Volume 326, number 1,2,3, 101–104 FEBS 12667 July 1993

Human cathepsin E produced in *E. coli*

Jeffrey Hill[a], Douglas S. Montgomery[b] and John Kay[a]

[a]*Department of Biochemistry, University of Wales College of Cardiff, Cardiff CF1 1ST, Wales, UK*

Received 26 April 1993

A cDNA for procathepsin E was generated from human gastric adenocarcinoma (AGS) cells, amplified by PCR and inserted into the T7 dependent vector pET 22b for expression in *E. coli*. Purification of the resultant product was accomplished simply, without the need to resort to column chromatography. The recombinant protein displayed comparable properties to those of its naturally occurring counterpart. The yield of homogeneous active enzyme obtained was ~ 3 mg per 40 g of cells. This is sufficient to permit crystallisation and structural analysis to begin and a mutagenesis programme to examine structure/activity relationships now to be undertaken.

Human cathepsin E; Cloning; PCR; Expression in *E. coli*; Recombinant cathepsin E; Purification; Characterisation; Chromogenic substrate hydrolysis; Inhibition

1. INTRODUCTION

The human body is known to contain five aspartic proteinases. Pepsin and gastricsin are both present in the stomach while the latter is also secreted from the prostate into seminal fluid [1,2]; renin is produced by a number of cells and tissues [3] while cathepsin D is a ubiquitous enzyme found in the lysosomes of most cells. The most recently characterised enzyme is cathepsin E [4], which has been implicated in the biogenesis of the vasoconstrictor peptide, endothelin [5] and to have a possible role in enabling infection of cells by the human immunodeficiency virus [6]. Recently, it has also been identified, for the first time unequivocally, to be the enzyme responsible for processing of a defined antigen (ovalbumin) in a defined lymphocyte system [7].

All five enzymes are synthesised in the form of zymogens. Production (of the precursors) of the secretory enzymes, pepsin and renin, and the membrane-encapsulated lysosomal cathepsin D in recombinant form [8–10] has greatly facilitated investigations into structure/function relationships in addition to providing sufficient material for 3-dimensional structural analysis by X-ray crystallography [11–13]. Cathepsin E is neither a secretory nor a lysosomal enzyme [7,14]; thus this intracellular aspartic proteinase is difficult to isolate in reasonable amounts. With its potential involvement in the processing reactions itemised above, it was considered of interest to produce cathepsin E in recombinant form to facilitate its molecular characterisation.

Correspondence address: J. Kay, Department of Biochemistry, University of Wales College of Cardiff, Cardiff CF1 1ST, Wales, UK. Fax: (44) (222) 87-4116.

2. MATERIALS AND METHODS

The human gastric adenocarcinoma AGS cell line (ATCC Designation, CRL 1739) which has been shown previously to express (pro)cathepsin E [15] was used as a source of total RNA. This was extracted using the lithium chloride precipitation based kit (RPN-1264, Amersham International plc, Bucks, UK). From this, messenger RNA was isolated using prepacked oligo(dT)-cellulose columns (Pharmacia Biotech. Ltd., Milton Keynes, UK) and cDNA was synthesised from 2 μg of mRNA using the cDNA Synthesis System Plus kit (Amersham International plc, Bucks, UK). Specific amplification of the procathepsin E cDNA was achieved using the PCR strategy described previously [16] utilising sense and antisense primers derived from the 5′-untranslated and 3′-coding regions of the human procathepsin E gene [15]. To facilitate cloning, these oligonucleotides (5′-GCGAGCATGCCATGGGATCCCTTCACAGGGTG-3′) and (5′-CACCGGTCGACTTAGGGGACTGCTGGGGC-3′) were designed to include *Nco*I and *Sal*I restriction sites respectively. Extraction and purification of the resulting 1,100 bp fragment and authentication of the oligonucleotide sequence by the dideoxy method was performed as described previously [16].

