

Identification of secreted and membrane proteins in the rat incisor enamel organ using a signal-trap screening approach

Pierre Moffatt¹, Charles E. Smith¹,
Roy Sooknanan², René St-Arnaud³,
Antonio Nanci¹

¹Laboratory for the Study of Calcified Tissues and Biomaterials, Faculté de Médecine Dentaire, Département de Stomatologie, Université de Montréal, Montréal, Québec, Canada; ²Alethia Biotherapeutics Inc., Montréal, Québec, Canada; ³Genetics Unit, Shriners Hospital for Children, Montréal, Québec, Canada

Moffatt P, Smith CE, Sooknanan R, St-Arnaud R, Nanci A. Identification of secreted and membrane proteins in the rat incisor enamel organ using a signal-trap screening approach. *Eur J Oral Sci* 2006; 114 (Suppl. 1): 139–146 © Eur J Oral Sci, 2006

The secretome represents the subset of proteins that are targeted by signal peptides to the endoplasmic reticulum. Among those, secreted proteins play a pivotal role because they regulate determinant cell activities such as differentiation and intercellular communication. In calcified tissues, they also represent key players in extracellular mineralization. This study was carried out to establish a secretome profile of rat enamel organ (EO) cells. A functional genomic technology, based on the signal trap methodology, was applied, starting with a library of 5'-enriched cDNA fragments prepared from rat incisor EOs. A total of 2,592 clones were analyzed by means of macroarray hybridizations and DNA sequencing. Ninety-four unique clones encoding a signal peptide were retrieved. Among those were 84 matched known genes, many not previously reported to be expressed by the EO. Most importantly, 10 clones were classified as being novel, with EO-009 identified as the rat homolog of human APin protein. These data indicate that many secreted and membrane-embedded EO proteins still remain to be identified, some of which may play crucial roles in regulating processes that create an optimal environment for the formation and organization of apatite crystals into a complex three-dimensional calcified matrix.

Antonio Nanci PhD, Université de Montréal, Faculté de Médecine Dentaire, PO Box 6128, Station Centre-Ville, Montréal, Québec, Canada H3C 3J7

Telefax: +1-514-3432233
E-mail: antonio.nanci@umontreal.ca

Key words: ameloblast; APin; enamel organ; secretome; signal-trap

Accepted for publication December 2005

The genome defines all proteins that cells can produce, which collectively comprise the proteome. The secretome represents the subset of proteins that are targeted by signal peptides to the endoplasmic reticulum, and includes secretory pathway resident proteins, membrane-embedded proteins, lysosomal proteins, and secreted proteins (1). The latter play a pivotal role because they regulate determinant cell activities, such as differentiation and intercellular communication, and also establish the environment around the cell. In calcified tissues, they help to determine when, where, and to what extent, the extracellular milieu will calcify. In teeth, this is underscored by mutations in genes encoding the main extracellular matrix proteins of the enamel that cause debilitating disorders such as *amelogenesis imperfecta* (2).

The various genome sequencing initiatives carried out in recent years have generated tremendous amounts of data on the expression and structure of genes. There is a marked interest in obtaining a comprehensive view of the secretome from many different species, including human, mouse, and others (3–5). The approaches taken mostly rely on computational analysis and genome-wide scanning for expressed sequence tags (ESTs) and/or full-length transcripts (Riken), sometimes coupled with biological screens. Albeit successful at annotating large databases of secreted proteins, the cell-specific site of

expression of a gene, and the function of the protein it encodes, cannot be easily predicted. Previous attempts to define the transcriptome of the tooth organ essentially applied broad sequencing approaches (6,7). The *ameloblastin* gene was identified from one such study (8). Differential display was also applied to discover novel cDNA fragments expressed specifically in the tooth organ (9). Collectively, however, these data have not always provided an insight into the true physiological role/purpose a protein serves. The main reason being that ESTs could not be used to predict accurately protein sequence and function as they are intrinsically 3' (C-terminally) biased.

Functional genomic tools, called 'signal-traps', have been developed to identify cDNAs encoding signal peptides in complex libraries, thus allowing the identification of function before sequence (10–14). One of these, conducted in mammalian cells, has previously been used successfully to start characterizing the secretome of bone and skin and to identify novel genes (15,16). In the present report, we have used this approach to look at the secretome of the rat enamel organ (EO), a component of the tooth about which only limited information is known with regard to secreted proteins and membrane-associated proteins. Apart from finding cDNAs encoding very well documented and important secreted proteins

expressed in teeth [e.g. amelogenin, ameloblastin, enamel, matrix metalloproteinase-20 (MMP-20)], this method allowed the identification of many other cDNAs encoding proteins with known functions in other tissues but as yet not suspected or known to be expressed in teeth. Last but not least, a few clones encoded entirely new proteins. Altogether, these data validate the screening strategy we have used as an efficient approach for identifying proteins secreted by the rat EO.

