THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 269, No. 15, Issue of April 15, pp. 11530–11536, 1994
Printed in U.S.A.

Human Cathepsin S: Chromosomal Localization, Gene Structure, and Tissue Distribution*

(Received for publication, December 15, 1993)

Guo-Ping Shi‡§, Andrew C. Webb¶, Karen E. Foster¶, Joan H. M. Knoll‖, Cynthia A. Lemere, John S. Munger§, and Harold A. Chapman‡§‡‡**

*From the ‡Physiology Program, Department of Environmental Health, Harvard School of Public Health, Boston, Massachusetts 02115, the §Department of Medicine and the **Center for Neurologic Diseases, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115, the ¶Department of Biological Sciences, Wellesley College, Wellesley, Massachusetts 02181, and the ‖Division of Genetics, Children's Hospital and Department of Pathology, Beth Israel Hospital and Harvard Medical School, Boston, Massachusetts 02115*

The human lysosomal cysteine proteinases, cathepsins H, L, and B, have been mapped to chromosomes 15, 9, and 8, respectively, and the genomic structures of cathepsins L and B have been determined. We report here the chromosomal localization and partial gene structure for a recently sequenced human cysteine proteinase, cathepsin S. A 20-kilobase pair genomic clone of the human cathepsin S gene was isolated from a human fibroblast genomic library and used to map the human cathepsin S gene to chromosome 1q21 by fluorescence *in situ* hybridization. This clone contains exons 1 through 5, introns 1 through 4, part of intron 5, and >7 kilobase pairs of the 5′-flanking sequence. The gene structure of human cathepsin S is similar to that of cathepsin L through the first 5 exons, except that cathepsin S introns are substantially larger. Sequencing of the 5′-flanking region revealed, similar to human cathepsin B, no classical TATA or CAAT box. In contrast to cathepsin B, cathepsin S contains only two SP1 and at least 18 AP1 binding sites that potentially could be involved in regulation of the gene. This 5′-flanking region also contains CA microsatellites. The presence of AP1 sites and CA microsatellites suggest that cathepsin S can be specifically regulated. Results of Northern blotting using probes for human cathepsins B, L, and S are consistent with this hypothesis; only cathepsin S shows a restricted tissue distribution, with highest levels in spleen, heart, and lung. In addition, immunostaining of lung tissue demonstrated detectable cathepsin S only in lung macrophages. The high level of expression in the spleen and in phagocytes suggests that cathepsin S may have a specific function in immunity, perhaps related to antigen processing.

Human cathepsin S (EC 3.4.22.27), cathepsin L (EC 3.4.22.15), cathepsin B (EC 3.4.22.1), and cathepsin H (EC 3.4.22.16) are cysteine proteinases belonging to the papain superfamily (Kirschke and Barrett, 1987; Kirschke *et al.*, 1986). They are involved in physiological protein degradation (Kirschke and Barrett, 1987) with endopeptidase, dipeptidylpeptidase, and aminopeptidase activities. Also they are considered to play roles in pathological tissue destruction and invasion (Barrett and Kirschke, 1981; Sloane *et al.*, 1991). The newly cloned human cathepsin S, like cathepsin L, has been shown to have strong elastase activity (Xin *et al.*, 1992; Shi *et al.*, 1992). Based on their reported cDNA sequences, human cathepsin S and cathepsin L share the highest degree of similarity in the cysteine proteinase family, about 49% amino acid homology (Gal and Gottesman, 1988; Joseph *et al.*, 1988; Ritonja *et al.*, 1988). They are the only two cysteine proteinases found in human alveolar macrophages with elastinolytic activity but cathepsin S differs from L in that it is stable and active *in vitro* at neutral pH (Kirschke *et al.*, 1989; Shi *et al.*, 1992; Mason *et al.*, 1986).

Cathepsins H, L, and B have been shown to have a broad tissue distribution (Barrett and Kirschke, 1981; Qian *et al.*, 1989, 1990). Recently, human cathepsin B 5′-flanking sequence was reported to contain a cluster of SP1 binding sites and high GC content (>80%) similar to housekeeping gene-like regulatory elements (Gong *et al.*, 1993). Cathepsin S differs from cathepsins H, L, and B in that it is reported to be focally expressed in rat tissues and is up-regulated by thyroid hormone in FRTL-5 cells (Petancesks and Devi, 1993). These observations imply that rodent cathepsin S may have its own unique regulatory pathway compared with cathepsins B and L. This may also be true in the human system. Human alveolar macrophages contain high levels of both cathepsin S mRNA and protein as demonstrated by Northern blotting and active site radiolabeling, respectively, whereas no cathepsin S mRNA or protein can be detected in myeloid cell lines such as HL60, U937, and THP-1. In contrast, human cathepsin B is expressed in all of these cell lines (Shi *et al.*, 1992).[1]

As a first step to further understand human cathepsin S regulation, we partially characterized the genomic organization and mapped its chromosomal location. Human tissue Northern blots and lung immunocytochemistry show that the cathepsin S gene is likely regulated in a tissue and cell type-specific manner, and moreover the analysis of its 5′-flanking sequence supports this possibility as well as suggesting the potential for polymorphism in the form of a CA repeat microsatellite.

MATERIALS AND METHODS

Genomic Library Screening—A human lung fibroblast WI-38 genomic library in the Lambda FIX II vector (Stratagene, La Jolla, CA) was screened as described previously (Samia *et al.*, 1990) for cathepsin S sequences using a 493-base pair (bp)[2] PCR fragment from nucleotides

* This work was supported by National Health Service Grants 18658, 48261, and 44712 and by a grant from the Council for Tobacco Research. 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.

‡‡ To whom correspondence should be addressed: Respiratory Division, Brigham and Women's Hospital, 75 Francis St., Boston, MA 02115. Tel.: 617-732-6074 Fax: 617-732-7421.

[1] G.-P. Shi, J. S. Munger, and H. A. Chapman, unpublished results.