The *Nco*I–*Sal*I fragment spanning the coding sequence for procathepsin E, was subcloned into the T7 expression vector pET 22b (AMS Biotech. Ltd., Witney, Oxon, UK) and expression studies were carried out in the *E. coli* strain BL21 (DE3) pLysS obtained from the same source. Small scale cultures (10 ml) were induced by the addition of 1 PTG (to 0.4 mM) and, at appropriate time intervals, samples were removed and the medium was separated by centrifugation from the *E. coli* cells. The pelleted cells were resuspended in 50 mM Tris-HCl buffer, pH 7.2, containing 0.15 M NaCl (Buffer A) and lysed by one freeze–thaw cycle followed by the addition of 1.25 mg lysozyme. The supernatant was separated from insoluble material by centrifugation at 16,000 × *g* for 30 min. The latter was washed in Buffer A, resedimented and then dissolved in 6 M urea in 100 mM Tris-HCl buffer, pH 8.0, containing 1 mM glycine, 1 mM EDTA and 50 mM mercaptoethanol (Buffer B), as described by Lin et al. [8].

Aliquots of medium, supernatant and urea-solubilised material were analysed by SDS-PAGE with visualisation by Coomassie blue staining and by Western blotting using a polyclonal antiserum raised against a synthetic peptide corresponding to the C-terminal sequence of the human enzyme (kindly provided by Dr. Anthony Davenport, University of Cambridge, UK). Proteolytic activity was measured against haemoglobin as substrate [14] in 0.17 M sodium citrate buffer,

pH 3.1, when crude preparations needed to be evaluated; or utilising peptide substrates containing *p*-nitrophenylalanine (Nph) in the P_1 position as a chromogenic reporter group [17] when purer enzyme samples were available. Kinetic analyses with two of these peptide substrates (generously provided by Dr. Ben M. Dunn, University of Florida, USA) were performed as described previously [4] except that the pH used was pH 3.1 at a final ionic strength of 0.1 M. Initial velocities were measured with at least six concentrations of each peptide substrate within an appropriate range in order to derive the kinetic constants, K_m and V_{max}. The estimated error for all measurements was always < 15%. Values for k_{cat} were derived from $V_{max} = k_{cat} \cdot [E]$ where the active concentration of individual preparations of enzyme was determined by active site titration against pepstatin as described previously [18]. The protein inhibitor from *Ascaris lumbricoides* and the synthetic inhibitor H-77 described previously [4,7,17], were kindly provided by Dr. R.J. Peanasky, University of South Dakota, USA and Dr. P.A. Charlton (formerly) of Glaxo, Greenford, UK, respectively.

N-Terminal sequence analysis by Edman degradation was kindly performed by Dr. Benne Parten (University of Florida, Gainesville, FL, USA) using an ABI Model 470 A sequenator.

3. RESULTS AND DISCUSSION

Analysis of the cDNA obtained from the human AGS cell line as described in section 2 produced a nucleotide sequence in precise ageement with that reported for human procathepsin E by Azuma et al. [15]. By means of the *Nco*I–*Sal*I ends engineered into the fragment, it was introduced thereby into a number of expression vectors under the control of *tac* or λpL promoters and attempts were made to produce recombinant procathepsin E in *E. coli*, either in its authentic form or as a fusion with such proteins as chloramphenicol acetyl transferase and maltose-binding protein. In these preliminary experiments, little authentic material resulted, because of plasmid instability/product toxicity in the *E. coli* cells, and in the cases of the fusion constructs, the resultant proteins accumulating in *E. coli* were completely insoluble, thus negating the advantage conferred by the presence of the fusion partner in providing a simple affinity chromatography procedure to facilitate purification.

