



HAL
open science

**Les phosphoprotéines sécrétées liant le calcium (SCPP)
impliquées dans la formation de l'émail dentaire :
expression chez le lézard *Anolis carolinensis* et évolution
chez les amniotes**

Barbara Gasse

► **To cite this version:**

Barbara Gasse. Les phosphoprotéines sécrétées liant le calcium (SCPP) impliquées dans la formation de l'émail dentaire : expression chez le lézard *Anolis carolinensis* et évolution chez les amniotes. Evolution [q-bio.PE]. Université Pierre et Marie Curie - Paris VI, 2015. Français. NNT : 2015PA066066 . tel-01154518

HAL Id: tel-01154518

<https://theses.hal.science/tel-01154518>

Submitted on 22 May 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE DE DOCTORAT DE L'UNIVERSITÉ PIERRE ET MARIE CURIE

Présentée par **Barbara GASSE**

Pour obtenir le grade de
DOCTEUR DE L'UNIVERSITE PIERRE ET MARIE CURIE

**Les phosphoprotéines sécrétées liant le calcium (SCPP) impliquées
dans la formation de l'émail dentaire : expression chez le lézard
Anolis carolinensis et évolution chez les amniotes**

Soutenue le 14 avril 2015

Devant le jury composé de:

Mme Ann HUYSEUNE	PU, Université de Gand	Rapportrice
M. Marc GIRONDOT	PU, Université Paris-Sud	Rapporteur
Mme Claire FOURNIER-THIBAUT	PU, UPMC	Examinatrice
M. Frédéric MARIN	DR, CNRS	Examinateur
Mme Ariane BERDAL	PU-PH, Univ. Paris Diderot	Examinatrice
M. Jean-Yves SIRE	DR, CNRS	Directeur de thèse

Ecole doctorale Complexité du vivant

UMR7138 Evolution Paris-Seine, Equipe Evolution et Développement du Squelette

7 quai Saint-Bernard, bâtiment A, 2e étage, 75005 Paris

REMERCIEMENTS

Je tiens tout d'abord à remercier mon directeur de thèse, Jean-Yves Sire, pour m'avoir accueilli dans son équipe et m'avoir donné l'opportunité de réaliser cette thèse. Merci de m'avoir encadré tout au long de celle-ci.

J'adresse mes sincères remerciements aux membres du jury pour avoir accepté d'évaluer mon travail de thèse. Merci tout particulièrement à Frédéric Marin qui a également suivi mes travaux lors des comités de thèse. J'associe à ces remerciements Catherine Chaussain pour avoir aussi participé à ces comités.

J'exprime toute ma gratitude à Agnès Bloch-Zupan qui m'a donné l'opportunité de participer au projet sur l'amélogénèse imparfaite qui a permis le financement de cette thèse.

Je remercie chaleureusement toute l'équipe "Evolution et Développement du Squelette" avec qui j'ai eu la chance de travailler.

À Tiphaine Davit-Béal, pour son aide et sa gentillesse. J'admire ton ambition et ta capacité à jongler avec la fac, l'hôpital, le labo, sans oublier la vie de famille.

À Jérémie Silvent, pour tous les bons moments passés au labo et dans les divers bureaux que l'on a partagés. Merci pour ton aide précieuse que tu continues à me donner malgré la distance.

À Amandine Leprévost, dont la présence rend les journées au bureau plus agréables. Merci de m'écouter, de me comprendre et de me supporter tous les jours.

À Marie-Claire Lajarille. Merci pour ton aide, en particulier pour avoir fait ces milliers de coupes de dents.

À Sidney Delgado, merci pour ta gentillesse et ton aide.

À Marion Chevrinai, et aux étudiants en master de passage au labo qui ont contribué à la bonne ambiance et sans qui les pauses bagel auraient eu moins de sens : Nicolas Cohadon, Natacha Delwarde, Suelen Paulino, Meriem Belheouane.

Je tiens également à remercier Anne-Gaëlle Lafont, pour m'avoir "passé le flambeau". Merci de m'avoir appris les bases de fonctionnement du labo qui m'ont été bien utiles.

TABLE DES MATIÈRES

REMERCIEMENTS	2
INTRODUCTION GÉNÉRALE	6
CHAPITRE 1 - REVUE DES CONNAISSANCES	12
I. Les phosphoprotéines sécrétées liant le calcium (SCPP)	13
II. Les SCPP de la dentine et de l'os, riches en acides aminés acides	15
III. Les SCPP riches en proline et en glutamine	15
1. Les SCPP de la salive et du lait	15
a) Les SCPP du lait	15
b) Les SCPP de la salive	16
2. Les SCPP de l'émail	16
a) Les protéines de la matrice de l'émail	17
i. Amélogénine	17
ii. Enaméline	21
iii. Améloblastine	23
b) Les SCPP jouant un rôle dans la maturation de l'émail et l'épithélium de jonction	28
i. ODontogenic, AMeloblast-associated protein	28
ii. Follicular Dendritic Cell-Secreted Protein	30
iii. Secretory Calcium-binding PhosphoProtein-Proline/Glutamine-rich 1	32
iv. Amélotine	34
CHAPITRE 2 - MATÉRIEL ET MÉTHODES	40
I. Matériel biologique	41
1. Espèces utilisées pour l'extraction d'acides nucléiques	41
2. Origine des échantillons pour l'extraction des acides nucléiques	42
a) Mammifères	42
b) Sauropsides	43
c) Amphibiens	43
3. Matériel pour l'hybridation <i>in situ</i>	43
a) <i>Anolis carolinensis</i>	43
b) <i>Pleurodeles waltl</i>	44
c) <i>Monodelphis domestica</i>	44
II. Méthodes	44
1. Histologie	44
a) Fixation et inclusion dans l'epon	44
b) Coupes et coloration	45
2. Biologie moléculaire	45
a) Extraction de l'ADN génomique	45
b) Extraction des ARN totaux de mâchoire	46
c) Conception des amorces pour la PCR	46
d) Polymerase Chain Reaction	47
i. Principe	47
ii. Protocole	48
e) RT-PCR en deux étapes	48
i. Synthèse de l'ADN complémentaire à partir des ARN	48

ii. Amplification des séquences d'intérêt	48
f) Electrophorèse sur gel d'agarose	49
i. Préparation du gel d'agarose 1%	49
ii. Dépôt et migration des échantillons	49
g) Purification des produits de PCR avant clonage	49
h) Clonage	50
i. Principe	50
ii. Ligation	50
iii. Transformation	50
iv. Culture et sélection des clones	51
i) Purification des plasmides	52
j) Linéarisation des plasmides et purification pour la transcription <i>in vitro</i>	52
k) Synthèse de sondes froides ARN	52
i. Transcription <i>in vitro</i> des sondes sens et antisens	52
ii. Purification des sondes	53
l) Préparation des échantillons pour l'hybridation <i>in situ</i>	53
i. Fixation et déminéralisation	53
ii. Déshydratation et inclusion	54
m) Hybridation <i>in situ</i> sur coupes	54
i. Traitement des coupes et hybridation	54
ii. Lavage des lames et immunoréaction	55
iii. Révélation	55
n) Composition des tampons pour l'hybridation <i>in situ</i>	56
i. Tampon d'hybridation	56
ii. Tampon de lavage	56
iii. MABT	56
iv. Solution de blocage	56
v. NTMT	57
3. Analyses <i>in silico</i>	57
a) Bases théoriques de l'analyse évolutive des protéines	57
b) Recherche des séquences dans les bases de données	58
c) Alignement de séquences	58
d) Modèle de substitution	58
e) Construction d'un arbre phylogénétique	59
f) Calcul de la séquence ancestrale	59
g) Calcul de la pression de sélection par la méthode « sliding window »	60
h) Calcul de la pression de sélection à chaque site par SLAC	60
i) Prédiction des peptides signaux	61
j) Prédiction des modifications post-traductionnelles	61

CHAPITRE 3 - Evolutionary analysis suggests that AMTN is enamel-specific and a candidate for AI **62**

Abstract	64
Introduction	64
Material et methods	65
Results	67
Discussion	73
Supplementary data	76

CHAPITRE 4 - Amelotin: an enamel matrix protein that experienced distinct evolutionary histories in amphibians, sauropsids and mammals	81
Abstract	83
Introduction	84
Material and methods	87
Results	91
Discussion	104
Conclusions	112
Supplementary data	115
CHAPITRE 5 - Amelotin gene structure and expression in the opossum <i>Monodelphis domestica</i>	123
Abstract	126
Introduction	127
Material and methods	128
Results	130
Discussion	134
CHAPITRE 6 - Comparative expression of the ameloblast-secreted genes, amelogenin, ameloblastin, enamelin and amelotin during amelogenesis in the lizard <i>Anolis carolinensis</i>	139
Abstract	141
Introduction	142
Material and methods	145
Results	147
Discussion	152
DISCUSSION - CONCLUSION	157
I. Evolution de la structure génique d'amélotine	158
II. Evolution de l'expression d'amélotine chez les tétrapodes	159
III. Expression des gènes codant les EMP chez <i>A. carolinensis</i>	161
IV. Amélotine était présent chez l'ancêtre commun des sarcoptérygiens	164
V. Amélotine est spécifique de l'émail	165
VI. Amélotine gène candidat pour l'amélognèse imparfaite ?	165
RÉFÉRENCES BIBLIOGRAPHIQUES	167
Liste des communications	187
I. Publications non incluses dans la thèse	187
II. Communication orales et affichées	187
1. Congrès nationaux	187
2. Congrès internationaux	188

INTRODUCTION GÉNÉRALE

Le recrutement des dents sur les mâchoires des premiers gnathostomes est une innovation majeure dans l'évolution des vertébrés, notamment grâce à leur rôle fondamental dans la prédation. Elles sont composées d'une cavité pulpaire entourée de dentine, recouverte, sauf exceptions, d'un tissu protecteur hyperminéralisé, généralement de l'émailloïde chez les chondrichthyens et les actinoptérygiens, ou de l'émail chez les sarcoptérygiens (Smith, 1989; Wakita, 1993; Donoghue & Sansom, 2002; Donoghue *et al.*, 2006).

Le développement dentaire se divise en quatre stades mettant en jeu des interactions épithélio-mésenchymateuses : bourgeon, capuchon, cloche et couronne (Fig. 1) (Thesleff, 2003). C'est lors de cette dernière étape que se déroule la formation de l'émail, ou l'amélogénèse, elle-même généralement répartie en deux phases principales : une phase de sécrétion, puis une phase de maturation (Moradian-Oldak, 2012).

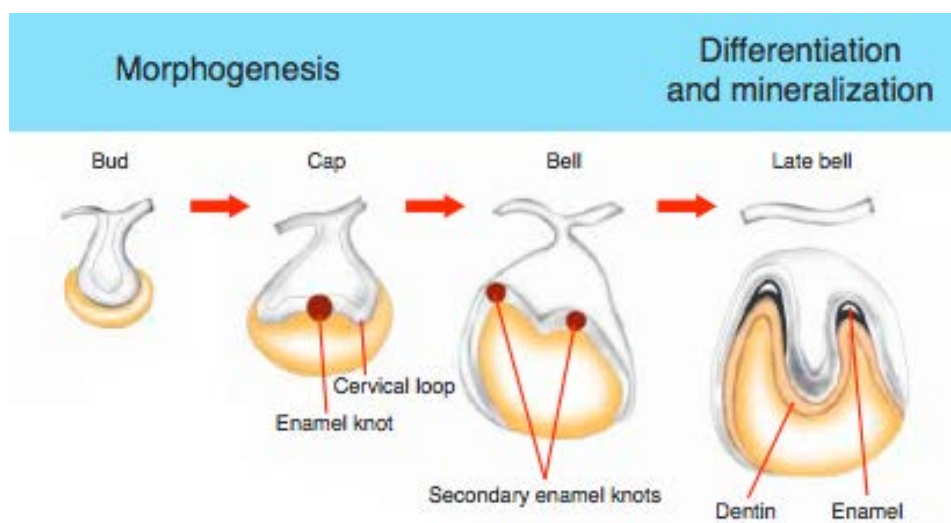


Figure 1. Les quatre principales étapes du développement dentaire.
D'après Thesleff, 2003.

C'est au contact de la première couche de dentine synthétisée par les odontoblastes, que les cellules de l'épithélium dentaire interne se différencient en améloblastes et commencent la synthèse de la matrice de l'émail : c'est la phase de sécrétion (Fig. 2a). La matrice de l'émail est principalement composée de trois protéines : amélogénine (AMEL), améloblastine (AMBN) et énaméline (ENAM). Ces protéines

ont la capacité d'initier la formation de cristaux de phosphate de calcium et d'en contrôler la forme et sont clivées, peu après leur sécrétion, par une métalloprotéinase, la MMP20. A la fin de l'étape de sécrétion, l'émail immature a atteint son épaisseur finale, mais n'est minéralisé qu'à environ 30% (Moradian-Oldak, 2012). Commence alors la phase de maturation de l'émail pendant laquelle 50% des améloblastes disparaissent par apoptose et les 50% restant se raccourcissent et s'élargissent (Smith & Warshawsky, 1977). Une lame basale contenant d'autres protéines secrétées par les améloblastes, Odontogenic, Ameloblast-associated protein (ODAM), Secretary Calcium-binding PhosphoProtein rich in Proline et Glutamine 1 (SCPP-PQ1) et amélotine (AMTN), est synthétisée entre la surface de l'émail et les améloblastes (Ganss & Abbarin, 2014). Une seconde protéase, la kallikréine 4 (KLK4), est synthétisée par les améloblastes et dégrade les fragments protéiques restant dans la matrice amélaire, permettant ainsi la croissance en largeur des cristaux (Fig. 2b). A la fin de la maturation, l'émail mature est composé de 96% d'hydroxyapatite, de 3,2% d'eau et de seulement 0,8% de matière organique (Smith, 1998).

Les cristaux d'hydroxyapatite de l'émail peuvent être arrangés en prismes ou non. Chez les mammifères la structure de l'émail est dite "prismatique" et elle est dite "non prismatique" chez les autres vertébrés (Fig. 3). La structure prismatique est due à la présence de prolongements cytoplasmiques au pôle distal des améloblastes, appelés prolongements de Tomes (Line & Novaes, 2005). Ces prolongements sont responsables de la formation de cristaux dans différentes orientations formant, d'une part, les prismes d'émail et, d'autre part, l'émail interprismatique. Chaque prisme est sécrété par un améloblaste unique, depuis la jonction énamélo-dentinaire jusqu'à la surface de la dent, traversant donc toute l'épaisseur de l'émail. En revanche, l'émail interprismatique est formé par plusieurs améloblastes voisins (Kallenbach, 1973). C'est l'imbrication des prismes et de l'émail interprismatique qui forme l'émail prismatique, plus résistant que l'émail non prismatique.

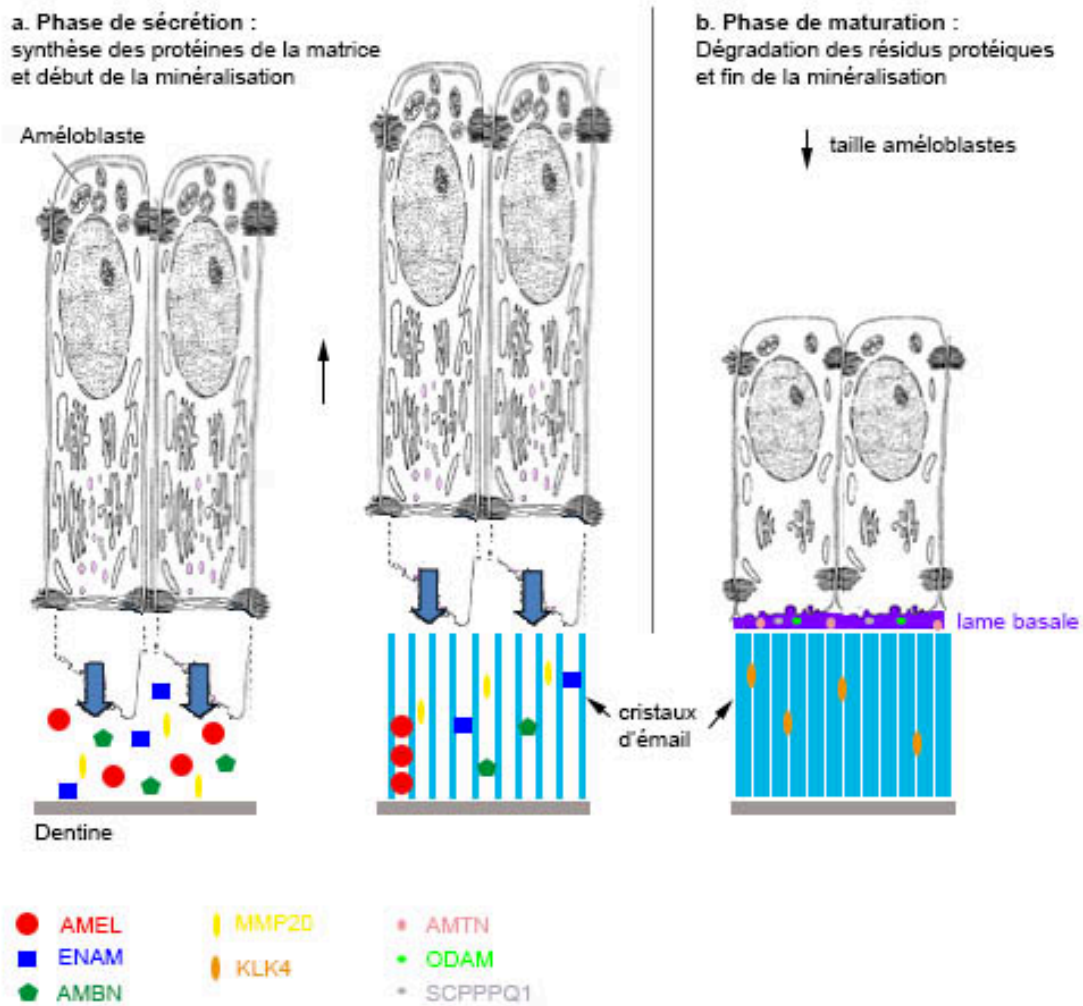
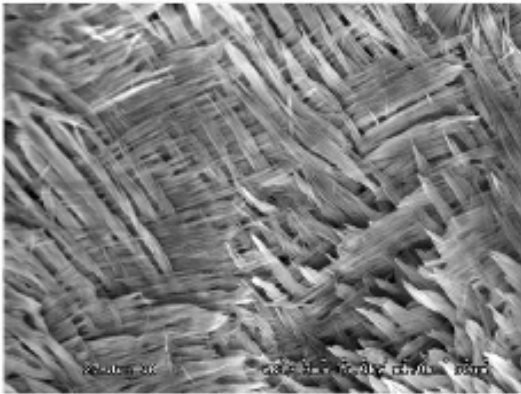


Figure 2. Schéma illustrant les deux phases de l'amélogénèse.

Chez les sarcoptérygiens non-mammaliens, les améloblastes sont dépourvus de prolongements de Tomes et l'émail n'est pas prismatique, ce qui est donc la condition ancestrale. De manière générale, les cristaux d'émail sont perpendiculaires à la surface de l'émail et parallèles entre eux (Fig. 3). Toutefois, en dépit de la différence de structure, les principales étapes de l'amélogénèse sont similaires chez tous les tétrapodes possédant de l'émail, qu'il soit prismatique ou non (Assaraf-Weill *et al.*, 2013).

Bos taurus



Rana pipiens

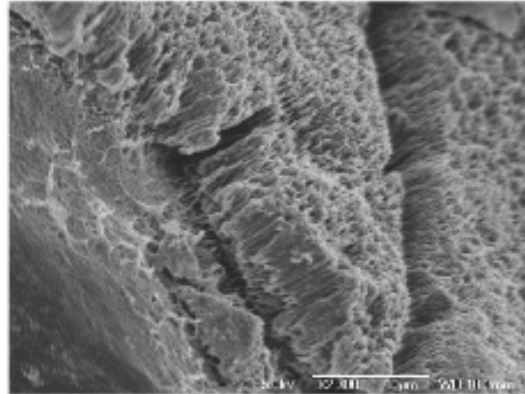


Figure 3. Images de microscopie électronique à balayage de l'émail de vache (*Bos taurus*) et de grenouille (*Rana pipiens*). (Diekwisch *et al.*, 2009)

La transition entre l'émail non prismatique et l'émail prismatique des mammifères a eu lieu très tôt dans l'évolution des mammifères sans pour autant que nous en ayons une explication biologique (Osborn & Hillman, 1979). Nous avons vu ci-dessus que les prolongements de Tomes en sont probablement les acteurs principaux mais nous ne savons pas la raison pour laquelle ces prolongements sont apparus. En revanche, la formation d'un émail prismatique chez les mammifères coïncide avec l'apparition de l'occlusion, permettant la mastication et ayant probablement contribué à la réduction du nombre de dents et de générations dentaires. On voit tout l'intérêt que confère dans ces conditions la présence d'un émail prismatique car il permet sur le long terme une plus grande résistance à l'abrasion et aux microfractures. Dans les dents des non-mammaliens, qui sont remplacées tout au long de leur vie et ne servent pas à la mastication, l'émail, peut être moins résistant car non prismatique, n'est pas soumis aux mêmes contraintes d'usure et de longévité.

Les protéines de la matrice de l'émail sécrétées pendant l'amélogenèse (AMEL, AMBN et ENAM) font partie d'une famille de protéines impliquées dans la minéralisation de l'os ou de la dent, les phosphoprotéines sécrétées liant le calcium (SCPP, pour Secretory calcium-binding PhosphoProtein) (Kawasaki & Weiss, 2003). L'équipe "Evolution et Développement du Squelette (EDS)", dirigée par Jean-Yves Sire, s'intéresse depuis de nombreuses années à l'origine et à l'évolution de cette famille de protéines et plus particulièrement celles qui sont sécrétées par les améloblastes.

Les trois protéines de la matrice de l'émail AMEL, AMBN et ENAM sont connues depuis plus de 30 ans et ont fait l'objet de beaucoup d'attention, majoritairement chez les mammifères, notamment en raison de leur implication dans une maladie génétique humaine, l'amélogénèse imparfaite. En effet, de nombreuses mutations ont été identifiées sur les gènes codant ces trois protéines et ont pour conséquence d'importants désordres, parfois très invalidants, lors de la formation de l'émail (Bartlett, 2013).

Plus récemment, d'autres protéines appartenant à la famille des SCPP, et secrétées par les améloblastes, ont été découvertes. Il s'agit d'AMTN, ODAM, et SCPP-PQ1. Ces protéines ont très peu été étudiées, et uniquement chez les rongeurs. Leur rôle dans l'amélogénèse reste à ce jour très peu connu (Ganss & Abbarin, 2014).

Dans ce contexte, l'objectif de l'équipe EDS est d'étendre les connaissances sur ces protéines et les gènes qui les codent, à toutes les lignées de mammifères et aux tétrapodes non mammaliens (amphibiens et sauropsides), afin de mieux comprendre leur évolution. Pour ma part, j'ai concentré mes recherches sur l'une de ces SCPP, l'amélotine, en utilisant un des modèles animaux du laboratoire, le lézard *Anolis carolinensis*.

Le premier chapitre de ma thèse est consacré à une revue des connaissances acquises à ce jour sur les SCPP secrétées par les améloblastes; il est suivi d'un chapitre détaillant les matériels et méthodes utilisés. Les résultats de mes recherches sont ensuite présentés dans quatre chapitres sous la forme d'articles, et je termine le manuscrit par une discussion-conclusion générale.

Le premier chapitre de résultats est consacré à l'étude de l'évolution de l'amélotine chez les mammifères. En effet, au début de ce travail de thèse, *AMTN* n'était connu que chez quelques mammifères (homme, rat et souris) et il était nécessaire d'accumuler des connaissances sur cette protéine dans cette lignée dont le génome de nombreux représentants était séquencé ou en cours de séquençage. L'analyse évolutive moléculaire d'*AMTN* à partir de 42 séquences de mammifères obtenues *in silico*, et représentant 220 millions d'années d'évolution, a permis de mieux connaître la structure du gène et d'identifier les acides aminés et les régions importantes pour la fonction de la protéine (Gasse *et al.*, 2012). Qu'en était-il chez les non mammaliens? *AMTN* était seulement identifiée dans l'unique génome de sauropside séquencé à ce moment là : *Anolis carolinensis*. Il était donc intéressant de confirmer cette séquence

à partir d'ADNc et de rechercher les séquences d'autres sauropsides et d'amphibiens, afin d'étudier, d'une part, l'évolution de ce gène chez les tétrapodes et, d'autre part, son expression par hybridation *in situ* au cours de l'amélogénèse chez un tétrapode non-mammalien. Les résultats de cette étude ont fait l'objet d'un deuxième article (Gasse *et al.*, 2015).

Récemment, Kawasaki et Amemiya (2014) ont prédit, à partir de recherches *in silico*, une séquence de l'opossum *Monodelphis domestica* différente de celle que j'avais prédite dans mon premier article sur l'analyse évolutive d'*AMTN* chez les mammifères (Gasse *et al.*, 2012). Je venais aussi de montrer que le patron d'expression d'*AMTN* chez les non-mammaliens est différent de celui qui avait été décrit chez la souris. Afin de (i) vérifier la séquence du transcrit et (ii) savoir si les modifications du patron d'expression ont été acquises chez un mammifère ancestral, ou plus tard dans la lignée des mammifères, j'ai pu récupérer la séquence d'*AMTN* dans un transcriptome de mâchoire séquencé dans l'équipe et j'ai réalisé des hybridations *in situ* chez un nouveau né de cette espèce de marsupial. Un article a été soumis sur ce sujet (Gasse *et al.*, soumis.).

Les différences importantes observées à la fois dans la structure et dans l'expression d'*AMTN* chez les mammifères et les tétrapodes non-mammaliens, m'ont conduit à vérifier si de telles différences existaient aussi pour les trois gènes codant les protéines de la matrice de l'émail, *AMEL*, *ENAM* et *AMBN*. J'ai donc réalisé une étude comparative de leur expression au cours de l'amélogénèse chez *A. carolinensis*. Les résultats ont été regroupés sous la forme d'un article (Gasse & Sire, en prep.).

Je tiens à m'excuser par avance auprès des lecteurs pour les inévitables redondances inhérentes au choix de la présentation des chapitres de résultats (3 à 6) de ma thèse sous la forme d'article.

CHAPITRE 1

REVUE DES CONNAISSANCES

I. Les phosphoprotéines sécrétées liant le calcium (SCPP)

L'émergence des tissus minéralisés tels que l'os et la dent représente une innovation importante dans l'évolution des vertébrés. La formation de ces tissus est liée à la présence d'une famille de protéines possédant la capacité de fixer le calcium, les phosphoprotéines sécrétées liant le calcium (SCPP, pour Secretory Calcium-binding PhosphoProtein).

Les événements de duplication de génome qui ont eu lieu très tôt dans la lignée des vertébrés, et aussi des duplications de gènes en tandem ont permis la diversification et l'évolution de cette famille. En effet, leurs gènes ont été recrutés à la suite de duplications à partir d'un gène ancestral, supposé être apparenté à SPARC-L1 (Secreted Protein Acidic and Rich in Cysteine - like 1) (Kawasaki *et al.*, 2004).

Bien que les protéines appartenant à cette famille ne présentent pas de similarité de séquences évidentes, elles partagent des caractéristiques qui témoignent de leur parenté. Comme ce sont des protéines sécrétées, elles possèdent un peptide signal, et celui-ci est très conservé au cours de l'évolution. De plus, elles possèdent un ou plusieurs sites de phosphorylation SXE (Ser-Xaa-Glu où Xaa représente n'importe quel acide aminé), l'un étant codé par la région 3' de l'exon 3 (Fig. 4). Ces protéines lient le calcium via leurs résidus phosphorylés.

La structure des gènes des SCPP possède d'autres caractéristiques. La région non traduite en 5' (5' UTR pour UnTranslated Region) est composée d'un (ou deux) exon suivi d'un exon (généralement l'exon 2) dont la structure est invariable chez les SCPP: une courte partie 5' non traduite, suivie d'une partie codant le peptide signal, puis les deux premiers acides aminés de la protéine mature. Tous les introns sont en phase 0 (Fig. 4) à quelques exceptions près, dans la région 3'. Les exons peuvent donc être épissés sans modifier la phase de lecture.

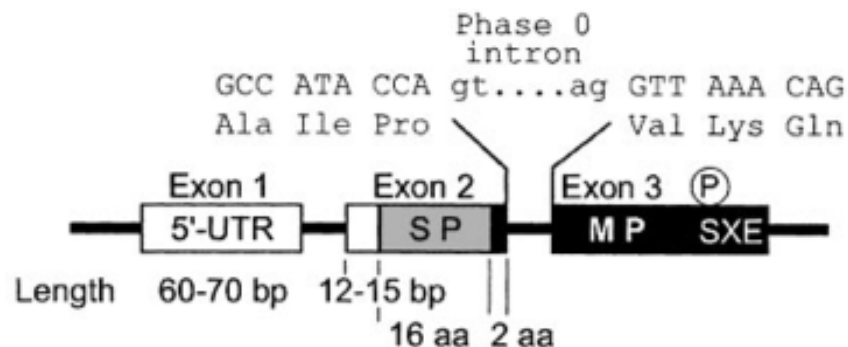


Figure 4. Caractéristiques communes aux SCPP (Kawasaki & Weiss, 2006).

Chez les mammifères les gènes des SCPP, à l'exception du gène *AMEL*, résident tous sur un même chromosome, et sont distribués en deux clusters de part et d'autre de *SPARC-L1* (Fig. 5). Ces deux clusters regroupent, d'une part, les gènes codant des protéines riches en acides aminés acides ("acid-rich SCPP cluster"), impliquées dans la minéralisation de l'os et de la dentine, et, d'autre part, des gènes codant des protéines riches en proline et glutamine ("P/Q-rich SCPP cluster"), comprenant les protéines impliquées dans la minéralisation de l'émail, les caséines du lait et des protéines salivaires (Kawasaki & Weiss, 2003, 2006). Chez les sauropsides (reptiles et oiseaux), les deux clusters résident sur des chromosomes différents (Fig. 5) (Al-Hashimi *et al.*, 2010).

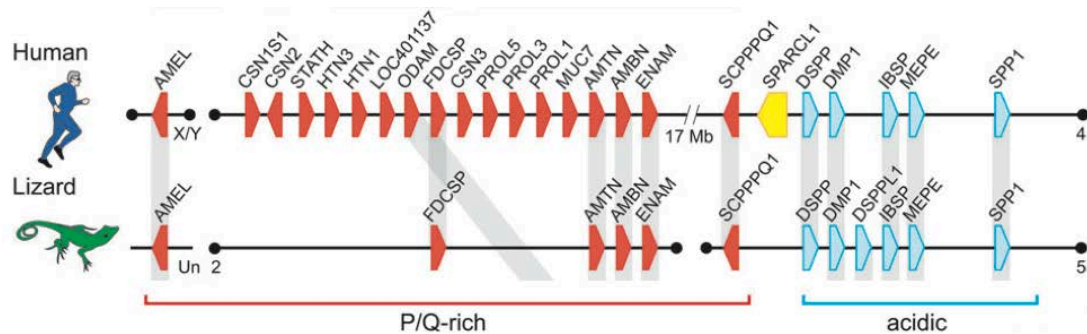


Figure 5. Localisation chromosomique des gènes des SCPP chez l'homme et le lézard. D'après Kawasaki et Amemiya, 2014.

II. Les SCPP de la dentine et de l'os, riches en acides aminés acides

Chez l'homme, cinq SCPP se trouvent dans cette catégorie: la sialophosphoprotéine dentinaire (Dentin sialophosphoprotein - DSPP), la phosphoprotéine acide 1 de la matrice dentinaire (Dentin Matrix acidic Phosphoprotein 1 - DMP1), la sialoprotéine liant les intégrines (Integrin-Binding SialoProtein - IBSP), aussi connue sous le nom de sialoprotéine de l'os (Bone SialoProtein - BSP), la phosphoglycoprotéine de la matrice extracellulaire (Matrix Extracellular, Phosphoglycoprotein - MEPE), et la phosphoprotéine 1 sécrétée (Secreted PhosphoProtein 1 - SPP1) aussi connue sous le nom d'ostéopontine (OPN). Ces protéines sont impliquées dans la minéralisation de l'os et/ou de la dentine et ont un pI très acide, en raison de leur composition riche en acides aminés acides (Glu et Asp) et à la présence de nombreux résidus phosphorylés (Kawasaki *et al.*, 2004). Elles sont également caractérisées par la présence d'un motif RGD (Arg-Gly-Asp) de liaison aux intégrines qui leur valut le nom de "Small

Integrin-Binding LIgand, N-linked Glycoprotein" (SIBLING) (Fisher *et al.*, 2001; Fisher & Fedarko, 2003).

Chez le lézard *Anolis carolinensis*, ces cinq gènes sont arrangés de la même manière que chez l'homme. Toutefois, un sixième gène similaire à DSPP, DSPP-Like1, se situe entre DMP1 et IBSP (Fig. 5) (Kawasaki, 2011).

III. Les SCPP riches en proline et en glutamine

Cette catégorie de SCPP regroupe les SCPP de la salive et du lait, et celles de l'émail. Il en existe 18 chez l'homme, et seulement six ont été identifiées à ce jour chez le lézard (Fig. 5) (Kawasaki, 2011). FDC-SP, bien qu'étant présente dans la salive chez l'homme (Guo *et al.*, 2006) et classifiée dans les SCPP de la salive et du lait par Kawasaki et Weiss (2006), est incluse dans cette revue dans les SCPP de l'émail. En effet, *FDCSP* est également exprimé dans les tissus dentaires et n'est pas exclusivement trouvé chez les mammifères, comme c'est le cas pour les SCPP de la salive et du lait.

1. Les SCPP de la salive et du lait

Ces protéines sont connues uniquement chez les mammifères. On en dénombre onze chez l'homme (Fig. 5).

a) Les SCPP du lait

Ce sont les caséines. Elles fournissent du phosphate de calcium aux nouveaux-nés pour le développement de leurs os et leurs dents. Il en existe de deux types : celles qui sont sensibles au calcium et celles qui ne le sont pas. Parmi les caséines sensibles au calcium on trouve les caséines α 1, α 2 et β codées par les gènes CSN1S1, CSN1S2 (invalidé chez l'homme) et CSN2 respectivement. Il n'y a qu'une seule caséine insensible au calcium, la caséine κ codée par le gène CSN3. Cette dernière a la capacité de stabiliser les complexes (caséines sensibles au calcium - phosphates de calcium) et de les maintenir en solution dans le lait (Kawasaki *et al.*, 2011). Tous les gènes des caséines ont leur dernier exon non traduit. Cette caractéristique est commune aux SCPP riches en P/Q situés entre CSN1S1 et CSN3 et à SCPP-PQ1 (Fig. 5) (Kawasaki, 2009). Les caséines sensibles au calcium dériveraient toutes de SCPP-PQ1, tandis que la caséine insensible au calcium dériverait de FDC-SP. SCPP-

PQ1 et FDC-SP provenant toutes deux d'ODAM, celle-ci ci serait donc l'ancêtre commun de toutes les caséines (Kawasaki *et al.*, 2011).

b) Les SCPP de la salive

Les SCPP de la salive protègent les dents en régulant la précipitation des sels de phosphate de calcium à la surface de l'émail. Certaines possèdent également des propriétés antibactériennes et antifongiques. Il en existe huit chez l'homme.

La stathérine (*STATH*), les histatines 1 et 3 (*HTN1* et *HTN3*), la mucine 7 (*MUC7*), les protéines Proline-rich 1 (*PROL1* ou *BPLP*), 3 (*PROL3* ou *SMR3B*), 5 (*PROL5* ou *SMR3A*) et 27 (*LOC401137* ou *PRR27*). *STATH*, *HTN1* et *HTN3* partagent, avec les gènes codants des SCPP du lait, un exon non traduit en 3'. Ces trois gènes dériveraient de *CSNIS2* (Kawasaki & Weiss, 2003). Ils n'ont été identifiés que chez les primates, leur duplication serait donc très récente (Sabatini *et al.*, 1993). Chez la souris le pseudogène de *STATH* a été identifié, indiquant que *STATH* a probablement dupliqué de *CSNIS2* avant la séparation de la lignée des rongeurs (Kawasaki & Weiss, 2003). La duplication entre *STATH* et *HTN* est estimée à 40-50 millions d'années (Ma) tandis que celle entre *HTN1* et *HTN3* est de 15-30 Ma (Sabatini *et al.*, 1993). Les gènes codant les protéines Proline-rich 1, 3, 5 et 27 sont exprimés à la fois par les glandes salivaires et les glandes lacrymales (Kawasaki & Weiss, 2006). Les gènes *PROL1*, *PROL3*, *PROL5* et *MUC7* partagent la même structure génique et sont situés entre *AMTN* et *CSN3* (Kawasaki & Weiss, 2003) tandis que *LOC401137*, situé entre *ODAM* et *HTN1*, a l'exon non traduit en 3' comme les caséines, *STATH*, *HTN1* et *HTN3* (Kawasaki & Weiss, 2006).

2. Les SCPP de l'émail

Les SCPP de l'émail sont au nombre de sept chez l'homme. Elles sont classées en deux catégories : les protéines de la matrice de l'émail (*EMP* pour "Enamel Matrix Protein"), *AMEL*, *ENAM* et *AMBN*, et les SCPP jouant un rôle dans la maturation de l'émail et l'épithélium de jonction : *AMTN*, *ODAM*, *SCPP-PQ1* et *FDC-SP* (Ganss & Abbarin, 2014). Chez le lézard, seulement six SCPP de l'émail ont été identifiés dans le génome (Fig. 5). En effet, *ODAM* n'a pas été trouvé dans le génome du lézard, alors qu'il est présent dans celui du xénope (Kawasaki, 2011).

a) Les protéines de la matrice de l'émail

i. Amélogénine

Caractérisation

Dès les années 1980, la séquence protéique complète de l'amélogénine (AMEL) a été déterminée chez la souris (Snead *et al.*, 1983), la vache (Takagi *et al.*, 1984), l'homme (Catalano-Sherman *et al.*, 1993), le rat (Bonass *et al.*, 1994), le cochon (Hu *et al.*, 1996a), l'opossum (Hu *et al.*, 1996b), et le hamster (Lyaruu *et al.*, 1998). AMEL représente 90% de la fraction protéique de l'émail bovin en formation (Termine *et al.*, 1980). C'est la seule donnée quantitative que nous possédons à ce jour.

Le gène de l'amélogénine réside, sur le brin antisens, dans un intron du gène *ARHGAP6* chez tous les sarcoptérygiens (Kawasaki & Amemiya, 2014). Chez les mammifères placentaires, il est localisé sur les chromosomes sexuels, tandis que chez les marsupiaux, les monotrèmes et les non-mammaliens il est autosomal (Sire *et al.*, 2005). *AMEL* est composé de sept exons chez les mammifères thériens (marsupiaux et placentaires). Cependant l'exon 4 n'est fonctionnel que chez certaines espèces de placentaires (homme, souris, rat, vache et cochon) et peut être épissé (Sire *et al.*, 2012). De plus, trois exons supplémentaires ont été identifiés dans le génome de rongeurs en aval de l'exon 7, les exons 4b, 8 et 9. Les exons 4b et 8 sont les homologues des exons 4 et 5 (Bartlett *et al.*, 2006). L'exon 4b n'a jamais été trouvé dans les transcrits. Les exons 8 et 9 ne sont présents que dans l'ARNm des muridés après épissage alternatif de l'exon 7 (Li *et al.*, 1998; Baba *et al.*, 2002; Papagerakis *et al.*, 2005; Sire *et al.*, 2012). L'épissage alternatif des transcrits d'*AMEL* est très fréquent et permet de générer une grande hétérogénéité dans les protéines qui en résultent (Lau *et al.*, 1992). En amont du gène, la région du promoteur d'*AMEL* a été caractérisée chez quelques mammifères (Salido *et al.*, 1992; Chen *et al.*, 1994). Un peptide signal de 16 acides aminés est codé par l'exon 2, et l'exon 3 code pour un site de phosphorylation SXE. Le grand exon 6 code pour une région hydrophobe contenant de nombreuses répétitions de triplets PXQ (Delgado *et al.*, 2005b). La région C-terminale est très chargée et possède un pI de 4,2, tandis que celui de la protéine entière est de 8,0 (Fincham *et al.*, 1999). AMEL ne possède qu'un seul site de modification post-traductionnelle : un site de phosphorylation sur Ser16 (Takagi *et al.*, 1984; Fincham & Moradian-Oldak, 1995). Une fois sécrétée la protéine est progressivement clivée par la métalloprotéinase 20 (MMP20) puis par la kallikréine 4 (KLK4) (Nagano *et al.*, 2009; Bartlett, 2013).

La structure génique d'*AMEL* est également connue chez quelques sauropsides et amphibiens après séquençage de l'ADNc. Elle est composée de six exons: 1, 2, 3, 5, 6 et 7 (Fig. 6) (Ishiyama *et al.*, 1998; Toyosawa *et al.*, 1998; Wang *et al.*, 2005, 2006; Delgado *et al.*, 2006; Assaraf-Weill *et al.*, 2013). Trois cas d'épissage alternatif ont été reportés chez un reptile et un amphibien. Dans le premier cas, un exon supplémentaire a été identifié entre les exons 5 et 6 (Wang *et al.*, 2012a). Les deux autres cas proviennent de la même espèce d'amphibien. Un des variants possède un nouvel exon entre les exons 2 et 3 tandis que l'exon 6 est absent dans l'autre variant (Wang *et al.*, 2013). La séquence d'*AMEL* a aussi été identifiée dans le génome du coelacanth *Latimeria chalumnae*. Chez ce dernier *AMEL* est probablement composé de sept exons mais l'homologue du premier exon non traduit n'a pas été identifié. De plus l'exon 3 a été dupliqué (Fig. 6) (Kawasaki & Amemiya, 2014). Dans le génome d'espèces édentées (oiseaux, fourmiliers, pangolins, baleines) ou dépourvue d'émail (paresseux, tatous, oryctérope), *AMEL* est invalidé (Sire *et al.*, 2008; Meredith *et al.*, 2013, 2014), ce qui indique qu'*AMEL* est une protéine spécifique de l'émail.

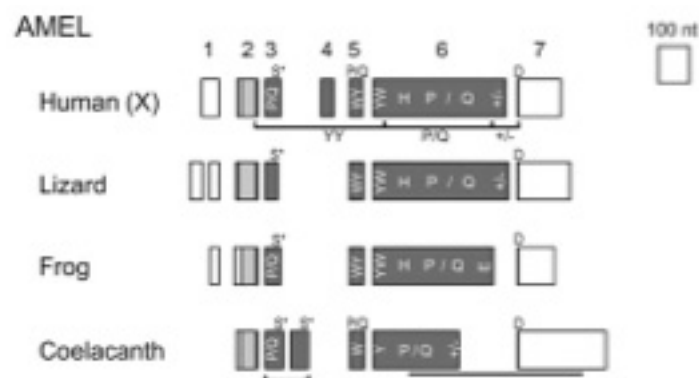


Figure 6. Structure génique d'*AMEL* chez les sarcoptérygiens.
D'après Kawasaki & Amemiya, 2014.

Expression

L'expression d'*AMEL* au cours de l'amélogenèse a été très étudiée chez les rongeurs. *AMEL* est exprimé dans les améloblastes à partir du stade de présécrétion jusqu'au début de la maturation de l'émail (Snead *et al.*, 1988; Nanci *et al.*, 1998; Bleicher *et al.*, 1999; Hu *et al.*, 2001a; Torres-Quintana *et al.*, 2005). Quelques études ont montré l'expression transitoire d'*AMEL* dans les odontoblastes de rat et de cochon (Oida *et al.*, 2002; Papagerakis *et al.*, 2003). Néanmoins, le niveau d'expression d'*AMEL* dans les odontoblastes est 1000 fois inférieur à celui des améloblastes (Oida *et al.*, 2002).

La localisation de la protéine est cohérente avec l'expression de l'ARNm dans les améloblastes. Chez le rat AMEL est localisée d'abord dans l'appareil de Golgi et dans les granules de présécrétion, puis dans les prolongements de Tomes et dans toute la couche de la matrice de l'émail pendant le stade de sécrétion et le début de la maturation de l'émail (Nanci *et al.*, 1998). Plus récemment, dans la dent humaine en formation, AMEL a été détectée dans les odontoblastes en cours de différenciation ainsi que dans la prédentine nouvellement déposée, et n'est plus exprimée une fois les odontoblastes matures. L'expression d'AMEL dans les odontoblastes est réactivée lors de la formation de dentine réparatrice après une plaie dentinaire (Mitsiadis *et al.*, 2014).

Chez les non-mammaliens, l'expression d'AMEL a été étudiée chez le lézard *Chalcides viridanus* (Fig. 7) (Delgado *et al.*, 2006a) et l'amphibien caudate *Pleurodeles waltl* (Assaraf-Weill *et al.*, 2013). Chez ces espèces, l'expression d'AMEL dans les améloblastes est similaire à celle qui a été observée chez les rongeurs. La localisation de la protéine a été étudiée chez des amphibiens et des squamates par immunohistochimie en utilisant des anticorps reconnaissants AMEL de mammifères (Herold *et al.*, 1989; Ishiyama *et al.*, 1998; Diekwisch *et al.*, 2002; Wang *et al.*, 2005), un anticorps reconnaissant spécifiquement AMEL de *Rana pipiens* (Diekwisch *et al.*, 2006), et ce même anticorps de grenouille chez trois squamates (Handrigan & Richman, 2011). Certaines de ces études ont également localisé la protéine dans les odontoblastes et les cellules pulpaire (Wang *et al.*, 2005; Diekwisch *et al.*, 2002, 2006; Handrigan & Richman, 2011). Bien qu'à ce jour, AMEL n'ait jamais été identifiée en dehors des sarcoptérygiens, Diekwisch *et al.* (2002) ont détecté la protéine et l'ARNm chez un requin en utilisant un anticorps et une sonde reconnaissants l'amélogénine de souris.

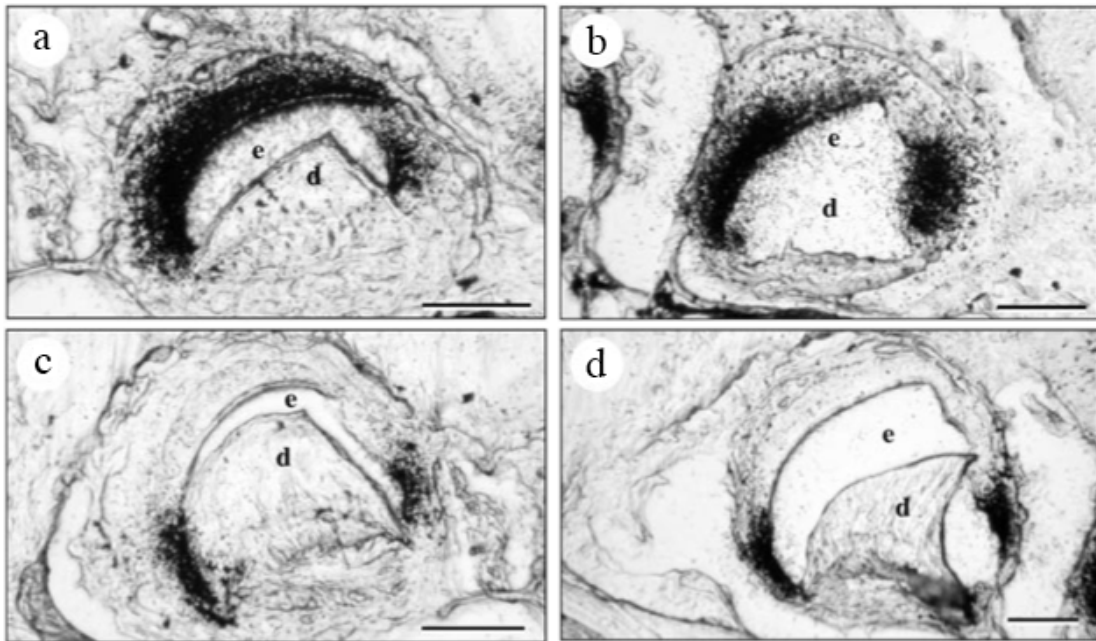


Figure 7. Expression d'AMEL par hybridation *in situ* au cours du développement dentaire chez le lézard *Chalcides viridanus*. D'après Delgado *et al.*, 2006.
 a. Début du dépôt de la matrice de l'émail. b. Dépôt et minéralisation de la matrice. c. Maturation. d. Fin de la maturation. .

Fonctions

AMEL possède un domaine de liaison AMEL-AMEL et est capable de s'auto-assembler pour former des "nanosphères" essentielles pour la formation de l'émail (Fincham *et al.*, 1995; Paine & Snead, 1997; Du *et al.*, 2005). AMEL interagit aussi avec AMBN, ODAM et ENAM *in vitro* (Ravindranath *et al.*, 2004; Fan *et al.*, 2009; Holcroft & Ganss, 2011).

Récemment, il a été montré que la région contenant les répétitions de polyprolines additionnée de la partie C-terminale chargée se lie fortement au phosphate de calcium et à l'hydroxyapatite, tandis que la partie N-terminale se lie préférentiellement au carbonate de calcium *in vitro* (Gopinathan *et al.*, 2014). De plus, ces auteurs ont montré que les parties C-terminale et celles qui sont riches en polyprolines agissent en synergie pour induire la formation de fins cristaux de phosphate de calcium, contrastant avec les larges cristaux fusionnés promus par la partie N-terminale (Gopinathan *et al.*, 2014). Les travaux de Guo *et al.* (2014) suggèrent qu'AMEL pourrait avoir un rôle dans la régulation du pH de l'émail en formation (Guo *et al.*, 2014). AMEL semble jouer un rôle dans l'organisation et la croissance en longueur des cristaux d'hydroxyapatite.

Les souris transgéniques chez lesquelles *AMEL* est inactivé présentent un émail hypoplasique désorganisé (Gibson *et al.*, 2001). Chez l'homme, des mutations sur *AMEL* entraînent une maladie génétique appelée amélogénèse imparfaite liée à l'X (AIH1) caractérisée par la formation d'un émail hypoplasique ou hypominéralisé (Wright *et al.*, 2003).

ii. Enaméline

Caractérisation

Le terme énaméline a été initialement utilisé pour désigner les protéines amélaïres de la catégorie des non-amélogénines, mais son usage est désormais restreint au gène de l'énaméline (*ENAM*) et la protéine qu'il code. *ENAM* est composé de dix exons dont huit sont codants (exons 3 à 10) (Fig. 8). L'UTR en 5' est composé des exons 1 et 2 et du début de l'exon 3, bien que l'exon 2 ne soit pas présent chez tous les mammifères placentaires (Hu *et al.*, 2001b; Al-Hashimi *et al.*, 2009). Chez les placentaires *ENAM* possède deux sites d'initiation de la traduction, l'un dans l'exon 3, l'autre, commun à tous les SCPP, dans l'exon 4. Lorsque la traduction débute au premier ATG, un long peptide signal de 39 acides aminés est codé par les exons 3 et 4 et diffère alors de celui des autres SCPP. Lorsqu'elle débute au second ATG situé dans l'exon 4, via l'épissage alternatif de l'exon 3, elle aboutit à un peptide signal de 16 acides aminés (Al-Hashimi *et al.*, 2009). L'existence de ces deux peptides signaux pourrait être utilisée pour réguler la vitesse de sécrétion d'*ENAM*.

Un exon supplémentaire situé entre les exons 8 et 9, nommé exon 8b, a été découvert chez deux marsupiaux et deux reptiles, mais pas chez l'ornithorynque ni la grenouille (Al-Hashimi *et al.*, 2010). Chez les non-mammaliens, l'exon 3 est absent et le site d'initiation de la traduction est celui de l'exon 4, plus commun (Al-Hashimi *et al.*, 2010). *ENAM* est un pseudogène dans le génome d'espèces ayant perdu leurs dents ou dont les dents sont dépourvues d'émail, ce qui indique que ce gène est spécifique de l'émail (Meredith *et al.*, 2009, 2013, 2014; Al-Hashimi *et al.*, 2010). *ENAM* a été identifié dans le génome du coelacanthe (Fig. 8) (Kawasaki & Amemiya, 2014).

La première publication de la séquence protéique d'*ENAM* de porc ne correspondait en fait qu'au fragment de 32kDa, le produit de clivage le plus abondant dans l'émail en formation (Tanabe *et al.*, 1990). En effet, la protéine *ENAM* est sécrétée d'abord sous forme d'un précurseur de 186 kDa qui subit rapidement une série de clivages protéolytiques par MMP20 (Yamakoshi *et al.*, 2006). Les clivages débutent dans la

partie C-terminale de la protéine en générant des fragments de 155, 142, 89, 34, 32 et 25 kDa (Fukae *et al.*, 1996). Le produit de clivage de 32 kDa est protégé de l'action de MMP20 et s'accumule dans toute l'épaisseur de la matrice de l'émail; il est dégradé par KLK4 au début du stade de maturation (Yamakoshi *et al.*, 2006). Trois sites de phosphorylation et trois sites de N-glycosylation sont conservés dans la partie N-terminale d'ENAM de mammifères. Un motif RGD est codé par l'exon 10 chez l'homme. Cependant, ce motif n'est pas présent chez tous les mammifères (Al-Hashimi *et al.*, 2009).

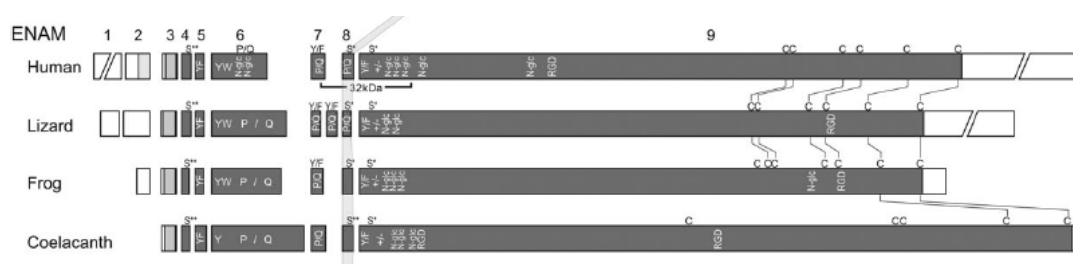


Figure 8. Structure génétique d'ENAM chez les sarcoptérygiens.
(Kawasaki & Amemiya, 2014)

Expression et fonction

Chez la souris l'expression du gène est détectée dans les améloblastes à partir du stade de présécrétion jusqu'au début de la maturation de l'émail, et n'a jamais été identifiée dans les odontoblastes (Hu *et al.*, 2000, 2001a). Chez le porc le fragment protéique de 32 kDa est retrouvé aux mêmes stades dans toute la matrice de l'émail (Dohi *et al.*, 1998). Aucune étude n'a montré d'expression génique ou protéique chez un non-mammalien. Plusieurs études ont suggéré que le fragment de 32kDa pourrait avoir un rôle nucléateur de la minéralisation des cristaux d'émail (Tanabe *et al.*, 1990; Bouropoulos & Moradian-Oldak, 2004). Ce fragment subit des modifications de conformation en fonction de la concentration en Ca^{2+} du milieu (Fan *et al.*, 2008) et est capable de se lier avec AMEL *in vitro* (Fan *et al.*, 2009).

Une lignée de souris transgéniques dans laquelle le gène *ENAM* a été inactivé et remplacé par le gène rapporteur *LacZ* présente une fine couche d'émail non minéralisé à la surface de ses dents (Hu *et al.*, 2008). Une seconde lignée de souris *ENAM*^{-/-} ne reproduit pas le même phénotype. En effet, chez ces souris aucun émail n'est produit et des kystes sont formés entre les améloblastes et la dentine (Sawada *et al.*, 2011). Chez l'homme, des mutations dans *ENAM* ont été identifiées comme étant

responsables d'une amélogénèse imparfaite autosomale dominante (AIH2) (Hu & Yamakoshi, 2003).

iii. Améloblastine

Caractérisation

L'améloblastine (AMBN) a été initialement identifiée chez le porc en 1995 par Uchida et collaborateurs sous le nom de sheath, puis sheathlin, comme étant une nouvelle protéine de la matrice de l'émail de la catégorie des non-amélogénines (Uchida *et al.*, 1995). La même année, un autre groupe identifia un nouveau transcrite dans les améloblastes d'incisive de rat (Matsuki *et al.*, 1995) qui fut caractérisé et nommé améloblastine (Krebsbach *et al.*, 1996). Deux variants, appelés amelin 1 et amelin 2 ont été identifiés dans une librairie d'ADNc de molaires de rat (Černý *et al.*, 1996). La séquence d'AMBN de la souris (Simmons *et al.*, 1998) et de l'homme (MacDougall *et al.*, 2000; Toyosawa *et al.*, 2000) fut séquencée, suivie par la publication de celle du cochon d'inde et de la vache (soumission directe à la base de donnée NCBI).

Chez l'homme, le gène est composé de treize exons, 1 à 13 (Fig. 9). L'analyse des cinq autres séquences géniques de mammifères publiées (souris, rat, cochon d'inde, vache et cochon) révèle seulement onze exons, indiquant que la séquence humaine est atypique avec deux exons supplémentaires. En effet, chez l'homme trois petits exons presque identiques de 39 pb chacun (exons 7, 8 et 9), sont présents tandis que seulement l'un d'entre eux est retrouvé dans les autres séquences. Pour des raisons pratiques, cet unique exon fut nommé exon 7, et les suivants les exons 8 et 9 furent considérés comme résultant de deux duplications successives de la région d'ADN contenant l'exon 7. Les autres exons d'AMBN sont alors numérotés de 1 à 7, puis de 10 à 13 chez les autres espèces. Deux transcrits alternatifs résultant d'un épissage intraexonique du début de l'exon 6 ont été identifiés chez le porc, la souris et l'homme (Hu *et al.*, 1997; Simmons *et al.*, 1998; MacDougall *et al.*, 2000). Le site d'initiation de la transcription et le promoteur d'AMBN ont été caractérisés chez la souris (Dhamija *et al.*, 1999). En amont du promoteur se situent de nombreux éléments cis impliqués dans la régulation de l'activité transcriptionnelle des cellules mésenchymateuses *in vitro* (Tamburstuen *et al.*, 2011b), ainsi qu'un site de liaison avec le facteur de transcription *Cbfa1*, essentiel pour la différenciation ostéoblastique (Dhamija & Krebsbach, 2001). Dans la région non traduite en 3' il existe trois signaux

de polyadénylation différents chez le rat et la souris, et deux chez l'homme (Hu *et al.*, 1997; Simmons *et al.*, 1998; MacDougall *et al.*, 2000; Toyosawa *et al.*, 2000).

En dehors des mammifères, seulement deux séquences de l'ADNc d'AMBN ont été publiées : celles d'un sauropside, le crocodile *Caiman crocodilus* (Shintani *et al.*, 2002), et d'un amphibien, le xénope *Xenopus laevis* (Shintani *et al.*, 2003). Le gène est constitué de onze et dix exons, respectivement (Fig. 9). AMBN a été récemment identifié dans le génome du coelacanthe *Latimeria chalumnae* (Kawasaki & Amemiya, 2014). Chez cette espèce le gène est constitué de douze exons. Dix exons sont homologues aux exons des AMBN de tétrapodes mais deux autres, situés entre les exons 11 et 12, ne sont homologues à aucun exon identifié chez les tétrapodes (Fig. 9) (Kawasaki & Amemiya, 2014).

Deux sites potentiels d'initiation de la traduction sont présents dans la région 5' des séquences d'AMBN de mammifères connues à ce jour (Krebsbach *et al.*, 1996; Dhamija *et al.*, 1999; Hu *et al.*, 1997). Il est possible que ces deux sites soient fonctionnels, comme cela a été proposé pour ENAM (Al-Hashimi *et al.*, 2009). Le peptide signal est alors soit composé de 26 acides aminés codés par l'exon 1 (5 acides aminés) et l'exon 2 (21 acides aminés), soit 16 acides aminés codés uniquement par l'exon 2. La séquence du crocodile ne possède que le deuxième site (Shintani *et al.*, 2002). Chez les rongeurs, AMBN est composée de 421 acides aminés (en considérant la première méthionine codée par l'exon 1) et a une masse moléculaire de 65 kDa (62 kDa une fois le peptide signal clivé) (Krebsbach *et al.*, 1996). AMBN possède de nombreux sites d'hydroxylation, de phosphorylation, de O-glycosylation et de myristoylation (Uchida *et al.*, 1997; Yamakoshi *et al.*, 2001; Hu *et al.*, 2005; Kobayashi *et al.*, 2007). Comme AMEL et ENAM, AMBN est clivée en plusieurs fragments par MMP20 puis par KLK4 (Simmer & Bartlett, 2004; Iwata *et al.*, 2007; Bartlett, 2013). La protéine prédite est faiblement acide avec un pI de 6,5 (Simmons *et al.*, 1998). La région N-terminale est basique et hydrophobe, la région C-terminale est acide et hydrophile, et le reste de la protéine est variable avec une alternance de régions hydrophiles et hydrophobes (Toyosawa *et al.*, 2000). AMBN est une protéine riche en proline (P, 15,2%), et plus particulièrement la région codée par les exons 5 et 6. AMBN contient également une grande proportion de leucine (L, 9,9%), d'alanine (A, 9%) et de glycine (G, 9,9%).

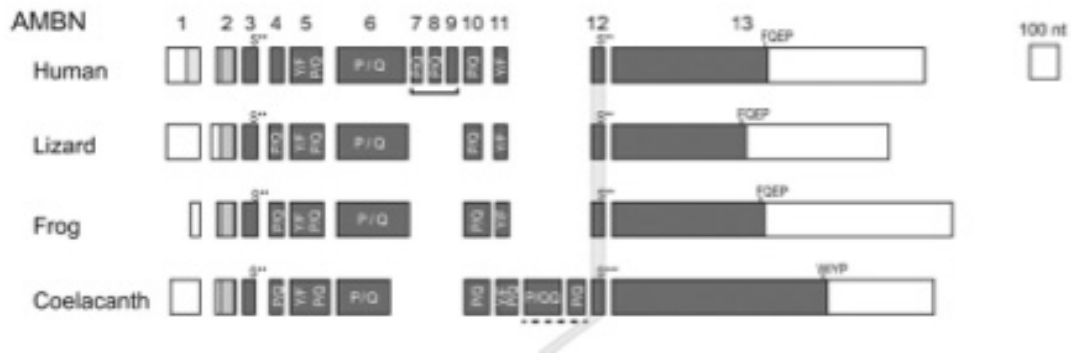


Figure 9. Structure génique d'AMBN chez les sarcoptérygiens.
(Kawasaki & Amemiya, 2014)

Expression et fonctions

Chez l'homme et les rongeurs, les ARNm d'AMBN sont détectés dans la partie distale du cytoplasme des améloblastes, avec une intense localisation dans les prolongements de Tomes à partir du stade présécrétoire, atteignant un pic d'expression pendant le stade de sécrétion, et diminuant pendant la maturation jusqu'à l'éruption (Lee *et al.*, 1996; Černý *et al.*, 1996; MacDougall *et al.*, 2000; Torres-Quintana *et al.*, 2005). Dans la matrice de l'émail en formation, la quantité de molécules d'AMBN représente environ 7% de celle d'AMEL (Černý *et al.*, 1996). La protéine a été localisée par immunohistochimie à la jonction amélo-dentinaire. De plus, des peptides de la région N-terminale, générés après clivage d'AMBN, ont été trouvés dans les prismes d'émail (Uchida *et al.*, 1997). De manière surprenante, dans des molaires de rat des transcrits d'AMBN ont été détectés dans les cellules de la gaine épithéliale de Hertwig avant la formation de l'émail, ce qui suggère un rôle dans la cémentogenèse et la formation des racines (Fong *et al.*, 1996; Ge *et al.*, 2013). Néanmoins, la protéine n'a pas été localisée dans le cément en formation mais uniquement à courte distance de celui-ci, à l'endroit où l'émail s'étend entre la dentine et le cément (Bosshardt & Nanci, 1998; Fong *et al.*, 1998). Il a été également montré qu'un court peptide de la partie N-terminale d'AMBN augmente l'expression de la phosphatase alcaline et de la BSP et stimule la différenciation des cellules du ligament parodontal et la minéralisation (Kitagawa *et al.*, 2011).

L'expression d'AMBN a été détectée transitoirement dans les pré-odontoblastes et les jeunes odontoblastes avant l'initiation de l'amélogénèse (Bègue-Kirn *et al.*, 1998; Fong *et al.*, 1998). Cela suggère qu'AMBN pourrait agir comme molécule de signalisation entre les améloblastes et les odontoblastes dans l'initiation de la

formation de l'émail mais aussi dans la formation de la dentine primaire (Spahr *et al.*, 2002; Nakamura *et al.*, 2006). Des transcrits d'*AMBN* ont été également identifiés dans une lignée de cellules odontoblastiques de rat *in vitro* (Hao *et al.*, 2005). Il est important de souligner que l'expression d'*AMBN* a aussi été détectée pendant le développement des os craniofaciaux chez le rat (Spahr *et al.*, 2006) et dans des ostéoblastes et ostéoclastes en culture (Tamburstuen *et al.*, 2010, 2011a; Iizuka *et al.*, 2011). Ces résultats suggèrent un rôle pour *AMBN* dans la formation osseuse et le remodelage osseux. Cependant, des souris KO pour le gène de l'*AMBN* ne présentent aucun phénotype osseux apparent (Fukumoto *et al.*, 2004). De plus, le gène est invalidé chez les oiseaux et chez les mammifères ayant perdu la capacité à former des dents ou de l'émail, comme les xénarthres, les pangolins, les baleines (Deméré *et al.*, 2008; Sire *et al.*, 2008). L'expression est restreinte aux dents chez les rongeurs et chez le crocodile (Fig. 10) (Krebsbach *et al.*, 1996; Shintani *et al.*, 2006).

