Identification, Cloning, and Characterization of Cystatin M, a Novel Cysteine Proteinase Inhibitor, Down-regulated in Breast Cancer*

(Received for publication, April 11, 1996, and in revised form, September 26, 1996)

Georgia Sotiropoulou‡§, Anthony Anisowicz‡, and Ruth Sager‡§

From the ‡Division of Cancer Genetics, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts 02115

A novel human cystatin gene was identified in a differential display comparison aimed at the isolation of transcriptionally regulated genes involved in invasion and metastasis of breast cancer. Messenger RNAs from primary and metastatic tumor cells isolated from the same patient were compared. A partial cDNA was isolated that was expressed in the primary tumor cell line but not in the metastatic line. The full-length cDNA was cloned and sequenced, and the inferred amino acid sequence was found to encode a novel protein, which we named cystatin M, with 40% homology to human family 2 cystatins and similar overall structure. Cystatin M is expressed by normal mammary cells and a variety of human tissues.

The mature cystatin M protein was produced in Escherichia coli as a glutathione S-transferase fusion protein using the pGEX-2T expression system and purified by affinity chromatography. The cystatin M fusion protein displayed inhibitory activity against papain. Native cystatin M protein of approximately 14.5 kDa was secreted and was immunoprecipitated from supernatants of mammary cell cultures using affinity-purified antisera raised against recombinant cystatin M. An N-glycosylated form of cystatin M of 20–22 kDa was co-immunoprecipitated and accounted for about 30–40% of total cystatin M protein. Both forms of native cystatin M also occurred intracellularly. Consistent with the mRNA differential expression, no cystatin M protein was detected in metastatic mammary epithelial tumor cells. Loss of expression of cystatin M is likely associated with the progression of a primary tumor to a metastatic phenotype.

Metastasis of a primary tumor is a multistage process involving aberrant functions of the tumor cell including increased local proteolysis, degradation of extracellular matrix components, invasion, migration, adhesion to the vascular basement membrane, migration through the vasculature, and proliferation at distant sites (1–4). Therefore, changes in the expression of multiple genes probably occur before tumor cells acquire the potential to metastasize. The identification of genes whose changes in expression determine the metastatic phenotype is essential in understanding the molecular mechanisms underlying metastasis and in the design of novel therapies to arrest progression of primary cancers.

In this study, we have applied differential display (5, 6) to follow changes in gene expression that arise during progression of a primary mammary cancer to the metastatic phenotype (7). Differential display is a PCR1-based method of differential expression cloning and offers the advantage of side-by-side comparisons of mRNAs from closely related cell populations displayed on sequencing gels as partial cDNAs. A novel cysteine proteinase inhibitor, cystatin M, was identified as being down-regulated in metastatic cells as compared with cells from the primary tumor.

Cystatins are endogenous inhibitors of mammalian lysosomal cysteine proteinases, such as cathespins B, L, H, and S, and the plant cysteine proteinases papain, acardin, and ficin. They function both intracellularly and extracellularly (8, 9). All inhibitory cystatins display structural and functional similarities and are members of a single protein superfamilly comprising three distinct families of closely related proteins: stefins, cystatins, and kininogens (reviews 8–12). Cystatin M is most closely related to family 2 cystatins, which consist of about 120 amino acid residues, and contain one or two disulfide loops near their C-terminal domain. Cystatins are postulated to control the activities of cysteine proteinases, which regulate protein turnover, as well as the processing of proenzymes and prohormones (8–12). Cystatins bind to their target peptidases very tightly but reversibly, forming high affinity (Kd = 10⁻⁹ to 10⁻¹² M) equimolar complexes in competition with their substrates (13, 14).

The mammalian lysosomal cathespins B, L, and S are inhibited to varying degrees by family 2 cystatins (15, 16). Cathespins B and L have been implicated in invasion and metastasis of tumor cells (17–19). Increased cathespin B and L activities have been reported in a variety of human and animal malignant tumors, which may reflect alterations in their expression, activation and processing, intracellular trafficking and delivery, as well as decreased regulation of these proteinases due to the reduced expression and activity of their endogenous inhibitors (17–19).

The isolation and characterization of cystatin M is described in this report. Our results suggest that loss of expression of this cysteine proteinase inhibitor in metastatic tumor cells probably

* This work was supported in part by National Institutes of Health Grant CA61253 (to R.S) and an award from the Fulbright Program (CIES, Council for International Exchange of Scholars) (to G.S.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) U62800.

‡ Present address: Dept. of Pharmacy, School of Health Sciences, University of Patras, Rion 261 10, Patras, Greece. Fax: 30-61-994941; E-mail: G.Sotiropoulou@upatras.gr.

‡ To whom correspondence should be addressed: Division of Cancer Genetics, Dana-Farber Cancer Institute, Harvard Medical School, 44 Binney St., Boston, MA 02115. Tel.: 617-632-3330; Fax: 617-632-4680.

1 The abbreviations used are: PCR, polymerase chain reaction; DD, differential display; rGST, recombinant glutathione S-transferase; rGST-cystatin M, recombinant GST-cystatin M fusion protein; Z-Phe-Arg-NHMec, carbobenzyoxyl-L-phenylalanyl-L-arginine-4-methylcoumaryl-7-amide; PAGE, polyacrylamide gel electrophoresis; DFCI, Dana-Farber Cancer Institute; DTT, dithiothreitol; kb, kilobase pair(s).
contributes to increased proteolytic potential, a feature of the metastatic phenotype. An invasion/metastasis suppressor function of cystatin M along the metastatic cascade is proposed.

