AMELOBLASTIN:
EVOLUTION, STRUCTURE AND FUNCTION IN ENAMEL FORMATION

by

Yong-Hee Patricia Chun

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Oral Health Sciences)
in The University of Michigan
2009

Doctoral Committee:

Professor Jan C.-C. Hu, Co-Chair
Professor James P. Simmer, Co-Chair
Professor Robert S. Fuller
Professor Paul H. Krebsbach
Professor Laurie K. McCauley

DEDICATION

To my parents
Prof. em. Dr.-Ing. Chung-Hwan and Dipl. Oecotroph. Chung-Hoon Chun

ACKNOWLEDGEMENTS

My doctoral studies in the Oral Health Science Program at the University of Michigan have been a very special experience to me. I would like to thank all individuals who encouraged and supported me throughout the doctoral training:

Dr. Jan Hu and Dr. James Simmer served as my research advisors and supported me to develop scientific and professional skills. They provided me with exceptional scientific resources, guidance and incredible scientific inspiration in relentless discussions. Their personal commitment to excellence in science, ethics and enthusiasm throughout my training is greatly appreciated.

I am very grateful to the members of my dissertation committee:

Dr. Robert Fuller, for his contributions as a thesis committee member, for his commitment to educating graduate students and for being a very inspiring teacher.

Dr. Paul Krebsbach, for support of my research from the Biologic and Materials Sciences Department. His scientific thoughts and discussions during courses, seminars, and journal clubs are highly appreciated.

Dr. Laurie McCauley, for her continuous support, generosity and mentoring since I came to Michigan in 2000. She has provided me with guidance in scientific and career decisions. Her approaches to science and her scientific insights are always valued.

The Oral Health Sciences Program directed by Dr. Charlotte Mistretta has provided me with an incredibly comprehensive training in science and professional development. Her

commitment to my education made a difference in the quality of training I have received. She constantly offered guidance through all stages of my PhD training balancing encouragement to explore science broadly and freedom to develop skills. I thank her for all the opportunities I had through involvement in the Michigan Oral Health Science Student Association and the Oral Health Sciences program.

I have received diverse support and training for conducting experiments and would like to thank the following individuals for their friendliness and generosity sharing knowledge and experimental material.

Present and past members of Dr. Hu and Dr. Simmer's laboratory:

Dr. Yasuo Yamakoshi, for providing fundamental teaching and collaboration in protein biochemistry.

Dr. Yuanyuan Hu, for support in molecular biology, mouse genetics and SEM analysis.

Dr. Takatoshi Nagano, for sharing his experimental experience and collaboration.

Dr. Takanori Iwata, for teaching me and providing support in molecular biology.

Dr. Yuhe Lu, for support and collaboration in molecular biology and mouse genetics.

Fumiko Yamakoshi, for technical support with cell culture.

Heechan Cha, Dr. Albert Hsun-Liang Chan, Dr. Sandra Hui-Chen Chan, Dr. Yumei Ding, Dr. Jung-Wook Kim, Dr. Kazuyuki Kobayashi, Pik Rangsikorn Lertlam, Dr. Suresh Madathilparambil, Dr. Petros Papagerakis, Amelia Richardson, Dr. Shuhei Tsuchiya for support in experiments and for creating a vibrant scientific environment.

Dr. Yoshihiko Yamada, at the National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, for providing the Ameloblastin null mouse.

Michigan State University Meat Laboratory (East Lansing, MI), for porcine teeth.

Dr. Taocong Jin, Molecular Biology Core Laboratory, for support with real-time PCR.

Dr. Yusuke Shiozawa and Dr. Russell Taichman, for sharing knowledge and support in real-time PCR experiments.

The Transgenic Animal Model Core, University of Michigan for generating transgenic mice.

Staff of the Animal Research Facility, School of Dentistry for technical support in animal maintenance.

I am very grateful to all Oral Health Sciences faculty for encouragement and commitment to teaching in many scientific discussions.

Dr. Jacques Nör served as my academic advisor and I am very thankful for his enthusiastic guidance and mentorship during my entire training.

Dr. Russell Taichman provided me with advice and mentorship during my periodontal and doctoral training.

Dean Peter Polverini has been very helpful during my doctoral training by creating a unique scientific environment in the School of Dentistry, and by providing financial support and mentorship.

I am very thankful to the Oral Health Sciences Research Office with Patricia Schultz, Misty Gravelin, Manette London, and Charlene Erickson for their administrative support and kindness.

All current and past Oral Health Sciences students are thanked for their comradeship and friendliness.

I would like to thank faculty and staff from the Periodontics & Oral Medicine Department and the Biologic and Materials Sciences Department for their support.

Dr. Hom-Lay Wang provided me with training and mentoring related to periodontics and professional development.

Dr. Martha Somerman inspired me to basic sciences and academic dentistry.

Dr. Barbara Smith introduced me to the University of Michigan and her true friendship is very valuable to me.

Finally, I would like to thank my parents.

PREFACE

This thesis includes significant work of Drs. Jan Hu, James Simmer, Yasuo Yamakoshi, Fumiko Yamakoshi, Takanori Iwata, Yuhe Lu, Yuanyuan Hu, Takatoshi Nagano, Jung-Wook Kim.

Description of my contribution to the work presented in this thesis:

I wrote all Chapters of this thesis. The work of Chapter 2 has been published in Eur J Oral Sci (2004, 114, 78-85). I have updated the references and added content to the discussion. Drs. Jan Hu and James Simmer contributed to the experimental design and thesis writing.

I wrote the 'Introduction' in Chapter 1 summarizing the specific aims, and background and significance on enamel formation.

Chapter 2 of this work has been published in Eur J Oral Sci (2004, 114, 78-85). The data on SPARC purification from porcine dentin has been generated by Dr. Yasuo Yamakoshi. The SPARC cDNA was cloned with assistance of Dr. Jung-Wook Kim. The modeling of SPARC was performed by myself under guidance of Drs. Jan Hu and James Simmer.

In Chapter 3 the ameloblastin construct for protein expression was cloned by Dr. Takanori Iwata. Fumiko Yamakoshi provided assistance with cell culture. Dr. Yasuo Yamakoshi provided fundamental teaching in protein biochemistry. He performed the experiment on deglycosylation of ameloblastin (Ambn). I performed the generation of

stable cell lines overexpressing Ambn, optimization of the large-scale culture and purification of recombinant Ambn.

In chapter 4, I produced recombinant Mmp-20. Klk4 was obtained from native sources by Dr. Yasuo Yamakoshi. The cleavage experiments were performed by myself. Dr. Yasuo Yamakoshi helped me to identify bands to be sent for N-terminal sequencing. The N-terminal sequencing was performed by Nancy Williams, Yale University. Dr. Takatoshi Nagano performed cleavage of the N-Ambn1 peptide by Mmp-20 and Klk4. The LC-MS/MS analysis was performed by David Allan, NextGen Science, Ann Arbor, MI. The work from Chapter 3 and 4 is not yet published.

In Chapter 5, transgenic mice were produced at Animal Core, University of Michigan. The transgene construct was prepared by Dr. Yuhe Lu. The Ambn null mouse was obtained from Dr. Yoshihiko Yamada, NIDCR. Dr. Yuanyuan Hu bred Ambn null mice for rescue experiments and generated the SEM images. Pik Lertlam supported the management of the mouse colony. Drs. Taocong Jin and Yusuke Shiozawa provided technical support in experimental design of real-time PCR experiments. I performed the matings, genotyping of the colony and real-time PCR experiments. The work from Chapter 5 is not yet published.

TABLE OF CONTENTS

DEDICATION ..... **ii**

ACKNOWLEDGEMENTS ..... **iii**

PREFACE… ..... **vii**

LIST OF FIGURES ..... **xi**

LIST OF TABLES ..... **xiv**

ABSTRACT ..... **xv**

CHAPTER 1 ..... **1**

INTRODUCTION

Problem Statement 1
General Hypothesis 2
Specific Aims 3
Background and Significance 5
References 15

CHAPTER 2 ..... **19**

EVOLUTION OF GENES IN ENAMEL FORMATION

Abstract 19
Introduction 20
Materials & Methods 26
Results 30
Discussion 39
References 52

CHAPTER 3 ..... **60**

RECOMBINANT PORCINE AMELOBLASTIN

Abstract 60
Introduction 61
Materials and Methods 67
Results 77
Discussion 83
References 85

CHAPTER 4 87

PROCESSING AND DEGRADATION OF AMELOBLASTIN

Abstract 87
Introduction 88
Materials and Methods 92
Results 103
Discussion 119
References 126

CHAPTER 5 130

AMELOBLASTIN FUNCTION *IN VIVO*

Abstract 130
Introduction 131
Materials & Methods 133
Results 146
Discussion 158
References 165

CHAPTER 6 169

CONCLUSION

Summary 169
Prospects 173
References 175

LIST OF FIGURES

CHAPTER 1

Figure 1.1: Stages of enamel formation. 7

Figure 1.2: Mineralization front formed by Tomes' process of the ameloblast and crystallites. 9

Figure 1.3: Scanning electron micrograph (SEM) of the surface of an acid-etched, ground section of mature mouse incisal dental enamel. 10

CHAPTER 2

Figure 2.1: Gene cluster of secreted calcium-binding phosphoproteins (SCPPs). 21

Figure 2.2: The gene structure is conserved in SCPPs. 23

Figure 2.3: Isolation of SPARC from porcine dentin. 30

Figure 2.4: Analysis of porcine SPARC cDNA and deduced amino acid sequences. 32

Figure 2.5: Analysis of aligned SPARC protein sequences. 34

Figure 2.6: Computer modeling of the porcine SPARC homodimer. 36

Figure 2.7: SCPP cluster in mammalian, avian, amphibian and fish species. 44

Figure 2.8: Alignment of signal peptides of human enamel SCPPs. 45

Figure 2.9: Conformation of Ambn changes in the presence of Ca^{2+}. 46

Figure 2.10: Multiple sequence alignment of Ambn using protein sequences of human, cattle, pig, rodents, caiman, and toad. 50

CHAPTER 3

Figure 3.1: Cross-section of porcine molars for immunolocalization of Ambn cleavage products. 62

Figure 3.2: Localization of Ambn in the sheath space..... 63

Figure 3.3: Ameloblastin structure from pig..... 65

Figure 3.4: Expression vector pEF6/V5-His-TOPO for porcine ameloblastin in mammalian cells. 69

Figure 3.5: Transfection of HEK293 with porcine ameloblastin results in expression of recombinant protein. 70

Figure 3.6: Large-scale production of recombinant ameloblastin. 79

Figure 3.7: HEK293 culture in the hollow fiber cartridge. 71

Figure 3.8: Analysis of expression and purification of recombinant porcine ameloblastin by SDS-PAGE. 80

Figure 3.9: Final purification of recombinant porcine ameloblastin. 81

Figure 3.10: Deglycosylation of rpAmbn by O-glycosidase and sialidase..... 82

CHAPTER 4

Figure 4.1: Mmp-20 and Klk4 mRNA detection in a day-28 mouse mandibular incisor. 89

Figure 4.2: Sequences of synthetic peptides simualting *in vivo* Ambn cleavage sites. 93

Figure 4.3: Expression vector for recombinant porcine Mmp-20 in *E. coli*. 95

Figure 4.4: Expression and purification of rpMmp-20 in *E. coli*..... 97

Figure 4.5: Fractionation of the catalytic domain of rpMmp-20 on a heparin affinity column. 98

Figure 4.6: Klk4 purified from native maturation stage enamel. 100

Figure 4.7: rpMmp-20 demonstrated enzymatic activity using rpAmel (P172) as substrate..... 104

Figure 4.8: Digestion of rpAmbn by Mmp-20 analyzed by SDS-PAGE and N-terminal sequencing. 107

Figure 4.9: Digestion of rpAmbn by Klk4 analyzed by SDS-PAGE und N-terminal sequencing. 110

Figure 4.10: Analysis of processing of N-Ambn1 by enamel proteinases. 114

Figure 4.11: Analysis of processing of C-Ambn2 by enamel proteinases..... 115

Figure 4.12: Analysis of processing of C-Ambn3 by enamel proteinases..... 116

Figure 4.13: RP-HPLC chromatograms of rpMmp-20 and Klk4 without substrate. 117

Figure 4.14: Summary of processing of ameloblastin. 119

Figure 4.15: Schematic of ameloblastin processing by Mmp-20. 123

CHAPTER 5

Figure 5.1: Enamel phenotype of the *Ambn* null mouse. 132

Figure 5.2: Strategy to generate transgene construct. 136

Figure 5.3: Design of primer sets to identify *Ambn* transgene incorporation.. 138

Figure 5.4: Generation of ameloblastin-null mice. 142

Figure 5.5: Mating of *Ambn* null with hemizygous transgenic *Ambn* mouse.. 144

Figure 5.6: PCR genotyping using primer sets at the *AmelX-5'Ambn* and *Ambn3'-AmelX* borders. 147

Figure 5.7: Overview of the tooth phenotype in adult $Ambn^{Tg}$ founder mice.. 148

Figure 5.8: Gene expression of $Ambn^{Tg}$, $Ambn^{WT}$ and $Amel^{WT}$ in dental tissues of founder offspring determined by real-time PCR..... 151

Figure 5.9: *Ambn* transgenic founder offspring selected for rescue experiment. 154

Figure 5.10: *Ambn* transgenic line derived from founder #2 crossed with *Ambn* null mouse. 155

Figure 5.11: SEM images of enamel from $Ambn^{+/+}$, $Ambn^{+-/Tg}$, $Ambn^{+/-}$, $Ambn^{-/-}$ animals on day 5. 158

LIST OF TABLES

CHAPTER 1

Table 1.1: Enamel matrix proteins. 6

CHAPTER 3

Table 3.1: Antibodies used in this study to detect recombinant porcine ameloblastin. 76

CHAPTER 5

Table 5.1: Generation of the transgene components by PCR: *AmelX promoter, Ambn* coding region, *AmelX3'*. 135

Table 5.2: Primers to amplify a product from the *AmelX-5'Ambn* border..... 138

Table 5.3: Primers to amplify a product from the *Ambn3'-AmelX* border..... 138

Table 5.4: Primer and probe sequences for real-time PCR on $Ambn^{Tg}$, $Ambn^{WT}$, $AmelX^{WT}$ and GAPDH. 140

Table 5.5: Primers to amplify a product from the neomycin phosphotransferase cassette. 143

Table 5.6: Primers to amplify a product from the *Ambn* intron 5. 143

Table 5.7: Overview of $Ambn^{Tg}$ founders and transmission rate of *Ambn* transgene. 149

Table 5.8: Summary of gene expression and tooth phenotype of transgene positive founder offspring. 153

ABSTRACT

Ameloblastin: Evolution, structure and function in enamel formation

by

Yong-Hee Patricia Chun

Co-Chairs: Jan C.-C. Hu and James P. Simmer

Dental enamel formation is dependent upon the secretion of three enamel proteins: amelogenin, enamelin, and ameloblastin (Ambn). These proteins contribute to a mineralization front on the secretory surface of ameloblasts. At the mineralization front, thin mineral ribbons grow in length, while enamelysin (Mmp-20) proteolytically processes the secreted enamel proteins. The absence of any one of these proteins results in major defects in the enamel layer. Once the enamel ribbons reach their final length, kallikrein 4 (Klk4) is secreted and degrades the residual enamel proteins, allowing the crystallites to widen and grow together. The goal of this work was to investigate the structure and function of Ambn and its evolution.

Ambn belongs to the secretory calcium-binding phosphoproteins, which descended from SPARC (secreted protein, acidic and rich in cysteine). All are involved in biomineralization. To better understand the structure of SPARC in mineralizing

systems, we isolated the SPARC cDNA and protein from developing pig teeth, and determined its quaternary structure by homology modeling after the human SPARC crystal structure.

Porcine Ambn is secreted as a 395 amino acid glycoprotein. Since Ambn is cleaved rapidly following its secretion, the intact protein has not been isolated from *in vivo* sources. We expressed recombinant porcine Ambn (rpAmbn) in mammalian (HEK293) cells. Recombinant Ambn was isolated by ammonium sulfate precipitation, metal- affinity chromatography, and reversed-phase high performance liquid chromatography. The purified, uncleaved protein migrated at 64 kDa on SDS-PAGE, was glycosylated, and based upon positive Stains-all staining, bound calcium.

Ambn processing is essential for enamel mineralization. *In vitro* Ambn cleavage sites generated by rpMmp-20 and native Klk4 were characterized using rpAmbn and synthetic fluorescent peptides. Mmp-20 cleaved rpAmbn at sites matching those of Ambn cleavage products isolated from immature enamel. In contrast, Klk4 degraded rpAmbn.

As enamel crystals cannot be synthesized *in vitro*, there are no functional assays for Ambn. We developed an *in vivo* functional assay by generating a transgenic mouse that expresses Ambn from the *AmelX* promoter. This transgenic mouse was crossed with the Ambn null mouse. The hybrid ($Ambn^{+/-Tg}$) showed normal enamel thickness and prisms similar to wildtype.

CHAPTER 1

INTRODUCTION

PROBLEM STATEMENT

During the secretory stage of dental enamel formation, the three major proteins secreted by ameloblasts are amelogenin, ameloblastin (Ambn), and enamelin. Proper mineralization of enamel depends upon the secretion of these proteins, as well as their being processed into smaller components and eventually degraded by proteases at maturation stage. Mice lacking expression of amelogenin, ameloblastin, enamelin, enamelysin (matrix metalloproteinase 20, Mmp-20) or kallikrein 4 (Klk4) exhibit severe enamel phenotypes. Human mutations in the genes encoding these enamel proteins (except Ambn) have been linked to amelogenesis imperfecta. The studies reported here were designed to advance our understanding of ameloblastin function during tooth enamel formation *in vitro* and *in vivo*.

The following Ambn post-translational modifications (PTMs) have been identified by the characterization of Ambn cleavage products isolated from porcine developing teeth: hydroxylation of Pro^{11} and Pro^{324}; O-linked glycosylations on Ser^{86} and Thr^{251}; and phosphorylation of Ser^{17}. Porcine ameloblastin is cleaved *in vivo* after Asn^{31}, Gln^{130}, Arg^{170}, Ala^{222}, Gly^{300}, Arg^{319}, and Ala^{342}. Uncleaved Ambn has never been isolated from developing teeth. A lack of pure, intact Ambn protein has precluded its structural and

functional characterization. Two ameloblastin isoforms are generated by alternative splicing. We have used a recombinant strategy to obtain intact Ambn for structural and functional studies. We focused on the larger, more abundant ameloblastin, which is comprised of 395 amino acids. Here we report the expression and isolation of recombinant ameloblastin from a eukaryotic host, the characterization of its major post-translational modifications, and spectrum of its cleavage products generated by Mmp-20 and Klk4 digestion. We have also initiated studies to assay ameloblastin structure and function *in vivo*. *Ambn* null mice have virtually no enamel. We have generated transgenic mice that specifically direct Ambn expression in ameloblasts using the amelogenin promoter and 3' untranslated region and have crossed these mice with the ameloblastin null mice to rescue the enamel phenotype.

GENERAL HYPOTHESIS

Ameloblastin is a member of the secretory calcium-binding phosphoprotein (SCPP) family of proteins, which trace their ancestry to SPARC (secreted protein acidic and rich in cysteine). Uncleaved ameloblastin has never been isolated or characterized from any source. Ameloblastin null mice display enamel agenesis, demonstrating that ameloblastin is critical for amelogenesis. We hypothesize that by investigating ameloblastin homologues, by assaying the structural and functional properties of full-length ameloblastin *in vitro*, by characterizing the mechanism of its proteolytic processing into smaller products that accumulate in the enamel matrix, and by establishing an *in vivo* system to assay ameloblastin structural and functional

relationships, we will gain new and important information on the roles played by ameloblastin in normal dental enamel formation.

SPECIFIC AIMS

Four specific aims are proposed to test these hypotheses:

SA 1 To characterize SPARC as the evolutionary origin of ameloblastin.

Ameloblastin resides within the locus of the secreted calcium-binding phosphoproteins (SCPP) on human chromosome 4. Secreted protein acidic and rich in cysteine (SPARC) was identified as the ancestor to the SCPP cluster (Kawasaki and Weiss, 2003). Structural properties of SPARC may provide insights into functional roles of SCPPs like ameloblastin.

SA 2 To produce recombinant porcine ameloblastin and characterize its post-translational modifications (PTM).

Uncleaved ameloblastin cannot be isolated from *in vivo* sources because it is rapidly processed following its secretion (Uchida *et al.*, 1991b). To obtain full-length, glycosylated ameloblastin for structural and functional studies we expressed Ambn in eukaryotic cells. Recombinant porcine ameloblastin (rpAmbn) is expressed from a stable cell line (HEK293), purified and characterized. Silver staining assesses its purity. Stains-all assesses calcium binding. Western blots using N- and C-terminal antibodies show the

protein is full-length. O-linked glycosylations are characterized by Emerald staining and by mobility shift assays, which use SDS-PAGE to disclose the reduction in Ambn apparent molecular weight of Ambn following enzymatic removal of its O-glycosylations.

SA 3 To determine the cleavage sites derived from Ambn generated by Mmp-20 and Klk4 *in vitro*.

The cleavage of enamel proteins by Mmp-20 and Klk4 are necessary steps for proper dental enamel formation. A better understanding of the roles of Mmp-20, Klk4, and Ambn during amelogenesis will be realized once we have determined how Ambn is processed or degraded by these enzymes. As Mmp-20 is primarily expressed in the secretory stages, we expect this enzyme cleaves recombinant Ambn *in vitro* at the same sites that generate the Ambn cleavage products that accumulate in immature enamel. Klk4, the maturation stage protease, is expected to cleave at many sites, and prepare the protein for reabsorption into the cell and final degradation in lysosomes.

Ambn (rpAmbn or synthetic Ambn peptides) will be incubated with rpMmp-20 or native pKlk4. The digestion products from rpAmbn will be characterized by SDS-PAGE, Western blotting, and by N-terminal sequencing. The cleavage products of the synthetic peptides will be identified by RP-HPLC and LC-MS/MS. The *in vitro* cleavage sites of Ambn will be compared with cleavage products of Ambn purified from porcine immature enamel *in vivo*.

SA 4 To determine if enamel rod organization and enamel thickness can be rescued in the *Ambn* null mice by Ambn transgene expression.

Enamel rod organization and enamel thickness are lost in the *Ambn* null mouse. Transgenic expression of *Ambn* from the amelogenin promoter is expected to recover the enamel phenotype caused by a lack of *Ambn* expression. The enamel will regain its thickness with normal rod organization. If this can be accomplished, it will be possible to introduce site-directed mutations in Ambn and to assay for Ambn function *in vivo*.

BACKGROUND AND SIGNIFICANCE

Tooth enamel is the most mineralized tissue of the human body consisting of hydroxyapatite to 95 % by weight and of a minor fraction of proteins and water. The mechanisms of enamel mineralization are highly regulated as amelogenesis starts without mineral with 100 % proteins, which are continuously secreted and processed. The mineralization events correlate to secretion, processing and reabsorption of enamel proteins. The "structural" (meaning not proteolytic) enamel proteins are amelogenin, ameloblastin and enamelin (Tab. 1.1). Their complexity is multiplied by alternative splicing and enzymatic processing. Each protein is thought to have specific function, localization and fate. Processing and degradation of enamel proteins is accomplished by Mmp-20 and Klk4.

Structural enamel proteins		
	MW	**Potential function**
Amelogenin (Amel) **85-90 %**	18-25 kDa	Regulation of crystal spacing Prevents crystals from fusing
Ameloblastin (Ambn) **3-5 %**	60-65 kDa	In sheath space Maintains rod boundaries
Enamelin (Enam) **1-3 %**	>200 kDa	Crystal elongation
Enamel proteinases		
	Expression	**Potential function**
Enamelysin (Mmp-20)	Secretory stage	Process enamel proteins
Kallikrein-4 (Klk4)	Transition/ Maturation Stages	Degrade enamel matrix

Table 1.1: Enamel matrix proteins.
During enamel formation 3 'structural' proteins and 2 proteinases are expressed and secreted. Amelogenin is the most abundant enamel protein with 85-90 %. Ameloblastin and enamelin are present in the matrix only at 1-5 %. The enamel proteinases are expressed sequentially, first Mmp-20 during secretory stage followed by Klk4 during maturation stage.

Enamel formation

The tooth arises from interactions between epithelial and mesenchymal cells. The inner enamel organ epithelium (EOE) proliferates and differentiates into polarized ameloblasts having tall columnar morphology. These cells undergo dramatic changes in morphology and function during their life cycle (Fig. 1.1). Cells of the inner enamel epithelium are attached to a basement membrane containing laminin (Salmivirta *et al.*, 1997), underneath which dentin matrix is deposited by odontoblasts. As the basement membrane disintegrates, the pre-secretory ameloblasts start secreting enamel proteins and Mmp-20. This enamel layer at the dentino-enamel junction mineralizes as aprismatic enamel. With the development of the apical Tomes' process, the ameloblasts begin producing and secreting enamel proteins in high quantity. The released proteins are processed immediately at the mineralization front. As the enamel layer gains thickness

through the extension of crystallites, the ameloblasts recede. During a short transition stage, the Tomes' process degenerates, and in maturation stage, ameloblasts oscillate between smooth-ended and ruffled-ended apical membranes. The ameloblasts secrete a basement membrane containing amelotin (Amtn) (Moffatt *et al.*, 2006) and Apin (other name ODAM) (Moffatt *et al.*, 2008). The Mmp-20 processed proteins are further digested by Klk4 and removed from the matrix.

