Archives of Oral Biology (2005) **50**, 205–212

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Archives of Oral Biology

www.intl.elsevierhealth.com/journals/arob

Amelogenin: lessons from evolution

Jean-Yves Sire[a,*], Sidney Delgado[a], Delphine Fromentin[a], Marc Girondot[b]

[a]*Equipe "Evolution and Développement du Squelette", CNRS FRE 2696, Université Paris 6-Pierre and Marie Curie, CNRS FRE 2696, Case 7077, 2, place Jussieu, 75251 Paris Cedex 05, France*
[b]*UMR 8079, Ecologie, Systématique et Evolution, Université Paris-Sud, Orsay, France*

Accepted 17 September 2004

KEYWORDS
Amelogenin;
Amino acids;
Evolution;
Tetrapods

Summary Amelogenin plays a crucial role in enamel structure and mineralization, but the function of its various domains is far to be understood. Evolutionary analysis seems to be a promising way to approach structure/function relationships. In this paper, we review the knowledge of amelogenin with a particular focus on what we have learnt from evolution, and we bring new data on the origin and evolution of this molecule.

The comparison of amniote (reptiles and mammals) amelogenin sequences reveals that, in contrast to the well-conserved C- and N-terminal domains, the central region (most of exon 6) is highly variable. The evolutionary analysis indicates that it was created by repeated insertion of three amino acids (triplets ProXGlu or ProXX). In several mammalian lineages a new run of triplet insertions and deletions has occurred independently in a locus considered a hot spot of mutation for mammalian amelogenin. In lizard and snake amelogenin evolves rapidly. Sequence alignment reveals that several residues in the N- and C-terminal regions were kept unchanged during 250 million years (MY), proving their importance for amelogenin structure and function. This alignment permits a rapid validation of the amelogenin mutations in human.

Genome sequencing and gene mapping permitted to refine the amelogenin story, in relation to the common location (chromosome 4 in human) of several genes coding for dental proteins and SPARCL1, a SPARC (osteonectin) relative. Amelogenin shares a similar organisation with these genes and a blast search in databanks indicates a strong relationship between amelogenin, ameloblastin and enamelin. Taken together these data suggest that amelogenin could have originated from either ameloblastin or enamelin, themselves being created from SPARCL1, which itself originated from a SPARC duplication, 600 millions years ago.

* Corresponding author. Tel.: +33 1 44 27 35 72; fax: +33 1 44 27 35 72.
E-mail address: sire@ccr.jussieu.fr (J.-Y. Sire).

0003-9969/$ – see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.archoralbio.2004.09.004

Introduction

Amelogenin, the major protein of forming enamel in mammals, is considered dental specific: (i) it is encountered in any other organ or tissue; (ii) amelogenin gene[a] (*AMEL*) was not found in toothless vertebrates, turtles and birds;[1] and (iii) in human, when *AMEL* is subjected to mutation, no phenotypes is observed out of enamel defects (*amelogenesis imperfecta*). In the course of an extensive comparative study of tooth and tooth-related tissues, we become interested in amelogenin because this lack of pleiotropy indicated this protein as a specific marker of enamel. Since our first study in 1998,[1] in our lab and in the literature numerous data have been accumulated, which bring new insights on the origin and evolution of *AMEL*. In the present paper we briefly summarise the reasons that have led our interest to amelogenin, then we review what we have learnt these last years on its origin and evolution. We add new information and we propose a scenario for amelogenin evolution.

From fish scale development to amelogenin evolution

Since the 1980s, extensive, comparative developmental studies of teeth and various elements of the dermal skeleton (odontodes, denticles, scales, and dermal bones) have been realized to understand how they are evolutionary related.[2–4] One of the main finding was to suggest that teleost fish scales derive from ancestral tooth-like elements, odontodes: (i) developmental features of tooth and scale are similar, (ii) the upper layer structure of scale is deposited by epithelial-derived cells as enamel, and (iii) elasmoid scales derive from ancestral odontocomplexes (see above reviews). However, during 100 million years (MY) of evolution the scale tissues were so deeply modified that it is difficult to infer their homology with dentin and enamel. Our efforts focused onto the well-mineralized, enamel-like tissue covering the scales, with the goal to demonstrate the expression of enamel specific proteins/genes during the formation of this layer. This would definitively close the debate on scale/tooth homology.

Because it is the major protein of forming enamel, we have chosen amelogenin, with the hope to clone its gene in teleost fish. However, the poor knowledge of this gene in vertebrates did not allow cloning *AMEL* in fish. Understanding *AMEL* evolution, and the location of sequence variations, which probably prevent cloning this gene in non-tetrapods, could help to define appropriate primers. We started, therefore, a study of amelogenin evolution in mammals, then extended our interest to reptiles and, recently, to amphibians. In parallel to these studies we have tried to understand the evolutionary origin of this protein.

