Identification and Structural and Functional Characterization of Human Enamelysin (MMP-20)[†,‡]

Elena Llano,[§,||] Alberto M. Pendás,[§,||] Vera Knäuper,[⊥] Timo Sorsa,[&] Tuula Salo,[$] Eduardo Salido,[+] Gillian Murphy,[⊥] James P. Simmer,[@] John D. Bartlett,[#] and Carlos López-Otín*[,§]

Departamento de Bioquímica y Biología Molecular, Facultad de Medicina, Universidad de Oviedo, 33006 Oviedo, Spain, Department of Cell and Molecular Biology, Strangeways Research Laboratory, Cambridge CB1 4RN, United Kingdom, Department of Periodontology, Institute of Dentistry, University of Helsinki, 00014 Helsinki, Finland, Departments of Oral Surgery and Pathology, University of Oulu, 90220 Oulu, Finland, Unidad de Investigación, Hospital Universitario de Canarias, 38071 Tenerife, Spain, Department of Pediatric Dentistry, University of Texas Health Science Center, San Antonio, Texas 78284, and Department of Biomineralization, Forsyth Dental Center, Boston, Massachusetts 02115

Received August 26, 1997; Revised Manuscript Received October 16, 1997[⊗]

ABSTRACT: A cDNA encoding a new human matrix metalloproteinase (MMP) has been cloned from RNA prepared from odontoblastic cells. The open reading frame of the cloned cDNA codes for a polypeptide of 483 amino acids and is extensively similar to the sequence of recently described porcine enamelysin, suggesting that the isolated cDNA codes for the human homologue of this enzyme. Human enamelysin (MMP-20) has a domain organization similar to other MMPs, including a signal peptide, a prodomain with the conserved motif PRCGVPD involved in maintaining enzyme latency, a catalytic domain with a Zn-binding site, and a COOH-terminal fragment similar to the sequence of hemopexin. The calculated molecular mass of human enamelysin is about 54 kDa, which is similar to that of collagenases or stromelysins. However, this human MMP lacks a series of structural features distinctive of these subfamilies of MMPs. The full-length human enamelysin cDNA has been expressed in *Escherichia coli*, and the purified and refolded recombinant protein is able to degrade synthetic peptides used as substrates of MMPs, confirming that human enamelysin belongs to this family of proteases. Furthermore, the recombinant human enamelysin is able to degrade amelogenin, the major protein component of the enamel matrix. On the basis of its degrading activity on amelogenin, and its highly restricted expression to dental tissues, we suggest that human enamelysin plays a central role in the process of tooth enamel formation. Finally, we have found that the human enamelysin gene (MMP-20) maps to chromosome 11q22, clustered to at least seven other members of the MMP gene family.

The matrix metalloproteinases (MMPs)[1] or matrixins constitute a group of structurally related zinc-dependent endopeptidases that mediate the degradation of extracellular matrix components. These proteolytic enzymes have been involved in a series of matrix-remodeling processes occurring in physiological conditions such as cytotrophoblast implantation, embryonic development, fetal ossification, angiogenesis, or wound healing. In addition, MMPs have been associated with the degradative processes accompanying pathological conditions such as arthritis, atherosclerosis, periodontitis, pulmonary emphysema, and tumor invasion and metastasis (*1*−*4*). At present, the family of human MMPs is composed of 15 members that can be classified into at least four different subfamilies—collagenases, gelatinases, stromelysins, and membrane-type MMPs (MT-MMPs)—although there are some enzymes such as stromelysin 3 (*5*), macrophage metalloelastase (*6*), and MMP-19 (*7*) that do not appear to fall into any of these subfamilies. Since it seems clear that MMPs play essential roles in both normal and pathological conditions, over the last few years, we have been interested in examining the possibility that additional yet uncharacterized members of the MMP family could be produced by human tissues (*7*−*9*).

One of the physiological situations in which MMPs could play a critical role takes place during the process of enamel formation (*10*−*12*). According to data from different groups, it is well-established that dental enamel is formed as a consequence of a matrix-mediated biomineralization process. During the secretory stage, an extracellular organic matrix is made by the enamel organ which encloses the developing tooth germ. As the apatitic crystallites are formed, enamel proteins are secreted and cleaved by a series of proteases from different classes and specificities (*13*−*18*). Later, as

[†] This work was supported by grants from Comisión Interministerial de Ciencia y Tecnología-Spain (SAF94-0892 and SAF97-0258), EU-BIOMED II (Spain) (BMH4-CT96-0017), Glaxo-Wellcome (Spain), The Finnish Dental Society, The Academy of Finland, and The Arthritis and Rheumatism Council (U.K.). A.M.P. is a recipient of the Severo Ochoa fellowship from Ayuntamiento de Oviedo-Asturias (Spain).

[‡] EMBL accession number Y12779.

* Send correspondence to Dr. Carlos López-Otín, Departamento de Bioquímica y Biología Molecular, Facultad de Medicina, Universidad de Oviedo, 33006 Oviedo, Spain. Telephone: 34-85-104201. Fax: 34-85-103564. E-mail: clo@dwarf1.quimica.uniovi.es.

[§] Universidad de Oviedo.

[||] These authors contributed equally to this work.

[⊥] Strangeways Research Laboratory.

[&] University of Helsinki.

[$] University of Oulu.

[+] Hospital Universitario de Canarias.

[@] University of Texas Health Science Center.

[#] Forsyth Dental Center.

[⊗] Abstract published in *Advance ACS Abstracts,* November 15, 1997.

[1] Abbreviations: bp, base pair(s); DAPI, diamidine-2-phenylindole dihydrochloride; FISH, fluorescent *in situ* hybridization; MMP, matrix metalloproteinase; PAGE, polyacrylamide gel electrophoresis; PCR, polymerase chain reaction; SDS, sodium dodecyl sulfate; STS, sequence-tagged site; TIMP, tissue inhibitor of metalloproteinases.

S0006-2960(97)02120-X CCC: $14.00

the enamel progresses through the transition and maturation stages of development, the organic matrix is removed, resulting in a mineralized tissue. Several groups have detected the presence of serine proteinases and metalloproteinases in the developing enamel matrix of porcine and bovine tissues which could be responsible for the degradation of extracellular matrix proteins, and particularly amelogenins, during the process of enamel maturation (*13*−*18*). Very recently, two of these proteins called ameloprotease I and enamelysin have been isolated from porcine tissues and characterized at the amino acid sequence level (*19, 20*). However, and despite the suggested central role of proteolytic enzymes in enamel formation, at present, very little is known about the nature and catalytic properties of the enzymes involved in this process in human tissues. In this work, we report the molecular cloning from odontoblastic cells and complete nucleotide sequence of a cDNA coding for a novel human MMP, MMP-20, which appears to be the homologue of porcine enamelysin. We also report the expression of this gene in prokaryotic systems and the preliminary enzymatic characterization of the recombinant enzyme. Finally, we report the chromosomal location of the human enamelysin gene and analyze its expression in human tissues. According to its restricted expression in dental tissues, and its ability to degrade amelogenins, we propose that human enamelysin plays a specific role in the process of tooth enamel formation.