[2] The abbreviations used are: bp, base pair(s); kb, kilobase pair(s); PCR, polymerase chain reaction; FISH, fluorescence *in situ* hybridiza-

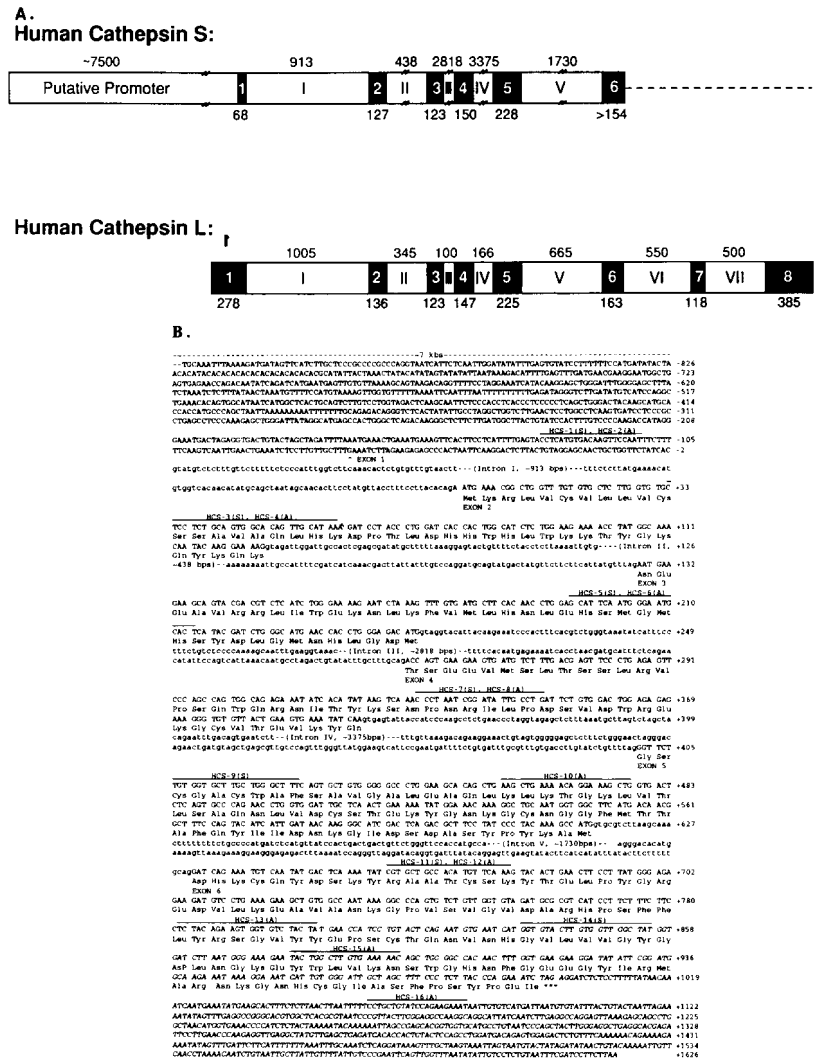

FIG. 1. **Human cathepsin S gene organization.** *A*, comparison of human cathepsin S and human cathepsin L (Chauhan *et al.*, 1993) gene organizations. *Open frames* represent the introns, and the exons are depicted by *shaded frames*, with their lengths in base pairs above or below the frame, respectively. *B*, nucleotide sequence of human cathepsin S gene and deduced amino acid sequence of the encoded product. Intron sequences are given in *lowercase letters* and intron numbers and sizes are indicated in *parentheses*. The transcription initiation site (exon 1) is shown by a *vertical hat* (^) below the nucleotide. *Overlines* represent the positions of primers used; sense primers are denoted by an *S* and antisense primers by an *A*. The nucleotide numbering in the *right-hand margin* is according to the cDNA sequence with the translational start site indicated as +1. The upstream unsequenced region was estimated at about 7 kb by PCR. The nucleotide sequence in italics after +782 is that of the cDNA sequence and its genomic structure has not been defined.

+406 to +894 (Fig. 1*B*) ^{32}P-labeled (DuPont NEN) by random hexamer extension (Multiprime, Amershan Corp.). Two positive clones were isolated from screening (three rounds) of $>10^6$ original plaques and were tested for the presence of 5′ sequences using an ^{32}P-end-labeled antisense primer from nucleotides +59 to +33 of our published cathepsin S cDNA sequence (HCS-4, Fig. 1*B*). Only one of these two clones hybridized to this primer. DNA prepared from this clone (Promega Magic-Maxiprep Kit, Madison, WI) was digested with *Sal*I and *Eco*RI (Lambda FIXTMII Polylinker restriction enzymes) separately and together. *Sal*I released the whole insert which is about 20 kb in size and *Eco*RI yielded five fragments (approximately 6000, 1400, 900, 800, and 700 bp). Digestion with both enzymes generated another two fragments of ~9000 and 1100 bp in size. All of these fragments were subcloned into pBluescript KSII(−) (Stratagene) and sequenced as double-stranded template with Sequenase in combination with dideoxynucleotide chain terminators according to the supplier's recommendation (U. S. Biochemical Corp.). DNA sequence was assembled and searched for transcriptional factor binding sites using the MacVector software package (Kodak IBI, New Haven, CT).

Intron Size Determination by PCR—To determine the intron sizes the original cloned genomic DNA was used as template to amplify the introns by PCR using primer pairs from adjacent exons as indicated in Fig. 1*B*. PCR was performed with Perkin-Elmer System 9600 in a final volume of 50 μl, containing 1 μg of DNA, 2 units of Taq DNA polymerase (Perkin-Elmer), 1 × Taq buffer (Perkin-Elmer), 3.5 mM MgCl2 (Promega), 1 mM dNTP (Promega) of each, 0.01% gelatin (Sigma), and 2 μg of T4 phage gene 32 protein (Boehringer Mannheim). The PCR conditions were 95 °C × 1 min, 55 °C × 2 min, 72 °C × 7 min, repeated for 35 cycles. Because the genomic clone did not contain the complete 3′-end of the cathepsin S gene, human peripheral blood mononuclear cell DNA prepared according Sambrook *et al.* (1989) was used as template in an attempt to amplify the remaining introns and exons. DNA was digested with *Eco*RI, *Sal*I, *Bam*HI, *Asp*700, *Cla*I, *Hin*dIII, and *Xho*I (Boehringer Mannheim), respectively. Digested DNAs were extracted with phenol/chloroform and precipitated with EtOH. Digested DNA (1 μg) from each sample was used as template with sense and antisense primers at the same PCR conditions as stated above. PCR products were subcloned directly into pCRTMII (Invitrogen Co., San Diego, CA) and sequenced with T7 and SP6 flanking vector primers as described (U. S. Biochemical Corp.).