Amelogenin evolution

Gene structure and supposed functions of the protein

In mammals, *AMEL* is said to possess seven exons, although exon 4 is lacking in most species studied, and two extra exons (8 and 9) are found in human and rodents[5,6] (Fig. 1). The C- and N-terminal regions are hydrophilic, while the central region (most of exon 6) is hydrophobic. *AMEL* is subjected to extensive alternative splicing.[7–9] The N-terminal region contains amino acids that are suspected to be involved in various functions: a phosphorylation site,[10] alpha helices,[11] a binding site to *N*-acetylglucosamine and keratins.[12] The amelogenin is post-secretory subjected to proteolysis, which ends by N-terminal cleavage leading to small TRAP peptides.[10] These cleavage steps are though to be important for enamel structure and for its proper mineralization. Amelogenin molecules aggregate into nanospheres (Fig. 2). The hydrophilic C- and N-terminal regions interact with the surrounding environment (cell processes and hydroxyapatite crystals) and the hydrophobic region (exon 6) forms the core of the nanospheres, which are responsible for the enamel microstructure.[13] The C-terminal region plays an important role in contributing to stabilize these nanospheres.[14]

In human, several mutations of *AMEL* lead to *amelogenesis imperfecta* (see review in Ref. [15]). Interestingly, aside large deletions or single nucleotide deletions resulting to frame shift alteration, some of these mutations are substitution of a single amino acid in the N-terminal region. These studies, which show a relationship between enamel defects and an amelogenin mutation, are important in that they highlight residues or domains which are essential for the proper functioning of the protein.[16]

Amelogenin Y evolution in eutherians

In eutherians, e.g., cattle, human, horse, black bear and monkeys, a copy of *AMEL* is located on each sex

[a] In the following, when not specified amelogenin gene (*AMEL*) means X-linked one.

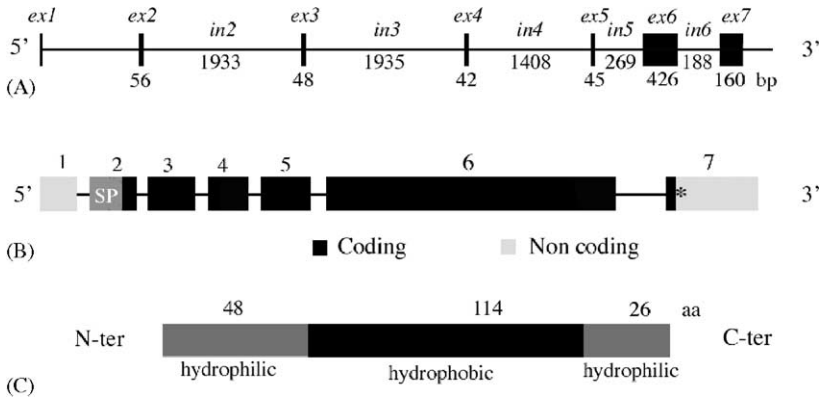

Figure 1 Amelogenin gene. (A) Gene structure. The five first exons (*ex1*—*ex5*) are small (42—56 bp) and exons 6 (426 bp) and 7 (160 bp) are large. (B) Coding and uncoding exons. Exon 1 is uncoding, most exon 2 sequence codes the signal peptide, and the only first three nucleotides of exon 7 are coding for the protein. (C) Linear representation of the native protein.

chromosome,[17] while the gene is autosomal in monotremes, marsupials and non-mammalian species. *AMELX* and *AMELY* show several differences (substitutions and indels), which can be explained by an independent evolution during millions years.[1] This also indicates that *AMELY* is not under strong functional constraint. *AMELY* is masked by *AMELX* and cannot provide protection when the latter is inactive.[18] The particular evolution of *AMELY* can be understood when considering sex chromosome evolution.[19,20] X and Y mammalian chromosomes have evolved from a pair of autosomal chromosomes of a common ''reptile-like'' ancestor (a therapsid), 250 million years ago. During mammalian evolution, four successive multigene inversions have occurred in the Y chromosome. They have led to restricted recombination of X and Y. *AMELY* being located close to the non-recombining inversion locus, this resulted in a low level of recombination.[20] *AMELY* may therefore be tending toward a pseudogene. The last inversion occurred early in eutherian evolution and the analysis of the few available *AMELY* sequences clearly indicates that the *AMELY* loci became non-recombining separately in each lineage, and in some species this occurred a long time after the eutherian diversification.

