Pig kidney legumain: an asparaginyl endopeptidase with restricted specificity

Pam M. DANDO*[1], Mara FORTUNATO*, Lorraine SMITH*, C. Graham KNIGHT†, John E. McKENDRICK‡ and Alan J. BARRETT*

*MRC Molecular Enzymology Laboratory, The Babraham Institute, Babraham, Cambridgeshire CB2 4AT, U.K., †Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QW, U.K., and ‡Department of Chemistry, University of Alberta, Edmonton, Alberta T6G 2G2, Canada

Legumain was recently discovered as a lysosomal endopeptidase in mammals [Chen, Dando, Rawlings, Brown, Young, Stevens, Hewitt, Watts and Barrett (1997) J. Biol. Chem. **272**, 8090–8098], having been known previously only from plants and invertebrates. It has been shown to play a key role in processing of the C fragment of tetanus toxin for presentation by the MHC class-II system [Manoury, Hewitt, Morrice, Dando, Barrett and Watts (1998) Nature (London) **396**, 695–699]. We examine here the specificity of the enzyme from pig kidney by use of protein, oligopeptide and synthetic arylamide substrates, all determinations being made at pH 5.8. In proteins, only about one in ten of the asparaginyl bonds were hydrolysed, and these were mostly predicted to be located at turns on the protein surface. Bonds that were not cleaved in tetanus toxin were cleaved when presented in oligopeptides, sometimes faster than an equivalent oligopeptide based on a bond that was cleaved in the protein. Legumain cleaved the bait region of rat α_1-macroglobulin and was 'trapped' by the macroglobulin, as most other endopeptidases are, but did not interact with human α_2-macroglobulin, which contains no asparagine residue in its bait region. Glycosylation of asparagine totally prevented hydrolysis by legumain. Specificity for arylamide substrates was evaluated with reference to benzyloxycarbonyl-Ala-Ala-Asn-aminomethylcoumarin, and the preference for the P3-position amino acid was Ala > Tyr (tertiary butyl) > Val > Pro > Phe = Tyr > Leu = Gly. There was no hydrolysis of substrate analogues containing mono- or di-*N*-methylasparagines, L-2-amino-3-ureidopropionic acid or citrulline in the P1 position. We conclude that mammalian legumain appears to be totally restricted to the hydrolysis of asparaginyl bonds in substrates of all kinds. There seem to be no strong preferences for particular amino acids in other subsites, and yet there are still unidentified factors that prevent hydrolysis of many asparaginyl bonds in proteins.

Key words: cysteine endopeptidase, limited proteolysis, lysosomal endopeptidase, protein processing.

INTRODUCTION

Legumain (EC 3.4.22.34) is a cysteine endopeptidase of peptidase family C13 [1] previously known only from plants [2–4] and the trematode *Schistosoma* [5], but recently discovered in mammals too [6,7]. Plant legumain has been found to be selective for hydrolysis of asparaginyl bonds, and the specificity of jack-bean legumain has been studied in detail in conjunction with the commercial supply of the enzyme for use in the amino acid sequencing of proteins [8]. It was reported that any of 20 amino acids could occur in the P1′ position. The biological functions of plant legumain include the post-translational processing of storage proteins, lectins and enzymes [3,9].

Mammalian legumain has recently been shown to be implicated in protein processing for the MHC class-II system, and specifically to play a key role in the processing of a bacterial antigen, tetanus toxin [10]. The substrate specificity of the enzyme has not previously been characterized in detail, but in the present paper we describe the action of legumain from pig kidney cortex on proteins including tetanus toxin, oligopeptides and synthetic arylamide substrates.

EXPERIMENTAL

Materials

Cattle albumin, cattle casein 1, chicken lysozyme C, horse myoglobin, horse ferritin, human transferrin and neurotensin were all obtained from Sigma Chemicals. Recombinant A and B chains of concanavalin were given by Dr. D. H. Jones, School of Biological Sciences, University of Wales, Swansea, U.K., and recombinant His-tagged tetanus toxoid C fragment, 9-fluorenylmethoxycarbonyl (Fmoc)-Ala-Glu-Asn-Lys-NH_2 and Fmoc-Lys-Asn-Asn-Glu-NH_2 were given by Dr. C. Watts, Department of Biochemistry, University of Dundee, Dundee, U.K. Peptides derived from tetanus toxoid C fragment (peptides 1–6, 8 and 10–12 in Table 2, see below) were synthesized by MWG-BIOTECH UK Ltd. (Milton Keynes, U.K.). Peptide 22 [QF27, (7-methoxycoumarin-4-yl)acetyl-Nle-Ala-Val-Lys-Tyr-Leu-Asn-Ser-Lys(Dnp)-Leu-Asp-D-Lys] was as described [11]. Other peptides based on tetanus toxoid C fragment, i.e. peptides 7 and 9 (see Table 2, below), and the N^{δ}-glucosaminyl-Asn[4] derivative of peptide 7, were synthesized by Genosys Europe (Cambridge, U.K.), as was (N^{δ}-glucosaminyl-Asn[5])neurotensin with acetyl protection of the carbohydrate hydroxyls. Peptides derived from ovalbumin were given by Dr. J. Stevens, Department of Immunology, The Babraham Institute, and were > 75% pure in HPLC. β-Amyloid protein 1–40 was provided by Dr. G. Christie, SmithKline Beecham Pharmaceuticals, Harlow, U.K., and the derived peptides 25–35 GSNKGAIIGLM and 35–25 MLGIIAGKNSG were purchased from Bachem (UK) Ltd. (Saffron Walden, Essex, U.K.). Rat liver tyrosine aminopeptidase and a peptide (residues 22–35, VNIGGRNSVQGRK) from the enzyme were given by Dr. J. Hargrove (University of Georgia, Atlanta, GA, U.S.A.). All other peptides were obtained from Sigma

