Isolation and sequence of a cDNA for human pro-(cathepsin L)

Susannah GAL* and Michael M. GOTTESMAN†

Laboratory of Molecular Biology, National Cancer Institute, National Institutes of Health, Building 37, Room 2E18, Bethesda, MD 20892, U.S.A.

The major excreted protein (MEP) of malignantly transformed mouse fibroblasts is the precursor to an acid proteinase with enzymic specificity similar to that of human cathepsin L. By cross-hybridization with a mouse MEP sequence, cDNA clones of the human form of MEP in an SV40 expression vector were isolated. A 1.6 kb cDNA showed 70 % deduced amino acid sequence identity with mouse MEP. The deduced amino acid sequence of the cloned human MEP was the same, except for two amino acids, as the *N*-terminal sequence of mature human cathepsin L, thereby establishing that human MEP is human pro-(cathepsin L). Use of this human pro-(cathepsin L) cDNA clone allowed the detection of a 1.6–1.8 kb pro-(cathepsin L) mRNA in human cells which was not detected with a mouse pro-(cathepsin L) probe.

INTRODUCTION

Mouse fibroblasts which are malignantly transformed or stimulated by growth factors or tumour promoters synthesize and secrete increased amounts of a 39 kDa glycoprotein with acid-proteinase activity (Gottesman, 1978; Gottesman & Sobel, 1980; Doherty *et al.*, 1985; Rabin *et al.*, 1986; Frick *et al.*, 1985; Nilson-Hamilton *et al.*, 1981; Gal & Gottesman, 1986a). This protein, termed 'MEP' (major excreted protein), is the precursor to two lower-*M_r* lysosomal proteins of 29 kDa and 21 kDa (Gal *et al.*, 1985), and contains the lysosomal recognition marker mannose 6-phosphate (Sahagian & Gottesman, 1982). Mouse MEP and human cathepsin L (Mason *et al.*, 1985) share amino acid bond cleavage specificities, catalytic constants and inhibitor susceptibilities (Gal & Gottesman, 1986b; Mason *et al.*, 1987). Comparison of the deduced amino acid sequences of mouse MEP, or a related protein called 'mouse cysteine proteinase', with a partial amino acid sequence for human cathepsin L, indicates strong similarity between these two sequences, but because of the species differences, does not prove identity (Troen *et al.*, 1987; Portnoy *et al.*, 1986; Denhardt *et al.*, 1986; Mason *et al.*, 1986).

In the present paper, we report the isolation and characterization of a full-length human MEP cDNA clone, isolated by cross-hybridization with a mouse MEP cDNA clone. The deduced amino acid sequence from this clone was identical, except for two amino acids, with that of the human cathepsin L sequence, proving that MEP is cathepsin L. This human cathepsin L cDNA can be used as a probe to determine the expression of cathepsin L mRNA in human cell lines and tissues.

MATERIALS AND METHODS

Isolation of a human cDNA clone for cathepsin L

The Okayama–Berg (1983) cDNA expression library termed 'GM637', made from mRNA isolated from SV40-virus-transformed human fibroblasts, was a gift

from Dr. Hirota Okayama (National Institute of Mental Health, National Institutes of Health, Bethesda, MD, U.S.A.). In all, 100 000 clones were screened on 20 filters by the technique of Troen (1987) for colony hybridization, using the 800-bp restriction-*Eco*RI fragment of mouse MEP (Troen *et al.*, 1987) labelled by nick translation (Lofstrand Laboratories, Gaithersburg, MD, U.S.A.). The filters were washed under moderate stringency conditions: twice with $2 \times \text{SSC}$ ($1 \times \text{SSC} = 0.15 \text{ M-NaCl}/0.015 \text{ M-sodium citrate}$)/1 % SDS and twice with $0.8 \times \text{SSC}/1 \%$ SDS at 60 °C, and three positive clones were obtained. The clone designated 'pHu-16' contained the largest insert (1.6 kb) and was used for all subsequent studies.