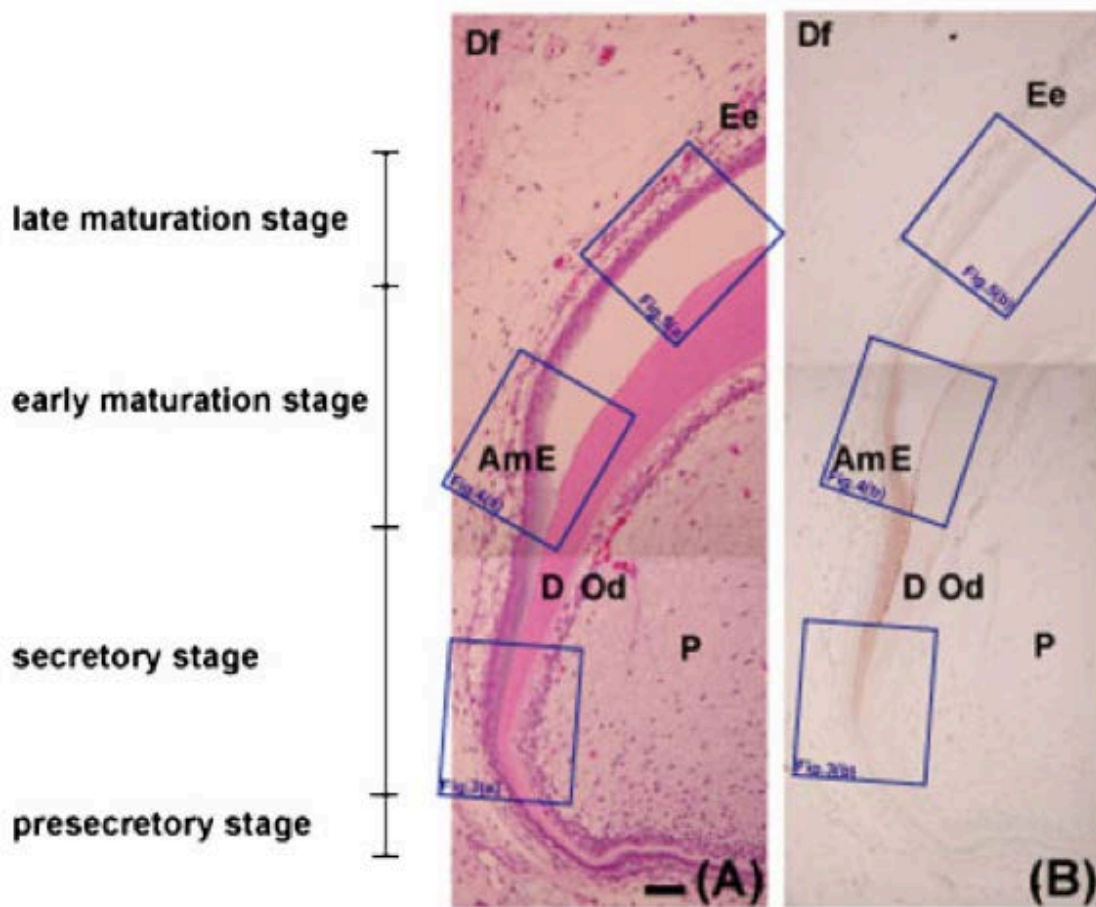


Figure 10. Localisation d'*AMBN* par immunohistochimie dans un bourgeon dentaire de *Caiman crocodilus*. A. Coloration à l'hématoxyline et à l'éosine. B. Immunomarquage. (Shintani *et al.*, 2006)

Dans sa région C-terminale, AMBN possède deux sites de liaison au calcium (Yamakoshi *et al.*, 2001; Vymětal *et al.*, 2008; Zhang *et al.*, 2011a) et un site de liaison à l'héparine (motif VTKG) (Černý *et al.*, 1996; Fukumoto *et al.*, 2005). AMBN régule l'attachement cellulaire via RhoA, une Rho-GTPase qui à son tour contrôle la prolifération cellulaire via p27, un inhibiteur de kinase de la famille Cip/Kip (Zhang *et al.*, 2011b). Des domaines d'interaction avec CD63, une glycoprotéine de surface, ont aussi été identifiés dans la région C-terminale d'AMBN. Ces domaines pourraient jouer un rôle dans la régulation de la différenciation cellulaire via la liaison de CD63 à l'intégrine beta1 (Iizuka *et al.*, 2011; Zhang *et al.*, 2011a). De plus, AMBN interagit avec le motif trityrosyl de la région N-terminale d'AMEL, formant un assemblage hétéromoléculaire qui promeut la minéralisation (Fong *et al.*, 1996; Ravindranath *et al.*, 2004). Il existe aussi potentiellement une interaction entre le fragment d'AMBN de 37 kDa (sans la partie N-terminale) et cette même région d'AMEL (Ravindranath *et al.*, 2007).

Les souris *AMBN*^{-/-} développent une hypoplasie sévère de l'émail et les améloblastes se détachent de la matrice. Ces défauts suggèrent qu'AMBN est essentielle pour la différenciation des améloblastes et la formation de l'émail (Fukumoto *et al.*, 2005). Chez les souris transgéniques double mutantes pour *AMBN* et *AMEL*, une fine couche d'émail se forme mais est très réduite en comparaison avec les souris simple mutantes. La surexpression d'*AMBN* chez les souris transgéniques perturbe l'agencement des cristaux de minéral, et la morphologie des prismes et interprismes est modifiée, résultant en un émail plus fin et poreux (Paine *et al.*, 2003). La surexpression de la protéine recombinante inhibe la prolifération de cellules d'améloblastome humain, et cette inhibition requiert les sites de liaison à l'héparine, suggérant une régulation du signal cellulaire (Sonoda *et al.*, 2009).

AMBN a été détecté dans certaines tumeurs odontogéniques et osseuses (Hirayama *et al.*, 2010; Iizuka *et al.*, 2011; Ren *et al.*, 2011, 2012; Crivelini *et al.*, 2012). A ce jour, une seule mutation homozygote dans le gène *AMBN*, impliquant la délétion de l'exon 6, a été détectée comme étant responsable d'une forme hypoplasique récessive de l'amélogénèse imparfaite (Poulter *et al.*, 2014).

b) Les SCPP jouant un rôle dans la maturation de l'émail et l'épithélium de jonction

i. ODontogenic, AMeloblast-associated protein

L'ARNm d'*ODAM* a été initialement identifié dans une librairie d'ADNc d'odontoblastes d'incisive de rat sous le nom OD-314 (Dey *et al.*, 2001). La protéine fut ensuite identifiée dans des dépôts d'amyloïde d'échantillons de tumeurs odontogéniques calcifiées humaines et nommée APin (pour Amyloid in Pindborg tumors) (Solomon *et al.*, 2003). Trois ans plus tard, l'ADNc fut cette fois identifié dans le sécrétome d'organe de l'émail d'incisive de rat (EO-009) et reconnu comme étant l'homologue d'Apin chez l'homme (Moffatt *et al.*, 2006a). Le gène est composé de douze exons, le premier et le dernier n'étant pas codants (Fig. 11). Un peptide signal est codé par l'exon 2. La protéine mature a un pI de 4,5 et est riche en glutamine (16%), proline (12%), leucine (9%) et serine (7%). La protéine possède un site de phosphorylation SXE codé par la région 3' de l'exon 3. Les introns sont tous en phase 0 (Moffatt *et al.*, 2006a).

ODAM est connu chez les mammifères, le xénope, le coelacanth, et même chez le poisson zèbre (Fig. 11) (Kawasaki, 2009; Kawasaki & Amemiya, 2014). En revanche, le gène n'a pas été trouvé dans le génome des sauropsides (crocodiles et squamates).

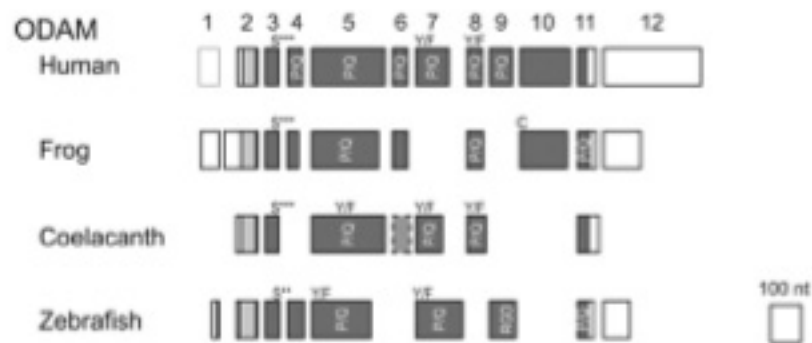


Figure 11. Structure génique d'*ODAM* chez les ostéichthyens.
(Kawasaki & Amemiya, 2014)

Expression et fonctions

Lors de l'amélogenèse, l'ARNm d'*ODAM* est exprimé dans les améloblastes. L'expression est faible pendant le stade de sécrétion, s'intensifie pendant la phase de transition et est forte pendant la maturation de l'émail (Park *et al.*, 2007; Moffatt *et al.*, 2008). La protéine est localisée dans le noyau des améloblastes pendant le stade

de sécrétion (Lee *et al.*, 2010), puis dans l'appareil de Golgi (Park *et al.*, 2007; Moffatt *et al.*, 2008) et à l'interface entre les améloblastes et la surface de l'émail dans la lame basale à partir du stade de transition post-sécrétoire et pendant la maturation de l'émail (Moffatt *et al.*, 2008; Dos Santos Neves *et al.*, 2012; Nishio *et al.*, 2013). La protéine est aussi détectée dans des cellules de la gaine épithéliale de Hertwig incrustées dans le ciment (Nishio *et al.*, 2010a). L'épithélium de jonction attaché à l'émail exprime également le gène (Park *et al.*, 2007) et la protéine (Moffatt *et al.*, 2008; Nishio *et al.*, 2010b, 2013). Après destruction mécanique de l'épithélium de jonction, l'expression d'ODAM diminue mais retrouve son niveau normal quand la régénération de l'épithélium de jonction est terminée. ODA M n'est donc pas impliquée dans la régénération de l'épithélium de jonction (Jue *et al.*, 2014). En revanche, après gingivectomie ODA M a été détecté dans les restes cellulaires épithéliaux de Malassez et les auteurs en déduisent un rôle possible dans la régénération parodontale (Nishio *et al.*, 2010a). ODA M serait impliquée dans la différenciation des améloblastes via une cascade de signalisation BMP2-BMPR-IB-ODAM-MAPK durant laquelle BMPR-IB phosphoryle un motif SXE dans la partie C-terminale d'ODAM en présence de BMP2 (Lee *et al.*, 2012a).

Des expériences *in vitro* ont montré que la surexpression et la sous-expression d'ODAM entraînent respectivement une augmentation et une diminution de l'expression de MMP20 et de la tufteline, une protéine ubiquitaire exprimée par les améloblastes lors de la formation de l'émail (Deutsch *et al.*, 2002). ODA M serait donc impliquée dans la régulation de l'expression de ces protéines (Park *et al.*, 2007). Lee *et al.* (2010) proposent un scénario dans lequel ODA M, dont l'expression est activée par Runx2, se fixe sur le promoteur de MMP20 pendant le stade de sécrétion et régulerait ainsi sa production. Pendant la maturation, ODA M n'est plus localisée dans le noyau et jouerait donc un autre rôle (Lee *et al.*, 2010). Des expériences de cribles double hybride ont permis de déterminer qu'ODAM interagissait avec elle-même, ainsi qu'avec AMTN, AMBN, plus faiblement avec AMEL mais pas avec ENAM (Holcroft & Ganss, 2011). En dehors des dents, ODA M est exprimé dans les glandes nasales et salivaires et dans l'épididyme de rat (Moffatt *et al.*, 2008). Chez l'homme, ODA M est détectée dans une multitude de tumeurs (Solomon *et al.*, 2003; Kestler *et al.*, 2008, 2011; Murphy *et al.*, 2008; Siddiqui *et al.*, 2009; Crivelini *et al.*, 2012; Lee *et al.*, 2012b) et pourrait jouer un rôle dans l'invasion des métastases (Lee *et al.*, 2010; Kestler *et al.*, 2011).

ii. Follicular Dendritic Cell-Secreted Protein

Caractérisation

Un ADNc codant pour une protéine jusqu'alors inconnue a été identifié dans des cellules dendritiques folliculaires d'amygdales humaines. La séquence protéique correspondante possède un peptide signal dans sa partie N-terminale. Cette protéine fut alors nommée Follicular Dendritic Cell-Secreted Protein (Marshall *et al.*, 2002). Le gène *FDCSP*, aussi appelé *C4orf7* (pour chromosome 4 open reading frame 7) est composé de cinq exons et s'étend sur 10 kb sur le chromosome 4 chez l'homme (Fig. 12). L'exon 1 et le début de l'exon 2 constituent la région non traduite en 5'. L'exon 2 code pour le peptide signal et les deux premiers acides aminés de la protéine mature. Les exons 3 et 4 codent le reste de la protéine tandis que l'exon 5 est non codant. La séquence protéique est constituée de 85 acides aminés chez l'homme et possède une région relativement riche en proline (18,8% chez l'homme, 21,4% chez la souris) (Marshall *et al.*, 2002). *FDCSP* a été identifié dans le génome de l'opossum *M. domestica* (Kawasaki *et al.*, 2007) et dans le génome du lézard *A. carolinensis* (Kawasaki, 2011). Il proviendrait d'une duplication d'*ODAM* (Kawasaki *et al.*, 2011).

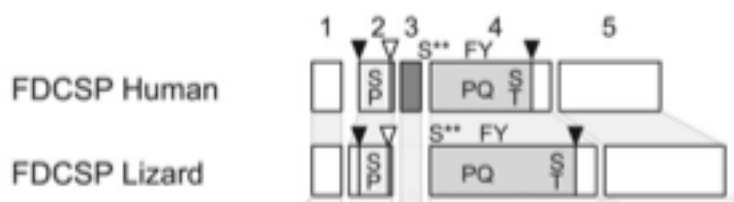


Figure 12. Structure génique de *FDCSP* chez l'homme et le lézard.

(Kawasaki *et al.*, 2011).

Expression et fonctions

L'expression de l'ARNm a été détectée par Northern blot dans les amygdales, les ganglions lymphatiques, la trachée et la prostate, et en plus faible quantité dans la thyroïde, l'estomac, le colon et la rate. Dans les amygdales humaines, l'expression est localisée dans les centres germinatifs, dans les cellules dendritiques folliculaires mais pas dans les lymphocytes B. En dehors des centres germinatifs, l'expression est détectée dans les cryptes d'amygdales enflammées (Marshall *et al.*, 2002). La génération de lignées de souris transgéniques dépourvues d'expression de *FDCSP* ou la surexprimant a montré que *FDCSP* agissait directement sur les lymphocytes B pour réguler la production d'IgA (Hou *et al.*, 2014).

L'ARNm a aussi été détecté dans les glandes salivaires et mammaires humaines et bovines (Rijnkels *et al.*, 2003), ainsi que dans le ligament parodontal et la glande parotide de souris (Nakamura *et al.*, 2005). L'exposition *in vitro* de cellules de ligament parodontal à la FDCSP maintient la population de fibroblastes et inhibe leur différenciation en cellules ostéogéniques. FDCSP pourrait donc être impliquée dans le maintien de l'état non minéralisé du ligament parodontal (Wei *et al.*, 2011; Xiang *et al.*, 2014a; b).

L'épithélium de jonction exprime aussi bien l'ARNm (Fig. 13) (Shinomura *et al.*, 2008) que la protéine (Oshiro *et al.*, 2012; Takahashi *et al.*, 2014). En revanche, l'expression de la protéine est localisée au niveau du cytoplasme des cellules de l'épithélium de jonction, en contradiction avec la présence d'un peptide signal. Par ailleurs, l'expression de la protéine dans les cellules de l'épithélium de jonction disparaît trois heures après une inflammation, avant de réapparaître trois jours plus tard. De plus, Takahashi *et al.* (2014) ont montré par qPCR une diminution de l'expression des ARNm de FDCSP concomitante à l'augmentation de l'expression d'IL-17 chez des rats atteints de parodontite. FDCSP jouerait donc un rôle dans le système de défense de la gencive via les interleukines 17, impliquées dans le recrutement des neutrophiles (Oshiro *et al.*, 2012; Takahashi *et al.*, 2014). D'autres auteurs ont montré que FDCSP pouvait s'adsorber sur des particules d'hydroxyapatite *in vitro* et ont suggéré qu'elle serait impliquée dans le système de défense contre des infections de la cavité buccale en recouvrant la surface du ciment, de l'os alvéolaire et de l'émail (Shinomura *et al.*, 2008). Chez l'homme, la protéine a été retrouvée dans la salive (Guo *et al.*, 2006) et est utilisée en tant que marqueur spécifique de la salive pour identifier les fluides corporels dans les analyses médico-légales (Park *et al.*, 2013). FDCSP est aussi détectée dans certaines tumeurs et jouerait un rôle dans l'invasion cellulaire des métastases en réduisant l'adhésion cellulaire par l'intermédiaire de la phosphorylation d'Akt et l'inhibition de l'E-cadhérine (Wang *et al.*, 2010; Wu *et al.*, 2010).

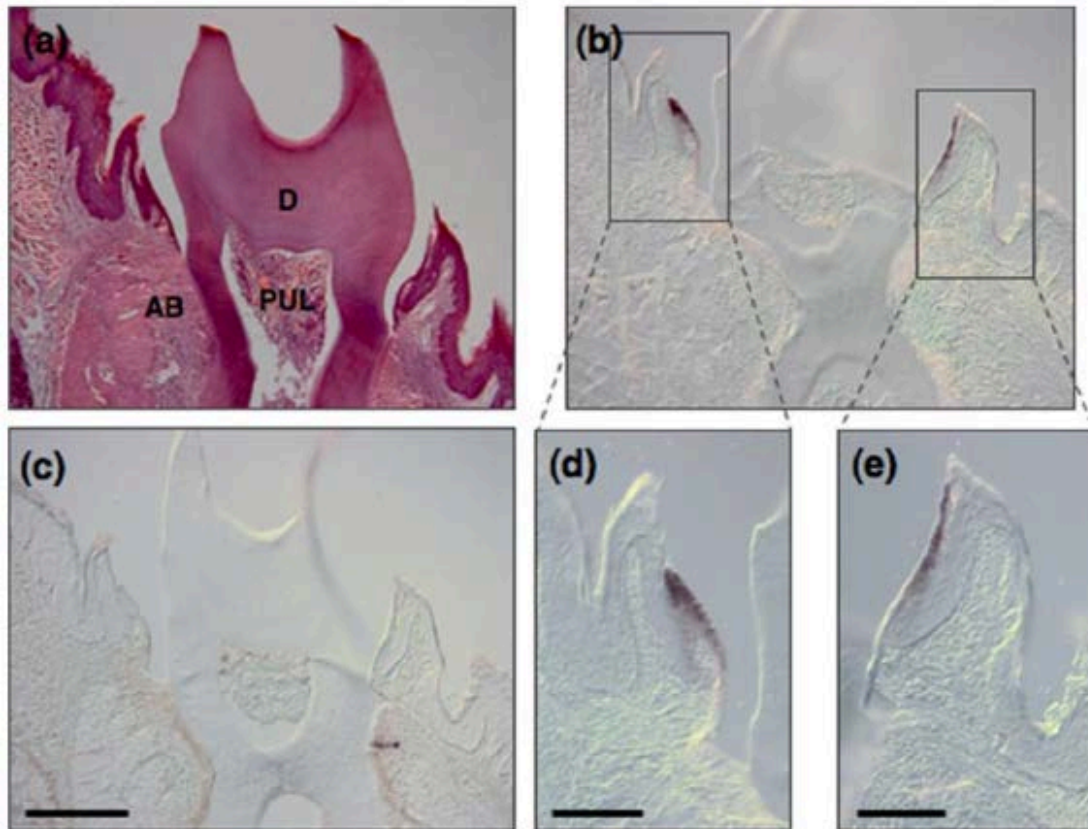


Figure 13. Expression de *FDCSP* par hybridation *in situ* dans l'épithélium gingival de molaire de souris. a. Coloration à l'hématoxyline. b. Expression de *FDCSP*. c. Contrôle négatif avec la sonde sens. d, e. Agrandissements des cadres de b. (Shinomura *et al.*, 2008).

iii. Secretory Calcium-binding PhosphoProtein-

Proline/Glutamine-rich 1

Caractérisation

SCPP-PQ1 a tout d'abord été identifiée en tant que EO-463 dans le sécrétome d'organe de l'émail du rat (Moffatt *et al.*, 2006a). L'ADNc a été cloné chez la souris, le rat et le lézard (Kawasaki, 2009; Kawasaki *et al.*, 2011; Moffatt *et al.*, 2014). Le gène est situé à côté de *SPARC-L1* chez l'homme (Kawasaki, 2009) et le rat (Moffatt *et al.*, 2014), et en amont des *SIBLING* chez le lézard (Fig. 5). Des recherches Blast à partir des séquences des exons de SCPP-PQ1 de l'homme ou du rat ont permis de prédire les séquences d'ADNc et de la protéine de seize autres mammifères. Le gène est composé de dix exons (Fig. 14). Le premier exon et le début de l'exon 2 ne sont pas traduits, comme chez la plupart des SCPP. Le clonage de l'ADNc de la souris a révélé la présence de deux transcrits de 45 pb de différence dans la région 5' UTR. L'exon 2 code pour un peptide signal de quinze résidus et les quatre premiers acides aminés de la protéine mature. La protéine est composée de 87 à 95 acides aminés.

Dépourvue de son peptide signal, elle a un pI proche de 6 et est riche en proline (20%), leucine (17%), glutamine (12%) et phénylalanine (10%). Elle est relativement hydrophobe (42% en masse de résidus AILFWV). Un site présumé de phosphorylation (motif SXXE), conservé chez la plupart des mammifères, est codé par la région 3' de l'exon 3 (Moffatt *et al.*, 2014). Les similarités de leur structure exonique et la distribution phylogénétique suggèrent que SCPP-PQ1, comme FDC-SP, dériverait d'ODAM (Kawasaki *et al.*, 2011).

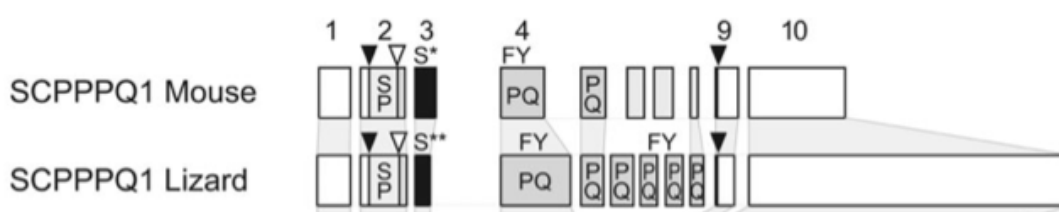


Figure 14. Structure génique de SCPP-PQ1 chez l'homme et le lézard.
(Kawasaki *et al.*, 2011)

Expression

Chez le rat les transcrits de *SCPP-PQ1* ont été détectés par Northern blot dans l'organe de l'émail au stade de maturation, dans la mandibule et dans la gencive. L'expression de *SCPP-PQ1* a été quantifiée par PCR quantitative. Elle est 20 fois plus élevée pendant la maturation tardive que pendant le début et le milieu de la maturation (Moffatt *et al.*, 2014). Dans l'incisive de rat et les molaires de souris, la protéine a été localisée par immunohistochimie dans les améloblastes et à l'interface entre leur surface apicale et l'émail pendant la maturation tardive de l'émail (Fig. 15). Les cellules de l'épithélium de jonction en formation sont également marquées. Ensuite, une fois l'épithélium formé, le marquage est localisé à l'interface entre les cellules et l'émail. Un immunomarquage aux particules d'or révèle la présence de la protéine dans la lame basale entre la surface de l'émail et les améloblastes au stade de maturation, et entre la surface de l'émail et l'épithélium de jonction après éruption (Moffatt *et al.*, 2014), comme ODAM et AMTN (Dos Santos Neves *et al.*, 2012).

La distribution spatiotemporelle de SCPP-PQ1 suggère un rôle dans la maturation tardive de l'émail, mais aussi dans le mécanisme d'attache des cellules épithéliales à la surface de la dent (Moffatt *et al.*, 2014).

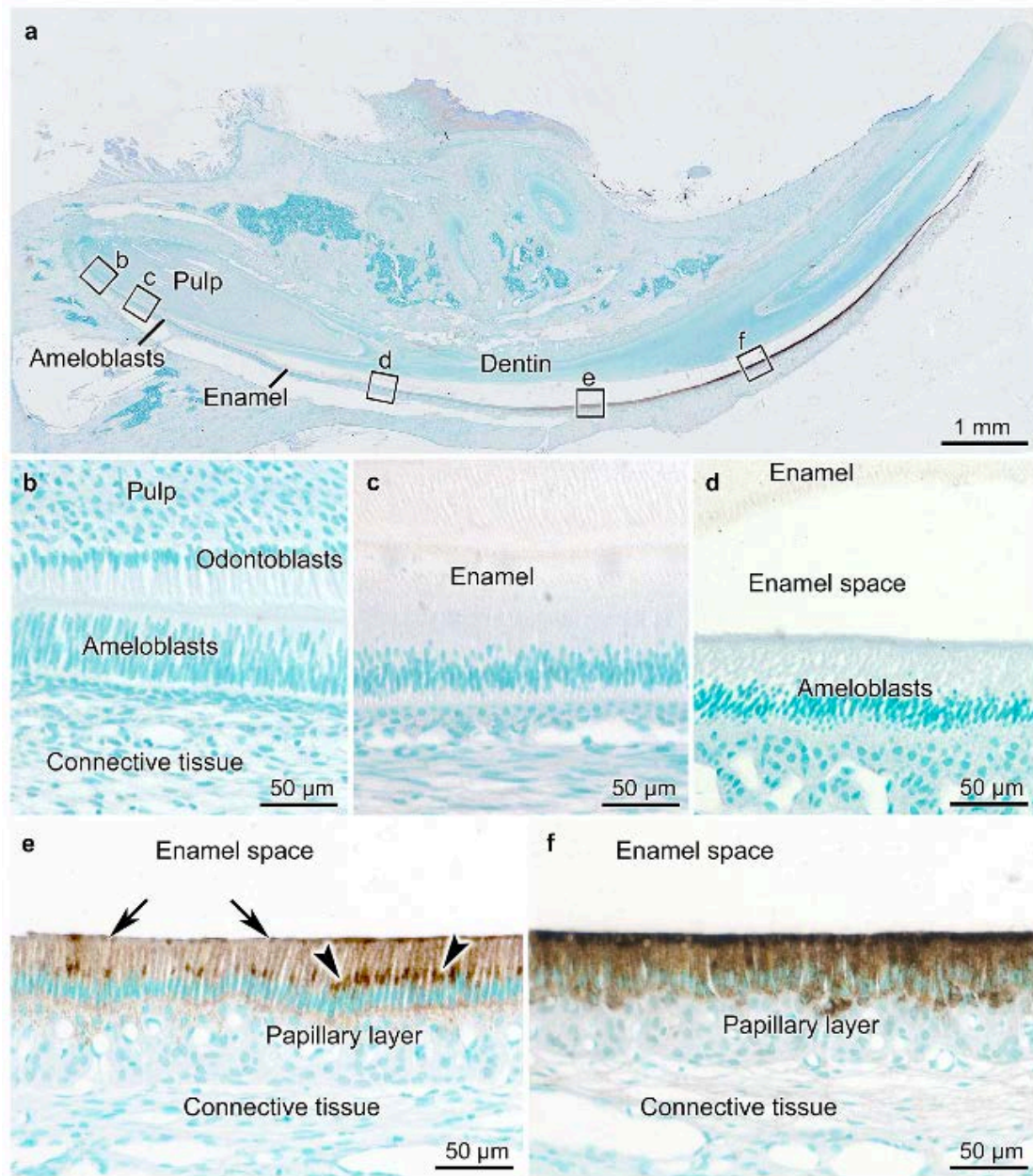


Figure 15. Localisation de SPP-PQ1 par immunohistochimie au cours de la formation de l'émail dans une incisive de rat. a. vue d'ensemble. b. stade de présécrétion. c. sécrétion. d. post-sécrétion. d. transition. e. début de la maturation. f. fin de la maturation.
(Moffatt *et al.*, 2014)

iv. Amélotine

Caractérisation

Amélotine (*AMTN*) a été découvert chez la souris par la technique de "differential display PCR" (DD-PCR) (Iwasaki *et al.*, 2005) et a été simultanément identifié dans le génome humain comme UNQ689 et son appartenance à la famille des SPP a été démontrée (Kawasaki & Weiss, 2006). En parallèle, l'équipe d'Antonio Nanci à

Montréal avait trouvé *AMTN* dans le sécrétome de l'organe de l'email de rat (EO-017) (Moffatt *et al.*, 2006a).

L'ARNm de l'*AMTN* de souris est composé de 979 nucléotides, codant 213 acides aminés. La masse moléculaire est de 22,1 kDa et son pI est prédit à 6,16 (Iwasaki *et al.*, 2005). Trois transcrits ont été identifiés chez le rat. Le plus long contient 1032 nucléotides. La protéine correspondante est composée de 212 acides aminés (pI 5,4). Les deux autres transcrits résultent de l'épissage de l'exon 7 pour l'un, et des exons 3 à 7 pour l'autre (Moffatt *et al.*, 2006b). La séquence protéique d'*AMTN* d'autres mammifères a été déterminée *in silico* : homme, chimpanzé, macaque, vache, cochon, chien, opossum. Chez toutes ces espèces le gène est composé de 9 exons (Moffatt *et al.*, 2006b; Trueb *et al.*, 2007). Seul le premier exon n'est pas codant. Les introns 1 à 7 sont en phase 0, mais le huitième est en phase 1. Un décalage de phase d'intron 0 à 1 a probablement été causé par une insertion ou délétion de nucléotides (Kawasaki & Weiss, 2006). Le promoteur du gène *AMTN* a été caractérisé chez la souris (Nakayama *et al.*, 2014a) et un site de liaison pour le miARN miR-1304 a été identifié dans la région non traduite en 3' chez les primates (Lopez-Valenzuela *et al.*, 2012). Par ailleurs, *AMTN* a aussi été identifié *in silico* dans le génome du lézard *Anolis carolinensis* (Kawasaki, 2011) et du coelacanthe *Latimeria chalumnae* (Kawasaki & Amemiya, 2014). Chez ces espèces, *AMTN* est prédit comme étant composé de onze et neuf exons, respectivement (Fig. 16). L'exon 1 n'a pas été identifié chez le coelacanthe. Deux exons supplémentaires par rapport à l'*AMTN* de mammifères, les exons 2b et 2c, ont été prédits entre les exons 2 et 3 chez ces deux espèces, et également un exon 3b situé entre les exons 3 et 4 chez le lézard. De plus, l'exon 9 est absent et l'exon 8 est beaucoup plus long que celui des mammifères. Cette découverte suggère l'apparition d'un site d'épissage dans l'exon 8, entraînant la perte de la région 3' de l'exon 8 et le recrutement de l'exon 9 chez les mammifères (Kawasaki & Amemiya, 2014). Ces auteurs proposent également une séquence d'*AMTN* d'opossum différente de celle qui avait été prédite par Moffatt *et al.* (2006). En effet, ils ont détecté la présence dans le génome d'un exon supplémentaire entre les exons 2 et 3, homologue à un exon de l'*AMTN* de lézard. De plus, ils prédisent que la séquence se termine avec le long exon 8 comme chez le lézard et le coelacanthe (Kawasaki & Amemiya, 2014). Le séquençage des transcrits permettrait de trancher entre ces deux hypothèses. Dans toutes les séquences d'*AMTN* de mammifères étudiées à ce jour, le peptide signal est codé par l'exon 2. Un site potentiel de phosphorylation SXXE

(SSEE) est présent dans la portion de la protéine codée par l'exon 7. Ce site de phosphorylation est prédit comme étant phosphorylé par la kinase CK2 (Casein kinase 2) (Moffatt *et al.*, 2006b). La kinase FAM20C phosphorylant *in vitro* le motif SXE présent dans d'autres SCPP (OPN, MEPE, IBSP et DMP1) (Ishikawa *et al.*, 2012; Tagliabracci *et al.*, 2012), il est donc possible qu'elle phosphoryle également AMTN. De plus, chez des souris knock-out pour le gène *FAM20C*, il a été montré que l'expression d'*AMTN* est abolie dans les améloblastes (Wang *et al.*, 2012b), ce qui suggère que FAM20C pourrait être aussi impliquée dans la régulation de la transcription d'*AMTN*. AMTN est riche en proline, leucine, glutamine et thréonine et contient des sites de O-glycosylation (Moffatt *et al.*, 2006b; Dos Santos Neves *et al.*, 2012).

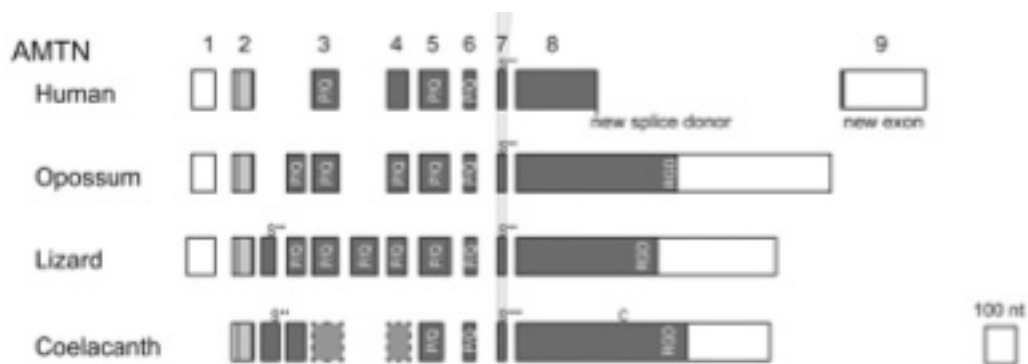


Figure 16. Structure génique d'*AMTN* chez les sarcoptérygiens.
(Kawasaki & Amemiya, 2014)

Expression et fonctions

L'expression de l'ARNm a été caractérisée par Northern blot à partir d'incisives de rat (Trueb *et al.*, 2007) et par hybridation *in situ* sur des coupes d'incisives et de molaires de souris (Iwasaki *et al.*, 2005). Ce gène est exprimé par les améloblastes au stade de maturation. Plus précisément, son expression débute lors de la transition entre les stades de sécrétion et de maturation, caractérisée par un raccourcissement des corps cellulaires des améloblastes et l'apparition d'une zone où les améloblastes se détachent de la matrice (Reith, 1961). L'expression décline en fin de maturation dans les améloblastes réduits (Fig. 17).

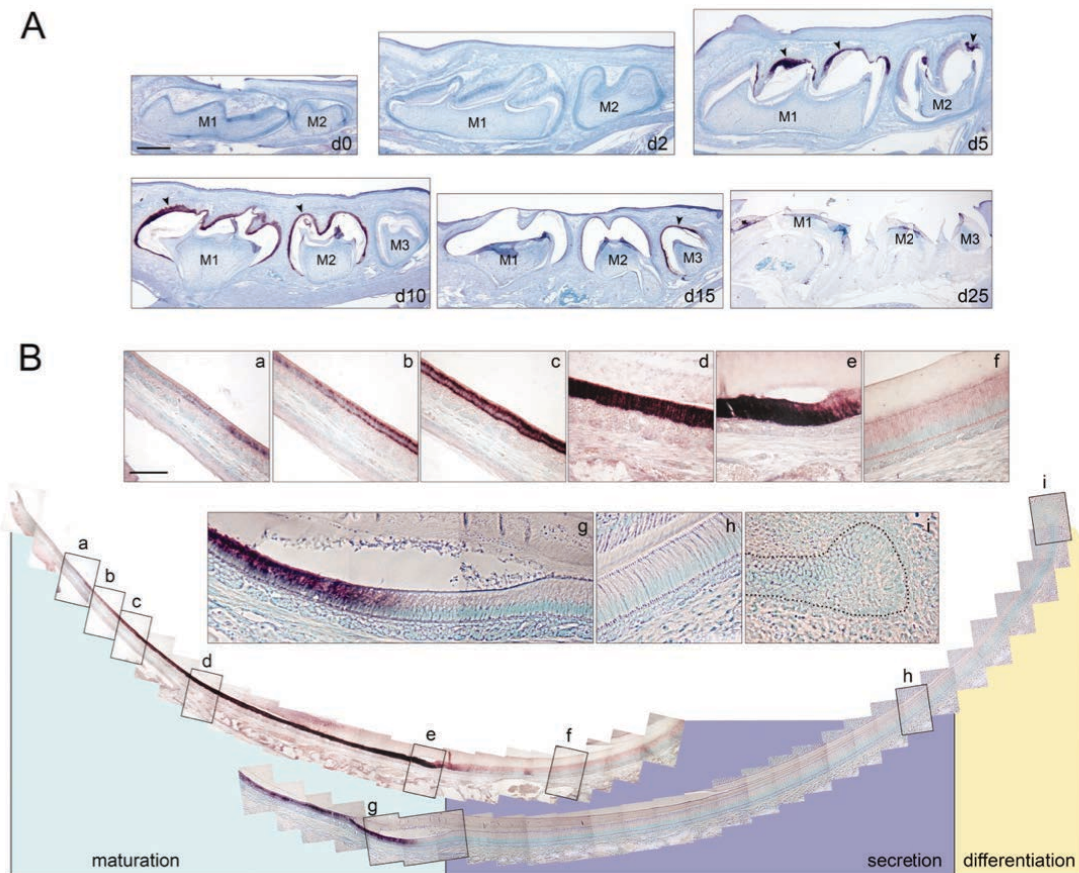


Figure 17. Expression d'AMTN par hybridation *in situ* dans des molaires (A) et incisives (B) de souris (Iwasaki *et al.*, 2005).

Chez le rat et la souris, la protéine est localisée d'abord dans les améloblastes pendant la transition sécrétion/maturation, puis dans la lame basale entre les améloblastes et la surface de l'émail pendant toute la maturation, mais elle n'est jamais retrouvée dans la matrice de l'émail (Moffatt *et al.*, 2006b). Ces résultats ont été contredits par l'étude de l'expression d'AMTN par immunohistochimie pendant le développement dentaire de la souris (Gao *et al.*, 2010). Ces auteurs montrent l'expression d'AMTN dans les améloblastes et la matrice de l'émail dès le stade de sécrétion jusqu'à la fin de la maturation, en utilisant un anticorps non purifié par chromatographie d'affinité. Cependant, la détection d'AMTN dans la matrice de l'émail est supposée artéfactuelle par Somogyi-Ganss *et al.* (2012) qui confirment les résultats des travaux de l'équipe d'Antonio Nanci (Moffatt *et al.*, 2006b) en utilisant un anticorps différent et purifié par chromatographie d'affinité. Mais par immunomarquage aux particules d'or la présence d'AMTN a été détectée au début de la maturation dans la couche superficielle d'émail de 200 à 500 nm, puis dans la lame basale à la jonction émail/améloblastes (20 nm) depuis la moitié jusqu'à la fin de la maturation (Dos

Santos Neves *et al.*, 2012; Somogyi-Ganss *et al.*, 2012). Les techniques de chromatographie d'exclusion et du double hybride ont permis de montrer que la protéine AMTN recombinante (23 kDa) interagit avec elle-même pour former des agrégats de 100 à 400 kDa, et aussi avec ODAM *in vitro* (Holcroft & Ganss, 2011). ODAM est également exprimée par les améloblastes au stade de maturation et est localisée dans la membrane basale (Dos Santos Neves *et al.*, 2012). Il est donc possible qu'AMTN et ODAM interagissent dans la lame basale. AMTN est aussi présente dans la lame basale entre l'épithélium de jonction de la gencive et l'émail (Moffatt *et al.*, 2006b). Pendant la formation de l'épithélium de jonction de molaires de rat, AMTN est présent à l'interface avec la surface de l'émail. Lors de sa régénération après gingivectomie, AMTN n'est plus détectée jusqu'à la reformation de la nouvelle lame basale entre la dent et l'épithélium. La présence d'AMTN semble donc être liée à la formation de l'attache épithéliale à la surface de la dent (Nishio *et al.*, 2010c). De plus, il a été montré que l'AMTN localisée dans la lame basale de l'épithélium de jonction provient des ARNm produits par les améloblastes au stade de maturation. Les cellules de l'épithélium de jonction n'expriment pas l'ARNm d'AMTN (Sawada *et al.*, 2014). Ces résultats sont en contradiction avec les travaux de Nakayama *et al.* (2014) qui montrent que la transcription d'AMTN est induite par des cytokines proinflammatoires dans des fibroblastes gingivaux humains de patients atteints de parodontite. Il est alors possible qu'AMTN joue un rôle dans l'inflammation gingivale (Nakayama *et al.*, 2014a; b).

L'équipe de Bernhard Ganss a généré une lignée de souris transgéniques exprimant AMTN sous le contrôle du promoteur d'AMEL. Ces souris expriment alors AMTN à un stade plus précoce de l'amélogenèse, dès la sécrétion de la matrice. L'émail de ces souris est altéré comparé à celui des animaux sauvages : il est plus fin d'environ 30% et présente une structure irrégulière. En effet, la structure prismatique habituellement observée dans l'émail est complètement désorganisée. De plus, les prolongements de Tomes ne se forment pas (Lacruz *et al.*, 2012). Les auteurs suggèrent qu'AMTN est impliquée dans la formation de la dernière couche d'émail, très fine, qui est non prismatique (Lacruz *et al.*, 2012).

Des expériences *in vitro* ont montré que la protéine recombinante humaine induisait la précipitation de l'hydroxyapatite. En revanche, la précipitation est moindre lorsque le motif phosphorylé SSEEL (codé par l'exon 7) est muté ou absent. De plus, un peptide de synthèse SSEEL phosphorylé est capable à lui seul d'induire une précipitation

d'hydroxyapatite, mais en moins grande quantité en comparaison à la protéine entière. D'autres parties de la protéine sont donc impliquées dans la minéralisation mais il reste encore à les identifier. AMTN est donc capable d'induire la minéralisation *in vitro* (Abbarin *et al.*, 2014).

Une lignée de souris $AMTN^{-/-}$, générée très récemment, présente dans son émail de surface une légère hypominéralisation. Le phénotype, très discret, ne permet pas malheureusement pas de préciser la fonction de la protéine (Nakayama *et al.*, 2015).

Par ailleurs, le gène *AMTN* a accumulé des mutations invalidantes chez les oiseaux, les tortues et les mammifères édentés ou dont les dents sont dépourvues d'émail (Meredith *et al.*, 2014). Ce gène est donc spécifique de la sphère dentaire.

Chez l'homme, *AMTN* est exprimée dans certaines tumeurs odontogéniques (Ren *et al.*, 2011, 2012; Crivelini *et al.*, 2012). A ce jour aucune mutation responsable d'une amélogénèse imparfaite n'a été identifiée.

CHAPITRE 2

MATÉRIEL ET MÉTHODES

I. Matériel biologique

1. Espèces utilisées pour l'extraction d'acides nucléiques

Nom vernaculaire	Genre et espèce	Classe, ordre, famille	Génome séquencé	Matériel disponible
Tatou	<i>Dasypus novemcinctus</i>	Mammalia, Placentalia, Xenarthra, Dasypodidae	✓	ADNg
Paresseux	<i>Choloepus hoffmanni</i>	Mammalia, Placentalia, Xenarthra, Megalonychidae	✓	ADNg
Opossum gris	<i>Monodelphis domestica</i>	Mammalia, Marsupialia, Didelphidae	✓	ADNc + transcriptome
Caïman à lunettes	<i>Caiman crocodilus</i>	Sauropsida, Crocodylia, Alligatoridae	✗	ADNc + transcriptome
Tarente du midi	<i>Tarentola mauritanica</i>	Sauropsida, Squamata, Phyllodactylidae	✗	ADNc + transcriptome
Lézard à longue queue	<i>Takydromus sexlineatus</i>	Sauropsida, Squamata, Lacertidae	✗	ADNc
Python royal	<i>Python regius</i>	Sauropsida, Squamata, Pythonidae	✗	ADNc
Anole vert	<i>Anolis carolinensis</i>	Sauropsida, Squamata, Dactyloidae	✓	ADNc
Xénope	<i>Xenopus tropicalis</i>	Amphibia, Anura, Pipidae	✓	ADNc
Pleurodèle	<i>Pleurodeles waltl</i>	Amphibia, Caudata, Salamandridae	✗	ADNc + transcriptome

Tableau 1 : Liste des espèces utilisées pour l'extraction d'ADNg ou d'ARN. Pour chaque espèce est indiqué le nom vernaculaire, le nom scientifique, la taxonomie et le matériel disponible.

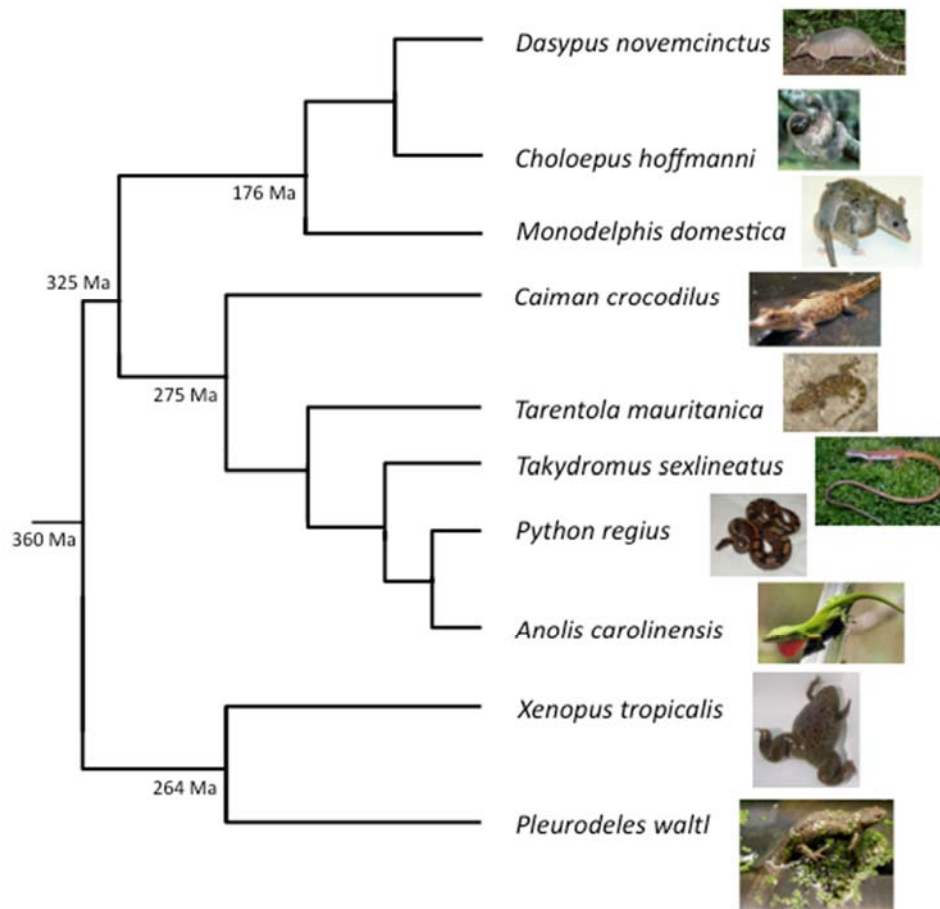


Figure 18. Arbre représentatif des relations de parenté entre les espèces utilisées dans cette thèse. Ma : millions d'années.

2. Origine des échantillons pour l'extraction des acides nucléiques

a) Mammifères

Les échantillons de tissus des deux xénarthres *C. hoffmanni* et *D. novemcinctus* m'ont été donnés préservés dans l'alcool. L'échantillon de tatou provient de Dr. W. Jim Loughry (Valdosta State University, Georgia, USA). Celui du paresseux vient de Dr. John Trupkiewicz (Philadelphia Zoo, PA 33694, USA). Ces échantillons nous ont été fournis par l'intermédiaire de Dr. F. Delsuc et Dr. F. Catseflis (ISEM, Université Montpellier II).

Pour l'opossum *M. domestica*, j'ai eu à ma disposition deux têtes de nouveaux-nés de deux jours (seize jours après fertilisation) fournies par le Dr Abigail Tucker (King's college, Londres). L'une d'elles était préservée dans du RNA later pour l'extraction des ARN, l'autre était fixée dans du paraformaldéhyde 4% et déminéralisée dans de l'EDTA 5% pour les hybridations *in situ*.

b) Sauropsides

Les squamates *T. sexlineatus*, *P. regius*, et *A. carolinensis* proviennent de dons de La Ferme Tropicale (Paris). Les spécimens de *T. mauritanica* ont été capturés sur le terrain, dans les Pyrénées orientales (arrêté de dérogation d'autorisation de capture concernant une espèce protégée n° 2014167-0004).

Un jeune (2 ans) *C. crocodilus* nous a été donné par le Dr Samuel Martin de la Ferme aux Crocodiles de Pierrelatte (Drôme).

Ces animaux ont été euthanasiés avec du chloroforme avant dissection des mâchoires.

c) Amphibiens

Le laboratoire possède de longue date un élevage de *P. waltl*. Ces animaux ont été euthanasiés avec une dose létale de tricaine avant dissection.

Le jeune *X. tropicalis* était un don du Muséum National d'Histoire Naturelle. Seule la mâchoire supérieure portant des dents a été utilisée.

3. Matériel pour l'hybridation *in situ*

a) *Anolis carolinensis*

Ce lézard est connu comme étant le premier reptile dont le génome a été entièrement séquencé en 2011. C'est une espèce polyphyodonte : les dents sont remplacées tout au long de la vie de l'animal. Sa dentition est hétérodonte : les dents antérieures ont une morphologie différente des dents postérieures. Elles présentent une unique pointe (cuspidé), tandis que les postérieures en ont trois (Fig. 19).

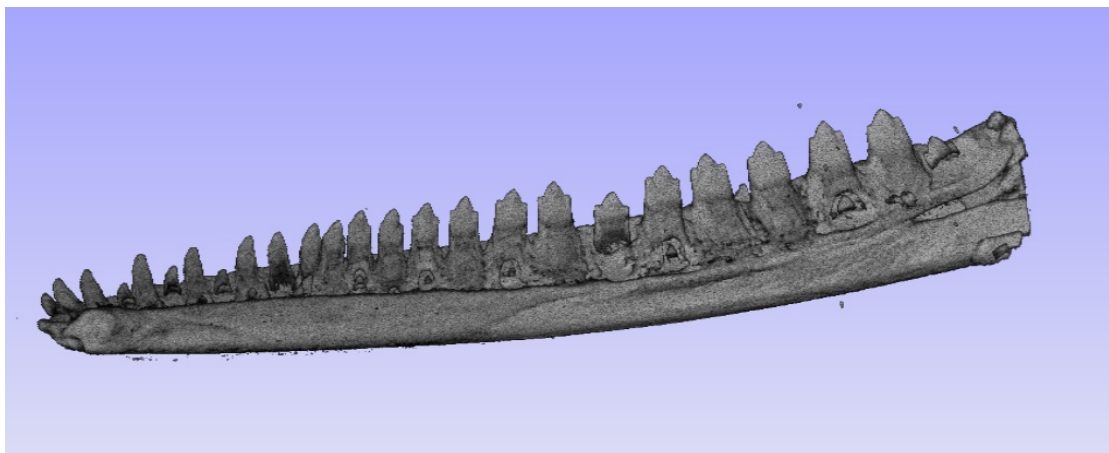


Figure 19. CT-Scan d'un quadrant de mâchoire inférieure d'*Anolis carolinensis* (Viscom, France).

b) *Pleurodeles waltl*

La dentition de *P. waltl* est polyphyodonte et subit un changement de morphologie à la métamorphose. Les dents des larves sont coniques, tandis qu'elles sont bicuspidées chez les spécimens après la métamorphose (Fig. 20) (Davit-Béal *et al.*, 2006). Une autre particularité des dents larvaires de *P. waltl* est la présence d'une couche d'émailloïde entre la dentine et l'émail (Davit-Béal *et al.*, 2007a).

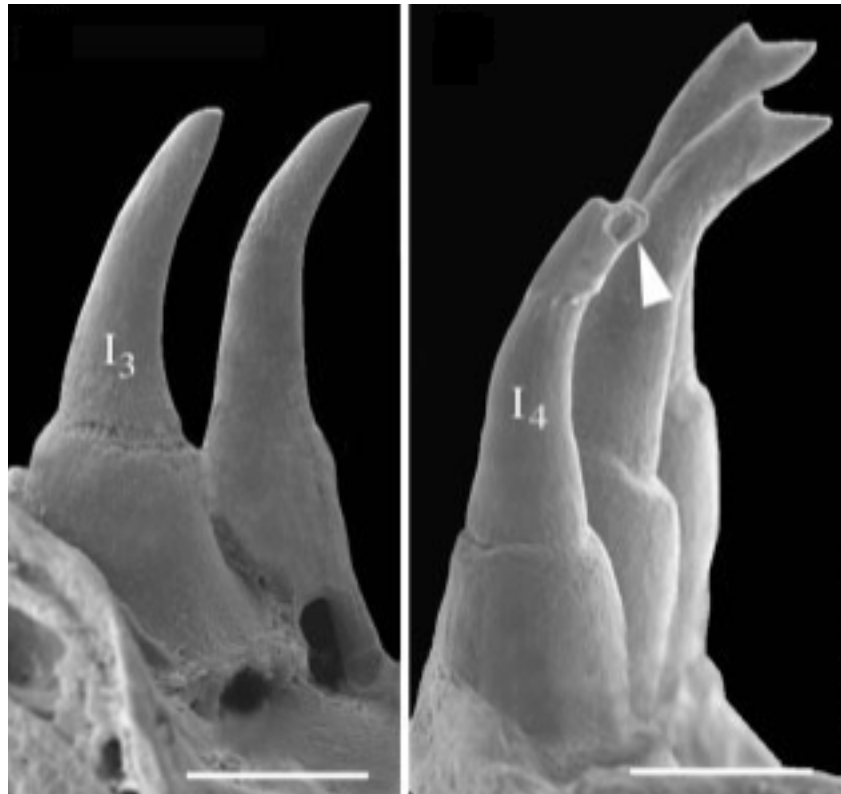


Figure 20. Images de microscopie électronique à balayage de dents chez une larve (gauche) et un juvénile *Pleurodeles waltl* (Davit-Béal *et al.*, 2006)

c) *Monodelphis domestica*

L'opossum possède une dentition hétérodonte. Sur chaque quadrant de mâchoire, il y a cinq incisives, une canine, trois prémolaires et quatre molaires. Seule la prémolaire 3 possède une dent de remplacement (Kozawa *et al.*, 1998).

II. Méthodes

1. Histologie

a) Fixation et inclusion dans l'epon

Après dissection, les mâchoires sont fixées pendant 2h dans une solution contenant 1,5% de glutaraldéhyde et 1,5% de paraformaldéhyde dans du tampon phosphate

(PBS pour Phosphate Buffered Saline). Les mâchoires sont ensuite déminéralisées pendant trois à quatre semaines à 4°C dans ce même fixateur dans lequel on ajoute 5% d'EDTA.

Lorsque les tissus sont déminéralisés, ils sont rincés toute une nuit dans du PBS, puis 2 fois 15 min dans du PBS/saccharose 10%. Les mâchoires sont post-fixées dans une solution de tétroxyde d'osmium 1%/PBS/saccharose pendant 1 h, puis rincées à nouveau dans du PBS/saccharose. Les tissus sont alors déshydratés dans des bains d'éthanol de degrés croissants puis dans un mélange d'éthanol et d'oxyde de propylène (50/50). La déshydratation est terminée par un bain d'oxyde de propylène pur puis l'imprégnation des échantillons commence progressivement par un bain d'oxyde de propylène/epon (50/50) pendant 2 jours dans un dessiccateur. Les tissus sont alors transférés dans de l'epon pur pendant 5 h à température ambiante. Enfin, les mâchoires sont disposés dans des moules remplis d'epon contenant un catalyseur pendant 48 h à 60°C, température à laquelle la résine polymérise.

b) Coupes et coloration

Les coupes semi-fines (1-2 µm) sont réalisées à l'aide d'un ultramicrotome (Reichert Austria OMU3 - Leica). Elles sont placées sur des lames de verres et colorées dans un mélange de bleu de toluidine et de bleu azur, séchées sur une plaque chauffante, puis montées avec de l'Eukitt.

2. Biologie moléculaire

a) Extraction de l'ADN génomique

L'extraction de l'ADN est réalisée à partir de tissus préservés dans l'alcool en utilisant le kit DNeasy Blood and Tissue (Qiagen). Le tissu (25 mg) est coupé en petits morceaux puis incubé dans le tampon ATL contenant de la protéinase K à 55°C pendant 1 à 3 heures. Lorsque le tissu est complètement lysé, 200 µl de tampon AL et 200 µl d'éthanol 96% sont ajoutés, puis le mélange est déposé sur une mini-colonne de filtration fournie par Qiagen et centrifugé 1 min à 6000 g. Lors de cette étape les acides nucléiques se sont fixés dans la membrane de la colonne. La membrane est alors lavée une première fois avec 500 µl de tampon RW1, puis avec 500 µl de tampon RW2, et est ensuite séchée par centrifugation à vide 3 min à 20 000 g. La membrane est incubée 1 min avec 200 µl de tampon AE puis centrifugée 1 min à

6000 g. Cette étape est répétée une seconde fois, permettant d'obtenir l'éluat final de 400 µl d'ADN génomique (ADNg) purifié.

b) Extraction des ARN totaux de mâchoire

Les mâchoires sont disséquées rapidement, coupées en petits morceaux puis placées dans un mortier rempli d'azote liquide. La congélation instantanée des tissus permet d'une part, de les rendre cassants et, d'autre part, de prévenir la dégradation des ARN par des RNases. Les tissus congelés sont alors réduits en fine poudre à l'aide d'un pilon. Cette poudre est transférée dans un tube contenant un tampon de lyse et du β-mercaptonéthanol afin de faire éclater les cellules. Ce tampon ainsi que la suite des réactifs utilisés ensuite font partie du kit de purification d'ARN "RNeasy Fibrous tissue" (Qiagen). La digestion des protéines est assurée par l'ajout de la protéinase K, 30 min à 55°C. Après une centrifugation de 3 min à 10 000 g, le culot contenant les débris cellulaires est éliminé et le surnageant est mélangé avec 1/2 volume d'éthanol à 96%, ce qui a pour conséquence de précipiter les acides nucléiques. Cette solution est placée dans une mini-colonne de filtration fournie par Qiagen et est centrifugée 30 s à 10 000 g. En passant à travers la membrane filtrante de la colonne, les acides nucléiques en solution vont être retenus par celle-ci. La membrane contenant les acides nucléiques est ensuite lavée avec le tampon RW1. Après élimination du tampon par centrifugation, l'ADN est dégradée par l'ajout de la DNase I pendant 15 min. Il ne reste alors plus que les ARN dans la membrane. Cette dernière est à nouveau lavée avec le tampon RW1, puis deux fois avec le tampon RPE. La membrane est ensuite séchée par centrifugation à vide pendant 1 min. L'ARN peut enfin être élué en centrifugeant la colonne 1 min à 10 000 g après l'ajout de 30 µl d'eau "RNase free". Cette étape est répétée une seconde fois et l'éluat final est aliquoté en tubes de 10 µl qui vont être conservés à -80°C.

c) Conception des amorces pour la PCR

Les amorces sont de courtes séquences monocaténares complémentaires de régions qui flanquent l'ADN à amplifier par PCR. Elles fonctionnent toujours par couple. Une d'elle est appelée amorce sens, ou forward, qui s'hybridera en 5' de la région d'ADN cible, et l'autre est l'amorce antisens, ou reverse, qui s'hybridera en 3' sur le brin complémentaire. Elles sont généralement composées d'une vingtaine de nucléotides.

Au cours de ma thèse, j'ai élaboré les couples d'amorces à l'aide d'un logiciel gratuit en ligne : Primer3 (Untergasser *et al.*, 2012). Celui-ci propose une sélection de couples à partir de la séquence qu'on lui donne et divers critères. Tout d'abord, il faut indiquer la taille de l'amplicon désirée, puis la taille des amorces souhaitée (entre 18 et 22 nucléotides), la température d'hybridation des amorces (entre 55 et 65°C) et enfin le pourcentage de GC (entre 40 et 60%). Une fois que la liste d'amorces est proposée par le logiciel, il faut choisir le meilleur couple. Pour cela, il convient de regarder le dernier triplet de l'amorce, car c'est par celui-ci que l'amorce va se fixer en premier. Idéalement, le triplet doit être de la composition suivante : C/G - A/T - C/G. Les amorces contenant des longues répétitions d'un même nucléotide sont à éviter. De même, les couples d'amorces présentant une complémentarité de séquence entre elles sont à proscrire.

d) Polymerase Chain Reaction

i. Principe

L'ADN extrait à partir d'un échantillon de tissu n'est pas directement analysable car il contient une masse trop importante de séquences nucléotidiques. Pour l'étude d'un gène donné, il convient donc d'isoler la séquence d'intérêt. Pour cela la Polymerase Chain Reaction (PCR) permet de sélectionner et de répliquer *in vitro* en un très grand nombre de copies (supérieur à un milliard) une séquence cible comprise entre deux régions d'ADN connues (les amorces). Une fois la réaction terminée, la quantité d'ADN de départ n'aura pas varié et sera négligeable devant celle de la séquence amplifiée qui pourra alors être exploitée.

La PCR est réalisée dans un mélange réactionnel qui comprend l'extrait d'ADN, une ADN polymérase (Taq) qui agit à haute température, les amorces, et les quatre désoxyribonucléotides triphosphates (dNTP : dATP, dCTP, dGTP et dTTP) en excès dans une solution tampon. L'amplification exponentielle est réalisée par une succession de cycles, composés chacun de trois étapes. La première étape, appelée étape de dénaturation, s'effectue à 94°C pendant 30 s, et permet de séparer les brins d'ADN bicaténaires et les amorces hybridées en rompant les liaisons hydrogène. La deuxième étape, l'hybridation, s'effectue pendant 30 s à une température comprise entre 50 et 65°C en fonction de la température d'hybridation des amorces. Pendant cette étape, la diminution de la température permet aux liaisons hydrogène de se reformer et donc aux amorces sens et anti-sens de s'hybrider spécifiquement avec

l'ADN. La dernière étape du cycle est une phase d'élongation au cours de laquelle la Taq polymérase synthétise le brin complémentaire de l'ADN cible à partir des dNTP présents dans le milieu réactionnel. Cette étape se déroule à 72°C pendant une durée proportionnelle à la taille du fragment d'ADN à amplifier (de quelques secondes à plusieurs minutes). La réalisation de 35 cycles permet d'obtenir une quantité d'ADN suffisante. A la fin de ces cycles, une dernière étape d'élongation à 72°C pendant 5 min permet de finaliser tous les amplicons formés.

ii. Protocole

Au cours de ma thèse, j'ai utilisé l'ADN polymérase GoTaq de Proméga. Pour un volume final de 50 µl on mélange 1 µl d'ADN (ADNg ou ADNc) à 10 µL de tampon GoTaq, 1 µL de dNTP, 1 µl des amorces sens et antisens, 35,75 µL d'eau et 0,25 µL de l'enzyme. La réaction de PCR est réalisée dans un thermocycleur « Mastercycler pro » d'Eppendorf.

e) RT-PCR en deux étapes

i. Synthèse de l'ADN complémentaire à partir des ARN

Les ARN précédemment purifiés sont rétro-transcrits en ADN complémentaire (ADNc) grâce à une transcriptase inverse d'origine virale capable de synthétiser de l'ADN monocaténaire à partir d'une matrice ARN. Des amorces non spécifiques oligo(dT)₁₈ sont utilisées afin de rétro-transcrire tous les ARNm présents dans le milieu réactionnel en se fixant à leur queue poly-A.

Pour cette réaction, la première étape consiste à mélanger 10 µl d'ARN à 1 µl d'oligo(dT)₁₈ et 1 µl de dNTP 10mM. Ce mélange est chauffé à 65°C pendant 5 min puis refroidi brièvement dans de la glace. Après centrifugation, on ajoute 4 µl du tampon 5X, 2 µl de DTT 0,1 M, 1 µl d'un inhibiteur de ribonucléase (RNasin, Proméga) et le mélange est incubé 2 min à 37°C. L'enzyme, M-MLV Reverse Transcriptase (Invitrogen), est alors ajoutée au mélange qui est ensuite incubé à 37°C pendant 50 min. La réaction est inactivée en chauffant à 70°C pendant 15 min. Finalement, la réaction aboutit à la formation d'ADN simple-brin complémentaire des ARNm, qui va être directement utilisé comme matrice d'ADN pour la PCR.

ii. Amplification des séquences d'intérêt

La deuxième étape consiste en une PCR « classique » en utilisant comme matrice l'ADNc et des amorces spécifiques du gène d'intérêt (cf. paragraphe « Polymerase Chain reaction »).

f) Electrophorèse sur gel d'agarose

Cette méthode permet de vérifier la présence du gène d'intérêt dans le produit de PCR. Les fragments d'ADN, chargés négativement, vont être soumis à un courant électrique et migrer, de l'anode vers la cathode, à travers les mailles d'un gel d'agarose. Les petits fragments migrant plus rapidement que les grands, on retrouvera, après migration, les ADN de petite taille en bas du gel, et ceux de grande taille en haut.

i. Préparation du gel d'agarose 1%

L'agarose est dissout à chaud dans un tampon de migration, le TBE (Tris-Borate-EDTA) à 0,5X. Après dissolution, la solution est laissée refroidir jusqu'à 50-60°C et du bromure d'éthidium (BET), un intercalant de l'ADN, est ajouté. Le BET, lorsqu'il est lié à l'ADN et soumis à un rayonnement ultraviolet, émet une fluorescence qui permet la visualisation des produits de PCR après migration. L'agarose dissout contenant le BET est coulé dans un moule contenant des peignes permettant de former des puits dans lesquels seront déposés les échantillons. Après polymérisation, le gel est placé dans une cuve contenant du TBE 0,5X.

ii. Dépôt et migration des échantillons

Les échantillons d'ADN sont d'abord mélangés avec du tampon de migration. Ce tampon contient du glycérol afin d'augmenter la densité de la solution d'ADN qui se placera ainsi au fond des puits, et un colorant permettant la visualisation de l'ADN pendant la migration. En parallèle des échantillons d'ADN, un marqueur de poids moléculaire est déposé afin de pouvoir estimer la taille des fragments d'ADN ayant migré dans le gel. La migration s'effectue pendant 25 min à 135 V.

