Complete Nucleotide and Deduced Amino Acid Sequences of Human and Murine Preprocathepsin L

An Abundant Transcript Induced by Transformation of Fibroblasts

Loren J. Joseph,* Louise C. Chang, Dorothy Stamenkovich, and Vikas P. Sukhatme
*The Howard Hughes Medical Institute, Departments of Medicine and *Pathology, University of Chicago, Chicago, Illinois 60637*

Abstract

Transfection of an activated ras oncogene into NIH3T3 fibroblasts leads to transformation and induction of a metastatic phenotype. To identify genes whose activation might mediate these processes, we used a differential screening strategy. A 1.5-kb transcript is induced fiftyfold, constitutes 1% of ras transformed cell messenger RNA (mRNA) and is the most abundantly induced message in these cells. Our sequence data shows that it encodes murine cathepsin L, a potent collagenolytic and elastinolytic lysosomal enzyme. The murine clone was used to isolate human cathepsin L complementary DNA (cDNA) clones. The complete nucleotide and deduced amino acid sequences of human and murine preprocathepsin L are presented and compared to other papain family cysteine proteinases. Northern analysis shows that both human and murine cathepsin L probes hybridize to a 1.5-kb transcript in several tissues, but also to a 4-kb transcript in human kidney. These clones will facilitate studies of the structure, expression, and function of cathepsin L, including its unexpected upregulation in transformation.

Introduction

Differential screening methodology has seen limited but increasing application to cancer biology (1, 2). For example, Matrisian et al. (3) identified a metalloproteinase secreted by polyoma-transformed fibroblasts but not by the parent line. Most recently Hanafusa et al. identified a novel protein whose expression is elevated after Rous sarcoma virus transformation of chick embryo fibroblasts (4).

We applied differential screening to a cDNA library made from Ha-ras (activated) transformed NIH 3T3 fibroblasts to identify genes that show altered levels of mRNA relative to nontransformed cells. Our choice of the ras system was based on two considerations: (*a*) there is substantial work relating tumorigenicity and metastatic potential to the level of ras expression (5–9); (*b*) ras expression occurs in several human malignancies (prostate, colon, breast, leukemic cell lines (10–12). It was our expectation that some of the cDNAs isolated using the NIH3T3-ras model system would encode products that might play a role not only in tumorigenesis but also in tumor progression, an area of clinical importance in which only limited advances have been made at the molecular level. We also hypothesized that expression of one or more ras-induced transcripts might correlate with metastatic potential in human cancers where ras is thought to play a causal role.

In this paper we report on the most abundant transcript induced by ras transformation. It predicts a protein of 334 amino acids with marked similarity to the papain cysteine proteinase superfamily (13). This murine sequence has also recently been reported as an activated mouse macrophage cysteine proteinase by Portnoy et al. (14), who, along with Mason et al. (15), suggested that it could be identical to the major excreted protein (MEP)[1] of transformed murine fibroblasts and to cathepsin L. The definitive identification of this sequence as murine cathepsin L was established by isolation and sequencing of the corresponding clone from a human kidney cDNA library. Partial sequencing of the human clone (16) predicted the partial amino acid sequence of human cathepsin L (HCL) recently reported by Mason et al. (15). (Amino acid sequence for the murine cathepsin L is not available.)

Cathepsin L is a lysosomal cysteine proteinase with a major role in intracellular protein catabolism (17, 18). Cathepsin L also shows the most potent collagenolytic and elastinolytic activity in vitro of any of the cathepsins (19, 20). Most recently human cathepsin L has been shown to proteolytically inactivate alpha-1 protease inhibitor, a major controlling element of human neutrophil elastase activity in vivo (21). Cathepsin L has been implicated in pathological processes including myofibril necrosis in myopathies (22) and in myocardial ischemia (23), and in the renal tubular response to proteinuria (24).

We also show the first complete nucleotide and predicted amino acid sequence for human preprocathepsin L, compare our human and murine sequences with those of other cysteine proteinases, provide the first demonstration of cathepsin L mRNA in a human tumor, and show evidence for a higher molecular weight human kidney transcript.

Methods

Cell culture. NIH3T3 mouse fibroblasts were kindly provided by S. Schwartz, and Kirsten ras transformed cells by Robert Bassin. All cells were maintained in Dulbecco's modified Eagle's medium (DMEM) (Gibco Laboratories, Grand Island, NY) with 10% fetal calf serum at 37°C in a humidified incubator containing 5% CO_2. Cells were routinely subcultured when they achieved 30–50% confluence so as to decrease the occurrence of spontaneous transformants.

Enzymes and reagents. Eco RI linkers, restriction endonucleases, Eco RI methylase and bacteriophage T4 DNA ligase were purchased from New England Biolabs, Beverly, MA. Reverse transcriptase from

Address reprint requests to Dr. Sukhatme, Howard Hughes Medical Institute, University of Chicago, Billings Hospital, S-508, 5841 S. Maryland Avenue, Chicago, IL 60637.

Received for publication 31 July 1987 and in revised form 10 November 1987.

J. Clin. Invest.
© The American Society for Clinical Investigation, Inc.
0021-9738/88/05/1621/09 $2.00
Volume 81, May 1988, 1621–1629

1. *Abbreviations used in this paper:* HCL, human cathepsin L; MCL, murine cathepsin L; MCP, murine cysteine proteinase; MEP, major excreted protein.

avian myeloblastosis virus was obtained from Molecular Genetic Resources, Tampa, FL. M13mp10 and M13mp11 vectors, as well as [alpha-^{32}P]dCTP (> 3,000 Ci/mmol), and 5'-[alpha-(^{35}S)thio]dATP (400–800 Ci/mmol) were purchased from Amersham Corp., Arlington Heights, IL. Polynucleotide kinase, M13 seventeen base sequencing primer, oligo (dT) cellulose, oligo $(dT)_{12-18}$, deoxy- and dideoxyribonucleotides were obtained from Pharmacia Fine Chemicals, Piscataway, NJ. S1 nuclease and the Klenow fragment of DNA polymerase used in the DNA sequencing reactions were purchased from Boehringer Mannheim, Indianapolis, IN. 17-mer oligonucleotide primers for DNA sequencing were synthesized by the methoxy phosphoramidite method on an Applied Biosystems (Foster City, CA) oligonucleotide synthesizer.

Transformation protocol. Using a standard calcium phosphate precipitation procedure (25), these cells were transfected with 20 μg of mouse L cell thymidine kinase negative (tk^-) fibroblast carrier DNA and 45 ng of DNA from plasmid pT24-C3 (26). This plasmid contains a 6.4 kb Bam HI insert encoding an activated human Ha-ras oncogene. 2 d after transfection, the medium was changed to DMEM with 3% FCS. Between days 10 and 20 after transfection, cloning cylinders were used to pick individual foci of morphologically transformed cells.

cDNA library. RNA was extracted from ras transformed, untransformed contact inhibited, and untransformed exponentially growing cells by the guanidinium thiocyanate method (27). Poly $(A)^+$ RNA was selected by oligo (dT)-cellulose chromatography (28). A lambda gt10 cDNA library was constructed (29) from 2 μg of poly $(A)^+$ RNA resulting in $\sim 5 \times 10^5$ recombinant phage with average insert size of ~ 700 bp. A human kidney cDNA library in lambda gt10 was kindly provided by Dr. G. Bell, Department of Biochemistry, University of Chicago.

Differential screening of cDNA library. ^{32}P-labeled cDNA probes were prepared from poly $(A)^+$ RNA as described (30) except that 100 μCi of [^{32}P]dCTP (> 3,000 Ci/mmol), 0.02 mM cold dCTP and 2–5 μg of poly $(A)^+$ RNA was used in each reaction. The mean size of the reverse transcribed probes as assessed by agarose gel electrophoresis and subsequent autoradiography was about 700 bases. Replica filter lifts (GeneScreenPlus; Dupont-New England Nuclear, Boston, MA) were prepared essentially as described by Benton and Davis (31) and $\sim 3 \times 10^6$ cpm of ^{32}P-cDNA was used per 90 mm diam filter. Hybridizations were carried out at 65°C in 1% SDS, 10% dextran sulfate, and 1 M NaCl as described in the GeneScreen protocol for a period of 16 h. The filters were washed twice for 20 min each time, first at room temperature in 2× SSC then at 65°C in 2× SSC, 1% SDS (low stringency wash), and finally at 65°C in 0.2× SSC (for high stringency wash). Exposure was for 18 h at −70°C with an intensifying screen.