Abbreviations used: Asn(Me), N^{δ}-methylasparagine; $Asn(Me)_2$, N^{δ},N^{δ}-dimethylasparagine; Aup, L-2-amino-3-ureidopropionic acid; Fmoc, 9-fluorenylmethoxycarbonyl; NHMec, 7-(4-methyl)coumarylamide; Z, benzyloxycarbonyl; QF27, (7-methoxycoumarin-4-yl)acetyl-Nle-Ala-Val-Lys-Tyr-Leu-Asn-Ser-Lys(Dnp)-Leu-Asp-D-Lys.

[1] To whom correspondence should be addressed.

Chemicals. N-terminal microsequencing was done by the Microchemical Facility at The Babraham Institute by use of an Applied Biosystems Procise 492. All other chemicals were of analytical or HPLC grade.

Purification of pig kidney legumain

Legumain (7 units/mg) was prepared from pig kidney cortex as described previously [6], except that Activated Thiol Sepharose 4B was used in place of Thiopropyl-Sepharose 6B in the covalent chromatography step.

Arylamide substrates

Z-Xaa-Ala-Asn-NHMec [in which Z is benzyloxycarbonyl, Xaa is Gly, Leu, Phe, Tyr or Val and NHMec is 7-(4-methyl)-coumarylamide], Z-Ala-Ala-Aup-NHMec (in which Aup is L-2-amino-3-ureidopropionic acid) and Z-Ala-Ala-Cit-NHMec (in which Cit is citrulline) were synthesized generally as described in [2,12]. Z-Asn(Me)-$NHPhNO_2$ [in which Asn(Me) is N^{δ}-methyl-asparagine, and $NHPhNO_2$ is *p*-nitroanilide] was synthesized by reaction of Z-Asp-t-butyl ester (in dichloromethane) with methylamine hydrochloride in the presence of triethylamine and isobutylchloroformate. The resulting Z-Asn(Me)-t-butyl ester was de-esterified with trifluoroacetic acid (20 %, v/v) at 0 °C and reacted with *p*-nitroaniline in the presence of phosphorus oxychloride. Z-$Asn(Me)_2$-$NHPhNO_2$ [in which $Asn(Me)_2$ is N^{δ},N^{δ}-dimethylasparagine] was synthesized by an analogous route from Z-Asp-t-butyl ester and dimethylamine hydrochloride. All the compounds gave satisfactory elemental analyses and mass spectra.

SDS/PAGE, transblots and N-terminal micro-sequencing

SDS/PAGE was as described [13] and electrophoretic transfer of separated proteins to PVDF membrane was done at 80 mA for 1.5 h in Caps buffer [3-(cyclohexylamino)propane-1-sulphonic acid], pH 11.

Action of legumain on proteins

The action of legumain on proteins was investigated with native substrates and substrates denatured by boiling in 1 % SDS for 5 min. Protein (20 μg in 50 μl) was incubated at 30 °C for up to 4 h with 5 m-units/ml of pig legumain in 50 μl of 39.5 mM citric acid/121 mM Na_2HPO_4, pH 5.8, containing 1 mM EDTA, 1 mM dithiothreitol and 0.01 % CHAPS (Buffer A). For assays with denatured protein the final SDS concentration was 0.02 %. The reactions were stopped by boiling for 5 min with an equal volume of SDS sample buffer. The polypeptide products were separated by SDS/PAGE (5 μg/gel lane), transblotted on to PVDF membrane and stained with Coomassie Brilliant Blue R. Peptide bands were excised and analysed by N-terminal microsequencing. Rat liver tyrosine aminotransferase was incubated with legumain (5 m-units/ml in Buffer A for 16 h).

Interaction of legumain with α-macroglobulins

Human plasma for the purification of α_2-macroglobulin was provided by the East Anglian Blood Centre, Addenbrookes Hospital, Cambridge, U.K. Human α_2-macroglobulin and rat α_1-macroglobulin were purified from plasma according to the methods of Barrett et al. [14] and Lonberg-Holm et al. [15], respectively. Each macroglobulin was incubated at 30 °C for 5 min with pig legumain in Buffer A supplemented with 0.2 M NaCl. The molar ratio (on a protein basis) of legumain to macroglobulin was approx. 1:2 (2.25 μM/5 μM respectively). The mixture was separated by gel filtration (SP Sepharose 100 HR) with monitoring at 280 nm. Fractions were assayed for legumain activity with the substrate Z-Ala-Ala-Asn-NHMec in the usual way, and those containing activity were subjected to SDS/PAGE. Bands were transblotted on to PVDF membrane, stained with Coomassie Brilliant Blue R, and cut out for N-terminal microsequencing.

