MMP-20 Is Predominately a Tooth-Specific Enzyme with a Deep Catalytic Pocket that Hydrolyzes Type V Collagen†

Benjamin E. Turk,‡ Daniel H. Lee,§ Yasuo Yamakoshi,|| Andreas Klingenhoff,⊥ Ernst Reichenberger,@ J. Timothy Wright,# James P. Simmer,|| Justin A. Komisarof,+ Lewis C. Cantley,◇ and John D. Bartlett*,§,●

Department of Pharmacology, Yale University, New Haven, Connecticut 06520, Department of Cytokine Biology, Forsyth Institute, Boston, Massachusetts 02115, University of Michigan Dental Research Laboratory, University of Michigan, Ann Arbor, Michigan 48108, Genomatix Software GmbH, Bayerstrasse 85a, 80335 Müchen, Germany, Center for Restorative Medicine and Skeletal Development, Department for Oral Rehabilitation, Biomaterials and Skeletal Development, University of Connecticut School of Dental Medicine, Farmington, Connecticut 06030, Department of Pediatric Dentistry, University of North Carolina, Chapel Hill, North Carolina 27599, Research Science Institute, Center for Excellence in Education, Vienna, Virginia 22124, Department of Medicine, Harvard Medical School, Beth Israel Deaconess Medical Center, Boston, Massachusetts 02115, and Department of Developmental Biology, Harvard School of Dental Medicine, Boston, Massachusetts 02115

Received November 3, 2005; Revised Manuscript Received February 7, 2006

ABSTRACT: Matrix metalloproteinase-20 (MMP-20, enamelysin) has a highly restricted pattern of expression. In healthy tissues, MMP-20 is observed in the enamel organ and pulp organ of developing teeth and is present only as an activated enzyme. To identify other tissues that may express MMP-20, we performed a systematic mouse tissue expression screen. Among the non-tooth tissues assayed, MMP-20 transcripts were identified only in minute quantities within the large intestine. The murine *Mmp20* promoter was cloned, sequenced, and assessed for potential tooth-specific regulatory elements. In silico analysis identified four promoter modules that were common to *Mmp20* and at least two of three coregulated predominantly tooth-specific genes that encode ameloblastin, amelogenin, and enamelin. We asked if the highly restricted MMP-20 expression pattern was associated with a broad substrate specificity that might preclude its expression in other tissues. An iterative mixture-based random doedecamer peptide library screen with Edman sequencing of MMP-20 cleavage products revealed that, among MMPs previously screened, MMP-20 had unique substrate preferences. These preferences indicate that MMP-20 has a deep and wide catalytic pocket that can accommodate substrates with large aromatic residues in the P1′ position. On the basis of matrices derived from the peptide library data, we identified and then confirmed that type V collagen is an MMP-20 substrate. Since type V collagen is not present in dental enamel but is an otherwise widely distributed collagen, and since only active MMP-20 has been observed in teeth, our data suggest that control of MMP-20 activity is primarily regulated by transcriptional means.

Dental enamel is unique because it is the only epithelially derived calcified tissue in vertebrates and it is the hardest substance in the body. Five predominantly enamel-specific gene products are secreted by ameloblasts of the enamel organ into the developing enamel matrix. These gene products are amelogenin, ameloblastin, enamelin, matrix metalloproteinase-20 (MMP-20),[1] and Kallikrein 4 (KLK-4). Of these five proteins, four are secreted during early enamel development as the enamel grows to full thickness and one, KLK-4, is secreted later in development as the thick enamel layer hardens (*1*). Both *AMBN* and *ENAML* localize to human chromosome 4 (*2*, *3*), whereas *MMP20* localizes to human chromosome 11q22 within the MMP gene cluster (*4*). The amelogenin gene is present only on the X chromosome in mice (*Amelx*), but the Y chromosome also carries an amelogenin gene (*AMELY*) in humans. In males, 90% of amelogenin transcripts are expressed from *AMELX* with only 10% expression from *AMELY* (*5*). *AMELY* does not compensate for loss of *AMELX*, and deletion of *AMELY* has no

† This work was supported by National Institute of Dental and Craniofacial Research Grants DE14084 (to J.D.B.) and DE 13237 (subproject 4 to J.D.B.), by NIH Grant GM56203 (to L.C.C.), and by a Special Fellow award from the Leukemia and Lymphoma Society (to B.E.T.) and was conducted in a laboratory renovated with NIH Grant CO6RR11244.

* To whom correspondence should be addressed. Telephone: (617) 262-5200, ext. 8388. Fax: (617) 892-8303. E-mail: jbartlett@forsyth.org.

‡ Yale University.

§ Forsyth Institute.

|| University of Michigan.

⊥ Genomatix Software GmbH.

@ University of Connecticut School of Dental Medicine.

University of North Carolina.

+ Research Science Institute.

◇ Harvard Medical School.

● Harvard School of Dental Medicine.

[1] Abbreviations: Abs, absorbance; MMP, matrix metalloproteinase; KLK-4, kallikrein-4; rh, recombinant human; RP-HPLC, reversed-phase high-pressure liquid chromatography; TFA, trifluoroacetic acid; TFBS, transcription factor binding site(s); TSS, transcription start site; HOMF, homeodomain transcription factors; BRNF, Brn POU domain factors; TBPF, TATA-binding protein factor; HOXF, factors with moderate activity for the homeodomain consensus sequence; HOXC, HOX-PBX complexes.

10.1021/bi052252o CCC: $33.50
Published on Web 03/04/2006

effect on teeth (*6*). Knockout mouse studies and/or human mutations have demonstrated that a loss of function in any one of the five enamel matrix proteins will cause diseased dental enamel termed *amelogenesis imperfecta* (*7*−*13*). Here we focus on MMP-20 by characterizing its tissue distribution, its promoter regulatory elements shared with coregulated genes, and its substrate specificity.

Matrix metalloproteinase-20 (MMP-20) was initially named enamelysin because it was isolated by PCR-based homology cloning from a porcine enamel organ-specific cDNA library, and Northern blot tissue analysis did not detect MMP-20 expression in other tissues (*14*). Later, ameloblasts of the enamel organ and odontoblasts of the dental papilla were demonstrated to express MMP-20 (*15*−*19*). Even so, MMP-20 expression proved to be highly restricted, especially when compared to that of other members of the MMP family. In one study, 51 different cell lines were assessed by RT-PCR for MMP-20 expression, and none tested positive (*20*). MMP-20 expression has been observed in pathologic tissues, such as in ghost cells of calcifying odontogenic cysts (*21*), and odontogenic tumors (*22*, *23*). MMP-20 expression was also observed in bradykinin-treated granulosa cells isolated from the follicles of porcine ovaries (*24*). Nevertheless, among healthy tissues, MMP-20 is predominantly expressed within enamel organ and the dental papilla. The phenotype in mice devoid of MMP-20 activity (*7*) and humans with homozygous *MMP20* mutation (*11*) supports the data demonstrating tooth-specific expression. MMP-20 null mice have thin brittle enamel that tends to delaminate from the underlying dentin, and individuals with the human mutation have soft, pigmented enamel that is similarly fragile. No other phenotypes were observed. Therefore, to date, MMP-20 is considered a tooth-specific MMP.

Here we take a defined approach to assessing the distribution of MMP-20 expression in the mouse. Although we did find that the large intestine expressed very low levels of MMP-20 that could not be detected by Northern blotting procedures, we consider MMP-20 to be an essentially tooth-specific matrix metalloproteinase. This conclusion led us to ask two more questions. (1) Can we identify the promoter elements directing MMP-20 tooth-specific expression? (2) Does MMP-20 possess a broad substrate specificity that might preclude its expression in other healthy tissues?

To definitively answer the first question, a cell line was required that expresses all the factors necessary for MMP-20 transcription. Although two cell lines were reported to express MMP-20 (*11*), we found these cells expressed so little MMP-20 that it precluded their use for promoter analyses. However, we did clone, sequence, and perform in depth in silico analysis of the murine MMP-20 promoter. The MMP-20 results were then compared with those of three other coregulated genes (*Amelx*, *Ambn*, and *Enam*) to identify transcription factor binding sites (TFBS) and/or promoter modules that likely play a role in tooth-specific gene expression.

To answer the second question, we sought to characterize the active site-mediated substrate specificity of MMP-20. An iterative mixture-based peptide library method was used (*25*). The approach involves treating peptide mixtures with MMP-20 and analyzing the product pool by amino-terminal Edman sequencing, which provides an indication of the relative selectivity for each amino acid at any of several positions relative to the cleavage site. The same method was previously applied to seven other MMPs, which allows direct comparison of the cleavage selectivity of MMP-20 with related proteases (*25*, *26*). The MMP-20 peptide library results were then used to generate a weight matrix that scores each amino acid at each position relative to the scissile bond (P4−P4′). This allowed us to search the protein sequence database with the goal of identifying new MMP-20 substrates. On the basis of the results from the peptide library screen, a new MMP-20 substrate was identified (type V collagen) that is widely distributed throughout the body and is present in bone and dentin (*27*, *28*) but is not present in enamel (*27*).

MATERIALS AND METHODS

Reverse Transcriptase-Polymerase Chain Reaction. Total RNA was extracted from day-old mouse tissues with Trizol reagent (Invitrogen), and the Access RT-PCR system (Promega) was used to reverse transcribe 1 μg of total RNA from each tissue and to amplify the resulting cDNAs. For MMP-20, mouse 5′ PCR primers annealed to exon 9 and 3′ primers annealed to exon 10. PCR primers were as follows: mMMP20 ex9 (sense), 5′-GGC CAT ATA GAT GCT GCT GTG GA-3′; MMP20 ex10 (antisense), 5′-GAG GCC AGT AGG AGA CAA AGA GGA C-3′; F-ms-β-actin (sense), 5′-TCG TGC GTG ACA TCA AGG AGA AGC-3′; and R-ms-β-actin (anitsense), 5′-CAG CAC TGT GTT GGC ATA GAG GTC-3′.