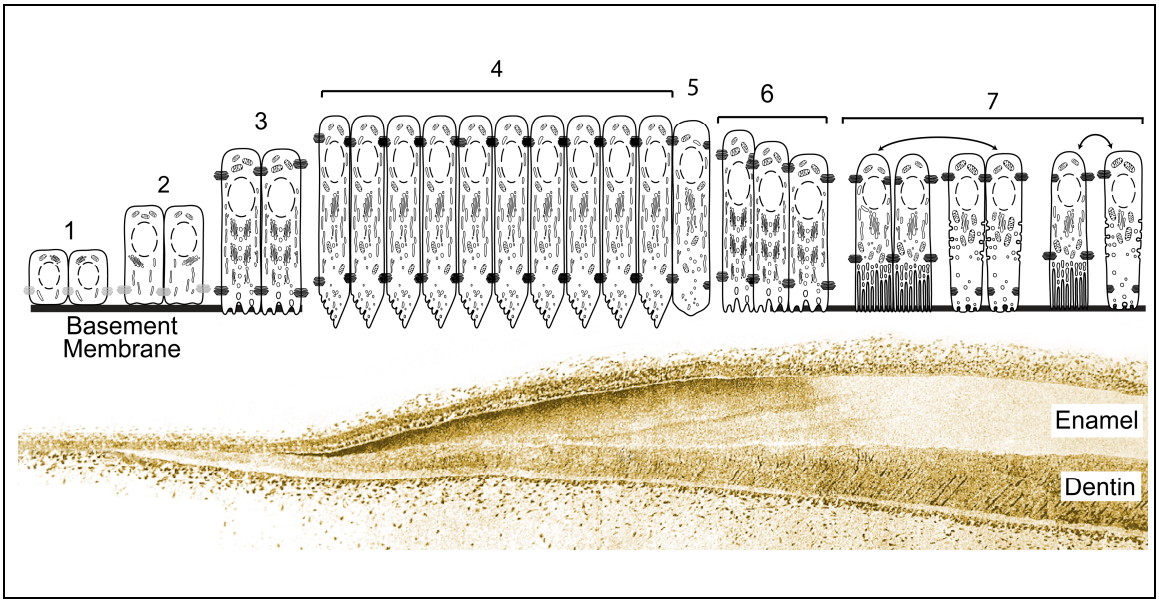

Figure 1.1: Stages of enamel formation.
Dental enamel formation is divided into pre-secretory, secretory, transition and maturation stages. Cells of the inner enamel epithelium (1) are separated from the dentin matrix by a basement membrane. When the ameloblasts start polarizing (2), the basement membrane disintegrates and projections (3) develop into an apical Tomes' process with a secretory and non-secretory face (4). During the secretory stage (4), enamel proteins are secreted, processed and crystals grow primarily in length. The Tomes' process redevelops (5), and in transition stage the ameloblasts reduce their length (6). During maturation stage, the basement membrane reforms and ameloblasts alternate between smooth-ended and ruffled-ended ameloblasts (7). As proteins are degraded and reabsorbed, the mineral content of enamel increases and crystals gain width. (Hu *et al.*, 2007).

Development of enamel structure

Enamel structure is dictated by the mineralization front with the onset of secretion. As crystallites span the entire thickness of the enamel layer, they reflect the dynamics at the mineralization front. The mineralization front is the functional unit consisting of the secretory face of the Tomes' process, the extracellular space between plasma membrane containing the organic material and the tip of the crystallite (Fig. 1.2) (Rönnholm, 1962). As enamel matrix proteins and enamel proteinases are released into the extracellular space, hydroxyl-apatite deposits rapidly on the crystallite tips and the crystallites lengthen rapidly in the direction of their c-axes. With the retraction of the Tomes' process and the cessation of protein secretion, the crystallites stop growing in length. During the maturation stage, enamel proteins are degraded and removed followed by mineral deposition exclusively on the sides of the crystallites. Crystallites grow in width and thickness until further growth is prevented by contact with adjacent crystals.

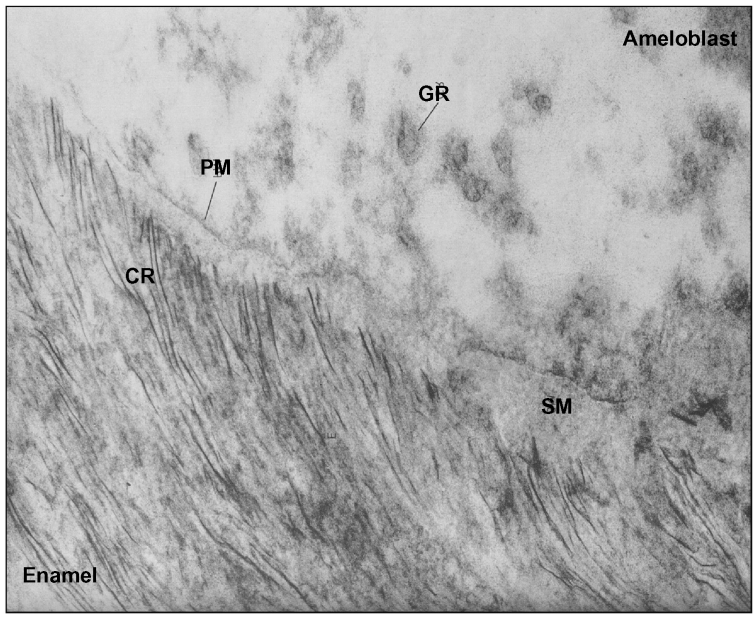

Figure 1.2: Mineralization front formed by Tomes' process of the ameloblast and crystallites.
The crystallites are adjacent to the plasma membrane and are surrounded by organic material. Crystallites grow at the tips along the c-axis at a high rate, and the Tomes' process recedes to accommodate the growing crystallites.
PM plasma membrane, GR granule, SM secreted matrix, CR crystallites. Magnification 140,000 x. (Rönnholm, 1962).

Structure of mature enamel

Histological images of cross-sections of undecalcified dental enamel reveal discrete rods and interrods that are separated by the narrow sheath space (Fig. 1.3). Rod and interrod enamel are structurally identical, but differ in orientation of crystallites (Warshawsky *et al.*, 1981). The rod crystals grow directly perpendicular to the secretory face of the Tomes' process of the ameloblast. In contrast, interrod crystals are inclined up to 40–65 degrees from the long axis of the rods and originate from the proximal portions of the Tomes' processes. The rods are partially surrounded by the sheath space. The sheath space was first discovered in cross-sections of undecalcified enamel etched with mild acid and stained with silver nitrate (Smreker, 1903).

The sheath space originates along the lateral surfaces of the Tomes' process, which are not associated with the extension of enamel crystallites. A distinct rod boundary of 30-50 Angstroms (Gustavsen and Silness, 1969) can only be defined on the convex side of each rod pointing occlusally or incisally. The sheath space extends from the dentino-enamel junction to the enamel surface and exits from early amelogenesis throughout maturation (Coolidge, 1965). In contrast to rod and interrod enamel, the sheath space contains few crystals but a high proportion of enamel proteins even at maturation stage. It is the first area to demineralize in the caries process (Meurman and Frank, 1991).

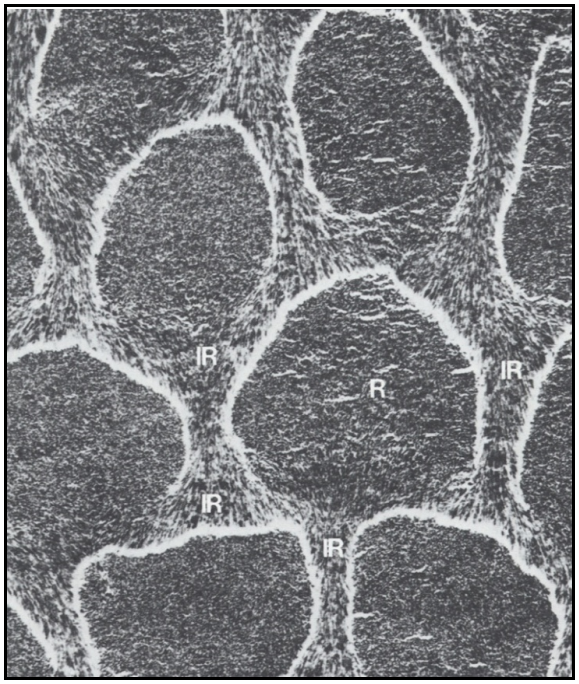

Figure 1.3: Scanning electron micrograph (SEM) of the surface of an acid-etched, ground section of mature mouse incisal dental enamel.
The ordered arrays of enamel rods are each constructed of parallel bundles of enamel crystallites. Enamel rods are present in all mammalian enamel, although the pattern of enamel rods varies in different organisms. Each rod consists of about 10,000 crystallites and follows the path taken by an individual ameloblast during its secretory stage migrations. Magnification 8,100 x. Courtesy of P.L. Glick in Feierskov and Thylstrup, 1986.

In immature enamel, the protein composition of the sheath space is distinct from rod and interrod regions. Rod and interrod proteins are soluble in neutral buffer, while sheath proteins are insoluble in neutral buffer, but soluble in dilute acetic acid. The sheath proteins contain less proline and histidine, but more serine, glycine, arginine (Glimcher *et al.*, 1964). The function of the sheath space is to serve as a path for enamel proteins to be removed from deeper layers (Glimcher *et al.*, 1964). The prismatic organization of enamel is unique, and the sheath space is an important aspect to it. Ameloblastin N-terminal cleavage products are located in the sheath space, but it is unknown how they localize so specifically and what their function might be in prism organization.

Enamel proteins

When tooth enamel matrix proteins were recognized as an important feature of enamel development, they were named according to their origin: enamel proteins from unerupted teeth were called amelogenins and proteins from mature teeth were called enamelins (Eastoe, 1979). Enamel proteins isolated from mature enamel with guanidine-HCl/EDTA buffer were categorized as amelogenins and non-amelogenins (Termine *et al.*, 1980). Simultaneously, nondissociative, pH-based extractions of enamel proteins from porcine teeth were developed (Fukae *et al.*, 1979; Fukae *et al.*, 1980; Shimizu *et al.*, 1979). These strategies have led to the discovery of amelogenin (Fukae *et al.*, 1979; Fukae *et al.*, 1980; Fukae and Shimizu, 1983), ameloblastin (Fukae and Tanabe, 1987a; Fukae and Tanabe, 1987b), enamelin (Tanabe *et al.*, 1990; Uchida *et al.*, 1991a) and enamel matrix serine proteinase 1 peptide sequences (Tanabe, 1984). The extraction protocols have been optimized and many enamel protein cleavage products have been

isolated during past 30 years (Yamakoshi *et al.*, 2003; Yamakoshi *et al.*, 2006). Cloning of the cDNA of amelogenin (Snead *et al.*, 1985), ameloblastin (Cerny *et al.*, 1996; Hu *et al.*, 1997a; Krebsbach *et al.*, 1996), enamelin (Hu *et al.*, 1997b) allowed deduction of their full-length amino acid sequences, which were compared against the N-terminal sequences of proteins isolated from developing teeth. It became clear that enamel proteins are processed by enzymes (Overall and Limeback, 1988; Shimizu *et al.*, 1979; Tanabe, 1984). The enamel proteinases were cloned and designated as enamelysin or Mmp-20 (Bartlett *et al.*, 1996) and enamel matrix serine proteinase 1, now kallikrein 4 (Klk4) (Simmer *et al.*, 1998). The gene mutations responsible for amelogenesis imperfecta in humans typically disrupt either the quantity or the quality of the gene product. Therefore, it is important to understand how the matrix proteins govern the fundamental mechanisms of amelogenesis.

Evolution and enamel formation

Human ameloblastin and enamelin genes have been localized to chromosome 4q13-21. This chromosomal region is linked to autosomal dominant amelogenesis imperfecta in humans (Karrman *et al.*, 1996). Upstream of Ambn are located two enamel genes, AMTN and APIN, which are expressed in the basement membrane during maturation stage of amelogenesis. Phylogenetic studies have discovered that the enamel proteins are products of a gene cluster termed secreted calcium-binding phosphoproteins (SCPP). Through whole genome wide duplication, secreted protein acidic and rich in cysteine (SPARC) duplicated as SPARC-like (SPARCL) from chromosome 5 to chromosome 4. Successive SPARCL duplications generated new genes that secreted their

calcium-binding protein products and adopted new functions. In mammalian evolution, enamel has been selected for as a highly mineralized tissue. *AMBN* is a member of the SCPP cluster and localizes between *SPARCL* and *ENAM*. The lack of constraint is reflected in the Ambn gene structure. Among mammalian species, Ambn has generally 11 exons, except in humans it has 13 exons. The two additional exons in human arose from tandem duplications of exon 7. Survival advantages through feeding strategies in birds and baleen whales have led to silencing and complete loss of enamel genes.

Ameloblastin

Ameloblastin is one of the enamel structural proteins and is critical for enamel formation. Deletion of *Ambn* in mice results in hypoplastic amelogenesis imperfecta (Fukumoto *et al.*, 2004). Instead of the highly organized prism pattern in the enamel, only a rudimentary amorphous material is deposited. The enamel is chalky and rapidly abraded. During tooth development, Ambn mRNA is expressed by the inner enamel epithelium from presecretory to late maturation stage (Lee *et al.*, 1996; Torres-Quintana *et al.*, 2005). In addition to ameloblasts, ameloblastin is also expressed by cells producing acellular and cellular cementum and by cells of the Hertwig's epithelial root sheath (Bosshardt and Nanci, 1998; Fong *et al.*, 1996).

Ameloblastin protein is rapidly processed after secretion. Full-length Ambn is only found adjacent to the non-secretory face of the Tomes' processes of the ameloblasts, while the lower molecular weight proteins are present in the sheath space and in the rods of the superficial layer (Hu *et al.*, 1997a; Murakami *et al.*, 1997; Uchida *et al.*, 1995). The N-terminal cleavage products (13, 15, 17 kDa) are stable and concentrate in the

prism sheath. In contrast, the C-terminal cleavage products (40, 50 kDa) are successively cleaved into smaller peptides (8, 13, 15, 27, 29 kDa) and lost from the immature enamel soon after secretion (Uchida *et al.*, 1998). The cleavage products of secretory stage enamel differ in structural properties, localization and function. However, the precise mechanism of how the cleavage products are generated and how they contribute to enamel formation is not known.

The main goal of the studies incorporated in this thesis is to learn about ameloblastin as one of the proteins essential for enamel formation. Our first approach was to investigate if Ambn's relationship to the better-studied SPARC (whose crystal structure was known) would yield valuable information concerning the structure and function of Ambn. This fundamental characterization of Ambn allows gathering information to better understand Ambn's role in enamel formation.

REFERENCES

Bartlett JD, Simmer JP, Xue J, Margolis HC, Moreno EC (1996). Molecular cloning and mRNA tissue distribution of a novel matrix metalloproteinase isolated from porcine enamel organ. *Gene* 183:123-128.

Bosshardt DD, Nanci A (1998). Immunolocalization of epithelial and mesenchymal matrix constituents in association with inner enamel epithelial cells. *J Histochem Cytochem* 46:135-142.

Cerny R, Slaby I, Hammarstrom 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.

Coolidge TB (1965). Biochemistry of the sheath spaces in caries. *Ann N Y Acad Sci* 131:884-892.

Eastoe JE (1979). Enamel protein chemistry--past, present and future. *J Dent Res* 58(Spec Issue B):753-764.

Feierskov O, Thylstrup A (1986). Dental enamel. In: Human oral embryology and histology. IA Mjor and O Feierskov editors. Copenhagen: Munksgaard, pp. 50-89.

Fong CD, Hammarstrom L, Lundmark C, Wurtz T, Slaby I (1996). Expression patterns of RNAs for amelin and amelogenin in developing rat molars and incisors. *Adv Dent Res* 10:195-200.

Fukae M, Ijiri H, Tanabe T, Shimizu M (1979). Partial amino acid sequences of two proteins in developing porcine enamel. *J Dent Res* 58:1000-1001.

Fukae M, Tanabe T, Ijiri H, Shimizu M (1980). Studies on porcine enamel proteins: a possible original enamel protein. *Tsurumi Univ Dent J* 6:87-94.

Fukae M, Shimizu M (1983). Amino acid sequence of the main component of porcine enamel proteins. *Jpn J Oral Biol* 25:29.

Fukae M, Tanabe T (1987a). Nonamelogenin components of porcine enamel in the protein fraction free from the enamel crystals. *Calcif Tissue Int* 40:286-293.

Fukae M, Tanabe T (1987b). ^{45}Ca-labeled proteins found in porcine developing dental enamel at an early stage of development. *Adv Dent Res* 1:261-266.

Fukumoto S, Kiba T, Hall B, Iehara N, Nakamura T, Longenecker G, Krebsbach PH, Nanci A, Kulkarni AB, Yamada Y (2004). Ameloblastin is a cell adhesion molecule required for maintaining the differentiation state of ameloblasts. *J Cell Biol* 167:973-983.

Glimcher MJ, Travis DF, Friberg UA, Mechanic GL (1964). The Electron Microscopic Localization of the neutral soluble proteins of developing bovine enamel. *J Ultrastruct Res* 10:362-376.

Gustavsen F, Silness J (1969). Crystal shape in the prism sheath region of sound human enamel. *Acta Odontol Scand* 27:617-629.

Hu CC, Fukae M, Uchida T, Qian Q, Zhang CH, Ryu OH, Tanabe T, Yamakoshi Y, Murakami C, Dohi N, Shimizu M, Simmer JP (1997a). Sheathlin: cloning, cDNA/polypeptide sequences, and immunolocalization of porcine enamel sheath proteins. *J Dent Res* 76:648-657.

Hu CC, Fukae M, Uchida T, Qian Q, Zhang CH, Ryu OH, Tanabe T, Yamakoshi Y, Murakami C, Dohi N, Shimizu M, Simmer JP (1997b). Cloning and characterization of porcine enamelin mRNAs. *J Dent Res* 76:1720-1729.

Hu JC, Chun Y-HP, Al-Hazzazzi T, Simmer JP (2007). Enamel formation and amelogenesis imperfecta. *Cells Tissues Organs* 186:78-85.

Karrman C, Backman B, Holmgren G, Forsman K (1996). Genetic heterogeneity of autosomal dominant amelogenesis imperfecta demonstrated by its exclusion from the AIH2 region on human chromosome 4Q. *Arch Oral Biol* 41:893-900.

Kawasaki K, Weiss KM (2003). Mineralized tissue and vertebrate evolution: the secretory calcium-binding phosphoprotein gene cluster. *Proc Natl Acad Sci U S A* 100:4060-4065.

Krebsbach PH, Lee SK, Matsuki Y, Kozak CA, Yamada KM, Yamada Y (1996). Full-length sequence, localization, and chromosomal mapping of ameloblastin. A novel tooth-specific gene. *J Biol Chem* 271:4431-4435.

Lee SK, Krebsbach PH, Matsuki Y, Nanci A, Yamada KM, Yamada Y (1996). Ameloblastin expression in rat incisors and human tooth germs. *Int J Dev Biol* 40:1141-1150.

Meurman JH, Frank RM (1991). Progression and surface ultrastructure of in vitro caused erosive lesions in human and bovine enamel. *Caries Res* 25:81-87.

Moffatt P, Smith CE, St-Arnaud R, Simmons D, Wright JT, Nanci A (2006). 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 CE, St-Arnaud R, Nanci A (2008). Characterization of Apin, a secreted protein highly expressed in tooth-associated epithelia. *J Cell Biochem* 103:941-956.

Murakami C, Dohi N, Fukae M, Tanabe T, Yamakoshi Y, Wakida K, Satoda T, Takahashi O, Shimizu M, Ryu OH, Simmer JP, Uchida T (1997). Immunochemical and immunohistochemical study of the 27- and 29-kDa calcium-binding proteins and related proteins in the porcine tooth germ. *Histochem Cell Biol* 107:485-494.

Overall CM, Limeback H (1988). Identification and characterization of enamel proteinases isolated from developing enamel. Amelogeninolytic serine proteinases are associated with enamel maturation in pig. *Biochem J* 256:965-972.

Rönnholm E (1962). An electron microscopic study of the amelogenesis in human teeth. I. The fine structure of the ameloblasts. *J Ultrastruct Res* 6:229-248.

Salmivirta K, Sorokin LM, Ekblom P (1997). Differential expression of laminin alpha chains during murine tooth development. *Dev Dyn* 210:206-215.

Shimizu M, Tanabe T, Fukae M (1979). Proteolytic enzyme in porcine immature enamel. *J Dent Res* 58:782-789.

Simmer JP, Fukae M, Tanabe T, Yamakoshi Y, Uchida T, Xue J, Margolis HC, Shimizu M, DeHart BC, Hu CC, Bartlett JD (1998). Purification, characterization, and cloning of enamel matrix serine proteinase 1. *J Dent Res* 77:377-386.

Smreker E (1903). Ueber die Darstellung der Kittsubstanz des Schmelzes menschlicher Zaehne. *Anat Anzeiger* 22:467-476.

Snead ML, Lau EC, Zeichner-David M, Fincham AG, Woo SL, Slavkin HC (1985). DNA sequence for cloned cDNA for murine amelogenin reveal the amino acid sequence for enamel-specific protein. *Biochem Biophys Res Commun* 129:812-818.

Tanabe T (1984). Purification and characterization of proteolytic enzymes in porcine immature enamel. (in Japanese). *Tsurumi Univ Dent J* 10:443-452.

Tanabe T, Aoba T, Moreno EC, Fukae M, Shimuzu M (1990). Properties of phosphorylated 32 kd nonamelogenin proteins isolated from porcine secretory enamel. *Calcif Tissue Int* 46:205-215.

Termine JD, Belcourt AB, Christner PJ, Conn KM, Nylen MU (1980). Properties of dissociatively extracted fetal tooth matrix proteins. I. Principal molecular species in developing bovine enamel. *J Biol Chem* 255:9760-9768.

Torres-Quintana MA, Gaete M, Hernandez M, Farias M, Lobos N (2005). Ameloblastin and amelogenin expression in posnatal developing mouse molars. *J Oral Sci* 47:27-34.

Uchida T, Tanabe T, Fukae M, Shimizu M (1991a). Immunocytochemical and immunochemical detection of a 32 kDa nonamelogenin and related proteins in porcine tooth germs. *Arch Histol Cytol* 54:527-538.

Uchida T, Tanabe T, Fukae M, Shimizu M, Yamada M, Miake K, Kobayashi S (1991b). Immunochemical and immunohistochemical studies, using antisera against porcine 25 kDa amelogenin, 89 kDa enamelin and the 13-17 kDa nonamelogenins, on immature enamel of the pig and rat. *Histochemistry* 96:129-138.

Uchida T, Fukae M, Tanabe T, Yamakoshi Y, Satoda T, Murakami C, Takahashi O, Shimizu M (1995). Immunological 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'. *Biomedical Research* 16:131-140.

Uchida T, Murakami C, Wakida K, Dohi N, Iwai Y, Simmer JP, Fukae M, Satoda T, Takahashi O (1998). Sheath proteins: synthesis, secretion, degradation and fate in forming enamel. *Eur J Oral Sci* 106 (Suppl 1) 308-314.

Warshawsky H, Josephsen K, Thylstrup A, Fejerskov O (1981). The development of enamel structure in rat incisors as compared to the teeth of monkey and man. *Anat Rec* 200:371-399.

Yamakoshi Y, Hu JC-C, Ryu OH, Tanabe T, Oida S, Fukae M, Simmer JP (2003). A comprehensive strategy for purifying pig enamel proteins. *Biomineralization (BIOMIN2001): formation, diversity, evolution and application.* 326-332.

Yamakoshi Y, Hu JC, Zhang H, Iwata T, Yamakoshi F, Simmer JP (2006). Proteomic analysis of enamel matrix using a two-dimensional protein fractionation system. *Eur J Oral Sci* 114 (Suppl 1) 266-71; discussion 285-286, 382.

CHAPTER 2

EVOLUTION OF GENES IN ENAMEL FORMATION

ABSTRACT

Many of the key genes that encode the major extracellular matrix proteins of enamel, bone and dentin are members of the SCPP (secretory calcium-binding phosphoprotein) family (Kawasaki and Weiss, 2003; Kawasaki and Weiss, 2006; Kawasaki *et al.*, 2007). The SCPP gene family arose by gene duplication of the secreted protein, acidic, cysteine-rich like 1 (SPARCL1) gene (Kawasaki *et al.*, 2004). SPARCL1 in turn, arose from *SPARC* (secreted protein, acidic and rich in cysteine) (Delgado *et al.*, 2001). The SCPPs are divided into two groups based upon their amino acid compositions: acidic SCPPs that contain more than 25 % glutamic and aspartic acids and phosphoserine residues, and the P/Q rich proteins that contain more than 20 % proline and glutamine. In mammals, acidic SCPPs are primarily involved in bone and dentin mineralization, whereas P/Q-rich SCPPs enable enamel mineralization (Kawasaki and Weiss 2008). To better understand the role of SPARC in mineralizing systems, we isolated SPARC from developing pig teeth, deduced the porcine SPARC primary structure from its cDNA sequence, and determined its quaternary structure by homology modeling after the human SPARC crystal structure.