Peptide hydrolysis by legumain

Peptides were incubated at 10 μM and/or 20 μM initial substrate concentration with 1 and/or 5 m-units/ml of legumain in Buffer A containing 1 mM Tris(carboxyethyl)phosphine instead of 1 mM dithiothreitol. Incubation was for up to 2 h at 30 °C. Samples were removed at intervals and the reaction stopped by addition of glacial acetic acid to 10 % (v/v). The products of hydrolysis were separated by HPLC on a Waters Discovery system equipped with photodiode array detection and a C_{18} 5μ, 300 Å column (2 mm × 150 mm). The gradient was linear from 5 to 55 % (v/v) acetonitrile containing 0.1 % trifluoroacetic acid over 45 min at 0.2 ml/min. The column effluent was monitored at 214 nm and peak areas were integrated by the Waters Millenium software. For each time point, the areas under the peaks representing products and unchanged substrate were recorded. The concentration of substrate remaining, [*S*], was calculated as:

$$[S] = [S_0]\left(\frac{\text{Area of peak for unchanged substrate}}{\text{Total area of new peaks} + \text{unchanged substrate}}\right) \quad (1)$$

where $[S_0]$ is the initial substrate concentration. Values of k_{cat}/K_m for each substrate were determined, on the assumption that $K_m \gg 20\ \mu$M for all peptides, by use of the equation:

$$k_{cat}/K_m = [2.303 \cdot \log(S_0/S)]/(t \cdot E_0) \quad (2)$$

where t is time and E_0 is enzyme concentration, assuming the specific activity of legumain to be 10 units/mg of protein. At least two separate determinations were made for each peptide, and the mean value is given.

Hydrolysis of arylamide substrates

Legumain was routinely assayed with Z-Ala-Ala-Asn-NHMec (10 μM) in Buffer A as described in [6]. Activity (1 unit) was defined as that releasing 1 μmol of product under the standard conditions. In the continuous fluorimetric assays, excitation and emission were at 360 and 460 nm for the peptidyl-NHMec substrates and 328 and 393 nm for QF27, respectively. The hydrolysis of *p*-nitroanilides was detected as an increase in absorbance at 410 nm during incubation of the substrate (100 μM) in Buffer A at 30 °C with 10 or 100 m-units of legumain/ml for up to 16 h. Values for k_{cat}/K_m for each substrate were determined according to the equation:

$$k_{cat}/K_m = v/(E_0 \cdot [S]) \quad (3)$$

RESULTS

Hydrolysis of proteins

Proteins were tested as substrates for legumain in the native and SDS-denatured states as described in the Experimental section. Cattle casein 1, human transferrin and concanavalin A were hydrolysed in the native state by legumain at 5 m-units/ml. These proteins were also tested after denaturation, and the same

Table 1 Points of cleavage of proteins by legumain

The bonds hydrolysed are indicated by the residue number of the P1 asparagine residue, as given in the SWISSPROT or PIR databases [29,30]. The proteins were cattle casein 1 (SP:P02662), cattle serum albumin (SP:P02787; SDS-denatured), chicken lysozyme C (SP:P00698; SDS-denatured), concanavalin A (SP:P08253), human transferrin (numbered as serotransferrin precursor SP:P02787) and rat α_1-macroglobulin (PIR:A42210). The residue numbers in the sequence of recombinant tetanus toxoid C fragment are as described previously [6].

Protein	Position of cleavage	P5	P4	P3	P2	P1	P1′	P2′	P3′	P4′	P5′
Casein 1	95	E	I	V	P	**N**	S	V	E	Q	K
Serum albumin	324	A	I	P	E	**N**	L	P	P	L	T
	404	Y	G	F	Q	**N**	A	L	I	V	R
Lysozyme C	62	T	Q	A	T	**N**	R	N	T	D	G
	64	A	T	N	R	**N**	T	D	G	S	D
Concanavalin A, A-chain	159	T	I	D	F	**N**	A	A	Y	N	A
	163	N	A	A	Y	**N**	A	D	T	I	V
Transferrin	95	L	A	P	N	**N**	L	K	P	V	V
	529	C	E	P	N	**N**	K	E	G	Y	Y
	574	A	K	N	L	**N**	E	K	D	Y	E
	603	A	R	A	P	**N**	H	A	V	V	T
Tetanus toxoid C fragment	26	R	H	I	D	**N**	E	E	D	I	D
	337	Y	T	P	N	**N**	E	I	D	S	F
	372	G	N	A	F	**N**	N	L	D	R	I
Rat α_1-macroglobulin	721	Y	R	S	S	**N**	I	R	T	S	S
	899	E	L	C	G	**N**	K	V	A	E	V