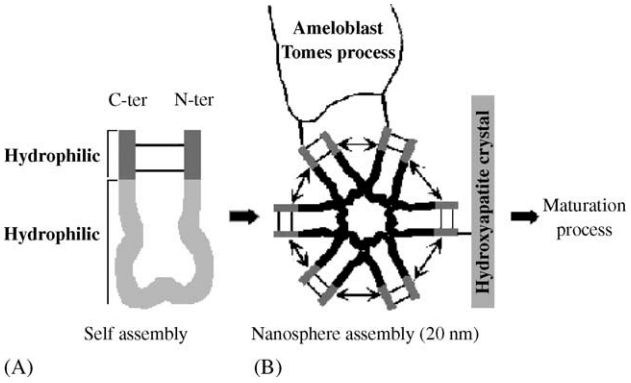

Figure 2 (A) Each amelogenin molecule self assembles by means of bindings at the N- and C-terminal, hydrophilic regions. (B) Several amelogenin molecules aggregate and constitute a nanosphere, which interacts both with the cell membrane of the ameloblast processes and the hydroxyapatite crystals (after Snead[13]).

Amelogenin evolution in tetrapods

To date, the entire or partial amelogenin (or cDNA) sequence is mainly known in mammals, with only a single sequence in amphibians (*Xenopus*) and crocodiles,[11] in a snake,[21] and a lizard.[22] In mammals, 26 amelogenin sequences are known,[17] and 21 new sequences are presented herein. Comparison of amniote (reptiles + mammals) amelogenins reveals that the C- (48 residues) and N- (26 residues) terminal regions show a high sequence similarity (Fig. 3). They evolve slowly because they are constrainted in relation to important functions of some amino acids. In contrast, most of the exon 6 is variable, which reflects low selection pressure. The number of substitutions is particularly important in lizard and snake, which avoids alignment (Fig. 3). Additional data are necessary, and the current analysis of 26 reptilian sequences will improve our knowledge on exon 6 evolution in amniotes. On the other hand, the evolutionary analysis of amelogenin using 26 mammalian sequences representative of the main lineages has permitted to calculate the putative ancestral amelogenin, as it existed in the mammalian ancestor, 200 MY ago[17] (Fig. 3). New insights were also brought in this study on the evolution of the variable region of exon 6. It has been created by

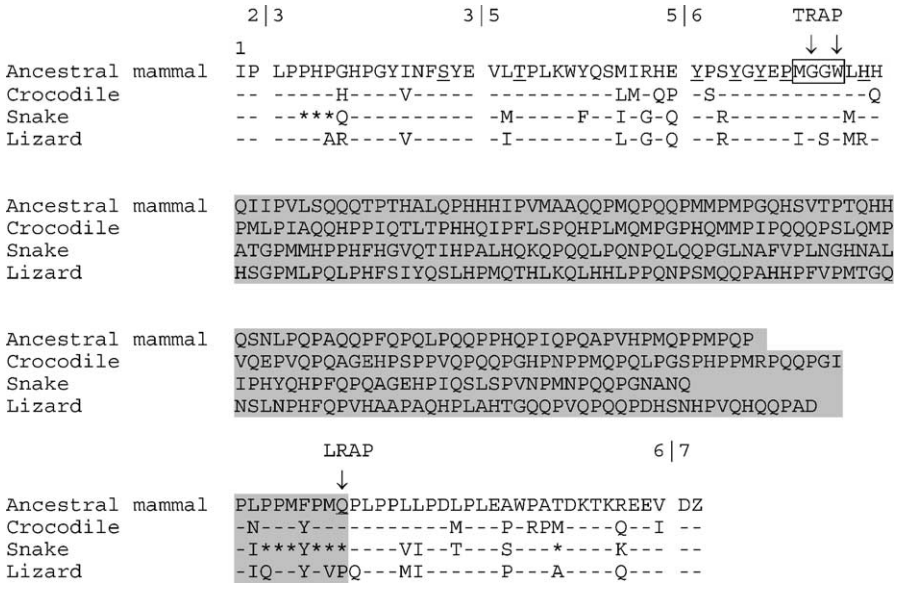

Figure 3 Alignment of the amino acid sequence of amniote amelogenins: the putative ancestral mammal (calculated from 26 sequences)[17], a crocodile,[11] a snake,[21] and a lizard.[22] The variable region (in grey) of exon6 sequences cannot be aligned. (-) identical residue; (*) deleted residue; 2|3 exon2—exon3 boundary; TRAP: proteolytic sites leading to the two tyrosine rich amelogenin peptides; LRAP: locus of the intra-exonic splicing site in mammals leading to the leucine rich amelogenin peptide.

numerous repeat insertions of three amino acids (PXX or PXQ), leading to the current richness in proline (P) and glutamine (Q). Several runs of repeats were identified, and they occurred long before mammalian divergence, and at least in a stem tetrapod, because such repeats are present in the same exon 6 region of reptile and frog amelogenins.