Nucleic acid methods. Phage particles were prepared by the plate lysate method and phage DNA was purified by CsCl equilibrium centrifugation (32). Plasmid DNA was isolated by the Triton-lysozyme lysis method followed by ethidium bromide/CsCl equilibrium centrifugation (32). Enzyme reactions were carried out as suggested by the suppliers. [^{32}P]dCTP labeled probes were made either by nick translation (33) or by random hexamer priming (34) using a Pharmacia kit. All Northern blots were annealed to radioactively labeled probe and washed as described in the previous paragraph.

DNA sequencing. The dideoxynucleotide chain termination method (35, 36) with M13 vectors mp10, mp11, mp18, and mp19 was used for the cDNA sequencing. Both strands of the insert of clone pRIT-1.2 and of the insert of clone pHCL-12.1 were fully sequenced, with the exception of 71 bp of 5'-untranslated and 74 bp of 3' untranslated sequence in pHCL-12.1 that were only sequenced in a single direction (Fig. 2 *B*). DNA sequence and deduced protein sequence data were processed using the MicroGenie program (Beckman Instruments, Inc., Palo Alto, CA).

Human tissues. Peripheral blood cells from patient M.L. were provided by Dr. R. Larson, Department of Medicine, University of Chicago. M.L. was a female with a monocytic leukemia that had evolved from a myelodysplastic syndrome. Mononuclear cells were separated from red blood cells (RBC) on a Ficoll-Hypaque gradient. After separation, the cells were washed four times in DMEM. The pellet was snap frozen in liquid nitrogen and stored at −70°C. Tissues from surgical pathology (kidney, breast) were provided by Dr. P. Dawson, Department of Pathology, University of Chicago. These tissues were received after sampling for surgical pathology. Sections were snap-frozen in liquid nitrogen and stored at −70°C. Selection of these particular cases was determined solely by availability at the time of the request.

Results

Isolation of ras-induced cDNA clones. After transfection of NIH3T3 cells with the activated c-Ha-ras oncogene, two clones, designated a-H-ras-1 and a-Ha-ras-2, showed the highest level of ras mRNA expression. These two cell lines also had the most spindle-shaped refractile appearance, grew most rapidly and were least adherent to the tissue culture dishes. Since the level of ras message might correlate with the invasive and metastatic behavior of these cells (5–9), we felt that use of the two clones with most abundant ras expression would lead to the identification of the largest number of ras regulated transcripts in a differential screening analysis. To decrease the incidence of false positives arising from alterations in gene expression due to passage in cell culture or due to insertional mutagenesis or DNA rearrangements secondary to the calcium phosphate method we constructed the cDNA library with RNA isolated from a-Ha-ras-1 cells and screened it differentially with cDNA from a-Ha-ras-2 cells and with cDNA from nontransformed NIH3T3 fibroblasts.

2,000 plaques screened in this manner led to the identification of 19 clones for which the hybridization signal was more intense with the probe from a-Ha-ras-1 than with the probe from the nontransformed cells. Sibship analysis revealed that 17 of the 19 clones cross-hybridized. These 17 were considered to represent a single cDNA designated RIT-1 (ras induced transcript-1). The longest RIT-1 insert, 800 base pairs, was subcloned into pBR322 and designated pRIT-1.1.

Northern analysis of mRNA size. To confirm that the mRNA corresponding to RIT-1 is markedly up-regulated in ras transformed fibroblasts, and to determine the size(s) of the mRNA, Northern blot analysis was performed on total RNA from exponentially growing ras-transformed 3T3 cells (a-Ha-ras-1 and -2), contact inhibited resting nontransformed 3T3 cells, and exponentially growing 3T3 cells. Fig. 1 *A* (3 h exposure) shows a band at 1.5 kb and shows that the RIT-1 transcript is induced ~ 50-fold (as measured by laser densitometry) relative to the nontransformed cells. Fig. 1 *B* shows the same lanes after a 3-d exposure, permitting better visualization of the band in the nontransformed cells. The RIT-1 mRNA is also highly induced in a Kirsten ras transformed cell line (37) (data not shown).

Since the Northern analysis indicated a transcript size of 1.5 kb, we used pRIT-1.1 to screen a mouse Ltk^- lambda gt10 cDNA library constructed previously (38). Several positively hybridizing clones were plaque purified. One of them, pRIT-1.2, had a 1.36-kb insert. The Bgl II-Hind III fragment containing this insert was subcloned into pUC19 and then sequenced in M13. This clone was shown by us to encode murine cathepsin L (MCL) (16). Later in this paper we show that it contains the entire coding region for the precursor form of this enzyme.

Restriction map and sequence strategy for MCL. A restriction map of the cDNA insert of pRIT 1.2 and the sequencing strategy is shown in Fig. 2 *A*. All restriction fragments were

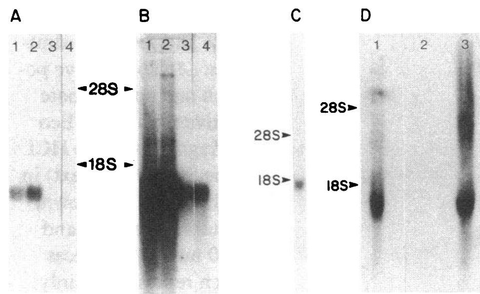

Figure 1. Northern analysis of mRNA transcript sizes. The position of ribosomal RNA size markers, indicated by the arrows labeled 18S and 28S, were determined for each gel by ethidium bromide staining. (*A*) 0.5 μg of poly (A)$^+$ RNA isolated from the following lines was loaded for each lane: a-H-ras-1 (lane *1*), a-H-ras-2 (lane *2*), confluent contact-inhibited nontransformed NIH3T3 cells (lane *3*), exponentially growing nontransformed NIH3T3 cells (lane *4*). Exposure was for 3 h at −70°C with an intensifying screen. (*B*) A 3-d exposure of the blot in A. (*C*) 10 μg of total RNA from PHA-stimulated lymphocytes was probed with oligolabeled pHCL-800.1. Exposure was overnight with a single intensifying screen. (*D*) All available total RNA was loaded for each of several surgical pathology specimens to maximize detection: 7 μg from a well-differentiated breast carcinoma (lane *1*), 11 μg from peripheral blood lymphocytes from a patient with monocytic leukemia arising from a myelodysplastic syndrome (lane *2*), 17 μg from a kidney undergoing chronic graft rejection (hypertensive vascular changes). This blot was probed with oligolabeled pHCL-800.1, and exposed overnight with a single intensifying screen.

subcloned into M13mp10 and M13mp11. The nucleotide sequence of this insert is shown in Fig. 3.

Isolation of cDNA clones for HCL. Plasmid pRIT 1.1 was used to screen a human kidney cDNA library. Hybridization was at low stringency. Of 20,000 clones screened, 5 were positive. All five clones contained an ~ 800 bp Eco RI fragment that hybridized at high-stringency to the 800 bp Eco RI fragment from the murine coding region. The longest insert, found in clone SL12.1, was ~ 4,000 bp. The 800-bp Eco RI fragment from SL12.1 was subcloned into pUC 13 and designated pHCL-800.1.