results were obtained. Cattle serum albumin and chicken lysozyme C were hydrolysed only after SDS denaturation, or by use of 50 m-units of legumain/ml. No cleavage of ferritin, myoglobin, the B chain of concanavalin A or tyrosine aminotransferase was seen, even after SDS denaturation or with the increased amount of enzyme. The recombinant tetanus toxoid C fragment was also hydrolysed while native, and data obtained previously [6] have been extended in the present study. SDS/PAGE showed that for each protein that was hydrolysed by legumain, several major fragments were formed and these then remained stable. This was an indication that the bonds that were not readily cleaved were not cleaved at all. The N-termini of the fragments were identified, and all new N-termini were found to have arisen from cleavage of asparaginyl bonds, as listed in Table 1. The specificity of legumain for asparaginyl bonds was thus confirmed, but it was striking that by no means all such bonds were hydrolysed. Indeed, of 163 asparaginyl bonds in the proteins tested, only 16 were cleaved. The possibility was considered that additional bonds had been cleaved but had yielded peptides too small for resolution in SDS/PAGE, but this was excluded by the finding that HPLC analysis of the reaction mixtures showed negligible amounts of small peptides (capable of elution from a C_{18} reverse-phase column). We concluded that legumain hydrolyses only asparaginyl bonds in the proteins tested, but that a high proportion (about 90%) of the asparaginyl bonds in these proteins were highly resistant to cleavage by the enzyme.

We considered whether any aspects of the amino acid sequences around the asparaginyl bonds in the test proteins could account for their resistance to hydrolysis. It can be seen in Table 1 that a wide variety of amino acids was present in each of the specificity subsites P5–P2 and P1′–P5′ (as defined in [16]); there was no evident requirement for specific amino acids in any subsite other than P1. It was noted, however, that half of the residues in the P3 position were either Pro or Ala (four of each), suggesting that these may be particularly favourable. Examination of the sequences around the asparaginyl bonds that were not cleaved similarly led to no definite conclusions, although it was observed that Gly had occurred much less frequently in the P3, P2, P1′ and P2′ positions of scissile bonds than it did around asparagine residues generally.

Next, we considered whether higher-level structure of the substrate proteins could account for the selective cleavage of asparaginyl bonds. For serum albumin, lysozyme and transferrin, crystallographic structures are to be found in the Brookhaven Protein Databank. These showed that of eight cleavages, four occurred at turns, two at the ends of α helices into turns, and two in β strands, all in surface locations. For those proteins with no published tertiary structure, secondary-structure predictions were made using the PEPTIDESTRUCTURE module of the Genetics Computer Group software package [17]. The predictions indicated that most of the asparaginyl bonds hydrolysed were on the surfaces of the substrate molecules, and six of eight were predicted to be at turns or at the ends of α helices.

Legumain and α-macroglobulins

The α-macroglobulins are high-molecular-mass plasma proteins that are able to form complexes with endopeptidases regardless of catalytic class. Binding is initiated by cleavage of the molecule in the 'bait region' and this leads to physical trapping of the peptidase molecule within that of the macroglobulin [18]. The bait regions of the macroglobulins are exquisitely sensitive to endopeptidase cleavage, being cleaved by the great majority of endopeptidases. Rat α_1-macroglobulin contains one asparagine residue in the bait region (residues 689–743; PIR:A42210, from the PIR database), whereas the bait region of human α_2-macroglobulin (residues 690–730; SP:P01023, from the SWISSPROT database) contains no asparagine (Figure 1).

When legumain was allowed to interact with rat macroglobulin, and the mixture was run in gel filtration as described in the Experimental section, about half the legumain activity was found to co-elute with the macroglobulin peak, the remainder of the activity being eluted in later fractions. N-terminal sequencing of polypeptide bands transblotted to PVDF membrane after SDS/PAGE separation of macroglobulin-containing fractions showed that rat α_1-macroglobulin had been cleaved by legumain at the asparagine residue of the bait region, Asn^{721}–Ile^{722}. With longer exposure to legumain there was a further cleavage of the bond Asn^{899}–Lys^{900}. The experiment was repeated with human α_2-macroglobulin, but with this protein all the legumain activity

a) Human α_2-macroglobulin (678-734)

-F T N S K I R K P K M C **P Q L Q Q Y I M H G P E G L R V G F Y E S D I M G**
R G H A R L V H V E E P H T *E T V R-*

b) Rat α_1-macroglobulin (677-747)

-F T N T T V H K P R Y C **P M Y Q A Y P P L P Y V G E P Q A L A M S A I P G**
A G Y R S S [N] **I R T S S M M M M G A S E V A Q E V E V R** *E T V R*

Figure 1 Amino acid sequences of bait regions of macroglobulins

The partial sequences are taken from the database entries SP:P01023 (human α_2-macroglobulin) and PIR:A42210 (rat α_1-macroglobulin). The residues printed in bold are those identified as forming the bait region by Sottrup-Jensen et al. [23]. It can be seen that a single asparagine residue (boxed) is present in the rat sequence, but none in the human. The new N-terminal amino acids identified after cleavage of the rat α_1-macroglobulin bait region by legumain are double underlined.

