Heterogeneity of Amelogenin mRNA in the Bovine Tooth Germ*

Hitoyata Shimokawa‡, Mark E. Sobel§, Makoto Sasaki†, John D. Termine¶, and Marian F. Young∥

From the ‡Bone Research Branch, National Institute of Dental Research, the §Laboratory of Pathology, National Cancer Institute, and the ¶Laboratory of Developmental Biology and Anomalies, National Institute of Dental Research, National Institutes of Health, Bethesda, Maryland 20892

The amelogenins are a complex mixture of hydrophobic proteins that are the major organic component of developing enamel. To study the molecular mechanisms underlying the heterogeneity of the amelogenins, we isolated cDNA clones encoding these proteins. The clones were definitively identified by hybrid-selected translation experiments and by comparison of the DNA sequence with the protein-derived amino acid sequence. Southern hybridization of bovine genomic DNA indicated that amelogenin is a single copy gene. However, Northern hybridization experiments distinctly showed two major species of mRNA, each of which were sufficiently large enough to encode the highest known molecular weight species of amelogenin proteins. Furthermore, immunoprecipitation of hybrid-selected translation products using isolated amelogenin cDNA showed multiple, translated protein products. These data are supportive of a differential mRNA processing mechanism involved in generating a heterogeneous family of amelogenin matrix proteins from a single gene.

The amelogenins are a complex mixture of hydrophobic proteins that are the major organic component of developing enamel. It has been suggested that these proline-rich polypeptides play a role in the mineralization and structural organization of developing enamel (1–4). Although purification and characterization of the amelogenins in immature enamel is difficult, dissociative extraction protocols have been successfully used to determine the sizes and compositions of the major molecular species (5–8). These studies indicated that the amelogenins are very heterogeneous and that during enamel mineralization, there is a shift in the average size of the matrix amelogenin molecules to lower molecular weight species (5, 7). While part of this lower molecular weight population can be attributed to proteolytic degradation, primary sequence analysis of the two major small molecular weight, bovine amelogenin species showed that one of them, a leucine-rich amelogenin peptide, had differences in primary structure from that of unprocessed higher molecular weight amelogenin molecules (6, 8). Amino acid sequences derived from one specific nucleotide sequence of mouse amelogenin cDNA (9) suggest that several amelogenins are synthesized as original gene products but have not completely resolved these differences. This suggests that some mechanism other than degradation of a single amelogenin translation product must be involved in generating the complex mixture of amelogenins in developing enamel.

To elucidate some of these apparent complexities in amelogenin gene expression, we have constructed a bovine ameloblast cDNA library and isolated two cDNAs encoding amelogenin. DNA sequence analysis described in this report will show that the two clones overlap extensively and that the protein they encode shows primary sequence differences from a previously reported protein-derived amino acid sequence of one of the 27,000-dalton forms of the bovine amelogenins (10). Gene copy number experiments suggest only one copy of the amelogenin gene/haploid genome. These and other data suggest that a differential mRNA processing mechanism may be involved in generating the heterogeneous proteins in the amelogenin peptide family.

MATERIALS AND METHODS

Enamel Preparation and Protein Extraction—Pure immature bovine enamel was obtained by removal of unerupted permanent incisor tooth germs from 18-month-old steers followed by fine dissection to remove adhering soft tissue. The developing enamel was scraped free of contaminating dentin from the outer tooth surface. The enamel samples were extracted with 0.5 M acetic acid as described previously (5, 11). All resulting protein fractions were dialyzed against cold 0.5 M acetic acid using Spectrapor-3 acetylated dialysis tubing (M, cut off ≈3500) and lyophilized. Extracts were analyzed by two-dimensional SDS–gel electrophoresis (12) using a mixture of 0.4%, 3.5–9.5, and 1.6%, 6–8 amphotolines (LKB) containing 8 M urea in the first dimension and a 15–30% polyacrylamide gradient containing 3% acrylamide stacking gel in the second dimension. Isoelectric focusing was performed at 300 V for 20 h. Proteins were visualized by silver stain (13).