**MATERIALS AND METHODS**

**Cell Strains and Cell Lines**—Normal human mammary epithelial cell strains (81N, 76N, and 70N) derived from reduction mammaplasties, as well as primary (21NT, 21PT) and metastatic (21MT-1, 21MT-2) tumor cell lines were established in this laboratory (7, 20). Metastatic tumor mammary epithelial cell lines MCF7, BT474, BT549, 47TD, ZK-75-1, MDA-MB-157, MDA-MB-231, MDA-MB-361, MDA-MB-435, and MDA-MB-436 were obtained from the American Tissue Culture Collection (ATCC, Rockville, MD). All cell cultures were routinely propagated in DFCI-1 medium (20). Cultures were grown in P100 dishes and harvested at about 75% confluence for RNA isolation and near confluence for DNA isolation. Immortal mammary epithelial cells were obtained by transfection of 76N cells with a plasmid containing the human papilloma virus-16 E6 gene (21). *Echerichia coli* K12 bacteria strain XL-1 Blue was obtained from Stratagene (La Jolla, CA). Tissue culture medium components were purchased from Life Technologies, Inc. or HyClone (Logan, UT).

**Differential Display of mRNA**—Total cellular RNAs (50 μg) from exponentially growing cell cultures were treated with DNase I in the presence of 10 mM ribonuclease inhibitor, to remove any residual DNA contamination as described elsewhere (22). Then, RNAs were extracted with phenol/chloroform, precipitated with ethanol, and redissolved in diethyl pyrocarbonate-treated water. The RNAs were subsequently reverse-transcribed by using a 3′-anchoring primer *Tn* (where *M* represents G, C, or A). The resultant partial cDNAs were amplified by PCR in the presence of 5′-S-dATP using *Tn*MA and OP2A (GCTTCCCTGAC) an arbitrary 10-mer primer, as the 5′-end primer (Operon Technologies, Inc.) and compared side-by-side on a 6% acrylamide/urea sequencing gel. These partial cDNA fragments correspond to the 3′-end of the mRNAs (5). A differentially displayed cDNA of ~0.3 kb (named 6A2 because it was amplified by using the T12MA and OP6A primers) was recovered from the dried gel, purified by a Millipore Ultracruz micro unit, reamplified by PCR, *P* labeled by the oligo-labeling method (23), and used as a probe for hybridization of Northern blots.

**Cloning and Sequencing of cDNAs and Data Base Analysis**—The 6A2 partial cDNA obtained from DD was cloned into the PCR2 vector using the TA cloning system (Invitrogen); a clone containing an insert that hybridized to a 0.6-kb transcript was isolated and sequenced on both strands with T7 and SP6 primers. A cDNA library from 21PT cells constructed in Lambda Zap II (Stratagene, San Diego, CA) was screened using the cloned PCR product as a probe; several full-length cDNA clones were isolated. Their differential expression was confirmed by Northern hybridizations of the same RNA samples used for DD, as well as RNAs from a number of normal and tumor cell lines. Three distinct full-length cDNA clones were sequenced on both strands. Sequencing was performed using an ABI automated sequencer. Model 373A, in the Molecular Biology Core Facility of the DFCI. Oligonucleotides were synthesized at the Molecular Biology Core Facility of the DFCI and by Amitof Inc. (Cambridge, MA). The BLAST algorithm was used for nucleic acid sequence comparisons (24). Protein sequence comparisons were performed by GCG with final alignments by PILEUP and PRETTYPLOT (24).

**Northern and Southern Analysis**—Total cellular RNA was purified by standard guanidinium isothiocyanate and cesium chloride centrifugation and analyzed as described (25). Genomic DNA was isolated and hybridized by standard methods (25). Hybridizations were performed in formamide at 37 °C overnight. The blots were washed at 65 °C for 1 h in 2 × SSC containing 0.1% SDS. The tissue blot (Human MTN Blot, Clontech, number 7760-1) was washed at 65 °C for 1 h in 0.5 × SSC containing 0.1% SDS. For a loading control, Northern blots were stripped and rehybridized to 36B4, a gene encoding a ribosomal protein, whose expression is not affected by growth conditions or estrogen treatment, as well as primary (21NT, 21PT) and metastatic (21MT-1, 21MT-2) tumor cell lines. Three distinct full-length cDNA clones were sequenced on both strands. Sequencing was performed using an ABI automated sequencer. Model 373A, in the Molecular Biology Core Facility of the DFCI. Oligonucleotides were synthesized at the Molecular Biology Core Facility of the DFCI and by Amitof Inc. (Cambridge, MA). The BLAST algorithm was used for nucleic acid sequence comparisons (24). Protein sequence comparisons were performed by GCG with final alignments by PILEUP and PRETTYPLOT (24).

