Human salivary cystatin S

Cloning, sequence analysis, hybridization *in situ* and immunocytochemistry

Libuse A. BOBEK,* Alfredo AGUIRRE and Michael J. LEVINE
Department of Oral Biology, School of Dental Medicine, State University of New York at Buffalo, Buffalo, NY 14214, U.S.A.

A human submandibular-gland (SMG) cDNA library was constructed in a λ gt11 Sfi-Not orientation-specific expression vector and then screened with antibody generated against human salivary cystatins. The clone C4-4 encoded an N-terminally truncated cystatin S, whereas the others encoded cystatin SN. The library was then rescreened with the C4-4, and the inserts of several positive clones were directly amplified from the eluted plaques by linear PCR and the PCR products analysed by Southern blotting and direct DNA sequencing. Two clones (C3 and C12) encoded a full-length secreted cystatin S and its leader peptide and included 5'- and 3'-untranslated regions. These clones showed a high degree of sequence similarity to cDNA clones encoding human salivary cystatin SN and genomic clones encoding cystatin SN and SA. Hybridization *in situ* of normal human SMG and parotid-gland (PG) tissue sections localized the cystatin-gene transcripts to the cytoplasm of serous acinar cells of both glands, with a much higher concentration of cystatin mRNA in the SMG. Immunocytochemistry localized the salivary cystatin gene products also to the serous cells, and the levels of cystatin protein correlated with the amount of cystatin mRNA, with a much stronger signal in the SMG than in the PG.

INTRODUCTION

Cystatins are potent inhibitors of cysteine proteinases such as ficin, papain and cathepsins B, H and L (Anastasi *et al.*, 1983; Abrahamson *et al.*, 1986; Barrett *et al.*, 1986*b*). Despite their structural diversity, this physiological function is common to all the members of the cystatin superfamily, suggesting that they are derived from a common ancestor (Muller-Esterl *et al.*, 1986; Saitoh *et al.*, 1988). Recently, Rawlings & Barrett (1990) constructed a scheme for the evolution of the diverse group of proteins that show sequence identity with chicken cystatin. Data from the comparison of amino acid sequences were supplemented with other evidence, such as the occurrence of disulphide bonds, the numbers of internal sequence repeats in the molecules and biological function. As a result, they suggested that additional families be recognized to accommodate the proteins that are not included in the previously established cystatin families 1, 2 and 3 (Barrett *et al.*, 1986*a,b*). For example, rice cystatin, which appears to be intermediate between family 1 and 2, might be assigned to a fourth family, whereas histidine-rich glycoprotein and α₂HS-glycoprotein might be members of new families; however, they point out that no data are available regarding the last two proteins as cysteine-proteinase inhibitors.

Human saliva contains multiple forms of cystatins which belong to family 2 of the cystatin superfamily. These were first isolated as cysteine-containing phosphoproteins by Shomers *et al.* (1982*a*). Isemura *et al.* (1984*a*, 1986, 1987) determined the amino acid sequence of cystatin S (then named SAP-1), cystatin SN (neutral pI) and cystatin SA (acidic pI) isolated from human whole saliva. Each protein contained 113 amino acids and showed 90% sequence identity with the other two salivary

cystatins. These cystatins had about 54% sequence identity with cystatin C, which also belongs to family 2, but is found in high amounts in cerebrospinal fluid and urine (Grubb & Lofberg, 1982; Brzin *et al.*, 1984; Ghiso *et al.*, 1986). It was suggested that the cystatins originally characterized from whole saliva were degraded by aminopeptidases, since later on cystatins S (Hawke *et al.*, 1987; Saitoh *et al.*, 1988) and SN (Al-Hashimi *et al.*, 1988) with an additional eight residues at the N-terminus (total of 121 residues) were isolated from glandular saliva. Genes encoding cystatins SN and SA have been cloned and sequenced (Saitoh *et al.*, 1987). They are composed of three exons and two introns, and one gene spans about 4.0 kbp of the genomic DNA. The proteins deduced from the coding sequences comprise a secretory signal peptide of 20 amino acids, followed by a secreted peptide of 121 amino acids. A cDNA clone, 594 bp long, encoding a full-length cystatin SN (then termed SA-I) was isolated and sequenced by Al-Hashimi *et al.* (1988). Saitoh *et al.* (1988) isolated and sequenced a truncated cDNA clone encoding 50 amino acid residues from the C-terminus of salivary cystatin, which is identical with the cystatin S sequence, except that amino acid residue 115 is asparagine rather than aspartic acid. The decoded sequence was also identical with the corresponding portion of cystatin SA, except at amino acid position 101, where glutamine replaced glutamic acid.

Although much attention has been given to the physiological functions and inhibitory mechanisms of cystatins, the precise structural requirements for these activities remain to be determined. A possible mode of interaction of chicken egg-white cystatin with papain has been proposed in which three parts of the inhibitor (the N-terminus and two hairpin loops containing residues 53–57 and 102–107) are important for the inhibition of

papain (Bode *et al.*, 1988). Studies on the structure–function relationships of salivary cystatins will be greatly facilitated by the availability of cDNA clones encoding the salivary cystatins. Accordingly, the present study describes the isolation and analysis of cDNA clones encoding human salivary cystatin S and SN from a submandibular-gland (SMG) cDNA expression library. In addition, we have examined the relative level and location of cystatin-gene transcripts in normal human salivary glands by *in situ* hybridization and have localized salivary cystatin-gene protein products by immunocytochemistry. The information obtained on patterns of cystatin-gene expression in normal human tissues should provide a basis for further studies on the pathogenesis of salivary-gland neoplasia.

MATERIALS AND METHODS

Materials

The following were purchased from Promega (Madison, WI, U.S.A.): RiboClone cDNA Synthesis System Kit, Packagene Lambda DNA Packaging System, Protoblot λ gt11 Immuno-screening System containing secondary antibody (goat anti-rabbit IgG), alkaline phosphatase conjugate and the necessary colour-development substrates (Nitroblue Tetrazolium and 5-bromo-4-chloroindol-3-yl phosphate), Nitroplus 2000 hybridization membrane, *Escherichia coli* strains Y1090 and JM109, λ gt11 Sfi-Not and pGem-11Zf vectors, λ forward and reverse primers, RNasin, and all restriction and other DNA-modifying enzymes. Genious Nonradioactive DNA Labeling and Detection Kit was from Boehringer-Mannheim Biochemicals (Indianapolis, IN, U.S.A.). Sequenase kit was from United States Biochemicals (Cleveland, OH, U.S.A.), Geneclean kit from Bio 101 (La Jolla, CA, U.S.A.). All radioisotopes were from New England Nuclear Corp. (Boston, MA, U.S.A.). A normal swine serum, biotinylated swine anti-rabbit IgG and avidin–biotin–alkaline phosphatase complex were from Dako, Glostrup, Denmark. Oligo(dT)–cellulose, naphthol AS-MX phosphate (free acid) and Fast-Red TR salt were from Sigma Chemical Co. (St. Louis, MO, U.S.A.). A temperature cycler was obtained from COY (Ann Arbor, MI, U.S.A.).

Human salivary-gland tissue

Normal SMG tissue was obtained after prophylactic radical neck dissection from a 72-year-old male afflicted with a naso-pharyngeal carcinoma. Parotid tissue was obtained from a patient with a history of chronic inflammatory disease of the lateral neck. Informed consent was obtained from patients before surgery. Immediately after excision, specimens were stored in liquid N_2 until used.