INTRODUCTION

Evolution of biomineralization

The development of biomineralization is a key event in evolution that led to its application in different mineralized tissues in vertebrates. All mineralized tissues such as teeth, bone and cartilage form within an extracellular matrix. Each mineralized tissue forms using a unique set of proteins, which is determined by the secretory cells lining the extracellular space. Fibrillar collagen is the predominant matrix component of many mineralizing systems, such as bone and dentin. During evolution, however, collagen-based extracellular matrices preceded the onset of biomineralization. The mineralization of extracellular matrices is associated with the evolution of genes encoding novel extracellular matrix proteins. SPARC (secreted protein acidic and rich in cysteine; also known as BM-40, osteonectin) is an ancient secreted protein that forms part of the basement membrane and also binds to collagen. Evolutionary genetics identified SPARC as the ancestral gene for secreted calcium-binding phosphoproteins (SCPP) that today mediate biomineralization in vertebrates (Kawasaki and Weiss, 2003; Kawasaki *et al.*, 2004; Kawasaki and Weiss, 2008). This cluster of genes on human chromosome 4q13-q21 encodes structural features for being targeted to the extracellular matrix, phosphorylated in the Golgi apparatus, and binding calcium. The SCPP family arose in a series of steps that included chromosomal and genome-wide duplications, as well as gene and exon duplications. About 6 % of all exons of the human genome arose from exon duplication indicating a high, non-random event to create genetic diversity (Fedorov *et al.*, 1998). This acquired genetic information has been exploited in different functional contexts, e.g. in milk, saliva, bone and tooth proteins.

SPARC as ancestral gene of biomineralization

The radiation of genes in a cluster is based on duplication events. The SCPP genes originate from a single ancestral gene, SPARC through iterative gene duplication (Kawasaki and Weiss, 2008). In a genome-wide duplication, SPARCL1 (SPARC-like1 protein) spawned from SPARC and translocated to a different chromosome (Kawasaki *et al.*, 2004). The timing of this divergence correlates into the onset of mineralized skeleton based on fossil records (Donoghue and Sansom, 2002).

SPARC and SPARCL1 are similar in their amino acid sequence of their signal peptide, and three functional domains. An important difference between SPARC and SPARCL1 is that SPARCL1 is phosphorylated by Golgi casein kinase. As a result of repeated tandem gene duplications SPARCL1 gave rise to subfamilies of enamel/milk/saliva genes and of dentin/bone genes located up- and down-stream of SPARCL1 (Fig. 2.1 (Kawasaki and Weiss, 2008)).

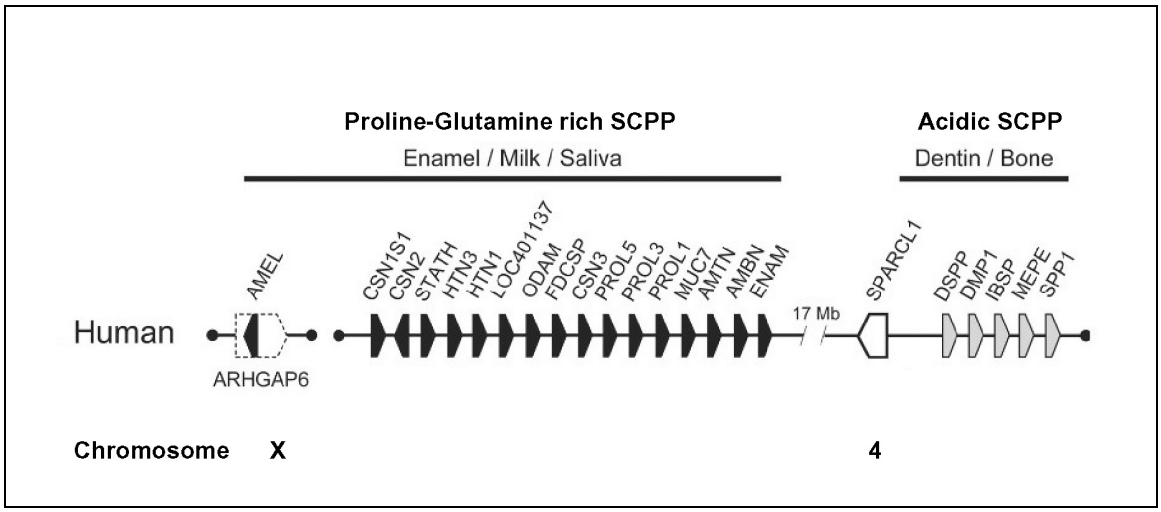

Figure 2.1: Gene cluster of secreted calcium-binding phosphoproteins (SCPPs).
Of the SCPP genes, all genes are located on chromosome 4 except for Amel located on the X chromosome. Gene duplications of SPARCL resulted in the acidic dentin and bone genes, and the Proline-Glutamine rich enamel, milk, saliva SCPP, including Amel on chromosome X. (Kawasaki and Weiss, 2008).

The enamel/milk/saliva proteins consist of 16 members and are rich in proline and glutamine. The dentin/bone SCPPs, also known as small integrin-binding ligand, N-linked glycoproteins (SIBLINGs) (Fisher *et al.*, 2001; Fisher and Fedarko, 2003) are acidic and have five members. Functionally, all SCPPs are targeted to the extracellular space and have the ability to bind calcium as means to control calcium concentration for mineralization (Kawasaki and Weiss, 2003).

Evidence for the descent of the enamel SCPPs from SPARCL1 is found in the 5' region of SPARCL1 (Kawasaki *et al.*, 2004; Kawasaki *et al.*, 2007). Strikingly, all SCPPs share the structure of exon 1 and the 5'-end of exon 2 constituting the 5'-untranslated region, followed by encoding of the signal peptide (16 amino acids) and 2 amino acids of the secreted protein (Fig. 2.2 (Kawasaki and Weiss, 2006)). The conservation of the signal peptide underscores the importance of targeting the protein to the extracellular space as a prerequisite for biomineralization. All SCPP genes have introns in phase 0, and thus coding triplets are not interrupted by intron-exon borders. Exon 3 encodes one or more SXE motifs in the context for phosphorylation by Golgi casein kinase 2 (Fig. 2.2). Other shared structural features of the SCPPs are glycosylation and processing by proteases. Together, these structural commonalities of the SCPPs are non-random and indicate tandem gene duplication events resulting in genetic diversity.

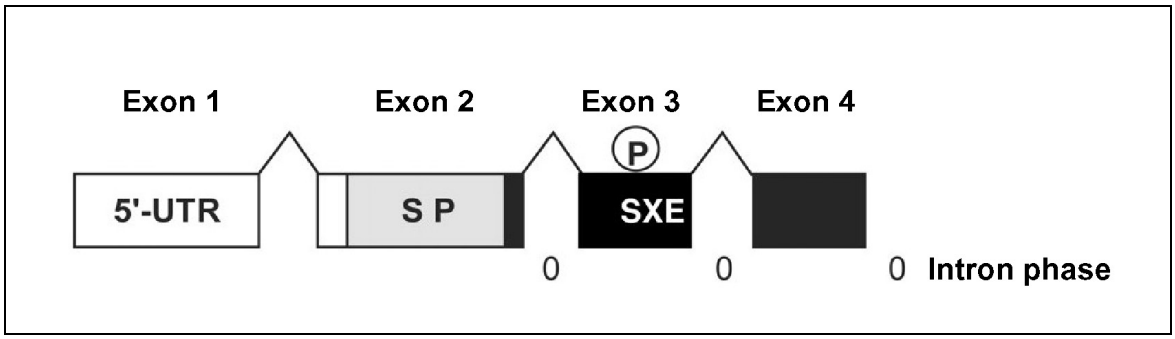

Figure 2.2: The gene structure is conserved in SCPPs.
All SCPP genes share structural features in the first 3 exons. Exon 1 contains the 5'-UTR, exon 2 encodes the signal peptide (SP), and exon 3 contains an SXE motif recognized by Golgi casein kinase for phosphorylation. The intron-exon borders in all genes are in phase 0. (Kawasaki and Weiss, 2003).

Evolution of enamel proteins

The enamel SCPPs are localized on human chromosome 4q13-q21 (*ENAM, AMBN, AMTN, APIN*) and on the X- chromosome (AMEL). Amel, Enam and Ambn are essential components of the enamel matrix and secreted during secretory stage by ameloblasts. During the maturation stage when enamel proteins are fully degraded and reabsorbed, AMTN and APIN are secreted as basement membrane proteins covering the outer enamel.

Comparing the genome of different lineages, allows insights in the origin and evolution of the SCPPs. Amel and Ambn are found in all tetrapods, but Enam is absent in amphibians (Kawasaki and Weiss, 2008). The avian genome is missing enamel, milk or saliva SCPP genes. The enamel genes degenerated in modern birds after they lost their teeth in Late Cretaceous (Peyer, 1968). Amel evolved from an autosomal location first to proto-X and proto-Y chromosomes in eutherians (placental mammals) and translocated later to X and Y chromosomes (Kawasaki and Weiss, 2006; Sire *et al.*, 2005). The

majority of the amelogenin is expressed from the X chromosome where the gene is inserted into the first intron of Rho GTPase activating protein 6 (ARHGAP6). Since Amel has maintained the structural features of the SCPPs, it is considered a member of the SCPP cluster.

The chromosomal region harboring the SCPP cluster has experienced extensive intra-chromosomal rearrangements revealed by comparison of chicken and human syntenic region (Kawasaki and Weiss, 2006). Most likely, the dentin/bone SCPPs duplicated first from SPARCL1. Then Ambn split from the dentin/bone SCPPs and gave rise to the milk/saliva/enamel SCPPs by tandem gene duplication in the mammalian lineage (Kawasaki and Weiss, 2006).

Functions of SPARC

SPARC is expressed and secreted in many mineralizing and non-mineralizing tissues during development and tissue maintenance. Other names for SPARC are *osteonectin*, referring to its purification from bone and its ability to bind hydroxyapatite (Termine *et al.*, 1981a; Termine *et al.*, 1981b), and *BM-40*, referring to its location in the basement membrane and apparent molecular weight of 40 kDa (Dziadek *et al.*, 1986).

During tooth development, SPARC is found in dentin (Bronckers *et al.*, 1989; Tung *et al.*, 1985), starting at the bell stage and continuing throughout adulthood (Reichert *et al.*, 1992). Its expression is highest immediately following crown completion and lowest at completion of the root (Fujisawa and Kuboki, 1989; Tung *et al.*, 1985). In rat teeth, SPARC expression was strongest in odontoblasts and pre-mature odontoblasts (Takano-Yamamoto *et al.*, 1994). In the initial stages of cytodifferentiation, and in the

non-mineralized, collagen-rich predentin, SPARC associated with collagen. In the mineral-rich, peritubular dentin, SPARC was expressed independently of collagen (Papagerakis *et al.*, 2002).

The SPARC gene maps to human chromosome 5q31-q33. It is 20 kb long and contains 10 exons that encode 303 amino acids. The secreted 40-kDa protein has four distinct domains (Swaroop *et al.*, 1988). The N-terminal end is highly acidic, binds 5-8 Ca^{2+} ions with a K_d of 10^{-5} to 10^{-3} M (Engel *et al.*, 1987), and binds hydroxyapatite *in vitro* (Fujisawa *et al.*, 1996). The middle part contains the follistatin-like (FS) domain with 5 intramolecular disulfide bridges, and the C-terminal extracellular (EC) calcium binding domain containing two structural calcium ions (K_d of 10^{-7}) each coordinated by EF-right hands (Busch *et al.*, 2000; Engel *et al.*, 1987). Although the complete SPARC protein has not been crystallized, the SPARC EC domain structure has been resolved to 2 Å and the combined FS-EC structure to 3.1 Å (Hohenester *et al.*, 1996; Hohenester *et al.*, 1997).

The SPARC gene and protein has been highly conserved during evolution. Human SPARC shares 99 to 92 % identities with SPARC from other mammals, and 79 % identity with a frog homologue. Conservation falls to 30 and 31 % in the invertebrates drosophila and nematode, respectively (Brekken and Sage, 2000). The FS and EC domains appear as modules in other proteins, such as SPARCL1, SC1, TSC-36, testican-1, -2, -3, SMOC-1, -2. Because its primary structure is so highly conserved and its crystal structure has been solved for the human and bovine homologues, reliable 3-D protein structures of SPARC homologues can be gained by homology modeling.

In the hope of identifying molecular properties common to all vertebrate mineralized tissues, we have endeavored to isolated and characterize the SPARC homologue from pig, which is an important animal model for dentin and enamel biomineralization studies. Porcine SPARC structure was characterized at the genetic level by cloning and at the protein level by isolation, multiple sequence alignment and computer-generated structural model. The results provide insight into the structural features that played a role at the dawn of vertebrate biomineralization.

MATERIALS & METHODS

All experimental procedures involving the use of animals were reviewed and approved by the Institutional Animal Care and Use Program at the University of Michigan.

Isolation of porcine SPARC

Tooth germs of permanent molars were surgically extracted from the maxillae and mandibles of six-month-old pigs at the Michigan State University Meat Laboratory (East Lansing, MI). The soft tissue and then part of the enamel layer of 55 molars were removed by scraping with a curette. Each molar crown was separately embedded in paraffin with the dentin side facing up. Each extraction buffer contains Protease Inhibitor Cocktail Set III (Calbiochem, La Jolla, CA, USA) and 1 mM 1,10-phenanthroline (Sigma, St. Louis, MO). Dentin proteins were extracted from the pulpal side of the teeth using sequential guanidine and guanidine/EDTA extractions as described previously

(Yamakoshi *et al.*, 2005). The EDTA (ethylene-diamine-tetra-acetic acid) extracts were pooled and fractionated by anion exchange chromatography using a Q-Sepharose Fast Flow column (Amersham Pharmacia, Uppsala, Sweden). SPARC was identified in the first peak (Q1), which was further fractionated by size exclusion chromatography with a Superdex 200 column (Amersham Biosciences Corp., Piscataway, NJ, USA).

Characterizing porcine SPARC protein

SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE) was performed using PAGE® Gold Precast Gels (4-20 % gradient Tris-Glycine gels) (Cambrex Bio Science Rockland Inc., Rockland, ME, USA). Samples were dissolved in Lammli sample buffer (Bio-Rad, Hercules, CA, USA). Following electrophoresis, the gels were stained with Bio-Safe Coomassie (Bio-Rad) or Stains-all (Sigma). The apparent molecular weights of protein bands were estimated by comparison with the SeeBlue® Plus2 Pre-Stained Standard (Invitrogen, Carlsbad, CA, USA). Proteins in the gel were transblotted to a Sequi-Blot PVDF Membrane (Bio-Rad) and strips containing SPARC were sent to the W. M. Keck Foundation Biotechnology Resource Laboratory at Yale University for automated Edman sequencing, which was performed using an Applied Biosystems Procise 494 cLC protein sequencer.

Cloning the porcine SPARC cDNA

The bovine SPARC cDNA sequence (Acc# NM_174464; (Bolander *et al.*, 1988)) was used to identify homologous porcine SPARC cDNA sequences in databases of

expressed sequence tags (ESTs) using BLASTN (Altschul *et al.*, 1997). Primers for the initial polymerase chain reaction (PCR) were designed using the porcine EST sequence (Acc# BU946733). A unidirectional cDNA library, which had been constructed from RNA extracted from the pulp organ of developing pig teeth (Hu *et al.*, 1996), was used as template. A segment of porcine SPARC cDNA was amplified using the primer pairs pSPARCf1 5'-ACCCACTAAGGAGCCCATTT and pSPARCr1 5'-GGGTGGAGGAACATGCTAGA, ligated into pCR2.1-TOPO (Invitrogen, Carlsbad, CA, USA) and transfected in *E. coli*. The PCR conditions used a 5 min denaturation at 94 ˚C, followed by 40 cycles each with denaturation at 94 ˚C for 30 sec, primer annealing at 60 ˚C, for 30 sec, and product extension at 72 ˚C for 30 sec, with a final extension at 72 ˚C for 7 min. DNA minipreps of the recombinant plasmid, which contained a 730 bp insert, were characterized by DNA sequencing using ABI Big Dye Terminator chemistry and an ABI 3100 automated DNA sequencer at the University of Michigan's DNA Sequencing Core. We used the insert sequence as a query to perform a BLASTN search of the pig EST database and designed a second set of PCR primers (pSPARCf2; 5'-ATGAGGGCCTGGATCTTCTT, pSPARCr2; 5'-TTCCACCCACACGTCACTAA), which allowed us to clone a 1192 bp amplification product containing the entire porcine SPARC coding region.

SPARC sequence analyses

The signal peptide cleavage site was predicted using Signal P 3.0 (Bendtsen *et al.*, 2004) at the Center for Biological Sequence Analysis Website (http://www.cbs.dtu.dk/services/SignalP/). The molecular weight and isoelectric point of

porcine SPARC was predicted using the Compute pI/Mw tool (http://ca.expasy.org/tools/pi_tool.html) on the Expasy molecular biology server (http://expasy.cbr.nrc.ca/tools/scnpsit1.html) at the Swiss Institute of Bioinformatics. Searches for potential phosphorylation, and O-linked and N-linked glycosylation sites were performed at the Centre for Biological Sequence Analysis at the Technical University of Denmark Website (http://www.cbs.dtu.dk/) using the NetNGlyc 1.0, (Blom *et al.*, 2004), NetOGlyc 2.0 (Hansen *et al.*, 1998), and NetPhos 2.0 (Blom *et al.*, 1999) prediction servers.

Computer modeling of the porcine SPARC 3-D structure

Pig SPARC was modeled by submitting the porcine amino acid sequence to Swiss-Model (http://swissmodel.expasy.org//SWISS-MODEL.html) (Schwede *et al.*, 2003) using human SPARC as a template (PDB:1bmo). Human SPARC molecules assemble into dimers (Maurer *et al.*, 1992). Since residues involved in dimerization are conserved between human and pig SPARC, the contacts of human SPARC were used for the pig dimer and assembled in Swiss-protein/deep view (http://us.expasy.org/spdbv/) (Guex and Peitsch, 1997). The model was then displayed using Protein-Explorer 2.45 beta (http://www.umass.edu/microbio/chime/pe/protexpl/frntdoor.htm) in Internet Explorer 6.0 and the plug-in, MDL Chime version 2.6 SP6 (http://www.mdl.com/).

RESULTS

Porcine SPARC was obtained from surgically extracted developing porcine molars (Fig. 2.3). SPARC was removed from dentin in the guanidine/EDTA extract.

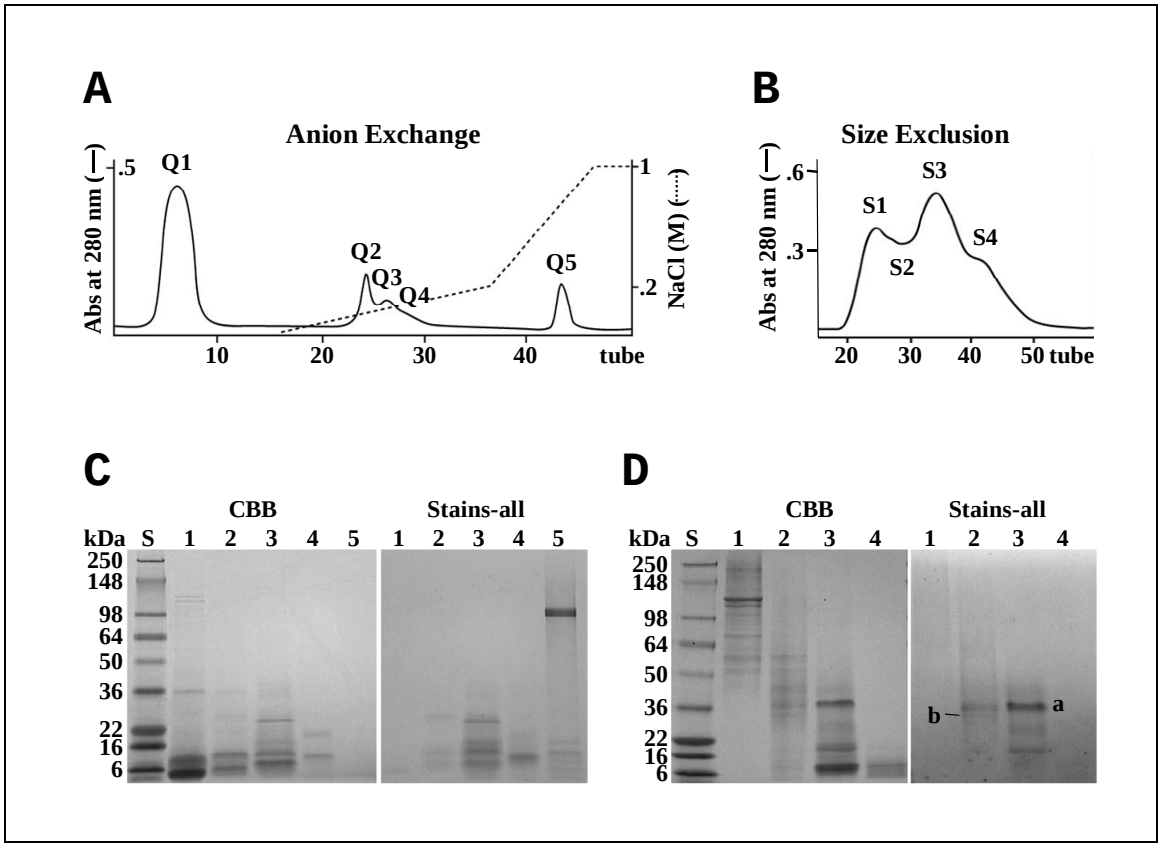

Figure 2.3: Isolation of SPARC from porcine dentin.
The guanidine/EDTA extraction of dentin was divided into 5 fractions by anion exchange chromatography (A). The anion exchange fraction (Q1) was then separated into 4 fractions by size exclusion chromatography (B). The contents of each fraction from the anion exchange and size exclusion separations were analyzed by SDS-PAGE stained with CBB and Stains-all (C and D, respectively). The size exclusion fractions were transblotted to a PVDF membrane and Stains-all positive bands ~ 20 kDa were characterized by Edman degradation. The 38-kDa band marked "a" in the S3 fraction gave the sequence: APQQEALPDETEV. A 35-kDa band marked "b" in the S2 fraction gave the sequence: DFEKNYNMYIFPV.

This extract was fractionated first by anion exchange chromatography (Fig. 2.3 A, 2.3 C) and then by size exclusion chromatography, being identified in the fractions corresponding to the first (Q1) and third (S3) chromatographic peaks respectively (Fig.

2.3 B, 2.3 D). The principle Stains-all positive protein in the Q1-S3 fraction migrated at 38 kDa, and gave the N-terminal sequence of APQQEALPDETEV. A lower abundance band was evident at 35 kDa in the Q1-S2 fraction and gave the N-terminal sequence of DFEKNYNMYIFPV. BLASTP searches of protein databases indicated that these sequences were likely to be derived from the porcine SPARC.

The porcine SPARC cDNA sequence was cloned and characterized from a porcine enamel organ epithelia-derived cDNA library. The cDNA and deduced amino acid sequences of porcine SPARC are shown in Fig. 2.4. Porcine SPARC contains 300 amino acids, including the 17 amino acid signal peptide. In the absence of post-translational modifications, the 283 aa secreted protein has a molecular mass of 32.3 kDa. Computer predictions indicate that porcine SPARC contains a single asparagine (Asn^{113}) in the appropriate context for glycosylation, while five threonines (Thr^{28}, Thr^{34}, Thr^{32}, Thr^{158} and Thr^{269}) and one serine (Ser^{245}) are potentially phosphorylated. Based upon the crystal structures of SPARC protein from other organisms, Asn^{113} is glycosylated, but no serine or threonine residues are likely to be phosphorylated.

The porcine SPARC deduced amino acid sequence was aligned to SPARC homologues from six other vertebrate species: *Homo sapiens*, *Bos taurus*, *Mus musculus*, *Gallus gallus*, *Xenopus laevis* and *Oryzias latipes* (Fig. 2.5). The alignment of SPARC sequences illustrates the remarkable conservation of primary amino acid sequence among widely divergent vertebrate species. Porcine SPARC, for instance, shares 96.2 % sequence identity with human SPARC.