**Production and Purification of Recombinant GST-Cystatin M**—The original vector (PGEX-2T) as well as the recombinant plasmid (PGEX-2T/cystatin M) were transformed into *E. coli* XL-1 Blue bacteria and propagated in Luria broth (LB) (25) in the presence of 100 μg/ml ampicillin for selection of the cells transformed with the PGEX-2T/cystatin M expression plasmid. The expression of recombinant protein was confirmed in exponentially growing bacteria (A600 = 0.8–1.0) with 0.2 μM isopropyl-1-thio-β-D-galactopyranoside for 1.5 h at 37 °C with vigorous agitation. The bacteria were harvested by centrifugation, washed twice with MTPBS (150 mM NaCl, 16 mM Na2HPO4, 4 mM NaH2PO4, pH 7.4) and resuspended in lysis buffer, MTPBS containing 1 mM DTT, 1 mM phenylmethylsulfonyl fluoride, and 2% Triton X-100. Cells were lysed on ice by mild sonication, and the suspension was centrifuged at 14,500 × g for 15 min. All subsequent purification steps were carried out at 4 °C. The fusion protein was purified from the clear lysate on glutathione-agarose beads under nondenaturing conditions, with an estimated yield of 3–5 mg per liter of bacterial culture. The glutathione-agarose column was washed with MTPBS containing 350 mM NaCl, and the fusion protein was eluted with 50 mM Tris-HCl, pH 8.0, containing 5 mM reduced glutathione, in order to cleave GST. The purified rGST-cystatin M was dialyzed against MTPBS containing 10% glycerol, sterilized by filtration through 0.22-μm filters (Costar, Cambridge, MA), and stored at −20 °C. The GST carrier was proteolytically cleaved from the fusion protein with thrombin and removed along with any uncleaved fusion protein by absorption on glutathione-agarose. Thrombin reaction was carried out at room temperature in the presence of 150 mM NaCl, 2.1 mM CaCl2, 0.2 NIH thrombin units/ml, and 1 mg/ml fusion protein. The reaction was stopped with 0.1 mM EGTDA, and cleavage was monitored by SDS-PAGE. Cleaved protein was dialyzed against MTPBS and stored at −20 °C. The purity of rGST-cystatin M was assessed by staining with Coomassie Brilliant Blue R-250 and Silver (Silver Stain Plus, Bio-Rad). The concentration of the purified protein was determined by Bradford assay using γ-globulin as a standard. Glutathione-agarose resin and other chemicals were purchased from Sigma, unless otherwise indicated. The reagents for SDS-PAGE and the Bradford assay were purchased from Bio-Rad.

**Reduction and Alkylation of Recombinant Cystatin M**—The protein (0.2 mg/ml) was denatured and reduced in 25 mM Tris-HCl, pH 8.0, containing 1 mM EDTA, 15 mM dithiothreitol, and 8 μl urea, at 37 °C for 30 min. Then the protein was alkylated with 50 mM iodoacetamide (Sigma) at room temperature for 20 min. Electrophoresis of the modified protein was carried out in nonreducing gels containing 7.5% acrylamide and 8 μl urea in 0.375 M Tris, pH 8.5.

**Polyclonal Antibodies**—The purified fusion protein was used to immunize New Zealand White rabbits. Antiserum raised against the fusion protein and cystatin M cleaved from the fusion protein by thrombin was used against GST and, subsequently, against GST-cystatin M-agarose column (28). The purified antibody was dialyzed against phosphate-buffered saline containing 50% glycerol, adjusted to 0.02% NaN3, and stored at 4 °C. The anti-cystatin M antibody specifically recognized cystatin M recombinant and native protein and did not cross-react with either cystatin C, stefin A, or stefin B on Western blots containing 10 μg of recombinant proteins.

**SDS-PAGE, Western Blotting, and Immunoprecipitation**—The recombinant protein and biological samples were denatured in SDS-PAGE sample buffer at 90 °C for 5 min and analyzed on 15% polyacrylamide gels. For immunoblot detection, the proteins were transferred to polyvinylidene difluoride (0.2 micron, Bio-Rad) and reacted with polyclonal antiserum (1:500) and preimmune serum (1:500). Anti-rabbit IgG horseradish peroxidase-linked whole antibody was used as secondary antibody (1:2000), and immunoreactive proteins were detected with the enhanced chemiluminescence system (Amersham Corp.). Transfer and quantitation of proteins were assessed by staining with 0.1% w/v amido black in 25% isopropyl alcohol and 10% acetic acid. Destaining solution contained 50% methanol and 7.5% acetic acid in H2O.

For the preparation of whole cell lysates, each cell was washed with phosphate-buffered saline, resuspended (8–10 × 106 cells/ml) in lysis buffer (50 mM Tris, pH 8.0, containing 120 mM NaCl, 0.5% Nonidet P-40, 5 μg/ml aprotinin, 50 μg/ml phenylmethylsulfonyl fluoride, 5 μg/ml leupeptin, 0.2 mM sodium orthovanadate, and 100 mM NaF). Lysed cells were rocked for 30 min and centrifuged at 14,500 × g for 15 min, and the supernatants were assayed immediately. All steps were performed at 4 °C.
FIG. 1. Discovery of cystatin M by two-way differential display. Total RNA from 21PT primary tumor cell line (P) and 21MT-1 metastatic tumor cell line (M) established from the same patient were reverse-transcribed with T<sub>7</sub>oligo, T<sub>3</sub>oligo, T<sub>7</sub>oligo, T<sub>3</sub>oligo, and T<sub>7</sub>-antichromatids anchoring primers (marked as G, A, T, and C, in 1st and 2nd, 3rd and 4th, 5th and 6th, and 7th and 8th lanes, respectively). The position of the 6A2 cDNA (~300 base pairs) present only in 21PT and absent in 21MT-1 cell line is marked with an asterisk below. Other examples of differentially displayed cDNAs are shown with arrowheads.

For immunoprecipitation, 3 ml of cell culture supernatant or 250 μl of fresh whole cell lysate were diluted 1:1 with 20 mM Tris, pH 8.0, containing 100 mM NaCl, 1 mM EDTA, and 0.5% Nonidet P-40; preimmune serum or affinity-purified antisera were added, respectively, to each sample at a 1:500 dilution, and the samples were incubated with mild agitation for 1 h. The immunoprecipitated proteins were then precipitated from 3 ml of 21 PT cell culture supernatant with ammonium sulfate, which was then removed by dialysis against 20 mM sodium phosphate, pH 7.2. The protein was digested by a 50-μl reaction by incubation at 37 °C for 24 h at 20 mM Na<sub>2</sub>HPO<sub>4</sub>, pH 7.2, containing 10 mM Na<sub>2</sub>HPO<sub>4</sub>, 50 mM EDTA, and 0.5% v/v Nonidet P-40, and 0.25 units of N-glycosidase F (Boehringer Mannheim) (29). Deglycosylation of cystatin M did not require previous denaturation of the protein. Deglycosylation of Native Cystatin M—Native cystatin M protein was precipitated from 3 ml of 21 PT cell culture supernatant with ammonium sulfate, which was then removed by dialysis against 20 mM Na<sub>2</sub>HPO<sub>4</sub>, pH 7.2. The protein was digested by a 50-μl reaction by incubation at 37 °C for 24 h at 20 mM Na<sub>2</sub>HPO<sub>4</sub>, pH 7.2, containing 10 mM Na<sub>2</sub>HPO<sub>4</sub>, 50 mM EDTA, and 0.5% v/v Nonidet P-40, and 0.25 units of N-glycosidase F (Boehringer Mannheim) (29). Deglycosylation of cystatin M did not require previous denaturation of the protein.