Southern Blot Analysis of RT-PCR Results. Thirty microliters of PCR product for each tissue analyzed by RT-PCR was loaded into an agarose gel, electrophoresed, transblotted onto a nylon membrane, and fixed to the membrane by incubation at 80 °C for 1 h. The Genius system (Boehringer Mannheim) was used to perform the Southern blotting procedures. In brief, primers nested to the RT-PCR primers were used in a PCR to label 339 bp of MMP-20 cDNA (20 ng/mL) with Digoxigenin-11-dUTP (DIG). The primers were as follows: F-South-MMP20 (sense), 5′-CTG GAT TGG CTG CTG AGT CGT-3′; and R-South-MMP20 (antisense), 5′-TGA GTG CAC ATA GTC CTT TT TCT-3′. The probe was incubated overnight at 50 °C in hybridization buffer consisting of 5× SSC, 0.1% *N*-lauroylsarcosine, 0.2% SDS, and 1.0% Genius system blocking reagent. Identification of cDNA bands resulted after incubation with anti-DIG-alkaline phosphatase followed by color development with NBT and X-phosphate solution as stated in the manufacturer's instructions.

qPCR Analysis of MMP-20 Expression in Mouse Tissues. A SuperScript III First-Strand Synthesis System for RT-PCR (Invitrogen) generated the cDNA for real-time PCR analysis. Primer sequences for MMP-20 expression analyses were as follows: fMMP20-rltm (sense), 5′-CCT CTG GCC TTG CTG TCC TT-3′; rMMP20-rltm (antisense), 5′-CCT GGG GGC CTC CTT TCT-3′. Internal control primers for mouse β-actin were as follows: fβ-actin-rltm (sense), 5′-TCA AGA TCA TTG CTC CTC CTG AGC-3′; and rβ-actin-rltm (antisense), 5′-TAC TCC TGC TTG CTG ATC CAC ATC-3′. The PCR mix (SYBR Green Supermix, Bio-Rad) was composed of 100 mM KCl, 40 mM Tris-HCl (pH 8.4), dNTPs (0.4 mM each), 50 units/mL iTaq DNA polymerase, 6 mM $MgCl_2$, SYBR Green I, 20 nM fluorescein, and stabilizers. The PCR temperature profile was 3 min 95 °C

for the initial melt, then 20 s at 95 °C and 30 s at 65 °C for 45 cycles, then 30 s at 95 °C, and then 1 min at 55 °C followed by stepwise temperature increases from 55 to 95 °C to generate the melt curve. Standard curves were generated with the primer set by use of a 10-fold dilution series ranging from 1000 ng/mL to 100 pg/mL. PCR efficiencies and relative expression levels of MMP-20 as a function of β-actin expression were calculated as previously described (*29*).

Cloning and in Silico Analysis of the Mmp20 5′-Flanking Region. The murine 129-strain genomic library cloned in the Lambda Fix II vector (Stratagene) was screened with a 441 bp MMP-20 cDNA probe to obtain genomic clones of MMP-20. The probe was DIG labeled (Boehringer Mannheim) as described above, and plaque lifting, prehybridization, hybridization, washings of the filters, and identification of plaques with the insert were performed according to the manufacturer's instructions. Putative positive clones were purified by secondary and tertiary screening rounds. Genomic DNA inserts were subcloned into pZero-2 plasmid vectors (Invitrogen) for identification of insert size and sequencing.

The amelogenin, ameloblastin, enamelin, and MMP-20 promoters were assessed for common TFBS by Genomatix software applications in the region spanning 500 bp upstream of the first or only transcription start site. As explained in detail elsewhere (*30*), TFBS can be described by position weight matrices. The matrices are mined from the literature, and the literature is cross-referenced to eliminate putatively erroneous TFBS from the matrices. This weight matrix pattern definition is superior to a fixed IUPAC consensus sequence because it represents the complete nucleotide distribution for each single position that defines the TFBS. With the nucleotide weight matrix, potential TFBS can also be quantified for their similarity to experimentally verified TFBS. The promoters were analyzed with these matrices by calculating individual conservation scores (C_i value) for each nucleotide of a potential TFBS. The calculation is as follows

$$C_i(i) = \frac{100}{\ln 5}\left[\sum_{\text{b}\in\text{A,C,G,T,gap}} P(i,\text{b}) \times \ln p(i,\text{b}) + \ln 5\right]$$

where $P(i,\text{b})$ is the relative frequency of nucleotide b at position i. This information is used by MatInspector to scan for nucleotide sequences that match patterns from a library. A matrix similarity score is then calculated which reaches 1 only if the test sequence corresponds to the most conserved nucleotide at each position of the matrix. The matrix similarity is calculated as follows

$$\text{mat\_sim} = \frac{\sum_{j=1}^{n} C_i(j) \times \text{score}(\text{b},j)}{\sum_{j=1}^{n} C_i(j) \times \text{max\_score}(j)}$$

where $C_i(j)$ is the C_i value of position j, n is the length of the matrix, score(b,j) is the matrix value for base b at position j, and max_score is the maximum score within a matrix column at position j (for details, see ref *31*).

Matrices for transcription factors (TFs) that bind to similar sites, as identified from different publications or research groups, are assigned to a family. Only the highest-scoring, most conserved match to one of these matrices is used for further analysis. The grouping of matrices into families condenses the output because redundant matches among related TFs are eliminated. Once the individual TFBS in the promoters were found, FrameWorker identified modules present in at least three of the four coregulated promoters. The modules consisted of at least two TFBS in common order, orientation, and distance from each other.

Primer Extension Analysis. Primer extension analysis to identify utilized transcription start sites was performed according to the method of Maniatis (*32*). Briefly, the antisense primer (5′-CAG TAG GTT GGG GTC TGC AGT GGC AAA CTT T) hybridized to a region located 47 bp downstream of the translation start site. The primer was labeled with [^{32}P]ATP, and 10^5 cpm was hybridized with 10 μg of total RNA isolated from mouse unerupted first molars. Reverse transcription was performed followed by degradation of single-stranded RNA with RNase A. cDNA fragments were separated on a 6% polyacrylamide gel next to a sequencing reaction generated by use of the same primer.

Expression and Purification of Recombinant Human MMP-20 (rhMMP-20). A 1466 bp cDNA fragment encompassing the entire hMMP-20 coding region was generated with PstI and was ligated in frame into the pRSET B *Escherichia coli* expression vector (Invitrogen) as previously described (*4*). rhMMP-20 was expressed in *E. coli* as an N-terminal His_6 fusion after induction with 0.1 mM 1-thio-β-D-galactopyranoside followed by overnight incubation at 30 °C as described previously (*4*). rhMMP-20 was solubilized from inclusion bodies and purified on a Ni column. The eluted protein was 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 as described previously (*4*), and the supernatant was concentrated 10-fold by use of a Microcon YM-10 filter (Amicon, Inc.). Enzymatically inactive control MMP-20 was prepared by point mutation of the glutamic acid at position 235 (within the HEXXH active site consensus) to an alanine (E235A).

Peptide Library Screen for MMP-20 Substrate Specificity. Analysis of MMP-20 cleavage specificity using peptide libraries was performed essentially as described previously (*25*) and is summarized below. To examine positions downstream of the cleavage site, we used an N-terminally acetylated random peptide dodecamer, Ac-XXXXXX-XXXXXX, in which each position X is an equimolar mixture of the 19 naturally occurring amino acids excluding cysteine. The random peptide (1 mM) was incubated with rhMMP-20 (1.5 μg/mL) for 4 h at 37 °C in MMP reaction buffer [50 mM HEPES (pH 7.4), 200 mM NaCl, and 10 mM $CaCl_2$]. A 10 μL aliquot of this digest was loaded directly onto an Applied Biosystems Procise 494 automated Edman sequencer. Because the N-termini of the substrate peptides are blocked, only the cleaved portion of the mixture is detected on the sequencer. Under these conditions, approximately 7.5% of the peptides are cleaved, meaning that 750 pmol of unblocked peptide is sequenced. On average, this corresponds to 40 pmol of each amino acid residue and is thus, as a pool, within the detection limits of standard Edman sequencing. Note that each individual peptide is present at quantities insufficient for detection; the method relies on analysis of the bulk properties of the substrate pool. The selectivity value

for each residue was calculated by dividing the molar amount of the residue in a given sequencing cycle by the amount present in the starting mixture and then normalizing the data to an average value of 1. Preferences at the upstream positions were obtained using the peptide mixtures MAXXXXX-LRGAARE(K-biotin) and MGXXPXXLRGGGEE(K-biotin). Peptides were digested with rhMMP-20 as described above, and the reactions were quenched by adding *o*-phenanthroline to a concentration of 20 mM. Digests (20 μL) were treated with 600 μL of avidin agarose in 600 μL of 50 mM NH_4HCO_3 with agitation for 1 h at room temperature to remove unreacted peptides and C-terminal fragments. The slurry was transferred to a column, and the unbound fraction was collected and pooled with 1.2 mL of 50 mM NH_4HCO_3 wash. This fraction, containing the N-terminal fragments of the cleaved peptides, was evaporated to dryness under reduced pressure with centrifugation, suspended in 200 μL of ddH_2O, and dried again. The residue was suspended in 30 μL of ddH_2O and applied to the Edman sequencer. Data were analyzed as described above. Identical reactions were run using the catalytically inactive E235A mutant of rhMMP-20, which did not detectably cleave any of the peptide libraries.

Collagen Digestion by MMP-20. Type III, IV, and V collagen (250 μg/μL) from human placenta (Sigma) were each incubated with human recombinant MMP-20 (15 μg/μL) in a buffer containing 100 mM Tris-HCl (pH 7.6), 200 mM NaCl, and 0.1% Brij-35 in a total volume of 200 μL. The reaction proceeded for 16 h at 37 °C and was stopped by the addition of EDTA at a final concentration of 20 mM. For SDS−PAGE analysis, sample loading buffer was added at a final concentration of 63 mM Tris-HCl, 2% SDS, 150 μL/mL glycerol, 0.05% (w/v) bromophenol blue, and 50 μL/mL dithiothreitol. Samples (20 μL) were run on 10% SDS−PAGE gels, and protein bands were stained with Coomassie Brilliant Blue.

RP-HPLC Fractionation of Type V Collagen. Human Type V collagen (Sigma) was fractionated by reversed-phase high-performance liquid chromatography (RP-HPLC) using the Beckman System Gold HPLC system equipped with a Discovery C18 column (4.6 mm inside diameter × 25 cm, Supelco, Bellefonte, PA). The column was equilibrated with 0.05% trifluoroacetic acid (buffer A) and eluted with a linear gradient of buffer B (from 0 to 100% B, over 50 min) containing 0.05% trifluoroacetic acid and 80% acetonitrile at a flow rate of 1.0 mL/min at room temperature. Type V collagen eluted in the second, third, and fourth peaks.

