Universidade de Aveiro Departamento de Química
2012

PATRÍCIA DE SOUSA PEREIRA

CARACTERIZAÇÃO DOS PÉPTIDOS SALIVARES DE MAMÍFEROS

CHARACTERIZATION OF MAMMAL SALIVARY PEPTIDES

Universidade de Aveiro Departamento de Química
2012

PATRÍCIA DE SOUSA PEREIRA

CARACTERIZAÇÃO DOS PÉPTIDOS SALIVARES DE MAMÍFEROS

CHARACTERIZATION OF MAMMAL SALIVARY PEPTIDES

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Bioquímica, especialização em Métodos Biomoleculares, realizada sob a orientação científica do Doutor Rui Vitorino, Investigador Auxiliar do Departamento de Química da Universidade de Aveiro e do Doutor Pedro Esteves, Investigador principal do CIBIO (Centro de Investigação em Biodiversidade e Recursos Genéticos), Universidade do Porto e professor coordenador do IPSN/CESPU

Dedico este trabalho aos meus pais.

o júri

presidente

Prof. Dr. Pedro Miguel Dimas Neves Domingues
professor auxiliar do Departamento de Química da Universidade de Aveiro

Dra. Joana Patrícia da Silva Abrantes
investigadora do CIBIO da Universidade do Porto

Dr. Rui Miguel Pinheiro Vitorino
investigador auxiliar do Departamento de Química da Universidade de Aveiro

Dr. Pedro José de Castro Esteves
investigador do CIBIO da Universidade do Porto e professor coordenador do IPSN/CESPU

agradecimentos

Para a concretização desta tese tive a ajuda e o apoio de várias pessoas às quais quero deixar o meu sincero agradecimento.

Em primeiro lugar, obrigado aos meus pais por todos os esforços feitos e pelo apoio e afeto com que sempre me acompanharam ao longo da vida.

Obrigado aos meus orientadores Doutor Rui Vitorino e Doutor Pedro Esteves pela orientação científica, pela disponibilidade, pela paciência e acima de tudo obrigado pelo incentivo que me fez querer chegar mais longe.

Um obrigado também muito especial à Doutora Rita Ferreira, pelos ensinamentos, pelo apoio e por ser uma verdadeira inspiração para mim. Agradeço à Joana por me ajudar a entrar no mundo da filogenia.

Agradeço também ao Doutor Bruno Colaço e ao Doutor Pedro Gomes pelas amostras disponibilizadas, sem as quais este trabalho não poderia ser feito.

Obrigado a todo o grupo de espectrometria de massa da Universidade de Aveiro pelo acolhimento e simpatia.

Por fim, obrigado a todos os meus colegas pelos momentos de descontração e obrigado aos meus amigos que me ajudaram em cada passo desta caminhada.

palavras-chave

Saliva, peptidos antimicrobianos, cistatinas, histatinas, proteinas-ricas em prolina, estaterina, timosinas, filogenia, espectrometria de massa

resumo

A saliva e os seus componentes desempenham diversas funções na cavidade oral, tais como lubrificação, proteção dos tecidos orais e ação antimicrobiana. Entre os componentes responsáveis por esses papéis estão diversos péptidos cuja evolução e presença na saliva de outras espécies de mamíferos não está clara.

No presente trabalho, duas classes destes péptidos, as cistatinas salivares e a timosina β4, foram analisadas usando ferramentas de genómica e de proteómica em conjunto. Para os estudos de proteómica foi colhida saliva de cão, rato, coelho e cordeiro, sendo a separação dos péptidos presentes feita por cromatografia liquida e a análise por espectrometria de massa tandem. Para os estudos de genómica foram pesquizadas bases de dados de sequências nucleotídicas e realizaram-se análises evolutivas. No que diz respeito à timosina β4 demonstrou que este péptido apresenta uma elevada conservação entre as diferentes espécies de mamíferos. Utilizando as sequências deste péptido encontradas no genoma dos diferentes mamíferos, foi possível identificar pela primeira vez por espectrometria de massa a timosina β4 na saliva do cão.

No caso da classe das cistatinas, nomeadamente cistatinas C, D e tipo-S (S, SA e SN), a análise evolutiva permitiu verificar que as cistatinas D e tipo-S são específicas dos primatas, o que sugere que terão emergindo após a grande separação dos mamíferos que ocorreu há cerca de 80-90 milhões de anos. Os resultados permitiram também verificar que algumas sequências presentes nas bases de dados encontram-se mal anotadas, incluindo a sequência atribuída à cistatina S encontrada no rato. Por outro lado, a análise filogenética demonstrou que a cistatina C está distribuída por várias classes de mamíferos. No entanto, permanece por compreender o mecanismo da sua secreção na saliva humana e a sua ausência na saliva de outras espécies de mamíferos.

Em conclusão, através da combinação da proteómica e filogenia podemos caracterizar e compreender a distribuição dos péptidos salivares em diferentes mamíferos e comparar com toda a informação existente para a saliva humana.

keywords

Saliva, antimicrobial peptides, cystatins, histatins, proline-rich proteins, statherin, thymosins, phylogeny, mass spectrometry

Abstract

Saliva and its components play several roles in the oral cavity, such as lubrication, protection of tissues and antimicrobial action. Among the components responsible for these roles are several peptides, which evolution and presence in other mammals' saliva is not clear.

In the present study, two peptide classes, salivary cystatins and thymosin β4, were analyzed using a combination of genomic and proteomic tools aiming the enlightening changes in the structure and distribution of these peptides between the different mammal species. For the proteomic analysis, saliva was collected from dog, rat, rabbit and lamb, being salivary peptides separated by chromatography and analyzed by tandem mass spectrometry. For the genomic studies, database of nucleotide sequences were searched and evolutionary analyses were performed. Regarding thymosin β4, the evolutionary analysis showed that this peptide is highly conserved through the collection of all peptide sequences from different mammals species genome, it was possible to identify for the first time by mass spectrometry the thymosin β4 in dog' saliva.

Respecting cystatins class, namely C, D and S-type cystatins (S, SA and SN), evolutionary analysis showed that D and S-type cystatins are Primate specific, which suggesting that these classes emerged after the great mammalian radiation at 80-90 million years ago. The results also showed errors in the annotation of these sequences in databases, in particular the sequence attributed to cystatin S detected in rat. In contrast, evolutionary analysis showed that cystatin C is widely distributed in several mammal classes. However, it is not clear their secretion mechanism to saliva and why its absence in saliva of other mammal' species.

In conclusion, using phylogenetic and proteomic approaches it will be possible to understand and characterize the distribution of these peptides in different mammal species and compare with what is known in the human saliva.

Table of contents

I. General Introduction 3

1. Salivary function 4
2. Salivary proteome and peptidome composition 5
3. Salivary peptides from glandular origin 8
   - 3.1. Contribution of salivary glands for saliva proteome composition 9
   - 3.2. Cystatins 10
     - 3.2.1. Structural features of cystatins 10
     - 3.2.2. Cystatins functions 11
     - 3.2.3. Cystatin family evolution 13
   - 3.3. Histatins and statherin 16
     - 3.3.1. Structural features of histatins 16
     - 3.3.2. Histatins function 18
     - 3.3.3. Structural features of statherin 19
     - 3.3.4. Statherin's function 20
     - 3.3.5. Statherin and histatin evolution 21
   - 3.4. Proline-rich proteins 22
     - 3.4.1. Structure of acidic PRPs 22
     - 3.4.2. Acidic PRPs function 23
     - 3.4.3. Structure of basic PRPs 24
     - 3.4.4. Basic PRPs function 25
     - 3.4.5. PRPs evolution 25
4. Other salivary peptides 26
   - 4.1. Defensins 27
   - 4.2. Cathelicidin 28
   - 4.3. Adrenomedullin 29
   - 4.4. β-Thymosins 30
   - 4.5. S100A proteins 30

5. Salivary composition analysis under an evolutionary perspective 31
6. Aim..... 37
II. Thymosin β4 in mammals' saliva..... 41
1. Introduction..... 41
2. Material and methods 42
2.1. Evolutionary analysis..... 42
2.2. Saliva collection 42
2.3. Peptide isolation and digestion..... 42
2.4. Peptide separation by LC 43
2.5. Mass spectrometry..... 43
3. Results..... 44
3.1. Evolutionary analysis..... 44
3.2. Saliva analysis by LC-MS/MS 45
4. Discussion 48
III. Evolution of C, D and S-type cystatins in mammals 53
1. Introduction..... 53
2. Material and methods 54
2.1. Proteomic analysis 54
2.1.1. Saliva collection and preparation..... 54
2.1.2. SDS-PAGE protein separation 55
2.1.3. In-gel protein digestion 55
2.1.4. Determination of the presence of cystatin S in saliva by Wetern-blotting ……..55
2.1.5. Peptide isolation and digestion..... 56
2.1.6. Peptide separation by LC 56
2.1.7. Mass spectrometry analysis 57
2.2. Evolutionary analysis..... 57
2.2.1. Cystatins sequences collection..... 57
2.2.2. Phylogenetic tree reconstruction 59

3. Results 59
3.1. Proteomic analysis 59
3.2. Evolutionary analysis 60
4. Discussion 67
IV. General discussion and conclusion 77
V. References 83

Abbreviations

ADM – Adrenomedulin

aPRP - Acidic proline-rich protein

BLAST - Basic Local Alignment Search Tool

bPRP - Basic proline-rich protein

CPs - cysteine peptidases

Db-s - Double band, slow

g-PRP - Glycosilated proline-rich protein

HNP - human neutrophil peptide

LC – Liquid chromatography

MALDI-TOF/TOF - matrix-assisted laser desorption/ionization–time-of-flight

MMP - Matrix metalloproteinase

MS/MS - tandem mass spectrometry

Pa - Parotid acidic protein

PIF-s - Parotid isoelectric-focusing variant, slow

PRPs – Proline-rich proteins

PTMs – Post-translational modifications

Tβ4 - thymosin β4

CHAPTER I – General Introduction

I. General Introduction

Many of the processes involved in digestion are common across animal species, including food maceration or enzymatic activity. However, some features in the digestive system are the result of adaptation to chemical, physical and nutritional properties of the diet, being closely related to the species' ecology [1]. The first portion of the digestive tract is the oral cavity, which comprehends the mouth and the structures enclosed within it. This cavity is bathed by saliva, a biofluid composed mainly of water (99.5%) and a complex mixture of proteins (0.3%) and inorganic substances (0.2%) [2,3]. The proteins present in saliva are mainly represented by glicoproteins, like mucins and lactoferrin, enzymes, like α-amylase, immunoglobulins and several peptides, like cystatins, statherin, histatins and proline-rich proteins [3]. The inorganic component is composed by the usual electrolytes found in others body fluids, like sodium, potassium and calcium, found in different concentrations in saliva and contributing to a hypotonic character [3]. These salivary components mainly result from salivary glands' secretion. In humans, parotid, submandibular and sublingual glands are responsible for the secretion of about 93% of the saliva volume, and less than 5% of the volume is secreted by a large number of minor glands present in the mucosa of the tongue (Von Ebner glands), cheeks, lips and palate [4,5]. However, some of the substances found in saliva are derived from mucosal and plasma exudates, gingival crevicular fluid, microorganisms, desquamated cells and food remains [4,5]. Indeed, oral cavity isn't sterile, containing about 700 species of bacteria, viruses and fungi [6,7], which seem to be host-specific [8].

Saliva, in particular human saliva has been the subject of several studies that attempt to find pathophysiological-related changes in its composition for diagnosis purposes and for the knowledge of diseases' pathogenesis [9]. Most of the recent studies in this field focus on the quantitative and/or qualitative alterations of salivary proteins and peptides [5,10-12]. However, the function and origin of some of these species are poorly comprehended and even less is known why some animal species, in particular mammal species, present a different salivary protein profile. So, the study of salivary protein families in a phylogenetic

perspective in mammals might reveal their relationship among different mammal species and relate their structural differences with their role in oral cavity.

1. Salivary function

Belonging to a large group of mucosal fluids, saliva has an important role in protecting the surface of the oral cavity against chemical, mechanical and microbial attack [13]. Saliva acts as a lubricant in the oral mucosa protection against friction. This lubricant action is mainly due to mucins and glycoproteins that represent about 16% of the total protein in whole saliva [14]. The high viscosity and elasticity that characterizes this class of proteins is also common to statherin and gPRPs, which not only contributes to the lubrication and hydration of the oral cavity but also, support mastication and swelling by the formation of the bolus [2,15]. With regard to food digestion, amylase has a key role in this process since it catalyzes the hydrolysis of glycosidic bonds from diet polysaccharides like starch [2].

Saliva also has a bactericidal action conferred by their protein species. For example, lactoferrin is a glycoprotein that is able to bind two atoms of iron, without which bacteria such as *Aggregatibacter actinomycetemcomitans* cannot survive [6,16]. Lysozyme is able to hydrolyze β-1,4 glycosidic linkages between N-acetyl-muramic acid and N-acetilglucosamine of the peptidoglycan found in bacteria's cell walls, leading to their lysis [2]. The antimicrobial action is also represented by immunoglobulins like sIgA, an important secretory immunoglobulin for host defense [17]. Some bacteria and virus could also produce cysteine proteinases that could damage salivary proteins and cause degradation of the oral tissues. However, cystatins present in saliva are able to minimize this proteolytic activity by acting as inhibitors of cysteine proteinases [18]. Moreover, the fungicidal activity of saliva is mainly due to salivary peptides called histatins [6], which also have antimicrobial activity against some strains of *Streptococcus mutans* [6]. Nevertheless, regular cleaning of the oral cavity regulates the balance between these pathogens and saliva antimicrobial activity [7].

Saliva also behaves like a buffer system, which protects the oral cavity from pathogenic microorganisms by denying their optimal colonization conditions and neutralizing the food acids or acids produced by acidogenic microorganisms, thus preventing tooth demineralization [9]. Salivary pH is controlled by specific buffers such as bicarbonate and phosphate buffer systems and also by certain proteins, keeping the human pH of the oral cavity approximately neutral, with higher pH values (about 7.4) observed when the salivary secretion increases and lower values (about 6.2) at low flow rates [2,3,9]. The bicarbonate system is mainly activated when the salivary flow rate increases and phosphate buffer acts when salivary flow is low [9]. Besides these buffer systems, proteins such as histatins or other alkaline products from the oral cavity also control saliva pH [9]. Despite the importance of the neutral pH in the human oral cavity, in other mammals the optimal pH is different. For instance, in ruminants the pH is higher (about 8.1) in order to buffer the volatile fatty acid produced during the digestive processes [19,20]. In frugivorous (fruit eater) Primates, saliva pH is lower than in folivorous (foliage eater) Primates; however, it is not clear whether this difference is a consequence of fruit acids that confer acidic saliva to frugivorous or if is due to buffering agents that maintain a more alkaline pH in folivores [1].

Besides these functions, saliva is also involved in the taste perception by its dissolution capacity allowing the gustatory buds to perceive different flavors associated with distinct substances [21]. All these important roles of saliva are a consequence of its composition, which seem to justify the considerable amount of published studies focused on saliva analysis in different scientific fields like health and life sciences [2,4,5,22,23].

2. Salivary proteome and peptidome composition

The latest studies focused in the salivary proteome shows the presence of more than 2000 different proteins in human saliva, most of which without a glandular origin [24]. In overall, salivary proteins and peptides might be divided in three distinct groups based on their distribution and origin, i.e. if they are saliva-specific or ubiquitous and if they are secreted by salivary glands or from external

sources. According to this classification, proteins secreted by salivary glands and present only in this biofluid are proteins like histatins and acidic proline-rich proteins [13]. Other proteins are common to other body fluids and to several tissues some of which are produced by secretory glands such as lysozymes, mucins, α-amylase and immunoglobulins, while others, such as albumin, are originated from other sources like plasma [13]. In table I.1, the contribution of the main salivary glands to the salivary proteome composition is summarized.

Table I.1 - Concentrations of salivary proteins in human parotid and submandibular-sublingual saliva. (Adapted from [2] and [5]).

Component		Origin*	Concentrations in whole saliva (µg/ml)**
Albumin		gingival crevicular fluid	60-1080
Amylase		Pr; Sm/Sl	380-500
Cystatin		Pr; Sm/Sl	50-280 (Cyst C: 0.9; Cyst D: 3.8; Cyst S: 53-116; Cyst SA: 78; Cyst SN: 39)
Histatin		Pr; Sm/Sl	2-33 (Hst1: 10.5-44.3; Hst3: 1.7-11.8; Hst5: 2.1-16.5)
Lactoferrin		All salivary glands	20
Lysozyme		Pr; Sm	40
Mucins		All salivary glands	MG1: 80-500; MG2: 10-200
Proline-rich proteins (PRPs)	**aPRPs**	Pr; Sm/Sl	267.2-720.6
	bPRPs	Pr	-
Statherin		Pr; Sm/Sl	2-12
sIgA		Pr; Sm/Sl	19-439
α-Defensins		GCF	HNP-1:0.1-10; HNP-2: 0.02-6.0; HNP-3: 0-2.7
β-Thymosins		GCF	Tβ4: 0.2–3.6
S100A proteins		various	S100A8: 1.93; S100A9: 1.93

***** Pr: Parotid gland; Sm: Submandibular gland; Sl: Sublingual gland; GCF: gingival crevicular fluid;
******Hst: histatin; Cyst: cystatin; MG: mucin; HNP: human neutrophil peptide

As can be depicted from Table I.1, the major characteristic components secreted by the human salivary glands are histatins, statherin, S-type cystatins, proline-rich proteins (PRPs), amylases, and mucins [25]. However several other

peptides can be found in saliva, and although some of them have non-glandular origin, most of them show antimicrobial activity [7].

The distribution of theses salivary components in body fluids and tissues could give some indication about its physiological role in the oral cavity. For instance, proteins present in more than one secretion are expected to have functions common to the mucosal secretions where they are found, like, for example, protection of the epithelial tissues from friction [13]. On the other side, it is likely that proteins only present in saliva have specific functions of saliva such as digestion or teeth protection [13].