Anti-amelogenin Antibodies Preparation—Chromatographically purified amelogenins (11) were injected intradermally and intramuscularly into rabbits in an emulsion of Freund's complete adjuvant containing 2.0 mg of protein. Booster injections, containing 1.0 mg of amelogenin in incomplete adjuvant, were administered similarly every 2 weeks for 6 weeks. Antibodies were purified from potential contaminants by affinity chromatography through Sepharose-protein A, Sepharose fetal calf serum, and Sepharose enamelines (non-amelogenin, enamel proteins) and finally, by affinity to Sepharose-bound purified amelogenins. The monospecificity of the resulting antibody preparation was tested by direct enzyme-linked immunosorbent assay and Western blots against various proteins including bovine serum, dentin extracts, purified enamelines, and amelogenins (14).

RNA Extraction and in Vitro Protein Synthesis—Total RNA was extracted from frozen ameloblast-rich tissue (7, 11) and translated in a nuclease-treated reticulocyte lysate system (New England Nuclear) using [35S]methionine (=1000 mCi/mmole) as radioactive tracer. Translation was terminated by the addition of 4 volumes of gel sample buffer (final concentration: 50 mM Tris-HCl (pH 6.8), 2.5% SDS, 10% glycerol, 0.5 mM β-mercaptoethanol, and 0.01% bromophenol blue) and used in the translation experiments.

The abbreviations used are: SDS, sodium dodecyl sulfate; bp, base pairs; PIPES, piperazine-N,N'-bis(2-ethanesulfonic acid).
blue) or by direct immunoprecipitation (15). Hybrid selected translation was carried out essentially as described (16) except that gel-purified insert DNA from the recombinant clone was used and that hybridization was performed with 300 \( \mu \)g/ml poly(A)+ mRNA in 60% formamide, 10 mM PIPES, and 0.4 mM NaCl at 42 °C for 16 h. Hybrid-selected mRNA was translated, immunoprecipitated, and separated by isoelectric focusing and reverse transcriptase (Life Science) in the presence of 0.1 mM Tris (pH 8.3), 0.14 mM KCl, 0.01 mM MgCl₂, and 0.8 mM each of unlabeled deoxyribonucleotide triphosphates (Pharmacia P-L Biochemicals). Then 10 \( \mu \)Ci each of [\( ^{32}P \)]dGTP and [\( ^{32}P \)]dATP were added to the reaction to evaluate the quantity and quality of first strand produced. This and subsequent reactions were terminated by adding EDTA (pH 8.0) and SDS to 23 mM and 0.47%, respectively. Enzymes and salts were removed by phenol/chloroform/isooamy alcohol (25:24:1) extractions followed by precipitation in 2 M ammonium acetate and 2 volumes of 100% ethanol. Precipitates were rinsed in 70% ethanol and dried by lyophilization.

For second strand synthesis single-stranded cDNA was resuspended in 20 mM Tris-HCl (pH 7.5), 4 mM MgCl₂, 10 mM (NH₄)₂SO₄, 100 mM KCl, 50 \( \mu \)g/ml bovine serum albumin (nuclease free, Bethesda Research Laboratories), 150 \( \mu \)M 3'-NAD, and 100 \( \mu \)M each of the four unlabeled deoxyribonucleotide triphosphates. After adding 10 \( \mu \)Ci each of [\( ^{32}P \)]dCTP and [\( ^{32}P \)]dATP as tracer, reactions were carried out with 1.5 units of E. coli DNA ligase and 10 units of DNA polymerase I (Boehringer Mannheim) for 1 h at 15 °C followed by incubation for 1 h at 25 °C.