SDS−PAGE. SDS−PAGE was performed using Novex 4 to 20% Tris-glycine gels (Invitrogen). Samples were dissolved in Laemmli sample buffer (Bio-Rad), and electrophoresis was carried out with a current of 20 mA for 1.5 h. The gels were stained with Bio-Safe Coomassie (Bio-Rad). The apparent molecular masses of protein bands were estimated by comparison with SeeBlue Plus2 prestained standards (Invitrogen).

Western Immunoblots. The proteins on electrophoresed gels were electrotransferred onto Hybond-P membrane (Amersham Biosciences). Polyclonal antibody for human type V collagen was used for 30 min at a dilution of 1:40000. The secondary antibody was HRP-conjugated anti-rabbit IgG (Bio-Rad). The membranes were immunostained by chemiluminescent detection with an ECL Advance Western blotting detection kit (Amersham Biosciences).

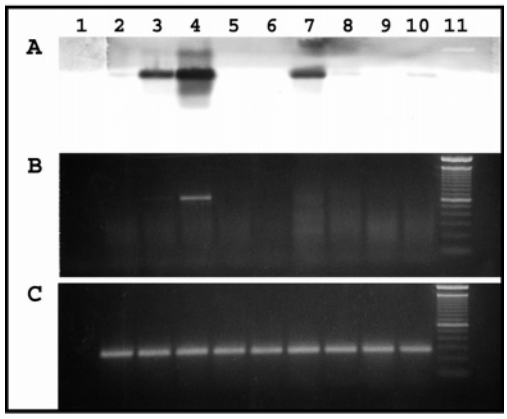

FIGURE 1: Identification of newborn mouse tissues expressing MMP-20. Tissues were dissected from newborn mice and processed for extraction of total RNA. (A) Southern blot, using a ^{32}P-labeled nested probe to confirm the tissue-specific MMP-20 RT-PCR results. (B) RT-PCR for identification of tissue-specific MMP-20 expression. This gel was processed for the Southern blot of panel A. (C) Positive RT-PCR control demonstrating that each tissue expresses β-actin: lane 1, negative control (no cDNA); lane 2, complete mouse pup with jaws removed (no teeth); lane 3, complete mouse pup; lane 4, jaw (with teeth); lane 5, stomach; lane 6, eye; lane 7, intestine; lane 8, bladder; lane 9, spleen; lane 10, thymus; and lane 11, 100 bp DNA ladder. Note that 30 cycles of PCR yielded an MMP-20 band only in the jaw containing the tooth buds (lane 4), while Southern blotting of the PCR results demonstrated MMP-20 expression in RNA extracted from the complete mouse pup (lane 3) and intestine (lane 7).

RESULTS

Analysis of Newborn Mouse Tissues for MMP-20 Expression. Day-old mice were sacrificed and processed for extraction of total RNA. RNA from bladder, eye, large intestine, spleen, stomach, thymus, and whole mandibles containing developing tooth buds was assessed for MMP-20 expression by RT-PCR procedures. In addition, complete day-old mice that either did or did not have their mandible and maxillas removed were processed for RT-PCR analysis of MMP-20 expression. After 30 PCR cycles, agarose gel electrophoresis revealed that only the jaw containing the developing tooth buds had appreciable levels of MMP-20 transcripts (Figure 1, middle panel). Nested primers were designed to amplify and clone the PCR product into the pPCR-Script AMP SK(+) vector for use in Southern blot procedures. Southern blotting of the PCR products revealed additional bands. The whole mouse, but not the whole mouse without jaws, had a distinct and well-defined band the same size as that present in the jaw-only lane (Figure 1, top panel). Interestingly, the presence of a band in the large intestine lane indicated that MMP-20 was also expressed in that tissue. The procedures were repeated with different mice at a later date with similar results. In an attempt to verify MMP-20 expression in the large intestine, 30 μg of total RNA extracted from the large intestine was electrophoresed for Northern blot procedures. However, although transcripts were clearly present in the lane containing RNA from the jaw, we were unable to detect MMP-20 expression by this approach in the large intestine (not shown).

To confirm that MMP-20 was expressed in the large intestine and to quantify any differences in the level of expression between tooth buds and large intestine, we performed quantitative real-time PCR (qPCR) according to

Table 1: Quantitative Real-Time PCR Analysis of MMP-20 Expression Levels in Mouse Tissues[a]

tissue assayed	cycle threshold	relative level of expression
large intestine	34.67	1.00
3-day-old EO	22.99	4.71×10^3
small intestine	nd	−
brain	nd	−
heart	nd	−
kidney	nd	−
liver	nd	−
lung	nd	−
pancreas	nd	−
spleen	nd	−
stomach	nd	−

[a] Cycle threshold represents the number of PCR cycles necessary for the start of logarithmic cDNA expansion. The relative level of expression was calculated as a function of the expression level observed in the large intestine.

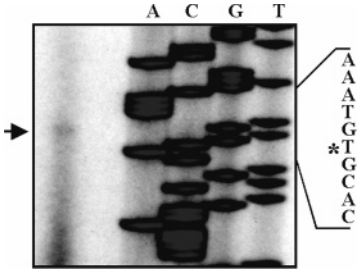

FIGURE 2: Location of the transcription start site in the murine MMP-20 gene. The transcription start was mapped by primer extension and in silico analysis. Each lane of the sequencing reaction was labeled with A, C, G, or T to identify the nucleotide sequenced within the genomic clone. The arrow designates the primer extension product. The asterisk designates the nucleotide position of the transcription start site that was identified by both primer extension and in silico analysis.

the method of Pfaffl (*29*) using β-actin as the internal reference gene. The tissues assessed for MMP-20 expression were 4-day-old first molar tooth buds in addition to adult brain, heart, kidney, liver, lung, pancreas, spleen, stomach, large intestine, and small intestine. Of the tissues that were assayed, only the tooth bud and the large intestine were positive for MMP-20 transcripts. However, the MMP-20 expression level observed in the large intestine was approximately 5000 times lower than that observed in the 4-day-old first molar tooth buds (Table 1). These results demonstrate that MMP-20 expression is primarily confined to tooth tissues and that MMP-20 is essentially a tooth-specific matrix metalloproteinase. To identify possible regulatory elements that direct tooth-specific expression, we cloned and characterized the murine MMP-20 promoter.

Cloning of the Mouse MMP-20 Promoter. A 441 bp probe spanning from 376 bp 5′ to the translation start site and extending into exon 1 was labeled with DIG and used to screen a mouse (129/SvJ) genomic DNA lambda fixII phage library. After three successive screening rounds, six clones were isolated, and phage was purified and assessed for the presence of probe sequence by PCR analysis. Inserts from four of the six clones were excised with NotI and subcloned into pZero-2. The insert orientation and length estimation of the 5′ noncoding region were assessed by PCR analysis. The 3′ primer originally used to make the probe for phage screening was paired with a primer specific for each vector arm. Agarose gel eletrophoresis of the PCR mixtures revealed insert orientation and demonstrated that each 5′ noncoding region was between approximately 8 and 10 kb in length. The clone with the longest 5′ noncoding region was sequenced. The entire genomic clone was 16 972 bp in length and was 99−100% identical to portions of the *Mus musculus* BAC clone RP23-338J1. The *Mmp20* ATG translation initiation codon was located at position 10 007 from our 5′ clone end (GenBank accession number DQ190458).

Identification of the Mmp20 Transcription Initiation Site. Primer extension was performed using a ^{32}P-labeled primer (5′-CAG TAG GTT GGG GTC TGC AGT GGC AAA CTT T) that hybridized 47 bp 3′ of the ATG translation initiation codon. The labeled primer annealed to the genomic clone to generate the sequencing ladder and was also used on total RNA isolated from mouse enamel organ (Figure 2) to identify the transcription start site (TSS). To independently confirm the enamelysin TSS, the entire 16 972 bp genomic fragment was assessed for consensus start sites by in silico analysis. As shown in Figure 2, good agreement exists between the experimental data and the computer-generated data. Thus, the entire 16 972 bp genomic fragment contained 9956 bp of enamelysin promoter upstream of the TSS, a 173 bp first exon which includes 50 bp of 3′ untranslated region and the translation start, the first intron of 6762 bp, and the remaining sequence represented the first 80 bp of the second exon. Next, we asked if the enamelysin promoter contains consensus transcription factor binding sites (TSBS) that direct its tooth-specific expression.

In Silico Analysis of the MMP-20 Promoter. No known cell line that expresses even moderate levels of MMP-20 exists, so traditional promoter analyses were not possible. However, as a necessary first step, we performed in silico analysis to identify candidate TFBS that may direct predominantly tooth-specific gene expression. Enamelysin, amelogenin, ameloblastin, and enamelin are primarily expressed in the tooth, and their promoters are coregulated (*15*, *16*, *19*, *29*, *33*). The murine promoter sequences were analyzed for TFBS common to at least three of the four promoters. Twenty-nine common elements were identified, 11 of which were common to all four promoters. However, on the basis of the pattern of binding sites that was detected, the sequences displayed no overall similarity. That is, the binding sites among the promoters were not conserved in number, orientation (location on the plus or minus strand), and/or relative distance from the TSS. Overall, no conserved patterns of individual TFBS were detectable among the four promoters. However, sequence subsets containing conserved modules were present. A module contains at least two TFBS in conserved order and at a conserved distance from each other and may enhance or antagonize specific activation or repression of a gene (*30*). Interestingly, although *Amelx* localizes to the X chromosome and *Ambn* is on mouse chromosome 5, they appeared to share several common patterns in their enhancer sequences. Both enhancers were enriched in binding sites for transcription factors related to developmental processes such as GATA, HOXF, and CLOX. The MMP-20 enhancer by contrast contained very few of these binding sites. Pairwise sequence analysis using FrameWorker demonstrated that the amelogenin, ameloblastin, and enamelin promoters had more enhancer modules in common with each other when compared to the MMP-20 promoter

