FEBS Letters 438 (1998) 114–118 FEBS 21039

Autocatalytic activation of human legumain at aspartic acid residues

Sherin Halfon, Sejal Patel, Felix Vega, Sandra Zurawski, Gerard Zurawski*

DNAX Research Institute, 901 California Avenue, Palo Alto, CA 94304-1104, USA

Received 15 September 1998

Abstract Human legumain was characterized following overexpression in a murine cell line as the C-terminal Ig-fusion protein. Upon acid treatment, the prolegumain autoproteolyzed distal to two aspartic acid residues to yield a highly active form. The ability of mature legumain to cleave after aspartic acid residues was confirmed with a small peptide substrate. Substitution of alanine for the putative catalytic cysteine, or for either of two strictly conserved histidine residues, partly or wholly eliminated autoactivation but not the ability of wild-type legumain to correctly process the variants to the properly sized proteins.

Key words: Cysteine protease; Legumain; Hematopoietic cell

1. Introduction

Asparagine-specific endoproteases (EC 3.4.22.34), a subfamily of cysteine proteases with no homology to other known proteases, are found in organisms as diverse as parasites, plants, and mammals [1–6]. They have been implicated in processes such as degradation and processing of stored plant vacuole proteins and blood fluke hemoglobin digestion. For these reasons, they have commonly been called legumains or hemoglobinases. We use the term legumain for the entire family to differentiate the schistosome homolog from digestive parasite proteases of the cathepsin family. Legumains are synthesized in prepro forms and are localized to intracellular acidic vacuoles (plants) or to the gut (schistosomes). However, little is known about the biochemistry of this family of proteases. For instance, although N-terminal sequencing and the size of the isolated native proteins indicate that maturation involves cleavage of both N- and C-terminal regions, the nature of this processing into active, mature forms is not well understood. Furthermore, the low sequence similarity of these proteins to other protease families has hampered the identification of the critical catalytic residues.

Here, we report the mouse preprolegumain sequence, and describe the production of recombinant human prolegumain – the first efficient production of a member of this protease family. In addition, we examine its autocatalytic maturation. This process reveals a previously unseen ability of this protease to cleave at aspartic acid as well as at asparagine residues.

2. Materials and methods

2.1. Protein expression and purification

Preprolegumain cDNA was identified as an EST from a human dendritic cell cDNA library and isolated as the full-length clone. PCR primers were designed to amplify the full preprolegumain coding region from the cDNA incorporating a *Hin*dIII site in the 5′ untranslated region, and to add residues encoding PRPPTGN (an IgA protease cleavage site) followed by an *Xho*I site distal to Tyr^{433}. The PCR product was inserted into a modified form of pCDM8 (Invitrogen) where the pCDM8 *Xho*I-*Not*I region was replaced with a fragment encoding human immunoglobulin G1, Fc (residues 74–768 of accession number X70421) preceded by 5′-CTCGAGC-3′ and flanked by 5′-GCGGCCGC-3′. The resulting expression plasmid was electroporated into COP5 cells [7]. Cells from five electroporations were used to seed 10-tray Cell Factories (Nunc). After 5–7 days in culture, the cells had reached confluence and the supernatant was harvested by filtration. The 94 kDa IgG-fusion protein was purified by HiTrapA chromatography (Pharmacia), and eluted with 100 mM glycine-HCl, pH 3.0. Typical yields were 1 mg/l. Following storage at 4°C for 48 h in pH 3 buffer, the autoactivated protein was dialyzed into 100 mM NaOAc, 100 mM NaCl, pH 5.0 and resubjected to HiTrapA chromatography, with collection of the flowthrough fractions.

Prolegumain was prepared from the Ig-fusion protein after the initial HiTrapA column by treatment with a 1/50 (by weight) ratio of IgA protease (Boehringer Mannheim) in PBS, 1 mM EDTA overnight at 37°C. Purification was effected by negative HiTrapA chromatography, with collection of the flowthrough.