Construction of an SMG cDNA library

RiboClone cDNA Synthesis System Kit was used for the synthesis of the double-stranded (ds) cDNA. Protocols described by Promega and Sambrook *et al.* (1989) were followed. Total SMG RNA was isolated from 1.0 g of tissue by extraction with guanidine isothiocyanate followed by centrifugation in a CsCl gradient. Poly(A)⁺ RNA was selected by affinity chromatography on an oligo(dT)–cellulose column. The λ gt11 Sfi-Not expression vector was used as a cloning vehicle. Directional cloning was achieved with a unique oligodeoxynucleotide primer-adaptor containing the recognition site for *Not*I upstream from an oligo(dT)₁₅ sequence which primed the first-strand synthesis. *Eco*RI adaptors were added to the double-stranded cDNA, which was followed by digestion with *Not*I and ligation of the cDNA to the *Eco*RI–*Not*I dephosphorylated λ gt11 Sfi-Not vector arms. Ligated DNA was packaged into the Packagene

Lambda DNA Packaging System. The titre of the library was determined on the Y1090 host strain (Young & Davis, 1983) to be 6×10^6 independent clones. One half of this library was amplified on the Y1090 host (Sambrook *et al.*, 1989) and the other half was left unamplified for the direct library screening.

Selection and analysis of cDNA clones encoding salivary cystatins

Screening of library with anti-cystatin antibody. The SMG cDNA expression library was plated on to a Y1090 bacterial host at a density of 10000 plaques/90 mm plate. Cystatin clones were identified by using a rabbit polyclonal antibody (whole rabbit serum) to human salivary cystatin (Shomers *et al.* 1982) and the Protoblot Immunoscreening System, following the manu-facturer's instructions. Several positive clones were purified by additional rounds of screening.

Screening of library with C4-4 probe and amplification of cDNA inserts by linear PCR (LPCR). An insert of cDNA clone, designated C4-4, which encodes the 5'-terminally truncated cystatin S (see the Results section) was used as a hybridization probe for selection of additional cystatin clones. DNA was labelled by random priming using the Genious Nonradioactive DNA Labeling and Detection Kit. Positive plaques were eluted, and aliquots of the eluted plaques were used directly for amplification *in vitro* by LPCR using a DNA thermocycler (Smith *et al.*, 1990). A typical reaction mixture contained 10 μl of eluted plaque (template DNA), 10 μl of 10 × *Taq* DNA polymerase buffer [500 mM-KCl/100 mM-Tris/HCl (pH 9.0)/ 15 mM-MgCl₂/0.1% gelatin/1.0% Triton X-100], 4 μl of a 2.5 mM mix of all four dNTPs (0.1 mM final concn.), 1 μl each of the λ forward and reverse primer (77 ng), 0.5 μl (2.5 units) of *Taq* DNA polymerase and distilled water to a final volume of 100 μl. The reaction mixture was overlaid with 100 μl of mineral oil. In all, 30 cycles of amplification were performed at a denaturation temperature of 94 °C for 1 min, an annealing temperature of 52 °C for 2 min and an extension temperature of 72 °C for 2 min.

DNA sequence analysis. Cystatin inserts of antibody-positive clones were first subcloned either in the M13 (Messing, 1983) and/or pGem11 vectors. For sequencing in M13, the λ cDNA clones were first cut with the *Not*I enzyme and the resulting cohesive ends were blunt-ended before cutting with the second enzyme, *Eco*RI. The cDNA insert fragments were then subcloned into the *Eco*RI/*Sma*I sites of the M13 mp18 and mp19 vectors. For sequencing in the pGem, the clones were digested with the *Sfi*I and *Not*I enzymes and the cDNA inserts were subcloned into the *Sfi*I/*Not*I sites of the pGem-11Zf (+) vector. The nucleotide sequences were determined by the dideoxynucleotide-chain-termination method of Sanger *et al.* (1977) using the Sequenase kit and [α-³⁵S]dATP. With PCR amplified DNA, direct sequencing was employed utilizing also the Sequenase kit and the λ forward and reverse primers, which were first labelled with [γ-³²P]ATP and polynucleotide kinase. The PCR-amplified DNA was gel-purified in order to remove the excess primers and the dNTPs, and then further purified with a Geneclean kit. The labelling, annealing and sequencing reactions were performed as described by Kusukawa *et al.* (1990), with some modifications. To the annealed mix of the primer and DNA template, 1 μl of 100 mM-DTT, 1 μl of manganese buffer (100 mM-MnCl₂/150 mM citric acid, pH 7.0) and 0.5 μl of double-distilled water were added. After mixing, 3.5 μl of the mixture were added to each of the four termination mixes (2.5 μl) in which the dNTPs/ddNTPs ratio was increased from 80:8 (found in the kit) to 240:8.

DNA sequence management. The University of Wisconsin Genetics Computer Group software was used to manage the

DNA sequence information (Devereux *et al.*, 1984). For DNA alignment, the GAP program was used (Needleman & Wunsh, 1970).

Hybridization *in situ*

Preparation of probes for hybridization *in situ*. The pGem-11Zf(+) plasmid vector containing the cDNA clone C4-4 was used as the template. The plasmid DNA was first linearized to produce 'run-off' transcripts derived from the insert sequence only, and the linearized DNA was phenol-extracted and ethanol-precipitated. For generation of the anti-sense RNA probe, about 250 ng of plasmid were linearized with *Sfi*I and the SP6 RNA polymerase was used for transcription, following the protocols described by Promega. For the negative control probe (RNA sense strand), the same amount of plasmid was linearized with *Not*I, and the T7 RNA polymerase was used for transcription. Typically, the transcription reactions contained linearized plasmid DNA (250 ng), 4.0 μl of 5 × transcription buffer [200 mM-Tris/HCl (pH 7.5) 30 mM-MgCl$_2$/10 mM-spermidine/50 mM-NaCl], 2.0 μl of 100 mM-DTT, 4.0 μl of 2.5 mM each of ATP, CTP and GTP, 8.0 μl of [α-^{35}S]UTP (sp. radioactivity 1350 Ci/mmol), 1 μl of RNasin (25 units) and 1 μl (10 units) of SP6 or T7 RNA polymerase. The reaction was allowed to proceed for 60 min at 37 °C.

Preparation of tissue sections for hybridization *in situ*. Frozen sections (6 μm thick) from parotid and submandibular glands were obtained at −20 °C. To reduce experimental variability, adjacent serial sections of parotid and submandibular tissues were prepared on single slides. Sections were then air-dried, fixed in 4 % paraformaldehyde, washed in PBS, dehydrated in graded alcohols, and stored at −20 °C until use for hybridization *in situ* (within 1 week).

Prehybridization treatment of tissue sections. Tissue slides were soaked in 0.2 M-HCl for 20 min at room temperature, rinsed in water for 5 min, soaked in 2 × SSC for 10 min at room temperature, and then heated in 0.1 × SSC at 60 °C for 10 min. Next, the slides were treated for 10 min with Pronase (250 μg/ml in 50 mM-Tris/HCl, pH 7.5, containing 5 mM-Na$_2$EDTA), after which the slides were rinsed for 1 min in glycine (2 mg/ml) in PBS. Slides were then rinsed twice for 1 min in PBS, followed by post-fixation in freshly prepared 4 % paraformaldehyde for 20 min and block fixation for 5 min in 3 × PBS. The slides were rinsed twice for 5 min in PBS and then treated with 0.25 % acetic anhydride in 100 mM-triethanolamine, pH 8.0, for 10 min to block polar and charged groups on tissue sections. Finally, the slides were rinsed four times in PBS for 5 min each and then dehydrated through 30, 60, 80, 95 and 100 % ethanol for 5 min each at room temperature.

Hybridization and washes. The slides were hybridized with 50 μl of hybridization mixture [50 % deionized formamide/1 × Denhardt's/300mM-NaCl/5mM-EDTA/20mM-Tris/HCl (pH 8.0)/100 mM-DTT/tRNA (50 μg/ml)] containing ~1 × 10^6 c.p.m. at 42 °C overnight in a tray with pre-warmed mineral oil. After rinsing in chloroform three times for 5 min each (in coplin jars with coverslips on), the slides were drained and air-dried. This was followed by washing with 4 × SSC at room temperature three times for 5 min each (coverslips removed). Next, the slides were treated with RNAase A (20 μg/ml) in 500 mM-NaCl/10 mM-Tris/HCl, pH 8.0, for 1 h at 37 °C and then washed in the same buffer without RNAase for 30 min at 37 °C. This was followed by washing the slides in 2 litres of 2 × SSC at 45 °C for 30 min (with stirring) and in 0.1 × SSC for 45 min. Last, the slides were dehydrated sequentially in 70 and 90 % ethanol/0.3 M-ammonium acetate for 10 min each and then air-dried.

