Cloning and expression of mouse legumain, a lysosomal endopeptidase

Jinq-May CHEN¹, Pam M. DANDO, Richard A. E. STEVENS, Mara FORTUNATO and Alan J. BARRETT MRC Peptidase Laboratory, The Babraham Institute, Babraham, Cambridge CB2 4AT, U.K.

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

INTRODUCTION

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

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

MATERIALS AND METHODS

Materials

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

Enzyme assays

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

Mouse tissues and subcellular fractionation of kidney

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

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

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

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

cathepsin B activities, and for determination of protein concentration.

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

Production of anti-legumain antibodies

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

Primary culture of rat kidney cells and immunolabelling

A kidney was removed from a 12-week old rat and the cortical tissue was dissected and minced under sterile conditions. The minced tissue was stirred at 30 °C in a solution containing 0.05 % trypsin and 0.02% EDTA to dissociate cells. The solution became turbid after about 20 min. Tissue aggregates were removed by sedimentation and discarded, and dissociated cells were collected by centrifugation at 400 g for 5 min. The trypsinized cells were plated on cover slips coated with rat-tail collagen to facilitate their attachment [14]. The cells were cultured in sodium pyruvate-free, high-glucose Dulbecco's modified Eagle's medium supplemented with 10% (v/v) fetal bovine serum. At 4 days after plating, cells were fixed in 3 % (w/v) paraformaldehyde in PBS (137 mM NaCl, 2.7 mM KCl, 4.3 mM Na₂HPO₄, 1.47 mM KH₂PO₄, pH 7.1) and permeabilized in 0.4 % Triton X-100 in PBS. Cells were incubated with the affinity-purified sheep anti-legumain antibody, followed by FITC-conjugated anti-(sheep immunoglobin) secondary antibody (Sigma). After immunostaining, cells were examined using a Zeiss Axiophot microscope equipped with an epifluorescence light source and a filter set for fluorescein, and photographed using a Zeiss camera system.

SDS/PAGE and immunoblotting

Electrophoresis was in the buffer system of Laemmli [15], and transfer of proteins to nitrocellulose was as described by Towbin et al. [16]. The blot was first incubated in blocking solution containing 5% horse serum in PBS for 2 h, and then probed with the affinity-purified sheep anti-legumain antibody ($10 \mu g/ml$ IgG). The incubation with primary antibody was at 20 °C for 2 h, followed by a PBS wash. Alkaline-phosphatase-conjugated secondary antibody (Sigma) was applied, and the colour reaction was developed using 0.005% 5-bromo-4-chloro-3-indolyl phosphate as the substrate in 100 mM Tris/HCl, pH 9.3, containing 0.005% Nitro Blue Tetrazolium, 50 mM MgCl₂ and 100 mM NaCl [17].

Database search and sequence analysis of mouse legumain clones

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

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

Expression of recombinant mouse legumain

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

medium was removed after 16 h and cells were subjected to shock with 15 % glycerol in Hepes-buffered saline (25 mM Hepes, 140 mM NaCl, 1.5 mM Na₂HPO₄, pH 7.1) for 1 min. The transfected cells were then rinsed once with PBS, and cultured in minimal Eagle's medium supplemented with 10 % (v/v) heatinactivated fetal bovine serum until harvest.

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

Cells were harvested and disrupted with the lysis buffer (0.1 M sodium citrate, pH 5.8, containing 1 mM EDTA and 1% n-octyl β -D-glucopyranoside). After three cycles of freezing (in solid CO₂) and thawing (30 °C), the cell lysate was centrifuged at 20000 g for 5 min. The supernatant was collected and stored at 4 °C. To purify the recombinant mouse legumain, the cell lysate from pCMVmsleg-transfected HEK 293 cells (ten 100 mm dishes) was further centrifuged at 100000 g for 1 h. The supernatant was diluted to 10 mM sodium citrate. The sample was then applied to a Mono S column (Pharmacia HR 5/5) equilibrated with 50 mM sodium citrate buffer, pH 5.5, containing 1 mM EDTA and 0.1% CHAPS. A step gradient of NaCl (0.15 M, 0.4 M and 1 M) in the same buffer was applied to wash the column and to elute bound proteins. The active recombinant mouse legumain was eluted from the column at 0.4 M NaCl.

HPLC analysis

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

RESULTS

Distribution of legumain and its subcellular localization in mouse tissues

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

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

To determine the subcellular localization of legumain within tissues, the procedure of Maunsbach [12] was used to prepare subcellular fractions from rat kidney, including brush border, mitochondria, lysosomes, microsomes and cytosol. Samples (6 μ g of total protein) from these fractions were analysed by immunoblotting with the anti-legumain antibody. The immunoreactive



Figure 1 Distribution of legumain in mouse tissues

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



Figure 2 Immunoblot analysis of rat kidney subcellular fractions

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

band of 35 kDa, presumably mature legumain, was found to be most intense in the lysosomal fraction (Figure 2, lane 5). Similar results were obtained with mouse kidney (not shown).

When the primary culture of rat kidney cells from the renal cortex was labelled with the anti-legumain antibody, the immuno114