Double-stranded cDNA was incubated in 250 mM NaCl, 50 mM NH₄ acetate (pH 4.5), 1 mM ZnCl₂, and 50 units of S1 nuclease (Bethesda Research Laboratories) for 30 min at 37 °C to remove hairpin loops. To assure that all cDNA molecules were blunt ended, the cDNA was incubated in 50 mM Tris-HCl (pH 8.0), 6 mM MgCl₂, 25 mM KCl, 100 \( \mu \)g/ml bovine serum albumin, 0.1 mM each of the four deoxyribonucleotide triphosphates, and 3 units of T₄ DNA polymerase for 30 min at 37 °C. Blunt ended cDNA was methylated by incubation in 0.1 M Tris-HCl (pH 8.0), 1 mM EDTA, 10 \( \mu \)M S-adenosylmethionine, 0.4 mg/ml bovine serum albumin, and 6 units of EcoRI methylase (New England Biolabs) for 20 min at 37 °C. A 0.5 unit of phosphorylated EcoRI linkers (New England Biolabs) was ligated to methylated cDNA by incubation in 25 mM Tris-HCl (pH 7.4), 10 mM MgCl₂, 10 mM dithiothreitol, 0.2 mM spermidine, 0.5 mM ATP, and 5 units of T₄ DNA ligase (New England Biolabs) for 18 h at 12 °C. Excess linkers were then cleaved by EcoRI (100 units) and separated from cDNA by chromatography on a Sepharose 4B column as described (11). Approximately 3.0 \( \mu \)g of EcoRI restricted λgt11 were added to the cDNA and incubated for 48 h under the ligation conditions described above. When recombinant cDNA was packaged into phage particles in vitro (Amersham Corp.), approximately 1 \times 10⁷ phage particles were obtained. About 30% of the phage contained cDNA insert. The ameloblast λgt11 cDNA library was screened as described by Young and Davis (17) using, for the first antibody, affinity purified rabbit anti-amelogenin antibodies at a dilution of 1:1000 and, for the second antibody, horseradish peroxidase-conjugated goat anti-rabbit antisera (Bio-Rad) at a dilution of 1:2000. Positive plaques were identified by reaction with 4-chloro-1-naphthol, amplified, and finally rescreened to purity as described by Benton and Davis (18).

**Northern Analysis—**3.5-\( \mu \)g aliquots of total RNA from bovine ameloblast and fetal bovine liver were electrophoresed in 1.2% formaldehyde-agarose gels and transferred to nitrocellulose as described (16). Hybrized gel-purified EcoRI restriction fragments previously liberated from the λgt11 vector were each nick translated (Amersham Corp.) with [\( ^{32}P \)]deoxyribonucleotides to a specific activity of 10⁷ cpm/\( \mu \)g and hybridized to the RNA bound to the filters. Hybridization was carried out at 37 °C in a mixture of 40% formamide (Fluka), 0.1 g/ml dextran sulfate, 5 \times SSC (75 mM sodium citrate, 750 mM sodium chloride), 10 mM Tris-HCl (pH 7.5), 4 \times Denhardt’s solution (19), and 0.1 mg/ml denatured salmon sperm DNA. The filters were washed 4 times for 10 min each in 2 \times SSC, 0.2% SDS at 25 °C and 4 times for 10 min in 0.1 \times SSC at 68 °C. Filters were autoradiographed by exposure of x-ray film to filters for 6 to 2 days.

**DNA Sequence of Amelogenin cDNA**—After amplification of the cloning fragment, the proteins were applied to the high pH end of the gel (the cathode). SDS-polyacrylamide gel electrophoresis was carried out in the second dimension using 15-30% (w/v) linear gradient polyacrylamide. Proteins were detected by silver staining (Fig. 1A), also immunoreacted with affinity purified anti-amelogenin protein antibodies to assure that all cDNA molecules were blunt ended, the cDNA was incubated in 50 mM Tris-HCl (pH 8.0), 6 mM MgCl₂, 25 mM KCl, 100 \( \mu \)g/ml bovine serum albumin, 0.1 mM each of the four deoxyribonucleotide triphosphates, and 3 units of T₄ DNA polymerase for 30 min at 37 °C. Blunt ended cDNA was methylated by incubation in 0.1 M Tris-HCl (pH 8.0), 1 mM EDTA, 10 \( \mu \)M S-adenosylmethionine, 0.4 mg/ml bovine serum albumin, and 6 units of EcoRI methylase (New England Biolabs) for 20 min at 37 °C. A 0.5 unit of phosphorylated EcoRI linkers (New England Biolabs) was ligated to methylated cDNA by incubation in 25 mM Tris-HCl (pH 7.4), 10 mM MgCl₂, 10 mM dithiothreitol, 0.2 mM spermidine, 0.5 mM ATP, and 5 units of T₄ DNA ligase (New England Biolabs) for 18 h at 12 °C. Excess linkers were then cleaved by EcoRI (100 units) and separated from cDNA by chromatography on a Sepharose CL-4B column as described (11). Approximately 3.0 \( \mu \)g of EcoRI restricted λgt11 were added to the cDNA and incubated for 48 h under the ligation conditions described above. When recombinant cDNA was packaged into phage particles in vitro (Amersham Corp.), approximately 1 \times 10⁷ phage particles were obtained. About 30% of the phage contained cDNA insert. The ameloblast λgt11 cDNA library was screened as described by Young and Davis (17) using, for the first antibody, affinity purified rabbit anti-amelogenin antibodies at a dilution of 1:1000 and, for the second antibody, horseradish peroxidase-conjugated goat anti-rabbit antisera (Bio-Rad) at a dilution of 1:2000. Positive plaques were identified by reaction with 4-chloro-1-naphthol, amplified, and finally rescreened to purity as described by Benton and Davis (18).
tween our cDNA-derived sequence and the 175-residue protein.