Emulsion coating and developing of slides. Liquid-emulsion

autoradiography was performed with Kodak NTB-2 emulsion diluted 1:1 with water at 4 °C for 10 days. The slides were then processed with Kodak D-19 developer. The slides were then dehydrated sequentially with 70 and 90 % ethanol/0.3 M-ammonium acetate for 5 min each, air-dried and stained with Harris hematoxylin and eosin Y.

Immunocytochemistry

Salivary-gland tissues were fixed for 10 h in 10 % buffered formalin and then paraffin-embedded. Sections (5 μm) were then de-paraffinized, hydrated and rinsed in TBS, pH 7.2, and then incubated with normal swine serum (1:5, v/v) for 20 min. The sections were next incubated with rabbit anti-salivary cystatin (1:200, v/v) for 30 min at room temperature in a humidified atmosphere. After the primary-antibody incubation and a TBS wash, slides were incubated for 30 min with a biotinylated swine anti-rabbit IgG (1:500, v/v). After another TBS wash, slides were incubated for 30 min with avidin–biotin–alkaline phosphatase prepared according to the manufacturers' directions. After a TBS wash, the chromogenic substrate solution (naphthol AS-MX phosphate and Fast-Red TR salt containing levamisole to block endogenous alkaline phosphatase activity) was added. Sections were counterstained with Mayer's hematoxylin. Negative controls consisted of replacing the primary antibody with non-immune rabbit serum (1:200, v/v).

RESULTS

Construction of the SMG cDNA library and selection and analysis of cDNA clones encoding salivary cystatins

In order to utilize both the nucleic acid and antibody probes for the library screening, λ gt11 Sfi-Not expression vector was used. Since this vector is also designed for orientation-specific cDNA cloning, this method doubles the in-frame orientation for the expression of inserts over conventional random-orientation cloning and thus increases the possibility of isolating specific cDNA using antibody probes. Screening of about 30000 recombinant clones from both the unamplified and amplified human SMG cDNA library with the anti-cystatin antibody revealed that about 1 % reacted positively. Nine positive clones were plaque-purified by additional rounds of screening and analysed further. Phage recombinant DNA was isolated from liquid minicultures of infected Y1090 (Sambrook *et al.*, 1989), digested with *Eco*RI/
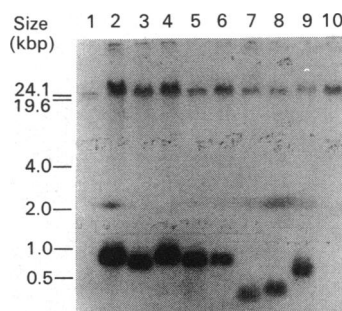

Fig. 1. Southern-blot analysis of cDNA clones which were identified as positive with anti-cystatin antibody

DNA was digested with *Eco*RI/*Not*I restriction enzymes, fragments separated on a 1.4 %-agarose gel, transferred on to a Nitroplus 2000 hybridization membrane and hybridized to the insert of cDNA clone C4-1 (lane 4). Lanes represent the following clones: lane 1, C1-2; lane 2, C1-1; lane 3, C1-3; lane 4, C4-1; lane 5, C4-3a; lane 6, C4-3b; lane 7, C4-2; lane 8, C4-4; lane 9, C4-5a; lane 10, C4-5b.
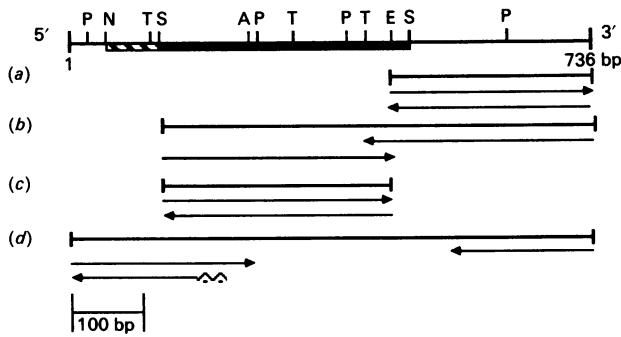
Fig. 2. Restriction-enzyme map of clone C3 and DNA sequence strategy for clones C3 and C4-4

Key to restriction enzymes: A, *Aha*II; E, *Eco*RI; N, *Nco*I; No, *Not*I; P, *Pst*I; S, *Sty*I; Sf, *Sfi*I; T, *Taq*I. The thin lines in the top diagram represent the 5'- and 3'-untranslated regions; the hatched bar, the putative signal peptide; the thick solid line, the secreted protein. The cDNA clone C4-4 subcloned in M13 mp18 and mp19 as the *Eco*RI–*Not*I fragment is represented in (*a*), as the *Sfi*I–*Not*I fragment in pGem-11Zf(+) in (*b*) and as the *Sfi*I/*Eco*RI fragment in pGem-11Zf(+) in (*c*). The PCR-amplified insert of clone C3 is represented in (*d*). Arrows indicate the direction and extension of sequencing. ∿ in (*d*) represents an internal cystatin SN primer.
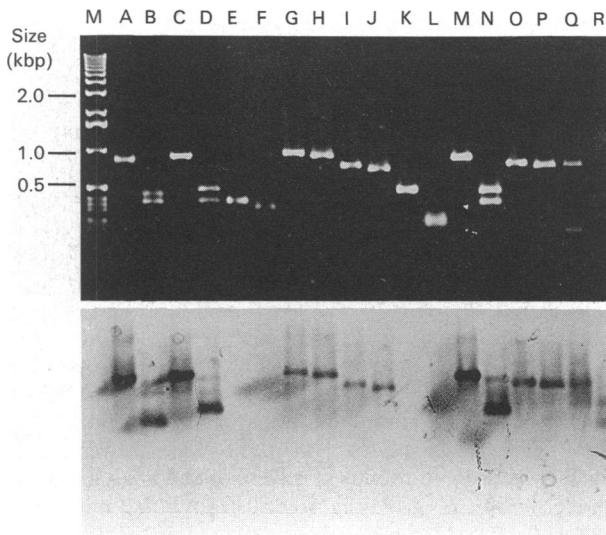

Fig. 3. Agarose-gel and Southern-blot analysis of PCR amplified inserts of cDNA clones selected with insert of clone C4-4

Undigested (lanes A, C, E, G, I, K, M, O and Q) and *Eco*RI-digested DNA (lanes, B, D, F, H, J, L, N, P and R) was separated by agarose-gel electrophoresis, stained with EtBr (top) and transferred to a Nitroplus 2000 hybridization membrane and probed with a *Sfi*I–*Eco*RI (5'-end) fragment of clone C4-4 (bottom). Lanes represent the following clones: lanes A and B, C4-4; lanes C and D, C3; lanes E and F, C4; lanes G and H, C5; lanes I and J, C7; lanes K and L, C10; lanes M and N, C12; lanes O and P, C16; and lanes Q and R, C17.