2.2. Preparation of $His^{47}Ala$, $Cys^{52}Ala$, and $His^{150}Ala$ variants

Mutagenic N-terminal and C-terminal PCR reactions were carried out on the cDNA sequence using the 5′- and 3′-terminal primers in conjunction with primers carrying the mutation. The two resulting reaction products (for each variant) were used as template for synthesis of the full-length product. Following confirmation of the mutation by DNA sequencing the variant proteins were produced as for wild-type.

2.3. Antibodies

Mouse monoclonal antibodies (mAb), which specifically recognized human legumain in Western blots, were generated by standard protocols from the Ig-fusion protein [8] using BALB/c mice (Simonsen). Goat anti-human-IgG horseradish peroxidase was used to visualize IgG bands.

2.4. Kinetic analysis

Fluorimetric determinations (SPEX Fluorolog) were calibrated using varying concentrations of AMC, for which a molar absorptivity of 18 000 at 354 nm was used. Concentrations of Z-AAN-AMC (1–100 μM, Enzyme Systems) were fluorimetrically standardized by complete legumain cleavage to Z-AAN and AMC. Reactions were carried out with 2.7 nM legumain in 100 mM sodium citrate, 50 mM NaCl, 10 mM DTT, 1% DMF, pH 6.0 at 25°C. Kinetic constants were determined with the Eadie-Hofstee plot using Kaleidagraph software (Synergy Software).

Kinetic analysis of succ-YVAD-AMC (Bachem, 100 μM) was carried out with 270 nM enzyme in 100 mM glycine-HCl, 10 mM DTT, 1% DMF, pH 3.0 at 25°C, and compared with Z-AAN-AMC under the same conditions.

3. Results

3.1. Overexpression of legumain

Human prolegumain fused to IgG was overexpressed in transiently transfected mouse COP5 cells utilizing the native secretion signal. A time course study showed that this material first appeared in the media 3 days post transfection, and that

*Corresponding author. Fax: (1) (650) 496-1214.
E-mail: gerard@dnax.org

PII: S0014-5793(98)01281-2

expression levels reached a plateau after approximately 5 days (data not shown). No changes in the level of prolegumain-IgG were seen with increased expression times, even though most of the COP5 cells had died by the end of 7 days, indicating that lysis of the cells led neither to an increase of legumain concentration from release of intracellular material (indeed, protein blots of cell pellets showed no evidence of intracellular material), nor to a decrease due to degradation. Furthermore,

