGAAGGACTACACCGGAGAGGA  420
 N  R  P  N  G  T  D  V  Y  K  G  V  L  K  D  Y  T  G  E  D    109

TGTGACTCCAGAGAATTTCCTCGCCGTGCTGAGAGGTGACGCAGAAGCTGTGAAGGGCAA  480
 V  T  P  E  N  F  L  A  V  L  R  G  D  A  E  A  V  K  G  K    129

AGGGTCTGGAAAAGTCTTGAAGAGTGGCCCCCGAGATCATGTCTTCATTTACTTCACCGA  540
 G  S  G  K  V  L  K  S  G  P  R  D  H  V  F  I  Y  F  T  D    149

CCACGGAGCCACCGGGATCCTGGTGTTTCCTAATGATGATCTTCATGTCAAGGACCTGAA  600
 H  G  A  T  G  I  L  V  F  P  N  D  D  L  H  V  K  D  L  N    169

TAAGACTATTCGCTACATGTATGAACACAAAATGTACCAGAAGATGGTGTTCTACATTGA  660
 K  T  I  R  Y  M  Y  E  H  K  M  Y  Q  K  M  V  F  Y  I  E    189

AGCTTGTGAGTCTGGCTCCATGATGAACCACCTGCCCGACGACATCAACGTTTATGCAAC  720
 A  C  E  S  G  S  M  M  N  H  L  P  D  D  I  N  V  Y  A  T    209

TACTGCGGCCAACCCCAAGGAGTCATCTTATGCCTGCTACTACGACGAGGAGAGGGGCAC  780
 T  A  A  N  P  K  E  S  S  Y  A  C  Y  Y  D  E  E  R  G  T    229

TTACCTGGGTGACTGGTACAGCGTCAACTGGATGGAAGACTCCGATGTGGAGGACCTGAC  840
 Y  L  G  D  W  Y  S  V  N  W  M  E  D  S  D  V  E  D  L  T    249

CAAAGAGACCCTTCACAAGCAGTACCACCTGGTCAAGTCCCACACCAACACCAGCCATGT  900
 K  E  T  L  H  K  Q  Y  H  L  V  K  S  H  T  N  T  S  H  V    269

CATGCAATATGGGAACAAATCTATCTCTACCATGAAAGTGATGCAGTTTCAGGGAATGAA  960
 M  Q  Y  G  N  K  S  I  S  T  M  K  V  M  Q  F  Q  G  M  K    289

GCACAGAGCCAGTTCCCCCATCTCCCTGCCTCCGGTCACACACCTTGACCTCACCCCCAG 1020
 H  R  A  S  S  P  I  S  L  P  P  V  T  H  L  D  L  T  P  S    309

CCCTGACGTGCCCCTGACCATCTTGAAGAGGAAGCTGCTGAGAACCAACGACGTGAAGGA 1080
 P  D  V  P  L  T  I  L  K  R  K  L  L  R  T  N  D  V  K  E    329

ATCCCAGAATCTCATTGGGCAGATCCAGCAATTTCTGGATGCCAGGCACGTCATTGAGAA 1140
 S  Q  N  L  I  G  Q  I  Q  Q  F  L  D  A  R  H  V  I  E  K    349

GTCTGTGCACAAGATCGTTTCCCTGCTGGCGGGATTTGGGGAAACTGCTGAGAGACATCT 1200
 S  V  H  K  I  V  S  L  L  A  G  F  G  E  T  A  E  R  H  L    369

GTCAGAGAGGACCATGCTCACAGCACATGACTGCTACCAGGAGGCTGTAACCCACTTCCG 1260
 S  E  R  T  M  L  T  A  H  D  C  Y  Q  E  A  V  T  H  F  R    389

CACACACTGCTTTAACTGGCACTCTGTCACGTACGAGCATGCCTTGCGGTACTTGTATGT 1320
 T  H  C  F  N  W  H  S  V  T  Y  E  H  A  L  R  Y  L  Y  V    409

GCTGGCCAATCTCTGTGAGGCACCATATCCGATTGACAGGATAGAGATGGCCATGGACAA 1380
 L  A  N  L  C  E  A  P  Y  P  I  D  R  I  E  M  A  M  D  K    429

AGTGTGTCTTAGTCACTACTGAACAGCTCGCTTCCCAATGAGTGAGCACAGTCCACTGGA 1440
 V  C  L  S  H  Y  *                                           435