Après migration, le gel est soumis à un rayonnement ultraviolet et photographié dans un analyseur Gel Doc (BioRad).

g) Purification des produits de PCR avant clonage

La purification des produits de PCR est effectuée en utilisant le kit « QIAquick Gel extraction kit » de Qiagen.

Lorsque la bande d'ADN à cloner est observée sur le gel d'agarose, elle est découpée à l'aide d'une lame de scalpel stérile, pesée puis déposée dans un tube contenant trois fois son volume de tampon QG (300 μ L de tampon pour 100 mg de gel). Le gel est dissout dans le tampon pendant 10 min dans un bain-marie à 50°C, puis la solution est déposée dans une colonne QIAquick, qui est alors centrifugée 1 min à 17 000 g. La colonne est lavée avec 500 μ L de tampon QG pour enlever d'éventuelles traces d'agarose, centrifugée puis lavée à nouveau avec 750 μ L de tampon PE. Une centrifugation à vide permet de sécher la membrane de filtration, et l'ADN est élué avec 30 μ L de tampon d'éluion (10mM Tris-Cl, pH 8,5) qui est incubé sur la membrane pendant 1 min avant centrifugation d'1 min.

h) Clonage

i. Principe

Le but du clonage est d'insérer un fragment d'ADN dans un vecteur (un plasmide) que l'on va introduire dans des bactéries afin qu'elles le répliquent. L'insertion d'ADN dans un plasmide présente aussi la possibilité de pouvoir ajouter en amont ou en aval de notre séquence d'intérêt des éléments régulateurs comme des promoteurs. C'est pourquoi, ayant pour objectif de synthétiser des sondes ARN, j'ai utilisé le kit pGEM-T de Proméga qui utilise un plasmide possédant de part et d'autre du site d'insertion deux promoteurs pour la transcription *in vitro*.

ii. Ligation

La première étape est d'insérer le fragment d'ADN d'intérêt dans le vecteur.

Lors de la PCR, la Taq polymérase ajoute aux extrémités 3' des amplicons une Adénine (A). Le vecteur pGEM-T linéarisé, fourni dans le kit, possède une Thymine (T) sortante à chaque extrémité 3'. Grâce à l'action d'une ligase, le fragment d'ADN possédant les extrémités A va venir s'accrocher aux extrémités T du vecteur linéarisé par complémentarité, et ainsi former un ADN circulaire que l'on va insérer dans des bactéries compétentes.

La réaction de ligation est mise en place en rassemblant dans un tube 3 μ L du produit de PCR purifié, 1 μ L du vecteur linéarisé pGEM-T, 5 μ L du tampon de ligation et 1 μ L de la ligase ADN T4. Le mélange est laissé à 4°C pendant toute une nuit.

iii. Transformation

La deuxième étape consiste à insérer le plasmide contenant l'insert dans des bactéries compétentes.

Le jour suivant la ligation, un aliquot de 50 μ L des bactéries JM109 est sorti du congélateur à -80°C et laissé décongeler 10 min dans de la glace. A ces bactéries sont ajouté 2 μ L du produit de ligation de la veille et le tout est mélangé doucement en tapotant le tube avec le doigt. Les bactéries sont ainsi laissées 20 min dans la glace avant d'être soumises à un choc thermique : elles sont incubées au bain-marie pendant 45 s à 42°C puis immédiatement remises dans la glace, pendant 2 min.

iv. Culture et sélection des clones

La suite des manipulations a pour but de répliquer les bactéries contenant le gène d'intérêt.

Les bactéries transformées sont mises en culture dans 950 μ L de milieu SOC pendant 1 h 30 à 37°C sous agitation. La culture est ensuite centrifugée 10 min à 1000 g, et le culot bactérien est repris dans 200 μ L de milieu SOC. Ce mélange est alors étalé dans une boîte de Pétri sur un milieu Luria Broth (LB)-Agar contenant de l'ampicilline, du Xgal (5-bromo-4-chloro-3-indolyl- β -D-galactoside) et de l'IPTG (isopropylthio- β -D-galactoside). L'ampicilline contenue dans le milieu de culture des boîtes de Pétri permet la sélection des bactéries transformées grâce à la présence du gène de résistance à l'ampicilline sur le plasmide. Le plasmide possède aussi le gène *lacZ*, permettant de sélectionner les bactéries contenant le plasmide et l'insert, et non pas celles ayant inséré un plasmide refermé sur lui même. En effet, la β -galactosidase produite par le gène *lacZ*, en présence de son substrat Xgal et de l'inducteur IPTG, induit une coloration bleue aux bactéries. Le gène *lacZ* étant situé au niveau du site d'insertion de l'ADN à cloner, la présence d'un insert le rend non-fonctionnel et les bactéries transformées restent blanches dans le cas où elles possèdent le plasmide contenant l'insert. Les clones bleus correspondent donc aux bactéries ayant inséré un plasmide vide.

Après incubation pendant une nuit à 37°C , une dizaine de colonies blanches qui ont poussé sur la boîte de Pétri sont testées par PCR pour la présence de l'insert.

Le mélange réactionnel pour la PCR est préparé au préalable avec des amorces universelles M13 (amorces qui s'hybrident à des sites présents sur l'ADN plasmidique de part et d'autre du site de clonage) et est réparti dans 10 tubes. Les dix colonies à tester sont ensuite repiquées une à une à l'aide d'un cure-dent stérile, d'abord sur une nouvelle boîte de Pétri puis dans le tube de PCR correspondant. De cette façon, les colonies testées positivement par PCR pour la présence de l'insert pourront être reprises à partir de la nouvelle boîte de Pétri, et mises en culture dans 5 mL de milieu

LB contenant 5 µL d'ampicilline pendant une nuit à 37°C sous agitation. Le jour suivant, les cultures bactériennes sont utilisées pour l'extraction des plasmides.

i) Purification des plasmides

Les plasmides contenus dans les bactéries sont extraits et purifiés à l'aide du kit Miniprep de Qiagen.

Les cultures bactériennes sont centrifugées pendant 10 min à 5000 g et le culot est resuspendu dans 250 µL de tampon P1, contenant la RNase A. Après l'ajout de 250 µL de tampon P2, puis de 350 µL du tampon N3, la solution est centrifugée 10 min à 17 000 g. Le culot contenant les débris cellulaires est éliminé et le surnageant est placé dans une colonne de filtration fournie par Qiagen. Une première centrifugation de 30 s permet de fixer les plasmides sur la membrane filtrante tandis que le liquide est évacué. La membrane est alors lavée une première fois par le tampon PB, puis par le tampon PE. La colonne est ensuite centrifugée à vide afin de sécher la membrane. Enfin, les plasmides purifiés sont élués dans 50 µL du tampon d'éluion (10mM Tris-Cl, pH 8,5).

j) Linéarisation des plasmides et purification pour la transcription

in vitro

Les plasmides sont linéarisés par PCR avec les amorces universelles M13 sens et antisens, puis le produit de PCR est purifié avec le kit QIAquick PCR purification (Qiagen). Dans ce dernier sont ajoutés 5 volumes du tampon PBI (250 µL pour 50 µL de produit de PCR) puis la solution est déposée dans une colonne QIAquick, et centrifugée 30 s à 17 000 g. La membrane de filtration est ensuite lavée par 750 µL de tampon PE, et séchée par centrifugation à vide pendant 1 min. Enfin, l'éluion est réalisée en centrifugeant 1 min après incubation d'1 min de 30 µL de tampon d'éluion (10mM Tris-Cl, pH 8,5).

k) Synthèse de sondes froides ARN

i. Transcription *in vitro* des sondes sens et antisens

Un promoteur d'ARN polymérase est situé de part d'autre de l'insert sur le plasmide linéarisé. Ces promoteurs vont permettre la transcription *in vitro* de l'ADN cloné en sonde ARN, sens ou antisens selon le promoteur utilisé. La sonde antisens va être complémentaire de l'ARNm ciblé lors de l'hybridation *in situ*. La sonde sens sera utilisée uniquement en tant que contrôle négatif, car elle ne s'hybridera pas avec les

ARNm. Les sondes sont marquées grâce à l'incorporation de bases Uracile (U) liées à la digoxigénine.

Le kit Riboprobe Combination System - Sp6/T7 de Proméga a été utilisé pour la synthèse des sondes. Dans un tube, on mélange 4 µL du tampon de transcription 5X, 2 µL de DTT à 100 mM, 1 µL de rATP à 10 mM, 1 µL de rCTP à 10 mM, 1 µL de rGTP à 10 mM, 0,5 µL de rUTP à 10 mM, 0,5 µL de UTP marqué à la digoxigénine, 8 µL de produit de PCR purifié, 1 µL d'inhibiteur de RNase et 1 µL d'ARN polymérase. Le mélange est incubé pendant 2 h à 37°C pour l'ARN polymérase T7 ou 40°C pour l'ARN polymérase Sp6. La matrice ADN est ensuite dégradée en incubant 15 min à 37°C avec 1 µL de DNase RQ1.

ii. Purification des sondes

Les sondes sont purifiées avec le kit Illustra ProbeQuant G-50 Micro Columns (GE Healthcare). La première étape consiste à ajouter 29 µL de tampon "Probe buffer type 1" au 21 µL du mélange précédent. La colonne de purification fournie par ce kit n'est pas prête à l'emploi, il faut d'abord la préparer. Pour cela, la colonne est d'abord vortexée pendant 20 s, puis son bout est coupé. Le bouchon de la colonne est ouvert d'un quart de tour et cette dernière est placée dans le tube de 2 mL fourni, qui est centrifugé 1 min à 2700 rpm pour éliminer le liquide contenu dans la colonne. La colonne, maintenant prête à être utilisée, est placée dans un tube de 1,5 mL et la solution de 50 µL (sonde + tampon) est déposée sur la résine. La colonne est à nouveau centrifugée pendant 2 min à 2700 rpm. L'éluat obtenu est la sonde purifiée, qui se conserve congelée à -20°C.

D) Préparation des échantillons pour l'hybridation *in situ*

i. Fixation et déminéralisation

Immédiatement après dissection, les mâchoires sont fixées dans du Formoy (30% formaldéhyde, 60% éthanol et 10% acide acétique) toute une nuit à 4°C. Il n'est pas possible d'effectuer des coupes de mâchoires directement car l'os et la dent sont des tissus trop durs. Il est alors nécessaire de déminéraliser les échantillons avant de pouvoir les sectionner. Pour cela, les mâchoires sont placées dans une solution d'acide acétique à 10%, à température ambiante sous agitation, pendant 4 semaines avec changement de la solution deux fois par semaine.

ii. Déshydratation et inclusion

Une fois suffisamment déminéralisées, les mâchoires sont déshydratées par une série de bains d'éthanol de degrés croissants : éthanol à 70% pendant 30 min, éthanol à 85% pendant 30 min, éthanol à 95% pendant 10 min, et éthanol à 100% pendant 2x10 min puis 30 min. Les échantillons sont ensuite plongés dans un bain de toluène jusqu'à ce qu'ils coulent (quelques secondes) puis dans un second bain pendant 1 min. Les mâchoires sont alors placées dans de la paraffine liquide dans une étuve à 58°C pendant quelques heures, puis dans un deuxième bain de paraffine pendant toute la nuit. Le lendemain, les mâchoires sont placées, dans l'orientation souhaitée, dans des moules contenant de la paraffine liquide, puis laissées refroidir à température ambiante. Une fois durcis, les blocs sont conservés à 4°C.

Les blocs de paraffine contenant les mâchoires sont sectionnés à une épaisseur de 8 à 10 µm à l'aide d'un microtome (Leica). Des séries d'une dizaine de coupes successives (appelées rubans) sont déposées dans un défriporteur à paraffine : un bain-marie dont l'eau est chauffée à 49°C, qui permet de "défroisser" les coupes. En effet, sous l'effet de la chaleur, les coupes s'étirent et elles sont alors récupérées directement avec la lame Superfrost Plus (Menzel-Glaser) que l'on plonge dans l'eau. Le ruban se colle instantanément sur la lame, qui est alors égouttée puis séchée à l'étuve à 37°C pendant 48 h minimum.

Alternativement, des coupes individuelles peuvent aussi être récupérées. Dans ce cas, la coupe est déposée délicatement à l'aide d'un pinceau sur une goutte d'eau RNase-free, sur une lame Superfrost Plus. Cette dernière est ensuite placée sur une plaque chauffante. Lorsque la paraffine s'étire, l'eau est retirée avec une pipette Pasteur. Les lames sont ensuite séchées comme décrit précédemment.

m) Hybridation *in situ* sur coupes

L'hybridation *in situ* est une manipulation qui se déroule sur trois jours. Elle consiste à hybrider spécifiquement des sondes ARN marquées sur un ARNm cible, afin de visualiser le lieu d'expression de cet ARNm dans la cellule. La composition des tampons indiqués par une * est détaillée dans le paragraphe n) ci-dessous.

i. Traitement des coupes et hybridation

Les lames sont placées dans portoir en verre, lui-même placé dans une cuve en verre. Le portoir possède une anse permettant de retirer les lames de la solution contenue dans la cuve.

La première étape consiste à déparaffiner les coupes en plongeant les lames dans deux bains de toluène de 5 min. Les coupes sont ensuite réhydratées dans des bains d'éthanol de degrés décroissants : deux bains de 5 min d'éthanol à 100%, puis un bain de 2 min éthanol à 90%, 2 min éthanol à 70%, 2 min éthanol à 30% et enfin dans du PBS. Les lames sont ensuite traitées à la protéinase K (0,6 µg/mL) pendant 5 min à 37°C, puis rincées dans du PBS. Une étape de post-fixation est réalisée dans du paraformaldéhyde 4%/PBS pendant 30 min. Les lames sont rincées dans du PBS, puis dans un tampon SSC 2X (Saline Sodium Citrate), avant d'être placées à plat dans une chambre humide. La sonde, diluée au 1/150e dans le tampon d'hybridation*, est déposée directement sur la lame (150 µL). La lame est alors recouverte d'une lamelle et la sonde est incubée toute la nuit dans la chambre humide dans une étuve à 65°C.

ii. Lavage des lames et immunoréaction

Les lames sont replacées dans la cuve pour être lavées avec le tampon de lavage* préalablement chauffé à 65°C. Un premier lavage de 30 min dans ce tampon permet de décoller les lamelles, puis deux lavages d'une heure permettent d'enlever l'excès de sonde non fixée, et de diminuer ainsi le bruit de fond. Les lames sont ramenées à température ambiante dans le tampon MABT*, puis sont à nouveau disposées à plat dans une chambre humide. Sur chaque lame sont ajoutés 300 µL de la solution de blocage* afin de bloquer les sites aspécifiques, pendant 2 h minimum. Les lames sont ensuite incubées avec l'anticorps anti-digoxygénine dilué au 1/1000e dans la solution de blocage pendant la nuit à température ambiante. L'anticorps reconnaît la digoxygénine des bases uraciles de la sonde, et est couplé à une enzyme, la phosphatase alcaline, qui va permettre la coloration lors de la révélation.

iii. Révélation

Les lames sont lavées dans quatre bains de 30 min de tampon MABT, puis équilibrées par le tampon NTMT*. La révélation est réalisée à 37°C, à l'abri de la lumière, dans le tampon NTMT dans lequel est ajouté du Nitro Blue Tetrazolium (NBT) à 22,5 µg/mL et du 5-Bromo-4-Chloro-3-Indolyl Phosphate (BCIP) à 115 µg/mL. Le BCIP est un substrat de la phosphatase alcaline (couplée à l'anticorps). Une fois dépourvu de son

phosphate, le BCI va être oxydé par le NBT et ainsi former un précipité bleu. La durée de la réaction et la formation de la coloration dépendent de la sonde et varient de quelques minutes à quelques heures. Lorsque le marquage est apparu, les lames sont lavées dans du PBS puis recouvertes d'un milieu de montage aqueux (Aquatex) et d'une lamelle.

n) Composition des tampons pour l'hybridation *in situ*

i. Tampon d'hybridation

Pour 10 ml :

2,5 ml SSC 20X
5 ml Formamide
2 ml Dextran sulfate 50%
250 µl ARNt (40mg/ml)
200 µl Denhardt's 50X
qsp H₂O

ii. Tampon de lavage

Pour 1 litre :

500 ml Formamide
50 ml SSC 20X
1ml Tween 20
qsp H₂O

iii. MABT

Pour 1 litre :

50 ml Acide maléique 2M
30 ml NaCl 5M
1 ml Tween 20
qsq H₂O

Ajuster le pH à 7,5 avec NaOH 10N.

iv. Solution de blocage

Pour 50 ml :

10 ml Sérum de chèvre
10 ml Blocking reagent (boehringer) 10%
30 ml MABT

v. NTMT

Pour 100 ml :

10 ml Tris-HCl 1M
2,5 ml MgCl₂ 2M
2 ml NaCl 5M
100 µl Tween 20
qsp H₂O

Ajuster le pH à 9,5 avec NaOH 10N.

3. Analyses *in silico*

a) Bases théoriques de l'analyse évolutive des protéines

Au cours des longues périodes de temps géologiques qui les séparent de leur origine, et qui se chiffrent en plusieurs centaines de millions d'années pour la plupart d'entre elles, les séquences nucléotidiques accumulent des mutations au hasard. Il existe différents types de mutations, qui entraînent des modifications plus ou moins importantes.

L'insertion ou la délétion d'un ou plusieurs nucléotides dans la séquence d'un gène peuvent entraîner un changement de cadre de lecture et avoir un effet dévastateur sur la protéine qu'il code. L'insertion ou la délétion d'un nombre de nucléotides multiple de trois se traduit par l'ajout ou la perte d'acides aminés, modifiant ainsi la longueur de la protéine. Le remplacement d'un nucléotide par un autre, appelé substitution, peut être soit synonyme, s'il n'entraîne pas de changement d'acide aminé grâce à la redondance du code génétique, ou non-synonyme s'il en entraîne un.

Les mutations peuvent être neutres, c'est à dire qu'elles n'ont aucun impact sur la fonction de la protéine, et sont donc conservées. Les mutations peuvent aussi entraîner un dysfonctionnement d'une protéine cruciale pour la survie d'un individu, celles-ci sont alors éliminées par sélection naturelle. Au contraire, certaines mutations peuvent apporter un avantage pour la survie d'un individu et seront conservées pour être transmises à la descendance. C'est pour cela que, même si tous les sites ont

théoriquement la même probabilité d'être mutés par le fruit du hasard, on constate que les protéines, ou certaines de leurs régions, n'évoluent pas au même rythme. En effet, les régions importantes pour la fonction ou la structure de la protéine accumulent peu, voire pas, de mutations et on dit que ces régions sont soumises à une sélection "purifiante". Ces régions évoluent très lentement et présentent un faible taux de substitution : le taux de substitutions synonymes est alors supérieur à celui des substitutions non-synonymes. En revanche, les régions moins importantes ne possèdent pas la même contrainte sélective et évoluent plus rapidement en retenant un grand nombre de mutations. Leur taux de substitutions synonymes est alors égal à celui des substitutions non-synonymes.

Lorsque l'on effectue l'analyse évolutive d'une protéine, on calcule le taux d'évolution (substitutions synonymes versus non-synonymes) pour tous les sites de la protéine et dans le plus grand nombre possible de lignées représentatives de longues périodes évolutives. Plus ce nombre est grand, plus la probabilité qu'un site soit conservé par hasard est faible. Ce type d'analyse permet de définir précisément tous les sites essentiels au bon fonctionnement d'une protéine.

b) Recherche des séquences dans les bases de données

Les séquences nucléotidiques ont été extraites des bases de données NCBI (National Center for Biotechnology Information : <http://www.ncbi.nlm.nih.gov>) et *Ensembl* (<http://www.ensembl.org>). Certaines séquences partielles ont été complétées en utilisant l'outil BLAST (Basic Local Alignment Search Tool) sur les données contenues dans les répertoires "Trace archives" ou "Whole-Genome Shotgun contigs" de NCBI.

c) Alignement de séquences

Les régions codantes ont été traduites en séquences d'acides aminés et alignées manuellement avec le logiciel Seaview v 4.3.3 (Gouy *et al.*, 2010).

d) Modèle de substitution

Les substitutions entre nucléotides se font préférentiellement par transition entre purines (A devient G, ou G devient A) ou entre pyrimidines (C remplacé par T, ou T par C). Le passage d'une purine à une pyrimidine, et *vice versa*, se fait par

transversion et est beaucoup plus rare. Les substitutions entre codons et acides aminés n'ont pas non plus toutes la même probabilité.

Il existe une multitude de modèles qui prennent en compte ces différences et qui permettent d'estimer le nombre de substitutions par site à partir d'un alignement de séquences apparentées en utilisant des algorithmes. Pour déterminer quel modèle de substitution s'applique le mieux à un jeu de données, un outil de sélection de modèle automatique disponible sur www.datamonkey.org est utilisé.

e) Construction d'un arbre phylogénétique

Tout d'abord, les régions manquantes et insertions sont retirées de l'alignement afin que toutes les séquences aient la même taille. Puis cet alignement est transféré dans le logiciel *MacClade 4.08*. À partir des espèces présentes dans l'alignement, le logiciel crée un arbre au hasard et il faut déplacer les branches afin que l'arbre reflète une phylogénie récente. Le fichier contenant l'alignement et l'arbre est enregistré au format Nexus puis est ouvert dans HYPHY (HYpothesis testing using PHYlogenies) qui va permettre de calculer les différentes longueurs de branches. Pour cela, le logiciel calcule la distance évolutive entre chaque taxon à partir de la distance calculée entre chaque séquence. La distance entre deux séquences apparentées est définie par le nombre de substitutions par site depuis leur divergence. Le calcul de distance est réalisé à partir du modèle de substitution défini précédemment, mais prend également en compte les mutations dites masquées, car ayant muté deux fois.

Dans HYPHY, l'arbre phylogénétique est obtenu de la manière suivante : le jeu de données est sélectionné ("select all"), et utilisé dans son intégralité pour l'analyse ("sélection => partition"). Ensuite, le type de données est sélectionné : nucléotides ou acides aminés, puis le modèle de substitution est précisé. La fonction du maximum de vraisemblance est alors construite ("build function") et les longueurs de branches sont obtenues ("optimize"). Les branches de petites tailles traduisent une vitesse d'évolution lente tandis que les branches plus longues indiquent une vitesse d'évolution plus rapide.

f) Calcul de la séquence ancestrale

La séquence ancestrale est obtenue en calculant les ancêtres à chaque nœud de l'arbre précédemment construit dans HYPHY. Pour cela il faut utiliser la fonction "ancestors" qui génère un alignement des séquences ancestrales possédant des

numéros de nœud qu'il est possible de retrouver sur l'arbre construit précédemment. Le nœud 0 représente l'ancêtre le plus ancien par rapport à l'arbre utilisé.

g) Calcul de la pression de sélection par la méthode « sliding window »

Le but de cette analyse est d'observer les contraintes évolutives appliquées à l'ensemble du gène en calculant le taux de substitution moyen sur l'ensemble de l'alignement. En effet, la méthode de la "fenêtre glissante" ou sliding window" permet de calculer la moyenne du taux de substitution sur une fenêtre de quelques bases, au lieu de calculer le taux de substitution à chaque site.

Pour réaliser cette analyse avec HYPHY, la fonction "Miscellaneous - Sliding window analysis" est sélectionnée dans l'onglet "Standard Analyses". Le fichier au format Nexus comprenant l'alignement de séquences associé à l'arbre est ensuite chargé, et le modèle de substitution est indiqué. Les paramètres de la fenêtre glissante sont ensuite définis : largeur de la fenêtre de 15 bases, pas de 5 bases. En clair, l'analyse se fait sur une sélection de 15 bases qui avance de 5 bases à chaque étape du calcul. A la fin de l'analyse, les résultats sont transférés dans un tableur afin de construire une courbe représentant le taux moyen de substitution. Plus les valeurs sur ce graphe sont grandes, plus l'évolution dans la région du gène correspondante est importante. A l'inverse, plus les valeurs sont faibles, plus l'évolution dans cette région est faible, ce qui indique des contraintes sélectives tendant à conserver ces nucléotides.

h) Calcul de la pression de sélection à chaque site par SLAC

La méthode SLAC (Single Likelihood Ancestor Counting) est utilisée dans le but d'estimer les pressions de sélections à chaque site. SLAC, disponible en ligne sur www.datamonkey.org, calcule à chaque position le rapport dN/dS dans lequel dN représente le nombre de changements non-synonymes et dS le nombre de substitutions synonymes.

$dN = NN/EN$ où NN est le nombre observé de substitutions non-synonymes et EN le nombre attendu de substitutions non-synonymes.

$dS = NS/ES$ où NS est le nombre de substitutions synonymes observé et ES le nombre de substitutions synonymes attendu.

Si $dN/dS = 1$, la région codante évolue de manière neutre, sans sélection.

Si $dN/dS < 1$, la région codante est sous sélection purifiante.

Si $dN/dS > 1$, la région codante est sous sélection positive.

Après avoir chargé sur le site internet le fichier Nexus contenant l'alignement et la phylogénie, le type de données (dans ce cas, l'analyse est faite sur les codons) ainsi que le modèle de substitution sont indiqués. Le niveau de significativité, appelé "p-value", est fixé à 0,1. Les résultats de l'analyse sont transférés dans un tableur afin de tracer une courbe représentant les pressions de sélections à chaque codon.

Les codons soumis à une sélection positive détectée par SLAC sont replacés sur l'arbre phylogénétique dans *MacClade* grâce à la fonction "Trace character". Ceux situés sur les branches terminales sont considérés comme étant non informatifs.

i) Prédiction des peptides signaux

La présence d'un peptide signal à l'extrémité N-terminale d'une protéine indique que la protéine est destinée à être excrétée hors de la cellule.

Les séquences codantes traduites en acides aminés sont analysées par le serveur SignalP 4.0 disponible en ligne (<http://www.cbs.dtu.dk/services/SignalP/>) (Petersen *et al.*, 2011a). Ce serveur est capable de prédire la présence d'un peptide signal en se basant sur l'analyse de ses trois régions caractéristiques.

j) Prédiction des modifications post-traductionnelles

Les sites potentiels de modifications post-traductionnelles sont identifiés à l'aide de la base de données *Prosite* (<http://prosite.expasy.org>) (Sigrist *et al.*, 2013). *Prosite* détecte entre autres les sites de phosphorylation et de N-glycosylation.

CHAPITRE 3

Evolutionary analysis suggests that AMTN is enamel-specific and a candidate for AI

Evolutionary analysis suggests that AMTN is enamel-specific and a candidate for AI

B. Gasse, J. Silvent et J-Y. Sire*

Université Pierre et Marie Curie, Institut de Biologie Paris-Seine, UMR7138 Evolution Paris-Seine, Paris, France

*corresponding author

Cet article a été publié en 2012 dans *Journal of Dental Research* 91(11):1085-1089.

Pour bien établir les bases de mon travail de thèse, il fallait disposer de bonnes connaissances sur l'amélotine (AMTN), ce qui n'était pas le cas car cette protéine n'était connue que chez quelques mammifères : l'homme, la souris et le rat. Mon premier objectif a donc consisté à déterminer si ce gène était présent chez tous les mammifères, et, si oui, comment il avait évolué dans cette lignée. La première étape a consisté à rechercher la séquence du gène *AMTN* dans les génomes de mammifères disponibles dans les bases de données. En tout, 42 séquences ont été extraites, provenant d'espèces représentatives des principales lignées de mammifères, ce qui représente 220 millions d'années d'évolution. Cette étude a permis d'étendre les connaissances de 3 à 45 espèces de mammifères et de déterminer les acides aminés et régions jouant un rôle important pour la fonction de la protéine. Ces derniers pourraient conduire à une amélogenèse imparfaite (AI) s'ils étaient substitués. Par ailleurs, l'accumulation de mutations dans les séquences d'AMTN d'espèces de mammifères dont les dents ne sont pas recouvertes d'émail indique que ce gène est spécifique de l'émail.

ABSTRACT

Molecular evolutionary analysis is an efficient method to predict and/or validate amino acid substitutions that could lead to a genetic disease and to highlight residues and motifs that could play an important role for the protein structure and/or function. We have applied such analysis for amelotin (AMTN), a recently identified enamel protein in rat, mouse and humans. In silico search for *AMTN* provided 42 new mammalian sequences that were added to the three published sequences to perform the analysis with a dataset representative of all lineages (circa 220 million years of evolution), including two enamel-less species, sloth and armadillo. During evolution, out of the 209 residues of human AMTN, 17 were unchanged and 34 were conservative of their chemical properties. Substituting these important residues could lead to AI. Also, AMTN possesses a well conserved signal peptide, two conserved motifs whose function is certainly important but unknown, and a putative phosphorylation site (SXE). In addition, the sequences of the two enamel-less species display mutations revealing that *AMTN* underwent pseudogeneization, which suggests that AMTN is an enamel-specific protein.

KEY WORDS: enamel matrix, evolutionary analysis, mammals, pseudogene, in silico analysis, genetic disease

INTRODUCTION

Amelogenesis imperfecta (AI) is a genetic disease characterized by abnormal formation and/or mineralization of enamel, which is prone to wear and breakage in both primary and permanent dentitions. Without other non-oral syndromic phenotypes, these disorders are mostly due to mutations of genes encoding functionally-important proteins involved in enamel formation (Witkop, 1988; Hart *et al.*, 2002, p. 200; Aldred *et al.*, 2003; Wright *et al.*, 2011). However, in some cases the genetic cause of AI is not found and other candidate genes have to be unmasked (Wright *et al.*, 2011). Among the currently identified genes responsible for AI, two encode enamel matrix proteins (EMP)— amelogenin (AMELX) and enamelin (ENAM) — and two encode enamel proteases — matrixmetalloproteinase 20 (MMP20) and kallikrein 4 (KLK4). This statement suggests that the other EMP genes, ameloblastin (*AMBN*), odontogenic ameloblast-associated (*ODAM*) and amelotin(*AMTN*), could be also candidates for AI. Amelotin (*AMTN*) was cloned first in the mouse (Iwasaki *et al.*, 2005b) then in the rat (Moffatt *et al.*, 2006b). In humans, *AMTN* is localized on chromosome 4, close to *AMBN* and *ENAM*, in a locus that was associated with AI. *In situ* hybridization and

immunohistochemistry experiments revealed the temporal and spatial localization of transcripts and protein during rodent odontogenesis, but the role of AMTN remains unclear (Iwasaki *et al.*, 2005; Moffatt *et al.*, 2006b; Trueb *et al.*, 2007; Gao *et al.*, 2010; Nishio *et al.*, 2010b; Somogyi-Ganss *et al.*, 2012). It has been suggested, however, that AMTN could play a role at the ameloblast-enamel interface during the late stages of enamel mineralization. Evolutionary analysis is an efficient method for (i) predicting and/or validating residue substitutions that could lead to a genetic disease, and (ii) highlighting amino acids and motifs that play an important role for the protein structure and/or function. We successfully applied this method to AMELX and ENAM (Delgado *et al.*, 2005b, 2007; Al-Hashimi *et al.*, 2009). Here, we have targeted AMTN using 43 sequences representing approximately 220 millions years (Ma) of mammalian evolution (Madsen, 2009). We also studied the AMTN sequence in two enamel-less species in order to know whether enamel loss in their ancestor resulted in the inactivation of the encoding gene.

MATERIAL AND METHODS

Sequence Dataset and Alignment

Three published (mouse, rat and human) and 42 unpublished mammalian *AMTNs* (i.e., computer predicted from sequenced genomes) were extracted from databases. Some sequences were completed by exploring the *NCBI Trace Archives* repository. A total of 45 nucleotide sequences, representative of all mammalian lineages, were obtained (Appendix 1). Nucleotide sequences were translated into protein sequences and aligned using *Clustal X 2.0* and *Se-Al v2.0a11*. Human AMTN was chosen as reference sequence in our alignment, which resulted in a 225 amino acid sequence, when including indels (Appendix 2). Only 0.6% data were missing. The two xenarthran sequences (armadillo and sloth) were sequenced de novo (see primers in Appendix 3) and separately analysed because they showed variations probably related to enamel lack in both species.

Evolutionary Analyses

The appropriate substitution model for our dataset — TN93 (Tamura and Nei, 1993) — was defined using an online automatic model selection tool (<http://www.datamonkey.org>: (Delpont *et al.*, 2010). In order to calculate the evolutionary distance between the 43 *AMTN* sequences and to test their relevance for the analysis, a distance matrix was established using the *HY*pothesis testing using *PHY*logenies (*Hyphy*) software (<http://www.datamonk3y.org/hyphy/doku.php>: (Kosakovsky Pond & Frost, 2005), and taking

into account (i) a recent phylogeny (Meredith *et al.*, 2011c) using *MacClade 4.08* (<http://macclade.org>) (Maddison & Maddison, 2005), (ii) the maximum likelihood method, and (iii) the TN93 substitution model. A distance tree was obtained as well as the putative ancestral AMTN sequence at each evolutionary node. Sloth and armadillo AMTN sequences were aligned with the ancestral placental sequence. A sliding window analysis of nucleotide sequence variability was conducted using *HyPhy* to identify strong functional constraints acting on *AMTN*. Ln of substitution rate per site along each branch of the mammalian phylogeny was estimated for a window of 15 bp with an overlap of 5 bp between windows. The selective pressures acting at each site of *AMTN* were identified with the codon-based *SLAC* method (*Single Likelihood Ancestor Counting*) in *HyPhy*. Using TN93 this method determines non synonymous substitutions (dN) in the alignment and compares them to the number of non synonymous changes expected at random in absence of selective pressure. We also used this method to determine the positions subjected to positive selection. One codon is considered positively selected when $dN > dS$. The test assumes that, under neutral hypothesis, the probability for a random substitution to be synonymous is $p = ES / (ES + EN)$, where EN and ES represent the expected number of normalized non-synonymous and synonymous substitutions. The result is considered significant when $p < 0.1$. Post-translationally modified sites on human and placental ancestral sequences were searched using the *Prosit* database (<http://prosite.expasy.org>: (Sigrist *et al.*, 2010). Putative signal peptides were analyzed with SignalP 4.0 (<http://www.cbs.dtu.dk/services/SignalP>: (Petersen *et al.*, 2011b).

RESULTS

Sequence Comparison

The sequences of the 42 new *AMTN* genes were validated through alignment with the mouse, rat and human sequences, including the intron splicing sites. Only three sequences — mouse lemur, hyrax and wallaby — were slightly incomplete, but with no consequence for our analyses (Appendix 2). The gene structure of *AMTN* with 9 exons, the first being non coding (not shown), is conserved in the 45 species with the exception of platypus, in which exons 4 and 5 were not found in the sequenced genome. The only variation concerns the C-terminal region, in which the stop codon is present at the end of the region encoded by exon 8 in baboon, macaque and kangaroo rat *AMTN*. Each exon has a similar length in each species, with a few exceptions resulting from insertions or deletions (indels) that have occurred independently. Most indels are located in the large and variable exon 8. Insertions are short (one or two residues), except for five amino acids inserted in dolphin *AMTN*. Deletions are larger (e.g., rabbit, pika, bushbaby). These indels contribute to sequence length variations from 195 residues in rabbit to 217 in dolphin (Appendix 2). In the primate lineage the first 3 residues encoded by exon 8 were lost in the common ancestor of the old world monkeys (Catarrhini).

Evolutionary Analyses

The distance tree indicates that the evolutionary rate of *AMTN* is relatively constant in all lineages, which means that mutations accumulate at a same rate, with a few exceptions: the high rate in platypus and rodent *AMTNs*, and the low rate in primates, except in lemurs (Fig. 1).

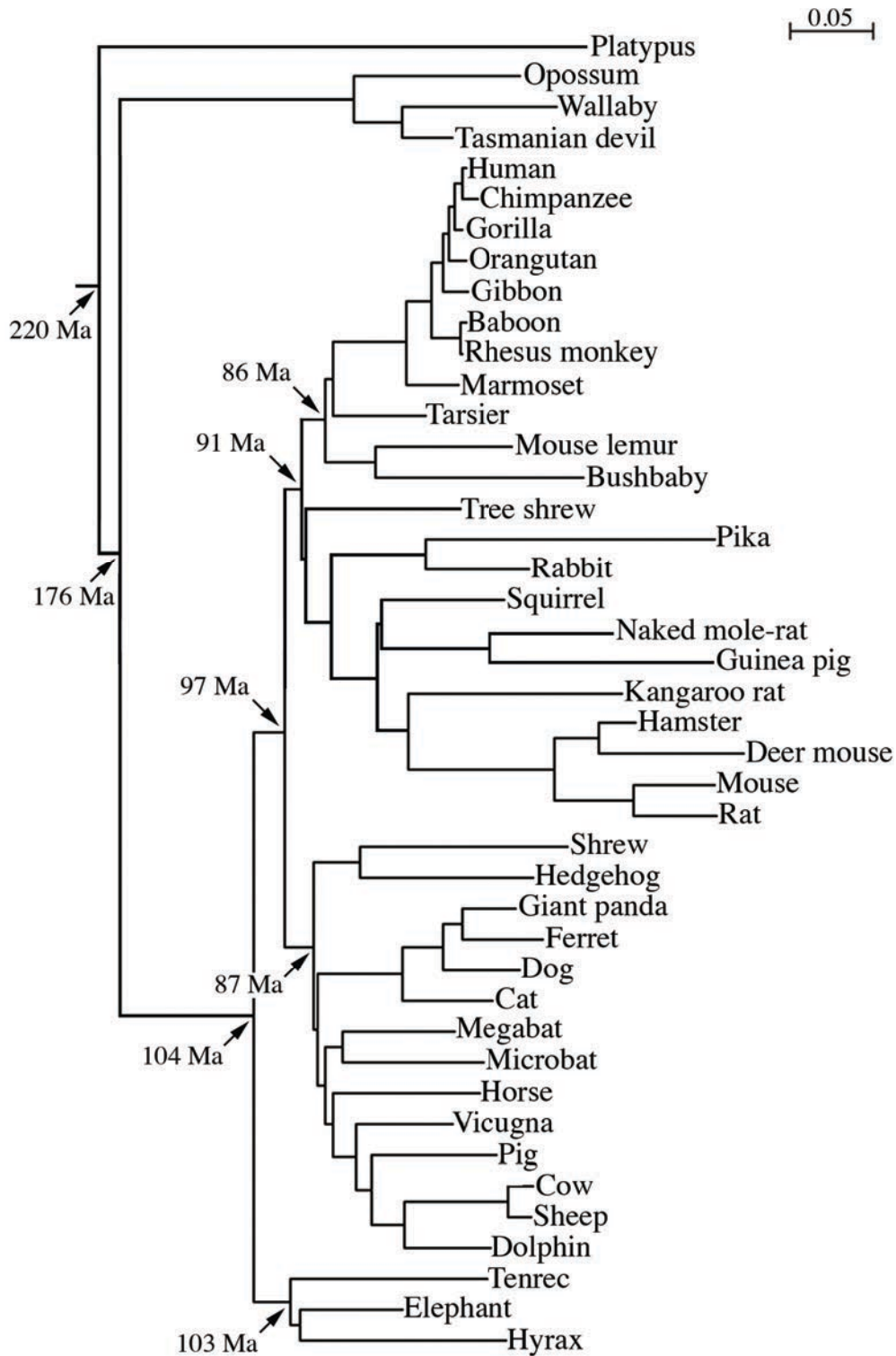


Figure 1. Distance tree of the 43 mammalian AMTNs obtained using *Hyphy* software. The branch length illustrates the evolutionary rate for each taxon. When the evolutionary rate of AMTN in a lineage leading to a taxon was high, this is represented with a long branch (e.g., in platypus, pika, guinea pig) while a slow evolutionary rate is indicated with a short branch (e.g., primates). The scale is indicated at the top. Divergence dates (Ma = million years) from (Madsen, 2009a; Murphy & Eizirik, 2009)

The analysis of purifying selection reveals 17 conserved positions, i.e., in which the amino acids were not changed during more than 220 Ma of mammalian evolution (Fig. 2; Appendix 2). These residues are of prime importance for AMTN structure and/or function. In addition, 34 positions are conservative, i.e., they can be substituted but only with an amino acid that does not change its functionality. If the chemical property of the replaced residue was different from that of the former one in that position, the substitution may result in a disorder. These 51 important positions are mainly located in regions encoded by exon 2 (10 positions out of 18), exon 4 (11 out of 22), exon 6 (7 out of 12) and exon 7 (6 out of 9). The other positions (158 in humans) are considered variable, i.e., the amino acid can be substituted with another residue without consequence for the protein. The intronic sequences that border the exons are well conserved, which indicates their crucial importance for intron splicing, and again would lead to a disorder when changed (Appendix 4). The analysis of the putative untranslated 5' *AMTN* sequences (not shown) did not reveal the presence of additional translation initiation sites (ATG) that would be in a correct context. The length of the putative signal peptide (SP) was the same (16 residues) for each AMTN sequence, and *SignalP 4.0* analysis indicated that the SP was correct. In each sequence the cleavage site is predicted to occur before the two last residues encoded by exon 2 (Fig. 2; Appendix 2).

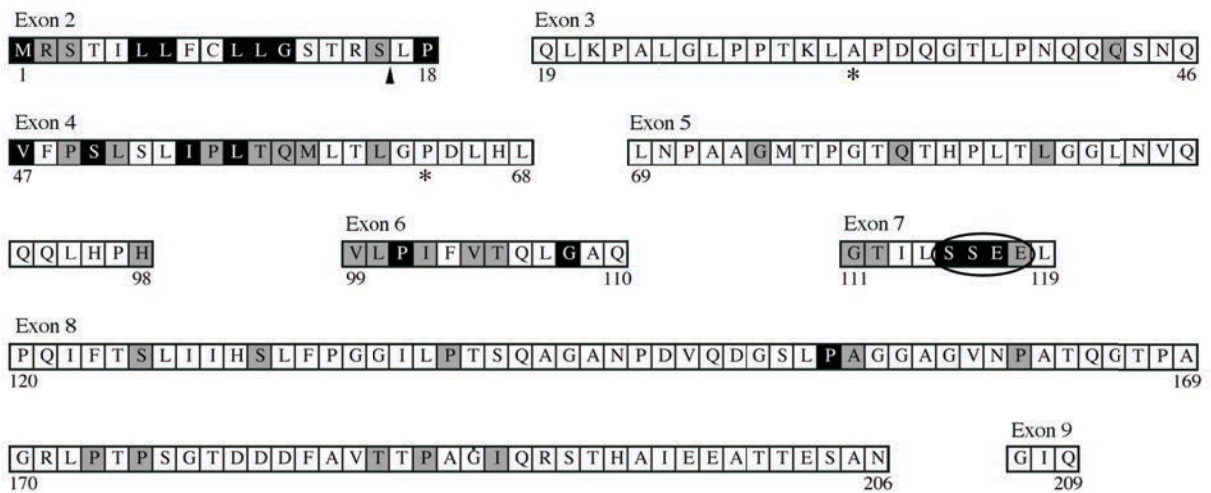


Figure 2. Amino acid sequence of human AMTN displaying the selected residues during mammalian evolution. The sequence is composed of 209 residues including the putative signal peptide encoded by exon 2. The arrow head points to the cleavage site of the signal peptide. Conserved positions, i.e., subjected to purifying selection during 220 Ma of mammalian evolution, are represented on a black background. Residues on a gray background are conservative positions where the amino acid can be replaced by a residue possessing the same properties. In humans, we predict that substitutions of unchanged and conservative residues could lead to amelogenesis imperfecta. The two positively selected residues identified in our analysis are indicated with an asterisk. Several positions are well conserved in the signal peptide and two conserved motifs are identified in the AMTN regions encoded by exon 4 and exon 6. The putative phosphorylation site (SXE) is well conserved (surrounded by an oval).

The sliding window analysis conducted using *HyPhy* identified several regions with strong functional constraints located in exons 2, 4, 6 and 7 (Fig. 3A). The selective pressures acting at each position were identified in the same regions using the *SLAC* method (Fig. 3B).

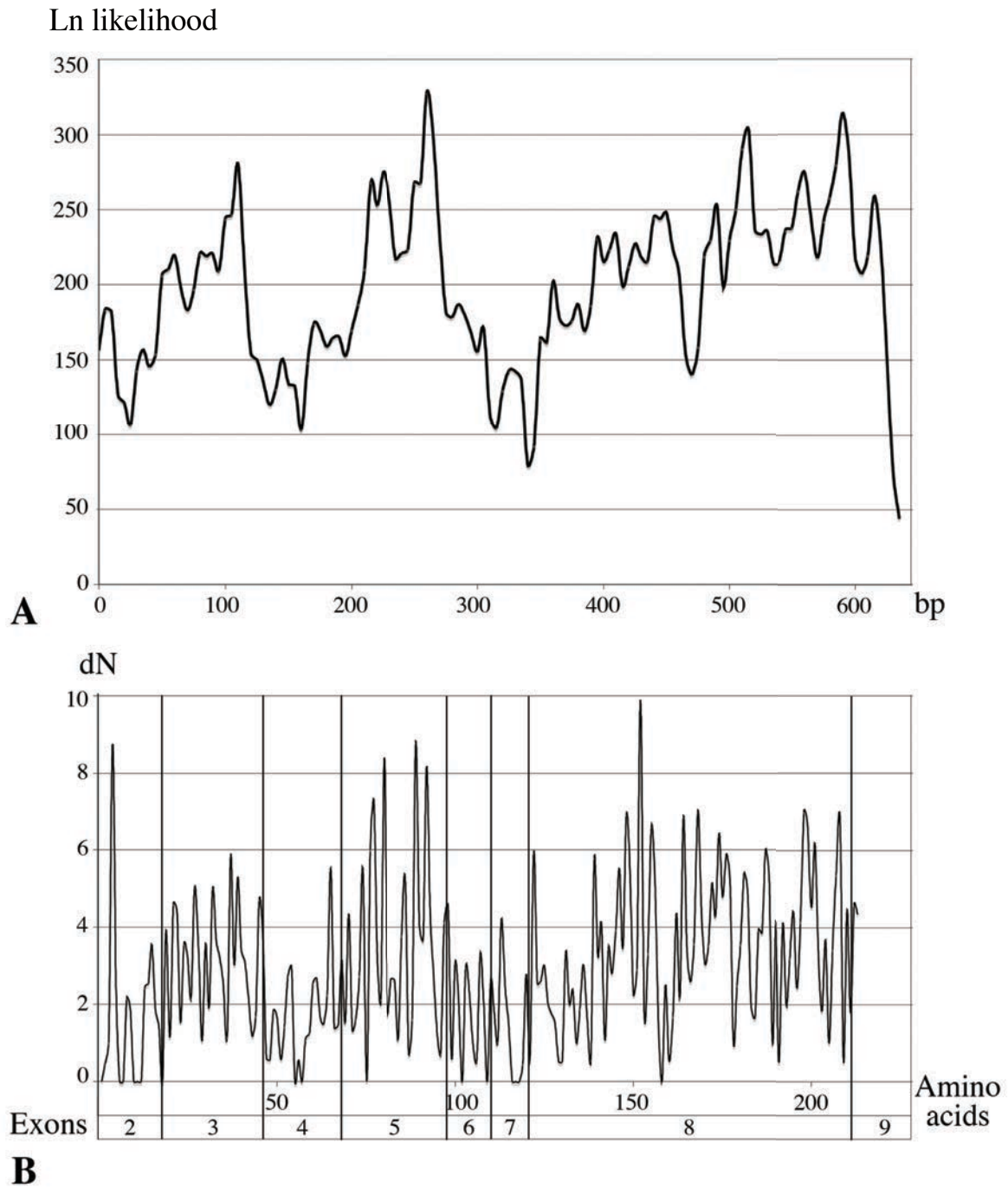


Figure 3. Substitution rate analysis of 43 mammalian AMTNs. (A) Ln of substitution rate per site along each branch of the mammalian phylogeny. Low substitution rates reflect unchanged base pairs. The peaks oriented downward indicate regions under strong functional constraints. (B) Analysis of non-synonymous substitution rate (dN) in the full-length amino acid AMTN sequences.

Therefore, the three methods of evolutionary analysis (purifying selection, sliding window and *SLAC*) pointed to the same, functionally important residues, most of them being regrouped into four, short, conserved regions of the protein: the hydrophobic region [LL(*FC*)LLG] of the SP; the region coded by the first half of exon 4 [V(*F*)PSL(*SL*)IPLTQM]; the region encoded by the 3' end of exon5 and most of exon 6 [HVLPI(*F*)VT(*QL*)G]; and the putative phosphorylation site in exon 7[GT(*IL*)SSEE]. These conserved regions were not found referenced in *Prosite* database as protein domains, and screening the AMTN sequences for post-translationally modified sites did not revealed conserved sites of importance, except the putative, well conserved phosphorylation site cited above. Several conservative residues are distributed along the AMTN sequence as, for instance, six prolines (P) in the sequence encoded by exon 8; one of them, P54 was unchanged during mammalian evolution. In addition, our analysis revealed two sites that were putatively positively selected (Fig. 2): position 32, in which the ancestral valine is replaced with alanine in carnivores and in primates; position 64, in which the ancestral serine is replaced with a proline in simiiform primates.

Enamel Less Mammals

The alignment of sloth and armadillo sequences with the ancestral AMTN of placental mammals clearly indicates that mutations have occurred in several positions that were found important for AMTN function and in some splice sites (Fig. 4). In addition, sloth *AMTN* displays a reading frameshift in exon 5 and armadillo *AMTN* in exon 8, both resulting in a premature stop codon. Our analysis indicates that these mutations have occurred independently in the two species lineages.

```

Ancestral  MKTTIDlFClLGSTQSLl KQLNPALGLPPAKLVDPDQATLLNQQQPNO VFFSLSLlIFLlTQMLTLGSDLQL
Sloth      =====
Armadillo  ===== .F.S..A..S.....G.E.....L... ..-T.....

Ancestral  LNPATGMAPGTlTLPLTLGGGLNVQQQLQFQl MLEIlIIVlAQLGAQl GAILlSSEEL PMAPQIFTGLLIQPLFFGAIL
Sloth      =====
Armadillo  ...SAEITSS...IA...D.YK..... I...L..... .I..... .V.....F...S.....

Ancestral  PlTSQAGANPDVQDGVLPAGQAGVNPlAIQGTPEGHLPTlPSGTDDEFGVlTTPAGIlQlRGlMHTTEETTESPN GIQ*
Sloth      ...E.....H.I.S.....LT.L...DYF...PE.G...AL.D..T.R...K..... ..EKAV*
Armadillo  ..T.PE.....C.....I..S.....YF.I.RAERMALE*

```

Figure 4. Alignment of the partial AMTN amino acid sequences of sloth (*Choloepus hoffmanni*) and armadillo (*Dasypus novemcinctus*), two xenarthran enamel-less species, with the ancestral AMTN sequence of placentals calculated using *HyPhy*. The crucial nucleotides of splice donor and acceptor sites are indicated in lowercase letters. Important amino acids for AMTN function and/or structure are indicated in black (unchanged positions) and grey (conservative positions) background on the ancestral sequence (see figure 2). Both xenarthran AMTN sequences display several substitutions on important positions. In sloth AMTN, a single nucleotide deletion occurred in exon 5 leading to a reading frameshift and to a premature stop codon generating a truncated protein. In addition, the splice acceptor site of exon 5 is mutated, and one amino acid of the conserved putative phosphorylation site in exon 7 is substituted. In armadillo AMTN, the splice donor site of exon 4 is mutated. Moreover, a reading frameshift in the C-terminal region of exon 8 generated a premature stop codon. Taken together, these observations indicate a relaxed functional pressure on these sequences, and thus pseudogeneization. (.) residue identical to the ancestral AMTN residue; (-) indel; (=) unknown amino acid; * stop codon.

DISCUSSION

Amelotin is Possibly a Candidate for Amelogenesis Imperfecta

This study extends our knowledge from 3 to 45 mammalian AMTN sequences. The well conserved gene structure indicates that this organization was present in the last common mammalian ancestor, and that the protein function was already defined 220 Ma (Madsen, 2009a). Given the important function of the region encoded by exon 4 (see below), the lack of exons 4 and 5 in platypus *AMTN* is surprising. Either AMTN function is ensured by another enamel matrix protein or, in this particular region, platypus genome assembly is not correct. Re-sequencing this region or sequencing *AMTN* in the echidnas, the sister lineage of platypus in monotremes, could answer this question. The distance tree does not reveal particular features. The constant substitution rate in most lineages indicates that AMTN evolved similarly and suggests the presence of variable positions, which explains the high evolutionary rates in species displaying rapid generation times (e.g., rodents). Our study

highlights four specific regions. The signal peptide sequence displays an identical length and five residues of the hydrophobic region are unchanged, indicating a strong functional constraint in AMTN secretion. The SP of the other EMPs studied, AMELX and ENAM are also conserved (Delgado *et al.*, 2007a; Al-Hashimi *et al.*, 2009) and this characteristic is shared by all members of the secretory calcium-binding phosphoprotein (SCPP) family (Kawasaki & Weiss, 2003). The putative casein kinase II phosphorylation site (SXE), already identified in previous studies, is also unchanged in the region encoded by exon 7 (SSE, positions 115-117 in human AMTN), which strongly suggests the crucial importance of this site. The presence of a SXE motif is also a characteristic of most SCPPs, in which, however, it is generally located in the N-terminal region (Kawasaki & Weiss, 2003). This indicates that the two phosphorylation sites are not homologous. Understanding the evolutionary origin of AMTN among SCPPs would reveal whether this site was inherited from an ancestral protein, from which it derived after duplication, or created in this region then recruited for its function. Two other conserved motifs were highlighted in our analyses; they certainly play an important functional role but they are not referenced as known protein domains in databases. A total of 51 positions (out of 209 in the human sequence) were found conserved or conservative during 220 Ma of mammalian evolution (purifying selection). These positions are supposed to be critical for the biological activity of AMTN. They were indicated on the human sequence and could be used as predictors of AMTN function and used as a proxy in the case AMTN is involved in AI. Intron splicing sites are also well conserved. The 158 non conservative, variable positions identify regions that are less critical for AMTN function. Except a few positions, the region encoded by exon8 largely accounts for this variability. In particular, the C-terminal extremity is highly variable and, therefore, appears not required for the protein function. Indeed, in some species a stop codon occurs prematurely, leading to a slightly different 3'UTR region and, in some cases, to the non-translation of exon 9. In contrast to the other EMPs studied, in the AMTN sequence there is neither amino acid repeats (as in AMELX: (Delgado *et al.*, 2005b; Sire *et al.*, 2007) nor exon repeats (ENAM, AMBN: (Sire *et al.*, 2007; Al-Hashimi *et al.*, 2009), which could indicate possible functional adaptation. However, in some lineages, the AMTN sequence may undergo subtle changes, that could be revealed through positive selection as for the two positions identified in this study. Therefore AMTN could be a candidate for AI and it should be included as a target gene in the diagnostic of this genetic disease. AI inheritance is probably autosomal recessive as for ENAM, MMP20 and KLK4 (Wright *et al.*, 2011). Two copies of the altered gene need to be inherited from the

parents, a probability that is weak, except for in-bred families and/or frequent substitutions of important AMTN residues in human populations.

Amelotin is Possibly Enamel Specific

The two xenarthrans studied lack enamel, which was lost more than 40 Ma and further back in time in sloth than in armadillo lineages (Davit-Béal *et al.*, 2009; Meredith *et al.*, 2009). Such species are true experiments in nature allowing to understand the fate of enamel-related protein genes. There is strong evidence that the two xenarthran *AMTN* are subjected to pseudogeneization, and that the invalidation started after the divergence of the two lineages from their common ancestor, which possessed enamel and undoubtedly functional *AMTN*. After *AMELX* (Sire *et al.*, 2008), *ENAM* and *AMBN* (Deméré *et al.*, 2008; Meredith *et al.*, 2009; Al-Hashimi *et al.*, 2010), and *MMP20* (Meredith *et al.*, 2011), *AMTN* is the fifth enamel-related protein gene that is being invalidated in enamel-less or edentulous species. *AMTN* function is therefore predicted to be enamel specific.

Acknowledgements. We thank the following colleagues for the generous gift of xenarthran tissues (via Drs F. Delsuc and F. Catseflis, ISEM, Université Montpellier II, France): Dr John Trupkiewicz (Philadelphia Zoo, 33694, USA) for *Choloepus hoffmanni* and Dr W. Jim Loughry (TX001, Texas, USA) for *Dasypus novemcinctus*. This study was financially supported by grants from CNRS and Université Pierre & Marie Curie to UMR7138, and by the 2008-2011 National "Programme Hospitalier de Recherche Clinique" (PHRC) from the French Ministry of Health, via the "Hôpitaux Universitaires de Strasbourg". The authors state there is no conflict of interest.

SUPPLEMENTARY DATA

Supplemental Appendix 1. Preferred common names, latin names, family and orders of the 45 species of mammals used for the evolutionary analysis of AMTN, with accession number in Genbank. The species are arranged in alphabetical order of common names. The three published sequences are in bold.

Common name	Genus & species	Order	Source
Alpaca	Vicugna pacos	Cetartiodactyla	ENSVPAT00000011876
Armadillo	Dasyus novemcinctus	Xenarthra	ENSDNOT00000010837
Baboon	Papio hamadryas	Primates	Blast Genome
Bushbaby	Otolemur garnettii	Primates	ENSOGAT00000003035
Cat	Felis catus	Carnivora	ENSFCAG00000007314
Chimpanzee	Pan troglodytes	Primates	XM_001160426.2
Chinese hamster	Cricetulus griseus	Rodentia	XM_003512878.1
Cow	Bos taurus	Cetartiodactyla	XM_870725.4
Deer mouse	Peromyscus maniculatus	Rodentia	GH509711 & GH509711
Dog	Canis familiaris	Carnivora	XM_849506.1
Dolphin	Tursiops truncatus	Cetartiodactyla	ENSTTRT00000008650
Elephant	Loxodonta africana	Afrotheria	XM_003414063.1
Ferret	Mustela putorius	Carnivora	Blast Genome
Giant panda	Ailuropoda melanoleuca	Carnivora	XM_002919515.1
Gibbon	Nomascus leucogenys	Primates	ENSNLET00000009808
Gorilla	Gorilla gorilla	Primates	Blast Genome
Guinea-pig	Cavia porcellus	Rodentia	ENSCPOT00000001120
Hedgehog	Erinaceus europaeus	Insectivora	ENSEEUT00000000206
Horse	Equus caballus	Perissodactyla	XM_001488949.1
Human	Homo sapiens	Primates	NM_212557.2
Hyrax	Procavia capensis	Afrotheria	ENSPCAT00000000142
Kangaroo rat	Dipodomys ordii	Rodentia	ENSDORT00000006077
Lesser hedgehog tenrec	Echinops telfairi	Afrotheria	ENSETET00000007095
Macaque	Macaca mulatta	Primates	XM_001106135.1
Marmoset	Callithrix jacchus	Primates	XM_002745764.1
Megabat	Pteropus vampyrus	Chiroptera	ENSPVAT00000004322
Microbat	Myotis lucifugus	Chiroptera	ENSMLUT00000012034
Mouse	Mus musculus	Rodentia	NM_027793.1
Mouse lemur	Microcebus murinus	Primates	ENSMICT00000005338
Opossum	Monodelphis domestica	Didelphimorpha	XM_001368673.1
Orangutan	Pongo pygmaeus	Primates	XM_002814834.1
Pig	Sus scrofa	Cetartiodactyla	ENSSSCT00000009788
Pika	Ochotona princeps	Lagomorpha	ENSOPRT00000010928
Platypus	Ornithorhynchus anatinus	Monotremata	XM_001512290.2
Rabbit	Oryctolagus cuniculus	Lagomorpha	ENSOCUT00000002587
Rat	Rattus norvegicus	Rodentia	NM_001044296.1
Sheep	Ovis aries	Cetartiodactyla	Blast Genome
Shrew	Sorex araneus	Insectivora	ENSSART00000011725
Sloth	Choloepus hoffmanni	Xenarthra	ENSCHOT00000000414
Squirrel	Spermophilus tridecemlineatus	Rodentia	ENSSTOT00000007939
Tarsier	Tarsius syrichta	Primates	ENSTSYT00000000264
Tasmanian devil	Sarcophilus harrisii	Dasyuromiiorpha	ENSSHAT00000021432
Tree shrew	Tupaia belangeri	Scandentia	ENSTBET00000017276
Wallaby	Macropus eugenii	Diprotodontia	ENSMEUT00000003096

Supplemental Appendix 2.

	[Exon2 ▼]	[Exon3]	[Exon4]	[Exon5]	[Exon6]
Human	MRSTILLFCLLGSTRSLP	-QLKPALGL--PPTKLAPDQGLTPNQQSNQ	VFPSLSLIPLTQMLTLGPDHL	LNPA-AGMTPGTQTHPLTLGGLNVQQQLHP-H	VLPIFVTQLGAQ 115
Chimpanzee	P.....E.....			
Gorilla	P.....E.....			M.....
Orangutan	.KT.....R.....			M.....
Gibbon	.KT.M.....		I.....	I.....
Baboon	.KTML.....Q.....		A.....V.L.....Q.....	M.....
Rhesus monkey	.KTM.....Q.....		A.....V.LP.....Q.....	M.....
Marmoset	.KT.....Q.....L.A.L.....Q.Q.....A.....	M.....H.....
Tarsier	.KT.....AQ.....	K.N.....V.L.....P.....S.Q.....A.....Q.....Q.....	M.....LA.....
Bushbaby	.KTM.....PAQ.....	KH.S.T.--Q.....EA..T..P..	.L.I.....S.R..	.S.-.LA.A.A.SPS.....-A..Q.-Q	M.....A.....
Mouse lemur	.KT.....Q.....	RH.N.....-T.....A.PLT..P..	.V...V.....S.....-V....L.A...P.....Q.-QIA.....
Mouse	.KTM...L....AQ... K.N.S.V--.A.PT.G.V.PLP...P..I.....L....S..P..	F...-..PH.AH.L.F...P..G...Q.-Q	M...I.A....	
Rat	.KTVV.L....AQ... R.S...A--.A.PT.G.V.PLT..P..I.....L....S..P..	F...-T..PH...L.F...P..G...Q.-Q	M...I.A....	
Deer mouse	.KT.....Q.....	K.IR.T..F--..TS....RLT.E.A..I.....A.L....S..Q..T.PAH...F.FP..P..RP...Q.-Q	M...I.AP....
Hamster	.KT.....Q.....	K.IS.T..--..-TA..VA.LT..P..I.....L....S..Q..-T.LAH.A.AL.F...P..R...QS-Q	M...I.A....
Kangaroo rat	.KTA.....Q.....	K.PN.T..Y--..PT.E.V..L....H..HLF...S..Q..	F.H.-T.IA...H.L.V...A..G...PS-Q	M...I.AS....
Guinea pig	.KTV...LG..WAQ... K.NS..EV--S...P...V.PLS..PT.S...L..AS..Q..	.T.-S.TK..A..LQ.S..A.QG.N..Q.-Q	L....A...P	
Squirrel	=====	=====S.....L....A..Q..	F...-I..A..A..L....A..G...QS-Q	M...I.A....
Rabbit	.KTA.....Q.....	K.N...E--..Q...E.PL...PT.A..A..L...EK.DG...Q.-Q	I...I.A...T.	
Pika	.KTA.....A.Q.....	K.N...VPL...QDR..E.PLK..HPS..S.....VST..-AAT.A..L...MEI.EA..K.QT-Q	M...I.A....
Tree shrew	.KTM.....Q.....	K.H==.S.--L-...=..=I.L...PK.H.....L....S..Q..V....A..L..A..R...H..Q.-Q	M...I.A....
Cow	.KAA.....L....	M.N...V--..V...A..L.P..P..S.....H....SN.Q..-...PS...L..A...K...Q.-Q	M..VI.AHF...
Sheep	.KAA.....L....	M.N...V--..V...A..L.P..PS.S.....H....S..Q..-...PS...L..A...K...Q.-Q	M..VI.AHF...
Dolphin	.KA.....L....	M.N...VV--..V.G.A..L..P..S.....Q.....-...A...L...K...Q.-Q	..VI.AH....
Pig	.KT.....L....	M.FN.V.--..V...A..R..P..S.....Q.....-I..V.SS..L...A...Q.-Q	MI.VI.AH....
Vicugna	.KTM.....L....	M.VN.T.--..A..V...A.PL..P.PS.....Q.....-...A..S..L...K...R.-Q	I..VI.AH..V.
Horse	.KTV.....L....	M.F--..--..PV...A.QL..P..N.....K..S..Q..-VA..A..L...KA...QS-Q	M....A.F...
Megabat	.KT.....QA.....	I.N...R--..V...A..L..P..GV.....S..Q.....-...A.AA..L...A...Q.-Q	M..VI.A....
Microbat	.KTV...Y....Q... M.....--..Q.V..T..L.R..PH.MVS..Q.....PP...A.AAH.L.A...Q.-Q	M..LI.A....	
Ferret	.KTM...LY...Q... T.N.V.--..P.H.A..L..P..L.....LF...S..Q..-...P..P..L..S..V..T...QS-Q	M..VI.AH...H	
Giant panda	.KTM...LY...Q... T.N.....--..P.H.AK.L..P..F...S..Q.....-...A..I..L..S..S..A...Q.-Q	M..VI.AH..TH	
Dog	.KTM...LY...Q... A.N.....--..T.H.A..L..P..F..AS..Q.....-...AS...L..S..V..T...Q.-Q	M..VI.AH...H	
Cat	.KTM...LY...Q... T.N.....--..L.A..L..L..I..A..Q.....P-..LA...L...T..P..QA-Q	M..VI.AH....	
Hedgehog	.KT.F.....HT... M.N.....--..S..V...A..LS.P.P.N.....S..Q.....T...P..L...Q.-Q	M...I.A....	
Shrew	.KT...L....Q... M.NH.....--..IN.V...A..-I...S..Q.....	R.-..-P-A..A..L..S..AV.G...Q.-Q	M..VI.A....	
Elephant	.KT.....Q.....	..N.....--SAA..V...A..L..P..L....S..QQ.....-T..A...F..N...TIK...QS-Q	L...I.A....
Hyrax	.KTM.....K.F..	..PN...F--..A.PV...A.PL...VS-	.L...V...F...S.PQ..	I...-..TAA.A.SL.VN..Q.TL-...Q.-Q	=====
Tenrec	.KTI.....Q.....	-P.N.....--..V..V.E.S..L.P..P..A..S..QM.....	.T.-T.VG...L.VN...VT...Q.-Q	M....A...T.
Opossum	.KTAV.....TIQ... -..Y.GV.--..P..GLE.PA.FTP...L..P.GQ...FSV.T.MQ	MT.-T.LL..I..IL.M...DTSIAP.VN.QQ	M..VL.A.I...		
Tasmanian devil	.KTAV.....TIQ... Q..HSGV.--..N.VL..P..LTP..PS	L..P..Q...S.IFSV.T.MP.	I...-T.LV.NI..IL.M..ADT.IA..VS.QQ	..VI.A.I..P	
Wallaby	.KTAV...G...TIQAF..-FV.GM.--..AN.VL..SA.LTP..P..	=====	I...-..LV.SI.II.M..TDA.TP..VS.-Q	..VI.A.I.T.	
Platypus	.KT...L....T.... -..VN..F.F--L.AQ.VV.HSI.QT..D...	-----	-----	-----	MI.LI.A.V...