Primer Extension—Human alveolar macrophage total RNA was isolated as described previously (Munger *et al.*, 1991). 0.1 μg of [γ-^{32}P]dATP kinase end-labeled antisense primer HCS-4 (Fig. 1*B*) and 50 μg of RNA were used for reverse transcription reaction. The primer extension experiments were carried out by standard protocols (Kingston, 1987). Primer extension reaction samples were boiled for 3 min and loaded onto a 6% sequencing gel alongside a conventional sequencing reaction ladder used as a size marker. After electrophoresis, the gel was dried and exposed to Kodak X-Omat film at −80 °C without an intensifying screen.

Fluorescence in Situ Hybridization (FISH)—Purified genomic clone phage DNA from above was labeled with digoxigenin-11-dUTP (Boehringer Mannheim) by nick translation and hybridized to human metaphase chromosomes prepared from phytohemagglutinin-stimulated peripheral lymphocytes of a male donor. Hybridization (150 ng of labeled DNA) and detection conditions on metaphase chromosomes were performed as described previously (Knoll *et al.*, 1993). Probe signals were detected with rhodamine-conjugated digoxigenin antibody (Boehringer Mannheim) and viewed through a triple band pass filter (fluorescein isothiocyanate/Texas Red/DAPI: Omega Optical Inc., Brattleboro, VT) using an epifluorescence microscope. For chromosome

tion; DAPI, 4′-6-diamidino-2-phenylindole; MBP, maltose-binding protein; PHA, phytohemagglutinin.

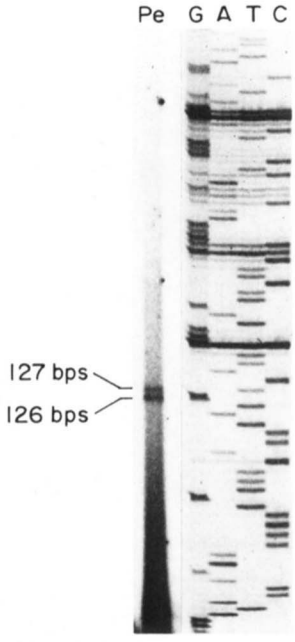

FIG. 2. **Identification of human cathepsin S transcription initiation site by primer extension.** Fifty micrograms of human alveolar macrophage total RNA was used for the primer extension (*Pe*) reaction with antisense primer from nucleotides +59 to +33 (*HCS-4*, Fig. 1*B*). The doublet primer extension products were determined to be 126 and 127 bp in length by denaturing polyacrylamide gel electrophoresis using the adjacent DNA sequencing reaction (*GATC*) as a size marker.

identification, cells were counterstained with DAPI and viewed through a DAPI filter set (Zeiss). Fifteen metaphases were scored for hybridization. Cells were photographed on Kodak Ektar 1000 color film.

Tissue Distribution—Human brain (from frontal cortex), liver, heart, lung, kidney, spleen, and skeletal muscle RNA were prepared. Twenty micrograms of each tissue total RNA was blotted onto Biodyne B (positively charged nylon) (Pall Biosupport, Glen Cove, NY) and UV-crosslinked (Bios Laboratories, New Haven, CT). Human cathepsin probes for S (493 bp), B (584 bp), and L (499 bp) were made by PCR using cDNA clones as template (Shi *et al.*, 1992) and labeled with [α-^{32}P]dATP as described above. Northern blots were prehybridized for 5 h and then hybridized with the above probes at 42 °C overnight and washed in 2 × SSC, 2% SDS for 2 × 15 min at room temperature, 1 × SSC, 1% SDS for 2 × 15 min at room temperature, 0.1 × SSC, 0.1% SDS 30 min at room temperature, and finally 0.1 × SSC, 0.1% SDS for 2–3 h at 42 °C. Filters were exposed to Kodak X-Omat film at −80 °C with an intensifying screen.

Preparation of Human Cathepsin S Fusion Protein—Human cathepsin S full-length cDNA (~1700 bp, 40 kDa) from pcDNA-I (Shi *et al.*, 1992) was fused in frame with a maltose-binding protein (MBP, 40 kDa, mal E gene) in the prokaryotic expression vector, pMAL-cRI (New England Biolabs, Beverly, MA). The construct was transformed into DH5a-competent cells (Life Technologies, Inc.). Cells were grown to ~0.4–0.5 OD_{500} and then stimulated with isopropyl-1-thio-β-D-galactopyranoside to a final concentration of 0.3 mM for 5 h at 37 °C. Cells were pelleted and suspended into hypertonic buffer, which contains 50 mM Tris, 25% sucrose, and 10 mM EDTA, pH 8.0. After suspension, lysozyme was added to a final concentration of 1 mg/ml, and a proteinase inhibitor mixture containing phenylmethylsulfonyl fluoride, leupeptin, pepstin, and aprotinin was added followed by 5% Triton X-100 according to Ronnett *et al.* (1984). The cell lysate was frozen and thawed three to four times, sonicated for 5 min, pelleted, and loaded onto a 10% SDS-polyacrylamide gel electrophoresis gel for electrophoresis. After staining the gel with a copper staining buffer (Bio-Rad) for 5 min, the fusion protein was excised from the gel and destained with destaining buffer (Bio-Rad) for ~30 min with three to four buffer changes. The gel slices were minced and loaded onto a electroeluter (model 422, Bio-Rad). About 1.5 mg of fusion protein was purified from 30 ml of bacterial culture.

Production and Purification of Anti-human Cathepsin S Antibody—Purified MBP-cathepsin S fusion protein was used to inject rabbits (East Acres Biologicals, Inc., Southbridge, MA). Monospecific antibody was made according to Haass *et al.* (1989). Expressed fusion protein was run on 10% SDS-polyacrylamide gel electrophoresis gel and electrotransferred onto nitrocellulose. Transferred proteins were viewed by Ponceau stain. The 80-kDa fusion protein was cut out, and the filter strip was blocked with 5% milk in phosphate-buffered saline. The strip was washed with phosphate-buffered saline and incubated with crude antiserum at 4 °C overnight with constant shaking. The nonspecific antibodies were removed by three 15-min 0.5 M Tris-buffered saline and 0.2% Triton X-100 washes. The monospecific antibody was then eluted with 0.2 M HCl for 1 min under shaking. In Western blotting, the human cathepsin S antibodies stained a single protein band consistent with the size of mature cathepsin S in lysates of human alveolar macrophages. To test the specificity of the antibodies, MBP-human cathepsin L fusion protein (containing full-length cathepsin L) was made and transferred onto nitrocellulose in the same way as cathepsin S. Aliquots of monospecific anti-MBP- cathepsin S antibodies were absorbed with nitrocellulose-bound cathepsin S or cathepsin L fusion protein strips overnight at 4 °C prior to use in immunostaining.