Table 2 Hydrolysis of asparagine-containing peptides derived from recombinant tetanus toxoid C fragment

All peptides were decapeptides in which the octapeptide sequence shown was extended by one Ala residue at either end. Residue numbers in the protein are as in [6]. The data are arranged in order of decreasing k_{cat}/K_m. Peptides 1–6, 8 and 10–12 correspond to some of the segments of the protein sequence that contain Asn residues but were not cleaved by legumain. Peptide 9 contains one of the three bonds that were hydrolysed in the protein, and peptide 7 is a form of peptide 9 modified to contain the potential glycosylation site, Asn-Glu-Ser. n/a, not applicable; n/p, not present.

		Sequence										
Peptide number	Asn position in protein	P4	P3	P2	P1	P1′	P2′	P3′	P4′	Hydrolysed in protein	Hydrolysed in peptide	k_{cat}/K_m for peptide (10^4 $M^{-1} \cdot s^{-1}$)
1	125	Y	G	T	**N**	E	Y	S	I	No	Yes	14.20
2	191	T	I	T	**N**	D	R	L	S	No	Yes	5.50
3	151	L	K	G	**N**	N	L	I	W	No	Yes	5.30
4	56	S	G	F	**N**	S	S	V	I	No	Yes	4.40
5	324	R	L	Y	**N**	G	L	K	F	No	Yes	3.40
6	297	Q	L	K	**N**	I	T	D	Y	No	Yes	3.10
7	n/p	H	I	D	**N**	E	S	D	I	n/p	Yes	0.64
8	425	G	T	H	**N**	G	Q	I	G	No	Yes	0.47
9	26	H	I	D	**N**	E	E	D	I	Yes	Yes	0.45
10	221	R	E	D	**N**	N	I	T	L	No	Yes	0.14
11	433	N	D	P	**N**	R	D	I	L	No	No	n/a
12	445	W	Y	F	**N**	H	L	K	D	No	No	n/a

was present in a late peak completely separate from the early-eluting macroglobulin fraction (results not shown). This could be accounted for by the fact that the bait region of human α_2-macroglobulin does not contain an Asn residue.

Action of legumain on peptides from tetanus toxoid

Of the 47 asparaginyl bonds in tetanus toxoid recombinant C fragment, three were readily cleaved by legumain but the remainder were unaffected even during prolonged incubation. To determine whether primary structure surrounding the resistant bonds prevented their hydrolysis, we obtained oligopeptides with asparaginyl residues in the same sequences. Each of ten decapeptides (peptides 1–6, 8 and 10–12) contained eight residues of the original sequence with the Asn residue in position 4, and an additional Ala residue at either terminus. The products of hydrolysis of these peptides by legumain were analysed by HPLC as usual, and specificity constants were calculated. It should be noted that peptides 3 and 5 each contained two Asn residues, and the HPLC analysis showed that both were cleaved in two positions. Table 2 shows that the efficiency of cleavage of the peptides ranged over more than two orders of magnitude, and only two of the peptides remained uncleaved. For further comparison, data were obtained for an octapeptide, peptide 9, containing one of the cleavage sites in the protein, Asn^{26}. Strikingly, specificity constants for this peptide and the closely similar peptide 7 were lower than those of several of the peptides that contained asparaginyl bonds not hydrolysed in the protein.

Hydrolysis of other oligopeptides

Table 3 shows comparative values for specificity constants of 26 other oligopeptides. Peptides 14, 15 and 21 each contained two Asn residues, and could be seen in the HPLC analysis to be cleaved twice. The three peptides most efficiently cleaved included two of the smallest in the series (peptides 13 and 14), and peptide 13 has been used successfully as a competing substrate to retard the hydrolysis of tetanus toxin [10]. It can be seen that cleavage can occur at an Asn residue with just one C-terminal amino acid if that is amidated (peptide 13), or with two C-terminal amino acids (peptides 28 and 33). Similarly, one blocked residue on the amino side is sufficient (peptide 14), as are two unblocked N-

Table 3 Relative rates of hydrolysis of other oligopeptides

The peptides were incubated at 10 or 20 μM (initial concentration) with 5 m-units of legumain/ml for various times at 30 °C as described in the text, and cleavage was monitored by HPLC. The amino acid sequences of the peptides are shown in the one-letter code, except that X is used for groups for which no one-letter code exists. In peptides 17, 19 and 27 the N-terminal X is pyroglutamyl. In peptides 27, 29 and 34 the C-terminal X is NH_2. Peptide 22, based on the sequence of peptide 19, is (7-methoxycoumarin-4-yl)acetyl-Nle-Ala-Val-Lys-Tyr-Leu-Asn-Ser-Lys(Dnp)-Leu-Asp-D-Lys, in which Nle = L-norleucine. Two sequences are truncated, as indicated by the ~ symbol; these are the oxidized insulin B chain, peptide 30, which is FVNQHLCGSHLVEALYLVCGERGFFYTPKA (in which C = cysteic acid), and the β-amyloid protein 1–40, peptide 20, which is DAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVV. VIP, vasoactive intestinal peptide.

