Cloning, Isolation, and Characterization of Mammalian Legumain, an Asparaginyl Endopeptidase*

Jinq-May Chen, Pam M. Dando, Neil D. Rawlings, Molly A. Brown, Nina E. Young, Richard A. Stevens, Eric Hewitt‡, Colin Watts‡, and Alan J. Barrett‡

From the Medical Research Council Peptidase Laboratory, Department of Immunology, The Babraham Institute, Babraham Hall, Babraham, Cambridgeshire CB2 4AT, United Kingdom and the ‡Department of Biochemistry, University of Dundee, Dundee DD1 4HN, United Kingdom

THE JOURNAL OF BIOLOGICAL CHEMISTRY Vol. 272, No. 12, Issue of March 21, pp. 8090–8098, 1997
© 1997 by The American Society for Biochemistry and Molecular Biology, Inc.

Legumain is a cysteine endopeptidase that shows strict specificity for hydrolysis of asparaginyl bonds. The enzyme belongs to peptidase family C13, and is thus unrelated to the better known cysteine peptidases of the papain family, C1 (Rawlings, N. D., and Barrett, A. J. (1994) Methods Enzymol. 244, 461–486). To date, legumain has been described only from plants and a blood fluke, Schistosoma mansoni. We now show that legumain is present in mammals. We have cloned and sequenced human legumain and part of pig legumain. We have also purified legumain to homogeneity (2200-fold, 83% yield) from pig kidney. The mammalian sequences have also purified legumain to homogeneity (2200-fold, 83% yield) from pig kidney. The mammalian sequences have now cloned and sequenced human legumain from a placenta library, and isolated and characterized pig legumain from kidney. We discuss the implications of the presence of an asparaginyl endopeptidase in mammalian cells.

EXPERIMENTAL PROCEDURES

Materials—Z-Ala-Ala-Asn-NHMec prepared as described (3) was supplied by Bachem. Sodium citrate buffer solutions were prepared by mixing equimolar solutions of citric acid and trisodium citrate, to the desired pH value. Sodium citrate/phosphate buffers in the range pH 3.0–7.5 were as described by McIlvaine (Ref. 6).

Ovocystatin (cystatin from chicken egg white) was prepared as described (7), and recombinant human cystatin C was the kind gift of Dr. Magnus Abrahamsson (Department of Clinical Chemistry, University of Lund, Sweden). Both cystatins were titrated with a solution of papain (Sigma, 2 × crystallized) that had been standardized with E64 (8, 9), and the concentrations quoted in the text relate to active inhibitor.

Azocasein was modified to include a hisidine tag (2).

A normal, full-term human placenta was kindly made available by the Rosie Maternity Hospital, Cambridge, and kidneys from freshly killed pigs were purchased from a local abattoir. Rats were Porton Wistar and rabbits New Zealand Whites.

Primer Design and PCR-based Cloning of Human and Pig Legu-

Cysteine peptidases form one of the major groups of proteolytic enzymes, and can be divided into about 30 separate families on the basis of their molecular structures (reviewed in Refs. 1 and 2). Three families of cysteine endopeptidases have been known to be represented in mammals. The most numerous are those of the papain family (C1), which include cathepsins B, H, L, S, and others. These are predominantly lysosomal enzymes, responsible for proteolysis in the lysosomal/endosomal system and also are secreted to act extracellularly. In the cytosolic fraction of the cell, there are members of the other two families of cysteine endopeptidases: the families of calpain (family C2) and caspase (previously interleukin 1-converting enzyme; C14). These peptidases mediate limited proteolysis of cytosolic substrates. We now report that the legumain family (C13) can be added to the list of mammalian cysteine endopeptidases.

Legumain is the name that was given by Kembhavi et al. (3) to an endopeptidase that is present in many leguminous and other seeds, after they had isolated and characterized the enzyme from Vigna aconitifolia (moth bean). Legumain is specific for the hydrolysis of asparaginyl bonds. The amino acid sequence of legumain from Ricinus communis (castor bean) showed it to be homologous with an enzyme from the fluke Schistosoma mansoni (4). At that time, the fluke enzyme was of unknown specificity, but it has now been shown also to be an asparaginyl endopeptidase (5), active on the test substrate that had been introduced by Kembhavi et al.