Interestingly, a new run of triplet insertions has occurred during mammalian evolution, but not in the other lineages. New PXQ repeats were inserted independently in several, unrelated lineages. These insertions (and deletions in some species) are always located in the same region, which is considered a hot spot of mutation for mammalian amelogenin.[17] The evolutionary analysis does not reveal the presence of this hot spot in the ancestral mammalian amelogenin. The presence of large indels in this region of the amelogenin could have an influence on the enamel microstructure, but this remains to be checked. Also, this highly variable region of exon 6 appears to be a favourable locus for *AMEL* polymorphism in human.

In the course of a current phylogenetic study of mammals using *AMEL*, we have added 21 new amelogenin sequences (mostly exon 6) to our previous data set (Fig. 4), increasing our data set to 47 mammalian sequences. The analysis of these sequences confirms largely the results of our previous study. Particularly, indels are present in the region of the hot spot of mutation.

The alignment of our data set of mammalian and reptilian amelogenins reveals that N- and C-terminal regions only present conserved residues, i.e., those kept unchanged during 250 million years of evolution. These results point to the important role that must play these residues for amelogenin. They are also predictable sites, which could lead to enamel defects when either deleted or substituted. The three amino acid positions known to lead to amelogenesis imperfecta (*AI*) when substituted are validated by this evolutionary analysis, i.e, they are present in all sequences studied so far.

Amelogenin origin

The evolutionary analysis reported above indicates that *AMEL* is unknown to date below the amphibian level. However, immunohistochemical studies (using antibodies against mammalian amelogenin) have detected amelogenin in non-tetrapods. Therefore, *AMEL* was probably present in the common ancestor of vertebrates, at least 450—500 MY ago. Where does it come from? Amelogenin is clearly a new gene, which certainly generated from duplication. But what could be this ancestral gene? We have shown that most part of exon 6 sequence is proper to *AMEL*, because it has been generated by the insertion of numerous triplet repeats (9 bases). In contrast, the C- and N-terminal regions, which have been highly conserved during evolution given their

