Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 2909–2913, May 1986
Cell Biology

Isolation of a cDNA clone for the human lysosomal proteinase cathepsin B

(acid hydrolase/synthetic oligonucleotides/thiol protease)

DUNNE FONG*, DAVID H. CALHOUN†, WANG-TING HSIEH‡, BENJAMIN LEE*, AND ROBERT D. WELLS‡

Departments of *Biology and ‡Biochemistry, University of Alabama at Birmingham, Birmingham, AL 35294; and †Department of Microbiology, Mount Sinai School of Medicine, New York, NY 10029

Communicated by J. T. Bonner, December 23, 1985

ABSTRACT **The cysteine proteinase cathepsin B is one member of the lysosomal acid hydrolases. Based on the peptide sequence of rat liver cathepsin B, an oligonucleotide mixture containing 128 different 17-mers was synthesized and used as a probe to screen adult and fetal human liver cDNA libraries. A recombinant clone with a 1540-nucleotide insert was identified from the fetal library, and DNA sequence analysis confirmed that this clone encodes human cathepsin B. The clone, designated pCB-1, has sequences for 81% of the coding region (for amino acid residues 50–252) together with ≈880 nucleotides of the 3′ untranslated region of the mRNA. The DNA sequence also shows that the predicted carboxyl terminus of the coding sequence is longer than the mature protein by 6 amino acid residues. Southern blot analysis of restriction enzyme digests of human placental DNA revealed a simple pattern of hybridizing fragments using the cathepsin B coding sequence as probe. The result suggests that there is a single copy of cathepsin B gene per haploid genome.**

Proteinases are present in all forms of living organisms. The events of general and limited proteolyses have been attributed to actions of these enzymes. In general proteolysis, proteinases digest nutrient proteins and participate in the turnover of cellular proteins, whereas in limited proteolysis, proteinases modify substrate proteins and alter the properties and physiological functions of these proteins. Thus, limited proteolysis can play regulatory roles during cell growth and differentiation (1, 2).

Based on the amino acid residues at their active sites, proteinases are classified into serine, cysteine, aspartate, and metalloproteinases. Members of the cysteine proteinase family have wide phylogenetic distribution, and examples include papain from the papaya plant and cathepsin B from mammalian tissues. Cathepsin B is a lysosomal enzyme that is functional during intracellular protein catabolism and may also be involved in the proteolytic processing of protein precursors such as proinsulin (3–5). Cathepsin B has been implicated in several disease states including muscular dystrophy (6), rheumatoid arthritis (7), and tumor metastasis (8–11). We and others have demonstrated that cathepsin B-like proteinases may also be involved during differentiation of the cellular slime mold *Dictyostelium discoideum* (12, 13).

To further investigate the role of cathepsin B in cellular functions, we have selected the molecular genetic approach. Here we report our results on the isolation and characterization of a cDNA clone for human cathepsin B.

MATERIALS AND METHODS

Synthesis of Oligodeoxyribonucleotides. A mixture of 17-nucleotide-long oligonucleotides encoding a portion of the rat liver cathepsin B protein sequence was synthesized by using a solid-phase phosphotriester method (14, 15). For amino acids specified by more than one codon, all possible nucleotides were inserted at the ambiguous positions. The synthesis was performed with a Vega Model 280 Automated Polynucleotide Synthesizer by a series of additions of dimers or dimer mixtures. After chemical synthesis, the oligonucleotide mixture was purified by high-performance liquid chromatography. These procedures were done with the assistance of David H. L. Bishop (University of Alabama at Birmingham Cancer Center Oligonucleotide Synthesis Core Facility).

