Eur J Oral Sci 2000; 108: 303–310
Printed in UK. All rights reserved

EUROPEAN JOURNAL OF
ORAL SCIENCES
ISSN 0909-8836

Cloning, characterization and immunolocalization of human ameloblastin

Mary MacDougall[1], Darrin Simmons[1], Ting Ting Gu[1], Kristina Forsman-Semb[2], Carina Kärrman Mårdh[2], Michael Mesbah[5], Nadine Forest[5], Paul H. Krebsbach[3], Yoishi Yamada[4], Ariane Berdal[5]

[1]University of Texas Health Science Center at San Antonio, Dental School, San Antonio, Texas, USA, [2]Departments of Clinical Genetics/Applied Cell and Molecular Biology, University of Umeå, Umeå, Sweden, [3]University of Michigan, School of Dentistry, Ann Arbor, Michigan, USA, [4]National Institute for Dental and Craniofacial Research, Bethesda, Maryland, USA, [5]Laboratory Biology-Odontology, Dental Faculty, University Paris VII, Cordeliers Institute, Paris, France

MacDougall M, Simmons D, Gu TT, Forsman-Semb K, Kärrman Mårdh C, Mesbah M, Forest N, Krebsbach PH, Yamada Y, Berdal A. Cloning, characterization and immunolocalization of human ameloblastin. Eur J Oral Sci 2000; 108: 303–310. © Eur J Oral Sci, 2000

Amelogenesis imperfecta is a broad classification of hereditary enamel defects, exhibiting both genetic and clinical diversity. Most amelogenesis imperfecta cases are autosomal dominant disorders, yet only the local hypoplastic form has been mapped to human chromosome 4q between D4S2421 and the albumin gene. An enamel protein cDNA, termed ameloblastin (also known as amelin and sheathlin), has been isolated from rat, mouse and pig. Its human homolog has been mapped to chromosome 4q21 between markers D4S409 and D4S400, flanking the local hypoplastic amelogenesis imperfecta critical region. Therefore, ameloblastin is a strong candidate gene for this form of amelogenesis imperfecta. To facilitate genetic studies related to this dental disease, we isolated and characterized a human ameloblastin cDNA. A human third molar cDNA library was screened and two ameloblastin clones identified. Nucleotide sequencing of these cDNAs indicated alternative splicing of the putative open reading frame, use of different polyadenylation signals, and a high degree of similarity to reported rat, mouse and porcine cDNAs. Immunohistochemistry studies on embryonic human teeth using an antibody to recombinant ameloblastin indicated ameloblastin expression by ameloblasts with localization in the enamel matrix associated with the sheath structures.

Dr. Mary MacDougall, University of Texas Health Science Center at San Antonio, Dental School, 7703 Floyd Curl Drive, San Antonio, TX 78229–3900, USA

Telefax: +1–210–5676603
E-mail: MacDougall@UTHSCSA.edu

Key words: ameloblastin; amelogenesis imperfecta; ameloblasts; alternative splicing; sheathlin; amelin

Accepted for publication April 2000

Amelogenesis imperfecta (AI) represents a broad spectrum of genetic diseases affecting enamel formation in both the primary and permanent dentitions. AI has been classified into 14 different subtypes according to the clinical appearance of the enamel and the Mendelian mode of inheritance (1). The combined prevalence of all forms of AI has been reported as 1:14,000 in the US (1), 1:8,000 in Israel (2), 1:4,000 in Sweden (3) and 1:700 in the Västerbotten county of Sweden (4). The enamel abnormalities have been categorized into three major groups, hypocalcified, hypomaturation, or hypoplastic, and the inheritance patterns reported include autosomal dominant or recessive, and X-linked dominant or recessive (1).

The autosomal dominant forms of AI are the most prevalent forms, representing over 95% of all reported cases and have been shown to be genetically heterogeneic (5). To date, a disease locus has been established for only one type of autosomal dominant AI, the local hypoplastic form (AIH2). AIH2 was initially mapped in three Swedish families to human chromosome 4q11–q21 within a 17.6 cM region (6). This locus was further refined to a 4-Mb region located between D4S2421 and the albumin gene by recombination breakpoint mapping using six Swedish families (7).

Ameloblastin (Ambn; also known as amelin and sheathlin) was initially isolated as the most abundant novel transcript from 400 random rat incisor cDNAs (8). The clone was 1925 bp, and contained multiple putative open reading frames (ORFs) with the longest encoded a 422 amino acid acidic protein with a reported pI of 5.54 and an M_r of 45 kDa. Using a similar approach, other investigators independently cloned two rat transcripts, termed amelin 1 and 2, which code for the same protein (9). *In situ* hybridization studies have shown Ambn to be expressed at high levels by the enamel-producing ameloblast cells (8–11) with

moderate expression by the Hertwig epithelial root sheath (12, 13) and transient expression by preodontoblasts (13). Immunohistochemistry has demonstrated localization to the ameloblast cell layer, with intense staining in the distal cytoplasm of ameloblasts, the Tomes processes of secretory ameloblasts, and within the matrix at the dentino-enamel junction (DEJ) (8, 10, 13–15) with limited expression in early preodontoblasts (13). Rat and mouse Ambn cDNAs (8, 9, 16) showed homology with the sequence of a 15-kDa enamel matrix sheath protein (14) identified by biochemical analysis of enamel matrix, as well as the corresponding porcine cDNA termed sheathlin (17).

The mouse *Ambn* locus has been mapped to chromosome 5, distal to the *Kit* gene (8). More precise mapping of the human AMBN locus using two-color fluorescence *in situ* hybridization has placed the gene on the long arm of human chromosome 4, between the sequence tagged site (STS) markers D4S400 and D4S409 (18). This region, 4q13–q21, has been shown to contain the gene locus for AIH2 (7). Furthermore, a human AMBN genomic P1 artificial chromosome (PAC) clone was shown to contain three STS markers, D4S2604, D4S2670 and D4S2609, previously placed within the critical AIH2 region as defined by haplotyping using six Swedish families (18). Therefore, AMBN appears to be a strong candidate gene for this form of AI.

