The amelogenin gene

SATOSHI SASAKI and HITOYATASHI SHIMOKAWA*
Department of Biochemistry, School of Dentistry, Tokyo Medical and Dental University, Tokyo, Japan

ABSTRACT Amelogenin is a major protein constituent of the developing enamel matrix. This protein is now well characterized from the data of amino acid sequences which have been shown to be at a high degree of homology between all species investigated to date. The gene structure of this protein has been demonstrated and it is confirmed that there are two amelogenin genes, one on the X-chromosome and the other on the Y-chromosome in humans. The mapping of human amelogenin gene in the p22 region on the X-chromosome and the gene on the Y-chromosome was established. It has been confirmed that several types of X-linked amelogenesis imperfecta are caused by structural defects in the amelogenin gene on the X-chromosome. The physiological importance of amelogenin in the enamel formation is suggested by the symptoms of this inherited disease in addition to inhibition experiments of amelogenin transcription and translation (Couwenhoven et al., J. Craniofac. Genet. Dev. Biol. 13:259-269, 1993; Diekwisch et al., Development 117:471-482, 1993). Recently, an attempt to synthesize recombinant mouse amelogenin by E. coli was also undertaken (Simmer et al., Calcif. Tissue Int. 54:312-319, 1994). The regulation of amelogenin expression is now under investigation (Chen et al., Dev. Dynamics 199:189-198, 1994) and the elucidation of this mechanism will contribute a great deal to the study of tooth development.

KEYWORDS: enamel protein, amelogenin, gene structure, phylogeny, amelogenesis imperfecta

Introduction

In the early stages of tooth development, internal enamel epithelial cells differentiate into ameloblasts which synthesize and secrete specific proteins as enamel matrix. Protein concentrations as high as 25-30% can be found in newly-secreted enamel. The matrix proteins then decrease during the maturation of the enamel and are replaced by apatite crystals. These proteins are thought to play important roles in the process of enamel mineralization.

There are two classes of proteins in the early enamel, amelogenin and enamelin. Amelogenin, which was termed by Eastoe (1965), is a discrete and major constituent unique to the developing enamel. This protein contains high concentrations of proline, glutamine, leucine and histidine. Another protein, enamelin, termed by Termine et al. (1980), is a minor protein component present in a form associated with the mineral phase in the developing enamel. Since neither mRNA nor the gene for enamelin has been identified to date, the nature of this protein has yet to be unequivocally established. On the other hand, amelogenin is now well characterized from amino acid sequencing data and the gene structure.

Identification of amelogenin gene

The amino acid sequence of purified bovine amelogenin was first reported by direct Edman sequencing (Takagi et al., 1984). Since then, the primary structures of amelogenin proteins of several mammalian species have been identified also indirectly by deduction from mRNA sequences (Shimokawa et al., 1987a; Gibson et al., 1992; Lau et al., 1992; Bonass et al., 1994). In addition, the presence of multiple amelogenin fractions was detected by gel filtration chromatography or electrophoresis and heterogeneity of matrix amelogenin in terms of molecular size and amino acid composition was demonstrated. Those findings prompted researchers to identify the nature of amelogenin gene(s) and to analyze the expression mechanism.

The amelogenin genes were identified on sex chromosomes in both mouse and man first by Lau et al. (1989). They constructed a recombinant cDNA for mouse amelogenin and used it as a specific hybridization probe. They found that there is only one amelogenin gene in the mouse genome, which is located on the X-chromosome. The analysis established the locus of the amelogenin gene to the distal portion of the mouse X-chromosome. On the other hand, two copies of the amelogenin genes were detected in the human male genome by Southern blot analysis of human-rodent cell hybrids. One copy of the amelogenin genes was located on the distal short arm of the X-chromosome in the p22.1-p22.3 region. The second copy of the amelogenin gene was localized near the centromere of the Y-chromosome. However, they did not determine DNA sequences for human and mouse amelogenin genes. At about the same time, human amelogenin gene, assumed to be

Abbreviations used in this paper: RT, reverse transcriptase; PCR, polymerase chain reaction.
The presence of at least four exons in the human gene derived from X-chromosome, was sequenced by Shimokawa et al. (1989). Thereafter, it has been confirmed that there are two amelogenin genes, one on the X-chromosome and the other on the Y-chromosome in some animals, and a single gene resides exclusively on the X-chromosome in other animals.