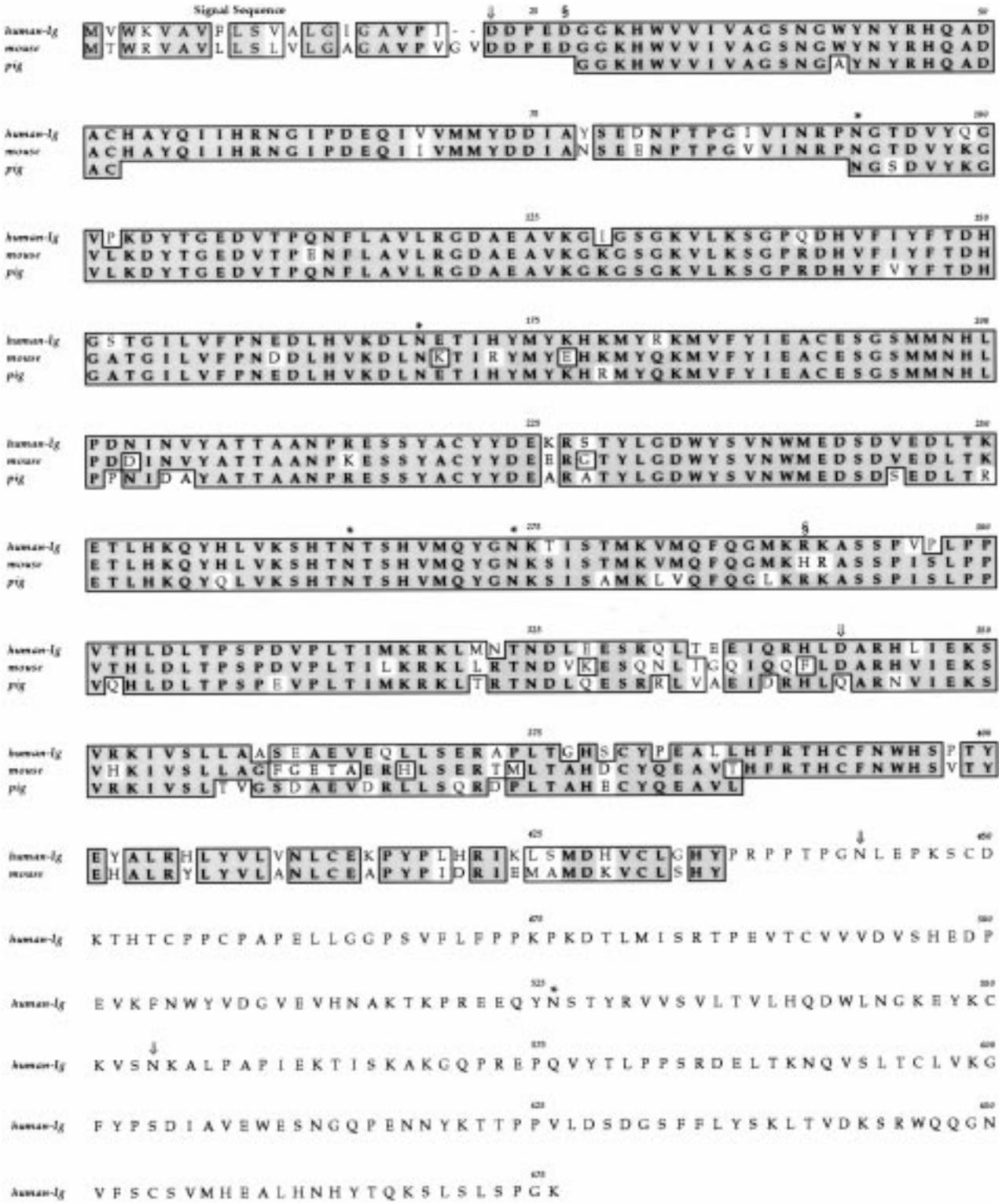

Fig. 1. Alignment of mammalian legumains. The human sequence shows the entire IgG-fusion used for overexpression, where amino acids 1–435 and 436–676 represent the preprolegumain and IgG sequences respectively. The mouse sequence is that determined as the consensus sequence of ESTs, as described in the text. The partial pig sequence is from [6]. Identical residues are boxed and shaded, while similar residues are boxed. The signal sequence comprises residues 1–17. ⇓ indicates observed autoactivation sites, § represents the predicted maturation sites, and * shows possible *N*-glycosylated Asn residues.

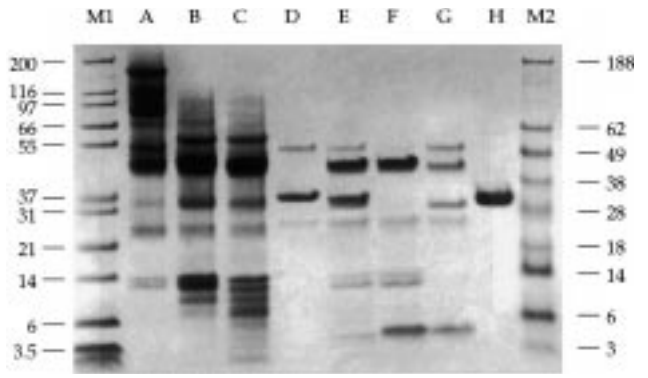

Fig. 2. Autoactivation and purification of legumain. Coomassie-stained SDS-PAGE. Lane A: Legumain-Ig fusion protein, eluted from a HiTrapA column; B: A, after storage at pH 3, 4°C for 3 days; C: A, after storage at pH 3, 4°C for 6 days. D: Purified IgG. E: Second HiTrapA column load. F: HiTrapA flowthrough. G: 2.5×Column Wash. H: 2.5×Eluate (Ig). M1 and M2 correspond to Mark12 and SeeBlue (Qiagen) molecular weight markers respectively.