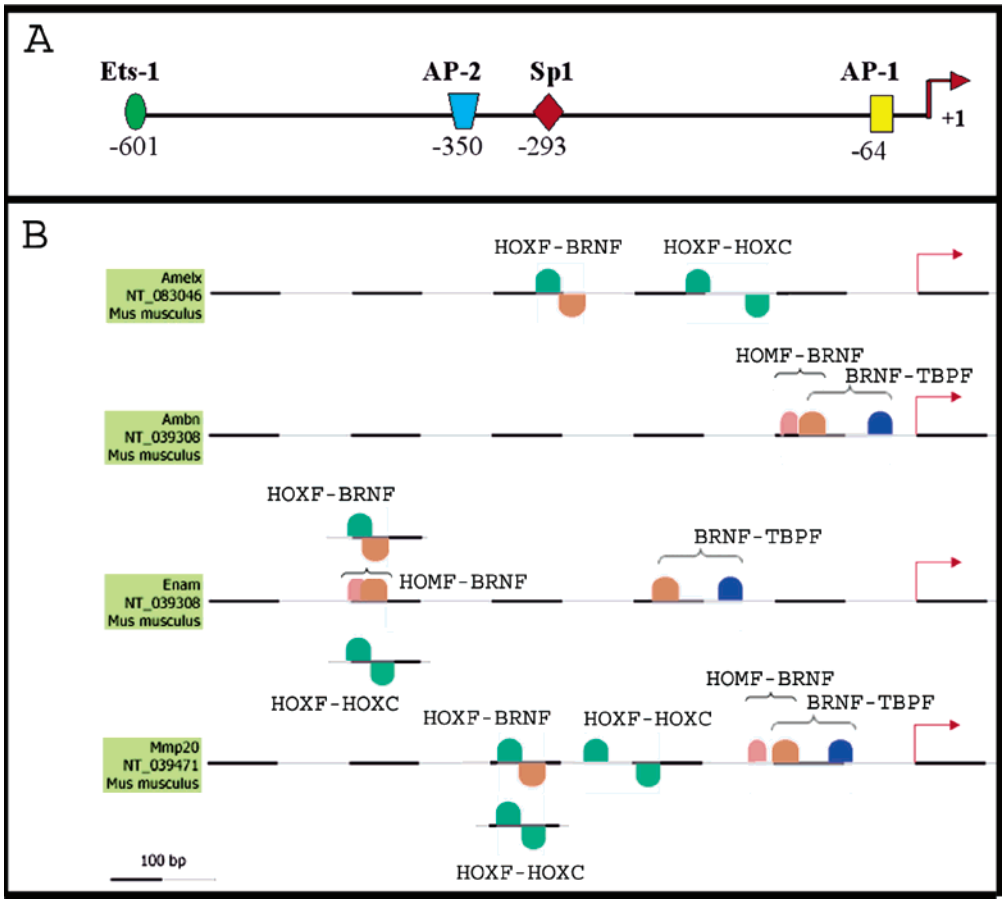

FIGURE 3: In silico analysis of the MMP-20 promoter and the coregulated amelogenin, ameloblastin, and enamelin promoters. (A) *Mmp-20* promoter schematic displaying the identity and distance from the transcription start site of four transcription factor binding sites (TFBS) commonly observed in the MMP family of enzymes. (B) Promoter analysis using FrameWorker to identify shared promoter modules in coregulated tooth-specific genes. Modules contain two or more TFBS and were identified on the basis of their probability of binding individual transcription factors and on the basis of each module's relative lack of abundance within the genome. Nine modules each with two TFBS were identified among the coregulated genes, and four of these were deemed likely to direct tooth-specific expression. Three of the four promoters contained each of the four identified modules (see Table 2). GenBank accession numbers are listed beneath the name of each promoter.

(not shown). These data lend support to the suggestion that *AMBN*, *AMELX*, and *ENAM* have arisen from an ancestral enamel matrix protein gene and that *AMELX* was subsequently translocated (*34*). Nine promoter modules were identified that were common to MMP-20 and to at least two of the three remaining tooth-specific genes. Four of these modules were highly conserved with respect to their distance relative to the TSS and with respect to the distance between the two TFBS present in the module. The modules that were identified were HOMF/BRNF, BRNF/TBPF, HOXF/BRNF, and HOXF/HOXC (Figure 3 and Table 2). The ameloblastin and MMP-20 BRNF binding site in the HOMF/BRNF module was the same BRNF site found in the BRNF/TBPF module (Figure 3, horizontal brackets). The two modules were unrelated in the enamelin promoter. The HOXF/BRNF module present in the amelogenin, enamelin, and MMP-20 promoters was independent of the two other modules containing a BRNF binding site. Interestingly, different forms of a HOXF/HOXC module were conserved either between amelogenin and MMP-20 or between enamelin and MMP-20 (Figure 3). Confirmation that each of these modules plays a role in tooth-specific gene expression awaits the development of a suitable cell line containing the appropriate transcription factors for natural expression of ameloblastin, amelogenin, enamelin, and MMP-20. However, the conserved orientation and distance between TFBS in each module and the relatively conserved distance of each module from their respective transcription start site are unique promoter characteristics that suggest involvement in tooth-specific gene expression.

Substrate Specificity of MMP-20. Mixture-based oriented peptide libraries (*25*) were used to assess MMP-20 cleavage site specificity. We started with a random dodecamer library, which was used to obtain the substrate preferences for the positions downstream of the cleavage site. The relative amounts of amino acid found at the four positions immediately C-terminal to the cleavage site are shown in Figure 4. As is typical for MMPs, MMP-20 is highly selective for hydrophobic residues at the P1′ position. The preferred amino acid at P1′ was leucine, but MMP-20 also strongly selected methionine and tyrosine to a comparable extent. While the so-called deep pocket MMPs such as MMP-3 can accommodate aromatic residues such as tyrosine at the P1′ position, aliphatic residues are consistently selected over aromatic ones (*25*, *35*). Thus, MMP-20 is unique in the extent to which it selects aromatic residues in the P1′ position. MMP-20 was relatively nonselective at the P2′ position and had a slight preference for smaller residues at P3′, which has been observed with some but not all MMPs (*25*, *26*, *35*−*37*).

To determine the cleavage specificity at sites N-terminal to the cleavage site, we used two different peptide mixtures

Table 2: MatInspector Promoter Analysis of Modules Present in Three of the Four Coregulated Tooth-Specific Genes[a]

	Ambn	*Amelx*	*Enam*	*Mmp20*
module 1	HOMF/BRNF		HOMF/BRNF	HOMF/BRNF
TFBS	TAAT**t**/**g**CACAt**tttt**		TAAT**t**/**at**agtTAAT	TAAT**t**/**ac**a**tg**aACTA***t*
strand specificity	+/+		+/+	+/+
bp separating TFBS	16		12	20
bp from TSS	64		375	83
module 2	BRNF/TPBF		BRNF/TPBF	BRNF/TPBF
TFBS	**g**CACAt**tttt**/**tt**TAAA		**at**agaTAAT/**a**TAAA	**ac**a**tg**aACTA**t**/**a**TAAA**a**
strand specificity	+/+		+/+	+/+
bp separating TFBS	48		46	39
bp from TSS	18		124	45
module 3		HOXF/BRNF	HOXF/BRNF	HOXF/HOXF/BRNF
TFBS		TAAA/**t**atATTT	TAAT/**at**agtTAAT	TAAA/**c**TAAT/**ttt**ATTT
strand specificity		+/−	+/−	+/+/−
bp separating TFBS		17	11	42, 16
bp from TSS		236	375	262
module 4		HOXF/HOXC	HOXF/HOXC	HOXF/HOXF/HOXC and HOXF/HOXC
TFBS		TTAAT**c**/**a**TGAT**tttt**	TAAT/TGATat**at**	TAAA/**c**TAAT/**t**TGAT**ttat** and TAAT/**gt**GATT**gac**
strand specificity		+/−	+/−	+/+/−, +/−
bp separating TFBS		42	17	44, 37, 18
bp from TSS		106	370	261, 181

[a] Modules denote the family of transcription factors with members that bind to the core binding site. TFBS nucleotides in capital letters denote the core sequence. Bold lowercase nucleotides denote base pairs that are observed in the TFBS with a frequency of >60%.

Table 3: Comparison of MMP-20 Peptide Library Cleavage Site Selectivity with Cleavage Sites Previously Demonstrated To Occur in Amelogenin[a]

amino acid position	P3	P2	P1	P1′	P2′	P3′
	A (2.4)	H (2.5)	S (3.2)	L (4.1)	R (1.6)	A (2.1)
	P (2.0)	Y (1.8)	Q (1.8)	M (3.1)	T (1.6)	S (2.0)
	V (1.4)	Q (1.5)	N (1.7)	Y (3.3)	V (1.4)	T (1.7)
	S (1.4)		I (2.1)			
Demonstrated Enamelysin Cleavage Sites for Porcine Amelogenin						
148	M	F	**S**	**M**	Q	S
147	**P**	M	F	S	M	Q
136	I	**Q**	P	**L**	L	P
107	**P**	L	P	A	Q	Q
105	N	L	P	**L**	P	A
63	S	H	A	**L**	Q	P
45	G	G	W	**L**	H	H

[a] In the top portion, data were generated by incubation of enamelysin with mixture-based oriented peptide libraries (see Figure 4) followed by amino acid sequencing. Numbers in the top row are preceded by a P to indicate the position of an amino acid relative to the scissile bond between P1 and P1′. Numbers in parentheses represent the selectivity values as defined in Figure 4. In the bottom portion, previously generated data identifying MMP-20 cleavage sites in porcine amelogenin of 173 amino acids in length are presented (*39*). Numbers in the first column identify the new amelogenin C-terminus after cleavage by MMP-20. Bold amino acids were selected by the mixture-based oriented library as promoting hydrolysis of the scissile bond.

of reduced complexity. Initially, the MAXXXXXLRGAARE-(K-biotin) library was used, which includes several fixed residues intended to direct MMP cleavage to the X−L bond. As demonstrated for other MMPs, data from this library indicated a preference for proline at the P3 position but also revealed an unusually strong preference for alanine (Figure 4). Specificity at the remaining upstream positions was determined using the MGXXPXXLRGGGEE(K-biotin) library in a similar manner. Fixing a proline residue at P3 directed a greater proportion of the cleavage to the intended site and thus increased the magnitude of the signal on the Edman sequencer relative to background for cycles corresponding to the less selective upstream positions. In contrast to most MMPs that prefer hydrophobic residues (*25*, *35*, *36*, *38*), MMP-20 displayed an unusual selection for histidine at the P2 position.