Material and methods

cDNA library construction, screening, and analysis

EOs were dissected from freeze-dried upper and lower incisors of 100-g male Wistar rats, as described in SMITH *et al.* (17). Total RNA was extracted using Trizol (Invitrogen, Mississauga, ON, Canada). The polyA-containing RNA ($\approx 4 \mu\text{g}$) was then isolated with oligodT beads using the Oligotex mRNA Midi kit (Qiagen, Mississauga, ON, Canada), and processed, essentially as previously described, for the construction of an oligo-capped library enriched in 5' fragments of cDNAs and the screening procedure (14). Briefly, the cDNA library was directionally cloned into the Sindbis virus-based signal-trap vectors and consisted of 750,000 clones. The library was then expression-screened in baby hamster kidney (BHK21) fibroblasts, which involved 4 basic steps: (i) electroporation, seeding, and growth of the cells for 16 h; (ii) accumulation of viral particles in media for 4 h; (iii) infection of a naïve BHK21 monolayer with the media; and (iv) collection of cellular RNA, including viral RNAs, after 20 h. The efficiency of transfection was estimated to be 10%, for a total of 3×10^6 expressing cells, which represents 4-times coverage of the library. The cDNA fragments found in the infected cells were amplified by reverse transcription-polymerase chain reaction (RT-PCR) with the Titan One Tube kit (Roche Diagnostics, Laval, QC, Canada) using primers flanking the site of insertion within the modified viral RNA. The amplified cDNA mixture was then transferred into a population of bacterial clones through shotgun cloning and analyzed by hybridization onto macroarrays and sequencing. A total of 2,592 bacterial colonies were individually picked, grown in 96-well plates, and gridded onto duplicate sets of nylon macroarray membranes (Osmonics, Westborough, MA, USA), each composed of 864 clones (9×96). Fragments corresponding to redundant clones after random sequencing were labeled with ^{32}P -dCTP by random priming and hybridized onto the macroarrays using Church buffer (18) at 65°C . Cycle-sequencing was then performed on the non-hybridizing clones using the DTCS Quick start kit and analyzed on a CEQ2000 automated DNA sequencer (Beckman Coulter, Mississauga, ON, Canada). Sequencing results were compared using BLASTN against GenBank non-redundant (nr) and EST databases. The coding sequence of cDNAs classified as being novel were further inspected for the presence of a signal peptide using the SignalP V3.0 server (19).

Northern blotting and full-length cloning of EO-009

Total RNA was extracted from 18 different freeze-dried rat tissues, as described above. The RNAs (10 or $20 \mu\text{g}$) were loaded on a MOPS-buffered 1.1% agarose/1.2% formaldehyde gel. After separation, the RNAs were transferred

from the gel onto a nylon membrane (Osmonics) with $20 \times$ sodium citrate (SSC) overnight, and fixed by UV crosslinking. The blot was prehybridized for 2 h in Church buffer at 65°C and hybridized under the same conditions overnight with the 430-bp EO-009 cDNA fragment labeled by random-priming with [^{32}P]dCTP (Amersham Biosciences, Baie d'Urfe, QC, Canada). After stringent washes with $0.2 \times \text{SSC}/0.1\%$ sodium dodecyl sulphate (SDS) at 65°C , the blot was autoradiographed. In order to obtain the full-length cDNA sequence for EO-009, we employed the 3'-RACE strategy, as described previously (15,16). Briefly, the EO total RNA was first reverse transcribed using SuperscriptII (Invitrogen) and a dT₁₅-tailed oligo (5'-gaga-tgaattcctcagactttttttttttt-3'). It was then amplified by polymerase chain reaction (PCR) with rTaq (Amersham) and the EO-009-specific forward primer (5'-gagaa-gtaaggtatcatctg-3') and the dT₁₅-tailed-specific reverse primer (5'-gagatgaattcctcagc-3'). The resulting products were cloned into pBluescriptKS (Stratagene, La Jolla, CA, USA) and sequenced as described above.

Results

In the present study, the starting material was EO caps (20), which contained predominantly epithelial cells with small amounts of adjacent connective tissue. The oligo-capped cDNA library comprised 750,000 clones, and its initial characterization indicated a prominent bias towards at least two clones: amelogenin and ameloblastin. This was not unexpected, as the library was not normalized, and their high rate of occurrence reflected the abundance of their mRNA in the starting RNA preparation. Because of that, their cDNA fragments were used as probes for hybridizing onto macroarrays of bacterial clones (≈ 2600) to maximize our chances of finding lower-abundance genes. Results indicated that a vast majority of clones hybridized with amelogenin and ameloblastin cDNAs ($\approx 63\%$ and $\approx 11\%$, respectively; data not shown). The remaining clones (≈ 600) were sequenced, yielding 91 unique species (Tables 1 and 2). Among those, 81 were identical or closely related to cDNAs encoding known proteins and are listed, according to function, in Table 1. Genes previously reported to be expressed in the EOs (*amelogenin*, *ameloblastin*, *enamelin*, *MMP-20*) were all found as part of this first screen.

A considerable proportion ($\approx 10\%$) of the clones found were classified as being unknown (Table 2). Cleavable signal peptides with high predictive values were found in all, except EO-525-a, which had a putative signal anchor. Eight of them had almost perfect matches on mouse Riken clones, and two did not relate to any publicly available sequences. Mainly based on the EST collection at UniGene, EO-560-a, -525-a, -081-a, -011, and -416-a had close to 500 total ESTs each and were derived from many different tissues (Table 2). The five remaining had limited (EO-009, -017, -063) or no (EO-014, EO-463) known ESTs. The expression profile of each, except EO-014, was further characterized by northern blotting. The results confirmed the selective expression of four genes (*EO-009*, *EO-017*, *EO-063*, and

Table 1

List of known clones retrieved from the screen of the enamel organ library for cDNAs encoding a signal peptide