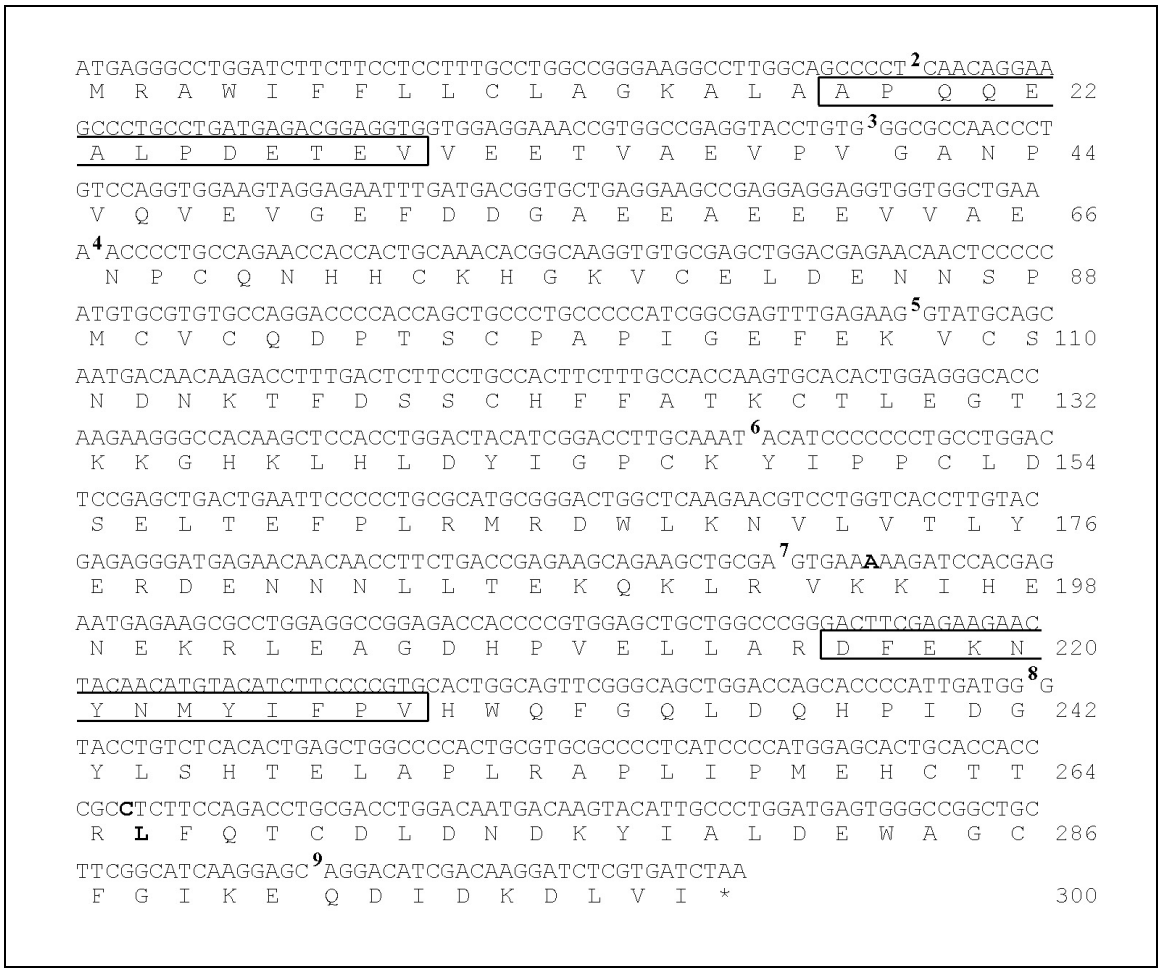

Figure 2.4: Analysis of porcine SPARC cDNA and deduced amino acid sequences.
The porcine SPARC coding region from exons 2 through 10 is shown. On the right are the numbers of the last amino acid in each row. There were two nucleotide sequence variations among the independently cloned cDNAs: the third position of codon 194 could be A or G, and the first position of codon 266 could be C or T. As a consequence, amino acid 266 could be Leu or Phe. The amino acid sequences determined by N-terminal sequencing of SPARC protein are boxed. *SignalP* predicted signal peptide cleavage site is after Ala^{17}, supporting the conclusion that the first boxed sequence (APQQ…) corresponds to the N-terminus of the secreted protein. The second boxed sequence (DFEK…) must have been generated by cleavage of the protein between Arg^{215} and Asp^{216}. The positions of introns, assigned by analogy to the human SPARC gene, are indicated by bold numbers. Introns 2, 3, 5, and 7 are phase 0; introns 4, 6, and 9 are phase 1, and intron 8 is phase 2. *NetNGlyc* predicted that Asn113 is glycosylated; the homologous amino acid was glycosylated in the human and bovine crystal structures. *NetPhos* predicted that threonines at positions 28, 34, 132, 158, 269, and Ser^{245} are phosphorylated. The human and bovine crystal structures however do not show any phosphorylated residues, although the N-terminal region has not been crystallized.

The crystal structure of human SPARC shows that all of that protein's cysteines form disulfide bridges (Hohenester *et al.*, 1996; Hohenester *et al.*, 1997). By analogy to human

SPARC, the 14 cysteines in the secreted porcine homologue are involved in the following disulfide bridge pairings: Cys^{69}:Cys^{80}; Cys^{74}:Cys^{90}; Cys^{92}:Cys^{127}; Cys^{98}:Cys^{120}; Cys^{109}:Cys^{146}; Cys^{152}:Cys^{262}; Cys^{270}:Cys^{286}.

SPARC is comprised of three functional domains: a flexible N-terminal domain, a follistatin-like (FS) domain, and an extracellular calcium-binding (EC) domain. The positions of these domains in the SPARC primary structure are indicated in Fig. 2.6.

The flexible N-terminal domain is an acidic, calcium-binding domain that is encoded by exons 3 and 4. This region of SPARC is the most variable across species and contains none of the disulfide bridges, which characterize the rest of the protein. The SPARC N-terminal domain has never been crystallized, so its 3D structure has not been determined. The porcine SPARC N-terminal domain extends from the Ala^{18} to Glu^{66}. Seventeen of its 50 amino acids have negatively charged side chains (Glu & Asp), while none are positively charged (Lys & Arg). Its calculated isoelectric point is 3.1.

The follistatin-like domain extends from Asn^{67} to Pro^{151}. Much of the FS domain is homologous to follistatin and serine protease inhibitors of the Kazal family. The segment extending from Val^{91} to Cys^{146} is described as the Kazal-like domain, although SPARC may not have protease inhibitor activity and its N-terminal domain might participate in the activation of MMP-2 at the cell surface (Gilles *et al.*, 1998).

The extracellular calcium binding (EC) domain extends from Cys^{152} to Ile^{300} at the C-terminus. The FS and EC domains are not entirely autonomous, and the carbohydrate attachment to Asn^{113} contacts Tyr^{148} in the interface between the FS and EC domains, possibly stabilizing the domain pair and preventing proteolysis of the peptide loop that connects them.

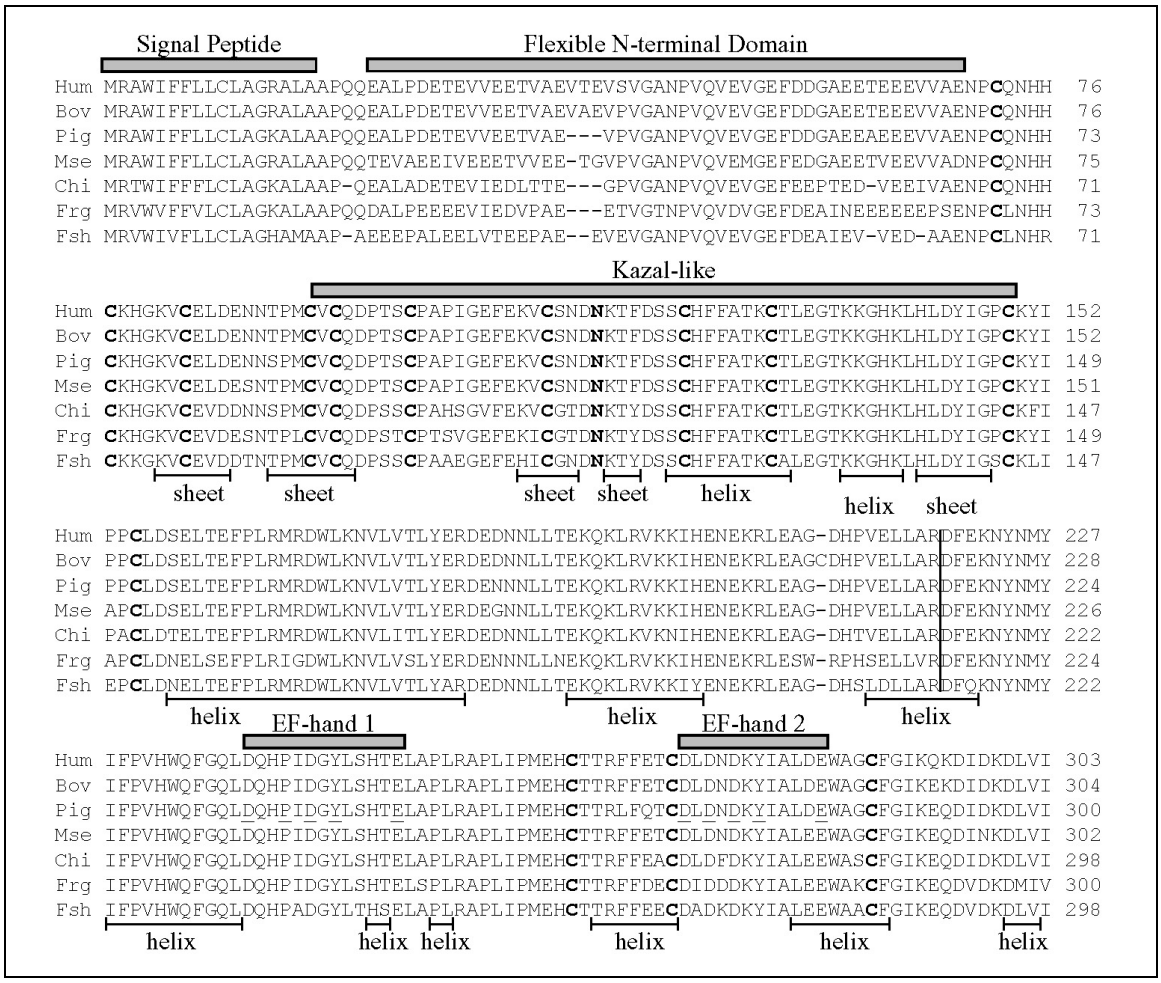

Figure 2.5: Analysis of aligned SPARC protein sequences.
The deduced SPARC amino acid sequences from Hum (human; *Homo sapiens* Acc# P09486), Bov, (bovine; *Bos taurus*, Acc# P13213), pig (*Sus scrofa*, Acc# AY963262), Mse (mouse; *Mus musculus*, Acc# P07214), Chi (Chicken, *Gallus gallus*, Acc# P36377), Frg (African clawed frog, *Xenopus laevis*, Acc# P36378), and Fsh (Japanese ricefish, *Oryzias latipes*, Acc# Q6PVV4). The domain structures of SPARC are indicated by bars above the alignment. In porcine, these include the signal peptide (Met^{1} to Ala^{17}), the flexible N-terminal domain (Ala^{18} to Glu^{66}), a follistatin (FS) like domain (not shown; Asn^{67} to Pro^{151}) containing the Kazal-like domain (Val^{91}-Cys^{146}), and an extracellular (EC) calcium binding domain (not shown; Cys^{152} to Ile^{300}) containing two EF-hands (Asp^{236}-Glu^{248} and Asp^{271}-Glu^{282}). Acidic amino acid side chains involved in coordinating the Ca^{+2} ion in the EF-hands are underlined (Asp^{236}, Asp^{241}, Glu^{248}, Asp^{271}, Asp^{273}, Asp^{275} and Glu^{282}) as are the amino acids whose backbone carbonyls help coordinate Ca^{+2} ions (Pro^{239}, Tyr^{243}, Tyr^{277}). The porcine SPARC amino acids involved in the coordination of Ca^{+2} ions by the EF-hands is by comparison to the analyses of human SPARC (Busch *et al.*, 2000). Post-translational modifications are indicated in bold. All 14 cysteines in the secreted protein are involved in disulfide bridges. The disulfide pairs are Cys^{69}:Cys^{80}; Cys^{74}:Cys^{90}; Cys^{92}:Cys^{127}; Cys^{98}:Cys^{120}; Cys^{109}:Cys^{146}; Cys^{152}:Cys^{262}; Cys^{270}:Cys^{286}. Asn^{113} is glycosylated. Sheet and helix secondary structures are indicated by line segments below the alignment. A vertical line shows the cleavage site we identified in porcine SPARC between Arg^{215} and Asp^{216}.

The defining feature of the EC domain is the presence two EF hands that strongly bind calcium. EF-hand 1 extends from Asp^{236} to Glu^{248}; EF-hand 2 extends from Asp^{271} to Glu^{282}.

Human SPARC served as template for the homologous modeling of porcine SPARC. Human SPARC (303 aa) and porcine SPARC (300 aa) differ by 7 amino acid substitutions and a single 3 aa deletion (Fig. 2.5). In the region of the computer modeling, the homology is even higher: porcine and human SPARC have the same length, and 229 out of 233 amino acids (>98 %) are identical. Two techniques were used to assess the precision of the protein homology modeling of porcine SPARC. The first was to generate a graphical model of the porcine SPARC dimers that colors each atom according to the temperature factors that had been calculated for it. The second was to generate a graphical model of the porcine SPARC dimer showing the regions of amino acid identity between porcine SPARC and its reference structure (human SPARC).

Temperature factors are a measure of the mobility or uncertainty of each atom's position. High values are colored in warmer (more red) and lower values in colder (more blue) colors. The more freedom that an amino acid has to move around, the higher will be its temperature factor. A graphic representation of the temperature factors in the porcine SPARC model is shown in Fig. 2.6 A. The temperatures are deep blue throughout the model. Small islands of green identify the 4 amino acids side chains that are not conserved. They are all located on the surface where their R-groups are easily accommodated and do not interfere with the positioning of adjacent amino acids, which remain a deep blue. Thus, the warmest colors in the temperature model (light green) do not extend beyond the positions of the substituted amino acids, which are shown in Fig.

2.6 B. This analysis suggests that the structure of the porcine SPARC homology model closely approximates the actual structure of the protein itself.

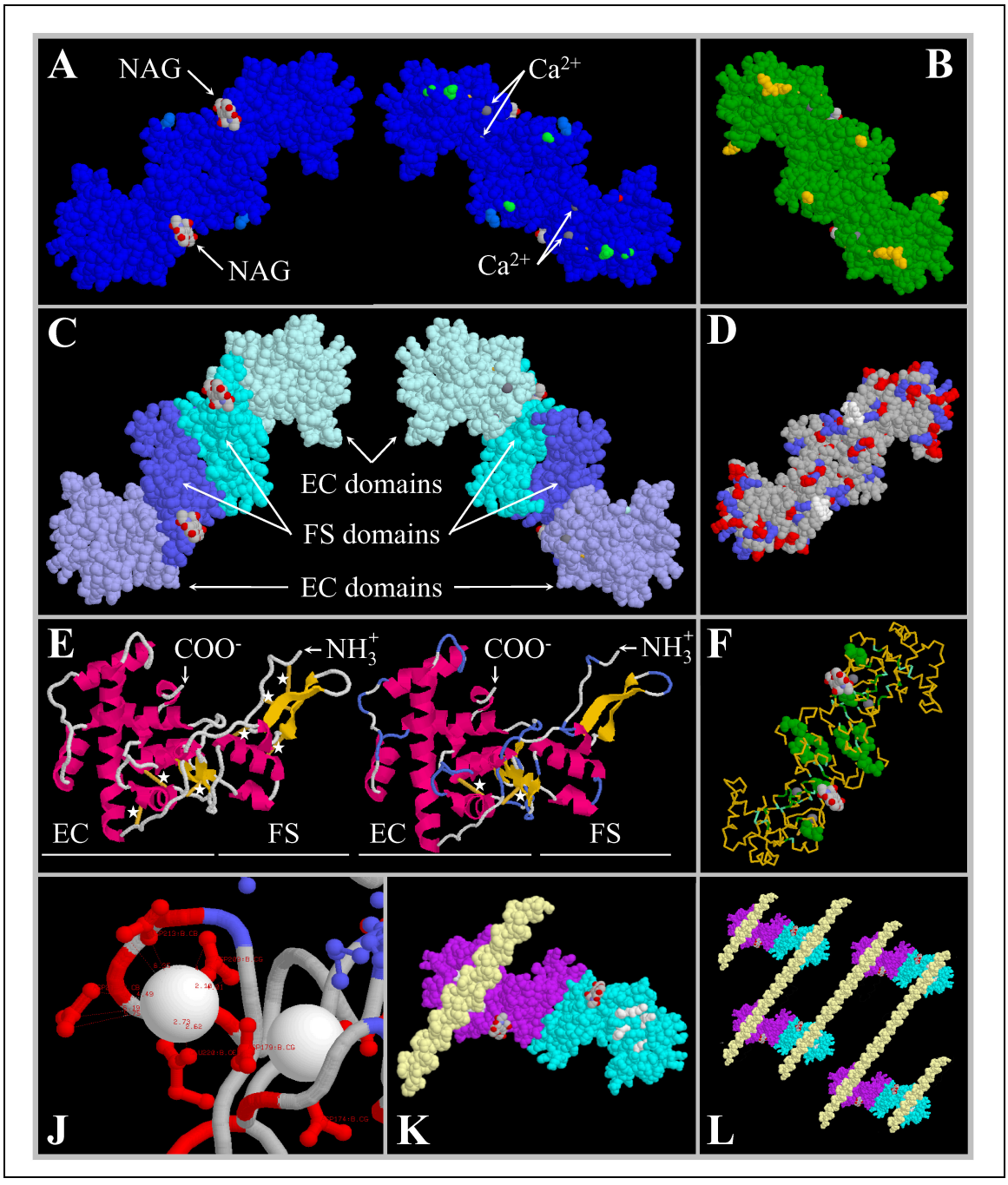

Figure 2.6: Computer modeling of the porcine SPARC homodimer.
The protein homology model of porcine SPARC was generated using the human SPARC structure (PDB:1bmo) as a template. Since the N-terminal acidic-rich domain has not been crystallized, only the FS domain and the EC domain were modeled (Hohenester *et al.*,

1997). Temperature analysis of the modeled SPARC dimer validated the porcine SPARC structure (A). Blue indicates atoms in the protein that lack of mobility, which is indicative of a tightly folded protein. Higher mobility is indicated in green and the most mobility by red. The N-acetylglucosamine (NAG) stubs of the glycosylations are shown in CPK colors (C, gray; O, red; N, blue). The calcium ions (gray) coordinated by the EF hands are labeled. Identity analysis (B) shows that most of the protein is identical with the human template (green), with the few amino acid changes (yellow) localizing to the protein surface. Domain analysis (C) shows SPARC dimer has an extended shape, with the EC domains on the ends (pale and light blue) and the FS domains in the center (blue and cyan) forming the interface between the monomers. Charge distribution analysis (D) shows acidic (red) and basic (blue) residues concentrate at the ends of the dimer. Some charged residues in the body of the dimer are evenly distributed in a linear pattern. NAG is displayed in white. Secondary structure analysis (E) shows α-helices (red ribbons), β-strands (yellow ribbons), turns (blue chords on porcine model) and disulfide bridges (yellow chords with stars). The human SPARC monomeric structure is on the left. The porcine SPARC monomeric structure is on the right. The human SPARC monomer (left) has 7 disulfide bridges with a bond distance of 3 Å. The porcine SPARC monomer (right) displays only 2 disulfide bonds. This discrepancy is due to limitations in the CHIME software and bond distances within 3 Å were confirmed for the seven cysteine pairs in the porcine model. Conservation analysis (F) was based upon alignment of the 18 known SPARC sequences in vertebrates and invertebrates. Amino acids identical in all 18 species are shown in green. Cysteines and the NAG attachment are space-filled. All of the cysteine pairs are conserved throughout phylogeny. The paired EF hands in the EC domain are shown in detail (J). The type I collagen triple helix (PDB:1cgd) binds to SPARC (K). SPARC participates in the organization of collagen networks (L). The SPARC monomers are shown in cyan and magenta, with amino acids known to participate in collagen (yellow) binding shown in white.

The computer modeling of porcine SPARC permitted the imaging of its domain structure (Fig. 2.6 C). The space-filled model illustrates the linear arrangement of the EC and FS domains in the SPARC dimer. The dimerization interface is formed by the FS domains. The N-linked glycosylation is located within the FS domain between the dimerization interface and the connecting loop between the FS and EC domains. The two carbohydrate attachments protrude from the same side of the dimer. The C-terminal EC domains are located on the ends of the dimer, which has a more highly charged surface than other regions of the protein (Fig. 2.6 D). Charged amino acids are conspicuously absent from the large grooves that run on opposite sides of the SPARC dimer along the EC-EC interface.

The secondary structure of human and porcine SPARC, which is mainly comprised of α-helices, β-strands and turns, is rigidly cross-linked by 7 disulfide bonds (Fig. 6 E). The FS domain is comprised of two subdomains. The N-terminal part is homologous to epidermal growth factor (EGF), having a β hairpin fold and 4 cysteines connected in a 1-3 and 2-4 pattern. The C-terminal part is a Kazal-like domain and connects the cysteines in a 5-9, 6-8, and 7-10 arrangement. These cysteines are conserved in an alignment of all known SPARC proteins, which includes sequences from invertebrates (Fig. 2.6 F). The EC domain is mainly helical except a small β sheet between the EF hands. The helices are sometimes targeted for cleavage by MMPs or other proteases (Mann *et al.*, 1987; Maurer *et al.*, 1997b; Sage *et al.*, 2003; Sasaki *et al.*, 1997). Cleavage exposes binding sites for type I and type IV collagens (Sasaki *et al.*, 1998). The main structural feature of the EC domain is a pair of EF hands that both bind a calcium ion in a pentagonal bipyramidal coordination using an acidic loop bracketed by two helices (Fig. 2.4 J). The calcium ions are closely bound, with distances ranging from 2.1 to 4.29 Å (Busch *et al.*, 2000).

Collagen and SPARC have an intimate, bidirectional relationship. SPARC plays an important role in regulation of collagen fibrillogenesis (Bradshaw *et al.*, 2002; Bradshaw *et al.*, 2003; Puolakkainen *et al.*, 2003). Collagen binding sites have been mapped to five residues within the SPARC EC domain (Sasaki *et al.*, 1998), covering an area correlating with the diameter of type I collagen. A SPARC dimer may interact with two collagen fibrils (one collagen per SPARC monomer), and organize them into a network (Fig. 2.6 K, 2.6 L).

DISCUSSION

SPARC and Structure & Function

SPARC is an ECM protein expressed in various tissues during development and tissue remodeling, where it serves important basic functions that have been conserved throughout evolution. The special conservation of exon 2 between SPARC and the noncollagenous proteins involved in vertebrate mineralization suggests that the common feature of these proteins is the presence of an appropriate targeting system for their secretion into the extracellular environment where mineralization occurs. All signal peptides are not equal. As the signal peptide emerges from the ribosome it attaches to a signal recognition particle (SRP), which mediates its binding and translocation through an aqueous pore (translocon) in the ER membrane (Luirink and Sinning, 2004; Schneider and Fechner, 2004). A translocating-chain-associated membrane protein (TRAM) is sometimes required, depending upon the signal peptide structure and length of its charged N-terminal region. Proteins associated with selected translocons, such as oligosaccharyl transferase (which adds N-linked carbohydrates to the nascent chain), can influence the post-translational modifications attached to the secreted protein (Ott and Lingappa, 2002). Conservation of the signal peptide on exon 2 is the central feature that links SPARC, SCPP including amelogenin in vertebrate mineralized tissues. This commonality indicates that preservation of a selective mechanism for the targeting of select proteins and the exclusion of all others is an essential element in the regulation of mineralization (Kawasaki and Weiss, 2003; Kawasaki and Weiss, 2008).

Structural analysis of the SPARC protein provides insight into the nature of extracellular matrix that preceded vertebrate biomineralization, and upon which bone,

dentin, cementum first formed. The structural/functional features that are most evident from the SPARC model are that its structure is particularly sturdy and tightly folded, with 7 disulfide bridges and 2 EF hands both coordinating Ca^{2+}, and that it forms dimers that bind and organize other extracellular matrix molecules, such as collagen. Mutation of one cysteine leads to the loss of SPARC conformation (Sage *et al.*, 1995) and consequently loss of its antiproliferative activity (Frizell *et al.*, 1995) and ability to disassemble focal adhesions (Murphy-Ullrich *et al.*, 1995). Other proteins that contain FS domains bind heparin or proteoglycans (Innis and Hyvonen, 2003; Inouye *et al.*, 1992). SPARC has important roles in basement membranes. The dermal skeleton of nematodes exhibited an uncoordinated (UNC) phenotype with SPARC overexpression (Schwarzbauer and Spencer, 1993). Reduction of SPARC expression in *C. elegans* by RNAi was lethal for some of the worms (Fitzgerald and Schwarzbauer, 1998). Viable worms had a 'clear' phenotype and were small in size.

Formation of the extracellular matrix (ECM) into large assemblies is essential for mechanical integrity and cell-matrix interactions. SPARC and collagen form an intimate relationship. SPARC-deficient mice showed impaired collagen fibrillogenesis (Bradshaw *et al.*, 2002; Bradshaw *et al.*, 2003; Puolakkainen *et al.*, 2003), even though at the mRNA level type I collagen expression was similar to the wildtype (Bradshaw *et al.*, 2002; Francki *et al.*, 1999; Puolakkainen *et al.*, 2003). Conversely, mice lacking type I collagen expression failed to deposit SPARC (Iruela-Arispe *et al.*, 1996). SPARC also has high affinity for collagen IV (Maurer *et al.*, 1997a; Pottgiesser *et al.*, 1994), which increased following treatment with MMP-3. Collagen IV associates with collagen I via decorin to form networks of micro-fibrillar structures with collagen IV coating collagen I to rather

than forming fibers (Keene *et al.*, 1998). Therefore, SPARC could participate in a variety of collagen networks that influence matrix assembly and organization.

Biomineralization evolved

The fundamental importance of SPARC is reflected by its role as the progenitor gene for the superfamily of secretory calcium-binding phosphoproteins (SCPPs) involved in core vertebrate feeding adaptations, the mineralized skeleton, and the evolution of saliva and lactation (Kawasaki and Weiss, 2003; Kawasaki and Weiss, 2008). The study of SPARC promises more insights into features common to all vertebrate mineralization systems, as well as a better understanding of how collagen extracellular matrices organize and provide the structural frameworks for mineralizing and nonmineralizing tissues.