On this basis, the necessity for a high yield expression system that was under very tight transcriptional control was indicated and, at this time, pET 22b, a supposedly secretory vector exhibiting just such characteristics, became available for use. In preliminary time course experiments on a 10 ml scale utilising cultures transformed with the pET 22b plasmid harbouring the procathepsin E gene, very little activity or immunoreactive material was detected in the medium or periplasmic space. However, a strongly positive band at approximately 43 kDa was observed (Fig. 1), when whole cell lysates (prepared in SDS) were examined by Western blotting with the anti-cathepsin E antiserum. This reached a maximum two hours after addition of the IPTG inducer. Separation of cell lysates into soluble and insoluble fractions indicated that the majority of the immunoreactive material was insoluble (Fig. 1).

Consequently, for larger scale purification, cultures (8 litres) were induced and after 2 h, the cells were harvested and lysed. The insoluble material was recovered, washed with Buffer A and solubilised in the 6 M urea-containing Buffer B (see section 2). Previous re-

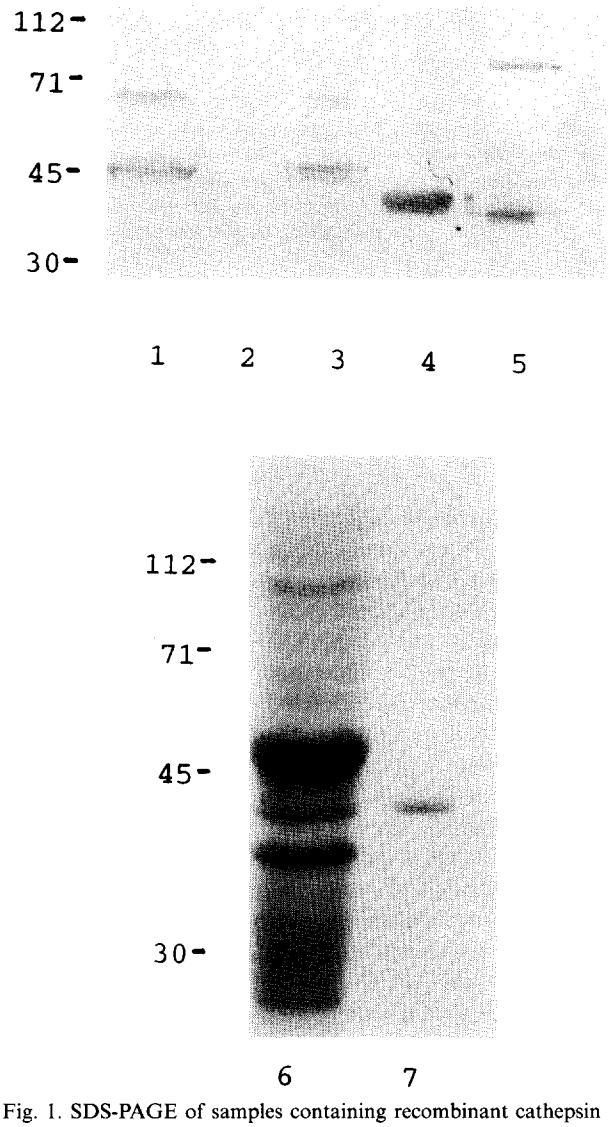

Fig. 1. SDS-PAGE of samples containing recombinant cathepsin E. (Upper panel) Lysate of *E. coli* cells harbouring the procathepsin E-containing plasmid (lane 1); soluble and insoluble fractions prepared therefrom (lanes 2 and 3, respectively); material obtained after pH 3.1 treatment, neutralisation and centrifugation (lanes 4 and 5). SDS-PAGE was performed under reducing conditions (in the presence of 100 mM mercaptoethanol) for lanes 1–4 and under non-reducing conditions (in the absence of mercaptoethanol) in lane 5. Visualisation after blotting onto nitrocellulose membrane was with an anti-cathepsin E antiserum. (The band at ~ 65 kDa in lanes 1 and 3 is an *E. coli* component that cross-reacts with this antiserum; this band was observed also on blots of lysates from cells harbouring a control plasmid without the procathepsin E insert). (Lower panel) Aliquots of the urea-solubilised pellet, that had been diluted forty-fold and reconcentrated, were analysed before (lane 6) and after (lane 7) acidification to pH 3.1 and re-neutralisation. SDS-PAGE gels were run under reducing conditions and stained with Coomassie blue. Mol. wt. markers (kDa) migrated as indicated.