Colony Screening with Oligonucleotides. The oligonucleotides were used to probe fetal and adult human liver cDNA libraries, kindly provided by Stuart H. Orkin (Harvard Medical School). The libraries were constructed by cloning the double-stranded cDNA into the *Pst* I site of plasmid pKT218 with GC homopolymer tailing (16, 17). High-density colony screening was performed (18) with a total of 50,000 colonies per nitrocellulose filter (127-mm Millipore Triton-free HATF) and four filters per library. Bacterial colonies were grown on the nitrocellulose filter placed in double-strength yeast extract/tryptone medium (2YT) containing tetracycline. Replica filters were made and placed on chloramphenicol plates for plasmid amplification. The replica colonies were lysed and the DNA was denatured and fixed onto the filters by baking 2 hr in a vacuum oven at 80°C (19). The nitrocellulose filters containing lysed bacterial colonies were washed for 0.5 hr at 65°C in a solution of 0.5% $NaDodSO_4$/0.1× standard saline citrate (NaCl/Cit; 1× is 0.15 M NaCl/0.015 M sodium citrate). Prehybridization was done at 65°C for 1 hr and then at 37°C for 1 hr in a solution of 0.5% $NaDodSO_4$/5× Denhardt's solution (1× Denhardt's solution is 0.02% bovine serum albumin/0.02% Ficoll/0.02% polyvinylpyrrolidone)/2× SSPE (1× SSPE is 0.18 M NaCl/10 mM NaH_2PO_4, pH 7.4/1 mM EDTA). Hybridization was performed at 37°C for 12–18 hr by the addition of ^{32}P-labeled oligonucleotides to fresh prehybridization solution at 1×10^6 cpm/ml. The oligonucleotides were 5′-end labeled by using [γ-^{32}P]ATP and T4 polynucleotide kinase (Bethesda Research Laboratories) at 20 units of enzyme to 100 pM oligonucleotides. After hybridization, the filters were washed at 37°C with several changes of 6× NaCl/Cit and exposed to x-ray films between two intensifier screens. The same filters were again washed in 6× NaCl/Cit at 45°C and re-exposed to films (20).

Isolation and Characterization of Plasmid DNA. After autoradiography, bacterial colonies that corresponded to positive signals on the x-ray films were isolated and re-screened with the labeled oligonucleotide probes. Plasmid DNA was extracted from bacterial colonies by using an alkaline lysis procedure (21) and purified by CsCl/ethidium bromide density-gradient centrifugation (19). A restriction

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: kb, kilobase(s).

```
                 69  70  71  72  73  74
PROTEIN:        CYS ASN GLY GLY TYR PRO

RNA:      5' UGU AAU GGU GGU UAU CCU 3'
               C   C   C   C   C   C
                       A   A       A
                       G   G       G

PROBE:    3' ACA TTA CCA CCA ATA GG  5'
               G   G   G   G   G
                       T   T
                       C   C
```

FIG. 1. Cathepsin B oligonucleotide sequences. The mixture consisted of 128 different 17-mers corresponding to amino acids 69–74 of rat liver cathepsin B (29).

map of pCB-1 was constructed by standard procedures (19) using pBR322 as control (22).

Sequence Analysis of Cathepsin B cDNA Clone. For DNA sequencing, purified pCB-1 plasmid DNA was first cut with one restriction enzyme, then treated with calf intestine alkaline phosphatase, end-labeled with T4 polynucleotide kinase, and restricted again with a second enzyme. The DNA fragments were separated in a 5% acrylamide gel and individual fragments were eluted for sequencing (23). In some cases, it was necessary to first isolate the fragments of interest by gel elution before end-labeling. Both regular and modified Maxam and Gilbert chemical cleavage methods were performed (24). In addition, for the 5′-terminal *Pst* I fragment, end-labeling with dideoxyadenosine-5′-[α-^{32}P]triphosphate was done using terminal deoxynucleotidyl transferase (25).

Blot Hybridization Analysis of Human Genomic DNA. Restriction enzyme digests of human placental DNA (a generous gift from Peter W. Melera, Sloan–Kettering Institute) were subjected to electrophoresis through 0.8% agarose gel and transferred to nitrocellulose filters by the method of Southern (26), and hybridized with nick-translated *Pst*/*Eco* insert fragment from the clone pCB-1 (27). Hybridization was done at 68°C and washes were performed sequentially at room temperature (2× NaCl/Cit with 0.5% $NaDodSO_4$) and 68°C (0.1× NaCl/Cit with 0.5% $NaDodSO_4$).

RESULTS

Screening Colonies with a Mixed Oligonucleotide Probe. Takio *et al.* (28) had reported that rat liver cathepsin B is composed of two polypeptide chains of molecular weights 25,000 and 5000 and that these chains are linked by disulfide bonds. Based on their partial amino acid sequence, a region corresponding to heavy chain residues 22–27 was selected for oligonucleotide synthesis. These residues are completely homologous to the plant proteinase papain. This choice was made because we planned to screen a human cDNA library using a probe derived from the rat amino acid sequence data. By selecting a region common to various cysteine proteinases, the chances of finding a human cathepsin B or related proteinase genes would be higher. However, after the completion of oligonucleotide synthesis, the complete cathepsin B protein sequence of rat liver was published by Takio *et al.* (29). The sequence revealed that the synthesized oligonucleotides really correspond to amino acid residues 69–74. Furthermore, because of codon ambiguity, the 17-nucleotide-long mixture has 128 different members (Fig. 1).