Peptide number	Identity	Total residues	Sequence	k_{cat}/K_m (10^4 $M^{-1} \cdot s^{-1}$)
13	Fmoc-AENK-NH_2	4	XAE**N**KX	34.5
14	Fmoc-KNNE-NH_2	4	XK**NN**EX	22.6
15	Ovalbumin 158–171	13	WVESQT**N**GII**N**VL	22.2
16	Ovalbumin 367–377	11	CIKHIAT**N**AVL	13.9
17	Neurotensin	13	XLYE**N**KPRRPYIL	11.0
18	Peptide 18 (22–35)	13	VNIGGR**N**SVQGRK	> 11.0
19	VIP 16–28	13	XMAVKKYL**N**SVLT	5.5
20	β-Amyloid 1–40	40	~ FAEDVGS**N**KGAIIGLMV ~	4.0
21	Ovalbumin 19–27	9	KVHHA**N**E**N**I	3.9
22	QF27	12	XXAVKYL**N**SXLDX	3.6
23	Ovalbumin 172–183	11	MVLV**N**AIEKGL	2.4
24	VIP 1–12	12	HSDAVFTD**N**YTR	2.2
25	Ovalbumin 52–62	10	QI**N**KVVRFDL	2.1
26	Ovalbumin 92–101	10	KP**N**DVYSFSL	1.3
27	Bombesin	14	XQRLG**N**QWAVGHLMX	1.1
28	Peptide T	8	ASTTT**N**YT	1.0
29	Allatotropin	13	GFK**N**VEMMTARGFX	0.5
30	Insulin B chain	30	FV**N**QHLCGSHLV ~	0.4
31	Ovalbumin 257–264	8	SII**N**FEKL	0.3
32	β-Amyloid 25–35	11	GS**N**KGAIIGLM	0.09
33	β-Amyloid 35–25	11	MLGIIAGK**N**SG	0.05
34	Ranakinin	11	KP**N**PERFYGLMX	Uncleaved
35	Eglin c-fragment methyl ester	4	T**N**VVX	Uncleaved
36	[Asn^1,Val^5,] angiotensin II	8	**N**RVYVHPF	Uncleaved
37	β-Amyloid 1–16	16	DAEFRHDSGYEVHHQK	Uncleaved
38	Peptide 38	9	PSLDDEFDL	Uncleaved

Table 4 Action on potential arylamide substrates

Values of k_{cat}/K_m were determined as described in the text. Asp(OMe), methyl ester of aspartic acid.

Substrate	k_{cat}/K_m (10^4 $M^{-1} \cdot s^{-1}$)
Z-Ala-Ala-Asn-NHMec	90
Z-Tyr(tBu)-Ala-Asn-NHMec	89
Z-Val-Ala-Asn-NHMec	75
Z-Pro-Ala-Asn-NHMec	57
Z-Phe-Ala-Asn-NHMec	52
Z-Tyr-Ala-Asn-NHMec	51
Z-Leu-Ala-Asn-NHMec	43
Z-Gly-Ala-Asn-NHMec	40
Z-Ala-Ala-Cit-NHMec	Uncleaved
Z-Ala-Ala-Aup-NHMec	Uncleaved
Asn-NHMec	Uncleaved
Bz-Asn-$NHPhNO_2$	0.4
Z-Asp(OMe)-$NHPhNO_2$	Uncleaved
Z-Asn(Me)-$NHPhNO_2$	Uncleaved
Z-$Asn(Me)_2$-$NHPhNO_2$	Uncleaved

terminal residues (peptides 25 and 26). Peptides 35 and 36 containing one and no residues N-terminal to the Asn residue, respectively, were not cleaved. Although peptides 32 and 33, containing only two amino acids to the N- and C-termini, respectively, were both hydrolysed, the longer β-amyloid-derived peptide 20 was hydrolysed much more efficiently. Peptide 38 contained an Asp–Asp sequence that has been reported to be cleaved by a plant form of legumain [19], but it was unaffected by the mammalian legumain. In our series, no Asn–Pro bond was found to be cleaved in either protein or oligopeptide.

Effect of glycosylation of an asparagine residue on legumain activity

The bioactive 13 residue peptide neurotensin is cleaved by legumain between Asn^5 and Lys^6 [6]. In order to determine whether legumain can cleave glycosylated asparagine residues, a derivative of neurotensin was synthesized containing Asn-(Ac_3AcNH-β-Glc) at residue 5. The glycosylated neurotensin was incubated at 30 °C with legumain (50 m-units/ml) for 16 h, but no evidence of hydrolysis was seen in the subsequent HPLC analysis. Similarly, the N^3-glucosaminyl-Asn^4 derivative of peptide 7 was resistant to hydrolysis.