The appearance of sequences homologous to legumain among the human expressed sequence tags (ESTs) in the databases alerted us to the presence of the enzyme in vertebrates, and accordingly, we made assays for asparaginyl endopeptidase activity in mammalian tissues. It was soon evident that legumain is present in human and other mammalian cells. We have now cloned and sequenced human legumain from a placenta library, and isolated and characterized pig legumain from kidney. We discuss the implications of the presence of an asparaginyl endopeptidase in mammalian cells.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

1 The abbreviations used are: EST, expressed sequence tag; CHAPS, 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonic acid; DTT, dithiotreitol; E64, trans-epoxysuccinyl-l-leucylamido(4-guanidino)butane; NHMec, 7-(4-methyl)coumarylamide; NHPhNO2, p-nitroanilide; VIP, vasoactive intestinal peptide; bp, base pair(s); PCR, polymerase chain reaction; PAGE, polyacrylamide gel electrophoresis; Tricine, N-tris(hydroxymethyl)methylglycine; Z, benzoyloxycarbonyl.

2 E. Hewitt, C. Watts, and A. J. Makoff, unpublished results.

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ To whom correspondence should be addressed. Tel.: 44-1223-832312; Fax: 44-1223-837952; E-mail: alan.barrett@bbsrc.ac.uk.
PCR reactions were in 50 μl, and each combination of the Hsl1f forward primer (500 nM) with each reverse primer was used for all amplification reactions. The PCR products were separated in 1% agarose gels and digested with EcoRI and HindIII, and cloned into pSP65 at 9-BamHI sites. Use of the nested dIII sites into Hsl1f and Hsl4r for cloning has been described in the literature (11). The 5′-rapid amplification of cDNA ends and primer Hsl4r, based on EST R17110) were sequenced by the method of 9-EcoRI. A partial cloning and sequencing of pig legumain cDNA was achieved by a similar PCR strategy, Two pig kidney cDNA libraries (gifts of Dr. Jerzy Adamski, Max-Planck-Institut für Experimentelle Endokrinologie, Hannover, Germany, and Dr. Claudia T. Evans, University of Texas Southwestern Medical Center, Dallas, TX) were used as templates for the amplification reactions. A 936-bp cDNA fragment was obtained when primers Hsl4f and Hsl4r were used. SDS-Polyacrylamide Gel Electrophoresis—Except when stated, SDS-polyacrylamide gel electrophoresis (SDS-PAGE) was in gels of 10% T, 2.6% C containing the Ammedei buffer system (14). Staining for protein was with Coomassie Brilliant Blue R, and glycoproteins were detected with the DIG glycan detection kit (Boehringer Mannheim). Transferrin and thimet oligopeptidase served as positive and negative controls, respectively. In some experiments, samples were run in a Tris-Tricine-buffered gel (Bio-Rad).

Fluorometric Assay for Legumain—Continuous fluorimetric assays with 2-Ala-Ala-Asn- NHSMe were as described (3), with slight modifications. The substrate (10 μM) was incubated at 30 °C in 2.5 ml of 39.5 mM citric acid, 121 mM Na2HPO4, pH 5.8, containing 1 mM DTT, 1 mM EDTA, and 0.1% CHAPS (assay buffer). The rate of formation of product was followed in a Perkin-Elmer fluorometer under the control of a IBM-compatible computer running the FLUSYS software (15). The excitation and emission wavelengths were 360 and 460 nm, respectively, and 1 unit of activity was defined as that releasing 1 μmol of product/min under the standard conditions.

For the assay of activity in a tissue sample, the tissue was homogenized in four volumes (w/v) of 0.1 mM sodium citrate buffer, pH 6.0, containing 1 mM EDTA and 2 mM 2-mercaptoethanol. Five complete strokes were used in a Braun-Potter homogenizer running at 1000 rpm. The homogenate was centrifuged (11,600 × g), and the supernatant was used in the enzyme assay as described above. Once a steady reaction rate had been determined (about 5 min), E64 was added to 10 μM. Two min later, 25 μl of 121 mM Na2HPO4 in 0.2 ml of 70% formic acid for 24 h at 25 °C. The mixture was diluted 10-fold with water and freeze-dried. The sample was then resuspended and run in Tris/Tricine-buffered SDS-PAGE, and the separated peptides were transblotted on to polyvinylidene difluoride membranes for N-terminal microsequence analysis. Amino Acid Analysis—The reaction was stopped by heating with 6 M HCl, 0.1% TFA, and the reaction was followed for an additional 5 min. Values for specific enzymatic activity were calculated by use of protein concentrations determined in the Bradford assay (16), with bovine serum albumin as standard.