Supplemental Appendix 2 (continued)

	[Exon7]	Exon8	[Ex9]
Human	GTILSSEEL	---PQIFTSLLIH-SLFPGGILPTSQAGAN----	PDVQDGLSLPAGGAGVNPATQGTPTAGRL-PTPSGTDD-DFAVTTTAPAGIQRSTHAIIEEATTESAN GIQ* 226
Chimpanzee	-----D.....R.....*
Gorilla	-----I.....S.....R.....T.....I.....*
Orangutan	-----G.....I.....R.....T.....*
Gibbon	-----I.....H.....N.....T.....*E*
Baboon	-----Q.....I.....T.....T.....*
Rhesus monkey	-----Q.....I.....T.....T.....*
Marmoset	-----N.....R.....Q.....S.....I.....H.....I.....S.....M.....T.....P.....*
Tarsier	-----TK-----
Bushbaby	-----QT.A.....I.....V.....E.....G.....S.....G.....TV.....PD.....*
Mouse lemur	=====	PVA.HF.G....-P.L.....PT.....-----	.G.GT....Q.....H...E.H....-TD...-E.G.....LR.GL.TT..T...PP. ===
Mouse	.AL.....	PLAS...G.L.-P...A.P.SG...TK-----	...N.V. TRQ .AKAVN...TP.HV-T.GV...D.YEMS...LR.A..TT.GT.IDPP. RT.*
Rat	.AL.....	PLAS...G.L.-P...A.Q.SG.T.K-----	...N.A. TRQ .AS.N.A.TP.H-T.AV...D.YEMS...L.A..TT.GT.MDPP. RTK*
Deer mouse	.AL.....	PLAS...G.L.HP.....S.T.K-----	...A.T.V. TRQ .AS.V...TQSQV-T.GV...D.YETS...MR.G..TT.GI.LDPP. RT.*
Hamster	.L.....	PLAS...G.L.-P.....S.T.K-----	...G.N.V. TRQ .A.V...TQ.HV-T.GV...D.YETS...R.G..TT.GT.IDTP. RT.*
Kangaroo rat	S.....	DS PLAS...G.L.-P...N.....T.PT-----	GA...VF...Q...K...I...TESHI-S...V...GG...R.GM.TTV.E..AD*
Guinea pig	AA.....	PLA...G.LLN-P...E.T.....S-----	NI.EEAI...QT...TNM.V.TE.Q.-S.GV...E...TL...KV..TP.KT..G.P. *
Squirrel	PLA...G.L.-P.....Q.N.....-----	A...V...QT...TN...SH.-A..T.I..DE.GE...R.G.QTTD.T...P. R.H*
Rabbit	V PAA.H...G.LL.-P.....TA.....-----	A.S.EV..T.Q.....IR-----A.A...IG.....PVT..T.S..PK ...*
Pika	.AV.....	PAA...L.G.LL.-PM.....GK...H-----	D...EV..T.K.E.S...R.AQE.QF-S...-V-N.-ALG..S.V...G.PPT..T...PK ...H*
Tree shrew	.V.....	PVA...G....-P.....S...T.....-----	E.HN.I...Q.....I...G.H....-D...-ELD.A...L.GM.TT..T...P. ...*
Cow	QGTS..L.G.F.-P...A.....-----	A.N.I...Q...A...A...EDPF-S.....S.....GRPTT..TP.G.PK ...*
Sheep	QGTS..L.G.F.-P...A.....-----	A.N.I...Q...A...A...EDPF-S.....A.....GRQTT..P.G.PK ...*
Dolphin	.A.....	PAT...L.GF.F.-P...A...C-PP.HLPTCP...N.I...Q...A...IR.NSE.PF-.T.....E...V...GIQTT..T.G.PK ...*	
Pig	.A.....	PATR..L.G.F.-T...A...P.P.--K-----	A.N.IH...Q...A...V...R.PF-.S.....D...L.G...T..T.G.P..M.*
Vicugna	PAA...G.F.-P...A.Q.....-----	N.I...QT...S.N...Q.PF-.V.....G..TM..T..P.P. ...*
Horse	PAA...AG..FQ-P.....-T-----	N.I...QG..K..I...SESP-.T.D...GG...GMRTT..TI.K.P. ...*
Megabat	PVA...AG..FQ-P.LS.A.....-----	N.I...Q...I...PIR.A.GSPF...AGN...G...GMLTTG.T.G.PS ...R*
Microbat	PVA...G.LF.-P.S.A...H...-----	I.NTI...Q...V.A.E.PS-.E...GM.IA..T.GFPH ...*
Ferret	.A.I.....	PVA.K...G.LFQ-P...S...P.-------	A.N.I...Q...I...AE.-FST...DI...G.A...KGM.TTQ.T.G.P. ...*
Giant panda	.A.....	PVA...G.LFQ-P...A...P.-------	A.N.I...Q...A...I...E.-FS...D...G.A...GM.TTQ.T.AGPP. ...*
Dog	.A.....	PGS...G..FQ-P...A...P.-------	A.N.I...Q...G...I...SE.-FST...D...G.A...G..TTQ.T.SGPP. .T.*
Cat	PMA...G..FQ-P...ST.N...-----	A.N.I...Q...M...I...SE.-FS...D...E.A.T.R.GM.TTQ.T.GPP. .N.*
Hedgehog	PVAS..L-G.LFQ-P.L.S.S.K.-------	T.N.I.S.QV.A...I.EDTEAPF-S...I...GM...GLQPT..T.G.P. .L.*
Shrew	PVASPLL.G.VFQ-P.LQ.A...N.I-------	LHNRI...Q.KP..LH...E.P...I...V...P.GL.LP..T...PE .F.*
Elephant	.A.....	PMA...AG.L.Q-P...A...L...T-----	E.E.I...Q...L...I.R.EKHP-S.S.D..S-V.G...L.GMRTTG.T...P. E.*
Hyrax	.A.....	PTT...G.L.Q-P...A.....VT-----	E.PE.A...Q...I...E...A.AM.-MSG..S...L...IYTT..T.D.PD ...*
Tenrec	.A.....	PVA..VY.G.L.-PW...AF.....ST-----	EG..A...Q.AAH.I...E.H....Y-LSG..A...L.DM.TT.....-D .T.*
Opossum	.AVR.....	PIA...G.L.Q-PFGT.AT...G.S.ID-----	ANT..AA...QT.G...IW..SE.Q...-SP...-V.EA.I.V...K---T.GS...AP. .TD*
Tasmanian devil	.AF.....	PMA...G.L.Q-PFGT.A...G.ERID-----	ANT..AA...QT.G.S.IW...E.Q...-P...-V.EA.N.V...---GS...AP. .TD*
Wallaby	=====	PMA...G.L.Q-PFGT.AM...G.EKI-----	ANA..TA...QT.G...IFW.A.E.QF...-P...-V.EA.I.D...---TA.GS...AP. .TD*
Platypus	.A.....	-----	G.F.Q-P.A..RMA.IG.D-------
			.ED..QV..T.QS.MS...DP.EDQP-.A..LL.-ESE..L..NL.GVSLTT..S...EPR .N.*

Supplemental Appendix 2. Alignment of the 43 amino acid sequences of mammalian AMTN used in the evolutionary analysis. The human sequence was used as a reference sequence for this alignment, which takes into consideration a recently published mammalian phylogeny (Meredith *et al.*, 2011c). This alignment leads to 226 positions (including all gaps). The signal peptide (exon 2) is underlined and its cleavage site is indicated by an arrow. Remarkably conserved residues and motifs are on gray background and squared, respectively (see Fig. 2). [] exon limits; (.) residue identical to human AMTN residue; (-) indel; (=) unknown amino acid; * stop codon.

Supplemental Appendix 3. Forward (F) and reverse (R) primers designed to amplify and sequence AMTN exons of *Dasyurus novemcinctus* (armadillo) and *Choloepus hoffmanni* (sloth) from genomic DNA extracted from ethanol-preserved tissues.

Armadillo			
Exons	Primers	Product size (bp)	Tm (°C)
3	F:CAGGGAATTA AAACTGCTCCA R:GGGAAGCCTCTCTCTCCTGT	348	58
4	F:TACCGCAAATTTGACCCATT R:TTGATTAGCATCCA ACTCTCC	404	58
5	F:GCAAATCCTCCCTTTAAACA R:TTCTAGAGAGGTTATGGGACAGC	400	56
6	F:CAGGCAATACATTAGTGGGAAA R:TGGCACACAGAAAGTTCTCAA	333	60
7	F:ACAGAGATCACAATGGCAGTT R:TGCTTCCATGGTTAGATATCAGG	377	60
8	F:GGCTCAGGAAAGGAAAGCAT R:CATCGAATGAGATTGGTGACA	648	60
9	F:TCTAGCACCAACAGGCAAAG R:TTCTCACATTGCTGGGTGTC	608	62

Sloth			
Exons	Primers	Product size (bp)	Tm (°C)
4	F:CACCAAGGGTCTTGAAAGGA R:TATTGAAGAAGGGCCAGGAG	507	62
5	F:TTCACAGTGGTCAGCTTTGC R:CTTTTCTCTTTCCCAACTCTT		59
6	F:TCTAAGTCCCCATCCAGCTC R:CTTGGCACACAGTAAGTGCTC	369	62
7	F:GGCAGCTTAACATACTGTTA R:GAAAGGATGGACTAATGAAAGC	328	56
8	F:GTTACCCTCCCCATTCTGT R:AAACACTTCCCCAGGTGCT	539	62
9	F2:AGCTCTTCCACCACCAACTG R2:CCTGTCTGAATGGAGGTGAC	699	58

Supplemental Appendix 4. Alignment of the untranslated DNA sequences (21 nucleotides) surrounding each AMTN exon in 12 species representative of the main mammalian lineages, with the human AMTN used as a reference sequence. The exons are represented with the only three first and last letters. The crucial nucleotides of splice acceptor (AG) and donor (GT) sites are in bold. (.)nucleotide identical to human AMTN nucleotide; (-) indel. Tasmanian d = Tasmanian devil.

Supplemental Appendix 4. Gasse et al.

Species	5'UTR	Exon 2	5'	Intron 2	3'	Exon 3	5'	Intron 3	
Human	ttgtaggttagcaatctgaaac	ATG // CCA		gtaagatggttatgtttgttt // atgtatttgttttcacaaaag		CAG // CAG		gtaagagtctctacaatatgga	
Tarsier	..a.....t.....	ATG // CCA	g..a.....cc.a... // ..a..c.....--c..		AAG // CAG		..t.....t.....g.gca.	
Tree shrew	..a.....t.....	ATG // CCA	g..a.....a... // ..ca..actg.....--c..		AAG // CAG	c...g..a...	
Mouse	..a.....atc....	ATG // CCA	g..a...c...ac.. // ..aa.....a.a.....--c..		AAG // CAG	c.c..ag...gaca.	
Rabbit	..ac.....t.....	ATG // CCA	gcta.g....a... // ..aa...g.....--c..		AAG // CAG	c.....g.a...	
Cow	a.t.c.....-.at.tg...	ATG // CCG	cacc....a... // ..aa.....a.....--c..		ATG // CAG	gct..	
Horse	..ac...c.....t.....	ATG // CCA	gcaa.....c... // ..aa.....a.....--c..		ATG // CAG	c...g.g.a.	
Dog	c.....t.t.....	ATG // CCG	gcta.cca.c.a... // ..taa.....ca.....--c..		GCA // CAG	c...-....g.gga..	
Microbat	..a.....t.....	ATG // CCA	gcaa.....a.c. // cctcgaaaa.g.t.tc.c..		ATG // CTG		..-.....c...c.g.ag	
Shrew	..ac....a...t.....	ATG // CCA	caa.ggt..at.c. // tct.tc..a.cacagatgc..		CTC // CAG	c...ga..c.ga..	
Elephant	..a.....t.....	ATG // CCG	c.cagc.....a... // ..aa.....a.....g...		CAG // CAG	t.....tt.g.ga..	
Tasmanian d	..t.....c...t.....	ATG // CCA	aa.t.tc.cc.aa // ..cc.g.gtg.att.ttg...		CAA // CAG	taat...t.ca.ctac	
Intron 3		3'	Exon 4	5'	Intron 4	3'	Exon 5	5'	Intron 5
Human	atcactttcaatcccctgcag	GTC // CTG		gtaagtgattggttatattat // gtcttctctccttgccctacag		TTA // CAT		gtaagttgaacagctggacct	
Tarsier	...a.....t...ca...	GTC // CTG	c.c.t.....		TTA // CAG	c...a...a..t.	
Tree shrew	...ac..t...at..ca...	GTC // CTG	ac.cc.....c... // t.....ta..t..tcg...		CTA // CAA	c..t.a.a...t.	
Mouse	cc.g.....g.....	GTC // CTG	ac.ct.....c.c // a.....ctga.....c...		TTC // CAG		...ac...ggcaag.ctt.c	
Rabbit	..c..at....gt...a...	GTC // CTG	ac.ca.....t.g... // a.....t...a.....g...		CTA // CAA	c...t.aactttag	
Cow	..t.a.....g.....	GTC // CTG	acc.t.-...c... // ..a.....c.a.....		CTA // CAA	c..agc..tact.a	
Horse	...a.....g..t....	GTC // CTG		..g...ac.ct.....c... // aaaa.....a.....		CTA // CAG	c..tgagg...tc	
Dog	c...a.....g...a...	GTC // CTG	acct.aa..ca.t. // a...c.c.g...t.....		CTA // CAA	c.g.cag.agttc	
Microbat	...a.....g.....	GTC // CTG	acc.t.....c... // acaac...t.....t....		CTA // CAA	-c...a.g...t.	
Shrew	...a.....gt..ca...	GTC // CTG	at.ac.catgtaat. // a.a.t..tta.tg..ct...		CGA // CAA	gccc...-..tt...	
Elephant	...a.....g.....	GTC // CAG	ac.ct.....c... // aaaaatc.ta...t.....		CTA // CAA	c...a..a.c.a.	
Tasmanian d	tc.tt...tt.gctt.tt..	CTC // CTG	t.ctc.ta.c.ga // a.gaaa..ta..at.....		ATA // CAA	cct.ggttctct.c	
Intron 5		3'	Exon 6	5'	Intron 6	3'	Exon 7	5'	Intron 7
Human	gttttctctcatttttccag	GTC // CAG		gtaaaaattatgcttaatat // tgtttgtcattttgtcttttag		GGC // TTG		gtaaaaaaataaaaaacta	
Tarsier	..a...a.a.ct.....	ATG // CAG		...g.t.at.tt..c.ata. //atcat.....		GGT // TTG	t..a.t...attac	
Tree shrew	ac.aaat.t.tc.c.....	ATG // CAG		...g.t.t.tt.at...at // ..a..a.tc.....t.....		GGT // TTG	t..ta.t.ttatta.	
Mouse	t.g.c.ct.t.....	ATG // CAG		...g.c.at..t.a..c.gc // ..tg..a.tg.....		GGT // TTG	tcct.t..gtc...	
Rabbit	c.gaa.t.t.gt.....	ATA // CAG		...g.t.at.tt.c.tatga // ..c..t.gtc.....		GGC // GTG	ttta.tc...ac.	
Cow	a.....tt.....	ATG // CAG		...g.t.at.tta.....g // gtg.a.t.t...c.....		GGT // CTG	c...a...tta.t.c	
Horse	a.....a..at.....	ATG // CAG		...g.t.at.tt...g.g // a.g...t...c.....		GGT // TTG	atttt..a.t	
Dog	..aa...t.a.tc.....	ATG // CAT		...g.t.at.cta...cg.a // gtg.a.t...c.....		GGT // TTG	a...t..aat	
Microbat	t.a.a.aatt.ct.....	ATG // CAG		...ggt.at.tca.tg..g // at...tgtaa.att.....		GGT // TTG	c.t.ttt.attac	
Shrew	aaa...tcact...g...	ATG // CAG		...g.taatt.taagtataa // ..t..at.at...c.....		GGT // TTG	t..c.tttatatt.t	
Elephant	a.g...g..t.....	CTG // CAG		...g.t.t.taa..tg..a // ..tg.at.t...c.....		GGT // TTG		...g....a.....tact	
Tasmanian d	a.a.a.gat.tt.....	GTC // CCG		...g.tgatcat..ttc.ga // ac...tg.tg.ctg.....		GGT // TTG	a.tg..ac...	
Intron 7		3'	Exon 8	5'	Intron 8	3'	Exon 9	5'	3' UTR
Human	tgtttgcttccctcattccag	CCA // ATG		gtaaatctcctagcttgaa // aggtggtttattttcttccag		GAA // TAA		gctgtttcaaatttttccac	
Tarsierc.g.....t..	CCA // ATG		...g.....c.tgcaa.t //c.....		GAA // TAA	c.....	
Tree shrew	...t.c.c.t.t...t..	CCA // ATG		...g.c.....gc.c...gc // ..ac.....		GAA // TAA		..t.....g..ca.ct	
Mousec.....-.....t..	CCG // ATA		...g.g.....gg.a... // ..a.ccac.ct.....t..		GAA // TAA	g..c.cc...g.	
Rabbitc.....	CCA // AGG		...g..tgt.c...g... // ..a...t.....		GAA // TAA	c.g.....	
Cow	..a...c.....t.....	CAA // AAG		...gg.....a...ca... // ..a.a.g...t.....		GAA // TAA	c...t.....	
Horsec.....	CCC // ATG		...g...t.c...aaaa. // ..aa.....		GAA // TAA	c.....	
Dog	ct...c...t.....	CCA // ATG	t..tg...ca... // ..a.....		GAA // TAA	tg.....	
Microbatc.....	CCA // ATG		...g.....c...ca... // ..a.....c.....		GAA // TAA	g.g.....	
Shrew	ctc..t.c.t...c.g...	CCA // AAG		...g...tg..ggc... // ..tgt.c.....		GAT // TAA		a.gattttt...caacta	
Elephant	...cac.....at..	CCA // ATG		...g...a.g.c... // ..a.....		AAA // TGA		..c...at..g.....	
Tasmanian d	ctc.gt....actc.c.t..	CCA // ACG		..ggg..t.t.caaaccat. // taa.a.c...g.....		GAA // TAA		.ag.c.aattt..aaaatcct	

CHAPITRE 4

Amelotin: an enamel matrix protein that experienced
distinct evolutionary histories in amphibians,
sauropsids and mammals

Amelotin: an enamel matrix protein that experienced distinct evolutionary histories in amphibians, sauropsids and mammals

Barbara Gasse¹, Ylenia Chiari², Jérémie Silvent^{1,3}, Tiphaine Davit-Béal¹ and Jean-Yves Sire^{1*}

¹ Université Pierre et Marie Curie, Institut de Biologie Paris-Seine, UMR7138 Evolution Paris-Seine, Paris, France

² University of South Alabama, Department of Biology, Mobile, 36688 AL, USA

³ Department of Structural Biology, Weizmann Institute of Science, Rehovot 76100, Israel

* Corresponding author

Cet article a été publié en mars 2015 dans *BMC Evolutionary Biology* doi: 10.1186/s12862-015-0329-x

Cet article s'inscrit dans la continuité du précédent et a pour objectif d'élargir au groupe des tétrapodes les connaissances obtenues sur amélotine des mammifères, toujours dans l'optique de mieux connaître les origines de cette protéine et son mode d'évolution. En 2012, année du début de ce travail, le génome d'*Anolis carolinensis* était le seul génome de reptile entièrement séquencé. C'est donc à partir de cette espèce que j'ai entrepris la recherche d'*AMTN* chez les non-mammaliens. Grâce au génome, des amorces spécifiques ont pu être définies et l'ADNc a été obtenu facilement. De plus, on disposait au laboratoire du séquençage des transcriptomes de mâchoire de *Caiman crocodylus* et de *Pleurodeles waltl*, ce qui m'a permis d'obtenir la séquence complète des transcrits d'*AMTN*. A partir de ces séquences, j'ai défini des amorces ciblant les régions conservées afin d'obtenir la séquence d'*AMTN* d'autres squamates et du xénope dont on possédait de l'ARN extrait de mâchoires. Alors que mon travail était terminé, *AMTN* a été identifié dans le génome du coelacanthe par un autre groupe, ce qui indique que ce gène était présent chez l'ancêtre des sarcoptérygiens il y a 430 millions d'années.

Les six séquences d'*AMTN* de reptiles et les deux séquences d'amphibiens ont été comparées à celles de quelques mammifères et du coelacanthe, révélant des différences dans la structure génique, comme la présence d'exons supplémentaires. La recherche d'*AMTN* dans le génome de deux sauropsides édentés, une tortue et un oiseau, indique que le gène n'est plus fonctionnel et confirme qu'*AMTN* est spécifique des dents.

Par ailleurs, l'expression de l'ARNm par hybridation *in situ* a été étudiée sur deux espèces : le lézard *A. carolinensis* et l'amphibien caudate *P. waltl*. Chez ces espèces, le patron d'expression diffère très largement de celui qui a été décrit chez la souris.

ABSTRACT

Background

Amelotin (AMTN) is an ameloblast-secreted protein that belongs to the secretory calcium-binding phosphoprotein (SCPP) family, which originated in early vertebrates. In rodents, AMTN is expressed during the maturation stage of amelogenesis only. This expression pattern strongly differs from the spatiotemporal expression of other ameloblast-secreted SCPPs, such as the enamel matrix proteins (EMPs). Furthermore, *AMTN* was characterized in rodents only. In this study, we applied various approaches, including *in silico* screening of databases, PCRs and transcriptome sequencing to characterize *AMTN* sequences in sauropsids and amphibians, and compared them to available mammalian and coelacanth sequences.

Results

We showed that (i) AMTN is tooth (enamel) specific and underwent pseudogenization in toothless turtles and birds, and (ii) the *AMTN* structure changed during tetrapod evolution. To infer AMTN function, we studied spatiotemporal expression of *AMTN* during amelogenesis in a salamander and a lizard, and compared the results with available expression data from mouse. We found that *AMTN* is expressed throughout amelogenesis in non-mammalian tetrapods, in contrast to its expression limited to enamel maturation in rodents.

Conclusions

Taken together our findings suggest that AMTN was primarily an EMP. Its functions were conserved in amphibians and sauropsids while a change occurred early in the mammalian lineage, modifying its expression pattern during amelogenesis and its gene structure. These changes likely led to a partial loss of AMTN function and could have a link with the emergence of prismatic enamel in mammals.

Keywords: Enamel, Amelogenesis, Tetrapods, Bioinformatics, Gene expression, Evolution

INTRODUCTION

In tetrapods, the forming enamel matrix is mainly composed of three enamel matrix proteins (EMPs) secreted by ameloblasts: amelogenin (AMEL), ameloblastin (AMBN), and enamelin (ENAM). Three ameloblast-secreted proteins are further known: odontogenic ameloblast-associated (ODAM), amelotin (AMTN) and secretory calcium-

binding phosphoprotein-proline-glutamine-rich 1 (SCPP-PQ1) (Iwasaki *et al.*, 2005; Moffatt *et al.*, 2006a, 2006b, 2008, 2014). These three proteins play a role during enamel maturation, but their precise function is still debated. Similar to EMPs, AMTN is a proline/glutamine (P/Q) rich protein that belongs to the secretory calcium-binding phosphoprotein (SCPP) family, which comprises proteins involved in mineralization of bone and tooth (dentin and enamel) tissues. In mammals only, this family also includes saliva and milk proteins (Kawasaki & Weiss, 2003, 2006; Kawasaki *et al.*, 2004). The SCPP family is originally related to the SPARC family that evolved through whole-genome duplications and gave rise to SPARC and SPARC-L1 in vertebrates (Dehal & Boore, 2005; Kawasaki *et al.*, 2007; Bertrand *et al.*, 2013). Further duplications of SPARC-L1 gave rise to the first members of bone, dentin and enamel SCPPs. The SCPP family expanded rapidly through tandem duplications as illustrated by the presence of ten SCPP genes (including *AMTN*, *AMEL*, *ENAM* and *AMBN*) in the coelacanth genome (Kawasaki & Amemiya, 2014). *AMTN* and the three EMPs were, therefore, present in the ancestral sarcopterygian genome 450 million years ago (Mya) (Hedges, 2009) and probably earlier. However, so far, even if a number of SCPPs were identified in the actinopterygian lineage (in teleost fishes), only two of them, *SPP1* and *ODAM*, were recognized as having an ortholog in sarcopterygians (Kawasaki, 2011).

In mammals, SCPP genes, with the exception of *AMEL*, are organized in two clusters (P/Q rich and acidic SCPPs) located on the same chromosome. In sauropsids (birds and reptiles), these clusters reside on two different chromosomes (Al-Hashimi *et al.*, 2010). In all genomes explored so far, the *AMTN* gene is always located upstream *AMBN* and *ENAM* (Huq *et al.*, 2005; Kawasaki *et al.*, 2007; Gasse *et al.*, 2012). Because the EMPs derive from duplication events, *AMEL*, *AMBN* and *ENAM* are phylogenetically related (Sire *et al.*, 2007). The encoding EMP genes have been identified in the genome of many tetrapods, amphibians, sauropsids and mammals (Toyosawa *et al.*, 1998; Shintani *et al.*, 2003; Al-Hashimi *et al.*, 2010), and recently in the coelacanth genome (Kawasaki & Amemiya, 2014). Although *AMTN* is known in all mammalian lineages (Gasse *et al.*, 2012), so far it was found only in one lizard (Kawasaki, 2011) and in one coelacanth (Kawasaki & Amemiya, 2014) genome. *AMTN* is therefore poorly known in non-mammalian tetrapods.

In mammals, *AMTN* is composed of nine exons, which are all encoding the protein except for the first one. As for all SCPP genes, the signal peptide is encoded by exon 2. The mature protein is encoded by the 3' end of exon 2 and the following exons (Iwasaki

et al., 2005). The translated protein is rich in proline, leucine, glutamine and threonine (52% of total by weight), and contains a conserved casein kinase 2 (CK2) phosphorylation site in the region encoded by exon 7 (SxxE/D, where x represents any amino acid). Two other transcripts were identified in rats. These splice variants lack either exon 7 or exons 3 to 7 (Moffatt *et al.*, 2006b).

A recent evolutionary analysis of AMTN in mammals has pointed to several amino acids unchanged for over 200 million years (Ma), indicating important functional sites (Gasse *et al.*, 2012). Furthermore, AMTN is also essential for enamel biology as demonstrated by the presence of pseudogenes in enamel-less mammals e.g., sloth and armadillo (Gasse *et al.*, 2012). However, in contrast to *AMEL*, *AMBN* and *ENAM*, in humans no *AMTN* mutation has been identified yet as being responsible for a genetic disease (*i.e.* amelogenesis imperfecta).

Studies in rodents have shown that AMTN is a secreted protein (Iwasaki *et al.*, 2005; Moffatt *et al.*, 2006b) that is post-translationally modified (Moffatt *et al.*, 2006b; Somogyi-Ganss *et al.*, 2012), probably by O-glycosylation (Dos Santos Neves *et al.*, 2012). In addition, AMTN interacts with itself and with ODAM, but not with AMBN, ENAM or AMEL (Holcroft & Ganss, 2011). *AMTN* is expressed by maturation-stage ameloblasts in mouse incisors (Iwasaki *et al.*, 2005). The protein is specifically localized to a basal lamina-like layer, between the ameloblasts and the enamel mineral surface in rat and mouse incisors (Moffatt *et al.*, 2006b; Nishio *et al.*, 2010b; Somogyi-Ganss *et al.*, 2012). In contrast, Gao *et al.* (2010) reports AMTN expression in the bulk enamel matrix of incisors and molars during early stages of mouse amelogenesis, although these results are thought to be artifactual (Somogyi-Ganss *et al.*, 2012).

The function of AMTN remains largely unknown and studies of its expression during amelogenesis in mice and rats led to various hypotheses, e.g. as being involved either in enamel matrix proteolysis (Iwasaki *et al.*, 2005; Gao *et al.*, 2010), or in adhesion of ameloblasts to the enamel surface (Moffatt *et al.*, 2006b), or in the establishment of the final prismless layer at the enamel surface (Somogyi-Ganss *et al.*, 2012). *In vitro* studies revealed that recombinant AMTN did not mediate attachment of any cell types (Somogyi-Ganss *et al.*, 2012), refuting any role in ameloblast adhesion. Transgenic mice that overexpressed *AMTN* under *AMEL* promoter (*AMEL* is expressed early during amelogenesis) showed a disorganized enamel, a feature consistent with a role of AMTN in disrupting the regular arrangement of hydroxyapatite prisms (Lacruz *et al.*, 2012). Moreover, a recent study has shown that a recombinant human AMTN promotes

hydroxyapatite precipitation and that the SSEEL motif in its phosphorylated form is necessary for the mineralizing property. The authors suggest that AMTN plays a role in the mineralization of the compact, nonprismatic, superficial enamel layer during maturation (Abbarin *et al.*, 2014).

Despite its importance in enamel maturation in mammals, very little is known on AMTN evolution (origin, relationships and evolutionary histories) and function in tetrapods. To address this issue, we (i) identified *AMTN* sequences in two amphibians and eight sauropsids, (ii) compared the gene structure and the amino acid composition of the protein with those from mammals and coelacanth, (iii) compared the spatiotemporal expression of *AMTN* during amelogenesis in a lizard and a salamander with the expression pattern described in rodents, and (iv) analyzed the differences in the light of tetrapod evolution.

MATERIAL AND METHODS

Amelotin sequences from publically available databases

AMTN sequences were obtained from databases (Ensembl and NCBI). See Additional file 1 for detailed information.

(i) Sequences of six representative mammalian lineages were used for comparison with non-mammalian *AMTN*: the published sequences from *Homo sapiens* (human, Primates) and *Mus musculus* (house mouse, Rodentia), and four sequences used in our previous study (Gasse *et al.*, 2012): *Bos taurus* (cow, Laurasiatheria), *Loxodonta africana* (African elephant, Afrotheria), *Monodelphis domestica* (gray short-tail opossum, Marsupiala), and *Ornithorhynchus anatinus* (platypus, Monotremata).

(ii) Computer-predicted sequence of the lizard *Anolis carolinensis* (green anole, Squamata) was used as template to find *AMTN* with BLAST (Basic Local Alignment Search Tool) in three additional sauropsid genomes available in the Whole Genome Shotgun (WGS) database: *Alligator mississippiensis* (American alligator, Crocodylia) and two snakes, *Ophiophagus hannah* (king cobra, Squamata) and *Python molurus* (Indian python, Squamata).

(iii) No hit was obtained when blasting the following genomes with sauropsid and mammalian *AMTN* sequences: the bird *Anas platyrhynchos* (mallard duck, Anseriformes), the turtle *Chelonia mydas* (green sea turtle, Testudines), the anuran *Xenopus (Silurana) tropicalis* (clawed frog, Amphibia) and the African coelacanth *Latimeria chalumnae* (Coelacanthiformes). However, as *AMTN* is located immediately

upstream *AMBN* in mammalian and reptilian genomes, we used a synteny-based approach. We extracted 100 kb of genomic DNA (gDNA) upstream *AMBN* in these four genomes and explored this region with UniDPlot, a software designed to screen DNA regions showing weak sequence similarity (Sire *et al.*, 2008). The search was performed using UniDPlot with mammalian and sauropsid *AMTN* sequences and provided a predicted *AMTN* sequence in these four species.

AMTN sequences obtained either from cDNA sequencing or from sequenced genomes available in public databases can be found in Additional file 2 and at <http://jysire.free.fr/>.

RNA extraction and PCR amplification of Amelotin transcripts

The lower jaw of a juvenile specimen of the following species was used for RNA extraction: *X. tropicalis*, *A. carolinensis*, *Python regius* (royal python, Squamata) and *Takydromus sexlineatus* (long-tailed lizard, Squamata). The primers were designed using Primer3 v.0.4.0 (Primer3 v.0.4.0, n.d.; Rozen & Skaletsky, 2000) taking into account the *AMTN* sequences of *X. tropicalis*, *A. carolinensis* and *P. molurus* obtained *in silico* (Additional file 3).

Immediately after dissection, the jaws were immersed in RNAlater (Qiagen). Total RNAs were extracted and purified (RNeasy Midi Kit; Qiagen, France), and converted into cDNA (RevertAid™ H Minus First Strand cDNA Synthesis Kit; Fermentas, France) using an oligo(dT)18 primer. *AMTN* transcripts were recovered by PCR amplification using GoTaq DNA polymerase (Promega, France), as previously described (Al-Hashimi *et al.*, 2010). Sequencing was performed by GATC Biotech.

Jaw transcriptome

Caiman crocodilus, 454 sequencing. - The lower jaw - *i.e.* mainly containing teeth, bone, cartilage, skin, muscle, and other tissues - was dissected from a 2 year-old *C. crocodilus*, cut into small pieces of about 5 mm-long each and flash frozen in liquid nitrogen. Tissues were finely ground in liquid nitrogen with a mortar and pestle and extracted following the protocol previously described (Gayral *et al.*, 2011). Total RNA quantity and quality was checked by NanoDrop spectrophotometer (Thermoscientific, Wilmington, DE, USA) and by a capillary electrophoresis in RNA 6000 Nano Lab-Chip (Agilent-Agilent Bioanalyzer). Four RNA extractions for a total of 17 µg were used for transcriptome sequencing. 454 sequencing (half a run) was commissioned to GATC

Biotech (Konstanz, Germany).

Other species, Illumina sequencing. - *P. waltl* (6 month-old), *T. mauritanica* (adult). The lower jaw was dissected, immediately immersed in RNA later (Qiagen), cut into small pieces, and frozen in liquid nitrogen. Samples were transferred into homogenization tubes containing 2.8 mm ceramic beads and 2 mL of RLT solution (Qiagen). Tissues were disrupted with the Minilys homogenizer (Ozyme), three times for 15 sec at 5,000 rpm. Total RNA was extracted following the protocol of RNeasy Fibrous tissue mini kit (Qiagen). The quantity and quality of RNA were analyzed by Experion RNA StdSens chip (Bio-Rad) and using an Experion Automated Electrophoresis system (Bio-Rad). Illumina sequencing [one run, 50 base pairs (bp) paired end] was commissioned to GATC Biotech.

Transcriptome screening

The sequenced jaw transcriptomes (see below "Jaw transcriptome sequencing") of the caudate amphibian, *Pleurodeles waltl* (sharp-ribbed salamander, Amphibia), and two sauropsids, *Caiman crocodilus* (spectacled caiman, Crocodylia) and *Tarentola mauritanica* (common wall gecko, Squamata) were assembled at ISEM-Montpellier 2 (France) and screened for *AMTN* transcripts with BLAST using the Montpellier Bioinformatics Biodiversity platform (<http://mbb.univ-montp2.fr/MBB/>).

Alignment, signal peptide and post-translational sites

The protein-coding regions of *AMTN* were translated into amino acid sequences and aligned to mammalian *AMTN* using Se-AL v2.0a11. The putative signal peptides were analyzed using SignalP 4.1 (Petersen *et al.*, 2011a) and putative remarkable sites were identified using Prosite database (Sigrist *et al.*, 2013).

Histological analyses

One adult *P. waltl*, and several young adults *A. carolinensis* (3 month- to 2 year-old) were used. Immediately after dissection, the lower jaw quadrants were fixed in a mixture of glutaraldehyde (1.5%) and paraformaldehyde (1.5%) in PBS for 2 h, at room temperature. They were demineralized for three weeks, under gentle agitation at 4°C, in the same fixative to which 5% EDTA was added. The solution was changed every two days. After washing overnight in PBS, tissues were post-fixed for 2 h in 1% osmium tetroxide, rinsed in PBS, dehydrated through a graded series of ethanol then immersed in

propylene oxide prior embedding in Epon 812 (EMS). 2 µm-thick sections were obtained using a Reichert OMU-3 ultramicrotome, deposited on a glass slide, stained with toluidine blue, and photographed with an Olympus BX61 microscope equipped with a QImaging camera using Image Pro Plus software (Media Cybernetics, Bethesda, MD).

***In situ* hybridization**

cDNA cloning and RNA probe synthesis. - *AMTN* transcript fragments were amplified from *A. carolinensis* (556 base pairs, bp) and *P. waltl* (703 bp) cDNA, using primers designed from the gDNA and transcriptome sequences, respectively (Additional file 3). Once each cDNA fragment of interest was recovered (QIAquick Gel Extraction Kit; Qiagen, France), 4 µl of the purified cDNA were inserted into a pCRII-TOPO plasmid vector containing T7 and SP6 promoters for *in vitro* RNA transcription (TOPO-TA cloning kit; Invitrogen, France), and transformed into competent *E. coli* TOP10F' bacteria. Colonies containing the vector and the insert were selected for plasmid purification (QIAprep Spin MiniPrep Kit; Qiagen, France). Purified plasmids were linearized by PCR using M13 forward and reverse primers and the sense and antisense RNA probes were synthesized using SP6 and T7 RNA polymerases (Riboprobe Combination System SP6/T7; Promega, France) in the presence of digoxigenin-UTP (Roche, France) and purified (ProbeQuant G-50 micro columns; GE Healthcare, France).

Tissue processing and in situ hybridization. - The jaws were dissected in quadrants and fixed overnight, at 4°C, in Formoy (30% formaldehyde 37%, 10% acetic acid and 60% ethanol) solution. Tissues were demineralized in 10% acetic acid for 20 to 30 days at room temperature and under gentle agitation. The solution was changed every two days. Samples were then dehydrated through an increasing series of ethanol, shortly immersed in toluene and embedded in Paraplast (Sigma, France). The sections (8 µm-thick) were obtained with a Leica RM2245 microtome and deposited on Superfrost PLUS slides (Fisher Scientific, France).

The sections were dewaxed in toluene, rehydrated through a decreasing series of ethanol, then in PBS, treated with proteinase K (0.6 µg/ml) for 5 min at 37°C, rinsed in PBS, post-fixed for 30 min in 4% paraformaldehyde, rinsed again in PBS and then in 2x SSC. The slides were incubated overnight, at 65°C, with the digoxigenin-labeled antisense probe (0.25ng/µl) in the hybridization buffer (50% formamide, 10% dextran

sulfate, 1x salt, 1x Denhardt, yeast tRNA 1 mg/ml). The following day, the slides were washed three times, at 65°C, in the washing buffer (50% formamide, 1x SSC, 0.1% Tween 20), and rinsed, at room temperature, in the Maleic Acid Buffer Tween (MABT), pH 7.5. Non-specific binding sites were blocked for 2h in a blocking solution (2% blocking reagent, 20% goat serum in MABT). Then, the slides were incubated overnight with the anti-digoxigenin antibody coupled to alkaline phosphatase (final concentration 750 mU/ml) in the blocking solution. After four baths of MABT, the slides were rinsed in NTM pH 9.5 (NaCl, TrisHCl, MgCl₂). The digoxigenin-labeled probes were revealed at 37°C using NBT/BCIP (nitro blue tetrazolium chloride/5-bromo-4-chloro-3-indoyl phosphate). The slides were mounted in Aquatex mounting medium (Merck, France), and photographed (Olympus BX61 microscope).

Ethics statement

All animal experiments conformed to the directives of the European parliament and of the council of 22 September 2010 on the protection of animals used for scientific purposes (Directive 2010/63/EU) and the French Rural Code (Article R214-87 to R214-137, Decree n° 2013-118 of 1st February 2013).

RESULTS

Amelotin gene structure in sarcopterygians

A total of 17 *AMTN* sequences of sarcopterygian species were obtained using *in silico* screening of public databases, RT-PCR on jaw cDNA, and screening of assembled jaw transcriptomes: six species representative of mammalian lineages (*Homo sapiens*, *Mus musculus*, *Bos taurus*, *Loxodonta africana*, *Monodelphis domestica* and *Ornithorhynchus anatinus*), eight sauropsids (two crocodiles: *Alligator mississippiensis* and *Caiman crocodylus*; three lizards: *Anolis carolinensis* - *AMTN* referred to as [GenBank gene ID: LOC100554538] -, *Tarentola mauritanica* and *Takydromus sexlineatus*; three snakes: *Ophiophagus hannah*, *Python molurus* and *P. regius*), two amphibians (one frog: *Xenopus tropicalis*; one salamander: *Pleurodeles waltl*), and one coelacanth (*Latimeria chalumnae*). Genbank accession numbers of *AMTN* sequences obtained from cDNA are indicated in Additional file 1. Most sequences included either complete or partial 5' and 3' untranslated regions (UTRs). No alternative splicing of *AMTN* was detected in the PCR products obtained from the cDNA of all non-mammalian species used in this study. In the gDNA sequences analyzed in GenBank,

AMTN was always found upstream *AMBN*, from 5 kb in *X. tropicalis* to 73 kb in *B. taurus*.

The exon-intron boundaries of nine non-mammalian *AMTN* sequences were inferred either from the comparison of cDNA and gDNA sequences obtained for the same species (*A. carolinensis* and *X. tropicalis*) or from sequence comparison of related species (*P. molurus* / *P. regius* / *O. hannah*, *A. mississippiensis* / *C. crocodilus*, and *T. sexlineatus* / *T. mauritanica* / *A. carolinensis*). The *AMTN* sequences in *P. molurus* and *P. regius*, on the one hand, and in *A. mississippiensis* and *C. crocodilus*, on the other hand, were nearly identical (98.8% and 97.3% nucleotide identity, respectively). For *H. sapiens*, *M. musculus*, *B. taurus*, *L. africana*, *M. domestica* and *O. anatinus* the exons were previously identified (Gasse *et al.*, 2012). For *P. waltil* *AMTN* (cDNA data only) the exon-intron boundaries were estimated from the alignment of all *AMTN* sequences. However, the limit between exon 4 and exon 5 could not be clearly defined because of the high variability of this region when compared to the homologous region in other *AMTNs*. For *L. chalumnae* (gDNA data only), the exons were identified with UniDPlot then aligned to other *AMTN* sequences.

The *AMTN* structure resulting from these comparisons is represented in Figure 1 for the main sarcopterygian lineages (Placentalia, Marsupialia, Monotremata, Crocrodilia, Squamata, Caudata, Anura and Actinistia). A total of 12 exons has been identified in the various sarcopterygian lineages, comprising the nine exons previously known in mammalian *AMTN* (exons 1 - 9), and three newly identified exons, *i.e.* two exons located between exons 2 and 3, and one exon inserted between exons 3 and 4. In order to conserve the current nomenclature of *AMTN* exons, we named 2b, 2c and 3b these three new exons, the former exons 2 and 3 becoming exons 2a and 3a (Figure 1). The largest number of exons identified in a single *AMTN* was eleven in squamates (lizards and snakes: exons 1, 2a, 2b, 2c, 3a, 3b, 4, 5, 6, 7 and 8; lack of exon 9), and the smallest was seven in *X. tropicalis* (lack of exons 2b, 3b, 4, 5 and 9). It is worth noting that the opossum *AMTN* (gDNA data only) may have either the same structure as in other mammals, *i.e.* the short exon 8 plus the exon 9 as suggested by Gasse *et al.* (Gasse *et al.*, 2012), or the structure in non-mammalian *AMTN*, *i.e.* only a large exon 8 as proposed by Kawasaki and Amemiya (Kawasaki & Amemiya, 2014). It is also possible that both *AMTN* sequences could be present in opossum as splice variants (Figures 1, 2). Similar features could exist in platypus, for which also only *AMTN* gDNA is available (Figure 1).

The 5' UTR of *AMTN* is composed of a variable number of nucleotides (nt), most in exon 1 plus 15 nt located at the beginning of exon 2a. In non-mammals, the sequence ends with exon 8 that is larger than in placental mammals and comprises the stop codon followed by the 3'UTR. The latter is also composed of a variable number of nucleotides. In placental mammals, the sequence ends with exon 9 that includes the termination codon and the 3'UTR. In non-placental mammals, platypus and opossum, both structures could occur (see below).

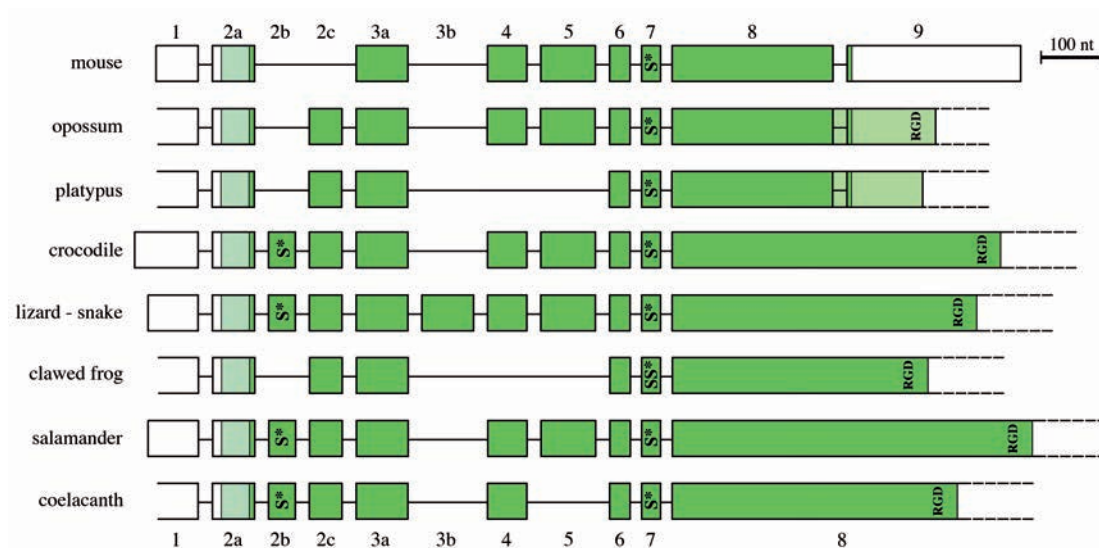


Figure 1. Amelotin structure in the main sarcopterygian lineages.

The origin of the data is indicated below the species names. cDNA means that the sequences were obtained from PCR sequencing and RNA-seq means that the sequences result from transcriptome sequencing. The exon-intron boundaries were defined using comparison of cDNA and gDNA sequences (mouse, crocodile, lizard and clawed frog), or only cDNA (salamander), or only gDNA (opossum, platypus and coelacanth) sequences. The lizard gene structure is also applicable to snakes (*O. hannah* and *P. regius*). Two alternative ends of the opossum and platypus *AMTN* are proposed (dark and light green) depending on whether or not intraexonic splicing occurs in exon 8 (see text). A total of 12 exons (blocks) were identified. Coding regions in dark green; signal peptide in light green; non-coding regions in white. Introns (lines) are not at scale. S*: putatively phosphorylated serine.

Particular features of Amelotin in sarcopterygians

A single signal peptide (SP) encoded by exon 2a was detected in the N-terminal region of all *AMTN* obtained in this study (Figure 2). The SP cleavage site was predicted to occur between position 16 (either Ala, Cys, Gly or Ser, depending on the species) and 17 (Leu, Lys or Phe) with a probability > 0.9. In most *AMTN*s, the SP is composed of 16 (15 in *X. tropicalis*) amino acids (aa), and the first two residues of the mature protein

are encoded by the last six nucleotides of exon 2a. Given the different *AMTN* structure in the various lineages, the length of the encoded protein varies from 252 aa in *X. tropicalis* to 355 aa in *P. waltl*.

Amino acid alignment indicated that the putative CK2 phosphorylation site (SxxE/D) previously identified in mammals in the region encoded by exon 7 is conserved in sarcopterygian *AMTN* (SseE) (Figure 2). In addition, in non-mammals a conserved SxE motif (putative phosphorylation site) was found in the region encoded by exon 2b, with the exception of *X. tropicalis* *AMTN* that lacks this exon. In the latter species, however, *AMTN* possesses a SxxD (SgsD: aa 78-81) and a SxE (SeE: aa 82-84) phosphorylation motifs adjacent one another in the region encoded by exon 7. A conserved Arg-Gly-Asp (RGD) motif (putative integrin-binding site) was found in the C-ter region encoded by exon 8 in non-mammals but it is missing in mammals, with the possible exception of marsupials (see below and Figure 2).

In addition to these features, in sauropsid *AMTN*, exon 2c and exon 3a encode the conserved motifs "PQQTQ" and "GLPPA", respectively; also, the motif "KLVPD" encoded by exon 3a is well conserved in squamates. In the latter no particular motif is found in the *AMTN* region encoded by exon 3b (only found in squamates) while a conserved motif "PQQL" is encoded by exon 5. Also, exon 6 encodes 12 amino acids that are well-conserved in all species studied. No particular function is known for these conserved motifs.

Figure 2. Amino acid alignment of the Amelotin sequences used in our study.

The sequences are ordered according to species relationships (see Additional file 1 for names and references). Non-mammalian tetrapod sequences are aligned against the human sequence and four *AMTN* sequences representative of the main mammalian lineages. The coelacanth (*Latimeria*) sequence is aligned (only gDNA sequence) at the bottom. Two alternative ends of the opossum *AMTN* (*Monodelphis* V1 and V2) are proposed depending on whether or not intraexonic splicing occurs in exon 8 (see text). The specific length of each sequence is indicated in brackets at its end. Signal peptide underlined. SxE and SxxE motifs (encoded by exon 2b and exon 7, respectively) and RGD motif (encoded by the end of exon 8) highlighted in grey. Vertical lines: exon limits; (.): residue identical to human *AMTN* residue; (-): indel; (?): unknown amino acid; *: stop codon.

Figure 2. Gasse et al.

	exon 2a	exon 2b	exon 2c	exon 3a
Homo	MRSTILLFCLLGGSTRSLP	-----	-----	-QLKPALGLPPTKLPADQ-GTLPNQQQSN-Q
Mus	.KTM...L....AQ...	-----	-----	K..N..S.V.A..PT.G.VTP...P.-
Bos	.KAA.....L...	-----	-----	M..N..V....V...-A..L.P..P.-
Loxodonta	.KT.....Q...	-----	-----	K..N....SAA..V...-A..L...P.-
Monodelphis	.KTAV.....TIQ...	-----	-----	Q..Y.GV...P..GLE..-PA.FTP...-
Caiman	.KTV...L...TAF...	VRQAGTFL---PVSNSRE	ILQLMRLYSSSLGNIPQQTQ	Q.VN.G....AQ.L...-TA.LA...A.-
Ophiophagus	.EVI...LS...L.F.T.	INHIGGSL---ATSSSKE	VLRLLRHYIAISNTPQQAQ	SRQR-.I...A..V..LNLPG...-PP.V-
Python	.KIV...L...LAF.K.	VNQAARIL---ATSNSRE	ILRLLRQYRATGNTPQQTQ	PRPRVVP...A..V..MNLPE...-LP.G-
Anolis	.KIV...LS...L.FC.	VNQFGRRF---ATSNSRE	ILRLMQRKYKAGNVPPQQTQ	QRSN.GI...A..V...H-P.A.-HIPSE-
Takydromus	????..LP..L.I...	LNQMGRIL---ATSNSRE	ILRLMQRKYKAMGNPNPQQTQ	QRPRLGA...A..V...NPSV...-VP.E.
Tarentola	.KIV..FLSF..L.IA..	ISPLERAI---ATSNSRE	ILRLMQRYRAMGNPNPQQTQ	QRSN.GF...A..V...-TA.A...-LP.E-
Xenopus	..VMF..L...RASLGF.	-----	MLHLMQMFNLLGRVPEQSQ	-RI..H.FVA.SQ.V-QLH..QLRYMHP-F.
Pleurodeles	.T...L...TSIAF.	VREAERIL---SASNSGE	MLQLMQLYNMGLNTPPEQTQ	..VH.GH...AQ.I..MTAQQTQ..L.QF.
Latimeria	.KTIM...F..TAVAV.	VHSGVRSVHGGVAASNSRE	RTPLLV---SQQPQLQQGQ	PRI-LQ.QTF.ER-V-QV-PVQ..I.-GQQ.

	exon 3b	exon 4	exon 5
Homo	-----	-----VFPSSLIPLTQMLTLGPD-L-HL	LNPAGMTPGTQTHPLTLGLGNLQQQLHP--H-
Mus	-----	...I...L...S...-P.	F....-PH.AH.L.F...P..G...Q.--Q-
Bos	-----	...S...H...SN--Q.	...PS...L..A...K...Q...Q--Q-
Loxodonta	-----	...L...S...-QQ	...T..A...F..N...TIK...QS--Q-
Monodelphis	-----	...L..P..GQ...F.SV.T.M--Q.	MT..T.LL..I..IL.M...DTSIAP.VN--QQ-
Caiman	-----	...T.N.F...RVNPF.A.I--N.	V.G...GV.I...L.TGA..A.TPH..V.-QQG
Ophiophagus	-----	...NL.V...-V.QN.IFPI--M--NY	.STLL.ILSTI..L.IVG..MT.P...-PGQA
Python	VTPFEWPTNPFVI---SQLA.EVGGPNLQ	...N..VF...-Q.N..YPI--W--NY	.TTLL.ILSTM.AL.AAG..MTIP...L.-GQP
Anolis	IPAFERPPFANYPILT--AQNPQELGSPTLQ	...N..GFNVL...VPI--I--NY	.ALLL.ALST...Q..PG...IP...L.PPQS
Takydromus	AVPVEWTFANFPVTSNFPQNPSELGGPSIQ	...NI.GYNVL.F...PM...-NT	.AILL.ALTL.P..L.GAGV.TTMP...L.-PQP
Tarentola	VTPLEWPTNPFMTP--PKIPSELGTTNLQ	...HI.GFN.L.F...IP...-I--NT	.TTLL.LGGLPS.QT.PV..V..P...L.PSQ-
Xenopus	-----	-----	-----
Pleurodeles	-----	PSENVFGNL.I.PPLSMPPGPFQS.G--NQ	..P.-VM.YP.AL.MNV.EMT.HHSMF.HP.H
Latimeria	-----	-----LL..YRFV...L.AQ..GVHAQQ	-----

	exon 6	exon 7	exon 8	exon 9
Homo	VLPIFVTQLGAQ	GTILS-SEEL	----PQIFTSIIHSL-----FPG-----	GILPQSQAGANPDV---QDGS-----L-PAGGA
Mus	M...I.A....	.AL...SEEL	-PLAS...G.L..P.	-A.P.SG...TK...N.V-----TRQ.
Bos	M..VI.AHF.	-QGTS..L.G..F.P.	-A.....A-----N.I-----Q.
Loxodonta	L...I.A....	.A....	-PMA...AG.L.QP.	-A...L..T.E..E.I-----Q.
Monodelphis	M..VL.A.I..	.AVR....	-PIA...G.L.QPF	-AT..G.S.IDANT...AA-----QT
Caiman	..LILG.M.P.	.LGV...M	-QTDA..LAG..L-LG	-F.QAN..SI.SEG--P.IV-----F...QT
Ophiophagus	M...I...M.P.	.AV...M	--SPS.F..G.L.-PG	--MQ-----AMF.SG..EPP.EA-----F-----V.Q.
Python	M...I.A.M.P.	.AV...M	--SPS.F..G.L.-PG	--MQ-----AL.SG..EPP.EG-----GEEEL-----Q.
Anolis	MF..I.F..M.P.	.V...M	--PT..VMAAFL-L-PG	--L-----LF.SG.S..L.ED-----EL-----QV
Takydromus	M...I.A.M.P.	.AV...M	--SPQS..V-GVL.-PG	--L-----FF.LG...L.EG-----ESL-----NQ.
Tarentola	I...IFA.M.P.	.AV...A	--QSPS.F..G.L.-PG	--MQ-----ILI.SG..E.L.EG-----E.L-----Q.
Xenopus	MS..MFA..RQ.	ALSG..D..N	-----VLPGV.MIPL	-----SA-----LDEN.ATVH...VII.NSILQEQPV-----N...QH
Pleurodeles	LP.LII..M.Q.	.LLM...M	-PFL..V.AGV.MPPFGPILPALVLP.VPAFAGLP.LP.LL.GA.R.AA	-----GA.ISLDRQEA-I..M.Q.
Latimeria	A.AVNIP..VP.	--ST...M	VQMSTGVLIPIVVDPA-----VP.	-----V.AIEMPV-----PG.I-----NP

	exon 8	exon 9
Homo	GVPNATQGTAGRLP--TPSGT-DDDFAVTTPAGIQRSTHAIIEEATTESAN	GIQ* (209)
Mus	.AKAVN...TP.HVT--GV.D...YEMS...LR.A..TT.GT.IDPP	RT.* (213)
Bos	.A...A...EDPFS-----S.....GRPTT..TP.G.PK	...* (211)
Loxodonta	.L...I.R..EKHPS--S.D.-SV.G...L..GMRTTG.T...P	E.* (213)
Monodelphis V1	.G...I.W..SE.Q...SP..V.EA.I.V...K--T.GS...AP.	.TD* (229)
Monodelphis V2	.G...I.W..SE.Q...SP..V.EA.I.V...K--T.GS...APN	-----GGFFQTHIMPLGPKETAEGSMRQKFLGDPSTLADM
Caiman	.M.QGA---P.Q...S.E..-A.VSGA.L...-KG-SQTP.SI..A.S	-----GVYTPPGFRQLDSEQV-TDASFAEP--AILLNM
Ophiophagus	-----N.E.....A...KVFPATDGLS.IPG	-----ELYPTPSGFRQSDTGTSG--LFLEP--TADINM
Python	.S.....E.....A...KVFPATDGLS.IPV	-----ELYPTPSGFRQSDVGTSG--MFVQP--TVDISM
Anolis	.T.....Q.N...-F.E.S.....A...KQSP.T.DGLNVAAG	-----AFYPTPSGFRQPGVVTNE--VFVEP--TGGIHM
Takydromus	.S.....Q.N...-F.E.S.....A.A..KISPTASDGLS.V.S	-----GVYPTPSGFRQPGVVTNG--V???????????
Tarentola	.S..N--S.Q...-F.E..-A.A--A...KVSPGMHDDLSTG..	-----GLNPTPSGFRQPGYNHGATQDPFAERTILVKLNM
Xenopus	.ST.I.NKILPE..WVI.TQA	-----AQNDVSV.GI IK
Pleurodeles	.L..NK.RETASPA.GV..AQDTGEVLAV.V..LRKGV-TGTD-S-AGLLQ	-----TFTPPAGLRQFPTEA-TEVMFVEPTGLVETET-
Latimeria	.GN.S...-DT.-I.GNP--VGGQP.V.T.VRKIAPTA.VSV.DAVNQGIFQPVAGATPTAAGLKKTVQDEDRPDIIVFVEPSPCDTIP	-----

	exon 8
Monodelphis V2	MPSNTRFPPSSVSELIPTGGNLMIPVAPQ-DPDSATSPMTLILRGDTHLPV* (312)
Caiman	EANEIQAPPTT---DIGTGTINSINLQT---QGYVNQIGVSPVRGDRALPLANRDVRLALREYKSRSA* (328)
Ophiophagus	EPSELREPPPTSVPKFDMSNMQSP-----HNLSQPLVRGDSHIPMNTVASKPLNLP* (294)
Python	EPSELREPPTSLVKLDMGNMQSP-----HNLSQPLVRGDSHILVNTVASKPLNVP* (325)
Anolis	EPSELREPPTSLVGLDKGNKQL-----QNLSPVVRGDSYMPMTTASPLKAP* (334)
Tarentola	EPSELREPPPTMARPEDDNSLGG-----HTLAQSPVRGDRHMPVTTVVSKLLREP* (348)
Xenopus	GRDNDLTVNAHTMGEENWILCNS-----YVPDLCELNHIVRGDGYIPKTPVYFKNGIQGNRNTESY* (252)
Pleurodeles	SPDGVDE-----RLPHSYRP-----QNRQFNVSARGDQNLAMAVPDLKPLRKHKKRRA* (355)
Latimeria	NGMKQLTFSSELITVSPAASITISSVTNSMPKKNWHIASKAGLRGDIAGSLKAAGKKPIQNY* (306)

Intraexonic splicing in mammalian Amelotin exon 8

In non-mammals, *AMTN* exon 8 is in average 975 bp-long and encodes a variable region of the protein, the C-terminal region, which houses a well-conserved RGD motif (Figure 2). Exon 8 also encodes the stop codon, and contains the 3'UTR. In contrast, in placental mammals, exon 8 is shorter than in the other lineages. A large part of the 3' region, including the region encoding the RGD motif, was lost and does not encode the termination codon, and is followed by intron 8 and exon 9. It has previously been suggested that the 3' end of exon 8 has been spliced out in placental mammals (Kawasaki & Amemiya, 2014). Here, we provide an alignment of the 3' extremity of mammalian exon 8 with the homologous intronic region in sauropsid sequences showing that nucleotide substitution in an ancestral mammalian exon 8 generated a favorable environment for an intraexonic splice donor site (Figure 3). This intraexonic splicing occurs within a codon, creating a phase 1 intron (*i.e.* the last nucleotide of exon 8 should be added to the two first nucleotides of the next exon to form a codon). A ninth exon possessing eight nucleotides followed by a stop codon and the 3'UTR was then recruited. In platypus and opossum, however, screening of the gDNA sequence provided two putative sequences depending on whether the intraexonic splicing occurs in exon 8 or not (Figure 2). In the former hypothesis the *AMTN* structure in the monotreme and marsupial lineages would be similar to that observed in placental mammals, *i.e.* a short exon 8 followed with exon 9 (Gasse *et al.*, 2012). If the intraexonic splicing did not occur, an RGD motif would be encoded in the C-ter region of the opossum *AMTN* (Kawasaki & Amemiya, 2014) but not in the platypus one (not shown). In opossum, the amino acids encoded by this exon 8 region can hardly be aligned to the homologous, non-mammalian sequences (Figure 2). In platypus alignment of this region is not possible (not shown).

		↓	
Homo	GAAGCCACCACAGAATCAGCAAAT	Ggt	aaattctcctagcttggaacatg
Mus	GGAACCACCATAGATCCACCAAAT	Ag	taagttgtcctagggtagaacatg
Bos	GAAACCCCCACAGGATCACCCAAA	Ggt	aggttctccaagccaggaagaca
Loxodonta	GAAACCACCACAGATCACCAAAT	Ggt	aaagttctccaaggctggaaaaatg
Monodelphis	GGGAGCACCACCGAGGCACCAAAT	Ggt	gggtttttccaacccatataat
Caiman	GAGAGCATCACTGAGGCAGCCAGT	GGGGT	TATACCACTCCACCTGGCTT
Python	GATGGACTGAGTAAAATCCAGAT	GAGTT	TACCAACTCCCTCTGGCTT
Anolis	GATGGCCTGAATGTAGCCCAGGT	GCCTT	TACCAACCCCTCTGGTTT
Takydromus	GATGGTCTAAGCGAGGTTGCAAGC	GGGGT	TACCAACCCCTCAGGCTT
Tarentola	GATGACCTGAGCGGACAGCAAAT	GGACT	GAATCCAACCTCCGCGATT

Figure 3. Intraexonic splicing of Amelotin exon 8 in mammals.

The nucleotidic region of the *AMTN* exon 8 of five mammals and five sauropsids were aligned. Note the intraexonic splice donor site (arrow) that appeared in an ancestral mammal. A favorable environment for splicing (highlighted in grey) was created when the sequence was changed into **Ggtg** (then conserved in monotreme and marsupial lineages as shown here) and then into **Ggta** in an ancestral placental. Coding sequence in bold and upper case letters; new intron in lower case letters.

Amelotin in toothless sauropsids

AMTN was searched in the sequenced genomes of the bird *Anas platyrhynchos* and of the turtle *Chelonia mydas* in order to check whether the gene is active or has been invalidated as a consequence of tooth loss in these two lineages. In the two species, our synteny-based approach allowed to identify φ *AMTN* in the gDNA region immediately upstream φ *AMBN*, as expected when taking into account the gene synteny observed in mammalian and sauropsid gDNA (Figure 4A). The genomic distance between the two pseudogenes is 8 kb in *A. platyrhynchos* and 1 kb in *C. mydas*. These findings support that *AMTN* is only a tooth protein in the sauropsid lineage.

In *A. platyrhynchos*, most of the φ *AMTN* sequence is hardly recognizable in the non-coding gDNA region of scaffold247. Only two exons (2b and 6) could be identified within the 15 kb genomic region upstream φ *AMBN*. They display many nucleotide substitutions and indels (Figure 4B). In *C. mydas*, *AMTN* is also a pseudogene found in contig173658. Numerous mutations have accumulated in the sequence, but several exons (2a, 2b, 4, 5, 6, and 7) were identified using UniDPlot within the 10 kb upstream φ *AMBN*. They display less variations compared to *A. platyrhynchos* *AMTN* (Figure 4B).