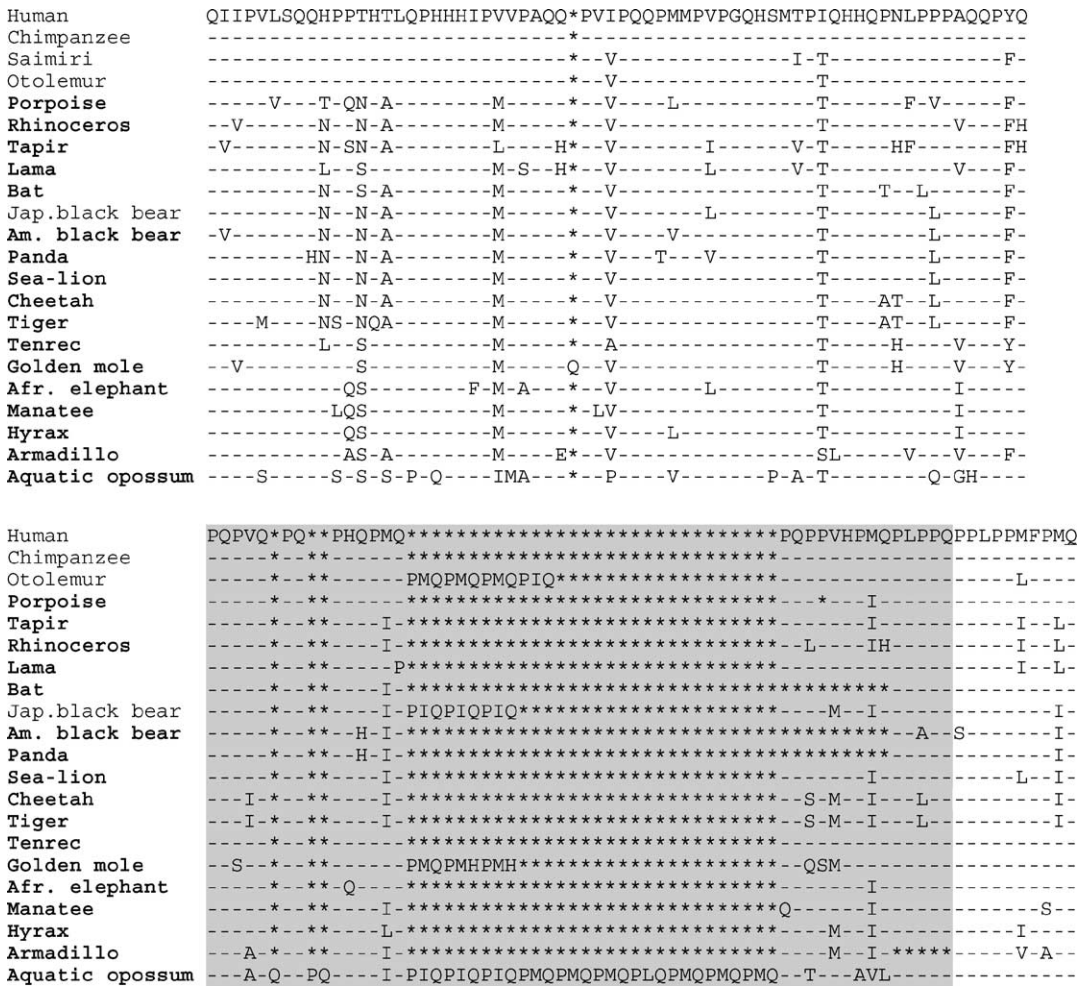

Figure 4 Alignment of the variable region of exon 6 amino acid sequence, from position 48 to the LRAP splicing site, for human and 21 mammalian amelogenins. In bold: 17 newly sequenced amelogenins. The hot spot of mutation is indicated in grey; four sequences (otolemur, Japanese black bear, golden mole, and aquatic opossum) show insertions of amino acid triplets (PXX), two sequences (American black bear and panda) show deletion of 3 triplets. (-) identical residue; (*) deleted/inserted residue.

important functions for cell—matrix interactions (the cell membrane of the ameloblast) and the hydroxyapatite crystals (Ca-binding phosphoprotein),[13] could have been retained in its ancestor (and in other derivatives if they exist).

In 1998, our attention was drawn by similarities existing between the signal peptides of AMEL and SPARC (osteonectin), a glycoprotein, known in protostomians and deuterostomians, involved in a large range of functions, notably Ca-binding and mediating cell—matrix interactions. Search in DNA banks indicated also similarities with the signal peptide of a SPARC-related protein, SC1 (hevin, QR1, SPARCL1). Testing similarities demonstrated a common evolutionary origin of exons coding the signal peptide (exon 1 in *SPARC* and *SPARCL1*, and exon 2 in *AMEL*). A phylogenetic analysis indicated that *AMEL* was probably duplicated from *SPARC*, 630 MY ago, i.e., long before ''Cambrian explosion''.[23]

These last years, advances in sequencing techniques have provided a large amount of data: human, mice, rat, fugu and zebrafish genomes have been sequenced, and numerous genes coding for dental and bone proteins have been cloned, characterized and mapped on chromosomes: enamel proteins (ameloblastin, AMBN; enamelin, ENAM); dentin and bone proteins (dentin sialophosphoprotein, DSPP; dentin matrix acidic phosphoprotein 1, DMP1; integrin-binding sialoprotein or bone sialoprotein, IBSP; matrix extracellular, phosphoglycoprotein or osteoregulin, MEPE; secreted phosphoprotein 1 or osteopontin, SPP1). Interestingly, all these dental and bone protein genes are located in two clusters on the long arm of human chromosome 4 (4q13—21): a cluster with genes