The ^{32}P-labeled oligonucleotides were used directly as *in situ* hybridization probes against replica filters containing lysed colonies of both fetal and adult human liver cDNA libraries. In an initial screening, ≈18 spots from the fetal library and 16 from the adult library exhibited hybridization after washing in 6× NaCl/Cit at 45°C. The temperature chosen was several degrees below the dissociation temperature (48°C) for the oligonucleotide mixture [calculated by the formula 2°C × (number of A·T base pairs) plus 4°C × (number of G·C base pairs), see ref. 30]. Regions corresponding to positive hybridization signals were lifted with toothpicks and the picked patches were rescreened with labeled oligonucleotide probes. Only three such patches remained positive in this hybridization, and single bacterial colonies from these patches were grown and screened again. A single colony from the fetal human liver cDNA library consistently gave a strong

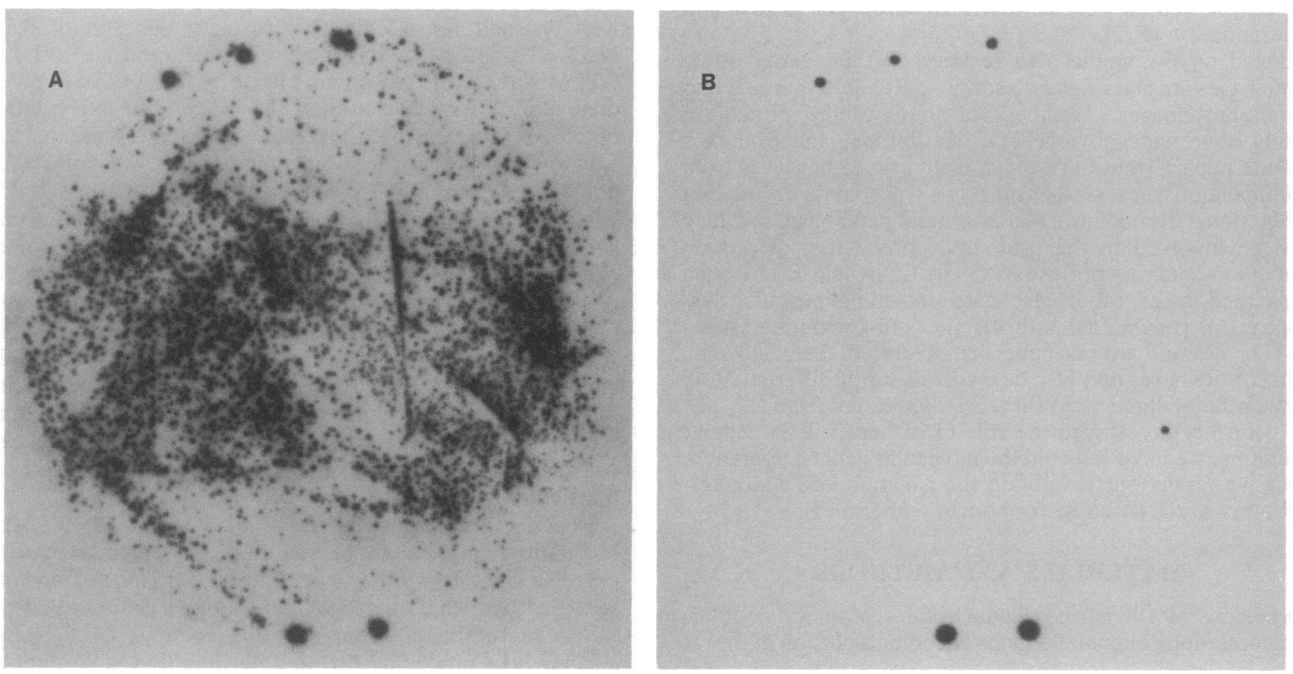

FIG. 2. Hybridization with the oligonucleotide probe in a high-density colony screening. The nitrocellulose filter was washed at 37°C (*A*) and subsequently at 45°C (*B*) after hybridization. The strong hybridization signal seen with the 45°C autoradiogram corresponded to the fetal human liver cathepsin B cDNA clone pCB-1. (The five dark spots on the edges are radioactive ink marks for orientation.)