Rate constants for irreversible inactivation were not found by non-linear regression analysis of the pseudo-first-order curves of inactivation by use of the FLUSYS software, giving kobs. The second order rate constant, kobs, was calculated as kobs/II, at [S] ≪ Km. Since the assay substrate was used at [S] ≪ Km, no correction for competition with the inhibitors was required.

Fluorometric Assay for Legumain—A temperature of 0–4 °C was maintained throughout the procedure. All buffers contained 1 mM EDTA and 2 mM 2-mercaptoethanol, and centrifugation was at 15,000 × g for 20 min, unless otherwise stated.

Cortical tissue (250 g) was dissected from pig kidney and homogenized with an equal weight of water for 15 s in a Waring blender. The crude homogenate was stored at −20 °C, or used immediately. The crude homogenate was mixed with an equal volume of 0.1 mM sodium citrate, pH 6.0, containing 0.4 mM ammonium sulfate. The homogenate was centrifuged, and the pellet was discarded. The supernatant was adjusted from 0.2 to 1.2 mM ammonium sulfate by addition of the solid, stirred for 30 min, and centrifuged. The pellet was discarded and the supernatant made 3.2 mM ammonium sulfate. After 30 min, the pellet containing the enzyme was collected by centrifugation.

The pellet was resuspended in water, adjusted to pH 5.0 with 0.2 mM citric acid, and dialyzed overnight against several changes of 10 mM sodium citrate, pH 5.0. During the dialysis, a heavy, brown precipitate formed, which contained the legumain. The precipitate was collected by centrifugation, dispersed in 50 mM sodium citrate, pH 5.0, by use of an Ultra Turrax homogenizer (Janke & Kunkel KG, Ika Werk, Staufen, M. The homogenate was stored at 20 °C, or used immediately. The crude homogenate was mixed with an equal volume of 0.1 mM sodium citrate, pH 5.0, containing 0.4 mM ammonium sulfate. The homogenate was centrifuged, and the pellet was discarded. The supernatant was adjusted to pH 5.0 with 0.1 mM trisodium citrate, and applied to a column (20-ml bed volume) of SP-Sepharose FF (Pharmacia) that had been pre-equilibrated with 50 mM sodium citrate, pH 5.5. The column was washed with two bed volumes of the equilibrating buffer before the enzyme was eluted with 0.4 M NaCl in the same buffer. Effluent fractions containing activity were combined and dialyzed into 10 mM sodium citrate, pH 5.0.

The solution was passed through 0.45- and 0.22-μm nitrocellulose filters in preparation for running on the Pharmacia Mono S FPLC column (type HR 10/10). A column (8-ml bed volume) of Thiopropyl-Sepharose 6B (Pharmacia) was activated with 2- pyridyl disulfide as described by the manufacturers. All buffers for use on the Mono S and Thiopropyl-Sepharose columns were mercaptoethanol-free, and were deoxygenated by sparging with nitrogen gas. The Mono S column was equilibrated with 50 mM sodium citrate, pH 5.5. The two columns were connected in tandem. The sample was first run on the Mono S column, which was then washed with three bed volumes of 50 mM sodium citrate, pH 5.5, 1 mM EDTA at a flow rate of 3 ml/min. When the A280 of the effluent had fallen nearly to zero, activity was eluted with a step of 0.4 M NaCl in 50 mM sodium citrate buffer, pH 5.5. As soon as protein was detected in the effluent, flow was redirected to the Thiopropyl-Sepharose column, and the flow rate was decreased to 0.25 ml/min. The protein from the Mono S column had run through the Thiopropyl-Sepharose column, the column was washed with five bed volumes of 50 mM sodium citrate, 0.2 M NaCl, 1 mM EDTA, pH 5.5, before being filled with the same buffer containing 10 mM cysteine, 0.1% CHAPS, pH 5.5. The column was allowed to stand overnight at 4 °C, and elution with the cysteine-containing buffer was resumed. Fractions containing activity were combined as the final product, which was stored at 4 °C or −20 °C in the eluting buffer.