Action of legumain on arylamide substrates

The activity of legumain was determined in fluorimetric assays with a series of peptidyl aminomethylcoumarin substrates in Buffer A at pH 5.8, 30 °C. Substrates were used at 2 μM concentration, which was in all cases far below the K_m value ($v \propto [S]$), so we were able to calculate values for the catalytic constant, k_{cat}/K_m, by use of eqn. (3). The results obtained (Table 4) show that replacement of Asn in P1 by two chemically similar amino acids, Aup and citrulline, prevented hydrolysis. In P3,

Tyr(tertiary butyl) gave a substrate that was hydrolysed almost as rapidly as the reference compound, Z-Ala-Ala-Asn-NHMec, but Val, Pro, Phe, Tyr and Gly gave progressively poorer substrates.

Spectrophotometric assays were also made with some potential *p*-nitroanilide substrates, containing asparagine, aspartic acid methyl ester, Asn(Me) or $Asn(Me)_2$ in the P1 position. Benzoyl-Asn-$NHPhNO_2$ was described previously as a substrate for a plant legumain [20]. We found (see Table 4) that the asparagine compound was hydrolysed, but those containing analogues of asparagine were again totally resistant to hydrolysis.

DISCUSSION

All lysosomal endopeptidases known prior to the discovery of mammalian legumain show broad action on proteins, so the strict specificity of legumain for asparaginyl bonds is striking. Not only is legumain restricted to hydrolysis of asparaginyl bonds but only about one in ten of these bonds was found to be hydrolysed even in substrate proteins that were SDS-denatured. We were not able to relate the selectivity for particular asparaginyl bonds to any characteristics of the primary structure. Most of the bonds cleaved in proteins were either known or predicted to be located on the surface of the molecule and at turns, but these are common locations for asparagine in these proteins.

Rat liver tyrosine aminotransferase was tested as a substrate of legumain because of the possibility that legumain might have been identical with an enzyme described previously as cathepsin T. Several of the properties described for this enzyme [21] resemble properties of legumain, and Peptide 18 from the susceptible N-terminal region of the aminotransferase was cleaved (Table 3). However, a clear characteristic of cathepsin T is its limited proteolysis of tyrosine aminotransferase [22]. We were not able to detect this activity with legumain, so we conclude that the two enzymes are different.

The interaction of legumain with human α_2-macroglobulin and rat α_1-macroglobulin was studied because the α-macroglobulins are proteins that have evolved as extremely efficient agents for the removal of endopeptidases from plasma and other extracellular body fluids. The bait regions of the macroglobulins are exquisitely sensitive to proteolysis, and thus allow the proteins to recognize the very great majority of endopeptidases regardless of their catalytic type, and almost regardless of their specificity [23]. The rat α_1-macroglobulin reacted with legumain in the normal way, being cleaved in the bait region at the single asparaginyl bond, but human α_2-macroglobulin was unaffected. This was attributable to the lack of an asparaginyl bond in the bait region of the human protein, and serves as an indication of the strict specificity of legumain.

One of the questions that we sought to answer with the synthetic oligopeptides was whether the selectivity for scissile bonds in the proteins had been dictated only by the fold of the proteins or possibly by some aspect of primary structure that we had failed to identify. Accordingly, ten oligopeptides were obtained that contained asparaginyl bonds that were not cleaved in the tetanus toxoid C fragment, and the effectiveness of these as substrates for legumain was compared with that of an oligopeptide based on one of the sites that are cleaved. We found that eight of the ten peptides were cleaved, and seven of them were cleaved more efficiently than the control peptide representing the site cleaved in the protein. This showed clearly the great importance of the protein fold.

The highest catalytic constants determined for legumain in the present study [e.g. with Z-Ala-Ala-Asn-NHMec and Z-Tyr-(tertiary butyl)-Ala-Asn-NHMec] were in the order of 90×10^4 $M^{-1} \cdot s^{-1}$. These may be compared with values of about 100×10^4 $M^{-1} \cdot s^{-1}$ for cathepsin B and 1000×10^4 $M^{-1} \cdot s^{-1}$ for cathepsin L with good substrates [24]. This may indicate that the active site of legumain is similar in catalytic efficiency to those of the C1 family enzymes, although quite different in structure [24a].

It is appropriate to consider whether the information we now have about the specificity of a mammalian form of legumain leads to insights into the possible biological functions of the enzyme. We found that the enzyme is a strict asparaginyl endopeptidase that is highly selective for a small minority of the asparaginyl bonds in proteins. As such, it can scarcely contribute much to the gross catabolism of proteins mediated by the many potent non-specific endopeptidases present in the lysosomal system as a whole, but could perform limited proteolysis in processing functions if segregated from the non-specific enzymes. The structure and properties of legumain have been markedly conserved in evolution since the divergence of plants and animals, and the enzyme still occurs in equivalent intracellular locations: in the vacuolar system in plants and in the lysosomal system in mammals. We should, therefore, consider the possibility that the functions of the enzyme in mammals are related to those in plants. Early work on both the plant form of legumain in seeds and on the legumain of *Schistosoma* as 'haemoglobinase' emphasized possible roles in the gross degradation of proteins, but as those enzymes have been further characterized it has come to seem more likely that their importance is in the limited proteolysis of other proteins, maturing them to their biologically active forms. Both in plants [3,25,26] and in *Schistosoma* [27] there is evidence for the activation of other proteolytic enzymes by legumain, and in plants a variety of other enzymes, lectins and storage proteins also are processed to their mature forms by the limited proteolytic action of legumain (often termed vacuolar-processing enzyme) [3,9]. Is it possible that there is a subcompartment of the lysosomal system in mammalian cells in which legumain can catalyse limited proteolysis without interference from less selective proteinases? One such compartment might be the MHC class-II assembly compartment [28]. The evidence implicating legumain in the processing of tetanus toxin by B cells [10] would clearly support this, but further work on the distribution of legumain between subcompartments of the lysosomal/endosomal system is needed.