Biomineralization has evolved in distant lineages independently. Through phenogenetic drift, diverse genotypes with similar phenotypes have been produced among mineralized tissues (Weiss, 2005). The trait of mineralized tissue has been retained, but the genome has changed significantly through duplication and deletion events. The 'birth-and-death" of genes is an important mechanism for expanding and shrinking gene clusters (Nei and Rooney, 2005). The evolutionary fates of new genes are either being silenced as pseudogenes or being fixed through evolution of redundancy, subfunctionalization or neofunctionalization (Walsh, 1995; Zhou and Wang, 2008). New genes usually show accelerated sequence/structure changes compared to their ancestors. New genes and altered expression pattern of new genes are exposed to positive selection and fixation within the genome. Gene redundancies increase the dosage of genes with the same function. In subfunctionalization, a part of the gene is maintained and optimized. In

neofunctionalization, the function of the gene is altered. Duplicated genes create redundancies, are prone to accumulating mutations and are eventually silenced as pseudogenes. During evolution, the SCPP cluster has gained and lost genes as the result of selection pressure for specific functions.

Evolution of dental mineralized tissues

Examples of pseudogene and loss of genes from the SCPP cluster are baleen whales and birds. Baleen whales exploit the filter feeding sparing the use of teeth. During early development, teeth are initiated, but are arrested in the bell stage. Odontoblasts start laying down dentin matrix, but ameloblast never differentiate and the tooth bud is resorbed (Ishikawa *et al.*, 1999). The enamel proteins Ambn and Enam are silenced in baleen whales through frameshift mutations introducing premature stop codons (Demere *et al.*, 2008).

In modern birds, the capacity to form teeth has been lost. Early in morphogenesis, tooth initiation is interrupted prior to forming an enamel knot. If signaling molecules are supplied by transplants, tooth development progresses (Harris *et al.*, 2006). However, enamel formation would not be possible since the enamel SCPP genes have been silenced through mutations or complete deletions. The Amel gene on chromosome 1 has been reduced to a pseudogene with several mutations. Ambn and Enam genes have been completely deleted following severe intrachromosomal arrangements in the chicken genome (Kawasaki and Weiss, 2006; Kawasaki and Weiss, 2008; Sire *et al.*, 2008) (Fig. 2.7). Inverse to the lack of teeth in birds is the development of specialized keratinized mucosa (rhamphotheca) and a beak as feeding strategies.

The combination of expansion of the SCPP cluster through duplication events and lack of natural selection pressure resulted in different strategies of mineralization (Donoghue *et al.*, 2006). Mineralized tissues in teeth consist of layers of dentin, cementum, enamel or enameloid. Enamel and enameloid are functionally equivalent as they cover the teeth, have higher degrees of mineral and are more resistant to wear, but they differ in their composition. Enameloid is secreted by odontoblasts and inner dental epithelium and its mineralization is based on a Col type I network. In the fish genome of fugu and zebrafish, enamel SCPPs including Amel are completely absent. SCPPs have expanded recently in the sarcopterygian lineage that diverges mammals and amphibians from fish (teleosts) (Kawasaki and Weiss, 2006). The tetrapod genome acquired enamel specific SCPPs. SCPP genes in teleosts and tetrapods have experienced independent gene duplication events. Therefore, the genetic origin in different mineralized tissues suggests convergent evolution. The diversity of mineralized tissues in teeth is the result of phenogenetic drift (Weiss, 2005). The constraints for genes for mineralization are very flexible facilitating gene duplications and giving rise to different strategies for mineralization.

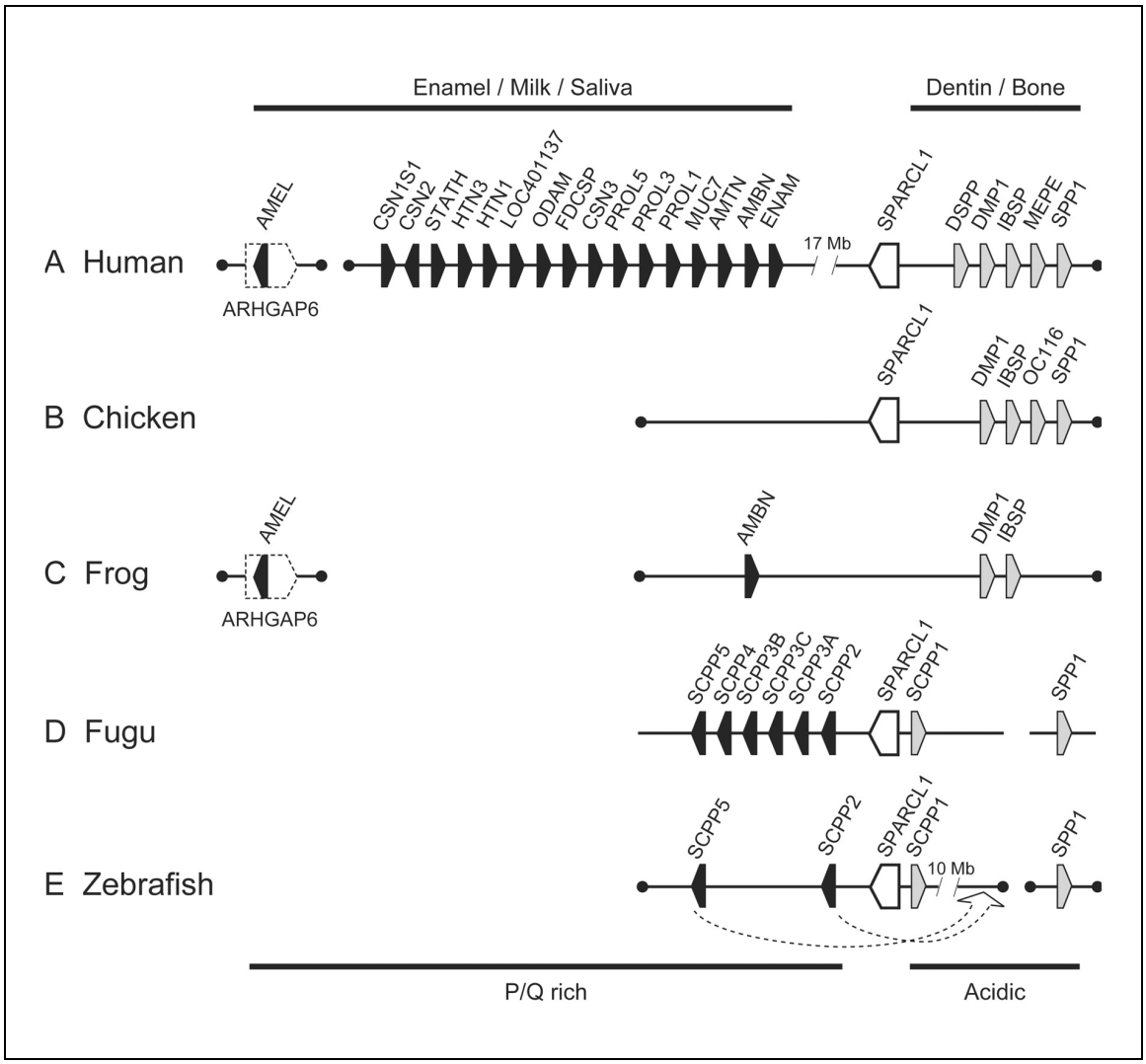

Figure 2.7: SCPP cluster in mammalian, avian, amphibian and fish species.
Human (A), chicken (B), frog (C), fugu (D), zebrafish (E) genomes (Kawasaki and Weiss, 2008)**.** Location and transcriptional orientation of SPARCL1 (open box), P/Q-rich SCPP (closedbox), acidic SCPP (shaded box), and ARHGAP6 (dashed-line box) genes are schematically illustrated. (A,C) Amel is located within ARHGAP6 on human chromosome X and in the frog genome. (A) Two large SCPP gene clusters, enamel/milk/saliva and dentin/bone classes, are separated by 17 megabases (Mb). (D) Linkage of SPP1 and the other SCPP genes is unknown. (E) SPP1 and the other SCPP genes reside on different chromosomes. SCPP2-SCPP5 and SPARCL1-SCPP1 clusters are separated by 10 Mb by an intrachromosomal re-arrangement.

The variation in the coding sequence of the SCPPs is opposed by the high degree of conservation in the signal peptide (Fig. 2.8).

The signal peptide at the N-terminus of the protein targets secreted proteins to the endoplasmatic reticulum (ER). It usually contains one or more positively charged amino acids followed by a hydrophobic core of 6-12 residues. The hydrophobic core is essential for crossing the ER membrane into the lumen. The conservation in the signal peptide between SPARC, SPARCL and the enamel SCPPs is very high. The ability to target the extracellular space offered novel opportunities to the SCPPs to diverge.

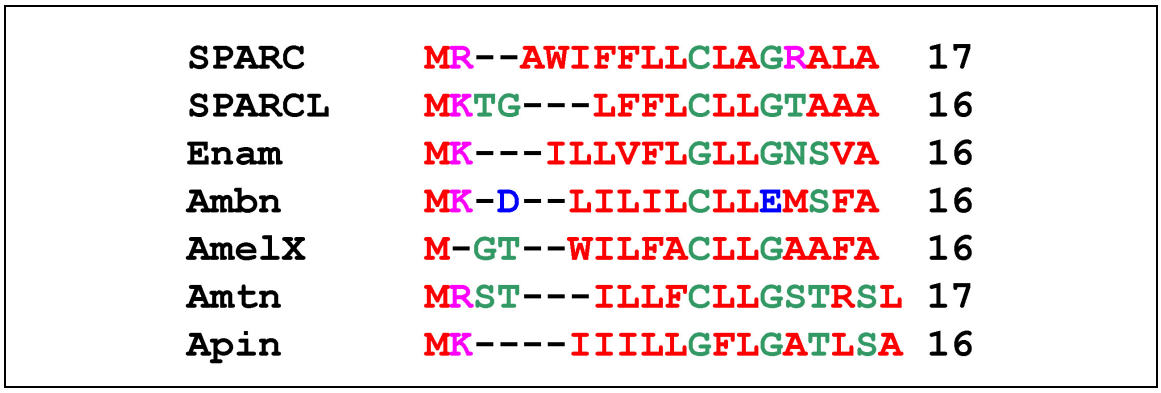

Figure 2.8: Alignment of signal peptides of human enamel SCPPs.
The signal peptide consists of 16 or 17 residues with a hydrophilic residue in the beginning followed by a hydrophobic core. The color code represents properties of amino acids: red hydrophobic, pink hydrophilic positive charge, blue hydrophilic negative charge, green hydrophilic uncharged.

Proteins without a native fold are also called intrinsically disordered or intrinsically unstructured (Dunker and Obradovic, 2001; Weinreb *et al.*, 1996). Regions without a tight three-dimensional fold may change conformation depend on specific environmental conditions. Ambn binds calcium ions from the surrounding environment and obtains a more compact fold that maintains potential cleavage sites exposed and accessible (Vymetal *et al.*, 2008) (Fig. 2.9).

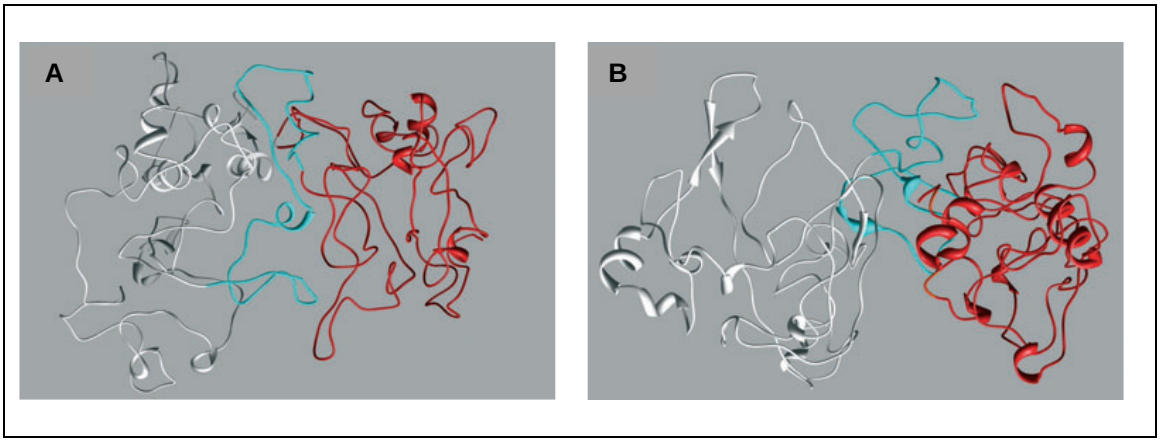

Figure 2.9: Conformation of Ambn changes in the presence of Ca^{2+}.
Human ameloblastin (AMBN) was simulated in water (A) and in the presence of 40 mM Ca^{2+} ions (B). Three-dimensional (3D) molecular structure of human AMBN was modeled using the ROBETTA server with the Rosetta *de novo* structure prediction method. The dynamic simulation of the protein conformation in water and Ca^{2+} ions was performed by using GROMACS. (Vymetal *et al.*, 2008).

Buried amino acids are highly constrained in tight conformation while unstructured regions are flexible. Residues that are facilitating stable conformation are less common in unstructured regions (e.g. C, W, Y, F, I, V, L). In contrast, residues with greater steric freedom are represented in higher proportion in unstructured proteins (M, K, R, S, Q, P, E). They tend to have post-translational modifications, alternatively spliced and important for signaling and regulatory processes (Dunker *et al.*, 2008). Unstructured regions of proteins without a tight fold evolve faster than ordered regions (Brown *et al.*, 2002). The unstructured regions may contain exposed residues that evolve faster. Conversely, proteins with structural or functional constraints evolve slower. The SCPP cluster is an example of this scenario with SPARC having a very compact fold dictated by its amino acid composition and coordination by calcium ions. From this ancestor the SCPP radiated maintaining the properties of secretion, calcium rich environment and phosphorylation into different functional contexts. Ameloblastin is representative for a

SPARC derived protein. Besides the features of the SCPP family, it acquired unstructured regions that could experience less selective constraints. Interestingly, the unstructured regions coincide with sites cleaved by enamel proteinases.

Ameloblastin and evolution of enamel prisms

The organization of enamel is characterized by presence or absence of prisms. The enamel crystallites are organized in rods and interrods, which are separated by the narrow sheath space (Coolidge, 1965; Smreker, 1903). The prism sheath separates a distinct change in crystal orientation (Sicher, 1962). As the rod interrods are composed of hydroxyapatite to >90 % (weight) providing strength for axial loading, enamel gains mechanical by breaking up the prisms by the sheath space and by arranging the prisms in an interwoven pattern. In contrast to rod and interrod enamel, the sheath of the rod contains few crystals but a high proportion of cleavage products of enamel proteins even at maturation stage (Meurman and Frank, 1991). The sheath space corresponds to the lateral surfaces of the Tomes' process, which are not secreting proteins and are not associated with the extension of enamel crystallites. During maturation stage, cleavage products of enamel proteins are directed to the sheath space where they concentrate. The sheath space extends from the dentino-enamel junction up to the enamel surface (Coolidge, 1965). The sheath space is thought to serve as an endocytosis path for enamel cleavage products in deeper enamel layers. In particular, Ambn N-terminal cleavage products concentrate in the sheath space during maturation stage.

Prism formation is clearly developed in mammals. Reptilians have non-prismatic enamel with a few exceptions. As most non-mammals lack prismatic organization, it has

been speculated that prismatic enamel has evolved from aprismatic enamel. The etiology for prismatic enamel has been intensely discussed to be related to prolonged occlusion, greater enamel thickness and location of stress points (Poole, 1956; Poole, 1957). Phylogenetic approaches may be helpful to gain insight into the development of the prism structure. Enamel organization and evolutionary lineage show inconsistencies. In allotherian (multituberculate) and metatherian (marsupial) mammals prism structure is missing and in the reptile *Gavialis gangeticus* enamel prisms are found (Sahni, 1987). The prismatic organization could have been acquired secondarily as the result of convergent evolution.

At the molecular level, it has been postulated that Ambn provides structure for the sheath space. In this context, aprismatic reptilian and amphibian Ambn gene structures have been studied (Shintani *et al.*, 2002; Shintani *et al.*, 2003) (Fig. 2.9). The caiman and frog Ambn have both 11 exon encoding a 407 or 408 amino acid residue proteins (Shintani *et al.*, 2002; Shintani *et al.*, 2003). The overall Ambn sequence similarity between reptiles and mammals is 38-47 %, and 29-37 % between amphibians and mammals, but higher for the N-terminal cleavage product and lower for the C-terminal cleavage product. Cleavage sites at the N-terminus are well conserved among mammals, reptiles and amphibians. The cleavage sites at the C-terminus are well conserved among mammals, but less between mammals and non-mammals. The post-translational modifications for phosphorylations by protein kinase C, tyrosine kinase, casein kinase II in exons 4, 6 and 10 (12) are highly conserved. Both sites for O-linked glycosylation in exon 6 and exon 11 (13) are conserved in mammal, amphibian and reptile sequence.

The conservation of cleavage sites in N-terminal products indicated the importance of generation of specific cleavage products in order to persist in mature enamel. Given that enamel proteinases are expressed in reptiles and amphibians (Shintani *et al.*, 2007), N-terminus of Ambn is very likely processed to cleavage products, which are not resorbed.

The transition from aprismatic to prismatic enamel is marked by a deviation in crystallite orientation. The discontinuity in crystallite orientation gradually increases from linear to planar patterns in an evolutionary period of 200 million years into pseudoprismatic enamel and prismatic enamel. The appearance of the convergence line was related to the development of a conical Tomes' process of the ameloblast (Lester and von Koenigswald, 1989). Enamel prisms developed in multituberculate mammals from circular prisms with complete prism sheath to arc-shaped prisms with incomplete prism boundaries (Krause and Carlson, 1986). The prism patterns differ in the number of ameloblasts contributing to it. The circular prisms are generated by one ameloblast, while the arc-shaped prisms arise from two to four ameloblasts (Boyde, 1969). Therefore, the prism patterns reflect the arrangement of the ameloblasts and the size and shape of the Tomes' process.

Since most modern reptiles or amphibians expressing Ambn but are lacking enamel prisms, Ambn may not be providing the structure of the sheath space. Ambn could be pushed into sheath space as a result of degradation of the enamel proteins and crystal maturation in width. As paleontological studies rely on mineral structures of fossil material, dynamic processes at the mineralization front adjacent to the ameloblast membrane are precluded from characterizations.

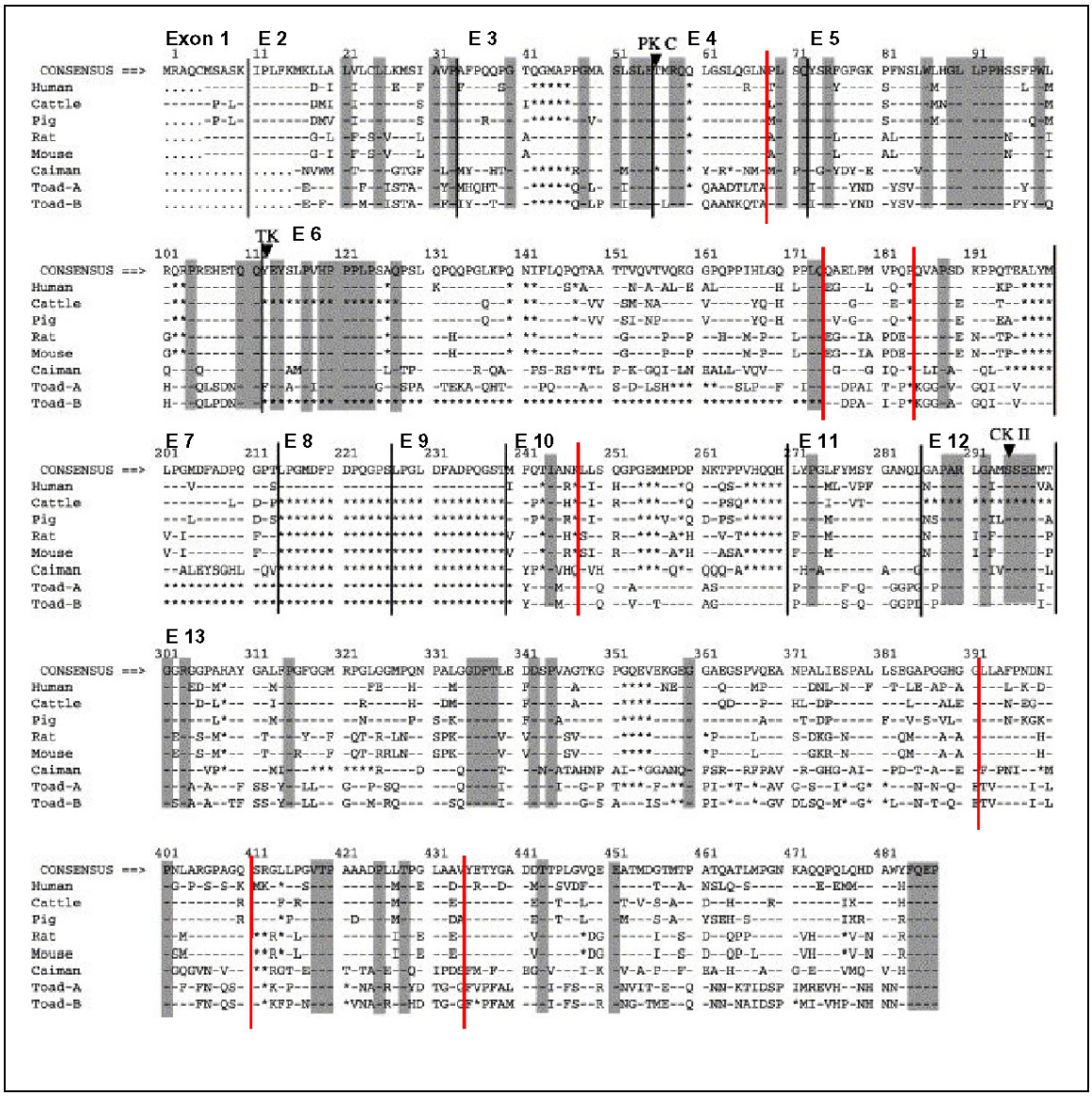

Figure 2.10: Multiple sequence alignment of Ambn using protein sequences of human, cattle, pig, rodents, caiman, and toad.
The sequence of Ambn is highly conserved in different species. Only in humans exons 8 and 9 were duplicated from exon 7. Dash – indicates sequence identity, star deletion, grey shaded boxes are highly conserved; potential phosphorylation sites are PKC protein kinase C, TK tyrosine kinase, CKII Golgi casein kinase. The red bars indicate cleavage sites; E indicates exons encoding amino acids. Modified from (Shintani *et al.*, 2003).

The Tomes' process of the ameloblast seems to be crucial for the orientation of crystals and sheath space. Disruption of the mineralization front results in loss of prism

organization. Consistent with the evolution of the SCPPs, structural organization of enamel has experience convergent development.

REFERENCES

Altschul S, Madden T, Schäffer A, Zhang J, Zhang Z, Miller W, Lipman D (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25:3389-3402.

Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004). Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 340:783-795.

Blom N, Gammeltoft S, Brunak S (1999). Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J. Mol. Biol.* 294:1351-1362.

Blom N, Sicheritz-Ponten T, Gupta R, Gammeltoft S, Brunak S (2004). Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* 4:1633-1649.

Bolander ME, Young MF, Fisher LW, Yamada Y, Termine JD (1988). Osteonectin cDNA sequence reveals potential binding regions for calcium and hydroxyapatite and shows homologies with both a basement membrane protein (SPARC) and a serine proteinase inhibitor (ovomucoid). *Proc Natl Acad Sci U S A* 85:2919-2923.

Boyde A (1969). Correlation of ameloblast size with enamel prism pattern: use of scanning electron microscope to make surface area measurements. *Z Zellforsch Mikrosk Anat* 93:583-593.

Bradshaw AD, Reed MJ, Sage EH (2002). SPARC-null mice exhibit accelerated cutaneous wound closure. *J Histochem Cytochem* 50:1-10.

Bradshaw AD, Puolakkainen P, Dasgupta J, Davidson JM, Wight TN, Helene Sage E (2003). SPARC-null mice display abnormalities in the dermis characterized by decreased collagen fibril diameter and reduced tensile strength. *J Invest Dermatol* 120:949-955.

Brekken RA, Sage EH (2000). SPARC, a matricellular protein: at the crossroads of cell-matrix. *Matrix Biol* 19:569-580.

Bronckers AL, Lyaruu DM, Woltgens JH (1989). Immunohistochemistry of extracellular matrix proteins during various stages of dentinogenesis. *Connect Tissue Res* 22:65-70.

Brown CJ, Takayama S, Campen AM, Vise P, Marshall TW, Oldfield CJ, Williams CJ, Dunker AK (2002). Evolutionary rate of heterogeneity in proteins with long disordered regions. *J Mol Evol* 55:104-110.

Busch E, Hohenester E, Timpl R, Paulsson M, Maurer P (2000). Calcium affinity, cooperativity, and domain interactions of extracellular EF-hands present in BM-40. *J Biol Chem* 275:25508-25515.

Coolidge TB (1965). Biochemistry of the sheath spaces in caries. *Ann N Y Acad Sci* 131:884-892.

Delgado S, Casane D, Bonnaud L, Laurin M, Sire JY, Girondot M (2001). Molecular evidence for precambrian origin of amelogenin, the major protein of vertebrate enamel. *Mol Biol Evol* 18:2146-2153.

Demere TA, McGowen MR, 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.

Donoghue PC, Sansom IJ (2002). Origin and early evolution of vertebrate skeletonization. *Microsc Res Tech* 59:352-372.

Donoghue PC, Sansom IJ, Downs JP (2006). Early evolution of vertebrate skeletal tissues and cellular interactions, and the canalization of skeletal development. *J Exp Zoolog B Mol Dev Evol* 306:278-294.