Deglycosylation with N-Glycosidase F—Legumain (15 μg in 0.25 ml) was digested with 0.1 mM citric acid, 0.2 mM Na2HPO4, 1 mM EDTA, 0.025% CHAPS, pH 7.2, and heated at 100 °C for 5 min. N-Glycosidase F (Boehringer Mannheim) was then introduced (0.125 milliunit), and the mixture was incubated at 37 °C for 24 h. The protein was precipitated from 10% trichloroacetic acid and taken up for SDS-PAGE.

Partial Amino Acid Sequencing—Legumain (90 μg) was digested with 50 μg of CNBr in 0.2 ml of 70% formic acid for 24 h at 25 °C. The mixture was diluted 10-fold with water and freeze-dried. The sample was then resuspended and run in Tris/Tricine-buffered SDS-PAGE, and the separated peptides were transblotted on to polyvinylidene difluoride membranes for N-terminal microsequence analysis. Amino Acid Analysis—The reaction was stopped by heating with 6 M HCl, 0.1% TFA, and the reaction was followed for an additional 5 min. Values for specific enzymatic activity were calculated by use of protein concentrations determined in the Bradford assay (16), with bovine serum albumin as standard.

Rate constants for irreversible inactivation were not found by non-linear regression analysis of the pseudo-first-order curves of inactivation by use of the FLUSYS software, giving kobs. The second order rate constant, kobs, was calculated as kobs/II, at [S] ≪ Km. Since the assay substrate was used at [S] ≪ Km, no correction for competition with the inhibitors was required.

Purification of Pig Legumain—A temperature of 0–4 °C was maintained throughout the procedure. All buffers contained 1 mM EDTA and 2 mM 2-mercaptoethanol, and centrifugation was at 15,000 × g for 20 min, unless otherwise stated.
0.1 M sodium chloroacetate in 0.1 M Tris/HCl, pH 8.0. Activity was expressed as increase in absorbance at 410 nm relative to a blank without enzyme. A linear plot of activity against ovocystatin concentration allowed the calculation of the molarity of the legumain solution, on the assumption of stoichiometric inhibition.

RESULTS

Cloning and Sequencing of Human and Pig Legumain—Human legumain cDNA clones, containing either the 5' or 3' ends with overlapping regions and the full-length cDNA, were obtained by PCR amplification from human placenta cDNA libraries, and sequenced as described under “Experimental Procedures.” The nucleotide sequence of the full-length cDNA (Fig. 1) comprises 1393 bp with an in-frame ATG codon at nucleotides 71–73 conforming to the Kozak (17) consensus for a translational initiation site. An open reading frame of 1302 bp, with the termination codon at nucleotides 1370–1372, encodes a protein of 433 amino acid residues and 49,406 Daltons.

A clone containing part of pig legumain cDNA was also obtained, by similar PCR-based cloning methods, when pig kidney cDNA libraries were used as templates. The pig legumain clone consists of 884 nucleotides encoding 294 amino acid residues. The deduced amino acid sequence of the pig legumain cDNA corresponds to the region Asn91–Leu384 in human legumain (Fig. 2). Comparison by use of BESTFIT (18) reveals 83% identity at the nucleotide level and 84% at the amino acid level.

Analysis of the N terminus of the deduced amino acid sequence of human legumain with the SIGCLEAVE program of EGCG (19) indicates the presence of a signal peptide, predicted to be cleaved at Ala17-Val18 (Fig. 1).

Detection of Legumain Activity in Mammalian Tissues—Legumain is a cysteine proteinase with a unique sequence motif RGD (residues 118–120). RGD sequences in cell-adhesive proteins such as fibronectin (20, 21) and the disintegrins (22) are responsible for the binding of the proteins to their cell-surface receptors, the integrins, but the significance of the RGD sequence in legumain has yet to be determined.

0.1 M sodium chloroacetate in 0.1 M Tris/HCl, pH 8.0. Activity was expressed as increase in absorbance at 410 nm relative to a blank without enzyme. A linear plot of activity against ovocystatin concentration allowed the calculation of the molarity of the legumain solution, on the assumption of stoichiometric inhibition.