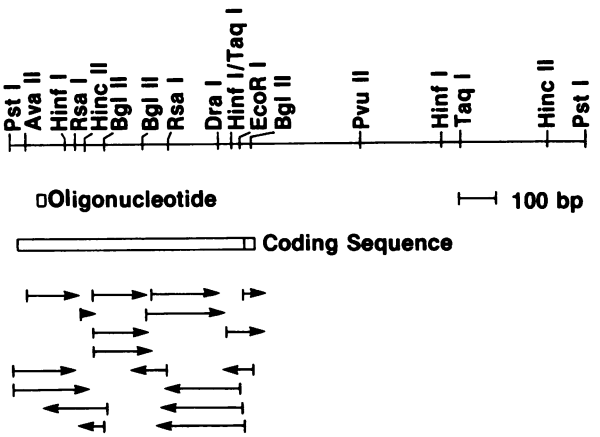

FIG. 3. Restriction map of the 1.54-kb *Pst* I insert of the cDNA clone pCB-1. Small bar denotes the oligonucleotide probe. Long bar denotes cathepsin B coding region together with the six-residue carboxyl-terminal extension. Horizontal arrows indicate the direction and extent of the chemical sequencing reactions for both anticoding and coding strands. A second *Hinf*I site lies 6 nucleotides in front of the *Hinf*I/*Taq* I site (not shown). bp, Base pairs.

hybridization signal after the procedures described above and thus was selected for further study. This clone was found to have a cathepsin B cDNA insert and has been designated pCB-1. (The clone corresponds to the strong hybridization spot shown in Fig. 2.)

Identification of a Human Cathepsin B cDNA Clone. Based on the published information on pKT218 and using pBR322 as a control, the pCB-1 plasmid DNA was analyzed by restriction enzyme digestion, Southern blotting, and hybridization to labeled oligonucleotide probes. Digestion with *Pst* I revealed a single insert of 1540 nucleotides. The insert was also found to have one *Eco*RI site, and labeled oligonucleotide probe hybridized to the 5′ or left fragment of the two *Eco*RI fragments. A restriction map of the insert is shown in Fig. 3.