Immunostaining of Frozen Human Lung Tissue—Unfixed cryosections of frozen human lung tissue obtained by surgical biopsy were stored at −20 °C. Morphologically normal areas of the biopsies were selected for sectioning. For immunocytochemistry, sections were brought to room temperature in Tris-buffered saline for 5 min. Tissue was fixed and endogenous peroxidase activity quenched with 0.3% hydrogen peroxide in methanol for 20 min. The sections were washed in water for 5 min and pretreated with 80% formic acid for 10 min at room temperature. Formic acid was removed from the sections in three 5-min washes in water. Nonspecific background staining was blocked by incubating sections in 10% goat serum in Tris-buffered saline for 20 min. Primary antibody (anti-human cathepsin S, 1:100) was applied for 1 h at room temperature. The horseradish peroxidase avidin-biotin complex system (Rabbit Elite ABC Kit, Vector Laboratories, Burlingame, CA) and diaminobenzidine were used to visualize bound antibody. Sections were counterstained with hematoxylin, dehydrated, cleared in Histoclear (National Diagnostics, Atlanta, GA), and coverslipped with Permount.

RESULTS

Genomic Library Screening, Subcloning, and DNA Sequencing—A 493-bp cDNA PCR fragment (from nucleotides +406 to +894, Fig. 1*B*) was used to screen a human lung fibroblast WI38 genomic library. A single clone hybridizing with multiple PCR fragments and oligonucleotides was characterized. This clone contains a 20-kb insert. The restriction enzymes *Sal*I and *Eco*RI were used to digest the DNA into seven fragments. To further characterize these fragments, they were subcloned into pBluescript KSII(−). DNA sequencing of restriction fragments of this genomic fragment subcloned into pBluescript revealed the sequences for the 5′ cDNA primer HCS-1 in a 9000-bp *Sal*I-*Eco*RI fragment and primer HCS-8 from nucleotides +349 to +330 in the 6000-bp *Eco*RI fragment. The remaining sequence in the seven subclones was completed by using the other exon primers shown in Fig. 1*B*. This genomic clone for human cathepsin S contains exons 1 through 5, introns 1 through 4, and part of intron 5 (Fig. 1, *A* and *B*).

Determination of Intron Sizes by Polymerase Chain Reaction—Using a modified PCR and adjacent exons primer pairs, we determined the sizes of introns 1 through 4 (Fig. 1, *A* and *B*). Based upon the high degree of homology between human cathepsin S and L cDNAs, it was speculated that the genomic organization of the two genes might also be similar. Therefore, primer pairs were selected using the published locations of the human cathepsin L gene introns (Chauhan *et al.*, 1993) as a guide. To determine the organization of the 3′-end of the human cathepsin S gene absent from our clone, we applied the same PCR strategy to amplify uncloned human genomic DNA. This strategy yielded two fragments: a 2000-bp fragment from nucleotides +406 to +677, which covers the whole of intron 5, and a 154-bp fragment from nucleotides +658 to +803 within intron 6. Those PCR fragments were subcloned into pCRII for sequencing as described above. All the other downstream primers did not yield PCR product. This may be because the 3′-end of this gene contains introns too large to PCR across. Since we

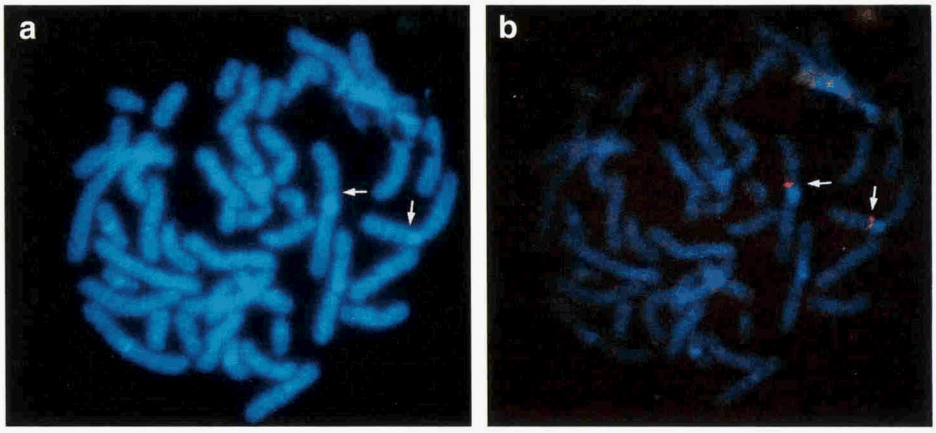

FIG. 3. **Localization of human cathepsin S gene by FISH.** *a*, partial metaphase of DAPI-banded chromosomes and *b*, corresponding double chromatid hybridization signals at human chromosome 1q21. *Arrows* indicate the chromosomal location and hybridizations of the human cathepsin S genomic clone probe.

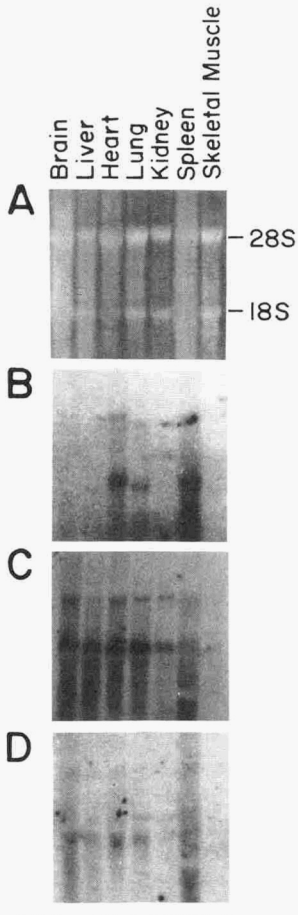

FIG. 4. **Northern blots of total RNA from various human tissues.** Twenty micrograms of total RNA from each human tissue were blotted onto Biodyne B nylon filter and stained with ethidium bromide (*A*). This Northern blot was hybridized sequentially with human cathepsins S (*B*), B (*C*), and L (*D*) cDNA probes (see "Materials and Methods") and exposed to Kodak X-Omat film at −80 °C.

did not find any intron sequence between nucleotides +658 to +803, exon 6 must be at least 154 bp in length (Fig. 1, *A* and *B*).