Dunker AK, Obradovic Z (2001). The protein trinity- linking function and disorder. *Nat Biotech* 19:805-806.

Dunker AK, Silman I, Uversky VN, Sussman JL (2008). Function and structure of inherently disordered proteins. *Curr Opin Struc Biol* 18:756-764.

Dziadek M, Paulsson M, Aumailley M, Timpl R (1986). Purification and tissue distribution of a small protein (BM-40) extracted from a basement membrane tumor. *Eur J Biochem* 161:455-464.

Engel J, Taylor W, Paulsson M, Sage H, Hogan B (1987). Calcium binding domains and calcium-induced conformational transition of SPARC/BM-40/osteonectin, an extracellular glycoprotein expressed in mineralized and nonmineralized tissues. *Biochemistry* 26:6958-6965.

Fedorov A, Fedorova L, Starshenko V, Filatov V, Grigor'ev E (1998). Influence of exon duplication on intron and exon phase distribution. *J Mol Evol* 46:263-271.

Fisher LW, Torchia DA, Fohr B, Young MF, Fedarko NS (2001). Flexible structures of SIBLING proteins, bone sialoprotein, and osteopontin. *Biochem Biophys Res Commun* 280:460-465.

Fisher LW, Fedarko NS (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.

Fitzgerald MC, Schwarzbauer JE (1998). Importance of the basement membrane protein SPARC for viability and fertility in Caenorhabditis elegans. *Curr Biol* 8:1285-1288.

Francki A, Bradshaw AD, Bassuk JA, Howe CC, Couser WG, Sage EH (1999). SPARC regulates the expression of collagen type I and transforming growth factor-beta1 in mesangial cells. *J Biol Chem* 274:32145-32152.

Frizell E, Liu SL, Abraham A, Ozaki I, Eghbali M, Sage EH, Zern MA (1995). Expression of SPARC in normal and fibrotic livers. *Hepatology* 21:847-854.

Fujisawa R, Kuboki Y (1989). Changes in levels of osteonectin in bovine dentine during tooth development. *Arch Oral Biol* 34:89-92.

Fujisawa R, Wada Y, Nodasaka Y, Kuboki Y (1996). Acidic amino acid-rich sequences as binding sites of osteonectin to hydroxyapatite crystals. *Biochim Biophys Acta* 1292:53-60.

Gilles C, Bassuk JA, Pulyaeva H, Sage EH, Foidart JM, Thompson EW (1998). SPARC/osteonectin induces matrix metalloproteinase 2 activation in human breast cancer cell lines. *Cancer Res* 58:5529-5536.

Guex N, Peitsch MC (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* 18:2714-2723.

Hansen JE, Lund O, Tolstrup N, Gooley AA, Williams KL, Brunak S (1998). NetOglyc: prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility. *Glycoconjugate J.* 15:115-130.

Harris MP, Hasso SM, Ferguson MWJ, Fallon JF (2006). The development of archosaurian first-generation teeth in a chicken mutant. *Curr Biol* 16:371-377.

Hohenester E, Maurer P, Hohenadl C, Timpl R, Jansonius JN, Engel J (1996). Structure of a novel extracellular Ca^{2+}-binding module in BM-40. *Nat Struct Biol* 3:67-73.

Hohenester E, Maurer P, Timpl R (1997). Crystal structure of a pair of follistatin-like and EF-hand calcium-binding domains in BM-40. *EMBO J* 16:3778-86.

Hu C-C, Bartlett JD, Zhang CH, Qian Q, Ryu OH, Simmer JP (1996). Cloning, cDNA sequence, and alternative splicing of porcine amelogenin mRNAs. *J. Dent. Res.* 75:1735-1741.

Innis CA, Hyvonen M (2003). Crystal structures of the heparan sulfate-binding domain of follistatin. Insights into ligand binding. *J Biol Chem* 278:39969-39977.

Inouye S, Ling N, Shimasaki S (1992). Localization of the heparin-binding site of follistatin. *Mol Cell Endocrinol* 90:1-6.

Iruela-Arispe ML, Vernon RB, Wu H, Jaenisch R, Sage EH (1996). Type I collagen-deficient Mov-13 mice do not retain SPARC in the extracellular matrix: implications for fibroblast function. *Dev Dyn* 207:171-183.

Ishikawa HH, Amasaki H, Dohguchi A, Furuya A, Suzuki K (1999). Immunohistological distributions of fibronectin, tenascin, type I, III and IV collagen, and laminin during tooth development and degeneration in fetuses of minke whale, *Balaenoptera acutorostrata*. *J Vet Med Sci* 61:227-232.

Kawasaki K, Weiss KM (2003). Mineralized tissue and vertebrate evolution: the secretory calcium-binding phosphoprotein gene cluster. *Proc Natl Acad Sci U S A* 100:4060-4065.

Kawasaki K, Suzuki T, Weiss KM (2004). Genetic basis for the evolution of vertebrate mineralized tissue. *Proc Natl Acad Sci U S A* 101:11356-11361.

Kawasaki K, Weiss KM (2006). Evolutionary genetics of vertebrate tissue mineralization: the origin and evolution of the secretory calcium-binding phosphoprotein family. *J Exp Zoolog B Mol Dev Evol* 306:295-316.

Kawasaki K, Buchanan AV, Weiss KM (2007). Gene duplication and the evolution of vertebrate skeletal mineralization. *Cells Tissues Organs.* 186:7-24.

Kawasaki K, Weiss KM (2008). SCPP gene evolution and the dental mineralization continuum. *J Dent Res* 87(6)*:*520-31.

Keene DR, Ridgway CC, Iozzo RV (1998). Type VI microfilaments interact with a specific region of banded collagen fibrils in skin. *J Histochem Cytochem* 46:215-220.

Krause DW, Carlson SJ (1986). The enamel ultrastructure of multituberculate mammals: A review. *Scan Electron Microsc* 4:1591-1607.

Lester KS, von Koenigswald W (1989). Crystallite orientation discontinuities and the evolution of mammalian enamel--or, when is a prism? *Scanning Microsc* 3:645-662; discussion 663.

Luirink J, Sinning I (2004). SRP-mediated protein targeting: structure and function revisited. *Biochim Biophys Acta* 1694:17-35.

Mann K, Deutzmann R, Paulsson M, Timpl R (1987). Solubilization of protein BM-40 from a basement membrane tumor with chelating agents and evidence for its identity with osteonectin and SPARC. *FEBS Lett* 218:167-172.

Maurer P, Mayer U, Bruch M, Jeno P, Mann K, Landwehr R, Engel J, Timpl R (1992). High-affinity and low-affinity calcium binding and stability of the multidomain

extracellular 40-kDa basement membrane glycoprotein (BM-40/SPARC/osteonectin). *Eur J Biochem* 205:233-240.

Maurer P, Gohring W, Sasaki T, Mann K, Timpl R, Nischt R (1997a). Recombinant and tissue-derived mouse BM-40 bind to several collagen types and have increased affinities after proteolytic activation. *Cell Mol Life Sci* 53:478-484.

Maurer P, Sasaki T, Mann K, Gohring W, Schwarzbauer JE, Timpl R (1997b). Structural and functional characterization of the extracellular calcium-binding protein BM-40/secreted protein, acidic, rich in cysteine/osteonectin from the nematode Caenorhabditis elegans. *Eur J Biochem* 248:209-216.

Meurman JH, Frank RM (1991). Progression and surface ultrastructure of in vitro caused erosive lesions in human and bovine enamel. *Caries Res* 25:81-87.

Murphy-Ullrich JE, Lane TF, Pallero MA, Sage EH (1995). SPARC mediates focal adhesion disassembly in endothelial cells through a follistatin-like region and the Ca^{2+}-binding EF-hand. *J Cell Biochem* 57:341-350.

Nei M, Rooney AP (2005). Concerted and birth-and-death evolution of multigene families. *Annu Rev Genet* 39:121-152.

Ott CM, Lingappa VR (2002). Integral membrane protein biosynthesis: why topology is hard to predict. *J Cell Sci* 115:2003-2009.

Papagerakis P, Berdal A, Mesbah M, Peuchmaur M, Malaval L, Nydegger J, Simmer J, Macdougall M (2002). Investigation of osteocalcin, osteonectin, and dentin sialophosphoprotein in developing human teeth. *Bone* 30:377-385.

Peyer B (1968). Comparative odontology. Chicago: The University of Chicago Press.

Poole DFG (1956). The structure of the teeth of some mammal-like reptiles. *Quart J Micr Sci* 97:303-312.

Poole DFG (1957). The formation and properties of organic matrix of reptilian tooth enamel. *Quart J Micr Sci* 98:349-367.

Pottgiesser J, Maurer P, Mayer U, Nischt R, Mann K, Timpl R, Krieg T, Engel J (1994). Changes in calcium and collagen IV binding caused by mutations in the EF hand and other domains of extracellular matrix protein BM-40 (SPARC, osteonectin). *J Mol Biol* 238:563-574.

Puolakkainen P, Bradshaw AD, Kyriakides TR, Reed M, Brekken R, Wight T, Bornstein P, Ratner B, Sage EH (2003). Compromised production of extracellular matrix in mice lacking secreted protein, acidic and rich in cysteine (SPARC) leads to a reduced foreign body reaction to implanted biomaterials. *Am J Pathol* 162:627-635.

Reichert T, Storkel S, Becker K, Fisher LW (1992). The role of osteonectin in human tooth development: an immunohistological study. *Calcif Tissue Int* 50:468-472.

Sage EH, Bassuk JA, Yost JC, Folkman MJ, Lane TF (1995). Inhibition of endothelial cell proliferation by SPARC is mediated through a Ca^{2+}-binding EF-hand sequence. *J Cell Biochem* 57:127-140.

Sage EH, Reed M, Funk SE, Truong T, Steadele M, Puolakkainen P, Maurice DH, Bassuk JA (2003). Cleavage of the matricellular protein SPARC by matrix metalloproteinase 3 produces polypeptides that influence angiogenesis. *J Biol Chem* 278:37849-37857.

Sahni A (1987). Evolutionary aspects of reptilian and mammalian enamel structure. *Scanning Microsc* 1:1903-1912.

Sasaki T, Gohring W, Mann K, Maurer P, Hohenester E, Knauper V, Murphy G, Timpl R (1997). Limited cleavage of extracellular matrix protein BM-40 by matrix metalloproteinases increases its affinity for collagens. *J Biol Chem* 272:9237-9243.

Sasaki T, Hohenester E, Gohring W, Timpl R (1998). Crystal structure and mapping by site-directed mutagenesis of the collagen-binding epitope of an activated form of BM-40/SPARC/osteonectin. *EMBO J* 17:1625-1634.

Schneider G, Fechner U (2004). Advances in the prediction of protein targeting signals. *Proteomics* 4:1571-1580.

Schwarzbauer JE, Spencer CS (1993). The Caenorhabditis elegans homologue of the extracellular calcium binding protein SPARC/osteonectin affects nematode body morphology and mobility. *Mol Biol Cell* 4:941-952.

Schwede T, Kopp J, Guex N, Peitsch MC (2003). SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res* 31:3381-3385.

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, Kamakura N, Toyosawa S, Ooshima T (2007). Identification and characterization of matrix metalloproteinase-20 (MMP20; enamelysin) genes in reptile and amphibian. *Gene* 392:89-97.

Sicher H (1962). In: Orbans"s oral histology and embryology. Saint Louis: Mosby, pp. 55-61.

Sire JY, Delgado S, Fromentin D, Girondot M (2005). Amelogenin: lessons from evolution. *Arch Oral Biol* 50:205-212.

Sire JY, Delgado SC, Girondot M (2008). Hen's teeth with enamel cap: from dream to impossibility. *BMC Evol Biol* 8:246.

Smreker E (1903). Ueber die Darstellung der Kittsubstanz des Schmelzes menschlicher Zaehne. *Anat Anzeiger* 22:467-476.

Swaroop A, Hogan BL, Francke U (1988). Molecular analysis of the cDNA for human SPARC/osteonectin/BM-40: sequence, expression, and localization of the gene to chromosome 5q31-q33. *Genomics* 2:37-47.

Takano-Yamamoto T, Takemura T, Kitamura Y, Nomura S (1994). Site-specific expression of mRNAs for osteonectin, osteocalcin, and osteopontin revealed by in situ hybridization in rat periodontal ligament during physiological tooth movement. *J Histochem Cytochem* 42:885-896.

Termine JD, Belcourt AB, Conn KM, Kleinman HK (1981a). Mineral and collagen-binding proteins of fetal calf bone. *J Biol Chem* 256:10403-10408.

Termine JD, Kleinman HK, Whitson SW, Conn KM, McGarvey ML, Martin GR (1981b). Osteonectin, a bone-specific protein linking mineral to collagen. *Cell* 26:99-105.

Tung PS, Domenicucci C, Wasi S, Sodek J (1985). Specific immunohistochemical localization of osteonectin and collagen types I and III in fetal and adult porcine dental tissues. *J Histochem Cytochem* 33:531-540.

Vymetal J, Slaby I, Spahr A, Vondrasek J, Lyngstadaas SP (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.

Walsh JB (1995). How often do duplicated genes evolve new functions? *Genetics* 139:421-428.

Weinreb PH, Zhen W, Poon AW, Conway KA, Landsbury PT (1996). NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded. *Biochemistry* 35:13709-13715.

Weiss KM (2005). The phenogenetic logic of life. *Nat Rev Genet* 6:36-45.

Yamakoshi Y, Hu JC, Fukae M, Iwata T, Kim JW, Zhang H, Simmer JP (2005). Porcine dentin sialoprotein is a proteoglycan with glycosaminoglycan chains containing chondroitin 6-sulfate. *J Biol Chem* 280:1552-1560.

Zhou Q, Wang W (2008). On the origin and evolution of new genes- a genomic and experimental perspective. *J Genet Genomics* 35:639-648.

CHAPTER 3

RECOMBINANT PORCINE AMELOBLASTIN

ABSTRACT

Ameloblastin is secreted primarily by secretory stage ameloblasts, and accounts for roughly 5 % of total enamel protein. Although its expression level is low relative to amelogenin, it is essential for physiological enamel formation. Since the ameloblastin cDNA was cloned (Cerny *et al.*, 1996; Hu *et al.*, 1997; Krebsbach *et al.*, 1996), knowledge about the gene product has not advanced much. A major obstacle to progress is that Ambn is rapidly degraded after its secretion into the enamel matrix. Full-length ameloblastin protein has never been isolated from *in vivo* sources. The production of recombinant ameloblastin is complicated because the protein contains several post-translational modifications, including at least two O-linked glycosylations. A major goal of my research was to express and purify recombinant Ambn for structural and functional studies.

The porcine Ambn cDNA encoding the 395 amino acid ameloblastin isoform was subcloned into pEF6/V5-His-TOPO and stably transfected into HEK293 cells. Relative expression levels of the secreted protein were determined by Western blotting. The highest expressing clone was selected for large-scale production. Secreted recombinant porcine ameloblastin (rpAmbn) was purified by ammonium sulfate precipitation, metal

affinity chromatography, and reversed phase-high performance liquid chromatography (RP-HPLC). Recombinant Ambn migrated at 64 kDa on SDS-PAGE and Western blots and proved to be the intact (uncleaved) ameloblastin glycoprotein.

INTRODUCTION

Ameloblastin

Ameloblastin (Ambn) cDNA was independently cloned by three groups (Cerny *et al.*, 1996; Hu *et al.*, 1997; Krebsbach *et al.*, 1996), which provided the deduced amino acid sequence of this secreted enamel protein. The Ambn primary structure is well conserved in mammals. Porcine Ambn shares 66 % identity with human and 65 % with mouse. The sequence similarity between reptile and mammalian Ambn decreases to 38-47 % (Shintani *et al.*, 2002). Between amphibian and mammalian ameloblastin sequence identities ranges between 29 to 37 % (Shintani *et al.*, 2003). The porcine ameloblastin cDNA encodes a signal peptide plus the 395 amino acid secreted protein that has a predicted molecular weight of 44 kDa. The apparent molecular weight of Ambn on SDS-PAGE is ~62 kDa (Uchida *et al.*, 1995), with the difference due to post-translational modifications (PTMs). Ambn is rich in proline (15 %), glycine (10 %), alanine (9 %), leucine (9 %) and has no cysteine.

Ambn null mice exhibit hypoplastic amelogenesis imperfecta (AI) (Fukumoto *et al.*, 2004). The enamel in the Ambn heterozygous mouse ($Ambn^{+/-}$) is nearly normal, suggestive of a recessive pattern of inheritance. No AI-causing *AMBN* mutations have yet been found in humans, although several non-synonymous polymorphisms have been

identified (Kim *et al.*, 2006). It is assumed that AI-causing mutations have not yet been identified in *AMBN* because its recessive pattern of inheritance makes it a rare disease. Based upon the *Ambn* null phenotype, Ambn is thought to facilitate enamel mineralization, and to potentially play a role in cell attachment and epigenetic signaling.

In the pig, N-terminal 13- and 15-kDa Ambn cleavage products concentrate in the sheath space (Hu *et al.*, 1997) (Fig. 3.2), producing a honeycomb staining pattern of the secretory stage enamel layer (Uchida *et al.*, 1991; Uchida *et al.*, 1995; Uchida *et al.*, 1997) (Fig. 3.1, Fig. 3.2).

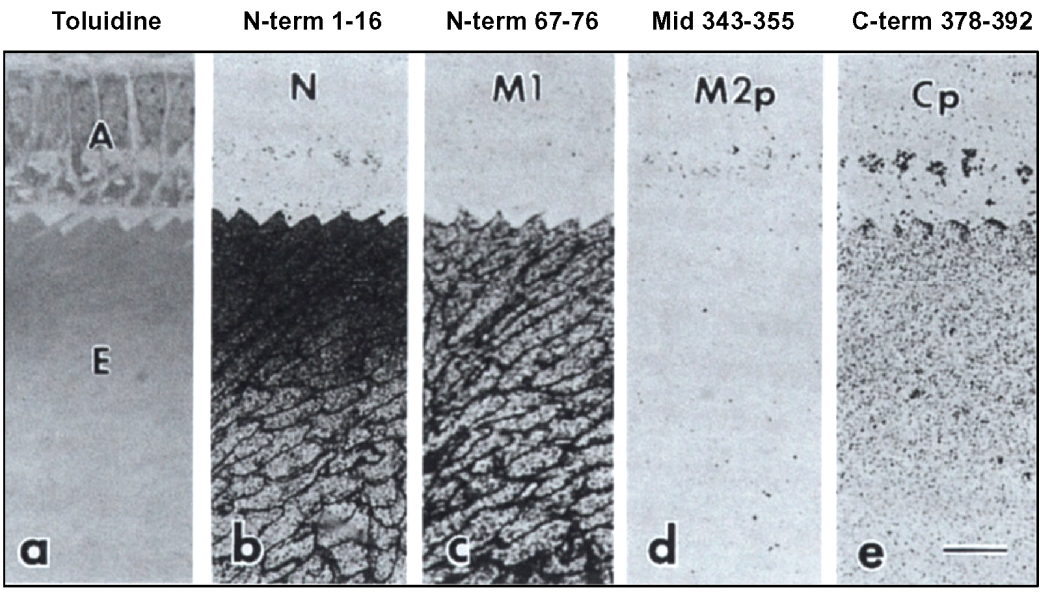

Figure 3.1: Cross-section of porcine molars for immunolocalization of Ambn cleavage products.
(a) toluidine stained, (b-e) anti-Ambn, (b) residues 1-16, (c) residues 67-76, (d) residues 343-355, (e) residues 378-392. Ambn N-terminal cleavage products concentrate in the sheath space, while intact Ambn localizes in the superficial rod and interrod space. A ameloblasts, E enamel layer. Bar represents 10 μm (Uchida *et al.*, 1998).

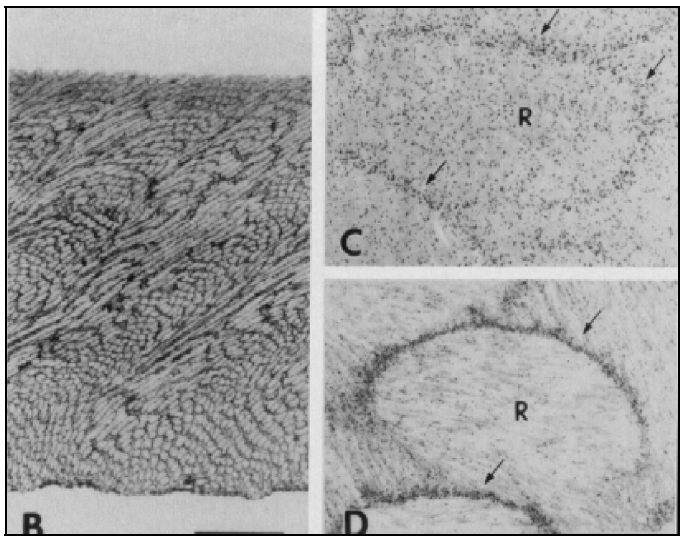

Figure 3.2: Localization of Ambn in the sheath space.
Immunolocalization, porcine molar, antibody against porcine antipeptide.
N-terminal Ambn is localized in to the sheath space resulting in a honeycomb pattern throughout the thickness of enamel (B). Ambn is diffusely localized in outer (C) enamel layers and concentrates to the sheath space in inner (D) enamel layers. Bar represents 2 μm. (Hu *et al.*, 1997).

Two alternatively spliced Ambn cDNA were identified in enamel organ epithelia (Cerny *et al.*, 1996; Hu *et al.*, 1997). The two mRNA differ by the inclusion or exclusion of a 45-nucleotide segment and by the use of alternative polyadenylation/cleavage sites. The secreted protein translated from the smaller alternatively spliced mRNA has 380 amino acids. The missing 15 amino acids correspond to residues Tyr^{73}-Gln^{87}, which are glycosylated (Kobayashi K *et al.*, 2007). Ser^{86} gave a blank cycle during N-terminal sequencing identifying it as the O-glycosylation site.

The C-terminal region of porcine Ambn is also glycosylated (at Thr^{361}). The oligosaccharides released from this site were N-acetylgalactosamine and galactose, indicating an O-linked glycosylation, which is sulphated and likely to bind calcium. The calcium–binding property of Ambn could be important to buffer the calcium ion

concentration in the extracellular enamel fluid, thereby binding or releasing calcium ions as needed for controlling crystal growth. Other PTM sites in porcine Ambn are Pro^{11} and Pro^{324}, which are hydroxylated, and Ser^{17} and Thr^{251}, which are phosphorylated. There are no putative N-linked glycosylation (NxS) or integrin binding (RGD) sites or signals for glycan attachment.

Ameloblastin cleavage products identified *in vivo*

Fragments of ameloblastin were first identified during protein extraction from unerupted pig teeth (Fukae and Tanabe, 1987a; Fukae and Tanabe, 1987b). Using pH based extraction protocols, these novel non-amelogenins had calcium-binding properties and were later identified as the Ambn C-terminal domain. Cleavage products were identified by using anti-peptide antibodies against the N-terminus, middle and C-terminus (Uchida *et al.*, 1995; Uchida *et al.*, 1997; Uchida *et al.*, 1998). Comprehensive isolation of enamel proteins has resulted in identification of many cleavage products of Ambn, as it is rapidly and extensively processed (Yamakoshi *et al.*, 2003; Yamakoshi *et al.*, 2006b). A total of 7 cleavage products have been isolated from porcine enamel (Fukae and Tanabe, 1987a; Fukae and Tanabe, 1987b; Yamakoshi *et al.*, 2006b). Cleavage of Ambn generates several lower molecular weight products (10 to 15 kDa) from the N- and C- terminus, and a larger middle piece of the protein (Fig. 3.3).

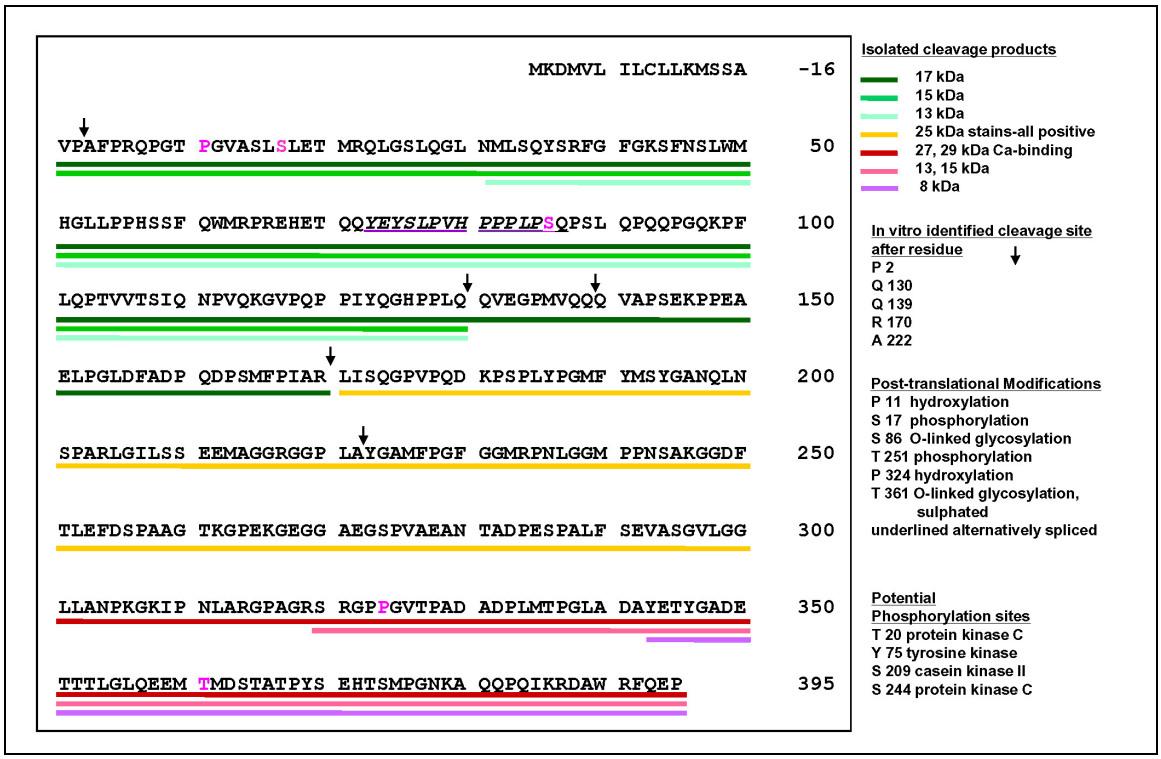

Figure 3.3: Ameloblastin structure from pig.
There are two porcine Ambn isoforms, which are expressed from alternatively spliced RNA transcripts. The most abundant isoform (shown above) has 395 amino acids. A smaller, 380 amino acid Ambn, is generated by deletion of the Tyr^{73}–Ser^{86} code during splicing. The identification of numerous cleavage products (lines in different colors) of Ambn in immature porcine enamel indicates several cleavage sites. Ambn has extensive post-translational modifications, including hydroxylations, O-glycosylations, and phosphorylations.