DNA sequence analysis was performed for the coding region of the pCB-1 insert by using the Maxam and Gilbert chemical method. The various restriction sites that were used for end-labeling are shown in Fig. 3, and the nucleotide sequence of the coding region is presented in Fig. 4. The sequence of pCB-1 begins at residue 50 in the amino acid sequence and extends ≈880 nucleotides into the 3′ untranslated region. Compared with the published amino acid sequence of human liver cathepsin B by Ritonja *et al.* (31), the predicted 203 residues are identical except for one difference: asparagine instead of aspartate at position 147. Furthermore, the nucleotide sequence also discloses the presence of six additional amino acid residues at the carboxyl terminus in comparison with the published cathepsin B sequence. Such a six-residue extension was first reported for the rat liver

```
50                                                  60
Val Glu Val Ser Ala Glu Asp Leu Leu Thr Cys Cys Gly Ser Met Cys Gly Asp Gly Cys
GTG GAG GTG TCG GCG GAG GAC CTG CTC ACA TGC TGT GGC AGC ATG TGT GGG GAC GGC TGT

70                                                  80
Asn Gly Gly Tyr Pro Ala Glu Ala Trp Asn Phe Trp Thr Arg Lys Gly Leu Val Ser Gly
AAT GGT GGC TAT CCT GCT GAA GCT TGG AAC TTC TGG ACA AGA AAA GGC CTA GTT TCT GGT

90                                                  100
Gly Leu Tyr Glu Ser His Val Gly Cys Arg Pro Tyr Ser Ile Pro Pro Cys Glu His His
GGC CTC TAT GAA TCC CAT GTA GGG TGC AGA CCG TAC TCC ATC CCT CCC TGT GAG CAC CAC

110                                                 120
Val Asn Gly Ser Arg Pro Pro Cys Thr Gly Glu Gly Asp Thr Pro Lys Cys Ser Lys Ile
GTC AAC GGC TCC CGG CCC CCA TGC ACG GGG GAG GGA GAT ACC CCC AAG TGT AGC AAG ATC

130                                                 140
Cys Glu Pro Gly Tyr Ser Pro Thr Tyr Lys Gln Asp Lys His Tyr Gly Tyr Asn Ser Tyr
TGT GAG CCT GGC TAC AGC CCG ACC TAC AAA CAG GAC AAG CAC TAC GGA TAC AAT TCC TAC

150                                                 160
Ser Val Ser Asn Ser Glu Lys Asp Ile Met Ala Glu Ile Tyr Lys Asn Gly Pro Val Glu
AGC GTT TCT AAT AGT GAG AAG GAT ATC ATG GCT GAG ATC TAC AAA AAC GGC CCC GTG GAG

170                                                 180
Gly Ala Phe Ser Val Tyr Ser Asp Phe Leu Leu Tyr Lys Ser Gly Val Tyr Gln His Val
GGA GCT TTC TCT GTG TAT TCG GAC TTT CTG CTT TAC AAG TCA GGA GTG TAC CAA CAC GTC

190                                                 200
Thr Gly Glu Met Met Gly Gly His Ala Ile Arg Ile Leu Gly Trp Gly Val Glu Asn Gly
ACC GGA GAG ATG ATG GGT GGC CAT GCC ATC CGC ATC CTG GGC TGG GGA GTG GAG AAT GGC

210                                                 220
Thr Pro Tyr Trp Leu Val Ala Asn Ser Trp Asn Thr Asp Trp Gly Asp Asn Gly Phe Phe
ACA CCC TAC TGG CTG GTT GCC AAC TCC TGG AAC ACT GAC TGG GGT GAC AAT GGC TTC TTT

230                                                 240
Lys Ile Leu Arg Gly Gln Asp His Cys Gly Ile Glu Ser Glu Val Val Ala Gly Ile Pro
AAA ATA CTC AGA GGA CAG GAT CAC TGT GGA ATC GAA TCA GAA GTG GTG GCT GGA ATT CCA

250     252                     258
Arg Thr Asp Gln Tyr Trp Glu Lys Ile
CGC ACC GAT CAG TAC TGG GAA AAG ATC TAA
```

FIG. 4. Sequence of pCB-1 insert corresponding to amino acids 50–252 of human cathepsin B (31). Amino acids 253–258 are the six-residue carboxyl-terminal extension. Sequence was determined by the chemical cleavage method of Maxam and Gilbert (23, 24). Numbers refer to amino acid positions.

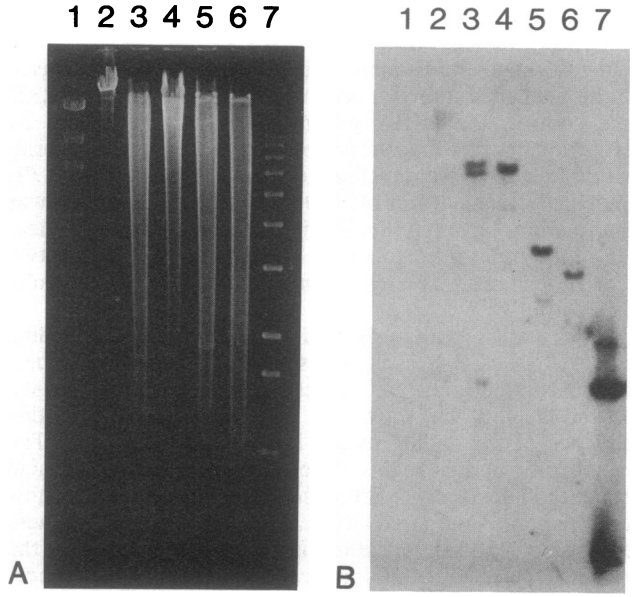

FIG. 5. Genomic representation of cathepsin B sequence. Human placental DNA (15 μg per sample) was digested with restriction enzymes (lanes 2–6 correspond to uncut DNA, and *Eco*RI, *Bam*HI, *Hin*dIII, and *Pst* I digests, respectively) and then separated on 0.8% agarose gel. (*A*) Gel banding pattern after ethidium bromide stain (molecular weight standards at lanes 1 and 7, *Hin*dIII digest of λ DNA, and 1-kb ladder from Bethesda Research Laboratories, respectively). (*B*) Autoradiogram after Southern blotting and hybridization using nick-translated *Pst* I/*Eco*RI insert fragment of pCB-1 (lane 7 shows some hybridization to pBR322 sequences within the 1-kb ladder).

cathepsin B by San Segundo *et al.* (32). The first three of the six residues are identical for both human and rat liver cathepsin B and this region may be conserved for the purpose of post-translational modification. Thus, our sequence confirms that of San Segundo *et al.* and extends the finding to human cathepsin B.