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

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



Figure 4 Dependence of legumain activity on disruption of organelles

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



Figure 5 Comparative distributions of legumain and cathepsin B

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

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

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

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

DNA sequencing and structural analysis of mouse legumain

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

Expression of recombinant legumain

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

Biochemical characterization of recombinant mouse legumain

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

GCT CGC	CTG. CCA	AGT GTC	CTG TTC	CGC TGT	GAC AGC	GCC(GGA(CGG. CAC	AAT GGG	TCC GTG	CACO	GT:	ICT(GAC(GCA CTG	GTC. GAG.	ACC AGT	GCG GGC	GCG. TGT	ATC GCT	ACC TCT	60 120
											M	Т	W	R	v	А	V	L	L	9
CAG	CCT	GGT	GCT	GGG	TGC	TGG	TGC	CGT	TCC	CGT	CGG	FGT	GGA	CGA	TCC	CGA	GGA	TGG	AGG	180
S	L	V	L	G	A	G	Α	v	Ρ	v	G	v	D	D	Ρ	Е	D	G	G	29
CAA	GCA	CTG	GGT	GGT	GAT	TGT	GGC	GGG	CTC	CAA'	rgg(CTG	GTA'	TAA	TTA	CCG.	ACA	CCA	GGC	240
К	Н	W	v	v	I	v	А	G	s	N	G	W	Y	Ν	Y	R	н	Q	A	49
AGA	CGC.	ATG	CCA	CGC	CTA	CCA	GAT	CAT	CCA	CCG	GAAG	CGG	GAT	TCC	TGA	CGA	GCA	GAT	CAT	300
D	А	С	н	А	Y	Q	Ι	Ι	н	R	N	G	Ι	Ρ	D	Е	Q	I	I	69
AGT	GAT	GAT	GTA	TGA	CGA	CAT	TGC	CAA	CTC	TGA	AGA	AAA	ccc	TAC	CCC	AGG	TGT	TGT	GAT	360
v	М	М	Y	D	D	Ι	A	Ν	S	Е	Е	N	Ρ	т	Ρ	G	v	v	I	89
CAA	CCG.	ACC	TAA	CGG	CAC.	AGA'	TGT.	ATA	CAA	GGG	AGT	CCT	GAA	GGA	CTA	CAC	CGG	AGA	GGA	420
Ν	R	Ρ	N	G	T	D	v	Y	ĸ	G	v	L	ĸ	D	Y	т	G	E	D	109
TGT	GAC	TCC	AGA	GAA	TTT	CCT	CGC	CGT	GCT	GAG	AGG	TGA	CGC	AGA	AGC	TGT	GAA	GGG	CAA	480
v	т	Ρ	Е	Ν	F	L	А	v	L	R	G	D	А	Е	А	v	к	G	ĸ	129
AGG	GTC	TGG	AAA	AGT	CTT	GAA	GAG	TGG	CCC	CCG	AGA	TCA	TGT	CTT	CAT	TTA	CTT	CAC	CGA	540
G	s	G	К	v	L	К	s	G	Ρ	R	D	н	v	F	I	Y	F	т	D	149
CCA	CGG	AGC	CAC	CGG	GAT	CCT	GGT	GTT	TCC	TAA	TGA	TGA	TCT	TCA	TGT	CAA	GGA	CCT	GAA	600
н	G	А	т	G	I	L	v	F	Ρ	N	D	D	L	н	v	К	D	L	Ν	169
																			autor (
таа	GAC	тат	TCG	CTA	CAT	GTA	TGA	ACA	CAA	AAT	GTA	CCA	GAA	GAT	GGT	GTT	CTA	CAT	TGA	660
к	т	I	R	Y	М	Y	Е	н	к	М	Y	Q	к	М	v	F	Y	I	Е	189
2272.000																				
AGC	TTG	TGA	GTC	TGG	CTC	CAT	GAT	GAA	CCA	CCT	GCC	CGA	CGA	CAT	CAA	CGT	TTA	TGC	AAC	720
A	С	Е	s	G	s	М	м	N	н	L	Ρ	D	D	I	Ν	v	Y	А	т	209
TAC	TGC	GGC	CAA	ccc	CAA	GGA	GTC	ATC	TTA	TGC	CTG	CTA	CTA	CGA	CGA	.GGA	GAG	GGG	CAC	780
т	A	A	N	P	к	Е	s	s	Y	А	С	Y	Y	D	Е	Е	R	G	т	229
-																				
тта	ССТ	GGG	TGA	CTG	GTA	CAG	CGT	CAA	CTG	GAT	GGA	AGA	CTC	CGA	TGT	GGA	GGA	CCT	'GAC	840
v	T.	G	D	W	Y	S	v	N	W	М	Е	D	s	D	v	Е	D	L	т	249
-	~	~			-	-														
CDD	AGA	GAC	ССТ	ጥሮል	CAA	GCA	GTA	CCA	CCT	GGT	CAA	GTC	CCA	CAC	CAA	CAC	CAG	CCA	TGT	900
CAA	AGA F	GAC T	CCT	TCA H	CAA K	GCA O	GTA V	.CCA H	CCT	GGT V	CAA K	GTC S	CCA H	CAC T	CAA N	.CAC T	CAG S	CCA H	TGT V	900 269
CAA K	AGA E	GAC T	CCT L	TCA H	ICAA K	gca Q	GTA Y	CCA H	CCT L	GGT V	CAA K	GTC S	CCA H	CAC T	CAA N	CAC T	CAG	CCA H	TGT V	900 269
CAA K	AGA E	GAC T ATA	CCT L	TCA H	CAA K	GCA Q ATC	GTA Y TAT	CCA H	CCT L TAC	GGT V CAT	CAA K GAA	GTC S AGT	CCA H GAT	CAC T	CAA <u>N</u> GTT	CAC T TCA	CAG	H H	V V GAA	900 269 960
CAA K CAT	AGA E GCA	GAC T ATA V	CCT L TGG	TCA H GAA	ICAA K ICAA K	GCA Q ATC	GTA Y TAT T	CCA H CTC	CCT L TAC	GGT V CAT M	CAA K GAA K	GTC S AGT V	CCA H GAT M	CAC T GCA	CAA <u>N</u> .GTT	CAC T TCA	CAG S .GGG	CCA H AAT	TGT V GAA K	900 269 960 289
CAA K CAT M	AGA E GCA Q	GAC T ATA Y	CCT L TGG G	TCA H GAA	ICAA K ICAA K	gca Q ATC <u>S</u>	GTA Y TAT I	CCA H CTC S	CCT L TAC T	GGT V CAT M	CAA K GAA K	GTC S AGT V	CCA H GAT M	CAC T GCA Q	CAA <u>N</u> .GTI F	T T TCA Q	CAG S .GGG G	H H AA1 M	TGT V GAA K	900 269 960 289
CAA K CAT M	AGA E GCA Q	GAC T ATA Y	CCT L TGG G	TCA H GAA N	ICAA K ICAA K	GCA Q ATC S	GTA Y TAT I	CCA H CTC S	CCT L TAC T	GGT V CAT M	CAA K GAA K	GTC S AGT V	CCA H GAT M	CAC T GCA Q	CAA <u>N</u> .GTT F	TCAC T TCA Q	CAG S GGG G	H H AA1 M	TGT V GAA K	900 269 960 289
CAA K CAT M GCA	AGA E GCA Q .CAG	GAC T ATA Y AGC	CCT L TGG G CAG	TCA H GAA N TTC	ICAA K ICAA K	GCA Q ATC S CAT	GTA Y TAT I CTC	CCA H CTC S CCT	CCT L TAC T GCC	GGT V CAT M TCC	CAA K GAA K GGT	GTC S AGT V CAC	CCA H GAT M ACA	CAC T GCA Q .CCT	CAA <u>N</u> .GTI F	CAC T TCA Q CCI	CAG S .GGG G 'CAC	H H AA1 M CCC	TGT V CAA K CAG	900 269 960 289 1020
CAA K CAT M GCA H	AGA E GCA Q .CAG R	GAC T ATA Y AGC A	CCT L TGG G CAG S	TCA H GAA N TTC	ICAA K ICAA K CCCC P	GCA Q ATC S CAT I	GTA Y TAT I CTC S	CCA H CTC S CCT L	CCT L TAC T GCC P	GGT V CAT M TCC P	CAA K GAA K GGT V	GTC S AGT V CAC T	CCA H GAT M ACA H	CAC T GCA Q .CCT L	CAA <u>N</u> .GTI F TGA D	TCA T Q L CCT	CAG S GGG G CAC T	AAT M CCC P	TGT V GAA K CAG S	900 269 960 289 1020 309
CAA K CAT M GCA H	AGA E GCA Q .CAG R	GAC T ATA Y AGC A	CCT L TGG G CAG S	TCA H GAA N TTC S	CAA K CAA K CCCC P	GCA Q ATC S CAT I	GTA Y TAT I CTC S	CCA H CTC S CCT L	CCT L TAC T GCC P	GGT V CAT M TCC P	CAA K GAA K GGT V	GTC S AGT V CAC T	CCA H GAT M ACA H	CAC T GCA Q .CCT L	CAA <u>N</u> .GTI F TGA D	CAC T TCA Q CCI L	CAG S G G CAC T	H H AA1 M CCC P	CAG CAG CAG S	900 269 960 289 1020 309
CAA K CAT M GCA H CCC	AGA E GCA Q .CAG R	GAC T ATA Y AGC A .CGT	CCT L TGG G CAG S	TCA H GAA N TTC S	CAA K CAA K CCCC P	GCA Q ATC S CAT I	GTA Y TAT I CTC S CTI	CCA H CTC S CCT L	CCT L TAC T GCC P .GAG	GGT V CAT M TCC P GAA	CAA K GAA K GGT V GCT	GTC S AGT V CAC T GCT	CCA H GAT M ACA H	CAC T GCA Q .CCI L SAAC	CAA N GTI F TGA D	CAC T TCA Q CCI L	CAG S GGG CAC T	CCA H AAI M CCCC P	CAG CAG S AGGA	900 269 960 289 1020 309
CAA K CAT M GCA H CCC P	AGA E GCA Q .CAG R TGA D	GAC T ATA Y AGC A CGT V	CCT L TGG G CAG S GCC P	TCA H GAA TTC S CCT L	CAA K CAA K CCCC P CGAC	GCA Q ATC S CAT I CAT I	GTA Y TAT I CTC S CTT L	CCA H CTC S CCT L GAA K	CCT L TAC T GCC P GCC R	GGT V CAT M TCC P GAA K	CAA K GAA K GGT V GCT L	GTC S AGT V CAC T GCT L	GAT M ACA H GAG R	CAC T GCA Q .CCT L SAAC T	CAA N GTI F TGA D CAA N	CAC T TCA Q CCI L CCA D	CAG GGG G CAC T CAC T V	CCA H AAI M CCC P GAF K	CAG CAG S CAG S CAG S CAG S CAG S	900 269 960 289 1020 309 1080 329
CAA K CAT M GCA H CCC P	AGA E GCA Q .CAG R TGA D	GAC T ATA Y AGC A .CGT V	CCT L TGG G CAG S GCC P	TCA H GGAA N TTCC S CCT L	CAA K CAA K CCCC P CGAC T	GCA Q ATC S CAT I CAT I	GTA Y TAT I CTC S CTT L	CCA H CTC S CCT L GAA K	CCT L TAC T GCC P GAG R	GGT V CAT M TCC P GAA K	CAA K GAA K GGT V GCT L	GTC S AGT V CAC T GCT L	CCA H GAT M ACA H GAC R	CAC T GCA Q .CCT L AAC T	CAA M GTT F TGA D CCAP	CAC T TCA Q CCT L CCA D	CAG S GGG CAC T CAC T CAC	CCA H AAT M CCCC P CGAF K	CAG CAG CAG CAG CAG CAA	900 269 960 289 1020 309 1080 329