Mapping of Transcription Initiation Site by Primer Extension—Our human macrophage-derived cathepsin S cDNA (Shi *et al.*, 1992) does not contain the first exon of its gene. However, others (Wiederanders *et al.*, 1992) have reported cathepsin S cDNA sequence from human testis, including the first exon. Based on this published exon sequence, we made two additional primers (sense and antisense nucleotides −111 to −131, *HCS-1* and *HCS-2*, Fig. 1*B*), and DNA sequencing confirmed the reported first exon sequence. A primer walking strategy was employed to obtain an additional ~800 bp of the 5′-flanking sequence. A primer extension experiment was then performed to determine the transcription start site. Fifty micrograms of human alveolar macrophage total RNA was reverse-transcribed after priming with 0.1 μg of ^{32}P-end-labeled antisense primer HCS-4. Using a DNA sequencing reaction as a size marker, the primer extension products were 126 and 127 bp from nucleotides +59 to the putative transcription start sites, respectively (Fig. 2). These data have been confirmed by primer extension using 50 μg of human lung cancer cell A549 total RNA with the same results as alveolar macrophage RNA (data not shown). Primer extension using another antisense primer further 5′ from these transcription start sites (nucleotides −111 to −131, HCS-2) produced no bands with either alveolar macrophage or A549 RNA (data not shown). Therefore, the transcription initiation site and exon 1 for human lung cathepsin S are predicted as shown in Fig. 1*B*.

Human Cathepsin S Gene Maps to Chromosome 1q21 by FISH—Fifteen human metaphases were analyzed following hybridization of the entire lambda clone containing the cathepsin S genomic DNA and all revealed single or double chromatid hybridization at 1q21 (Fig. 3).

Tissue Distribution—cDNA probes for human cathepsins B, L, and S generated by PCR were used sequentially to probe Northern blots of human RNA from brain (frontal cortex), liver, heart, lung, kidney, spleen, and skeletal muscle. Only spleen, lung, and heart have detectable cathepsin S mRNA (Fig. 4*B*). However, cathepsin B and L mRNA are found in all of the tissue tested although there is only a faint signal in skeletal muscle (Fig. 4, *C* and *D*).

Human Lung Tissue Immunocytochemistry—Human cathepsin S full-length cDNA was expressed as recombinant protein fused with MBP in pMAL-cRI. Expressed fusion protein was purified and used to raise anti-cathepsin S antibody. Crude antibody was immunoaffinity-purified and then adsorbed with either a recombinant cathepsin L fusion protein expressed in pMAL-cRI or with the cathepsin S fusion protein prior to immunostaining lung tissue sections. Affinity-purified antibodies adsorbed with a cathepsin L fusion protein stained virtually all alveolar macrophages and a few interstitial cells, which may also be macrophages, but not airway or vascular lining cells (Fig. 5*A*). In contrast, adsorption of the antibodies with nitrocellulose-bound recombinant cathepsin S completely removed the lung immunostaining (Fig 5*B*).

A CA Microsatellite Located Upstream of Cathepsin S Gene—The CA microsatellites of human cathepsin S were found by sequencing the 5′-flanking region of this gene. These repeats span 36 base pairs with few interruptions (Fig. 6). In an attempt to find other CA repeats, we hybridized a [γ-^{32}]ATP-end-labeled $(CA)_{20}$ repeat oligonucleotide to a Southern blot made from the cathepsin S genomic clone *Sal*I and *Eco*RI digestion