*Not*I restriction enzymes and analysed by Southern blotting using one of the longest inserts (C4-1) as a hybridization probe (Fig. 1). This probe cross-hybridized with seven of the nine clones identified. Owing to the extensive antibody and nucleic acid cross-reactivity among the isoforms of salivary cystatins, a partial DNA sequence analysis was performed to determine which cystatins were encoded by these antibody-positive clones. Six clones encoded cystatin SN, indicating that the mRNA encoding cystatin SN is much more abundant in the SMG than

```
  1 GGCTCTCACCCTCCTCTCCTGCAGCTCCAGCTCTGTGCTCTGCCTCTGAG 50
    |||||||||||||||||||||||||||||||  ||||||||||||||||||
  1 GGCTCTCACCCTCCTCTCCTGCAGCTCCAGCTTTGTGCTCTGCCTCTGAG 50

            -20.          .            -10.          .
          MetAlaArgProLeuCysThrLeuLeuLeuLeuMetAlaThrLeu
 51 GAGACCATGGCCCGGCCTCTGTGTACCCTGCTACTCCTGATGGCTACCCT 100
    |||||||||||||| || |||| |||||||||  ||||||  |||| |||||
 51 GAGACCATGGCCCAGCATCTGAGTACCCTGCTGCTCCTGCTGGCCACCCT 100
          MetAla GlnHisLeuSerThrLeuLeuLeuLeuLeuAlaThrLeu

            .    -1 1.           .            10    .
          AlaGlyAlaLeuAlaSerSerSerLysGluGluAsnArgIleIleProGly
101 GGCTGGGGCTCTGGCCTCGAGCTCCAAGGAGGAGAATAGGATAATCCCAG 150
    |||| ||| |||||||   |||| |||||||||||||| ||||||||||| |
101 AGCTGTGGCCCTGGCCTGGAGCCCCAAGGAGGAGGATAGGATAATCCCGG 150
          AlaValAlaLeuAla TrpSerProLysGluGluAspArgIleIleProGly

                            .          20              .
          GlyIleTyrAspAlaAspLeuAsnAspGluTrpValGlnArgAlaLeu
151 GTGGCATCTATGATGCAGACCTCAATGATGAGTGGGTACAGCGTGCCCTT 200
    ||||||||||| |  |||||||||||||||||||||||||||||||||||
151 GTGGCATCTATAACGCAGACCTCAATGATGAGTGGGTACAGCGTGCCCTT 200
          GlyIleTyr AsnAlaAspLeuAsnAspGluTrpValGlnArgAlaLeu

            30        .          .          40          .
          HisPheAlaIleSerGluTyrAsnLysAlaThrGluAspGluTyrTyrArg
201 CACTTCGCCATCAGCGAGTACAACAAGGCCACCGAAGATGAGTACTACAG 250
    |||||||||||||||||||||||| ||||||||| |||||| |||||||||
201 CACTTCGCCATCAGCGAGTATAACAAGGCCACCAAAGATGACTACTACAG 250
          HisPheAlaIleSerGluTyrAsnLysAlaThr LysAspAsp TyrTyrArg

            .          50          .          .      60
          ArgProLeuGlnValLeuArgAlaArgGluGlnThrGlnGlyValAlaAsn
251 ACGCCCGCTGCAGGTGCTGCGAGCCAGGGAGCAGACCTTTGGGGGGGTGA 300
    ||| ||||| ||| ||  ||||||  | ||||||||| ||||||||||||
251 ACGTCCGCTGCGGGTACTAAGAGCCAGGCAACAGACCGTTGGGGGGGTGA 300
          ArgProLeu ArgValLeuArgAlaArg GlnGlnThr ValGlyGlyValAsn

                        .          70          .          .
          TyrPhePheAspValGluValGlyArgThrIleCysThrLysSerGln
301 ATTACTTCTTCGACGTAGAGGTGGGCCGCACCATATGTACCAAGTCCCAG 350
    ||||||||||||||||||||||||||||||||  ||||||||||||||||
301 ATTACTTCTTCGACGTAGAGGTGGGCCGAACCATATGTACCAAGTCCCAG 350
          TyrPhePheAspValGluValGlyArgThrIleCysThrLysSerGln

            80          .          .          90          .
          ProAsnLeuAspThrCysAlaPheHisGluGlnProGluLeuGlnLysLys
351 CCCAACTTGGACACCTGTGCCTTCCATGAACAGCCAGAACTGCAGAAGAA 400
    ||||||||||||||||||||||||||||||||||||||||||||||||||
351 CCCAACTTGGACACCTGTGCCTTCCATGAACAGCCAGAACTGCAGAAGAA 400
          ProAsnLeuAspThrCysAlaPheHisGluGlnProGluLeuGlnLysLys

            .          100         .          110         .
          GlnLeuCysSerPheGluIleTyrGluValProTrpGluAspArgMetSer
401 ACAGTTATGCTCTTTCGAGATCTACGAAGTTCCCTGGGAGGACAGAATGT 450
    |||||| ||||||||||||||||||||||||||||||||||| |||||| ||
401 ACAGTTGTGCTCTTTCGAGATCTACGAAGTTCCCTGGGAGAACAGAAGGT 450
          GlnLeuCysSerPheGluIleTyrGluValProTrpGlu AsnArg Arg Ser

            .          120         .          .          .
          LeuValAsnSerArgCysGlnGluAlaEnd
451 CCCTGGT GAATTCCAGGTGTCAAGAAGCCTAGGGGTCTGTGCCAGGCCAG 500
    |||||||| |||||||||| |||||| ||||||||| |||||||||||||
451 CCCTGGTGAAATCCAGGTGTCAAGAATCCTAGGGATCTGTGCCAGGCCAT 500
          LeuVal LysSerArgCysGlnGlu SerEnd

            .          .          .          .          *
501 TCACACCGACCACCACCCACTCCCACCCCCTGTAGTGCTCCCACCCCTGG 550
    || |||| ||||||||||||||||||||||||||||||||||||||||||
501 TCGCACCAGCCACCACCCACTCCCACCCCCTGTAGTGCTCCCACCCCTGG 550

                .          .          .          .          .
551 ACTGGTGGCCCCCACCCTGCGGGAGGCCTCCCCATGTGCCTGTGCCAAGA 600
    |||||||||||||||||||||||||||||||||||||||||||| |||||||
551 ACTGGTGGCCCCCACCCTGCGGGAGGCCTCCCCATGTGCCTGCGCCAAGA 600

            .          .          .          .          .
601 GACAGACAGAGAAGGCTGCAGGAGTCCTTTGTTGCTCAGCAGGGCGCTCT 650
    |||||||||||||||||||||||||||||||||||||||||||||||||||
601 GACAGACAGAGAAGGCTGCAGGAGTCCTTTGTTGCTCAGCAGGGCGCTCT 650

            .          .          .          .          .
651 GCCCTCCCTCCTTCCTTCTTGCTTCTAATAGACCTGGTACATGGTACACA 700
    ||||||||||||||||||||||||||| |||| ||||||||||||||||||
651 GCCCTCCCTCCTTCCTTCTTGCTTCTCATAGCCCTGGTACATGGTACACA 700

701 CACCCCCACCTCCTGC AATTAAA CAGTAGCATCGCC 736
    |||||||||||||||||||||||||||||||||||||
701 CACCCCCACCTCCTGC AATTAAA CAGTAGCATCGCC 736
```

Fig. 4. Comparison of nucleotide and deduced amino acid sequences of cDNA inserts encoding cystatin S (clone C3 and C4-4, top lines) and cystatin SN (clone C4-1, bottom lines)

The amino acids in SN which differ from those of S are underlined; the *Eco*RI site in cystatin S (nucleotides 458–463) is underlined once and the polyadenylation sites in both clones are underlined twice. * Indicates the last common nucleotide of clones C4-1 and 10G5.1 (see the text for details).