To determine the genomic organization of human AMBN and thereby facilitate genetic studies related to AIH2, we have isolated and characterized human Ambn cDNA clones. In addition, the spatial expression pattern of the translated protein during early human tooth development has been analyzed.

Material and methods

Labeled Ambn probe

The rat Ambn cDNA (1.9 kb) (8) was labeled to a high specific activity using the Prime-It II random priming kit (Stratagene, La Jolla, CA, USA) and [α-^{32}P]dCTP. The probe was purified using a Sephadex G-50 column as previously described (16).

Human tooth cDNA library screening

A previously characterized human third molar cDNA library (19) was plated with host cell XL1-Blue MRF′ at a density of 5×10^4 plaques per 150-mm LB plate. Plaques were transferred to duplicate nylon filters, denatured, neutralized and rinsed as previously described (19). Damp filters were ultraviolet (UV)-cross-linked (Stratalinker; Stratagene) using the auto setting, rinsed in distilled water, and prehybridized for 2 h at 68°C in Quik-Hyb hybridization solution (Stratagene) plus 100 μg/ml of sheared salmon sperm DNA.

Filters were hybridized with labeled rat Ambn as previously described (16). Blotted filters were placed on Biomax MR-1 film (Kodak, Rochester, NY, USA) overnight at −70°C with intensifying screens. Putative positive clones (~25 plaques) were selected and confirmed by polymerase chain reaction (PCR) amplification with an Ambn-specific primer set as previously described (16). Samples which were positive for Ambn amplicons were plated at a density of approximately 100 plaques/100 mm plate and a secondary screen performed as described above using a mouse full length Ambn cDNA probe (16). Six tertiary screened plaques were picked, and tested for insert length using PCR with flanking vector-specific primers generated to the T3 and T7 RNA polymerase promoter sequences. These clones were rescued to Bluescript SK(−) phagemids using the Ex-assist Helper Phage kit (Stratagene).

DNA sequence and analysis

DNA sequence was determined using automated sequencing (Applied Biosystems, Model 370A) at the Human Genome DNA Core Laboratory (University of Texas Health Science Center, San Antonio, TX, USA). Unresolved DNA sequence was determined manually using Sequenase 2.0 kit according to the suppliers instructions (Amersham USB, Arlington Heights, IL, USA). A contiguous DNA sequence was generated using the Assemblylign program (Kodak) using data from multiple overlapping sequence reactions. Open reading frame, protein subsequence, and amino acid content were analyzed using MacVector 6.0 (Oxford Molecular, Campbell, CA, USA). Protein alignments were generated using the Clustal W (1.4) feature of MacVector 6.0 using default program settings. Isoelectric points, and isotope averaged molecular weight were calculated using MacBiospec (Perkin-Elmer Sciex, Thornhill, Ontario, Canada).

Immunohistochemistry

Mandibles from fetuses ($n = 4$; 21–26 wk of gestation) were obtained after pathological diagnosis for medical reasons, with parental informed signed consent. This study was performed on a limited collection under the approved direction of the French National Ethics committee. Samples were fixed in buffered formaldehyde, decalcification with nitric acid (Labonor, Paris, France), dehydrated, paraffin-embedded and histological sections (5 μm) prepared. Primary and secondary antibodies, as

well as conjugated peroxidase, were diluted in a solution of 0.05 M Tris NaCl, pH 7.6 containing 1% normal goat serum (NGS; Amersham, les Ulis, France). Washes were performed in 0.05% Tris NaCl solution, pH 7.6, containing 1% bovine serum albumin (Sigma, le Verpilliere, France). Endogenous peroxidases in the tissues were inactivated by treatment with 3% H_2O_2 for 15 min. The tissue was then blocked with 10% NGS for 24 h at 4°C. The primary antibody (1:5,000 dilution) was incubated for 18 h at 4°C, biotinylated secondary antibody (1:800, Sigma) for 30 min, and streptavidin peroxidase (1:300, Vector, Paris, France) for 30 min. Diaminobenzidine was used as the peroxidase substrate (Sigma). The sections were counterstained with Harris hematoxylin (Shandon, Paris, France), dehydrated and mounted. Micrographs were taken with a Zeiss Orthoplan microscope.

Results

In order to facilitate determination of AMBN gene structure and mutational analysis for AIH2, we screened a human third molar cDNA library using full-length rat and mouse Ambn cDNA clones. Initial screening of approximately 5×10^4 recombinants resulted in 25 primary clones. Small aliquots of these primary isolates were initially confirmed by PCR amplification using an Ambn primer set based rat cDNA sequence (rAmbnY274S 5′AGATTCCACTTTTCAAAATG3′ and rAmbnY274AS 5′TTCTGGGGTGTGACCCTG3′). Positive isolates were then subjected to a single secondary screening at low density using radiolabeled full-length mouse Ambn cDNA clone (16). Isolated positive single clones were again amplified by PCR using vector-specific primers to the insert flanking T3 and T7 RNA polymerase promoter segments of the Bluescript SK + vector flanking the insert in order to determine the approximate cDNA insert size. Six clones containing the largest inserts were chosen for further DNA sequence analysis. These human Ambn clones were rescued to phagemid, and the DNA sequence of both strands determined. The longest clone (human 1A2) was 1983 bp exclusive of the poly A tail, and contained an open reading frame encoding 447 amino acids (Genebank accession number AF209780). This clone contained a long untranslated 3′ region with three potential polyadenylation signals (AATAAA) (Fig. 1). The putative start site (ATG) was at base 84, with a potential signal peptide of 26 amino acids. Selection of this initial methionine (Met) as the start codon was based on previously published rat and mouse Ambn sequences (8, 16). Additional clones identified varied by the inclusion or exclusion of a 45-bp segment encoding 15 amino acids (98–112) and the length of the 3′ untranslated region (arrow, Fig. 1) by the use of the second or third polyadenylation signals beginning at nucleotides 1615 or 1963, respectively. The 15-amino acid (45-bp) segment YEYSLPVHPPPLPSLQ of the human Ambn clone was 100% conserved with a segment shown to be alternately spliced at the mRNA level in rat and mouse Ambn and in pig sheathlin (8, 4, 17). This segment contained 8.0% of the total proline residues found in the mature human protein.