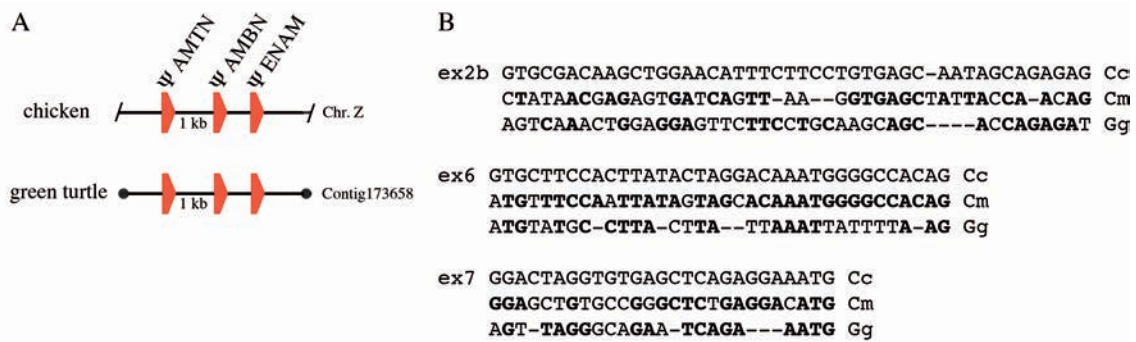


Figure 4. Amelotin pseudogenization in two toothless sauropsids.

A. Location of pseudogenized *AMTN* (φ *AMTN*) in the duck and green turtle genomes. In these two genomes, φ *AMTN* was found (synteny-based approach) upstream φ *AMBN* (ameloblastin) and φ *ENAM* (enamelin), all EMP genes being pseudogenized in toothless sauropsids.

B. Alignment of the nucleotide sequences of φ *AMTN* exons 2b and 6 of the green turtle *Chelonia mydas* (Cm) and the duck *Anas platyrhynchos* (Ap) with the corresponding coding exons of the crocodile *Caiman crocodilus* (Cc), their closest toothed relative. Note the few sequence variations in turtle φ *AMTN* exon 6. Conserved nucleotides in bold.

Searching for functional and invalidated exons in sarcopterygian gDNA sequences

Our results indicated the presence of a variable number of exons in sarcopterygian *AMTNs* (Figures 1, 2). To verify whether exons were present but unidentified in some species, we screened target regions (intronic region between two exons) in sequenced genomes available in Genbank.

Mammalian Amelotin

In mammals, many *AMTN* sequences were previously obtained by blasting sequenced genomes, and the presence of nine exons was confirmed (Gasse *et al.*, 2012), but possibly eight in marsupials and seven in monotremes, depending on the intraexonic splicing in exon 8 described above. We looked for the putative presence of exons 2b, 2c and 3b in mammalian sequences by screening introns 2 and 3 of mammalian gDNA. We confirmed the presence of a putative exon 2c in *M. domestica* *AMTN* gDNA sequence (Figures 1, 2), then in the homologous genomic region of two other marsupials (Tasmanian devil: *Sarcophilus harrisii*; tammar wallaby: *Macropus eugenii*) and of a monotreme (platypus: *Ornithorhynchus anatinus*) (not shown). This exon 2c (57 bp) possessed high nucleotide identity with sauropsidian exon 2c and displayed correct putative splice sites, which suggests that exon 2c is functional in marsupial and monotreme *AMTN*. Exon 2c was not found in placental mammals but weak nucleotide similarity in the homologous genomic region of various mammalian species (not shown)

suggests that exon 2c was present in an ancestral placental *AMTN*, then invalidated early, before the diversification of current placental lineages. In contrast, neither functional nor pseudoexons 2b and 3b were found in any mammalian genomes studied to date, which suggests that these two exons were either not present in the *AMTN* sequence of the last common ancestral mammal or invalidated early in the mammalian lineage, and no longer recognizable in gDNA.

Sauropsid Amelotin

In sauropsids, *AMTN* is composed of 11 exons in squamates (no exon 9) and of 10 exons in crocodiles (lack of exons 3b and 9) (Figures 1, 2). The exon 9 found in all mammalian *AMTN* was identified neither in sauropsid gDNA nor in cDNA, all sequences ending with exon 8. Blasting the genomic region (*i.e.*, 2 kb downstream exon 8) of sauropsids (*A. mississippiensis*, *A. carolinensis*, *P. molurus*, and *O. hannah*) using mammalian exon 9 sequence did not provide valuable hit. Exon 3b was identified in squamates only. We checked whether this exon was missed in *A. mississippiensis* gDNA, but no hit (even as pseudoexon) was obtained.

Amphibian Amelotin

In amphibians, *AMTN* was found to be composed either of 10 exons (absence of exons 3b and 9) in *P. waltl* (caudates) or of seven exons (absence of exons 2b, 3b, 4, 5 and 9) in *X. tropicalis* (anurans) (Figures 1, 2). Exon 9 was identified neither in *P. waltl* nor in *X. tropicalis* cDNA, the two sequences ending with exon 8. Blasting the genomic region (*i.e.*, 2 kb downstream exon 8) of *X. tropicalis* did not provide valuable hit. Similarly, exons 2b, 3b, 4 and 5 were not identified in the target genomic regions (introns 2 and 3) of *X. tropicalis* and no pseudoexons were found.

Coelacanth Amelotin

Exons 3b, 5 and 9 were not found in coelacanth *AMTN* (Figures 1, 2). We did not identify any of them, even as pseudoexons, when blasting the target gDNA regions (introns 3 and 4, and 2 kb downstream exon 8).

Amelotin expression during amelogenesis in lizard and salamander

Lizards and salamanders have polyphyodont dentitions - their teeth are continuously shed and replaced during the lifetime of the animal. Therefore, in a few, serially sectioned jaw quadrants of *Anolis carolinensis* and of *Pleurodeles waltl* we found various stages of tooth development, allowing the study of *AMTN* expression during various steps of amelogenesis in the two species (Figures 5, 6).

Anolis carolinensis

In the late cytodifferentiation stage, ameloblasts have differentiated from the inner dental epithelium of the enamel organ and they are facing mesenchymal cells of the dental papilla differentiated into odontoblasts. As soon as a thin layer of pre-dentin was deposited between the ameloblasts and odontoblasts, *AMTN* mRNAs were detected in the ameloblasts surrounding the forming teeth (Figure 5A, B). *AMTN* expression signal increased in intensity as enamel matrix was deposited and its mineralization started (Figure 5C-F). In tricuspid teeth, enamel matrix was deposited first on the central major cusp, then on the two adjacent minor cusps. *AMTN* expression weakened in the ameloblasts facing the major cusp at the time enamel maturation progresses, while *AMTN* transcripts were strongly labelled in ameloblasts surrounding the minor cusps (Figure 5G, H). *AMTN* expression was still detected in reduced ameloblasts during late maturation stage just before the tooth root attaches to the dental bone (Figure 5I, J). After eruption, when the tooth became functional, *AMTN* transcripts were no longer detected in the inner dental epithelial cells (not shown).

Figure 5. Amelotin expression during amelogenesis in the iguanid lizard *Anolis carolinensis*.

Longitudinal (A-H, K) and transverse (I, J) sections of demineralized lower jaws (oral cavity on the top). In this lizard the dentition is heterodont: anterior teeth are monocuspid (A-F, I, J) and posterior teeth are tricuspid (G, H, K). A, C, E, G, I: Histology, toluidine blue staining; B, D, F, H, J, K: *in situ* hybridization with *AMTN* antisense mRNA probe; K: Negative control with *AMTN* sense mRNA probe.

A, B: Late cytodifferentiation. *AMTN* transcripts are weakly labelled in the ameloblasts lining the pre-dentin matrix. C, D: Early deposition of enamel matrix. *AMTN* transcripts are detected in the well-polarized ameloblasts facing the enamel matrix (dark blue in C). E, F: Enamel deposition. The ameloblasts along the deposited enamel matrix are well labeled. G, H: Enamel maturation. *AMTN* mRNAs are strongly identified in the ameloblasts facing the enamel layer. The latter is well mineralized with the exception of the regions covering the minor mesial and distal cusps. I, J: *AMTN* expression is still detected during late maturation stage before the tooth attachment process begins.

am: ameloblast, db: dentary bone; de: dentine; en: enamel; es: demineralized enamel space; ide: inner dental epithelium; od: odontoblasts; ode: outer dental epithelium; oe: oral epithelium; pc: pulp cavity; pd: pre-dentin; sr: stellate reticulum. Scale bars = 50 μ m.

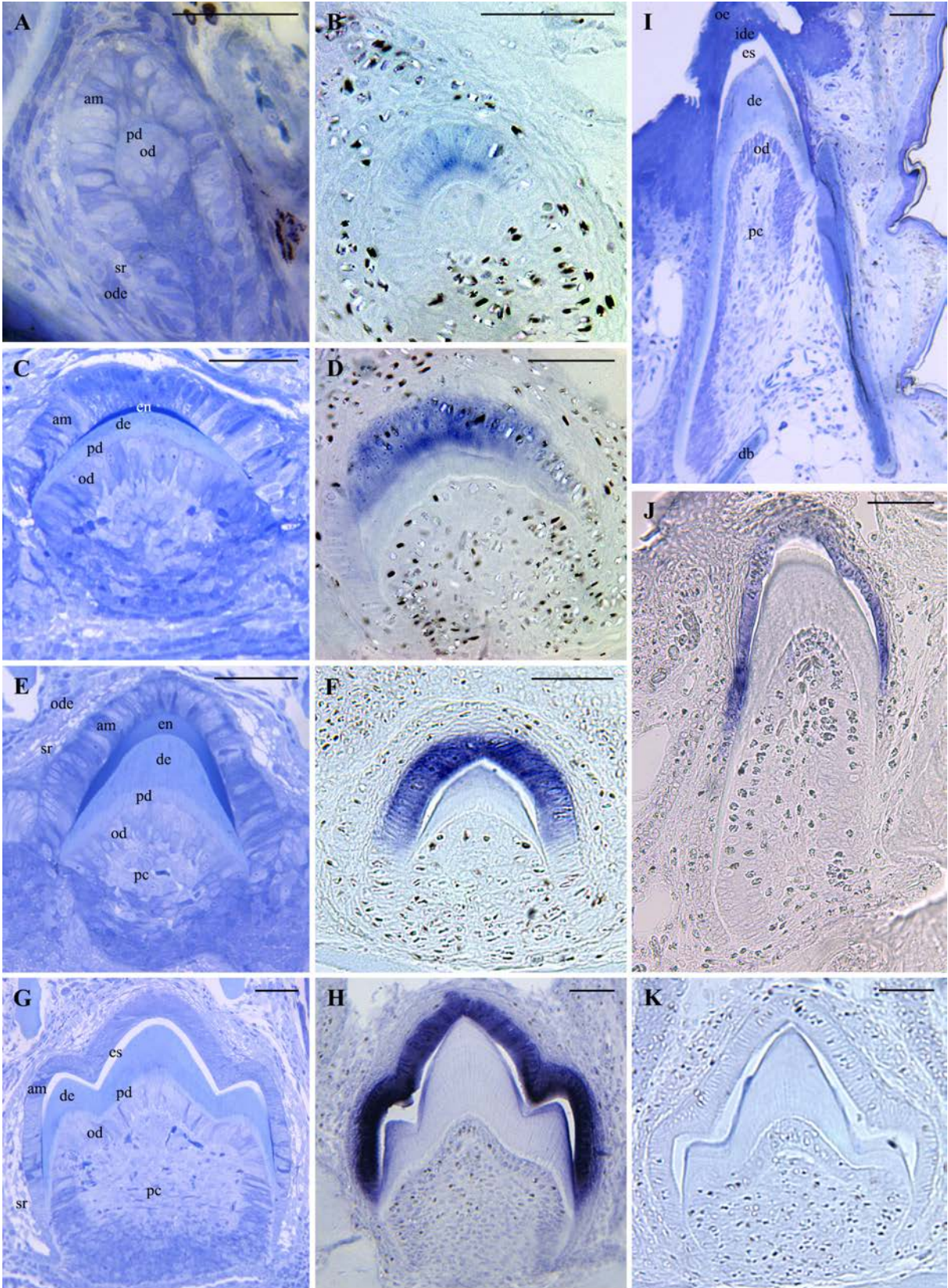


Figure 5. Gasse et al.

Pleurodeles waltl

AMTN expression was monitored in larvae (from stage 34 onwards) and in adult specimens (Figure 6). In larval teeth dentin is covered first with enameloid then with enamel, while in juvenile and adult teeth, dentin is covered with enamel only. The onset of *AMTN* expression is detected earlier in larval than in adult teeth, but only during early deposition of enamel matrix (Figure 6A). In contrast, in adults labelling appears when a thick layer of dentin is already present (Figure 6F, G). During enamel maturation, *AMTN* expression is roughly similar in both larval and adult teeth: transcripts are identified in the ameloblasts around the tooth tip and in those facing the recently deposited enamel matrix along the crown base (Figure 6B, H). Once enamel is mature, *AMTN* expression is no longer detected in the ameloblasts around the upper region of the tooth but is detected in the cells of the inner dental epithelium facing the mid region of the tooth shaft (Fig 6C, I). Before tooth attachment to the dentary bone *AMTN* is still expressed in the inner dental epithelial cells facing the dentin base and the forming pedicel (Figure 6D, J). *AMTN* is no longer detected in the latter region after tooth eruption (not shown).

Figure 6. Amelotin expression during amelogenesis in the salamander *Pleurodeles waltl*.

A-D: larval stages 34, 35, 36 and 38, respectively (Gallien & Durocher, 1957); E-K: three year-old specimens.

Transverse (A-D, E, J) and longitudinal (F-I, K) sections of demineralized jaws. A-D, F-J: *in situ* hybridization; E: histology; K: negative control. A: Enamel deposition on top of enameloid. The ameloblasts surrounding the enameloid cap express *AMTN*. B: Enamel maturation. Ameloblasts are strongly labelled along the crown base where enamel was recently deposited while labelling is weaker at its top where enamel is well mineralized. C: Late maturation stage. *AMTN* is no longer detected in the ameloblasts around the crown, but is identified in the cells at the surface of the dentin shaft, in a region lacking enamel. D: Tooth undergoing attachment process. Along the tooth base *AMTN* is still expressed in the cells of the inner dental epithelium. E: Differentiated tooth showing polarized ameloblasts. F: Early deposition of enamel matrix. *AMTN* transcripts are not detected, while thin layers of enamel and predentin matrix are observed. G: Enamel matrix deposition. *AMTN* mRNAs are clearly identified in the ameloblasts around the tooth top. H: Early enamel maturation. A strong *AMTN* expression is observed in the ameloblasts both facing the mineralized enamel and the recently deposited enamel matrix along the tooth shaft. I: Late enamel maturation. The ameloblasts are no longer labelled at the top of the tooth, where enamel is mature (empty space after demineralization), but are labelled in the cells facing the tooth matrix along the tooth shaft. J: Tooth in process of attachment. *AMTN* expression is visible in the inner dental epithelium cells facing the tooth matrix along the tooth base. am: ameloblasts; de: dentin; es: demineralized enamel space; od: odontoblasts; ode: outer dental epithelium. Scale bars: A-D = 25 μ m; E-K = 50 μ m.

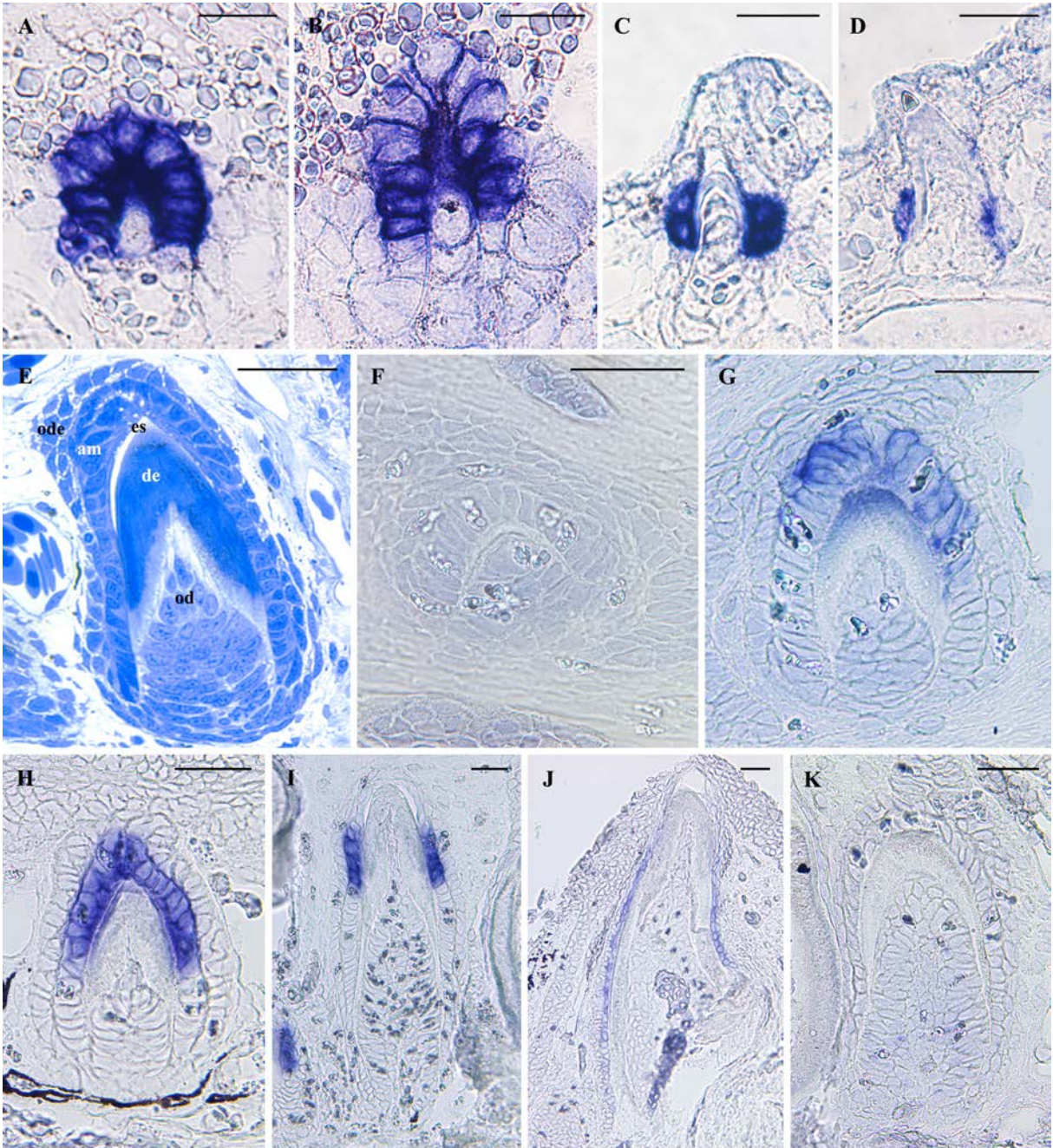


Figure 6. Gasse et al.

DISCUSSION

In this study, we combined RNA-seq data (transcriptomics), genomic data screening, isolation, characterization and expression of the *AMTN* gene to investigate the evolution of this SCPP in tetrapods. Our results suggest that *AMTN* was a component of the enamel matrix in the last common ancestor of tetrapods and that its structure changed and the function most likely changed in mammals, probably as a consequence of drastic events that occurred in the coding and/or regulatory region of the gene in the last common ancestor of mammals.

The evolutionary history of Amelotin reveals important structural modifications in mammals

According to its genomic position, *AMTN* could be related to *ENAM* and *AMBN* because *AMTN* is located close to *AMBN* and *ENAM* in the same DNA region, and always upstream *AMBN* in the genome of sarcopterygian species annotated to date (Sire *et al.*, 2007; Kawasaki & Amemiya, 2014). This vicinity suggests that *AMTN* and *AMBN* could be closely related SCPP genes, an hypothesis which is supported by the presence of remarkable motifs in the two proteins, in particular a SxxE motif encoded by a small exon in the middle region of the protein (exon 7 in *AMTN*, exon 12 in *AMBN*) (Kawasaki & Amemiya, 2014). Furthermore, the expression of *AMTN* and *AMBN* (but not that of *AMEL* or *ENAM*) is modified by the *FAM20C* kinase in mice (Wang *et al.*, 2012b). Nevertheless, further data are needed in current representatives of basal actinopterygian lineages, *e.g.* polypterids and lepisosteids, and of chondrichthyans to better understand the evolutionary history of these EMPs in gnathostomes.

The comparison of *AMTN* structure in sarcopterygian lineages suggests that the ancestral sequence was composed of either nine or 10 exons, depending on whether exon 5 was present in the ancestor and then lost in the coelacanth lineage (or not found in gDNA) or acquired in the lineage leading to tetrapods (Figure 7). To date *AMTN* transcripts, from which the *AMTN* structure could be accurately defined, are available neither in the coelacanth (only genomic sequence), nor in Dipnoi (lungfishes), nor in representatives of the basal actinopterygian lineages (polypterids, lepisosteids, acipenserids and amiids). Furthermore, when only gDNA is available coding exons could be missed, as illustrated by the discovery of the putative exon 2c in monotreme and marsupial *AMTN*, after this exon was identified in sauropsids. Indeed, this exon was

not detected in opossum *AMTN* when using comparison with mammalian sequences that do not possess exon 2c (Gasse *et al.*, 2012) but was later found by Kawasaki and Amemiya (Kawasaki & Amemiya, 2014) when using comparison to sauropsid sequences that include this exon.

Our results suggest that some coding exons present in the coelacanth *AMTN* and in most of the tetrapods analyzed here derive from those of an ancestral SCPP, from which *AMTN* and EMPs originated after duplication. Indeed, *AMTN* exons 2a and 2b are homologous to exons 2 and 3 (exon 4 and 5 in *ENAM*) of all SCPPs, respectively. In addition, *AMTN* shares several other characteristics with all EMPs, *i.e.* the signal peptide encoded by exon 2a, the conserved Ser-X-Glu motif (putative phosphorylation site) encoded by exon 2b, and the presence of a variable region encoded by a large exon, as previously observed by Kawasaki and Weiss (Kawasaki & Weiss, 2003).

The presence of *AMTN* exon 2b in salamander cDNA and its absence in clawed frog, for which gDNA and cDNA sequences are available, indicates that this exon has been lost either early in the anuran lineage (no *AMTN* sequences available in other frog lineages), or later in the lineage leading to Pipidae (Figure 7). Exon 2b also lacks in mammalian *AMTN*, a finding indicating that this exon was lost in an ancestral mammal, after mammalian and sauropsid lineages diverged. This would indicate that the phosphorylation site (SxE motif) encoded by exon 2b, the presence of which characterizes most SCPPs, is not required for the correct *AMTN* function during amelogenesis in clawed frog and mammals. This leads to the question of whether or not the phosphorylation site (SxxE motif) encoded by exon 7 compensates, in mammals, for the lack of that encoded by exon 2b.

Exon 2c was found in non-mammalian sarcopterygian *AMTN* and in non-placental mammals. The presence of this exon in representative species of the four sarcopterygian lineages (Actinistia, Amphibia, Sauropsida and Mammalia) indicates that its origin is to be found before the divergence of these lineages. On the other hand, its absence in placental mammals, suggests that exon 2c was probably lost in the common ancestor of placentals.

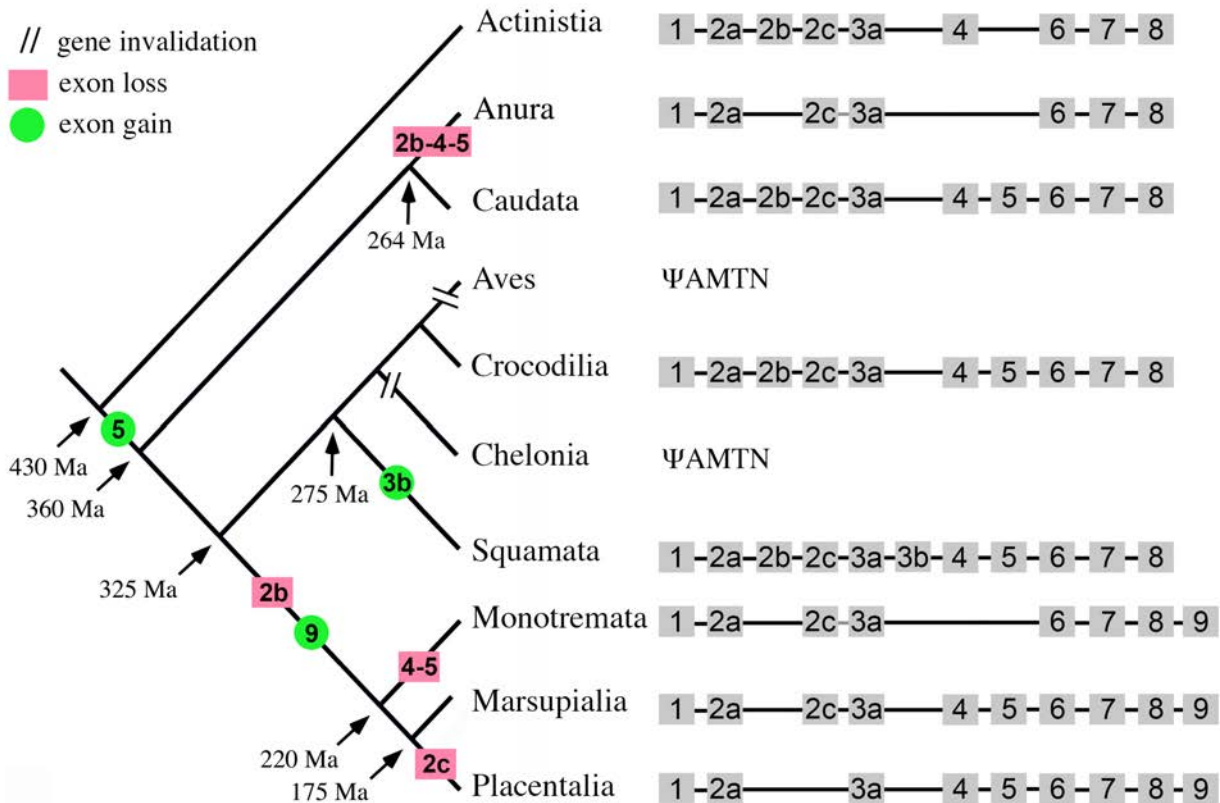


Figure 7. Amelotin evolution in tetrapods.

Exons are shown as grey squares (not at scale). Species names used in this study are given in Additional file 1. *AMTN* was present in the last common ancestor of sarcopterygians and was either composed of 9 or 10 exons depending on the presence of exon 5 before or after the actinistian lineage (coelacanthiformes) separated from the other sarcopterygians. Distinct loss and gain of exons occurred in the various lineages explaining the current gene structure. *AMTN* was invalidated (Ψ *AMTN*) independently in chelonian (turtles) and aves (modern birds) after the capability to develop teeth was lost in both lineages. Two important events have been identified in the mammalian lineage: loss of exon 2b that encodes a phosphorylation site, and recruitment of exon 9 after exon 8 was shortened through intraexonic splicing, leading to the loss of the integrin-binding RGD motif. We can establish a parallel between these structural changes and the differences observed in *AMTN* expression during amelogenesis in mice and non-mammalian tetrapods (see text), suggesting important changes in the protein function. Estimation dates for lineage divergence and phylogenetic relationships are from (Hedges, 2009a; TimeTree, n.d.; Madsen, 2009b; Cannatella *et al.*, 2009; Shedlock & Edwards, 2009).

Exon 3b was found in squamate *AMTN* only, a finding which suggests that this exon was acquired either in the lepidosaurian lineage after its divergence from other sauropsid lineages or soon after Squamata diverged from the sphenodontian lineage (Figure 7).

Exon 4 is absent in the clawed frog *AMTN* sequence (gDNA and cDNA available), a finding which indicates that, as proposed for exon 2b this exon was lost either early in

the anuran lineage or later, in the lineage leading to Pipidae. We did not find this exon in platypus *AMTN*, indicating that exon 4 was secondarily lost in this lineage. However, in absence of *AMTN* cDNA in this lineage, the lack of exon 4 could also result from a genome assembly artefact.

Exon 5 has not been found in coelacanth, clawed frog and platypus genomes, and is not present in clawed frog cDNA either. As discussed above we cannot know whether this exon was present in the sarcopterygian ancestor or was recruited later in a tetrapod ancestor. Parsimonious analysis suggests that exon 5 was secondarily lost in clawed frog (not found in clawed frog cDNA and gDNA sequence, but present in salamander *AMTN* cDNA) and in platypus, although the lack in the latter could result from a badly assembled genomic sequence.

Exon 7 encodes an SxxE motif (putative phosphorylation site) that is found in all *AMTN* sequences analyzed so far. When phosphorylated, this motif promotes hydroxyapatite precipitation *in vitro* (Abbarin *et al.*, 2014). Such motif is also found in two other EMPs, *AMBN* (encoded by exon 12) and *ENAM* (exon 9) (Kawasaki & Amemiya, 2014). These findings suggest that (i) the ancestral sarcopterygian *AMTN* exon 7 possessed this SxxE motif, (ii) the motif is functional (conserved during more than 430 Ma), and (iii) *AMTN* exon 7 is homologous to *AMBN* exon 12 and to *ENAM* exon 8, all of them preceding the large exons 8, 13 and 10, respectively. In clawed frog *AMTN*, the exon 7 not only encodes a SxxE/D motif, but also codes for a second phosphorylation site. Because the SxE motif encoded by exon 2b was lost in the pipid (or anuran) lineage, we suggest that the second SxE motif, encoded by exon 7, could compensate for the loss of the former. The presence of two phosphorylation sites, that could be located in different regions of the protein, was probably essential for *AMTN* to fulfill its function in non-mammalian sarcopterygians.

Exon 8 experienced two different evolutionary histories in sarcopterygians. In non-mammals, exon 8 is large and encodes a variable region of the protein, the C-terminal region of which includes a well-conserved Arg-Gly-Asp (RGD) integrin-binding motif that likely plays a crucial role for the protein function. RGD motifs are recognized by integrins and mediate the attachment of the cell to the extracellular matrix (Ruoslahti, 1996). As for the last exon of all EMPs and of many SCPPs, exon 8 encodes also the termination codon and contains the 3'UTR. In contrast, in all placentalia lineages, *AMTN* exon 8 is shorter than in the other lineages and has lost a large part of the 3' region, including the region encoding the RGD motif in non-mammals (Kawasaki

& Amemiya, 2014), and hence the loss of the adhesive function provided by this motif (Ruoslahti, 1996). In addition, placentalia *AMTN* possesses an exon 9, which encodes the three last amino acids of the protein followed by the stop codon and the 3'UTR. Sequence comparison showed that a mutation in exon 8 generated an intraexonic splice site in an ancestral placentalia *AMTN*. This intraexonic splicing occurred within a codon, leading to the recruitment of exon 9 encoding an appropriate stop codon. It appears that encoding the stop codon could be the only role of exon 9 as several mammalian species lack this last exon because a mutation generated a stop codon at the extremity of exon 8 (Gasse *et al.*, 2012).

In monotremes and marsupials, the question of the efficient presence of the intraexonic splicing is posed. Indeed, the unavailability of cDNA data in both lineages has led to various interpretations. The first possibility we already proposed (Gasse *et al.*, 2012) is that intraexonic splicing occurred early in mammalian evolution, prior to current mammalian lineages differentiated. This finding is supported by (i) platypus and opossum *AMTNs* displaying a similar environment favorable for intraexonic splicing as observed in the homologous region of placental mammals, and (ii) exon 9 sequence, similar to that in placental mammals, present in both gDNA. The second possibility, suggested by Kawasaki and Amemiya (Kawasaki & Amemiya, 2014), is that the intraexonic splicing in *AMTN* exon 8 did not occur in an ancestral marsupial but took place later in an ancestral placental. This finding is supported by the conservation of an RGD motif in the extremity of opossum *AMTN*, as in non-mammalian species. Such a long lasting conservation of RGD motif, although the rest of the sequence is less conserved, could indicate that this region is functional. A third interpretation could be that both *AMTN* variants are present in marsupials, the one with a short exon 8 and exon 9 (no RGD encoded), the other with the large exon 8 only (encoding an RGD). In platypus, the interpretation could be the same. However, if the intraexonic splicing does not occur and a large exon 8 transcribed, the latter does not encode an RGD. This suggests a different evolution of *AMTN* in this lineage. We previously showed in mammals that *AMTN* is enamel specific (Gasse *et al.*, 2012). In platypus, enamel is no longer present in adults, meaning that the functional pressure on enamel proteins is relaxed (Davit-Béal *et al.*, 2009). This could have led to some changes in *AMTN* sequence, including mutations in the sequence encoding the RGD motif.

A RGD motif is present in all members of the acidic SCPPs (dentin and bone SCPPs) and is also found in the ENAM sequence of several mammalian species (Al-

Hashimi *et al.*, 2009). This suggests that the RGD motif of AMTN was probably inherited from an ancestral SCPP. Human and mouse recombinant AMTN do not mediate attachment of any cell types (Somogyi-Ganss *et al.*, 2012a), consistent with the absence of RGD motif in mammalian AMTN.

Taken together our results suggest that in non-mammalian sarcopterygians AMTN may play an important role in ameloblast adhesion to the enamel matrix. This role in ameloblast adhesion was then probably lost, either in the last common mammalian ancestor or later in the last common ancestor of placentals, as a consequence of a mutation generating an intraexonic splice site in exon 8, suppressing the encoded RGD motif. Could the absence of this RGD motif explain the late expression of AMTN, *i.e.* from the maturation phase onwards, previously reported during murine amelogenesis (Iwasaki *et al.*, 2005; Moffatt *et al.*, 2006b; Nishio *et al.*, 2010b)? The comparative study of *AMTN* expression in two selected tetrapods has answered partially this question.

Drastic changes of Amelotin expression during amelogenesis occurred in early mammals

In lizard and salamander, tooth formation is roughly similar to that described in other vertebrates, but unlike mammals, ameloblasts do not exhibit Tomes' processes, rod and interrod enamel structures are absent, and hence enamel is not prismatic (Delgado *et al.*, 2005a, 2006; Davit-Béal *et al.*, 2007b). Therefore important changes (Tomes' processes, prismatic enamel) appeared in the mammalian lineage after it diverged from sauropsids, and before the divergence of marsupials, because the main ameloblast features and enamel structure are observed in current marsupial representatives, *e.g.* opossum (Sasagawa & Ferguson, 1991). These changes could be related to variation in spatial-temporal expression of the ameloblast-secreted proteins/genes during enamel matrix formation in mammals compared to non-mammalian tetrapods. In previous studies, we have shown that *AMEL* expression during amelogenesis was similar in all tetrapods (Delgado *et al.*, 2006; Assaraf-Weill *et al.*, 2013). No data are currently available in the literature for the expression of other ameloblast-secreted genes such as *AMBN*, *ENAM*, or *MMP20* in non-mammalian tetrapods.

Differences in the spatial-temporal expression of *AMTN* appear when comparing lizard or salamander amelogenesis to the one of rodents. In the incisors and molars of the latter, *AMTN* is predominantly expressed in maturation-stage ameloblasts (Iwasaki *et*

al., 2005). In lizard, *AMTN* expression is detected in ameloblasts as soon as enamel matrix deposition has started, and is maintained from secretion to late maturation stage. *AMTN* is no longer expressed in the reduced ameloblasts after tooth eruption. In salamander larvae, *AMTN* expression is detected in early stages of amelogenesis while in adults, the onset of expression is delayed and appears slightly later than in the lizard, during enamel matrix deposition. Then, in both larvae and adult salamanders, *AMTN* expression is maintained in ameloblast until maturation stage. The disappearance of *AMTN* signal in ameloblasts occurs earlier than in lizard but it persists longer in the inner dental epithelium along the dentin cone. This late *AMTN* expression in the cells facing the region lacking enamel along the tooth base could be related to the formation of the amphibian pedicel (Davit-Béal *et al.*, 2006, 2007b). Future studies should investigate this further. In erupted teeth of rats, a late expression of *AMTN* was detected in the internal basal lamina of the junctional epithelium (Moffatt *et al.*, 2006b; Nishio *et al.*, 2010b), although no mRNA expression has been detected in the junctional epithelium of post-eruptive mouse molars (Iwasaki *et al.*, 2005; Sawada *et al.*, 2014). In rodents, neither expression of the *AMTN* gene nor the protein was ever detected beyond the cemento-enamel junction in the epithelial cells along the tooth root. Nevertheless, *AMTN* was found in Hertwig's epithelial root sheath cells entrapped in cementum (Nishio *et al.*, 2010a). In lizard and salamander, *AMTN* expression is detected earlier in secretory stage ameloblasts, and later in the cells of the inner dental epithelium along the tooth root. Immunolocalization of the protein should be performed to better understand the significance of the mRNA expression pattern.

AMTN being expressed during the various steps of amelogenesis in lizard and salamander, we concluded that (i) *AMTN* must be considered an enamel matrix protein (EMP), as are *AMBN*, *ENAM* and *AMEL*, and (ii) this status was the ancestral condition in non-mammalian tetrapods. When comparing lizard and salamander to mammals, one could hypothesize that the differences observed both in the timing of *AMTN* expression and in the gene structure could be responsible for the differences in enamel microstructure (non-prismatic versus prismatic, respectively) in these lineages. The absence of both Tomes's process and of an organized rod/interrod enamel structure in lizard and salamander, on the one hand, and in mice overexpressing *AMTN* under *AMEL* promoter, on the other hand, support this hypothesis (Lacruz *et al.*, 2012).

An evolutionary scenario for Amelotin

The putative evolutionary scenario we propose to explain how the spatiotemporal *AMTN* expression and the gene structure changed in mammals could be visualized as the succession of several events that occurred early in mammalian evolution. One event probably occurred at the promoter level (mutation of the promoter, change in transcription factor, etc.) resulting in late *AMTN* expression during enamel mineralization. Two other, probably independent events resulted in the loss of exon 2b encoding the SxE motif and in the intraexonic splicing of exon 8 eliminating the RGD motif. The occurrence of both events (in any order) may have led to important changes in ameloblast structure (differentiation of Tomes' processes) and in enamel organization (formation of prisms).

Amelotin was tooth specific in early amniotes

In birds, pseudogenization of EMP genes is related to tooth loss (Sire *et al.*, 2008; Al-Hashimi *et al.*, 2010; Meredith *et al.*, 2013). In *A. platyrhynchos*, we found that *AMTN* is also a pseudogene. As observed for EMP genes, *AMTN* underwent pseudogenization after the capability to develop teeth was lost in the common ancestor of modern birds, 80-100 Mya (Davit-Béal *et al.*, 2009; Louchart & Viriot, 2011). Numerous mutations were accumulated (substitutions and indels), producing a φ *AMTN* sequence that is now hardly recognizable from the non-coding gDNA. In turtles, the capability to build teeth was lost in the last common ancestor of chelonians, 250 Mya (Davit-Béal *et al.*, 2009). In *C. mydas*, *AMTN* also underwent pseudogenization, producing a φ *AMTN*, in which a few exons could be identified. These findings indicate that in the ancestors of these two sauropsid species *AMTN* had no other functions than being involved in amelogenesis and is, therefore, an enamel specific protein in toothed sauropsids. In addition, it appears that, after the loss of function, mutations in this gene accumulated more rapidly in birds than in turtles although teeth were lost three folds earlier in turtle than in the bird ancestor (Davit-Béal *et al.*, 2009). This could be related to a slower evolution of chelonians during a long geological period (long generation time) compared to rapid evolution of neornithes (short generation time) (Chiari *et al.*, 2012; Meredith *et al.*, 2013).

In a previous study, we showed similar *AMTN* pseudogenization in two xenarthran mammals, armadillo and sloth, in which teeth are devoid of enamel cover (Gasse *et al.*,

2012). These results therefore suggest that *AMTN* was tooth/enamel specific in the last common amniote ancestor, similarly to what was observed for other EMPs.

CONCLUSIONS

This is the first comparative study of *AMTN* sequence in representatives of various mammalian and non-mammalian vertebrates. *AMTN* was present in the last common sarcopterygian ancestor more than 420 Mya. Various loss or gain of coding exons were identified in coelacanth, amphibians, reptiles and non-placental mammals, indicating different evolutionary pathways in sarcopterygian lineages. *In situ* hybridization of *AMTN* during tooth formation in lizard and salamander revealed that *AMTN* is expressed in ameloblasts during all stages of amelogenesis (with some differences in both species). These findings suggest that *AMTN* is an essential EMP for amelogenesis in these two species. The differences in spatiotemporal expression of *AMTN* described in murine vs lizard and salamander teeth would then suggest that in sauropsids and amphibians *AMTN* could have different functions than in mammals.

List of abbreviations

Aa: amino acids

AMBN : ameloblastin

AMEL: amelogenin

AMTN: amelotin

BLAST: basic local alignment search tool

bp: base pair

CK2: casein kinase 2

EMPs: enamel matrix protein

ENAM : enamelin

Ma: million years

MABT: maleic acid buffer tween

MBB: Montpellier bioinformatics biodiversity

Mya: Million years ago

NBT/BCIP: nitro blue tetrazolium chloride/5-bromo-4-chloro-3-indodyl phosphate

Nt: nucleotides

NTM: NaCl, TrisHCl, MgCl₂

ODAM : odontogenic ameloblast-associated

RGD: arg-gly-asp

SCPP: secretory calcium-binding phosphoprotein

UTRs: untranslated regions

WGS: whole genome shotgun

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

BG and JYS conceived and designed the experiments. BG carried out the molecular genetic experiments. YC and JS performed the jaw RNA extractions for the transcriptomic study. TDB participated in the salamander study. JYS performed in silico data recovery and sequence alignment. BG and JYS wrote the manuscript.

All authors read and approved the final manuscript.

Acknowledgements

We are grateful to Samuel Martin, Director of "La ferme des crocodiles", Pierrelatte, France, for giving us a young *Caiman crocodilus*; to K. Daouès, director of "La ferme tropicale", Paris, France, for giving us a juvenile *Python regius*, a juvenile *Tachydromus sexlineatus*, and several *Anolis carolinensis*. This study was financially supported by grants from CNRS, Université Pierre & Marie Curie, and from the ANR-12-BSV7-020, project "Jaws". We thank Marie-Claire Lajarille for histological preparations.

SUPPLEMENTARAY DATA

Additional file 1. Scientific and common names of the species used in our study, and sources of Amelotin sequences.

Last access to databases was on February 2014. The sequences obtained from sequenced genomes (names in bold) are available at <http://jysire.free.fr/>.

Genus Species	Common name	Family and Order	Sources
Mammalia			
<i>Bos taurus</i>	Cow	Bovidae, Artiodactyla	XM_870725
<i>Homo sapiens</i>	Human	Hominidae, Primates	NM_212557
<i>Loxodonta africana</i>	African elephant	Elephantidae, Proboscidea	XM_003414063
<i>Monodelphis domestica</i>	Opossum	Didelphidae, Didelphimorphia	XM_007495514
<i>Mus musculus</i>	Mouse	Muridae, Rodentia	NM_027793.1
<i>Ornithorhynchus anatinus</i>	Platypus	Ornithorhynchidae, Monotremata	OANA5
Sauropsida			
<i>Alligator mississippiensis</i>	American alligator	Alligatoridae, Crocodylia	allMis0.2
<i>Anolis carolinensis</i>	Green anole	Iguanidae, Squamata	XM_008103694 / KM069435
<i>Caiman crocodylus</i>	Spectacled caiman	Alligatoridae, Crocodylia	KM069444
<i>Chelonia mydas</i>	Green sea turtle	Cheloniidae, Testudines	CheMyd_1.0
<i>Anas platyrhynchos</i>	Mallard duck	Anatidae, Anseriformes	BGI_duck_1.0
<i>Ophiophagus hannah</i>	King cobra	Elapidae, Squamata	SAMN02439592
<i>Python molurus</i>	Indian python	Pythonidae, Squamata	PytMol_1.0
<i>Python regius</i>	Royal python	Pythonidae, Squamata	KM069438
<i>Takydromus sexlineatus</i>	Long-tailed lizard	Lacertidae, Squamata	KM069441
<i>Tarentola mauritanica</i>	Common wall gecko	Gekkonidae, Squamata	KM069446
Amphibia			
<i>Pleurodeles waltl</i>	Sharp-ribbed salamander	Salamandridae, Caudata	KM069436
<i>Xenopus (Silurana) tropicalis</i>	Clawed frog	Pipidae, Anura	KM069437
Coelacanthid			
<i>Latimeria chalumnae</i>	Coelacanth	Coelacanthidae, Coelacanthiformes	LatCha1

Additional file 2. Amelotin sequences.

In bold: from cDNA sequences. ?: unknown residues. *: Stop codon.

MAMMALIA

Primates

>AMTN *Homo sapiens* [NM_212557.2]

MRSTILLFCLLGSTRSLPQLKALGLPPTKLPADQGTLPNQQSNQVFPSSLSLIPLTQMLTLGPDHLHLLNPAAGMTPGTQTHPLTLGG
LNVQQQLHHPVLP IFVTVQLGAQGTILSSEELPQIFTSLI IHS LFPGGILP T SQAGANPDVQDGS LPAGGAGVNPATQGT PAGR LPTPS
GTDDDFAVTTPAGIQRSTHAIIEEATTESANGIQ*

>AMTN *Pan troglodytes*

MRSTILLFCLLGSTRSLPQLKALGLPPTKLPADQGTLPNQQSNQVFPSSLSLIPLTQMLTLGPDHLHLLNPAAGMTPGTQTHPLTLGG
LNVQQQLHHPVLP IFVTVQLGAQGTILSSEELPQIFTSLI IHS LFPGGILP T SQAGANPDVQDGS LPAGGAGVNPATQGT PAGR LPTPS
GTDDDFAVTTPADIQRSTRAIEEATTESANGIQ*

>AMTN *Pan paniscus*

MRSTILLFCLLGSTRSLPQLKALGLPPTKLPADQGTLPNQQSNQVFPSSLSLIPLTQMLTLGPDHLHLLNPAAGMTPGTQTHPLTLGG
LNVQQQLHHPVLP IFVTVQLGAQGTILSSEELPQIFTSLI IHS LFPGGILP T SQAGANPDVQDGS LPAGGAGVNPATQGT PAGR LPTPS
GTDDDFAVTTPADIQRSTRAIEEATTESANGIQ*

>AMTN *Gorilla gorilla*

MRSTILLFCLLGSTRSLPQLKALGLPPTKLPADQGTLPNQQSNQVFPSSLSLIPLTQMLTLGPDHLHLLNPAAGMTPGTQTHPLTLGG
LNVQQQLHHPMLP IFVTVQLGAQGTILSSEELPQIFTSLI IHS LFPGGILP T SQAGANPDVQDGS LPAGGAGVNPATQGT PAGR LPTPS
GTDDDFAVTTPAGIQRSTRAIEETTESANGIQ*

>AMTN *Pongo pygmaeus*

MKTTILLFCLLGSTRSLPQLKALGLPPTKRAPDQGTLPNQQSNQVFPSSLSLIPLTQMLTLGPDHLHLLNPAAGMTPGTQTHPLTLGG
LNVQQQLHHPMLP IFVTVQLGAQGTILSSEELPQIFTSLI IHS LFPGGILP T SQAGANPDVQDGS LPPGGAGVNPATQGT PAGR LPTPS
GTDDDFAVTTPAGIQRSTRAIEETTESANGIQ*

>AMTN *Pongo abelii*

MKTTILLFCLLGSTRSLPQLKALGLPPTKRAPDQGTLPNQQSNQVFPSSLSLIPLTQMLTLGPDHLHLLNPAAGMTPGTQTHPLTLGG
LNVQQQLHHPMLP IFVTVQLGAQGTILSSEELPQIFTSLI IHS LFPGGILP T SQAGANPDVQDGS LPPGGAGVNPATQGT PAGR LPTPS
GTDDDFAVTTPAGIQRNTRAIETTESANGIQ*

>AMTN *Nomascus leucogenys*

MKTTMLLFCLLGSTRSLPQLKALGLPPTKLPADQGTLPNQQSNQVFPSSLSLIPLTQMLTLGPDHLHLLNPAAGMTPGTQTHPLTLGG
LNIQQQLHHPHLP IFVTVQLGAQGTILSSEELPQIFTSLI IHS LFPGGILP T SQAGANPDVQDGS LPAGGAGVNPATQGT PAGR LPTPS
GTDDDFAVTTPAGIQRNTRAIETTESANGIE*

>AMTN *Macaca mulatta*

MKTMILLFCLLGSTQSLPQLKALGLPPTKLPADQGTLPNQQSNQVFPSSLSLIPLTQMLTLGPDHLHLLNPAAGMTPGTQAHPLTLGV
LNLPPQLQPHMLP IFVTVQLGAQGTILSSEELPQIFTSLI IHS LFPGGILP T SQAGANPDVQDGS LPAGQAGVNPATQGT PAGR LPTPS
GTDDDFAVTTPAGIQRSTHATEETTE*

>AMTN *Macaca fascicularis*

MKTMILLFCLLGSTQSLPQLKALGLPPTKLPADQGTLPNQQSNQVFPSSLSLIPLTQMLTLGPDHLHLLNPAAGMTPGTQAHPLTLGV
LNLPPQLQPHMLP IFVTVQLGAQGTILSSEELPQIFTSLI IHS LFPGGILP T SQAGANPDVQDGS LPAGQAGVNPATQGT PAGR LPTPS
GTDDDFAVTTPAGIQRSTHATEETTE*

>AMTN *Nasalis larvatus*

MKTMILLFCLLGSTQSLPQLKALGLPPTKLPADQGTLPNQQSNQVFPSSLSLIPLTQMLTLGPDHLHLLNPAAGMTPGTQAHPLTLGV
LNLQHQLQPHMLP IFVTVQVGAQGTILSSEELPQIFTSLI IHS LFPGGILP T SQAGANPDVQDGS PPAGQAGVNPATQGT PAGR LPTPS
GTDDDFAVTTPAGIQRSTHATEETTESPNGIQ*

>AMTN *Papio anubis*

MKTMILLFCLLGSTQSLPQLKALGLPPTKLPADQGTLPNQQSNQVFPSSLSLIPLTQMLTLGPDHLHLLNPAAGMTPGTQAHPLTLGV
LNLQQQLQPHMLP IFVTVQLGAQGTILSSEELPQIFTSLI IHS LFPGGILP T SQAGANPDVQDGS LPAGQAGVNPATQGT PAGR LPTPS
GTDDDFAVTTPAGIQRSTHATEETTE*

>AMTN *Papio hamadryas*

MKTMILLFCLLGSTQSLPQLKALGLPPTKLPADQGTLPNQQSNQVFPSSLSLIPLTQMLTLGPDHLHLLNPAAGMTPGTQAHPLTLGV
LNLQQQLQPHMLP IFVTVQLGAQGTILSSEELPQIFTSLI IHS LFPGGILP T SQAGANPDVQDGS LPAGQAGVNPATQGT PAGR LPTPS
GTDDDFAVTTPAGIQRSTHATEETTE*

>AMTN *Callithrix jacchus*

MKTTILLFCLLGSTQSLPQLKALGLPPTKLPADQATLLNQQSNQVFPSSLSLIPLTQMLTLGQDLQLVKSCCRNDGWYDPPIDPGR
VECTTATATTCLMPLIFVTVHLGAQGTILSSEELPLAPQIFTG LIIHSLFPGGILP T SQAGANPNVQDGR LPAGQAGVNSAIQGT PAGR LPTPS
PIPSSTDDDFAMTTPAGIQRSTHAMEETTESPNGIQ*

>AMTN *Saimiri boliviensis*

MKTTILLFCLLGSTQSLPQLKALGLPPTKLPADQATLLNQQSNQVFPSSLSLIPLTQVTLGQDLQLLNPAAGMTPGAQTHPLTLGG
LNVQQQLQPHMLP IFVTVHLGAQGTILSSEELPLAPQIFTG LIIHSLFPGGILP T SQAGANPNVQDGR LPAGQAGVNPATQGT PAGR LPTPS
TPSSTDDDFAMTTPAGIQRSTHATEETTESNGIQ*

>AMTN *Microcebus murinus*

MKTTILLFCLLGSTQSLPRHLNPAALGLPPTKLPADQATPLTQQPNQVFPSSLSVIPLTQMLTLGSDLHLLNPAAGMTPGALQAHPLPLG
GLNVQQQLQPVLP IFVTVQLGAQ????????PVAPHFFTGLIIHPLLPGGILPPTQAGANPDGQGTSLPAGQAGVNPATHTPGEHL
PTPTDDEDFGVTPAGLRRGLHTTEETTEPPNE???

>AMTN *Daubentonia madagascariensis*

MKTTILLFCLLGSTQSVKHLNPAALGLPPTKLPADQATLLNQQPNQVFPSSLSLIPLTQMFALGSDLHLLNPAAGMTPGALQAHPLPLG
GLNVQQQLQSMPLIFVTVQLGAQGTILSSEELPVAPQIFTG LIIHPLFPGGILP T SQVGANPPGGQGS LPAGQAGVNPATHTPGEHH
PTPRDDEDFGVTPAGIHRGHTTEETTESPNGIQ*

>AMTN *Tarsius syrichta*

MKTTILLFCLLGSTQSLPQLKALGLPPTKLPADQVTLNQQPNQVFPSSLSLIPLTQMLTLGSDLQLLNPAAGMTPGAQTHPLTLGG
NVQQQLQPMPLIFVTVQLGAQGTILSSEELPVGSPFFTG LIIHPLFPGGILP T SQAGTKPDVQDGS LPAGQAGVNPATHTPAGLPTP
SVTDDDFGVTPASIQRGHTTEETTESPDGIQ*

>AMTN *Otolemur garnettii*

MKTMILLFCLLGPASQSLPKHLSPTLGLQPTKLAPDEATLPTQQPNQVLPSSISLIPLTQMLTLGSDLRLLSPAAGLAPGAQAHSPSLG
GLNAQQLQPQMLPIFVAQLGAQGTILSSEELPVAPQIFFTGLIIHPLFPGGVLPASQARANLDVQGENLPAGTAEHLTPRDDDES
TPPAGIRSGTSTGSDTTTELPNGIQ*

Dermaptera

>AMTN Galeopterus variegatus
MKTAILLFCLLGSTQSLPKQLNPAALGLPPTKVALNQATPLNQQQPNQVFPSSLSLIPLTQMLTLGSDLRLLHPTAGMTPGAQTLPLTLG
ALNVQHHLQPQMLPIFVAQLGAQGTILSSEELPVAPQIFFTGLIIHPLFPGGVLPASQARANLDVQGENLPAGTAEHLTPRDDDES
GTDDDFEVATPAGIQRGLHTTEETTTESPNGIQ*

Euarchontoglires

Scandentia

>AMTN Tupaia chinensis
MKTMILLFCLLGSTQSLPKQHNPASGLPPTKVALNQATPLNQQQPNQVFPSSLSLIPLTQMLTLGSDLRLLHPTAGMTPGAQTLPLTLG
RLNVQHHLQPQMLPIFVAQLGAQGTILSSEELPVAPQIFFTGLIIHPLFPGGVLPASQARANLDVQGENLPAGTAEHLTPRDDDES
PTPSDDELDVATPAGILRGMHTTEETTTESPNGIQ*

>AMTN Tupaia belangeri
MKTMILLFCLLGSTQSLPKQHNPASGLPPTKVALNQATPLNQQQPNQVFPSSLSLIPLTQMLTLGSDLRLLHPTAGMTPGAQTLPLTLG
RLNVQHHLQPQMLPIFVAQLGAQGTILSSEELPVAPQIFFTGLIIHPLFPGGVLPASQARANLDVQGENLPAGTAEHLTPRDDDES
PTPSDDELDVATPAGILRGMHTTEETTTESPNGIQ*

Rodentia

>AMTN Mus musculus [NM_027793]
MKTMILLFCLLGSASQSLPKQLNPAASGLPPTKVALNQATPLNQQQPNQVFPSSLSLIPLTQMLTLGSDLRLLHPTAGMTPGAQTLPLTLG
LNGQQQLQPQMLPIFVAQLGAQGTILSSEELPVAPQIFFTGLIIHPLFPGGVLPASQARANLDVQGENLPAGTAEHLTPRDDDES
TPGVTDDDDYEMSTPAGLRRTHTTEGTTIDPPNRTQ*

>AMTN Rattus norvegicus [NM_001044296]
MKTVVLLFCLLGSASQSLPKQLNPAALGLPPTKVALNQATPLNQQQPNQVFPSSLSLIPLTQMLTLGSDLRLLHPTAGMTPGAQTLPLTLG
LNGQQQLQPQMLPIFVAQLGAQGTILSSEELPVAPQIFFTGLIIHPLFPGGVLPASQARANLDVQGENLPAGTAEHLTPRDDDES
PAVTDDDDYEMSTPAGLRATHHTTEGTTMDDPPNRTK*

>AMTN Cricetulus griseus
MKTILLFCLLGSTQSLPKQISPTLGLPPTKVALNQATPLNQQQPNQVFPSSLSLIPLTQMLTLGSDLRLLHPTAGMTPGAQTLPLTLG
LNRQQQLQPQMLPIFVAQLGAQGTILSSEELPVAPQIFFTGLIIHPLFPGGVLPASQARANLDVQGENLPAGTAEHLTPRDDDES
TPGVTDDDDYETSTPAGIRRGHTHTTEGTTIDTPNRTQ*

>AMTN Mesocricetus auratus
MKTILLFCLLGSASQSLPKQISPTLGLPPTKVALNQATPLNQQQPNQVFPSSLSLIPLTQMLTLGSDLRLLHPTAGMTPGAQTLPLTLG
LNRQQQLQPQMLPIFVAQLGAQGTILSSEELPVAPQIFFTGLIIHPLFPGGVLPASQARANLDVQGENLPAGTAEHLTPRDDDES
TPGVTDDDDYEMSTPAGIRRGHTHTTEGTTIDTPNRTQ*

>AMTN Peromyscus maniculatus
MKTILLFCLLGSASQSLPKQIRPTLGLPPTKVALNQATPLNQQQPNQVFPSSLSLIPLTQMLTLGSDLRLLHPTAGMTPGAQTLPLTLG
LNRQQQLQPQMLPIFVAQLGAQGTILSSEELPVAPQIFFTGLIIHPLFPGGVLPASQARANLDVQGENLPAGTAEHLTPRDDDES
TTPGVTDDDDYETSTPAGMRRGHTHTTEGTTIDTPNRTQ*

>AMTN Microtus ochrogaster
MKTILLFCLLGSASQSLPKQINPALGLPPTKVALNQATPLNQQQPNQVFPSSLSLIPLTQMLTLGSDLRLLHPTAGMTPGAQTLPLTLG
LSRSQQQLQPQMLPIFVAQLGAQGTILSSEELPVAPQIFFTGLIIHPLFPGGVLPASQARANLDVQGENLPAGTAEHLTPRDDDES
TPAVTDDDDYEMSTAAGMRRGHTHTTEGTTIDTPNRTKDLDD*

>AMTN Jaculus jaculus
MRTILLFCLLGSASQSLPKQLNPAALGLPPTKVALNQATPLNQQQPNQVFPSSLSLIPLTQMLTLGSDLRLLHPTAGMTPGAQTLPLTLG
ALNGQQQLQPQMLPIFVAQLGAQGTILSSEELPVAPQIFFTGLIIHPLFPGGVLPASQARANLDVQGENLPAGTAEHLTPRDDDES
TDDDFAMTTPAGLRKGVHTTTEGTTPESTSGIH*

>AMTN Dipodomys ordii
MKTAILLFCLLGSTQSLPKQPNPTLGYPPPTKPTPEQVTLNQQQSHQVFPSSLSLIPLTHLFTLGSDLQLFNHATGIAPGHTLPLVTLG
ALNGQQQLPSQMLPIFVAQLGAQGTILSSEELPVAPQIFFTGLIIHPLFPGGVLPASQARANLDVQGENLPAGTAEHLTPRDDDES
STPSVTDDDDFGGTPAGIRRMHTTVEETTAD*

>AMTN Cavia porcellus
MKTVILLFCLLGSASQSLPKQLNSALEVSPPTKVALNQATPLNQQQPNQVFPSSLSLIPLTQMLTLGSDLRLLHPTAGMTPGAQTLPLTLG
ALQGNQQLQPQMLPIFVAQLGAPAAALSSEELPVAPQIFFTGLIIHPLFPGGVLPASQARANLDVQGENLPAGTAEHLTPRDDDES
STPGVTDDDDFEVTTTLGIQKVTHTPEKTTTGSPPN*

>AMTN Octodon degus
MKTAILLFCLLGSASQSLPKQLNSALEVSPPTKVALNQATPLNQQQPNQVFPSSLSLIPLTQMLTLGSDLRLLHPTAGMTPGAQTLPLTLG
TVNGQNQLQPQMLPIFVAQLGAQGTILSSEELPVAPQIFFTGLIIHPLFPGGVLPASQARANLDVQGENLPAGTAEHLTPRDDDES
STSGVTDDDDFEVTTTAPGIQRATHHTTKQITGSPNRIH*

>AMTN Chinchilla laniger
aMKTAILLFCLLGSTQALPKQLNSALEVSPPTKVALNQATPLNQQQPNQVFPSSLSLIPLTHLFTLGSDLQLFNHATGIAPGHTLPLVTLG
GALNGQNQLQPQMLPIFVAQFGAQAALSSEELPVAPQIFFTGLIIHPLFPGGVLPASQARANLDVQGENLPAGTAEHLTPRDDDES
LSTPGVTDDDDFEVTTTAPGIQKATHHTTKGITTGSPNRIH*

>AMTN Heterocephalus glaber
MKTMILLFCLLGTQSLPKQLNSALEVSPPTKVALNQATPLNQQQPNQVFPSSLSLIPLTQMLTLGSDLRLLHPTAGMTPGAQTLPLTLG
ALNGQNQQQPQMLPIFVAQIGAQTILSSEELPVAPQIFFTGLIIHPLFPGGVLPASQARANLDVQGENLPAGTAEHLTPRDDDES
LTPGVTDDDDFEATPAGIQRATHHTTKGITTGSPNGTQ*

>AMTN Fukomys damarensis
MKTILLFCLLGSASQSLPKQLNSALEVSPPTKVALNQATPLNQQQPNQVFPSSLSLIPLTQMLTLGSDLRLLHPTAGMTPGAQTLPLTLG
ALNGQNQLQVLPFVTVQGAQATMLSSSEELPVAPQIFFTGLIIHPLFPGGVLPASQARANLDVQGENLPAGTAEHLTPRDDDES
TPGVTDDDDFEATPAGIQRATHHTTKGITTGSPNGIH*

>AMTN Nannospalax galili
MKTILLFCLLGSASQSLPKQLNSALEVSPPTKVALNQATPLNQQQPNQVFPSSLSLIPLTQMLTLGSDLRLLHPTAGMTPGAQTLPLTLG
SLNGQQQLQPQMLPIFVARLGAQGTILSSEELPVAPQIFFTGLIIHPLFPGGVLPASQARANLDVQGENLPAGTAEHLTPRDDDES
PTPGVTDDDDYEMSTPAGMRRGHTHTTEGTTIDTPNRTQ*

>AMTN Ictidomys (Spermophilus) tridecemlineatus
MKTILLFCLLGSASQSLPKQFNGLAPPTKVALNQATPLNQQQPNQVFPSSLSLIPLTQMLTLGSDLRLLHPTAGMTPGAQTLPLTLG
NGQQQLQPQMLPIFVAQLGAQGTILSSEELPVAPQIFFTGLIIHPLFPGGVLPASQARANLDVQGENLPAGTAEHLTPRDDDES
PTGIDDDDEFGETTAPGIQRGTQTDETTTESPNIH*

Lagomorpha

>AMTN *Oryctolagus cuniculus*

MKTAAILLFCLLGSLTQSLPKQLNPALELPPTKQAPDQETPLNQOQPTQVFPSSLISLIPLTQMLTLLNPAAGMAPGAQTLPLTLEKLDGQQ
QLQPQILPIIVAQLGTQGTILSSEEVPAAPHIFTGLLHPLFTSQTAANADSQDEVLPQTGQAGVNPAPAIRATPADDIGVTTTAPAGIQR
TPVTEETTSESPKGIQ*

>AMTN *Ochotona princeps*

MKTAAILLFCLLGATQSLPKQLNPALELGPVLPPTKQDRDQETPLKQHQPSQVFPSSLISLIPLTQMLTLLNPAAGAAATGAQTLPLTM
EILEAQKQLQTMPLPIIVAQLGAQAVLSSEELPAAPQIILTGLLHPMTFGKAGAHDDVQDEVLPQGAEVSPATRGAQEGGFSTPVN
DALGVTSVPGIQRGTPPTTEETTSPKGIH*

Laurasiatheria

Carnivora

>AMTN *Felis catus*

MKTMILLLYLLGSLTQSLPTQLNPALELGLPTKALDQATLLNQOQLNQVFPSSLISLIPLTQMITLQADLQLLNPPAGLAPGTQTLPLTLGG
LNTQQPLQAQMLPVIIVAHGAQGTILSSEELPMAPQIFFTGLIFQPLFPSTLPNSQANPDAQNGILPAGQAGMNPATQGTSEGFSTP
SDTDDDFEVTAPTGIIRGMHTTQETTTGPPNGNQ*

>AMTN *Panthera tigris*

MKTMILLLYLLGSLTQSLPTQLNPALELGLPTKALDQATLLNQOQLNQVFPSSLISLIPLTQMITLQADLQLLNPPAGLAPGTQTLPLTLGG
LNTQQPLQAQMLPVIIVAHGAQGTILSSEELPMAPQIFFTGLIFQPLFPSTLPNSQANPDAQNGILPAGQAE??AIQGTSEGFSTP
SDTDDDFEVTAPAGIIRGMHTTQETTTGLPNGNQ*