Northern analysis of HCL mRNA. pHCL-800.1 was used to probe Northern blots containing total RNA from several human tissues, as shown in Fig. 1 *C* and *D*. A single band at 1.5 kb is detected in PHA-activated human peripheral blood lymphocytes, also in a moderately well differentiated adenocarcinoma of the breast with minimal inflammation, and in RNA from a rejected human kidney transplant. The kidney tissue was processed after delay for surgical pathology examination, which showed hypertensive vascular disease but no evidence of inflammation. There is also a partially degraded but intense band at ~ 4 kb. It is not known if this represents alternative splicing, or a transcript from a homologous gene. Its strong intensity and absence in other tissues suggests that it is not a prespliced precursor. The presence of a larger transcript is consistent with the fact that clone SL12.1, itself derived from kidney, is shown by sequencing to code for cathepsin L, and has an insert size of 4,000 bases. 257 nucleotides of coding sequence, entirely from the pre- and proregions occur within a 1,600-bp Eco RI fragment (see Fig. 2 *B*). Lane 2 in Fig. 1 *D* contains total RNA from the PBL of a patient with monocytic leukemia that evolved from a myelodysplastic syndrome. Although the amount of RNA loaded, 11 μg, was intermediate between that loaded for breast and for kidney, and did not appear degraded as assessed by EtBr staining, no band is seen. This result was not unexpected since MCL appears to be highly expressed in only some macrophage/monocyte systems (14).

Restriction map and sequence strategy for human cDNA. Restriction mapping showed that SL 12.1 has an ~ 1,600 bp Eco RI fragment immediately 5′ to the 800-bp fragment, as well as an ~ 1,300-bp Eco RI fragment 5′ to the 1,600 bp piece (and forming the 5′ terminus of SL12.1). Two smaller Eco RI fragments are present at the 3′ terminus, analogous to the map for MCL. The restriction map and sequencing strategy are shown in Fig. 2 *B*. The sequencing strategy used both restriction fragment subcloning and the use of synthetic oligonucleotide primers. The nucleotide sequence is shown in Fig. 3.

Nucleotide and amino acid sequence analysis with comparison of murine and human clones. Examination of both nucleotide sequences shows that in each case the reading frame shown is the only one without multiple termination codons. In each the initiator codon represents the most 5′ of all in frame methionines, and is preceded by a termination codon located one codon upstream. These initiator codons are the only ones surrounded by a sequence in partial agreement with the Kozak consensus sequence for initiation of translation (ACCATGG) (39). Neither sequence shows a poly A tail; however they both show a consensus sequence for a polyadenylation signal (40). There is 76% identity at the nucleotide level, which is almost entirely confined to the coding and 3′ noncoding region.

The complete deduced amino acid sequences for MCL and HCL are shown in Fig. 4. Counting only matches of identical amino acids, there is 72% identity. We have recently shown (16) that the predicted human sequence from clone SL12.1 definitively identifies it as human cathepsin L since 38 out of 40 amino acids match the partial N-terminal sequence of the heavy chain of cathepsin L as reported by Mason et al. (15). Our complete deduced amino acid sequence, reported here, shows complete agreement (amino acids 292–333) with the entire 42 amino acid sequence for the light chain also reported by Mason (15) and establishes that his group has sequenced the entire light chain.

Both murine and human predicted amino acid sequences show similar organization. Each begins with a 17 amino acid long hydrophobic sequence that ends in Ala-X-Ala, a proposed consensus cleavage site for signal peptides (41). These are typical features of a "signal" sequence involved in targetting the newly synthesized protein to the endoplasmic reticulum. In HCL the proposed leader sequence is followed by a 96 amino acid propeptide, absent from the heavy chain in the mature protein, as shown by the sequence data of Mason et al. (15). Murine cathepsin L has been shown to undergo processing, reflected in abrupt decreases in size of the protein; however, the amino acid sequence is not available to verify the predicted analogous cleavage sites.

Mason et al. have shown that mature functional HCL is composed of a heavy chain and a light chain, derived from a single chain precursor (15). We cannot rule out the possibility that several amino acids are split off the carboxy terminal of the heavy chain; however, the site of interchain cleavage must occur after the His-276 (corresponding to His-159 in papain) active site, setting a limit on the extent of cleavage. Purification of human cathepsin L, has required acid activation (20) that could result in artificial generation of two chains; how-

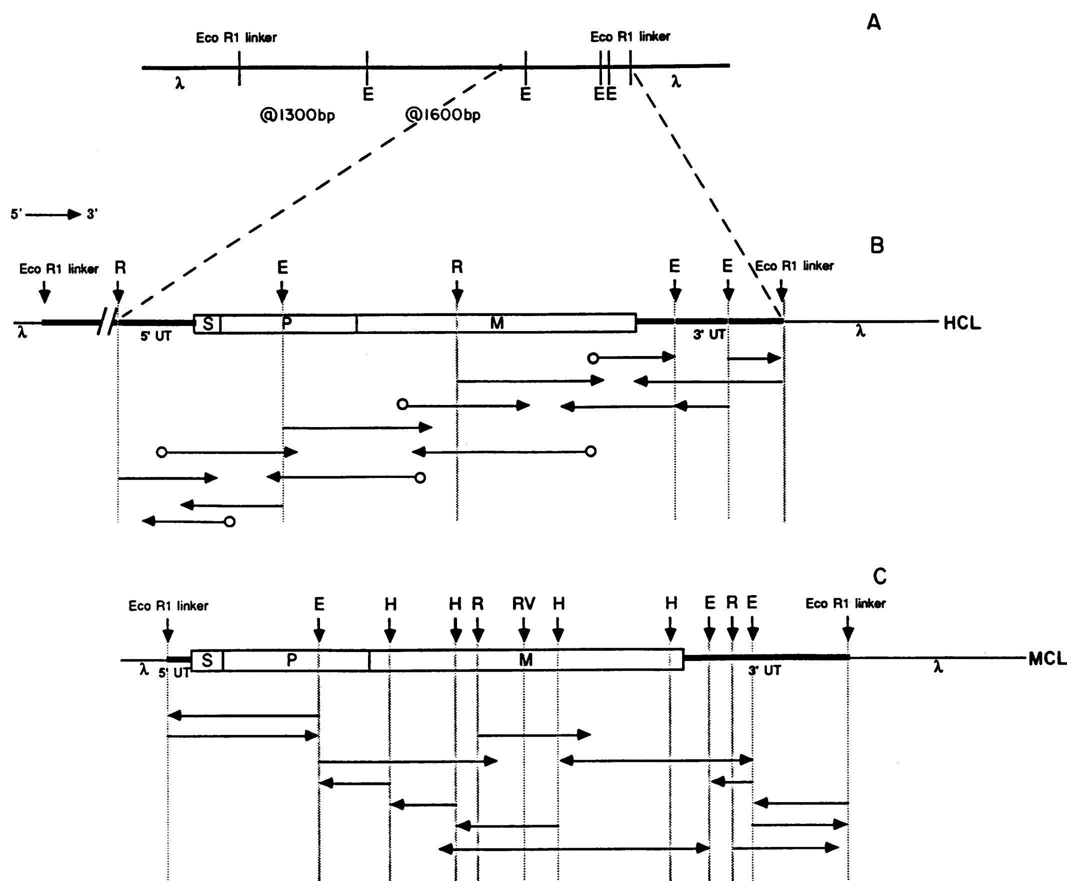

Figure 2. Restriction maps and sequence strategies. (*A*) The relative position and approximate relative sizes of the Eco RI fragments of the HCL clone SL12.1 (see text) in lambda gt10 are shown, including the 1300 and 1600 nucleotide pieces which represent mainly untranslated 5′ sequence. The region sequenced, including the entire coding portion of HCL, is demarcated by the dashed lines, which lead to an expanded map of this region as shown in B. (*B* and *C*). The open boxes represent the predicted protein coding regions, the dark lines 5′ and 3′ untranslated (ut) sequences, and the light lines are lambda gt10 arms. The coding region is divided into a signal sequence (S), a proposed "pro" region (P), and the proposed mature protein (M). Although each coding region is the same size as its counterpart in the other species, the scale for MCL (*C*) is slightly different than that for HCL in order to accommodate clearly the greater number of restriction map markings. (E = EcoRI, H = Hae III, R = RSA I, RV = Eco RV). The Eco RI sites that connect the phage arms to the insert cDNA correspond to synthetic oligonucleotide linkers used in the construction of the cDNA library. Arrows indicate orientation and length of the cDNA sequences determined. Arrows that have open circles at their base indicate sequences primed from oligonucleotides synthesized based on already established sequence in the cDNA.

ever, MCL undergoes similar cleavage intracellularly, suggesting the processing is not an artifact in human enzyme preparation (14).