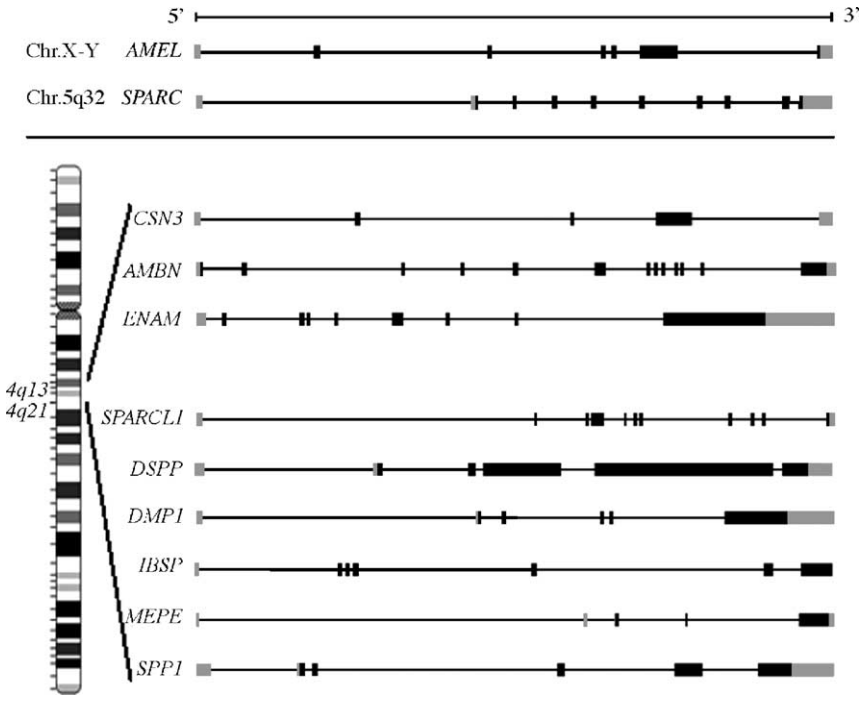

Figure 5 Mapping of the genes coding for dental and bone proteins on human chromosome 4 (4q13—21). See text for abbreviations. The organization of these gene is similar to each other and to that of SPARC (chromosome 5, q32) and of AMEL (X and Y chromosome). In grey: non coding sequence.

coding for enamel proteins (*ENAM*, *AMBN*), milk caseins and salivary proteins (statherins and histatins), and a cluster containing five genes coding for dentin and bone extracellular matrix proteins (*DSPP, DMP1, IBSP, MEPE, SPP1*). Surprisingly, *SPARCL1*, which was identified previously as being related to amelogenin[23] is mapped between these two clusters (Fig. 5).

Two recent studies[24,25] have analysed the correlation between the common chromosome location of these genes, their exon structure and the possible similar functions of the proteins (e.g., Ca-binding phosphoproteins). The genes of the second cluster, i.e., coding for proteins expressed in bone and teeth have a similar exon structure, but their sequences show a low level of conservation, except for the signal peptide. These genes belong to a single family, called SIBLING (Small Integrin-Binding LIgand N-linked Glycoproteins) by Fisher and Fedarko.[24] Duplications, which have generated these genes, occurred a long time ago from a common ancestor. Each gene has subsequently acquired its own function through mutations favouring specialized adaptive functions. Similarly, Kawasaki and Weiss[25] considered the similarity of exon structure of the genes in the two clusters as supporting a common ancestry (milk casein, e.g. *CSN3*, and salivary protein genes being secondarily duplicated from one of the two enamel protein genes, *ENAM* or *AMBN*). These authors proposed to include *AMEL* in their analysis. They also identified *SPARC* (on human chromosome 5) to have possibly originated the primordial dental/bone protein gene.

These new important data modify slightly our previous analysis on the evolutionary origin of *AMEL*. For instance, *AMEL* exon structure, and particularly exons 2 and 3, is more similar to *AMBN* and *ENAM* than to the other genes in the clusters.[25] We used Psi-blast method to search for similarities of signal peptides of the genes. This strategy consisted in performing a Blast search in the database taking into account the possible variation of sequence during evolution (Table 1). Although this analysis cannot be used to support a duplication chronology, the presence of SPARCL1 and SPARC in the results confirms our previous study, which demonstrated a relationship between AMEL and SPARC.[23] However, the new data suggest that AMEL is closer to SPARCL1 than to SPARC, and even closer to AMBN and ENAM than to SPARCL1. Moreover, the common chromosome location indicates that SPARCL1 could be the ancestor of AMBN and ENAM.

These findings lead us to propose a putative scenario, in which SPARC duplicated into SPARCL1 early in deuterian evolution, then SPARCL1 was copied several times on the same chromosome, giving rise to the dental and bone protein gene

Table 1 Results of psi blast search in GenBank (July 2004) using exon 2 signal peptide from SPARC, SPARCL1, AMEL, ENAM and DSPP. See text for abbreviations.

Exon 2 signal peptide (order of appearance of sequences with similarities using psi blast search)				
SPARC	SPARCL 1	AMEL	ENAM	DSPP
SPARC	SPARCL 1	AMEL	ENAM	DSPP
	IBSP	AMBN		
	DMP	SPARCL 1		
	SPP1	SPARC		
	AMBN	IBSP		
	AMEL	ENAM		
		CSN 3		

PAM-30 substitution matrix was chosen instead of BLOSUM one because short but strong alignments are more easily detected using a matrix with a higher relative entropy.[26]