Function*	Clone ID	Name of hit	GenBank accession number	Protein type [†]	Cellular localization	
Cell-cell interaction/ adhesion	EO-027	Epithelial V-like antigen	XM_236197	I	Cell membrane	
	EO-490-a	PECAM	CK364601	I	Cell membrane	
	EO-298	Connexin43	NM_012567	mp	Cell membrane	
Extracellular matrix (components)	EO-081-a	Endomucin-1	XM_215710	s	Extracellular	
	EO-003	Ameloblastin	NM_012900	s	Extracellular	
	EO-015-s	Amelogenin	U51195	s	Extracellular	
	EO-086-a	Biglycan	NM_017087	s	Extracellular	
	EO-195-a	Collagen I a1	XM_213440	s	Extracellular	
	EO-064	Collagen I a2	NM_053356	s	Extracellular	
	EO-016	Decorin	XM_343201	s	Extracellular	
	EO-361-a	Enamelin	U82698	s	Extracellular	
	EO-202	Extracellular matrix protein 1	XM_215645	s	Extracellular	
	EO-237-a	Lumican	X84039	s	Extracellular	
	EO-132	Matrix gamma-carboxyglutamate protein	J03026	s	Extracellular	
	EO-020	Mimcan	XM_214441	s	Extracellular	
	EO-047	Osteonectin	D28875	s	Extracellular	
	EO-318	Osteopontin	M14656	s	Extracellular	
	Extracellular matrix (metabolism)	EO-090	Alkaline phosphatase	NM_013059	GPI	Cell surface
EO-041-a		Plasma glutamate carboxypeptidase	AF131077	I	Cell membrane	
EO-110-a		Ectonucleotide pyrophosphatase/ phosphodiesterase 2	NM_057104	II	Cell membrane	
EO-082-a		Lysyl hydroxylase-2	AJ430861	s	ER [‡]	
EO-338-a		Acid phosphatase	NM_020072	s	Extracellular	
EO-002-a		Carbonic anhydrase 6	XM_216584	s	Extracellular	
EO-300-a		Cathepsin K	AF010306	s	Extracellular	
EO-062-a		Matrix metalloproteinase 19	XM_222317	s	Extracellular	
EO-224		Matrix metalloproteinase 20	XM_235796	s	Extracellular	
EO-214-a		Peroxiredoxin 4	BC059122	s	Extracellular	
EO-393-a		Phospholipase A2, group 5	U03763	s	Extracellular	
EO-028-a		Plasminogen activator, urokinase	NM_013085	s	Extracellular	
Secretory pathway (related to)		EO-337-a	Signal sequence receptor beta	XM_215619	I	ER
		EO-545-a	Signal sequence receptor delta	BC058482	I	ER
		EO-472-a	FK506-binding protein 10	XM_340901	s	ER
	EO-386-a	Glucose regulated protein 78-kDa	BC062017	s	ER	
	EO-533-a	Mesoderm development candidate 2	XM_218854	s	ER	
	EO-286-a	Protein disulfide isomerase-related protein ERp72	BC061535	s	ER	
	EO-005	Reticulocalbin	AJ001929	s	ER	
	EO-121	Lysosomal membrane glycoprotein 2	NM_017068	I	Lysosomal	
	EO-034	Mannose-6-phosphate receptor, cation-dependent	BC079226	I	Lysosomal	
	EO-098	ERG-28/C14ORF1/RIKEN 1190004E09	XM_216756	mp	Lysosomal	
	EO-001-a	Lysosomal-associated protein transmembrane 4 alpha	NM_199384	mp	Lysosomal	
	EO-553-a	Carboxypeptidase C	NM_008906	s	Lysosomal	
	EO-193	Serglycin	K02934	s	Extracellular	
	Signaling/growth factors (related to)	EO-269	Coxsackie virus and adenovirus receptor	XM_340962	I	Cell membrane
		EO-075-a	FXYD domain-containing ion transport regulator 3	NM_172317	I	Cell membrane
EO-066-a		FXYD domain-containing ion transport regulator 5	NM_021909	I	Cell membrane	
EO-085-a		Integrin, beta 6	BC079069	I	Cell membrane	
EO-514-a		Toll-like receptor 8	NM_133212	I	Cell membrane	
EO-079-a		Transferrin receptor	XM_340999	II	Cell membrane	
EO-009-a		Apoptosis-related protein-3	XM_216650	mp	Cell membrane	
EO-366-a		Mid-1-related chloride channel 1	NM_133414	mp	Cell membrane	
EO-247		p53 apoptosis effector related to PMP-22 (Perp)	XM_214953	mp	Cell membrane	
EO-234-a		Follistatin-related protein	U06864	s	Extracellular	
EO-194		Midkine	NM_030859	s	Extracellular	
EO-013-a		Monokine induced by gamma interferon (Cxcl9)	AF537208	s	Extracellular	
EO-156-a		Neural stem cell-derived neuronal survival protein	AF475282	s	Extracellular	
EO-550-a		Neurotensin/Neuromedin N	XM_216884	s	Extracellular	
EO-161		Odorant-binding protein 1F	NM_138903	s	Extracellular	
EO-152-a	Periostin	XM_342245	s	Extracellular		
EO-388-a	Pleiotrophin	NM_017066	s	Extracellular		
EO-123-a	Stem cell growth factor	XM_344888	s	Extracellular		
EO-171-a	Stromal cell-derived factor 2	XM_213277	s	Extracellular		