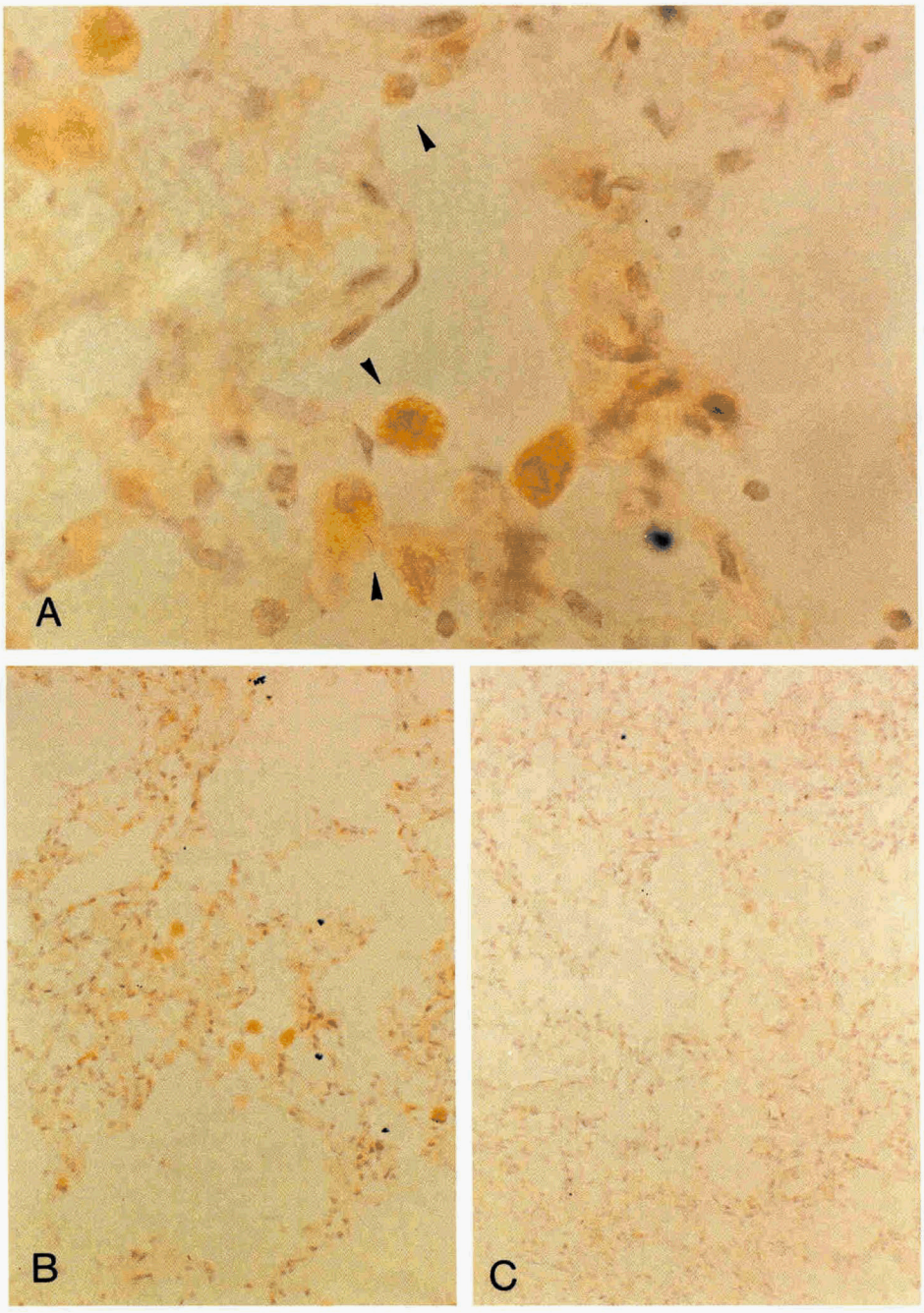

FIG. 5. **Immunostaining of human lung tissue with anti-human cathepsin S antibodies.** Photographs of immunostaining of human lung tissues with affinity-purified cathepsin S antibodies preadsorbed with either immobilized cathepsin L fusion protein (*A* and *B*) or after immunostaining of an adjacent section with cathepsin S antibodies preadsorbed with immobilized cathepsin S. Lung sections were incubated with primary antibodies and then developed with peroxidase-coupled secondary antibodies as described in the text. *Arrows* in *A* indicate positive staining in alveolar macrophages at × 100 magnification; *B* illustrates focal cellular staining consistent with macrophages at × 40 magnification. *C* shows that this immunostaining was abrogated by preadsorption with recombinant cathepsin S.

fragments and the 2-kb intron 5 PCR product and probed with the ^{32}P-labeled $(CA)_{20}$ oligonucleotide. Only the 9-kb 5′-end fragment demonstrated hybridization to this probe.

DISCUSSION

In this study we have utilized a 20-kb genomic clone to map the chromosomal location of the cathepsin S gene and partially characterize its gene structure. Cathepsin S appears to be a single gene that clearly maps to chromosome 1q21 (Fig. 3). This is in marked contrast to the multiple copy pseudogene arrangement recently reported for human cathepsin L (Chauhan *et al.*, 1993). This distinction was also revealed in our parallel screening of this genomic library with a human cathepsin L cDNA probe, which yielded nine clones compared with the two we isolated for cathepsin S (data not shown). As predicted from the high degree of amino acid homology derived from the cDNAs for human cathepsin S and cathepsin L (Shi *et al.*, 1992), we find that the cathepsin S gene has a similar exon-intron arrangement (Fig. 1) to the recently reported human cathepsin L gene organization (Chauhan *et al.*, 1993), except cathepsin S has substantially larger introns (determined by PCR). Despite the similarities among the known cysteine proteinases having strong active site sequence homology, *i.e.* cathepsins H, L, B, and S, the mapping results demonstrate that these genes all reside on separate human chromosomes; namely, chromosomes 15, 9, 8, and 1, respectively (Bouyon *et al.*, 1989; Chanhan *et al.*, 1993; Fan *et al.*, 1989; Fong *et al.*, 1992; Gong *et al.*, 1993; Wang *et al.*, 1987).

Sequencing upstream of the cathepsin S gene reveals distinct differences compared with the constitutive gene product, cathepsin B. The reported human cathepsin B 5′-flanking region is GC-rich (>80%), has no classical TATA box, and contains 15 SP1 transcription factor binding sites (Gong *et al.*, 1993). All of these genomic features are typical of a weakly regulated gene product and are consistent with the widespread tissue distribution of cathepsin B reported here and previously (Gong *et al.*,

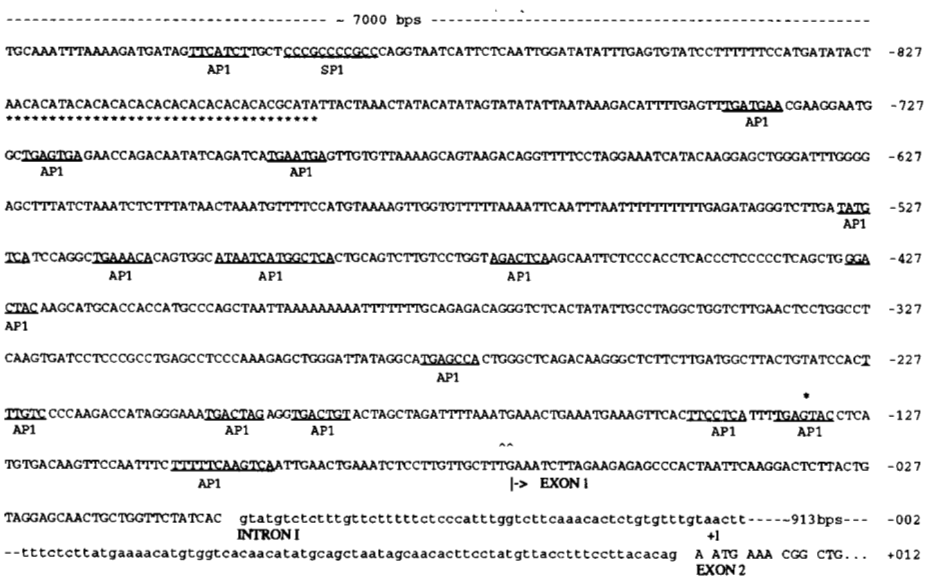

FIG. 6. **Nucleotide sequence of the 5′-flanking region of the human cathepsin S gene.** First and second exons and first intron are indicated. The two apparent transcription initiation sites are shown by "*up hats*". The *arrow* indicates the predicted preferred transcription initiation site at the purine (Breathnach and Chambon, 1981) and start of exon 1. The *asterisk* is positioned over the 5′-end of reported human testis cathepsin S cDNA, indicating the putative first exon start in this tissue (Wiederanders *et al.*, 1992). The predicted AP1 transcription factor binding sites are indicated by *single underlines*. Only two SP1 motifs were found and also are *underlined*. *Stars* underneath the sequence indicate the CA microsatellite repeats.