Alignment of the predicted amino acid sequence for human Ambn with published sequences for pig, mouse and rat cDNA clones is shown in Fig. 2. This data indicates that Ambn is a highly conserved protein between the four species studied. The human Ambn amino acid sequences was more similar to that of the pig sheathlin than the two reported rodent clones. Between human and pig, we found 65% identity with an additional 9% conservative substitution. The unique five-amino acid deletion (AQGMA) found in pig sheathlin beginning at amino acid 35 is also absent in the human Ambn clone, as well as a single proline after amino acid 165. In addition, three single amino acid insertions of glycine (aa 322), phenalanine (aa 408), and glutamic acid (aa 436), and one two-amino acid insert of lysine–glycine (aa 371–372) are found in the human Ambn sequence. These insertions are again more similar to the sequence of pig sheathlin. Most interesting is the insertion of 26 amino acids found unique to the human AMBN cDNA that is not contained in any other species. This segment SLPGMDFPDPQGPSLPGLDFADPQGS) contains an almost perfect double repeat of a 13 amino acid sequence (amino acids 190–215, underlined). The potential post-translational modification of human Ambn by casein kinase II at residue 262 is conserved, as is the tyrosine kinase site at residue 98.

The predicted mature human protein would have a calculated isotope-averaged molecular weight of 45.6 kDa and would be weakly acidic with a pI of 6.1. Molecular weight, pI, and amino acid content of human Ambn is similar to rat, mouse and porcine proteins (Table 1). Human protein is rich in proline (15%), glycine (9%), alanine (7.6%), and leucine (9.7%), with no cysteine found as previously reported for other species.