Recently, emphasis has been given to the peptidome analysis of human saliva [3,26-29]. Polypeptide molecules with molecular weight below 15kDa are considered peptides and they are not necessarily fragments of larger proteins but instead might result from gene expression and subsequent maturation [4]. These peptides and other salivary proteins might also be the target of complex molecular processes. After their biosynthesis they may undergo intracellular pos-translational modifications (PTMs) and, once delivered in the oral cavity, they become target of many enzymes being continually modified [30]. Beyond PTMs like glycosylation, phosphorylation and sulfation, salivary proteins might undergo several cleavage processes by endo- and exo-proteases [5]. The proteases present in the oral cavity are mainly derived from leukocytes of the gingival exudates, microorganisms and epithelial cells [4,31]. It is also worth of note that salivary proteins are not equally susceptible to proteolytic breakdown, being histatins, statherin, acidic PRPs, and basic PRPs the most susceptible ones [26,30,32,33]. So far, 21 proteases were recognized to be involved in saliva proteolysis, namelly kallikrein 1, carboxipeptidase E, cathepsin L, cathepsin D, serine and aspartic proteases and MMP proteases [4,34]. Nevertheless, the proteolytic activity is to some degree non-specific and all salivary proteins could be the target of this process, depending mostly on their type and structure [4].

3. Salivary peptides from glandular origin

Cystatins, histatins, proline-rich proteins (PRPs) and statherin are considered the four major salivary protein families and are known as low molecular weight proteins with established primary structures in humans and unique in the fact that they present multiple physiological activities dependent on the polypeptide chain's region. Indeed, 40-50% of the total secreted proteins in saliva are peptides like cystatins, histatins, statherin and PRPs [4]. These families have multiple activities in saliva, which may be related to their polypeptide chain structure and also to the location and structure of their genes, possibly reflecting evolutionary aspects. The genes encoding these peptides are clustered on chromosomes 3, 4, 12, 20 and 21 (Figure I.1) in the human genome [2].

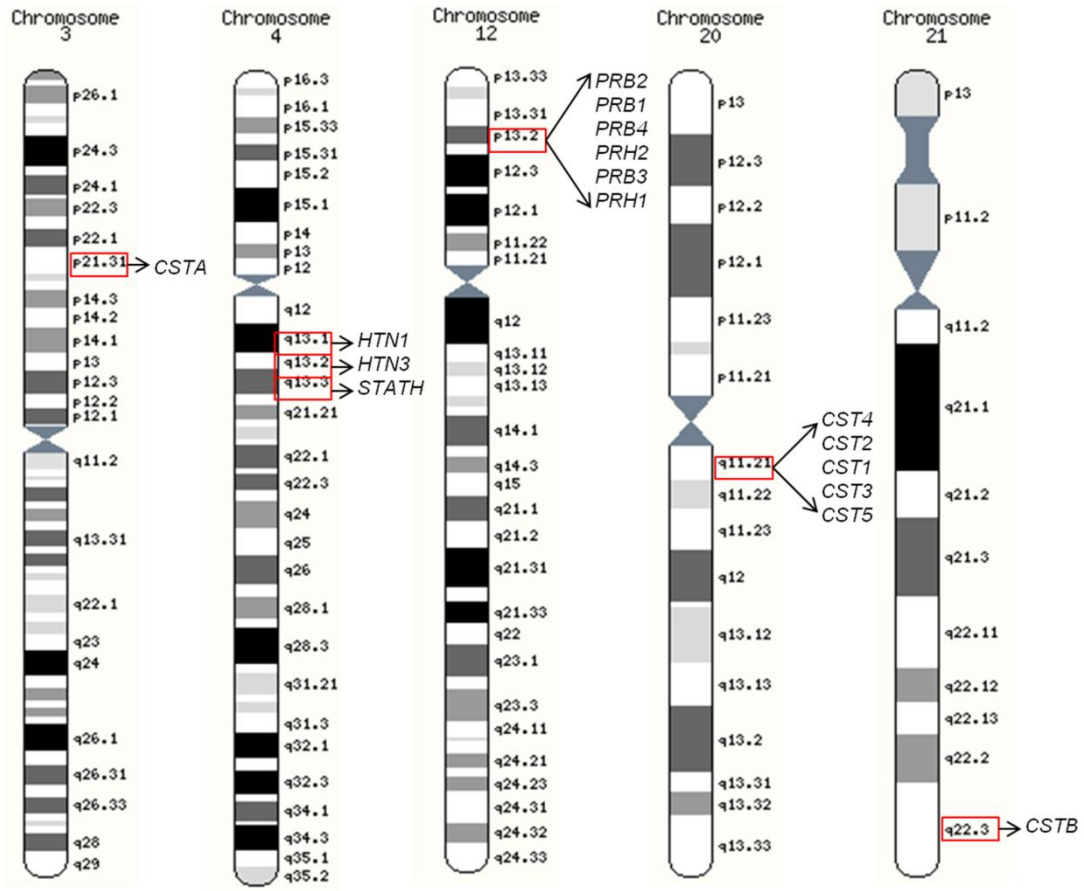

Figure I.1 - Location on human chromosomes 3, 4, 12, 20 and 21 of the genes encoding for statherin, histatins, PRPs and cystatins found in saliva. The information was obtained on Ensembl and NCBI.

The analysis of their evolutionary features in mammals might allow the establishment of the relation between structure and function for the different protein families, revealing their importance in the maintenance of oral health.

3.1. Contribution of salivary glands for saliva proteome composition

Salivary glands show considerable anatomic variability in terms of their size and location among different mammals, which might be understood as a lack of evolutionary restraint [35]. In opposition, organs such as liver and pancreas are cytologically and histologically very conserved among mammals, suggestive of conserved gene expression and regulation, perhaps associated with their fundamental role in the organism [35,36]. The low conservation of salivary glands in evolutionary lineages seems to be related to speciation events, possibly reflecting the different nutritional strategies of mammals [35]. In humans, the morphology of the salivary glands and their content in acinar and ductal cells is well established [37]. The acinar cells are responsible for the production of the gland content. After protein synthesis in these cells, proteins transit to the Golgi apparatus to be stored at secretory granules, from which they are released into the duct system [4,37]. Ductal cells are responsible for driving the content of the glands into the oral cavity [35]. Each gland has its own kind of acinar cells, and so, a characteristic type of secretion. For example, the secretion of the parotid gland is mainly serous, whereas the sublingual gland has a mucous secretion and the submandibular gland has a mixture of both [4,5]. The sublingual gland has many excretory ducts and some of them are joined to form the major sublingual duct, known as duct of Bartholin [5]. The Wharton’s duct, or submandibular duct, opens near the sublingual duct, so the content of this two glands reach the oral cavity as a mixture [5]. At last, the Stensen duct is responsible for carry the content of parotid to the oral cavity [5].

The contribution of each of these glands to saliva composition varies depending on several factors like age, alimentary habits, gender, and also on circadian rhythms and hormonal stimuli [3,38-40].

3.2. Cystatins

Cystatins belong to a heterogeneous family of cysteine proteinase inhibitors that can be divided into three main families [41]: family I include intracellular cystatins A and B, also called stefins; family II is composed mainly by extracellular cystatins such as cystatin E, F, S, SN, SA, C, D, among others; and family III includes intravascular cystatins known as kininogens. Cystatins S, SA, SN (S-type cystatins), C and D are usually found in human saliva, and cystatins A and B were also identified in saliva of children aged up to six months [22,42]. Whereas cystatin C and D are usually found in several body fluids, the S-type cystatins are characteristic of saliva being secreted mostly by submandibular and sublingual glands, although they can also be found in tears [13,22]. Cystatins A and B are usually found in the cytoplasm of epithelial cells but can be secreted or released under specific circumstances [42,43].

3.2.1. Structural features of cystatins

Cystatins S, SA, SN, C, D, A and B are encoded by *CST4*, *CST2*, *CST1*, *CST3*, *CST5*, *CSTA* and *CSTB* genes, respectively [43]. The genes expressing S-type cystatins are clustered on human chromosome 20 at the locus 20p11.2, together with the loci of cystatin C and D known as *CST1-5* locus [23]. S-type cystatins are 13–14 kDa proteins with 121 amino acid residues in length, lacking the 20 residues of the signal peptide [43]. The amino acid sequence of these cystatins shows about 88% homology and all of the S-type cystatins have a conserved glycine at position 11 [43]. Cystatin C is a 13.4 kDa protein with 120 amino acid residues in length, after the removal of the 26 residue signal peptide, and is a basic protein (pI = 9.3), contrasting with other family members [41]. Finally, cystatin D is a 13.8 kDa protein, composed by 120 amino acids after the removal of the 20 amino acid residues signal peptide, with its sequence only showing about 55% homology with other salivary cystatins [41]. These peptides and other type-II cystatins present disulfide bonds at the carboxyl-terminal and

cystatin S can be found phosphorylated at Ser-3 giving rise to cystatin S1, or diphosphorylated at Ser-1 and Ser-3 originating cystatin S2 [23].

Human cystatins A and B are expressed by genes placed at loci 3q21 and loci 21q22.3, respectively [2]. These proteins have about 11 kDa and present almost 100 amino acid residues in length, being different from type-II cystatins in several ways [44]. There are no signal peptide sequences and disulfide bonds in cystatins A and B and they have a different amino acid sequence from type-II cystatins [43]. Cystatins A and B are slightly different in their amino acid sequence showing only 50% of identity [44].

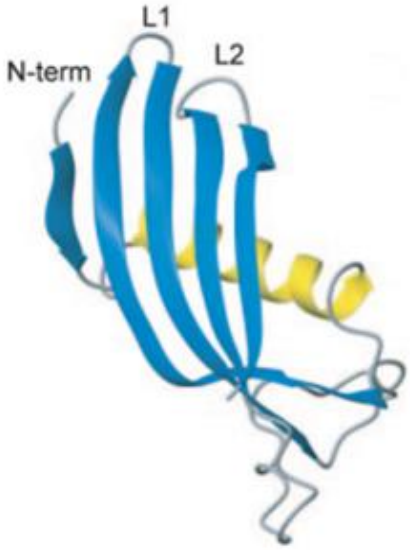

Figure I.2 - Representation of the structure of human cystatin D. The α-helix is marked in yellow, and the β-sheet is marked in blue. The three segments involved in cysteine proteases binding, formed by the N-terminal segment (N-term) and the first and second hairpin loops (L1 for QXVXG segment and L2 for PW residues) are indicated [45].

Although type-I and type-II cystatins present unrelated amino acid sequences, their tertiary structures have similar folds including five stranded anti-parallel β-sheet wrapped around a five-turn α-helix (Figure I.2) [45,46]. These cystatins share three conserved regions, the N-terminal segment, the conserved QXVXG region folded into a β-hairpin loop, and another hairpin loop with the PW conserved residues, all of which important for the inhibition of papain-like cysteine proteases (Figure I.2) [47].

3.2.2. Cystatins functions

These classes of cystatins act as inhibitors of mammals cysteine proteinases of the papain superfamily like mammalian lysosomal cathepsins B, C, H and L, by binding tightly to the enzyme and blocking substrate binding [48]. Most of the proteolytic enzymes are endopeptidases, but cathepsin B presents a carboxypeptidase activity, and cathepsin H has strong aminopeptidase and limited endopeptidase activity [48]. These kind of cathepsins are involved in the normal protein turnover of intracellular and endocytosed proteins not only in the oral cavity

but also in other tissues [48]. As lysosomal enzymes, they also acts in phagocytosis and can be released to the extracellular space by immune cells, where they can be involved in damaging and remodeling of the extracellular matrix and tissues [48]. Cathepsin B is the predominant human cysteine proteinase and it is inhibited by cystatin C and S-type cystatins whereas cystatin A and B mainly control exogenous cysteine proteases [13].

Hall et al. (1995) suggested the N-terminal region of cystatin C, in particular residues 8, 9 and 10, to be crucial for the substrate-binding pockets of cathepsins B, L and H (Figure I.3) [43,49]. In turn, the loop region QXVXG contributes 40-60% to the binding to cathepsins B, H and L binding, and the PW region is mainly important for cathepsin B binding (Figure I.3) [43]. Thus, cystatins can place N-terminal residues and hairpin loop regions along the active site in the same orientation as a substrate, inhibiting cysteine protease [48].

MAGPLRAPLLLLAILAVALAVSPAAGSSPGKPPRLVGGPMDASVEEEGVRRALDFAVGEYNKASN
DMYHSRALQVVRARKQIVAGVNYFLDVELGRTTCTKTQPNLDNCPFHDQPHLKRKAFCSFQIYAV
PWQGTMTLSKSTCQDA

Figure I.3 – Amino acid sequence of human cystatin C with important segments to cathepsins binding highlighted in red, and the signal peptide highlighted in blue. (Sequence was obtained in the UniProt database)

S-type and D cystatins are poorer cathepsin inhibitors when compared to cystatin C, which inhibits cathepsin B and cathepsin L. This fact may be related to the contribution of the different conserved segments to their ability to bind cysteine proteases. As the structure of cysteine proteases interfere with the binding of polypeptide substrates, they can also interfere with the binding of cystatins [43,48]. S-type cystatins are also components of the enamel pellicle, binding to hydroxyapatite through negatively charged regions, such as the N-terminal region of the α-helix, and phosphorylated residues, inhibiting crystal growth and calcium phosphate precipitation [43,50].

Finally cystatins have antimicrobial activity against some pathogens and are able to inhibit viral replication probably by inhibiting necessary cysteine proteases [43,51]. Though unable to inhibit many bacterial species, S-type cystatins and cystatin C can inhibit growth of *Porphyromonas gingivalis* [43,52].

Regarding antiviral activity, it was found that human S-type cystatins and cystatin C are able to block herpes simplex virus-1 (HSV-1) replication [43]. Moreover, cystatin C and also cystatin D act as inhibitors of coronavirus replication, which can cause acute gastroenteritis when present above physiological levels [43].

3.2.3. Cystatin family evolution

The diversity of biological functions of the cystatin super-family members suggests the cystatin domain has evolved from a primordial cysteine protease inhibitory domain into a more diverse protein–protein interaction module [53]. Muller-Ester *et al.* [54] first proposed that cystatins emerged from a stefin-like precursor protein by a fusion of separate exons but with only the one encoding the N-terminal sequence related to the stefins. From it, the kininogens and other molecules with cystatin domains, like fetuins, could have evolved separately or in parallel, and retained or lost their protease-inhibitory activity and active site sequences [55].

In 1990 Rawlings and Barrett [56] suggested that the archetypal cystatin had no disulfide bonds. They also proposed that chicken cystatin diverged from the other type-II cystatins around ~300 million years ago, corresponding to the time of the earliest reptiles and the divergence of birds and mammals [56]. The divergence of cystatins C and S was suggested to occur during early mammalian evolution, about 180 million years ago [56]. In 2002, Dickinson [43] proposed that cystatins from plants and animals diverged from a common ancestor at about 1.6 billion years ago, being phytocystatins the best representative of the ancestral cystatin. This author also suggested that α-helix structure and the first disulfide bond were acquired about 1.2 billion years ago, and the second disulfide bond together with the other characteristics of type-II cystatins must have evolved about 1 billion year ago. A few smaller and evolutionarily younger orthologous families have emerged in some mammalian orders, such as the Primate-specific subfamily of salivary cystatins (cystatins S, SA, SN and D). The phylogenetic tree proposed to the vertebrate type-II cystatins shows that D and S-type cystatins form a monophyletic clade with quite good confidence (84%) (Figure I.4). According to

Dickinson [43], there is no reason to separate cystatin D and S-type cystatins, since they seem to have evolved from a common cystatin C-like ancestor at about 100 million years ago, and should be collectively referred as SD type cystatins (Figure I.4).

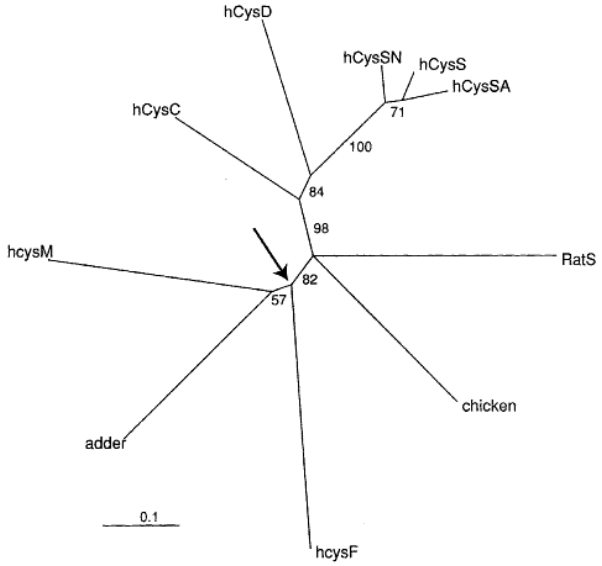

Figure I.4 - Phylogenetic tree of some vertebrate type-II cystatins. The large arrow indicates a possible position for the root of the tree, with proteins on the cystatin C branch of the tree to the top. Abbreviations used: hCys, human cystatin; rat S, rat salivary cystatin S; and adder, puff adder (*Bitis arietans*) venom cystatin. [43]

In this tree, the branch representing the rat cystatin S is not grouped with the human SD-type cystatins, which indicates an independent origin. However this branch is not positioned with high confidence, so it is unclear if the rat cystatin S is a highly divergent ortholog of the human cystatins or if represents a case of independent evolution (Figure I.4) [43].

The latest phylogenomic analysis of prokaryotic and eukaryotic cystatins has shown the presence of only two ancestral lineages for the cystatin family, stefins and cystatins. They are referred as ancient eukaryotic paralogs because they were probably formed by duplication prior to the divergence of the principal eukaryotes lineages. The work of Kordis and Turk [47] also suggested an intracellular progenitor of the cystatin superfamily lacking a signal peptide and disulfide bridges, and so similar to the extant Giaardia cystatin. In contrast to plant cystatins, cystatins from vertebrates and placental mammals' suffered great diversification during the evolution. Cystatin and stefins were then created by a primordial gene duplication and while stefins remained as a single gene or as small multigene family in eukaryotes, cystatins underwent a more complex and dynamic evolution through several gene and domain duplications (Figure I.5) [47].