MATERIALS AND METHODS

Materials. A high-density gridded human P1 artificial chromosome (PAC) genomic library was kindly supplied by the Human Genome Mapping Resource Center (Cambridgeshire, U.K.). Northern blots containing polyadenylated RNAs from different human tissues were purchased from Clontech (Palo Alto, CA). Oligonucleotides were synthesized in an Applied Biosystems (Foster City, CA) model 381A DNA synthesizer. Restriction endonucleases and other reagents used for molecular cloning were from Boehringer Mannheim (Mannheim, Germany). Double-stranded DNA probes were radiolabeled with [^{32}P]dCTP (3000 Ci/mmol) using a commercial random-priming kit purchased from Amersham (Amersham, U.K.).

Screening of a Human Genomic Library. A human P1 artificial chromosome (PAC) genomic library was hybridized at 50 °C in 5× SSPE (1× = 150 mM NaCl, 10 mM NaH_2PO_4, and 1 mM EDTA at pH 7.7), 1% SDS, 0.5% dried milk, and 6% PEG 8000, with a partial cDNA for porcine enamelysin (*20*). This probe was 780 bp long, and extended from positions 300 to 1080 of the porcine cDNA. Posthybridization washes were performed under low stringency conditions (2× SSC and 0.1% SDS at 50 °C). Finally, the membranes were exposed to XAR-5 film (Kodak) at −70 °C with intensifying screens. One positive clone was identified and characterized by endonuclease restriction analysis and nucleotide sequencing.

RNA Isolation from Human Odontoblasts and Pulpal Tissues. Intact third molars free of caries lesions as well as of developmental lesions were extracted during normal surgical treatment of systemically healthy young adult patients (18−25 years old). Immediately after tooth extraction, the teeth were stored in phosphate-buffered saline (PBS) at pH 7.4 and subsequently used for RNA extraction. The teeth were swabbed with 70% alcohol, and a horizontal grow was gently cut around the surface 2 mm deep about 5 mm apically from the cemento−enamel junction with a thin-sectioning machine (Gillings-Hamco, Hamco-Machines Inc., Rochester, NY). Care was taken not to reach the pulp. The root was dissected from the crowns with pliers. The pulpal tissue of the teeth was gently removed with forceps, collected in 120 μL of Trizol Reagent (Life Technologies, Paisley, U.K.), and extracted briefly by vortexing. The odontoblasts lining the pulpal chambers were scrabbed by sterile knife and placed in 30 μL of the same reagent. Total RNA was isolated from the pulpal and odontoblast tissues by following the instructions of the manufacturer of Trizol Reagent. The RNA concentration and purity were determined by spectrophotometry. Ten micrograms of RNA was reverse-transcribed by using the SuperScript kit (Life Technologies) using oligo(dT16) primers.

PCR Amplification of cDNA from Human Tissues. DNA was isolated from a series of cDNA libraries prepared from human tissues (liver, heart, kidney, muscle, placenta, brain, breast, breast carcinoma, ovary, testis, lung, pancreas, leukocytes, pulp, and odontoblasts) and PCR-amplified with two specific oligonucleotides: 5′-GTGAGGAGATGAAGGTGCTCCC-3′ (enamel 1) and 5′-CCTCCCAGGCCTTCTCCAGGA-3′ (enamel 4). The nucleotide sequence of these primers was deduced from the sequence of two putative exons identified in the genomic clone isolated for human enamelysin. The PCR was carried out in a GeneAmp 2400 PCR system from Perkin-Elmer/Cetus for 40 cycles of denaturation (94 °C, 15 s), annealing (58 °C, 15 s), and extension (72 °C, 45 s). The PCR products were phosphorylated with T4 polynucleotide kinase and cloned in the *Sma*I site of pUC18. The material amplified from odontoblast cells was extended toward the 3′-end by following a conventional RACE strategy, using cDNA from human odontoblasts and a specific oligonucleotide (5′-CCTCGGAAAGTATTCCTGG GGA-3′) derived from the most 3′ exon identified after sequence analysis of the human enamelysin genomic clone. Finally, the full-length cDNA was obtained by PCR amplification using the Expand Long PCR kit (Boehringer-Mannheim) (35 cycles consisting of 15 s at 94 °C, 15 s at 58 °C, and 4 min at 68 °C) with primers enamel 1 (including the putative start codon) and 5′-CTAGGCTTTTCTATTTAGCAACC-3′ (containing the stop codon). One microliter from this amplified material was further PCR-amplified under the same conditions using a sense nested primer (5′-ATGAAGGTGCTCCCTGCATCTGG-3′) and the same antisense primer (5′-CTAGGCTTTTCTATTTAGCAACC-3′). Following gel purification, the amplification product was cloned and sequenced as described above.

DNA Sequencing. Selected DNA fragments were inserted in the polylinker region of phage vector M13mp19 (*21*) and sequenced by the dideoxy chain termination method (*22*) using either the M13 universal primer or cDNA-specific primers and the Sequenase Version 2.0 kit (U.S. Biochemicals, Cleveland, OH). All nucleotides were identified in both strands. Sequence ambiguities were solved by substituting dITP for dGTP in the sequencing reactions according to the instructions of the manufacturer. Computer analysis of DNA and protein sequences was performed with the GCG software package of the University of Wisconsin Genetics Computer Group (*23*).