the mRNAs encoding the other types of salivary cystatin. The seventh clone (C4-4, Fig. 1, lane 8) encoded only the last seven amino acids of a secreted cystatin, as well as the 3'-untranslated
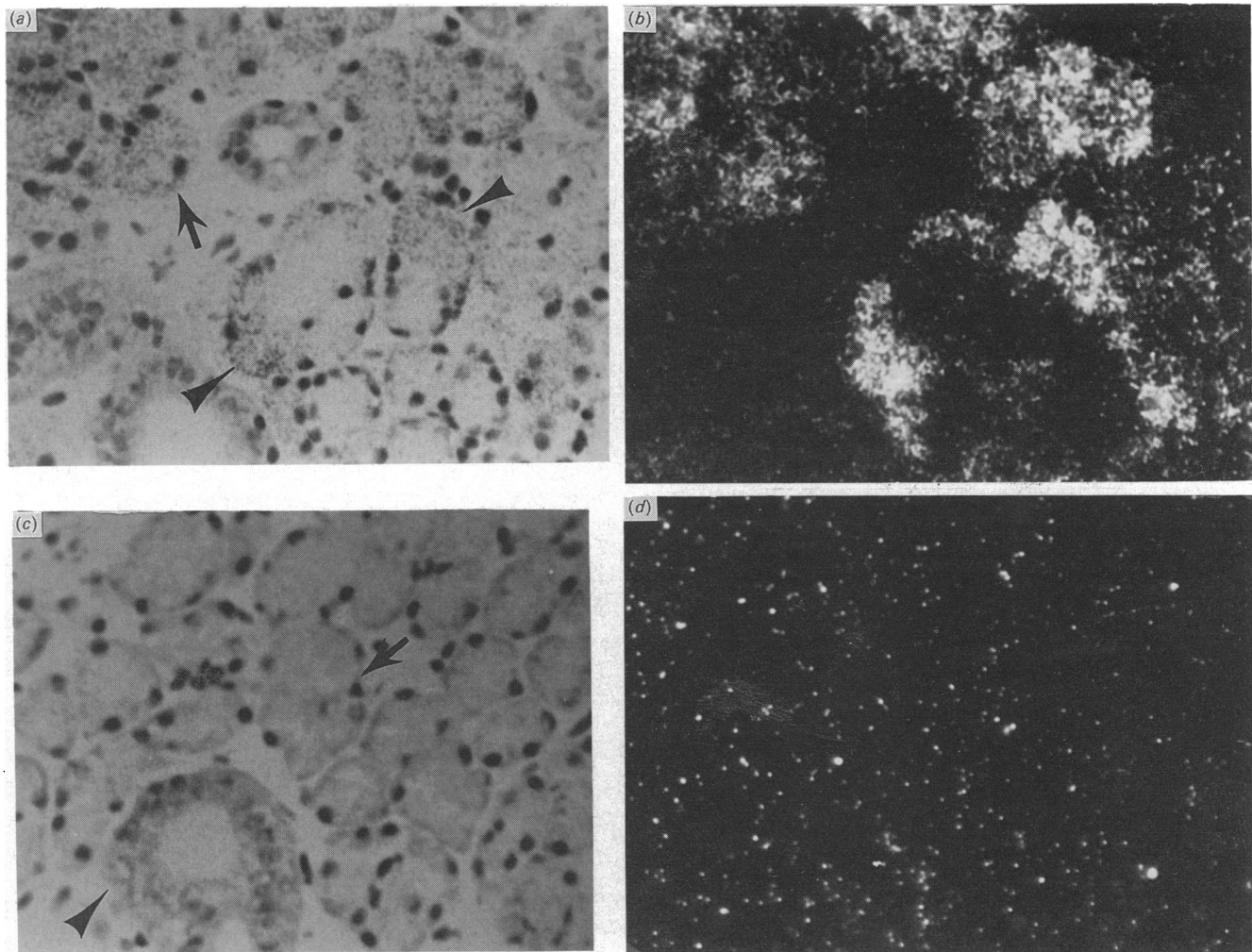
Fig. 5. Hybridization analysis *in situ* of human SMG

(*a*) SMG section hybridized to the anti-sense cRNA probe. A strong signal in the acinar serous cells (arrow) and serous demilunes (arrowhead) was observed (bright-field illumination; magnification × 200). (*b*) Dark-field illumination of (*a*) showing intense hybridization signal in the salivary acini and no signal in the ductal system. (*c*) SMG section hybridized to the sense cRNA probe; numerous acini (arrow) and an intralobular duct (arrowhead) are shown (bright-field illumination; magnification × 200). (*d*) Dark-field illumination of (*c*), demonstrating only background hybridization.

region (see Figs. 2 and 4). The last amino acid (residue 121) differed from that of cystatin SN (alanine as against serine). This suggested that clone C4-4 encoded either cystatin SA or S and contained an internal *Eco*RI recognition site and that by sub-cloning the *Eco*RI–*Not*I fragment most of the original insert was lost. The entire insert of the λ C4-4 clone was therefore cloned as a *Sfi*I–*Not*I fragment into the *Sfi*I/*Not*I sites of the pGem-11Zf(+) vector and sequenced (see Fig. 2*b*). Sequencing data confirmed that the insert of clone C4-4 contained an *Eco*RI site, that it encoded cystatin S and showed that this was not a full-length cDNA clone; since it lacked the 5′-untranslated region, the leader peptide region and the first four nucleotides of the secreted peptide. However, it contained the entire 3′-untranslated region, including the polyadenylation signal and the poly(A) tail. To confirm the sequence of the *Sfi*I–*Eco*RI fragment on the other strand, the *Sfi*I–*Eco*RI fragment was subcloned into the pGem-11Zf(+) vector and sequenced (Fig. 2*c*). The nucleotide sequence of an entire insert of clone (C4-1) encoding cystatin SN was also determined (Fig. 4, lower sequence) and compared with the nucleotide and decoded amino acid sequences of cDNA clones C4-4 and C3, both encoding cystatin S (see below).

Isolation and sequence analysis of a full-length cDNA encoding cystatin S

The SMG cDNA library was rescreened using the insert of clone C4-4 as the hybridization probe. Twenty positive clones were plaque purified, and several of these (with C4-4 as a positive control) were used directly for amplification *in vitro* of DNA by LPCR, thus eliminating the need to isolate λ DNA (as described in the Materials and methods section). To determine which inserts of the amplified clones encoded cystatin S (or SA) and which were larger in size than the clone C4-4, the PCR products were digested with *Eco*RI. The undigested and digested DNA was analysed by agarose-gel electrophoresis (Fig. 3, upper panel). Four cDNA inserts out of eight amplified (besides the control C4-4 clone, lane B) contained an *Eco*RI recognition site and therefore putatively encoded cystatin S or SA. Two of these undigested inserts (C3, lane C; and C12, lane M) were larger than the insert of the C4-4 clone (lane A). To determine whether the C3 and C12 inserts were longer at their 5′-end [rather than in their poly(A)⁺ tail length], a Southern blot of the same gel was probed with a cystatin 5′-end-specific probe (*Sfi*I–*Eco*RI fragment of the C4-4 insert). The hybridization results, shown
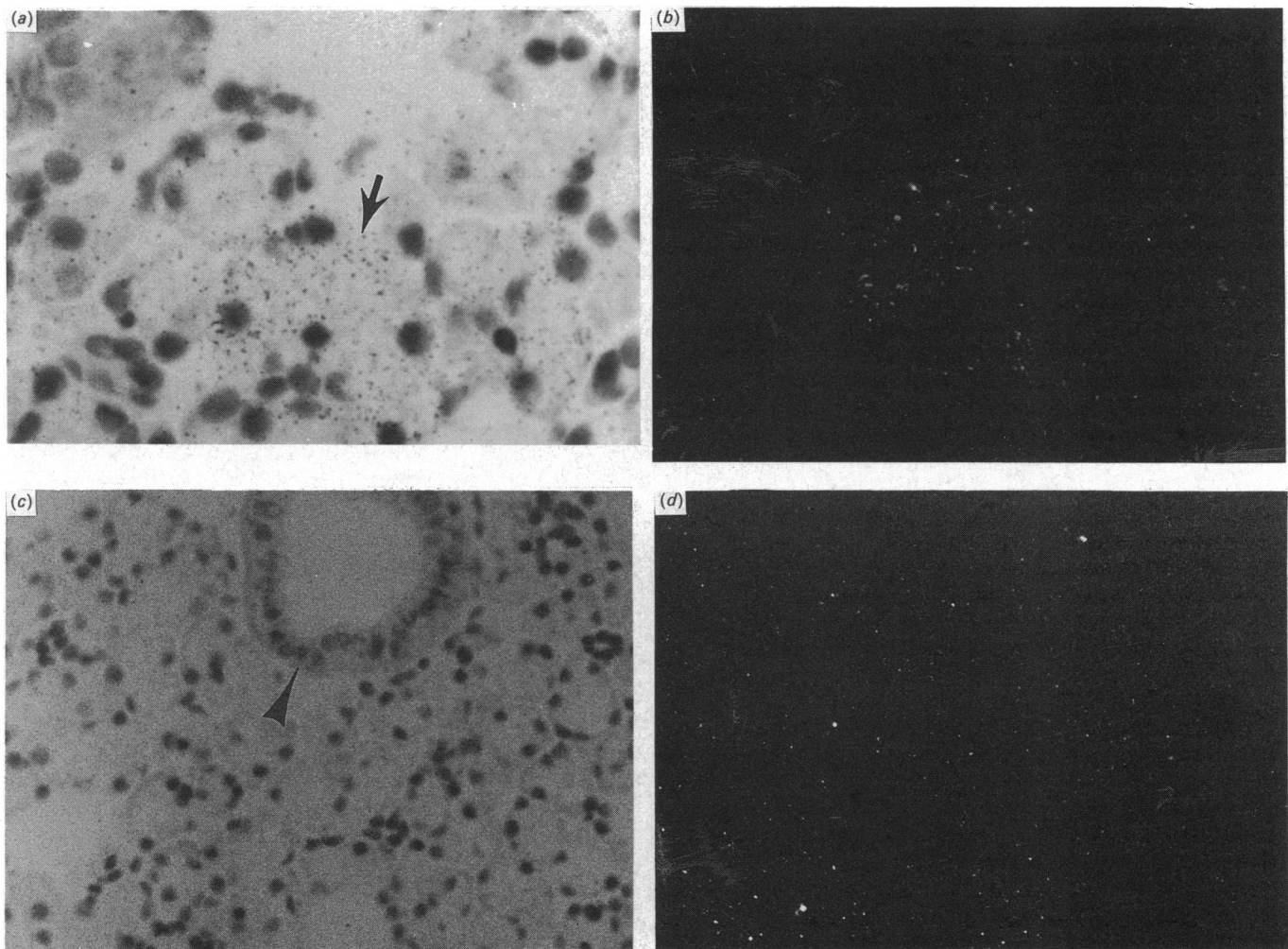
Fig. 6. Hybridization analysis _in situ_ of human PG