Analysis of Cystatin M Inhibitory Activity by Papain Assays—Papain activity was assayed in 125 mM phosphate buffer, pH 6.8, containing 4 mM DTT, 1 mM EDTA, and 0.05% Brij-35 using the fluorogenic synthetic SDS-PAGE. Proteinase inhibitors were added to the supernatants immediately after collection to prevent proteolytic degradation of cystatin M.

Deglycosylation of Native Cystatin M—Native cystatin M protein was precipitated from 3 ml of 21 PT cell culture supernatant with ammonium sulfate, which was then removed by dialysis against 20 mM Na<sub>2</sub>HPO<sub>4</sub>, pH 7.2. The protein was digested by a 50-μl reaction by incubation at 37 °C for 24 h at 20 mM Na<sub>2</sub>HPO<sub>4</sub>, pH 7.2, containing 10 mM Na<sub>2</sub>HPO<sub>4</sub>, 50 mM EDTA, and 0.5% v/v Nonidet P-40, and 0.25 units of N-glycosidase F (Boehringer Mannheim) (29). Deglycosylation of cystatin M did not require previous denaturation of the protein.

Deglycosylation of Native Cystatin M—Native cystatin M protein was precipitated from 3 ml of 21 PT cell culture supernatant with ammonium sulfate, which was then removed by dialysis against 20 mM Na<sub>2</sub>HPO<sub>4</sub>, pH 7.2. The protein was digested by a 50-μl reaction by incubation at 37 °C for 24 h at 20 mM Na<sub>2</sub>HPO<sub>4</sub>, pH 7.2, containing 10 mM Na<sub>2</sub>HPO<sub>4</sub>, 50 mM EDTA, and 0.5% v/v Nonidet P-40, and 0.25 units of N-glycosidase F (Boehringer Mannheim) (29). Deglycosylation of cystatin M did not require previous denaturation of the protein.

Cloning and Sequencing of a Full-length cDNA Clone—A 21PT cDNA library was screened with the cloned partial cDNA obtained from DD. Several positive clones were selected, hybridized to total RNAs from normal and tumor cell lines, and all displayed confirmatory differential expression on Northern blots as on the DD gel. The longest clones were sequenced on both strands. Sequence comparison with the Genbank database using the BLAST program (24) revealed regions of homology to cystatins.

RESULTS

Differential Expression Cloning of a cDNA Encoding a Novel Human Cystatin—Messenger RNAs from a mammary epithelial primary tumor cell line, 21PT, and from a metastatic cell line, 21MT-1, derived from the same patient were compared by DD. Each lane contained 50–100 bands most of which were similar in size and intensity between the two cell populations. A small number of bands (1–2%) appeared in only one of the two lanes. A cDNA of about 0.3 kb was detected, which was present in 21PT primary, but absent in 21MT-1 metastatic cells (Fig. 1, lanes A, P, and M, respectively). This partial PCR product was used as a probe to hybridize Northern blots containing the same RNAs used for DD. A differentially expressed transcript of 0.6 kb was detected. The partial cDNA obtained from DD was cloned into a plasmid vector; Northern hybridizations were repeated with the cloned cDNA as a probe and sequenced on both strands. Sequence comparison with the Genbank data base using the BLAST program (24) revealed regions of homology to cystatins.

Primary Structure Analysis of Cystatin M and Alignments with Other Cystatins—Assuming that translation starts at the first ATG codon which lies in a Kozak consensus sequence, the full-length cystatin M cDNA contains an open reading frame of 447 nucleotides (24–470), a short 5'-untranslated sequence (1–23), and a 3'-untranslated sequence of 128 nucleotides, with...
Cystatin M, a Novel Cysteine Proteinase Inhibitor

Cystatin M, a novel cysteine proteinase inhibitor, shares three conserved domains with other family 2 cystatins. The homology between cystatins from different species is high, indicating a possible evolutionary relationship. The conserved cystatin motifs are underlined in bold. Sequences were retrieved from Genbank using the following accession numbers: cystatin C (A33400) (30), cystatin D (A47142) (31), cystatin S (S17667) (32), cystatin SN (A29632) (33), cystatin SA (B29632) (33), and chicken cystatin (A34456) (34).

Fig. 3. Sequence alignment of cystatin M compared with other human family 2 cystatins and chicken cystatin. Preprotein sequences were compared in the above alignment. Dashes indicate gaps introduced to optimize the alignment. The numbering refers to cystatin M with number one being the first amino acid of the signal peptide. Amino acid residues conserved in any six of the seven sequences are marked with brackets. The conserved cystatin motifs are underlined in bold. A potential N-linked glycosylation site, Asn137-Asp138-Glu139-Gly140, is also conserved.