To confirm that the peptide library results accurately reflect the true MMP-20 substrate specificity, we compared these results to those for previously identified MMP-20 cleavage sites present in the most abundant enamel matrix protein, amelogenin. Previously, recombinant porcine MMP-20 was incubated with recombinant porcine amelogenin, and the resulting cleavage products were identified by mass spectrometry (*39*). Presented in Table 3 are the preferred amino acids for positions P3−P3′ that direct hydrolysis at the scissile bond as identified by the peptide library screen (top portion of Table 3). Also presented in Table 3 are the MMP-20-generated amelogenin cleavage sites that were previously identified by mass spectrometry (bottom portion). Five of the seven sites have a hydrophobic residue at P1′ (most have leucine), which supports this position as the most selective for MMP-20. The two sites that lack a larger hydrophobic residue at this position have a highly selected proline at the P3 position. This presumably compensates for the lack of a large hydrophobic residue at P1′. Though none of the cleavage sites within amelogenin closely match the consensus as defined using peptide libraries at all positions, inspection of the full amelogenin primary sequence revealed no potential sites that appeared to be more optimal than the authentic cleavage sites.

Prediction and Identification of a Novel MMP-20 Substrate. We identified several novel potential substrates by searching the SWIS-PROT database using Scansite (*40*) with matrices derived from the MMP-20 library data. Among the substrates identified were collagens III and V. MMP-20 was predicted to cleave the collagen V α1 chain precursor (GPRG-LLGPK), the collagen V α3 chain precursor (PRLG-LQGP), and the collagen III α1 chain precursor (GPQG-LQGL). Recombinant human MMP-20 was assessed for its ability to cleave type III, IV, or V collagen. Type IV collagen

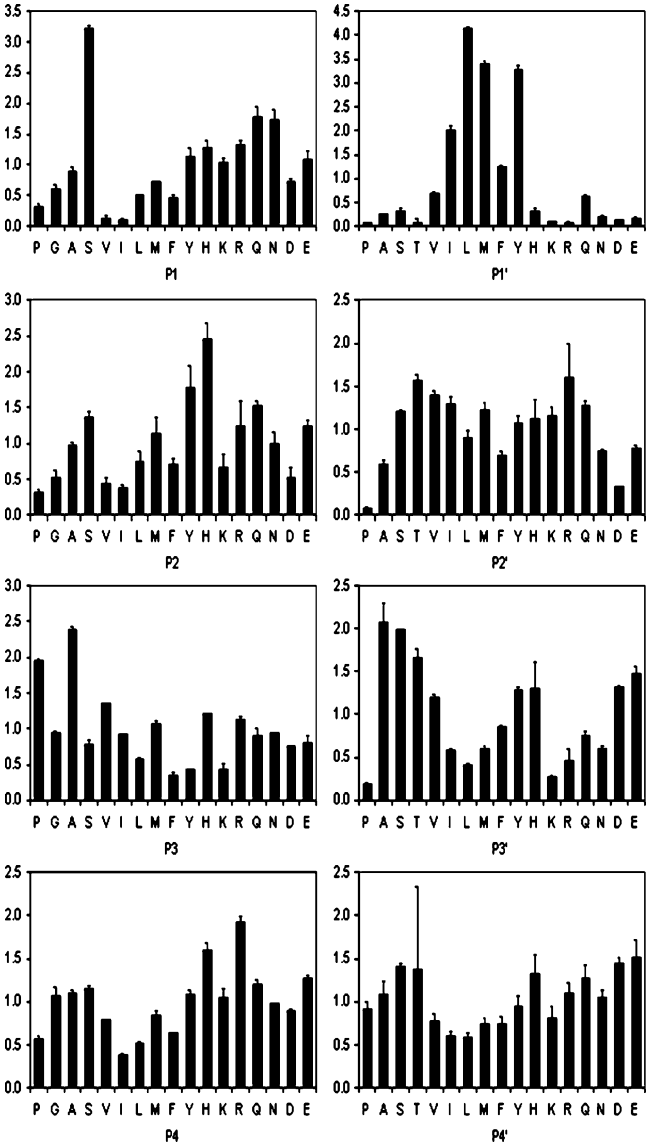

FIGURE 4: Cleavage site specificity of MMP-20. The graphs in the right column show the relative distributions of amino acid residues at positions C-terminal (P1′−P4′) to the MMP-20 cleavage site. These sites were identified by sequencing an N-terminally acetylated random peptide dodecamer (Ac-XXXXXXXXXXXX). Data were normalized so that a value of 1 corresponded to the average quantity per amino acid in a given sequencing cycle and would indicate no selectivity. Tryptophan was not included in the analysis because of poor yield during sequencing. The graphs in the left column represent the positions amino-terminal to the MMP-20 cleavage site. For the P3 position, data were generated using the MAXXXXXLRGAARE(K-biotin) library. For all other positions, the P3 proline MGXXPXXLRGGGEE(K-biotin) library was used. Glutamine and threonine were omitted in some cycles because of the high background on the sequencer. Data were normalized as for the primed sites.

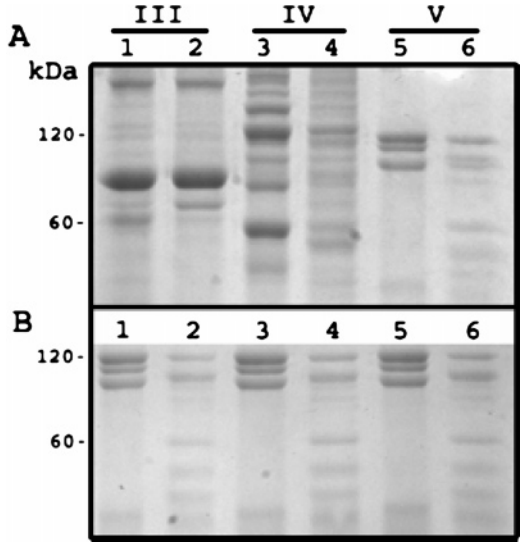

FIGURE 5: Hydrolysis of type V, but not type III, collagen by MMP-20. SDS−PAGE analysis of potential MMP-20 collagen substrates. The odd-numbered lanes did not include enzyme. In even-numbered lanes, collagens (250 μg/mL) were incubated with MMP-20 (15 μg/mL) at 37 °C for 16 h followed by Coomassie Brilliant Blue staining. (A) Type III, IV, and V collagens were electrophoresed in lanes 1 and 2, 3 and 4, and 5 and 6, respectively. Type IV collagen served as the positive control. Although type III collagen was not cleaved by MMP-20, it appeared that type V was cleaved. (B) To confirm that the type V collagen cleavage results in panel A were not a result of protein underloading or artifact, the same experiment was repeated with the same procedure on three more type V controls and corresponding enzyme-treated samples. All three sample sets confirmed the results shown in panel A.

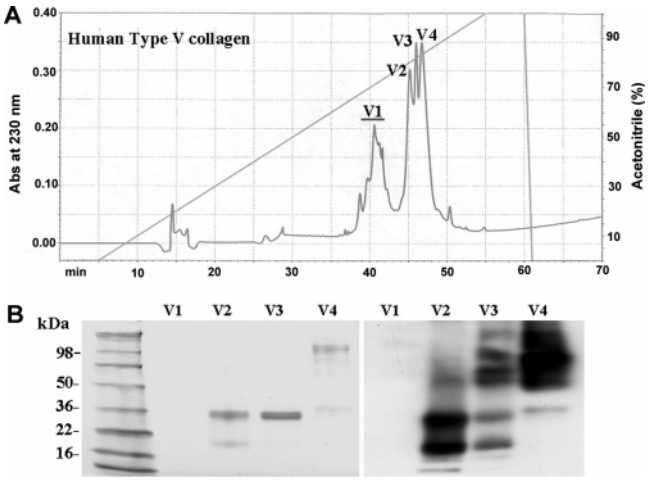

FIGURE 6: Separation and identification of type V collagen fractions. (A) Chromatogram of type V collagen fractionated using a Discovery C18 RP-HPLC column, showing four chromatographic peaks. (B) SDS−PAGE gel (4 to 20%) stained with CBB (left) and a Western blot identifying the type V collagen bands (right). Note that fraction V3 contains a mixture of bands that were also present in either V2 or V4.

was previously identified as an MMP-20 substrate (*41*), so it served as the positive control. SDS−PAGE analysis demonstrated that type V, but not type III, collagen was cleaved by MMP-20 (Figure 5, top panel). The type V collagen results were repeated with three more MMP-20 digestions to verify the original result (Figure 5, bottom panel).

To confirm that MMP-20 cleaves type V collagen and not a contaminate present within the collagen preparation, the type V collagen preparation was divided into four fractions (Figure 6, top panel) by reversed-phase high-pressure liquid chromatography (RP-HPLC). Each fraction was electrophoresed on a 4 to 20% SDS−PAGE gel where the protein either was stained or underwent Western blotting procedures with a collagen type V-specific antibody. The first protein fraction was identified as contaminate(s) because it did not stain with Coomassie Brilliant Blue and was not recognized by the type V collagen antibody (Figure 6, bottom panel). The three remaining fractions, however, did contain type V collagen. Fraction 3 appeared to be an intermediate com-

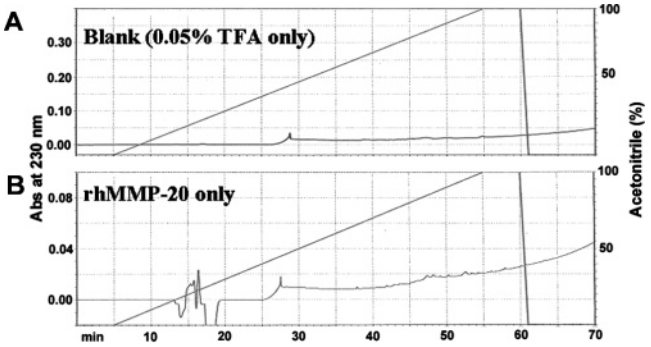

FIGURE 7: Blank and rhMMP-20 chromatograph controls. (A) Blank chromatograph containing just 0.05% trifluoroacetic acid demonstrating the absence of contaminates. (B) Chromatograph of rhMMP-20 in the absence of type V collagen. No peaks occurred in the 45−50 min range where the type V collagen fractions were previously identified.