1993). Although the human cathepsin S 5′-flanking region reported here also shows no conventional TATA or CAAT boxes typical of most RNA polymerase II transcriptional units, this ~800-bp region contains only two SP1 sites, a much lower GC content (~40%), and at least 18 AP1 binding sites (Fig. 6). These regulatory elements are more typical of promoters associated with genes subject to coordinate control such as by phorbol diester or growth factor receptor pathways (Nicholson *et al.*, 1990). These findings suggest that cathepsin S is not a constitutive gene product and the restricted tissue and cellular distribution of the enzyme (Figs. 4 and 5) is consistent with this conclusion. Signals which regulate the expression of human cathepsin S are unknown, but preliminary data in macrophage progenitors such as U937 and THP-1 human mononuclear cell lines indicate the cathepsin S gene is not readily inducible by phorbol diesters.[1]

Primer extension reactions employing human alveolar macrophage and human lung adenocarcinoma cell RNA as template were used to determine the transcriptional start site. In both these cell types, two adjacent start sites were identified 67 bp 5′ to the initiator AUG. Interestingly, Wiederanders *et al.* (1992) reported a human testis cathepsin S cDNA, which contains 157 bp before the first encoded methionine, suggesting that the first exon should be at least 156 bp in length. We have confirmed our results by two additional experiments. First, antisense primers positioned −111 to −131 (HCS-2) to the start AUG did not yield primer extension products when either macrophage or A549 cell RNA was used as template. Second, only sense primers beyond the −67-bp start site yielded PCR product on cDNA first strand template reverse transcribed from macrophage RNA using various 3′ antisense primers (data not shown). Since the 157-bp 5′ cDNA sequence reported by Wiederanders *et al.* (1992) is identical to that observed upstream in our genomic clone (Fig. 1), all of these data are consistent with the conclusion that different tissues utilize different transcriptional start sites for the initiation of cathepsin S gene expression. This conclusion may be reflected in the Northern blotting shown in Fig. 4, where the size of the cathepsin S transcript (*B*) in lung tissue appears slightly smaller than that observed with heart and spleen. The presence of multiple transcriptional initiation sites may reflect the lack of TATA or CAAT sequences in the 5′-flanking sequence (Aronow *et al.*, 1989; Bonham and Fujita, 1993; Johnson *et al.*, 1988; Mavrothalassitis *et al.*, 1990). Whether the differential initiation sites in different tissues reflect different elements promoting gene expression remains to be defined.

Microsatellites consisting of CA and other dinucleotide repeats have been shown to be highly polymorphic and useful in genetic linkage analysis (Litt and Luty, 1989; Smeets *et al.*, 1989; Tautz, 1989; Weber and May, 1989). These repeats may occur adjacent to the coding regions, in introns within genes, or even in 5′- or 3′-untranslated regions (Hamada *et al.*, 1984; Weber and May, 1989). CA microsatellites of six or more repeats appear on average in human genomic DNA every 30 kb and 12% of the CA microsatellites are over 20 repeats (Beckman and Weber, 1992; Stallings *et al.*, 1991). The human cathepsin S gene contains an 18-CA repeat ~600 bp prior to the initiation sites (Fig. 6). CA microsatellites within 5′-flanking sequence of other genes have been shown to regulate expression, possibly by the formation of Z-form DNA (Delic *et al.*, 1991; Schroth *et al.*, 1992), although to date no human disease has been specifically linked to altered expression of cathepsin S.

Restricted tissue distribution of cathepsin S has been reported previously in other species. However there are notable differences between rodent and bovine expression of cathepsin S and our observations with human cathepsin S (Kirschke *et al.*, 1989; Petanceska and Devi, 1993). In the rat and the cow, for example, there is widespread expression of cathepsin S in brain. In contrast, human cathepsin S mRNA is not widely distributed in normal human brain (Fig. 4); this conclusion has been confirmed by additional immunoctyochemical staining of brain sections with our cathepsin S antibodies (not shown). We have demonstrated recently that co-transfection of the human amyloid precursor protein and cathepsin S genes in embryonic kidney cells generates amyloid Aβ peptides *in vitro* and that cathepsin S is demonstrable in neurons of Alzheimer's Disease brain.[3] Thus a systematic study of cathepsin S expression in

[3] J. S. Munger, C. Haass, G. P. Shi, C. Lemere, F. Wong, D. Tepelow, D. Selkoe, and H. A. Chapman, submitted for publication.

normal and diseased human brain is in progress and will be published separately.[4]

A consistent finding among species is the high level of expression in the spleen; indeed the initial purification of cathepsin S from bovine spleen underlies its common name. Our observations of limited expression of lung cathepsin S within macrophages coupled with its high expression in spleen suggest a specific role for cathepsin S in immunity. Given the broad substrate specificity of cathepsin S, induction of this enzyme may be involved in antigen processing. Such a role in protein hydrolysis would be consistent with prior proposals that cathepsin S in thyroid contributes to thyroglobulin hydrolysis important to thyroid hormone production (Petancesks and Devi, 1993). These observations raise the possibility that the normal limited expression of human cathepsin S follows cellular need for a powerful protease for specific intracellular degradative functions. The ability of cathepsin S to mediate the generation of peptides such as Aβ with potentially toxic properties (Yankner *et al.*, 1990) suggests that cathepsin S expression may also be involved in disease processes where altered proteolytic processing is a prominent part of the pathobiology.

Acknowledgments—We thank Dr. David Beier (Division of Genetics, Brigham and Women's Hospital, Boston, MA) for helpful recommendations and discussion, Dr. Benjamin Neel (Beth Isreal Hospital, Boston, MA) for the A549 cell line, and Drs. Mary Sunday and Cathleen Haley (Department of Pathology, Brigham and Women's Hospital, Boston, MA) for human lung tissue sections. The contribution from members of the Wellesley College Molecular Biology course in screening the human genomic library is gratefully acknowledged.