Nucleotide sequences of amelogenin genes

Human amelogenin gene was first isolated and sequenced by Shimokawa et al. (1989). A human genomic DNA library in the phage-vector was screened by plaque hybridization using a bovine amelogenin cDNA probe. One clone containing a 13-kb insert, probably from X-chromosome, was obtained. Restriction enzyme fragments were subcloned and sequenced by the dideoxy chain termination method. Some segments of the isolated human genomic DNA corresponded to the bovine amelogenin sequence. At that time, the presence of at least four exons in the human gene structure was suggested. First coding exon was not identified, although nucleotides of three putative exons (corresponding to exons 3, 5 and 6 by Salido et al., 1992; Fig. 1) were sequenced and the deduced amino acid sequence was reported.

Nakahori et al. (1991a) reported two genomic sequences for human amelogenin from the X- and Y-chromosomes. They isolated a Y-specific clone encoding amelogenin which was used as a probe. Southern hybridization analysis of male and female genomic DNAs detected a male-specific band on the Y-chromosome and a band common to male and female on the X-chromosome. By digestion analysis with restriction enzymes, the difference between the X- and Y-chromosome sequences was confirmed. They sequenced the nucleotides in the homologous regions from the X- and Y-chromosomes and aligned them with the best possible match. There was 88.9% homology between the X and Y nucleotide sequences. First exon was not identified and three presumed exons (exons 3, 5 and 6 by Salido et al., 1992; Fig. 1) both on the X and Y (93% homology) were recognized and amino acid sequences were deduced. However, it was not certain whether the human Y amelogenin gene is transcriptionally active or not.

Fincham et al. (1991) reported the differences in human amelogenin components according to the sex of the individual. They extracted enamel protein samples from fetal and postnatal human deciduous dentitions and the proteins were analyzed by electrophoresis. Amelogenin bands specific for male-derived specimens were detected and a sex-linked dimorphism of amelogenin was indicated. Later, in 1992, Salido et al. reported that human amelogenin genes are expressed from both the X- and Y-chromosomes. A human genomic library was screened with a probe of mouse amelogenin cDNA, and positive clones from the X- and Y-chromosomes were isolated. Both genes spanned more than 9 kb and 7 exons were identified in the sequences of both the X- and Y-derived clones of amelogenin genes (Fig. 1). Complete sequences of human amelogenin protein including signal peptide of 16 residues on both X- and Y-chromosomes were identified.

The mature human amelogenin from the X-chromosome skipping exon 4 is a protein of 19.8 kDa with 175 amino acid residues and which from Y is a 20 kDa protein with 176 amino acids. The methionine at residue 29 of the Y-chromosome amelogenin is absent from the X amelogenin, due to a 3-base deletion in the human X amelogenin gene. The overall sequence similarity between the X- and Y-derived cDNAs is 91%. The protein coding regions (exons 2-6) are highly conserved, with a similarity index between 93% and 100%, whereas the 5' and 3' untranslated regions (exons 1 and 7) are much less conserved.

A major transcript was shown to be derived from the X-chromosome by the technique of RT-PCR. However, faint Y-derived transcripts which hybridized only with the Y-specific probe were detected. The result indicated that the amelogenin sequence on the Y-chromosome is also transcribed, although at a level much lower (about 10% of the total) than that of the X-homolog.

Gibson et al. (1991b, 1992) isolated and characterized bovine amelogenin genes. They isolated bovine X and Y amelogenin genomic clones containing six exons encoding the entire protein structure. Within the coding regions of the two genes, 93% of the nucleotides were identical. However, there was deletion of 63 nucleotides in exon 5 of the Y-chromosome in addition to 41 single nucleotide differences as compared to the X-specific region (Fig. 1). Accordingly, a loss of 21 amino acids is found from mature region of Y-specific amelogenin protein (176 residues) as compared to that of X-specific amelogenin (197 residues), and there is a total of 11 amino acid differences in the mature protein portion plus 2 differences in the signal sequence (16 amino acids). The part of the 21 amino acids lacking in the Y-amelogenin is the (Gln-Pro-X) repetitive sequence, which has been postulated to have a β-spiral structure (Renugopalakrishnan et al., 1986). This structural feature is found only in bovine X-chromosomal amelogenin but is absent from all other known amelogenin sequences. In that sense, bovine X-specific amelogenin is unique in its structure. In bovine tooth development, the amelogenin gene on the Y-chromosome may also be actively transcribed because Y-responsible cDNAs were identified. Comparison of human and bovine amelogenin genes on the X- and Y-chromosomes is shown in Fig. 1. Since exon 4 in human amelogenin genes is usually skipped when they are transcribed, there may be a possibility that an equivalent exon exists in bovine amelogenin genes even though it was not identified by the study of Gibson et al. (1992).