CAA K CAT M GCA H CCC P ATC	AGA E GCA Q CAG R TGA D	GAC T ATA Y AGC A CGT V GAA	CCT L TGG G CAG S GCC P TCT	TCA H GAA TTC S CCT L CCT	CAA K CAA K CCCC P CGAC T TGG	GCA Q ATC S CAT I CAT I GCA	GTA Y TAT I CTC S CTT L GAT	CCA H CTC S CCT L GGAA K	CCT L TAC T GCC P GAG R GCA	GGT V CAT M TCC P GAA K ATT	CAA K GAA K GGT V GCT L TCT	GTC S AGT V CAC T GCT L GGA	CCA H GAT M ACA H GACA R TGC	CAC T GCA Q CCT L EAAC T CAG	CAA <u>N</u> GTT F TGA D CCAA N SCAA	CAC T TCA Q CCT L CCT D CCT	CAG S GGG G CAC T CAC V CAT	CCCA H RAAT M CCCCC P CGAF K TGA	GAA K CAG S AGGA E	900 269 960 289 1020 309 1080 329 1140
CAA K CAT M GCA H CCCC P ATC S	AGA E GCA Q CAG R TGA D CCCA Q	GAC T ATA Y AGC A CGT V GAA N	CCT L TGG G CAG S GCC P TCT L	TCA H GGAA TTC S CCT L CCT I	CAA K CAA K CCCC P CGAC T TGG G	GCA Q ATC S CAT I CAT I GCA Q	GTA Y TAT I CTC S CTT L GAT I	CCA H CTC S CCT L CCA K Q Q	CCT L TAC T GCC P GAG R GCA Q	GGT V CAT M TCC P GAA K ATT F	CAA K GAA K GGT V GCT L TCT L	GTC S AGT V CAC T GCT L GGA D	CCA H GAT M ACA H GAG R TGC A	CAC T GCA Q CCT L CAG R	CAA N GTT F TGA D CAA N GCA H	CAC T TCA Q CCT L CCA D CCA V	CAG S GGG G CAC T CAC V CAT I	CCCA H AAAT M CCCC P CGAF K TGA E	GAA K CAG S AGGA E AGAA K	900 269 960 289 1020 309 1080 329 1140 349
CAA K CAT M GCA H CCCC P ATC S	AGA E GCA Q CAG R TGA D CCCA Q	GAC T ATA Y AGC A CGT V GAA N	CCT L TGG G CAG S GCC P TCT L	H H GAA TTC S CCT L CAT I	CAA K CAA K CCCC P CGAC T TGG G	GCA Q ATC S CAT I CAT I GCA Q	GTA Y TAT I CTC S CTT L GAT I	CCA H CCTC S CCCT L GGAA K CCCA Q	CCT L TAC T GCC P GAG R GCA Q	GGT V CAT M TCC P GAA K ATT F	CAA K GAA K GGT V GCT L TCT L	GTC S AGT V CAC T GCT L GGA D	CCA H GAT M ACA H GAC R TGC A	CAC T GCA Q CCT L CAG R	CAA <u>N</u> GTT F TGA D CAA N GCA H	CAC T TCA Q CCT L CCT D CCT V	CAG GGGG G CAC T CAC T V CAT I	CCCA H HAATI M CCCC P CGAF K TGA E	CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG	900 269 960 289 1020 309 1080 329 1140 349
CAA K CAT M GCA H CCCC P ATC S GTCC	AGA E GCA Q CAG R TGA D CCCA Q TGT	GAC T ATA Y AGC A CGT V GAA N GCA	CCT L TGG G CAG S GCC P TCT L CAA	TCA H GAA TTC S CCT L CCAT I GAT	CAA K CAA K CCAA P CGAC T TGG G CGT	GCA Q ATC S CAT I CAT I GCA Q TTC	GTA Y TAT I CTC S CTT L GAT I CCT	CCA H CTC S CCT L GAA K CCA Q Q GCT	CCT L TAC T GCC P GAG R GCA Q GGC	GGT V CAT M TCC P GAA K ATT F GGG	CAA K GAA K GGT V GCT L TCT L	GTC S AGT V CAC T GGT L GGA D TGG	CCA H GAT M ACA H GAG R TGC A GGA	CAC T GCA Q CCT L CAG R CAG R AAC	CAA N GTT F TGA D CCAA N GCA H TGC	CAC T TCA Q CCT L CCA D CCA V TGA	CAG S GGG G CAC T CAC V CAT I GAG	CCCA H RAAT M CCCC P CGAP K TGA E	TGT V CAA CCAG S AGGA E GAA K TCT	900 269 960 289 1020 309 1080 329 1140 349
CAA K CAT M GCA H CCCC P ATC S GTC S	AGA E GCA Q CAG R TGA D CCCA Q TGT V	GAC T ATA Y AGC A CGT V GAA N GCA H	CCT L TGG G CAG S GCC P TCT L CAA K	TCA H GGAA TTC S CCT L CCT I CCAT I GAT I	CAA K CAA K CCAA P CGAC T TGG G CGT V	GCA Q ATC S CAT I CAT I GCA Q TTC S	GTA Y TAT I CTC S CTT L GAT I CCT	CCA H CTC S CCT L GAA K CCA Q GCT L	CCT L TAC T GCC P GCA R GCA Q GCA Q GGC A	GGT V CAT M TCC P GAA K ATT F GGG	CAA K GAA K GGT L TCT L ATT F	GTC S AGT V CAC T GCT L GGA D TGG G	CCA H GAT M ACA H GAG R TGC A GGA E	CAC T GCA Q CCT L EAAC T CAG R AAC T	CAA N GTT F TGA D CCAA N GCA H TGC A	CAC T TCA Q CCT L CCT D CCT V TGA E	CAG S GGGG G CAC T V CAT I GAG R	CCCA H AAT M CCCC P CGAF K TGA E TGA H	TGT V CGAA K CCAG S AGGA E GAAA K TTCT L	900 269 960 289 1020 309 1080 329 1140 349 1200 369
CAA K CAT M GCA H CCCC P ATC S GTC S	AGA E GCA Q CCAG R TGA D CCCA Q TGT V	GAC T ATA Y AGC A CGT V GAA N GCA H	CCT L TGG G CAG S GCC P TCT L CAA K	TCA H GAA TTC S TTC S CCT L CCT I CAT I I GAT	CAA K CAA K CCCC P CGAC T TGG G CCGT V	GCA Q ATC S CAT I GCA Q TTC S	GTA Y TAT I CTC S CTI L GAT I CCT L	CCA H CTC S CCT L GAA K CCA Q GCT L	CCT L TAC T GCC P GCA R GCA Q GCA Q GGC A	GGT V CAT M TCC P GAA K ATT F GGG G	CAA K GAA K GGT V GCT L TCT L ATT F	GTC S AGT V CAC T GCT L GGA D TGG G	CCA H GAT M ACA H GGAC R TGC A GGA E	CAC T GCA Q CCT L CAG R CAG R AAC T	CAA N GTT F TGA D CAA N GCAA H TGC A	CAC T Q CCT L CCT D CCT V TGA	CAG S GGG C CAC T V CAT I GAG R	CCA H H AAAT M CCCC P GAA K TGA K TGA H ACA H	TGT V CGAA K CCAG S AGGA E GAA K TCT L	900 269 960 289 1020 309 1080 329 1140 349 1200 369
CAA K CAT M GCA H CCCC P ATC S GTC S GTC	AGA E GCA Q CAG R TGA D CCCA Q TGT V X AGA	GAC T ATA Y AGC A CGT V GAA N GCA H GAG	CCT L TGG G CAG S GCC P TCT L CAA K GAC	TCA H GAA TTCC S CCT L CCT I CAT I SCAT	CAA K CAA CCCC P CGAC T TGG G CCGT V V CGCT	GCA Q ATC S CAT I GCA Q TTC S CAC	GTA Y TAT I CTC S CTI L GAT I CCT L AGC	CCA H CTC S CCT L GGAA K CCA Q GCT L ACA	CCT L TAC T GCC P GAG R GCA Q GGC A TGA	GGT V CAT M TCC P GAA K ATT F GGG G CTG	CAA K GAA K GGT U GCT L TCT L ATT F	GTC S AGT V CAC T GGA D GGA G G CCA	CCA H GAT M ACA H GGAC R TGC A GGA E GGA	CAC T GCA Q CCT L CAG R CAG R T GGC	CAA N GTT F TGA D CCAA N GCA H TGC A	CCT L CCT L CCT L CCT V TGA E	CAG S GGG G CAC T V CAT I GAG R CAT	CCA H H AAAT M CCCC P GAA K TGA K TGA H ACA H	TGT V CGAA K CCAG S AGGA K K TCT L	9000 2655 9600 2889 10200 3059 10800 3259 11400 3459 12000 3659
CAA K CAT M GCA H CCCC P ATC S GTC S GTC S	AGA E GCA Q CAG R TGA D CCCA Q TGT V V AGA E	GAC T ATA Y AGC A CGT V GAA H GAG R	CCT L TGG G CAG S GCC P TCT L CAA K GAC T	TCA H GGAA TTC S CCT L CCT I I GGAT I S CCAT M	CCAA K CCAA K CCCC P CCCC T TGG G CCCT V CCCT L	GCA Q ATC S CAT I GCA Q TTC S CAC T	GTA Y TAT I CTC S CTI L GAT I CCT L AGC A	CCA H CTC S CCT L GGAA K CCA Q GCT L ACA H	CCT L TAC T GCC P GGAG R GGA Q GGC A TGA D	GGT V CAT M TCC P GAA K ATT F GGG G CTG C	CAA K GAA K GGT U GCT L TCT L ATT F CTA Y	GTC S AGT V CAC T GGA D GGA G G CCA Q	CCA H GAT M CACA H GGAC R TGC A GGA E GGA E	CAC T GCA Q CCT L CAG R CAG R AAC T GGC A	CAA <u>N</u> GTT F TGA D CAA N GCAA H TGC A TGC V	CCT L CCT L CCT L CCT L CCT V TGA E YAAC T	CAG S GGG G CAC T V CAT I GAG R CCA R H	CCCA H AAAT M CCCC P CGAA K TGA K TGA H ACA H CTT F	TGT V CGAA K CCAG S AGGA K K TCT L CCCG R	900 265 960 285 1020 309 1080 329 1140 349 1200 365 1220 389
CAA K CAT M GCA H CCCC P ATC S GTC S GTC S	AGA E GCAG R TGA D CCCA D CCCA Q TGT V X AGA E	GAC T ATA Y AGC A CGT V GAA N GCA H GAG R	CCT L TGG G CAG S GCC P TCT L CAA K GAC T	TCA H GGAA TTC S CCT L CCT I CCT I I CCT I I CCT I I M	CAA K CAA K CCAA F CCCC T T CCCC T T CCCCT V CCCT L	GCA Q ATC S CAT I GCA Q TTC S CAC T	GTA Y TAT I CTC S CTT L GAT I CCT L AGC A	CCA H CTC S CCT L CCA K CCA K CCA Q GCT L ACA H	CCT L TAC T GCC P GGC R GCA Q GGC A TGA D	GGT V CAT M TCC P GGAA K F GGG G C C	CAA K GAA K GGT L GCT L TCT L ATT F CTA Y	GTC S AGT V CAC T GGA D GGA G G CCA Q	CCA H GAT M CACA R TGC R TGC A GGA E GGA	CAC T GCA Q CCT L CAG R CAG R A A CAG C A	CAA N GTT F TGA D CCAA N GCA H TGC A TGT V	CAC T TCA Q CCT L CCT D CCT V TGA E TGA T	CCAG GGGG G CCAC T V CCAT I GAG R CCAA H	CCCA H AAT M CCCC P CGAF K TGA E ACA H CTT F	TGT V GAA K CAG S AGGA E GAA K TCT L CCCG R	900 269 960 289 1020 309 1080 329 1140 349 1200 369 1260 389