Table 1
Continued

Function*	Clone ID	Name of hit	GenBank accession number	Protein type [†]	Cellular localization
Miscellaneous	EO-018-a	Thymus cell antigen 1, theta	NM_012673	GPI	Cell surface
	EO-300	HLA class II histocompatibility antigen, DQ(2) alpha	X14879	I	Cell membrane
	EO-291-a	CD74 antigen	BC059152	II	Cell membrane
	EO-053-a	Small membrane protein 1	XM_216545	mp	Cell membrane
	EO-479-a	Tumor differentially expressed 1, like	NM_182951	mp	Cell membrane
	EO-010-a	Transmembrane protein 4/MSAP	CF110471	II	ER
	EO-281	GlcNAc-1-P transferase	BC063184	mp	ER
	EO-072-a	Epididymal major secretory protein ME1	BC058132	mp	Lysosomal
	EO-394-a	$\alpha_2\mu$ -globulin	AB039828	s	Extracellular
	EO-005-s	Apolipoprotein E	NM_138828	s	Extracellular
	EO-030	β 2-microglobulin	Y00441	s	Extracellular
	EO-002	Casein kappa	K02598	s	Extracellular
	EO-015	Clusterin	NM_012679	s	Extracellular
	EO-011-a	Complement component factor h	NM_130409	s	Extracellular
	EO-159	Glycoprotein 38	NM_019358	s	Extracellular
	EO-227	Lecithin:cholesterol acyltransferase	U62803	s	Extracellular
	EO-105-a	Milk fat globule-EGF factor 8 protein	NM_012811	s	Extracellular
	EO-095	Prion protein	XM_346677	s	Extracellular
	EO-161-a	Prosaposin	NM_013013	s	Extracellular
	EO-256-a	RAR-responder protein 2 (TIG2)	XM_216142	s	Extracellular

*Non-mutually exclusive.

[†]I, Type I; II, Type II; mp, multipass; s, secreted; GPI, GPI-anchored.[‡]ER, endoplasmic reticulum.

Table 2

List of unknown clones retrieved from the screen of the enamel organ library for cDNAs encoding a signal peptide

Clone ID	Name of hit	GenBank accession number	Signal peptide prediction*	Expression data			Site [‡]
				ESTs [†]			
				Human	Mouse	Rat	
EO-560-a	Riken 1810020G14, similar to	XM_216763	0.919	359	129	46	Ubiquitous (IS)
EO-525-a	Riken 2310033H11, similar to	XM_237786	0.046 [¶]	378	231	50	Wide (IS)
EO-081	Riken 2610001J05, similar to	AK011259	1.000	474	136	33	Ubiquitous (IS)
EO-011	Riken 2610003J06, similar to	XM_340772	0.805	133	246	36	Ubiquitous (E; IS)
EO-416-a	C11orf3, similar to	XM_215787	1.000	366	227	35	Ubiquitous (IS)
EO-009 ^{††}	Riken 2310011G06, similar to	XM_132137	0.997	34	16	3	EOs (E)
EO-063	Riken 5430401F13, similar to	XM_342772	1.000	0	9	6	EOs (E)
EO-017 ^{††}	Riken 5430427O21, similar to	XM_284166	1.000	4	8	5	EOs (E)
EO-463-a	Novel [§]	–	1.000	0	0	0	EOs (E)
EO-014	Novel	–	0.986	0	0	0	ND**

*Cleavable signal peptide probability according to SignalP 3.0 HMM.

[†]ESTs: expressed sequenced tags as compiled in Unigene (NCBI).[‡]E: experimental (northern blotting); IS: *in silico* (bioinformatics).[§]Indicates no hit on any databases.[¶]Signal anchor.

**Not determined.

^{††}Sequences have been deposited at GenBank, accession numbers DQ198380 and DQ198381.

EO-463) in rat EOs (Table 2). An example of northern blotting for *EO-009* is shown in Fig. 1. A predominant and abundant transcript, of ≈ 1.3 kb in size, was detected almost exclusively in EOs (Fig. 1, closed arrow head). A much fainter and larger transcript (≈ 2.2 kb) was also observed (open arrow head). The smaller transcript was also seen in periodontal ligaments but at a much weaker intensity ($\leq 2\%$). None of the other tissues, including muscles or bones, showed detectable expression level by northern blotting. The site of

expression of the *EO-009* gene was further refined on RNA samples extracted from individual secretory (S) and maturation (M) regions of the EOs (Fig. 1, right panel). The *EO-009* gene was mostly expressed in the M region, with a pattern identical to that described for the whole EOs.

In order to gain more information on the protein encoded by the rat *EO-009* gene, its full-length cDNA was obtained (Fig. 2). Starting at the first ATG (Fig. 2B, boxed), which was in an adequate Kozak context, the

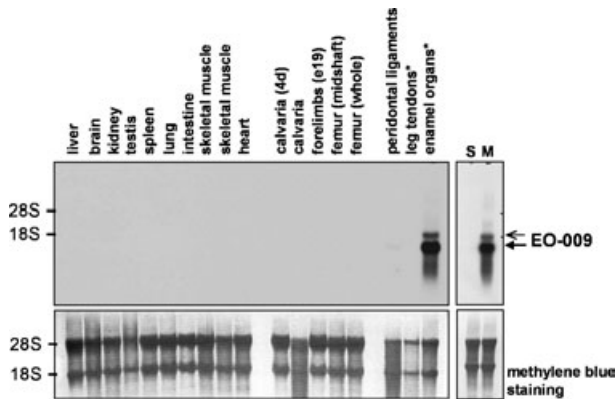


Fig. 1. Northern blotting analyses of the expression of rat *EO-009* in selected tissues. Total RNA was from specified adult rat tissues unless otherwise indicated (4 d, 4 d old; e19, embryo at d 19). Ten (*) or 20 μ g of RNA was separated on agarose gels and processed for hybridization. The probe used was the *EO-009* cDNA fragment (see Fig. 2A) labeled with 32 P. Staining of the membranes with methylene blue indicates overall integrity and equal loading of samples. The 28S and 18S ribosomal RNA are indicated at the left of the blot. S and M indicate RNA isolated from the secretory and maturation stages of rat enamel organs (EOs), respectively. The arrows point to the two different transcripts detected in EOs.