**FIG. 1.** Nucleotide and deduced amino acid sequences of human legumain. Upper lines show the nucleotide sequence. Sequences corresponding to the forward and reverse primers used in PCR are underlined, and their designations are written above. Lower lines represent the deduced amino acid sequence, and the hypothetical N-terminal Gly is marked. The predicted signal peptide is underlined, and the termination codon (TGA) is marked with an asterisk.
gumain is unusual among mammalian proteolytic enzymes in being unaffected by E64 but inhibited by cystatin, and these characteristics were used to confirm the specificity of the assays. E64 is a potent inhibitor of the cysteine peptidases of family C1 such as cathepsins B, H, and L (9), but 10 \( \mu M \) E64 had no significant effect on the rates of hydrolysis of Z-Ala-Ala-Asn-NHMec by tissue extracts. In all cases, the activity was totally inhibited by 50 nM ovocystatin, however. It was noted that with these crude samples, the full activity of legumain was expressed even in the absence of exogenous thiol activator.

The results of assays with a variety of mammalian tissues are shown in Table I. Activity was greatest in kidney, among the rat, pig, and rabbit tissues examined. A low level of activity (0.29 milliunit/mg) was unambiguously detected in human placentae (Fig. 3), but none was detected in blood leukocytes.

**Purification of Legumain from Pig Kidney**—Pig kidney was selected as source for purification of legumain, and the results of a typical preparation are summarized in Table II. We found it necessary to devise an unconventional procedure for the purification of the enzyme, because legumain is stable only in the range pH 3–6 (see below), and at low salt concentrations it tends to be adsorbed to any solid material present. When the 1.2–3.2M ammonium sulfate fraction (stage B) was dialyzed at pH 5.0, a heavy precipitate formed, to which the legumain was adsorbed. The enzyme eluted by raising the salt concentration at pH 6.0 showed a large increase in specific activity (stage C). The enzyme was also bound exceptionally tightly to SP-Sepha-

**FIG. 2.** Alignment of amino acid sequences of legumains. Sequences are numbered according to human preprolegumain, and residues identical to those in human legumain are shown in white on black. Key to sequences: a, human preprolegumain; b, pig legumain (from protein sequencing); c, pig legumain (from cDNA); d, *S. mansoni* prepro- "hemoglobinase" (EMBL accessions M17423 and M21308); and e, *C. ensiformis* preprolegumain (D31787). The alignment was constructed by use of the PILEUP program.
thiols, and bound to Thiopropyl-Sepharose activated with 2-pyridyl disulfide. After overnight exposure to 10 mM cysteine, it was eluted in satisfactory yield. The overall purification factor was about 2200-fold, with a recovery of 8% (stage E).

The final product ran as a single, somewhat diffuse band in SDS-PAGE, at about 34 kDa (Fig. 4). The enzyme was stable to storage at pH 5.8, showing little loss of activity over several months. This material was used for all further characterization of the enzyme.

Partial Amino Acid Sequences of Pig Legumain—N-terminal and internal amino acid sequence data were obtained for the purified pig kidney legumain. An N-terminal sequence of 25 residues was obtained, together with an internal sequence of 19 residues (Fig. 2).

Pig Legumain Is N-Glycosylated—Legumain was run in SDS-PAGE, with and without treatment with N-glycosidase F. As can be seen in Fig. 4, the original band at about 34 kDa was converted to a band at 31 kDa, consistent with deglycosylation. Proteins from a parallel gel were transblotted on to a polyvinylidene difluoride membrane, and stained for glycan. Legumain gave a clear positive reaction, which was lost following treatment with N-glycosidase F, and the positive and negative controls gave the expected results (data not shown).

Dependence of Activity and Stability of Pig Legumain on pH—Legumain solutions in sodium phosphate/citrate, pH 3.0–7.2, were incubated at 30 °C, and samples were removed for assay at 2 and 4 h. It was found (Fig. 5A) that there was no appreciable loss of activity in the range pH 4.2–5.5, and even during 24 h there was little loss of activity in this range (data not shown). Below pH 4.2, stability fell off gradually, but above pH 6.0, it fell off precipitously, and no activity survived 2 h at pH 6.6. The rate of loss of activity above pH 6.0 was strongly dependent on temperature, so that at 25 °C linear rates could be obtained in brief, continuous assays up to pH 7.0.

The pH dependence of activity was determined in the same buffer system, at 25 °C, with the result shown in Fig. 5B. It can be seen that maximal activity occurred at pH 6.4. However, routine assays were more conveniently made at 30 °C for longer periods, and pH 5.8 was selected as the most suitable value for these.