ATATGAACCAACCAGAGACTGAAAGGGCGGACCAGAGGCAGCACCCGCGCCCCTGGCCCC 1500
CAGGAATACTGCCCGCCCACCCCAGGGCTTGCTTTTTGAAGATACCTGCTTACTAAGAAG 1560
CCAGTTTGGGTGGGTAAAGCTCTCTGGAAGAAGGAACTTTGCTTCTTAGGAGTTTTTTTG 1620
TTTGTTTGTTTTGGTTTGTTTTGTTGTCCATTAGCTTTCAAGAGCAAATTCCCCGCGGCT 1680
TCTCTAGCCAGGGAAGGAATCGTCTGAGAAATTCAAAGCTGAAACCTCTTGCCGTCTTCA 1740
CAGTGATTTCACTGAAGAAGAGGGTGGAAAGCAAGCCCCTATGGGAGAATTATTTTTAGA 1800
ATTATATAATTTTTGATTGCTTTTATATTTTATTCTGTAATAATGGATGTTTTAAAACAA 1860
ATAAGTGAAGTGAAAAAAAAAAAAAAAAA                                 1889
```

Figure 6 Nucleotide and deduced amino acid sequence of mouse legumain

The mouse legumain amino acid sequence includes a putative leader peptide (underlined), four potential N-linked glycosylation sites (double underlined) and an RGD motif (boxed). In the nucleotide sequence, the initiation codon and the polyadenylation signal are in **bold** type.

N-phenylmaleimide, were inhibitory, but the more selective inhibitors of cysteine peptidases of the papain family, i.e. Z-Phe-Ala-CHN_2, E-64 and leupeptin, were without effect. General

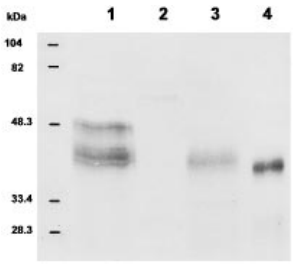

Figure 7 Western immunoblot of recombinant legumain in lysate of HEK 293 cells transfected with mouse legumain cDNA

The immunoblot was probed with the polyclonal anti-legumain antibody. The antibody immunoreacted with the recombinant legumain of 35 kDa (mature enzyme) and a putative 48 kDa pro-form. Samples in the blot are: lane 1, lysate of HEK 293 cells transfected with the mouse legumain construct; lane 2, lysate of HEK 293 cells transfected with the vector plasmid pCMV-SPORT2 and pAdVAntage; lane 3, mouse kidney tissue homogenate; lane 4, purified legumain from pig kidney.

inhibitors of serine peptidases (benzamidine, PMSF), aspartic peptidases (pepstatin) and metallopeptidases (1,10-phenanthroline, EDTA) did not affect the enzymic activity of recombinant legumain significantly.

Recombinant legumain was capable of cleaving both oligopeptide and protein substrates. The peptide neurotensin (pGlu-Leu-Tyr-Glu-Asn-Lys-Pro-Arg-Arg-Pro-Tyr-Ile-Leu) was incubated with 0.3 m-unit of recombinant legumain/ml for 2 h at 30 °C and the reaction product was separated by HPLC. Analysis of the peptide products proved that recombinant mouse legumain cleaved the oligopeptide exclusively at the Asn–Lys bond.

The C-fragment of tetanus toxoid was used to test whether recombinant legumain degrades protein substrates. The enzyme generated three fragments, and N-terminal micro-sequencing of these showed that they were identical to those produced by pig legumain.

DISCUSSION

In the present work, we have used the mouse as a model system to study the tissue distribution and subcellular localization of legumain, a cysteine proteinase recently discovered in mammals. We have also sequenced mouse legumain cDNA and expressed the recombinant protein in order to elucidate its molecular properties.

We found that legumain activity was at least 3-fold higher in the kidney and placenta than in other tissues examined, and that legumain was detected as a 35 kDa mature form in tissue extracts (Figure 1). Studies on subcellular localization in both mouse and rat kidney showed that legumain, like cathepsin B, is located in the lysosomal fraction (Figures 2 and 4). The punctate immunostaining pattern in cultured rat kidney cells (Figure 3) and the acidic-acting nature of the enzyme are consistent with this view. We therefore conclude that legumain is a lysosomal enzyme. Although legumain and cathepsin B are lysosomal, the ratio of activities of the two enzymes differed among tissues, with kidney lysosomes being particularly rich in legumain (Figure 5). This raises the possibility that legumain may have a specific role in the physiology of the kidney.

The mouse legumain cDNA sequence consists of 1305 nucleotides encoding 435 amino acids with a putative leader peptide of 17 residues (Figure 6). On the basis of alignment with the mature pig legumain N-terminal sequence [7], Gly-28 is likely to be the first residue of the mature mouse enzyme. Despite the fact that