Cathepsin B-Like Sequences in Human Genomic DNA. To test the number of copies of cathepsin B gene sequences in human DNA, we analyzed human placental DNA by Southern blotting after digestion with various restriction enzymes. The genomic fragments that hybridized to the cathepsin B probe (nick-translated *Pst*/*Eco* 593-nucleotide DNA fragment) are shown in Fig. 5. Whereas the *Pst* I digest showed a single band at 3.1 kilobases (kb), both *Bam*HI and *Hin*dIII digests showed one intense and one faint band each (6.0 and 4.8 kb for *Bam*HI, 3.7 and 2.8 kb for *Hin*dIII). The *Eco*RI digest showed two strong bands (6.1 and 5.8 kb) and a faint band (1.7 kb). Intense hybridization to single bands indicates that there may be only one copy of the human cathepsin B gene per haploid genome. The faint bands may arise because of cross-hybridization to genes of related cysteine proteinases.

DISCUSSION

In the present study, we report the isolation of a cDNA clone for human cathepsin B from a human fetal liver cDNA library. Both fetal and adult human liver were screened because mammalian livers have traditionally been a source for the purification of the cathepsin B enzyme (28, 31, 33). The probe was an oligonucleotide mixture composed of 128 different 17-mers that correspond to amino acid residues 69–74 of rat liver cathepsin B. This probe was selected because it had been demonstrated previously that 23-mers with 384 different members were successful in screening for the complement C4 protein clones (34). Although we were successful in finding the human liver cathepsin B plasmid pCB-1, there are alternative approaches. Instead of an oligonucleotide mixture, deoxyinosine can be placed as substitutes at ambiguous codon positions (35), or a single longer "optimal" probe can be made based on codon utilization frequency (36). We have used a washing temperature that was derived from the calculated dissociation temperature of the oligonucleotide mixture (30). However, hybridization independent of base composition can be achieved, provided that tetramethylammonium chloride is included in the wash solution (37).

Recently, San Segundo *et al.* (32) isolated two rat liver cathepsin B cDNA clones from a rat liver cDNA library constructed in λgt11. These investigators reported no success with short oligonucleotide mixtures of 14-mers and 17-mers, and they had to synthesize longer probes of 50-mer and 54-mer for their screening. In contrast, we used an oligonucleotide mixture, a 17-mer with 128 members, for our experiments.

Both the data of San Segundo *et al.* (32) and ours show that rat and human cathepsin B have a six-residue carboxyl-terminal extension, respectively. It is interesting to note that Faust *et al.* (38) found that human cathepsin D, an aspartic proteinase, also has a two-residue extension at the carboxyl terminus by DNA sequencing. Thus both cathepsins B and D seem to undergo carboxyl-terminal post-translational modifications. Indeed, such proteolytic processing during cathepsin D biosynthesis in porcine kidney cells has been reported by Erickson and Blobel (39).

Results presented here and those of others (29, 31, 32) lead one to conclude that human and rat liver cathepsin B are very similar, at both the protein and the nucleotide level. Both proteins are 252 residues long and contain only 41 different residues (31). Furthermore, Kirschke *et al.* (40) reported that there are almost no species differences in the catalytic properties of rat, bovine, and human cathepsin B, and that a comparison of the substrate specificity of cathepsin B from the three species also showed no major difference. In summary, cathepsin B must be well conserved among the mammalian species.

Cathepsin B is known to have isozymes (41). For example, a minor cathepsin B isozyme with a different carbohydrate side chain, and at least one different amino acid residue (serine instead of cysteine at position 117) has been reported for porcine spleen (42). The recently selected cathepsin B cDNA clones may be useful as probes for determining the different cathepsin B isozyme species.

Cathepsin B has been implicated in tumor metastasis (8–11). However, there is controversy as to whether the tumor-associated cathepsin B-like proteinase is just a high molecular weight latent cathepsin B (43) or is a related enzyme derived from a different gene locus (44). With the cathepsin B cDNA probes on hand, this problem can be addressed.