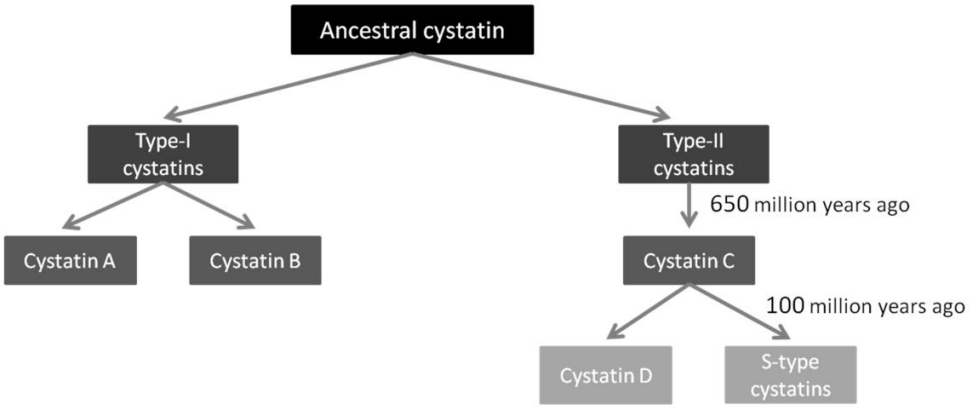

Figure I.5 - Evolutionary scheme of type-I and type-II cystatins

About 650 mya, the ancestor of vertebrates' cystatins experienced the first diversification, giving rise to four orthologous family, including the progenitors of cystatin C. From the ancestor bony vertebrates (Euteleostomi) two novel orthologous families emerged, including the current cystatin C [47]. From the ancestor of land vertebrates (Tetrapoda) two novel orthologous families appeared and finally from placental mammals six orthologous families, that include cystatins 8, 9, 11, 12, 13 and like-1 were formed [47]. In the ancestor of Amniota the duplication of the stefin gene has produced stefins A and B [47]. These authors noticed a large difference between placental (Eutheria) and ancestral (Prototheria and Metatheria) mammals, with the latter including 11 orthologous cystatin-derived families in addition of stefins A and B, and the former have 17 orthologous as well as stefins A and B [47]. So, ancestral mammals are closer to other land vertebrates such as amphibians that also have approximately 11 orthologous families. Kordis and Turk [47] also reported the partial loss of ancestral inhibitory activity in vertebrate orthologous families with the acquisition of novel functions in innate immunity. For example, human salivary cystatins are more recent than other type-II cystatins and they are less active against the host lysossomal cathepsins than others such as cystatin C [43].

3.3. Histatins and statherin

Histatins belong to a family of small peptides with 3 to 4 kDa containing multiple histidine residues, being these peptides slightly basic [13,57]. These peptides are secreted by parotid, submandibular and sublingual glands and were first described in the early 1970s as peptides that enhance the glycolytic activity of microorganisms [58,59]. Only later it was described their potent bactericidal and fungicidal properties [60,61]. The concentrations of histatins in whole saliva are considerably lower than the concentrations found in parotid or in submandibular/sublingual secretions. This fact is mainly due to proteolytic breakdown, involving several enzymes with various proteolytic specificities, such as serine proteases [30]. So far, 135 different histatin family peptides have been acknowledge in whole saliva resulting from fragmentation of histatins 1 and 3 [4,30]. These peptides have already been found in saliva of humans, macaques, and a number of other Primates representing various Anthropoids [62].

Histatins genes show nearly identical overall gene structures with statherin, suggesting these genes belong to a single gene family exhibiting accelerated evolution between the histatin and statherin coding sequences [62]. Human statherin is a 5.38kDa singular salivary peptide rich in tyrosine, proline and glutamic acid secreted by parotid and submandibular glands, also emerging in other body fluids besides saliva [13,63].

3.3.1. Structural features of histatins

Histatins 1, 3 and 5 are the most prominent histatins in human saliva [4]. Histatin 1 and histatin 3 have very similar sequences and are encoded by genes *HTN1* and *HTN3*, also known as *HIS1* and *HIS2*, respectively, both located on human chromosome 4 at the loci 4q13 [64-66]. These two genes show 89% overall sequence identity, with exon sequences exhibiting 95% homology, which have probably emerged by a gene duplication event in hominoids around 15-30 million years ago, after their divergence from cercopithecoids and prior to the hominid divergence [62].

Human histatin 1 is composed by 38 amino acid residues without the signal peptide (first 19 amino acid residues), emerging phosphorylated on Ser-2 and poly-sulfated on the four terminal tyrosines by the tyrosylprotein sulfotransferase (Figure I.6) [65,67,68]. Proteolytic cleavage originates histatin 2, which consists on the fragment sequence 12-38 [66].

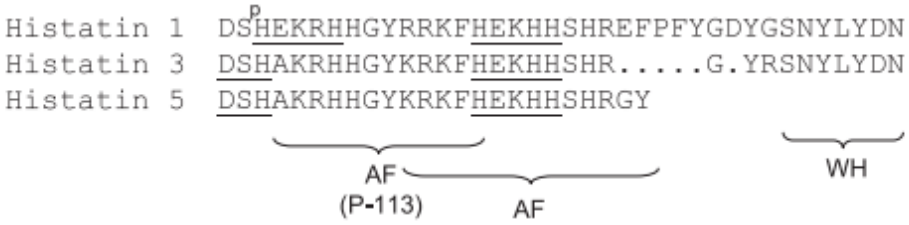

Figure I.6 - Amino acid sequence of human histatins 1, 3 and 5 and indication of functionally important peptide regions: DSH, metal-binding motif; HEXXH, zinc-binding region; AF, antifungal domain; WH, wound-healing domain [68]

Histatin 3 presents 32 amino acid residues in length after removal the signal peptide (Figure I.6) [66]. Due to the high susceptibility of motif RGYR to proteolysis, histatin 3 sequential cleavages give rise to many fragments including histatin 5, a fragment composed by the first 24 amino acid residues present in histatin 3 and resulting from tryptic like cleavage after Tyr-24 [33,65]. The substitution of Glu residue present in histatin 1 at position 4 by Ala in histatin 3 abolished the kinase recognition site preventing the Ser-2 phosphorylation [65]. The first three residues of histatin 3 and 5 (DSH) comprehend a conserved motif that complexes Cu^{2+} and Ni^{2+} [69]. Histatin 1 also has two specific motifs (HEXXH) capable of binding Zn^{2+}, as well as histatin 3 and 5, that show an equal one (Figure I.6) [69].

Similar proteins to histatin exist in macaques, cattle and most likely in other mammals [69,70]. In the specie *Macaca fascicularis*, it was found one peptide assigned as histatin 1 which showed 89% and 91% of similarity with human histatins 1 and 3 sequences, respectively. In addition, it exhibited comparable fungicidal effects against *Candida albicans* with human histatins [71]. However, in cattle their structure and function are not well established yet [70].

3.3.2. Histatins function

Histatin 1 seems to be involved in the maintenance of tooth enamel mineral and pellicle formation by binding to hydroxyapatite, being able to inhibit crystal growth of calcium phosphate salts [69]. The negative charge of the phosphate group is a determining factor in the interaction of histatin 1 with tooth enamel mineral [30]. Histatin 1 also shows wound-healing activity both in oral primary cells and in non-oral cells [72,73]. However, histatin 2 appears to be more active in wound-closure process than histatin 1 [69]. It is though that for this function histatins are internalized by the epithelial cells and activate the extracellular signal-regulated kinases pathway, enhancing epithelial migration to the wound [69]. Histatin 3 also presents a small wound-healing activity, suggesting the last seven amino acid residues of histatins 1, 2 and 3 are those responsible for this activity (Figure I.6) [69].

By complexing ions like Zn^{2+}, Cu^{2+} and Ni^{2+} histatins eliminate cofactors for enzymes or essential elements for bacterial growth [69]. In this regard, histatins are able to kill *Streptococcus mutans*, a bacteria responsible for dental carries and histatin 5 shows a great ability to inhibit the human matrix metalloproteases MMP-2 and MMP-9 and appears to be the most potent histatin against yeast *Candida albinans*, *Cryptococcus neoformans* and *Aspergillus fumigatus* [59,69,74]. Residues 12-25 from histatin 3 are the main responsible for their fungicial activity and they are present in almost all of the longer fragments, suggesting the proteolytic process would not reduce the antimicrobial properties of these peptides [30]. Histatin 5 is also essential for the salivary antimicrobial activity and, like other cationic peptides, can adopt α-helice secondary structures [75]. These amphiphilic structures seem to form channels or pores in model lipid bilayers [57]. However, it has been suggested the mechanism of action of histatin 5 against microorganisms differs from that of most helical proteins that usually lyse pathogen's lipid membranes. In *C. albicans* it is thought that this peptide is internalized and reaches the mitochondrion where causes loss of transmembrane potential and inhibits the respiratory chain at coenzyme Q level [76]. All these steps are reflected in an energetic collapse due to decreased ATP synthesis and increased

levels of reactive oxygen species that damage biologically important molecules, leading consequently to loss of cell integrity [5]. Histatin 5 also has antifungal domains, which are located in the N-terminal and middle region of these molecules (Figure I.6) [68].

3.3.3. Structural features of statherin

Human statherin is composed by 43 amino acid residues without the signal peptide and is phosphorylated at Ser-2 and Ser-3 [13,77]. In humans, statherin is encoded by the gene *STATH* localized on chromosome 4q.13.3, near to histatin genes [23]. Three other variants of statherin, known as SV1, SV2 and SV3 were also characterized [78]. SV2 lacks residues 6-15 present on statherin sequence due to alternative splicing of a exon coding for the missing residues [78]. SV1 and SV3 have the same sequence as statherin and SV2, respectively, but lack the carboxyl-terminal phenylalanine residue; however, it is not known if the removal of this residue affects these statherin's functions [30,78]. Statherin is one of the peptides more prone to fragmentation and besides the three variants found in human saliva, 87 different fragments were also identified [4]. Raj *et al.*, in 1992 [79], reported the importance of the negative charge density given by the first 15 amino acid residues and of the helical conformation at the N-terminal region for statherin interaction with hydroxyapatite. The first five amino acid residues are negatively charged and are followed by some positively charged residues, where the segment Glu5-Gly15 forms an amphipathic α-helix with the hydrophobic and hydrophilic residues on opposite sides of the helix , isolating hydrophobic residues at positions 7, 8, 11, and 14 from the aqueous medium (Figure I.7) [63,80]. The remaining sequence is composed of hydrophobic and mostly uncharged polar residues like glutamine, proline, and tyrosine [79]. When attached to hydroxyapatite statherin

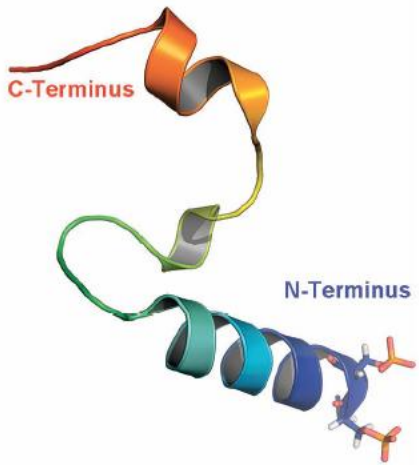

Figure I.7 – Structure of human statherin with sticks drepresenting the two phosphoserine side chains Sep2 and Sep3 [81]

shows a structure represented by an α-helical structure at the N-terminal region (residues 1–16), a polyproline type II (PII) helix in the intermediate region (residues 19–35), and a 3_{10} helix in the C-terminal region (residues 36–43) (Figure I.7) [81,82].

3.3.4. Statherin's function

The statherin peptide shows great affinity for calcium phosphate minerals and is able to inhibit precipitation and crystal growth of hydroxyapatite from supersaturated solutions of calcium phosphate [79]. The first polar 6 amino acid residues (Asp-pSer-pSer-Glu-Glu-Lys) of statherin are essential for binding hydroxyapatite and the first 15 residues are the main responsible for crystal growth's inhibition (Figure I.8) [80]. Like for other anionic proteins, it is suggested that the inhibition of calcium phosphate precipitation might proceed via binding to calcium ions [79].

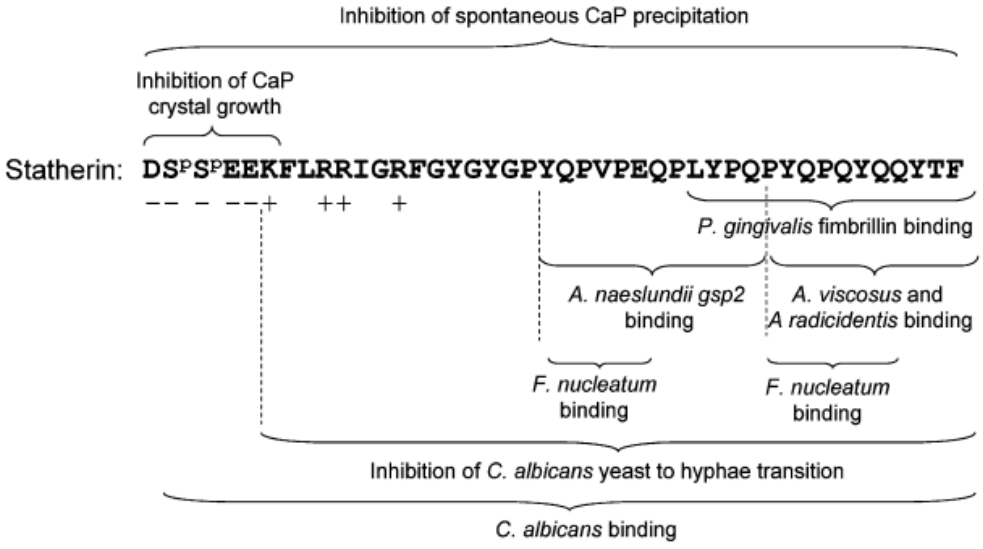

Figure I.8 – Amino acid sequence of the human statherin and their functional characteristics [83]

When adsorbing to the tooth surface, the C-terminal domain undergoes transition from random coil to a helical conformation, unmasking oral bacterial binding epitopes that were not accessible in the protein while in solution [83]. Thus, statherin can promote the adherence of oral bacteria such as

Porphyromonas Gingivalis and *Actinornyces viscosus* [13]. In this regard, it is thought that bacterial adhesins may recognize two PQ segments (residues 31-32 and 36-37) located at C-terminal (Figure I.8) [80]. Statherin showed also the ability to inhibit the conversion of *Candida albicans* blastoconidia into the more virulent hyphal growth form [84].

Finally, together the polar N-terminal end, the relatively nonpolar C-terminal end and the structure presented are necessary to form an amphipathic oriented film at the enamel interface providing the lubricant properties also expressed by statherin [80].

3.3.5. Statherin and histatin evolution

Histatin and statherin exhibit little similarity in their amino acid sequence and are not considered members of the same family [62]. However, the histatin cDNAs exhibit an unexpected similarity to statherin cDNAs in the 5' and 3' untranslated regions (UTRs) and signal peptide sequences, and their genes are localized on position 4q11-13 of the human chromosome, further suggesting a possible evolutionary relationship [62]. Comparison of the *HTN* sequences with the *STATH* sequence suggests that *STATH*, *HTN1* and *HTN3* arose by gene-duplication events (Figure I.9) [62].

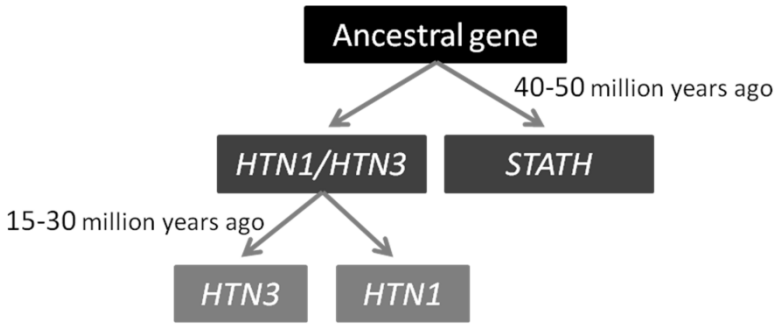

Figure I.9 - Evolutionary scheme of the statherin/histatin family.

The *HTN1* and *STATH* genes show nearly identical overall gene structures, exhibit 77%-81% sequence identity in the intronic regions and 80%-88% sequence identity in noncoding exons but only 38%-43% sequence identity in the protein-coding regions of exons 4 and 5 [62]. Together with their chromosomal location,

this suggests that *HTN*, and *STATH* belong to a single gene family with accelerated evolution between the *HTN* and *STATH* coding sequences [62]. The time of the initial gene-duplication event preceding the divergence of the *STATH* and *HTN* sequences has not been clearly defined. However, Li and Tanimura [85] suggested that *HTN* gene duplication occurred about 15-30 mya in hominoids after their divergence from cercopithecoids and prior to the hominid divergence. Thus, authors estimated by the rate in intron sequences substitutions that the initial gene-duplication event occurred about 40-50 million years ago [62].

3.4. Proline-rich proteins

Proline-rich proteins (PRPs) account for more than 60% in weight of the total salivary proteome being characterized by a predominance of amino acids proline, glycine, glutamine and glutamate [13,23]. This heterogeneous group is usually divided in acidic (aPRPs), basic (bPRPs) and glycosylated (gPRPs) proline-rich proteins and are encoded by a multigene family of 6 genes, giving rise to more than 20 PRP species due differential mRNA splicing and proteolytic cleavage after secretion [13]. In this sense, the concentration of acidic PRPs is lower in whole saliva than in glands due to proteolytic process [30]. Finally, glycosilated PRPs belongs to basic PRPs group, being its glycosilated form closely related to acidic PRPs, with similar amino acid sequences and close chromosomal locations [13]. The order of their genes in the cluster is most likely to be 5' *PRB2-PRB1-PRB4-PRH2-PRB3-PRH1* 3' [86].