no anti-legumain or anti-IgG immunoreactive bands at lower molecular weight were observed on protein blots. The prolegumain-IgG was purified from the supernatant by protein A affinity chromatography, eluting with pH 3 buffer. Initially, we rapidly neutralized the eluted fractions and cleaved the secreted fusion protein with IgA protease at an engineered site lying in the linker between the legumain and the IgG moieties. The resulting 56 kDa product was further purified by negative protein A absorption to yield prolegumain. Treatment of this material with *N*-glycosidase F resulted in a protein of approximately 40 kDa by Western blot (data not shown), consistent with the size expected from the cDNA sequence, which indicates that the difference in size was due to *N*-glycosylation at some or all of the four potential sites. N-terminal amino acid sequence analysis confirmed that this secreted product began at residue 18 of the open reading frame (ORF), as predicted by PSORT [9] and Chen et al. [6] (Fig. 1).

By analogy with other family members, maturation of human legumain is expected to require removal of both an ~11–14 residue N-terminal and a ~145 residue C-terminal prodomain. Therefore, treatment of both the IgG-fusion and the Igase-cleaved material with a number of proteases was explored in order to generate the active protein. Proteolysis of the 94 kDa fusion protein with trypsin conjugated to agarose beads at pH 8 resulted in a 34 kDa species that reacted with the anti-legumain antibody with little or no further degradation, consistent with the size expected for the mature, glycosylated protein. Nevertheless, no proteolytic activity toward any substrate could be detected with this preparation, consistent with the reported instability of porcine legumain to neutral or basic pH [6]. However, when the trypsin activation was carried out at pH 5.5, the resultant 34 kDa protein had limited proteolytic activity towards a large number of protein substrates following extended incubations (data not shown).

By removal of the neutralization step following chromatography, highly active human legumain could be prepared from the prolegumain-IgG fusion protein: storage in pH 3 buffer at 4°C for 2 days resulted in autoactivation (see below). The mature protein was isolated by dialysis into pH 5 sodium acetate buffer followed by negative protein A affinity chromatography. Both SDS-PAGE (Fig. 2) and gel filtration chromatography (data not shown) indicated a molecular weight of approximately 48 kDa.

3.2. Autoactivation of legumain

Affinity-purified prolegumain-IgG stored at pH 3 undergoes autocatalytic activation. Autoactivation could not be effected by incubating unpurified prolegumain-IgG (i.e. the cell supernatant) at pH 3, indicating that affinity purification was necessary for proteolysis to occur. No efforts were made to optimize the activation conditions. Activation was followed both by changes in the appearance of the material on SDS-PAGE (Fig. 2), and by the increased catalytic activity towards Z-AAN-AMC over time (data not shown). Bands on SDS-PAGE were identified by Western blot versus both anti-IgG and anti-legumain antibodies, and confirmed with amino acid sequencing: the 94 kDa protein, which was recognized by both antibodies, was cleaved to a 48 kDa band, reactive only with anti-legumain, and a 36 kDa protein reactive only with anti-IgG. Other low molecular weight protein bands did not react with either antibody. Amino acid sequencing was used to identify all the cleavage sites (Table 1), and revealed that although the IgG chain was cleaved at at least two asparagine residues, cleavage within the legumain chain occurred only at aspartic acid residues.