There is a discrepancy concerning sexual dimorphism of tooth size. In humans, male teeth are usually larger than female teeth, but the molars of male mice are smaller than those of females. This
discrepancy could be accounted for by the difference between human amelogenin genes on both the X- and Y-chromosomes and mouse gene only on X-linked locus (Blecher, 1992). However, it was concluded that smaller molars in male mice than in females was due to hormonal rather than chromosomal factors. Comparison of the amelogenin primary sequences of other animals such as mouse (Lau et al., 1992), porcine (Yamakoshi et al., 1994) and rat (Bonass et al., 1994) provides evidence for extreme levels of conservation. Therefore, the amelogenin gene structures of those animals may be similar to those of human and bovine, but have not yet been isolated and characterized.

Mapping of human amelogenin gene loci on the X- and Y-chromosomes

Lau et al. (1989) mapped human amelogenin genes on the short arm of the X-chromosome in the p22.1-p22.3 region and in the pericentric region of the Y-chromosome. The amelogenin sequence on the human Y-chromosome was tentatively mapped onto the proximal long arm in the Yq11 region, on the basis of retention of the Y locus in a fibroblast cell line GM-1709 which was derived from a phenotypic female with an intact X-chromosome and an isochromosome for the long arm of the Y-chromosome (Lau et al., 1989). However, a controversy arose over the localization of human amelogenin gene on the Y-chromosome. Nakahori et al. (1991b) mapped the amelogenin sequence on the short arm of Y-chromosome using newly cloned Y-specific fragment probes. Bailey et al. (1992) also reported that the location of the Y amelogenin sequence was found in the short arm, Yp. According to Salido et al. (1992), aberrant Y-chromosome in GM-1709 was redefined as dicentric with a breakpoint in the p11.2 region, thus raising the possibility that the Y locus may be on the short arm of the Y-chromosome. They mapped the Y locus by fluorescent in situ hybridization using a Y-genomic clone as a probe. Metaphase Y-chromosomes with hybridization signals were examined and the amelogenin locus was found clearly on the short arm of the Y-chromosome in the p11.2 region.

Recently, precise mapping of human amelogenin gene (AMG) in the p22 region on the X-chromosome (Schaefer et al., 1993) and the homolog (AMGL) on the Y-chromosome (Foote et al., 1992) were reported. The genetic map of human amelogenin on the X- and Y-chromosomes is shown in Fig. 2.

With regard to mouse amelogenin, Chapman et al. (1991) analyzed the amelogenin locus on X-chromosome by the restriction site variation of genomic DNA. The amelogenin gene was established to be at the most distal locus in the genetic map of the mouse X-chromosome.

Heterogeneity of amelogenin mRNA due to alternative splicing

The heterogeneity of amelogenin proteins in the enamel matrix in terms of molecular size and amino acid composition has long been well recognized as described above. This is the result of a number of different processes. One reason is a heterogeneous population of mRNAs for amelogenin produced by alternative splicing. They are translated into a mixture of nascent amelogenin proteins and peptides.

The pluralism of amelogenin mRNA was first reported by Shimokawa et al. (1987a,b). Two species of mRNA encoding amelogenin in the extract from bovine ameloblast-rich tissue were identified and several other hybrid-selected translation products from cDNA clones were shown by immunoprecipitation. Also, four transcripts in different lengths having bovine amelogenin message were detected by Young et al. (1987). These data suggested that amelogenin heterogeneity arises from the production of multiple mRNAs by alternative splicing.