Chromosomal Mapping. DNA isolated from a panel of 24 monochromosomal somatic cell hybrids containing a single human chromosome in a mouse or hamster cell line background (kindly provided by the Human Genome Mapping Resource Center) was PCR-screened for the presence of a genomic sequence-tagged site (STS) generated by primers 5′-TGAGGTCAACGTCCAAGCCAC-3′ and 5′-CCCATCTCCTTCTCCTAACTGG-3′ obtained from a noncoding region of the human enamelysin gene. Amplification conditions were identical with those previously described for the generation of the cDNA probe. To perform chromosomal location by fluorescent in situ hybridization (FISH), DNA from the isolated PAC DNA clone was obtained with the standard alkaline lysis method using QIAGEN columns (QIAGEN Inc., Chatsworth, CA). Then, 2 μg of the PAC DNA was nick-translated with biotin 16−dUTP, whereas the centromeric probe for chromosome 11 was labeled with digoxigenin 11−dUTP (Boehringer Mannheim). Both probes were hybridized to normal male metaphase chromosomes obtained from phytohemagglutinin-stimulated cultured lymphocytes (*24*), and detected using two avidin−fluorescein layers (*25*). Chromosomes were diamidine-2-phenylindole dihydrochloride (DAPI)-banded and images were captured in a Zeiss axiophot fluorescent microscope equipped with a CCD camera (Photometrics).

Northern Blot Hybridization. Northern blots containing 2 μg of poly(A) plus RNA of different human tissue specimens were prehybridized at 42 °C for 3 h in 50% formamide, 5× SSPE (1× = 150 mM NaCl, 10 mM NaH_2PO_4, and 1 mM EDTA at pH 7.4), 10× Denhardt's, 2% SDS, and 100 mg/mL denatured herring sperm DNA and then hybridized with radiolabeled enamelysin full-length cDNA for 20 h under the same conditions. Filters were washed with 0.1× SSC and 0.1% SDS for 2 h at 50 °C and exposed to autoradiography. RNA integrity and equal loading were assessed by hybridization with an actin probe.

Construction of an Expression Vector for Human Enamelysin, Expression in Escherichia coli, and Refolding of the Recombinant Enzyme. The 1466 bp fragment of the enamelysin cDNA generated by *Pst*I cleavage was ligated in frame into the pRSET B *E. coli* expression vector (Invitrogen) previously cleaved with *Pst*I. The correct orientation of the fragment in the vector was confirmed by restriction analysis with *Eco*RI. The expression vector was transformed into BL21(DE3)pLysS competent *E. coli* cells and grown on agar plates containing chloramphenicol and ampicillin. Single colonies were used to inoculate 100 mL cultures in 2YT medium supplemented with 33 μg/mL chloramphenicol and 50 μg/mL ampicillin. Five hundred microliters of the corresponding overnight culture was used to inoculate 500 mL of 2YT medium containing the above antibiotics. After the culture reached an A_{600} of 0.6, expression was induced by addition of isopropyl 1-thio-β-D-galactopyranoside (IPTG) at a final concentration of 0.1 mM followed by overnight incubation at 30 °C. Human enamelysin was expressed under these conditions and showed an apparent molecular mass of 57 000 Da, which is in excellent agreement with the calculated molecular mass of the enzyme. The inclusion bodies containing the enzyme were extensively washed, solubilized in 20 mM Tris/HCl (pH 8.0), 6 M urea, and 1 mM β-mercaptoethanol, and further purified by gel filtration chromatography using S-200 HR equilibrated in 20 mM Tris/HCl (pH 8.0) and 3 M urea. Fractions containing the enzyme were refolded by dialysis against 20 mM Tris/HCl (pH 7.5), 5 mM $CaCl_2$, 150 mM NaCl, 50 μM $ZnCl_2$, and 0.01% NaN_3. The precipitated protein was removed by centrifugation, and the remaining enzyme solution contained active enamelysin which was used for the further characterization of the enzymatic activity.

Active Site Titration and Substrate Specificity Analysis. Initial analysis of the refolded recombinant enamelysin revealed that the synthetic fluorescent substrate McaPLGLDpaARNH$_2$ was hydrolyzed. It was therefore possible to perform active site titrations using recombinant TIMP-2 with a known concentration essentially as described by Willembrock *et al*. (*26*), and our data revealed that the enzyme was inhibited by TIMP-2 in a concentration-dependent fashion. We then determined the k_{cat}/K_m values for the hydrolysis of McaPLGLDpaARNH$_2$ and McaPLANvaDpaARNH$_2$ in routine assays at 37 °C and at substrate concentrations of 0.5 and 1.5 μM. Murine recombinant amelogenin (*27*) was incubated with recombinant active enamelysin at substrate concentrations of 200 ng/μL followed by analysis of the degradation products by SDS−PAGE. Parallel experiments were performed in the presence of TIMP-2.

RESULTS AND DISCUSSION

Identification, Cloning, and Sequence Analysis of the Human Enamelysin Gene. To examine the possibility that the human genome could contain genes related to that encoding the recently described porcine enamelysin, a human PAC genomic library was screened using as a probe a partial cDNA for porcine enamelysin (*20*). After hybridization under low stringency conditions, a single positive clone was identified. The insert contained in this clone was characterized by endonucleotide restriction analysis and nucleotide sequencing of selected fragments showing positive hybridization with the probe. This analysis allowed the identification of six putative exons whose sequence was significantly similar to those previously determined for other MMPs, including porcine enamelysin. As a previous step to obtain a complete cDNA sequence for this putative novel human MMP, we first synthesized two specific oligonucleotides, sense (enamel 1) and antisense (enamel 4), derived from the sequences determined for exons 1 and 4, respectively. These primers were used to screen for the possible expression of the identified human gene in different tissues. After PCR amplification using as a template cDNAs from a variety of human tissues, a band of the expected size (610 bp) was only amplified from dental tissues (odontoblasts and pulpal tissue). Northern blot analysis of RNAs obtained from multiple human tissues (leukocytes, colon, small intestine, ovary, testis, prostate, thymus, spleen, pancreas, kidney, skeletal muscle, liver, lung, placenta, brain, and heart) failed to show any positive hybridization signal with human enamelysin probes, thus confirming the restricted expression pattern of this gene (data not shown). Nucleotide sequence analysis of the PCR-amplified material from dental tissues revealed that it had a sequence identical with that derived from analysis of exons 1−4 of the isolated genomic clone. To extend the partial cDNA sequence toward the 3′-end, we performed 3′-RACE experiments using a specific oligonucleotide deduced from exon 6 and the cDNA from human odontoblasts as a template. These experiments led us finally to obtain a fragment long enough to contain the entire coding information for an archetypical MMP.