Catalytic Activity of Isolated Pig Legumain—Active site titration of a 1 unit/ml solution of legumain with 3.0 mM ovocystatin showed 2.0 μM reactive sites. For a 34-kDa protein, this would imply a specific activity of 14.7 units/mg. The actual specific activity of combined fractions of our final enzyme preparation in the standard assay, with Z-Ala-Ala-Asn-NHMec at pH 5.8 and 30 °C, was only 7.7 (Table I), but values of 10 or more were seen with individual fractions on several occasions.

Substrate Specificity of Pig Legumain—The enzyme hydrolyzed Z-Ala-Ala-Asn-NHMec and Bz-Asn-NHPHNO₂, but showed no action on Asn-NHMec or Asp-NHMec.

Kinetic parameters for the hydrolysis of Z-Ala-Ala-Asn-NHMec by pig kidney legumain were: $k_{cat}$ 46 s⁻¹, $K_m$ 50 μM, and $k_{cat}/K_m$ 920,000 s⁻¹ M⁻¹. These values are of similar magnitude to those for cathepsins B and L acting on their preferred aminomethylcoumarin substrates (10).

Legumain was tested for hydrolysis of a range of oligopeptide substrates (from Sigma). Each peptide (50 μM in 100 μl) was incubated with 0.3 milliunit of legumain/ml in 50 mM sodium citrate buffer, pH 6.0, containing 1 mM EDTA, 1 mM DTT, for 2 h at 30 °C. Neurotensin, human VIP fragment 1–12, and chicken VIP fragment 16–28 were cleaved. The products were isolated by high performance liquid chromatography, and identified by amino acid analysis, revealing cleavage positions as shown in Fig. 6. It is evident that in each case, cleavage was at an asparaginyl bond.

Other peptides that were cleaved, but for which the products were not identified, were neurotensin fragment 1–11, pig VIP fragment 1–28, bombesin, somatostatin, and allatotropin. All of these peptides contain asparaginyl bonds.

Not all asparagine-containing peptides were cleaved, however; under the conditions of assay, there was no detectable hydrolysis of [Asn¹,Val⁰]angiotensin II, Ala-Ser-Thr-Thr-Asn-Tyr-Thr, or Gly-Ser-Asn-Lys-Gly-Asn-Ile-Ile-Gly-Leu-Met. Other peptides that were not cleaved were bradykinin, substance P, dynorphin A fragment 1–13, and angiotensin I, none of which contains an asparaginyl bond.

Action of Legumain on Proteins—Azocasein (1.5%, w/v) in 50 mM sodium citrate buffer, pH 5.5, containing 1 mM EDTA and 5 mM DTT, was incubated with legumain (5 milliunits/ml) for 1 h at 30 °C. The reaction was stopped by making the mixture 3% (w/v) trichloroacetic acid, and soluble peptides were separated by filtration for quantification by A₂₈₀. The enzyme caused an increase in A₂₈₀ of 0.128 under these conditions. Under the same conditions, cleavage of both bovine serum albumin and human transferrin was evident in SDS-PAGE.
There was no obvious enhancement of the rate of hydrolysis of these proteins by prior denaturation at 100 °C for 5 min.

The recombinant C-fragment of tetanus toxoid (50 μg in 100 μl) was cleaved at three bonds by legumain (5 milliunits, 1 h), under the conditions used for azocasein (Fig. 7). Again, only asparaginyl bonds had been hydrolyzed, but many other asparaginyl bonds in the protein were unaffected. The bonds that were cleaved were in some of the most hydrophilic parts of the molecule, as identified by the Kyte and Doolittle (23) hydrophy prediction algorithm.

**Inhibitors**—Potential inhibitors of pig legumain were tested in the standard assay (Table III). General inhibitors of peptidases of serine, aspartic, and metallo-catalytic type (including phenylmethanesulfonyl fluoride, pepstatin, and 1,10-phenanthroline) had no effect on the enzyme. In contrast, phenylmethylsulfonyl fluoride, pepstatin, and 1,10-phenanthroline were reasonably effective inhibitors even in the presence of 1 mM DTT. E64 had no effect on legumain, even at 100 μM. This value is in the same order of magnitude as the Kᵢ for cathepsin B, but 2 or 3 orders of magnitude higher than those for papain.