>AMTN *Canis familiaris*

MKTMILLLYLLGSLTQSLPAQLNPALELGLPTKLPHTQATLLNQOQPNQVFPSSLISLIPLTQMTFLASDLQLLNPAAGMASGTQTLPLSLG
VLNTQQQLQPQMLPVIIVAHGAQGTILSSEELPMAPQIFFTGLIFQPLFPSTLPNSQANPDAQNGILPAGQAGGNPAIQGTPEFSFT
PSDIDDDFVGTAPAGIQRGTHHTTQETTTGPPNGTQ*

>AMTN *Mustela putorius* [JP005519]

MKTMILLLYLLGSLTQSLPTQLNPALELGLPTKLPHTQATLLNQOQPNQVFPSSLISLIPLTQMTFLASDLQLLNPAAGMPPGPQTLPLSLG
VLNTQQQLQPQMLPVIIVAHGAQGTILSSEELPMAPQIFFTGLIFQPLFPSTLPNSQANPDAQNGILPAGQAGVNPATQGTAEFSFT
PSDIDDDFVGTAPAGIQRGTHHTTQETTTGPPNGTQ*

>AMTN *Ailuropoda melanoleuca*

MKTMILLLYLLGSTRSLPTQLNPALELGLPTKLPHTQAKLLNQOQPNQVFPSSLISLIPLTQMTFLASDLQLLNPAAGMAPGIQTLPLSLG
SLNAQQQLQPQMLPVIIVAHGAQGTILSSEELPMAPQIFFTGLIFQPLFPSTLPNSQANPDAQNGILPAGQAGANPAIQGTPEGSFT
PSDIDDDFVGTAPAGIQRGTHHTTQETTTGPPNGTQ*

>AMTN *Ursus maritimus*

MKTLILLLYLLGSTRSLPTQLNPALELGLPTKLPHTQAKLLNQOQPNQVFPSSLISLIPLTQMTFLASDLQLLNPAAGMAPGTQTLPLSLG
GLNAQQQLQPQMLPVIIVAHGAQGTILSSEELPMAPQIFFTGLIFQPLFPSTLPNSQANPDAQNGILPAGQAGASPAIQGTPEGSFT
PSDIDDDFVGTAPAGIQRGTHHTTQETTTGPPNGTQ*

>AMTN *Odobenus rosmarus*

MKTMILLLYLLGSLTQSLPTQFNALGLPTKLPHTQAKLLNQOQPNQVFPSSLISLIPLTQMTFLASDLQLLNPAAGMAPGTQTLPLNLG
GLNTQQQLQPQMLPVIIVAHGAQGTILSSEELPMAPQIFFTGLIFQPLFPSTLPNSQANPDAQNGILPAGQAGVNPATQGTPEGSFT
PSNTDDDFVGTAPAGIQRGTHHTTQETTTGPPNGTQ*

>AMTN *Leptonychotes weddellii*

MKAMILLLYLLGSLTQSLPTQLNPALELGLPTKLPHTQAKLLNQOQPNQVFPSSLISLIPLTQMTFLASDLQLLNPAAGMAPGTQTLPLNLG
GLNTQQQLQPQMLPVIIVAHGAQGTILSSEELPMAPQIFFTGLIFQPLFPSTLPNSQANPDAQNGILPAGQAGVNPATQGTPEGSFT
PSDIDDDFVGTAPAGIQRGTHHTTQETTTGPPNGTQ*

Cetartiodactyla

>AMTN *Bos taurus*

MKAAILLFCLLGSLTSLPMQLNPALELGLPTKLPVDPQATLLNQOQPNQVFPSSLISLIPLTHMLTLGSLNQLLNPAAGMPSGTQTLPLALG
GLKVQQQLQPQMLPVIIVAHGAQGTILSSEELQGTSQLTGLIFHPLFPSTLPNSQANPDAQNGILPAGQAGANPAAQGTPEDPFST
PSGTDDDFASTTPAGIQRGRPTTEETPTGSPKGIQ*

>AMTN *Bos grunniens*

MKAAILLFCLLGSLTSLPMQLNPALELGLPTKLPVDPQATLLNQOQPNQVFPSSLISLIPLTHMLTLGSLNQLLNPAAGMPSGTQTLPLALG
GLKVQQQLQPQMLPVIIVAHGAQGTILSSEELQGTSQLTGLIFHPLFPSTLPNSQANPDAQNGILPAGQAGANPAAQGTPEDPFST
PSGTDDDFASTTPAGIQRGRPTTEETPTGSPKGIQ*

>AMTN *Bos indicus*

MKAAILLFCLLGSLTSLPMQLNPALELGLPTKLPVDPQATLLNQOQPNQVFPSSLISLIPLTHMLTLGSLNQLLNPAAGMPSGTQTLPLALG
GLKVQQQLQPQMLPVIIVAHGAQGTILSSEELQGTSQLTGLIFHPLFPSTLPNSQANPDAQNGILPAGQAGANPAAQGTPEDPFST
PSGTDDDFASTTPAGIQRGRPTTEETPTGSPKGIQ*

>AMTN *Pantholops hodgsonii*

MKAAILLFCLLGSLTSLPMQLNPALELGLPTKLPVDPQATLLNQOQPNQVFPSSLISLIPLTHMLTLGSLNQLLNPAAGMPSGTQTLPLALG
GLKVQQQLQPQMLPVIIVAHGAQGTILSSEELQGTSQLTGLIFHPLFPSTLPNSQANPDAQNGILPAGQAGANPATQGTSEDPFST
PSGTDDDFAAATTPAGIQRGRPTTEETPTGSPKGIQ*

>AMTN *Capra hircus*

MKAAILLFCLLGSLTSLPMQLNPALELGLPTKLPVDPQATLLNQOQPNQVFPSSLISLIPLTHMLTLGSLNQLLNPAAGMPSGTQTLPLALG
GLKVQQQLQPQMLPVIIVAHGAQGTILSSEELQGTSQLTGLIFHPLFPSTLPNSQANPDAQNGILPAGQAGANPATQGTSEDPFST
PSGTDDDFAAATTPAGIQRGRPTTEETPTGSPKGIQ*

>AMTN *Ovis aries*

MKAAILLFCLLGSLTSLPMQLNPALELGLPTKLPVDPQATLLNQOQPNQVFPSSLISLIPLTHMLTLGSLNQLLNPAAGMPSGTQTLPLALG
GLKVQQQLQPQMLPVIIVAHGAQGTILSSEELQGTSQLTGLIFHPLFPSTLPNSQANPDAQNGILPAGQAGANPATQGTSEDPFST
EDPFSTPSGTDDDFAAATTPAGIQRGRPTTEETPTGSPKGIQ*

>AMTN *Sus scrofa*

MKTTILLFCLLGSLTSLPMQFNPLVGLPPTKLPVDPQATLLNQOQPNQVFPSSLISLIPLTQMLTLGSLNQLLNPAAGMPSQTPLTLG
ALNVQQQLQPQMLPVIIVAHGAQGTILSSEELPATRQILTGLIFHPLFPSTLPNSQANPDAQNGILPAGQAGANPAVQGTFRGPFPT
SSGTDDDFDVTTAPAGLQRGTHATEETTTGSPNGMQ*

>AMTN *Vicugna pacos*

MKTMILLFCLLGSLTSLPMQVNPVTLGLPPTKLPVDPQATLLNQOQPNQVFPSSLISLIPLTQMLTLGSLNQLLNPAAGMPSQTPLTLGGL
NVKQQLRPQILPVIIVAHGAQGTILSSEELPAAPQIFFTGLIFHPLFPSTLPNSQANPDAQNGILPAGQAGVNSANQGTFRGPLPTSS
GTDDDFVTTAPAGIQRGTHATEETTTGSPNGMQ*

>AMTN *Camelus ferus*

MKTTILLFCLLGSSTLSLPMQLNPTLGLPPAKLVDPQATPLNQQQPNQVFPSSLSLIPLTQMLTLGSDLQLLNPGVAPGSQTLPLTLGGL
 NMKQQQLQPQILPVIVAHLVGQGTILSSEELPAAPQIFFTGLIFHPLFPGAIQPTSQANPDVQNGILPAGQAGVNSANQGTWPGLPPTPS
 GTDDDFVVTTPAGIQRGTHTEETTTSPNGIQ*

>AMTN *Orcinus orca*
 MKATILLFCLLGSSTLSLPMQLNPLVVPPTKLVDPQATLLNQQQPNQVFPSSLSLIPLTQMLTLGSDLQLLNPAAGMAPGTQTLPLTLG
 GLKVQQQLQPQVLPVIVAHLGAQGAILLSSEELPATPQILTFGIFHPLFPGAILPTSQATPDVQNGILPAGQAGANPAIRGNSEGGFPPT
 TSGTDDDDFAETTPVGIQRGIHTTEETTTGSPKGI*

>AMTN *Tursiops truncatus*
 MKATILLFCLLGSSTLSLPMQLNPLVVPPTKLVDPQATLLNQQQPNQVFPSSLSLIPLTQMLTLGSDLQLLNPAAGMAPGTQTLPLTLG
 GLKVQQQLQPQVLPVIVAHLGAQGAILLSSEELPATPQILTFGIFHPLFPGAILPTCPPAHLPTCPDQVQNGILPAGQAGANPAIRGNS
 EGGFPPTTSGTDDDDFAETTPVGIQRGIHTTEETTTGSPKGI*

>AMTN *Physeter catodon*
 MKTMILLFCLLGSSTLSLPMQLNPLVLPPTKLVDPQATRLNQQQPNQVFPSSLSLIPLTQMLTLGSDLQLLNPAAGMAPGSQTLPLTLG
 GLKVQQQLQPQVLPVIVAHLGAQGTILSSEELPATPQILTFGIFHPLFPGAILPTSQATPDVQNGILPAGQAGANPAIQGTPEGGFPPT
 TSGRDDNFVVTTPAGLQRGTHTEETTTGSPKGI*

>AMTN *Balaenoptera acutorostrata*
 MKTTILLFCLLGSSTLSLPMQLNPLVLPPTKLVDPQATLLNQQQPNQVFPSSLSLIPLTQMLTLGSDLLNPVAGMAPGTQTLPLTLGGL
 KVQQQLQPQVLPVIVAHLGAQGAILSSDELPAATPQILTFGIFLPLFPGAILPTSQATPDVQNGILPAGQAGANPAIQGTPEGGFPPTT
 GTDDDDFAETTPAGIQRGTHTEETTTGSPKGI*

Eulipotyphla

>AMTN *Erinaceus europaeus*
 MKTTFLLFCLLGSSTHTLPMQLNPLGPPSKLVDPQATLLSQPNQVFPSSLNLIPLTQMLTLGSDLQLLNPTAGMTPGPQTLPLTLG
 GLNVQQQLQPQMLPIIVAQLGAQGTILSSEELPVASQILGLLFPQLLPGSILSTSKANPDQNGILPAGQAGANPAIQGTPEGGFPPT
 SGIDDDDFMTTPAGIQRGLQPTTEETTTGSPNGLQ*

>AMTN *Sorex araneus*
 MKTTILLFCLLGSSTQSLPMQLNHALGLPPLINLVDPQATLLNQQQPNQVFPSSLSLIPLTQMLTLGSDLQRLNAPAPGAQTLPLSLGAVNG
 QQQLQPQMLPVIVAQLGAQGTILSSEELPVASPLLTGLVFPQLLQGAILPNSQINPDLHNRILPAGQAKPNPALHGTPEGLPPTPSGI
 DDDFVVTTPAGIPRGLHLPEETTTESPEGFQ*

>AMTN *Condylura cristata*
 MKTTILLFCLLGSSTQSLPMQLNPLVLPPTKLVDPQATLLNQQQPTQVFPSSLSLIPLTQMLTLGSDLHLLHPTAGITPGTQTLPLTLG
 GLNAQQQLQPQMLPVIVAQLGAQGTILSSEELPVSOHIFAGLIFQPLFPGAILPTSQANPDQNGILPAGQVGINPAIQGTPEGGFPPT
 PSAIDDDYGVTTTPAGIQRGLHTTEETTTGLPDGIQ*

Chiroptera

>AMTN *Eptesicus fuscus*
 MKTVILLFCLLGSSTQSLPMQFKPALGLPPTQLVDPQATRLNQQQPNQVFPSSLSLIPLTQMLTMGSDLQLLNPPMAPAAHTLPLTL
 GGLNVQQQLQPQMLPLIVAQLGAQGTILSSEELPVAPQIFFTGLIFHPLFPGAILPTSHANPDVQNTILPAGQAGVNPVAVQGAPEGGPSP
 TPGTDDDDFEVTTTPAGIQRGMHIAEETTTGFPNGIQ*

>AMTN *Myotis davidii*
 MKTVILLFCLLGSSTQSLPMQKPALGLPPTQLVDPQATLLNQQQPNQVFPSSLSLIPLTQMLTMVSDLQLLNPPMAPAAHTLPLAL
 GGLNVQQQLQHQVLPVIVAQLGAQGTILSSEELPVAPQIFFTGLIFHPLLSGAILPTSQANPDQNTILPAGQAGVNPVAVQGAPEGGPSP
 TPSVTDDDDFEVTTTPAGIQRGMHIAEETTTGFHHGIQ*

>AMTN *Myotis lucifugus*
 MKTVILLFCLLGSSTQSLPMQKPALGLPPTQLVDPQATLLNQQQPNQVFPSSLSLIPLTQMLTMVSDLQLLNPPMAPAAHTLPLAL
 GGLNVQQQLQPQMLPLIVAQLGAQGTILSSEELPVAPQIFFTGLIFHPLFPGAILPTSHANPDQNTILPAGQAGVNPVAVQGAPEGGPSP
 TPGTDDDDFEVTTTPAGIQRGMHIAEETTTGFHHGIQ*

>AMTN *Pteropus alecto*
 MKTTILLFCLLGSSTQALPIQLNPLALRPLPPTKLVDPQATLLNQQQPNQVFPSSLGVIPLTQMLTLGSDLQLLNPAAGMAPAAQTLPLTLV
 ALNVQQQLQPQMLPVIVAQLGAQGTILSSEELPVAPQIFFTGLIFHPLFPGAILPTSQANPDVQNGILPAGQAGINPPIRRAPGSPFPPT
 PSGAGNDFGVTTTPAGIQRGMHTTEETTTGSPSGIR*

>AMTN *Pteropus vampyrus*
 MKTTILLFCLLGSSTQALPIQLNPLALRPLPPTKLVDPQATLLNQQQPNQVFPSSLGVIPLTQMLTLGSDLQLLNPAAGMAPAAQTLPLTLG
 ALNVQQQLQPQMLPVIVAQLGAQGTILSSEELPVAPQIFFTGLIFHPLFPGAILPTSQANPDVQNGILPAGQAGINPPIRRAPGSPFPPT
 PSGAGNDFGVTTTPAGIQRGMHTTGETTTGSPSGIR*

Perissodactyla

>AMTN *Equus caballus*
 MKTVILLFCLLGSSTLSLPMQFPALGLPPTKVPDPQATQLNQQQPNQVFPSSLNLIPLTQMLKLGSDLQLLNPAAGVAPGAQTLPLTLGG
 LKAQQQLQSQMLPIFVAQFQAQGTILSSEELPAAPQIFAGLIFQPLFPGGILPTSQATPDVQNGILPAGQGGVKPAIQGTSESPPLPTT
 SDTDDDDFGTTPAGIQRGMRTTEETITKSPNGIQ*

>AMTN *Ceratotherium simum*
 MKTAILLFCLLGSSTLSLPMQLSPLGLPPTKLVDPQATLLNQQQPNQVFPSSLMPLTQMLTLGSDRQQLNPAAGVAPGAQTLPLTLG
 GLKAQQQLQSQMLPVFVAQIGAQGTILSSEELPVARQIFFTSLIFQPLFPGAILPTSQANPDVQNGILPAGQGGVKPAIQGTSESPFPPT
 PSDPDDDDLGGTTPAGIQRGTHTTKETTTRSPNGIQ*

Xenarthra

>AMTN *Dasyopus novemcinctus* [JX523656] (pseudogene)
 ??????????????????KQFNSALALPSAKLVDPQATLLNQQQPNQVFPPLTLIPLTQMLTLGSDLQLLNPSAEITSSQTIALTLGD
 LYKQQQLQPQILPIILAQLGAQGTILSSEELPVAPQIFFTGLIFHPLFPGAILPTTQPEANPDVQCGVLPAGQAGINPSIQGTPEGYFP
 IPRAERMALE*PFLAASKGTHHTPLRKPP*SYQVEFIK

>AMTN *Choloepus hoffmanni* [JX523657] (pseudogene)
 ??????????????????LFSLSLPLTQMLAVESDLQLVNLSARMVTGVQTLPLTQGT
 *TYSSSSCHKCYQLL*HNLESRVQS*QTRNCQRLHKSSQPSLSSSQEPSCPPVRQELIQMCSMESFLPKDQE*ILPSWEHQKTTSONP
 VAQKMDLE*PPLQISKGADTPPRKPPQSHQMGLKLFK

Afrotheria

>AMTN *Loxodonta africana*
 MKTTILLFCLLGSSTQSLPQLNPLGLSAAKLVDPQATLLNQQQPNQVFPSSLSLIPLTQMLTLGSDLQLLNPATGMAPGTQTFPLNLGG
 LTIKQQQLQSQMLPIIVAQLGAQGAILLSSEELPMAPQIFAGLLIQPLFPGAILPTSLAGATPEVQEGILPAGQAGLNPAIQRTPEKHPS
 TSSDTSVFGVTTTPAGLQRMRTTGETTTESPNIEQ*

>AMTN *Procavia capensis*

MKTMILLFCLLGSTKSFQPNPALGFPAPKVPDQATPLNQQQVSVLPSLSVIPLTQMFTLGSQPQLINPAAGTAAGAQLPVLNLGQ
LTLQQLQ????????????GAILSSSEELPTTPQIFTGILLIQLPLFPGAILPTSQAGVTPPEVPEGALPAGQAGVNPAPIQGTPEGRPL
APSAMDDMSGVTPSAGLQRSIYTTEETTTDSDPGIQ*

>AMTN *Trichetus manatus*
MKTMILLFCLLGSTQSLPQLNLTALGPLPAKLVPDQATLLNQQQPDQVFPSSLISLIPLTQLLTPGSDLQLLNTATGMAPGAQTLPLNLG
GLAVQQQLQSLPLIIIAQLGAQGVILSSEELAVAPQIFAGLLIQLPLFPGAILPTSQAGATPEVQEGILPAGQAGVNPAPIQRTPEGHL
PTPSGTDMMYGVTPSGLQRMHTTEETTTDSDPGIQ*

>AMTN *Echinops telfairi*
MKTIILLFCLLGSTQSLPPLNPALGLPVPKLVPEQSTLLNPQPNQVFPSSLISLIPLTQMLALGSDLQMLNTATGVFGPTQTLPLNLGG
VTVQQQLQPLPFIWAQLGTQGAILSSSEELPVAPQVYTGILLIHPWFPGAFLLPTSQAGSTPEGQDQALPAGQAAAHPAIQGTPEGHL
TPSGTDYLSGVTPAPAGLQRMHTTEETTTDSDPGIQ*

>AMTN *Chrysochloris asiatica*
MKTMILLFCLLRSTQSLPQPNPALGGLPAKLVDQATQLNHQQPNQVFPSSLISLIPLTQMFTLGSDDLQLLNPASGMAPGAQTHPLNLG
GLTVQQQLHPQMLPVIWAQLGAQGVILSSEELPVAPQIFAGLLIHPFPVGMPLPTSQAGPTPEIHEGILPAGQAGVNSAIQGTLEGHL
PTPSGTDYMMGVTPAPAGLQRTGYTTEETTTESPNGIQ*

>AMTN *Elephantulus edwardii*
MKTIILLFLLSTQSLPQLNLPALGLPQAKLVDPQTTLLNQQQPNQVFTSLGLIPLTQLLTLGSDLHLLNPSTGLAAGAQTLPPLNLG
ALTQQQLQSQMLPIIWAQLGAQGVILSSEELPVAPQIFAGLLIQLPLYPGAILPTSQAGATPELQEGTLTLAGQAGVNSAIQRTPEHDL
PTPTGTDVDFGTTTASLQSRMYTTEETTTESPNGI*

>AMTN *Orycteropus afer* (pseudogene)
??
????????????MLPIIWAQLGDQGTIINSEELPVAPQIFAGLLIQLPLFQEPCLPIRQGLLQRSRTESFPQNKQE*IMPSSKPKQKVT
QLPVAQMICLE*PLLLTSKGTCTPLRKLPLNHQMRHTE

Marsupialia

>AMTN *Monodelphis domestica*
MKTAVLLFCLLGSTQSLPILQLMSMPVPGHGQQQAQQLYPGVGLPPPGLGLEQALFTPQQSNQLFPPLGQIPLTQMFSVGTDMQLM
TPATGLLPGIQLPMTLGDTSIAPQVNPQMLPVLVAQIGAQAVRSSEELPIAPQIFTGILLIQLPFGTGATLPTGQSGIDANTQDAAL
PAGQTGGNPAIWGTSEGLPPTSPDDVFEATIPVGIQKATEGSTTEAPNGTD*

>AMTN *Sarcophilus harrisii*
MKTAVLLFCLLGSTQSLPILQLMRMPVQGFQQVQQLHSGVGLPPTNLVLDQPTLLTPQQPSQLFPPLSQIPLSQIFSVGTDMPL
INPATGLVPIIQLPMTLADTNIAQQVSPQVLPVIWAQIGAPGAFSSSEELPMAPQIFTGILLIQLPFGTGAILPTGQERIDANTQDAA
LPAGQTGGNSAIWGTPEGQLPTSPGDDVFEATNPVGIQRIEGSTTEAPNGTD*

>AMTN *Macropus eugenii*
MKTAVLLFGLLGTIQAFFILQLMRDVSSTRPWPSTASTGIQFVPMGLPPANLVLDQSALLTPQQPNQ????????????????????
?INPAAGLVPISQIILPMTLTDANTPQQVSPQVLPVIWAQIGTQ????????PMAPQIFTGILLIQLPFGTGAMLPQGEKINANAQDTA
LPAGQTGGNPFIWGAPEGQFPPTSPGDDVFEATIPDGIQRTAEGSTTEAPNGTD*

Monotremata

>AMTN *Ornithorhynchus anatinus*
MKTIIILLFCLLGTTRSLPMLQILRMAFRDNIPQQIQVNPAPFGFLPAQLVVDHSILQTQQDSNQMIPLIWAQVGAQGAILSSSEELIF
TGLFIQPLAPGRMAPIGQDNPEQDQVLPPTGQSGMSPATQDPPEDQPPPTASGLLDESEVTLPANLQGVSLTTEESTTEPRNGN*

SAUROPSIDA

Crocodylia

>AMTN *Alligator mississippiensis*
MKTVILLFCLLGTAFSLPVRQAGTLPVNSREILQLMRLYSSSLGNIPQQTQQQVNPGLGLPQAQLLPDQTALLAQANQVFPPLNFI
LTVWPFADINLVNGAAGVPIITQTLPTGAGGANTQHLVLPQQGVLPILGQMGPPQGLGVSSEEMQTDQILAGLILLGLQGLLANQ
AVANSEGLDIVFPAGQTGVNQGSPGLPSPEGTADMSGATLPAGIKKGSQTPESITEAASGVYTTTTPGFRQLDSEQVTDIFAEPAIL
LNMEANEIQAPPTTDIGTGTIVHSINLHTQGYDNKIGVSPVRGDRALPLANIDVKALREYKRSRA*

>AMTN *Alligator sinensis*
MKTVILLFCLLGTAFSLPVRQAGTLPVNSREILQLMRLYSSSLGNIPQQTQQQVNPGLGLPQAQLLPDQTALLAQANQVFPPLNFI
LTVWPFADINLVNGAAGVPIITQTLPTGAGGANTQHLVLPQQGVLPILGQMGPPQGLGVSSEEMQTDQILAGLILLGLQGLFQANQ
AVNSEGPDIVFPAGQTGVNQGSPGLPSPEGTADMSGATLPAGIKKGSQTPESITEAASGVYTTTTPGFRQLDSEQVTDIFAEPAIL
LNMEANEIQAPPTTDIGTGTIHSINLHMQGYDNKIGVSPVRGDRALPLANIDVKALREYKRSRA*

>AMTN *Caiman crocodilus* [KM069444]
MKTVILLFCLLGTAFSLPVRQAGTLPVNSREILQLMRLYSSSLGNIPQQTQQQVNPGLGLPQAQLLPDQTALLAQANQVFPPLNFI
PLTRVWPFADINLVNGAAGVPIITQTLPTGAGGANTPHLVPQQGVLPILGQMGPPQGLGVSSEEMQTDQILAGLILLGLQGLFQANQ
QASINSEGPDIVFPAGQTGMNQGAPQLPSPEGTADVSGATLPAGIKKGSQTPESITEAASGVYTTTTPGFRQLDSEQVTDIFAEPAIL
LNMEANEIQAPPTTDIGTGTINSINLQTYGVNQIGVSPVRGDRALPLANRDRALREYKRSRA*

Squamata

>AMTN *Ophiophagus hannah*
MEVILLFCLLGLTFSTPINHIGSSLATSSKEVLRLLHRYIAISNTPPQAQSRQRAILPPAKLVDPDLNLPGNQPNVNLVLSVPQN
QIFPIDMNYLSTLLGILSTIQTLPVGGGMPVPPQPPGQAMLP IIVTQMGPPQAVLSSEEMSPSQFFTGILLIPGMQGAMFPSPGQAE
PPEAQDGLPVGANPEGTAAGIQKVFPTATDGLSEIPGELYPTPSGFRQSDTGTSGFLLEPTADINMEPSELREPTSPVKFDMNSM
QSPHNLSQPLVRGDSHIPMNTVASKPLNLP*

>AMTN *Python molurus*
MKIVILLFCLLGLAFSKPVNQAARILATSNREILRLLQRYRATGNTPPQTPRPRVVPGLPQAQLLPDQTALLAQANQVFPPLNFI
TNFPVVISQLAEVGGPNLQNFVFSIPQNQMYPIDWNYFTSLGLILSTMQALPAAGGGMTPQQLLPQGMPLPIIWAQMGPPQAVLSSE
EMSPSQFFTGILLIPGMQGALLPSGQAEPPEGGEELLPAQAGSNPEGTAAGIQKVFPTTTDGLSEIPDELYPTPSGFRQPDVGTSGM
FIQPTVDISMPESELREPTSLVKLDMGNMQSPHNLSQPLVRGDSHILVNTVASKPLNLP*

>AMTN *Python regius* [KM069438]
MKIVILLFCLLGLAFSKPVNQAARILATSNREILRLLQRYRATGNTPPQTPRPRVVPGLPQAQLLPDQTALLAQANQVFPPLNFI
TNFPVVISQLAEVGGPNLQNFVFSIPQNQMYPIDWNYFTSLGLILSTMQALPAAGGGMTPQQLLPQGMPLPIIWAQMGPPQAVLSSE
EMSPSQFFTGILLIPGMQGALLPSGQAEPPEGGEELLPAQAGSNPEGTAAGIQKVFPTTTDGLSEIPVELYPTPSGFRQPDVGTSGM
FVQPTVDISMPESELREPTSLVKLDMGNMQSPHNLSQPLVRGDSHIPVNTVARKPLNVP*

>AMTN *Crotaphytus insularis* (partial) [KM069443]
??
ANFP IISPNLELGNPNVQNLPGFVSLPPLTQMVPIDINSLAALLGVLSATQIQPLPGGGITGPQQLLPAQSMPLPIIFTQMGPPQVTL
SSEEMPSQVFAFLVPMQGLLPSGQAGTIPGQEEPLPAGQAGTNQNLPPYPESTAAGIQKGSPTTGDGLREATSISYSPSGFR
QPVVNTNEVFIPTVGINMEPSELRETT????????????

>AMTN *Crotalus mitchellii*
MKIILLLLSLLGLTFSIPINHVGRIILATSSSKEILRLLQRYIAVGNTPQQAQSRQRPVLPAPKLVDPDPSLPGNQPPNVNFPVFSLPQN
QVFPIDMTYLSNLLQILSTMQTLPFVGGGMTVPQQLPPGQAMLPIIVTQMGPQGAVLSSEEMSPSQFFTGILLPGMQGAMFPGQAEF
PPEGQDGLLPVQGANPEGTAAGIQKVFPTATDGLSEIPGELYPTPSGFRQPDAGTSDMFVEPTADINMEPSELREPPTSLVKFDMGNM
QSPHNLSQSPVVRGDSPIPMNTVASKPLNV*

>AMTN *Anolis carolinensis* [KM069435]
MKIVILLLLSLLGLTFCCLPVNQFGRRFATSNREILRLMQRYKAQGNVPPQQTQQRSNPGIPLPAKLVDPDQHPLANHIPSEIPAFERPF
ANYPIILTAQNPELGSPTLQNFPGFNVLPLTQMVPIDINYLALLLGGALSTTQTQPLPGGLNIPQQLLPPQSMFPIIFTQMGPQGT
LSSEEMPTPQVMAAFLPLGLPGGLFPGSGALPEDQDELLPAGQVGTNQGNLPPFESTAAGIQKGPATEDGLNVAPGAFYPTPSGF
RQPGVVTNEVFVEPTGGIHMEPSELREPPTSLVGLDKGNQKLNLSQSPVVRGDSYMPMTTASKPLKAP*

>AMTN *Timon tangitanus* (partial) [KM069439]
??
RDVQRYKAMGNPPQQTQQRPRLGGLPAPKLVDPDQNPVSNQVNEQVAVLEWT
FANFPVTSNFPQNPSELGGPNMQNFPNGYNVLPFMDLNSLAILLGALTLPTLPVAGVGTMPQQLLPPQSMPLPIIVAQMGPQGAVLS
SEEMPPQSQIVGVLVPGGGFFPLGQAGALPEGQEGLLPANQAGSNQGNLPPFDSTAAAIQKISPTASDGLSEVPSGLYPTPSGFRQPG
PVTNGVFAEPTAGNMMEPSELQGT???

>AMTN *Takydromus sexlineatus* (partial)[KM069441]
????ILLLLPLGLTISLPLNQMGRIILATSNREILRLMQRYKAMGNPPQQTQQRPRLGAGLPPAKLVDPDQNPVSNQVNEQVAVPEWT
FANFPVTSNFPQNPSELGGPSIQNIPGYNVLPFTQMLPMDLNTLAILLGGALTLPTLPVAGVGTMPQQLLPPQSMPLPIIVAQMGPQGAVL
AVLSSEEMSPQSQIVGVLIPGGGFFPLGQAGALPEGQESLLPANQAGSNQGNLPPFEGTAAAIQKISPTASDGLSEVAVSGV????

>AMTN *Tarentola mauritanica* [KM069446]
MKIVILFLSFLGLTIALPISPLERAIATSNREILRLMQRYRAMGNPPQQTQQRSNPGFGLPAPKLVDPDQTALANQLPNEVTPLEWPF
TNFPMTPKIPSELGTTNLQHIPGFNLLPFTQMLPMDLNTLAILLGGALTLPTLPVAGVGTMPQQLLPPQSMPLPIIVAQMGPQGAVL
SSEEAQSPSQFFTLILPQMGIILIPSGQAEALPEGQEGLLPAGQAGSNPANSAGQLPFEGTADATAPAGIQKVSFPMHDDLSGTAN
GLNPTPSGFRQPGYDHGATQDPFAERTIFVKLNMEPSELREPPTMARPEDDNLVGHGTLAQSLVGRDRHMPVTTVVSKLLREP*

AMPHIBIA

Anura

>AMTN *Xenopus (Silurana) tropicalis* [KM069437]
MVMFLLCLLRASLGFPLHLMQMFNLLGRVPEQTQRIPHFLVAPSQLVQLHGTQLRYMHPFQMSPIFAQLRQALSGSDSEENVLP
GLIMIPLSAGLDENIATVHQAGVIDNSILQEQPVLNPAGQHGVSSTAIQNKILPELPWVPTQAAQNDVSVTGIKGSYTDPLPTNP
DIDLNAPEPTMSMRDNDLTVNAHTMGEENWILCNSYVVDLCELHNIVRGDGYIPKTPVYFKNGQIQGNRNTESY*

Caudata

>AMTN *Pleurodeles waltl* [KM069436]
MRTTILLCLLGTSTIAFPVREAEILSASNSGEMQLMQLYNMLGNTPEQTQVHHPGHGLPAPQLIPDMTAQQTQQQLSQFQPSNVF
GNLPIILPPLSMPPGFPPQSPGNQPPGVMPYPQALPMNVGEMTVHHSMFPHHPLIITQMGGQLLMSSEEMFPLQVFAVIMPPF
GPILPALPVLPGVPAFAGLPLPLLQGAARPAAGAGISLDEQEAIPMGQAGLNPKNQRE TASPAPGVTPAQDTGEVLAVTVPAGLRK
GVTGTD SAGLLQFTTPPAGLRQFPTE SATEVMFVEPTGLVTETSPDGVDERLPHSYRPNQRFN SVARGDQNLAMAVPDLKPLRKHKK
RRA*

COELACANTHIFORMES

>AMTN *Latimeria chalumnae*
MKTIMLLFCFLGTAVAVPVHSGVRSVHGGVAASNSRERTPLLVSQPQLQGGQPRILQLQTFPERVQVPVQPNIQGQQQLLPSYRFVP
LTQLLAQGPVHQAALAVNIPQLVPPQSTSSSEEMVQMSGTGLIPVVDPAVPGVPAIEMPVPGGILPNPGGNPSTQGGDTPIPGNPVG
GQPTVPTGVRKIAPTAEVSVTDVAVNQGIFQPVAGATPTAAGLKKTVQDEDRPDVIFVEPSPPCDIPNGMKQLTFSEELITVSPA
EASITISSVTNSMPKNKWHIASKAGLRGDIAGSLKAAGKKPIQNY*

Additional file 3. Forward (F) and reverse (R) primers used to amplify Amelotin cDNA of various species and to build *AMTN* probes.

Target regions and annealing temperatures are indicated.

Species	Primers	Target region	Annealing temperature
<i>Anolis carolinensis</i>	F: CAACAGCAAGATACCAATTTGG R: GCAGTGACACACATCATTGG	5' UTR to 3' UTR	50°C
	Probe: F: GGTTCAGATCAGCATCCAT R: ATTCAGGCCATCTTCAGTCG	exon 3a to exon 8	58°C
<i>Python regius</i>	F: AGTCGTGGGGTAAGAGATGC R: TGCTTCCCTTTCTGAGCTG	5' UTR to 3' UTR	59°C
<i>Takydromus sexlineatus</i>	F: ATGAAGATTGTAATTTTACT R: CATTCCTCAGAGCTGAGCA	exon 2a to exon 7	50°C
	F: ATATTACGGCTGWTGCAAAG R: GGTGGTTCCCGAAGTTCACCTTG	exon 2c to exon 8 (3' end)	50°C
<i>Tarentola mauritanica</i>	F: CTACAAGCAACAGCAGAGA R: CTRTCTCCTCTCACARGTG	exon 2b to exon 8 (3' end)	50°C
<i>Caiman crocodilus</i>	F: TTTACCAGTGCGACAAGCTG R: TCTCCTCTCACAGGGGATACAC	exon 2b to exon 8 (3' end)	60°C
<i>Xenopus tropicalis</i>	F: GTTCCACAGAAATATATGGTCATG R: GCAACTCTAAACAAATCTACAGAC	5' UTR to 3' UTR	55°C
<i>Pleurodeles waltl</i>	F: AGAAACATTCGGGACTGGTG R: TTGCCAGATTGTGATCGT	5' UTR to 3' UTR	60°C
	Probe: F: ACCAGGCTTTCCACAGTCAC R: ATCACCTCTCGCTACGCTGT	exon 3b to exon 8 (3' end)	62°C

CHAPITRE 5

Amelotin gene structure and expression in the
opossum *Monodelphis domestica*

Amelotin gene structure and expression in the opossum *Monodelphis domestica*

Barbara Gasse (1), Xi Liu (1, 2), Erwan Corre (2) & Jean-Yves Sire (1) *

(1) UMR 7138-Evolution Paris-Seine, Institut de Biologie Paris-Seine, Université Pierre et Marie Curie, France

(2) Plateforme ABiMS (Analyses and Bioinformatics for Marine Science), Station Biologique de Roscoff, France

*Corresponding author

Statut actuel : Cet article a été soumis pour publication dans *PLOS ONE* en mars 2015.

Les différences observées dans la structure génique et dans d'expression spatio-temporelle de l'ARNm d'*AMTN* entre la souris et les tétrapodes non-mammaliens (chapitre 4), nous avaient conduits à formuler l'hypothèse qu'un évènement important s'était produit, probablement très tôt, au cours de l'évolution des mammifères, comme l'indiquait la conservation de la structure génique chez ces derniers (chapitre 3). Cependant, nous ne possédions que la souris comme point de comparaison pour les données d'expression et la plupart des séquences mammaliennes étaient des séquences génomiques.

Par ailleurs, dans une publication récente, Kawasaki et Amemiya (2014) avaient prédit par des recherches *in silico* une séquence d'*AMTN* de l'opossum (*Monodelphis domestica*) différente de celle que nous avons prédite dans notre premier article (chapitre 3), c'est-à-dire proche de celle des tétrapodes non-mammaliens (i.e. se terminant par un long exon 8) alors que nous la supposions semblable à celle de tous les autres mammifères (i.e. se terminant par un court exon 8 suivi d'un exon 9).

Nous avons saisi l'opportunité de pouvoir disposer d'échantillons de mâchoire de jeunes opossums pour obtenir la séquence complète de l'ARNm d'*AMTN* et ainsi vérifier la structure génique et son expression spatio-temporelle au cours de l'amélogénèse chez ce marsupial. Si le domaine d'expression est très semblable à celui

qui a été décrit chez la souris (expression au cours des stades tardifs de l'amélogenèse) nous avons trouvé deux variants correspondant pour 70% au transcrit "mammalien" et 30% au transcrit "non-mammalien" ce qui nous a permis de mieux préciser le mode d'évolution d'AMTN chez les mammifères.

ABSTRACT

Amelotin (AMTN) is an ameloblast-secreted protein that belongs to the secretory calcium-binding phosphoprotein family, which also includes the enamel matrix proteins. Although AMTN is supposed to play an important role in enamel formation, data on its expression (gene and protein) during amelogenesis were long limited to the rodents, in which it is expressed during the maturation stage. Recent comparative studies revealed that (i) *AMTN* was expressed earlier in sauropsids and amphibians, i.e. as soon as ameloblasts are depositing the enamel matrix, and (ii) the gene structure was different compared to that deduced from many mammalian genomes using *in silico* approaches. In the present study, the question was whether the differences in gene structure and expression observed between rodents and non-mammalian tetrapods dated back to an early ancestral mammal or were acquired later in mammalian evolution. Here, we sequenced, assembled and screened the jaw transcriptome of a neonate opossum *Monodelphis domestica* (marsupial). We found two *AMTN* transcripts. Variant 1, representing 70.8% of the total amount, displayed the *AMTN* structure known in mammals, whereas variant 2 (29.2%) exhibited the non-mammalian-like structure. In addition, we studied *AMTN* expression during amelogenesis in a neonate specimen. We obtained similar data as those reported in rodents. These findings indicate that (i) the important changes in *AMTN* structure leading to variant 1 occurred in an ancestral mammal, more than 180 Ma before the divergence of marsupials and placentals, (ii) the resulting isoform 1, lacking the RGD motif, was selected in the mammalian lineage although the ancestral isoform 2 was still present at a lesser amount, and (iii) there are probably relationships between the gene structure and its spatiotemporal expression during amelogenesis.

Key words. Amelogenesis, Enamel, Amelotin, In situ hybridization, Evolution

INTRODUCTION

Amelotin (AMTN) belongs to the secretory calcium-binding phosphoprotein (SCPP) family, and more precisely to the Pro-Glu rich sub-family, which includes among others the three currently known enamel matrix proteins (EMPs), amelogenin (AMEL), ameloblastin (AMBN) and enamelin (ENAM) (Kawasaki *et al.*, 2007). All of them play an important role during enamel matrix formation, mineralization and maturation. The three EMP genes and *AMTN* were found in tetrapod and coelacanth (*Latimeria chalumnae*) genomes (Kawasaki & Amemiya, 2014). However, only a few data are available on *AMTN* expression and on the supposed function(s) of the encoded protein. In mammals, our knowledge of *AMTN* gene expression is limited to the rodents (Iwasaki *et al.*, 2005), while in non-mammalian vertebrates it was recently studied in a lizard and a salamander (Gasse *et al.*, 2015). Comparison of *AMTN* expression revealed important spatiotemporal differences in rodents versus non-mammalian tetrapods. During amelogenesis, *AMTN* was described as being expressed by ameloblasts from maturation stage onwards (Iwasaki *et al.*, 2005) while in lizard and salamander *AMTN* transcripts were detected from early enamel matrix formation to late enamel maturation, with mRNA expression extending deep along the dentin shaft (Gasse *et al.*, 2015). These findings suggest different functions of *AMTN* during amelogenesis in non-mammalian tetrapods and rodents. In the latter it is generally admitted that *AMTN* is not an EMP, and that it could play a role in the adherence of ameloblasts and the junctional epithelium to the enamel surface, or/and in the mineralization of the outer aprismatic enamel layer (Moffatt *et al.*, 2006b; Lacruz *et al.*, 2012; Somogyi-Ganss *et al.*, 2012a; Nishio *et al.*, 2013). Nevertheless, our recent data in non-mammalian tetrapods clearly indicate that the ancestral *AMTN* was an EMP, which possessed several functional motifs that were lost during mammalian evolution (Gasse *et al.*, 2015). It is clear that important changes occurred in an ancestral mammalian *AMTN*, but the only available data in rodents were not appropriate to identify when and how this event took place during mammalian evolution.

In a previous study, using a large number of genomic sequences (gDNA) we demonstrated that the *AMTN* structure was conserved in all mammalian lineages, and we suggested that the origin of this gene structure took place in the last common mammalian ancestor, more than 200 millions years ago (Ma) (Gasse *et al.*, 2012). This finding could indicate that the *AMTN* function, as proposed in rodents, dated

back to this ancient geological period (310-200 Ma), during which early mammals acquired dental occlusion and prismatic enamel (Wood *et al.*, 1999). However, the hypothesis of structural/functional changes in an ancestral *AMTN* prior the differentiation of current mammalian lineages was recently questioned. Kawasaki and Amemiya (2014) suggested that the 3' extremity of the coding *AMTN* sequence in the opossum *Monodelphis domestica* (Marsupialia) was similar to that described in non-mammals, i.e. the sequence ends with a larger exon 8 and encodes a RGD motif which is missing in rodent *AMTN*. This finding was in contrast with the *AMTN* structure of the opossum we deduced from comparison with many mammalian sequences, i.e. the sequence has a short exon 8 lacking the RGD motif and ends with exon 9 (Gasse *et al.*, 2012). Furthermore, Kawasaki and Amemiya (2014) revealed the presence of an extra exon located between exons 2 and 3 of opossum *AMTN*. This exon has been found in all *AMTN* sequences of non-mammalian tetrapods, but is lacking in rodent *AMTN* cDNA. We also recently suggested that this exon (that we called exon 2c) was functional and should be present in marsupial *AMTN* cDNA (Gasse *et al.*, 2015). However, these hypotheses were proposed through screening the gDNA of *M. domestica* and accurate data based on mRNA sequencing were necessary to confirm or infirm them.

The present study was undertaken in *M. domestica* with the aims to (i) define the correct *AMTN* structure in this species and the protein composition, and (ii) analyze the spatiotemporal expression of *AMTN* transcripts during amelogenesis. Fulfilling these objectives was crucial to better understand when structural changes took place in *AMTN* during mammalian evolution and whether they resulted in different spatiotemporal expression.

MATERIAL AND METHODS

The heads of two opossum (*Monodelphis domestica*) neonates (16 days post fertilization, dpp) in which teeth are forming were used in our study. The material was a gift from the King's College London facility lab (courtesy Dr Abigail Tucker). One head was preserved in RNALater for RNA extraction and sequencing, the other was fixed in PFA 4% and demineralized in EDTA 5% for *AMTN* mRNA expression study.

RNA extraction and jaw transcriptome sequencing

The lower jaw was dissected, cut into small pieces and frozen in liquid nitrogen. Total RNA was extracted and aliquoted. For a detailed description the readers are referred to Gasse et al. (2015). Illumina sequencing [one run, 50 base pair (bp) paired end] was commissioned to GATC Biotech. The sequenced transcriptome was assembled at ISEM-Montpellier 2 (France) using Trinity (release Apr13, 2014) the genome-independent transcriptome assembler using the default parameters (Grabherr *et al.*, 2011). Then the 192,056 sequences were screened for *AMTN* transcripts with BLAST using the Montpellier Bioinformatics Biodiversity (MBB) platform at [http://mbb.univ-montp2.fr/MBB/html_pise/BlastDB.html].

PCR

To check the results obtained from the jaw transcriptome sequencing, primers were designed with Primer3 v.0.4.0 (<http://bioinfo.ut.ee/primer3-0.4.0/primer3/>; (Untergasser *et al.*, 2012)) using the *AMTN* sequences obtained in the transcriptome. Two sets of primers were designed to obtain the 3' end of the transcripts, depending on the presence of either a large exon 8 alone (Reverse 1) or a short exon 8 followed with exon 9 (Reverse 2).

Forward: GTCTCCTAGGAACAATCCAATCA (exon 2a)

Reverse 1: ACTGGCAGATGAGTGTCTCC (3' end of sauropsid-like exon 8)

Reverse 2: CTTGTGGGGCAGATTAGAGG (3' end: UTR of putative exon 9)

Total RNAs were routinely extracted, purified (RNeasy Midi Kit; Qiagen, France), and converted into cDNA (M-MLV Reverse Transcriptase, Invitrogen) using an oligo(dT)18 primer. *AMTN* transcripts were recovered by PCR amplification (annealing temperature: 58°C) using GoTaq DNA polymerase (Promega, France), as previously described (Al-Hashimi et al., 2010). Sequencing was performed by GATC Biotech. We used SeaView 4.5.2 for sequence alignment (Gouy *et al.*, 2010).

For sequence comparison we used the published *AMTN* cDNA of the rodent *Mus musculus* (Genbank accession NM_027793.1), the crocodile *Caiman crocodilus* (KM069444) and the lizard *Anolis carolinensis* (KM069435).

Transcript quantification

Relative abundances of the two transcripts were estimated by remapping the reads on the transcriptome assembly with Bowtie2 (Langmead *et al.*, 2009) and performing RSEM (RNA-Seq by Expectation Maximization) abundance estimation (Li & Dewey, 2011). For each transcript the FPKM (Fragments Per Kilobase Of Exon Per Million Fragments Mapped) values and abundance estimation of variants were computed.

***In situ* hybridization**

AMTN transcript fragments were amplified (706 bp) as previously described (Gasse *et al.*, 2015), using the primer set 2. Therefore, the probe possesses a common sequence in their N-ter region and differed in their C-ter region.

Briefly: purified cDNAs were inserted into a vector containing T7 and SP6 promoters for *in vitro* RNA transcription, and transformed into competent *E. coli*. Colonies containing the vector and the insert were selected, plasmids purified and linearized by PCR, and the antisense RNA probe was synthesized in the presence of digoxigenin.

The jaws were dissected, dehydrated, and embedded in paraplast. The sections (8 µm-thick) were deposited on slides, dewaxed in toluene, rehydrated, treated with proteinase K, post-fixed in paraformaldehyde, rinsed and incubated overnight with the digoxigenin-labeled antisense probe in the hybridization buffer. The following day, the slides were washed three times, rinsed in the maleic acid buffer tween, non-specific binding sites blocked, and incubated overnight with the anti-digoxigenin antibody coupled to alkaline phosphatase. Then, the slides were rinsed, the digoxigenin-labeled probes revealed by immunocytochemistry, mounted and photographed.

RESULTS

Amelotin gene structure

Screening the assembled jaw transcriptome using the putative *AMTN* sequence of opossum provided two transcripts with high e-value. They were 1138 bp and 1730 bp long. Alignment with the putative gDNA sequence revealed that these transcripts corresponded to two *AMTN* variants, variant 1 (V1) and variant 2 (V2) respectively, and included the 3' and 5' untranslated region (UTR). The transcripts share the same sequence, including a 119 bp-long 5' UTR, and differ in their 3' region. The shortest transcript, V1, possesses a short exon 8 and ends with exon 9, encoding the three last

residues, the stop codon and a 326 bp-long 3' UTR. The coding cDNA sequence was deposited in Genbank under accession number KP184475. The largest transcript, V2, possesses a large exon 8, encoding a RGD motif, the stop codon and a 669 bp-long 3' UTR (accession number KP184476). PCR sequencing confirmed the presence of these two variants. In addition, cloning the transcripts for probe synthesis resulted in identifying a third variant, not found in the transcriptome. This variant 3 (V3) ends with exon 9 but lacks exon 7 (accession number KP184477). Skipping exon 7 (V3) results in the absence of the putative phosphorylation site (SxE motif) in the encoded protein. The cDNA sequences were translated into amino acid sequences and aligned to the mouse, crocodile and lizard *AMTN* sequences (Fig. 1).

The three variants possess the exon 2c previously identified as putatively functional in the opossum gDNA, but lack exon 2b, which is present in non-mammalian tetrapods and encodes an SxE (SRE) motif. Therefore variants 1 and 2 possess a single SxE motif (encoded by exon 7), while V3 lacks this phosphorylation site. Exon 3b is lacking, which confirms its absence in mammalian *AMTN*. The RGD motif, encoded by all non-mammalian *AMTNs*, is only encoded by the 3' region of V2 in opossum.

Relative abundance of amelotin mRNA variants

The FPKM values (number of normalized reads) for variants 1 and 2 (variant 3 not identified in the transcriptome) were 6.39 and 2.64, respectively. FPKM%, i.e. the percentage of the two *AMTN* transcripts related to all transcripts, were 0.00115 and 0.00047, respectively, which means that V1 represented 70.79% vs 29.21% for V2 of all *AMTN* transcripts in the sequenced jaw transcriptome.

	exon 2a	exon 2b	exon 2c	exon 3a
Mus	<u>MKTMILLLCLLGS</u> QSLP	-----	ILQLMSMPVPGHGQQQAQ	KQLNPASGVPATKPTPGQVTPLP-QQQPNQ
Monodelphis V1	...AV.....TI....	-----	ILQLMSMPVPGHGQQQAQ	-..Y.GV.L.PP.LGLE.-PA.FTP....
Monodelphis V2	...AV.....TI....	-----	ILQLMSMPVPGHGQQQAQ	-..Y.GV.L.PP.LGLE.-PA.FTP....
Caiman	...V.....T.F...	VRQAGTFLPVNSRE	ILQLMRLYSSLGNIPQQTQ	Q.V..GL.L.PAQLL.D.-.A.LA-..A..
Anolis	..IV....S...LTFC..	VNQFGRRFATSNSRE	ILRLMQRYKAQGNVPQQTQ	QRS..GI.L.PA.LV.D.-H..A.-HI.SE
	exon 3b	exon 4	exon 5	
Mus	-----	VFPSISLIPLTQLLTLGSDLPL	FNPAAG-PHGAHTLPFTLGPLNGQQQLQP--Q-	
Monodelphis V1	-----	L..PLGQ.....MFSV.T.MQ.	MT..T.LLP.IQI..M...DTSIAP.VN.-Q.-	
Monodelphis V2	-----	L..PLGQ.....MFSV.T.MQ.	MT..T.LLP.IQI..M...DTSIAP.VN.-Q.-	
Caiman	-----	...TLNF.....RVWPF.A.IN.	V.G...GVPITQ...TGA.GA.TPH-.V.-Q.G	
Anolis	IPAFERPFANYPILTAQNPELGSPTLQ	N..GFNVL....MVPI--.INY	LALLL.ALSTTQ.Q.LPG.G..IP...L.PP.S	
	exon 6	exon 7	exon 8	
Mus	MLPIIVAQLGAQ	GALLSSEEL	PLASQIFTGLLIHPLFPGAIPPSGQAGTKPDVQNGVLPTRQAGAKAVNQGTTPGHVTTTPG	
Monodelphis V1	...VL...I...I...	..VR.....I.P.....Q.FGT..TL.T..S.IDANT.DAA..AG.T.GNPATW..SE.QLP..S		
Monodelphis V2	...VL...I...I...	..VR.....I.P.....Q.FGT..TL.T..S.IDANT.DAA..AG.T.GNPATW..SE.QLP..S		
Caiman	V..L.LG.M.P..LGV...M	QTDA..LA..IL-LGLQ..-FLQAN..SINSEGPDI.F.AG.T.MNQA----.QLPS.E		
Anolis	.F...FT.M.P..TV....M	-PTP.VMAAF.L-.GL..GLF....S.AL.ED..EL..AG.V.TN-----Q.NLPP.F.E		
	exon 8	exon 9		
Mus	VTDDDDYEMSTPAGLRRATHTEGTTIDPPN	RTQ* (213)		
Monodelphis V1	SP-..VF.ATI.V.IQK---I...S.ESA..	..D* (229)		
Monodelphis V2	SP-..VF.ATI.V.IQK---I...S.ESA..	GGFFQTHIMPLG // PTMLILRGDTHLPV* (312)		
Caiman	G.-A.VSGATL...I-KG-SQ.PESITEAASGVYTPPGFRQL	// IGVSPPVRGDRALPLANRDVRLREYKRSRA* (328)		
Anolis	S.-----A..IQKSPA..DGLNVAPGAFYPTPSGFRQP	// LSQSPVRGDSYMPMTTTASKPLKAP* (334)		

Figure 1. Alignment of the two main *AMTN* variants (V1, V2) of *Monodelphis domestica* with rodent (*Mus musculus*), crocodile (*Caiman crocodylus*) and lizard (*Anolis carolinensis*) sequences. Functionally important motifs (SxE motifs encoded by exon 2b and exon 7, and the RGD motif encoded at the end of exon 8) are highlighted in grey. The signal peptide is underlined. Sequence length is indicated in brackets. Exon limits are indicated by vertical lines; (.): residue identical to *M. musculus* *AMTN* residue; (-): indel; *: stop codon. (//): for convenience of presentation the sequences were shortened.

Expression of amelotin gene during amelogenesis

Several teeth were forming in the jaws of the 16 dpp neonate we used for *in situ* hybridization (Fig. 2). Serial, longitudinal sections of demineralized upper and lower jaws revealed various stages of amelogenesis depending on the section level. Most teeth were well developed and exhibited features indicating that ameloblasts were well differentiated (Fig. 2B-D). Most, if not all, tooth sections revealed the presence of a well developed dental organ with differentiated, polarized odontoblasts facing predentin matrix, and a well developed enamel organ, with prominent stellate reticulum and differentiated, polarized ameloblasts facing either a thin layer of enamel matrix or the dentin layer. *In situ* hybridization did not reveal the presence of *AMTN* transcripts in the differentiated ameloblasts located along the dentin shaft, nor in the ameloblasts located upper and facing recently deposited enamel matrix (Fig. 2B). In contrast, in sections through the upper, more advanced region of some teeth, *AMTN* mRNA were clearly identified in the ameloblasts facing well mineralized or mature enamel (Fig. 2C and 2D, respectively), recognizable as an empty space devoid of matrix resulting from the demineralization process. Careful examination of all serial sections of upper and lower jaws did not reveal other labelled cells, neither odontoblasts nor cells located along the tooth base. These findings indicate that *AMTN* transcription is activated late in the ameloblasts during opossum amelogenesis, from the transition stage (advanced mineralization) onwards.

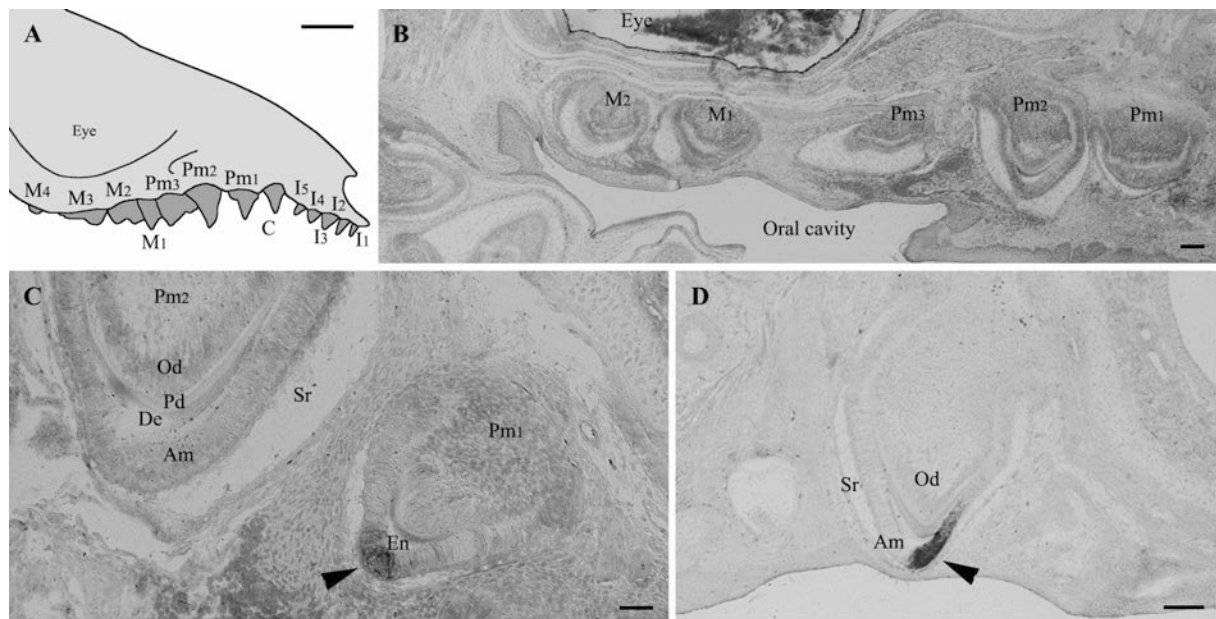


Figure 2. **A:** Schematic drawing of the upper jaw of a *Monodelphis domestica* neonate (lateral view). **B-D:** *In situ* hybridization of longitudinal tooth sections in the upper jaw of a 16dpp neonate using *AMTN* probe. **B:** Low magnification of a section showing five developing teeth. No labelling of *AMTN* transcripts can be observed although the teeth are already formed and present well-differentiated ameloblasts with Tomes' processes. Enamel matrix is present but not yet undergoing maturation process, at least in this section level of premolar 2. **C:** Section of premolars 1 and 2. A clear labelling of *AMTN* transcripts (arrowhead) is observed in a few ameloblasts located at the tooth tip and facing mineralized enamel of premolar 1. The ameloblasts of premolar 2, which are well-differentiated and display Tomes' processes, are not labelled with *AMTN* probe. **D:** Section through the upper region of premolar 1 showing a clear *AMTN* expression in the ameloblasts facing the well-mineralized enamel only. Am: ameloblasts; De: dentin; Od: odontoblasts; Pd: predentin; Sr: stellate reticulum. C: canine; I: incisor; M: molar; Pm: premolar. Scale bars: A = 2 mm; B, D= 100 μ m; C: 50 μ m.

DISCUSSION

Amelotin gene structure changed during mammalian evolution

The discovery of three *AMTN* variants in the opossum jaw transcriptome brings further clues to *AMTN* evolution in mammals and answers the initial question, on which transcript was present in marsupials (Gasse *et al.*, 2015). The presence of variants 1 and 2, i.e. both placental and non-mammal sequences, respectively, demonstrates that the mutation creating a novel, intra-exonic splice site resulting in V1 occurred in *AMTN* exon 8 before the differentiation of Marsupialia and Placentalia, i.e. more than 180 Ma (Madsen, 2009). Also, this finding indicates that splicing is alternative because the ancestral sequence, i.e. variant 2, is also present in

the transcriptome. The donor splice site resulting in skipping the 3' region of exon 8 (hence the stop codon) took place using a GTG sequence, which changed later in placental *AMTN*, into the more common GTA sequence (Gasse *et al.*, 2015). The relative expression level of these two variants provided by the estimated FPKM values, i.e. 70.8 vs 29.2% respectively, although providing no accurate data due to the lack of replicates, could indicate that selection of V1 occurred either in the marsupial lineage or earlier, in an ancestral mammal. Indeed, alternative splicing is no longer present in placental *AMTN*, as for instance in rodents (Moffatt *et al.*, 2006b), in which V2 was not found. We believe that (i) the intra-exonic splicing was favored when GTG changed into GTA, and (ii) the selective pressure on V2, encoding the RGD motif in the C-ter region of the protein in non-mammals, was relaxed.

In our previous evolutionary analysis of *AMTN*, we concluded that the gene structure in all mammalian lineages was similar to the rodent sequence, i.e. the V1 sequence, the 5' region of which encodes a protein lacking the RGD motif (Gasse *et al.*, 2012). Here, we show that this conclusion was not entirely correct because the opossum *AMTN* sequence can alternatively end with a large exon 8, as in non-mammals. Such *AMTN* structure was predicted by Kawasaki and Amemiya (2014) and by Genbank (ENSMODT00000015781). In absence of cDNA from the platypus *Ornithorhynchus anatinus* we cannot complete the mammalian *AMTN* story, although a similar intraexonic splice site was predicted to exist in exon 8 of this species (Gasse *et al.*, 2012).

Variant 3 sequence, i.e. the *AMTN* transcript lacking exon 7, was only found when cloning PCR products and was not present in the transcriptome assembly. However, PCR and transcriptome sequencing were performed from the same RNA extraction. This means that V3 transcripts were present but probably at a too low amount to be assembled after transcriptome sequencing. V3 sequence was probably obtained at random when cloning. It is worth noting that such variant is known in rat *AMTN* (Moffatt *et al.*, 2006b), which indicates that such alternative splicing can occur.

In the present study we also confirm the presence of exon 2c in the three variants. This exon was previously predicted to be present through screening of opossum gDNA and sequence comparison among tetrapods (Gasse *et al.*, 2015; Kawasaki & Amemiya, 2014). Exon 2c was present in the last common ancestral tetrapod, conserved in early mammals, including monotremes and marsupials, and

then lost in placentals. The loss of this exon indicates that the encoded region was not important for the protein function. The presence of exon 2c in marsupial *AMTN* could be considered a remnant of the ancestral sequence in tetrapods, having probably a limited function.

More interestingly, at the functional point of view, we also confirm the lack of exon 2b, as previously observed in rodent *AMTN* cDNA (Iwasaki *et al.*, 2005; Moffatt *et al.*, 2006b). Indeed, in non-mammalian tetrapods exon 2b encodes an SxE motif. Such motif is known to play an important role as a phosphorylation site and it characterizes many SCPP family members (Kawasaki & Weiss, 2003). The loss of this motif has certainly important consequences for *AMTN* function during amelogenesis in mammals compared to non-mammalian tetrapods, but this hypothesis remains to be tested. However, as described in rodents the opossum *AMTN* possesses another putative phosphorylation site encoded by exon 7 (Iwasaki *et al.*, 2005; Moffatt *et al.*, 2006b). A recent study has shown that a synthetic phosphopeptide pSpSEEL was able to promote hydroxyapatite precipitation and that the two Ser of this motif are crucial for the mineralizing ability of *AMTN* (Abbarin *et al.*, 2014). Conservation of this SxE motif through mammalian evolution while the other SxE was lost suggests different function for these two sites during amelogenesis. Here also further functional studies are needed to understand the respective role of these two sites. In particular, one can wonder which function has, if translated, the minor *AMTN* transcript that lacks exon 7 and would encode a protein devoid of any SxE motif in opossum and rat. The short phosphopeptide pSpSEEL promoted HA precipitation from the solution, albeit to a lesser extent than the full length recombinant human *AMTN* protein, while the non-phosphorylated peptide had no effect on mineralization. Other mineral interacting sites should exist on the sequence that are involved in HA mineralization and required for a fully functional protein (Abbarin *et al.*, 2014).

A late spatio-temporal amelotin expression was defined early in mammalian evolution

The availability of the jaws of one *M. domestica* neonate, in which teeth were at various stages of development allowed us to demonstrate that during amelogenesis *AMTN* expression occurs late during amelogenesis, when enamel matrix is well mineralized. Such expression pattern is similar to that described in mice, in which

AMTN transcripts were identified from transition to late maturation stages (Iwasaki *et al.*, 2005). This finding drastically contrasts with the results obtained in amphibians and sauropsids, in which *AMTN* is expressed from early enamel matrix deposition to late maturation stage, and even along the dentin shaft when teeth are well formed (Gasse *et al.*, 2015). In opossum, the *AMTN* probe never labelled differentiated ameloblasts at early stages of enamel matrix deposition, nor the inner dental epithelial cells located along the tooth base after enamel maturation. Such results were similar to those reported in mice (Iwasaki *et al.*, 2005). Taken together these data indicate that in ancestral tetrapods *AMTN* was involved in processes more extended spatially and temporally during tooth development than in marsupials and placentals. *AMTN* expression was restricted early in mammalian evolution, playing only a role during late stage of amelogenesis. These changes occurred at least 180 Ma before the two lineages, Marsupialia and Placentalia diverged (Madsen, 2009).

Both gene organization and expression support structure-function relationships

By demonstrating an ancient origin of the current *AMTN* structure and expression pattern in mammalian evolution, our results reinforce our previous hypothesis of a relation between gene (and protein) structure and the expression pattern (Gasse *et al.*, 2015). Without elaborating exaggerated speculations it is clear that when *AMTN* possesses the two SxE (encoded by exons 2b and 7) and the RGD (encoded by the 5' region of exon 8) motifs as in lizards and salamanders, the expression pattern is spatially and temporally extended, and that these features were the ancestral condition in tetrapods. In contrast, when *AMTN* is lacking the first SxE and the RGD motifs, the expression pattern is restricted to the late stages of enamel formation, a feature which could be related to the loss of the above motifs. Further experiments should be designed to reveal the function of these *AMTN* motifs during amelogenesis and the reasons why they were lost in an ancestral mammal. In the same field of research, studies in our laboratory are currently performed to check whether the expression patterns of EMPs (amelogenin, ameloblastin and enamelin) are similar or different in rodents and sauropsids as observed for *AMTN*.