Immunolocalization studies were performed using a previously characterized Ambn antibody generated against a recombinant rat COOH region protein (aa 175–348) and therefore able to recognize both isoforms (8). This antibody has been tested for specificity by Western blot using

```
CACGCACAGTCCTGAAAAAAATTTTAATCTTCTTTTCTTAGAACTATCTTGGTTGGCATCATCAGGCCCTGAGAGCACAGTGC   83

ATG TCA GCA TCT AAG ATT CCA CTT TTC AAA ATG AAG GAC CTG ATA CTG ATC CTA TGC CTC CTG  146
Met Ser Ala Ser Lys Ile Pro Leu Phe Lys Met Lys Asp Leu Ile Leu Ile Leu Cys Leu Leu   21

GAA ATG AGT TTT GCA GTG CCG TTC TTT CCT CAG CAA TCT GGA ACA CCG GGT ATG GCT AGT TTG  209
Glu Met Ser Phe Ala Val Pro Phe Phe Pro Gln Gln Ser Gly Thr Pro Gly Met Ala Ser Leu   42

AGC CTT GAG ACA ATG AGA CAG TTG GGA AGT CTG CAG AGA TTA AAC ACA CTT TCT CAG TAT TCT  272
Ser Leu Glu Thr Met Arg Gln Leu Gly Ser Leu Gln Arg Leu Asn Thr Leu Ser Gln Tyr Ser   63

AGA TAC GGC TTT GGA AAA TCA TTT AAT TCT TTG TGG ATG CAC GGT CTC CTC CCA CCA CAT TCC  335
Arg Tyr Gly Phe Gly Lys Ser Phe Asn Ser Leu Trp Met His Gly Leu Leu Pro Pro His Ser   84

TCT CTT CCA TGG ATG AGG CCA AGA GAA CAT GAA ACT CAA CAG TAT GAA TAT TCT TTG CCT GTG  398
Ser Leu Pro Trp Met Arg Pro Arg Glu His Glu Thr Gln Gln Tyr Glu Tyr Ser Leu Pro Val  105

CAT CCC CCA CCT CTC CCA TCA CAG CCA TCC TTG AAG CCT CAA CAG CCA GGA CTG AAA CCT TTT  461
His Pro Pro Pro Leu Pro Ser Gln Pro Ser Leu Lys Pro Gln Gln Pro Gly Leu Lys Pro Phe  126

CTC CAG TCT GCT GCT GCA ACC ACC AAC CAG GCC ACA GCA CTG AAA GAA GCA CTT CAG CCT CCA  524
Leu Gln Ser Ala Ala Ala Thr Thr Asn Gln Ala Thr Ala Leu Lys Glu Ala Leu Gln Pro Pro  147

ATT CAC CTG GGA CAT CTG CCC TTG CAG GAA GGA AAA CTG CCT CTG GTT CAG CAG CGG GTG GCA  587
Ile His Leu Gly His Leu Pro Leu Gln Glu Gly Lys Leu Pro Leu Val Gln Gln Arg Val Ala  168

CCA TCA GAT AAG CCA CCA AAG CCT GAG CTC CCA AGA GTA GAT TTT GCC GAT CCA CAA GGT CCA  650
Pro Ser Asp Lys Pro Pro Lys Pro Glu Leu Pro Arg Val Asp Phe Ala Asp Pro Gln Gly Pro  189

TCA CTC CCA GGA ATG GAT TTT CCT GAT CCA CAA GGT CCA TCA CTC CCA GGA TTG GAT TTT GCT  713
Ser Leu Pro Gly Met Asp Phe Pro Asp Pro Gln Gly Pro Ser Leu Pro Gly Leu Asp Phe Ala  210

GAT CCA CAA GGT TCA ACA ATT TTC CAA ATA GCC CGT TTG ATT TCT CAC GGA CCA ATG CCA CAA  776
Asp Pro Gln Gly Ser Thr Ile Phe Gln Ile Ala Arg Leu Ile Ser His Gly Pro Met Pro Gln  231

AAT AAA CAA TCT CCA CTT TAT CCA GGA ATG TTG TAC GTG CCT TTT GGA GCA AAT CAA TTG AAT  839
Asn Lys Gln Ser Pro Leu Tyr Pro Gly Met Leu Tyr Val Pro Phe Gly Ala Asn Gln Leu Asn  252

GCC CCT GCC AGA CTT GGC ATC ATG AGT TCA GAA GAA GTG GCA GGC GGG AGA GAA GAC CCA ATG  902
Ala Pro Ala Arg Leu Gly Ile Met Ser Ser Glu Glu Val Ala Gly Gly Arg Glu Asp Pro Met  273

GCC TAT GGA GCC ATG TTT CCA GGA TTT GGA GGC ATG AGG CCC GGC TTT GAG GGA ATG CCC CAC  965
Ala Tyr Gly Ala Met Phe Pro Gly Phe Gly Gly Met Arg Pro Gly Phe Glu Gly Met Pro His  294

AAC CCA GCT ATG GGC GGT GAC TTC ACT CTG GAA TTT GAC TCC CCA GTG GCT GCC ACC AAA GGC 1028
Asn Pro Ala Met Gly Gly Asp Phe Thr Leu Glu Phe Asp Ser Pro Val Ala Ala Thr Lys Gly  315

CCT GAG AAC GAA GAA GGA GGT GCA CAA GGC TCC CCT ATG CCG GAG GCC AAC CCA GAC AAT CTA 1091
Pro Glu Asn Glu Glu Gly Gly Ala Gln Gly Ser Pro Met Pro Glu Ala Asn Pro Asp Asn Leu  336

GAA AAC CCA GCT TTC CTT ACA GAG CTA GAA CCT GCT CCC CAC GCA GGG CTC CTT GTT TTC CTT 1154
Glu Asn Pro Ala Phe Leu Thr Glu Leu Glu Pro Ala Pro His Ala Gly Leu Leu Val Phe Leu  357

AAG GAT GAC ATT CCC GGC CTG CCA AGG GAC CCT TCA GGG AAG ATG AAG GGA CTC CCC AGC GTC 1217
Lys Asp Asp Ile Pro Gly Leu Pro Arg Asp Pro Ser Gly Lys Met Lys Gly Leu Pro Ser Val  378

ACC CCA GCA GCT GTT GAC CCA CTG ATG ACC CCT GAA TTA GCT GAT GTT TAT AGG ACC TAC GAT 1280
Thr Pro Ala Ala Val Asp Pro Leu Met Thr Pro Glu Leu Ala Asp Val Tyr Arg Thr Tyr Asp  399

GCT GAC ATG ACC ACA TCC GTG GAT TTT CAG GAA GAA GCA ACC ATG GAT ACC ACG ATG GCC CCA 1343
Ala Asp Met Thr Thr Ser Val Asp Phe Gln Glu Glu Ala Thr Met Asp Thr Thr Met Ala Pro  420

AAC TCT TTG CAA ACA TCC ATG CCA GGA AAC AAA GCC CAG GAG CCC GAG ATG ATG CAT GAC GCA 1406
Asn Ser Leu Gln Thr Ser Met Pro Gly Asn Lys Ala Gln Glu Pro Glu Met Met His Asp Ala  441