The murine sequence displays one more carboxy terminal amino acid than does HCL. It is not known if this is present in the mature protein or is processed. MCL has two cysteine residues lacking in HCL: Cys-125 and Cys-146 of the mature protein. It is not clear if this results in intrachain disulfide bonding. MCL shows two potential N-glycosylation sites, Asn-221 and Asn-268. HCL shows only one, Asn 221. The overall similarity of the L cathepsins with several of the other known cysteine proteinases of the papain superfamily is shown in Fig. 5. Papain is used as the prototype (and for numbering), gaps are inserted in papain as well as other members to maximize identity. This multiple sequence alignment is an elaboration of that presented by Takio et al. for rat cathepsins B and H (42). Conservative changes of amino acids are also noted. The regions under comparison include most of the mature protein for each enzyme, but not the leader or propeptide pieces since these have no counterpart in papain. A pairwise application of the protein alignment program of MicroGenie (Beckman Instruments, Inc.), again, applied only to the mature proteins, and counting only matches of identical amino acids, was also performed. MCL and HCL show 26–28% identity with the murine and human B cathepsins (43), 40–42% identity with papain, and 45–46% identity with rat cathepsin H (42).

The three-dimensional structure of papain has been determined (44). Cys-25 and His-159 (numbering based on papain) are the active sites. The cysteine proteinases show near identity in an 11 amino acid sequence centered on Cys-25, less conservation around the His-159. Other amino acids have been implicated in determining substrate preferences, typically those that can contact substrate side chains or help determine the size of the "pocket" into which the substrate fits. For example, Ser-205 is near the opening of the pocket. In cathepsin B it is replaced by Glu. This substitution has been proposed by Barrett and Kirschke (45) as the reason cathepsin B can bind both Z-Phe-Arg and Z-Arg-Arg comparably. This hypothesis is consistent with the presence of Ala-205 in both HCL and MCL, which could limit interaction with the second basic sidegroup and explain the marked preference of cathepsin L for Z-Phe-Arg over Z-Arg-Arg.

Discussion

This report presents the complete nucleotide sequences and the corresponding deduced amino acid sequences for human and murine preprocathepsin L. The murine sequence is nearly

```
-222            GGGCTGCAGGTCGACTCTCGAGGATCCCCACCTCCACGTGCC HCL
-180   CTGTTTTTCTGGAGGCACATCCTTGGCCTCTTCCACAGTCCTTGGGTAAATGCTTGGGAG HCL
-120   AATAATTTAAATATTTTTATTCTACCATGGTGGCCCTAATTTTTCAGGGGGCAGTAAGAT HCL

-45                CGGCAGACTTCTTGTGCGCACGTAGCCGCCTCAGGTGTTTGAACC MCL
-60    GGCTTTTTAGGATTGGTCTAATCAGATCCTCATTTTTGTTCCCTTCCTAGGTTTTAAAAC HCL

MCL 1  ATGAATCTTTTACTCCTTTTGGCTGTCCTCTGCTTGGGAACAGCCTTAGCTACTCCA
HCL    ATGAATCCTACACTCATCCTTGCTGCCTTTTGCCTGGGAATTGCCTCAGCTACTCTA

58     AAATTTGATCAAACCTTTAGTGCAGAGTGGCACCAGTGGAAGTCCACGCACAGAAGACTG
       ACATTTGATCACAGTTTAGAGGCACAGTGGACCAAGTGGAAGGCGATGCACAACAGATTA

118    TATGGCACGAATGAGGAAGAGTGGAGGAGAGCGATATGGGAGAAGAACATGAGAATCATC
       TACGGCATGAATGAAGAAGGATGGAGGAGAGCAGTGTGGGAGAAGAACATGAAGATGATT

178    CAGCTACACAACGGGGAATACAGCAACGGGCAGCACGGCTTTTCCATGGAGATGAACGCC
       GAACTGCACAATCAGGAATACAGGGAAGGGAAACACAGCTTCACAATGGCCATGAACGCC

238    TTTGGTGACATGACCAATGAGGAATTCAGGCAGGTGGTGAATGGCTATCGCCACCAGAAG
       TTTGGAGACATGACCAGTGAAGAATTCAGGCAGGTGATGAATGGCTTTCAAAACCGTAAG

298    CACAAGAAGGGGAGGCTTTTTCAGGAACCGCTGATGCTTAAGATCCCCAAGTCTGTGGAC
       CCCAGGAAGGGGAAAGTGTTCCAGGAACCTCTGTTTTATGAGGCCCCCAGATCTGTGGAT

358    TGGAGAGAAAAGGGTTGTGTGACTCCTGTGAAGAACCAGGGCCAGTGCGGGTCTTGTTGG
       TGGAGAGAGAAAGGCTACGTGACTCCTGTGAAGAATCAGGGTCAGTGTGGTTCTTGTTGG

418    GCGTTTAGCGCATCGGGTTGCCTAGAAGGACAGATGTTCCTTAAGACCGGCAAACTGATC
       GCTTTTAGTGCTACTGGTGCTCTTGAAGGACAGATGTTCCGGAAAACTGGGAGGCTTATC

478    TCACTGAGTGAACAGAACCTTGTGGACTGTTCTCACGCTCAAGGCAATCAGGGCTGTAAC
       TCACTGAGTGAGCAGAATCTGGTAGACTGCTCTGGGCCTCAAGGCAATGAAGGCTGCAAT

538    GGAGGCCTGATGGATTTTGCTTTCCAGTACATTAAGGAAAATGGAGGTCTGGACTCGGAG
       GGTGGCCTAATGGATTATGCTTTCCAGTATGTTCAGGATAATGGAGGCCTGGACTCTGAG

598    GAGTCTTACCCCTATGAAGCAAAGGACGGATCTTGTAAATACAGAGCCGAGTTCGCTGTG
       GAATCCTATCCATATGAGGCAACAGAAGAATCCTGTAAGTACAATCCCAAGTATTCTGTT

658    GCTAATGACACAGGGTTCGTGGATATCCCTCAGCAAGAGAAAGCCCTCATGAAGGCTGTG
       GCTAATGACACCGGCTTTGTGGACATCCCTAAGCAGGAGAAGGCCCTGATGAAGGCAGTT

718    GCGACTGTGGGGCCTATTTCTGTTGCTATGGACGCAAGCCATCCGTCTCTCCAGTTCTAT
       GCAACTGTGGGGCCCATTTCTGTTGCTATTGATGCAGGTCATGAGTCCTTCCTGTTCTAT

778    AGTTCAGGCATCTACTATGAACCCAACTGTAGCAGCAAGAACCTCGACCATGGGGTTCTG
       AAAGAAGGCATTTATTTTGAGCCAGACTGTAGCAGTGAAGACATGGATCATGGTGTGCTG

838    TTGGTGGGCTATGGCTATGAAGGAACAGATTCAAATAAGAATAAATATTGGCTTGTCAAG
       GTGGTTGGCTACGGATTTGAAAGCACAGAATCAGATAACAATAAATATTGGCTGGTGAAG

898    AACAGCTGGGGAAGTGAATGGGGTATGGAAGGCTACATCAAAATAGCCAAAGACCGGGAC
       AACAGCTGGGGTGAAGAATGGGGCATGGGTGGCTACGTAAAGATGGCCAAAGACCGGAGA

958    AACCACTGTGGACTTGCCACCGCGGCCAGCTATCCTGTCGTGAATTGA-TGGGTAGCGGT
       AACCATTGTGGAATTGCCTCAGCAGCCAGCTACCCCACTGTGTGAGCTGTGGACGGTGAT

1,017  AATG-AGGACTT-A-TGGACACTATG--TCCAAAG---GAATTCAGCTTAAAACTGACCA
1,018  GAGGAAGGACTTGACTGGGGATGGCGCATGCATGGGAGGAATTC---TTCAGTCT-ACC-

1,069  AACCCTTATTGAGTCAAACCATGGTACTTGAATCATTGAGGATCCAAGTCATGATTTGAA
1,073  AGCCCCCGCTGTGTCGGA-TA-CACACTCGAATCATTGAAGATCCGAGT-GTGATTTGAA

1,129  TTCTGTTGCCATTTTTACATGGGTTAAATGTTACCACTACTTAAAACTCCTGTTATAAAC
1,130  TTCTG-TGATATTTTCACACTGG-TAAATGTTACCTCTA-TTTTAATTACTGCTATAAAT

1,189  AGCTTTATAATATTGAAAACTTAGTGCTTAATTCTGAGTCTGGAATATTTGTTTTATATA
1,187  AGGTTTATATTATTG---ATTCA---CTTA---CTGACT-TTGCATTTTCGTTTT-TA-A

1,249  AAGGTTGTATAAAACTTTCTTTACCTCTTAAAAATAAATTTTAGCTCAGTGTGTGTGTCG
1,235  AAGGATGTATAAA--TT--TTTACCTGTTTAAATAAAATCG
```