Sequencing and blotting

Sequencing was done using the Sanger *et al.* 1977) dideoxy method with the Promega sequencing kit (Promega Biotec, Madison, WI, U.S.A.) and deoxyadenosine [α - ^{35}S]thiotriphosphate at a specific radioactivity of 500 Ci/mmol (NEN, Boston, MA, U.S.A.). pHu-16 was sequenced by using a combination of two techniques: (1) subcloning regions of pHu-16 DNA into Promega vectors pGEM-3 and pGEM-4 with Sp6 and T7 primers (Promega), and (2) direct sequencing using 50 ng of synthetic oligonucleotides (Applied Biosystems) as primers and the entire pHu-16 cDNA as template.

Detection of human procathepsin L RNA

RNA was isolated from tissue-culture cells as previously described (Chirgwin *et al.*, 1979; Maniatis *et al.*, 1982), and Northern blots were as described by Shen *et al.* (1986). Northern blots were probed with nick-translated *Eco*RI fragments from pHu-16 cDNA or pMMEP-14 cDNA (Troen *et al.*, 1987, 1988).

RESULTS AND DISCUSSION

Isolation and sequence of a human cathepsin L cDNA

We selected the human MEP clone by screening a cDNA library made from SV40-virus-transformed

Abbreviations used: MEP, major excreted protein.

* Present address: Friedrich-Miescher-Institut, P.O. Box 2543, CH-4002 Basel, Switzerland.

† To whom correspondence and reprint requests should be addressed.

Fig. 1. (a) Sequencing strategy for pHu-16 and (b) the entire sequence of pHu-16 with the longest open reading frame translated underneath beginning at nucleotide 305 and terminating at nucleotide 1304

(a) Bold lines are sequences determined using subcloned fragments in pGEM3 or pGEM4 primed with Sp6 and T7 oligonucleotides. The dotted lines indicate sequences derived from the entire pHu-16 plasmid using synthetic oligonucleotides (*) as primers. The orientation of the clone is shown along with some of the restriction sites; V = *PvuII*, T = *TaqI*, E = *EcoRI*, R = *RsaI*, S = *SphI*. (b) The poly(A) addition signal is underlined.

human fibroblasts at low stringency using an 800-bp *EcoRI* fragment from a mouse MEP cDNA clone (see the Materials and methods section). One clone, termed 'pHu-16', contained a 1600-bp cDNA insert. The restriction map, sequencing strategy, complete nucleotide sequence and deduced amino acid sequence for pHu-16 are shown in Fig. 1.

The sequence of the human MEP cDNA revealed a single long open reading frame coding for a protein of M_r 37522. The deduced amino acid sequence was compared with the mouse MEP sequence (Fig. 2) and

Fig. 2. Homology of mouse and human MEP sequences

Complete amino acid sequences deduced for pHu-16 (human MEP) and pMMEP-14 [mouse MEP (Troen *et al.*, 1987)] are shown. The underlined sequences of human MEP correspond to protein sequences of human cathepsin L obtained by Mason *et al.* (1986). The underlined sequence of mouse MEP corresponds to the N-terminal sequence of secreted MEP purified from mouse tissue-culture medium (Troen *et al.*, 1987). Arrows indicate where processing occurs, and asterisks indicate potential N-glycosylation sites. The proposed active-site amino acids based on papain, Cys-138 and His-276, are shown by @ and # respectively. Homology is indicated as exact matches (|) and conservative substitutions (:).

was found to be 70% identical over the entire sequence, with some regions having 90% identity (e.g. between amino acids 115 and 168 and between amino acids 217 and 249). Over 98% sequence identity was found between the N-terminal amino acid sequence of mature human cathepsin L (Mason *et al.*, 1986) and human MEP (shown as the underlined sequence in Fig. 2). The two amino acids which are different from our predicted sequence and that published by Mason *et al.* (Pro for Glu at position 148 and Tyr for Arg at position 153) may be due to problems with the amino acid analysis or to polymorphic differences in the proteins isolated from two different human sources. These results show that MEP is pro-(cathepsin L). Similar conclusions have been reached by Joseph *et al.* (1987), who reported the partial sequence of the human analogue of a *ras*-oncogene-induced protein which appears to be identical with MEP/cathepsin L.