TGG CAT TTC CAA GAG CCC TGACAGCTCTAAGATATTAGCTACTTTCTGTATGCACAAGCTTCCCAGCTTTGTCCCCA 1483
Trp His Phe Gln Glu Pro  *                                                           447

CAGTGTACCTTTTTGCTAAAACACTTATTACCCTTCTGCAGCAAAGGCATTAAAAGCGCTAAGCATATATTAATAAATGCAAG 1566
TGGCTAGAAATAGTGTAGGTCCCCTTCTTGCTTTCAATATCTTGTTGAAATAAAATGTGTCAATTGTC⇓TCTGTGATTTAGAA 1648
ACACTATTAATAACATCAGAGCAAGGTTCTAAGGGTCTCAGCATTTGATCATCACTTTTTACTTAGCTGTCTTAAGCATTATA 1731
GAATTTCTCTTACCAGCATGACACTATTATATTCAGGAAACATGGCACTGCTTTTTTCTCTAAGCAAAGGCAAATATCCTCAT 1814
AATTCTAAGCTAATTCATTTAAACTTTATTAAATGGGGATTGGTGGGAAACTCCCGACTGGTATTACTGGGTTTGGTATAATT 1897
TAATTTAAAATTCTCATTTATAGAGTATTTTATTTAATCTAGTATAAAAAATCTAGCCTAATTTAAATAAAGAATTTTTCTCA 1980
CTGAAAAAAAAAAAAAAAAAAAAAAAAA                                                         2008
```

```
Human         1 MSASKIPLFKMKDLILILCLLEMSFAVPFFPQQSG-----TPGMASLSLE  45
Pig           1 MPALKIPLFKMKDMVLILCLLKMSSAVPAFPRQPG-----TPGVASLSLE  45
Mouse         1 MSASKIPLFKMKGLILFLSLVKMSLAVPAFPQQPGAQGMAPPGMASLSLE  50
Rat           1 MSASKIPLFKMKGLLLFLSLVKMSLAVPAFPQRPGGQGMAPPGMASLSLE  50
                * * ******** ..* *.*. ** *** **.. *      **.******

Human        46 TMRQLGSLQRLNTLSQYSRYGFGKSFNSLWMHGLLPPHSSLPWMRPREHE  95
Pig          46 TMRQLGSLQGLNMLSQYSRFGFGKSFNSLWMHGLLPPHSSFQWMRPREHE  95
Mouse        51 TMRQLGSLQGLNALSQYSRLGFGKALNSLWLHGLLPPHNSFPWIGPREHE 100
Rat          51 TMRQLGSLQGLNALSQYSRLGFGKALNSLWLHGLLPPHNSFPWIGPREHE 100
                ********* ** ****** ****. ****.******* *  *. *****

Human        96 TQQYEYSLPVHPPPLPSQPSLKPQQPGLKPFLQSAAATTNQATALKEALQ 145
Pig          96 TQQYEYSLPVHPPPLPSQPSLQPQQPGQKPFLQPTVVTSIQNPVQKGVPQ 145
Mouse       101 TQQYEYSLPVHPPPLPSQPSLQPHQPGLKPFLQPTAATGVQVTPQKPGPQ 150
Rat         101 TQQYEYSLPVHPPPLPSQPSLQPHQPGLKPFLQPTAATGVQVTPQKPGPH 135
                ********************.*.*** ***** .  *  *    *   .

Human       146 PPIHLGHLPLQEGELPLVQQ-QVAPSDKPPKPELPGVDFADPQGPSLPGM 194
Pig         146 PPIYQGHPPLQQVEGPMVQQ-QVAPSEKPPEAELPGLDFADPQDP----- 189
Mouse       151 PPMHPGQLPLQEGELIAPDEPQVAPSENPPTPEVPIMDFADPQFP----- 195
Rat         136 PPMHPGQLPLQEGELIAPDEPQVAPSENPPTPEVPIMDFGDPQFP----- 180
                **.  *. ***. *     . *****. **  *.* .** *** *

Human       195 DFPDPQGPSLPGLDFADPQGSTIFQIARLISHGPMPQNKQSPLYPGMLYV 244
Pig         190 ----------------------SMFPIARLISQGPVPQDKPSPLYPGMFYM 218
Mouse       196 ----------------------TVFQIARSISRGPMAHNKASAFYPGMFYM 224
Rat         181 ----------------------TVFQIAHSLSRGPMAHNKVPTFYPGMFYM 209
                                    ..* **. .*.**. . *    **** *.

Human       245 PFGANQLNAPARLGIMSSEEVAGGREDPMAYGAMFPGFGGMRPGFEGMPH 294
Pig         219 SYGANQLNSPARLGILSSEEMAGGRGGPLAYGAMFPGFGGMRPNLGGMPP 268
Mouse       225 SYGANQLNAPARIGFMSSEEMPGERGSPMAYGTLFPRFGGFRQTLRRLNQ 274
Rat         210 SYGANQLNAPGRIGFMSSEEMPGERGSPMGYGTLFPGYGGFRQTLRGLNQ 259
                 .******.* *.* .****. * *  *. **..** .** *     .

Human       295 NPAMGGDFTLEFDSPVAATKGPENEEGGAQGSPMPEANPDNLENPAFLTE 344
Pig         269 NSAKGGDFTLEFDSPAAGTKGPEKGEGGAEGSPVAEANTADPESPALFSE 318
Mouse       275 NSPKGGDFTVEVDSPVSVTKGPEKGEG-PEGSPLQEANPGKRENPALLSQ 323
Rat         260 NSPKGGDFTVEVDSPVSVTKGPEKGEG-PEGSPLQEPSPDKGENPALLSQ 308
                *   *****.* *** . *****  **  .***. *      * **  ..

Human       345 LEPAPHAGLLALPKDDIPGLPRDSPSGKMKGLPSVTPAAVDPLMTPELAD 394
Pig         319 VASGVLGGLLANPKGKIPNLAR-GPAGRSRGPPGVTPADADPLMTPGLAD 368
Mouse       324 MAPGAHAGLLAFPNDHIPSMAR-GPAGQR--LLGVTPAAADPLITPELAE 371
Rat         309 IAPGAHAGLLAFPNDHIPNMAR-GPAGQR--LLGVTPAAADPLITPELAE 356
                .      **** *   ** . *  *.*.      ****  ***.** **.

Human       394 VYRTYDADMTTSVDFQEEATMDTTMAPNSLQTSMPGNKAQEPEMMHDAWH 444
Pig         368 AYETYGADETTTLGLQEEMTMDSTATPYSEHTSMPGNKAQQPQIKRDAWR 418
Mouse       371 VYETYGADVTTPLG-DGEATMDITMSPDTQQPLLPGNKVHQP-QVHNAWR 419
Rat         356 VYETYGADVTTPLG-DGEATMDITMSPDTQQPPMPGNKVHQP-QVHNAWR 404
                * ** ** ** .    * *** * .* . .  .**** ..*   . **.*

Human       444 FQEP 447
Pig         418 FQEP 421
Mouse       419 FQEP 422
Rat         404 FQEP 407
                ****
```