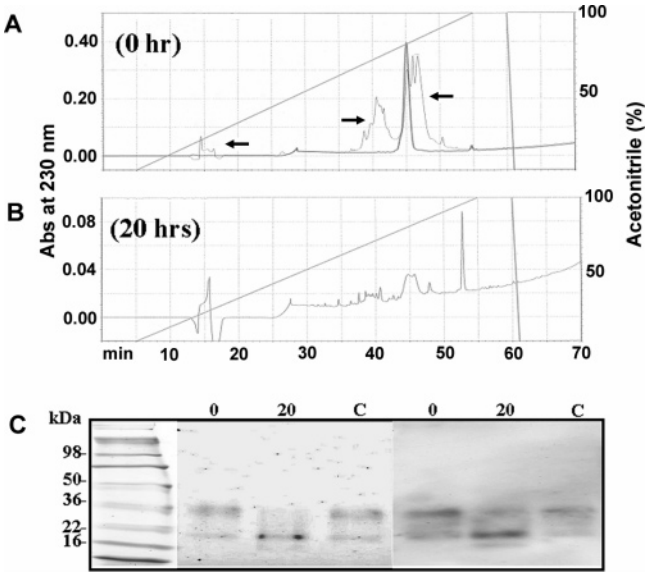

FIGURE 8: Type V collagen from fraction V2 is an MMP-20 substrate. (A) Chromatogram of purified fraction V2 demonstrating a high level of V2 purity. Note that arrows point to a background trace from the chromatogram in Figure 6, demonstrating that the peak that is shown was from the original V2 fraction. (B) Chromatogram of fraction V2 after digestion with rhMMP-20 for 20 h. Note that the peak at approximately 45 min was substantially diminished in magnitude after the digestion. (C) SDS−PAGE gel (left) and Western blot (right) demonstrating that fraction V2 type V collagen was digested after being exposed to rhMMP-20 for 20 h (lane 20) as compared to the 0 h digestion (lane 0) and type V collagen without rhMMP-20 (lane C).

prised of proteins from both fractions 2 and 4 (Figure 6, bottom panel). Therefore, only fractions 2 and 4 were selected for further analysis. Initially, a blank run and fractionation of just rhMMP-20 were performed to demonstrate the absence of contaminates and to demonstrate that MMP-20 was not a significant component of the type V collagen fractions (Figure 7). Next, fractions 2 and 4 of the original type V collagen preparation were individually incubated with rhMMP-20 and were assessed by SDS−PAGE and Western blotting for proteolysis (Figures 8 and 9). MMP-20 cleaved type V collagen from both fractions. These results demonstrate conclusively that type V collagen is an MMP-20 substrate, and they further validate the peptide library as a useful tool for identifying novel enzyme substrates (*25*).

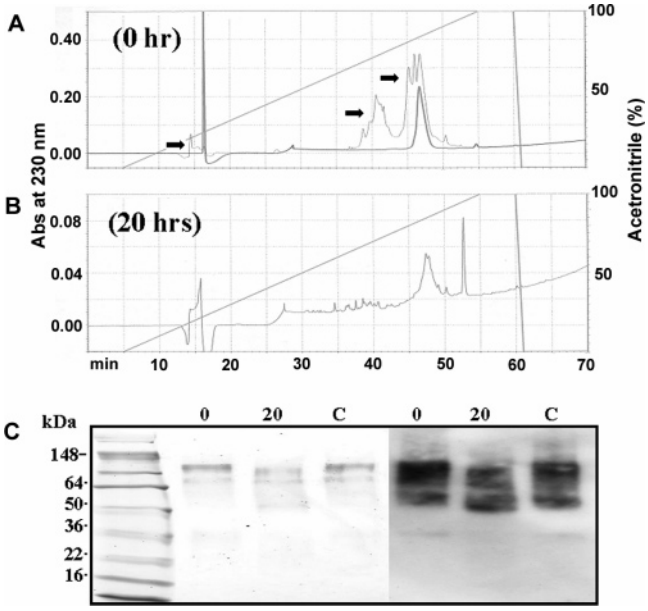

FIGURE 9: Type V collagen from fraction V4 is an MMP-20 substrate. (A) Chromatogram of purified fraction V4 demonstrating a high level of V4 purity. Note that arrows point to a background trace from the chromatogram in Figure 6 demonstrating that the peak shown was from the original V4 fraction. (B) Chromatogram of fraction V4 after digestion with rhMMP-20 for 20 h. Note that the peak at approximately 47 min was diminished in magnitude after the digestion. (C) SDS−PAGE gel (left) and Western blot (right) demonstrating that fraction V4 type V collagen was digested after being exposed to rhMMP-20 for 20 h (lane 20) as compared to the 0 h digestion (lane 0) and type V collagen without rhMMP-20 (lane C).

DISCUSSION

Since MMP-20 has a highly restricted tooth-specific pattern of expression, we sought to take a more defined approach to identifying other tissues that might express MMP-20. RT-PCR followed by Southern blotting did detect MMP-20 transcripts in the large intestine (Figure 1). However, the expression levels in the intestine were so low that they could not be detected by Northern blot analysis and were approximately 5000 times lower than the levels observed in 4-day-old tooth buds (Table 1). Recently, a dot blot MMP array was used to demonstrate that MMP-20 is expressed in human fetal and adult lung (*42*). However, these results were not confirmed by any other means, and we have previously demonstrated by Northern blot analysis of porcine lung (*14*) and in this study by qPCR of mouse lung that no detectable MMP-20 transcripts were observed. So, although the possibility that human lung expresses MMP-20 exists, we remain confident that MMP-20 is essentially a tooth-specific MMP. This conclusion led us to ask if we could identify putative promoter elements that direct MMP-20 tooth-specific expression and to ask if MMP-20 possesses a broad substrate specificity that might necessitate its limited expression in other tissues. Interestingly, the MMP-20 zymogen has not been observed on zymograms of extracted porcine enamel proteins. Zymography did reveal two bands of approximately 46 and 41 kDa that resulted from MMP-20 activity (*18*). However, when highly purified porcine MMP-20 was resolved into the 46 and 41 kDa bands by SDS−PAGE and casein zymography followed by transblotting for Edman sequencing and mass spectrometry, both bands were demonstrated to represent active MMP-20 (no

propeptide) with the correct active form N-terminus (*43*). These data demonstrated that the porcine MMP-20 zymogen, if present, is present in very small quantities within developing enamel. Furthermore, the results suggest that MMP-20 either is very easily activated or autoactivates soon after its secretion into the enamel matrix. If these activation properties were generally applicable to other tissues, then restricted expression would be the primary means of controlling MMP-20 enzymatic activity without the associated inhibition of other MMPs which would occur if tissue inhibitors of MMPs were induced. Since this suggests that MMP-20 activity is regulated primarily at the transcriptional level, we were interested in characterizing the MMP-20 promoter to identify potential tooth-specific regulatory elements.

Since amelogenin, ameloblastin, enamelin, and MMP-20 are coregulated predominantly tooth-specific gene products, we assessed their promoters for shared regulatory elements. Conserved patterns of individual TFBS were identified using a large library of quality-checked nucleotide weight matrices that are compiled by Genomatix. This approach was used previously to link diabetes-associated genes to insulin/glucose signaling pathways. Without a priori knowledge, legitimate coregulated genes were identified (*44*). Although no pattern could be identified that was conserved among all four of the tooth gene promoters, four modules containing two TFBS with conserved orientations and distances from each other were identified in three of the four sequences. The modules were HOMF/BRNF, BRNF/TBPF, HOXF/BRNF, and HOXF/HOXC.

The HOMF family is composed of homeodomain transcription factors, including *Dlx* and *Msx*, both of which are important factors necessary for tooth development. Deletion of *Dlx1* and *Dlx2* from the mouse genome results in the loss of maxillary molars (*45*), whereas the expression of *Dlx5* and *Dlx6* compensates for this loss and allows molar development in the mandibles of these animals **(**reviewed in ref *46*). A *Dlx3* mutation was demonstrated to cause defective dental enamel (*amelogenesis imperfecta*) and enlarged pulp chambers (taurodontism) in humans (*47*). *Dlx* and *Msx* will heterodimerize (*48*), and both *Msx1* and *Msx2* are also expressed in developing teeth. Mutations in the human and mouse *MSX1* gene cause orofacial clefting and tooth agenesis, while *Msx2*-deficient mice have enamel organs that fail to develop properly (reviewed in ref *49*). Notably, *Msx2* was demonstrated to play a role in the spatiotemporal expression of the amelogenin gene (*50*). Therefore, members of the HOXF family of transcription factors are essential for proper tooth development and have even been demonstrated to regulate amelogenin gene expression.

The HOXF family is also composed of homeodomain transcription factors, including *Barx2* and *goosecoid*. Although *Barx2* expression has not yet been demonstrated in tooth tissues, it is expressed in several mesenchymal and epithelial tissues during development and was demonstrated to serve as a regulator of hair follicle remodeling (*51*). In *goosecoid* null mice the teeth developed normally, but the underlying skeletal elements that support the teeth were absent (*52*). The HOXC family includes the HOX-PBX complexes that play a role in specifying regional identity along the anteroposterior axis of developing embryos (*53*). The BRNF family contains Brn POU domain factors that are widely distributed and are predominately associated with the differentiation, regulation, and proliferation of neural stem cells and neurons (reviewed in ref *54*). The TBPF family contains the TATA-binding protein factors that are commonly found in regulated genes. Thus, the in silico analysis did identify potential regulatory modules that were shared among the amelogenin, ameloblastin, enamelin, and MMP-20 promoters, indicating that a complex network of regulatory elements may be responsible for each gene's coordinated expression during enamel development. The identified modules are ideal candidates for in vitro mutation analysis for characterization of tooth-specific expression. However, it will be necessary to identify a suitable cell line for these experiments as none is now known to exist.