3.3. Kinetic analysis

All members of the legumain protease family which have been characterized have been found to be active only towards N-terminally protected non-glycosylated asparagine-containing substrates. The recombinant human protein has maximal activity at pH 6.1 (which was dependent on the buffer used), with the kinetic constants k_{cat}, 4 s^{-1}, K_m, 20 μM, k_{cat}/K_m, 200 000 s^{-1} M^{-1}, similar to the activities for other family members [6,10]. We could detect no activity toward a glutamine-containing substrate (Z-GAQ-AMC), confirming the specificity of legumain towards asparagine, nor did we detect any effect of trypsin treatment on the size of the active material or on its activity. The contrasting ability of trypsin to generate a 34 kDa legumain product from prolegumain-IgG versus its inability to cleave the 48 kDa autocatalytically proc-

Table 1
Autocleavage sites of prolegumain-Ig

Molecular weight (kDa)	Sequence	Identity[a]
55	P-X	Unknown[b]
50[c]	DPEDGGXH	Legumain, aa 23–
36	LEPXSXDTH	IgG, aa 442–
17	WEXET	Unknown[b]
15	KALPAPXE	IgG, aa 555–
14	KALPAPXE	IgG, aa 555–
10	ARHLI	Legumain, aa 343–
5–6	L/W-P/V-ETI	Unknown
3	XENDEQD	Unknown

Sequences are of bands shown in lane E, Fig. 2. Indicated sequences are as determined by N-terminal sequencing of PVDF-immobilized bands. The identities show the locations of the N-terminus of each sequence. Those that are listed as unknown could not be matched.
[a]Amino acids refer to numbering of Fig. 1.
[b]These bands are visible in recombinant expression of Ig alone, and probably represent medium contaminants.
[c]This sequence was determined from two different preparations, with identical results.

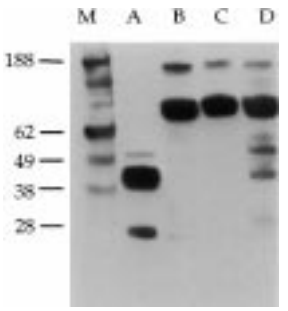

Fig. 3. Legumain mutant proteins. Western blot of approximately 500 ng legumain. The predominant band in the mutant lanes is the Ig-fusion following storage at pH 3, 4°C for periods of time of at least a month, whereas the major band for wild-type is the mature protein. Lane M: molecular weight markers. A: Wild-type. B: His^{47}Ala. C: Cys^{52}Ala. D: His^{150}Ala.

essed form likely indicates a conformational change in legumain upon maturation.

Consistent with the products observed during autoactivation, we found a low, but detectable, activity at pH 3 for an aspartic acid-containing substrate: 100 μM succ-YVAD-AMC was cleaved at a rate of 8 nmol/s by 270 nM enzyme while an equimolar amount of Z-AAN-AMC was cleaved at a rate of 167 nmol/s under the same conditions. Although aspartic acid cleavage is quite inefficient, it is more than sufficient to account for the observed autoactivation. Activity toward succ-YVAD-AMC was absent at the typical pH 6 assay conditions, nor was it observed with the cell supernatant, or by purified, activated protease exposed to neutral pH. Moreover, the activity was fully inhibited by chicken egg white cystatin (data not shown). These observations are consistent with the activity being entirely due to legumain.

3.4. *His^{47}Ala, Cys^{52}Ala, and His^{150}Ala variants*

Since the legumains display little similarity to other known protease families, the active site residues have not been determined. From their patterns of inhibition [2,6], it is thought that legumains are cysteine proteases. As Cys^{52} is the only completely conserved cysteine within the mature sequence, it is likely that this residue is the reactive nucleophile. In addition, cysteine proteases commonly contain a catalytically important histidine. Four histidine residues are completely conserved within the legumain family and are therefore potential catalytic residues: His^{47}, His^{53}, His^{150}, and His^{268}. We have produced variant legumains with alanine substituted at positions 47 and 150 in order to test the roles of these residues, and have also produced the Cys^{52}Ala variant (Fig. 3). N-terminal sequencing of all three variant fusion proteins gave the sequence VPIDD indicating that the signal sequence of the mutant proteins was correctly processed. Incubation of these variants at pH 3 for extended periods of time resulted in little or no change in size as determined by SDS-PAGE (Fig. 3), consistent with reduced autocatalytic processing. Furthermore, although processing could be effected for all three variants to the properly sized material by the addition of a small amount of wild-type protein (data not shown), no activity of the mutants toward Z-AAN-AMC was observed.