**Presence of Legumain in Mammalian Tissues**—The extracts of mammalian tissues hydrolyzed the fluorometric substrate Z-Ala-Ala-Asn-NHMec, and essentially all of the activity had been in kidney, with appreciable activities also in rat spleen and pig liver. A low but significant level of activity was detected in kidney, with appreciable activities also in rat spleen and pig liver. A low but significant level of activity was detected in kidney, with appreciable activities also in rat spleen and pig liver. A low but significant level of activity was detected in kidney, with appreciable activities also in rat spleen and pig liver. A low but significant level of activity was detected in kidney, with appreciable activities also in rat spleen and pig liver. A low but significant level of activity was detected in kidney, with appreciable activities also in rat spleen and pig liver. A low but significant level of activity was detected in kidney, with appreciable activities also in rat spleen and pig liver. A low but significant level of activity was detected in kidney, with appreciable activities also in rat spleen and pig liver.
in human placenta. In view of the countless assays of the proteolytic activities of mammalian tissues that have been made over past decades, it came as a surprise to find a novel enzyme, particularly one that can be detected with a simple protein substrate such as azocasein.

**Catalytic Activity**—Pig legumain was found to be very labile at neutral pH. In this respect it resembles such lysosomal cysteine endopeptidases of family C1 as cathepsins B and L. Legumains of the plants *Phaseolus vulgaris* and *C. ensiformis* are also unstable at pH 7.5 (28, 29), and pH optima are in the region of 5.5 (29, 30). Legumain of *Schistosoma* differs in this respect, with its pH optimum of 6.8, as has been shown directly by Dalton *et al.* (5).

There seems little doubt that the legumains are cysteine peptidases. The purified pig enzyme is active only in the presence of thiol compounds, but the enzyme in crude extracts of mammalian tissues did not require a thiol activator, and the same has been reported for *Schistosoma* (5).

The inhibition properties of pig legumain (Table III) were consistent with a cysteine peptidase. With low nanomolar $K_i$ values, the two cystatins tested were more potent by about 1000-fold than they are against legumains of *V. aconitifolia* and *C. ensiformis*, however (3, 33).

The rate constants for inactivation of pig legumain by iodoacetate and iodoacetamide (Table IV) are very low compared to those for a typical cysteine endopeptidase of family C1 such as papain (31), but are of similar magnitude to those for glycy endopeptidase and bromelain (32). It was notable that the maleimides were much more reactive, especially N-phenylmaleimide. 3,4-Dichloroisocoumarin is normally regarded as an inhibitor specific for serine peptidases, so it was not expected to inhibit legumain, but it has been found to inhibit calpains (family C2) and caspase 1 (C14).3

Legumain of *C. ensiformis* is highly specific for the cleavage of asparaginyl bonds (33), and all indications are that the pig enzyme is equally selective. However, the legumains may cleave only some of the asparaginyl bonds in a polypeptide substrate. This was suggested by the fact that several of the asparagine-containing oligopeptides tested were not hydrolyzed, and only 3 of the 47 asparaginyl bonds in the C-fragment of tetanus toxoid were cleaved. This implies the existence of additional determinants of specificity that have yet to be identified. The specificity of plant legumain in its action on seed proteins seems to include a preference for hydrophilic surface loops (34), and a similar preference would be consistent with the action of pig legumain on the C-fragment of tetanus toxoid reported here.

There has been no report of a peptidase with strict specificity for asparaginyl bonds apart from those for species variants of legumain, so there are no direct precedents as a basis for speculation about the structural mechanism of this specificity. Perhaps the closest analogy is the strong preference of picornains for cleavage of -Gln-Gly- bonds in their processing of picornaviral polyproteins, and in these enzymes, a histidine side-chain seems primarily responsible for the interaction with glutamine (35).

**Possible Biological Functions**—Legumain was previously known from plants and a simple animal, *Schistosoma*, and we now know that it is present also in mammals. The enzyme was evidently present in the protozoan ancestor of plants and animals at the time of their divergence about 1000 million years ago, and since the strict specificity for hydrolysis of asparaginyl bonds is seen in both plants and animals, there is little doubt that it was also exhibited by the archetypal enzyme. The functional conservation of the enzyme across such a wide spread of organisms suggests that there may be some fundamental biological importance in the hydrolysis of asparaginyl bonds. The
The protein was incubated with legumain. If confirmed, this finding will be consistent with the properties of the enzyme, since it is synthesized with a signal peptide, and is N-glycosylated, and moreover has a requirement for an acidic environment. Plant legumain is present in the vacuole, an analog of the lysosome (34, 40), and the enzyme processes a papain-family endopeptidase, SH-EP, in Vigna mungo (mung bean; Ref. 41).