the calculated molecular mass of the mouse legumain proenzyme is 47634 Da, a major band of 35 kDa was detected in both endogenous and recombinant mouse legumain (Figures 1, 2 and 7). This suggests that post-translational processing takes place during the biosynthesis of mouse legumain. Plant legumains [3] and also legumain from pig kidney are processed at both their N- and C-termini, so this is probably a general characteristic.

The deduced amino acid sequences of mouse and human legumain show 83 % identity. Both sequences have a putative leader sequence, four potential N-linked glycosylation sites and an Arg-Gly-Asp (RGD) motif. In the N-terminal propeptide region, however, a difference was observed between the mouse and human sequences, in that there are 10 residues (VPVGV-DDPED) in mouse compared with eight (VPIDDPED) in human [7].

The RGD motif in cell-adhesive proteins such as fibronectin and the disintegrins is responsible for the binding of the proteins to their cell-surface receptors, the integrins [24,25]. The RGD sequence is also present and conserved in the Homo B domain of the proprotein convertases of the mammalian subtilisin/kexin subfamily (see [26] for a review). It has been shown that mutation of the RGD residues in intracellular proprotein convertase (PC1) resulted in loss of processing activity. This was attributable to an increased rate of degradation of PC1 molecules and a decreased rate of processing of the PC1 proenzyme [27]. It is conceivable that, in legumain, the RGD sequence is also involved in cellular trafficking and proenzyme processing during biosynthesis.

The effects of inhibitors show that both mouse and pig legumains are cysteine endopeptidases. Nevertheless the sequences are completely unrelated to those of cathepsin B and the other known lysosomal cysteine endopeptidases, all of which are relatives of papain (in peptidase family C1), whereas the legumains belong to peptidase family C13 [28]. There are only three cysteine residues in the mature mammalian legumains, and only two are conserved from plants to mammals. It therefore appears that the catalytic residue of mouse legumain is either Cys-52 or Cys-191 (numbered according to the mouse prepro-enzyme). Neither of these occurs in an amino acid sequence motif recognizably similar to that around a known catalytic cysteine residue in any other family of cysteine peptidases.

The biochemical properties of recombinant mouse legumain are very similar to those of pig legumain isolated from kidney cortex. Like pig legumain, the mouse recombinant protein is acid-acting, and it loses activity at neutral pH. This is consistent with the lysosomal localization of legumain (Figures 2–4), since most of the lysosomal homologues of papain also are denatured at neutral pH, unlike papain itself. This property may be an adaptation of lysosomal endopeptidases that serves to restrict their activities to the acidic environment within the lysosomal system. Its appearance in another family of endopeptidases appears to be a striking example of convergent evolution.

Like the naturally occurring enzyme, recombinant mouse legumain hydrolysed both the 13-residue peptide neurotensin and the 53 kDa C-fragment of tetanus toxoid exclusively at asparaginyl bonds. The previously described lysosomal endopeptidases are of rather broad specificity [29–32], and the appearance of an additional enzyme that is highly specific for hydrolysis at asparaginyl bonds was not anticipated. The occurrence of legumain with retention of this specificity in mammalian lysosomes and the lysosome-like vacuoles of plant cells [3] shows that it has been conserved at least since the divergence of plants and animals perhaps 1×10^9 years ago. The implication is that there may be similarly conserved substrates that remain to be discovered.