In summary, in an ancestral mammal, an unknown event occurred in the regulatory region of *AMTN*, resulting in the late expression of *AMTN* during amelogenesis. Then, more than 180 Ma, i.e. before the differentiation of the marsupial lineage, a mutation in the large exon 8, characteristic of *AMTN* in the last common

ancestral tetrapod (Gasse *et al.*, 2015) occurred. This mutation resulted in the creation of an alternative, intra-exonic splicing leading to the disappearance of the RGD motif encoded by this 5' region of the gene, and probably to drastic change in the protein function. Both transcripts first coexisted as alternative variants, the shortest being progressively selected to finally become the only transcript (putting aside the variants concerning other exons) present in placental mammals.

Acknowledgments. We are grateful to Dr Abigail Tucker and Neal Anthwal (King's College London) for sending us the heads of the two opossum neonates. This study was financially supported by CNRS, Université Pierre et Marie Curie, and by the Project "Jaws" (ANR-12-BSV7-020).

CHAPITRE 6

Comparative expression of the ameloblast-secreted genes, amelogenin, ameloblastin, enamelin and amelotin during amelogenesis in the lizard *Anolis carolinensis*

**Comparative expression of the ameloblast-secreted genes,
amelogenin, ameloblastin, enamelin and amelotin during amelogenesis
in the lizard *Anolis carolinensis***

Barbara Gasse and Jean-Yves Sire*

UMR7138 Evolution Paris-Seine, Institut de Biologie Paris-Seine, Université Pierre et Marie Curie, France

*Corresponding author:

Statut actuel : cet article est en cours de préparation et sera soumis pour publication prochainement.

Les différences observées à la fois dans la structure génique et dans le patron d'expression d'*AMTN* entre le lézard *Anolis carolinensis* et les mammifères, nous ont conduit à nous demander ce qu'il en était pour les trois gènes codant les protéines de la matrice de l'émail amélogénine (*AMEL*), améloblastine (*AMBN*) et énaméline (*ENAM*). Seule l'expression d'*AMEL* avait déjà été étudiée par notre équipe chez un lézard (*Chalcides viridanus*) et un amphibien (*Pleurodeles waltl*), et était très semblable à celle que l'on connaît chez les mammifères. Toutefois, ni l'expression d'*ENAM* ni celle d'*AMBN* n'avait été étudiée chez un tétrapode non-mammalien.

Nous avons donc comparé l'expression des quatre gènes *AMEL*, *AMBN*, *ENAM* et *AMTN* au cours de l'amélogenèse chez *A. carolinensis*, avec celle qui est connue chez les mammifères. Chez ce lézard, seule l'expression d'*AMTN* est différente de celle des mammifères, ce qui pourrait suggérer un lien entre les modifications importantes de la structure génique et du domaine d'expression d'*AMTN* et la transition entre l'émail non prismatique des tétrapodes non mammaliens et l'émail prismatique des mammifères.

ABSTRACT

In a recent study, we have demonstrated that amelotin (*AMTN*) gene structure and expression have changed during tetrapod evolution. Indeed, this ameloblast-secreted gene is expressed throughout amelogenesis in non-mammalian tetrapods, while in mammals its expression is restricted to the transition and maturation stages of enamel formation. Previous studies about amelogenin (*AMEL*) gene expression in a lizard and an amphibian have shown similar expression pattern to that in mammals. To our knowledge, there are no data regarding ameloblastin (*AMBN*) and enamelin (*ENAM*) gene expression in non-mammalian species. The present study aims to look at, and compare, the gene structure and expression of the ameloblast-secreted genes, *AMEL*, *AMBN*, *ENAM* and *AMTN* during amelogenesis in the lizard *Anolis carolinensis*.

Here, we provide the full-length cDNA sequence of *A. carolinensis* *AMEL* and *AMBN*, and show for the first time the gene expression of *ENAM* and *AMBN* in a non-mammalian species.

Gene expression of *AMEL*, *AMBN* and *ENAM* in ameloblasts during amelogenesis in *A. carolinensis* is similar to that described in mammals. It is noteworthy that *AMEL* and *AMBN* expression is also found in odontoblasts.

Our findings indicate that *AMTN* is the only ameloblast-secreted gene that is differentially expressed in ameloblasts between mammals and sauropsids. *AMTN* may be the key gene to explain the structural differences between mammalian and reptilian enamel i.e., presence versus absence of prisms.

INTRODUCTION

The process of enamel formation, amelogenesis, has been extensively studied in mammals, mostly in rodents (Simmer & Fincham, 1995; Bartlett, 2013). Six ameloblast-secreted proteins (ASPs) are expressed during amelogenesis and are involved either in enamel matrix formation or in the control of the mineralization process, or in both. Three of them, amelogenin (AMEL), ameloblastin (AMBN) and enamelin (ENAM) are expressed from the first stages of amelogenesis onwards and are named enamel matrix proteins (EMPs), with AMEL representing circa 90% of the forming enamel matrix (Termine *et al.*, 1980). EMPs have important function as demonstrated by the important enamel disorders when encoding genes are invalidated in mice (Gibson *et al.*, 2001; Smith *et al.*, 2009) and by the numerous mutations in humans leading to amelogenesis imperfecta (Wright, 2006; Poulter *et al.*, 2014). In rodents, the three other ASPs, odontogenic ameloblast-associated protein (ODAM), amelotin (AMTN) and secretory calcium-binding phosphoprotein proline and glutamine rich 1 (SCPPPQ1) are only expressed during the late stages of enamel formation (Moffatt *et al.*, 2006b, 2008, 2014) and their role is not well known (Ganss & Abbarin, 2014). To date, in humans no mutation in the genes coding for these three proteins has been reported to result in any enamel disorder. *ODAM* and *SCPPPQ1* genes map to the locus for AIH2 on the chromosome 4q (Kärman *et al.*, 1997), therefore they could be good candidate genes as shown for *AMTN* (Gasse *et al.*, 2012).

The six ASPs belong to the Secretory Calcium-binding Phosphoprotein (SCPP) family, whose genes arose from an ancestral gene through tandem duplications (Kawasaki & Weiss, 2003; Kawasaki *et al.*, 2004). ASP encoding genes were found in the coelacanth genome (Kawasaki & Amemiya, 2014). This finding indicates that all these ASP genes were already present in the genome of the last common ancestor of sarcopterygians. Indeed, the evolutionary origin of ASPs probably dates back to early vertebrates, around 530 million years ago (Ma) (Sire *et al.*, 2007; Hedges, 2009).

For years, our research group has been interested in the story of SCPPs, more particularly in their relationships, and in the relation between changes in gene structure and putative changes in protein function (Gasse *et al.*, 2015; Al-Hashimi *et al.*, 2010). However, data on ASP expression during amelogenesis accumulated mostly in mammals, in which enamel is prismatic, and this large amount of

information contrasts with the poor knowledge on ASP expression during enamel formation in non-mammalian tetrapods, in which teeth are covered with non-prismatic enamel (Delgado *et al.*, 2005a; Davit-Béal *et al.*, 2007a). Current knowledge on the expression pattern of the six ASPs is briefly summarized below.

AMEL is expressed in rodent ameloblasts from pre-secretory to early maturation stages of amelogenesis (Nanci *et al.*, 1998; Bleicher *et al.*, 1999; Hu *et al.*, 2001a; Torres-Quintana *et al.*, 2005). Studies in pigs and rats revealed transient *AMEL* expression in odontoblasts (Oida *et al.*, 2002; Papagerakis *et al.*, 2003). In amphibians and reptiles, immunohistochemical studies have also shown *AMEL* localization in the odontoblasts and in pulp cells (Wang *et al.*, 2005; Diekwisch *et al.*, 2006; Handrigan & Richman, 2011). Our research group has performed investigations on the spatio-temporal expression of *AMEL* in two non-mammalian tetrapods, a lizard and a salamander we used as representatives of reptilian and amphibian lineages, respectively: in the scincid lizard *Chalcides viridanus* by Delgado *et al.* (2006) and in the salamander *Pleurodeles waltl* by Assaraf-Weill *et al.* (2013). Similarities of spatio-temporal expression of the major EMP gene, *AMEL*, in the ameloblasts of these two species with that described in rodents, led us to conclude that in non-mammalian tetrapod amelogenesis does not differ from that described in mammals in spite of the lack of Tomes' processes and prisms in non-mammalian enamel (Assaraf-Weill *et al.*, 2013).

AMBN expression in rodents extends from the pre-secretory to the maturation stage and transcripts are detected at reduced levels in late maturation stage-ameloblasts (Lee *et al.*, 1996; Nanci *et al.*, 1998; Torres-Quintana *et al.*, 2005). *AMBN* is also transiently expressed by pre-odontoblasts (Bègue-Kirn *et al.*, 1998; Fong *et al.*, 1998). In non-mammals, *AMBN* cDNA has been characterized in a crocodile and a frog (Shintani *et al.*, 2002, 2003), but *AMBN* expression has only been studied in the caiman by immunohistochemistry (Shintani *et al.*, 2006). The distribution of caiman *AMBN* in ameloblasts and in immature enamel matrix was similar to that described in mammals, challenging the hypothesis of *AMBN* relationship with enamel prismatic structure in mammals (Hu *et al.*, 1997; Shintani *et al.*, 2006).

ENAM expression in mice is uninterrupted during pre-secretory, transition and early maturation stages of amelogenesis. *ENAM* expression terminates slightly earlier than *AMEL* in maturation stage-ameloblasts and was never detected in odontoblasts

(Hu *et al.*, 2000, 2001a). *ENAM* cDNA has been characterized in a lizard, a crocodile and a frog (Al-Hashimi *et al.*, 2010). However, to our knowledge, *ENAM* expression was not described in non-mammalian species.

SCPPPQ1 expression has only been studied by immunochemistry in rat teeth (Moffatt *et al.*, 2014), while its cDNA has been characterized in mouse, rat and lizard (Kawasaki, 2009; Kawasaki *et al.*, 2011; Moffatt *et al.*, 2014). To date *ODAM* was not found in reptilian genomes but was identified in amphibians, the coelacanth and in teleost fishes (Kawasaki, 2009; Kawasaki & Amemiya, 2014). *ODAM* expression has been monitored in rat, frog and zebrafish teeth (Park *et al.*, 2007; Kawasaki, 2009). Both SCPPPQ1 and *ODAM* have been identified in late maturation stages of amelogenesis (Moffatt *et al.*, 2008, 2014).

AMTN is expressed late during amelogenesis in the mouse (Iwasaki *et al.*, 2005). The protein localizes to the basal lamina between the ameloblasts and the enamel surface (Moffatt *et al.*, 2006b; Dos Santos Neves *et al.*, 2012) and is believed to be involved in the formation of the final non-prismatic enamel layer (Somogyi-Ganss *et al.*, 2012; Abbarin *et al.*, 2014). In contrast to previous data obtained for *AMEL*, a recent study performed in the lizard *Anolis carolinensis* and in the salamander *Pleurodeles waltl* revealed that both *AMTN* structure and spatio-temporal expression strongly differed from mammalian ones. Indeed, during amelogenesis, *AMTN* is expressed in ameloblasts from secretion stage onwards in both species (Gasse *et al.*, 2015), whereas in mouse it is not detected before the transitional stage (Iwasaki *et al.*, 2005). These findings suggest that *AMTN* was an EMP in the last common ancestor of tetrapods. For an unknown reason, early in mammalian lineage the expression pattern of *AMTN* drastically changed as well as its structure (Gasse *et al.*, 2015; Gasse *et al.*, submit.).

The present study is a logical follow-up of our previous results on the structure and expression pattern of *AMTN* during enamel formation in *A. carolinensis*. As reported above, in mammals the initial expression of *AMEL*, *AMBN* and *ENAM* occurs nearly synchronously in pre-secretory ameloblasts and their expression ends at various steps of the maturation stage. In contrast, *AMTN* expression does not start before the transition stage but carries on until late enamel maturation stage. To date, in non-mammalian tetrapods, data on the gene expression pattern have only been documented for *AMEL* (scincid lizard: Delgado *et al.*, 2006; salamandrid amphibian: Assaraf-Weill *et al.*, 2013) and *AMTN* (dactyloid lizard and salamandrid amphibian:

Gasse *et al.*, 2015). *AMEL* structure and expression were found similar to that described in mammals, whereas those of *AMTN* strongly differ, findings which challenge our previous findings that amelogenesis in non-mammalian tetrapods was similar to that described in mammals (Assaraf-Weill *et al.*, 2013). But what is the structure of *AMBN* and the spatio-temporal expression of *ENAM* and *AMBN* in the lizard, and are they different from those reported in rodents, as revealed recently for *AMTN* (Gasse *et al.*, 2015)? To answer these questions we compared the structure of the four *A. carolinensis* ASP genes with that of representative tetrapod sequences, and compared the expression patterns of these ASP genes during lizard amelogenesis with those previously described in rodents.

MATERIAL AND METHODS

Biological material

Anolis carolinensis specimens were a gift from the pet shop "La Ferme tropicale" in Paris. Several juvenile specimens were sacrificed according to the guidelines of the French Ethics Committee. Lower and upper jaws were dissected. One sample was used for RNA extraction, and the other jaws were divided into two quadrants and fixed for *in situ* hybridization experiments.

mRNA sequences

Most mRNA sequences of tetrapodan ASPs used in our study were available in GenBank either as published mRNA or as computer-predicted sequences from sequenced genomes. They were:

AMEL: *Homo sapiens* (NM_182680.1); *Mus musculus* (NM_001081978.2); *Monodelphis domestica* (XM_003341802.2); *Paleosuchus palpebrosus* (AF095568.1); *A. carolinensis* (XM_008122480.1); *Xenopus tropicalis* (NM_001113681.1); *Pleurodeles waltl* (JX508595.1).

AMBN: *H. sapiens* (NM_016519); *M. musculus* (NM_009664); *M. domestica* (XM_007495519.1); *Caiman crocodilus* (AY043290); *A. carolinensis* (XM_008103732); *X. laevis* (NM_001090020.1).

ENAM: *H. sapiens* (NM_031889); *M. musculus* (NM_017468); *M. domestica* (XM_007495648.1); *Crocodylus niloticus* (GU344683); *A. carolinensis* (GU198361); *X. tropicalis* (EU642606).

AMTN: *H. sapiens* (NM_212557); *M. musculus* (NM_027793.1); *M. domestica* (XM_007495514); *C. crocodilus* (KM069444); *A. carolinensis* (KM069435); *X. tropicalis* (KM069437); *Pleurodeles waltl* (KM069436).

RNA extraction, PCR and probe synthesis

Immediately after dissection the sample was immersed in liquid nitrogen and reduced to powder. Total RNAs were extracted and purified using RNeasy fibrous tissue Mini kit (Qiagen, France). RNAs were converted into cDNA by RT-PCR (RevertAid™ H Minus First Strand cDNA Synthesis Kit; Fermentas, France) using oligo(dT)18 primers.

Rapid Amplification of cDNA End (RACE)-PCR was performed to recover the full length (including 5' and 3' ends) of *AMEL* and *AMBN* mRNA using SMARTer™ RACE cDNA Amplification Kit and Advantage 2 PCR Kit (Clontech) as previously described (Al-Hashimi *et al.*, 2010). Primers were designed from the computer-predicted genomic sequences of these two genes in *A. carolinensis* using Primer3 v.0.4.0 (<http://frodo.wi.mit.edu/primer3/>) (Untergasser *et al.*, 2012).

For probe synthesis, specific primers were designed from the cDNA sequences of the four ASPs:

AMEL : sense 5'-TTTGCTATTCCATTGCCACA-3'; antisense 5'-GGCCATGATTCAAGAGGTGT-3';

AMBN : sense 5'-ATGTTCTGCTCTGCCGCTAT-3'; antisense 5'-GCAGCTCCTTGGTTTGCTAC-3';

ENAM : sense 5'-CAGCCTACATTTCCCCTTCA-3'; antisense 5'-CTGTGCCACTCCATTTTCCTT-3';

AMTN : see Gasse *et al.*, 2015.

cDNA was amplified by PCR (GoTaq polymerase, Promega, France), inserted into a pCRII-TOPO vector containing T7 and SP6 promoters for *in vitro* RNA transcription (TOPO-TA cloning kit; Invitrogen, France), and transformed into competent *E. coli* TOP10F' bacteria. Plasmids were purified (QIAprep Spin MiniPrep Kit; Qiagen, France) and linearized by PCR using M13 universal primers. Antisense RNA probes were synthesized using SP6 and T7 RNA polymerases (Riboprobe Combination System SP6/T7; Promega, France) in the presence of digoxigenin-UTP (Roche, France) and purified (ProbeQuant G-50 micro columns; GE Healthcare, France).

Sequence alignments were performed using the SeaView 4.3.3 software (Gouy *et al.*, 2010).

***In situ* hybridization on sections**

Samples were fixed overnight at 4°C in Formoy's solution (30% formaldehyde 37%, 10% acetic acid and 60% ethanol), and demineralized in 10% acetic acid for 1 month at room temperature under gentle agitation. Samples were then dehydrated in ethanol, shortly immersed in toluene and embedded in Paraplast (Sigma, France). Eight μm -thick sections were obtained with a Leica RM2245 microtome, deposited on Superfrost PLUS slides (Fisher Scientific, France) and dried. They were then dewaxed in toluene, rehydrated through a decreasing series of ethanol then in PBS, treated with proteinase K (0.6 $\mu\text{g}/\text{ml}$) for 5 min at 37°C, rinsed in PBS, post-fixed for 30 min in 4% paraformaldehyde, rinsed again in PBS and then in 2X SCC. The slides were incubated overnight, at 65°C, with the digoxigenin-labeled antisense probe (dilution: 1/150) in the hybridization buffer (50% formamide, 10% dextran sulfate, 1X salt, 1X Denhardt, yeast RNA 1 mg/ml). The following day, the slides were washed three times, at 65°C, in the washing buffer (50% formamide, 1X SCC, 0.1% Tween 20), and rinsed, at room temperature, in the Maleic Acid Buffer Tween (MABT), pH 7.5. Non-specific binding sites were blocked for 2h in a blocking solution (2% blocking reagent, 20% goat serum in MABT). Then, the slides were incubated overnight with the anti-digoxigenin antibody coupled to alkaline phosphatase (dilution: 1/1000) in the blocking solution. The next day, the slides were rinsed four times in MABT, then in NTM (NaCl, TrisHCl, MgCl₂) buffer pH 9.5. The digoxigenin-labeled probes were revealed at 37°C using NBT/BCIP (nitro blue tetrazolium chloride/5-bromo-4-chloro-3-indodyl phosphate). The slides were mounted in Aquatex mounting medium (Merck, France), and photographed (Olympus BX61 microscope).

RESULTS

ASP gene structure in *A. carolinensis* and representative tetrapods (Fig. 1)

In previous publications we provided the full-length cDNA sequence of *ENAM* (4807 base pairs, bp) and *AMTN* (1501 bp), along with gene structure and sequence comparison with other tetrapods (Gasse *et al.*, 2015; Al-Hashimi *et al.*, 2010). In summary: in *ENAM*, structural changes concern (i) the 5' extremity, which displays variations both in the UTR (either one or two non non-coding exons) and in the signal peptide (short in non-mammals, while either short or long in mammals due to the presence of two translation initiation sites), (ii) the presence of an additional coding

exon 8b in sauropsids and non-placental mammals, and (iii) a RGD motif encoded at the end of exon 10 in the only non-mammals (Al-Hashimi *et al.*, 2010). The differences in *AMTN* structure are more obvious as (i) the occurrence of a splice site in mammals led to exon 8 truncation, disappearance of an encoded RGD motif, and recruitment of exon 9, although a transitional condition was observed in the opossum (Gasse *et al.* *subm.*), (ii) the loss of exon 2b in mammals and of exon 2c in placentals, and (iii) the presence of an additional exon 3b in lizards (Gasse *et al.*, 2015).

AMEL and *AMBN* mRNA sequences were available in Genbank as computer-predicted sequences from the sequenced genome of *A. carolinensis*. RACE-PCR provided full-length sequence of these two transcripts. No alternative splice variant was found for either gene in the PCR products. The gene structure was defined using cDNA/gDNA comparisons and sequences were aligned to representative tetrapod sequences.

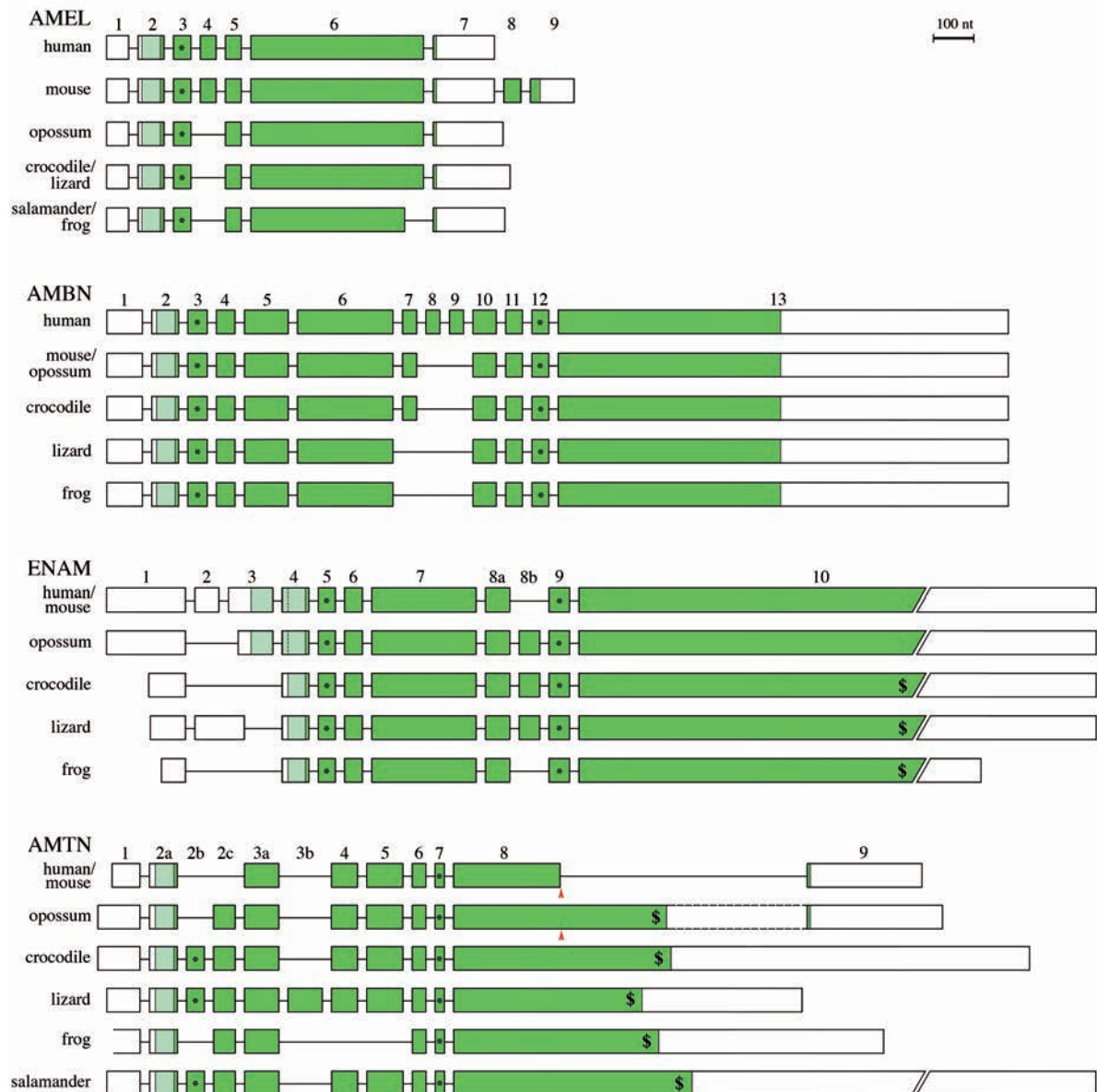


Figure 1. Gene structure of amelogenin (*AMEL*), ameloblastin (*AMBN*), enamelin (*ENAM*) and amelotin (*AMTN*) in representatives of the main tetrapod lineages.

Exons and introns are represented by boxes and lines, respectively (not at scale). Coding regions are in dark green; signal peptide in light green; non-coding regions in white; * : potential phosphorylated Ser residue; \$: RDG motif. In mammalian *AMTN*, a splice site appeared in the middle region of the large exon 8 (red arrowheads), resulting in skipping half of the 3' region of exon 8, and in exon 9 recruitment. In opossum, the two 3' end configurations do occur as alternative transcripts.

The full-length sequence of *AMEL* transcript in *A. carolinensis* comprises 837 bp, including UTRs. The computer-predicted coding sequence and the 3'UTR were confirmed. Exon 1 sequence of 5'UTR was different from the predicted sequence. In this lizard the *AMEL* structure is similar to that previously reported in tetrapods. The

gene lacks exon 4, which is only found in placental *AMEL* and belongs to a minor transcript. *AMEL* cDNA sequence of *A. carolinensis* has been deposited in Genbank (accession number KP792754).

The *AMBN* transcript of *A. carolinensis* is 1787 bp long, including UTRs. The 5' UTR consists of exon 1 (118 bp) and the beginning of exon 2 (32 bp). The gene structure is similar to other tetrapod *AMBN* sequences with the exception of exon 7 that is lacking. The computer-predicted mRNA sequence was quite different from the transcript in (i) being 272 bp shorter, (ii) starting the coding sequence with the ATG in exon 3, and (iii) lacking exon 4 and 6 bp at the end of the 3' UTR. *AMBN* cDNA sequence of *A. carolinensis* has been deposited in Genbank (accession number KP792753).

***In situ* hybridization of ASP genes during amelogenesis (Fig. 2)**

The expression of the four ASP genes was monitored in replacement teeth of *Anolis carolinensis* during four stages of amelogenesis: (1) early enamel matrix deposition, when a thin layer of predentin was deposited by the odontoblasts and when ameloblasts were already differentiated around the tooth tip (Figs 2A, E, I, M); (2) enamel matrix deposition and mineralization, when ameloblasts are active along the upper part of the teeth (Fig. 2B, F, J, N); (3) enamel matrix maturation, when mineralization increases from the tooth tip towards the tooth base, a stage during which enamel proteins are degraded by proteases (Fig. 2C, 2G, 2K, 2O); and (4) when most enamel is matured, tooth nearly to become functional and a new replacement tooth already formed (Fig. 2D, 2H, 2L, 2P).

Amelogenin

AMEL expression is first detected in secretory ameloblasts located at the tooth tip during early enamel matrix deposition (Fig. 2A), then transcripts are identified in the whole ameloblast layer facing enamel matrix deposition (Fig. 2B). During the maturation stage *AMEL* expression is no longer seen in the ameloblasts at the tooth tip but is still present in those towards the tooth base on which enamel matrix is still deposited. At this stage, *AMEL* transcripts are also detected in the odontoblasts facing predentin matrix deposition (Fig. 2C). When tooth is nearly functional *AMEL* transcripts are no longer detected in the reduced ameloblasts, but are still, faintly present in odontoblasts in the upper region of the pulp cavity (Fig. 2D).

Ameloblastin

AMBN expression is detected in the ameloblasts during early enamel matrix deposition (Fig. 2E). As matrix deposition progresses, *AMBN* transcripts are strongly labelled in the well-polarized ameloblasts (Fig. 2F). At the maturation stage, *AMBN* mRNA are strongly labelled in ameloblasts facing both the maturing (tooth tip) and the mineralizing (towards the tooth base) enamel matrix. Expression is also identified in a few odontoblasts located in the upper region of the pulp cavity (Fig. 2G). During late maturation stage prior to tooth eruption *AMBN* is still detected in the ameloblasts near the dentin-enamel junction where enamel maturation is not completed, while it is no longer detected in mature enamel. *AMBN* transcripts are still detected in a few odontoblasts (Fig. 2H).

Enamelin

ENAM transcripts are detected early during tooth development, in ameloblasts at the early secretory stage (Fig. 2I). During enamel matrix deposition, *ENAM* expression is localized in the ameloblasts around the tooth tip (Fig. 2J). Later, *ENAM* mRNA are no longer labelled in the ameloblasts facing the maturing enamel at the tooth tip, while still present in the ameloblasts facing the immature enamel towards the tooth base, near the dentin-enamel junction (Fig. 2K). At the late maturation stage, *ENAM* expression was no longer identified in the reduced ameloblasts, including those located towards the tooth base (Fig. 2L). *ENAM* transcripts were never detected in odontoblasts or other cells of the dental organ.

Amelotin

AMTN is expressed in the ameloblasts as early as a thin layer of predentin is deposited, then the expression is maintained throughout amelogenesis in the whole ameloblast layer until the late maturation stage (Fig. 2M-P). *AMTN* transcripts were never detected in the odontoblasts.

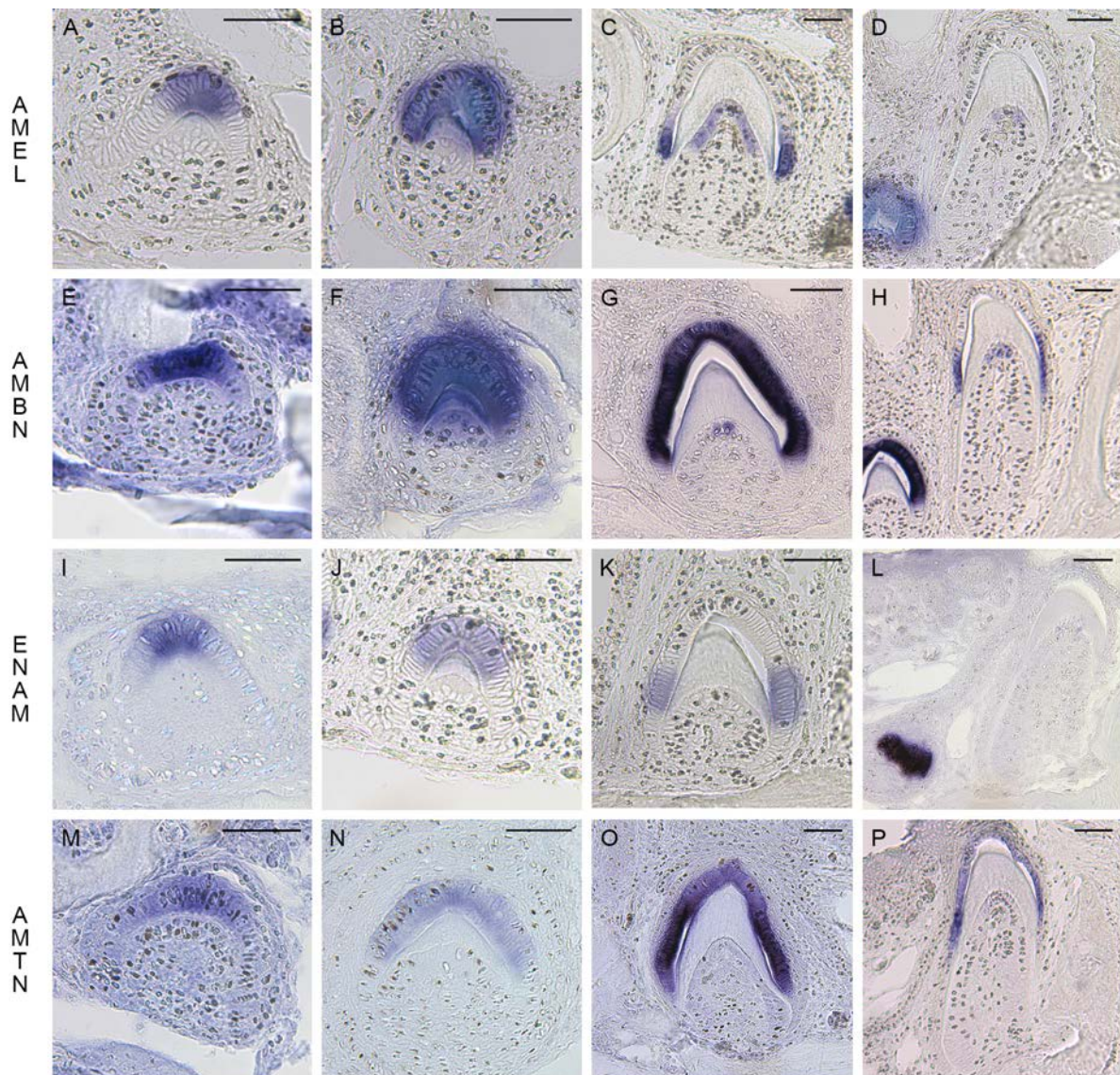


Figure 2. Comparison of *AMEL*, *AMBN*, *ENAM* and *AMTN* expression using *in situ* hybridization on tooth sections during amelogenesis of the iguanid lizard *Anolis carolinensis*. Early secretory stage (A, E, I, M), secretory stage (B, F, J, N), early maturation stage (C, G, K, O) and late maturation stage (D, H, L, P). Scale bars = 50 μ m.

DISCUSSION

In the present study, we provide the full-length cDNA sequences of *AMEL* and *AMBN* in the lizard *A. carolinensis*, which, along with previous data on *ENAM* and *AMTN* sequences obtained in our research group (Gasse *et al.*, 2015; Al-Hashimi *et al.*, 2010), allowed accurate comparison to other ASP cDNA sequences in tetrapods. This is also the first description of *AMBN* and *ENAM* expression during amelogenesis in a non-mammalian species, which allowed to compare the four ASP gene expression one another in the lizard and with published data in the mouse.

Conservation and variations of tetrapodan ASP gene structure

Sequencing mRNA allowed us to clarify the gene structure of *AMEL* and *AMBN* in *A. carolinensis*, especially in the 5' region, including UTRs. This was quite expected because these non-coding regions are highly variable and are difficult to find using computer-prediction, as they are not always conserved through evolution. Our results also confirm previous computer-predicted coding regions.

The 5' UTR of *AMEL* was different from the two predicted sequences in GenBank while the 3'UTR is identical. The structure of the coding sequence, composed of five exons (2, 3, 5, 6 and 7), is similar to that previously reported in other reptiles *Iguana iguana* (Wang *et al.*, 2006) and *Chalcides viridanus* (Delgado *et al.*, 2006), in amphibians *Rana pipiens* (Wang *et al.*, 2005) and *Pleurodeles waltl* (Assaraf-Weill *et al.*, 2013) and in the coelacanth (Kawasaki & Amemiya, 2014). Recently, an additional exon was found between *AMEL* exons 5 and 6 in another iguanid lizard *Ctenosaura similis* (Wang *et al.*, 2012a). This exon was not found in *A. carolinensis*, which indicates independent exon recruitment in the *Ctenosaura* lineage. *AMEL* structure in reptiles differs, however, from that in mammals in lacking exon 4, which confirms the origin of this exon in mammals (Sire *et al.*, 2012). Nevertheless, in mice, exon 4 does not belong to the major *AMEL* transcript, and its role is not understood to date. Moreover, lizard *AMEL* does not exhibit various splicing variants as described in the mouse, with additional exons 8 and 9 encoded in minor isoforms (Sire *et al.*, 2012).

In GenBank, the computer-predicted *AMBN* sequence of *A. carolinensis* was not complete. The full length sequence of *AMBN* mRNA, including both 5' and 3' UTR, indicates the presence of ten exons, a gene structure similar to the one predicted by Kawasaki and Amemiya (2014) in this lizard and to that of the frog sequence (Shintani *et al.*, 2003). The *AMBN* structure of the lizard and frog differs from that described in the crocodile *Caiman crocodilus* (Shintani *et al.*, 2002) and in mammals (Delsuc *et al.*, submit.) in lacking exon 7. This finding suggests that exon 7 probably appeared in an ancestral amniote then was lost in the squamate lineage. In mammals, exon 7 sequence is variable and the encoded peptide does not contain important residues or motifs (Delsuc *et al.*, submit.). Moreover, this exon is known to have duplicated several times, and independently in some mammalian species (Delsuc *et al.*, submit.; MacDougall *et al.*, 2000; Toyosawa *et al.*, 2000). Although alternative splicing was reported in rat, mouse, human and pig *AMBN* (Hu *et al.*, 1997; Simmons

et al., 1998; MacDougall *et al.*, 2000; Lee *et al.*, 2003), only a single lizard transcript was found in the PCR products.

ENAM structure in *A. carolinensis* was previously reported in the course of an evolutionary analysis of *ENAM* in tetrapods (Al-Hashimi *et al.*, 2010). The main differences deduced from the comparison with mammalian *ENAM* structure are the absence of exon 3 and the presence of the additional exon 8b in the lizard sequence. In mammals, the former houses a putative, correct translation initiation site (ATG) suggesting the presence of two alternative (either long or short) signal peptides, the short signal peptide being ancestral. Exon 8b is also found in crocodile and marsupial *ENAM*, but absent in frog and lost in placental mammals. An RGD motif corresponding to a cell attachment sequence is present in the C-terminal region of all non-mammalian tetrapod *ENAM*. This motif is only present in some mammalian species, a finding which suggests limited functional constraints.

AMTN structure has been recently studied in tetrapods, including in *A. carolinensis* (Gasse *et al.*, 2015). In sauropsids and amphibians the gene structure greatly differs from that described in mammals. Compared to the mouse, in lizard *AMTN* displays three additional exons, ends with a large exon 8 that encodes a RGD motif (shorter exon 8 and no RGD in the mouse) and does not possess exon 9 found in all mammalian sequences (Gasse *et al.*, 2012, 2015).

Taken together, our comparative analyses of ASP gene structure in *A. carolinensis* versus mammals, including *Mus musculus*, indicate that *AMEL*, *AMBN* and *ENAM* structure are roughly similar in both lizard and mice, but reveal a drastically different structure for *AMTN*.

Similarities and differences in ASP mRNA expression during amelogenesis in lizard and mouse

During *A. carolinensis* amelogenesis, the expression patterns of *AMEL*, *AMBN*, *ENAM* are spatio-temporally similar to those described in mammals, i.e. *AMEL* and *ENAM* are predominantly expressed in ameloblasts during the secretory and transition stages of amelogenesis, while *AMBN* displays a broader distribution from secretory to late maturation stages, its expression being maintained the longest of the three (Somogyi-Ganss *et al.*, 2012). In contrast to what was previously described during amelogenesis in rodents (Iwasaki *et al.*, 2005), in lizard *AMTN* expression starts nearly simultaneously with the three other ASP genes, in secretory stage ameloblasts

and goes on until late maturation stages, i.e. beyond the expression of the three other ASP genes including *AMBN*. This difference in spatio-temporal expression of *AMTN* could be explained as important changes in the function of the protein during evolution, from a wide expression pattern throughout amelogenesis in non-mammals to more restricted pattern in mammals (Gasse *et al.*, 2015).

During enamel matrix formation in *A. carolinensis*, the expression of two ASP genes *AMEL* and *AMBN* was detected both in odontoblasts and ameloblasts.

In a previous study of *AMEL* expression during amelogenesis in a scincid lizard, *Chalcides viridanus* transcripts were detected in ameloblasts, from the secretory stage until the onset of the maturation process, but were never identified in odontoblasts (Delgado *et al.*, 2006). Different expression patterns of *AMEL* in two lizards showing similar amelogenesis are quite surprising but could be related to the evolutionary distance that separates the two species. Indeed, scincid and dactyloid lineages have diverged around 166 millions years ago (Hedges & Vidal, 2009), a period long enough for some changes to occur in *A. carolinensis* amelogenesis. In the latter the onset of *AMEL* expression in odontoblasts matches the onset of enamel maturation stage, and these cells remain labelled after *AMEL* expression has stopped in late maturation stage ameloblasts. More investigations are however needed to understand the reasons why *AMEL* is expressed by these odontoblasts. How is activated *AMEL*? Is the protein secreted? Does it play a role in reinforcing the dentin-enamel junction? In mammals, several studies have detected *AMEL* expression in pre-odontoblasts and recently differentiated odontoblasts but never in later stages (Oida *et al.*, 2002; Papagerakis *et al.*, 2003). The short distance between ameloblasts and odontoblasts at the onset of *AMEL* expression in mammals compared to the large distance between these cells in lizard, as they are separated by the dentine and enamel layers, could mean that the process leading to *AMEL* activation are different in lizards and mice.

In *A. carolinensis*, *AMBN* transcripts are detected concomitantly with *AMEL* ones in the odontoblasts located at the upper region of the pulp cavity. This suggests that the two genes are activated simultaneously, probably by means of the same process. Here, we have to keep in mind that *AMEL* and *AMBN* are related genes phylogenetically (Sire *et al.*, 2007). In mammals, *AMBN* expression in odontoblasts has also been reported but only during early stages of amelogenesis (Bègue-Kirn *et al.*, 1998; Fong *et al.*, 1998). This again indicates that both genes probably have relationships.

To conclude, our study reveals that among the four ASP genes *AMTN* is the only gene that displays major differences both in structure and in expression pattern during amelogenesis in lizard versus mice. The slight variations observed in the structure of the three other ASP genes seemed to have no consequence in their spatio-temporal expression during amelogenesis. In the amphibian, *Pleurodeles waltl* the expression pattern of *AMTN* during amelogenesis was similar to that described in the lizard (Gasse *et al.*, 2015). A recent study in the mouse suggested that *AMTN* is involved in the establishment of the non-prismatic surface enamel layer and promotes calcium phosphate mineralization (Abbarin *et al.*, 2014). In non-mammalian tetrapods, enamel is not prismatic while it is prismatic in mammals. Nevertheless, the final surface layer that forms concomitantly to *AMTN* expression in the facing ameloblasts is prismless. Given the important role played by the enamel matrix proteins in the high organization of the enamel structure and its mineralization, and given that our comparison of the four lizard and mouse ASP genes pointed to only important changes in *AMTN* structure and expression, we conclude that these *AMTN* changes could be related to the transition between the non-mammalian, non-prismatic enamel, to the prismatic enamel in mammals.

Acknowledgements

We thank K. Daouès, director of "La ferme tropicale", Paris, France, for giving us several *Anolis carolinensis*. We are grateful to Nawfal Al-Hashimi for performing RACE-PCRs. We thank Marie-Claire Lajarille for histological preparations. This project was financially supported by Université Pierre et Marie Curie, CNRS, and ANR-12-BSV7-020, project "Jaws".

DISCUSSION - CONCLUSION

Mes travaux de thèse s'inscrivent dans la continuité des travaux de l'équipe « Evolution et Développement du Squelette (EDS) » sur l'évolution de la famille des SCPP, et plus particulièrement de la sous-famille des SCPP riches en P/Q, impliquées dans la formation de l'émail : amélogénine (Delgado *et al.*, 2005b, 2006a; Assaraf-Weill *et al.*, 2013), énaméline (Al-Hashimi *et al.*, 2009, 2010) et améloblastine (Delsuc *et al.*, submit.).

Lors de ma thèse, je me suis intéressée à un autre membre de cette famille qui avait reçu moins d'attention, l'amélotine. Mes travaux de thèse ont apporté des faits nouveaux concernant l'évolution de ce gène et de la protéine qu'il code. Les principaux résultats concernent, d'une part, la structure du gène et, d'autre part, son patron d'expression. Ensemble, ces deux aspects permettent de retracer une partie de l'histoire évolutive de l'amélotine chez les tétrapodes.

I. Evolution de la structure génique d'amélotine

L'analyse des séquences d'*AMTN* de différentes espèces de tétrapodes a permis de mieux visualiser les changements au niveau de la structure génique au cours de l'évolution des tétrapodes (amphibiens, sauropsides et mammifères).

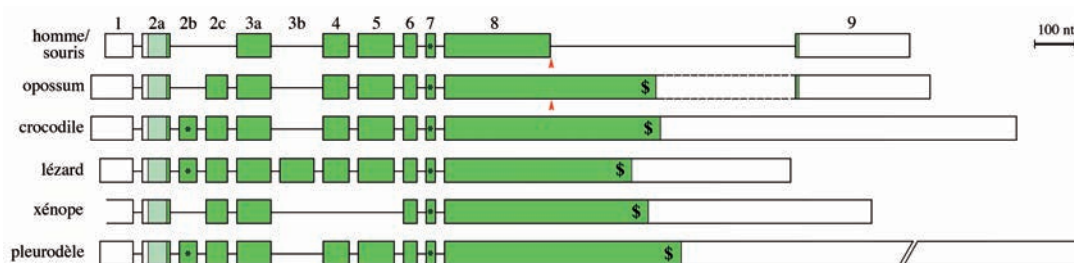


Figure 21. Structure génique d'amélotine chez les tétrapodes. Les rectangles représentent les exons et les lignes les introns (pas à l'échelle). Les régions non traduites sont en blanc, le peptide signal en vert clair et la partie codante en vert foncé. Le site d'épissage intraexonique est représenté par la flèche rouge. *: motif SXE; \$: motif RGD.

Mes travaux de thèse ont mis en évidence, pour la première fois, la présence d'*AMTN* chez un amphibien. La séquence a été identifiée dans le génome et les transcrits de l'amphibien anoure *Xenopus tropicalis*, puis dans le transcriptome du caudate *Pleurodeles waltl*.

L'obtention du gène chez ces deux espèces est importante car la séquence du xénope étant très variée, la comparaison avec celle du pleurodèle a permis de déterminer les exons qui avaient été perdus secondairement dans la lignée des anoures. En effet, le

gène du xénope comporte sept exons, contre dix chez le pleurodèle (Fig. 21). Néanmoins, chez ce dernier certaines limites intron/exon sont incertaines en raison de la grande variabilité des séquences et de l'absence d'un génome séquencé. Le gène de l'ancêtre commun des amphibiens possédait dix exons.

Chez les sauropsides, j'ai pu obtenir la séquence d'*AMTN* de six espèces : un crocodylien et cinq squamates, dont deux serpents. La comparaison de ces séquences a permis de mettre en évidence l'apparition d'un exon supplémentaire uniquement chez les squamates (Fig. 21).

Par ailleurs, l'analyse évolutive chez les mammifères a permis d'étendre les connaissances de 3 à 45 espèces de mammifères, représentant 220 millions d'années d'évolution. L'ensemble des mammifères ne partage pas la même structure génique. Le gène est composé de neuf exons chez les mammifères placentaires. Chez les marsupiaux, la structure génique diffère. En effet, *AMTN* possède un exon supplémentaire (exon 2c) par rapport aux placentaires, mais qui est aussi présent chez les non-mammaliens (Fig. 21). Par ailleurs, j'ai également montré chez l'opossum l'existence d'un épissage alternatif de la partie 3' du gène. Soit le gène finit avec l'exon 9 comme chez les placentaires, ou soit il se termine avec l'exon 8 qui est plus long comme chez les non-mammaliens (Fig. 21).

Chez l'ornithorynque, pour qui nous ne possédons que des données génomiques, l'exon 2c est aussi présent. Dans le génome, les deux fins alternatives du gène sont également possibles mais le transcrite se terminant par un long exon 8 serait très variable et ne coderait pas le motif RGD. Le séquençage de l'ARNm permettrait de savoir si les deux variants sont retrouvés dans les transcrits. Par ailleurs, les exons 4 et 5 sont absents dans le génome. Ici encore le séquençage de l'ARNm permettrait de savoir s'ils sont vraiment absents ou s'il s'agit juste d'un problème d'assemblage du génome.

Chez le dernier ancêtre commun des mammifères, le gène était donc composé de neuf ou dix exons, suivant que l'épissage intraexonique dans l'exon 8 était présent ou non. L'exon 2c a ensuite été perdu et le variant contenant l'exon 9 a été sélectionné pour être le seul conservé chez les placentaires.

II. Evolution de l'expression d'amélotine chez les tétrapodes

Le deuxième aspect de l'histoire évolutive d'*AMTN* étudié au cours de cette thèse concerne son patron d'expression.

Chez la souris, *AMTN* est exprimé assez tardivement dans les améloblastes à partir du stade de transition (situé entre la sécrétion et la maturation), puis pendant la maturation de l'émail (Iwasaki *et al.*, 2005a). L'étude de l'expression d'*AMTN* au cours de l'amélogénèse chez le pleurodèle, avant et après métamorphose, ainsi que chez l'anolé vert, révèle une expression différente de celle qui a été décrite chez la souris. En effet, chez ces espèces *AMTN* est détecté dans les améloblastes dès le début de la formation de l'émail et pendant toute l'amélogénèse jusqu'à la fin de la maturation de l'émail. Ces résultats, partagés par les amphibiens et les dauropsides, suggèrent que l'expression du gène pendant toute l'amélogénèse était la condition ancestrale chez l'ancêtre commun des tétrapodes et elle aurait été modifiée dans la lignée des mammifères.

Lors de cette thèse, l'étude de l'expression d'*AMTN* chez les larves de pleurodèle n'a été réalisée qu'à partir du stade 34 pendant lequel le tissu de recouvrement produit est l'émail et non l'émailloïde, dont la formation se termine lors du stade 33 (Davit-Béal *et al.*, 2007a). Il serait intéressant d'étudier l'expression d'*AMTN* au stade 33 pour voir s'il est aussi synthétisé lors de la formation de l'émailloïde, ce qui pourrait être probable étant donnée son expression précoce lors de l'amélogénèse. Une telle démonstration, associée à celle de l'expression d'*AMBN* et d'*ENAM*, consoliderait les résultats montrant qu'*AMEL* est exprimé à la fin de la formation de l'émailloïde (Assaraf-Weill *et al.*, 2013) et appuierait les hypothèses évolutives concernant la transition entre émailloïde et émail.

L'étude de l'expression du gène chez l'opossum montre un patron d'expression tardif, similaire à celui qui a été observé chez la souris. Ce résultat permet de situer l'événement qui modifia le patron d'expression de "précoce" à "tardif" à au moins 175 millions d'années, date de la divergence des lignées marsupiales et placentaires. Cet événement pourrait être l'apparition d'une mutation sur le promoteur, ou bien un changement dans la régulation par des facteurs de transcription. Il est important de souligner que le changement de patron d'expression a pu être l'événement qui a favorisé le recrutement d'un des variants apparus lors de la formation du site d'épissage intraexonique au milieu de l'exon 8 chez un ancêtre mammalien. Cela signifierait que les contraintes fonctionnelles s'exerçant sur la région C-terminale (contenant notamment le motif RGD) auraient été relâchées lorsque le gène n'a plus été exprimé lors des premiers stades de l'amélogénèse. Le motif RGD de l'*AMTN* aurait donc une importance fonctionnelle chez les tétrapodes non mammaliens.

III. Expression des gènes codant les EMP chez *A. carolinensis*

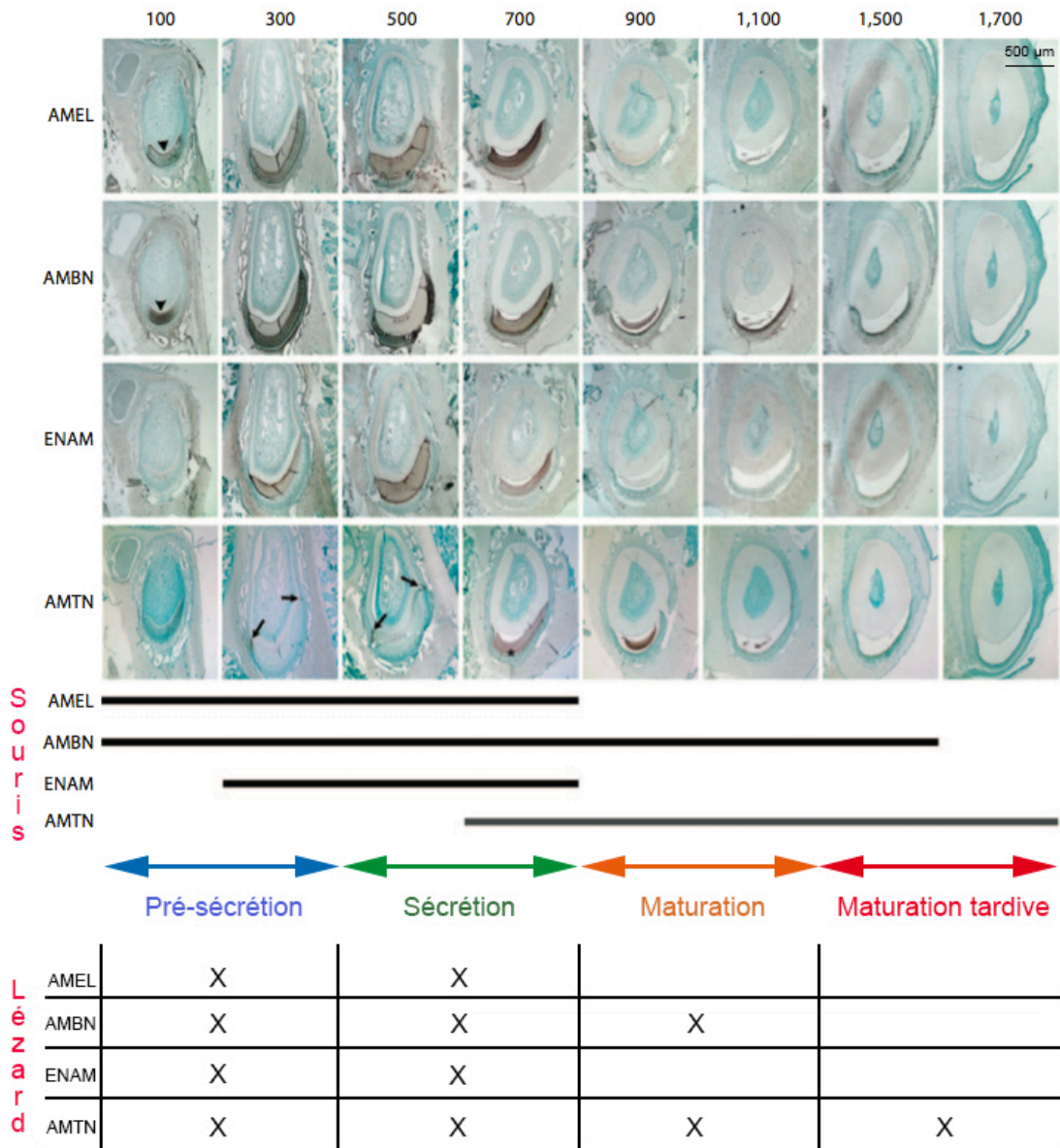


Figure 22. Comparaison des profils d'expression d'*AMEL*, *AMBN*, *ENAM* et *AMTN* chez la souris, en haut (d'après Somogyi-Ganss et al., 2012) et le lézard, en bas.

L'expression des gènes codant les protéines de la matrice de l'émail *AMEL*, *AMBN* et *ENAM* dans les améloblastes pendant l'amélogénèse chez *A. carolinensis* est similaire à celle de leurs orthologues chez la souris (Fig. 22). Toutefois, *AMEL* et *AMBN* sont également exprimés dans les odontoblastes. Chez les rongeurs, l'expression de ces deux gènes a été identifiée dans les pré-odontoblastes ou les jeunes odontoblastes dans quelques études mais jamais à des stades si avancés (Bègue-Kirn *et al.*, 1998;

Fong *et al.*, 1998; Oida *et al.*, 2002; Papagerakis *et al.*, 2003). Le rôle de ces transcrits dans les odontoblastes est inconnu. Il serait intéressant de savoir s'ils sont traduits.

Au cours de l'amélogenèse d'*A. carolinensis*, *AMTN* est le seul gène, sur les quatre, à être exprimé de manière différente de son orthologue mammifère (Fig. 22). Or, la différence principale entre l'émail d'un reptile et celui d'un mammifère est sa structure qui est non prismatique chez les tétrapodes non-mammaliens, prismatique chez les mammifères, une organisation à laquelle sont associés des prolongements de Tomes des améloblastes (Sander, 2000; Line & Novaes, 2005).

Chez les reptiles, en effet, l'émail le plus souvent décrit est non prismatique. Les cristallites sont parallèles entre eux et perpendiculaires à la surface. Ce type d'émail était considéré comme une caractéristique de l'émail reptilien jusqu'à ce que Cooper et Poole (1973) montrent la présence de prismes chez un agamidae (*Uromastix*). De plus, un émail pré-prismatique ou pseudo-prismatique a été décrit chez des reptiles et des thérapsides fossiles (Sahni, 1987; Torii, 1998). Dans l'émail pseudo-prismatique, les cristallites forment des structures en colonne mais sans la présence de substance interprismatique. Il est alors généralement accepté que l'émail pré-prismatique, présent chez les premiers synapsides (groupe frère des sauropsides), représente la transition entre l'émail non prismatique typique des amniotes ancestraux et l'émail prismatique des mammifères, les derniers représentants vivants des synapsides (Osborn & Hillman, 1979). Cet émail pseudo-prismatique serait formé par des améloblastes possédant de petits prolongements de Tomes (Fig. 23) (Line & Novaes, 2005). Toutefois, à ma connaissance il n'existe pas d'images convaincantes de la présence de ces extensions cytoplasmiques chez un reptile actuel.

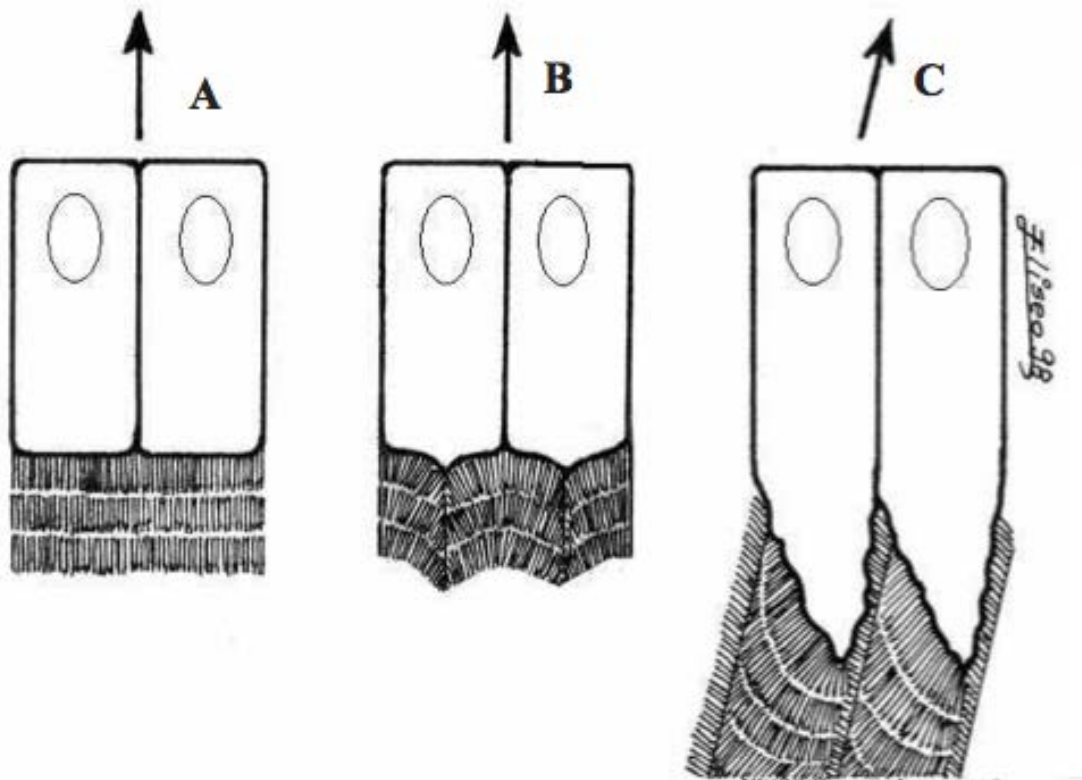


Figure 23. Evolution de la structure de l'émail en fonction de la morphologie des améloblastes. A. Email non prismatique produit par des améloblastes dépourvus de prolongement de Tomes. B. Email pseudo-prismatique produit par des améloblastes possédant de petits prolongements de Tomes. C. Email prismatique produit par des améloblastes possédant des prolongements de Tomes. La flèche indique le mouvement de migration des améloblastes. D'après Line et Novaes, 2005.

Chez tous les mammifères, l'émail est prismatique mais il existe cependant quelques différences, comme par exemple, la présence de tubules dans l'émail des marsupiaux (Sasagawa & Ferguson, 1991).

Les différences que j'ai observées dans la structure et l'expression d'*AMTN* entre lézard/salamandre et mammifères, associées à l'absence *versus* présence de prisme dans l'émail, laissent penser qu'elles pourraient expliquer la transition entre l'émail non prismatique et l'émail prismatique chez les premiers mammifères.

Cette hypothèse est confortée par les travaux de l'équipe de Bernhard Ganss à Toronto qui montrent qu'*AMTN* est impliquée dans la minéralisation de la couche finale d'émail non prismatique chez la souris (Lacruz *et al.*, 2012; Abbarin *et al.*, 2014; Nakayama *et al.*, 2015).

C'est dans cet objectif que nous avons commencé une collaboration avec cette équipe et lancé la génération d'une lignée de souris transgéniques exprimant *AMTN* d'A.

carolinensis sous le contrôle du promoteur d'*AMEL* qui initiera la transcription d'*AMTN* en même temps que celle d'*AMEL*, très tôt lors de l'amélogénèse, comme chez le lézard (promoteur utilisé par Lacruz *et al.*, 2012). Nous voulons savoir si la présence, au début de la formation de la matrice amélaire, des deux motifs fonctionnels (SXE et RGD) de l'*AMTN* de lézard, conduira, chez la souris, à la formation d'un émail non prismatique et à l'absence de prolongements de Tomes. Pour rappel la surexpression, chez la souris, d'*AMTN* de souris sous le contrôle du même promoteur d'*AMEL* a conduit à la formation d'un émail d'épaisseur réduite et possédant une structure profondément désorganisée (Lacruz *et al.*, 2012).

IV. Amélotine était présent chez l'ancêtre commun des sarcoptérygiens

La découverte des gènes des quatre protéines de la matrice de l'émail *AMEL*, *AMBN*, *ENAM* et *AMTN* dans le génome du coelacanthe *Latimeria chalumnae* par Kawasaki et Amemiya (2014) indique que ces gènes étaient déjà présents chez l'ancêtre commun des sarcoptérygiens, il y a 430 millions d'années, et suggère fortement leur présence chez les premiers ostéichthyens. Sire *et al.* (2007) ont émis l'hypothèse, à partir des données connues chez les tétrapodes, selon laquelle *ENAM* serait le gène le plus ancien, et qu'il serait à l'origine des trois autres à la suite de duplications (*ENAM* ->*AMBN* ->*AMEL*). Delgado *et al.* (2001) avaient déjà proposé qu'au moins un des gènes de l'émail (*AMEL*) était présent bien avant l'avènement des vertébrés (>500 Ma), une hypothèse consolidée par le fait qu'une structure très semblable à l'émail a été décrite à la surface du squelette dermique de vertébrés très anciens (Sansom *et al.*, 1992). Il n'est cependant pas possible, à l'heure actuelle, de savoir lequel des quatre gènes est le plus ancien. Nous recherchons actuellement les orthologues de ces gènes dans le génome et le transcriptome de mâchoires des chondrichthyens (requins et raies). Les premiers résultats sont très encourageants.

A ce jour aucun de ces quatre gènes n'a été identifié chez les actinoptérygiens, mais il est probable que la grande distance évolutive rende leur identification difficile. Toutefois de nombreux membres des SCPP ont été identifiés chez les téléostéens mais sans pouvoir définir les orthologies. Seuls *ODAM* et *SPPI* ont pu être identifiés chez ces téléostéens (Kawasaki, 2009), ce qui indique qu'ils étaient présents également chez les premiers ostéichthyens. Ils ont été aussi identifiés chez le coelacanthe (Kawasaki & Amemiya, 2014).

L'absence d'*ODAM* chez les sauropsides, alors qu'il a été identifié chez les téléostéens, le coelacanthe, les amphibiens et les mammifères est un fait intrigant. Quant à *SCPP-PQ1*, il n'a été identifié que chez les mammifères et les reptiles. Chez la souris, *SCPP-PQ1* est exprimé tardivement, comme *ODAM* et *AMTN* (Moffatt *et al.*, 2014). Il est alors envisageable que chez les reptiles, l'absence d'*ODAM* soit compensée par celle de *SCPP-PQ1*. A ce jour, très peu d'études ont été menées sur ce gène récemment découvert. Nous en étudions actuellement l'expression chez le lézard.

V. Amélotine est spécifique de l'émail

Lorsqu'une protéine n'est plus utile, le gène qui la code n'est plus soumis à contrainte fonctionnelle et accumule des mutations. C'est le cas de nombreux gènes impliqués dans la formation de la dentine et l'émail des dents dont la perte des dents est advenue de façon indépendante chez plusieurs espèces (mammifères, oiseaux, tortues,...). Lorsqu'un même gène est invalidé chez plusieurs espèces ayant perdu les dents, cela signifie que la seule fonction de la protéine était liée à la présence de dents car si elle possédait d'autres fonctions le gène resterait fonctionnel.

La recherche de pseudogènes dans les génomes n'est pas chose aisée car ils ne sont pas facilement identifiables par Blast. Nous avons tout de même réussi à identifier certains exons d'*AMTN* dans le génome de deux mammifères dont les dents ne sont pas recouvertes d'émail : le tatou et le paresseux (Gasse *et al.*, 2012), et chez deux sauropsides édentés : le canard et la tortue (Gasse *et al.*, 2015). Ces résultats indiquent qu'*AMTN* est spécifique de l'émail. Ces résultats ont été confirmés par une étude publiée très récemment (Meredith *et al.*, 2014). *AMTN* est le cinquième gène lié à l'émail qui est invalidé dans des espèces édentées ou sans émail. En effet, *AMEL*, *ENAM*, *AMBN*, et *MMP20* sont également invalidés chez ces espèces (Deméré *et al.*, 2008; Sire *et al.*, 2008; Meredith *et al.*, 2009, 2011, 2013, 2014; Al-Hashimi *et al.*, 2010).