Fig. 2. Amino acid alignment using the Clustal W (1.4) software of human, pig sheathlin, mouse and rat Ambn. The blosum identity matrix was used, with an open gap penalty equal to 10 and an extended gap penalty equal to 0.1. The numbering of the amino acids for each species is shown on the right. Identities are denoted by an asterisk (*), conserved similarities by a dot (·), and deletions by a dash (-). The potential NH_2-terminus is indicated by an arrow and the conserved alternative splice segment is shown in bold.

Fig. 1. Composite DNA sequence of human Ambn clones and the deduced amino acid sequence (Genebank accession number AF209780). The numbering of protein and DNA sequence is shown on the right. The signal peptide is underlined with a single line, the stop codon is noted by an asterisk. Three polyadenylation signals (AATAAA) are shown in bold, and the 45-bp (15 amino acid) segment which is alternatively spliced is bold and bold underlined.

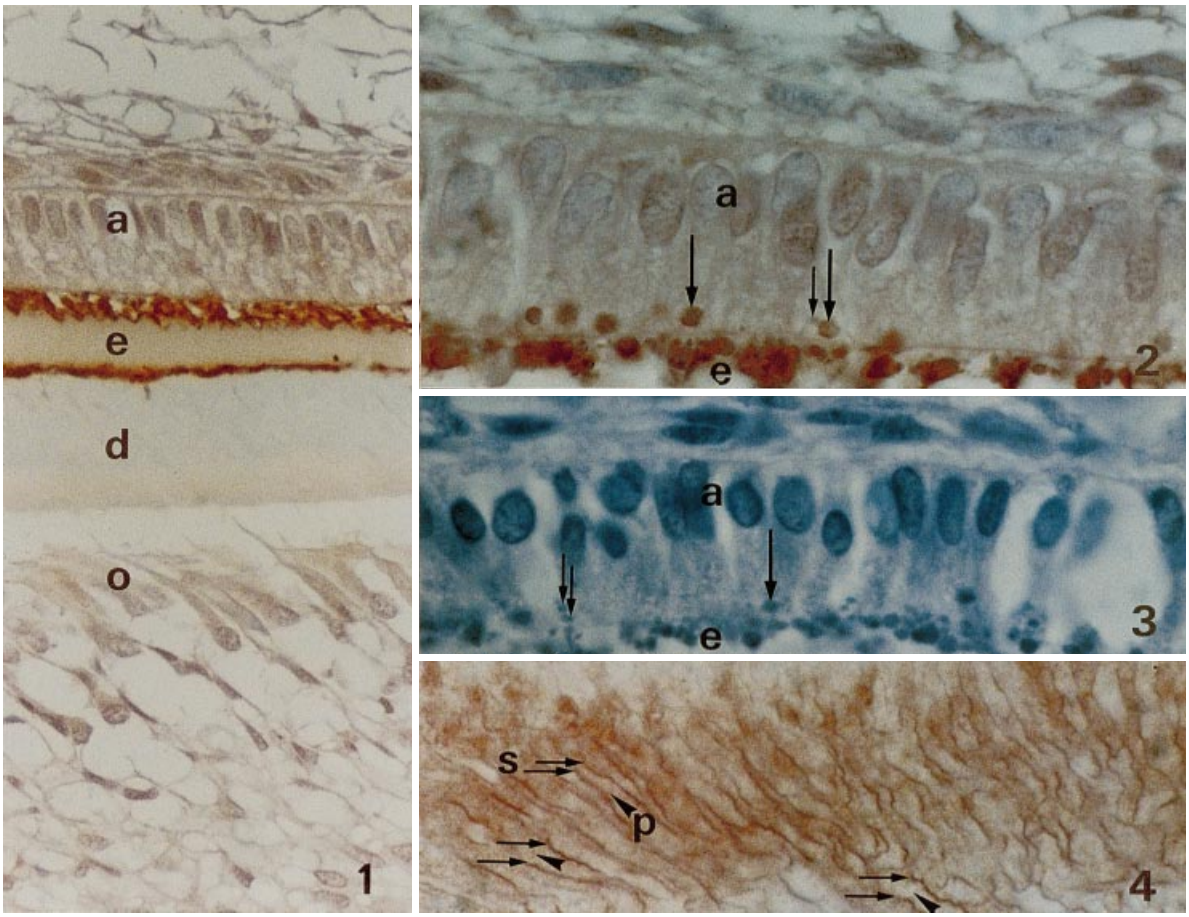

Fig. 3. Immunolocalization of Ambn in a pre-natal human developing tooth germ at 26 weeks. Positive immunoperoxidase staining using a 1:5000 dilution of the anti-ameloblastin antibody is shown in panels 1 (40 ×), 2 (100 ×) and 4 (100 ×). No staining was seen using a 1:5000 dilution of control IgGs (panel 3, 100 ×). Ambn immunoreactivity appears located intracellularly within the distal part of the ameloblast cell body (panels 1 and 2). Extracellularly, positive labeling is seen within the newly forming enamel next to the ameloblasts and distally at the enamel dentinal junction. Notably, light, sporadic staining of the distal ends of the dentinal tubules could be seen (panel 1). At high magnification (100 ×), within the ameloblasts staining appeared associated with distal secretory granules (panel 2, long arrows). In the extracellular matrix, a more intense staining was found at the prism borders referred to as ''enamel sheaths'', while lighter more diffuse staining was observed in the remainder of the enamel matrix. Abbreviations: a, ameloblasts; e, enamel; d, dentin; o, odontoblasts; s, sheaths; and p, prisms.

recombinant Ambn protein, extracts of various tissues and enamel extracellular matrix (Y. Yamada, unpublished data). At the prenatal stages examined, Ambn protein was only evident by immunocytochemistry within developing tooth organs. Ambn immunoreactivity first appeared in the secretion stage ameloblasts. Inside the cell compartment, Ambn epitopes were exclusively located inside large, rounded structures at the distal end of the cell body. Extracellularly, immunoreactive enamel was present nearby the secretory pole of ameloblasts and also in the vicinity of the dentinal-enamel junction. In the deep layers of forming enamel, immunoreactivity varied, the prism sheaths were associated with highest apparent concentrations with diffused staining between these structures. Faint immunostaining was also present in the predentin region and within the developing odontoblast cell layer facing mantle dentin. However, staining was not detectable within forming bone, Meckels cartilage, buccal epithelium, muscles, or salivary glands.