REFERENCES

Aronow, B., Lattier, D., Silbiger, R., Dusing, M., Hutton, J., Jones, G., Stock, J., McNeish, J., Potter, S., Witte, D., and Wigington, D. (1989) *Genes & Dev.* **3,** 1384–1400

Barrett, A. J., and Kirschke, H. (1981) *Methods Enzymol.* **80,** 535–561

Beckman, J. S., and Weber, J. L. (1992) *Genomics* **12,** 627–631

Bonham, K., and Fujita, D. J. (1993) *Oncogene* **8,** 1973–1981

Bouyon, C., Vignon, X., Ferrara, M., De Larocque, A., Geneix, A., and Malet, P. (1989) *Cytogenet. Cell Genet.* **51,** 966

Breathnach, R., and Chambon, P. (1991) *Annu. Rev. Biochem* **50,** 349–383

Chauhan, S. S., Popescu, N. C., Ray, D., Fleischmann, R., Gottesman, M. M., and Troen, B. R. (1993) *J. Biol. Chem.* **268,** 1039–1045

Delic, J., Onclercq, R., and Moisan-Coppey, M. (1991) *Biochem. Biophys. Res. Commun.* **180,** 1273–1283

Fan, Y. S., Byers, M. G., Eddy, R. L., Joseph, L. J., Sukhatme, V. P., Chan, S. J., and Shows, T. B. (1989) *Cytogenet. Cell Genet.* **51,** 996

Fong, D., Chan, M. M.-Y., Hsieh, W.-T., Menninger, J. C., and Ward, D. C. (1992) *Hum. Genet.* **89,** 10–12

[4] C. Lemere, G. P. Shi, J. S. Munger, H. A. Chapman, and D. Selkoe, manuscript in preparation.

Gal, S., and Gottesman, M. M. (1988) *Biochem. J.* **253,** 303–306

Gong, Q., Chan, S. J., Bajkowski, A. S., Steiner, D. F., and Frankfater, A. (1993) *DNA Cell Biol.* **12,** 299–309

Haass, Ch., Pesold-Hurt, B., Multhaup, G., Beyreuther, K., and Kloetzel, P.-M. (1989) *EMBO J.* **8,** 2373–2379

Hamada, H., Petrino, M. G., Kakunaga, T., Seidman, M., and Stollar, B. D. (1984) *Mol. Cell. Biol.* **4,** 2610–2621

Johnson, A. C., Jinno, Y., and Merlino, G. T. (1988) *Mol. Cell. Biol.* **8,** 4174–4184

Joseph, L. J., Chang, L. C., Stamenkovich, D., and Sukhatme, V. P. (1988) *J. Clin. Invest.* **81,** 1621–1629

Kingston, R. E. (1987) in *Current Protocols in Molecular Biology,* Vol. 1, pp 4.8.1–4.8.3, Greene Publishing Associates and Wiley-Interscience, New York

Kirschke, H., and Barrett, A. J. (1987) in *Lysosomes: Their Role in Protein Breakdown* (Glaumann, H., and Ballard, F. J., eds) pp. 193–238, Academic Press, London

Kirschke, H., Schmidt, I., and Wiederanders, B. (1986) *Biochem. J.* **240,** 455–459

Kirschke, H., Wiederanders, B., Bromme, D., and Rinne, A. (1989) *Biochem. J.* **264,** 467–473

Knoll, J. H. M., Sinnett, D., Wagstaff, J., Glatt, K., Schantz, W., Whiting, P. M., Wingrove, P., Sikela, J. M., and Lalande, M. (1993) *Hum. Mol. Genet.* **2,** 183–189

Litt, M., and Luty, J. A. (1989) *Am. J. Hum. Genet.* **44,** 397–401

Mason, R. W., Johnson, D. A., Barrett, A. J., and Chapman, H. A., Jr.(1986) *Biochem. J.* **233,** 925–927

Mavrothalassitis, G. J., Watson, D. K., and Papis, T. S. (1990) *Oncogene* **5,** 1337–1342

Munger, J. S., Shi, G.-P., Mark, E., Chin, D. T., Gerard, C., and Chapman, H. A. (1991) *J. Biol. Chem.* **266,** 18832–18838

Nicholson, R. C., Mader, S., Nagpal, S., Leid, M., Rochette-Egly, C., and Chambon, P. (1990) *EMBO J.* **9,** 4443–4454

Petanceska, S., and Devi, L. (1993) *J. Biol. Chem.* **267,** 26038–26043

Qian, F., Bajkowski, A., Steiner, D. F., Chan, S. J., and Frankfater, A. (1989) *Cancer Res.* **49,** 4870–4875

Qian, F., Frankfater, A., Miller, R. V., Chan, S. J., and Steiner, D. F. (1990) *Int. J. Biochem.* **22,** 1457–1464

Ronnett, G. V., Knutson, V. P., Kohanski, R. A., Simpson, T. L., and Lane, M. D. (1984) *J. Biol. Chem.* **259,** 4566–4575

Ritonja, A., Popovic, T., Kotnik, M., Machleidt, W., and Turk, V. (1988) *FEBS Lett.* **228,** 341–345

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual,* 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Samia, J. A., Alexander, S. J., Horton, K. W., Auron, P. E., Byers, M. G., Shows, T. B., and Webb, A. C. (1990) *Genomics* **6,** 159–167

Schroth, G. P., Chou, P.-J., and Ho, P. S. (1992) *J. Biol. Chem.* **267,** 11846–11855

Shi, G.-P., Munger, J. S., Meara, J. P., Rich, D. H., and Chapman, H. A. (1992) *J. Biol. Chem.* **267,** 7258–7262

Sloane, B. F., Rozhin, J., Krepela, E., Ziegler, G., and Sameni, M. (1991) *Biomed. Biochim. Acta* **50,** 549–554

Smeets, H. J. M., Brunner, H. G., Ropers, H. H., and Wieringa, B. (1989) *Hum. Genet.* **83,** 245–251

Stallings, R. L., Ford, A. F., Nelson, D., Torney, D. C., Hildebrand, C. E., and Moyzis, R. K. (1991) *Genomics* **10,** 807–815

Tautz, D. (1989) *Nucleic Acids Res.* **17,** 6463–6471

Wang, X., Chan, S. J., Eddy, R. L., Byers, M. G., Fukushima, Y., Henry, W. M., Haley, L. L., Steiner, D. F., and Shows, T. B. (1987) *Cytogenet. Cell Genet.* **46,** 710–711

Weber, J. L., and May, P. E. (1989) *Am. J. Hum. Genet.* **44,** 388–396

Wiederanders, B., Bromme, D., Kirschke, H., von Figura, K., Schmidt, B., and Peters, C. (1992) *J. Biol. Chem.* **267,** 13708–13713

Xin, X.-Q., Gunesekera, B., and Mason, R. W. (1992) *Arch. Biochem. Biophys.* **299,** 334–339

Yankner, B., Duffy, L., and Kirschner, D. (1990) *Science* **250,** 279–282