1189-bp long cDNA isolated had an open reading frame of 278 amino acids. The size of the full-length cDNA is in accordance with that of the lower size transcript detected by northern blotting (Fig. 1). A consensus polyadenylation signal was found just preceding the polyA stretch (Fig. 2A,B). BLAST search on the latest rat genome at Ensembl (release RGSC 3.4) indicated that the *EO-009* gene is composed of 12 exons and spans \approx 9 kb. The resulting *EO-009* mature protein (devoid of signal peptide) has a computed pI of 4.5 and is particularly rich in glutamine (16%), proline (12%), leucine (9%), and serine (7%) (Fig. 2B). Protein sequences for human (279 amino acids) and mouse (273 amino acids), derived from available sequences (UniGene Hs.143811 and XM_216763, respectively), were compared to the rat *EO-009* protein (Fig. 3). The alignment from the three species showed an overall high degree of similarity: 87% for rat vs. mouse; 72% for rat vs. human; and 75% for mouse vs. human. A portion of the deduced human *EO-009* protein (residues 127–279), conserved in rat and mouse, was identical to a protein named APin (NP_060325).

Discussion

The present study was aimed at characterizing the secretome of the rat EOs using a previously described viral-based signal-trap procedure (14). This method utilizes mammalian cells as filters to allow identification of cDNA fragments only encoding signal peptides from very complex cDNA libraries. As part of this single screen, we found the main four secreted gene products known to be expressed in the EOs (amelogenin,

ameloblastin, enamelin, MMP-20). One notable exception was kallikrein-4 (KLK-4), which may have escaped detection because our analysis was finite. Splice variants of the amelogenin transcript were also identified (data not shown). Intersection of clones found in the present study and in a random sequencing project on partial cDNAs isolated from the pulp and apical part of rat EOs (6), included collagen1 α 1, collagen1 α 2, decorin, osteoglycin, osteonectin, clusterin, phospholipase A2-5, ERp72, and LAPT4. The majority of clones found, however, encoded proteins with known functions in other tissues but not as yet suspected or known to be expressed in teeth. Carbonic anhydrase 6 is one such example and has been described by SMITH *et al.* (17). Despite being reported, several other hit sequences still encode proteins (MSAP, TIG2, APR-3, c11orf3) with ill-defined functions, especially in the context of tooth physiology (21–24). As a whole, they typify a wide spectrum of different topologies (type I, type II, multipass, glycosylphosphatidylinositol-anchor, and secreted) and cellular locations (endoplasmic reticulum, lysosomal, extracellular, cell membrane, and cell surface). While the vast majority of clones could reasonably be assigned to epithelial cells, a minority was more typical of mesenchymal cells (e.g. cathepsin K, collagens) and may have originated from adjacent tissues dissected along with EOs.

Most importantly, 10 cDNA fragments found were considered as being unknown, based on the paucity of information regarding the function and sequence of the putative proteins they encode. A detailed *in silico* analysis through UniGene in human, mouse, and rat provided a rapid view on their abundance and expression pattern. Half had a vast repertoire of ESTs and thus were qualified as being either widely or ubiquitously expressed (Table 2). A confirmation that the predicted expression pattern is valid came from northern blotting with the *EO-011* fragment, which showed uniform expression in a panel of more than 50 adult and embryonic mouse RNAs (data not shown). Although they were not examined further, these genes nonetheless represent equally interesting targets to investigate in the tooth context. In fact, the ubiquitous expression of a gene does not necessarily imply that the protein it encodes is a non-tissue-specific housekeeping molecule (25). Considering the now almost complete sequencing and annotation of genomes from several species, it was surprising to find two genes (*EO-014* and *EO-463*) that did not hit on any ESTs publicly available. This finding validates the use of the present functional genomic approach to help identify novel genes.

Closer inspection of the matching Riken clones 2310011G06 (*EO-009*) and 5430401F13 (*EO-063*) revealed that they correspond, respectively, to clones AF020442 (9) and R47092 (6), both sequenced from cDNAs originating from rat EOs. Because the latter were partial sequences not comprising their naive 5' extremity (where the signal peptide usually lies), they were not defined as being secreted proteins, highlighting the usefulness of the present expression-screening approach. The full-length cloning of the *EO-009* cDNA