The peptide library screen indicated that while MMP-20 was typical of MMPs by its primary preference for hydrophobic residues at the P1′ position, the high selectivity at this position for the aromatic tyrosine residue was unique, indicating that MMP-20 has a deep catalytic pocket (Figure 4). Crystal structures and mutagenesis of various MMPs have largely implicated a single residue, corresponding to Arg214 of human MMP-1, in controlling the depth of the S1′ pocket (*55*, *56*). Shallow pocket MMPs have a relatively large residue at this position (arginine or tyrosine) that restricts access of larger residues to the pocket, whereas the deep and intermediate pocket MMPs have leucine at this position. MMP-20 by contrast has a threonine residue at the analogous position, suggesting it may have a wider S1′ pocket more accommodating to aromatic residues. These observations support recent work indicating that a continuum of shapes and sizes exists for the S1′ pocket among various MMPs (*57*). MMP-20 secondary selections were also distinctive and included the unusual preference for histidine at the P2 position. To confirm that the peptide library results were representative of cleavages that occur in vivo, these results were compared to the known MMP-20 cleavage sites present in amelogenin. Although some features of the library-derived MMP-20 cleavage motif were represented in amelogenin cleavage sites, no individual site was a particularly close match to the consensus (Table 3). This may indicate that amelogenin cleavage sites are suboptimal and that relatively high levels of MMP-20 activity are required to degrade its physiological substrate. This notion supports the observation of the "laddering effect" of amelogenin in the enamel matrix where proteolytic cleavage gradually reduces the size of amelogenins as new uncleaved amelogenins are secreted into the matrix (*58*). Thus, SDS−PAGE of enamel matrix proteins reveals a laddering of amelogenin bands that persists until MMP-20 is no longer secreted into the matrix. Alternatively, efficient cleavage of amelogenin may require docking interactions distal from the active site that are not revealed by our peptide library studies. It is also possible that fixing the P′ positions in the library prior to identifying the P positions biases the selection of preferred amino acids for the unknown P positions. This would occur if cooperativity between subsites existed, which was demonstrated for several other MMPs (*26*, *37*, *59*, *60*). It should also be noted that the library results were generated with recombinant human MMP-20 while the amelogenin cleavage sites were previously identified by use of recombinant porcine MMP-20. Exceptions aside, the peptide library screen and the previously published amelogenin data (*39*) did predominantly

coincide, indicating that MMP-20 has a unique substrate specificity among the MMP family of enzymes.

On the basis of the peptide library results, we searched for a novel MMP-20 substrate. Both type III and type V collagens were identified as potential MMP-20 substrates. Type III collagen is essential for collagen I fibrillogenesis and for normal cardiovascular development (*61*), and type V collagen controls the initiation of collagen fibril assembly. Murine deficiency of the major type V collagen chain, pro-α1(V), causes death in early embryogenesis and is associated with a virtual absence of collagen fibrils (*62*). Therefore, the presence of type I collagen is associated with the presence of type V collagen. Dentin and bone contain both of these collagens (*27*, *28*). However, enamel contains neither (*27*). Although odontoblasts do express low levels of MMP-20 (*15*), it is unclear what role MMP-20 plays in processing type V collagen in dentin, because dentin contains other MMPs, such as MMP-2, -9, and -10 (*63*) that are also capable of cleaving type V collagen. We would argue that since MMP-20 is easily activated and has a broad substrate specificity, its restricted pattern of expression is necessitated to prevent unrestrained matrix degradation.

Taken together, the results demonstrate that MMP-20 has a highly restricted tooth-specific pattern of expression, has a large catalytic pocket with a broad substrate specificity, and is capable of cleaving a widely distributed collagen that is responsible for collagen fibril assembly, and its activity is likely regulated by transcriptional means.

ACKNOWLEDGMENT

We thank Jun Xue for her technical expertise.

REFERENCES

1. Bartlett, J. D., and Simmer, J. P. (1999) Proteinases in developing dental enamel, *Crit. Rev. Oral Biol. Med. 10*, 425−441.
2. Hu, C. C., Hart, T. C., Dupont, B. R., Chen, J. J., Sun, X., Qian, Q., Zhang, C. H., Jiang, H., Mattern, V. L., Wright, J. T., and Simmer, J. P. (2000) Cloning human enamelin cDNA, chromosomal localization, and analysis of expression during tooth development, *J. Dent. Res. 79*, 912−919.
3. MacDougall, M., DuPont, B. R., Simmons, D., Reus, B., Krebsbach, P., Karrman, C., Holmgren, G., Leach, R. J., and Forsman, K. (1997) Ameloblastin gene (AMBN) maps within the critical region for autosomal dominant amelogenesis imperfecta at chromosome 4q21, *Genomics 41*, 115−118.
4. Llano, E., Pendas, A. M., Knauper, V., Sorsa, T., Salo, T., Salido, E., Murphy, G., Simmer, J. P., Bartlett, J. D., and Lopez-Otin, C. (1997) Identification and structural and functional characterization of human enamelysin (MMP-20), *Biochemistry 36*, 15101−15108.
5. Salido, E. C., Yen, P. H., Koprivnikar, K., Yu, L. C., and Shapiro, L. J. (1992) The human enamel protein gene amelogenin is expressed from both the X and the Y chromosomes, *Am. J. Hum. Genet. 50*, 303−316.
6. Lattanzi, W., Di Giacomo, M. C., Lenato, G. M., Chimienti, G., Voglino, G., Resta, N., Pepe, G., and Guanti, G. (2005) A large interstitial deletion encompassing the amelogenin gene on the short arm of the Y chromosome, *Hum. Genet. 116*, 395−401.
7. Caterina, J. J., Skobe, Z., Shi, J., Ding, Y., Simmer, J. P., Birkedal-Hansen, H., and Bartlett, J. D. (2002) Enamelysin (matrix metalloproteinase 20)-deficient mice display an amelogenesis imperfecta phenotype, *J. Biol. Chem. 277*, 49598−49604.
8. Fukumoto, S., Kiba, T., Hall, B., Iehara, N., Nakamura, T., Longenecker, G., Krebsbach, P. H., Nanci, A., Kulkarni, A. B., and Yamada, Y. (2004) Ameloblastin is a cell adhesion molecule required for maintaining the differentiation state of ameloblasts, *J. Cell Biol. 167*, 973−983.
9. Gibson, C. W., Yuan, Z. A., Hall, B., Longenecker, G., Chen, E., Thyagarajan, T., Sreenath, T., Wright, J. T., Decker, S., Piddington, R., Harrison, G., and Kulkarni, A. B. (2001) Amelogenin-deficient mice display an amelogenesis imperfecta phenotype, *J. Biol. Chem. 276*, 31871−31875.
10. Hart, P. S., Hart, T. C., Michalec, M. D., Ryu, O. H., Simmons, D., Hong, S., and Wright, J. T. (2004) Mutation in kallikrein 4 causes autosomal recessive hypomaturation amelogenesis imperfecta, *J. Med. Genet. 41*, 545−549.
11. Kim, J. W., Simmer, J. P., Hart, T. C., Hart, P. S., Ramaswami, M. D., Bartlett, J. D., and Hu, J. C. (2005) MMP-20 mutation in autosomal recessive pigmented hypomaturation amelogenesis imperfecta, *J. Med. Genet. 42*, 271−275.
12. Lagerstrom, M., Dahl, N., Nakahori, Y., Nakagome, Y., Backman, B., Landegren, U., and Pettersson, U. (1991) A deletion in the amelogenin gene (AMG) causes X-linked amelogenesis imperfecta (AIH1), *Genomics 10*, 971−975.
13. Rajpar, M. H., Harley, K., Laing, C., Davies, R. M., and Dixon, M. J. (2001) Mutation of the gene encoding the enamel-specific protein, enamelin, causes autosomal-dominant amelogenesis imperfecta, *Hum. Mol. Genet. 10*, 1673−1677.
14. Bartlett, J. D., Simmer, J. P., Xue, J., Margolis, H. C., and Moreno, E. C. (1996) Molecular cloning and mRNA tissue distribution of a novel matrix metalloproteinase isolated from porcine enamel organ, *Gene 183*, 123−128.
15. Bartlett, J. D., Ryu, O. H., Xue, J., Simmer, J. P., and Margolis, H. C. (1998) Enamelysin mRNA displays a developmentally defined pattern of expression and encodes a protein which degrades amelogenin, *Connect. Tissue Res. 39*, 101−109.
16. Bègue-Kirn, C., Krebsbach, P. H., Bartlett, J. D., and Butler, W. T. (1998) Dentin sialoprotein, dentin phosphoprotein, enamelysin and ameloblastin: Tooth-specific molecules that are distinctively expressed during murine dental differentiation, *Eur. J. Oral Sci. 106*, 963−970.
17. Caterina, J., Shi, J., Sun, X., Qian, Q., Yamada, S., Liu, Y., Krakora, S., Bartlett, J. D., Yamada, Y., Engler, J. A., Birkedal-Hansen, H., and Simmer, J. P. (2000) Cloning, characterization, and expression analysis of mouse enamelysin, *J. Dent. Res. 79*, 1697−1703.
18. Fukae, M., Tanabe, T., Uchida, T., Lee, S. K., Ryu, O. H., Murakami, C., Wakida, K., Simmer, J. P., Yamada, Y., and Bartlett, J. D. (1998) Enamelysin (matrix metalloproteinase-20): Localization in the developing tooth and effects of pH and calcium on amelogenin hydrolysis, *J. Dent. Res. 77*, 1580−1588.
19. Hu, J. C., Sun, X., Zhang, C., Liu, S., Bartlett, J. D., and Simmer, J. P. (2002) Enamelysin and kallikrein-4 mRNA expression in developing mouse molars, *Eur. J. Oral Sci. 110*, 307−315.
20. Grant, G. M., Giambernardi, T. A., Grant, A. M., and Klebe, R. J. (1999) Overview of expression of matrix metalloproteinases (MMP-17, MMP-18, and MMP-20) in cultured human cells, *Matrix Biol. 18*, 145−148.
21. Takata, T., Zhao, M., Nikai, H., Uchida, T., and Wang, T. (2000) Ghost cells in calcifying odontogenic cyst express enamel-related proteins, *Histochem. J. 32*, 223−229.
22. Takata, T., Zhao, M., Uchida, T., Wang, T., Aoki, T., Bartlett, J. D., and Nikai, H. (2000) Immunohistochemical detection and distribution of enamelysin (MMP-20) in human odontogenic tumors, *J. Dent. Res. 79*, 1608−1613.
23. Väänänen, A., Tjaderhane, L., Eklund, L., Heljasvaara, R., Pihlajaniemi, T., Herva, R., Ding, Y., Bartlett, J. D., and Salo, T. (2004) Expression of collagen XVIII and MMP-20 in developing teeth and odontogenic tumors, *Matrix Biol. 23*, 153−161.
24. Kimura, A., Kihara, T., Ohkura, R., Ogiwara, K., and Takahashi, T. (2001) Localization of bradykinin B(2) receptor in the follicles of porcine ovary and increased expression of matrix metalloproteinase-3 and -20 in cultured granulosa cells by bradykinin treatment, *Biol. Reprod. 65*, 1462−1470.
25. Turk, B. E., Huang, L. L., Piro, E. T., and Cantley, L. C. (2001) Determination of protease cleavage site motifs using mixture-based oriented peptide libraries, *Nat. Biotechnol. 19*, 661−667.
26. Park, H. I., Turk, B. E., Gerkema, F. E., Cantley, L. C., and Sang, Q. X. (2002) Peptide substrate specificities and protein cleavage sites of human endometase/matrilysin-2/matrix metalloproteinase-26, *J. Biol. Chem. 277*, 35168−35175.
27. Bronckers, A. L., Gay, S., Lyaruu, D. M., Gay, R. E., and Miller, E. J. (1986) Localization of type V collagen with monoclonal antibodies in developing dental and peridental tissues of the rat and hamster, *Collagen Relat. Res. 6*, 1−13.
28. Lukinmaa, P. L., and Waltimo, J. (1992) Immunohistochemical localization of types I, V, and VI collagen in human permanent teeth and periodontal ligament, *J. Dent. Res. 71*, 391−397.