(a) Parotid-gland section hybridized to the anti-sense cRNA probe. A focal scarce deposition of silver grains in the cytoplasm of serous acinar cells (arrow) was observed in the salivary-gland parenchyma (bright-field illumination; haematoxylin/eosin stain; magnification × 400). (b) Dark-field illumination of (a) to depict the weak hybridization signal observed in the PG tissues. (c) PG section hybridized to the sense cRNA probe, showing an intralobular duct (arrowhead) amid serous acini (bright-field illumination; haematoxylin/eosin stain; original magnification × 200). (d) Dark-field illumination of (c), showing only background hybridization.

in Fig. 3 (lower panel), confirmed that both C3 (lane D) and C12 (lane N) contained a larger hybridizing fragment than C4-4 (lane B).

Direct sequencing of the double-stranded PCR-amplified cDNA inserts of C3 and C12 was performed as described in the Materials and methods section (Kusukawa _et al._, 1990). Problems associated with direct sequencing of double-stranded DNA generated by PCR can include carry-over of primers and template re-annealing. These were overcome by purification of PCR products on a low-melting agarose, use of a Geneclean Kit (which also removes dNTPs), use of ^{32}P-labelled sequencing primers and annealing of primer to template by a snap-cooling procedure. When the termination mixes supplied in the Sequenase kit were used without modification, the sequencing reactions terminated very quickly, resulting in a short sequencing ladder. After increasing the dNTP to ddNTP ratio 3-fold (from 80:1 to 240:1), the sequencing ladder was readable for about 200 nucleotides (results not shown). Both C3 and C12 inserts encoded cystatin S. C3 is the same length as C1-1 and C4-1 (which both encode cystatin SN), whereas C12 is 20 bp shorter at its 5′-end. The composite nucleotide sequences and the decoded amino acid sequences of the C3 and C4-4 clones are shown in Fig. 4 (upper

sequence), where they are compared with clone C4-1 encoding cystatin SN (lower sequence). Clone C3 is 736 bp long [excluding the poly(A) tail] and consists of 56 bp of the 5′-untranslated region, followed by 60 bp encoding a putative signal peptide of 20 amino acids, 366 bp encoding the secreted protein of 121 amino acids and the stop codon, and 254 bp of the 3′-untranslated region containing a polyadenylation signal AATTAAA located 13 bp before the poly(A) tail. The sequence of C4-1, encoding cystatin SN, agrees with the sequence of the clone designated 10G5.1, previously described by Al-Hashimi _et al._ (1988) with some exceptions. The first 14 nucleotides found at the 5′-end of clone 10G5.1 was not found in any of the clones analysed in the present study, and the sequence of C4-1 loses its similarity to 10G5.1 in the 3′-untranslated region after a nucleotide marked by an asterisk. The 3′-untranslated region of clone C4-1 agrees with that of genomic DNA clone encoding cystatin SN (Saitoh _et al._, 1987). Therefore the part of the 3′-untranslated region of the clone 10G5.1 may represent a cDNA cloning artefact. The C4-1 clone is also longer than clone 10G5.1; it contains the entire 3′-untranslated region (254 bp versus 98 bp in 10G5.1). Three other cystatin clones analysed in the present study have one nucleotide substitution (A for G) in a sequence corresponding
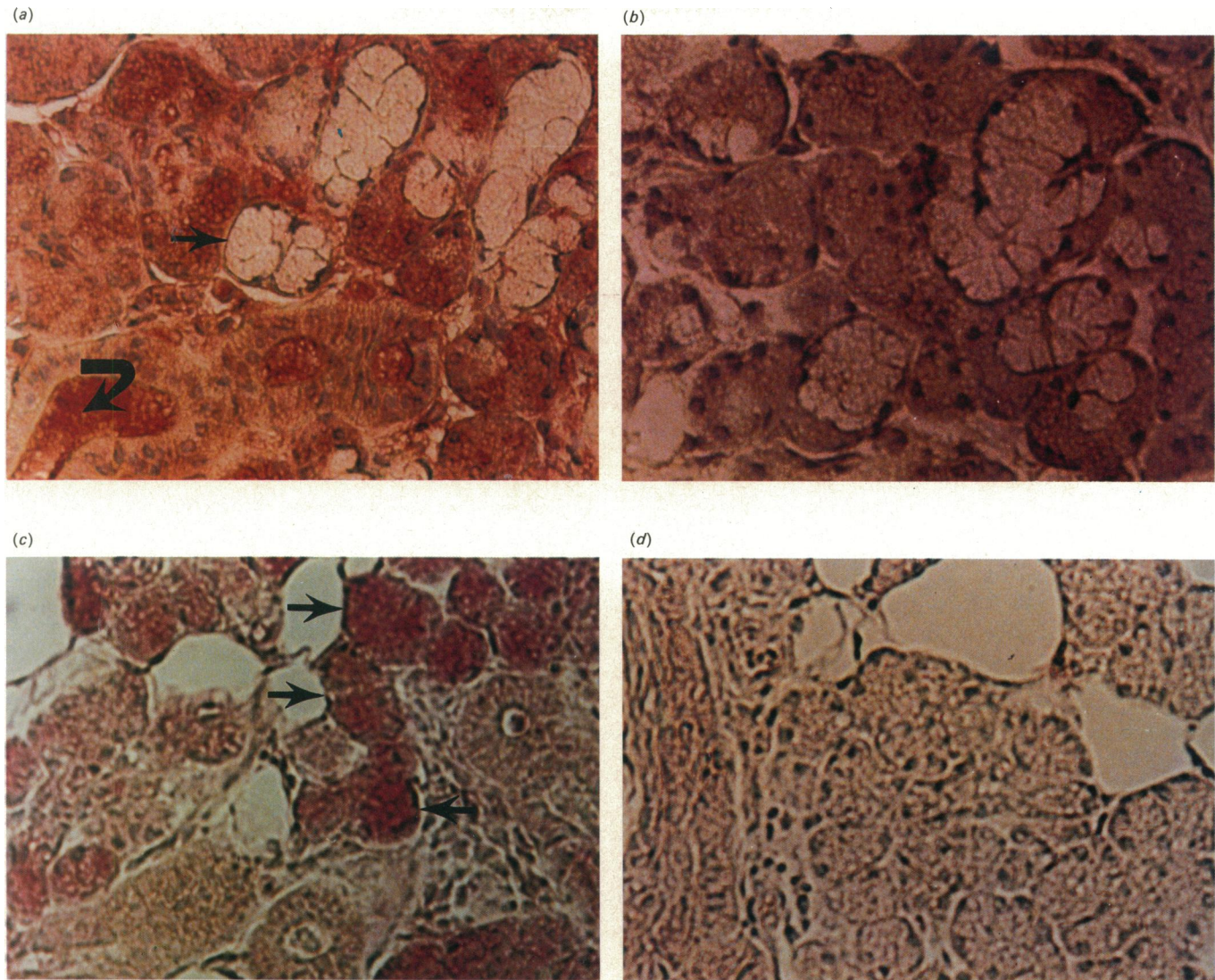
Fig. 7. Immunocytochemistry of human SMG and PG (Mayer's haematoxylin counterstain; magnification × 200)