The initial Ambn cleavages are nearer to the N-terminus and generate smaller N-terminal (15, 17 kDa) and larger C-terminal Ambn polypeptides (40, 50 kDa). The C-terminal products are subsequently cleaved and cannot be detected except at the enamel surface, suggesting that reabsorption of the cleaved products from the Ambn C-terminus is very rapid (Brookes *et al.*, 2001). The localization and the behavior of the N- and C-terminal cleavage products are very distinct and their roles in enamel mineralization could be very specific.

The 27- and the 29-kDa calcium-binding proteins are derived from the C-terminus of Ambn and share the same N-terminus: at Leu^{301} (Murakami *et al.*, 1997). Anti-peptide antibodies specific for the Ambn C-terminal region immunostained only the enamel rods (not the sheath space) near the enamel surface, producing a reverse honeycomb pattern that did not extend into the deeper layers (Fig. 3.1). Immunopositive Golgi apparatus, secretory granules and large lysosomes suggest that Ambn is secreted, and the C-terminal region is reabsorbed into ameloblasts by endocytosis.

The N-terminal 17-kDa porcine Ambn has 170 amino acids, extending from Val^{1} to Arg^{170}. The 15-kDa porcine Ambn has 130 amino acids, extending from Val^{1} to Gln^{130} (Hu *et al.*, 1997). The 13-kDa protein contains 99 residues, extending from Met^{32} to Gln^{130}. The calculated isoelectric points of these N-terminal proteins are between 10 and 11, which is significantly higher than the parent protein (pI 6.6). The pI of the Ambn C-terminus, in contrast, is 4.5.

The ameloblastin protein has multiple post-translational modifications. Two O-linked glycosylations are present (on Thr^{361} and Ser^{86}). The Thr^{361} modification is sulphated and binds calcium (Kobayashi *et al.*, 2007; Yamakoshi *et al.*, 2001). Pro^{11} and Pro^{324} are hydroxylated. Ser^{17} and Thr^{251} are phosphorylated. There are no putative N-linked glycosylation or glycan attachment sites, or integrin binding (RGD) sites.

MATERIALS AND METHODS

Cloning of ameloblastin expression construct

Porcine enamel organ epithelium was obtained from 6-month-old unerupted molars. The mRNA was extracted using Trizol (Invitrogen) and reverse transcribed using Superscript III reverse transcriptase (Invitrogen). Porcine ameloblastin cDNA was amplified from the cDNA library with the oligonucleotide primers 5'-atcaccaggccctgagagca and 5'-gggctcttggaaacgccacg annealing at 65 ºC. The reaction was performed with AccuPrime Pfx SuperMix (Invitrogen) in 35 cycles with a final extension at 68 ºC for 30 min. The amplified sequence was confirmed by DNA sequencing. For TOPO TA cloning, A-overhangs were added at the 3' end of the amplified sequence using Taq polymerase and dATP at 72 ºC for 10 min. The product was electrophoresed on a 2 % agarose gel and band was cut out for DNA extraction. The DNA was extracted and purified by Wizard SV gel and PCR clean-up System (Promega, Madison, WI, USA).

The purified DNA was ligated into pEF6/V5-His-TOPO (Invitrogen, Corp., Carlsbad, CA, USA) and transformed into chemically competent *E. coli* (TOP10 One Shot, Invitrogen) at 42 ºC with 200 mM NaCl and 10 mM $MgCl_2$. The reaction was incubated on ice for 30 min. The cells were heat-shocked for 30 seconds at 42 ºC without shaking and immediately transferred on ice for 2 min. SOC medium (250 µl) was added and the transformation mixture was shaken at 37 ºC, 200 rpm for 1 hour. Twenty µl of the transformation was spread out on an LB plate and incubated at 37 ºC overnight. Colonies were selected from LB plates and amplified using T7 (forward) and Ambn (reverse)

primers for analysis of orientation. Colonies were expanded in LB broth (2 ml) and plasmid DNA was purified by QIAprep Miniprep (Qiagen, Valencia, CA, USA). Plasmid DNA was analyzed by sequencing using T7 (forward) and BGH (reverse) primers for orientation of the Ambn insert.

Production of recombinant porcine Ambn

The pEF6/V5-His-TOPO vector contains enhancer and promoter elements of the human elongation factor EF-1 α to initiate gene expression (Fig. 3.4). At the 3' end are encoded a V5 epitope and 6 histidine residues accounting for 4.86 kDa of the molecular weight of the expressed protein. Incorporation of the vector can be selected for using the blasticidin resistance gene.

For the expression of ameloblastin, HEK293-H cells (Human embryonic kidney cells, Invitrogen Corp., Gibco, Grand Island, NY) were chosen. HEK293-H strain is commonly used for protein expression. Their human and epithelial origin is closely related to pig ameloblasts and was assumed to express enzymes responsible for post-translational modifications. The cells were cultured in Dulbecco's modified Eagle's medium (DMEM) (Invitrogen Gibco) with 10 % (v/v) fetal bovine serum (FBS) (Invitrogen) in a 5 % CO_2 humidified chamber.

HEK293-H cells were transiently transfected using Lipofectamine 2000 (Invitrogen) and recombinant pig ameloblastin was detected in the medium as a secreted 440-amino-acid protein with a 45-amino-acid V5-epitope and poly-histidine tag at the C-terminus (KGNSADIQHSGGRSSLEGPRFEGKPIPNPLLGLDSTRTGHHHHHH). The

C-terminal histidine residues are exploited for purification by immobilized metal ion affinity chromatography (IMAC).

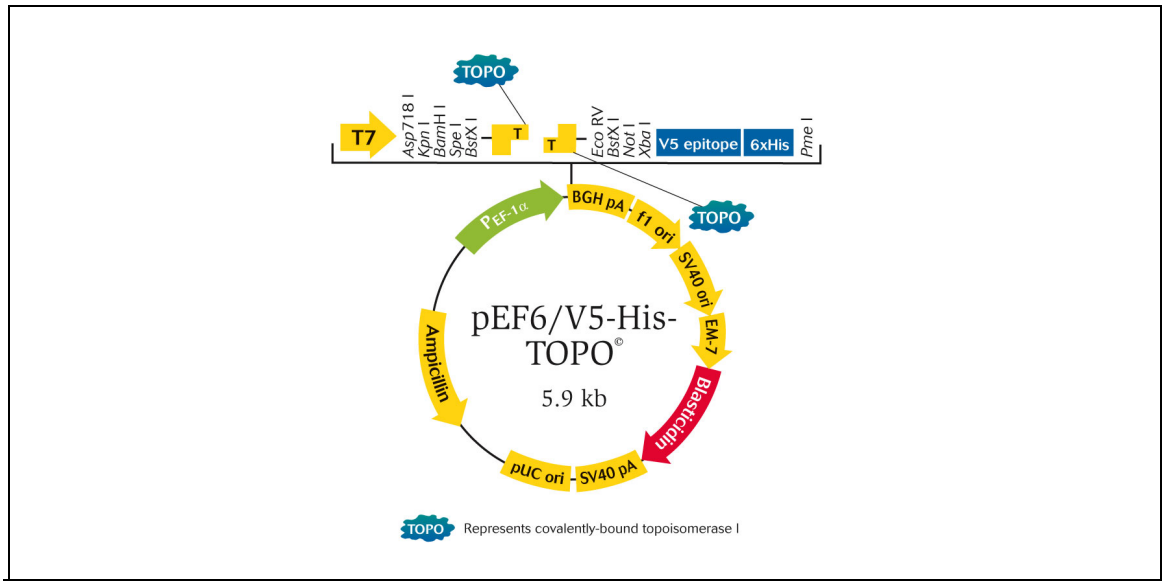

Figure 3.4: Expression vector pEF6/V5-His-TOPO for porcine ameloblastin in mammalian cells.
The pEF6/V5-His-TOPO (Invitrogen) uses the human elongation factor I alpha promoter to express proteins. The coding region is inserted between the TOPO sites. The C-terminus encodes a V5 epitope and six histidine residues. Stable mammalian integrants are selected by exploiting the blasticidin resistance gene.

Transient transfection of the HEK293-H cells with Lipofectamine 2000 (Invitrogen) resulted in detection of a V5 positive signal on Western blots using the V5-HRP antibody (Invitrogen) (Fig. 3.5 A).

The goal was to create a stable cell line overexpressing Ambn. Stably transfected cells carrying the blasticidin resistance gene were selected from single cell clones using 0.3 µg/ml blasticidin S HCl (Invitrogen). Fifty-four clones were expanded for screening of Ambn expression level by Western blot analysis using the V5 antibody. The highest expressing clone number 21A (Fig. 3.5 B) was selected for expansion in large-scale cell

culture. Western blots against the N-terminus and C-terminus of Ambn confirmed the presence of ameloblastin.

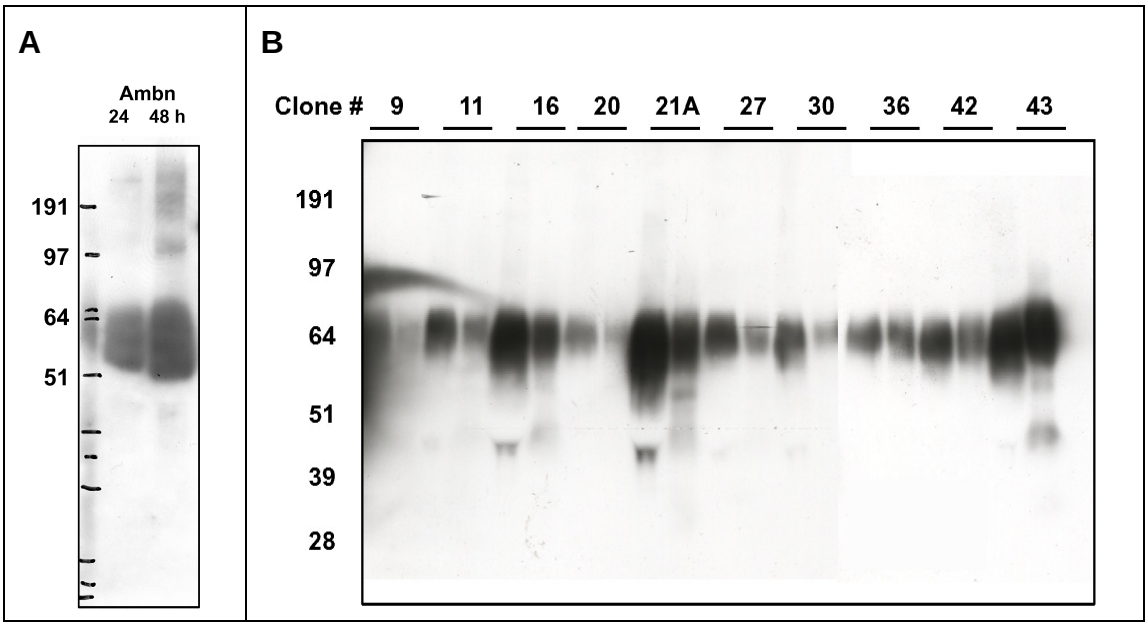

Figure 3.5: Transfection of HEK293 with porcine ameloblastin results in expression of recombinant protein.
A) Western blot analysis of medium from HEK293 cells that were transiently transfected with Ambn-His. The antibody recognizes the V5 epitope at the C-terminus of recombinant Ambn-His.
B) Screening of HEK293 single cell clones for highest Ambn expression by Western blot (anti-V5 antibody).

Large-scale cell culture

The large-scale cell culture of Ambn overexpressing cells was performed as monolayer culture in two 10-level CellStacks and one 1-level CellStack (Corning, Corning, NY) at a cell density of 7.8 x 10^4 cells/cm^2 (Fig. 3.6). Cells were cultured in DMEM supplemented with 10 % FBS at 37 ºC and 5 % CO_2 for 2 days to reach confluency. On day 3 the medium was switched to DMEM supplemented with 1 % SerumPlus (SAFCO Biosciences, Lenexa, Kansas, USA). The medium of 525 ml volume

was collected every 24 h. The total cell culture surface area equaled 13,356 cm^2, corresponding to 178 T75 flasks. Cells were passaged weekly.

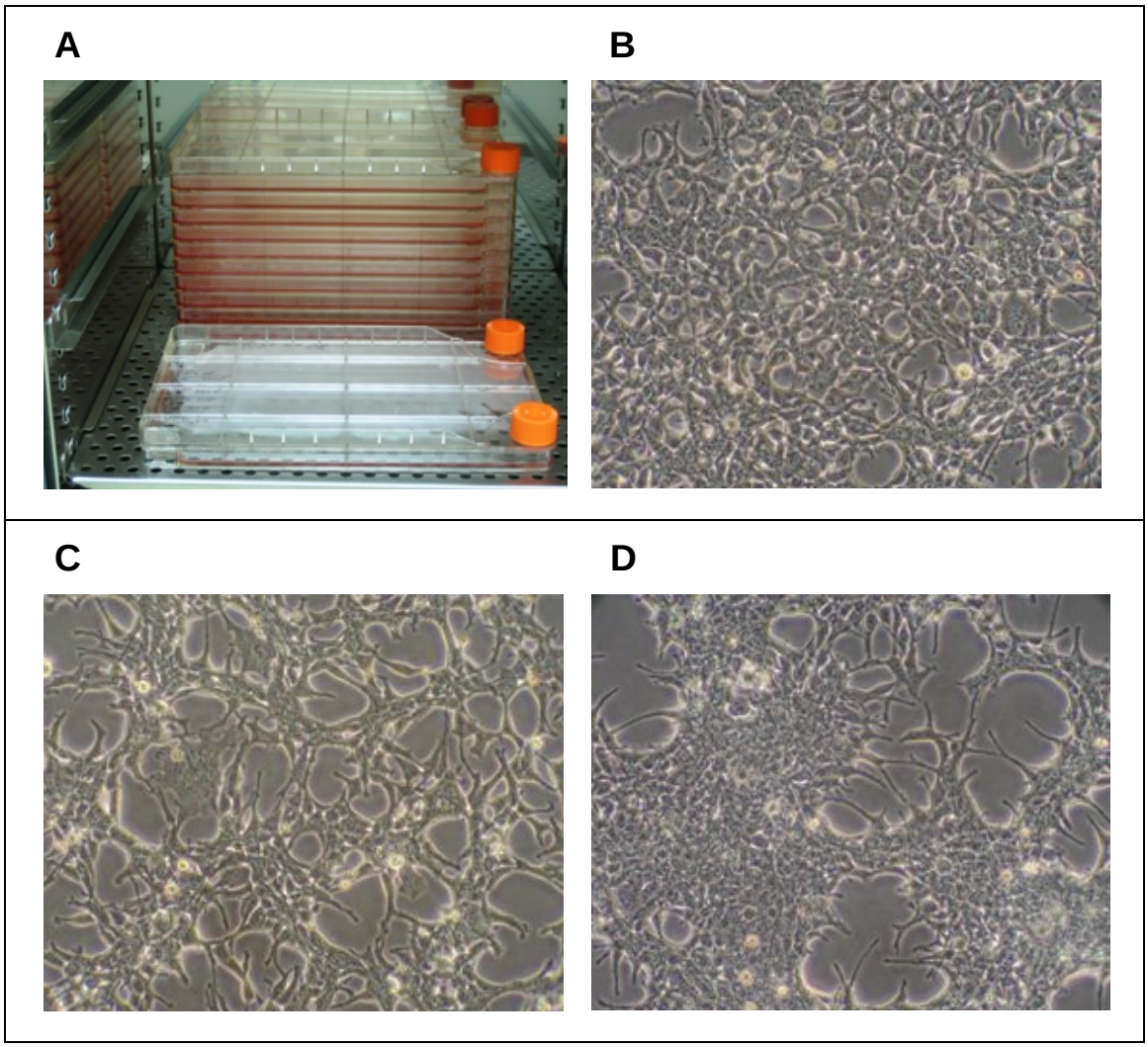

Figure 3.6: Large-scale production of recombinant ameloblastin.
A) Recombinant Ambn was produced by stably transfected HEK293 cells (Line 21A) in monolayer. To increase the cell number, cells were grown in cell-stacks.
B) Light microscopy of cells at confluency, day 3. Cells were supplied with DMEM and 10 % FBS.
C) Cell culture at day 4. After reaching confluency, the cells were switched to DMEM and 1 % SerumPlus. 24 h after switching the medium the cells reduced their size and rate of proliferation. Recombinant Ambn secreted into the medium was collected every 24 h for further purification.
D) At day 8, 5 days after switching the medium to 1 % SerumPlus cells were attached and were growing in a monolayer.

Purification of Ambn-His

Non-specific proteinases inhibitors (EDTA 4 mM pH 7.0, PMSF 0.5 mM) were added immediately after collection of the medium after 24 hours. Ambn-His was precipitated with ammonium sulfate at 50 % saturation overnight on ice. The precipitate was centrifuged at 8,000 rpm and resuspended in sodium phosphate buffer (50 mM sodium phosphate, 300 mM NaCl, pH 7.4). Next, Ambn-His was extracted by IMAC (Talon Metal Affinity Resin, Clontech, Mountain View, CA, USA). After 20 min binding time the supernatant was applied to new beads to bind unbound Ambn-His. The charged beads were washed with sodium phosphate buffer and stored at 4 ºC. The Ambn-His bound to the beads was accumulated over 4-5 days and eluted with 200 mM imidazole and sodium phosphate buffer. The elutant was desalted by using a centrifugal filter device with a 10-kDa molecular size cutoff (Amicon Ultra -15, Millipore, Billerica, MA, USA) and lyophilized. The protein was resuspended in 0.05 % trifluoroacetic acid (TFA) and loaded onto reversed phase high performance liquid chromatography (RP-HPLC) (HPLC; Beckman Coulter, System Gold, Fullerton, CA, USA) using a C18 column, 25 cm x 10 mm x 5 um (Discovery C18, Supelco, Bellefonte, PA) at 230 nm. As starting buffer 0.05 % TFA was used. 0.05 % TFA + 80 % acetonitrile was used as elution buffer with a 2 % gradient from 0-100 %. The flow rate was 1.0 ml/min. The peaks on the chromatogram were monitored at 230 nm and separately collected. After lyophilization the collected fractions were analyzed by SDS-PAGE (NuPAGE 4-12 % Bis-Tris, MES or MOPS buffer, Invitrogen) followed by the universal protein stains coomassie brilliant blue (CBB) (SimplyBlue, Safestain, Invitrogen Corp., Carlsbad, CA) and silver (ProteoSilver Silver Stain Kit, Sigma-Aldrich, St. Louis, MO). Stains-all staining was performed to stain calcium–binding proteins. Western blot analysis was performed for

specific detection of proteins. For detection of recombinant ameloblastin and specific regions of the protein, four antipeptide antibodies and one polyclonal antibody were used (Tab. 3.1). The apparent molecular weights of protein bands were estimated by comparison with SeeBlue®Plus2 Pre-Stained Standard (Invitrogen).

SDS-PAGE electrophoresis

Ameloblastin SDS-PAGE samples were prepared with Sample Buffer (Bio-Rad Laboratories, Hercules, CA, USA) in a 1:1 ratio and run on 4-12 % Bis-Tris Nu-PAGE gel (Invitrogen) with MES or MOPS buffer (Invitrogen). Electrophoresis was performed at a constant current of 200 V for 38-42 min.

Coomassie Brilliant Blue staining

After SDS-PAGE, the gel was washed in water three times for 10 min. The gel was stained with Coomassie brilliant blue (CBB) (SimplyBlue Safestain, Invitrogen) for 1 h. The gel was de-stained with water to increase the clarity of the bands and to reduce the background staining.

Silver staining

The silver staining (ProteoSilver Silver Stain Kit, Sigma-Aldrich, St. Louis, MO) was performed to analyze proteins bands that are not stained by CBB and to visualize proteins of lower quantity. The gel was first fixed in 50 % ethanol and 10 % acetic acid

for 20 min and then washed in 30 % ethanol. After one wash with water, the gel was sensitized with ProteoSilver sensitizer for 10 min. The gel was washed and then equilibrated in silver solution. After a brief washing step, the gel was developed until bands became visible. The development was stopped, and the gel washed with water.

Stains-all staining

Cationic Stains-all staining stains proteins differentially. Highly acidic proteins and calcium-binding proteins stain blue, proteoglycans stain purple, and less acidic proteins stain purple. To remove SDS after electrophoresis, the gel is washed in 25 % methanol for 30 min 3 times. The gel is soaked in Stains-all solution (30 mM Tris-HCl, pH 8.8, 7.5 % formamide (v/v), 25 % methanol (v/v), 0.025 % (w/v) Stains-all. Due to the photosensitivity of Stains-all gels, gels were incubated in containers covered with aluminum foil on an orbital shaker for overnight.

Zymography

The zymogram was performed using a 12 % SDS-acrylamide gel containing 0.1 % casein as substrate. The protein sample was prepared with sample buffer (Laemmli Sample Buffer, Bio-Rad Laboratories, Hercules, CA, USA) in a 1:1 ratio and run at constant 125 V under simultaneous cooling using Tris-Glycine SDS running buffer.

After electrophoresis, the gel was washed in 2.5 % Triton X for 30 min twice and incubated in developing buffer (50 mM Tris-HCl pH 7.4; 10 mM $CaCl_2$) for 24 h at 37 ºC. The gel was stained with CBB R-250 solution (0.23 % CBB-250, 10 % acetic acid, 50

% methanol) for 30 min and destained with 7 % acetic acid 10 % methanol until clear bands appeared.

Western blot analysis and antibodies

After gel electrophoresis the proteins were transblotted onto a PVDF membrane (Hybond-P, GE Healthcare, Buckinghamshire, UK) for Western blot analysis and blocked with 5 % nonfat milk (BioRad, Hercules, CA, USA) dissolved in TBS-tween (20 mM Tris, 500 mM sodium chloride, pH 7.5 (TBS) with tween 20, BioRad, Hercules, CA, USA).

The antibodies used were raised against specific N- and C-terminal aspects of Ambn to obtain information of the molecular weight and the integrity of recombinant Ambn. The 5 different antibodies included two polyclonal anti-peptide antibodies against the N-terminal region: (1) Ambn-63 (Iwata *et al.*, 2007) and (2) Ambn-73. Ambn-73 recognized an alternatively spliced region of 14 residues. The C-terminal region was recognized by an anti-peptide antibody raised against the C-terminal region (3) Ambn-381 (Iwata *et al.*, 2007), (4) one polyclonal anti-center, and (5) anti-V5 antibody recognizing a C-terminal epitope outside the ameloblastin coding region (Tab. 3.1).

Blots were immersed with chemiluminescence (ECL Plus, GE Healthcare) for 5 min and exposed to a film (Super RX, Fujifilm, Tokyo, Japan). Western blots were washed in TBS-tween and incubated with anti-rabbit IgG secondary antibody (Pierce, Rockford, IL, USA). Immunostaining was detected by chemiluminescence (ECL Plus) exposed to a film (Super RX, Fujifilm).

Name		Sequence	Condition
Ambn-63	Polyclonal	63-QWMRPREHETQQYEYS-76	1:2000
Ambn-73	Polyclonal	73-YEYSLPVHPPPLP-85	1:2000
Ambn-center	Polyclonal	131-QVEGPMVQQQVAPSEKPPEAELPGLDFADPQDPS MFPIARLISQGPVPQDKPSPLYPGMFYMSYGANQLNSP ARLGILSSEEMAGGRGGPLAYGAMFPGFGGMRPNLGG MPPNSAKGGDFTLEFDSPAAGTKGPEKGEGGAEGSPV AEANTADPESPALFSEVASGVL-298	1:10000
Ambn-381	Polyclonal	381-QQPQIKRDAWRFQE-394	1:5000
V5-HRP	Monoclonal	GLPIPNPLLGLDST	1:10000

Table 3.1: Antibodies used in this study to detect recombinant porcine ameloblastin. The antibodies recognize specific regions of Ambn and serves as a tool to characterize and distinguish different cleavage products.