3.4.1. Structure of acidic PRPs

Acidic PRPs are proteins of about 16kDa, only expressed by salivary glands by two loci, *PRH-1* and *PRH-2* located on human chromosome 12p13.2 [13,23]. *PRH1* codes for three alleles that express PIF-s (parotid isoelectric-focusing variant, slow), Db-s (double band, slow) and Pa (parotid acidic protein) isoforms [5]. *PRH2* are bi-allelic and codes for PRP-1 and PRP-2, two proteins similar to PIF-s protein with 150 residues in length (without the 16 amino acid residues

signal peptide) [5,50]. Before being stored in the secretory granules, these proteins are phosphorylated at Ser-8 and Ser-22 and sometimes at Ser-17 and could undergo cyclization at the *N*-terminal due to the formation of pyroglutamate from glutamate residue [5]. Usually Pa protein is detected as a dimeric derivative originated by disulfide bonding involving Cys-103 residue [32]. After the secretion, the isoforms PRP-1, PRP-2, PIF-s and Db-s can also be partially cleaved near to the C-terminal end at the Arg_{103}-Pro-Pro-Arg_{106} ↓ motif, being the site of the cleavage located after the residue Arg-106 (Arg-127 in Db-s protein), resulting in four truncated isoforms known as PRP-3, PRP-4, PIF-f with 106 residues in length and Db-f, and a C-terminal fragment called P-C peptide with 44 residues [50,87]. The Pa isoform does not undergo proteolysis because the first Arg residue is replaced by a Cys residue, abolishing the protease recognition site [30]. Besides these, another 77 fragments from aPRPs have already been found [4]. It has been suggested the cleavage of aPRPs is under the action of a proprotein convertase which recognizes the RXXR consensus sequence, like furin [32,88].

aPRPs are typically composed by a polar and negatively charged amino terminus followed by an extended nonpolar sequence [50]. The differences between the most similar aPRPs are on residues 4 and 50. PRP-1 and PRP-3 have an acid aspartic on position 4 and an asparagine on position 50, PRP-2 and PRP-4 have acid aspartic residues in both positions, and PIF-s and PIF-f have asparagines on position 4 and aspartic acid on position 50 [50]. Pa has also 150 residues in length but have different residues at positions 26 and 103, and Db-s has 171 residues in length, due to a 21-residue insert after Gly83 and have a different residue at position 26 [50].

It is also worth of note the three uncommon allelic products of PRH1 found in Asiatic population, where At and Aw proteins are allelic products of the PRH l locus and Au is another allele of PRH2 [89].

3.4.2. Acidic PRPs function

The aPRPs N-terminal 30 residues are able to bind to hydroxyapatite and bind calcium ions, thus inhibiting the crystal growth of calcium phosphate in

supersaturated solutions (Figure I.10) [50]. It is thought that the adsorption of aPRPs to hydroxyapatite causes a conformational change which together with the acidic amino acid and phosphoserine residues promotes the calcium binding ability [50]. This conformational change also allows the attachment of some oral bacteria such as *Actinomyces viscosus* to the aPRPs carboxyl terminal [50]. Also for this function seems to be important the several segments of P-Q present along the amino acid sequence (Figure I.10) [90].

MLLILLSVALLAFSSAQDLDEDVSQEDVPLVISDGGDSEQFIDEERQGPPLGGQQSQPSAGDGNQ
DDGPQQGPPQQGGQQQQGPPPPQGKPQGPPQQGGHPPPPQGRPQGPPQQGGHPRPPRGRP
QGPPQQGGHQQGPPPPPPGKPQGPPPQGGRPQGPPQGQSPQ

Figure I.10 – Amino acid sequence of the human acidic PRP with first 30 residues responsible for binding hydroxyapatite and calcium ions highlighted in green, the several segments PQ important for the bacterial attachment highlighted in red and the signal peptide highlighted in blue. (Sequence was obtained in the UniProt database)

Finally, some studies reveal that *Streptococcus mutans* binds more strongly to larger aPRPs than to the smaller ones. Thus, larger aPRPs seems to promote bacterial attachment whereas smaller aPRPs allow the reduction of the bacterial attachment through the lack of the main binding sites [50].

3.4.3. Structure of basic PRPs

bPRPs are present in several body fluids and at the oral cavity level they are only expressed by parotid gland and represent the major component of adult saliva [5]. These peptides are expressed by four loci, *PRB1-PRB4* located at 12p13.2 chromosomal position. *PRB1* and *PRB3* loci express four alleles, small (S), medium (M), large (L) and very large (VL) while *PRB2* and *PRB4* loci express three alleles, S, M and L [5]. The size of these alleles depends on the number of 63-bp tandem repetitions found in these genes [86]. bPRPs are very polymorphic due to individual insertions or deletions, tandem repeats, alternative splicing and complex PTMs [23]. So, the proteins expressed by these genes appear only as multiple fragments of larger pro-proteins and present several PTMs such as phosphorylation and glycosylation [4,91]. There are some issues about the

nomenclature of these peptides. Isemura *et al.* [92] and Saitoh *et al.* [93-95] classified some of them as P-D, P-E, P-F and P-H. Later Kauffman *et al.* [30] classified these and other fragments as IB5, IB9, IB8c, IB4 (or P-D, P-E, P-F and P-H, respectively), IB1, IB6, IB7, IB8a, II-1 and II-2. Recently, Amado *et al.* [4] assigned short names according IUPAC recommendations to these same peptides, bPRP4L 225/294 to IB5, bPRP1L 75/136 to IB9, bPRP2L 283/343 to IB8c, bPRP2L 345/416 to IB4, bPRP2L 1/96 to IB1, bPRP1L 259/376 to IB6, bPRP2L 97/155 to IB7, bPRP2L 159/277 to IB8a, bPRP4L 1/174 to II-1 and bPRP1L 1/75 to II-2 [4]. However, other fragments were also found and sometimes overlap the peptides mentioned, which creates chaos in the nomenclature [4]. In glycosylated PRPs the proteolytic process seems susceptible to the presence and spatial distribution of carbohydrate chains, whereas in non-glycosylated basic PRPs the cleavage occurs mostly in Arg-Ser-Xaa-Arg↓ segment (the arrow indicates the site after which cleavage occurs), a motif similar to that in aPRPs [30].

3.4.4. Basic PRPs function

The biological roles of bPRPs are not completely known. Nevertheless, these proteins seem to prevent the toxic effects of food tannins by blocking their absorption from the gut [96]. The complexes are formed by the association between the tannin polyphenol ring and the pyrrolidone ring of proline [96]. A role in taste perception has also been suggested for bPRPs since the complete expression of these peptides is only reached after puberty, the age at which this sense is fully developed [5]. Finally, gPRPs have the main function of lubricating oral tissues [13].

3.4.5. PRPs evolution

PRB and *PRH* genes have a similar organization, with 4 exons and 3 introns [86]. The first exon has 64 bp in length and encodes the 16 amino acid residues signal peptide and the first 5 residues of the N-terminal region [86]. The

second exon contains 36 bp and code for the next 12 amino acid residues [86]. Exon 3 is largest than the previous and is composed by multiple 63 bp tandem repeats, varying in number and sequence among the genes. *PRH* genes have an additional 102 bp portion in this exon [86]. The last exon is untranslated and contains a poly(A) sequence [86].

In 1990 Kim *et al.* [97] proposed an evolutionary relationship based on a physical map of the chromosomal location of human PRP genes and their partial gene sequences. According to this model, initial gene duplication occurred followed by three unequal but homologous cross overs (Figure I.11) [86].

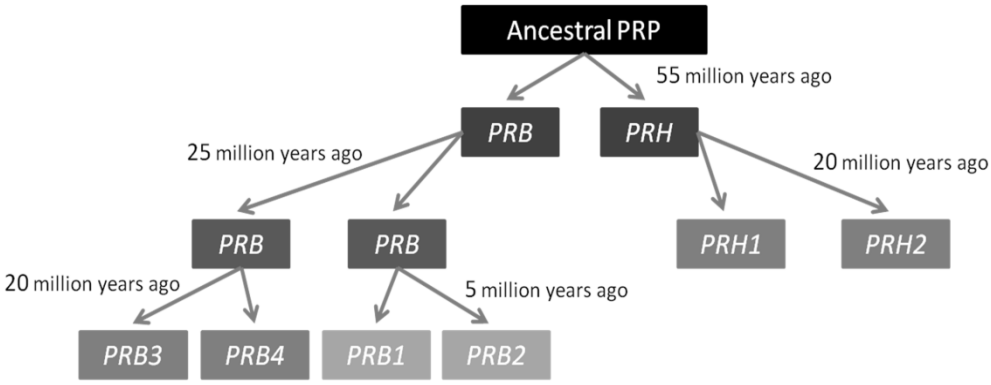

Figure I.11 - Evolutionary scheme of the PRPs family.

The first duplication is thought to have occurred around 55 mya and gave rise to the ancient *PRB* and *PRH* genes [86]. The second duplication arose around 25 mya and originated two related precursor genes from the ancient *PRB* gene. At about 20 mya, the genes *PRH1* and *PRH2* appeared from the ancient *PRH* gene and the genes *PRB3* and *PRB4* emerged by duplication of one of the precursor *PRB* genes [86]. Finally, about 5 mya the other *PRB* precursor gene was duplicated and originated *PRB1* and *PRB2* [86].

4. Other salivary peptides

The oral epithelium is not only a passive cover, but a barrier against infection, playing an active role in innate host defense [98]. Epithelial cells, responding to the contact of oral bacteria and their products, produce natural

antimicrobial peptides and proteins which are part of the innate immune system [99]. These peptides complement the antimicrobial factors of saliva, such as the histatins, lysozyme, and salivary immunoglobulins [99].

The clinical use of these peptides may be beneficial for the early and non-invasive detection of oral diseases, such as periodontitis, where the concentration of some of these peptides is altered [7]. Besides, its use was already suggested to the management of oral pathological conditions, since they have shown a broad-spectrum activity, mainly against bacteria even at low levels, and are able to minimize the inflammatory state [7].

4.1. Defensins

Defensins are the most prominent mammalian antimicrobial peptides and could be divided in two main subfamilies known as α-defensins and β-defensins both presenting antibacterial, antifungal and antiviral activities [69]. Human α-defensins are composed by six peptides, HNP (human neutrophil peptide)-1 to -4, HD-5 and HD-6, encoded by five genes, where the HNP-2 is a truncated form of HNP-1 or HNP-3 [69]. β-defensins are composed by four peptides, hBD-1 to -4, whose genes are located together with the five α-defensins genes in a cluster on human chromosome 8q23 [100].

The HNP-1 to -4 peptides are mainly located in azurophilic granules of the neutrophils, which enter the oral cavity mainly through the junctional epithelium surrounding the teeth, while β-defensins are produced in epithelial cells and mucosa [69]. However, in human saliva, only five members of defensin's family – HNP-1, HNP-2, HNP-3, hBD-1 and hBD-2 – have been identified [2,101]. Also, the distribution of these peptides in the various tissues of different animal species is not homogeneous. For example, α-defensins are not present in mouse neutrophils like is common in human and in species such as bovin; these immunologic cells contain β-defensins instead of α-defensins [100]. Additionally, another class of defensins, the θ-defensins were found, but only in species of Old World monkeys and in orangutans (not found in humans) [69].

This family of peptides present high variable amino acid sequences with only six characteristic cysteine motif [100]. Regarding defensins evolution it has been suggested the ancestral vertebrate defesin was a β-defensin and at some point in mammalian evolution a duplication event gave rise to the α-defensin gene (Figure I.12) [102,103]. Further duplication of α-defensin gene occurred after species divergence, since α-defensin genes cluster seems to be specific for each specie [103]. The presence of β-defensins orthologues in several different mammalian species suggests the occurrence of duplication of the ancestral β-defensin gene even before the mammalian appearance [102]. However, it was also suggested that these genes were under an ongoing evolutionary process after mammalian divergence, originating species-specific β-defensin genes [102].

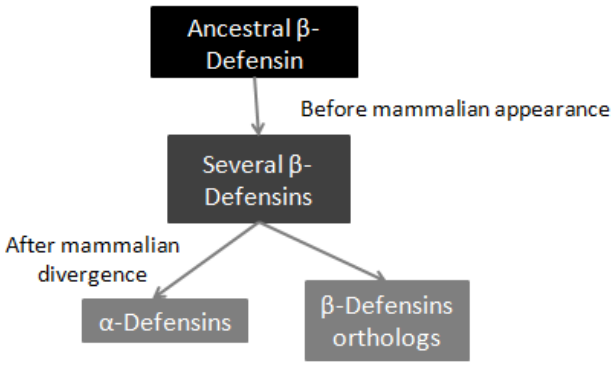

Figure I.12 - Evolutionary scheme of the defensins family.

4.2. Cathelicidin

Cathelicidin is an 18 kDa cationic antimicrobial protein produced in neutrophils and epithelial cells. The proteolytic cleavage at the C-terminal originates different heterogeneous peptides that can be found in saliva and gingival crevicular fluid [104,105]. These peptides are linked at the N-terminal prosequence, known as “cathelin” domain, and can range in size from 12 to 100 amino acid residues, being the most common linear peptides of 23-37 amino acid residues with amphipathic α-helices conformation [106].

Despite the presence of cathelicidin family components in several mammalian species, the most known cathelicidin peptides are human LL-37,

porcine PR-39, and bovin BMAP-28 [106]. Regarding the oral cavity, the LL-37 peptide was identified in human saliva and gingival crevicular fluid, where it acts as an antibacterial agent and is also able to neutralize the lipopolysaccharides produced by bacteria like Gram-negative bacteria, playing an important role in maintaining oral health [7,99].

The evolutionary history of these molecules is not clear, besides not being found in non-mammal species [106]. Additionally, the genes encoding this protein family are under different selective pressures, existing different sets of related genes in each species [106]. In bovines, there are eleven cathelicidin genes clustered on chromosome 22, while in sheep there are eight different genes on chromosome 19. In turn, humans and mice contain one typical cathelicidin domain each, being the genes located at chromosome 3 and 9, respectively [106]. Curiously, these genes are located at regions of conserved synteny but most peptide sequences are remarkably different, being only highly conserved the cathelin domain [106].

4.3. Adrenomedullin

Adrenomedulin (ADM) is a small peptide with 52 amino acid residues resulting from a precursor protein codified by a gene located on human chromosome 11 [99]. In turn, ADM is a member of calcitonine superfamily and present considerable homology with other family members, such as amylin or calcitonin [107].

Despite, the presence of adrenomedullin in saliva from the submandibular and parotid glands, its concentration is higher in whole saliva, suggesting the oral epithelium as the main contributor of this peptide [108]. Regarding its function, this peptide was first considered as a vasodilator and later was recognized its antibacterial activity against Gram positive and Gram negative bacteria in the oral cavity [99]. Moreover, it is though the C-terminus is the main antimicrobial part, and the formation of postsecretory fragments could enhance the antimicrobial activity, manly against *Escherichia coli* and *Staphylococcus aureus* [69]. However

it was suggested that low concentration of adrenomedullin in saliva of healthy subjects cause no significant antimicrobial action [108].

4.4. β-Thymosins

The β-thymosins are a family of highly conserved 5kDa, that in man is composed by thymosin β4 (Tβ4), thymosin β10 (Tβ10) and thymosin β15 (Tβ15) [109]. The main function of this peptide family is to regulate the dynamics of the cytoskeleton by binding to cytoskeletal G-actin (monomers form) [110]. Among these, Tβ4 is the most studied, being present at high concentration in almost every cell [109]. This peptide shows several functions, since it promotes cell migration and differentiation, upregulate metalloproteinase activity, promotes tissue repair and angiogenesis and is also involved in tumor metastasis [110-112]. Despite being highly conserved, different β-thymosins have different functions, for example, while Tβ4 seems to promote angiogenesis and cell migration, Tβ10 inhibit this pathway and is a barrier to cell migration [113].

Tβ4 can also be found in saliva and other body fluids. In pre-term newborn saliva, the high levels of this peptide found in the oral cavity are mainly due to secretion by salivary glands, but in adult saliva, Tβ4 originate mainly from gingival crevicular fluid [114,115]. In the oral cavity, the presence of this peptide is associated with the development of several cranio-facial organs but also with its antimicrobial, anti-inflammatory and anti-apoptotic activity [116,117].

Regarding the evolution of this peptide family, very few data are available, and the existence information only focus on the conserved domain that bind to G-actin [118].

4.5. S100A proteins

S100A8 (calgranulin A) and S100A9 (calgranulim B) form a dimer known as calprotectin, which is a heterodimeric protein able to bind calcium and zinc, for which is essential the conserved metal binding His-X-X-X-His motif [7,99]. These peptides can be found in neutrophils, monocytes, macrophages and mucosal

keratinocytes, and are also expressed in cells of stratified oral epithelia and in cultured gingival epithelial cells [99].

These peptides show antimicrobial activity, mainly by sequestering zinc which is essential for the growing of some microbial species [99]. At high concentrations, the calprotectin heterodimer is able to inhibit the growth of *Escherichia coli*, *Staphylococcus aureus* and *Staphylococcus epidermidis*, and even at lower concentrations (4-32 µg/ml) it can inhibit growth of *Candida albicans* [119].

Regarding its evolution it is though that two ancestor members, S100A1 and S100B, originated during the vertebrate evolution gave rise to all the S100 protein family, which shown conservation of the N- and C-terminal regions [120,121]. The thirteen human S100 genes are clustered on chromosome 1q21, and it was found in mouse a syntenic region on chromosome 3 containing the S100 genes cluster, which shows a conservation pattern during evolution [120].