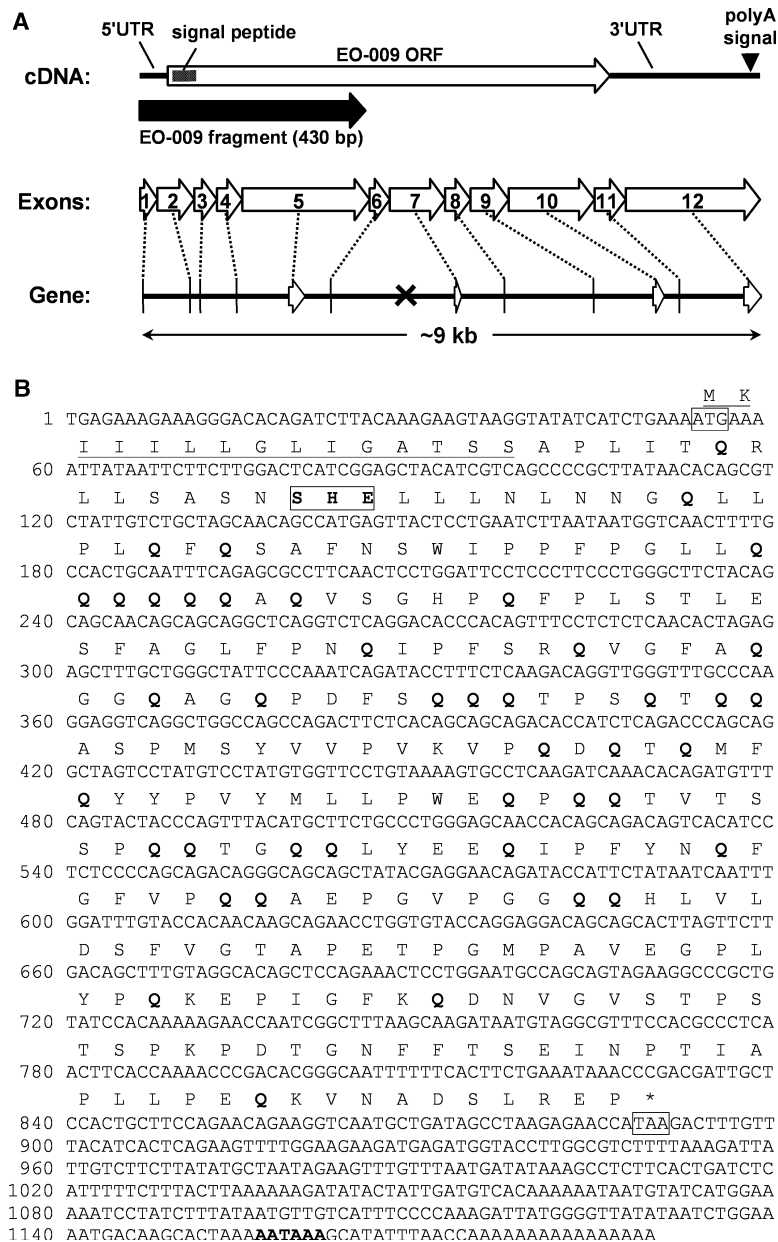


Fig. 2. Cloning of the rat EO-009 full-length cDNA. (A) Schematic representation of the rat EO-009 cDNA with its open reading frame (ORF). The locations of the initial EO-009 fragment (black arrow), as well as the exons (open arrows) comprising the cDNA, are illustrated. The architecture of the *EO-009* gene on chromosome 14, with positions of the exons, is shown at the bottom. The cross denotes a gap in the rat genome sequence. (B) Nucleotide sequence and corresponding ORF for EO-009. The ATG and stop codon are boxed, and the polyadenylation signal is in bold. The signal peptide sequence is underlined, and the abundant glutamine (Q) residues are shown in bold. The putative phosphorylation site (SHE) is boxed. UTR, untranslated region.

(Fig. 2) revealed that it represents a longer form of a 153-amino acid protein identified by SOLOMON *et al.* (26) from calcifying epithelial odontogenic (Pindborg) human tumor samples, which they named APin. In that work, polypeptides corresponding to residues 127–170 and 127–172 of human EO-009 were extracted and sequenced from the characteristic amyloid deposits present in such tumors. Although these fragments were postulated to be derived from a longer secreted precursor protein, the identified 153-amino acid APin protein does not have a signal peptide at its N terminus

(26). Because the size of the full-length EO-009 cDNA cloned herein closely matches the lower size, more abundant transcript detected in rat EOs by northern blotting, it probably represents the full-length sequence. The deduced human EO-009 protein from reconstituted full-length cDNA, and its alignment with that of rat and mouse, showed that it is highly conserved across species. The most salient features of the mature protein sequence are its very acidic pI (4.5), and the abundance of glutamine and proline (total 28%) and hydrophobic (30%) residues. Apart from potential serine/threonine

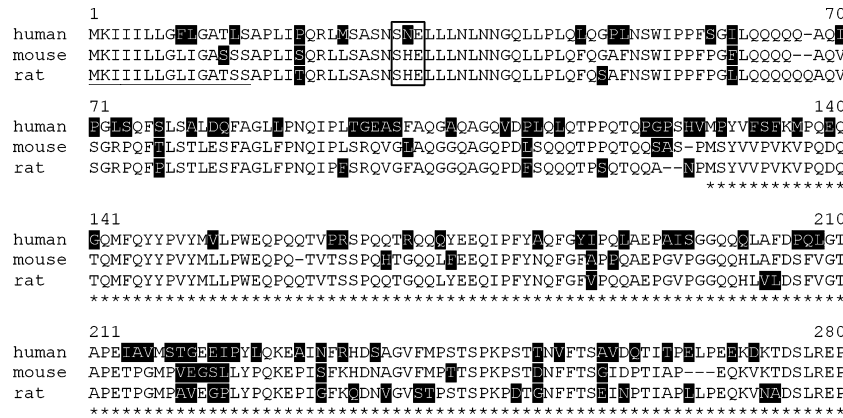


Fig. 3. Protein sequence alignment for the rat, mouse, and predicted human EO-009. The human and mouse proteins were derived from the genomic sequence and XM_132137, respectively. Non-identical residues are highlighted in black. The signal peptide region (line) is similar and present in the three species. The sequence corresponding to the short form of the human APin protein is marked by asterisks. The numbering on top refers to the human sequence.

phosphorylation sites, the molecule contains no other obvious motifs or post-translational modifications. Another level of complexity, not present in rat or mouse, seems to exist in human where a splice variant, lacking exon 2, is supported by several ESTs (27). The shorter isoform would still be secreted but be missing residues 18–31.