To generate a full-length cDNA molecule useful for subsequent *in vitro* expression experiments, two primers covering the start and stop codon of the cDNA were synthesized and used for PCR amplification using a commercial extralarge PCR Kit. A product of the expected size (about 1.5 kb) was obtained and its identity established by nucleotide sequencing. Analysis of this sequence (Figure 1) showed the presence of an open reading frame coding for a protein of 483 amino acids with a predicted molecular mass of 54.4 kDa. Amino acid sequence comparison of the identified sequence with that of porcine enamelysin revealed a high percentage of identities between them (89%), strongly suggesting that the cloned cDNA corresponds to the human homologue of this enzyme. Following the nomenclature system for MMPs (*28*), we would assign number 20 to this human metalloprotease, MMP-19 corresponding to the most recently described human MMP, cloned and characterized from a liver cDNA library (*7*). A computer search for amino acid sequence similarities between the human enamelysin (MMP-20) sequence and those of the different human MMPs characterized to date demonstrated a limited sequence similarity, with values of identities ranging from 45% with collagenase 3 to 33% with MMP-19. However, and despite these low values of sequence identities with human MMPs, the deduced amino acid sequence for human enamelysin presents all the structural features characteristic of members of this family of proteolytic enzymes. Thus, it contains the three domains that are conserved among all of them, including the predomain with a hydrophobic leader sequence which targets these enzymes to the secretory pathway, the prodomain containing the conserved PRCGVPD motif (at positions 98−104) involved in maintaining the latency of these proteinases, and the catalytic domain with the consensus sequence HE*X*GH*XX*G*XX*H (at positions 226−236) containing the three histidine residues involved in the coordination of the zinc atom at the active site. The catalytic domain also shows a methionine residue, seven residues C-terminal to the Zn-binding site, completely conserved in all MMPs, and proposed to play an essential role in the structure of the active sites of these enzymes (*29*). The deduced sequence for human enamelysin also contains a COOH-terminal domain of about 200 residues that is similar to the sequence of hemopexin and present in all family members with the exception of matrilysin and the soybean leaf MMP (*30*, *31*). Finally, the alignment of MMP-20 with the remaining MMPs characterized to date allows speculation about the putative cleavage site of the proenzyme form of the molecule for rendering the mature proteinase. According to this amino acid comparison around the activation site, the active processed form of MMP-20 would start at the Tyr residue located at position 108, and would have a calculated molecular mass of 42.6 kDa.

We next performed a more detailed analysis of the amino acid sequence of human enamelysin in an attempt to classify this enzyme in one of the main MMP subfamilies, including collagenases, stromelysins, gelatinases, and membrane-type MMPs. However, this sequence analysis revealed that human enamelysin cannot be assigned to any of these MMP subclasses. Thus, this enzyme lacks two of the three residues (Tyr-214, Asp-235, and Gly-237 in collagenase 23 numbering) that are conserved in all collagenases characterized to date and that have been proposed as essential determinants of collagenase specificity (*8*, *32*, *33*). The equivalent residues in human enamelysin at these positions are Phe-218, Asp-239, and Ser-241 (Figures 1 and 2A). The human enamelysin sequence also shows an insertion of eight amino acids in the COOH-terminal region of its catalytic domain, similar in size to the nine-residue insertion present in the same region in the stromelysins, but absent in collagenases. The introduction of these nine residues in the equivalent region of collagenases results in complete loss of the collagenolytic activity of the chimeric enzymes (*34*) (Figure 2B). Taken together, these structural comparisons make unlikely the possibility that human enamelysin is a collagenase. We next examined the possibility that human enamelysin could belong to the stromelysin subfamily of MMPs. However, these enzymes contain close to the Zn-binding site three conserved residues (Thr-210, Asn-231, and Glu-233 in stromelysin 1 numbering) which are distinct from the equivalent residues in human enamelysin (Phe-218, Asp-239, and Ser-241) (Figure 2B). Similarly, and despite the observed similarities in the length of the hinge region of both human enamelysin and stromelysins, a more detailed analysis of the respective amino acid sequences of these regions revealed marked differences between them. Thus, the enamelysin insertion at this region is rich in basic residues and is therefore positively charged. In contrast, the nine-residue insertion of stromelysins is markedly hydrophobic, whereas that present in MMP-19 is extremely acidic. On this basis, and considering previous studies on the influence of these hinge insertions in MMP substrate specificity (*34*), it is tempting to speculate that the basic insertion identified in the hinge region of human enamelysin may provide unique catalytic properties to this protein. Finally, the structural analysis of the amino acid sequence of human enamelysin revealed that it lacks the fibronectin-like domain present in all gelatinases, the COOH-terminal extension rich in hydrophobic residues characteristic of MT-MMPs, and the furin activation consensus sequence (R-X-R/X-R) mediating the intracellular activation of MT-MMPs and stromelysin 3 (*35*−*40*). It is also remarkable that human enamelysin does not contain any potential site of N-glycosylation in its amino acid sequence, including the one absolutely conserved in the catalytic domain of collagenases, macrophage metalloelastases, stromelysins 1 and 2, gelatinases B, and MT1-, -2-, and 3-MMPs, and whose effective glycosylation has been demonstrated in some of them. In summary, and according to all these structural comparisons, human enamelysin cannot be classified in any of the previously described subfamilies of MMPs, which could be consistent with a putative role of this enzyme in the degradation of specific substrates during the process of enamel formation.