Figure 3. Comparison of the nucleotide sequences for murine and human cathepsin L. The nucleotide corresponding to the start of the coding sequence (initiator methionine) for each protein is designated as number 1 and also indicated by a heavy arrow. The end of each coding region is designated by an arrowhead. The proposed consensus polyadenylation signal is bracketed for each. Hyphens are introduced to maximize alignment in the 3′ untranslated regions, no attempt was made to align the 5′ untranslated region.

```
MCL 1       Met Asn Leu Leu Leu Leu Leu Ala Val Leu Cys Leu Gly Thr Ala Leu Ala Thr Pro
             |   |           |   :   |   |           |   |   |       |       |   |
HCL         Met Asn Pro Thr Leu Ile Leu Ala Ala Phe Cys Leu Gly Ile Ala Ser Ala Thr Leu

   20   Lys Phe Asp Gln Thr Phe Ser Ala Glu Trp His Gln Trp Lys Ser Thr His Arg Arg Leu
             |   |       :           |       |           |   |   :       |       |   |
        Thr Phe Asp His Ser Leu Glu Ala Gln Trp Thr Lys Trp Lys Ala Met His Asn Arg Leu

   40   Tyr Gly Thr Asn Glu Glu Glu Trp Arg Arg Ala Ile Trp Glu Lys Asn Met Arg Ile Ile
         |   |       |   |   |       |   |   |   |   :   |   |   |   |   |   :   :   |
        Tyr Gly Met Asn Glu Glu Gly Trp Arg Arg Ala Val Trp Glu Lys Asn Met Lys Met Ile

   60   Gln Leu His Asn Gly Glu Tyr Ser Asn Gly Gln His Gly Phe Ser Met Glu Met Asn Ala
             |   |   |       |   |           |       |       |   :   |       |   |   |
        Glu Leu His Asn Gln Glu Tyr Arg Glu Gly Lys His Ser Phe Thr Met Ala Met Asn Ala

   80   Phe Gly Asp Met Thr Asn Glu Glu Phe Arg Gln Val Val Asn Gly Tyr Arg His Gln Lys
         |   |   |   |   |       |   |   |   |   |   |   :   |   |   :               |
        Phe Gly Asp Met Thr Ser Glu Glu Phe Arg Gln Val Met Asn Gly Phe Gln Asn Arg Lys

  100   His Lys Lys Gly Arg Leu Phe Gln Glu Pro Leu Met Leu Lys Ile Pro Lys Ser Val Asp
             :   |   |   :   :   |   |   |   |   |                   |   :   |   |   |
        Pro Arg Lys Gly Lys Val Phe Gln Glu Pro Leu Phe Tyr Glu Ala Pro Arg Ser Val Asp

  120   Trp Arg Glu Lys Gly Cys Val Thr Pro Val Lys Asn Gln Gly Gln Cys Gly Ser Cys Trp
         |   |   |   |   |       |   |   |   |   |   |   |   |   |   |   |   |   |   |
        Trp Arg Glu Lys Gly Tyr Val Thr Pro Val Lys Asn Gln Gly Gln Cys Gly Ser Cys Trp

  140   Ala Phe Ser Ala Ser Gly Cys Leu Glu Gly Gln Met Phe Leu Lys Thr Gly Lys Leu Ile
         |   |   |   |   :   |       |   |   |   |   |   |       |   |   |   :   |   |
        Ala Phe Ser Ala Thr Gly Ala Leu Glu Gly Gln Met Phe Arg Lys Thr Gly Arg Leu Ile

  160   Ser Leu Ser Glu Gln Asn Leu Val Asp Cys Ser His Ala Gln Gly Asn Gln Gly Cys Asn
         |   |   |   |   |   |   |   |   |   |   |           |   |   |       |   |   |
        Ser Leu Ser Glu Gln Asn Leu Val Asp Cys Ser Gly Pro Gln Gly Asn Glu Gly Cys Asn

  180   Gly Gly Leu Met Asp Phe Ala Phe Gln Tyr Ile Lys Glu Asn Gly Gly Leu Asp Ser Glu
         |   |   |   |   |   :   |   |   |   |   :       :   |   |   |   |   |   |   |
        Gly Gly Leu Met Asp Tyr Ala Phe Gln Tyr Val Gln Asp Asn Gly Gly Leu Asp Ser Glu

  200   Glu Ser Tyr Pro Tyr Glu Ala Lys Asp Gly Ser Cys Lys Tyr Arg Ala Glu Phe Ala Val
         |   |   |   |   |   |   |           |   |   |   |               :   :   |
        Glu Ser Tyr Pro Tyr Glu Ala Thr Glu Glu Ser Cys Lys Tyr Asn Pro Lys Tyr Ser Val

             v
  220   Ala Asn Asp Thr Gly Phe Val Asp Ile Pro Gln Gln Glu Lys Ala Leu Met Lys Ala Val
         |   |   |   |   |   |   |   |   |   |       |   |   |   |   |   |   |   |   |
        Ala Asn Asp Thr Gly Phe Val Asp Ile Pro Lys Gln Glu Lys Ala Leu Met Lys Ala Val
             ^

  240   Ala Thr Val Gly Pro Ile Ser Val Ala Met Asp Ala Ser His Pro Ser Leu Gln Phe Tyr
         |   |   |   |   |   |   |   |   |   :   |   |       |       |           |   |
        Ala Thr Val Gly Pro Ile Ser Val Ala Ile Asp Ala Gly His Glu Ser Phe Leu Phe Tyr

                                         v
  260   Ser Ser Gly Ile Tyr Tyr Glu Pro Asn Cys Ser Ser Lys Asn Leu Asp His Gly Val Leu
                 |   |   |   :   |   |       |   |   |           :   |   |   |   |   |
        Lys Glu Gly Ile Tyr Phe Glu Pro Asp Cys Ser Ser Glu Asp Met Asp His Gly Val Leu

  280   Leu Val Gly Tyr Gly Tyr Glu Gly Thr Asp Ser Asn Lys Asn Lys Tyr Trp Leu Val Lys
         :   |   |   |   |   :   |       |   :   |           |   |   |   |   |   |   |
        Val Val Gly Tyr Gly Phe Glu Ser Thr Glu Ser Asp Asn Asn Lys Tyr Trp Leu Val Lys

  300   Asn Ser Trp Gly Ser Glu Trp Gly Met Glu Gly Tyr Ile Lys Ile Ala Lys Asp Arg Asp
         |   |   |   |       |   |   |   |       |   |   :   |   :   |   |   |   |
        Asn Ser Trp Gly Glu Glu Trp Gly Met Gly Gly Tyr Val Lys Met Ala Lys Asp Arg Arg

  320   Asn His Cys Gly Leu Ala Thr Ala Ala Ser Tyr Pro Val Val Asn End
         |   |   |   |   :   |   :   |   |   |   |   |       |
        Asn His Cys Gly Ile Ala Ser Ala Ala Ser Tyr Pro Thr Val End
```

Figure 4. Comparison of the deduced amino acid sequences for murine and HCL. The translated sequence is given in three letter code. A proposed 17-amino acid signal sequence occurs in each at the same position (amino acids 1–17), the predicted cleavage site is indicated by an arrow. The proposed "pro" region in each begins after the arrow and ends at the site marked by an arrowhead (amino acids 18–113). The sequence for the proposed "mature" protein extends from the arrowhead to the last amino acid. The active site cysteine and histidine are boxed. Potential N-linked glycosylation sites are marked by carats (^). Identical amino acids in corresponding positions are marked by a solid vertical line, amino acids reflecting conservative changes but occurring in corresponding positions are marked by : .

identical to that recently reported by Portnoy et al. (14) for a cDNA expressing a murine cysteine proteinase (MCP) in a murine macrophage line. These differences occur in the 5′ and 3′ untranslated region. Denhardt et al. (46) have reported a partial cDNA sequence for the MEP of transformed murine fibroblasts, which is in full agreement with our data. Thus, MCP and MEP are in fact MCL.