5. Salivary composition analysis under an evolutionary perspective

For nearly a century, mammalian phylogenetics was dominated by comparative anatomy and palaeontology, allowing the division of living mammals into two subclasses, Prototheria and Theria. Prototheria includes the egg-laying monotremes, such as platypus, whereas Theria includes two major groups of viviparous mammals, marsupials and placentals [122]. Regarding placentals Kumar and Hedges [123] estimated intraordinal divergences for Primates, Carnivores and Rodents and found that these were all at 65.5 (±0.3) million years ago (mya) except for Rodents, where intraordinal divergences were as far back as 112 (±3.5) mya. The subsequent evolution of the distinct orders can vary consistently with species body size, population dynamics, lifestyle and location [124] (Figure I.13). Indeed, contrary to hopes that molecular evolution would be clock-like, variation in evolutionary rates between species appears to be the rule, rather than the exception [125]. In this sense, small-bodied mammal species tend to have faster rates of molecular evolution than their larger relatives, mostly

because smaller mammals have a larger population size and more generations *per* unit time due a shorter lifespan, accumulating more DNA copy errors and evolving faster [124].

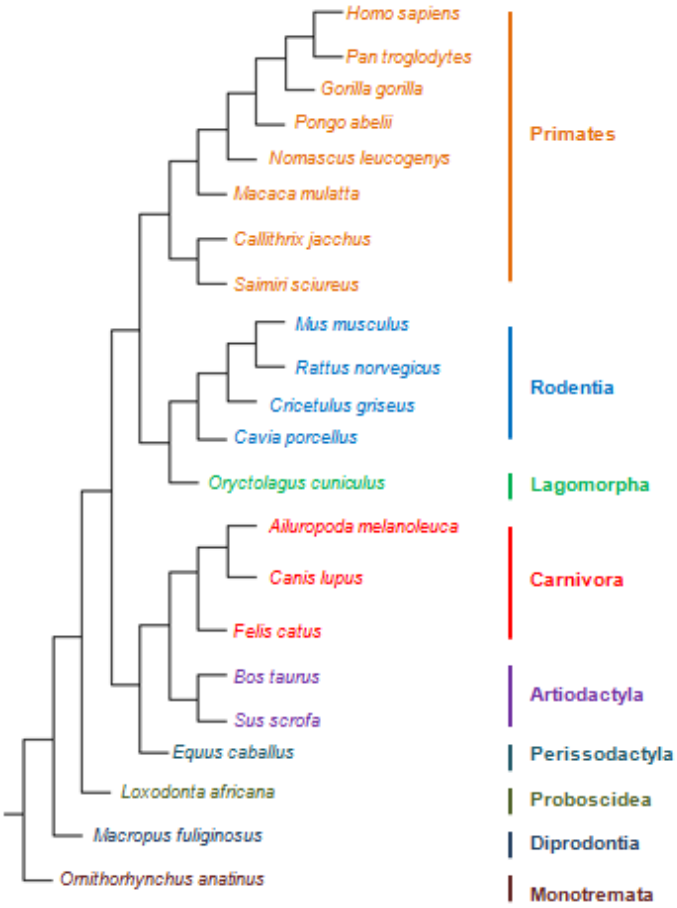

Figure I.13 - Representative tree of the placental mammal evolution (data from http://www.timetree.org/)

Nowadays, the evolutionary relatedness among groups of organisms might be studied by phylogenetic analysis, employing progressive alignment of nucleic acid and/or protein sequences of several organisms to describe their evolution using a model that consists of two components, fist a phylogenetic tree showing the inferred evolutionary relationships among the species under study; second the description of the way individual sequences evolve by nucleotide or amino acid replacement along the branches of that tree [126,127]. These replacements are usually described as the products of chance mutation events, and their occurrence at each sequence site is mathematically modeled to produce the phylogenetic tree

[126]. Thus, any change across successive generations in the heritable characteristics of biological populations culminates in evolution and differentiation between species and might be reflected in protein structure, fold and function. Phenomenon such as fusion of duplicated domains and divergence through mutation could affect both the stability of protein folds and the requirements of protein function [128]. A recent example is the evolution of amylase gene, important for saliva function. Perry *et al.* [129] specifically looked at the effect of dietary starch on the number of copies of *AMY1* in the human genome, the gene that encodes the salivary amylase enzyme. They found that populations with high-starch diets have, on average, more *AMY1* copies than those with traditionally low-starch diets, which probably is associated with the improvement of the digestion of starchy foods [129]. This study allowed the recognition the environment's influence on the evolution of amylase, which certainly also affect other salivary proteins adapting them to the environment of different mammal species. Beyond amylase, other proteins present in saliva also exhibits great composition variation in nature, which my represent adaptations to dietary habits [19]. It is known that some characteristics such as pH are quite different between humans and other kind of mammals [1,20], but few alterations in the composition of salivary proteome among different species are known.

Despite the importance of human salivary peptides, in some ruminant species as sheep and goat, there is no evidence of the presence of cystatins, proline-rich proteins and histatins in parotid saliva proteome [19]. However, the saliva of ruminants contributes largely to the maintenance of the inorganic composition and the fluid volume of the rumen contents, important for maximum cellulose digestion [130]. So, the diversities of the digestive system and diet between ruminants and other mammals could explain some differences in their saliva composition. Nonetheless, it is thought the group of genes encoding statherin, histatins and some other proteins like casein, mapped on human chromosome 4 (in mouse on chromosome 5, in rat on chromosome 14, and in cow on chromosome 6), has been subjected to conserved chromosomal synteny during mammalian evolution, i.e., the co-localization of these genes was kept within the chromosomes of different mammalian species (Figure I.14) [131]. These

genes form a kind of secretory calcium-binding phosphoprotein gene cluster, being all of them associated with mineralized tissue [131]. Some of the genes clustered on this location are orthologous genes, e.g. genes present in different species originated by vertical descent from a single gene of the last common ancestor (Figure I.14). However, in cattle it was found a new gene which had about 30-50% amino acid sequence identity to human histatin and statherin, and since no direct ortholog was found in the human genome, the gene was named histatherin [132,133]. Studies about their role and activity suggest that histatherin is a ruminant-specific gene that plays a role in host defense in the oral cavity and milk in cattle [133]. Remnants of these genes are also present in mouse and rat at the genomic DNA level; however, no mouse transcripts have been identified yet [70]. Statherin was already identified in the pig parotid secretion, showing that this protein is not unique to Primates [134].

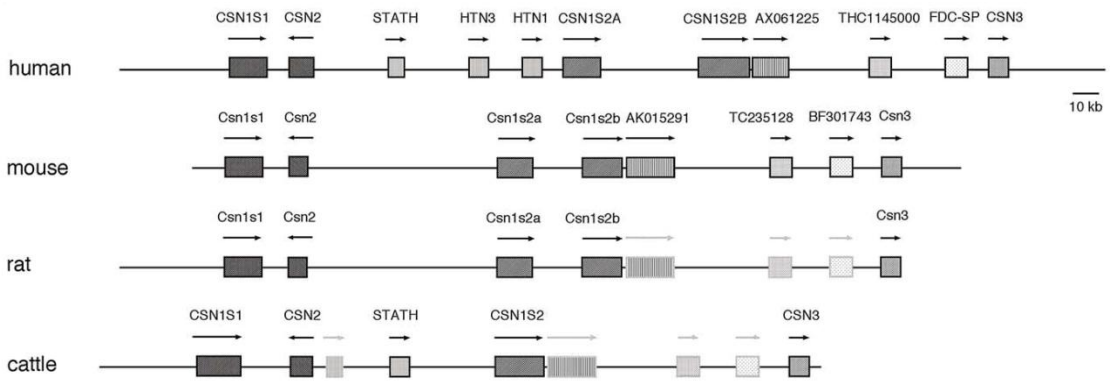

Figure I.14 - Overview of organization of the genes identified in the casein gene cluster region in human, mouse, rat, and cow. Orthologous genes are indicated by identical shading, genes whose presence is predicted based on comparative analysis but have not been verified by expression analyses or presence of matching sequences in the databases are depicted in light gray [70].

Regarding PRPs, the motifs rich in proline seem similar among mammals. In swine parotid gland there are several fragments of proline-rich peptides but the intact pro-protein was not found, suggesting that, as in humans, the cleavage occurs before granule storage [5]. These cleavage products are shorter than human peptides (12-66 amino acid residues) and present more sequential proline residues [5]. Moreover, the consensus sequences recognized by convertase are

very different from those in humans, suggesting new protease activities [5]. The distinction of basic and acidic PRPs loci is not possible in swine, but only the SP-A peptide was identified with acidic *pI*, being the remained basic fragments [5]. In other species like mice and rats the aPRPs are more similar to those in humans, keeping an initial acidic segment followed by several basic and proline residues [5]. However, in salivary glands of rats, mouse and hamster PRPs are normally not detected or are present in very low amounts, but they are greatly induced by dietary tannins, highlighting their importance in the neutralization of these and other polyphenols [135].

Currently, there are several nucleic acid and protein sequences available in databases such as Uniprot, NCBI and Ensembl, including information regarding the salivary peptidome, allowing the filogenetic analysis of this peptidome. Based on these databases and in published works, it was possible to establish in which mammal species the typical human salivary peptides in study were already identified.

As can be observed in Table I.2, which presents a list of the mammal species where salivary peptides have been identified based on databases such as Uniprot, NCBI and Ensembl (http://www.uniprot.org/; http://www.ncbi.nlm.nih.gov/; http://www.ensembl.org/index.html) or described in the literature, few studies have focused on the identification of salivary protein families in other mammals than humans being most of these peptide sequences assigned as predicted in databases like Swissprot and Trembl (http://www.ebi.ac.uk/uniprot/). Genomic and proteomic analyses will be crucial not only to confirm the existence of all predicted peptides on saliva of several mammal species but also their position on the phylogenetic tree allowing a better understanding of the salivary proteins taxonomy.

Table I.2- Peptides identified in some mammal species. Information based on [62,70,71,86,134,136-151] Ensembl, NCBI and Uniprot

Specie	S-type cystatins	Cystatin D	Cystatin C	Statherin	Histatins	PRPs	References
Homo sapiens	✓	✓	✓	✓	✓	✓	Saitoh *et al.* (1987); Freije (1991); Saitoh *et al.* (1989); Sabatini *et al.* (1990); Sabatini *et al.* (1993); Azen *et al.* (1996); Kim *et al.* (1993)
Macaca mulatta	✓		✓	✓		✓	Yang *et al.* (2011); Wei *et al.* (1996)
Macaca fascicularis	✓			✓	✓		Yang *et al.* (2011); Oppenheim *et al.* (1982); Xu *et al.* (1990)
Gorilla gorilla				✓ *	✓		Padovan *et al.* (unpublished)
Callithrix jacchus				✓ *			
Nomascus leucogenys				✓ *	✓		Padovan *et al.* (unpublished)
Pan troglodytes					✓ *		
Pongo abelii				✓ *	✓ *		
Trachypithecus cristatus					✓		Padovan *et al.* (unpublished)
Chlorocebus aethiops					✓		Padovan *et al.* (unpublished)
Cricetulus griseus	✓						Xu *et al.* (2011)
Rattus norvegicus	✓	✓	✓			✓ *	Cox and Shaw (1992); Esnard *et al.* (1990)
Mus musculus		✓	✓			✓	Frygelius *et al.* (2007); Huh *et al.* (1995); Lopez-Solis and Kemmerling (2005)
Sus scrofa				✓ *		✓	Zhang *et al.* (2005); Manconi *et al.* (2010)
Bos taurus				✓			Rijnkels *el al.* (2003)

* Sequences obtained from ENSEMBL

6. Aim

Saliva functions like protection against microorganisms, lubrication and mineralization of teeth are attributable, at least partially, to peptides families like statherin, histatins, cystatins and PRPs, and other antimicrobial peptides. However, very few is known about their presence, structure and function in other animal species. In this sense, this study will mainly focus on the salivary peptides of several mammal species. Using a proteomic approach, the salivary peptidome of some different mammals' species will be assessed in an attempt to know its composition and understand the differences when compared to human saliva. The interpretation of the data under a phylogenetic perspective will allow enlightening changes in the structure and distribution of these peptides between the different mammal species, under the influence of natural selection, genetic drift, mutation and gene flow, eventually explaining phenomena such as adaptation and speciation.

To fulfill the main goal two cases will be the target of this work:

a) in the first case, a phylogenetic analysis will be used to support the mass spectrometry search of β-thymosin peptide in different mammal species,

b) in the second case, the phylogenetic approach will be used to explain the absence of C, D and S-type cystatins identifications using proteomics approach in several mammals' saliva.

CHAPTER II – Thymosin β4 in mammals' saliva

II. Thymosin β4 in mammals' saliva

1. Introduction

Thymosins belong to a family composed by three distinct classes of low molecular weight proteins (*α*, *β*, and *γ*) isolated from calf thymus [109,152]. The 16 known members of the beta-thymosin class are highly conserved acidic peptides that bind G-actin and control the assembly and disassembly of actin filaments which regulate the dynamics of the cytoskeleton [110]. From these, thymosin β4 (Tβ4) is the most abundant isoform and probably the most active, being present in all cells, except for the erythrocytes, and in several body fluids such as plasma, tears and saliva [109-111]. This 5kDa peptide are composed by 43 amino acid residues and present several functions, such as influence in cell migration and differentiation, angiogenesis promotion in many tissues and regulation of metalloproteinase (MMP) activity [110-112].

Since Tβ4 lacks a signal peptide that allow its secretion, its presence in body fluids only is possible due to damaged cells, and its concentration in human saliva ranged from 0.2–3.6 μg/mL, being the gingival crevicular fluid the main contributor [111,115,153]. This biofluid, besides its importance in food maceration, has also an important role in protecting the surface of the oral cavity against chemical, mechanical and microbial attack [13]. In turn, the presence of Tβ4 in saliva could be related with the development of several cranio-facial organs, participating for example in the initiation, growth and differentiation of tooth germ, by promoting cell proliferation and differentiation [114,116]. Furthermore, the antimicrobial, anti-inflammatory and anti-apoptotic activity of Tβ4 in oral cavity cells has inferred its use as a therapeutic treatment of oral disorders, such as periodontal disease [117,153].

Despite the evident importance of Tβ4 in saliva, it has only been clearly reported in humans [153], remaining unknown its presence in other mammals saliva. Taking this in consideration, identification of Tβ4 in other species will contributes to clarify its importance in saliva and the mechanism behind its secretion.

2. Material and methods

2.1. Evolutionary analysis

NCBI (http://www.ncbi.nlm.nih.gov), Ensembl (http://www.ensembl.org), and SwissProt (http://www.ebi.ac.uk/uniprot/) databases were used iteratively for the collection of thymosin β4 sequences. The Basic Local Alignment Search Tool (BLAST) [154] was also used to complete the collection of amino acid sequences for thymosin β4 and multiple sequence alignments were performed using the BioEdit Sequence Alignment Editor [155].

2.2. Saliva collection

Dog (mongrel) and lamb saliva were collected during 4 minutes using salivettes (Sterile Saliva Collection Devices by Sarstedt) where the swab was placed underneath the animal's tongue. After the collection the salivette was centrifuged at 1000g for 2min for recover the saliva sample. For rat (Wistar) and rabbit (White New Zealand) saliva collection the animals were first injected with the sialagogue pilocarpine (4µmol/kg). All the samples were then centrifuged at 12000g for 10min at 4ºC to be collected the supernatant.

2.3. Peptide isolation and digestion

The total protein content of each sample was quantified by the DC protein assay (BioRad). 100µL of each sample was mixed with ammonium hydrogenocarbonate 25mM (1:1) and filtered in filter units of 50kDa (Vivaspin 500 - 50 kDa, Sartorius Stedim Biotech) by centrifugation at 10000g for about 20min at approximately 12ºC. The peptides present in the filtrate were then digested overnight with trypsin (12.5ng/mL) at 37ºC. The digested peptides were dried in a speedvac (SpeedVac Plus SC 210 A, Thermo Savant).

2.4. Peptide separation by LC

Dried peptides were ressuspended in solvent A, (water/acetonitrile/trifluoroacetic acid (98:2:0.05 v/v/v)) and ten microliters of each sample was separated using an Ultimate 3000 (Dionex) using a capillary column (Pepmap100 C18; 3 µm particle size; 0.75 µm internal diameter, 15 cm in length). A gradient of solvent A to solvent B (water/acetonitrile/trifluoroacetic acid (10:90:0.045, v/v/v)) was used. The separation was performed using a linear gradient (5-55 % B for 30 min, 55-80 % B for 10 min and 70-5% A for 5 min) with a flow rate of 0.3µL/min. The eluted peptides were applied directly on a MALDI plate in 15 sec fractions using an automatic fraction collector Probot (Dionex, Amsterdam).

2.5. Mass spectrometry

Mass spectra were obtained on a matrix-assisted laser desorption/ionization–time-of-flight MALDI-TOF/TOF mass spectrometer (4800 Proteomics Analyzer, Applied Biosystems, Foster City, CA, USA) in the positive ion reflector mode and in the mass range from 700-4500 Da with 800 laser shots. A data-dependent acquisition method was created to select the 16 most intense peaks in each sample spot for subsequent tandem mass spectrometry (MS/MS) data acquisition. GluFib (Glu-1-fibrinopeptide B) (*m/z* 1570.6) was used for internal calibration of the mass spectra.

MS/MS data was searched against the Swissprot, Trembl and NCBI protein databases for all species using paragon algorithm from ProteinPilot™ software (version 4.0, Applied Biosystems, USA) and Mascot software (v.2.1.0.4, Matrix Science Ltd, U.K.). An MS tolerance of 30 ppm was selected for precursor ions and 0.3 Da for fragment ions. Confidence levels upper to 99% were used as positive protein identification criteria. In order to estimate the false discovery rate (FDR) a reverse decoy database was created for all SwissProt resulting in 5% of FDR (false positive peptides/(false positive peptides + total peptides))*100. Unique peptides retrieved from FDR search were considered for analysis.