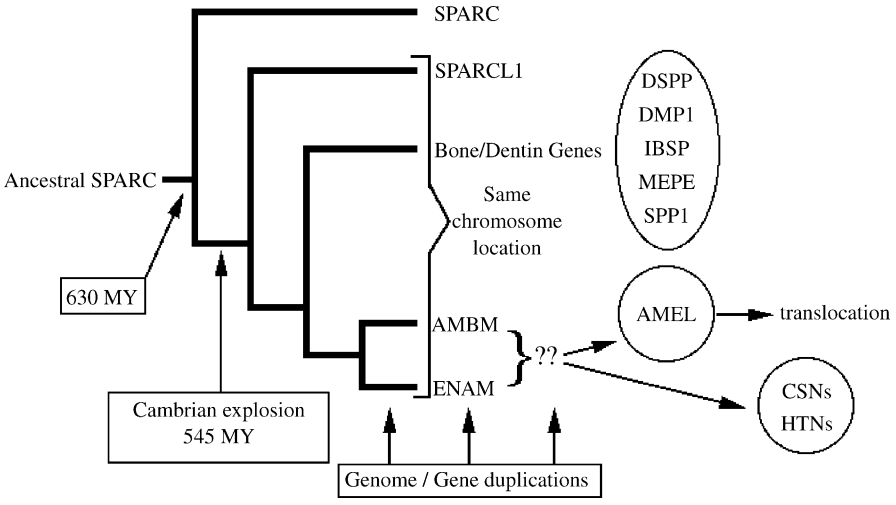

Figure 6 A probable scenario on the origin and evolution of amelogenin. Early in deuterian evolution (630 MY ago), *SPARC* duplicated into *SPARCL1*. During successive phases of genome duplication *SPARCL1* was copied several times on the same chromosome, giving rise to two clusters: the enamel protein genes (*AMBN—ENAM*) and the bone-dentin protein genes (*DSPP, DMP1, IBSP, MEPE, SPP1*). *AMEL* duplicated from either *AMBN* or *ENAM*, and was translocated elsewhere. *AMEL* differentiated slowly in the 5′ and 3′ regions, but rapidly in the central region where amino acid repeats were added. A new function was acquired in enamel formation. See text for abbreviations. CSNs: caseins; HTHs: histatins.

clusters, and finally either AMBN or ENAM duplicated into AMEL, which was translocated elsewhere and differentiated separately (Fig. 6).

Conclusion and clues for the future

This review paper demonstrates how the evolutionary analysis could be helpful (i) to reveal domains and residues, which play important roles for the structure/function of a protein, and (ii) to validate human genetic disease (e.g., *AI*). This implies that an important amount of work has to be done in the future to produce large datasets of sequences representative of the vertebrate phylogeny. Indeed, our knowledge is limited in general to a few mammalian species (e.g., human, mouse, rat, pig, bovine) and to a single species in other tetrapod lineages when available (crocodile, chicken, xenopus, etc.). Even for amelogenin, for which we know mammalian (50), reptilian (26) and amphibian (1) sequences, we are not able yet to find this gene in bony and cartilaginous fish.

However, large dataset analysis can (i) bring information on gene evolution, (ii) lead to ancestral sequences, (iii) help to understand gene relationships, and (iv) be used to build phylogenies. For instance, relationships between AMEL, AMBN and ENAM could be confirmed in building a phylogeny as presented partially by Kawasaki and Weiss,[25] but based on a larger set of data. Indeed, the current data avoid full-length sequence alignment and exons 2 and 3 are too small to be used as significant markers. However, if our hypothesis that AMEL

derives from AMBN or ENAM is right, the comparison of the ancestral sequences should bring interesting information.

Another possibility would be to look for syntheny of the 4q clusters in other vertebrate lineages. Indeed, we know that syntheny is conserved in human (4q13—21), rat (14p21—22) and mouse (5E). The current sequenced genomes (bovine, pig, chicken, etc.) will indicate whether such syntheny is conserved in another mammalian lineage (artiodactyls) and in archosaurs (birds). As far as we know, any genes belonging to these clusters have been mapped in zebrafish and fugu chromosomes.