Gottesman et al. have extensively investigated the relationship of MEP (MCL) to transformation, cell cycle progression, growth factor stimulation, and transcriptional control (47–51). After transformation of murine NIH or BALB 3T3 fibroblasts with K-ras, intracellular MCL rises from less than 0.02 to 1% of the total intracellular protein. The unprocessed MCL is secreted and accounts for up to 30% of the excreted proteins. Similar results apply to transformation by Ha-ras, MSV, SV40, and methylcholanthrene. Stimulation of nontransformed 3T3 fibroblasts by plating in serum at low density did not significantly elevate the basal level. Denhardt examined MEP (MCL) mRNA expression at different points in the cell-cycle (stimulated by serum) and found less than a twofold change (46). MEP (MCL) can be induced to high levels in nontransformed fibroblasts by stimulation with PDGF or TPA. Insulin, somatomedin C, and plasma are without effect (52).

We have presented the first demonstration of cathepsin L mRNA expression in a human tumor, specifically breast carcinoma. Rat mammary glands in culture have been shown to secrete a high molecular weight form of cathepsin L (53). Cell lines such as MCF-7, which are derived from human breast carcinomas, are known to secrete several proteases (54). Transfection of v-Ha-ras into MCF-7 breast cancer cells has been shown to have multiple effects, including increased inva-

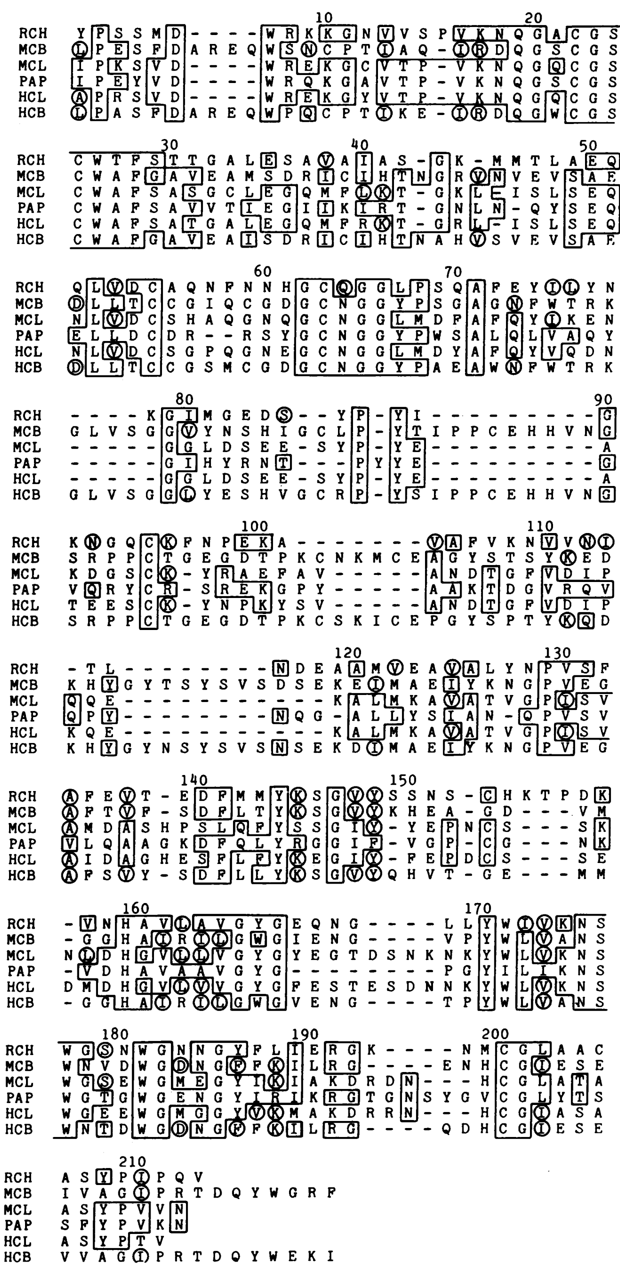

Figure 5. Multiple sequence alignment of several cysteine proteinases. Papain serves as the prototype and the basis for the numbering. Corresponding amino acids that are identical to papain are boxed, those that reflect conservative changes are circled. Gaps have been introduced to maximize alignment. RCH, rat cathepsin H; MCB, murine cathepsin B; PAP, papain; HCB, human cathepsin B.

siveness and enhanced adhesion to a laminin substrate (55, 56). It is difficult to determine an appropriate reference tissue to determine if cathepsin L mRNA expression in breast carcinoma is "elevated": "normal" breast tissue is primarily composed of adipocytes. A study comparing levels in other breast lesions and the MCF-7 line before and after v-H-ras transfection is in progress.

It remains to be determined if an elevated level of cathepsin L is essential to transformation by ras in 3T3 or other cells. A marked elevation in cathepsin L is not indispensable to proliferation since it is not significantly elevated in exponentially growing nontransformed fibroblasts or several tumor cell lines. It could still be critical to transformation by ras. Cathepsin L might be required to perform a specific proteolytic function, such as has been mentioned with alpha-one protease inhibitor (21), distinct from its role in "general" protein catabolism. Alternatively, some change induced by ras might fortuitously result in increased synthesis and/or accumulation of cathepsin L, perhaps by interrupting a negative feedback loop. "Ras" or a "ras"-like protein has recently been implicated in the control of exocytosis (57). Rat cathepsins D and C have been shown to be present in exocytic coated vesicles in nontransformed liver cells (58). Similar studies have not yet been done for cathepsin L or for the other cathepsins in transformed cells. Another broad set of explanations involve alterations in protein sorting. Glycosylation abnormalities, which have been extensively described for the cell surface in malignancy (59), could alter sorting of cathepsin L, resulting in increased secretion and altered intracellular distribution. An alteration in glycosylation, affecting a specific protein, has been described for uteroferrin, a lysosomal enzyme, which has a "covered" mannose-6-phosphate and results in constitutive secretion (60). An altered transferase in I cell disease prevents expression of the mannose-6 phosphate marker and results in massive secretion of many lysosomal constitutents (61). Some cells, such as J774.2, appear to lack a properly functioning mannose 6 phosphate receptor (215 kD receptor), again resulting in marked secretion of multiple lysosomal species (62). Finally, if increased transcription/translation of cathepsin L is the primary event, the quantity of cathepsin L synthesis might overwhelm the normal sorting mechanism. Examination of the level of extracellular secretion of other lysosomal enzymes, and other studies are under way to distinguish among these possibilities.

Whether or not cathepsin L is essential to ras transformation, as a protease with potent collagenolytic and elastinolytic activity it could play a role in local invasion or metastasis. The role of cathepsin L in tumor progression has not been explicitly studied. Studies of metastasis using inhibitors for the "thiol proteinases" are not straightforward. Extracellular activity presumably requires an acidic pH. Such a pH can occur at several sites (reviewed by Weiss [63]). Release of lysosomes or translocation of proton ATPases to the cell surface (64) could transiently provide a locally favorable pH at any site. We have used RNA from a series of transformed rat fibroblasts described by Pozzatti et al. (65) to assay for cathepsin L mRNA expression. The c-Ha-ras-transformed lines are highly metastatic in nude mice, and c-Ha-ras +E1a (adenovirus type 2) double transformants are not metastatic. These lines express similar levels of ras mRNA but, interestingly, the level of murine cathepsin L mRNA correlates sharply with their metastatic potential (data not shown). These studies are consistent with an extracellular role for this enzyme.