CAA K CAT M GCA H CCCC P ATC S GTC S GTC S CAC	AGA E GCAG R TGA D CCCA Q TGT V V ZAGA E	GAC T Y AGC A CGT V GAA N GCA H GAG R CTG	CCT L TGG G CAG S GCC P TCT L CAA K GAC T	TCA H GAA S TTC S CCT L CCT I CCT I I GAT I M TAA	CAA K CAA CCAA CCCC P CGAC T CGAC T CGT V CGCT L	GCA Q ATC S CAT I GCA T CAC S CAC T GCA	GTA Y TAT I CTC S CTT I GAT I CCT A CCT A CTC	CCA H CTC S CCT L GGAA K CCA Q GCT L GCT L H TGT	CCT L TAC T GCC P GGC R GCA GGC A TGA D CAC	GGT V CAT M TCC P GAA K GGG G C C GTA	CAA K GAA K GGT L CTA TCT F CTA Y CGA	GTC S AGT V CAC T GGA D GGA CCA Q GCA	CCA H GAT M ACA H CGAC R TGC GGA E GGA E TGC	CAC T GCA Q CCT L CAG R CAG R A A CAG C T CAG C T	CAAA <u>N</u> GGTT F TGAA N GCAAA H TGC A TGC V GCG	CACC T Q CCT L CCT D CCT V TGA C CT T CCT T CCT T CCT T CCT T CCT T CCT T CCT T CCT T CCT T C CCT T C A CCT L C C C T L C C C T L C C C C C C C	CCAG GGGG G CCAC T V CCAT I GAG R CCCA H CCCA	CCCA H AAAI M CCCC P CAAF K TGA K TGA H ACA H CTTI F	TGT V GAAA K CAGG S AGGA K K TCT L CCCG R TGT	900 269 960 289 1020 309 1080 329 1140 349 1200 369 1200 389
CAA K CAT M GCA H CCCC P ATC S GTC S GTC S CACC T	AGA E GCA Q CAG R TGA D CCCA Q TGT V V AGA E ACA	GAC T ATA Y AGC A CGT V GAA N GCA H GAG R CTG C	CCT L TGG G CAG S GCC P TCT L CAA K GAC T F	TCA H GAA S TTC S CCT L CCT I CCT I S CAT I M M TAA N	CAA K CAA CCAA CCCC P CGAC T CCGT V CGCT L CCTG W	GCA Q ATC S CAT I CAT I GCA Q TTC S CAC T GCA H	GTA Y TAT I CTC S CTI I CCT I CCT A CTC S	CCA H CTC S CCT L GAA K CCA K CCA Q GCT L H TGT V	CCT L TAC T GCC P GGAG R GCA Q GGC A TGA D CAC T	GGT V CAT M CAT P GAA K ATT F GGG C C C GTA Y	CAA K GAA K GGT L CTA TCT F CTA Y CGA E	GTC S AGT V CAC T GGA D GGA Q CCA Q GCA	CCA H GAT M ACA H CGAC R TGC A GGA E GGA E TGC A	CAC T GCA Q CCT T CAG R CAG R CAG C T CAG T CAG T CAG T L	CAAA N GTT F TGAA D CCAAA N GCAA H TGC A TGC A TGT V GCG R	CAC T TCA Q CCT L CCGA D CCGT V TGA CCGT T CCGT T CCGT Y	CCAG GGGG G CCAC T CCAT I GAG R CCAT L	CCCA H AAAI M CCCC P CAAF K TGAA E ACCA H CTTI F CTTI F	TGT V GAAA K CCAG S AGGA K TCT L CCCG R TGT V	900 265 960 289 1020 329 1140 349 1200 369 1260 389 1260 389
CAA K CAT M GCA H CCCC P ATC S GTC S GTC S CAC T	AGA E GCA Q CAGG R TGA D CCCA Q TGT V V AGA E ACA H	GAC T ATA Y AGC A CGT V GAA N GCA H GAG R CTG C	CCT L TGG G CAG S GCC P L CAA K GAC T CTT F	TCA H GAA TTC S CCT L CCT I CCT I I CCT M TAA N	CAA K CCAA K CCCC P CGAC T CGAC T G CGAC T CGAC T CGAC T CGAC T CGAC T CGAC T CGAC T CGAC T CGAC T CGAC C T CGAC C T CGAC C T CGAC C C C C C C C C C C C C C C C C C	GCA Q ATC S CAT I CAT I GCA Q TTC S CAC T GCA H	GTA Y TAT I CTC S CTT L GAT L CCT A GCT S	CCA H CTC S CCT L GAA K CCA Q GCT L ACA H TGT V	CCT L TAC T GCC P GGAG R GCA Q GGC A TGA D CAC T	GGT V CAT M TCC P GAA K ATT F GGG C C GTA Y	CAA K GAA K GGT L GCT L TCT L ATT F CTA Y CGA E	GTC S AGT V CAC T GGA D TGG G CCA Q CCA P	CCA H GAT M GACA R TGC A GGA E GGA E TGC A	CAC T GCA Q CCT T CAG R AAC T GGC A CTT L	CAA N GTT F TGA D CCAA N GCAA H TGC A TGC A TGT V GCG R	CACC T Q CCT L CCGA D CCGT V TGA CCGT T GGTA Y	CAG GGG G CAC T V CAT I GAG R CCA H CCAT L	CCCA H M CCCC P CGAA K TGA E CTI F GTA Y	TGT V GAAA K CCAG S AGGA E GAAA K TCT L CCCG R TGT V	900 265 960 289 1020 309 1080 329 1140 349 1200 369 1260 389 1260 389
CAA K CAT M GCA H CCC P ATC S GTC S GTC S CAC T GCT	AGA E GCA Q CAG R TGA D CCA Q TGT V TGT V V AGA A H V GGC	GAC T ATA Y AGC A CGT GAA R GCA R CTG C CAA	CCT L TGG G CAG S GCC P TCT L CAA K GAC T T TCT	TCA H SGAA TTC S CCT L CCT I CCAT I S CCAT M N S CCT G	CAA K CCAA CCCC P CGAC T CGAC T CGGT CGT C CGT L CCGT L CCGG CGT C CGT C CGT C CGT C CGT C CGT C CGT C CGT C C CGC C C C	GCA Q ATC S CAT I CAT I GCA T CAC T CAC T GCA H GGC	GTA Y TAT I CTC S CTT L GAT L CCT A CCT S ACC	CCA H CTC S CCT L GAA K CCA Q GCT L CCA K CCA TGT T TGT V ATA	CCT L TAC T GCC P GGC R GCA GGC A TGA D CAC T CAC	GGT V CAT M TCC P GAA K ATT F GGG G C C GTA Y GAT	CAA K GAA K GGT L CTA F CTA F CTA CGA E TGA	GTC S AGT V CAC T GCT GGA G GGA CCA G CAG	CCA H GAT M GACA R TGC A GGA E GGA TGC A GAT	CAC T GCA Q CCT T CAG R AAC T CAG A CAG A CTT L	CAA N GTT F TGA D CCAA N GCA H TGC A TGC A TGT V GCG R GAT	CCAC T TCA Q CCT L CCT D CCGT V TGA CCGT V TGA C GGA Y CGGC	CAG GGG G CAC T V CAC T I GAG R CCAT I CCAT L CCAT	GGAA GGAA CCCC P GGAA K TGA K TGA K CTT F GGA Y GGA	TGT V CGAA K CCAG S AGGA K TCT L CCCG R TCT V V CCAA	900 265 960 285 1020 309 1080 329 1140 349 1200 369 1200 389 1320 409
CAA K CAT M GCA H CCC P ATC S GTC S GTC S CAC T L	AGA E GCA Q CAG R TGA D CCCA Q TGT V TGT V XAGA E XAGA H CGGC A	GAC T Y AGC A CGT V GAA A GCA R CTG C CAA N	CCT L TGG G CAG S GCC P TCT L CAA K GAC T T T T T T T L	TCA H GAA TTC S CCI L CAI I GAI M TAA N TAA N CCI C	CAA K CAA CCAA P CCCC P CCCC T CCCC T CCCC T C CCCCC T C CCCC T C CCCCT T C CCCCT T C CCCCT T C CCCCT T C CCCCT T C CCCCT T C CCCCT T C CCCCT T C CCCCT T C CCCCT T C CCCCT T C CCCCT T C CCCCT T C CCCCT T C C CCCTT T C CCCTT T C CCCTT T C CCCTT T C CCCTT T C CCCTT T C CCCTT T C CCCTT T C CCCTT T C CCCTT T C CCCTT T C CCCTT T C CCCTT C CCCTT T C CCCTT C C CCCTT C C CCCTT C C CCCTT C C CCCTT C C CCCTT C C C CCCTT C	GCA Q CAT I CAT I GCA T CAT S CAC T CAC T CAC A GCA A	GTA Y TAT I CTC S CTT L GAT L CCT A CCT A CCC A CCC P	CCA H CTC S CCT L GAA K Q GCT L CCA Q GCT L H TGT V X ATA Y	CCT L TAC T GCC P GGC R GGC R GGC A C GGC A TGA D C A C A C T C A C C C C C C P C C C P C C C P C C P C C P C C P C C P C C C P C C C P C C C P C C C P C C C P C C C C P C C C P C C C C P C C C P C C C P C C C P C C C P C C C P C C C P C C C C P C C C P C C C C P C C C C C P C	GGT V CAT M TCC P GAA K GGA C GGA C GTA Y GAT I	CAA K GAA K GGT L GCT L TCT L CTA F CTA Y CGA E TGA D	GTC S AGT V CAC T GCT GGA GGA CCA G GCA H CAG R	CCA H GAT M GACA TGC A GGAC E GGA E TGC A GGAT I	CAC T GCA Q CCT T CAG R AAC T CAG R AAC T CAG A CTT L AGA E	CAA N GTT F TGA D CAA N GCAA H TGC A TGC A TGC A CAA M	CCAC T TCA Q CCT L CCT D CCA D CCA T CCA T CCA T CCA C C T C CCA C C C C	CAG GGG G CAC T V CAT I GAG R CAT I CAT L CAT L	CCCA H AAAT M CCCC P CGAA K TGAA E CTT F CACA H CTT F CGTA Y CGGA D	TGT V CGAA K CCAG S AGGA K TCT L CCCG R CCCG R TGT V V CCAA K	900 265 960 289 1020 309 1080 349 1200 369 1260 389 1260 389 1260 389 1320 409