Fig. 3. Northern blots of RNA from tissue-culture cells and human tissues

(a) and (b), Northern blots of 10 μ g of total RNA from tissue-culture cells using two different probes. The sources of RNA were human KB cells (a), multidrug-resistant KB cells (b), mouse NIH 3T3 cells (c) and multidrug-resistant NIH 3T3 cells (d) as previously described (Shen *et al.*, 1986). In (a) hybridization was with the human pHu-16 *Eco*RI fragment and in (b) hybridization was with the *Eco*RI fragment from the mouse MEP clone pMMEP-14 (Troen *et al.*, 1987).

As for mouse MEP and mouse cysteine proteinase (Troen *et al.*, 1987; Portnoy *et al.*, 1986; Denhardt *et al.*, 1986), significant homologies were also observed when the human MEP sequence was compared with those of other cysteine proteinases, including cathepsin B, cathepsin H, actinidin and papain. Mouse and human forms of MEP show nearly identical sequences around amino acids Cys-138 and His-276 (Fig. 2), which are putative active sites on the basis of similarities to papain (Kamphuis *et al.*, 1985). Comparison of our sequence with those of several other human lysosomal proteins shows no obvious conserved signals for the lysosomal-protein-specific transfer of UDP-*N*-acetylglucosamine to the high-mannose chain to yield the lysosomal mannose 6-phosphate marker.

The *N*-terminal sequence of secreted mouse MEP indicates where pre-MEP is processed to yield MEP (Troen *et al.*, 1987). The deduced amino acid sequence of human prepro-(cathepsin L) also has the canonical Ala-Xaa-Ala sequence which precedes the leader-sequence cleavage site (Perlman & Halvorson, 1983) (Fig. 2). The positions of the *N*-terminal peptides of mature human cathepsin L within the deduced amino acid sequence of human MEP indicate the sites at which this precursor to human cathepsin L is processed to the lower M_r mature forms.

Because the full-length pro-(cathepsin L) cDNA which we isolated is cloned downstream from the SV40 promoter (Okayama & Berg, 1983), it is possible to express this cDNA in transfected cells. Recent experiments indicate that the human pro-(cathepsin L) cDNA is expressed in mouse NIH 3T3 cells as a 42 kDa protein that is secreted when it is overproduced (Kane *et al.*, 1988).

Use of the human procathepsin L cDNA to detect a procathepsin L mRNA in human cells

Fig. 3 shows the results of Northern blots with human KB (HeLa) cell and NIH 3T3 mouse-cell RNA probed with the *Eco*RI 800-bp fragment from the mouse and human MEP cDNA clones. The human probe recognized a 1.6–1.8-kb pro-(cathepsin L) message in human KB cells which migrated just below the 18 S RNA marker (Fig. 3a) and appeared to be the same size as MEP mRNA from mouse cells (Fig. 3b). At high stringency, the mouse probe recognized a 1.6–1.8-kb RNA in mouse cells but no message in human cells (Fig. 3b). These data indicate that the human pro-(cathepsin L) cDNA described here (pHu-16) can be used to detect a human pro-(cathepsin L) mRNA, and that there is only one major pro-(cathepsin L) mRNA in cultured human KB cells. Preliminary results indicate that all normal human tissues tested express this pro-(cathepsin L) mRNA (results not shown). The availability of a full-length cDNA probe for human pro-(cathepsin L) will make it possible to assess relative levels of pro-(cathepsin L) mRNA in different tissues and in tumours.