3. Results

3.1. Evolutionary analysis

Searching databases allowed to retrieve several amino acid sequences for Tβ4 accounting 52 entries in the Swiss-Prot/TrEMBL (http://www.uniprot.org/) and more than 100 sequences in NCBI database. Some of these primary amino acid sequences were related to species distributed in several mammalian orders such as Primates, Rodents, Carnivores, Lagomorphs, Artiodactyls and Diprotodonts. The alignment of several Tβ4 amino acid sequences from different mammal species allowed to recognize a high conservation region of this peptide among all the species (Figure II.1).

Figure II.1 - Amino acid sequence alignment of thymosin β4 from several species: *Canis lupus familiaris*(1) (XP_003432557); *Canis lupus familiaris*(2) (XP_003435498); *Homo sapiens* (HUMAN_TMSB4, P62328); *Pongo abelii* (PONAB_TMSB4, Q5R7H8); *Bos Taurus* (BOVIN_TMSB4, P62326); *Equus caballus* (HORSE_TMSB4, P62327); *Sus scrofa* (PIG_TMSB4, Q95274); *Rattus norvegicus* (RAT_TMSB4, P62329); *Mus musculus* (MOUSE_TMSB4, P20065-2); *Chinchilla villidera* (CHILA_TMSB4, Q6S9C5);*Oryctolagus cuniculus* (RABBIT_TMSB4, P34032); *Macropus eugenii* (MACEU_ TMSB4, Q7YRC3).

The high degree of conservation observed in Tβ4 does not allowed to infer the evolutionary history of this peptide in the mammalian class. However, the high similarity between distant mammal species allows predicting its existence in organisms where Tβ4 has not yet been identified.

3.2. Saliva analysis by LC-MS/MS

Saliva from mammal species including dog, rat, rabbit and lamb was collected for analysis by mass spectrometry. Thus, the low molecular weight fraction was obtained through the utilization of microfilters and analyzed under digestion with trypsin or without digestion. The analysis by LC-MS/MS of the undigested fraction did not allow the identification of any peptide belonging to Tβ4. However, when digested with trypsin, several peptides were identified. It should be pointed that these identifications were achieved by using an internal database where all amino acids sequences were grouped for phylogenetic analysis. For instance, the amino acid sequence of Tβ4 from lamb or sheep is not available in SwissProt whereas in case of dog, several Tβ4 amino acid sequences have been detected and assigned as predicted. Using the Mascot software with internal database containing all sequences retrieved from phylogenetic approach, four fragments (from 10 hypothetical) were assigned in dog saliva after tryptic digestion, presenting high similarity with human Tβ4 and mouse Tβ4 with a confidence of 100%. When Paragon algorithm was used, nine fragments were identified having high similarity with the swine Tβ4 (sp|Q95274|TYB4_PIG) with a confidence of 97.7% (Table II.1). The amino acid sequence of Tβ4 from pig (*Sus scrofa*) is 100% identical with one of the predicted amino acid sequences from dog (*Canis lupus familiaris* (1) – XP_003432557) found in the evolutionary analysis (Figure II.1). However, in the undigested samples the analysis LC-MS/MS not identified this peptide, showing no fragments caused by other proteolytic enzymes.

In addition, from those six fragments identified, four presented post-translational modifications such as methionine oxidation and serine acetylation. An example is shown in the MS/MS spectrum of figure II.2, where an acetylation at Ser-1 (plus 42.010565Da) and oxidation at Met-6 (plus 15.994915Da) were identified.

Table II.1 – Theorical mass values of tryptic fragments from predicted *Canis lupus familiaris* thymosin β4 (XP_003432557) and the mass values of tryptic fragments determined by LC-MALDI-MS/MS in dog saliva and identified as thymosin β4 peptide.

Fragment sequence	Position	Theorical tryptic peptides		Experimental tryptic peptides	
		Monoisotopic Mass $[M+H]^+$	Average Mass	Monoisotopic Mass $[M+H]^+$	Observations
KTETQEKNPLPSKETIEQEKQAGES	20-44	2829.4112	2831.0817	2829.435	
TETQEKNPLPSKETIEQEKQAGES	21-44	2701.3163	2702.9066	2701.336	
LKKTETQEKNPLPSKETIEQEK	18-39	2598.3985	2599.9594	2598.422	
MSDKPDMAEIEKFDKSKLKK	1-20	2368.2251	2369.8222	-	Not found
KTETQEKNPLPSKETIEQEK	20-39	2357.2195	2358.6238	2357.219	
MSDKPDMAEIEKFDKSKLK	1-19	2240.1301	2241.6470	-	Not found
TETQEKNPLPSKETIEQEK	21-39	2229.1245	2230.4486	2229.123	
MSDKPDMAEIEKFDKSK	1-17	1998.9511	2000.3113	-	Not found
NPLPSKETIEQEKQAGES	27-44	1984.9822	1986.1560	1984.939	
MSDKPDMAEIEKFDK	1-15	1783.8241	1785.0575	-	Not found
[Ac]SDKPDMAEIEKFDK	2-15	-	-	1694.7797	N-acetylserine
[Ac]SDKPDMAEIEKFDKSK	2-17	-	-	1909.9177	N-acetylserine
[Ac]SDKPDMAEIEKFDKSKLK	2-19	-	-	2151.1077	N-acetylserine
[Ac]SDKPDM[Oxi]AEIEKFDK	2-15	-	-	1710.7767	N-acetylserine; Met-oxidation
[Ac]SDKPDM[Oxi]AEIEKFDKSK	2-17	-	-	1925.9156	N-acetylserine; Met-oxidation

According the amino acid sequences of Tβ4 found in databases for the other species here in study (Figure II.1), it would be expected similar fragments in their saliva. However, none of the peptides found in dog saliva were identified in rat, rabbit or lamb saliva's.

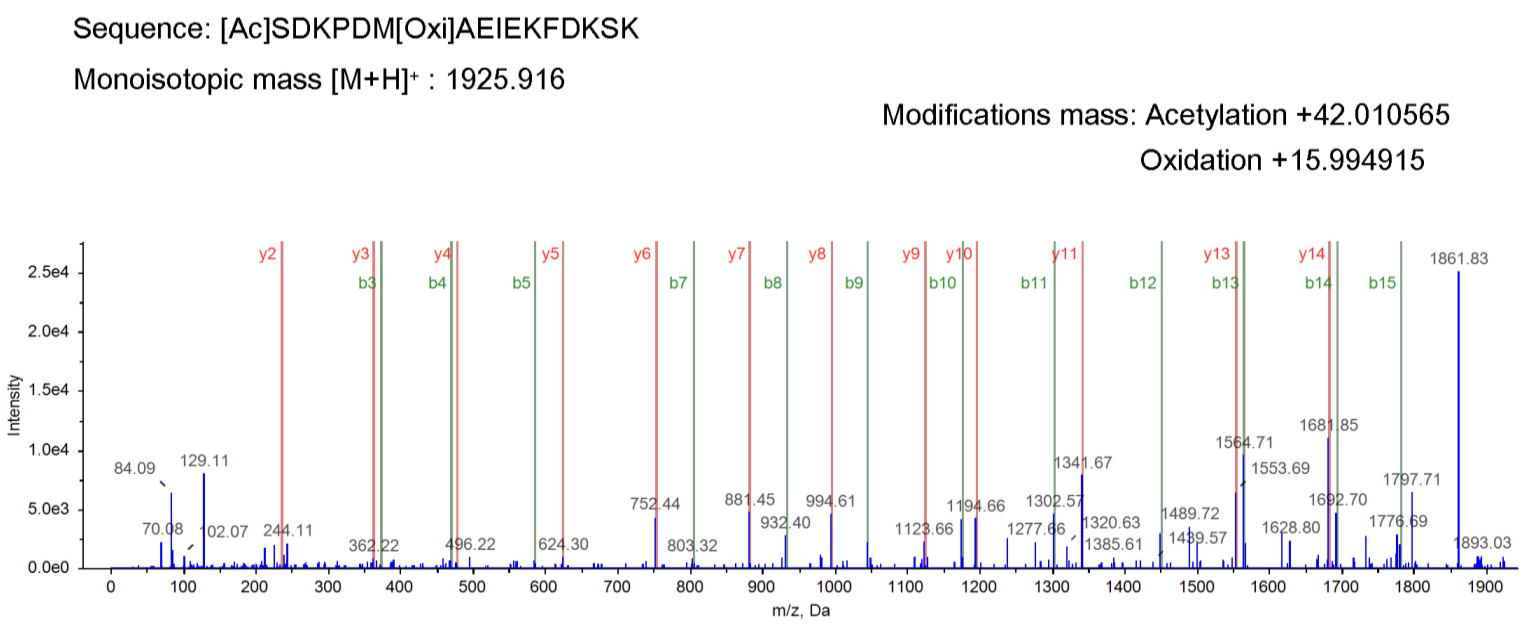

Figure II.2 – MS/MS spectra of the fragmentation of the ion with m/z= 1925.916 revealing the fragment [Ac]SDKPDM[Oxi]AEIEKFDKSK present in thymosin β4 of *Canis lupus familiaris*, where the first S residue and M residue presents a deviation of 42.010565Da and 15.994915Da corresponding to an acetylation and oxidation, respectively. The red peaks corresponding to the mass ion peaks from b series and the green peaks representing the mass ion peaks from y series.

4. Discussion

Thymosin β4 could be found in several members of the vertebrate phylum and is widely distributed throughout the mammalian tissues [156-158]. In mammals the gene that expresses Tβ4 is usually located at the X chromosome, although paralogues of this gene could be found in human and chimpanzee chromosome Y, and also in humans it was found the inclusion of pseudogenes in different chromosomes [159-161]. Moreover, the gene encoding Tβ4 displays high conservation among different species, not only by the high sequence homology but also by the similar arrangement of their genes, suggesting the importance of the structure conservation throughout the evolution of this gene [161]. With respect to saliva, this small peptide has only been found in humans [153], detected during the development of mouth's structures, such as the tooth and salivary glands in mouse and humans [114,116,162,163].

The high conservation observed for Tβ4, even regarding distant mammal species such as Primates and the Diprotodonts (order of marsupial mammals) that diverge at more than 160mya (http://www.timetree.org), reveals a negative or purifying selection pressure on the conserved amino acids of Tβ4. This suggests that it has maintained its multiple functions and is not involved in adaptive processes. However, its high conservation is useful for allowing the clear identification of this peptide even in mammal species where it has not yet been fully characterized. Since this peptide has already been identified in human saliva, its similarity among species suggested not only a functional conservation but also a similar distribution and so, it could be found in saliva of other species.

In the present study, we used the information retrieved from phylogenetic approach to perform the characterization of Tβ4 in saliva from mammals' species through LC-MS/MS analysis of tryptic fragments. Using different software and gathering amino acid sequences in an internal database, it was possible to identify Tβ4 but only in dog saliva. Originally, it was expected the presence of this peptide mainly in rat saliva, since it was already identified in the oral cavity of this animal. However no identification of Tβ4 was obtained in saliva from rat, rabbit or lamb.

Despite the fragments found in dog saliva being identified as Tβ4 are from human, mouse and pig, it is important to note that these three isoforms present 100% amino acid identity (Figure II.1). Although in databases there are several predicted amino acid and nucleic acid sequences of dog thymosin β4, most of the similar fragments found correspond to the gene product located at LOC100683370 on chromosome 1 (XM_003432509), whose protein product (XP_003432557) is also 100% identical to human, mouse and pig thymosin β4 peptides (Figure II.1). As can be observed in human thymosin β4, the peptide identified in dog saliva also lacks the initial methionine and appears acetylated at the N-terminal serine [164,165]. However, the influence of this modification in Tβ4 function has not been determined yet [164]. Furthermore, the two identified oxidized peptides evidenced the presence of thymosin β4 sulfoxide in dog saliva which is occasionally detected in human saliva [115]. About 5% of Tβ4 appears oxidized in human gingival crevicular fluid, having this sulfoxide specie a 20-fold lower affinity to G-actin which may regulate its actin-sequestering function [115,166]. In the present work the relative amount of thymosin β4 sulfoxide is not available, but the presence of Tβ4 and its sulfoxide in dog saliva shows the importance of both species to the actin binding mechanism.

Despite the different chromosomal location of Tβ4 gene identified in dog and in humans, as expected from phylogenetic analysis, the primary amino acid structure remained conserved throughout evolution. In fact, the entire molecule is involved in the functional activities carried out by this peptide, and distinct active sites play distinct functions: the segment 1-4 is important for anti-inflammation and angiogenesis, the segment 17-22 is crucial for actin binding and wound healing and the segment 1-15 confers cytotoxity protection [111]. Thus, being this amino acid sequence conserved in several mammal species, including Tβ4 from dog, it seems that similar function will be found in all of these species. In addition, it is worthy of note that no fragments were identified when LC-MS/MS analysis was performed without digestion, suggesting a high protection against proteolytic activity of saliva when compared to other salivary peptides such as histatins or PRPs that are the target of diverse salivary proteases, namely, kallikreins, cathepsins or metalloprotease, originating a large number of fragments [4,34].

Furthermore, the cells and mechanism responsible for Tβ4 release in saliva are unknown. However, it is thought that this peptide has its main origin in the gingival crevicular fluid and its presence is probably connected to a role in the regulation of the junctional epithelium angiogenesis [115,167].

The elucidation of Tβ4 function in saliva and the mechanism behind its secretion is crucial for its application as a therapeutic treatment for oral disorders, such as periodontal disease, mainly due its antimicrobial, anti-inflammatory and anti-apoptotic activity [117,153]. Besides that, Tβ4 presents cytoprotective properties that could protect gingival fibroblasts from apoptosis induced by chlorhexidine, the compound currently used as bacteriostatic and bactericidal agent in gingivitis and periodontitis management [117]. Thus, further studies are necessary to elucidate the importance of the presence of Tβ4 in this body fluid, and to understand why this peptide is only found in human and, now, in dog saliva.

CHAPTER III – Evolution of C, D and S-type cystatins in mammals

III. Evolution of C, D and S-type cystatins in mammals

1. Introduction

Cystatins belong to a superfamily of low molecular weight proteins that usually act as inhibitors of cysteine peptidases (CPs) and are frequently involved in the normal protein turnover of intracellular and endocytosed proteins. Four main cystatin families are known, including type-II cystatins, that are characterized as proteins with 120-125 amino acid residues and two disulfide bonds [43]. Some proteins of this group are expressed in human saliva, namely, cystatin C, cystatin D and S-type cystatins (S, SN, and SA), which are encoded by the genes *CST3*, *CST5*, *CST4*, *CST1* and *CST2* (Table III.1), respectively, located on human chromosome 20 at the *CST1-5* locus [23,26,43].

Table III.1 - Genes of "salivary cystatins" and their respective proteins in humans

Genes	Proteins
CST1	Cystatin SN
CST2	Cystatin SA
CST3	Cystatin C
CST4	Cystatin S
CST5	Cystatin D

Cystatin C-like proteins could be found in several vertebrates and evolutionary studies focused on this superfamily suggested that the ancestral cystatin C gave rise to the Primate-specific subfamily of salivary cystatins (cystatin D, S, SA and SN) [43,47].

This protein family performs functions such as direct inhibition of endogenous and exogenous CPs, modulation of the immune system, antimicrobial activity and maintenance of the tooth surface [43]. However, they do not all present the same activity, i.e. D and S-type cystatins are poorer inhibitors of papain-like CPs when compared to cystatin C and cystatin C is less effective in the

bactericidal activity than cystatin S [43,51,168]. Indeed, there is a correlation between the activity of each cystatin and their evolution since successive generations of salivary cystatins are progressively less active against the host lysosomal cathepsins B, H, and L [43]. Additionally, human cystatin S presents relatively poor inhibitory activity and high affinity to the tooth surface, which suggests a main role in the maintenance of the mineralized surfaces [43].

It was already suggested that D and S-type cystatins evolved in Primates to protect the oral cavity from dietary and environmental CPs presenting poor inhibition of endogenous CPs [43]. However the evolution of these proteins in Primates is not clear, making essential the understanding of their evolution to better interpret their function in the oral cavity. Furthermore, despite some previous studies reporting D and S-type cystatins as Primate specific, in databases such as Swissprot or NCBI, sequences from non-Primate species annotated as cystatin S or cystatin D can be found. A good example is the rat cystatin S, although not clear whether if this cystatin is specific to Primates or if it can also be found in other mammals, particularly in its saliva.

Thus, it is important to understand the distribution of this specific group of proteins in saliva from different mammal species, using proteomic and evolutionary methods together to accomplish this aim.

2. Material and methods

2.1. Proteomic analysis

2.1.1. Saliva collection and preparation

Dog (mongrel) and lamb saliva were collected during 4 minutes using salivettes (Sterile Saliva Collection Devices by Sarstedt) where the swab was placed underneath the animal's tongue. After the collection the salivette was centrifuged at 1000g for 2min to recover the saliva sample. For rat (Wistar) and rabbit (White New Zealand) saliva collection the animals were first injected with

the sialagogue pilocarpine (4µmol/kg). All the samples were then centrifuged at 12000g for 10min at 4ºC to collect the supernatant.

2.1.2. SDS-PAGE protein separation

Proteins present in saliva samples were separated by gel electrophoresis under denaturing conditions to be analysed by western blotting and by LC-MS/MS. For this, a 12,5% SDS-PAGE gel was made according to Laemmli [169] to separate the proteins by their molecular weight. The gel was then fixed in 40% methanol/10% acetic acid for 1 hour and stained with colloidal coomassie blue overnight. Finally, the gel was distained with 25% methanol until the background color was removed.

2.1.3. In-gel protein digestion

For mass spectrometry analysis, protein bands located around the 15kDa area were excised to be further digested. To the in-gel digestion, the gel sections were washed twice with 50mM ammonium bicarbonate and 100% acetonitrile and washed with 100% acetonitrile. After the gel pieces were dried in vacuum, 25µl of 10µg/ml trypsin (diluted in 25mM ammonium bicarbonate) was added and after one hour, the gel pieces were covered with 50mM ammonium bicarbonate. The samples were incubated overnight at 37ºC and the tryptic peptides were then extracted from the gel pieces with 10% (v/v) formic acid. These fragments were finally dried in vacuum and ressuspended in 20µl of 50% acetonitrile and 0.1% formic acid.