3.5. *Determination of mouse preprolegumain cDNA sequence*

BLASTN searches of the dbEST database using the human cDNA sequence as a probe revealed very high similarity to overlapping mouse ESTs with the following accession numbers: AA000373, AA000961, AA107459, AA122993, AA615503, AA688513, W11820, W14837, W35784, W35821, W36795, W83121. A consensus compilation of these yielded a putative full-length mouse cDNA sequence, with a corresponding 83% amino acid identity to the human protein (Fig. 1). Nucleotides following position 322 were confirmed by sequencing IMAGE clone number 578262, corresponding to the EST with accession number AA122993. This sequence has been submitted to the GenBank database (accession number AF044266).

4. Discussion

In this paper, we describe the production of milligram quantities of active human legumain in eukaryotic cell culture. Since previous efforts to produce recombinant legumain family members have met with limited success, this represents a significant advance for this family of proteases. The activity of the mature recombinant human legumain towards the tripeptide substrate Z-AAN-AMC was consistent with that of other native family members. However, at low pH, activity towards aspartic acid residues was also detected. As the pK_a of an isolated aspartic acid is approximately 4, we would expect these residues to be largely protonated at pH 3, and therefore functionally similar to asparagine. This similarity in structure is likely responsible for the additional activity observed at low pH.

We observed autoactivation of prolegumain-IgG following affinity purification and acidification. Several lines of evidence point to these cleavage reactions being autocatalytic. First, we see no processing with unpurified product, indicating that impurities from the cell culture medium is unlikely to be responsible for this reaction. This observation further suggests that an inhibitor must be removed preceding maturation and/or that the reaction is intermolecular and only occurs when the protein is concentrated on the affinity resin. Second, mutation of putative active site residues prevents autoprocessing (partially or fully depending on the mutant), but still allows processing by wild-type protein. Third, the sites of cleavage observed are consistent with the specificity of legumain for cleavage of peptide bonds following asparagine and aspartic acid residues. This mechanism may be responsible for the observation that acidification of native legumain during purification increased its activity 65% [10]. Autocatalytic activation following acidification of many proteases with low pH optima has been reported [11–16]. Acid-pH triggered processing may be a common mode for activation of lysosomal proteases, by which they mature only when they reach the vesicles in which they are to act, preventing inappropriate processing within the cell. Interestingly, protonation of an aspartic acid residue within a conserved motif is postulated to be a switch in the autoactivation of papain by triggering a conformational change in the proenzyme [17]. A similar motif is seen in many of the cathepsins, indicating that they may also be regulated in this manner. Furthermore, many of these proenzymes have been shown to be stable at neutral pH ([18], for example), but are deactivated rapidly when the mature enzyme is exposed to neutral or basic pHs. This behavior is also seen for legumain.

The cleavage sites we observe are not the same as those predicted to be the in vivo processing sites by alignment with related legumains [6]. Nevertheless, the autoactivated

material is nearly as active as the native porcine protein. This may indicate that other proteases are partly or wholly responsible for maturation of legumain in vivo, possibly by further processing the autocatalytically cleaved material.

Legumain does not cleave any asparagine residues within its own sequence, although it does cleave within the IgG portion of the fusion protein: while there are several Asn residues within legumain that are not part of glycosylation consensus sequences, these appear to be unavailable to the enzyme active site. The inability to cleave Asn residues within the legumain sequence likely has a role in stabilizing the protease in vivo. Non-reactivity of some Asn residues has been observed previously for other members of this family [6,19]. Additional factors which determine why some asparagine residues and not others are recognized are not currently understood since no consensus sequence is apparent.