Functional Characterization of Recombinant Human Enamelysin Produced in E. coli. As a preliminary step to perform the biochemical characterization of human enamelysin and to examine its functional relevance, we undertook the production of a recombinant form of this enzyme in *E. coli*. To do that, a 1.5 kb fragment containing the entire open reading frame for MMP-20 (proMMP-20) was first subcloned into the expression vector pRSETB. The resulting plasmid was then transformed into *E. coli* BL21(DE3)pLysS, and the transformed bacteria were induced to produce the recombinant protein by treatment with IPTG. Finally, extracts were prepared from the induced bacteria and analyzed by SDS−PAGE (data not shown). According to the data obtained, the insoluble fraction of the bacteria transformed with the

```
   1  CTACTGTGAGGGGATGAAGGTGCTCCCTGCATCTGGCCTTGCTGTCTTCCTCATCATGGCTTTGAAGTTTTCCAC   75
   1                M  K  V  L  P  A  S  G  L  A  V  F  L  I  M  A  L  K  F  S  T    25

  76  TGCAGCCCCCTCCCTAGTTGCAGCCTCCCCCAGGACCTGGAGGAACAACTACCGCCTCGCACAGGCGTATCTTGA   150
  26   A  A  P  S  L  V  A  A  S  P  R  T  W  R  N  N  Y  R  L  A  Q  A  Y  L  D    50

 151  CAAATATTACACAAATAAAGAAGGACACCAGATTGGTGAGATGGTTGCAAGAGGAAGCAATTCCATGATAAGGAA   225
  51   K  Y  Y  T  N  K  E  G  H  Q  I  G  E  M  V  A  R  G  S  N  S  M  I  R  K    75

 226  GATTAAGGAGCTACAAGCGTTCTTTGGCCTCCAAGTCACCGGGAAGTTAGACCAGACCACAATGAACGTGATCAA   300
  76   I  K  E  L  Q  A  F  F  G  L  Q  V  T  G  K  L  D  Q  T  T  M  N  V  I  K   100

 301  GAAGCCTCGCTGTGGAGTTCCTGATGTGGCCAATTATCGCCTCTTCCCTGGTGAACCCAAATGGAAAAAAAATAC   375
 101   K  P  R  C  G  V  P  D  V  A  N  Y  R  L  F  P  G  E  P  K  W  K  K  N  T   125

 376  TTTGACATACAGAATATCTAAATACACACCTTCCATGAGTTCTGTCGAGGTGGACAAAGCAGTGGAGATGGCCTT   450
 126   L  T  Y  R  I  S  K  Y  T  P  S  M  S  S  V  E  V  D  K  A  V  E  M  A  L   150

 451  GCAGGCCTGGAGTAGCGCCGTCCCTCTGAGCTTTGTCAGAATAAACTCAGGAGAAGCGGATATTATGATATCTTT   525
 151   Q  A  W  S  S  A  V  P  L  S  F  V  R  I  N  S  G  E  A  D  I  M  I  S  F   175

 526  TGAAAATGGAGATCACGGGGATTCCTATCCATTCGATGGGCCTCGGGGGACTCTAGCCCATGCATTTGCTCCTGG   600
 176   E  N  G  D  H  G  D  S  Y  P  F  D  G  P  R  G  T  L  A  H  A  F  A  P  G   200

 601  AGAAGGCCTGGGAGGAGATACACATTTCGACAATCCTGAGAAGTGGACTATGGGAACGAATGGTTTTAATTTGTT   675
 201   E  G  L  G  G  D  T  H  F  D  N  P  E  K  W  T  M  G  T  N  G  F  N  L  F   225

 676  TACCGTTGCTGCTCATGAATTTGGCCATGCCCTGGGCCTGGCCCATTCCACAGACCCATCAGCACTGATGTACCC   750
 226   T  V  A  A  H  E  F  G  H  A  L  G  L  A  H  S  T  D  P  S  A  L  M  Y  P   250

 751  AACTTATAAGTACAAGAATCCCTATGGATTCCACCTCCCCAAAGATGATGTGAAAGGGATCCAGGCATTATACGG   825
 251   T  Y  K  Y  K  N  P  Y  G  F  H  L  P  K  D  D  V  K  G  I  Q  A  L  Y  G   275

 826  ACCTCGGAAAGTATTCCTGGGGAAGCCCACTCTGCCCCATGCCCCCCATCACAAGCCATCCATCCCTGACCTCTG   900
 276   P  R  K  V  F  L  G  K  P  T  L  P  H  A  P  H  H  K  P  S  I  P  D  L  C   300

 901  TGACTCCAGCTCATCCTTTGACGCTGTGACAATGCTGGGGAAGGAGCTCCTGCTCTTCAAGGACCGGATTTTCTG   975
 301   D  S  S  S  S  F  D  A  V  T  M  L  G  K  E  L  L  L  F  K  D  R  I  F  W   325

 976  GAGACGGCAGGTTCACTTGCGGACAGGAATTCGGCCCAGCACTATTACCAGCTCCTTCCCCCAGCTCATGTCCAA  1050
 326   R  R  Q  V  H  L  R  T  G  I  R  P  S  T  I  T  S  S  F  P  Q  L  M  S  N   350

1051  TGTGGATGCAGCTTACGAAGTGGCTGAGAGGGGCACTGCTTACTTCTTCAAAGGTCCCCACTACTGGATAACAAG  1125
 351   V  D  A  A  Y  E  V  A  E  R  G  T  A  Y  F  F  K  G  P  H  Y  W  I  T  R   375

1126  AGGATTCCAAATGCAAGGTCCTCCTCGGACTATTTATGACTTTGGATTTCCAAGGCACGTGCAGCAAATAGATGC  1200
 376   G  F  Q  M  Q  G  P  P  R  T  I  Y  D  F  G  F  P  R  H  V  Q  Q  I  D  A   400

1201  TGCTGTCTACCTCAGGGAGCCACAGAAGACCCTTTTCTTTGTGGGAGATGAATACTACAGCTACGACGAAAGGAA  1275
 401   A  V  Y  L  R  E  P  Q  K  T  L  F  F  V  G  D  E  Y  Y  S  Y  D  E  R  K   425

1276  AAGGAAAAATGGAAAAAGACTATCCAAAGAATACTGAAGAAGAATTTTCAGGAGTAAATGGCCAAATCGATGCTGC  1350
 426   R  K  M  E  K  D  Y  P  K  N  T  E  E  E  F  S  G  V  N  G  Q  I  D  A  A   450

1351  TGTAGAATTAAATGGCTACATTTACTTCTTTTCAGGACCAAAAACATACAAGTATGACACAGAGAAGGAAGATGT  1425
 451   V  E  L  N  G  Y  I  Y  F  F  S  G  P  K  T  Y  K  Y  D  T  E  K  E  D  V   475

1426  GGTTAGTGTGGTGAAATCTAGTTCCTGGATTGGTTGCTAAATAGAAAAGCCTAGTCTTCTCAAGCAATGAGGATG  1500
 476   V  S  V  V  K  S  S  S  W  I  G  C  *                                       500

1501  ACTACAAGCAGCCTCTAACTGGATCTTAAGGACTAAAGCAGAATGTAGGAGAGGGATTCTTCCAAAGGCCTTCAA  1575

1576  ATCAAATTAGAATTCACTGAGAATAATAATACTTCCAATTTTTTTCATAGTTGTATAATCAGAATTTCAATCCAC  1650

1651  ATTAGAAAAGTTTTTATATGGGCA  1674
```

FIGURE 1: Nucleotide sequence and deduced amino acid sequence of human enamelysin. Nucleotide and amino acid residues are numbered.