Interestingly, the *EO-009/APin* gene resides on a locus of rat chromosome 14 comprising the well-defined cluster of casein and salivary gland genes, as well as *ameloblastin* and *enamelin* (27,28). In fact, the *EO-009* gene lies next to the novel *follicular dendritic cell-secreted gene* (29), with its immediate neighbor being the *casein kappa* gene that we also found in the present screen (Table 1). The *EO-009* gene structure possesses two of the characteristics present in most genes of the locus (28), namely phase-0 introns (codons are not disrupted by their boundaries) and the SXE (where X is any amino acids) phosphorylation site encoded by the end of exon 3 (Fig. 2). A very similar chromosomal architecture is also observed in the human (chromosome 4) and mouse (chromosome 5). The northern blotting experiment convincingly showed that the *EO-009* transcript is highly and specifically expressed in the M stage of rat EOs. It is still possible, however, that lower expression of *EO-009* may occur in other tissues (thymus, lung, heart, trachea, lachrymal glands, colon) as indicated by the source of matching sequences in UniGene (Hs.143811 and Mm.79700).

In summary, the present data indicate that secreted and membrane-embedded proteins still remain to be identified in EOs. One of the upcoming challenges will be to assign functions to proteins possessing unknown motifs and to determine whether there are diseases associated with them. A more definitive answer as to whether they play crucial roles in the EO formation and integrity should come, at least in part, through *in situ* hybridization and immunolocalization studies, and ultimately transgenesis and targeted gene disruption in mouse.

Acknowledgements – This work was supported by the Canadian Institutes of Health Research.

References

1. TJALMA H, ANTELMANN H, JONGBLOED JD, BRAUN PG, DARMON E, DORENBOS R, DUBOIS JY, WESTERS H, ZANEN G, QUAX WJ, KUIPERS OP, BRON S, HECKER M, VAN DIJL JM. Proteomics of protein secretion by *Bacillus subtilis*: separating the ‘secrets’ of the secretome. *Microbiol Mol Biol Rev* 2004; **68**: 207–233.
2. MACDOUGALL M. Dental structural diseases mapping to human chromosome 4q21. *Connect Tissue Res* 2003; **44**: 285–291.
3. CLARK HF, GURNEY AL, ABAYA E, BAKER K, BALDWIN D, BRUSH J, CHEN J, CHOW B, CHUI C, CROWLEY C, CURRELL B, DEUEL B, DOWD P, EATON D, FOSTER J, GRIMALDI C, GU Q, HASS PE, HELDENS S, HUANG A, KIM HS, KLIMOWSKI L, JIN Y, JOHNSON S, LEE J, LEWIS L, LIAO D, MARK M, ROBBIE E, SANCHEZ C, SCHOENFELD J, SESHAGIRI S, SIMMONS L, SINGH J, SMITH V, STINSON J, VAGTS A, VANDLEN R, WATANABE C, WIEAND D, WOODS K, XIE MH, YANSURA D, YI S, YU G, YUAN J, ZHANG M, ZHANG Z, GODDARD A, WOOD WI, GODOWSKI P, GRAY A. The secreted protein discovery initiative (SPDI), a large-scale effort to identify novel human secreted and transmembrane proteins: a bioinformatics assessment. *Genome Res* 2003; **13**: 2265–2270.
4. GRIMMOND SM, MIRANDA KC, YUAN Z, DAVIS MJ, HUME DA, YAGI K, TOMINAGA N, BONO H, HAYASHIZAKI Y, OKAZAKI Y, TEASDALE RD. The mouse secretome: functional classification of the proteins secreted into the extracellular environment. *Genome Res* 2003; **13**: 1350–1359.
5. KLEE EW, CARLSON DF, FAHRENKRUG SC, EKKER SC, ELLIS LB. Identifying secretomes in people, pufferfish and pigs. *Nucleic Acids Res* 2004; **32**: 1414–1421.
6. MATSUKI Y, NAKASHIMA M, AMIZUKA N, WARSHAWSKY H, GOLTZMAN D, YAMADA KM, YAMADA Y. A compilation of partial sequences of randomly selected cDNA clones from the rat incisor. *J Dent Res* 1995; **74**: 307–312.
7. UNDA FJ, IEHARA N, DE VS, LA FM, VILAXA A, COBOURNE M, SHARPE PT, YAMADA Y. Analysis of cDNAs from a mouse embryo tooth library: identification of novel genes during tooth development. *Connect Tissue Res* 2002; **43**: 176–179.
8. KREBSBACH PH, LEE SK, MATSUKI Y, KOZAK CA, YAMADA KM, YAMADA Y. Full-length sequence, localization, and