- **Cystatin M**: ALA GLY GLN LEU THR ... GLN 149
- **Cystatin C**: ALA GLY GLN LEU THR ... GLN 149
- **Cystatin S**: ALA GLY GLN LEU THR ... GLN 149
- **Cystatin SN**: ALA GLY GLN LEU THR ... GLN 149
- **Cystatin SA**: ALA GLY GLN LEU THR ... GLN 149
- **Chicken Cys**: ALA GLY GLN LEU THR ... GLN 149

The homology at the nucleotide level is 40–45% to all previously characterized cystatins and 42% to chicken cystatin.

Cystatin M shows the closest homology to cystatin C. The two proteins share 33% identical and 38% conserved amino acid residues. The homology between cystatins from different species, including chicken, mouse, rat, and fowl adder, is 39–48% for conserved amino acid residues. All previously characterized cystatins contain about 120 amino acid residues and two intrachain disulfide bridges. Cystatin M indeed contains four cysteine residues near the C-terminal domain, Cys38, Cys113, Cys126, and Cys146. Since cystatin M displays the characteristic structural features of family 2 cystatins, it constitutes a new member of this family. Following the international accepted nomenclature (35), this novel cystatin was designated cystatin M, because it was cloned from mammary epithelial cells.

Hopp and Woods' (36) hydrophilicity plot (not shown) revealed the presence of a hydrophobic sequence consisting of 20 residues containing only one charged amino acid in the N-terminal region (Arg3) and a Pro-Trp135 motif near the C-terminal end, conserved in all three normal mammary cell strains tested, 76N, 70N (Fig. 4), and 81N (not shown), but was absent in many metastatic mammary tumor cell lines including estrogen receptor (+) and (-) lines, BT549, MCF7, T47D, ZR-75–1, BT-474, MDA-MB-361 (Fig. 4), and 21MT-1, MDA-MB-157, MDA-MB-435, MDA-MB-436, while trace transcript levels were detected in MDA-MB-231 (not shown). Cystatin M mRNA was not expressed by 56NF (normal human fibroblasts), FS3 human foreskin fibroblasts (not shown), or by normal human leukocytes.

**Northern Analysis**—Northern blots containing total RNA from exponentially growing normal and tumor mammary epithelial cell lines were hybridized against a full-length cystatin M cDNA (Fig. 4). The cystatin M mRNA of 0.6 kb was detected in all three normal mammary cell strains tested, 76N, 70N (Fig. 4), and 81N (not shown), but was absent in many metastatic mammary tumor cell lines including estrogen receptor (+) and (-) lines, BT549, MCF7, T47D, ZR-75–1, BT-474, MDA-MB-361 (Fig. 4), and 21MT-1, MDA-MB-157, MDA-MB-435, MDA-MB-436, while trace transcript levels were detected in MDA-MB-231 (not shown). Cystatin M mRNA was not expressed by 56NF (normal human fibroblasts), FS3 human foreskin fibroblasts (not shown), or by normal human leukocytes (Fig. 4). Although all normal human mammary epithelial cell lines expressed a clearly detectable cystatin M transcript, its abundance was lower than in the overexpressing 21PT, 21NT, and 21MT-2 tumor cell lines. However, the highly invasive 21MT-1 cell line from the same tumor progression series did not express the cystatin M transcript. The cystatin M mRNA levels in human papilloma virus-immortalized normal 76N cells (21) were comparable with the levels of its expression in the corresponding normal cells (not shown).

**Southern Analysis**—A single major band hybridizing with the cystatin M full-length cDNA was detected in DNAs from a series of normal and tumor mammary epithelial cell lines.
cleaved with HindIII (~15.0 kb) (Fig. 5, upper) or EcoRI (~7.0 kb) (Fig. 5, lower). Similarly, NcoI (~3.1 and 1.1 kb) and PvuII (~2.9 and 2.8 kb) digests showed uniform patterns (not shown). Based on these results, the cystatin M gene does not appear grossly rearranged or deleted in tumor cell lines.

**Tissue Distribution of Cystatin M**—The expression of cystatin M was studied in normal human tissues (Fig. 6). Relatively high levels of cystatin M mRNA were present in placenta, lung, skeletal muscle, kidney, and pancreas. Transcripts larger in size were detected in skeletal muscle (1.0 kb) and kidney (0.85 kb) (Fig. 6). A second transcript of slightly larger size was detectable in all the above tissues. A low abundance message was seen in heart tissue. Whether cystatin M transcript is present in brain tissue is not conclusive from this blot, since this lane was significantly underloaded. Trace amounts of a larger transcript of 1.8 kb can be seen, but it is not clear whether this transcript originates from cystatin M or from a closely related gene. Cross-hybridization of cystatin M with cystatin C is unlikely, since the corresponding cDNAs displayed relatively low homology, and nucleotide sequence alignments showed that no extended contiguous stretches of homologous segments are present (not shown). Similarly to cystatin C, which is widely distributed in tissues and biological fluids (38, 39), cystatin M is expressed in many tissues. The secretory cystatins S, SN, and SA have been reported in saliva, seminal plasma, and tears (32, 33), whereas cystatin D displays a tissue-restricted expression to parotid gland (31).

**Cystatin M Is Down-regulated in Human Cancers**—The expression of cystatin M was tested in other human cancers. No cystatin M message was detected in the following cell lines: PC-3 prostate adenocarcinoma, A549 and Calu-1 lung carcinomas, MIA Pa-Ca-2 pancreatic carcinoma, A2058, G-361 and SKME30 malignant melanomas, T24 bladder transitional cell carcinoma, MIA Pa-CA-2 pancreatic carcinoma, A2058, G-361 and PC-3 prostate adenocarcinoma, A549 and Calu-1 lung carcinoma, SCC-25 tongue squamous cell carcinoma, HuTu80 duodenal adenocarcinoma, OAT4 small cell lung carcinoma, SCC-25 tongue squamous cell carcinoma, whereas a relatively low message was detected in WiDr and SW480 colon adenocarcinomas (not shown). These results indicate that cystatin M might be down-regulated in other epithelial cancers, although this speculation should be confirmed by studies employing matched normal and tumor cells.