One example is the leucine-rich amelogenin peptide (LRAP) of 46 amino acid residues isolated from the developing bovine enamel (Fincham et al., 1981). Gibson et al. (1991a) showed that a precursor of LRAP (59-residue peptide) is initially translated from a mRNA produced by alternative splicing involving exon skipping of the amelogenin gene on the X-chromosome (Fig. 3). The 46 residue LRAP is then generated by proteolytic cleavage after being secreted into the matrix. In the case of rat, Northern blot analysis of the incisor mRNA showed four amelogenin transcripts at secretory stage (DenBesten and Li, 1992). In mouse, a total of five cDNAs derived from alternatively spliced amelogenin mRNA have been cloned and sequenced by Lau et al. (1992).

While structures of human and bovine amelogenin gene on the X- and Y-chromosomes have well been characterized as mentioned above, patterns of alternative splicing are revealed in comparison to the sequences of various mRNAs. From the tooth bud of a human male fetus, Salido et al. (1992) extracted cellular RNA which was reverse-transcribed into cDNA and amplified by PCR using a primer common to both the X- and Y-chromosomes. Reverse transcriptase PCR amplification of the 5' portion of the amelogenin transcripts revealed several alternatively spliced products (Fig. 3). The splicing pattern was different between the X- and Y-derived mRNAs. Three different species of X-derived cDNA clones resulting from alternative splicing were isolated.

Exon 4 of amelogenin gene on the human X-chromosome was present in only one of twenty-eight X-derived cDNA clones isolated. The majority of the cDNA clones contained the remaining six exons, while 16% of the cDNA clones lacked exon 3. In the case of Y-derived cDNA, exon 4 was lacking from all the clones isolated. In Fig. 3, alternative splicing patterns of amelogenin transcription are shown.
Fig. 3. Alternatively spliced mRNAs for human and bovine amelogenin. (A) Human amelogenin gene structure is shown at the top of the figure. Four kinds of amelogenin mRNAs (1)-(4) have been identified (Salido et al., 1992). A full-length mRNA (1) containing a total of 7 exons is rather exceptional and exon 4 is absent from the majority of the mRNAs (2)-(4). In other mRNAs, (3) and (4), exon 3 is lacking, too. (B) Bovine amelogenin gene structure is shown at the top of the figure. Four kinds of amelogenin mRNAs (1)-(4) have been identified (Gibson et al., 1992). (1) A major full-length mRNA. (2) Precursor of a peptide LRAP. (4) A peptide SAP2 named by Gibson et al. (1992). Part of exon 5 is removed by exon skipping in the mRNAs (3) and (4).

Promoter of amelogenin gene and regulation of expression

It has been well documented that a stage- and space-specific expression of amelogenin is found in ameloblasts during tooth development (Snead et al., 1988; Amar et al., 1989; Inage et al., 1989). Thus, there may be a strict regulatory mechanism in amelogenin gene expression. From the study of Salido et al. (1992), a putative TATA box at position -25 and a reversed CCAAT box at -51 were discovered in the 5' promoter region of human amelogenin genes on both X- and Y-chromosomes. Such a reversed CCAAT box has already been found at the 5' promoter region of another extracellular matrix protein, α1(IV) collagen (Klllen et al., 1988).

Chen et al. (1994) investigated the regulation of amelogenin gene expression during tooth development. They constructed transgenic mice in order to identify the segment of the amelogenin gene required for specific expression in ameloblasts. A 3.5 kb fragment of the putative promoter segment from the bovine X-chromosomal amelogenin gene was isolated. This fragment, which included a TATA box, an inverted CCAAT box, partial glucocorticoid receptor binding consensus sequences and the transcription initiation site, was linked to the β-galactosidase gene. This transgene expression vector was injected into fertilized mouse eggs. Newborn transgene-positive mice expressed β-galactosidase activity in developing teeth when treated with the X-gal substrate. Foci of molar ameloblasts were positive in newborn mice and the stain intensity and number of positive ameloblasts increased in 1-day and 2-day postnatal mice. Some of the adjacent stratum intermedium cells were also positive in the later stages. Thus, the enamel organ- and stage-specific expression of amelogenin was exhibited in transgenic animals. However, the specific promoter sequence for the regulation of amelogenin gene expression was not identified. Moreover, it was unable to detect blue stain in the incisal ameloblasts. The regulation mechanism of amelogenin expression may be different between the molar and incisor in rodents.