Characterization of PTM of rpAmbn

Porcine Ambn is known to have O-linked glycosylations at Ser^{86} and Thr^{361}. To determine if rpAmbn has O-linked glycosylations, it was digested with O-glycosidase and sialidase (QA-Bio, San Mateo, CA, USA) to remove the carbohydrate chain. As O-glycosidase only cleaves unsubstituted Gal-β(1-3)GalNAc-alpha chains attached to Ser or Thr, the sialic acid was first removed. rpAmbn was incubated with sialidase and O-glycosidase at 37 ºC for 2 h without additional protease inhibitors. The digest was analyzed on SDS-PAGE stained with CBB, silver, Stains-all. Western blot analysis was performed using antibodies recognizing the N- or C-terminal region of Ambn.

Glycoprotein staining

Pro-Q Emerald 488 (Molecular Probes, Eugene, OR, USA) was used to stain glycoproteins. After SDS-PAGE gel were fixed in 50 % methanol and 5 % glacial acetic acid for 1 hour on a shaker. The gel was washed in 3 % glacial acetic acid twice for 20 minutes and oxidized in oxidizing solution containing periodic acid and 3 % acetic acid. After washing, the gel was stained with Pro-Q Emerald 488 with gentle agitation for 2 hours. The gel is washed in 3 % glacial acetic acid and proteins were visualized at 470-500 nm.

RESULTS

For the characterization of full-length ameloblastin with appropriate post-translational modifications, the protein must be expressed in mammalian cells as a recombinant protein. Full-length ameloblastin is not available from developing teeth, since it is rapidly processed by proteases. My goal was to obtain full-length ameloblastin protein in sufficient quantity for structural experiments. Because Ambn has extensive post-translational modifications, a mammalian cell line was chosen for recombinant protein expression. The post-translational modifications of recombinant Ambn were characterized and compared with the ones known from *in vivo* porcine Ambn cleavage products.

For large-scale expression of porcine ameloblastin, the ameloblastin coding region was cloned and transfected stably into HEK293 cells. As the protein yield is

proportional to the cell number, an expression system that could hold the largest cell number was desirable.

Optimization of cell culture system

As a promising culture system appeared to be the hollow fiber cell culture system cartridge (FiberCell Systems, Frederick, MD, USA), in which cells are harbored between two membranes (Fig. 3.7 A). This system used 2 L of serum-free medium (HyQ SFM4HEK293, HyClone, Logan, UT, USA) per day circulating through the cartridge to supply the cells with nutrients and to remove waste products. The secreted protein was removed daily through lateral ports of the cartridge in a 50 ml volume. Western blot analysis using anti-peptide antibodies raised against the N-terminal and C-terminal aspects of Ambn revealed that the recombinant protein was degraded rapidly and progressively over time (Fig. 3.7 B). To characterize the enzymes active in the expression system, zymograms were performed. Casein zymogram (12 % Tris Glycine, 0.05 % casein, Invitrogen) was negative, ruling out enzymes with collagenase like activity. Gelatin zymogram (10 % Tris Glycine, 0.1 % gelatin, Invitrogen) and Western blot analysis revealed MMP-2 with a molecular weight of 62 kDa in its active form and 72 kDa in its latent form (Fig. 3.7 B) as the major proteinase produced endogenously by HEK293-H cells. At lower levels, MMP-1, -3, -8, -9, 10, -13, MT1-MMP, MT3-MMP and TIMP1, -2, and -3 are expressed (Liu and Wu, 2006).

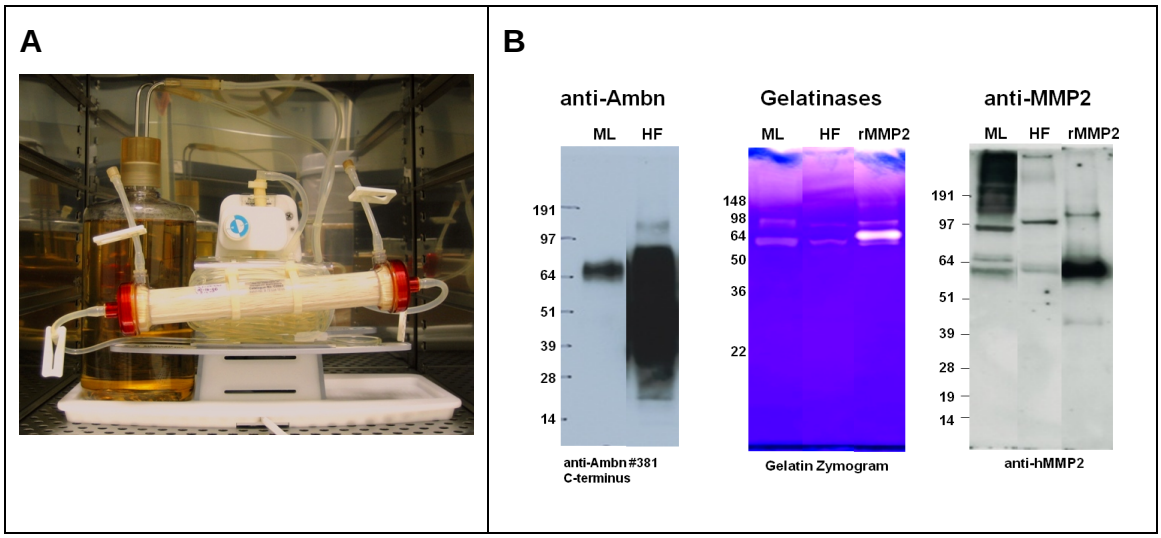

Figure 3.7: HEK293 culture in the hollow fiber cartridge.
A) The hollow fiber cartridge is attached to the pump and placed inside the incubator together with the medium reservoir. The cells are located inside the hollow cartridge between the red filters. The medium continuously travels through the cartridge to deliver nutrients and to remove waste products. The culturing conditions for the cells are 37 ºC, 8 % CO_2, humidified atmosphere and serum-free medium.
B) Analysis of the cartridge medium containing secreted proteins. Ameloblastin Western blot analysis, gelatin zymogram and MMP-2 Western blot analyses were performed. Recombinant Ambn is rapidly degraded in the hollow fiber culture, but stable in a monolayer culture. In the hollow fiber cartridge, Mmp-2 is present in its activated form. ML monolayer, HF hollow fiber.

In contrast to the three-dimensional cell culture, the recombinant ameloblastin protein maintained its full-length from monolayer culture due to lower proteinase expression. Therefore, the large-scale production was switched to a monolayer culture system requiring weekly passaging of the cells (Fig. 3.6). Until cells reached confluency, they were cultured with 10 % FBS, but during protein collection, the media was changed to 1 % SerumPlus, to reduce the load of enzymes and contaminating proteins.

Purification of recombinant porcine ameloblastin

Recombinant porcine ameloblastin was expressed and isolated from monolayer culture. After ammonium precipitation and IMAC, a 64-kDa protein was purified that maintained the molecular weight as indicated by Western blot analysis using N- and C-terminal antibodies. CBB staining and silver staining revealed impurities in the concentrated fraction (Fig. 3.8, lane 11).

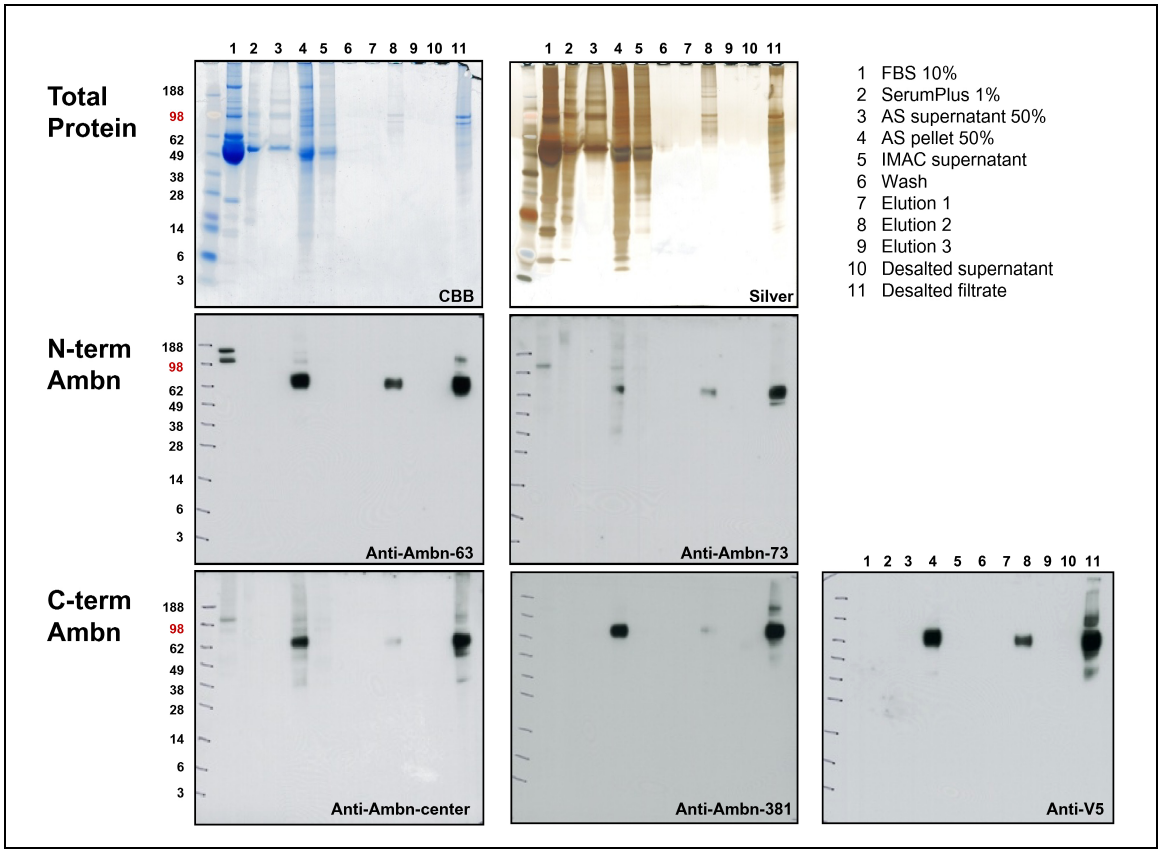

Figure 3.8: Analysis of the expression and purification of recombinant porcine ameloblastin by SDS-PAGE.
The expression and purification by ammonium sulfate precipitation and IMAC were evaluated by SDS-PAGE. In gels stained for total protein by CBB and silver staining, the quality of purification is analyzed. The load of contaminating proteins is reduced from the starting material (lane 2) to the desalted filtrate (lane 11). Western blot using antibodies specific for Ambn (anti-63, -73, -381, -center) and the V5 epitope demonstrate the size and integrity of the recombinant protein.

RP-HPLC was performed on the concentrated fraction using a C18 column (Fig. 3.9). The peak 2 corresponded to single bands on CBB and silver staining indicating purity of >95 %. Western blot analysis using the N- and C-terminal Ambn antibodies yielded an immunopositive signal at 64 kDa. In addition, dimers at ~120 kDa and trimers at ~180 kDa were noted on Western blots. Since the protein sequence does not encode cysteine residues, these interactions could be facilitated through non-covalent interactions even under denaturing conditions of SDS-PAGE. From a total of 2.5 L collected in a week, 0.25 to 0.5 mg rpAmbn was purified, corresponding to 100-200 µg/L medium.

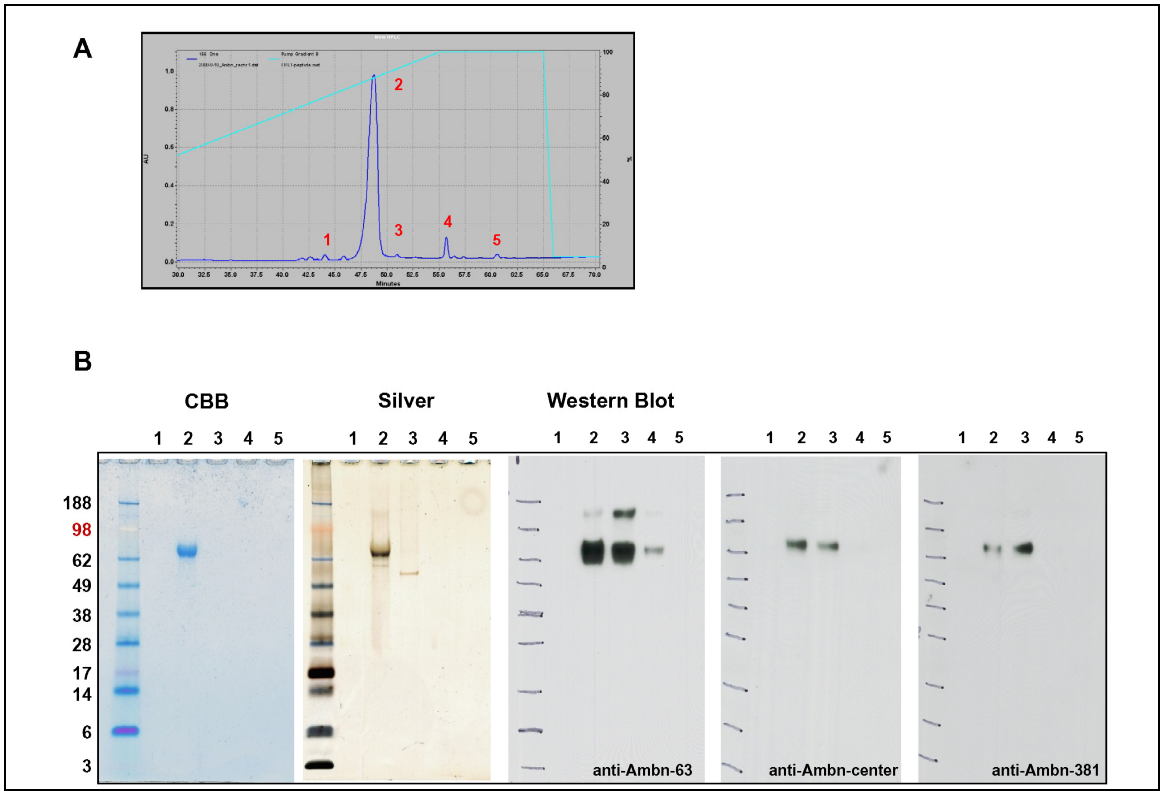

Figure 3.9: Final purification of recombinant porcine ameloblastin.
RP-HPLC using a C18 column was performed to purify Ambn. Proteins were eluted with acetonitrile containing 0.05 % TFA as buffer at a 2 % gradient, 1.0 ml/min flow rate, and monitored at 230 nm. The chromatogram revealed 5 peaks that were collected and analyzed by CBB, silver staining, and Western blot analysis. The peaks 1-5 in the RP-HPLC chromatogram correspond to lanes 1-5 on SDS-PAGE. Fraction 2 was collected from a single peak and produced a single band of 64 kDa on CBB stained gels and Western blots.

Characterization of rpAmbn PTMs

The double digestion of rpAmbn by sialidase and O-glycosidase was analyzed by SDS-PAGE, in-gel stains, and Western blots on transblotted membranes (Fig. 3.10). All gels and membranes demonstrated a shift of the starting material from above 64 kDa to 64 kDa after digestion. The signal was reduced in intensity after digestion. The loss of the Stains-all positive band after digestion shows that calcium-binding property is linked to O-linked glycosylation. The Pro-Q Emerald 488 staining detected a glycoprotein, whose molecular weight and signal intensity were decreased after digestion. Western blots using N- and C-terminal antibodies demonstrated that the rpAmbn was full-length after deglycosylation. Together, rpAmbn is O-linked glycosylated and binds calcium (Fig. 3.10).

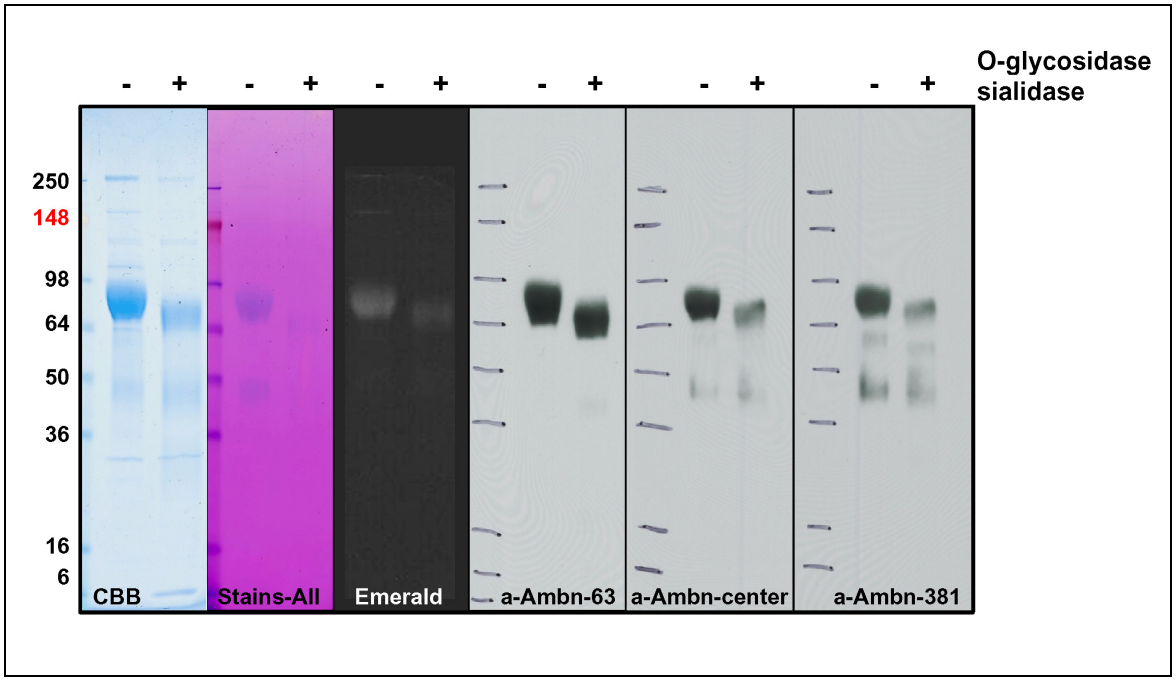

Figure 3.10: Deglycosylation of rpAmbn by O-glycosidase and sialidase.
After SDS-PAGE (4-20 % Tris-Glycine gel) CBB. Stains-all and Emerald stained gels and Western blots display a shift in band size and decreased in signal intensity. Lanes marked with + or – indicate presence or absence of O-glycosidase and sialidase in the reaction.

DISCUSSION

Intact Ambn has never been isolated from developing teeth because proteases are co-secreted with enamel proteins and it is expressed in low quantity. Therefore, knowledge on the structure and function of full-length Ambn is scarce. Ambn is a secreted protein that contains several post-translational modifications. The post-translational modifications are O-linked glycosylations at the N-terminal Ser^{86} and the C-terminal Thr^{361}. Thr^{361} is further sulphated and binds calcium (Yamakoshi *et al.*, 2001). Phosphorylations are present on Ser^{17} and Thr^{251}. Pro^{11} and Pro^{324} are hydroxylated. In total 7 cleavage sites have been identified in Ambn *in vivo*.

In the past, several groups have attempted to produce recombinant ameloblastin using different strategies. However, most of these recombinant proteins lacked PTMs due to use of a non-mammalian expression system (Lee *et al.*, 1996; Nakamura *et al.*, 2006; Ravindranath *et al.*, 2004; Shintani *et al.*, 2006) or they lacked sufficient yields for *in vitro* experiments (Iwata *et al.*, 2007; Zeichner-David *et al.*, 2006). Rapid degradation was a common complication noted by several groups. Therefore, it was necessary to revisit the production of recombinant ameloblastin to address concerns of protein quality and quantity.

In this study, full-length porcine Ambn was expressed in a eukaryotic host and was O-glycosylated, as is true of the protein in enamel PTMs. The recombinant Ambn produced in this study had an apparent molecular weight of about 64 to 66 kDa including a histidine-tag and V5 site, which added 4.86 kDa to the Ambn C-terminus. The apparent weight of the recombinant protein corresponds to the native protein, which has been describes as a 62-kDa band from immature pig enamel (Uchida *et al.*, 1991). PTMs on

the recombinant protein were immediately evident, as its apparent molecular weight on SDS-PAGE greatly exceeded the deduced molecular weight of its primary amino acid sequence. Glycosylations were demonstrated by Pro-Q Emerald 488 staining and deglycosylation of the O-linked sugar chains. Calcium-binding property was demonstrated by Stains-all positive signal. Phosphorylations and hydroxylations were not determined in the recombinant protein.

The quality of the recombinant porcine ameloblastin resembles closely the native protein. The yield of the proteins is low but sufficient for structural studies.

REFERENCES

Brookes SJ, Kirkham J, Shore RC, Wood SR, Slaby I, Robinson C (2001). Amelin extracellular processing and aggregation during rat incisor amelogenesis. *Arch Oral Biol* 46:201-208.

Cerny R, Slaby I, Hammarstrom 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.

Fukae M, Tanabe T (1987a). Nonamelogenin components of porcine enamel in the protein fraction free from the enamel crystals. *Calcif Tissue Int* 40:286-293.

Fukae M, Tanabe T (1987b). ^{45}Ca-labeled proteins found in porcine developing dental enamel at an early stage of development. *Adv Dent Res* 1:261-266.

Fukumoto S, Kiba T, Hall B, Iehara N, Nakamura T, Longenecker G, Krebsbach PH, Nanci A, Kulkarni AB, Yamada Y (2004). Ameloblastin is a cell adhesion molecule required for maintaining the differentiation state of ameloblasts. *J Cell Biol* 167:973-983.

Hu CC, Fukae M, Uchida T, Qian Q, Zhang CH, Ryu OH, Tanabe T, Yamakoshi Y, Murakami C, Dohi N, Shimizu M, Simmer JP (1997). Sheathlin: cloning, cDNA/polypeptide sequences, and immunolocalization of porcine enamel sheath proteins. *J Dent Res* 76:648-657.

Iwata T, Yamakoshi Y, Hu JC, Ishikawa I, Bartlett JD, Krebsbach PH, Simmer JP (2007). Processing of ameloblastin by MMP-20. *J Dent Res* 86:153-157.

Kim JW, Simmer JP, Lin BP, Seymen F, Bartlett JD, Hu JC (2006). Mutational analysis of candidate genes in 24 amelogenesis imperfecta families. *Eur J Oral Sci* 114 (Suppl 1) 3-12; discussion 39-41, 379.

Kobayashi K, Yamakoshi Y, Hu JC, Gomi K, Arai T, Fukae M, Krebsbach PH, Simmer JP (2007). Splicing determines the glycosylation state of ameloblastin. *J Dent Res* 86:962-967.

Krebsbach PH, Lee SK, Matsuki Y, Kozak CA, Yamada KM, Yamada Y (1996). Full-length sequence, localization, and chromosomal mapping of ameloblastin. A novel tooth-specific gene. *J Biol Chem* 271:4431-4435.

Liu CH, Wu PS (2006). Characterization of matrix metalloproteinase expressed by human embryonic kidney cells. *Biotechnol Lett* 28:1725-1730.

Murakami C, Dohi N, Fukae M, Tanabe T, Yamakoshi Y, Wakida K, Satoda T, Takahashi O, Shimizu M, Ryu OH, Simmer JP, Uchida T (1997). Immunochemical and

immunohistochemical study of the 27- and 29-kDa calcium-binding proteins and related proteins in the porcine tooth germ. *Histochem Cell Biol* 107:485-494.

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.

Uchida T, Tanabe T, Fukae M, Shimizu M, Yamada M, Miake K, Kobayashi S (1991). Immunochemical and immunohistochemical studies, using antisera against porcine 25 kDa amelogenin, 89 kDa enamelin and the 13-17 kDa nonamelogenins, on immature enamel of the pig and rat. *Histochemistry* 96:129-138.

Uchida T, Fukae M, Tanabe T, Yamakoshi Y, Satoda T, Murakami C, Takahashi O, Shimizu M (1995). Immunological 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'. *Biomedical Research* 16:131-140.

Uchida T, Murakami C, Dohi N, Wakida K, Satoda T, Takahashi O (1997). Synthesis, secretion, degradation, and fate of ameloblastin during the matrix formation stage of the rat incisor as shown by immunocytochemistry and immunochemistry using region-specific antibodies. *J Histochem Cytochem* 45:1329-1340.

Uchida T, Murakami C, Wakida K, Dohi N, Iwai Y, Simmer JP, Fukae M, Satoda T, Takahashi O (1998). Sheath proteins: synthesis, secretion, degradation and fate in forming enamel. *Eur J Oral Sci* 106 (Suppl 1) 308-314.

Yamakoshi Y, Tanabe T, Oida S, Hu CC, Simmer JP, 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.

Yamakoshi Y, Hu JC-C, Ryu OH, Tanabe T, Oida S, Fukae M, Simmer JP (2003). A comprehensive strategy for purifying pig enamel proteins. *Biomineralization (BIOMIN2001): formation, diversity, evolution and application.* 326-332.

Yamakoshi Y, Hu JC, Zhang H, Iwata T, Yamakoshi F, Simmer JP (2006). Proteomic analysis of enamel matrix using a two-dimensional protein fractionation system. *Eur J Oral Sci* 114 (Suppl 1) 266-271; discussion 285-286, 382.

Zeichner-David M, Chen LS, Hsu Z, Reyna J, Caton J, Bringas P (2006). Amelogenin and ameloblastin show growth-factor like activity in periodontal ligament cells. *Eur J Oral Sci* 114 (Suppl 1) 244-253; discussion 254-6, 381-382.