Financial support was received from the Medical Research Council. We thank Dr. G. Christie, Dr. J. Hargrove, Dr. H. Jones, Dr. J. Stevens and Dr. C. Watts for providing materials (see the Experimental section for details).

REFERENCES

1 Barrett, A. J. and Rawlings, N. D. (1996) Perspect. Drug Discov. Design **6**, 1–11
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., Kinoshita, T., Hiraiwa, N. and Nishimura, M. (1998) J. Plant Physiol. **152**, 668–674
4 Ishii, S., Abe, Y., Matsushita, H. and Kato, I. (1990) J. Protein Chem. **9**, 294–295
5 Dalton, J. P. and Brindley, P. J. (1998) in Handbook of Proteolytic Enzymes (Barrett, A. J., Rawlings, N. D. and Woessner, J. F., eds.), pp. 749–754, Academic Press, London
6 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
7 Chen, J.-M., Dando, P. M., Stevens, R. A. E., Fortunato, M. and Barrett, A. J. (1998) Biochem. J. **335**, 111–117
8 Ishii, S. (1994) Methods Enzymol. **244**, 605–615
9 Sheldon, P. S., Keen, J. N. and Bowles, D. J. (1996) Biochem. J. **320**, 865–870
10 Manoury, B., Hewitt, E. W., Morrice, N., Dando, P. M., Barrett, A. J. and Watts, C. (1998) Nature (London), **396**, 695–699
11 Anastasi, A., Knight, C. G. and Barrett, A. J. (1993) Biochem. J. **290**, 601–607

12 Knight, C. G. (1995) Methods Enzymol. **248**, 18–34
13 Bury, A. F. (1981) J. Chromatogr. **213**, 491–500
14 Barrett, A. J., Brown, M. A. and Sayers, C. A. (1979) Biochem. J. **181**, 401–418
15 Lonberg-Holm, K., Reed, D. L., Roberts, R. C., Hebert, R. R., Hillman, M. C. and Kutney, R. M. (1987) J. Biol. Chem. **262**, 438–445
16 Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (1992) Enzyme Nomenclature 1992, pp. 371–372, Academic Press, Orlando
17 Genetics Computer Group (1994) Program Manual for the Wisconsin Package, Version 8, University of Madison, Wisconsin
18 Barrett, A. J. and Starkey, P. M. (1973) Biochem. J. **133**, 709–724
19 Hiraiwa, N., Nishimura, M. and Hara-Nishimura, I. (1997) Plant J. **12**, 819–829
20 Cornel, F. A. and Plaxton, W. C. (1994) Physiol. Plant. **91**, 599–604
21 Gohda, E. and Pitot, H. C. (1998) in Handbook of Proteolytic Enzymes (Barrett, A. J., Rawlings, N. D. and Woessner, J. F., eds.), pp. 774–775, Academic Press, London
22 Hargrove, J. L., Gohda, E., Pitot, H. C. and Granner, D. K. (1982) Biochemistry **21**, 283–289
23 Sottrup-Jensen, L., Sand, O., Kristensen, L. and Fey, G. H. (1989) J. Biol. Chem. **264**, 15781–15789
24 Kirschke, H., Barrett, A. J. and Rawlings, N. D. (1998) Lysosomal Cysteine Proteinases, Oxford University Press, Oxford
24a Chen, J.-M., Rawlings, N. D., Stevens, R. A. E. and Barrett, A. J. (1998) FEBS Lett. **441**, 361–365
25 Hiraiwa, N., Kondo, M., Nishimura, M. and Hara-Nishimura, I. (1997) Eur. J. Biochem. **246**, 133–141
26 Okamoto, T. and Minamikawa, T. (1995) Eur. J. Biochem. **231**, 300–305
27 Dalton, J. P. and Brindley, P. J. (1996) Parasitol. Today **12**, 125–125
28 Glickman, J. N., Morton, P. A., Slot, J. W., Kornfeld, S. and Geuze, H. J. (1996) J. Cell Biol. **132**, 769–785
29 Bairoch, A. and Apweiler, R. (1998) Nucleic Acids Res. **26**, 38–42
30 Barker, W. C., Garavelli, J. S., Haft, D. H., Hunt, L. T., Marzec, C. R., Orcutt, B. C., Srinivasarao, G. Y., Yeh, L. S. L., Ledley, R. S., Mewes, H. W. et al. (1998) Nucleic Acids Res. **26**, 27–32

Received 18 November 1998/4 January 1999; accepted 27 January 1999