In C. ensiformis, legumain apparently catalyzes the transpeptidation of concanavalin A (42), but whether a protein-splicing function could be significant for mammalian legumain remains to be established.

It is perhaps too early in the study of legumain in mammals to have clear ideas about possible disease involvement, but a significant report is that of Sharma et al. (43), who have detected an enzyme in lens that may be responsible for an age-dependent degradation of α-crystallin. The enzyme hydrolyzes an Asn-Glu bond and is inhibited by N-ethylmaleimide, but not E64, and therefore possibly is legumain.

Legumain may well be secreted from cells. The instability of the enzyme at neutral pH would potentially restrict its extracellular activity, except perhaps in an acidic pericellular environment. The interactions of the RGD sequence with membrane components may help to retain it in this environment.

In summary, we have shown that mammalian tissues contain a cysteine endopeptidase of a family not previously known to be present in mammals, and with a distinctive specificity. By analogy with the important functions of the other families of cysteine peptidases found in mammals (families C1 (cathepsin B, C2 (calpains), and C14 (caspase)), family C13 may also be present in mammals, and with a distinctive specificity. By analogy with the important functions of the other families of cysteine peptidases found in mammals (families C1 (cathepsin B, C2 (calpains), and C14 (caspase)), family C13 may also prove to be of considerable biomedical interest.

**Acknowledgments**—We thank John Coadwell (The Babraham Institute) for much help with the bioinformatics.

**REFERENCES**

1. Rawlings, N. D., and Barrett, A. J. (1994) Methods Enzymol. 244, 461–486
2. Barrett, A. J., and Rawlings, N. D. (1996) Perspect. Drug Discov. Design 6, 1–11
3. Kembhavi, A. A., Buttle, D. J., Knight, C. G., and Barrett, A. J. (1993) Arch. Biochem. Biophys. 303, 208–213
4. Hara-Nishimura, I., Takeuchi, Y., and Nishimura, M. (1993) Plant Cell Physiol. 34, 1651–1659
5. Dalton, J. P., Hata-Jamriska, L., and Brindley, P. J. (1995) Parasitology 111, 575–580
6. Dawson, R. M. C., Elliott, D. C., Elliott, W. H., and Jones, K. M. (1987) Data
7. 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
8. Barrett, A. J. (1981) Methods Enzymol. 80, 771–778
9. Barrett, A. J., Kembhavi, A. A., Brown, M. A., Kirschke, H., Knight, C. G., Tamai, M., and Hanada, K. (1982) Biochem. J. 201, 189–198
10. Barrett, A. J., and Kirschke, H. (1981) Methods Enzymol. 80, 535–561
11. Lundberg, K. S., Shoemaker, S. D., Adams, M. W. W., Short, J. M., Sorge, J. A., and Mathur, E. J. (1991) Gene (Amst.) 108, 1–6
12. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 5463–5467
13. Frohman, M. A., Dush, M. K., and Martin, G. R. (1988) Proc. Natl. Acad. Sci. U. S. A. 85, 8998–9002
14. Bury, A. F. (1981) J. Chromatogr. 213, 491–500
15. Rawlings, N. D., and Barrett, A. J. (1990) Comput. Appl. Biosci. 6, 118–119
16. Bradford, M. M. (1976) Anal. Biochem. 72, 248–254
17. Kozak, M. (1991) J. Biol. Chem. 266, 19867–19870
18. Genetics Computer Group (1994) Program Manual for the Wisconsin Package, Version 8, University of Wisconsin, Madison, WI
19. Lipman, D. J., and Pearson, W. R. (1985) Science 227, 1435–1441
20. Alonso, J. M., and Granell, A. (1995) Plant Physiol. 109, 541–547
21. 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
22. Ceoma, C., and Polgar, L. (1984) Biochem. J. 222, 769–776
23. Koyama, H., Iwata, K., and Nishimura, M. (1991) FEBS Lett. 294, 89–93
24. Bottart, A., Cope, A., Galleschi, L., Jopova, A., and Savitski, F. (1996) Physiol. Plant. 97, 475–480
25. Daou, D., and Kestere, K., and Elser, N. (1996) Biochem. Biophys. Res. Commun. 231, 365–370
26. Makoff, A. J., Ballantine, S. P., Smallwood, A., and Fairweather, N. F. (1989) Biotechnology 7, 1043–1046