ports [8,10] have demonstrated that the homologous human (pro)enzymes, pepsinogen and procathepsin D, also accumulate in inclusion bodies in *E. coli* but dissolve readily in 6 M urea and can be renatured therefrom by forty-fold dilution of the denaturant. This procedure was thus applied to the material solubilised by 6 M urea and the resultant solution was then re-concentrated by ultrafiltration in an Amicon stirred cell fitted with a PM-30 membrane. Proteolytic activity towards haemoglobin as substrate at pH 3.1 was readily measured in the concentrate and since this was completely inhibited by inclusion of 100 nM pepstatin in the assays, it must be attributed to the action of an aspartic proteinase. Since *E. coli* is not known to produce this type of proteinase, it would thus appear that active, recombinant (pro)cathepsin E had been generated.

Numerous attempts were made to purify this material with potential activity by successive chromatographic procedures including DEAE-cellulose and FPLC on Mono Q. However, the (pro)cathepsin E was always eluted in comparable positions to the residual *E. coli* proteins. Therefore, a simpler procedure was devised. The partially-purified material, after solubilisation with 6 M urea, re-folding by 40-fold dilution, and concentration by ultrafiltration as described above, was acidified to pH 3.1 by the addition of 1 M sodium citrate buffer. After incubation for 10 min at 25°C, the material was centrifuged and the supernatant was re-neutralised to pH 8.2. This solution was centrifuged again. Analysis of the resultant supernatant indicated that all of the activity (100%) towards haemoglobin as substrate had remained in solution, whereas all of the undesired protein had precipitated and had been removed by centrifugation (Fig. 1; compare lanes 6 and 7). By this simple purification scheme, homogeneous material had been generated without the need to resort to column chromatography.

Analysis by SDS-PAGE under reducing conditions, followed by staining with Coomassie blue revealed a single band of approximately 38 kDa (Fig. 1; lower panel). When the sample was electrophoresed under non-reducing conditions, bands of approximately 38 and 76 kDa reacted positively on Western blots with the antiserum to cathepsin E (Fig. 1; upper panel). This tendency to form dimers linked by an inter-chain disulphide bond has been observed previously for naturally-occurring cathepsin E. Samples of the concentrated recombinant enzyme were sequenced after electroblotting onto PVDF membrane for direct introduction into the sequenator. Through 25 cycles, the following unique sequence was obtained:

Thr-Glu-Ser-Cys-Met-Asp-Gln-Ser-Ala-Lys-Glu-Pro-Leu-Ile-Asn-Tyr-Leu-Asp-Met-Glu-Tyr-Phe-Gly-Thr~

This is exactly coincident with that predicted for mature cathepsin E from the nucleotide sequence [15]. Depending on its source of origin (stomach or red blood cells), naturally-occurring, mature cathepsin E has been shown to be heterogeneous [19,20] at its N-terminus, consisting of two closely similar isozymes [4] which differ by only three residues at their respective N-termini, viz.

Ile-Gln-Phe-Thr-Glu-Ser-Cys~

and

Thr-Glu-Ser-Cys~

It is evident that the acidification procedure used in the purification of the recombinant protein had resulted in activation of the 43 kDa proform to generate the shorter (38 kDa) of these two isozymes of mature cathepsin E in a homogeneous form.

The concentration of active enzyme in each preparation was determined by active site titration using isovaleryl-pepstatin [18]. On this basis, it was estimated that the yield of homogeneous, active enzyme was approximately 3,000 μg from an 8 litre culture. This amount is substantially in excess of any material that can be purified from naturally-occurring human cells or tissues.