Phylogeny and evolution of amelogenin gene

As described above, it is well established that there are two amelogenin genes, one on the X- and the other on the Y-chromosome in some mammals, with a single gene residing exclusively on the X-chromosome in other vertebrates. Male and female genomic DNA samples from a variety of animals were investigated using the "Noah's ark blot" by Nakahori et al. (1991a) and also using the PCR method by Bailey et al. (1992). The results of the two reports are summarized in the Table.

It seems to be reasonable that the amelogenin genes reside on both the X- and Y-chromosomes in anthropoids (gorilla, chimpanzee and orangutan) as well as in man. Some old-world monkeys (Japanese monkey, rhesus monkey and crab-eating macaque) have amelogenin genes on the X- and Y-chromosomes, while other old-world monkeys (baboon, patas monkey, green monkey and talapoin) have no Y-homolog. A new-world monkey, tamarin, has only one amelogenin gene on the X-chromosome, while the capuchin monkey, another new-world variety, has both the X and Y genes. Only one amelogenin gene locus resides on the X-chromosome in the mouse and rat, but bovine genes are on the X- and Y-chromosomes. The data of chromosomal distribution of

<table>
<thead>
<tr>
<th>TABLE 1 AMELOGENIN GENE DISTRIBUTION ON SEX CHROMOSOMES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amelogenin genes on</td>
</tr>
<tr>
<td>X- and Y-chromosomes</td>
</tr>
<tr>
<td>Human</td>
</tr>
<tr>
<td>Gorilla</td>
</tr>
<tr>
<td>Chimpanzee</td>
</tr>
<tr>
<td>Orangutan</td>
</tr>
<tr>
<td>Old world</td>
</tr>
<tr>
<td>monkeys</td>
</tr>
<tr>
<td>Japanese monkey</td>
</tr>
<tr>
<td>Rhesus monkey</td>
</tr>
<tr>
<td>Crab-eating macaque</td>
</tr>
<tr>
<td>Talapoin</td>
</tr>
<tr>
<td>New world</td>
</tr>
<tr>
<td>monkeys</td>
</tr>
<tr>
<td>Tufted capuchin monkey</td>
</tr>
<tr>
<td>Cattle</td>
</tr>
<tr>
<td>Rat</td>
</tr>
<tr>
<td>Mouse</td>
</tr>
</tbody>
</table>

(Nakahori et al., 1991a; Bailey et al., 1992)
amelogenin gene(s) shown in the Table cannot simply be attributed to single gene duplication or translocation via phylogenetic lineage. Multiple duplication or deletion is required to conform to the phylogenetic distribution of the species.

It is of particular interest that the amelogenin gene homologs are located on the autosomes but not on sex chromosomes in the primate forms of mammals, monotremes (platypus) and marsupials (wallaby), according to the study of Watson et al. (1992). These authors determined that amelogenin gene homologs were located on chromosomes 5q and 1q in wallaby, and on chromosomes 1 and 2 in platypus (duckbill) by Southern blot analysis. The adult duckbill has no true teeth, although the hatching has an egg tooth and very young duckbills transiently possess teeth with typically mammalian characteristics. And this primitive egg-laying mammal has amelogenin gene-like loci on autosomes. However, the authors were not able to determine whether these homologous sequences are active amelogenin genes or pseudogenes. They surmised that the amelogenin gene region must have been translocated to X- and Y-chromosomes before the divergence of primes and other mammals.

Teeth are phylogenetically old structures and have existed in almost every recent species of the vertebrate classes. However, by immunohistochemical analysis, only enamels are present in lower vertebrates such as sharks, bony fish and larval amphibians, while both enamels and amelogenins are present in higher vertebrate and adult amphibian teeth (Herold et al., 1989). Lyngstadaas et al. (1990) investigated DNA sequences encoding amelogenin gene in fish and other vertebrate species. Hybridization assay using a mouse amelogenin cDNA probe was performed with restriction enzyme fragments of genomic DNA. Genomic DNA from wolfish (Anarrhichaslupus) hybridized with the mouse amelogenin probe under stringent conditions, exhibiting one band of 18 kb. This result suggested that a DNA sequence homologous to the mouse amelogenin gene is present even in the bony fish. However, it is not known whether the DNA sequence in this fish is expressed or not.