We thank Dr. D.-w. Shen for RNA samples, Dr. H. Okayama for the human cDNA library, Dr. B. Troen, Mr. S. Smith and Dr. S. Kane for helpful discussions, Mr. S. Neal for photographic assistance, and Ms. J. Sharrar for secretarial help.

REFERENCES

- Chirgwin, J., Prezbyla, A., MacDonald, R. & Rutter, W. J. (1979) *Biochemistry* **18**, 5294–5299
- Denhardt, C. T., Hamilton, R. T., Parfett, C. L. F., Edwards, D. R., St. Pierre, R., Waterhouse, P. & Nilsen-Hamilton, M. (1986) *Cancer Res.* **46**, 4590–4593
- Doherty, P. J., Hua, L., Liau, G., Gal, S., Graham, D., Sobel, M. & Gottesman, M. M. (1985) *Mol. Cell. Biol.* **5**, 466–473
- Frick, K. K., Doherty, P. J., Gottesman, M. M. & Scher, C. D. (1985) *Mol. Cell. Biol.* **5**, 2582–2589
- Gal, S. & Gottesman, M. M. (1986a) *J. Biol. Chem.* **261**, 1760–1765
- Gal, S. & Gottesman, M. M. (1986b) *Biochem. Biophys. Res. Commun.* **139**, 156–162
- Gal, S., Willingham, M. C. & Gottesman, M. M. (1985) *J. Cell Biol.* **100**, 535–544
- Gottesman, M. M. (1978) *Proc. Natl. Acad. Sci. U.S.A.* **75**, 2767–2771
- Gottesman, M. M. & Sobel, M. E. (1980) *Cell* (Cambridge, Mass.) **19**, 449–455
- Joseph, L., Lapid, S. & Sukhatme, V. (1987) *Nucleic Acids Res.* **15**, 3186
- Kamphuis, I. G., Drenth, J. & Baker, E. N. (1985) *J. Mol. Biol.* **182**, 317–329
- Kane, S. E., Troen, B. R., Gal, S., Ueda, K., Pastan, I. & Gottesman, M. M. (1988) *Mol. Cell. Biol.*, in the press

- Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) in *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York
- Mason, R. W., Green, G. D. J. & Barrett, A. J. (1985) *Biochem. J.* **226**, 233–241
- Mason, R. W., Walker, J. E. & Northrop, F. D. (1986) *Biochem. J.* **240**, 373–377
- Mason, R. W., Gal, S. & Gottesman, M. M. (1987) *Biochem. J.* **248**, 440–454
- Nilsen-Hamilton, M., Hamilton, R. T., Allen, W. R. & Mas-soglia, S. L. (1981) *Biochem. Biophys. Res. Commun.* **101**, 411–417
- Okayama, H. & Berg, P. (1983) *Mol. Cell. Biol.* **3**, 280–289
- Perlman, D. & Halvorson, H. O. (1983) *J. Mol. Biol.* **167**, 391–409
- Portnoy, D. A., Erikson, A. H., Kochan, J., Ravetch, J. V. & Unkless, J. C. (1986) *J. Biol. Chem.* **261**, 14697–14703
- Rabin, M. S., Doherty, P. J. & Gottesman, M. M. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 357–360
- Sahagian, G. G. & Gottesman, M. M. (1982) *J. Biol. Chem.* **257**, 11145–11150
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467
- Shen, D.-w., Fojo, A., Roninson, I. B., Chin, J. E., Soffir, R., Pastan, I. & Gottesman, M. M. (1986) *Mol. Cell. Biol.* **6**, 4039–4044
- Troen, B. R. (1987) *Methods Enzymol.* **151**, 416–426
- Troen, B. R., Gal, S. & Gottesman, M. M. (1987) *Biochem. J.* **246**, 731–735
- Troen, B. R., Ascherman, D., Atlas, D. & Gottesman, M. M. (1988) *J. Biol. Chem.* **263**, 254–261

Received 10 February 1988/6 April 1988; accepted 25 April 1988