recombinant plasmid contained a protein of the expected size which was not present in the control extracts. Following solubilization and purification under denaturing conditions, we refolded the recombinant enamelysin as described in

```
A MMP-1      209 EYNLHRVAAHELGHSLGLSHSTDIGA 234
  MMP-8      208 NYNLFLVAAHEFGHSLGLAHSSDPGA 233
  MMP-13     213 GYNLFLVAAFEFGHSLGLDHSKDPGA 238
  MMP-3      209 GTNLFLVAAHEIGHSLGLFHSANTEA 234
  MMP-10     208 GTNLFLVAAHELGHSLGLFHSANTEA 233
  MMP-19     203 GVNLRIIAAHEVGHALGLGHSRYSQA 228
  MMP-20     217 GFNLFTVAAHEFGHALGLAHSTDPSA 242

B MMP-1      260 YGRSQNPVQ----------------PIGPQTPKACD 279
  MMP-8      261 YGLSSNPIQ----------------PTGPSTPKPCD 280
  MMP-13     266 YGPGDEDPN----------------PKHPKTPDKCD 285
  MMP-3      209 YGPPPSDPE-------TPLVPTEPVPPEPPFTPANCD 291
  MMP-10     262 YGPPPASTE-------EPLVPTKSVPSGSEMPAKCD 290
  MMP-19     255 YGKKSPVIRDEEEEETELPTVPPVPTEPSPMPDPCS 290
  MMP-20     270 YGPRKVFLG--------KPTLPHAPHHKPSIPDLCD 297
```

FIGURE 2: Comparison of the amino acid sequence of human enamelysin around the Zn-binding site and the hinge region with those of other human MMPs. The amino acid sequences of fibroblast collagenase (MMP-1), neutrophil collagenase (MMP-8), collagenase 3 (MMP-13), stromelysin 1 (MMP-3), stromelysin 2 (MMP-10), and MMP-19 were extracted from the SwissProt data base and used for multiple alignment. (A) Amino acid sequence of the region close to the Zn-binding site containing distinctive residues between collagenases and stromelysins. Distinctive residues are in bold. (B) Amino acid sequence around the hinge region of MMPs. For comparison purposes, numbering in each protein starts in the initiator methionine. Gaps are indicated by hyphens. Basic residues present in the MMP-20 hinge are in bold.

Materials and Methods. During the refolding process, enamelysin autoproteolytically proceeded to the active enzyme form, and subsequently, all further experiments were performed with the active enzyme. Active recombinant enamelysin hydrolyzed the synthetic quenched fluorescent substrates McaPLGLDpaARNH$_2$ and McaPLANvaDpaARNH$_2$ with distinct k_{cat}/K_m values of 4.55×10^4 and 5.88×10^4 M^{-1} s^{-1}. This activity was inhibited by human TIMP-2, thereby confirming that human enamelysin is an authentic member of the MMP family. We furthermore assessed whether recombinant human enamelysin was able to cleave its putative natural substrate amelogenin. Amelogenin (200 ng/μL) was incubated with 3.9 nM human enamelysin for 18 h at 37 °C, and SDS−PAGE analysis revealed that four major fragments were generated showing molecular masses of 24 000, 23 000, 22 000, and 20 000 Da, respectively (Figure 3, lane 1). The extent of amelogenin degradation was low, likely because the high dilution of the active recombinant enamelysin recovered after the refolding process. Nevertheless, the cleavage of amelogenin by enamelysin was completely inhibited by TIMP-2 (Figure 3, lane 2), providing additional evidence of the specificity of this proteolytic reaction mediated by human enamelysin.

The finding that human enamelysin is able to degrade amelogenin together with its restricted pattern of expression in human tissues has prompted us to speculate that this MMP may play a highly specialized role during enamel biomineralization. The formation of dental enamel is a dynamic process involving the removal of most of the protein components of the extracellular enamel matrix and their replacement by mineral ions, calcium and phosphorus, finally leading to a fully mineralized mature tissue (*10*−*12*). The characterization of the proteolytic pathways involved in this massive degradation of proteins occurring during enamel development and mineralization has been hampered by the extremely low amounts of proteases present in the maturing enamel as well as by their marked instability (*13*−*18*). In

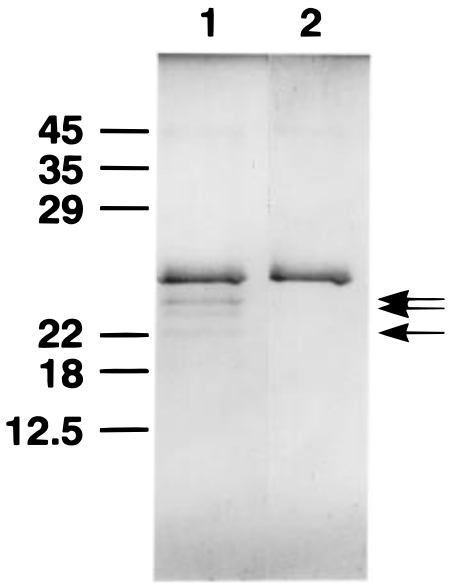

FIGURE 3: Enzymatic activity of recombinant human enamelysin. The recombinant protein produced in *E. coli* was refolded and its proteolytic activity assayed against purified murine amelogenin (200 ng/mL) in the absence (lane 1) or in the presence (lane 2) of TIMP-2.

fact, to our knowledge, this work represents the first study in which a human dental protease has been identified and characterized at the molecular level. The precise functional role of human enamelysin during tooth formation is yet unclear. One likely possibility is that this enzyme may function within the fluid phase of enamel to modify selectively the molecular masses of some proteins, such as amelogenins, thus resulting in a profound alteration of their biochemical properties and/or their molecular interactions with other components in the developing enamel. In addition, human enamelysin may also play a less specific role as a digestive enzyme involved in the removal of intact or processed matrix proteins so that initially formed crystals can expand during the maturation stage. Further studies with other non-amelogenin substrates, including some of the recently characterized enamelins such as tuftelin and ameloblastin (*41*, *42*), will be required to define the biochemical and enzymatic properties of this human enzyme as well as its precise role in the unique process of formation of dental enamel.