2.1.4. Determination of the presence of cystatin S in saliva by Wetern-blotting

After the protein separation by SDS-PAGE the proteins were transferred to a nitrocellulose membrane (Whatman®, Protan®) in a buffer with 25mM Tris, 192mM Glycine with a pH 8,3 and 20% methanol during 2hours at 200mA. The membrane was then blocked with 5% (w/v) dry non-fat milk in TBS-T (buffer with

20mM Tris at pH 7,5 and 150mM NaCl and 0,1% Tween 20) for 1 hour and incubated overnight at 4ºC with a solution of anti-cystatin S antibody (rabbit monoclonal (Ab58515 Abcam) in a dilution 1:500 in blocking solution. The membrane was washed three times during 10minutes with TBS-T and was incubated with an anti-rabbit antibody (F9887 Sigma) in a dilution of 1:1000 during 1 hour. The membrane was washed again with TBS-T and detection was made with enhanced chemiluminescence (ECL-Plus, Amersham Pharmacia) and exposed to a photographic film (Kodak Sigma). The film was revealed and fixed.

2.1.5. Peptide isolation and digestion

The total protein content of each sample was quantified by the DC protein assay (BioRad). 100µL of each sample was mixed with ammonium hydrogenocarbonate (1:1) and filtered in filter units of 50kDa (Vivaspin 500 - 50 kDa, Sartorius Stedim Biotech) by centrifugation at 10000g for about 20min at approximately 12ºC. The peptides present in the filtrate were then digested overnight with trypsin (12.5ng/mL) at 37ºC. The digested peptides were dried in a speedvac (SpeedVac Plus SC 210 A, Thermo Savant).

2.1.6. Peptide separation by LC

Both the filtrated samples and the samples separated by gel electrophoresis, after obtaining the tryptic fragments by digestion, were separated by LC. For this, dried peptides were ressuspended in solvent A, (water/acetonitrile/trifluoroacetic acid (98:2:0.05 v/v/v)) and ten microliters of each sample was separated using an Ultimate 3000 (Dionex) using a capillary column (Pepmap100 C18; 3 µm particle size; 0.75 µm internal diameter, 15 cm in length). A gradient of solvent A to solvent B (water/acetonitrile/trifluoroacetic acid (10:90:0.045, v/v/v)) was used. The separation was performed using a linear gradient (5-55 % B for 30 min, 55-80 % B for 10 min and 70-5% A for 5 min) with a flow rate of 0.3µL/min. The eluted peptides were applied directly on a MALDI plate

in 15 sec fractions using an automatic fraction collector Probot (Dionex, Amsterdam).

2.1.7. Mass spectrometry analysis

Mass spectra were obtained on a matrix-assisted laser desorption/ionization–time-of-flight MALDI-TOF/TOF mass spectrometer (4800 Proteomics Analyzer, Applied Biosystems, Foster City, CA, USA) in the positive ion reflector mode and in the mass range from 700-4500 Da with 800 laser shots. A data-dependent acquisition method was created to select the 16 most intense peaks in each sample spot for subsequent tandem mass spectrometry (MS/MS) data acquisition. GluFib (Glu-1-fibrinopeptide B) (*m/z* 1570.6) was used for internal calibration of the mass spectra in the analysis made in fragments originated from the filtration, and trypsin autolysis peaks were used for internal calibration of the mass spectra from the fragments originated from the in-gel digestion, allowing a mass accuracy of more than 25ppm.

MS/MS data was searched against the Swissprot, Trembl and NCBI protein databases for all species using paragon algorithm from ProteinPilot™ software (version 4.0, Applied Biosystems, USA) and Mascot software (v.2.1.0.4, Matrix Science Ltd, U.K.). An MS tolerance of 30 ppm was selected for precursor ions and 0.3 Da for fragment ions. Confidence levels upper to 99% were used as positive protein identification criteria. In order to estimate the false discovery rate (FDR) a reverse decoy database was created for all SwissProt resulting in 5% of FDR (false positive peptides/(false positive peptides + total peptides))*100. Unique peptides retrieved from FDR search were considered for analysis.

2.2. Evolutionary analysis

2.2.1. Cystatins sequences collection

For the search of nucleotide sequences from cystatins C, D, S, SA and SN from several mammals the NCBI (http://www.ncbi.nlm.nih.gov) and Ensembl

(http://www.ensembl.org) databases were used by searching iteratively for the different query names (example: “cystatin S”, “CST4”, etc). Basic Local Alignment Search Tool (BLAST) [154] was also used to complete the collection of amino acid sequences for C, D and S-type cystatins (S,SA and SN). The accession numbers of the cystatin sequences retrieved are shown in Table III.2.

Table III.2 – Accession numbers from the sequences found in databases of cystatins C, D, S, SA and SN and their chromosomal localization.

	Species (common name)	Accession number	Chromosome
Cystatin C	*Canis lupus familiaris* (dog)	XR_134509	23
	Felis catus (domestic cat)	AB562976	Unknown
	Loxodonta africana (elephant)	XM_003411620	Unknown
	Bos taurus (cattle)	NM_174029	13
	Sus scrofa (pig)	NM_001044602	17
	Ailuropoda melanoleuca (panda)	XM_002931133	Unknown
	Nomascus leucogenys (gibbon)	XM_003280847	Unknown
	Callithrix jacchus (marmoset)	XM_002747514	5
	Pongo abelli (orangutan)	ENSPPYT00000012515	20
	Saimiri sciureus (squirrel monkey)	U52028	Unknown
	Homo sapiens (human)	NM_000099	20
	Macaca mulatta (rhesus monkey)	U51912	10
	Oryctolagus cuniculus (rabbit)	NM_001082706	4
	Rattus norvegicus (rat)	NM_012837	3
	Mus musculus (mouse)	NM_009976	2
	Cavia porcellus (guinea pig)	XM_003476403	Unknown
	Cricetulus griseus (chinese hamster)	XM_003499974	Unknown
Cystatin D	*Gorilla gorilla*	ENSGGOT00000016631	20
	Nomascus leucogenys (gibbon)	XM_003278498	Unknown
	Callithrix jacchus (marmoset)	ENSCJAT00000001226	5
	Homo sapiens (human)	NM_001900	20
	Rattus norvegicus (rat)	NM_001108961	3
	Macaca mulatta (rhesus monkey)	XM_001097898	10
	Pan troglodytes (chimpanzee)	XM_001147326	20
Cystatin SN	*Homo sapiens* (human)	NM_001898	20
	Pan troglodytes (chimpanzee)	XM_001147668	20
	Gorilla gorila	ENSGGOT00000002801	20
	Nomascus leucogenys (gibbon)	XM_003278497	Unknown
	Pongo abelli (orangutan)1	XM_002830032	20
	Pongo abelli (orangutan)2	XM_002834542	20
	Pongo abelli (orangutan)3	XM_002834568	20

	Pongo abelli (orangutan)4	XM_002834995	Unknown
	Pongo abelli (orangutan)5	XM_002834557	20
	Pongo abelli (orangutan)6	XM_002834559	20
Cystatin S	*Homo sapiens* (human)	NM_001899	20
	Pan troglodytes (chimpanzee)	XM_514553	20
	Callithrix jacchus (marmoset)	XM_002747520	5
	Rattus norvegicus (rat)1	M75281	Unknown
	Rattus norvegicus (rat)2	XM_002726241	3
	Rattus norvegicus (rat)3	ENSRNOT00000055522	3
Cystatin SA	*Homo sapiens* (human)	NM_001322	20
	Pan troglodytes (chimpanzee)	XM_001147822	20
	Macaca mulatta (rhesus monkey)	XM_001097284	10
	Chicken cystatin	J05077	

2.2.2. Phylogenetic tree reconstruction

Multiple sequence alignments were performed in BioEdit Sequence Alignment Editor [155] by using Clustal W [170] and by manual alignment. The evolutionary analysis was performed with the program MEGA 4.0 [171] and the phylogenetic tree was constructed using the neighbor-joining (NJ) method [172], considering complete deletion for gaps and being the consistency of the tree branches assessed by 1000 bootstrap replicates. The nucleotide coding sequence of chicken cystatin was used as an outgroup.

3. Results

3.1. Proteomic analysis

The western-blotting technique was first used to detect the presence of cystatin S in saliva from human, rat, lamb, rabbit and dog samples. The results obtained were clear, since it was only detected the presence of cystatin S in the human saliva sample. Regarding the other saliva samples, the antibody used did not recognize the presence of this particular protein. However, the saliva analysis by tandem mass spectrometry allowed the identification of several proteins such as cystatins A, B and even cystatins M/E and 10. Two initial approaches were

used. On one hand, the salivary protein content was separated by gel electrophoresis and the bands detected near the 15kDa region were excised to be digested and the proteins identified. In the other approach, saliva samples were filtered to only obtain the proteins under 50kDa. These two strategies were used mainly to overcome the possibility of retention of the proteins of interest in the gel or in the filter, which would prevent their identification. However, the LC-MS/MS results only showed the presence of few type-II cystatins in saliva from rat, dog, rabbit and lamb (Table III.3). Still, this approach allowed identifying the presence of other type-II cystatins that were not expected in saliva and even to identify cystatins A and B (type-I cystatins) in dog and lamb saliva (Table III.3).

Table III.3 – Cystatins found by LC-MS/MS from dog, rat, lamb and rabbit saliva. The searches that allowed these identifications were made using the mascot software against the Swissprot database. The Swissprot accession numbers are indicated in the table.

	Cystatin-A	Cystatin-B	Cystatin-6	Cystatin-10	Cystatin-D
Dog	CYTA_HUMAN	CYTB_HUMAN	CYTM_HUMAN	-	-
Rat	-	-	-	D4AAU9_RAT	G3H702_CRIGR
Lamb	CYTA_BOVIN	CYTB_SHEEP			
Rabbit	-	-	-	D4AAU9_RAT	-

3.2. Evolutionary analysis

Saliva characterization from different mammal species by a proteomic approach did not allowed the identification of C and S-type cystatins in any of the species studied. However, in rat saliva, only cystatin-D was identified, which did not enable the understanding of their distribution in the mammalian class. In this sense, the phylogenetic analysis of these specific proteins could clarify the absence of these proteins in saliva of species other than humans and explain the ambiguity in their classification.

The collection of C, D and S-type cystatins from the databases showed that cystatins such as cystatin C had been already identified in several mammal species while the sequences assigned as D and S-type cystatins are almost exclusive to Primates (Table III.2). However, in many mammal species these proteins have not yet been identified and, in some cases, the sequences available

are incorrectly annotated or even barely assembled. In humans, the genes encoding these proteins are well characterized, located on the chromosome 20 in the locus p11.21 and separated between each other by less than 0,1Mb. Regarding other mammals species, the locus containing the genes that encode type-II cystatins are often located in a syntenic region, being surrounded by the *NXT1* and *CZF1* genes and having upstream the *ACSS1* gene (Figure III.1).

By looking for *CST3* gene it is possible to note that, among the genes here in study, this gene is the only one present in several genomes. However it has not been identified in the *Gorilla gorilla* and *Pan troglodytes* genomes. Regarding the dog genome, the conserved synteny is not observed because this region seems to be split between chromosomes 23 and 24 and, despite dog *CST3* being located on chromosome 23, its chromosomal position is not clear. According to the gene organization in the other mammal species, *CST3* is probably located upstream the *CST9* gene. Besides that, the genes on chromosome 24 are positioned in a reverse way, i.e., the transcriptional orientation has an inverse direction but the gene organization remains the same.

The other four genes analyzed are present mainly in Primate genomes. The chromosomal location of *CST3*, *CST1* and *CST5* genes showed that *CST3* and *CST5* genes are separated by ~200kb, while CST1 is located between ~100Kb apart from both. Regarding the S-type cystatins (*CST1*, *CST2* and *CST4* genes), in the human genome they are organized in the order *CST4-CST2-CST1*, which could also be seen in the *Pan troglodytes* genome. However, while *CST1* gene may be retrieved from several Primate genomes, *CST2* and *CST4* genes are only found in *Homo sapiens* and *Pan troglodytes* genomes. In addition, there is also a gene assigned as *CST2* in the rhesus monkey genome. However, whereas in human and chimpanzee this gene is located downstream of the *CST1* gene, in rhesus monkey genome it was located upstream among the *CST9* and *CST3* genes. It is also important to note that in the rat genome there is one gene assigned as *CST5* and three genes annotated as *CST4* located ~1Mb apart from the *CST3* gene, but their chromosomal position is not consistent with that of Primates.

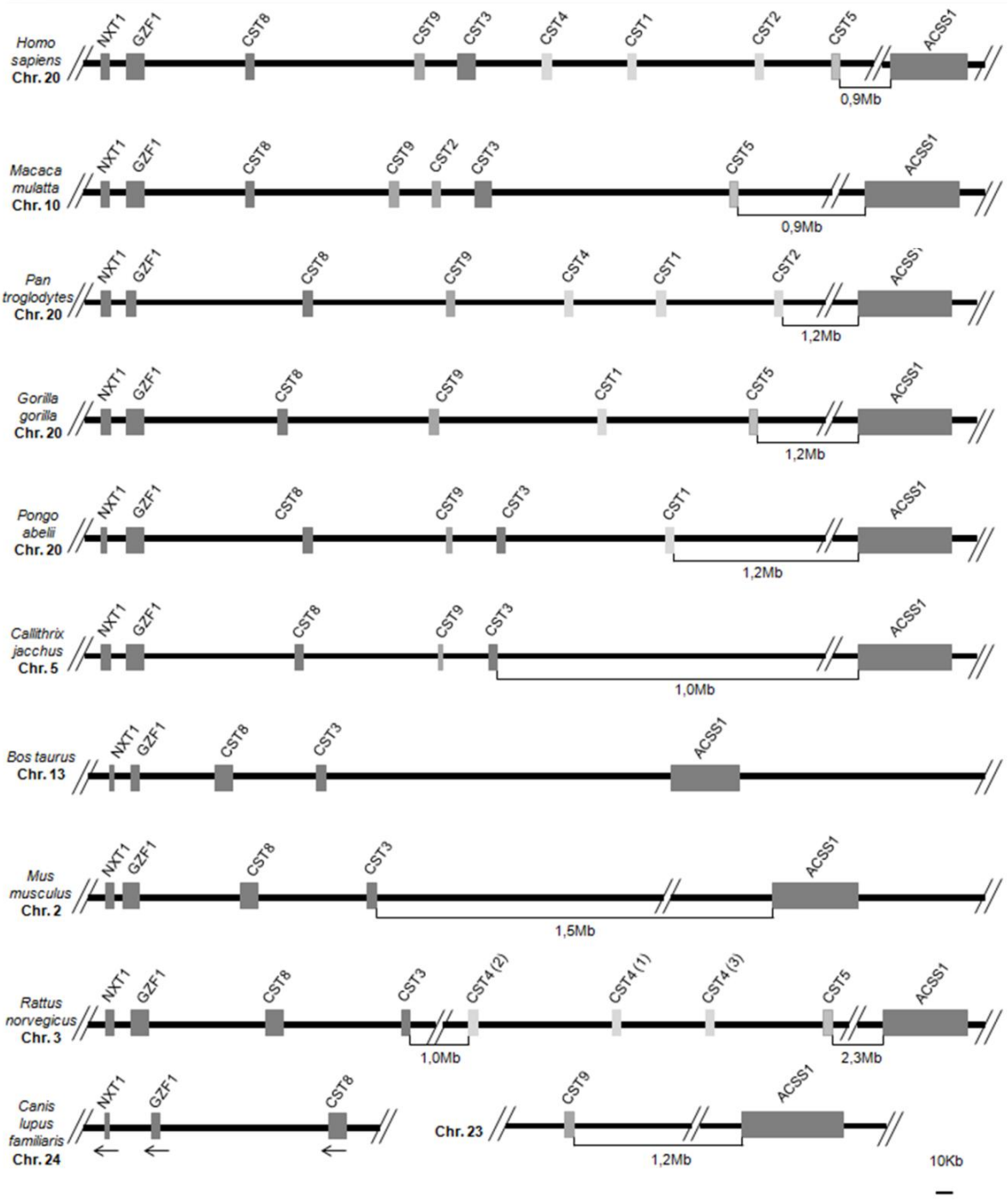

Figure III.1 – Type-II cystatin gene cluster region overview, showing the organization of the CST1-5 genes in human (*Homo sapiens*), Rhesus monkey (*Macaca mulatta*), Chimpanzee (*Pan troglodytes*), Gorilla (*Gorilla gorilla*) Orangutan (*Pongo abelii*), Marmoset (*Callithrix jacchus*), Cattle (*Bos taurus*), mouse (*Mus musculus*) and rat (*Rattus norvegicus*). Orthologous genes are indicated by identical shading. The arrows indicate the reverse position and transcriptional orientation of the genes (data from Ensembl).

The alignment of the nucleotide sequences from C, D and S-type cystatins allowed the construction of a phylogenetic tree aiming the understanding of phylogenetic relationships between these type-II cystatins in mammals. The presented tree is a rooted neighbor-joining tree containing mammal sequences of C, D and S-type cystatins (Figure III.2). To better understand the positioning of the cystatins' sequences in the branches of the tree, the alignment of their amino acid sequences is shown (Figure III.3) allowing the detection of differences in the amino acid composition that could be relevant to their functional role.