**Isolation and Characterization of Recombinant Cystatin M**—The cDNA encoding the mature cystatin M protein was amplified and expressed in *E. coli* as a GST-cystatin M fusion protein (27). Single-stranded cDNA from the full-length clone was PCR-amplified using a pair of gene-specific synthetic oligonucleotides corresponding to sequences on the sense strand upstream to the ATG start site and to the antisense strand downstream to the stop codon. The amplified region of the cDNA sequence does not contain the hydrophobic signal peptide, and Leu"^22" is its N-terminal amino acid. A single PCR product of the anticipated size was subcloned into the PGEX-2T expression vector. The resulting plasmid contained the coding sequence for the putative mature cystatin M in frame with the sequence coding for a thrombin site, at the C-terminal of the GST sequence. The fusion protein was purified by affinity chromatography and was eluted as a single band of 40.5 kDa (Fig. 7A, lane 4). This band was not present in control extracts, which contained only the rGST protein of 26 kDa (Fig. 7A, lane 2).

The rGST carrier was completely cleaved from the purified fusion protein by proteolytic digestion at the thrombin site (Fig. 7A, lanes 5–8). Recombinant cystatin M protein was further purified by absorption of rGST and any traces of uncleaved rGST-cystatin M on immobilized glutathione and eluted as a single band with a molecular mass of ~14.5 kDa (not shown). Cleaved cystatin M showed a tendency to aggregate resulting in a low yield purification.

The possibility that a contaminating bacterial protease could be co-purified with the cystatin M fusion protein was examined by overloading the protein preparation on nonreducing sub- strate gels (zymograms) containing either 0.2% casein or 0.1% gelatin, in parallel with purified papain and trypsin as positive controls. No protease activity was detected for the cystatin M protein preparation (not shown).

The fusion protein migrated mainly as a single band, and only 1–2% of the protein appeared as a dimer on a nonreducing, nondenaturing acrylamide gel (not shown). In order to determine whether recombinant cystatin M contained the predicted
disulfide bonds, the protein was completely unfolded in 8 M urea and reduced with excess DTT, and the reduced thiols were carboxymethylated with iodoacetamide. The electrophoretic mobility of the untreated protein, in nonreducing urea/acrylamide gels at pH 8.5, was compared with the mobility of the modified protein (Fig. 7B). The slower migration of the reduced protein (Fig. 7B, lanes 2–3) indicates the presence in the untreated protein of intramolecular disulfides, which restrict its flexibility. The slowing in migration was more pronounced when the protein was alkylated with the neutral iodoacetamide, which blocked the negative charges on the reduced thiols (Fig. 7B, lane 4). The reduced-alkylated protein should rather be compared with the control, since it carries the same net charge. The band shift was specifically due to the modification of the cysteine residues involved in disulfide bonds and not to the modification of other residues since, when the protein was alkylated without being previously reduced, it migrated like the untreated control (Fig. 7B, lane 1). These results show that disulfide bridges have formed in the recombinant protein. The fact that the recombinant protein displays inhibitory activity against cysteine proteinases (see below) establishes that the protein is properly folded and that at least one disulfide bond is present. A previous study with chicken cystatin has shown that only one disulfide bond (proximal to the C terminus) is enough for maintaining the conformation of the inhibitor required for binding of target proteinases, and absence of it completely destroys the inhibitory activity (40).

Detection of Cystatin M in Vivo—Native cystatin M protein secreted into 21PT cell culture supernatants was immunopre-

FIG. 7. A, purification and thrombin cleavage of recombinant GST-cystatin M fusion protein. Samples were subjected to 15% SDS-PAGE at reducing conditions and stained with Coomassie Brilliant Blue R-250 (lane 1, control; lane 2, 6 μg of affinity-purified GST; lane 3, 85 μg of total protein extracted from E. coli transformed with pGEX-2T vector; lane 4, 6 μg of affinity-purified rGST-cystatin M fusion protein; lanes 5–8, thrombin cleavage of rGST-cystatin M fusion protein for 0, 2, 30, and 90 min, respectively; each lane contains 14 μg of total protein. Left, molecular mass markers: phosphorylase b, 97.4 kDa; serum albumin, 66.2 kDa; ovalbumin, 45 kDa; carbonic anhydrase, 31 kDa; trypsin inhibitor, 21.5 kDa; lysozyme, 14.4 kDa. B, reduction and alkylation of cystatin M fusion protein. Protein was unfolded in 8 M urea and alkylated in the absence of DTT (lane 1). Unfolded protein (lane 2), was reduced by DTT (lane 3) and, subsequently, alkylated with iodoacetamide (lane 4). Each lane contains 5 μg of protein. Electrophoresis was carried out in a nonreducing gel containing 7.5% acrylamide and 8 M urea in 0.375 M Tris at pH 8.5. Gel was stained with Coomassie Brilliant Blue R-250.