The cathepsin L cDNA clones generated in this work will prove useful in the study of the structure, processing, expression, and function of this lysosomal enzyme and its possible role in malignancy.

Acknowledgments

We are grateful to G. Toback, S. Gluck, and G. Jamieson for helpful discussions; S. Lapid and C. Stoklosa for expert technical assistance, and Kathy Cobb for typing the manuscript, and P. Gardner in D.

Steiner's laboratory for synthesizing the oligonucleotides. Dr. R. Hoover kindly provided Northern blots from PHA activated lymphocytes.

This work was supported by the Howard Hughes Medical Institute, by a grant from the University of Chicago Cancer Research Foundation and Development Funds from the Department of Medicine, and the Biological Sciences Division of the University of Chicago.

References

1. Scott, M. R. D., K. H. Westphal, and P. W. J. Rigby. 1983. Activation of mouse genes in transformed cells. *Cell.* 34:557–567.

2. Brickell, P. M., D. S. Latchman, D. Murphy, K. Willison, and P. W. J. Rigby. 1983. Activation of a Qa/T1a class I major histocompatibility antigen gene is a general feature of oncogenesis in the mouse. *Nature (Lond.).* 306:756–760.

3. Matrisian, L. M., N. Glaichenhaus, M. C. Gesnel, and R. Breathnach. 1985. Epidermal growth factor and oncogenes induce transcription of the same cellular mRNA in rat fibroblasts. *EMBO (Eur. Mol. Biol. Org.) J.* 4:1435–1440.

4. Sugano, S., M. Y. Stoeckle, and H. Hanafusa. 1987. Transformation by Rous sarcoma virus induces a novel gene with homology to a mitogenic platelet protein. *Cell.* 49:321–328.

5. Grieg, R. G., T. P. Koestler, D. L. Trainer, S. P. Corwin, L. Miles, T. Kline, R. Sweet, S. Yokoyama, and G. Poste. 1985. Tumorigenic and metastatic properties of "normal" and ras-transfected NIH/3T3 cells. *Proc. Natl. Acad. Sci. USA.* 82:3698–3701.

6. Bradley, M. O., A. R. Kraynak, R. D. Storer, and J. B. Gibbs. 1986. Experimental metastasis in nude mice of NIH 3T3 cells containing various ras genes. *Proc. Natl. Acad. Sci. USA.* 83:5277–5281.

7. Muschel, R. J., J. E. Williams, D. R. Lowy, and L. A. Liotta. 1985. Harvey ras induction of metastatic potential depends upon oncogene activation and the type of recipient cell. *Am. J. Pathol.* 121:1–8.

8. Egan, S. E., G. A. McClarty, L. Jarolim, J. A. Wright, I. Spiro, G. Hager, and A. H. Greenberg. 1987. Expression of H-ras correlates with metastatic potential: Evidence for direct regulation of the metastatic phenotype in 10T 1/2 and NIH 3T3 cells. *Mol. Cell. Biol.* 7:830–837.

9. Geiser, A. G., J. D. Channing, C. J. Marshall, and E. J. Stanbridge. 1986. Suppression of tumorigenicity with continued expression of the c-Ha-ras oncogene in EJ bladder carcinoma-human fibroblast hybrid cells. *Proc. Natl. Acad. Sci. USA.* 83:5209–5213.

10. Viola, M. V., F. Fromowitz, S. Oravez, S. Deb, G. Finkel, J. Lundy, P. Hand, A. Thor, and J. Schlom. 1986. Expression of ras oncogene p21 in prostate cancer. *N. Engl. J. Med.* 314:133–137.

11. Hand, P. H., A. Thor, D. Wunderlich, R. C. Muraro, A. Caruso, and J. Schlom. 1984. Monoclonal antibodies of predefined specificity detect activated ras gene expression in human mammary and colon carcinomas. *Proc. Natl. Acad. Sci. USA.* 81:5227–5231.

12. Andreeff, M., D. E. Slater, J. Bressler, and M. E. Furth. 1986. Cellular ras oncogene expression and cell cycle measured by flow cytometry in hematopoietic cell lines. *Blood.* 67:676–681.

13. Barrett, A. J., J. Martin, J. Nicklin, and N. D. Rawlings. 1984. The papain superfamily of cysteine proteinases and their protein inhibitors. *Symposia Biologica Hungarica.* 25:203–217.

14. Portnoy, D. A., A. H. Erickson, J. Kochan, J. V. Ravetch, and J. C. Unkeless. 1986. Cloning and characterization of a mouse cysteine proteinase. *J. Biol. Chem.* 261:14697–14703.

15. Mason, R. W., J. W. Walker, and F. D. Northrop. 1986. The N-terminal amino acid sequences of the heavy and light chains of human cathepsin L. *Biochem. J.* 240:373–377.

16. Joseph, L., S. Lapid, and V. Sukhatme. 1987. The major ras-induced protein in NIH3T3 cells is cathepsin L. *Nucleic Acids Res.* 15:3186.

17. Bohley, P., H. Kirschke, S. Schaper, and B. Wiederanders. 1984. Principle of the regulation of intracellular proteolysis. *Symposia Biologocia Hungarica.* 25:101–118.

18. Mayer, R. J., and F. Doherty. 1986. Intracellular protein catabolism: state of the art. *FEBS (Fed. Eur. Biochem. Soc.) Lett.* 198:181.

19. Mason, R. W., D. A. Johnson, A. J. Barrett, and H. A. Chapman. 1986. Elastinolytic activity of human cathepsin L. *Biochem. J.* 233:925–927.

20. Kirschke, H., A. A. Kembhavi, P. Bohely, and A. J. Barrett. 1982. Action of rat liver cathepsin L on collagen and other substrates. *Biochem. J.* 201:367–372.

21. Johnson, D. A., A. J. Barrett, and R. W. Mason. 1986. Cathepsin L inactivates alpha-$_1$-proteinase inhibitor by cleavage in the reactive site region. *J. Biol. Chem.* 261:14748–14751.

22. Kominami, E., K. Ii, and N. Katunuma. 1987. Activation of the intramyofibral autophagic lysosomal system in muscular dystrophy. *Am. J. Pathol.* 127:461–466.

23. Tsuchida, K., H. Aihara, K. Isogai, K. Hanada, and N. Shibata. 1986. Degradation of myocardial structural proteinsin myocardial infarcted dogs is reduced by Ep459, a cysteineproteinase inhibitor. *Biol. Chem. Hoppe-Seyler.* 367:39–45.

24. Olbricht, C. J., J. K. Cannon, L. C. Garg, and C. C. Tisher. 1986. Activities of cathepsins B and L in isolated nephron segments from proteinuric and nonproteinuric rats. *Am. J. Physiol.* 250:F1044–F1062.

25. Wigler, M., R. Sweet, S. Gek Kee, B. Wold, A. Pellicer, E. Lacy, T. Maniatis, S. Silverstein, and R. Axel. 1979. Transformation of mammalian cells with genes from procaryotes and eucaryotes. *Cell.* 16:777–785.

26. Santos, E., S. R. Tronick, S. A. Aaronson, S. Pulciani, and M. Barbacid. 1982. T24 human bladder carcinoma oncogene is an activated form of the normal human homologue of BALB- and Harvey MSV transforming genes. *Nature (Lond.).* 298:343–347.

27. Chirgwin, J. M., A. E. Pryzbala, R. J. MacDonald, and W. J. Rutter. 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry.* 18:5294–5299.

28. Aviv, H., and P. Leder. 1972. Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. *Proc. Natl. Acad. Sci. USA.* 69:1408–1412.

29. Huynh, T. V., R. A. Young, and R. W. Davis. 1985. Constructing and screening cDNA libraries in lambda gt10 and lambda gt11. *In* DNA Cloning. D. M. Glover, editor. IRL Press, Oxford, England. 49–78.