Discussion

We have identified six human Ambn cDNA clones, exhibiting alternative splicing and the use of two different polyadenylation signals. Analysis of the isolated clone showed there was no direct correlation with the inclusion or exclusion of the 45-bp (15 aa) alternatively spliced segment with utilization of a particular polyadenylation signal. The two alternatively spliced transcripts we report agree with the Northern blot findings of two rat Ambn transcripts (1.6 and 2.0 kb) (8). The amino acid content of high proline (15%), glycine (10%), alanine (9%), and leucine (9%) for human Ambn

Table 1

Comparison of the amino acid compositions of Human Ameloblastin clone IA2 with mouse Ambn, rat Ambn and porcine sheathlin. Analysis is for the secreted protein lacking the signal peptide using MacBiospec software

	Human ameloblastin		*Mouse ameloblastin*		*Rat ameloblastin*		*Pig sheathlin*	
Amino acid	Res/Mol	Res/1000	Res/Mol	Res/1000	Res/Mol	Res/1000	Res/Mol	Res/1000
Alanine	32	76	34	86	32	81	31	78
Arginine	13	31	15	38	11	28	14	35
Asparagine	13	31	14	35	14	35	11	28
Aspartic acid	20	48	11	28	12	30	13	33
Cysteine	0	0	0	0	0	0	0	0
Glutamine	28	67	31	78	30	76	31	78
Glutamic acid	25	59	23	58	23	58	23	58
Glycine	38	90	39	98	41	104	46	116
Histidine	11	26	11	28	13	33	6	15
Isoleucine	6	14	9	23	8	20	7	18
Leucine	41	97	36	91	35	88	33	84
Lysine	13	31	9	23	9	23	12	30
Methionine	22	52	15	38	16	40	16	41
Phenylalanine	20	48	16	40	15	38	15	38
Proline	63	150	61	154	63	159	59	149
Serine	29	69	25	63	24	61	31	78
Threonine	20	48	19	48	20	51	18	46
Tryptophan	3	7	3	8	3	8	3	8
Tyrosine	9	21	9	23	10	25	11	28
Valine	15	36	16	40	17	43	15	38
Total	421	1000	396	1000	396	1000	395	1000
pI	6.1		6.5		6.3		6.6	
M_r	45,595		42,521		42,358		42,010	

remains consistent with that found in other species (8, 4, 16) as does the calculated pI of 6.1.

The presence of two alternatively spliced transcripts of human Ambn is comparable to that reported for mouse Ambn (16) and pig sheathlin sequence (17). The alternatively spliced 45-bp segment, amino acid sequence YEYSLPVHPPP-LPSQ, is 100% conserved in all species studied to date. This homology across vertebrate evolution suggests a significant functional role for this segment during amelogenesis. The presence of a unique 5 amino acid insert in mouse and rat Ambn is also absent in human as reported for pig (17). Most interesting is the 26 amino acid insertion found exclusively in the human Ambn sequence. Most likely, this segment arose through the insertion and duplication of a smaller 13 amino acid sequence based on the degree of amino acid similarity. This segment increases the molecular weight of both human Ambn isoforms, as well as increasing the overall acidity of the human protein.

One of the slight ambiguities related to the Ambn cDNA clones identified to date is the assignment of the start codon between the first, second or third Met. The first Met is the start codon ($\sim$90% of eukaryotic mRNAs), and this initial codon is preceded by a purine (adenine or guanine) in the (-3) position ($\sim$97% of eukaryotic sequences) (20). This leads to a conflict for all reported Ambn sequences, where the first Met does not have a purine at the -3 position. This would support the assignment of the second Met as the start codon which does have an adenine in the correct position, leading to a shorter 16 amino acid signal peptide. However, at the amino acid level the sequence between the first and second Met is highly conserved (8 out of 10 residues) between species, suggesting utilization of the first Met.

The present investigation demonstrates immunological cross-reactivity between a polyclonal recombinant rat Ambn antibody and native human Ambn epitopes. This evidence supports the existence of homologies in these polypeptide regions between both species. Previously, studies have shown this same cross-reactivity on human forming tooth germs at the early bell stage of development (10). In contrast to the early expression of Ambn mRNAs at E-14 by RT-PCR analysis (16), the protein was only detectable in differentiated ameloblasts. However, the detection of early protein synthesis may be under the threshold of immunohistochemistry sensitivity, since decalcification procedures may delete more than 50% of the immunoreactive species (21), and human samples are not ideally preserved in contrast to experimental approaches using animal models where

intracardiac perfusions are usually performed (10, 14, 21, 22).

Comparative data on the cellular and extracellular distribution of Ambn (human, the present study and rat incisor; 8, 22, 23) shows conservation among mammals for the respective compartmentalization of matrix proteins during the process of enamel formation. In our study, immunoreactivity for Ambn proteins appeared concentrated in the prism sheaths of forming enamel, as shown previously for the NH_2-terminal region of rat Ambn (22) and porcine sheathlin (14). However, the distribution of immunoreactive species at the secretory pole of ameloblasts was not related to the formation of enamel sheath as also seen in rat (22, 23), but in contrast to localization in pig (14). The present data support the notion that Ambn may constitute a functional group of proteins which contribute to the formation of prism sheaths.

The characterization of the Ambn cDNA sequence will facilitate determination of the genomic organization of the gene and future mutational analyses in patients presenting with the local hypoplastic form of AI.

Acknowledgements – This work is supported by NIDCR/NIH research grant DE09875 (M.M.) and Hospital Clinical Research Program grant AOM 96067 (A.B.).