CAA K CAT M GCA H CCC P ATC S GTC S GTC S CAC T CAC T L	AGA E GCA Q CAG TGA TGA CCA Q TGT V V AGA E ACA H CGGC A	GAC T ATA Y AGC A CGT V GAA N GCA R GCA CTG C CAA N	CCT L TGG G CAG S GCC P TCT L CAA K GAC T F TCT L	TCA H GAA TTC S CCI L CCI I GAI I CCAI M TAA N N CCI C	CAA K CCAA CCCC P CCCC T CCCC T CCCC C CCCC C CCCC C CCCC C CCCC C C CCCC	GCA Q ATC S CAT I GCA Q TTC S CAC S CAC T GCA H GGC A	GTA Y I CTC S CTT L GAT L CCT L CCT C A CC C C C C C C C C C C C C C C	CCA H CTC S CCT L CCA K CCA K CCA C CCA K CCA T G C L CCA T G C T C C C C C C C C C C C C C C C C	CCT L TAC GCC P GGC R GGC R GGC R GGC R C C C C C C C C	GGT V CAT P GAA K GGGG C GTA GTA Y GAT I	CAA K GAA K GGT L GCT L TCT L TCT F CTA Y CGA E TGA D	GTC S AGT V CAC T GGA D TGG G GCA Q GCA H CAG R	CCA H GAT M GACA TGC A GGAC E GGA E TGC A GGAT I	CAC T GCA Q CCT T CAG R AAC T GGCC A CCTT L VAGA E	CAA N GTT F TGA D CAA N GCAA H TGC A TGC A TGC C A TGC R GCG R GAT M	CGAC TCA Q TCA Q CCT L CGA D CGA TGA CGTA T GGCA	CCAG G G CCAC T CCAT I GAG R CCAT L CCAT L CCAT L CCAT M	CCCA H AAAT M CCCC P CGAF K TGA E TGA H CTT F CTT F CGTA Y CGGA D	TGT V CAAG S CAGG S CAGGA E GAAA K TCT L CCCG R TGT V CCAA K	900 265 960 289 1020 309 1080 329 1140 349 1200 369 1260 389 1320 409
CAA K GCA H CCCC P ATC S GTC S GTC S CAC T GCT L AGT	AGA E GCA Q TGA D CCCA D CCCA Q TGT V V AGA E ACA H GGCC A	GAC T ATA Y AGC A CGT V GAA N GCA R GCA CTG C CAA N TCT	CCT L TGG G CAG S GCC P TCT L CAA K GAC T CTT F L TCT L	TCA H GGAA TTC S CCT L CCAT I GGAT I CCAT M TAA N CCTC C C	CAA K CAA K CCAA C CAC T CCC T C C C C C C C C C C	GCA Q ATCCS CAT I CAT I GCA CAC T CAC T CAC CAC C CAC C CAC C CAC C CAC C CAC C CAT I CAT I CAT I CAT I CAT C CAT I CAT C C CAT C C CAT C C CAT C C CAT C C C C	GTA Y TAT I CTC S CTT L GAT I CCT L AGC A CTC S AAC	CCA H CTC S CCT L GCA K CCA K CCA C CA A C CA A CA CA CA CA CA CA CA	CCT L TAC T GCC P GGC R GGA R GGA C A TGA T CAC T C TCC P TCG	GGT V CAT M TCC P GAA K ATT F GGG G C C GTA Y GAT I CTT	CAA K GAA K GGT L CTA F CTA F CTA CGA CCA	GTC S AGT V CAC T GGA D TGG G G CCA G CCA R CAG R AAT	CCA H GAT ACA H GGAG R GGA E GGA TGC A GGA TGC A GGAT I GAG	CAC T GCA Q CCT E AAC T GGC A CAG R CAG R CAG C T CAG C A C T CAG C C T C CAG C T T CAG C C T T C CAG C T T C C C T T C C C T C C C T C	CAA N GTT F TGA D CCAA N GCAA N TGC A TGC R GCG R GCA M GCA	CAC T Q TCA Q CCT L CCT V CCA T CCA T CCA T CCA T CCA T CCA C CAC D CCA T C CAC T C CAC T C CCA C C C C C C	CAG G G CAC T CAT I GAG R CAT I GAG R CAT I CAT I CAT I CAT	CCCA H AAAT M CCCC P CGAA K TGAA K TGAA H CTT F GTA Y CGGA D	TGT V GAAA K CAGGS AGGA K TCT L CCCG R TCT T V V CCCG R TGT V V CCAA K CCAA CCCAA R CCCG	900 265 960 289 1020 309 1080 329 1140 365 1200 365 1200 389 1320 409 1320 409
CAA K CAT M GCA H CCCC P ATC S GTC S GTC S GTC S CAC T CAC T L AGT V	AGA E GCA Q CAG R TGA D CCCA D CCCA Q TGT V V AGA E H CGGC A C C	GAC T ATA Y AGC A CGT V GAA C GAA C C C C C C C C C C C C C C C	CCT L TGG G CAG S GCC P TCT L CAA K GAC T CTT F L TCT L TAG S	TCA H GAA TTC S CCT L CCT I CCAT I CCAT M TAA N CCTC C C TCA H	CAAA K CCAA K CCCC P CGAC T TGG G CGGT L CCGG V CGCT L CCGGT K CCGGT L CCGGT L CCGGT K CCGGT C CGCT C CGCC C CGCC T CGAC C CCCC P CGAC C T CGAC C T CGAC C T CCCC P C CCCC P C CCCC P C CCCC P C CCCC P C CCCC P C CCCC P C CCCC P C CCCC P C CCCC C C C C C C C C C C C C C C C C	GCA Q ATCC S CAT I CAT I GCA Q TTCC S CAC T CAC S CAC A CTCC A CTCC S CAC CAC S CAC S CAC S CAT I I S CAT I I S CAT I I S CAT S I S CAT S I S CAT S I S CAT S I S CAT S I S CAT S I S CAT S I S CAT S I S CAT S I S CAT S I S CAT I I S CAT I I S CAT S I S CAT I I S CAT S I S CAT S I S CAT S I S CAT S I S CAT S I S S CAT S S CAT S S S S S S S S S S S S S S S S S S S	GTA Y TAT I CTC S CTI L GAT I CCT L AGC A CTC S AACC	CCA H CTC S CCT L GAA K CCA K CCA C CCA K CCA K CCA TGT L CCA S CCT S S CCT S S CCT S S CCT S S CCT S S CCT C S S CCT S S CCT S S CCT S S CCT S S CCT S S CCT S S CCT S S CCT S S CCT S S S CCT S S S CCT S S S CCT S S S CCT S S S S	CCT L TAC T GGCG P GGC R GGC R GGC A CAC T TCAC T TCC TCC	GGT V CAT M TCC P GAA K ATT F GGG C GTA C TG C TG C T I CTT	CAA K GAA K GGT L CTA F CTA F CTA F CGA D CCCC	GTC S AGT V CAC T GCT L GGA CCA G GCA H CAG R AAT	CCA H GAT ACA H GGAG R TGC A GGA TGC A GGA I GAG	CAC T GCA Q CCT L CAG R CAG R CAG C T CAG C CAG C T CAG C CAG C T CAG C CAG C CAG C CAG C CAG C CAG C C CAG C C C C	CAA N GTT F TGA D CAA N GCA H TGC A TGC A TGC R GCG R GCA M GCA	CGAC T Q TCA Q CCT L CGA CGA T CGA CGTA Y CGGC A CAG	CCAG G G CCAC T CCAT I GAG R CCAT L CCAT L CCAT M CCAT CCAT	CCCA H AAAT M CCCC P GGAA K TGAA K TGAA H CTT F GGAA H S GGAA D CTT F CGTA T CTT F	V CAAA K CCAG S AGGA E AGGA K CTCT L CCCG R UTGT V CCAA K CCAA K	900 265 960 285 1020 305 1080 325 1140 349 1200 365 1260 405 1380 405 1380 425
CAA K CAT M GCA H CCCC P ATC S GTC S GTC S GTC S CAC T GCT L AGT V	AGA E GCA Q CAG R TGA D CCA Q TGT V AGA E C C C C C C	GAC T ATA Y AGC A CGT V GAA C GAA C C C C C C C C C C C C C C C	CCT L TGG G CAG S GCC P TCT L CAA K GAC T TCT F TCT L TAG S	TCA H GAA TTC S CCT L CCAT I GAT M TAA N CCTG C C TCA H	CAAA K CCAA K CCCC P CGAC T TGG G CGT C CGT L CCTG W TGA E X TGA X TGA X TGA X TGA X TGAC T T CGT T L	GCA Q ATC S CAT I GCA T CAT S CAC T CAC T GCA C T GCA C T C CAC T S CAC T C CAT I S CAT I S CAT I CAT I S C C C T C C T C S C T C C T C S C C C T C S C S	GTA Y TAT I CTC S CTT L GAT L CCT A CCC A A CTC S A A CCC A A C C C C A	CCA H CTC S CCT L GAA K CCA K CCA C CCA K CCA K CCA TGT V S AGC	CCT L TAC T GGCG P GGC R GGC R GGC A CAC T TCC P TCC TCC	GGT V CAT M TCC P GAA K ATT F GGG C C GTA Y GAT I CTT	CAA K GAA K GGT L TCT L TCT CTA F CTA Y CGA E TGA D CCC	GTC S AGT V CAC T GCT L GGA CCA G GCA H CAG R AAT	CCA H GAT ACA H GAC R TGC R GGA E GGA TGC A GGA T GAG	CAC T GCA Q CCT L CAG CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C C CAG C C C C	CAA N GTT TGA D CCAA N GCA H TGC A TGC A TGC R GCA M GCA	CAC T Q TCA Q CCT L CCT L CCT V CCT A CCT T CCT A C CCT A C CCT A C CCT A C C C C	CAG S GGG G CAC T V CAT I GAG R CCAT L CCAT L CCAT S CCAT	GCCA H AAAI M CCCC P GAA K TGA K TGA H CTI F GTA Y GGA D CACI	TGT V GAAA K CCAG S AGGA E GAAA K TCT L CCCG R TCT V V CCAA K S GGA	900 265 960 285 1020 309 1080 329 1140 349 1200 369 1200 389 1320 409 1380 429
CAA K CAT M GCA H CCCC P ATC S GTC S GTC S GTC S GTC S GTC S CAC T CAC T CAC A GTC A A GTC A A C A C A T A C A T A C A T A C A T A C A T A C A C	AGA E GCA Q CAG R TGA D CCA Q TGA V CAGA E V V AGA A H CGGC C TGA	GAC T ATA Y AGC A CGT V GAA N GGA R GAG R CTG C CAA TCT L ACC	CCT L G CAG S CAG S CAG T L CAA K GAC T T CTT F TCT L S S AAC	TCA H GAA TTC S CCT L CCAT I CCAT I CCAT N CCTG C C TCA H	CAAA K CAAA CCAA CCAA T CCAA T CCAT C CCAT C CCAA CCAA CCAA CCAA CCAA CCAA CCAA CCAA CCAAA CCCC C C CCC C C C C C C C C C C C C C	GCA Q ATC S CAT I GCA T CAT S CAC T GCA H GGC A CTG CTG	GTA Y TAT I CTC S CTI L GAT L CCT A ACC P AACA	CCA H CTC S CCT L GGAA K CCA Q GCT L ACA H TGT V ATA Y GGG	CCT L TAC T GCC P GGC R GCA C C GCA C T CAC T C C GCC C C GCC C C GCC C C GCC C C C	GGT V CAT M GGAA K GGAA CTG GGG C GTA Y GAT I CTT ACC	CAA K GAA K GGT L TCT L TCT CTA TCT CTA TCT CCA CCC AGA	GTC S AGT V CAC T GCA D CCA G GCA H CAG R CAG R AAT GGC	CCA H GAT ACA H GAC R TGC R GGA E GGA TGC A GAT I GAG	CAC T GCA Q Q CCT L CAG R CAG R CAG C T CAG A CTT L CAG A CTT L CAG C A CAC T CAG C C C C C C C C C C C C C C C C C C	CAA N GTT TGA D CCAA N GCA H TGC A TGC A GCA R GCG R GCA CGC	CAC T TCA Q CCT L CCA D CCA T CCA T CCA T CCA C CCA C CCA C CCA C CCA C CCA C CCA C CCA C CCA C CCA C	CCAG S GGG G CCAC T V CCAT I GAG R CCAT L CCAT L CCAT L CCAT CCAT CCAT CCA	CCCA H AAAI M CCCC P GGAA K TGAA K TGAA H CTT F GGAA D CACI CGGA	TGT V GAAA K CCAG S GGAA K TCT L CCCG R TGT V V CCAA K CCCC	900 265 960 289 1020 305 1080 329 1140 369 1200 369 1320 409 1380 429 1380 429 1440 435