- chromosomal mapping of ameloblastin. A novel tooth-specific gene. *J Biol Chem* 1996; **271**: 4431–4435.
9. BONASS WA, KIRKHAM J, SHORE RC, BROOKES SJ, GODFREY CL, ROBINSON C. Identification of rat enamel organ RNA transcripts using differential-display. *Connect Tissue Res* 1998; **38**: 249–256.
 10. CHEN HLEDER P. A new signal sequence trap using alkaline phosphatase as a reporter. *Nucleic Acids Res* 1999; **27**: 1219–1222.
 11. JACOBS KA, COLLINS-RACIE LA, COLBERT M, DUCKETT M, GOLDEN-FLEET M, KELLEHER K, KRIZ R, LAVALLIE ER, MERBERG D, SPAULDING V, STOVER J, WILLIAMSON MJ, MCCOY JM. A genetic selection for isolating cDNAs encoding secreted proteins. *Gene* 1997; **198**: 289–296.
 12. KLEIN RD, GU Q, GODDARD A, ROSENTHAL A. Selection for genes encoding secreted proteins and receptors. *Proc Natl Acad Sci USA* 1996; **93**: 7108–7113.
 13. KOJIMA TKITAMURA T. A signal sequence trap based on a constitutively active cytokine receptor. *Nat Biotechnol* 1999; **17**: 487–490.
 14. MOFFATT P, SALOIS P, GAUMOND MH, ST-AMANT N, GODIN E, LANCTOT C. Engineered viruses to select genes encoding secreted and membrane-bound proteins in mammalian cells. *Nucleic Acids Res* 2002; **30**: 4285–4294.
 15. MOFFATT P, SALOIS P, ST-AMANT N, GAUMOND MH, LANCTOT C. Identification of a conserved cluster of skin-specific genes encoding secreted proteins. *Gene* 2004; **334**: 123–131.
 16. THOMAS G, MOFFATT P, SALOIS P, GAUMOND MH, GINGRAS R, GODIN E, MIAO D, GOLTZMAN D, LANCTOT C. Osteocrin, a novel bone-specific secreted protein that modulates the osteoblast phenotype. *J Biol Chem* 2003; **278**: 50563–50571.
 17. SMITH CE, NANJI A, MOFFATT P. Evidence by signal peptide trap technology for expression of carbonic anhydrase 6 in rat incisor enamel organ. *Eur J Oral Sci* 2006; **114** (Suppl.1): 147–153.
 18. CHURCH GM, GILBERT W. Genomic sequencing. *Proc Natl Acad Sci USA* 1984; **81**: 1991–1995.
 19. BENDTSEN JD, NIELSEN H, VON HG, BRUNAK S. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 2004; **340**: 783–795.
 20. NANJI A, ZALZAL S, SMITH CE. Application of backscattered electron imaging and lectin-gold cytochemistry to visualize the distribution of glycoconjugates in a basal lamina. *Scanning Microsc* 1987; **1**: 1963–1970.
 21. BORNHAUSER BC, OLSSON PA, LINDHOLM D. MSAP is a novel MIR-interacting protein that enhances neurite outgrowth and increases myosin regulatory light chain. *J Biol Chem* 2003; **278**: 35412–35420.
 22. MEDER W, WENDLAND M, BUSMANN A, KUTZLEB C, SPODSBERG N, JOHN H, RICHTER R, SCHLEUDER D, MEYER M, FORSSMANN WG. Characterization of human circulating TIG2 as a ligand for the orphan receptor ChemR23. *FEBS Lett* 2003; **555**: 495–499.
 23. O'BRIEN KP, TAPIA-PAEZ I, STAHL-BACKDAHL M, KEDRA D, DUMANSKI JP. Characterization of five novel human genes in the 11q13-q22 region. *Biochem Biophys Res Commun* 2000; **273**: 90–94.
 24. ZHU F, YAN W, ZHAO ZL, CHAI YB, LU F, WANG Q, PENG WD, YANG AG, WANG CJ. Improved PCR-based subtractive hybridization strategy for cloning differentially expressed genes. *Biotechniques* 2000; **29**: 310–313.
 25. YANG X, KARSENTY G. ATF4, the osteoblast accumulation of which is determined post-translationally, can induce osteoblast-specific gene expression in non-osteoblastic cells. *J Biol Chem* 2004; **279**: 47109–47114.
 26. SOLOMON A, MURPHY CL, WEAVER K, WEISS DT, HRNCIC R, EULITZ M, DONNELL RL, SLETTEN K, WESTERMARK G, WESTERMARK P. Calcifying epithelial odontogenic (Pindborg) tumor-associated amyloid consists of a novel human protein. *J Lab Clin Med* 2003; **142**: 348–355.
 27. RIJNKELS M, ELNITSKI L, MILLER W, ROSEN JM. Multispecies comparative analysis of a mammalian-specific genomic domain encoding secretory proteins. *Genomics* 2003; **82**: 417–432.
 28. KAWASAKI K, WEISS KM. Mineralized tissue and vertebrate evolution: the secretory calcium-binding phosphoprotein gene cluster. *Proc Natl Acad Sci USA* 2003; **100**: 4060–4065.
 29. MARSHALL AJ, DU Q, DRAVES KE, SHIKISHIMA Y, HAYGLASS KT, CLARK EA. FDC-SP, a novel secreted protein expressed by follicular dendritic cells. *J Immunol* 2002; **169**: 2381–2389.