30. St. John, T. P., and R. W. Davis. 1979. Isolation of glactose-inducible DNA sequences from Saccharomyces cerevisiae by differential plaque filter hybridization. *Cell.* 16:443–452.

31. Benton, D., and R. W. Davis. 1977. Screening lambda gt recombinant clones by hybridization to single plaques in situ. *Science (Wash. DC).* 196:180–182.

32. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular Cloning. Cold Spring Harbor Laboratory, New York.

33. Rigby, P. W. J., M. Diekmann, C. Rhodes, and P. Berg. 1977. Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. *J. Mol. Biol.* 113:237–251.

34. Feinberg, A. P., and B. Vogelstein. 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* 132:6–13.

35. Sanger, F., S. Nicklen, and A. Coulsen. 1977. DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* 74:5463–5467.

36. Biggin, M. D., T. J. Gibson, and G. F. Hong. 1983. Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. *Proc. Natl. Acad. Sci. USA.* 80:3963–3967.

37. Noda, M., Z. Selinger, E. M. Scolnick, and R. H. Bassin. 1983. Flat revertants isolated from Kirsten sarcoma virus-transformed cells are resistant to the action of specific oncogenes. *Proc. Natl. Acad. Sci. USA.* 80:5602–5606.

38. Kavathas, P., V. P. Sukhatme, L. A. Herzenberg, and J. R. Parnes. 1984. Isolation of the gene encoding the human T-lymphocyte differentiation antigen Leu-2(T8) by gene transfer and cDNA subtraction. *Proc. Natl. Acad. Sci. USA.* 81:7688.

39. Kozak, M. 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell.* 44:283–292.

40. Perlman, D., and H. Halvorson. 1983. A putative signal peptidase recognition site and sequence in eukaryotic and prokaryotic signal peptides. *J. Mol. Biol.* 167:391–409.

41. von Heijne, G. 1984. How signal sequence maintain cleavage specificity. *J. Mol. Biol.* 173:243–251.

42. Takio, K., T. Towatari, N. Katunuma, D. C. Teller, and K. Titani. 1983. Homology of amino acid sequences of rat liver cathepsins B and H with that of papain. *Proc. Natl. Acad. Sci. USA.* 80:3666–3670.

43. Chan, S. J., B. San Segundo, M. B. McCormick, and D. F. Steiner. 1986. Nucleotide and predicted amino acid sequences of cloned human and mouse preprocathepsin B cDNAs. *Proc. Natl. Acad. Sci. USA.* 83:7721–7725.

44. Kamphuis, I. G., J. Drenth, and E. N. Baker. 1985. Thiol proteases: comparative studies based on the high-resolution structures of papain and actinidin, and on amino acid sequence information for cathepsins B and H and stem bromelain. *J. Mol. Biol.* 182:317–329.

45. Barrett, A. J., and H. Kirschke. 1982. Cathepsin B, cathepsin H, and cathepsin L. *Methods Enzymol.* 80:535–561.

46. Denhardt, D. T., R. T. Hamilton, C. L. J. Parfett, D. R. Edwards, R. St. Pierre, P. Waterhouse, and M. Nilsen-Hamilton. 1986. Close relationship of the major secreted protein of transformed murine fibroblasts to thiol-dependent cathepsins. *Cancer Res.* 46:4590–4593.

47. Gottesman, M. M. 1978. Transformation-dependent secretion of a low molecular weight protein by murine fibroblasts. *Proc. Natl. Acad. Sci. USA.* 75:2767–2771.

48. Gottesman, M. M., and M. E. Sobel. 1980. Tumor promoters and Kirsten Sarcoma virus increase synthesis of a secreted glycoprotein by regulating levels of translatable mRNA. *Cell.* 19:449–455.

49. Gal, S., M. C. Willingham, and M. M. Gottesman. 1985. Processing and lysosomal localization of a glycoprotein whose secretion is transformation stimulated. *J. Cell Biol.* 100:535–544.

50. Frick, K. K., P. J. Doherty, M. M. Gottesman, and C. D. Scher. 1985. Regulation of the transcript for a lysosomal protein: evidence for a gene program modified by platelet-derived growth factor. *Mol. Cell. Biol.* 5:2582–2589.

51. Rabin, M. S., P. J. Doherty, and M. M. Gottesman. 1986. The tumor promoter phorbol 12-myristate 13-acetate induces a program of altered gene expression similar to that induced by platelet-derived growth factor and transforming oncogenes. *Proc. Natl. Acad. Sci. USA.* 83:357–360.

52. Gal, S., and M. M. Gottesman. 1986. The major excreted protein of transformed fibroblasts is an activable acid-protease. *J. Biol. Chem.* 261:1760–1765.

53. Recklies, A. D., and J. S. Mort. 1985. Rat mammary gland in culture secretes a stable high molecular weight form of cathepsin L. *Biochem. Biophys. Res. Commun.* 131:402–407.

54. Capony, F., M. Morissett, A. J. Barrett, J. P. Capony, P. Broquet, F. Vignon, M. Chambon, P. Louisot, and H. Rochefort. 1987. Phosphorylation, glycosylation, and proteolytic activity of the 52 kd estrogen-induced protein secreted by MCF-7 cells. *J. Biol. Chem.* 104:253–262.

55. Albini, A., J. Graf, G. T. Kitten, H. K. Kleinman, G. R. Martin, A. Veilette, and M. E. Lippman. 1986. 17beta-estradiol regulates and v-Ha-ras transfection constitutively enhances MCF-7 breast cancer cell interactions with basement membrane. *Proc. Natl. Acad. Sci. USA.* 83:8182–8186.

56. Dickson, R. B., A. Kasid, K. K. Huff, S. E. Bates, C. Knabbe, D. Bronzert, E. P. Gelmann, and M. E. Lippman. 1987. Activation of growth factor secretion in tumorigenic states of breast cancer induced by 17 beta-estradiol or v-Ha-ras oncogene. *Proc. Natl. Acad. Sci. USA.* 84:837–841.

57. Burgoyne, R. D. 1987. Control of exocytosis. *Nature (Lond.).* 328:112–113.

58. Lemansky, P., A. Hasilik, K. von Figura, S. Helmy, J. Fishman, R. E. Fine, N. L. Kedersha, and L. H. Rome. 1987. Lysosomal enzyme precursors in coated vesicles derived from the exocytic and endocytic pathways. *J. Cell Biol.* 104:1743–1748.

59. Dennis, J. W., S. Laferte, C. Waghorne, M. Breitman, and R. S. Kerbel. 1987. Beta 1-6 branching of Asn-linked oligosaccharides is directly associated with metastasis. *Science (Wash. DC).* 236:582–585.

60. Baumbach, G. A., P. T. K. Saunders, F. W. Bazer, and R. M. Roberts. 1984. Uteroferrin has N-asparagine-linked high-mannose-type oligosaccharides that contain mannose-6-phosphate. *Proc. Natl. Acad. Sci. USA.* 81:2985–2989.

61. Kornfeld, S. 1986. Trafficking of lysosomal enzymes in normal and disease states. *J. Clin. Invest.* 77:1–6.

62. Gabel, C. A., D. E. Goldberg, and S. Kornfeld. 1983. Identification and characterization of cells deficient in the mannose 6-phosphate receptor: evidence for an alternate pathway for lysosomal enzyme targeting. *Proc. Natl. Acad. Sci. USA.* 80:775–779.

63. Weiss, L. 1985. Principles of Metastasis. Academic Press, New York. 12–13.

64. Gluck, S., C. Cannon, and Q. Al-Awqati. 1982. Exocytosis regulates urinary acidification in turtle bladder by rapid insertion of H^+ pumps into the luminal membrane. *Proc. Natl. Acad. Sci. USA.* 79:4327–4331.

65. Pozzatti, R., R. Muschel, J. Williams, R. Padmanabhan, B. Howard, L. Liotta, and G. Khoury. 1986. Primary rat embryo cells transformed by one or two oncogenes show different metastatic potentials. *Science (Wash. DC).* 232:223–227.