References

1. Girondot M, Sire J-Y. Evolution of the amelogenin gene in toothed and tooth-less vertebrates. *Eur J Oral Sci* 1998;**106**(Suppl. 1):501—8.
2. Huysseune A, Sire J-Y. Evolution of patterns and processes in teeth and tooth-related tissues in non-mammalian vertebrates. *Eur J Oral Sci* 1998;**106**(Suppl. 1):437—81.
3. Sire J-Y, Davit-Beal T, Delgado S, Van Der Heyden C, Huysseune A. The first generation teeth in non-mammalian lineages: evidence for a conserved ancestral character? *Microsc Res Technol* 2002;**59**:408—34.
4. Sire J-Y, Huysseune A. Formation of skeletal and dental tissues in fish: a comparative and evolutionary approach. *Biol Rev* 2003;**78**:219—49.
5. Li W, Mathews C, Gao C, DenBesten PK. Identification of two additional exons at the 3′ end of the amelogenin gene. *Arch Oral Biol* 1998;**43**:497—504.
6. Baba O, Takahashi N, Terashima T, Li W, DenBesten PK, Takano Y. Expression of alternatively spliced RNA transcripts of amelogenin gene exons 8 and 9 and its end products in the rat incisor. *J Histochem Cytochem* 2002;**50**:1229—36.
7. Gibson C, Golub E, Herold R, Risser M, Ding W, Shimokawa H, et al. Structure and expression of the bovine amelogenin gene. *Biochemistry* 1991;**30**:175—9.
8. Simmer JP. Alternative splicing of amelogenins. *Connect Tissue Res* 1995;**32**:131—6.
9. Yuan ZA, Chen E, Gibson CW. Model system for evaluation of alternative splicing: exon skipping. *DNA Cell Biol* 2001;**20**:807—13.
10. Fincham AG, Hu Y, Lau EC, Slavkin HC, Snead ML. Amelogenin post-secretory processing during biomineralization in the postnatal mouse molar tooth. *Arch Oral Biol* 1991;**36**:305—17.
11. Toyosawa S, O'HUigin C, Figueroa F, Tichy H, Klein J. Identification and characterization of amelogenin genes in monotremes, reptiles, and amphibians. *Proc Natl Acad Sci USA* 1998;**95**:13056—61.
12. Ravindranath RMH, Basilrose RM, Ravindranath NH, Vaitheesvaran B. Amelogenin interacts with cytokeratin-5 in ameloblasts during enamel growth. *J Biol Chem* 2003;**278**:20293—302.
13. Snead M. Amelogenin protein exhibits a modlar design: Implications for form and function. *Connect Tissue Res* 2003;**44**(Suppl. 1):47—51.
14. Moradian-Oldak J, Simmer JP, Sarte PE, Zeichner-David M, Fincham AG. Specific cleavage of a recombinant murine amelogenin at the carboxy- terminal region by a proteinase fraction isolated from developing bovine tooth enamel. *Arch Oral Biol* 1994;**39**:647—56.
15. Hart PS, Hart TC, Simmer JP, Wright JT. A nomenclature for X-linked amelogenesis imperfecta. *Arch Oral Biol* 2002;**47**:255—60.
16. Wright JT, Hart PS, Aldred MJ, Seow K, Crawford PJM. Hong SP, et al. Relationship of phenotype and genotype in X-linked amelogenesis imperfecta. *Connect Tissue Res* 2003;**44**(Suppl. 1):72—8.
17. Delgado S, Girondot M, Sire J-Y. Molecular evolution of amelogenin in mammals, *J Mol Evol*, in press.
18. Lagerstrom M, Dahl N, Nakahori Y, Nakagome Y, Backman B, Landegren U, et al. A deletion in the amelogenin gene (AMG) causes X-linked amelogenesis imperfecta (AIH1). *Genomics* 1991;**10**:971—5.
19. Lahn BT, Pearson NM, Jegalian K. The human Y chromo some, in the light of evolution. *Nat Rev Genet* 2001;**2**:207—16.
20. Iwase M, Satta Y, Hirai Y, Hirai H, Imai H, Takahata N. The amelogenin loci span an ancient pseudoautosomal boundary in diverse mammalian species. *Proc Natl Acad Sci USA* 2003;**100**:5258—63.
21. Ishiyama M, Mikami M, Shimokawa H, Oida S. Amelogenin protein in tooth germs of the snake Elaphe quadrivirgata, immunohistochemistry, cloning and cDNA sequence. *Arch Histol Cytol* 1998;**61**:467—74.
22. Delgado S, Couble ML, Zalzal S, Nanci A, Magloire H, Sire J-Y. Cloning, sequencing and expression of amelogenin in the scincid lizard Chalcides (Squamata), *Dev Gen Evol*, in press.
23. Delgado S, Casane D, Bonnaud L, Laurin M, Sire JY, Girondot M. Molecular evidence for Precambrian origin of amelogenin, the major protein of vertebrate enamel. *Mol Biol Evol* 2001;**18**:2146—53.
24. Fisher LW, Fedarko NS. Six genes expressed in bones and teeth encode the current members of the sibling family of proteins. *Connect Tissue Res* 2003;**44**(Suppl. 1):33—40.
25. Kawasaki K, Weiss KM. Mineralized tissue and vertebrate evolution: the secretory calcium-binding phosphoprotein gene cluster. *Proc Natl Acad Sci USA* 2003;**100**:4060—5.
26. Altschul SF. Amino acid substitution matrices from an information theoretic perspective. *J Mol Biol* 1991;**219**:555—65.