CAA K CAT M GCA H CCC P ATC S GTC S GTC S GTC S CAC T L AGT V ATA CAG	AGA E GCA Q CAG R TGA D CCA Q TGT V V CAGA E C C C C C C C C C C	GAC T ATA Y AGC A CGT V GAA H GAG R GAA H GAG C C C C A C T C T C T C T C C A	CCT L TGG G CAG S CAG P TCT L CAA K GAC T TCT F TAG S AAC	TCA H GAA TTCC S CCT L CAT I CAT N TAA N CCTCC C C TCA H CCACCCCCCCCCCCCCCCCCCCC	CAAA K CCAA CCAA CCAA T CCAT T CCAT C CCAT C CCAT C CCAT C CCAT C CCAT C CCAA T C CCAA T C CCAA T C CCAA C T C CCAC T C CCAC T C CCCT C C CCCT C	GCA Q ATC S CAT I GCA Q T T CAT S CAC T GCA T CAC S CAC C C C C C C C C C C C C C C C	GTA Y IAT I CTC S CTT L GAT L AGC A ACC P AAAA CCA	CCA H CTC S CCT L CCA K CCA K CCA C Q CCA K CCA K CCA S CCT L CCA S CCA S CCT S S CCT S S S CCT S S S CCT S S S CCT S S S S	CCT L TAC T GGCC P GGC R GGA C C C C C C C C C C C C C C C C C C	GGT V CAT M TCC P GAA K G G G C G G C G G TA Y GAT I CTT ACC GCT	CAA K GAA K GGT V GCT L TCT L CTA Y CGA E TGA D CCC AGA	GTC S AGT V CAC T GCT L GGA D CCA G CCA Q GCA H CAG R AAT GGC TGA	CCA H GAT M GACA R GGAG C C A GGA C C A GGA TGC A GAT I GAG A GAG	CAC T GCA Q Q CCT T CAG CAG C T CAG C T CAG C T C CAG C T C CAG C T C CAG C T C CAG C T C CAG C T C CAG C T C C C T C C C C T C C C C C C C C	CAA N GTT F TGA D CCAA N GCAA H TGC A TGC A TGC A GCAA CCAA N GCAA CCAA N GCAA CCAA N GCAA CCAA N CCAA CCAA CCAA N CCAA N CCAA N CCAA N CCAA	CGAC T TCA Q CCT L CCT T CCGT T CCGT T CCGT T CCGT T CCGT C CCT	CAG S GGG G CAC T V CAT I GAG R CCAT L CCAT L CCAT L CCAT CAT CAT CAT CAT CAT CAT CAT CAT CA	CCCA H RAAT M CCCC P CAA K TGAA K TGAA K TGAA H CTT F GGAA U CTT F GGAA CTT F CGCA CTT F CCCC P CCCA P CCCC P CCCC P CCCC P CCCC P CCCC P CCCC P C CCCC P C CCCC P C CCCC P C CCCC P C CCCC P C	UGAA K CAG S AGGA K CAG S AGGA K CTCT L CCCG R CCCCG K CAAA K CCAA S CCCAA CAAA CAAA CAAA CA	900 265 960 289 1020 325 1140 349 1200 365 1260 389 1320 409 1380 425 1380 425 1380 435
CAA K CAT M GCA H CCC P ATC S GTC S GTC S GTC S CAC T L AGT V ATA CAC CCA	AGA E GCA Q TGA TGA Q TGA C C AGA C C C TGA C C TGA C C TGA C C TGA C	GAC T ATA Y AGC A CGT V GAA H GCA R C C C C C A C TCT L ACC TGG	CCT L TGG G CAG P TCT L CAA K GAC T CTT F TCT T CAA S GAC GTG	TCA N CCT CCT CCT CCT CCT CCT CCT	CCAA K CCAA CCAA CCAA T CCCC T TGG G CCGT V CCGT C CCTG W CCCA X CCCA X CCCA X CCCA X CCCA X CCCA X CCCC C C CCCC T CCCC T CCCC T C CCCC T C CCCC T C CCCC T C CCCC T C CCCC T C CCCC T C CCCC T C CCCC T C CCCC T C CCCC T C CCCC T C CCCT T C CCCT C C CCCC T C CCCT C C CCCC T C CCCT C C CCCC T C CCCC T C CCCT C	GCA Q ATC S CAT I GCA Q TTC S CAC T GCA T CAC C C C C C C C C C C C C C C C C	GTA Y TAT I CTC S CTI I GAT I CCT A ACC A ACC P AAA CCAC CCC CCC	CCA H CTC S CCT L GGAA K CCA K CCA C Q CCA K CCA K C CA S GGCT C C C C C C C C C C C C C C C C C	CCT L TAC T GGCC P GGC R GGC R GGC A C GGC T TCA TCC T C CCC T C CCGC CCTT AAG	GGT V CAT M TCC P GAA K ATT F GGG G G G G TA Y GAT I CTT I CTT ACC GCT AAC	CAA K GAA K GGT L TCT F CTA TCT F CTA Y CGA CCC AGA TTT GAA	GTC S AGT V CAC T GCA D GGA CCA G GCA H CAG R AAT GGC TGA CTT	CCA H GAT M GACA R TGC A GGA E GGA E TGC A GAT I GAG A GAT	CAC T GGCA Q CCT T CAG R CAG R CAG C T CAG C T C CAG T C CAG T C CAG T C CAG T C CAG C T C CAG C T C CAG C C C T C CAG C C C C C C C C C C C C C C C C C	CAA N GTT F TGA D CCAA N GCAA H TGC A TGC A GCA R GCA M GCA CCCC CCTC	CGAC T TCA Q CCT L CGA CGT T CGA CGT C CGC C CGC C CGC C CGC C CGC C CGC C CGC C CGC C CGC C CGC C CGC C CGC C CGC C C C C C C C C C C C C C C C C C C C	CAG S GGG G CAC T V CAT I GAG R CCAT L CCAT L CCAT CCAT CCAT CCAT CCAT	CCA H RAAT M CCCC P GGAA K TGA K TGA H CTT F GGAA Y CGGA QGGA C TTT	V V CGAA K CCAG S AGGA K CTCT L CCCG R TTCT V V CCAA K CCCAA K CCCCC CCCCCCCCCCCCCC	900 265 960 285 1020 309 1200 349 1200 365 1200 389 1320 405 1380 405 1380 425 1440 435 1500 1560
CAA K CAT M GCA H CCC P S GTC S GTC S GTC S GTC S CAC T T L AGT V V ATA CAA CAT TTT	AGA E GCA Q CAG R TGA D CCCA Q C C C C C C C C C C C C C C C C	GAC T ATA Y AGC A CGT V GAA N GCA R GAG C C C C C A C TCT L CCAA TCT TGT	CCT L TGG G CAG S GCC P TCT L CAA K GAC T TCT TAG S CTT TAG S GTT	TCA H GAA TTC S CCI L CCI I CCI C CCI C C TAA N CCI C C C C C C C C C C C C C C C C C	CCAA K CCAA K CCCC P CGCT T C CCGT C CCGT C CCGT C CCGT C CCGT C CCGT C CCGT C CCGT C CCGT C CCGT C CCGT C CCGT C CCGT C CCCC C C C	GCA Q ATC S CAT I GCA T CAC S CAC T GCA H GGC A CTG A CTG A CTG C T T T	GTA Y TAT I CTC S GAT I GAT I CCT I CCT A ACC A AAA ACCA CTC TGT	CCA H CTC S CCT L CCA K CCA K CCA C CCA C CCA K CCA C CCA C CCA C CCA C CCA C CCA C CCA C CCA C CCA C	CCTT L TAC T GGCC P GGC R GGC R GGC A GGC A C C C C C C C C C C C C C C	GGT V CAT M TCC P GAA K ATT F GGG G G C C GTA Y GAT I CTT ACC GCT AACC GCTA	CAA K GAA K GGT L TCT L TCT L TCT F CTA Y CGA CCC AGA TTT GAA GCT	GTC S AGT V CAC T GGT L GGA D TGG G CCA G CCA R CAG R AAT GGCC TGA CTT TTC	CCA H GAT M GACA R TGC A GGA E GGA E TGC A GGA TGC A AGA TGC AAGA	CAC T GGA Q CCT L CAG R CAG C T CAG C T CAG C T CAG C T CAG C T C CAG C T C CAG C T C CAG C T C CAG C T C CAG C T C CAG C T C C C T C C C C T C C C C C T C C C C C T C	CAA N GTT F TGA D CCAA N GCAA H TGC A CCAA N GCAA C GCAA M GCAA C GCAA C CCAA C CCAA C CCAA C CCAA C CAA N D C C CAA N C CAA C C C CAA N C CAA C C C C	CCAC T TCA Q CCT L CCT T CCAC T CCAC T CCAC T CCAC CCAC	CCAG S GGG G CCAC T V CCAT I GAG R CCAT I GAG R CCAT L CCAT L CCAT CCAT CCAT CCAT CCAT	CCA H BAAT M CCCC P GAA K TGA K TGA H CTT F GTA CTT F GGA CTT F GGA CTT T GGA CTT T CGGC	TGAT V CGAA K CCAG S GGAA K CTCT L CCCG R TCT V CCCG R CCCG R CCCG C CCCG C CCCG C CCCG C CCCG C CCCG C CCCG C CCCG C CCCA C CCAG C CCCG C C CCCG C C CCCG C C CCCG C C CCCG C C CCCG C C CCCG C C CCCG C C CCCG C C CCCG C C CCCG C C CCCG C C CCCG C C CCCG C C CCCG C C CCCG C C CCCG C C CCCG C C C CCCG C C C CCCG C C C CCCG C C C CCCG C C C C C CCCG C C C C CCCG C	900 269 900 289 1020 309 1080 349 1200 369 1260 389 1320 409 1380 429 1380 429 1380 429 1440 435