(a) SMG tissue section incubated with rabbit anti-cystatin antibody. An intense signal was detected in numerous serous cells and the remaining intraductal saliva (curved arrow). The mucous acini (arrow) were not stained. (b) SMG tissue section incubated with rabbit normal serum (negative control). No signal was detected in the sections examined. (c) PG section incubated with rabbit anti-cystatin antibody. The signal was detected in limited numbers of serous acini (arrows). (d) PG-tissue section incubated with rabbit normal serum (negative control). No signal was observed in the sections examined.

to amino acid 20 of the secreted peptide. This substitution changes asparagine to serine and could represent a genetic variation or an artefact of cDNA synthesis.

Hybridization in situ

Hybridization in situ to SMG tissue sections demonstrated that the vast majority of serous acinar cells and serous demilunes strongly hybridized with the ^{35}S-labelled anti-sense RNA probe (riboprobe), indicating a high concentration of cystatin mRNA. By contrast, the mucous acini did not express the cystatin genes. Salivary cystatin mRNA was also absent in the intercalated, striated and excretory ductal cells, as well as myoepithelial cells (Figs. 5a and 5b). Negative controls consisted of tissue sections hybridized with the sense riboprobe (Figs. 5c and 5d). Similarly, mRNA in serous acinar cells of the PG hybridized to the same probe, whereas no hybridization signal specific for cystatin was observed in parotid ductal and myoepithelial cells (Figs. 6a and 6b). However, the hybridization signal in parotid tissue was less intense than in the submandibular tissue. Moreover, the signal

was not present in all the acinar cells, but only in discrete areas that showed a weak signal, slightly above background. This difference in signal hybridization intensity appears real, since submandibular and parotid tissue sections were run on the same slides, thereby diminishing the likelihood of experimental variation. Again, negative controls consisted of tissue sections hybridized with the sense riboprobe (Fig. 6c and 6d).

Immunocytochemistry

The immunocytochemical findings correlated with the results of hybridization in situ. Again, only glandular serous cells gave a detectable signal, whereas the mucous, myoepithelial, and ductal cells (intercalated, striated and excretory) were negative (Fig. 7a). Moreover, there was a stronger signal in the submandibular (Fig. 7a) than in the parotid tissues (Fig. 7c). In addition, different staining characteristics were observed in the tissues examined. The submandibular gland showed a homogeneous and intense staining of the serous cells, whereas the parotid tissue displayed both a focal and less intense staining of

the serous elements. All the sections incubated with the non-immune rabbit serum displayed no immunoreactivity (Figs. 7b and 7d).

DISCUSSION

The cDNA clones C3, C12 and C4-4 encoding salivary cystatin S show a high degree of identity with the cystatin SN cDNA clone C4-1. A high degree of identity was also found with the 3′ terminally truncated cystatin SN cDNA described by Al-Hashimi et al. (1988) and with the genomic clones, CST1 and CST2, encoding cystatin SN and SA respectively (Saitoh et al., 1987). The overall identity between clones C3 (S) and C4-1 (SN) was 94.6%. Interestingly, DNA identity was higher in the 5′- (98.2%) and 3′-untranslated regions (96.9%) than in the protein-coding regions (85.0% and 93.7% identities in the leader and secreted peptide regions respectively). Generally, protein-coding regions show higher identities than the flanking regions, owing to the higher selective pressure on the coding sequences. The decoded amino acid sequence of the secreted protein agrees with the sequence of cystatin S determined by Isemura et al. (1984a), except that residue 115 is asparagine rather than aspartic acid. Asparagine is also found in this position in cystatin SA (in both the protein isolated from saliva and decoded amino acid sequence of the genomic clone), whereas lysine is found in cystatin SN. In addition, the DNA sequence of this region in cystatin SA and S contains a distinguishing feature from cystatin SN, namely a recognition site for the restriction enzyme EcoRI.

Since in situ hybridization allows detection and localization of specific nucleic acid sequences directly within a cell or a tissue, the specific mRNAs can be detected in heterogeneous cell populations where only one of multiple sets of cells may be transcribing a particular message (Lum, 1986; Singer et al., 1986). The human SMG contains a mixture of serous and mucous acinar cells, whereas the PG contains just serous cells. Hybridization in situ localized the cystatin mRNA to the serous acinar cells of both SMG and PG, but a much higher concentration of salivary cystatin mRNA was found in the SMG. Higher levels of the salivary cystatin mRNAs in the SMG then in the PG were also found by Sabatini et al. (1989), who studied tissue distribution of cystatin mRNAs in human and macaque-monkey tissues by Northern-blot analyses using a human salivary cystatin partial cDNA clone as a hybridization probe. None of the other tissues they examined showed detectable hybridization signal.

The immunocytochemistry data indicated that the levels of salivary cystatin expression correlated with the amount of salivary cystatin mRNA and also localized the salivary cystatins to the serous cells of the submandibular and parotid glands. These results agree with the immunohistochemical findings of Isemura et al. (1984b).

Although both salivary cystatins and cystatin C belong to family 2 of the cystatin superfamily, their expression may be regulated differently. The expression of human salivary cystatins seems to be restricted to salivary tissues with far more mRNA found in SMG than in PG (see above). Cystatin C was expressed in every human tissue examined, including kidney, liver, pancreas, intestine, stomach, lung and placenta with the highest expression seen in seminal vesicles (Abrahamson et al., 1990); human salivary-gland tissues, however, were not examined. The authors postulated that this apparently non-tissue specific expression may be related to the structure of the 5′-flanking region of the cystatin C gene. In particular, the promoter region lacks a typical CAAT box and contains binding sites (GC-rich elements) for transcription factor Sp1, a property associated with the promoter region of 'housekeeping' genes. In contrast, the 5′-flanking

regions of cystatin SN and SA genes lack the Sp1 binding sites and contain the CAAT box (Saitoh et al., 1987).

In general, an increase in local proteinases and proteinase inhibitors have been observed in the presence of inflammation (Keilova & Tomasek, 1977; Jarvinen, 1978), acquired immuno-defficiency syndrome (Alavaikko et al., 1985) and malignancy (Sloane & Honn, 1984; Okumichi et al., 1984). In particular, salivary cystatins may participate in modulation of dental-plaque-mediated diseases by inhibiting cysteine proteinases of oral bacterial origin. Indeed, the potential anti-bacterial activity of salivary cystatins was first suggested by Obenauf et al. (1984). Also, Babnik et al. (1988) demonstrated the presence of the three families of cysteine-proteinase inhibitors (kininogens, stefin A, and cystatin C) in inflamed human gingiva. Skaleric et al. (1989) found an inverse correlation between cystatin C concentration in gingival samples and average pocket depth around the teeth of periodontal patients. Salivary cystatins are involved in the formation of the enamel pellicle acquired in vivo (Levine et al., 1985; Al-Hashimi & Levine, 1989) and may play a role in mineralization processes (Shomers et al., 1982b; Johnsson et al., 1991). Cystatins are also being investigated as potential anti-viral agents (Korant, 1988; Korant et al., 1988). Earlier studies demonstrated the ability of chicken cystatin to alter intracellular proteolytic processing of poliovirus proteins, resulting in a reduction of virus yield (Korant et al., 1985). Bjorck et al. (1990) also found that a recombinant human cystatin C displayed strong inhibitory effects on Herpes simplex virus type 1 replication, with no significant effect on poliovirus type 1.