The ability of the recombinant enzyme to hydrolyse two chromogenic substrates was examined and kinetic parameters (K_m, k_{cat}) for their hydrolysis are listed in Table I. Both have been demonstrated previously to be excellent substrates for naturally-occurring cathepsin E from both gastric and erythrocyte sources [4]. The values derived for the recombinant enzyme were directly comparable to those kinetic parameters measured previously for the naturally occurring enzymes. We have also shown previously that one feature which distinguishes cathepsin E from its lysosomal counterpart, cathepsin D, is its susceptibility to inhibition by a 17 kDa protein

Table I

Kinetic constants for the interactions of chromogenic substrates and inhibitors with recombinant human cathepsin E

Substrate or inhibitor	K_i (nM)	K_m (μM)	k_{cat} (s^{-1})
1	–	65	120
2	–	50	115
Ascaris protein	5	–	–
H-77	150	–	–
Isovaleryl-pepstatin	0.3	–	–

All measurements were carried out in 0.1 M sodium formate buffer, pH 3.1, at 37°C.

Substrate 1 = Pro-Pro-Thr-Ile-Phe*Nph-Arg-Leu
Substrate 2 = Lys-Pro-Ile-Glu-Phe*Nph-Arg-Leu

The latter was used for evaluation of the inhibitors, which were all competitive in nature. The estimated precision of all values obtained was in the range 10–15%.

inhibitor from the parasitic worm, *Ascaris lumbricoides* [4,7,14,18]. Consequently, an inhibition constant (K_i) was determined for the interaction of this protein with the recombinant enzyme. The value obtained (Table I) was essentially identical to those determined previously [14,18] for naturally-occurring cathepsin E from different tissues/cells/species.

This comparability between the recombinant enzyme and its naturally-occurring counterpart was substantiated further by evaluation of the interactions of two low molecular weight inhibitors of widely different potencies. The synthetic inhibitor H77 (= D-His-Pro-Phe-His-Leu-ψ[CH_2-NH]-Leu-Val-Tyr), which is much less potent than the *Ascaris* inhibitor, was comparably effective against the recombinant enzyme (Table I) to the K_i reported earlier against cathepsin E of natural origin [14]. Isovaleryl pepstatin was a very tight-binding inhibitor (Table I), as might be expected towards an aspartic proteinase, such as cathepsin E [14].

It would appear that the recombinant cathepsin E thus produced was comparable in its behaviour to the enzyme isolated from natural sources. The latter, however, is notorious for its heterogeneity: (a) at its N-terminus because of the presence of two scissile peptide bonds that are readily cleaved autocatalytically during activation of the precursor under acidic conditions [20]; (b) in being a glycoprotein [14,19]; (c) in having an 'additional' cysteine residue (compared to other archetypal aspartic proteinases) close to the N-terminus (residue 4) which forms an interchain disulphide bond so that an equilibrium mixture of monomeric and dimeric enzymes results. These facets, together with the paucity of available material, have not permitted crystallisation and structural determination to be undertaken so far. However, with this demonstration of the generation of large quantities of homogeneous cathepsin E in, and its facile purification from *E. coli*, these two difficulties have been surmounted. A programme to replace the cysteine residue (with Ser) together with mutagenesis of other appropriate residues is now underway to address this situation and to enable evaluation of structure/function relationships in this important enzyme.

Acknowledgements: This work was supported by a CASE studentship (to J.H.) from the UK Science and Engineering Research Council. We are most grateful to Drs. Callum Campbell, Christine Grinham and Colin Dykes, in Glaxo; Ray Jupp (formerly of this Department), Ben Dunn and Benne Parten, University of Florida; and Anthony Davenport, University of Cambridge, for their valuable contributions to this research.