The biological functions of mammalian legumain have not yet been determined. However, the fact that legumain activity is particularly abundant in the renal cortex suggests that legumain may contribute to the processing of macromolecules that pass across the glomerular filter and are re-absorbed by proximal tubule cells through pinocytosis. Knowledge of the nucleotide sequence of mouse legumain described here will allow site-directed mutagenesis to be carried out in order to probe functionally important residues, and may also lead to the creation of a legumain-deficient strain of mouse that could shed further light on this matter.

We thank Dr. Neil Rawlings and Mr. John Coadwell for much help with the bioinformatics and for helpful comments.

REFERENCES

1 Takeda, O., Miura, Y., Mitta, M., Matsushita, H., Kato, I., Abe, Y., Yokosawa, H. and Ishii, S. (1994) J. Biochem. (Tokyo) **116**, 541–546
2 Kembhavi, A. A., Buttle, D. J., Knight, C. G. and Barrett, A. J. (1993) Arch. Biochem. Biophys. **303**, 208–213
3 Hara-Nishimura, I., Takeuchi, Y. and Nishimura, M. (1993) Plant Cell **5**, 1651–1659
4 El Meanawy, M. A., Aji, T., Phillips, N. F. B., Davis, R. E., Salata, R. A., Malhotra, I., McClain, D., Aikawa, M. and Davis, A. H. (1990) Am. J. Trop. Med. Hyg. **43**, 67–78
5 Abe, Y., Shirane, K., Yokosawa, H., Matsushita, H., Mitta, M., Kato, I. and Ishii, S. (1993) J. Biol. Chem. **268**, 3525–3529
6 Dalton, J. P., Hola-Jamriska, L. and Brindley, P. J. (1995) Parasitology **111**, 575–580
7 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
8 Anastasi, A., Brown, M. A., Kembhavi, A. A., Nicklin, M. J. H., Sayers, C. A., Sunter, D. C. and Barrett, A. J. (1983) Biochem. J. **211**, 129–138
9 Barrett, A. J. and Kirschke, H. (1981) Methods Enzymol. **80**, 535–561
10 Rawlings, N. D. and Barrett, A. J. (1990) Comput. Appl. Biosci. **6**, 118–119
11 Bradford, M. M. (1976) Anal. Biochem. **72**, 248–254
12 Maunsbach, A. B. (1974) Methods Enzymol. **31**, 330–339
13 Chen, J.-M., Changco, A., Brown, M. A. and Barrett, A. J. (1995) Exp. Cell Res. **216**, 80–85
14 Chen, J.-M. and Little, C. D. (1985) J. Cell Biol. **101**, 1175–1181
15 Laemmli, U. K. (1970) Nature (London) **227**, 680–685
16 Towbin, H., Staehelin, T. and Gordon, J. (1979) Proc. Natl. Acad. Sci. U.S.A. **76**, 4350–4354
17 Blake, M. S., Johnston, K. H., Russell-Jones, G. J. and Gotschlich, E. C. (1984) Anal. Biochem. **136**, 175–179
18 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) J. Mol. Biol. **215**, 403–410
19 Genetics Computer Group (1994) Program Manual for the Wisconsin Package, Version 8, September 1994, University of Madison, Wisconsin
20 Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd edn., Cold Spring Harbor Laboratory Press, Cold Spring Harbor
21 Graham, F. L., Smiley, J., Russell, W. C. and Nairn, R. (1977) J. Gen. Virol. **36**, 59–72
22 Gorman, C. M., Gies, D. R. and McCray, G. (1990) DNA Protein Eng. Tech. **2**, 3–10
23 de Duve, C., Pressman, B. C., Gianetto, R., Wattiaux, R. and Appelmans, F. (1955) Biochem. J. **60**, 605–617
24 D'Souza, S. E., Ginsberg, M. H. and Plow, E. F. (1991) Trends Biochem. Sci. **16**, 246–250
25 Ruoslahti, E. and Pierschbacher, M. D. (1986) Cell **44**, 517–518
26 Nakayama, K. (1997) Biochem. J. **327**, 625–635
27 Lusson, J., Benjannet, S., Hamelin, J., Savaria, D., Chrétien, M. and Seidah, N. G. (1997) Biochem. J. **326**, 737–744
28 Rawlings, N. D. and Barrett, A. J. (1994) Methods Enzymol. **244**, 461–486
29 Barrett, A. J. (1992) Ann. N. Y. Acad. Sci., **674**, 1–15
30 Bohley, P. and Seglen, P. O. (1992) Experientia **48**, 151–157
31 Berg, T., Gjøen, T. and Bakke, O. (1995) Biochem. J. **307**, 313–326
32 Storrie, B. (1995) in Principles of Medical Biology, vol. 3 (Bittar, E. E. and Bittar, N., eds.), pp. 1–18, JAI Press, Greenwich, CT

Received 9 April 1998/22 June 1998; accepted 28 July 1998