In addition to cathepsin B, mammalian tissues have other cysteine proteinases. Examples include the low molecular weight cathepsin B-like enzymes, cathepsins H, L, and S (40, 45), and the high molecular weight cysteine proteinases, cathepsins J and K (46). The available cathepsin B probes may be useful in screening for recombinant clones containing these other cysteine proteinases.

In addition to the isolation of the rat (32) and human liver cathepsin B cDNA clones, several reports of the isolation of other cysteine proteinase clones have been made recently. Gay and Walker (47) found a partial bovine cathepsin cDNA clone unexpectedly during their search for the bovine mitochondrial H^+-ATPase. Williams *et al.* (13) were analyzing the developmentally regulated cDNA clones of the cellular slime mold. They sequenced one clone and, by homology searches, discovered that they had the cysteine proteinase 1 of *D. discoideum*. Ohno *et al.* (48) found cDNA clones for chicken

calpain, the calcium-dependent protease. DNA sequence analysis revealed that calpain contains a domain related to cysteine proteinases and another with homology to the calmodulin-like calcium-binding proteins. This finding suggests that calpains may have arisen as a result of the fusion of genes for proteins of completely different functions and evolutionary origins. Sequence of a cDNA insert derived from a developmentally regulated barley cysteine proteinase, aleurain, has shown that the enzyme is closely related to mammalian cathepsin H (49). This result indicates that similar cysteine proteinases may be involved in important roles in both plants and animals.

The study of cathepsin B and other cysteine proteinases by the molecular genetic approach will be useful in investigating the regulation of proteinase gene expression during cell growth and differentiation and during cellular pathological states. Isolation of cathepsin B cDNA clones is a first step toward this goal.

We thank Ms. Marion Chan and Drs. Gillian Air, John Bonner, Christian de Duve, Daniel Jones, and Wallace LeStourgeon for their support. We are grateful to Dr. Stuart Orkin for cDNA libraries, Dr. David Bishop for oligonucleotide synthesis, Dr. Peter Melera for human placental DNA, and Ms. Sandra Roper for manuscript preparation. D.F. wishes to thank Ms. Micaela Caserta, Mr. Petros Hantzopoulos, and Drs. Jeffrey Engler, Carol Gibbs, David Irving, and Michael Kilpatrick for their advice and help. This work was supported by Grants PCM-8409201 and PCM-8308644 from the National Science Foundation to D.F. and R.D.W., respectively, and GM-30822 from the National Institutes of Health to R.D.W.