CAA K CAT M GCA H CCC P ATC S GTC S GTC S GTC S CAC T L AGT V ATA CAG CCA TTT TCT	AGA E GCA Q TGA TGA CCCA Q TGT V CCCA C C C C C C C C C C C C C C C C	GAC T ATA Y AGC A CGT V GAA N GCA C C C C C C C C C C C C C C C C C C	CCT L TGG G CAG S CAG T L CAA K GAC T TCT F TAG S AAC GTG GTG TTT AGG	TCA H GGAA TTC S CCT L CCT I CCAT I CCAT I CCAT N CCCG C C C C C C C C C C C C C C C C	CCAA K CCAA CCAA CCCA P CCCC T CCCC T CCCC C CCC C CCCC C CCCC C CCCC C CCCC C	GCA Q ATC S CAT I GCA Q CTC GCA T T CAC S CAC C C C C C C C C C C C C C C C	GTA Y TAT I CTC S CTT L GAT L CCT L CCT L AGC AAC CTC S AAC CTC S AAC CTC GTC	CCA H CTC S CCT L S GCA K CCA C CCA C CCA C CCA C C C C C C C	CCT L TAC T GGCC P GGC R GGC R GGC A C GGC T TCC TCC TCC TCC CGG CCAC CCAC	GGT V CAT M TCC P GAA K F GGG G CTG GTA Y GAT I CTT I CTT AACC GCTA	CAA K GAA K GGT L TCT L TCT F CTA T TT T GAA GCT CCA	GTC S AGT V CAC T GGA D GGA G GGA H CAG R CAG R AAT GGCA CTT TCA AGC	CCA H GAT M GACA H GGAG C C C C C C C C C C C C C C C C C	CAC T GGA Q CCT T CAG R AAC T CAG A CCT T L CAG A CCT T L CAG A CCT T L CAG A CCT T L CAG A C C C C	CAA N GTT F TGA D CCAA N GCAA H TGC A TGC A CCAA C C C C C C C C C C C C C C C	CAC T TCA Q CCT L CCT T CCA T CCA T CCA C CAC T C CAC C T C CAC C C C	CCAG S GGG G CCAC T V CCAT I GGAG R CCAT L CCAT L CCAT CCAT CCCAT CCCAT CCCAT CCCAT CCCAT CCCAT CCCAT CCCAT CCCAT CCCAT CCCAT CCCAT CCAC C CCAC C CCAC C CCAC C CCAC C CCAC C CCAC C CCAC C CCAC C CCAC C CCAC C CCAC C CCAC C CCAC C CCAC C CCAC C CCAC C CCAC C CCAC C CCAC C CCAC C CCAC C CCAC C CCAC C CCAC C CCAC C CCAC C CCAC C CCAC C CCAC C CCAC C CCAC C CCAC C CCAC C CCAC C CCAC C CCAC C CCCCAC C CCCCAC C CCCCAC C CCCCAC C CCCCAC C CCCCAC C CCCCAC C CCCCAC C CCCCAC C CCCCAC C CCCCAC CCCCAC CCCCAC CCCCAC CCCCAC CCCCCAC CCCCAC CCCCAC CCCCAC CCCCAC CCCCAC CCCCAC CCCCCAC CCCCCAC CCCCCAC CCCCCC	CCCA H HAAT M CCCCC P GGAA K TGAA K TGAA K CTT F GGAA CTT F GGAA CTT T GGGA TTT GGGC TTTT GGCG	TGAT V CGAA CCAG S CAGGA E GGAA K TCT L CCCG R CCCG R CCCA K CCCA K CCCA K CCCA CCAA C CCAA C CCCA C CCAA C CCAA C C CAA C C CAA C	900 265 960 285 1020 325 1080 349 1200 365 1240 409 1380 409 1380 425 1440 435 1560 1560 1680
CAA K CAT M GCA H CCC P ATC S GTC S GTC S GTC S GTC S GTC S CAC T L AGT V ATA CAG CA TTT TCAC	AGA E GCA Q TGA D CCA Q TGT T V AGA C C A GGC A GGC C C TGA GGT GTT C C TGA	GAC T ATA Y AGC A CGT V GAA N GCA C C GAA C C C C A C C C A C C C C A C C C C	CCT L TGG G CAG S GCC P TCT L CAA K GAC T TCT TCT L TAG S GAC TTGC GTGT GTGT GAC	TCA H GGAA TTC S CCI L CCI I CCI I CCI I CCI N TAA N CCI C C C C C C C C C C C C C C C C C	CCAA K CCAA CCAA CCCC P CCCC T CCCC T CCCC CCCC	GCA Q ATC S CAT I GCA Q TTC S CAC T TTC S CAC T CAC S CAC C C C C C C C C C C C C C C T C C C C	GTA Y TAT I CTC S CTT L GAT I CCT L AGC A ACC A ACC TGT GGG	CCA H CTC S CCT L CCA K CCA Q GCT L CCA Q CCA K CCA Q CCA C Q CCA K CCA C C C C C C C C C C C C C C C	CCT L TAC T GCC P GCC P GCC A GCC T CAC T CAC T CAC T C CCC A CAC C C CAC C C CAC C C CAC C C CAC C C C C C C C C C C C C C C C C C C C	GGT V CAT M TCC P GAA K F GGG G C C GTA C C C C C C C C C C C C C C C C C C C	CAA K GAA K GGT L GGT L CTA F CTA F CTA T GAA CCC AGA TTTA GAA GCTA AGA	GTC S AGT V CAC T GGA D CGA G GCA H CAG R CAG R AAT GGC TGA CTT TTC CCC	CCA H GAT M GACA H GGAG R GGAG C A GGA C A GGAG TGC A A GAG A GAG TGC A A GAG TGC TGC A C A GAG TGC A C A C A C A C A C A C A C A C A C A	CAC T GGA Q CCT L CAG R AAC T CAG A CTT L CAG A CTT L CAG A CTT L CAG C A CTT L CAG C A C CAG C A C C T C A GGC A C C T C CAG C A C C T C C A C C T C C A C C T C C A C C T C C A C C T C C A C C T C C A C C T C C A C C T C C A C C T C C A C C T C C A C C T C C A C C T C C A C C T C C A C C T C C C C	CAA N GTT F TGA D CCAP N GCA H TGC N GCA TGC N GCA R GCA C GCA C C C C C C C C C C C C C C C	CAC T TCA Q TCA Q CCT L CCT T CCA T CCA T CCA T C CCA T C CCA T C CCA T C CCA C C C T C CCA C C C C	CAG S S GGG G CAC T CAC T CAC T CAC T CAC T CAC T CAC T CAC T CAC T CAC T CAC T CAC T CAC T CAC T CAC C T CAC C T CAC C T CAC C T CAC C T CAC C T CAC C T CAC C T CAC C T CAC C C C	CCCA H HAAT M CCCCC P GGAA K TGAA K TGAA CTT F GGAA CTT F GGAA D CTT GGCA TTT TTT	V CGAA K CCAG S AGGA E GAA K TCT L CCCG R TCT V V CCCA K S CCCC R C CCCA K S CCCC C R C CCCA C R C CCCA C R C CCCA C R C CCCA C R C C C C	900 265 960 285 1020 309 1200 349 1200 369 1200 389 1320 409 1380 409 1380 409 1380 409 1380 409 1440 435 1500 1560 1620 1620 1620
CAA K CAT M GCA H CCCC P ATC S GTC S GTC S GTC S CAC T U ATA ATT TCT CAC	AGA E GCA Q TGA TGA CCAG R TGA Q TGA CCA Q CCA Q CCA Q CCA Q CCA Q CCA Q CCA Q CCA Q CCA Q CCA G C TGA C C C C C C C C C C C C C C C C C C C	GAC T ATA Y AGC A CGT V GAA H GCA R CTG C C A CTG C C A CTG T TGG TGG TGG TGG TGT T TAA	CCT L TGG G CAG S CAG TCT L CAA K GAC TCT TAG GTG GTG GTG GTG CAC TTT	TCA H GAA TTC S CCI L CCI I CCI I CCI I CCI I CCI C C C C	CCAA K CCAA K CCCC P CCCC T TGG G CCGT V V CCTG CCGT C CCGT C CCGT C CCGT C CCGT C CCGT C CCGT C CCGT C CCGT C CCGT C CCGT C CCGT C CCCC C C CCCC C C C	GCA Q ATC S CAT I GCA T CAT S CAC T CAC T CAC T CAC C CAC T CAC T CAC T CAC T CAC T CAT I CAT I CAT I CAT I CAT I CAT I CAT I CAT I CAT C T CAT C CAT I CAT C CAT I CAT C CAT I CAT C CAC C C C	GTA Y TAT I CTC S CTT L GAT L CCT L CCT L CCT S CCT S CCT C CCT C CCT T L AGC S CCT T T T T T T T T T T	CCA H CTC S CCT L CCA K CCA K CCA K CCA K CCA K CCA K CCA S CCT C CCA S CCT C C CCA C C C C C C C C C C C C C C	CCT L TAC T CGCC P GGCC P GGC R GGC R GGC A CAC T CCC T CCC T CCC T CCC CCC T CCC CCC	GGT V CAT M TCC P GAA K G GAA CTG GGG C GTA CTT I CTT ACC AAG CTT AACC AAG TTAT	CAA K GAA K GGT L TCT L CTA TCT F CTA TCT CTA CCA CCC AGA TTT GAA GCT CCC AGA TTT GAA CCCC	GTC S AGT V CAC T GGA D GGA C Q GCA G C Q GCA R C AGC R CAG R CAG C CAG C C G CAG C G C C C C C C C C	CCA H GAT M GAC R TGC R GGA E GGA E TGC A GGA I GAG I GAG TGC A AGA TGC A AGA TGC	CAC T GGA Q CCT L CAG R CAG R CAG C T CAG C T CAG C T CAG C T CAG C T CAG C T CAG C T C AAC T C AAC C T C AGCA T C AGCA T C CAG C A C C T C CAG C A C C T C CAG C A C C T C CAG C A C C T C CAG C A C C T C CAG C A C C T C CAG C A C C T C CAG C A C C T C CAG C A C C T C CAG C A C C T C CAG C A C C T C C C C T C C C C T C C C C T C C C C T C C C C T C C C C T C C C C C T C C C C T C C C C T C C C C T C C C C T C C C C C T C C C C T C C C C T C C C C T C C C C T C C C C T C C C C T C C C C C C T C C C C C T C C C C C T C C C C C T C C C C C C T C C C C C T C C C C C T C C C C C T C	CAA N GTT F D CAA D CAA N GCA H TGC A CAA CAA CAA CAA CAA CAA CAA CAA CAA	CCAC T TCA Q TCA Q CCT L CCT T CCGA T CCGA T CCGA T CCGA T CCGA CCCT CCGA CCCT CCGA CCCT CCGA CCGA	CAG S GGG G CAC T CAC T CAT I GAG R CCAT L CCAT L CCAT M CCAT CAT CCAT CCAT T CCAT T CCAT T CCAT T CCAT T CAC T T CAC T T CAC T T CAC T T CAC T T CAC C T CAC T T CAC T T CAC T CAC T CAC T CAC T CAC T CAC T CAC T CAC T CAC T CAC T CAC T CAC T CAC T CAC T C CAT T C C C C	CCCA H BAAT M CCCC P CACA K TGA K TGA K CTT F GGA CTT GGC TTT GGC TTT GGC TTTT AAA	V GAA K CAG S GAA K CAG K CCCC R CCCC CCC CCC CCC CCCC CCCC CC	900 269 900 289 1020 309 1080 349 1200 369 1200 429 1380 429 1380 429 1440 435 1500 1560 1680 1860