Physical Mapping of the Human Enamelysin Gene. According to the genomic organization and structural data obtained for human enamelysin, this protein presents the overall domain distribution characteristic of other MMPs, such as collagenases and stromelysins, which are encoded by genes clustered within the long arm of chromosome 11 (*43*, *44*). However, human enamelysin shows also some structural differences with respect to these MMPs, such as the absence in its catalytic domain of the distinctive residues of collagenases and stromelysins. In addition, the tissue expression pattern and substrate specificity of enamelysin appear to be absolutely unique among all human MMPs characterized to date. One hypothesis for explaining this structural and functional divergence may be that the gene encoding human enamelysin belongs to a different evolutionary MMP subfamily located out of the MMP cluster identified at the 11q22 chromosomal region of the human genome. A similar explanation has been suggested for the case of the recently described MMP-19, which exhibits

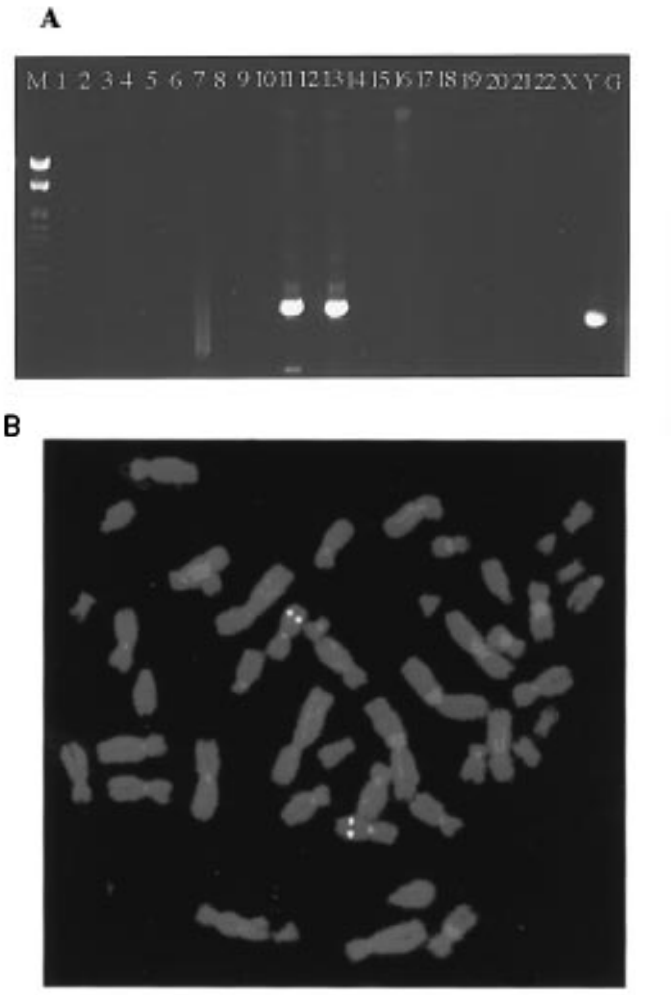

FIGURE 4: Chromosomal mapping of the human enamelysin gene. (A) One hundred nanograms of total DNA from the 24 monochromosomal somatic cell lines was PCR-amplified with primers 5′-TGAGGTCAACGTCCAAGCCAC-3′ and 5′-CCCATCTCCT-TCTCCTAACTGG-3′ as described in Materials and Methods. pBR322 digested with *Hae*III (Boehringer Mannheim) was used as a size marker (M). Amplification of human genomic DNA (G) was used as a control. (B**)** FISH hybridization with a centromeric probe of chromosome 11 (red) and with a probe specific for human enamelysin (yellow). Metaphase cells were counterstained in blue with DAPI.

unique structural and enzymatic properties and maps to chromosome 12q14 (*7*). To clarify this question, we tried to establish the chromosomal location of the enamelysin gene in the human genome. These studies could also be of interest in evaluating the possible linkage of this gene to some of the hereditary forms of amelogenesis imperfecta, a human disease characterized by extensive alterations in the process of enamel formation (*45*). As a previous step for these studies, we designed a PCR-based strategy directed to screen a panel of monochromosomal cell hybrids containing a single human chromosome in a rodent background. The sequence-tagged site (STS) specific for the human enamelysin gene was generated by using two primers whose sequence was derived from data obtained from noncoding regions of the gene. As shown in Figure 4A, a DNA fragment was amplified from hybrids containing human chromosomes 11 and 13. However, the somatic cell hybrids containing human chromosome 13 also carry fragments from chromosome 11 (*46*), strongly suggesting that the human enamelysin gene mapped to this latter chromosome. To confirm this localization as well as to provide a more precise localization of the human enamelysin gene within chromosome 11, we carried out FISH experiments on metaphase spreads using the biotinylated PAC clone containing the human enamelysin gene (yellow color), and a digoxygenated centromeric probe specific for chromosome 11 as an anchor marker (red color). A total of 50 metaphase spreads were examined, and the results of a representative experiment are shown in Figure 4B. In agreement with the human−rodent somatic hybrid studies, yellow fluorescent signals were located on chromosome 11 and no other chromosome site was labeled above background (Figure 4B). After DAPI-banding of metaphases showing hybridization in both chromosomes 11, the fluorescent signal was assigned to the q22.3-23 region of this chromosome. According to these results, we conclude that the human enamelysin gene (MMP-20) is clustered to the seven members of the MMP family (MMP-1, -3, -7, -8, -10, -12, and -13) previously mapped at the long arm of chromosome 11 (*43*, *44*). Furthermore, these results suggest that MMP-20 is not a candidate gene for any of the forms of the amelogenesis imperfecta syndrome whose loci have been identified at chromosomes X and 4 (*47*, *48*). The availability of MMP-20-derived reagents, including the specific probes generated in this work, will be very helpful to future works directed at evaluating the precise role of this novel human enzyme in the process of enamel formation in both normal and pathological conditions.

ACKNOWLEDGMENT

We are very grateful to Dr. P. Morán for advice in FISH studies, Drs. J. C. de Vicente, M. Balbín, and G. Velasco for helpful comments, Drs. F. Vizoso and I. Díez-Itza for providing tissue samples, and S. Alvarez for excellent technical assistance. Monochromosomal somatic cell hybrid DNAs were kindly provided by the Human Genome Mapping Resource Center (Cambridgeshire, U.K.).