REFERENCES

[1] Kay, J. (1985) in: Aspartic Proteinases and Their Inhibitors (Kostka, V. ed.) pp. 1–17, Walter de Gruyter, Berlin.
[2] Reid, W.A., Vongsorasak, L., Svasti, J., Valler, M.J. and Kay, J. (1984) Cell Tissue Res. 236, 597–600.
[3] Dhanaraj, V., Dealwis, C.G., Frazao, C., Badasso, M., Sibanda, B.L., Tickle, I.J., Cooper, J.B., Driessen, H.P.C., Newman, M., Aguilar, C., Wood, S.P., Blundell, T.L., Hobart, P.M., Geoghegan, K.F., Ammirati, M.J., Danley, D.E., O'Connor, B.A. and Hoover, D.J. (1992) Nature 357, 466–472.
[4] Jupp, R.A., Richards, A.D., Kay, J., Dunn, B.M., Wyckoff, J.B., Samloff, I.M. and Yamamoto, K. (1988) Biochem. J. 254, 895–898.
[5] Lees, W.E., Kalinka, S., Meech, J.D., Capper, S.J., Cook, N.D. and Kay, J. (1990) Biochem. J. 273, 99–102.
[6] Clements, G.J., Price-Jones, M.J., Stephens, P.E., Sutton, C., Schulz, T.F., Clapham, P.R., McKeating, J.A., McClure, M.O., Thomson, S., Marsh, M., Kay, J., Weiss, R.A. and Moore, J.P. (1991) AIDS Res. 7, 3–16.
[7] Bennett, K., Levine, T., Ellis, J.S., Peanasky, R.J., Samloff, I.M., Kay, J. and Chain, B.M. (1992) Eur. J. Immunol. 22, 1519–1524.
[8] Lin, X., Wong, R.N.S. and Tang, J. (1989) J. Biol. Chem. 264, 4483–4489.
[9] Harrison, T.M., Chidgey, M.A.J., Brammar, W.J. and Adams, G.J. (1989) Proteins 5, 259–265.
[10] Conner, G.E. and Richo, G. (1992) Biochemistry 31, 1142–1147.
[11] Sielecki, A.R., Fedorov, A.A., Boodhoo, A., Andreeva, N.S. and James, M.N.G. (1990) J. Mol. Biol. 214, 143–170.
[12] Sielecki, A.R., Hayakawa, K., Fujinaga, M., Murphy, M.E.P., Fraser, M., Muir, A.K., Carilli, C.T., Lewicki, J.A., Baxter, J.D. and James, M.N.G. (1989) Science 243, 1346–1351.
[13] Scarborough, P.E., Guruprasad, K., Topham, C., Richo, G., Conner, G.E., Blundell, T.L. and Dunn, B.M. (1992) Protein Sci. 2, 264–276.
[14] Samloff, I.M., Taggart, R.T., Shiraishi, T., Branch, T., Reid, W.A., Heath, R., Lewis, R.W., Valler, M.J. and Kay, J. (1987) Gastroenterology 93, 77–84.
[15] Azuma, T., Pals, G., Mohandas, T.K., Couvreur, J.M. and Taggart, R.T. (1989) J. Biol. Chem. 264, 16748–16753.
[16] Kille, P., Stephens, P.E. and Kay, J. (1991) Biochim. Biophys. Acta 1089, 407–410.
[17] Kay, J. and Dunn, B.M. (1992) Scand. J. Clin. Lab. Invest. 52 (Suppl. 210), 23–30.
[18] Valler, M.J., Kay, J., Aoyagi, T. and Dunn, B.M. (1985) J. Enzyme Inhib. 1, 77–82.
[19] Athauda, S.B.P., Matsuzaki, O., Kageyama, T. and Takahashi, K. (1990) Biochem. Biophys. Res. Commun. 168, 878–885.
[20] Athauda, S.B.P., Takahashi, T., Kageyama, T. and Takahashi, K. (1991) Biochem. Biophys. Res. Commun. 175, 152–158.