Kollar and Fisher (1980) reported that intraocular grafts of "chick epithelium" combined with mouse molar mesenchyme produced dental structures including perfectly formed crown with differentiated ameloblasts depositing enamel matrix. This report suggested retention of quiescent genetic information about tooth formation including amelogenin synthesis during evolution of birds. However, no hybridization bands were detected in the DNA sample from chicken in the study of Lyngstadaas et al. (1990). A loss of genetic coding for amelogenin in birds was indicated.

**Amelogenesis imperfecta**

Amelogenesis imperfecta is an inherited disease affecting the formation of tooth enamel in primary and permanent dentition. In general, there are two different phenotypes in clinical manifestations, the hypoplastic and the hypominalization. The former is characterized by an enamel that does not attain normal thickness during development, whereas in the latter the enamel is softer than normal. Genetically, amelogenesis imperfecta contains heterogeneous and autosomal dominant, autosomal recessive and X-linked forms with division of these categories according to the perceived phenotype (Witkop, 1989). Affected (hemizygous) males and (heterozygous) females in the same family have different clinical manifestations, with the teeth of heterozygous females exhibiting vertical markings of the enamel. Therefore, it was suggested that this type of amelogenesis imperfecta is caused by a structural alteration in the amelogenin gene.

Lagerstrom et al. (1991) reported a study of a Swedish family of X-linked amelogenesis imperfecta (AIH1). They found a deletion extending over 5 kb of the amelogenin gene in inherited males. Later, the deletion in the amelogenin gene was analyzed in detail by their group (Lagerstrom-Fermer et al., 1993). They isolated the breakpoints of a 5-kb deletion in the amelogenin gene on the basis of nucleotide sequence information located upstream of the lesion, using a PCR technique. The deletion removed five of the seven exons, spanning from the second intron to the last exon. Only the first two codons for the mature protein remained, consistent with the relatively severe phenotype of affected individuals. The mutation appeared to have arisen as an illegitimate recombination event.

Also, Aldred et al. (1992a) investigated a family with X-linked amelogenesis imperfecta having a nonsense mutation in exon 5 of the amelogenin gene. This mutation was a single base (C) deletion which altered the reading frame and introduced a TGA stop codon into the sequence of the amelogenin gene. Recently, an identical mutation in exon 5 of the amelogenin gene causing AIH1 was reported by Lench et al. (1994).

Three families with X-linked amelogenesis imperfecta were investigated by Aldred et al. (1992b) using polymorphic DNA markers flanking the position of amelogenin X-gene. Using two-point and multipoint linkage analyses, locations of the amelogenin gene or related mutations were identified. The study demonstrated that there were at least two different loci on the X-chromosome causing amelogenesis imperfecta, one on the distal short arm (AIH1) due to mutations in the amelogenin X-gene and another on the long arm (AIH3) as a result of alterations in a gene in the Xq22-q28 region.
Application of amelogenin gene detection to human sex test

A simple method of human sex determination is frequently required in forensic analyses or in the identification of human remains. We have developed a simple and rapid sex test by amplifying both X- and Y-specific amelogenin sequences. DNAs were extracted from male and female human blood samples and X- and Y-amelogenin genes were amplified by PCR using two common primers. Amplified products were electrophoresed in agarose gel. The result is shown in Fig. 4. Both amelogenin gene X-specific 1.5 kbp and Y-specific 1.3 kbp fragments were detected in male samples and only the X-specific fragment (1.5 kbp) was observed in female samples. This method may also have medical applications in characterizing sex chromosome abnormalities such as Klinefelter's syndrome (XXY). Similar methods were reported by Akane et al. (1992) and by Sullivan et al. (1993).

Acknowledgments

We are grateful to Drs. Yukata Nakahori and Yuzo Takagi for their helpful suggestions and valuable discussions during the preparation of this manuscript. We also thank Dr. Hisako Katano for her help in preparing the manuscript and figures.

References


Amelogenin gene