REFERENCES

1. Woessner, J. F. (1991) *FASEB J. 5*, 2145−2154.
2. Matrisian, L. M. (1992) *BioEssays 14*, 455−463.
3. Stetler-Stevenson, W. G., Aznavoorian, S., and Liotta, L. A. (1993) *Annu. Rev. Cell Biol. 9*, 541−573.
4. Birkedal-Hansen, H., Moore W. G. I., Bodden, M. K., Windsor, L. J., Birkedal-Hansen, B., DeCarlo, A., and Engler, J. A. (1993) *Crit. Rev. Oral Biol. Med. 4*, 197−250.
5. Basset, P., Bellocq, J. P., Wolf, C., Stoll, I., Hutin, P., Limacher, J. M., Podhajcer, O. L., Chenard, M. P., Rio, M. C., and Chambon, P. (1990) *Nature 348*, 699−704.
6. Belaaouaj, A., Shipley, J. M., Kobayashi, D. K., Zimonjic, D. B., Popescu, N., Silverman, G. A., and Shapiro, S. D. (1995) *J. Biol. Chem. 270*, 14568−14575.
7. Pendás, A. M., Knäuper, V., Puente, X. S., Llano, E., Mattei, M. G., Apte, S., Murphy, G., and López-Otín, C. (1997) *J. Biol. Chem. 272*, 4281−4286.
8. Freije, J. P., Díez-Itza, I., Balbín, M., Sánchez, L. M., Blasco, R., Tolivia, J., and López-Otín, C. (1994) *J. Biol. Chem. 269*, 16766−16773.
9. Puente, X. S., Pendás, A. M., Llano, E., Velasco G., and López-Otín, C. (1996) *Cancer Res. 56*, 944−949.
10. Simmer, J. P., and Fincham, A. G. (1995) *Crit. Rev. Oral Biol. Med. 6*, 84−108.
11. Deutsch, D., Catalano-Sherman, J., Dafni, S., David, S., and Palmon, A. (1995) *Connect. Tissue Res. 32*, 97−107.
12. Smith, C. E., and Nanci, A. (1996) *Anat. Rec. 245*, 186−207.
13. Moe, D., and Birkedal-Hansen, H. (1979) *J. Dent. Res. 58*, 1012−1013.
14. Overall, C. M., and Limeback, H. (1988) *Biochem. J. 256*, 965−972.

15. Carter, J., Smillie, A. C., and Shepherd, M. G. (1989) *Arch. Oral Biol. 34*, 195−202.
16. DenBesten, P. K., and Heffernan, L. M. (1989) *Arch. Oral Biol. 34*, 399−404.
17. Tanabe, T., Fukae, M., Uchida, T., and Shimizu, M. (1992) *Calcif. Tissue Int. 51*, 213−217.
18. Tanabe, T., Fukae, M., and Shimizu, M. (1994) *Arch. Oral Biol. 39*, 277−281.
19. Moradian-Oldak, J., Leung, W., Simmer, J. P., Zeichner-David, M., and Fincham, A. G. (1996) *Biochem. J. 318*, 1015−1021.
20. Bartlett, J. D., Simmer, J. P., Xue, J., Margolis, H. C., and Moreno, E. C. (1996) *Gene 183*, 123−128.
21. Messing, J., and Vieira, J. (1982) *Gene 19*, 269−276.
22. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463−5467.
23. Devereux, J., Haeberli, P., and Smithies, O. (1984) *Nucleic Acids Res. 12*, 387−395.
24. Pendás, A. M., Matilla, T., Uría, J. A., Freije, J. P., Fueyo, A., Estivill, X., and López-Otín, C. (1994) *Cytogenet. Cell Genet. 66*, 263−266.
25. Pendás, A. M., Morán, P., Freije, J. P., and García-Vázquez, E. (1994) *Cytogenet. Cell Genet. 67*, 31−36.
26. Willembrock, F., Crabbe, T., Slocombe, P. M., Sutton, C. W., Docherty, A. J. P., Cockett, M. I., O'Shea, M. I., Brocklehurst, K., Philips, I. R., and Murphy, G. (1993) *Biochemistry 32*, 4330−4337.
27. Hu, C. C., Ryu, O. H., Qian, Q, Zhang, C. H., and Simmer, J. P. (1997) *J. Dent. Res. 76*, 641−647.
28. Okada, Y., Nagase, H., and Harris, E. D., Jr. (1986) *J. Biol. Chem. 261*, 14245−14255.
29. Bode, W., Gomis-Rüth, F. X., and Stöcker, W. (1993) *FEBS Lett. 331*, 134−140.
30. Gaire, M., Magbanua, Z., McDonell, S., McNeill, L., Lovett, D. H., and Matrisian, L. M. (1994) *J. Biol. Chem. 269*, 2032−2040.
31. Pak, J. H., Liu, C. Y., Huangpu, J., and Graham, J. S. (1996) *FEBS Lett. 404*, 283−288.
32. Sánchez-López, R., Alexander, C. M., Behrendtsen, O., Breathnach, R., and Werb, Z. (1993) *J. Biol. Chem. 268*, 7238−7247.
33. Stolow, M. A., Bauzon, D. D., Li, J., Sedgwick, T., Liang, V. C. T., Sang, Q. A., and Shi, Y. B. (1996) *Mol. Biol. Cell 7*, 1471−1483.
34. Hirose, T., Patterson, C., Pourmotabbed, T., Mainardi, C. L., and Hasty, K. A. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 2569−2573.
35. Sato, H., Takino, T., Okada, Y., Cao, J., Shinagawa, A., Yamamoto, E., and Seiki, M. (1994) *Nature 370*, 61−65.
36. Strongin, A. Y., Collier, Y., Bannikov, G., Marmer, B. L., Grant, G. A., and Goldberg, G. I. (1995) *J. Biol. Chem. 270*, 5331−5338.
37. Will, H., and Hinzmann, B. (1995) *Eur. J. Biochem. 231*, 602−608.
38. Takino, T., Sato, H., Shinagawa, A., and Seiki, M. (1995) *J. Biol. Chem. 270*, 23013−23020.
39. Pei, D., and Weiss, S. J. (1995) *Nature 375*, 244−247.
40. Santavicca, M., Noel, A., Angliker, H., Stoll, Y., Segain, J. P., Anglard, P., Chretien, M., Seidah, N., and Basset, P. (1996) *Biochem. J. 315*, 953−958.
41. Deutsch, D., Palmon, A., Fisher, L. W., Kolodny, N., Termine, J. D., and Young, M. F. (1991) *J. Biol. Chem. 266*, 16021−16028.
42. Krebsbach, P. H., Lee, S. K., Matsuki, Y., Kozak, C. A., Yamada, K. M., and Yamada, Y. (1996) *J. Biol. Chem. 271*, 4431−4435.
43. Pendás, A. M., Matilla, T., Estivill, X., and López-Otín, C. (1995) *Genomics 26*, 615−618.
44. Pendás, A. M., Santamaría, I., Alvarez, M. V., Pritchard, M., and López-Otín, C. (1996) *Genomics 37*, 266−269.
45. Witkop, C. J., Jr. (1989) *J. Oral Pathol. 17*, 547−553.
46. Zhong, S., Nolf, C. R., and Spurr, N. K. (1992) *Hum. Genet. 90*, 435−439.
47. Lagerström, M., Dahl, N., Nakahori, Y., Nakagome, Y., Bäckman, B., Landegren, U., and Pettersson, U. (1991) *Genomics 10*, 971−997.
48. Kärrman, C., Bäckman, B., Dixon, M., Holmgren, G., and Forsman, K. (1997) *Genomics 39*, 164−170.

BI972120Y