The DNA sequence of the overlapping clones, XAm 11, was characterized extensively (insert sizes 700 and 300 bp, respectively). Restriction analysis showed both clones contained an internal HindIII site and that the longer clone also contained three PstI sites (Fig. 2). A DNA sequencing strategy based on these mapping data is shown in Fig. 2.

**Derived Amino Acid Sequence of XAm 11 and λAm 16—**The DNA sequence of λAm 11 and XAm 11 was determined and the derived amino acid sequence compared to the complete amino acid sequence reported (10) for one purified bovine amelogenin. Fig. 3 shows that our longest clone, λAm 16 (insert size 700 bp), lacks DNA sequences encoding amino acids 1-32 but contains sequence for the rest of the protein. Eight disperse amino acid discrepancies were observed between our cDNA-derived sequence and the 175-residue protein sequence previously reported (10) for one purified bovine amelogenin. Fig. 3 brackets, residues 1-33 and 164-174). The last 12 derived amino acid residues in the cDNA were not detected in the published bovine protein sequence, but were identical to the carboxyl terminus of the cDNA-derived amino acid sequence recently reported for a mouse amelogenin (9). The DNA sequence of the overlapping regions of λAm 11 and λAm 16 were identical.

**A Comparison of Mouse and Bovine Amelogenin cDNA—**The DNA sequence from our bovine amelogenin cDNA was compared to the DNA sequence previously reported for a mouse amelogenin (9) (Fig. 4). Most striking is the high degree of homology in the first 180 bases (87% homology) and last 165 bases (86% homology) of the coding regions. It should be noted that this homologous region at the 3' end encodes the protein sequence of divergence (residues 139-163) discussed above. More DNA digression was detected internally, however, in the two cDNAs; even with optimum alignment, a stretch of 21 bp (residues 280-300) appears to be absent in the mouse sequence.

**Heterogeneity and Homology of Amelogenin mRNA—**The detection of a heterogeneous population of cell-free translation products from ameloblast RNA immunoprecipitable by anti-amelogenin antisera, suggests that an analogous heterogeneous population of amelogenin mRNA exists. To determine whether the heterogeneous population of amelogenin mRNA shares sequence homology, hybrid-selected translation experiments were performed. Under stringent hybridization conditions (see "Materials and Methods") DNA from λAm 16 specifically hybridized ("selected") mRNA encoding at least four species of protein (Fig. 5, Lane 3). These proteins were all shown to be amelogenins based on their immunoprecipitation with antibodies to amelogenin (Fig. 5, Lane 4). Identical results were obtained using λAm 11 (data not shown).

**Size of Amelogenin mRNA—**Total RNA from purified bovine ameloblast and bovine liver was separated by electrophoresis, transferred to nitrocellulose, and allowed to hybridize to radiolabeled insert from XAm 16 and λAm 16. Extensive hybridization was found to ameloblast but not liver mRNA (Fig. 6). The mRNA encoding bovine amelogenin was two sizes, 14 S and 10 S (Fig. 6, Lanes 1 and 3). With prolonged exposure times, a larger minor species of mRNA (28 S) was also detected.