FIG. 8. Detection of cystatin M in vivo. A, immunoprecipitates with preimmune serum and affinity-purified antisera from culture supernatants of 21PT (lanes 1–2) and MDA435 cell lines (lanes 3 and 4), respectively. The arrow indicates an immunoreactive protein with the size of cystatin M. B, deglycosylation of cystatin M using N-glycosidase F. Cystatin M protein precipitated from 21PT cell culture supernatant was incubated at 37°C for 24 h in the absence (lane 1) or presence (lane 2) of N-glycosidase F. Cystatin M cleaved from rGST-cystatin M fusion protein by thrombin migrates with an apparent size of 14.5 kDa (lane 3). The arrow on the right indicates glycosylated cystatin M. For Western detection proteins were immunoblotted with anti-rGST-cystatin M serum. Bound antibody was detected by enhanced chemiluminescence (Amersham). The molecular masses of the immunoreactive proteins were estimated based on their mobility relative to prestained molecular mass markers shown in kilodaltons on the left.
Cystatin M, a Novel Cysteine Proteinase Inhibitor

Fig. 9. Inhibition of papain by rGST-cystatin M. Titration of papain activity in the absence and presence of increasing concentrations of cystatin M fusion protein (FP) in continuous rate assays. The concentration of papain was 100 nM, and the Z-Phe-Arg-NHMec substrate was 25 μM.

40% of the total cystatin M protein was estimated to occur in the glycosylated form. Both forms of cystatin M were present in 21PT whole cell lysates (not shown). However, no cystatin M was detected in lysates from MDA435, MDA157, and BT549 metastatic tumor cell lines nor in the corresponding supernatants (not shown), as expected because these cells do not express a transcript for cystatin M. The amounts of cystatin M protein secreted by 70N cells are very low, whereas no intracellular protein was detected (not shown). This result is in accordance with the low abundance of cystatin M transcript in 76N and 70N normal cells (Fig. 4, lanes 1–2).

Inhibitory Activity of Cystatin M—The inhibitory profile of cystatin M fusion protein against papain was studied in continuous rate in vitro assays by incubation of papain with increasing concentrations of the inhibitor and assessment of the residual activity assayed in the presence of the fluorogenic substrate Z-Phe-Arg-NHMec. Papain activity was almost completely inhibited in the presence of 2 nM of the inhibitor (Fig. 9). The inhibition of papain activity in these assays was completely inhibited in the presence of E-64 (trans-epoxysuccinyl-L-leucylamido(4-guanidino)-butane), a specific inhibitor of cysteine protease activity. Similar results were obtained by stopped-flow papain assays. The fusion protein had no inhibitory activity against trypsin. Neither rGST nor bovine serum albumin proteins had any effect on papain activity when added at concentrations similar to or much higher than the inhibitory concentrations of rGST-cystatin M. Cleaved cystatin M also displayed inhibitory activity against papain (not shown), but its concentration could not be determined accurately, because cleaved cystatin M was partially aggregated under the purification conditions applied. These results suggest that native cystatin M is an active cysteine protease inhibitor.

DISCUSSION

The differential display method was applied to the isolation of transcriptionally regulated genes involved in metastasis of a primary mammary tumor. The isolation of a novel gene likely associated with cancer invasion/metastasis is described in this report. The gene encodes cystatin M, a new member of human family 2 cystatins, expressed by normal and 21PT primary breast tumor cells but absent in metastatic cells. Cystatin M mRNA is abundantly present in a variety of normal human tissues. Cystatin M protein contains all the functional motifs conserved among thiol proteinase inhibitors, which suggests that cystatin M is an active inhibitor and could play a role in a variety of cellular functions (8–12). Indeed, recombinant GST-cystatin M fusion protein efficiently inhibited papain in in vitro assays.

A unique feature of cystatin M is that 30–40% of the native protein is glycosylated, whereas all other known human family 2 cystatins are not glycosylated. Whether the carbohydrate moiety modifies the inhibitory activity and specificity or the subcellular localization of cystatin M requires further investigation. If the carbohydrate moiety affects the conformation and inhibitory activity of cystatin M, the recombinant protein might not display optimal inhibitory activity, since it is not glycosylated. However, the fact that the non-glycosylated recombinant protein efficiently inhibited papain indicates that the sugar is not indispensable for the inhibitory activity, consistent with the inhibitory activity being determined in an additive manner by independent affinity contributions from three different domains of the cystatin proteins.

Structural Similarities and Differences Between Cystatin M and Other Family 2 Cystatins—Cystatin M contains the three structural motifs associated with cysteine protease inhibitory activity. A structural element unique to cystatin M is a five-residue insertion, Arg102 to Asp106, which is located between the cysteine residues forming the first interchain disulfide bridge and is absent in all other cystatins (41). Data from x-ray crystallography and NMR spectroscopy of phosphorylated and unphosphorylated chicken cystatin (42–43) revealed that the corresponding segment comprising residues Cys71 to Met89 is a structurally variable region containing the disulfide bridge (Cys71 to Cys81) and the phosphorylation site Ser80. The presence of this insertion in an unstructured region is not likely to cause significant changes in conformation, like phosphorylation of Ser80 had no significant effect on the structure of chicken cystatin (42). In addition, based on the secondary structure described for chicken cystatin (42), this part of the molecule lies on the opposite site to the conserved hairpin loop segments, which interact with the cysteine proteinase, and most likely is not important for the inhibitory activity of cystatin M, although it could possibly play a role in targeting of the protein.

In mammalian cells cystatin M may target any of the lysosomal cathepsins B, L, H, and S or an unidentified cysteine proteinase with papain-like activity. Unlike serine proteases (44), cystatins do not bind covalently to target proteinases but rather block their active site (45) and display a broad specificity. The importance of the binding segments for this interaction varies with the target cysteine proteinase due to structural differences in the active-site region of the proteinase (15, 16). In general, the N-terminal segment is essential for the tight-binding inhibitory properties of cystatins (13, 46, 47) and contains an evolutionarily conserved Gly residue, which confers flexibility to the N-terminal segment, a prerequisite for optimal enzyme binding. Interactions between side chains of the residues Val10, Leu9, and Arg8, preceding the conserved Gly11, also contribute to tight binding (45, 48).