References

1. Witkop C, Sauk JJ. Heritable Defects of Enamel. In: Steward RE, Prescott GH, eds. *Oral facial genetics*. St. Louis: Mosby, 1976; 151–226.
2. Chosack A, Eidelman E, Wisotski I, Cohen T. Amelogenesis imperfecta among Israeli Jews and the description of a new type of local hypoplastic autosomal recessive amelogenesis imperfecta. *Oral Surg Oral Med Oral Pathol* 1979; **47**: 148–156.
3. Bäckman B, Holmgren G. Amelogenesis imperfecta: A genetic study. *Hum Hered* 1988; **38**: 189–206.
4. Bäckman B, Holm AK. Amelogenesis imperfecta: prevalence and incidence in a Northern Swedish County. *Comm Dent Oral Epidemiol* 1986; **14**: 43–47.
5. Kärrman C, Bäckman B, Holmgren G, Forsman K. Genetic heterogeneity of autosomal dominant amelogenesis imperfecta: demonstrated by its exclusion from the AIH2 region on human chromosome 4q. *Arch Oral Biol* 1996; **41**: 893–900.
6. Forsman K, Lind L, Bäckman B, Westermark E, Holmgren G. Localization of a gene for autosomal dominant amelogenesis imperfecta (ADAI) to chromosome 4q. *Hum Mol Genet* 1994; **3**: 1621–1625.
7. Kärrman C, Bäckman B, Dixon M, Forsman K. Mapping of the locus for autosomal dominant amelogenesis imperfecta (AIH2) to a 4 Mb YAC contig on chromosome 4q11–q21. *Genomics* 1997; **39**: 164–170.
8. Krebsbach PH, Lee SK, Matsuki Y, Kozak CA, Yamada K, Yamada Y. Full-length sequence, localization, and chromosomal mapping of ameloblastin: a novel tooth-specific gene. *J Biol Chem* 1996; **271**: 4431–4435.
9. Cerny R, Slaby I, Hammarström L, Wurtz T. A novel gene expressed in rat ameloblasts codes for proteins with cell binding domains. *J Bone Miner Res* 1996; **11**: 883–891.
10. Lee SK, Krebsbach PH, Matsuki Y, Nanci A, Yamada KM, Yamada Y. Ameloblastin expression in rat incisors and human tooth germs. *Int J Dev Biol* 1996; **40**: 1141–1150.
11. Fong CD, Hammarström L, Lundmark C, Wurtz T, Slaby I. Expression patterns of RNAs from amelin and amelogenin in developing rat molars and incisors. *Adv Dent Res* 1996; **10**: 195–200.
12. Fong CD, Slaby I, Hammarström, L. Amelin: An enamel-related protein, transcribed in the cells of epithelial root sheath. *J Bone Miner Res* 1996; **11**: 892–898.
13. Fong CD, Cerny R, Hammarström L, Slaby I. Sequential expression of an amelin gene in mesenchymal and epithelial cells during odontogenesis in rats. *Eur J Oral Sci* 1998; **106**: 324–330.
14. Uchida T, Fukae M, Tanabe T, Yamakoshi Y, Satoda T, Murakami C, Takahashi O, Shimizu M. Immunochemical and immunocytochemical study of a 15 kDa non-amelogenin and related proteins in the porcine immature enamel: proposal of a new group of enamel proteins sheath proteins. *Biomed Res* 1995; **16**: 131–140.
15. Uchida T, Murakami C, Wakida K, Dohi N, Iwai Y, Simmer JP, Fukae M, Satoda T, Takahaashi O. Sheath proteins: synthesis, secretion, degradation and fate in forming enamel. *Eur J Oral Sci* 1998; **106**: 308–314.
16. Simmons D, Gu TT, Krebsbach PH, Yamada Y, MacDougall M. Identification and characterization of a cDNA for mouse amelobastin. *Connect Tissue Res* 1998; **39**: 3–12[307–316].
17. Hu C-C, Fukae M, Uchida T, Qian Q, Zhang CH, Ryu OH, Tanabe T, Yamakoshi Y, Murakami C, Dohi N, Shimizu M, Simmer JP. Sheathlin: cloning, cDNA/polypeptide sequences, and immunolocalization of porcine enamel proteins concentrated in the sheath space. *J Dent Res* 1997; **76**: 648–657.
18. MacDougall M, DuPont BR, Simmons D, Reus B, Krebsbach P, Karrman C, Holmgren G, Leach RJ, Forsman K. Ameloblastin gene (AMBN) maps within the critical region for autosomal dominant amelogenesis imperfecta at chromosome 4q21. *Genomics* 1997; **41**: 115.
19. MacDougall M, Simmons D, Zianghong L, Nydegger J, Feng J, Gu TT. Dentin phosphoprotein and dentin sialoprotein are cleavage products expressed from a single transcript coded by a gene on human chromosome 4. *J Biol Chem* 1997; **272**: 835–842.
20. Kozak M. Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucl Acids Res* 1984; **12**: 857–872.
21. Berdal A, Nanci A, Smith CE, Ahluwalia JP, Thomasset M, Cuisinier-Gleizes P, Mathieu H. Differential expression of calbindin-D28k in rat incisor ameloblasts throughout enamel development. *Anat Rec* 1991; **230**: 149–163.
22. Uchida T, Murakami C, Dohi N, Wakida K, Satoda T, Takahashi O. Synthesis, secretion, degradation, and fate of ameloblastin during matrix formation stage of the rat incisor as shown by immunocytochemistry and immunochemistry using region-specific antibodies. *J Histochem Cytochem* 1997; **45**: 1329–1340.
23. Nanci A, Zalzal S, Lavoie P, Kunikata M, Chen W-Y, Krebsbach PH, Yamada, Y, Hammarström L, Simmer JP, Fincham AG, Snead ML, Smith CE. Comparative immunochemical analyses of the developmental expression and distribution of ameloblastin and amelogenin in rat incisors. *J Histochem Cytochem* 1998; **46**: 911–934.