**Amelogenin Genomic DNA—**To determine whether multiple copies of the amelogenin gene are present in the genome, bovine genomic DNA was fragmented with various restriction enzymes and analyzed by Southern blotting. With all the enzymes tested, radiolabeled λAm 16 hybridized to a single genomic band (Fig. 7, Lanes 1-3). The detection of one HindIII genomic fragment in a gene with an internal HindIII site (see restriction map of λAm 11 and λAm 16) may represent two fragments with similar gel mobility or, alternatively, lack of detection due to the resolution of the gel (fragments too large or small). The banding pattern for human was similarly simple but different in the absolute lengths of hybridizable fragments (Fig. 7, Lanes 4-6).

**DISCUSSION**

Analysis of the composition of developing dental enamel indicates that the major protein family, the amelogenins, contains related yet multifarious peptide members. This heterogeneity was first demonstrated at the protein level (Refs. 5-8 and the present study) and has now been extended to include possible heterogeneity of the mRNA. When ameloblast mRNA was translated in a cell-free system, at least 10 amelogenin protein species were observed. In order to demonstrate unequivocally that this amelogenin diversity arises from a diverse population of mRNA, we have constructed and isolated two cDNAs encoding a bovine amelogenin. Analysis of these cDNAs indicates that several amelogenin mRNAs may contribute to the intricate pattern of amelogenin gene expression during ameloblast differentiation.

Analysis of the derived amino acid sequence obtained from the amelogenin cDNA has revealed several unusual features about these novel proteins. First, several scattered, minor discrepancies were observed between our cDNA-derived se-
FIG. 3. Derived amino acid sequence of a bovine amelogenin. a, derived sequence obtained from DNA sequence of XAm 11 and XAm 16; b, amino acid sequence of one of the M, = 27,000 forms of amelogenin previously determined by Takagi et al. (10), aligned with optimum homology. Numbering starts with the first amino acid reported by Takagi et al. (10). Dashed line represents sequence not determined from cDNA. Boxed sequences show the discrepancies between a and b where divergent sequences deduced from cDNA were found. The stars (*) at residues 93 and 158-161 show residues present in our cDNA but absent in the previously reported amino acid sequence (10). Bracket sets delineate sequences corresponding to the leucine-rich amelogenin peptide (6, 8). Dots are shown at cDNA-derived sequences not previously reported from bovine amelogenin amino acid sequence, identical to the carboxyl terminus of a cDNA-sequenced murine amelogenin (9).

FIG. 4. Comparison of mouse and bovine amelogenin cDNA sequences. The numbering system starts with the first DNA residue of XAm 16. Stars show bases that diverge between mouse (9) and bovine cDNA. Dashes at residues 262-279 show sequences not determined from the bovine cDNA subclones. Open triangles show a region in the bovine sequence which has no corresponding homology to the mouse sequence.

sequence and the amino acid sequence previously reported for a single purified bovine amelogenin protein which had been isolated from enamal of a similar developmental stage. These discrepancies may be accounted for either by errors in amino acid sequencing or, even more likely, by subtle differences between the peptide encoded by this particular mRNA (cDNA) and the particular amelogenin peptide used for amino acid sequence analysis. In addition to these minor acid differences, a major cluster of divergence, also seen in a mouse amelogenin species (9), was observed between residues 139 and 163. This major block of divergence was distinctly amelogenin-like in sequence character and is unlikely to result solely due to amino acid sequencing errors. This may indicate that these cDNA encode a novel species of amelogenin which is similar but not identical to the peptide previously reported by Takagi et al. (10). Similar, subtle primary sequence divergence has previously been demonstrated for the small molecular weight (M, ~ 5000) tyrosine-rich amelogenin peptides and leucine-rich amelogenin peptides (6, 8). These peptides are enriched in more heavily mineralized enamel matrix, known to undergo substantial proteolytic degradation (6, 8). In each of these two small peptides, the first 33 residues are identical to each other and to the reported amino terminus of a larger (M, = 27,000) form of amelogenin. Conversely, residues 34-45 of the tyrosine-rich amelogenin peptides and residues 34-46 of the leucine-rich amelogenin peptides were different (6, 8). Residues 34-45 of the tyrosine-rich amelogenin peptides corresponded to identical residues (34-45) of the M, = 27,000 amelogenin, suggesting that this peptide (tyrosine-rich amelogenin peptide) is derived from the amino terminus of the larger protein (6, 8). Quite interestingly, we note that residues 37-46 of the leucine-rich amelogenin peptides have a nearly identical sequence to residues 160-169 of both the bovine amelogenin sequence reported here and by Takagi (10).