1. Neurath, H. (1984) *Science* **224,** 350–357.
2. North, M. J. (1982) *Microbiol. Rev.* **46,** 308–340.
3. Barrett, A. J. (1984) *Biochem. Soc. Trans.* **12,** 899–902.
4. Khairallah, E. A., Bond, J. S. & Bird, J. W. C. (1985) *Intracellular Protein Catabolism* (Liss, New York).
5. Docherty, K., Hutton, J. C. & Steiner, D. F. (1984) *J. Biol. Chem.* **259,** 6041–6044.
6. Katunuma, N. & Kominami, E. (1983) *Curr. Top. Cell. Regul.* **22,** 71–101.
7. Mort, J. S., Recklies, A. D. & Poole, A. R. (1984) *Arthritis Rheum.* **27,** 509–515.
8. Strauli, P., Barrett, A. J. & Baici, A. (1980) *Proteinases and Tumor Invasion* (Raven, New York).
9. Sloane, B. F., Honn, K. V., Sadler, J. G., Turner, W. A., Kimpson, J. J. & Taylor, J. D. (1982) *Cancer Res.* **42,** 980–986.
10. Koppel, P., Baici, A., Keist, R., Matzka, S. & Keller, R. (1984) *Exp. Cell Biol.* **52,** 293–299.
11. Mullins, D. E. & Rifkin, D. B. (1985) in *Developments in Cell Biology: Secretory Processes*, eds. Dean, R. T. & Stahl, P. (Butterworths, London), Vol. 1, pp. 159–177.
12. Fong, D. & Bonner, J. T. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 6481–6485.
13. Williams, J. G., North, M. J. & Mahbubani, H. (1985) *EMBO J.* **4,** 999–1006.
14. Stawinski, J., Hozumi, T., Narang, S. A., Bahl, C. P. & Wu, R. (1977) *Nucleic Acids Res.* **4,** 353–371.
15. Duckworth, M. L., Gait, M. J., Goelet, P., Hong, G. F., Singh, M. & Titmas, R. C. (1981) *Nucleic Acids Res.* **9,** 1691–1706.
16. Michelson, A. M., Markham, A. F. & Orkin, S. H. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 472–476.
17. Woods, D. E., Markham, A. F., Ricker, A. T., Goldberger, G. & Colten, H. R. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 5661–5665.
18. Hanahan, D. & Meselson, M. (1980) *Gene* **10,** 63–67.
19. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
20. Wallace, R. B., Johnson, M. J., Hirose, T., Miyake, T., Kawashima, E. H. & Itakura, K. (1981) *Nucleic Acids Res.* **9,** 879–894.
21. Birnboim, H. C. & Doly, J. (1979) *Nucleic Acids Res.* **7,** 1513–1523.
22. Talmadge, K. & Gilbert, W. (1980) *Gene* **12,** 235–241.
23. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560.
24. Bencini, D. A., O'Donovan, G. A. & Wild, J. R. (1984) *BioTechniques* **2,** 4–5.
25. Yousaf, S. I., Carroll, A. R. & Clarke, B. E. (1984) *Gene* **27,** 309–313.
26. Southern, E. M. (1975) *J. Mol. Biol.* **98,** 503–517.
27. Fong, D., Wallach, M., Keithly, J., Melera, P. W. & Chang, K. P. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 5782–5786.
28. Takio, K., Towatari, T., Katunuma, N. & Titani, K. (1980) *Biochem. Biophys. Res. Commun.* **97,** 340–346.
29. Takio, K., Towatari, T., Katunuma, N., Teller, D. C. & Titani, K. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 3666–3670.
30. Meinkoth, J. & Wahl, G. (1984) *Anal. Biochem.* **138,** 267–284.
31. Ritonja, A., Popovic, T., Turk, V., Widenmann, K. & Machleidt, W. (1985) *FEBS Lett.* **181,** 169–172.
32. San Segundo, B., Chan, S. J. & Steiner, D. F. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 2320–2324.
33. Barrett, A. J. (1973) *Biochem. J.* **131,** 809–822.
34. Whitehead, A. S., Goldberger, G., Woods, D. E., Markham, A. F. & Colten, H. R. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 5387–5391.
35. Ohtsuka, E., Matsuki, S., Ikehara, M., Takahashi, Y. & Matsubara, K. (1985) *J. Biol. Chem.* **260,** 2605–2608.
36. Lathe, R. (1985) *J. Mol. Biol.* **183,** 1–12.
37. Wood, W. I., Gitschier, J., Lasky, L. A. & Lawn, R. M. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 1585–1588.
38. Faust, P. L., Kornfeld, S. & Chirgwin, J. M. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 4910–4914.
39. Erickson, A. H. & Blobel, G. (1983) *Biochemistry* **22,** 5201–5205.
40. Kirschke, H., Locnikar, P. & Turk, V. (1984) *FEBS Lett.* **174,** 123–127.
41. Smith, R. E. (1984) *J. Histochem. Cytochem.* **32,** 1265–1274.
42. Takahashi, T., Schmidt, P. G. & Tang, J. (1984) *J. Biol. Chem.* **259,** 6059–6062.
43. Mort, J. S., Leduc, M. S. & Recklies, A. D. (1983) *Biochim. Biophys. Acta* **755,** 369–375.
44. Olstein, A. D. & Liener, I. E. (1983) *J. Biol. Chem.* **258,** 11049–11056.
45. Lynn, K. R. & Labow, R. S. (1984) *Can. J. Biochem. Cell. Biol.* **62,** 1301–1308.
46. Liao, J. C. R. & Lenney, J. F. (1984) *Biochem. Biophys. Res. Commun.* **124,** 909–916.
47. Gay, N. J. & Walker, J. E. (1985) *Biochem. J.* **255,** 707–712.
48. Ohno, S., Emori, Y., Imajoh, S., Kawasaki, H., Kisaragi, M. & Suzuki, K. (1984) *Nature (London)* **312,** 566–570.
49. Rogers, J. C., Dean, D. & Heck, G. R. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 6512–6516.