The broad distribution, acidic catalytic optimum, and instability of the mature protein to neutral pH is consistent with legumain being localized to acidic intracellular compartments such as lysosomes or endosomes. Targeting of soluble proteins to these compartments is normally mediated by the mannose-6-phosphate receptor following phosphorylation of carbohydrate chains. The fact that we could observe no intracellular legumain in COP5 cells may be due to a number of factors. First, the lysosomal sorting mechanism may simply be overwhelmed by the overexpressed protein, or the phosphotransferase may be sterically blocked by the IgG moiety. In these cases, the default pathway would be the secretory one. Second, some transformed mouse cells have previously been shown to have an altered lysosomal trafficking pattern ([20], and references therein), and COP5 cells may well have this defect. Third, it is possible that legumain is normally secreted. This would not rule out a lysosomal function if the proprotease is then endocytosed or pinocytosed.

In conclusion, we have produced active recombinant human legumain. This system was used to study the autocatalytic process and will be useful for more detailed investigation of the structure-function properties of legumain. Although various studies have shown that this cysteine endoprotease preferentially cleaves following asparagine residues, it was found to autocatalytically cleave only after aspartic residues, resulting in a highly active protease with the expected asparaginase specificity.

Acknowledgements: We thank Allison Helms, Connie Huffine, and Anh Quan for DNA sequencing, Vicki Saylor and John Abrams for antibody production, Debbie Liggett for oligonucleotide synthesis, and Fernando Rock and Jessica Foster for assisting with chromatography. We thank the scientists at the Laboratory of Immunological Research, Dardilly, France for their contributions in the initial studies examining dendritic cell cDNAs, from which we initially identified the human legumain cDNA. DNAX is supported by Schering-Plough Corporation.

References

[1] Rawlings, N.D. and Barrett, A.J. (1994) Methods Enzymol. 244, 461–486.
[2] Ishii, S. (1994) Methods Enzymol. 244, 604–615.
[3] Bairoch, A. (1998) http://expasy.hcuge.ch/cgi-bin/lists?peptidas.txt.
[4] Rawlings, N.D. and Barrett, A.J. (1998) http://www.bi.bbsrc.ac.uk/Merops/Merops.htm.
[5] Tanaka, T., Inazawa, J. and Nakamura, Y. (1996) Cytogenet. Cell Genet. 74, 120–123.
[6] Chen, J.M. et al. (1997) J. Biol. Chem. 272, 8090–8098.
[7] Tyndall, C., La Mantia, G., Thacker, C.M., Favaloro, J. and Kamen, R. (1981) Nucleic Acids Res. 9, 6231–6250.
[8] Kohler, G. and Milstein, C. (1975) Nature 256, 495.
[9] Nakai, K. and Kanehisa, M. (1992) Genomics 14, 897–911.
[10] Kembhavi, A.A., Buttle, D.J., Knight, C.G. and Barrett, A.J. (1993) Arch. Biochem. Biophys. 303, 208–213.
[11] Mach, L., Mort, J.S. and Glossl, J. (1994) J. Biol. Chem. 269, 13030–13035.
[12] McQueney, M.S. et al. (1997) J. Biol. Chem. 272, 13955–13960.
[13] Menard, R., Carmona, E., Takebo, S., Dufour, E., Plouffe, C., Mason, P. and Mort, J.S. (1998) J. Biol. Chem. 273, 4478–4484.
[14] Bromme, D. et al. (1993) J. Biol. Chem. 268, 4832–4838.
[15] Kageyama, T. et al. (1992) J. Biol. Chem. 267, 16450–16459.
[16] Van Den Hazel, H., Wolff, A.M., Kielland-Brandt, M.C. and Winther, J.R. (1997) Biochem. J. 326, 339–344.
[17] Vernet, T. et al. (1995) J. Biol. Chem. 270, 10838–10846.
[18] McDonald, J.K. and Emerick, J.M. (1995) Arch. Biochem. Biophys. 323, 409–422.
[19] Abe, Y., Shirane, K., Yokosawa, H., Matsushita, H., Mitta, M., Kato, I. and Ishii, S. (1993) J. Biol. Chem. 268, 3525–3529.
[20] Isidoro, C., Demoz, M., De Stefanis, D., Baccino, F.M., Hasilik, A. and Bonelli, G. (1997) Int. J. Cancer 70, 310–314.