In order to determine the possible relationship of our cDNA to other amelogenin mRNA, hybrid-selected translation ex-
Amelogenin mRNA Heterogeneity

FIG. 5. Hybrid selected translation. A 700-bp fragment from \( \lambda \)Am 16 was immobilized on diazobenzyloxymethylcellulose paper and hybridized to total ameloblast RNA. Hybrid selected mRNA was thermally eluted, translated in a rabbit reticulocyte lysate, and electrophoresed in 10% polyacrylamide gels. Lane 1, endogenous translation in the absence of RNA; Lane 2, total translation products of ameloblast mRNA; Lanes 3 and 4, ameloblast mRNA hybrid selected with \( \lambda \)Am 16 before (Lane 3) and after (Lane 4) immunoprecipitation with anti-amelogenin antisera.

FIG. 6. Northern hybridization. Total RNA was extracted from bovine ameloblast (Lanes 1 and 3) and fetal bovine liver (Lanes 2 and 4) as described under "Materials and Methods." RNA was electrophoresed in 1.2% agarose gels and transferred to nitrocellulose. The filter was hybridized to \( ^{32}P \) labeled EcoRI insert from \( \lambda \)Am 11 (Lanes 1 and 2) and \( \lambda \)Am 16 (Lanes 3 and 4).

Amelogenin RNA heterogeneity experiments were performed. The rationale for these studies was based on the assumption that minimal processing of protein occurs in the rabbit reticulocyte lysate. In these studies, at least four different polypeptides were translated from the RNA(s) hybridized to our cloned amelogenin cDNA. This observation implies that if multiple RNAs for amelogenin are responsible for generating multiple forms of amelogenin, these RNAs share extensive homology to \( \lambda \)Am 16 and most likely to each other, as well. It is unlikely that a single amelogenin mRNA was selected by \( \lambda \)Am 16 and upon translation in the rabbit reticulocyte lysate, generated multiple protein species. To determine more precisely the nature of the mRNAs encoding amelogenin, Northern blot experiments were performed.

Two species of amelogenin message were apparent with sizes of 14 S and 10 S. In contrast, the mouse amelogenin mRNA appears to be homogeneous (25), although its size of 13 S is similar to those detected in the bovine system. In the bovine mRNA, both the 10 S and 14 S species would be sufficient in length to encode one protein species of \( M_r = 27,000 \). Our translation and Northern data suggest that, in addition to the apparent differences in their overall length, it is likely that bovine amelogenin RNAs differ internally as well.

Based on the apparent diversity of the amelogenin protein family, it was of interest to determine whether this diversity possibly arises from an equally diverse set of genes within the bovine genome. Multiple copies of similar, but nonidentical genes of a different group of proline-rich proteins have recently been reported (26). Hybridization of radiolabeled proline-rich protein cDNA to genomic restriction digest yields an intricate banding pattern suggesting multiple copies of these salivary gland genes (26). Conversely, when a similar experiment was performed using radiolabeled \( \lambda \)Am 16 as probe, a single band of hybridization was detected for both bovine and human genomic DNA implying that only one copy of the amelogenin gene exists per haploid genome. It is possible that part of the diversity of the amelogenin peptide family may arise by a differential splicing mechanism as has been demonstrated for several other heterogeneous protein families (27-29).

Acknowledgments — We would like to thank C. Ramis for help with Northern and Southern blotting and A. Day for advice on hybrid-selected translation experiments.

REFERENCES


Fig. 7. Genomic restriction digests. Genomic DNA from bovine (Lanes 1-3) and human (Lanes 4-6) liver was digested with various restriction enzymes, electrophoresed in 8% agarose gels, and transferred to nitrocellulose. Insert DNA from \( \lambda \)Am 16 was radiolabeled with \( ^{32}P \) by nick translation and hybridized to the blot under stringent conditions. Autoradiograms were obtained after exposure of x-ray film for 3 days.
Amelogenin mRNA Heterogeneity