VI. Amélotine gène candidat pour l'amélognèse imparfaite ?

L'analyse évolutive d'*AMTN* chez les mammifères a révélé que, chez l'homme, 51 positions sur 209 ont été conservées pendant 220 millions d'années. La conservation de ces acides aminés sur une si longue période indique qu'ils sont cruciaux pour la structure ou la fonction de la protéine. De telles études, réalisées depuis de nombreuses années dans notre équipe EDS, sont utilisées pour prédire et valider des

mutations pouvant causer des maladies génétiques (Delsuc *et al.*, submit.; Delgado *et al.*, 2007; Al-Hashimi *et al.*, 2009; Silvent *et al.*, 2014), un axe de recherche récemment nommé phylomédecine (Kumar *et al.*, 2011). Par exemple, la conservation du site potentiel de phosphorylation SXXE codé par l'exon 7, non seulement chez tous les mammifères, mais aussi chez les non-mammaliens, indique que ce motif est essentiel pour la fonction de la protéine. Cette affirmation est confirmée par les travaux d'Abbarin *et al.* (2014) qui montrent que ce peptide est essentiel à la minéralisation. Il est alors plus que probable qu'une mutation sur ce motif entraînerait un mauvais fonctionnement de la protéine, pouvant causer des défauts dans la formation de l'émail. Cependant, le phénotype des souris knockout pour *AMTN* est discret : seule la dernière couche superficielle d'émail est affectée. Celle-ci présente une légère hypominéralisation et des dépôts de matrice, indiquant une maturation incomplète (Nakayama *et al.*, 2015). Il est donc probable que les défauts d'émail résultant d'une mutation du gène de l'amélotine chez l'homme soient discrets et n'aient pas encore été rapportés. De plus, si le mode de transmission est récessif, la fréquence d'apparition de la maladie sera plus faible, comme le démontre la découverte très récente de la première mutation dans *AMBN* causant l'amélogénèse imparfaite (Poulter *et al.*, 2014).

RÉFÉRENCES BIBLIOGRAPHIQUES

- Abbarin, N., Miguel, S.S., Holcroft, J., Iwasaki, K. & Ganss, B. 2014. The Enamel Protein Amelotin is a Promoter of Hydroxyapatite Mineralization. *J. Bone Miner. Res.*, doi: 10.1002/jbmr.2411.
- Aldred, M., Savarirayan, R. & Crawford, P. 2003. Amelogenesis imperfecta: a classification and catalogue for the 21st century. *Oral Dis.* **9**: 19–23.
- Al-Hashimi, N., Lafont, A.-G., Delgado, S., Kawasaki, K. & Sire, J.-Y. 2010. The Enamelin Genes in Lizard, Crocodile, and Frog and the Pseudogene in the Chicken Provide New Insights on Enamelin Evolution in Tetrapods. *Mol. Biol. Evol.* **27**: 2078–2094.
- Al-Hashimi, N., Sire, J.-Y. & Delgado, S. 2009. Evolutionary Analysis of Mammalian Enamelin, The Largest Enamel Protein, Supports a Crucial Role for the 32-kDa Peptide and Reveals Selective Adaptation in Rodents and Primates. *J. Mol. Evol.* **69**: 635–656.
- Assaraf-Weill, N., Gasse, B., Al-Hashimi, N., Delgado, S., Sire, J.-Y. & Davit-Béal, T. 2013. Conservation of Amelogenin Gene Expression During Tetrapod Evolution. *J. Exp. Zool. B Mol. Dev. Evol.* **320**: 200–209.
- Baba, O., Takahashi, N., Terashima, T., Li, W., DenBesten, P.K. & Takano, Y. 2002. Expression of Alternatively Spliced RNA Transcripts of Amelogenin Gene Exons 8 and 9 and Its End Products in the Rat Incisor. *J. Histochem. Cytochem.* **50**: 1229–1236.
- Bartlett, J.D. 2013. Dental Enamel Development: Proteinases and Their Enamel Matrix Substrates. *ISRN Dent.* **2013**.
- Bartlett, J.D., Ball, R.L., Kawai, T., Tye, C.E., Tsuchiya, M. & Simmer, J.P. 2006. Origin, Splicing, and Expression of Rodent Amelogenin Exon 8. *J. Dent. Res.* **85**: 894–899.
- Bègue-Kirn, C., Krebsbach, P.H., Bartlett, J.D. & Butler, W.T. 1998. Dentin sialoprotein, dentin phosphoprotein, enamelysin and ameloblastin, tooth-specific molecules that are distinctively expressed during murine dental differentiation. *Eur. J. Oral Sci.* **106**: 963–970.
- Bertrand, S., Fuentealba, J., Aze, A., Hudson, C., Yasuo, H., Torrejon, M., *et al.* 2013. A dynamic history of gene duplications and losses characterizes the evolution of the SPARC family in eumetazoans. *Proc. R. Soc. Lond. B Biol. Sci.* **280**: 20122963.
- Bleicher, F., Couble, M.L., Farges, J.C., Couble, P. & Magloire, H. 1999. Sequential expression of matrix protein genes in developing rat teeth. *Matrix Biol.* **18**: 133–143.
- Bonass, W.A., Robinson, P.A., Kirkham, J., Shore, R.C. & Robinson, C. 1994. Molecular Cloning and DNA Sequence of Rat Amelogenin and a Comparative Analysis of Mammalian Amelogenin Protein Sequence Divergence. *Biochem. Biophys. Res. Commun.* **198**: 755–763.

- Bosshardt, D.D. & Nanci, A. 1998. Immunolocalization of Epithelial and Mesenchymal Matrix Constituents in Association with Inner Enamel Epithelial Cells. *J. Histochem. Cytochem.* **46**: 135–142.
- Bouropoulos, N. & Moradian-Oldak, J. 2004. Induction of Apatite by the Cooperative Effect of Amelogenin and the 32-kDa Enamelin. *J. Dent. Res.* **83**: 278–282.
- Cannatella, D.C., Vieitesb, D.R., Zhangb, P., Wakeb, M.H. & Wakeb, D.B. 2009. Amphibians (Lissamphibia). In: *The timetree of life*, pp. 353–356. Hedges SB and Kumar S.
- Catalano-Sherman, J., Palmon, A., Burstein, Y. & Deutsch, D. 1993. Amino acid sequence of a major human amelogenin protein employing Edman degradation and cDNA sequencing. *J. Dent. Res.* **72**: 1566–1572.
- Černý, R., Slaby, I., Hammarström, L. & Wurtz, T. 1996. A novel gene expressed in rat ameloblasts codes for proteins with cell binding domains. *J. Bone Miner. Res.* **11**: 883–891.
- Chen, E., Piddington, R., Decker, S., Park, J., Yuan, Z.-A., Abrams, W.R., *et al.* 1994. Regulation of amelogenin gene expression during tooth development. *Dev. Dyn.* **199**: 189–198.
- Chiari, Y., Cahais, V., Galtier, N. & Delsuc, F. 2012. Phylogenomic analyses support the position of turtles as the sister group of birds and crocodiles (Archosauria). *BMC Biol.* **10**: 65.
- Crivelini, M.M., Felipini, R.C., Miyahara, G.I. & de Sousa, S.C.O.M. 2012. Expression of odontogenic ameloblast-associated protein, amelotin, ameloblastin, and amelogenin in odontogenic tumors: immunohistochemical analysis and pathogenetic considerations. *J. Oral Pathol. Med.* **41**: 272–280.
- Davit-Béal, T., Allizard, F. & Sire, J.-Y. 2006. Morphological variations in a tooth family through ontogeny in *Pleurodeles waltl* (Lissamphibia, Caudata). *J. Morphol.* **267**: 1048–1065.
- Davit-Béal, T., Allizard, F. & Sire, J.-Y. 2007a. Enameloid/enamel transition through successive tooth replacements in *Pleurodeles waltl* (Lissamphibia, Caudata). *Cell Tissue Res.* **328**: 167–183.
- Davit-Béal, T., Chisaka, H., Delgado, S. & Sire, J.-Y. 2007b. Amphibian teeth: current knowledge, unanswered questions, and some directions for future research. *Biol. Rev.* **82**: 49–81.
- Davit-Béal, T., Tucker, A.S. & Sire, J.-Y. 2009. Loss of teeth and enamel in tetrapods: fossil record, genetic data and morphological adaptations. *J. Anat.* **214**: 477–501.
- Dehal, P. & Boore, J.L. 2005. Two Rounds of Whole Genome Duplication in the Ancestral Vertebrate. *PLoS Biol* **3**: e314.

- Delgado, S., Couble, M.-L., Magloire, H. & Sire, J.-Y. 2006. Cloning, Sequencing, and Expression of the Amelogenin Gene in Two Scincid Lizards. *J. Dent. Res.* **85**: 138–143.
- Delgado, S., Davit-Béal, T., Allizard, F. & Sire, J.-Y. 2005a. Tooth development in a scincid lizard, *Chalcides viridanus* (Squamata), with particular attention to enamel formation. *Cell Tissue Res.* **319**: 71–89.
- Delgado, S., Girondot, M. & Sire, J.-Y. 2005b. Molecular Evolution of Amelogenin in Mammals. *J. Mol. Evol.* **60**: 12–30.
- Delgado, S., Ishiyama, M. & Sire, J.-Y. 2007. Validation of amelogenesis imperfecta inferred from amelogenin evolution. *J. Dent. Res.* **86**: 326–330.
- Delport, W., Poon, A.F.Y., Frost, S.D.W. & Kosakovsky Pond, S.L. 2010. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinforma. Oxf. Engl.* **26**: 2455–2457.
- Delsuc, F., Gasse, B. & Sire, J.-Y. submit. Evolutionary analysis of selective constraints identifies ameloblastin (AMBN) as a potential candidate for amelogenesis imperfecta. *BMC Evol. Biol.*
- Deméré, T.A., McGowen, M.R., Berta, A. & Gatesy, J. 2008. Morphological and Molecular Evidence for a Stepwise Evolutionary Transition from Teeth to Baleen in Mysticete Whales. *Syst. Biol.* **57**: 15–37.
- Deutsch, D., Leiser, Y., Shay, B., Fermon, E., Taylor, A., Rosenfeld, E., *et al.* 2002. The human tuftelin gene and the expression of tuftelin in mineralizing and nonmineralizing tissues. *Connect. Tissue Res.* **43**: 425–434.
- Dey, R., Son, H.-H. & Cho, M.-I. 2001. Isolation and partial sequencing of potentially odontoblast-specific/enriched rat cDNA clones obtained by suppression subtractive hybridization. *Arch. Oral Biol.* **46**: 249–260.
- Dhamija, S. & Krebsbach, P.H. 2001. Role of *Cbfa1* in Ameloblastin Gene Transcription. *J. Biol. Chem.* **276**: 35159–35164.
- Dhamija, S., Liu, Y., Yamada, Y., Snead, M.L. & Krebsbach, P.H. 1999. Cloning and Characterization of the Murine Ameloblastin Promoter. *J. Biol. Chem.* **274**: 20738–20743.
- Diekwisch, T.G.H., Berman, B.J., Anderton, X., Gurinsky, B., Ortega, A.J., Satchell, P.G., *et al.* 2002. Membranes, minerals, and proteins of developing vertebrate enamel. *Microsc. Res. Tech.* **59**: 373–395.
- Diekwisch, T.G.H., Jin, T., Wang, X., Ito, Y., Schmidt, M., Druzinsky, R., *et al.* 2009. Amelogenin Evolution and Tetrapod Enamel Structure. *Front. Oral Biol.* **13**: 74–79.
- Diekwisch, T.G.H., Wang, X., Fan, J.L., Ito, Y. & Luan, X. 2006. Expression and characterization of a *Rana pipiens* amelogenin protein. *Eur. J. Oral Sci.* **114**: 86–92.

- Dohi, N., Murakami, C., Tanabe, T., Yamakoshi, Y., Fukae, M., Yamamoto, Y., *et al.* 1998. Immunocytochemical and immunochemical study of enamelin, using antibodies against porcine 89-kDa enamelin and its N-terminal synthetic peptide, in porcine tooth germs. *Cell Tissue Res.* **293**: 313–325.
- Donoghue, P.C.J. & Sansom, I.J. 2002. Origin and early evolution of vertebrate skeletonization. *Microsc. Res. Tech.* **59**: 352–372.
- Donoghue, P.C.J., Sansom, I.J. & Downs, J.P. 2006. Early evolution of vertebrate skeletal tissues and cellular interactions, and the canalization of skeletal development. *J. Exp. Zool. B Mol. Dev. Evol.* **306B**: 278–294.
- Dos Santos Neves, J., Wazen, R.M., Kuroda, S., Francis Zalzal, S., Moffatt, P. & Nanci, A. 2012. Odontogenic ameloblast-associated and amelotin are novel basal lamina components. *Histochem. Cell Biol.* **137**: 329–338.
- Du, C., Falini, G., Fermani, S., Abbott, C. & Moradian-Oldak, J. 2005. Supramolecular assembly of amelogenin nanospheres into birefringent microribbons. *Science* **307**: 1450–1454.
- Fan, D., Du, C., Sun, Z., Lakshminarayanan, R. & Moradian-Oldak, J. 2009. In vitro study on the interaction between the 32 kDa enamelin and amelogenin. *J. Struct. Biol.* **166**: 88–94.
- Fan, D., Lakshminarayanan, R. & Moradian-Oldak, J. 2008. The 32 kDa enamelin undergoes conformational transitions upon calcium binding. *J. Struct. Biol.* **163**: 109–115.
- Fincham, A.G. & Moradian-Oldak, J. 1995. Recent advances in amelogenin biochemistry. *Connect. Tissue Res.* **32**: 119–124.
- Fincham, A.G., Moradian-Oldak, J., Diekwisch, T.G.H., Lyaruu, D.M., Wright, J.T., Bringas Jr., P., *et al.* 1995. Evidence for Amelogenin “Nanospheres” as Functional Components of Secretory-Stage Enamel Matrix. *J. Struct. Biol.* **115**: 50–59.
- Fincham, A.G., Moradian-Oldak, J. & Simmer, J.P. 1999. The Structural Biology of the Developing Dental Enamel Matrix. *J. Struct. Biol.* **126**: 270–299.
- Fisher, L.W. & Fedarko, N.S. 2003. Six genes expressed in bones and teeth encode the current members of the SIBLING family of proteins. *Connect. Tissue Res.* **44 Suppl 1**: 33–40.
- Fisher, L.W., Torchia, D.A., Fohr, B., Young, M.F. & Fedarko, N.S. 2001. Flexible Structures of SIBLING Proteins, Bone Sialoprotein, and Osteopontin. *Biochem. Biophys. Res. Commun.* **280**: 460–465.
- Fong, C.D., Cerný, R., Hammarström, L. & Slaby, I. 1998. Sequential expression of an amelogenin gene in mesenchymal and epithelial cells during odontogenesis in rats. *Eur. J. Oral Sci.* **106 Suppl 1**: 324–330.

- Fong, C.D., Slaby, I. & Hammarström, L. 1996. Amelin: An enamel-related protein, transcribed in the cells of epithelial root sheath. *J. Bone Miner. Res.* **11**: 892–898.
- Fukae, M., Tanabe, T., Murakami, C., Dohi, N., Uchida, T. & Shimizu, M. 1996. Primary structure of the porcine 89-kDa enamelin. *Adv. Dent. Res.* **10**: 111–118.
- Fukumoto, S., Kiba, T., Hall, B., Iehara, N., Nakamura, T., Longenecker, G., *et al.* 2004. Ameloblastin is a cell adhesion molecule required for maintaining the differentiation state of ameloblasts. *J. Cell Biol.* **167**: 973–983.
- Fukumoto, S., Yamada, A., Nonaka, K. & Yamada, Y. 2005. Essential Roles of Ameloblastin in Maintaining Ameloblast Differentiation and Enamel Formation. *Cells Tissues Organs* **181**: 189–195.
- Gallien, L. & Durocher, M. 1957. Table chronologique du développement chez *Pleurodeles waltlii* Michah. *Bull Biol Fr Belg* **91**: 97–114.
- Ganss, B. & Abbarin, N. 2014. Maturation and beyond: proteins in the developmental continuum from enamel epithelium to junctional epithelium. *Craniofacial Biol.* **5**: 371.
- Gao, Y., Wang, W., Sun, Y., Zhang, J., Li, D., Wei, Y., *et al.* 2010. Distribution of Amelotin in Mouse Tooth Development. *Anat. Rec. Adv. Integr. Anat. Evol. Biol.* **293**: 135–140.
- Gasse, B., Chiari, Y., Silvent, J., Davit-Béal, T. & Sire, J.-Y. 2015. Amelotin: an enamel matrix protein that experienced distinct evolutionary histories in amphibians, sauropsids and mammals. *BMC Evol. Biol.* doi: 10.1186/s12862-015-0329-x.
- Gasse, B., Silvent, J. & Sire, J.-Y. 2012. Evolutionary Analysis Suggests That AMTN is Enamel-specific and a Candidate for AI. *J. Dent. Res.* **91**: 1085–1089.
- Gayral, P., Weinert, L., Chiari, Y., Tsagkogeorga, G., Ballenghien, M. & Galtier, N. 2011. Next-generation sequencing of transcriptomes: a guide to RNA isolation in nonmodel animals. *Mol. Ecol. Resour.* **11**: 650–661.
- Ge, Y., Kong, Z., Guo, Y., Tang, W., Guo, W. & Tian, W. 2013. The role of odontogenic genes and proteins in tooth epithelial cells and their niche cells during rat tooth root development. *Arch. Oral Biol.* **58**: 151–159.
- Gibson, C.W., Yuan, Z.-A., Hall, B., Longenecker, G., Chen, E., Thyagarajan, T., *et al.* 2001. Amelogenin-deficient Mice Display an Amelogenesis Imperfecta Phenotype. *J. Biol. Chem.* **276**: 31871–31875.
- Gopinathan, G., Jin, T., Liu, M., Li, S., Atsawasuwan, P., Galang, M.-T., *et al.* 2014. The expanded amelogenin polyproline region preferentially binds to apatite versus carbonate and promotes apatite crystal elongation. *Craniofacial Biol.* **5**: 430.

- Gouy, M., Guindon, S. & Gascuel, O. 2010. SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. *Mol. Biol. Evol.* **27**: 221–224.
- Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., *et al.* 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**: 644–652.
- Guo, J., Lyaruu, D.M., Takano, Y., Gibson, C.W., DenBesten, P.K. & Bronckers, A.L.J.J. 2014. Amelogenins as Potential Buffers during Secretory-stage Amelogenesis. *J. Dent. Res.* 0022034514564186.
- Guo, T., Rudnick, P.A., Wang, W., Lee, C.S., DeVoe, D.L. & Balgley, B.M. 2006. Characterization of the Human Salivary Proteome by Capillary Isoelectric Focusing/Nanoreversed-Phase Liquid Chromatography Coupled with ESI-Tandem MS. *J. Proteome Res.* **5**: 1469–1478.
- Handrigan, G.R. & Richman, J.M. 2011. Unicuspid and bicuspid tooth crown formation in squamates. *J. Exp. Zool. B Mol. Dev. Evol.* **316B**: 598–608.
- Hao, J., He, G., Narayanan, K., Zou, B., Lin, L., Muni, T., *et al.* 2005. Identification of differentially expressed cDNA transcripts from a rat odontoblast cell line. *Bone* **37**: 578–588.
- Hart, P.S., Aldred, M.J., Crawford, P.J.M., Wright, N.J., Hart, T.C. & Wright, J.T. 2002. Amelogenesis imperfecta phenotype–genotype correlations with two amelogenin gene mutations. *Arch. Oral Biol.* **47**: 261–265.
- Hedges, S.B. 2009. Vertebrates (Vertebrata). In: *Timetree of Life*, pp. 309–314. Hedges SB and Kumar S.
- Hedges, S.B. & Vidal, N. 2009. Lizards, snakes, and amphisbaenians (Squamata). In: *The Timetree of Life*, pp. 383–389. Oxford University Press.
- Herold, R., Rosenbloom, J. & Granovsky, M. 1989. Phylogenetic distribution of enamel proteins: Immunohistochemical localization with monoclonal antibodies indicates the evolutionary appearance of enamelin prior to amelogenins. *Calcif. Tissue Int.* **45**: 88–94.
- Hirayama, K., Miyasho, T., Ohmachi, T., Watanabe, T., Yokota, H. & Taniyama, H. 2010. Biochemical and Immunohistochemical Characterization of the Amyloid in Canine Amyloid-Producing Odontogenic Tumor. *Vet. Pathol. Online* **47**: 915–922.
- Holcroft, J. & Ganss, B. 2011. Identification of Amelotin- and ODAM-interacting enamel matrix proteins using the yeast two-hybrid system. *Eur. J. Oral Sci.* **119 Suppl 1**: 301–306.
- Hou, S., Landego, I., Jayachandran, N., Miller, A., Gibson, I.W., Ambrose, C., *et al.* 2014. Follicular dendritic cell secreted protein FDC-SP controls IgA production. *Mucosal Immunol.* **7**: 948–957.

- Hu, C.-C., Bartlett, J.D., Zhang, C.H., Qian, Q., Ryu, O.H. & Simmer, J.P. 1996a. Cloning, cDNA sequence, and alternative splicing of porcine amelogenin mRNAs. *J. Dent. Res.* **75**: 1735–1741.
- Hu, C.C., Fukae, M., Uchida, T., Qian, Q., Zhang, C.H., Ryu, O.H., *et al.* 1997. Sheathlin: cloning, cDNA/polypeptide sequences, and immunolocalization of porcine enamel sheath proteins. *J. Dent. Res.* **76**: 648–657.
- Hu, C.-C., Hart, T.C., Dupont, B.R., Chen, J.J., Sun, X., Qian, Q., *et al.* 2000. Cloning Human Enamelin cDNA, Chromosomal Localization, and Analysis of Expression during Tooth Development. *J. Dent. Res.* **79**: 912–919.
- Hu, C.-C., Zhang, C., Qian, Q., Ryu, O.H., Moradian-Oldak, J., Fincham, A.G., *et al.* 1996b. Cloning, DNA sequence, and alternative splicing of opossum amelogenin mRNAs. *J. Dent. Res.* **75**: 1728–1734.
- Hu, J.C.-C., Hu, Y., Smith, C.E., McKee, M.D., Wright, J.T., Yamakoshi, Y., *et al.* 2008. Enamel Defects and Ameloblast-specific Expression in Enam Knock-out/lacZ Knock-in Mice. *J. Biol. Chem.* **283**: 10858–10871.
- Hu, J.C.-C., Sun, X., Zhang, C. & Simmer, J.P. 2001a. A comparison of enamel and amelogenin expression in developing mouse molars. *Eur. J. Oral Sci.* **109**: 125–132.
- Hu, J.C.-C. & Yamakoshi, Y. 2003. Enamelin and Autosomal-dominant Amelogenesis Imperfecta. *Crit. Rev. Oral Biol. Med.* **14**: 387–398.
- Hu, J.C.-C., Yamakoshi, Y., Yamakoshi, F., Krebsbach, P.H. & Simmer, J.P. 2005. Proteomics and Genetics of Dental Enamel. *Cells Tissues Organs* **181**: 219–231.
- Hu, J.C.-C., Zhang, C.H., Yang, Y., Kärrman-Mårdh, C., Forsman-Semb, K. & Simmer, J.P. 2001b. Cloning and Characterization of the Mouse and Human Enamelin Genes. *J. Dent. Res.* **80**: 898–902.
- Huq, N.L., Cross, K.J., Ung, M. & Reynolds, E.C. 2005. A review of protein structure and gene organisation for proteins associated with mineralised tissue and calcium phosphate stabilisation encoded on human chromosome 4. *Arch. Oral Biol.* **50**: 599–609.
- Iizuka, S., Kudo, Y., Yoshida, M., Tsunematsu, T., Yoshiko, Y., Uchida, T., *et al.* 2011. Ameloblastin Regulates Osteogenic Differentiation by Inhibiting Src Kinase via Cross Talk between Integrin β 1 and CD63. *Mol. Cell. Biol.* **31**: 783–792.
- Ishikawa, H.O., Xu, A., Ogura, E., Manning, G. & Irvine, K.D. 2012. The Raine Syndrome Protein FAM20C Is a Golgi Kinase That Phosphorylates Bio-Mineralization Proteins. *PLoS ONE* **7**: e42988.
- Ishiyama, M., Mikami, M., Shimokawa, H. & Oida, S. 1998. Amelogenin protein in tooth germs of the snake *Elaphe quadrivirgata*, immunohistochemistry, cloning and cDNA sequence. *Arch. Histol. Cytol.* **61**: 467–474.

- Iwasaki, K., Bajenova, E., Somogyi-Ganss, E., Miller, M., Nguyen, V., Nourkeyhani, H., *et al.* 2005. Amelotin--a Novel Secreted, Ameloblast-specific Protein. *J. Dent. Res.* **84**: 1127–1132.
- Iwata, T., Yamakoshi, Y., Hu, J.C.-C., Ishikawa, I., Bartlett, J.D., Krebsbach, P.H., *et al.* 2007. Processing of Ameloblastin by MMP-20. *J. Dent. Res.* **86**: 153–157.
- Jue, S.-S., Kim, J.-Y., Na, S.-H., Jeon, K.-D., Bang, H.-J., Park, J.H., *et al.* 2014. Localization of ODAM, PCNA, and CK14 in regenerating junctional epithelium during orthodontic tooth movement in rats. *Angle Orthod.* **84**: 534–540.
- Kallenbach, E. 1973. The fine structure of Tomes' process of rat incisor ameloblasts and its relationship to the elaboration of enamel. *Tissue Cell* **5**: 501–524.
- Kawasaki, K. 2011. The SCPP Gene Family and the Complexity of Hard Tissues in Vertebrates. *Cells Tissues Organs* **194**: 108–112.
- Kawasaki, K. 2009. The SCPP gene repertoire in bony vertebrates and graded differences in mineralized tissues. *Dev. Genes Evol.* **219**: 147–157.
- Kawasaki, K. & Amemiya, C.T. 2014. SCPP genes in the coelacanth: Tissue mineralization genes shared by sarcopterygians. *J. Exp. Zool. B Mol. Dev. Evol.* **322**: 390–402.
- Kawasaki, K., Buchanan, A.V. & Weiss, K.M. 2007. Gene Duplication and the Evolution of Vertebrate Skeletal Mineralization. *Cells Tissues Organs* **186**: 7–24.
- Kawasaki, K., Lafont, A.-G. & Sire, J.-Y. 2011. The Evolution of Milk Casein Genes from Tooth Genes before the Origin of Mammals. *Mol. Biol. Evol.* **28**: 2053–2061.
- Kawasaki, K., Suzuki, T. & Weiss, K.M. 2004. Genetic basis for the evolution of vertebrate mineralized tissue. *Proc. Natl. Acad. Sci. U. S. A.* **101**: 11356–11361.
- Kawasaki, K. & Weiss, K.M. 2006. Evolutionary genetics of vertebrate tissue mineralization: the origin and evolution of the secretory calcium-binding phosphoprotein family. *J. Exp. Zool. B Mol. Dev. Evol.* **306B**: 295–316.
- Kawasaki, K. & Weiss, K.M. 2003. Mineralized tissue and vertebrate evolution: the secretory calcium-binding phosphoprotein gene cluster. *Proc. Natl. Acad. Sci. U. S. A.* **100**: 4060–4065.
- Kestler, D., James S. Foster, Charles T. Bruker, John W. Prenshaw, Stephen J. Kennel, Jonathan S. Wall, *et al.* 2011. ODAM Expression Inhibits Human Breast Cancer Tumorigenesis. *Breast Cancer Basic Clin. Res.* **73**.
- Kestler, D.P., Foster, J.S., Macy, S.D., Murphy, C.L., Weiss, D.T. & Solomon, A. 2008. Expression of Odontogenic Ameloblast-Associated Protein (ODAM) in Dental and Other Epithelial Neoplasms. *Mol. Med.* **14**: 318–326.

- Kitagawa, M., Kitagawa, S., Nagasaki, A., Miyauchi, M., Uchida, T. & Takata, T. 2011. Synthetic ameloblastin peptide stimulates differentiation of human periodontal ligament cells. *Arch. Oral Biol.* **56**: 374–379.
- Kobayashi, K., Yamakoshi, Y., Hu, J.C.-C., Gomi, K., Arai, T., Fukae, M., *et al.* 2007. Splicing Determines the Glycosylation State of Ameloblastin. *J. Dent. Res.* **86**: 962–967.
- Kosakovsky Pond, S.L. & Frost, S.D.W. 2005. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinforma. Oxf. Engl.* **21**: 2531–2533.
- Kozawa, Y., Iwasa, Y. & Mishima, H. 1998. Degeneration of tooth germ in the developing dentition of the gray short-tailed opossum (*Monodelphis domestica*). *Eur. J. Oral Sci.* **106 Suppl 1**: 509–512.
- Krebsbach, P.H., Lee, S.K., Matsuki, Y., Kozak, C.A., Yamada, K.M. & Yamada, Y. 1996. Full-length Sequence, Localization, and Chromosomal Mapping of Ameloblastin A NOVEL TOOTH-SPECIFIC GENE. *J. Biol. Chem.* **271**: 4431–4435.
- Kumar, S., Dudley, J.T., Filipski, A. & Liu, L. 2011. Phylomedicine: an evolutionary telescope to explore and diagnose the universe of disease mutations. *Trends Genet.* **27**: 377–386.
- Lacruz, R.S., Nakayama, Y., Holcroft, J., Nguyen, V., Somogyi-Ganss, E., Snead, M.L., *et al.* 2012. Targeted Overexpression of Amelotin Disrupts the Microstructure of Dental Enamel. *PLoS ONE* **7**: e35200.
- Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**: R25.
- Lau, E.C., Simmer, J.P., Bringas, P., Hsu, D.D., Hu, C.C., Zeichner-David, M., *et al.* 1992. Alternative splicing of the mouse amelogenin primary RNA transcript contributes to amelogenin heterogeneity. *Biochem. Biophys. Res. Commun.* **188**: 1253–1260.
- Lee, H.-K., Lee, D.-S., Ryoo, H.-M., Park, J.-T., Park, S.-J., Bae, H.-S., *et al.* 2010. The odontogenic ameloblast-associated protein (ODAM) cooperates with RUNX2 and modulates enamel mineralization via regulation of MMP-20. *J. Cell. Biochem.* **111**: 755–767.
- Lee, H.-K., Park, J.-T., Cho, Y.-S., Bae, H.-S., Cho, M.-I. & Park, J.-C. 2012a. Odontogenic ameloblasts-associated protein (ODAM), via phosphorylation by bone morphogenetic protein receptor type IB (BMPR-IB), is implicated in ameloblast differentiation. *J. Cell. Biochem.* **113**: 1754–1765.
- Lee, H.-K., Park, S.-J., Oh, H.-J., Kim, J.-W., Bae, H.-S. & Park, J.-C. 2012b. Expression pattern, subcellular localization, and functional implications of ODA in ameloblasts, odontoblasts, osteoblasts, and various cancer cells. *Gene Expr. Patterns* **12**: 102–108.

- Lee, S.K., Kim, S.M., Lee, Y.J., Yamada, K.M., Yamada, Y. & Chi, J.G. 2003. The Structure of the Rat Ameloblastin Gene and Its Expression in Amelogenesis. *Korean Soc. Mol. Cell. Biol.* 216–225.
- Lee, S.K., Krebsbach, P.H., Matsuki, Y., Nanci, A., Yamada, K.M. & Yamada, Y. 1996. Ameloblastin expression in rat incisors and human tooth germs. *Int. J. Dev. Biol.* **40**: 1141–1150.
- Li, B. & Dewey, C.N. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**: 323.
- Line, S.R.P. & Novaes, P.D. 2005. The development and evolution of mammalian enamel: structural and functional aspects. *Braz J Morphol Sci* **22**: 67–72.
- Li, W., Mathews, C., Gao, C. & Denbesten, P.K. 1998. Identification of two additional exons at the 3' end of the amelogenin gene. *Arch. Oral Biol.* **43**: 497–504.
- Lopez-Valenzuela, M., Ramírez, O., Rosas, A., García-Vargas, S., Rasilla, M. de la, Lalueza-Fox, C., *et al.* 2012. An Ancestral miR-1304 Allele Present in Neanderthals Regulates Genes Involved in Enamel Formation and Could Explain Dental Differences with Modern Humans. *Mol. Biol. Evol.* **29**: 1797–1806.
- Louchart, A. & Viriot, L. 2011. From snout to beak: the loss of teeth in birds. *Trends Ecol. Evol.* **26**: 663–673.
- Lyaruu, D.M., Hu, C.C., Zhang, C., Qian, Q., Ryu, O.H., Moradian-Oldak, J., *et al.* 1998. Derived protein and cDNA sequences of hamster amelogenin. *Eur. J. Oral Sci.* **106 Suppl 1**: 299–307.
- MacDougall, M., Simmons, D., Gu, T.T., Forsman-Semb, K., Kärman Mårdh, C., Mesbah, M., *et al.* 2000. Cloning, characterization and immunolocalization of human ameloblastin. *Eur. J. Oral Sci.* **108**: 303–310.
- Maddison, D. & Maddison, W. 2005. MacClade 4: Analysis of phylogeny and character evolution. Version 4.08a.
- Madsen, O. 2009. Mammals. In: *The time tree of life*, pp. pp. 459–461. Hedges SB, Kumar S.,
- Madsen, O. 2009c. Mammals (Mammalia). In: *The TimeTree of Life*, pp. 459–461.
- Marshall, A.J., Du, Q., Draves, K.E., Shikishima, Y., HayGlass, K.T. & Clark, E.A. 2002. FDC-SP, a novel secreted protein expressed by follicular dendritic cells. *J. Immunol.* **169**: 2381–2389.
- Matsuki, Y., Nakashima, M., Amizuka, N., Warshawsky, H., Goltzman, D., Yamada, K.M., *et al.* 1995. A compilation of partial sequences of randomly selected cDNA clones from the rat incisor. *J. Dent. Res.* **74**: 307–312.

- Meredith, R.W., Gatesy, J., Cheng, J. & Springer, M.S. 2011a. Pseudogenization of the tooth gene enamelysin (MMP20) in the common ancestor of extant baleen whales. *Proc. R. Soc. B Biol. Sci.* **278**: 993–1002.
- Meredith, R.W., Gatesy, J., Cheng, J. & Springer, M.S. 2011b. Pseudogenization of the tooth gene enamelysin (MMP20) in the common ancestor of extant baleen whales. *Proc. R. Soc. Lond. B Biol. Sci.* **278**: 993–1002.
- Meredith, R.W., Gatesy, J., Murphy, W.J., Ryder, O.A. & Springer, M.S. 2009. Molecular Decay of the Tooth Gene Enamelin (ENAM) Mirrors the Loss of Enamel in the Fossil Record of Placental Mammals. *PLoS Genet* **5**: e1000634.
- Meredith, R.W., Gatesy, J. & Springer, M.S. 2013. Molecular decay of enamel matrix protein genes in turtles and other edentulous amniotes. *BMC Evol. Biol.* **13**: 20.
- Meredith, R.W., Janecka, J.E., Gatesy, J., Ryder, O.A., Fisher, C.A., Teeling, E.C., *et al.* 2011c. Impacts of the Cretaceous Terrestrial Revolution and KPg Extinction on Mammal Diversification. *Science* **334**: 521–524.
- Meredith, R.W., Zhang, G., Gilbert, M.T.P., Jarvis, E.D. & Springer, M.S. 2014. Evidence for a single loss of mineralized teeth in the common avian ancestor. *Science* **346**: 1254390.
- Mitsiadis, T.A., Filatova, A., Papaccio, G., Goldberg, M., About, I. & Papagerakis, P. 2014. Distribution of the amelogenin protein in developing, injured and carious human teeth. *Craniofacial Biol.* **5**: 477.
- Moffatt, P., Smith, C.E., Sooknanan, R., St-Arnaud, R. & Nanci, A. 2006a. Identification of secreted and membrane proteins in the rat incisor enamel organ using a signal-trap screening approach. *Eur. J. Oral Sci.* **114**: 139–146.
- Moffatt, P., Smith, C.E., St-Arnaud, R., Simmons, D., Wright, J.T. & Nanci, A. 2006b. Cloning of rat amelotin and localization of the protein to the basal lamina of maturation stage ameloblasts and junctional epithelium. *Biochem. J.* **399**: 37–46.
- Moffatt, P., Smith, C.E., St-Arnaud, R. & Nanci, A. 2008. Characterization of Apin, a secreted protein highly expressed in tooth-associated epithelia. *J. Cell. Biochem.* **103**: 941–956.
- Moffatt, P., Wazen, R.M., Neves, J.D.S. & Nanci, A. 2014. Characterisation of secretory calcium-binding phosphoprotein-proline-glutamine-rich 1: a novel basal lamina component expressed at cell-tooth interfaces. *Cell Tissue Res.* **358**: 843–855.
- Moradian-Oldak, J. 2012. Protein-mediated enamel mineralization. *Front. Biosci. J. Virtual Libr.* **17**: 1996–2023.
- Murphy, C.L., Kestler, D.P., Foster, J.S., Wang, S., Macy, S.D., Kennel, S.J., *et al.* 2008. Odontogenic ameloblast-associated protein nature of the amyloid found

- in calcifying epithelial odontogenic tumors and unerupted tooth follicles. *Amyloid* **15**: 89–95.
- Murphy, W. & Eizirik, E. 2009. Placental mammals (Eutheria). In: *The time tree of life*, pp. pp. 471–474. Hedges SB, Kumar S,.
- Nagano, T., Kakegawa, A., Yamakoshi, Y., Tsuchiya, S., Hu, J.C.-C., Gomi, K., *et al.* 2009. Mmp-20 and Klk4 Cleavage Site Preferences for Amelogenin Sequences. *J. Dent. Res.* **88**: 823–828.
- Nakamura, S., Terashima, T., Yoshida, T., Iseki, S., Takano, Y., Ishikawa, I., *et al.* 2005. Identification of genes preferentially expressed in periodontal ligament: Specific expression of a novel secreted protein, FDC-SP. *Biochem. Biophys. Res. Commun.* **338**: 1197–1203.
- Nakamura, Y., Slaby, I., Spahr, A., Pezeshki, G., Matsumoto, K. & Lyngstadaas, S.P. 2006. Ameloblastin Fusion Protein Enhances Pulpal Healing and Dentin Formation in Porcine Teeth. *Calcif. Tissue Int.* **78**: 278–284.
- Nakayama, Y., Holcroft, J. & Ganss, B. 2015. Enamel Hypomineralization and Structural Defects in Amelotin-deficient Mice. *J. Dent. Res.*, doi: 10.1177/0022034514566214.
- Nakayama, Y., Takai, H., Matsui, S., Matsumura, H., Zhou, L., Kato, A., *et al.* 2014a. Proinflammatory cytokines induce amelotin transcription in human gingival fibroblasts. *J. Oral Sci.* **56**: 261–268.
- Nakayama, Y., Takai, H., Matsui, S., Zhou, L., Abiko, Y., Ganss, B., *et al.* 2014b. Transcriptional regulation of amelotin gene by proinflammatory cytokines in gingival fibroblasts. *Connect. Tissue Res.* **55**: 18–20.
- Nanci, A., Zalzal, S., Lavoie, P., Kunikata, M., Chen, W.-Y., Krebsbach, P.H., *et al.* 1998. Comparative Immunochemical Analyses of the Developmental Expression and Distribution of Ameloblastin and Amelogenin in Rat Incisors. *J. Histochem. Cytochem.* **46**: 911–934.
- Nishio, C., Wazen, R., Kuroda, S., Moffatt, P. & Nanci, A. 2010a. Disruption of periodontal integrity induces expression of apin by epithelial cell rests of Malassez. *J. Periodontal Res.* **45**: 709–713.
- Nishio, C., Wazen, R., Kuroda, S., Moffatt, P. & Nanci, A. 2010b. Expression pattern of odontogenic ameloblast-associated and amelotin during formation and regeneration of the junctional epithelium. *Eur Cell Mater* **20**: 393–402.
- Nishio, C., Wazen, R., Moffatt, P. & Nanci, A. 2013. Expression of odontogenic ameloblast-associated and amelotin proteins in the junctional epithelium. *Periodontol. 2000* **63**: 59–66.
- Oida, S., Nagano, T., Yamakoshi, Y., Ando, H., Yamada, M. & Fukae, M. 2002. Amelogenin Gene Expression in Porcine Odontoblasts. *J. Dent. Res.* **81**: 103–108.

- Osborn, J.W. & Hillman, J. 1979. Enamel structure in some therapsids and mesozoic mammals. *Calcif. Tissue Int.* **29**: 47–61.
- Oshiro, A., Iseki, S., Miyauchi, M., Terashima, T., Kawaguchi, Y., Ikeda, Y., *et al.* 2012. Lipopolysaccharide induces rapid loss of follicular dendritic cell-secreted protein in the junctional epithelium. *J. Periodontal Res.* **47**: 689–694.
- Paine, M.L. & Snead, M.L. 1997. Protein Interactions During Assembly of the Enamel Organic Extracellular Matrix. *J. Bone Miner. Res.* **12**: 221–227.
- Paine, M.L., Wang, H.-J., Luo, W., Krebsbach, P.H. & Snead, M.L. 2003. A Transgenic Animal Model Resembling Amelogenesis Imperfecta Related to Ameloblastin Overexpression. *J. Biol. Chem.* **278**: 19447–19452.
- Papagerakis, P., Ibarra, J.M., Inozentseva, N., DenBesten, P. & MacDougall, M. 2005. Mouse Amelogenin Exons 8 and 9: Sequence Analysis and Protein Distribution. *J. Dent. Res.* **84**: 613–617.
- Papagerakis, P., MacDougall, M., Hotton, D., Bailleul-Forestier, I., Oboeuf, M. & Berdal, A. 2003. Expression of amelogenin in odontoblasts. *Bone* **32**: 228–240.
- Park, J.-C., Park, J.-T., Son, H.-H., Kim, H.-J., Jeong, M.-J., Lee, C.-S., *et al.* 2007. The amyloid protein APin is highly expressed during enamel mineralization and maturation in rat incisors. *Eur. J. Oral Sci.* **115**: 153–160.
- Park, S.-M., Park, S.-Y., Kim, J.-H., Kang, T.-W., Park, J.-L., Woo, K.-M., *et al.* 2013. Genome-wide mRNA profiling and multiplex quantitative RT-PCR for forensic body fluid identification. *Forensic Sci. Int. Genet.* **7**: 143–150.
- Petersen, T.N., Brunak, S., von Heijne, G. & Nielsen, H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **8**: 785–786.
- Poulter, J.A., Murillo, G., Brookes, S.J., Smith, C.E.L., Parry, D.A., Silva, S., *et al.* 2014. Deletion of ameloblastin exon 6 is associated with amelogenesis imperfecta. *Hum. Mol. Genet.* **23**: 5317–5324.
- Ravindranath, H.H., Chen, L.-S., Zeichner-David, M., Ishima, R. & Ravindranath, R.M.H. 2004. Interaction between the enamel matrix proteins amelogenin and ameloblastin. *Biochem. Biophys. Res. Commun.* **323**: 1075–1083.
- Ravindranath, R.M.H., Devarajan, A. & Uchida, T. 2007. Spatiotemporal Expression of Ameloblastin Isoforms during Murine Tooth Development. *J. Biol. Chem.* **282**: 36370–36376.
- Reith, E.J. 1961. THE ULTRASTRUCTURE OF AMELOBLASTS DURING MATRIX FORMATION AND THE MATURATION OF ENAMEL. *J. Cell Biol.* **9**: 825–839.

- Ren, C., Amm, H.M., DeVilliers, P., Wu, Y., Deatherage, J.R., Liu, Z., *et al.* 2012. Targeting the Sonic Hedgehog Pathway in Keratocystic Odontogenic Tumor. *J. Biol. Chem.* **287**: 27117–27125.
- Ren, C., Diniz, M.G., Piazza, C., Amm, H.M., Rollins, D.L., Rivera, H., *et al.* 2011. Differential Enamel and Osteogenic Gene Expression Profiles in Odontogenic Tumors. *Cells Tissues Organs* **194**: 296–301.
- Rijnkels, M., Elnitski, L., Miller, W. & Rosen, J.M. 2003. Multispecies comparative analysis of a mammalian-specific genomic domain encoding secretory proteins. *Genomics* **82**: 417–432.
- Rozen, S. & Skaletsky, H. 2000. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol. Clifton NJ* **132**: 365–386.
- Ruoslahti, E. 1996. Rgd and Other Recognition Sequences for Integrins. *Annu. Rev. Cell Dev. Biol.* **12**: 697–715.
- Sabatini, L.M., Ota, T. & Azen, E.A. 1993. Nucleotide sequence analysis of the human salivary protein genes HIS1 and HIS2, and evolution of the STATH/HIS gene family. *Mol. Biol. Evol.* **10**: 497–511.
- Sahni, A. 1987. Evolutionary aspects of reptilian and mammalian enamel structure. *Scanning Microsc.* **1**: 1903–1912.
- Salido, E.C., Yen, P.H., Koprivnikar, K., Yu, L.C. & Shapiro, L.J. 1992. The human enamel protein gene amelogenin is expressed from both the X and the Y chromosomes. *Am. J. Hum. Genet.* **50**: 303–316.
- Sander, P.M. 2000. Prismless enamel in amniotes: terminology, function and evolution. In: *Development, Function and Evolution of Teeth*, pp. 92–106. Mark F. Teaford, Moya Meredith Smith & Mark W. J. Ferguson.
- Sansom, I.J., Smith, M.P., Armstrong, H.A. & Smith, M.M. 1992. Presence of the earliest vertebrate hard tissue in conodonts. *Science* **256**: 1308–1311.
- Sasagawa, I. & Ferguson, M.W. 1991. The development of enamel tubules during the formation of enamel in the marsupial *Monodelphis domestica*. *J. Anat.* **179**: 47–58.
- Sawada, T., Sekiguchi, H., Uchida, T., Yamashita, H., Shintani, S. & Yanagisawa, T. 2011. Histological and immunohistochemical analyses of molar tooth germ in enamelin-deficient mouse. *Acta Histochem.* **113**: 542–546.
- Sawada, T., Yamazaki, T., Shibayama, K., Kumazawa, K., Yamaguchi, Y. & Ohshima, M. 2014. Expression and localization of laminin 5, laminin 10, type IV collagen, and amelotin in adult murine gingiva. *J. Mol. Histol.* **45**: 293–302.
- Shedlock, A.M. & Edwards, S.V. 2009. Amniotes (amniota). In: *The timetree of life*, pp. 375–379. Hedges SB and Kumar S.

- Shinomura, T., Nakamura, S., Ito, K., Shirasawa, S., Höök, M. & Kimura, J.H. 2008. Adsorption of Follicular Dendritic Cell-secreted Protein (FDC-SP) onto Mineral Deposits APPLICATION OF A NEW STABLE GENE EXPRESSION SYSTEM. *J. Biol. Chem.* **283**: 33658–33664.
- Shintani, S., Kobata, M., Toyosawa, S., Fujiwara, T., Sato, A. & Ooshima, T. 2002. Identification and characterization of ameloblastin gene in a reptile. *Gene* **283**: 245–254.
- Shintani, S., Kobata, M., Toyosawa, S. & Ooshima, T. 2003. Identification and characterization of ameloblastin gene in an amphibian, *Xenopus laevis*. *Gene* **318**: 125–136.
- Shintani, S., Kobata, M., Toyosawa, S. & Ooshima, T. 2006. Expression of ameloblastin during enamel formation in a crocodile. *J. Exp. Zool. B Mol. Dev. Evol.* **306B**: 126–133.
- Siddiqui, S., Bruker, C.T., Kestler, D.P., Foster, J.S., Gray, K.D., Solomon, A., *et al.* 2009. Odontogenic Ameloblast Associated Protein as a Novel Biomarker for Human Breast Cancer. *Am. Surg.* **75**: 769–775.
- Sigrist, C.J.A., Castro, E. de, Cerutti, L., Cuche, B.A., Hulo, N., Bridge, A., *et al.* 2013. New and continuing developments at PROSITE. *Nucleic Acids Res.* **41**: D344–D347.
- Sigrist, C.J.A., Cerutti, L., de Castro, E., Langendijk-Genevaux, P.S., Bulliard, V., Bairoch, A., *et al.* 2010. PROSITE, a protein domain database for functional characterization and annotation. *Nucleic Acids Res.* **38**: D161–166.
- Silvent, J., Gasse, B., Mornet, E. & Sire, J.-Y. 2014. Molecular Evolution of the Tissue-Nonspecific Alkaline Phosphatase Allows Prediction and Validation of Missense Mutations Responsible for Hypophosphatasia. *J. Biol. Chem.* jbc.M114.576843.
- Simmer, J.P. & Bartlett, J.D. 2004. Kallikrein 4 Is a Secreted Protein. *Cancer Res.* **64**: 8481–8483.
- Simmer, J.P. & Fincham, A.G. 1995. Molecular Mechanisms of Dental Enamel Formation. *Crit. Rev. Oral Biol. Med.* **6**: 84–108.
- Simmons, D., Gu, T.T., Krebsbach, P.H., Yamada, Y. & MacDougall, M. 1998. Identification and characterization of a cDNA for mouse ameloblastin. *Connect. Tissue Res.* **39**: 3–12; discussion 63–67.
- Sire, J.-Y., Delgado, S., Fromentin, D. & Girondot, M. 2005. Amelogenin: lessons from evolution. *Arch. Oral Biol.* **50**: 205–212.
- Sire, J.-Y., Davit-Béal, T., Delgado, S. & Gu, X. 2007. The origin and evolution of enamel mineralization genes. *Cells Tissues Organs* **186**: 25–48.
- Sire, J.-Y., Delgado, S.C. & Girondot, M. 2008. Hen's teeth with enamel cap: from dream to impossibility. *BMC Evol. Biol.* **8**: 246.

- Sire, J.-Y., Huang, Y., Li, W., Delgado, S., Goldberg, M. & DenBesten, P.K. 2012. Evolutionary Story of Mammalian-specific Amelogenin Exons 4, “4b”, 8, and 9. *J. Dent. Res.* **91**: 84–89.
- Smith, C.E. 1998. Cellular and Chemical Events During Enamel Maturation. *Crit. Rev. Oral Biol. Med.* **9**: 128–161.
- Smith, C.E. & Warshawsky, H. 1977. Quantitative analysis of cell turnover in the enamel organ of the rat incisor. Evidence for ameloblast death immediately after enamel matrix secretion. *Anat. Rec.* **187**: 63–98.
- Smith, C.E., Wazen, R., Hu, Y., Zalzal, S.F., Nanci, A., Simmer, J.P., *et al.* 2009. Consequences for enamel development and mineralization resulting from loss of function of ameloblastin or enamelin. *Eur. J. Oral Sci.* **117**: 485–497.
- Smith, M.M. 1989. Distribution and variation in enamel structure in the oral teeth of sarcopterygians: Its significance for the evolution of a protoprismatic enamel. *Hist. Biol.* **3**: 97–126.
- Snead, M.L., Luo, W., Lau, E.C. & Slavkin, H.C. 1988. Spatial- and temporal-restricted pattern for amelogenin gene expression during mouse molar tooth organogenesis. *Development* **104**: 77–85.
- Snead, M.L., Zeichner-David, M., Chandra, T., Robson, K.J., Woo, S.L. & Slavkin, H.C. 1983. Construction and identification of mouse amelogenin cDNA clones. *Proc. Natl. Acad. Sci.* **80**: 7254–7258.
- Solomon, A., Murphy, C.L., Weaver, K., Weiss, D.T., Hrcic, R., Eulitz, M., *et al.* 2003. Calcifying epithelial odontogenic (Pindborg) tumor-associated amyloid consists of a novel human protein. *J. Lab. Clin. Med.* **142**: 348–355.
- Somogyi-Ganss, E., Nakayama, Y., Iwasaki, K., Nakano, Y., Stolf, D., McKee, M.D., *et al.* 2012a. Comparative Temporospatial Expression Profiling of Murine Amelotin Protein during Amelogenesis. *Cells Tissues Organs* **195**: 535–549.
- Sonoda, A., Iwamoto, T., Nakamura, T., Fukumoto, E., Yoshizaki, K., Yamada, A., *et al.* 2009. Critical Role of Heparin Binding Domains of Ameloblastin for Dental Epithelium Cell Adhesion and Ameloblastoma Proliferation. *J. Biol. Chem.* **284**: 27176–27184.
- Spahr, A., Lyngstadaas, S.P., Slaby, I., Haller, B., Boeckh, C., Tsoulfidou, F., *et al.* 2002. Expression of amelin and trauma-induced dentin formation. *Clin. Oral Investig.* **6**: 51–57.
- Spahr, A., Lyngstadaas, S.P., Slaby, I. & Pezeshki, G. 2006. Ameloblastin expression during craniofacial bone formation in rats. *Eur. J. Oral Sci.* **114**: 504–511.
- Tagliabracci, V.S., Engel, J.L., Wen, J., Wiley, S.E., Worby, C.A., Kinch, L.N., *et al.* 2012. Secreted Kinase Phosphorylates Extracellular Proteins That Regulate Biomineralization. *Science* **336**: 1150–1153.

- Takagi, T., Suzuki, M., Baba, T., Minegishi, K. & Sasaki, S. 1984. Complete amino acid sequence of amelogenin in developing bovine enamel. *Biochem. Biophys. Res. Commun.* **121**: 592–597.
- Takahashi, S., Fukuda, M., Mitani, A., Fujimura, T., Iwamura, Y., Sato, S., *et al.* 2014. Follicular dendritic cell-secreted protein is decreased in experimental periodontitis concurrently with the increase of interleukin-17 expression and the Rankl/Opg mRNA ratio. *J. Periodontal Res.* **49**: 390–397.
- Tamburstuen, M.V., Reppe, S., Spahr, A., Sabetrasekh, R., Kvalheim, G., Slaby, I., *et al.* 2010. Ameloblastin promotes bone growth by enhancing proliferation of progenitor cells and by stimulating immunoregulators. *Eur. J. Oral Sci.* **118**: 451–459.
- Tamburstuen, M.V., Reseland, J.E., Spahr, A., Brookes, S.J., Kvalheim, G., Slaby, I., *et al.* 2011a. Ameloblastin expression and putative autoregulation in mesenchymal cells suggest a role in early bone formation and repair. *Bone* **48**: 406–413.
- Tamburstuen, M.V., Snead, M.L., Reseland, J.E., Paine, M.L. & Lyngstadaas, S.P. 2011b. Ameloblastin upstream region contains structural elements regulating transcriptional activity in a stromal cell line derived from bone marrow. *Eur. J. Oral Sci.* **119**: 286–292.
- Tanabe, T., Aoba, T., Moreno, E.C., Fukae, M. & Shimuzu, M. 1990. Properties of phosphorylated 32 kd nonamelogenin proteins isolated from porcine secretory enamel. *Calcif. Tissue Int.* **46**: 205–215.
- Termine, J.D., Belcourt, A.B., Christner, P.J., Conn, K.M. & Nysten, M.U. 1980. Properties of dissociatively extracted fetal tooth matrix proteins. I. Principal molecular species in developing bovine enamel. *J. Biol. Chem.* **255**: 9760–9768.
- Thesleff, I. 2003. Epithelial-mesenchymal signalling regulating tooth morphogenesis. *J. Cell Sci.* **116**: 1647–1648.
- Torii, S. 1998. Origin of enamel prisms and Hunter-Schreger bands in reptilian enamel. *Connect. Tissue Res.* **38**: 45–51; discussion 73–79.
- Torres-Quintana, M.A., Gaete, M., Hernandez, M., Farfías, M. & Lobos, N. 2005. Ameloblastin and amelogenin expression in postnatal developing mouse molars. *J. Oral Sci.* **47**: 27–34.
- Toyosawa, S., Fujiwara, T., Ooshima, T., Shintani, S., Sato, A., Ogawa, Y., *et al.* 2000. Cloning and characterization of the human ameloblastin gene. *Gene* **256**: 1–11.
- Toyosawa, S., O’Hugin, C., Figueroa, F., Tichy, H. & Klein, J. 1998. Identification and characterization of amelogenin genes in monotremes, reptiles, and amphibians. *Proc. Natl. Acad. Sci. U. S. A.* **95**: 13056–13061.

- Trueb, B., Taeschler, S., Schild, C. & Lang, N.P. 2007. Expression of phosphoproteins and amelotin in teeth. *Int. J. Mol. Med.* **19**: 49–54.
- Uchida, T., Fukae, M., Tanabe, T., Yamakoshi, Y., Satoda, T., Murakami, C., *et al.* 1995. 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.* **16**: 131–140.
- Uchida, T., Murakami, C., Wakida, K., Satoda, T., Dohi, N. & Takahashi, O. 1997. Synthesis, Secretion, Degradation, and Fate of Ameloblastin During the Matrix Formation Stage of the Rat Incisor as Shown by Immunocytochemistry and Immunocytochemistry Using Region-specific Antibodies. *J. Histochem. Cytochem.* **45**: 1329–1340.
- Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., *et al.* 2012. Primer3—new capabilities and interfaces. *Nucleic Acids Res.* **40**: e115–e115.
- Vymětal, J., Slabý, I., Spahr, A., Vondrášek, J. & Lyngstadaas, S.P. 2008. Bioinformatic analysis and molecular modelling of human ameloblastin suggest a two-domain intrinsically unstructured calcium-binding protein. *Eur. J. Oral Sci.* **116**: 124–134.
- Wakita, M. 1993. Current studies on tooth enamel development in lower vertebrates. *Kaibogaku Zasshi* **68**: 399–409.
- Wang, C., Zhou, L., Li, S., Wei, J., Wang, W., Zhou, T., *et al.* 2010. C4orf7 contributes to ovarian cancer metastasis by promoting cancer cell migration and invasion. *Oncol. Rep.*, doi: 10.3892/or_00000939.
- Wang, X., Deng, X. & Zhang, X. 2012a. Identification of a Novel Splicing Form of Amelogenin Gene in a Reptile, *Ctenosaura similis*. *PLoS ONE* **7**.
- Wang, X., Fan, J.-L., Ito, Y., Luan, X. & Diekwisch, T.G.H. 2006. Identification and characterization of a squamate reptilian amelogenin gene: *Iguana iguana*. *J. Exp. Zool. B Mol. Dev. Evol.* **306B**: 393–406.
- Wang, X., Ito, Y., Luan, X., Yamane, A. & Diekwisch, T.G.H. 2005. Amelogenin sequence and enamel biomineralization in *Rana pipiens*. *J. Exp. Zool. B Mol. Dev. Evol.* **304B**: 177–186.
- Wang, X., Wang, S., Lu, Y., Gibson, M.P., Liu, Y., Yuan, B., *et al.* 2012b. FAM20C Plays an Essential Role in the Formation of Murine Teeth. *J. Biol. Chem.* **287**: 35934–35942.
- Wang, X., Xing, Z., Zhang, X., Zhu, L. & Diekwisch, T.G.H. 2013. Alternative Splicing of the Amelogenin Gene in a Caudate Amphibian, *Plethodon cinereus*. *PLoS ONE* **8**.
- Wei, N., Yu, H., Yang, S., Yang, X., Yuan, Q., Man, Y., *et al.* 2011. Effect of FDC-SP on the phenotype expression of cultured periodontal ligament cells. *Arch. Med. Sci. AMS* **7**: 235–241.

- Witkop, C.J. 1988. Amelogenesis imperfecta, dentinogenesis imperfecta and dentin dysplasia revisited: problems in classification. *J. Oral Pathol.* **17**: 547–553.
- Wood, C.B., Dumont, E.R. & Crompton, A.W. 1999. New Studies of Enamel Microstructure in Mesozoic Mammals: A Review of Enamel Prisms as a Mammalian Synapomorphy. *J. Mamm. Evol.* **6**: 177–213.
- Wright, J.T. 2006. The molecular etiologies and associated phenotypes of amelogenesis imperfecta. *Am. J. Med. Genet. A.* **140A**: 2547–2555.
- Wright, J.T., Hart, P.S., Aldred, M.J., Seow, K., Crawford, P.J.M., Hong, S.P., *et al.* 2003. Relationship of phenotype and genotype in X-linked amelogenesis imperfecta. *Connect. Tissue Res.* **44 Suppl 1**: 72–78.
- Wright, J.T., Torain, M., Long, K., Seow, K., Crawford, P., Aldred, M.J., *et al.* 2011. Amelogenesis Imperfecta: Genotype-Phenotype Studies in 71 Families. *Cells Tissues Organs* **194**: 279–283.
- Wu, M., Han, L., Shi, Y., Xu, G., Wei, J., You, L., *et al.* 2010. Development and characterization of a novel method for the analysis of gene expression patterns in lymphatic endothelial cells derived from primary breast tissues. *J. Cancer Res. Clin. Oncol.* **136**: 863–872.
- Xiang, L., Ma, L., He, Y., Wei, N. & Gong, P. 2014a. Osteogenic differentiation of human periodontal ligament cells after transfection with recombinant lentiviral vector containing follicular dendritic cell secreted protein. *J. Periodontal Res.* **49**: 554–562.
- Xiang, L., Ma, L., He, Y., Wei, N. & Gong, P. 2014b. Transfection With Follicular Dendritic Cell Secreted Protein to Affect Phenotype Expression of Human Periodontal Ligament Cells. *J. Cell. Biochem.* **115**: 940–948.
- Yamakoshi, Y., Hu, J.C.-C., Fukae, M., Yamakoshi, F. & Simmer, J.P. 2006. How do enamelysin and kallikrein 4 process the 32-kDa enamelin? *Eur. J. Oral Sci.* **114**: 45–51.
- Yamakoshi, Y., Tanabe, T., Oida, S., Hu, C.-C., Simmer, J.P. & Fukae, M. 2001. Calcium binding of enamel proteins and their derivatives with emphasis on the calcium-binding domain of porcine sheathlin. *Arch. Oral Biol.* **46**: 1005–1014.
- Zhang, X., Diekwisch, T.G.H. & Luan, X. 2011a. Structure and function of ameloblastin as an extracellular matrix protein: adhesion, calcium binding, and CD63 interaction in human and mouse. *Eur. J. Oral Sci.* **119**: 270–279.
- Zhang, Y., Zhang, X., Lu, X., Atsawasuwana, P. & Luan, X. 2011b. Ameloblastin regulates cell attachment and proliferation through RhoA and p27. *Eur. J. Oral Sci.* **119**: 280–285.

LISTE DES COMMUNICATIONS

I. Publications non incluses dans la thèse

- Silvent J, **Gasse B**, Mornet E, Sire J-Y. 2014. Molecular evolution of the tissue-nonspecific alkaline phosphatase allows prediction and validation of missense mutations responsible for hypophosphatasia. *J. Biol. Chem.* 289(35):24168-24179.
- Assaraf-Weill N, **Gasse B**, Silvent S, Bardet C, Sire J-Y, Davit-Béal T. 2014. Ameloblasts express type I collagen during amelogenesis. *Journal of Dental Research* 93(5):502-507.
- Chaussain C, Bouazza N, **Gasse B**, Laffont A-G, Opsahl Vital S, Davit-Béal T, Moulis E, Chabadel O, Hennequin M, Courson F, Droz D, Vaysse F, Laboux O, Tassery H, Carel J-C, Alcais A, Treluyer J-M, Beldjord C, Sire J-Y. 2014. Dental caries and enamel haplotype. *Journal of Dental Research* 93(4):360-365.
- Assaraf-Weill N, **Gasse B**, Al-Hashimi N, Delgado S, Sire J-Y, Davit-Béal T. 2013. Conservation of amelogenin gene expression during tetrapod evolution. *J. Exp. Zool. (Mol. Dev. Evol.)* 320B:200-209.
- Gasse B**, Grabar S, Lafont A-G, Quinquis L, Opsahl Vital S, David-Béal T, Moulis E, Chabadel O, Hennequin M, Courson F, Droz D, Vaysse F, Laboux O, Tassery H, Al-Hashimi N, Boillot A, Carel J-C, Treluyer J-M, Jeanpierre M, Beldjord C, Sire J-Y, Chaussain C. 2013. Common SNPs of amelogeninX (AMELX) and dental caries susceptibility. *Journal of Dental Research* 92(5):418-424.
- Gasse B**, Karayigit E, Mathieu E, Jung S, Garret A, Huckert M, Morkmued S, Schneider C, Vidal L, Hemmerlé J, Sire J-Y, Bloch-Zupan A. 2013. Homozygous and compound heterozygous MMP20 mutations in amelogenesis imperfecta. *Journal of Dental Research.* 92(7):598-603

II. Communication orales et affichées

1. Congrès nationaux

- Gasse B**, Davit-Béal T, Sire J-Y. 2014. Evolution de la structure génique et de l'expression de l'amélotine chez les tétrapodes. *16e Journées Françaises de Biologie des Tissus Minéralisés, Limoges.* Poster
- Gasse B**, Delsuc F, Sire J-Y. 2013. Analyse évolutive de l'améloblastine chez les mammifères. *15e Journées Françaises de Biologie des Tissus Minéralisés, Poitiers.* Poster
- Gasse B**, Mornet E, Sire J-Y. 2012. L'étude évolutive de la phosphatase alcaline osseuse permet de valider et de prédire les mutations responsables de l'hypophosphatasie. *14e Journées Françaises de Biologie des Tissus*

Minéralisés, Bordeaux. Communication orale. **Prix meilleurs oraux étudiants**

Gasse B, Al-Hashimi N, Sire J-Y. 2011. Origine, évolution et expression d'une nouvelle protéine de l'émail, l'amélotine. *13e journées françaises de Biologie des Tissus Minéralisés, Paris*. Communication orale

2. Congrès internationaux

Gasse B, Sire J-Y. 2013. History of an enamel protein, amelotin, which experienced distinct evolutionary pathways in sarcopterygian lineages. *11th International Conference on Tooth Morphogenesis and Differentiation, La Londe Les Maures, France*. Communication orale

Gasse B, Mornet E, Sire J-Y. 2012. Evolutionary analysis of ALPL validates and predicts human mutations leading to hypophosphatasia. *6th International Alkaline Phosphatase Symposium, Huningue, France*. Communication orale