29. Pfaffl, M. W. (2001) A new mathematical model for relative quantification in real-time RT-PCR, *Nucleic Acids Res. 29*, e45.
30. Cartharius, K., Frech, K., Grote, K., Klocke, B., Haltmeier, M., Klingenhoff, A., Frisch, M., Bayerlein, M., and Werner, T. (2005) MatInspector and beyond: Promoter analysis based on transcription factor binding sites, *Bioinformatics 21*, 2933−2942.
31. Quandt, K., Frech, K., Karas, H., Wingender, E., and Werner, T. (1995) MatInd and MatInspector: New fast and versatile tools for detection of consensus matches in nucleotide sequence data, *Nucleic Acids Res. 23*, 4878−4884.
32. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1998) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Plainview, NY.
33. Hu, J. C., Sun, X., Zhang, C., and Simmer, J. P. (2001) A comparison of enamelin and amelogenin expression in developing mouse molars, *Eur. J. Oral Sci. 109*, 125−132.
34. Kawasaki, K., and Weiss, K. M. (2003) Mineralized tissue and vertebrate evolution: The secretory calcium-binding phosphoprotein gene cluster, *Proc. Natl. Acad. Sci. U.S.A. 100*, 4060−4065.
35. Nagase, H., and Fields, G. B. (1996) Human matrix metalloproteinase specificity studies using collagen sequence-based synthetic peptides, *Biopolymers 40*, 399−416.
36. Deng, S. J., Bickett, D. M., Mitchell, J. L., Lambert, M. H., Blackburn, R. K., Carter, H. L., III, Neugebauer, J., Pahel, G., Weiner, M. P., and Moss, M. L. (2000) Substrate specificity of human collagenase 3 assessed using a phage-displayed peptide library, *J. Biol. Chem. 275*, 31422−31427.
37. Kridel, S. J., Chen, E., Kotra, L. P., Howard, E. W., Mobashery, S., and Smith, J. W. (2001) Substrate hydrolysis by matrix metalloproteinase-9, *J. Biol. Chem. 276*, 20572−20578.
38. McGeehan, G. M., Bickett, D. M., Green, M., Kassel, D., Wiseman, J. S., and Berman, J. (1994) Characterization of the peptide substrate specificities of interstitial collagenase and 92-kDa gelatinase. Implications for substrate optimization, *J. Biol. Chem. 269*, 32814−32820.
39. Ryu, O. H., Fincham, A. G., Hu, C. C., Zhang, C., Qian, Q., Bartlett, J. D., and Simmer, J. P. (1999) Characterization of recombinant pig enamelysin activity and cleavage of recombinant pig and mouse amelogenins, *J. Dent. Res. 78*, 743−750.
40. Yaffe, M. B., Leparc, G. G., Lai, J., Obata, T., Volinia, S., and Cantley, L. C. (2001) A motif-based profile scanning approach for genome-wide prediction of signaling pathways, *Nat. Biotechnol. 19*, 348−353.
41. Väänänen, A., Srinivas, R., Parikka, M., Palosaari, H., Bartlett, J. D., Iwata, K., Grenman, R., Stenman, U. H., Sorsa, T., and Salo, T. (2001) Expression and regulation of MMP-20 in human tongue carcinoma cells, *J. Dent. Res. 80*, 1884−1889.
42. Ryu, J., Vicencio, A. G., Yeager, M. E., Kashgarian, M., Haddad, G. G., and Eickelberg, O. (2005) Differential expression of matrix metalloproteinases and their inhibitors in human and mouse lung development, *Thromb. Haemostasis 94*, 175−183.
43. Yamada, Y., Yamakoshi, Y., Gerlach, R. F., Hu, C. C., Matsumoto, K., Fukae, M., Oida, S., Bartlett, J. D., and Simmer, J. P. (2003) Purification and Characterization of Enamelysin from Secretory Stage Pig Enamel, *Archives of Comparative Biology of Tooth Enamel (Japan) 8*, 21−25.
44. Dohr, S., Klingenhoff, A., Maier, H., Hrabe de, A. M., Werner, T., and Schneider, R. (2005) Linking disease-associated genes to regulatory networks via promoter organization, *Nucleic Acids Res. 33*, 864−872.
45. Thomas, B. L., Tucker, A. S., Qui, M., Ferguson, C. A., Hardcastle, Z., Rubenstein, J. L., and Sharpe, P. T. (1997) Role of Dlx-1 and Dlx-2 genes in patterning of the murine dentition, *Development 124*, 4811−4818.
46. Tucker, A., and Sharpe, P. (2004) The cutting-edge of mammalian development: How the embryo makes teeth, *Nat. Rev. Genet. 5*, 499−508.
47. Dong, J., Amor, D., Aldred, M. J., Gu, T., Escamilla, M., and MacDougall, M. (2005) DLX3 mutation associated with autosomal dominant amelogenesis imperfecta with taurodontism, *Am. J. Med. Genet. A 133*, 138−141.
48. Zhang, H., Hu, G., Wang, H., Sciavolino, P., Iler, N., Shen, M. M., and bate-Shen, C. (1997) Heterodimerization of Msx and Dlx homeoproteins results in functional antagonism, *Mol. Cell. Biol. 17*, 2920−2932.
49. Alappat, S., Zhang, Z. Y., and Chen, Y. P. (2003) Msx homeobox gene family and craniofacial development, *Cell Res. 13*, 429−442.
50. Zhou, Y. L., Lei, Y., and Snead, M. L. (2000) Functional antagonism between Msx2 and CCAAT/enhancer-binding protein α in regulating the mouse amelogenin gene expression is mediated by protein−protein interaction, *J. Biol. Chem. 275*, 29066−29075.
51. Olson, L. E., Zhang, J., Taylor, H., Rose, D. W., and Rosenfeld, M. G. (2005) Barx2 functions through distinct corepressor classes to regulate hair follicle remodeling, *Proc. Natl. Acad. Sci. U.S.A. 102*, 3708−3713.
52. Yamada, G., Mansouri, A., Torres, M., Stuart, E. T., Blum, M., Schultz, M., De Robertis, E. M., and Gruss, P. (1995) Targeted mutation of the murine goosecoid gene results in craniofacial defects and neonatal death, *Development 121*, 2917−2922.
53. Peterson, R. L., Papenbrock, T., Davda, M. M., and Awgulewitsch, A. (1994) The murine Hoxc cluster contains five neighboring AbdB-related Hox genes that show unique spatially coordinated expression in posterior embryonic subregions, *Mech. Dev. 47*, 253−260.
54. Andersen, B., and Rosenfeld, M. G. (2001) POU domain factors in the neuroendocrine system: Lessons from developmental biology provide insights into human disease, *Endocr. Rev. 22*, 2−35.
55. Bode, W., Fernandez-Catalan, C., Tschesche, H., Grams, F., Nagase, H., and Maskos, K. (1999) Structural properties of matrix metalloproteinases, *Cell. Mol. Life Sci. 55*, 639−652.
56. Welch, A. R., Holman, C. M., Huber, M., Brenner, M. C., Browner, M. F., and Van Wart, H. E. (1996) Understanding the P1′ specificity of the matrix metalloproteinases: Effect of S1′ pocket mutations in matrilysin and stromelysin-1, *Biochemistry 35*, 10103−10109.
57. Park, H. I., Jin, Y., Hurst, D. R., Monroe, C. A., Lee, S., Schwartz, M. A., and Sang, Q. X. (2003) The intermediate S1′ pocket of the endometase/matrilysin-2 active site revealed by enzyme inhibition kinetic studies, protein sequence analyses, and homology modeling, *J. Biol. Chem. 278*, 51646−51653.
58. Smith, C. E., Pompura, J. R., Borenstein, S., Fazel, A., and Nanci, A. (1989) Degradation and loss of matrix proteins from developing enamel, *Anat. Rec. 224*, 292−316.
59. Chen, E. I., Kridel, S. J., Howard, E. W., Li, W., Godzik, A., and Smith, J. W. (2002) A unique substrate recognition profile for matrix metalloproteinase-2, *J. Biol. Chem. 277*, 4485−4491.
60. Kridel, S. J., Sawai, H., Ratnikov, B. I., Chen, E. I., Li, W., Godzik, A., Strongin, A. Y., and Smith, J. W. (2002) A unique substrate binding mode discriminates membrane type-1 matrix metalloproteinase from other matrix metalloproteinases, *J. Biol. Chem. 277*, 23788−23793.
61. Liu, X., Wu, H., Byrne, M., Krane, S., and Jaenisch, R. (1997) Type III collagen is crucial for collagen I fibrillogenesis and for normal cardiovascular development, *Proc. Natl. Acad. Sci. U.S.A. 94*, 1852−1856.
62. Wenstrup, R. J., Florer, J. B., Brunskill, E. W., Bell, S. M., Chervoneva, I., and Birk, D. E. (2004) Type V collagen controls the initiation of collagen fibril assembly, *J. Biol. Chem. 279*, 53331−53337.
63. Palosaari, H., Pennington, C. J., Larmas, M., Edwards, D. R., Tjaderhane, L., and Salo, T. (2003) Expression profile of matrix metalloproteinases (MMPs) and tissue inhibitors of MMPs in mature human odontoblasts and pulp tissue, *Eur. J. Oral Sci. 111*, 117−127.

BI052252O