A complete understanding of the function(s) of human salivary cystatins requires a detailed structural knowledge of these molecules. The availability of the cystatin cDNA clones will facilitate these studies. For example, recombinant salivary cystatins with different structural variations can be used to investigate their diverse biological properties. In fact, we have recently been able to express large amounts of biologically active cystatin SN and S in an E. coli system (Bobek et al., 1990, 1991; L. A. Bobek, X. Wang & M. J. Levine, unpublished work). Similarly, molecular biological reagents can be used to examine regulatory elements involved in the cystatin gene expression in both normal and diseased salivary tissues.

REFERENCES

Abrahamson, M., Barrett, A. J., Salvesen, G. & Grubb, A. (1986) J. Biol. Chem. 261, 11282–11289
Abrahamson, M., Olafson, I., Palsdottir, A., Ulvsback, M., LundWall, A., Jensson, O. & Grubb, A. (1990) Biochem. J. 268, 287–294
Alavaikko, M., Aine, R., Rinne, A., Jarvinen, M., Blanco, G., Apaja-Sarkkinen, M. & Hopsu-Havu, V. K. (1985) Int. J. Cancer 35, 319–325
Al-Hashimi, I., Dickinson, D. P. & Levine, M. J. (1988) J. Biol. Chem. 263, 9381–9387
Al-Hashimi, I. & Levine, M. J. (1989) Arch. Oral Biol. 34, 289–295
Anastasi, A., Brown, M. A., Kembhavi, A. A., Nicklin, M. J. H., Savers, C. A., Sunter, D. & Barrett, A. J. (1983) Biochem. J. 211, 129–138
Babnik, J., Curin, V., Lah, T., Turk, V. & Skaleric, U. (1988) Biol. Chem. Hoppe-Seyler 369, (Suppl.) 271–276
Barrett, A. J., Fritz, H., Grubb, A., Isemura, S., Jarvinen, M., Katunuma, N., Machleidt, W., Muller-Esterl, W., Sasaki, M. & Turk, V. (1986a) Biochem. J. 236, 312

Barrett, A., Rawlings, N., Davies, M., Machleidt, W., Salvesen, G. & Turk, V. (1986b) in Proteinase Inhibitors (Barrett, A. & Salvesen, G., eds.), pp. 515–569, Elsevier, Amsterdam

Bjorck, L., Grubb, A. & Kjellen, L. (1990) J. Virol. 64, 941–943

Bobek, L., Wang, X., Bobek, M. & Levine, M. J. (1990) J. Dent. Res. 69, (Special Issue) 168 (abstr. 468)

Bobek, L. A., Wang, X. & Levine, M. J. (1991) J. Dent. Res. 70 (Special Issue) 500 (abstr. 1876)

Bode, W., Engh, R., Musil, D., Thiele, U., Huber, R., Karshikov, A., Brzin, J., Kos, J. & Turk, V. (1988) EMBO J. 7, 2593–2599

Brzin, J., Popovic, T., Turk, V., Borchart, U. & Machleidt, W. (1984) Biochem. Biophys. Res. Commun. 118, 103–109

Devereux, P., Haeberli, P. & Smithies, O. (1984) Nucleic Acids Res. 12, 387–391

Ghiso, J., Jensson, O., & Frangione, B. (1986) Proc. Natl. Acad. Sci. U.S.A. 83, 2974–2978

Grubb, A. & Lofberg, H. (1982) Proc. Natl. Acad. Sci. U.S.A. 79, 3024–3027

Hawke, D. H., Yuan, P. M., Wilson, K. J. & Hunkapiller, M. W. (1987) Biochem. Biophys. Res. Commun. 145, 1248–1253

Isemura, S., Saitoh, E. & Sanada, K. (1984a) J. Biochem. (Tokyo) 96, 489–498

Isemura, S., Saitoh, E., Ito, S., Isemura, M. & Sanada, K. (1984b) J. Biochem. 96, 1311–1314

Isemura, S., Saitoh, E. & Sanada, K. (1986) FEBS Lett. 198, 145–149

Isemura, S., Saitoh, E. & Sanada, K. (1987) J. Biochem. (Tokyo) 102, 693–704

Jarvinen, M. (1978) J. Invest. Dermatol. 71, 114–118

Johnsson, M., Richardson, C. F., Campbell, A. A., Bergey, E. J., Levine, M. J. & Nancollas, G. H. (1991) Arch. Oral Biol., in the press

Keilova, H. & Tomasek, V. (1977) Acta Biol. Med. Ger. 36, 1873–1881

Korant, B. (1988) CRC Crit. Rev. Biotechnol. 8, 149

Korant, B., Brzin, J. & Turk, V. (1985) Biochem. Biophys. Res. Commun. 127, 1072–1076

Korant, B., Towatari, T., Kelley, M., Brzin, J., Lenarcic, B. & Turk, V. (1988) Biol. Chem. Hoppe-Seyler (Suppl.) 369, 281–286

Kusukawa, N., Uemori, T., Asada, K. & Kato, I. (1990) BioTechniques 9, 66–72

Levine, M. J., Tabak, L. A., Reddy, M. & Mandel, I. D. (1985) in Molecular Basis of Oral Microbial Adhesion (Mergenhagen, S. & Rosan, B., eds.), pp. 125, American Society of Microbiologists, Washington

Lum, J. B. (1986) BioTechniques 4, 32–39

Messing, J. (1983) Methods Enzymol. 101, 20–78

Muller-Esterl, W., Iwanaga, S. & Takanishi, S. (1986) Trends Biochem. Sci. 11, 336–339

Needleman, S. B. & Wunsh, C. (1970) J. Mol. Biol. 48, 443–453

Obenauf, S. D., Fisher, R. H., Cowman, R. A. & Fitzgerald, R. J. (1984) Infect. Immun. 46, 797–801

Okumichi, T., Nishiki, M., Takasugi, S., Yamane, M. & Ezaki, H. (1984) Hiroshima J. Med. Sci. 33, 801

Rawlings, N. D. & Barrett, A. J. (1990) J. Mol. Evol. 30, 60–71

Sabatini, L. M., Warner, T. F., Saitoh, E. & Azen, E. A. (1989) J. Dent. Res. 68, 1138–1145

Saitoh, E., Kim, H.-S., Smithies, O. & Maeda, N. (1987) Gene 6, 329–338

Saitoh, E., Isemura, S., Sanada, K., Kim, H.-S., Smithies, O. & Maeda, N. (1988) Biol. Chem. Hoppe-Seyler 369 (Suppl.), 191–197

Sambrook, J., Fritsch, E. & Maniatis, T. (1989) Molecular Cloning, a Laboratory Manual (2nd edn.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Sanger, F., Nicklen, S. & Coulson, A. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 5463–5467

Shomers, J. P., Tabak, L. A., Levine, M. J., Mandel, I. D. & Ellison, S. A. (1982a) J. Dent. Res. 61, 973–977

Shomers, J. P., Tabak, L. A., Levine, M. J., Mandel, I. D. & Hay, D. I. (1982b) J. Dent. Res. 61, 397–399

Singer, R. B., Bentley-Lawrence, J. & Villnave, C. (1986) BioTechniques 4, 230–250

Skaleric, V., Babnik, J., Curin, V., Lah, T. & Turk, V. (1989) Arch. Oral Biol. 34, 301–305

Sloane, B. F. & Honn, K. V. (1984) Cancer Mets. Rev. 3, 249–263

Smith, D. P., Johnstone, E. M., Little, S. P. & Hsiung, H. M. (1990) BioTechniques 9, 48–54

Young, R. A. & Davis, R. W. (1983) Proc. Natl. Acad. Sci. U.S.A. 80, 1194–1198