The mechanism of interaction of cystatin M with cysteine proteinases is likely to be similar to the mechanism described for other family 2 cystatins. However, the biological specificity of cystatin M for lysosomal cathepsins remains to be determined. Two of the three residues preceding the conserved Gly in cystatin M are identical with residues at corresponding positions in cystatin C, except that in cystatin M a Met appears at the position of Leu9 which confers selectivity in cystatin C (16).

Role of Cystatin M—Cystatin M mRNA and protein are absent
in metastatic as well as in the BT474 primary breast tumor cell lines. In the 21PT primary tumor cells, however, cystatin M is expressed at levels higher than in 76N and 70N normal cells grown under the same conditions. In recent unpublished studies, we2 have found that cystatin M is highly expressed in normal luminal epithelial cells, which line the ducts, produce milk, and give rise to cancer cells but at low levels in myoepithelial cells. From other evidence, it is likely that cultured normal mammary epithelial cells (e.g., 76N and 70N) resemble myoepithelial more than luminal cells, which may account for the low cystatin M expression found in these cells.

The cystatin M gene may be regulated at the level of transcription, possibly via the retinoic acid β-receptor/retinoic acid pathway. In recent experiments, the expression of cystatin M mRNA was induced by retinoic acid in metastatic mammary epithelial tumor cell lines transfected with the retinoic acid β-receptor.3 The mechanism of down-regulation of cystatin M in tumor cells will require further investigation, including promoter analysis.

Loss of expression of cystatin M is associated with the progress of a primary tumor to a metastatic phenotype, suggesting a putative metastasis suppressor function of the protein. The molecular basis of this function is likely to be the inhibition of cathepsin B and/or L activities. Indeed, cystatin M possesses strong inhibitory activity against both lysosomal cathepsins B and L.4 Loss of expression of cystatin M in metastasizing tumor cells could, at least partially, underlie their aberrant proteolytic function (3). Increased proteolytic potential of metastatic cells results from the combined aberrant regulation of proteolytic enzymes and their endogenous inhibitors (2, 3, 17, 18), involving aberrant expression, processing, stability, activity, intracellular trafficking, and localization of cathepsins B and L (17, 18, 49). Lysosomal cathepsins B and L normally act only intracellularly but, when overexpressed in tumor cells, are secreted or associated with the plasma membrane, where they probably act cooperatively in increasing local proteolysis and directly degrading components of the extracellular matrix and basement membrane (17, 18). The cystatin M protein is mainly secreted and could act to inhibit the extracellular or near-surface activities of cathepsins B and L in malignant cells and block invasion and metastasis to distant sites. Cystatin M also occurs intracellularly at low levels, as shown by steady state and pulse-chase metabolic labeling of 21PT cells.5

Endogenous inhibitors of cathepsins constitute the ultimate level of regulation of the overall cellular cysteine proteasome activity. Decreased cysteine proteinase inhibitory activity in tumor cells contributes to malignant progression (17, 18) and results from decreased protein levels, as well as expression of defective and less active protein forms of cystatin inhibitors (50, 51). Stef A has been proposed to be a tumor suppressor because its expression and activity correlates reversibly with malignant tumor progression (49–51). However, the hypothesis that cysteine proteinase inhibitors are tumor suppressors has been questioned (12). Among known cytostats, cystatin M might be most closely associated with tumor suppression. Our results shed light on the possible roles of cytostats in cancer and indicate that down-regulation of cystatin M during growth of the primary tumor may contribute to aberrant proteolysis in metastasizing tumor cells.

Acknowledgments—We thank Dr. Bonnie Sloane for critical reading of the manuscript, Tom Graf for assistance with sequence analysis, and Drs. Othon Hiopoulos and Robert Liddington for critical suggestions.

REFERENCES
1. Poste, G., and Fidler, I. J. (1980) Nature 283, 138–146
2. Liotta, L. A., Steeg, P. S., and Fidler, I. J. (1990) Cell 64, 327–336
3. Miguitti, P., Robbins, E., and Rifkin, D. B. (1989) Cell 47, 487–489
4. Stettler-Schneider, W. G., Annariorion, S., and Liotta, L. A. (1993) Ann. Rev. Cell Biol. 9, 541–573
5. Liang, L., and Pardee, A. B. (1992) Science 257, 967–970
6. Liang, L., Averbouk, L., and Pardee, A. B. (1993) Nature Cell Biol. 21, 1277–1275
7. Band, V., Zajchowski, D., Swisshelm, K., Trask, D., Kulesa, V., Cohen, C., Connolly, J., and Sager, R. (1990) Cancer Res. 50, 7351–7357
8. Barrett, A. J. (1987) Trends Biochem. Sci. 12, 193–196
9. Rich, D. H. (1988) in Protease Inhibitors (Barrett, A. J. and Salvesen, G., eds) Vo1. 12, pp. 153–178, Elsevier Science Publishers B.V., Amsterdam
10. Turk, V., and Bode, W. (1991) FEBS Lett. 285, 213–219
11. Bobek, L. A., and Levine, M. J. (1993) Crit. Rev. Oral Biol. Med. 3, 307–332
12. Calkins, C. C., and Sloane, B. F. (1988) J. Biol. Chem. 263, 3157–3164
13. Lah, T. T., Kokalj-Kunovar, M., Strukelj, B., Punguerca, J., Barli-Magenja, D., Drobnik-Kosorok, M., Kastelic, L., Babnik, J., Golohu, R., Vrhovec, I., and Turk, V. (1993) Int. J. Cancer 50, 36–44
14. Larini, M., Maie, L., and Hepsa-Novaco, V. K. (1987) Acta Histochem. 82, 5–18
15. Bobek, L. A., Helm, K. M., Day, N. A., Moin, K., Honn, K. V., Crisvonec, J. D., and Sloane, B. F. (1989) Biochem. Biophys. Acta 1003, 63–73