Figure 6 Nucleotide and deduced amino acid sequence of mouse legumain

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

N-phenylmaleimide, were inhibitory, but the more selective inhibitors of cysteine peptidases of the papain family, i.e. Z-Phe-Ala-CHN₂, E-64 and leupeptin, were without effect. General



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

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

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

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

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

DISCUSSION

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

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

The mouse legumain cDNA sequence consists of 1305 nucleotides encoding 435 amino acids with a putative leader peptide of 17 residues (Figure 6). On the basis of alignment with the mature pig legumain N-terminal sequence [7], Gly-28 is likely to be the first residue of the mature mouse enzyme. Despite the fact that the calculated molecular mass of the mouse legumain proenzyme is 47634 Da, a major band of 35 kDa was detected in both endogenous and recombinant mouse legumain (Figures 1, 2 and 7). This suggests that post-translational processing takes place during the biosynthesis of mouse legumain. Plant legumains [3] and also legumain from pig kidney are processed at both their N-and C-termini, so this is probably a general characteristic.

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

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

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

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

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

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

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

REFERENCES

- Takeda, O., Miura, Y., Mitta, M., Matsushita, H., Kato, I., Abe, Y., Yokosawa, H. and Ishii, S. (1994) J. Biochem. (Tokyo) **116**, 541–546
- 2 Kembhavi, A. A., Buttle, D. J., Knight, C. G. and Barrett, A. J. (1993) Arch. Biochem. Biophys. 303, 208–213
- 3 Hara-Nishimura, I., Takeuchi, Y. and Nishimura, M. (1993) Plant Cell 5, 1651–1659
- 4 El Meanawy, M. A., Aji, T., Phillips, N. F. B., Davis, R. E., Salata, R. A., Malhotra, I.,
- McClain, D., Aikawa, M. and Davis, A. H. (1990) Am. J. Trop. Med. Hyg. 43, 67–78
 Abe, Y., Shirane, K., Yokosawa, H., Matsushita, H., Mitta, M., Kato, I. and Ishii, S. (1993) J. Biol. Chem. 268, 3525–3529
- 6 Dalton, J. P., Hola-Jamriska, L. and Brindley, P. J. (1995) Parasitology **111**, 575–580
- 7 Chen, J. M., Dando, P. M., Rawlings, N. D., Brown, M. A., Young, N. E., Stevens, R. A., Hewitt, E., Watts, C. and Barrett, A. J. (1997) J. Biol. Chem. 272, 8090–8098
- 8 Anastasi, A., Brown, M. A., Kembhavi, A. A., Nicklin, M. J. H., Sayers, C. A., Sunter, D. C. and Barrett, A. J. (1983) Biochem. J. **211**, 129–138
- 9 Barrett, A. J. and Kirschke, H. (1981) Methods Enzymol. 80, 535-561
- 10 Rawlings, N. D. and Barrett, A. J. (1990) Comput. Appl. Biosci. 6, 118-119
- 11 Bradford, M. M. (1976) Anal. Biochem. 72, 248–254
- 12 Maunsbach, A. B. (1974) Methods Enzymol. 31, 330-339
- 13 Chen, J.-M., Changco, A., Brown, M. A. and Barrett, A. J. (1995) Exp. Cell Res. 216, 80-85
- 14 Chen, J.-M. and Little, C. D. (1985) J. Cell Biol. 101, 1175–1181
- 15 Laemmli, U. K. (1970) Nature (London) 227, 680–685
- 16 Towbin, H., Staehelin, T. and Gordon, J. (1979) Proc. Natl. Acad. Sci. U.S.A. 76, 4350–4354
- 17 Blake, M. S., Johnston, K. H., Russell-Jones, G. J. and Gotschlich, E. C. (1984) Anal. Biochem. **136**, 175–179
- 18 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) J. Mol. Biol. 215, 403–410
- 19 Genetics Computer Group (1994) Program Manual for the Wisconsin Package, Version 8, September 1994, University of Madison, Wisconsin
- 20 Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd edn., Cold Spring Harbor Laboratory Press, Cold Spring Harbor
- 21 Graham, F. L., Smiley, J., Russell, W. C. and Nairn, R. (1977) J. Gen. Virol. 36, 59–72
- 22 Gorman, C. M., Gies, D. R. and McCray, G. (1990) DNA Protein Eng. Tech. 2, 3-10
- 23 de Duve, C., Pressman, B. C., Gianetto, R., Wattiaux, R. and Appelmans, F. (1955) Biochem. J. 60, 605–617
- 24 D'Souza, S. E., Ginsberg, M. H. and Plow, E. F. (1991) Trends Biochem. Sci. 16, 246–250
- 25 Ruoslahti, E. and Pierschbacher, M. D. (1986) Cell 44, 517-518
- 26 Nakayama, K. (1997) Biochem. J. 327, 625-635
- 27 Lusson, J., Benjannet, S., Hamelin, J., Savaria, D., Chrétien, M. and Seidah, N. G. (1997) Biochem. J. **326**, 737–744
- 28 Rawlings, N. D. and Barrett, A. J. (1994) Methods Enzymol. 244, 461-486
- 29 Barrett, A. J. (1992) Ann. N. Y. Acad. Sci., 674, 1-15
- 30 Bohley, P. and Seglen, P. O. (1992) Experientia 48, 151-157
- 31 Berg, T., Gjøen, T. and Bakke, O. (1995) Biochem. J. 307, 313-326
- 32 Storrie, B. (1995) in Principles of Medical Biology, vol. 3 (Bittar, E. E. and Bittar, N., eds.), pp. 1–18, JAI Press, Greenwich, CT