Looking to the amino acid sequences presented in Figure III.3, it is clearly visible amino acid segments that allow distinguishing the different cystatins. Indeed, *CST5* sequences present several typical amino acid patterns relatively to *CST3* sequences and some of them correspond to basic to acidic amino acid substitutions (for example RKA residues in *CST3* and EEE residues in *CST5*). *CST1*, *CST2* and *CST4* exhibit some amino acid substitutions in common, being difficult to identify differences between the sequences that allow their division into the different cystatin types. From the amino acid alignment it is also possible to observe the highly divergent rat cystatins (*CST4* and *CST5*) and *Cricetulus grisues CST3*, being substantially different from the other *CST4*, *CST5* and *CST3* sequences. Besides that, there are also several amino acid substitutions in the signal peptide regions which could influence the mechanism of their secretion. In addition, the signal peptide from some cystatin sequences is not known and can vary among different animal species, for example, the signal peptide from human *CST3* is composed by the first 26 amino acid residues, while the signal peptide of rat and rabbit *CST3* are composed by the first 20 and 28 amino acid residues, respectively.

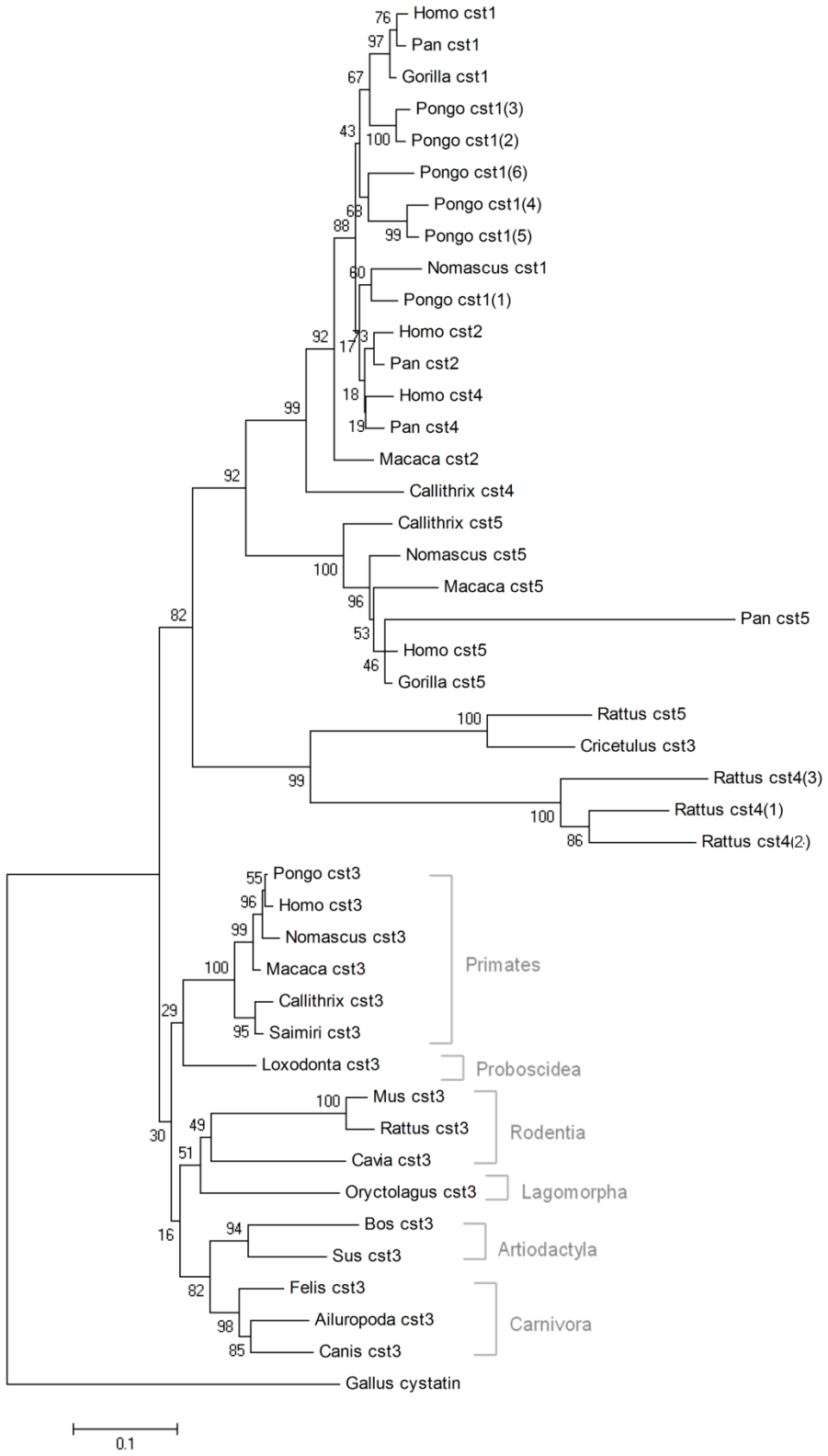

Figure III.2 – Rooted neighbor-joining tree representing the bootstrap consensus following 1000 replicates. The nucleotide sequences were obtained from ENSEMBL and NCBI databases and their respective accession numbers are reported in table III.2. In the tree there is also the indication of the mammalian classes associated with the *CST3* clusters.

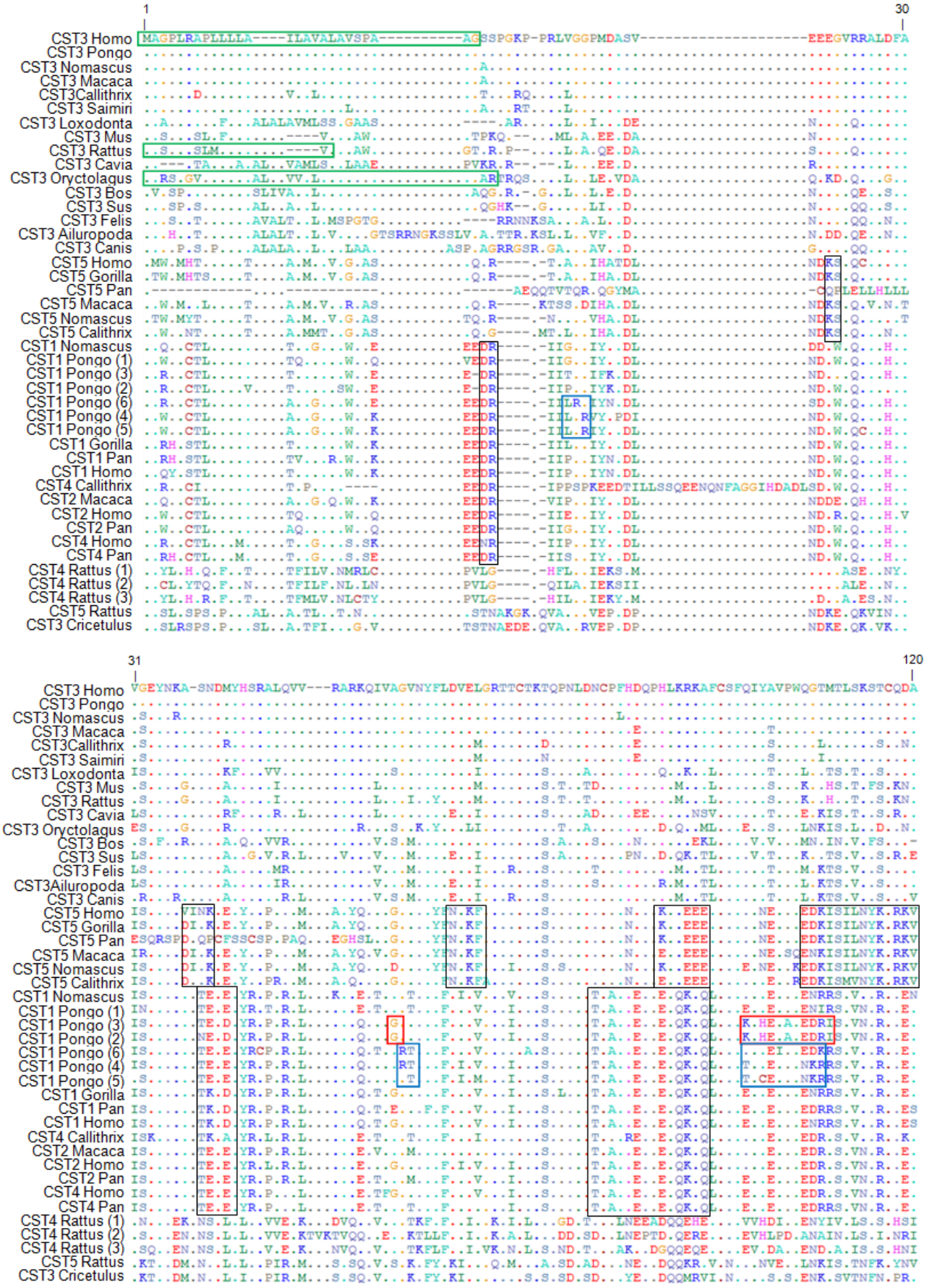

Figure III.3 – Amino acid composition of the cystatins used for the phylogenetic tree reconstruction. The boxed amino acids indicate the main amino acid substitutions in the different cystatins; red and blue boxes indicate amino acid differences that divide the *CST1* (cystatin SN) copies from *Pongo abelii* in two main groups; green boxes indicated the signal peptide.

Several different copies of *CST1* gene were detected for *Pongo abelii.* These copies appear distributed in different clusters in the phylogenetic tree (Figure III.2). All the sequences are from genes located on chromosome 20 of *Pongo abelii*, except for *CST1* (4) (XM_002834995) for which the chromosomal location remains unknown. The copy *CST1* (1) (XM_002830032) is located near other type-II cystatins in *Pongo abelli* at chromosome 20 whereas the other genes are located downstream in the same chromosome at the telomeric region (Figure III.4). These sequences show several amino acid substitutions suggesting that they are not the result of a single duplication from the ancestral gene (Figure III.3).

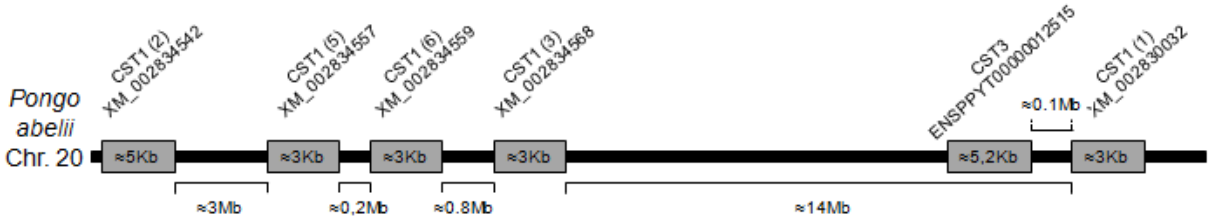

Figure III.4 – Chromosomal location of *CST1* and *CST3* of *Pongo abelli* in Chromosome 20 (data from Ensembl).

The phylogenetic tree also revealed an unexpected position of the branch that includes the sequences of rat cystatin S (*CST4*), rat cystatin D (*CST5*) and cystatin C (*CST3*) from *Cricetulus griseus*. Moreover, the pairwise distances, i.e. the path-distance between taxa on a phylogram, revealed large distances between these sequences and the other cystatin sequences (Table III.4). With regard to the sequences of rat cystatin S (*CST4*) the genetic distances between these sequences and the human cystatin S are higher than the distance to the chicken cystatin, the sequence used as an outgroup. Additionally, the search with Blastn [154] with these sequences only identified with high confidence proteins also assigned as cystatin S of rat and the alignment made with other proteins, even other type of cystatins, showed query coverage under 50% and a percentage of identity below 70%.

Table III.4 – Pairwise distances between human cystatins and cystatin S and cystatin D from Rattus norvegicus, cystatin C from Cricetulus griseus and chicken cystatin. The distance values of Rodent genes corresponding to genes in the human genome are highlighted in grey.

	Homo cst4	*Homo cst5*	*Homo cst3*
Chicken cystatin	0,683	0,731	0,612
Ratus cst4 (1)	0,713	0,619	0,580
Ratus cst4 (2)	0,720	0,657	0,605
Ratus cst4 (3)	0,711	0,639	0,641
Rattus cst5	0,640	0,560	0,529
Cricetulus cst3	0,602	0,531	0,522

4. Discussion

Despite the importance of cystatins in the protection of the oral cavity, in some mammal species there is no evidence of their presence in saliva. By a proteomic approach of rat, dog, rabbit and lamb saliva only the presence of cystatin D in rat was confirmed. None of the other type-II cystatins here in study were found. Surprisingly, in dog and lamb saliva the presence of cystatins A and B, also known as stefins or type-I cystatins, was observed. These cystatins have been already identified in human saliva from healthy adult individuals [26], although more commonly found in higher concentrations in infant saliva being the main cystatins in saliva at this age [42]. The presence of type-I cystatins instead type-II cystatins in dog and lamb saliva could indicate a higher importance of the intracellular inhibition of papain and cathepsins L, S and H and the development of squamous epithelia in the oral cavity of these mammals since this involves the action of cystatin A and B [42,173]. Yet, it is not clear why these cystatins are only found in dog and lamb saliva and are not found in the other saliva's in study. Nevertheless, by looking at the evolution of mammals (Figure I.13), the dog (Carnivora order) and sheep (Artiodactyla order) species are closely related, which could suggest that during the evolution these groups suffered alterations in these cystatins that allowed their secretion in saliva. Yet in dog saliva, the presence of cystatin-6, a type-II cystatin also known as cystatin-M/E, was identified. This cystatin, besides being a glycoprotein, only shares 35% homology with the other

type-II cystatins [174] and is overall expressed in the cutaneous epithelia. However, due to its putative signal peptide it can be secreted and has already been found in human sweat [174]. In addition, small amounts of cystatin M/E may be detected in the esophagus, oropharyngeal tissues and lung [174], and its expression has already been verified in rat salivary glands [175]. Like other cystatins, this protein is involved in the inhibition of cysteine proteases, controlling mainly the activity of cathepsin V, cathepsin L and legumain [174]. As type-II cystatins are normally inhibitors of papain-like cystein proteases (C1 family), the ability of cystatin M/E to inhibit also legumain (family C13) [174] could reveal an adaptive strategy adopted by the dog to protect the oral cavity against the degradation caused by this protease.

Cystatin-10, another type-II cystatin was found in rat and rabbit saliva. The expression of this cystatin has already been observed in rat salivary glands being secreted mostly by the parotid gland [175], showing that this cystatin is also expressed in soft tissues and is not specific to cartilage as it was originally thought [176]. However, its function in the oral cavity is not yet clear. Additionally, it is important to refer that the approach used for rat and rabbit saliva collection involved the use pilocarpine, a sialagogue that stimulates salivation and may affect the pattern of protein secretion [177,178]. This could explain, for example, why cystatin-M/E, previously detected in rat salivary glands [175], was not identified in rat saliva, and why cystatin-10 was only found in rat and rabbit saliva.

Among the cystatins normally found in human saliva, only a protein assigned as cystatin D was identified in this study for rat saliva. The remaining cystatins, i.e., cystatin C and S-type cystatins were not identified in rat, rabbit, lamb or dog saliva's using the approach described in section 2.1 of this chapter, remaining to understand why these proteins are not present in saliva from the studied species. In this context, a phylogenetic interpretation of the proteomic results could clarify this question and allow understanding the distribution of these proteins among mammals.

The analysis of the chromosomal location of cystatins in several mammal genomes showed a shared synteny among Primates, Rodents and other genomes (Figure III.1) showing the same gene order in the chromosomes. By looking at the

genomes of these species it is also evident the presence of the *CST3* gene in almost of all of the analyzed genomes, which is consistent with the previous idea that *CST3* (cystatin C) is the most ancestral gene among the genes here in study. From those genomes analyzed, this gene is located upstream the *CST9* and *CST8* genes, other type-II cystatins genes also located in this chromosomal region. However, the synteny is not observed in the dog genome, being the cystatin locus split between two chromosomes. This could be the result of a chromosome translocation due the rearrangement of the genome, without the cystatin gene cluster being lost.

In the phylogenetic tree (Figure III.2), the cystatin C sequences of the different mammalian orders are clustered according to the molecular tree of placental mammals (Figure I.13), suggesting that the gene tree for this gene reflects the mammal evolution tree and allows to clearly distinguish the branches of Primates, Rodents, Carnivores, Artiodactyla and Lagomorphs. This supports previous studies that suggested cystatin C was at the origin of the other type-II cystatins [43,47], being the most ancestral cystatin here in study. Indeed, two of the main conserved domains in cystatins, the Q_{55}-X-V-X-G_{59} and P_{105}-W_{106} segments [47], are conserved in all the cystatin C sequences analyzed (Figure III.3). Thus, the absence of a high degree of differentiation between cystatin C proteins in the different mammal species may suggest that this protein has conserved its function. Although not being a major cystatin in the human saliva [22], cystatin C was not found in any of the animal saliva's analyzed by the proteomic approach. Thus, the mechanism behind its secretion that may explain its presence in human saliva but not in other mammal species namely rat, rabbit, dog and sheep, remain to be understood. However, it is important to note that the secretion of cystatins in saliva is dependent on their signal peptide. As can be seen from the Figure III.3, the signal peptides from cystatin C (*CST3*) sequences of the animal species analyzed by the proteomic approach are quite different when compared to the human signal peptide which might influence its secretion.

Regarding cystatin D (*CST5*), the phylogenetic tree showed a cluster comprising only D cystatins from Primates (Figure III.2). The other cystatin assigned as cystatin D from *Rattus norvegicus* is placed in a different branch

along with three other sequences annotated as rat cystatin S and Chinese hamster cystatin C, which may indicate a wrong annotation. Cystatin D seems to have had its origin before the Primate divergence, which occurred roughly at 85 mya [179], since it is present in species from Old and New World Monkeys. Cystatin SN (*CST1*) has emerged also only in Primates but presents a more complex evolution (Figure III.2). Indeed, in *Pongo abelii* several sequences of cystatin SN were found, which represent successive duplications of the gene *CST1* in this species. These sequences are distributed in different branches in the tree and according to their distribution and key amino acids in their composition (Figure III.3), the most ancestral cystatin SN in *Pongo abelii* is represented by *CST1* (1), which is clustered with *Nomascus CST1*. In turn, by duplication, this cystatin gene originated two distinct groups. One group is represented by *CST1* (2) and (3), and the other contains *CST1* (6), (5) and (4). The evolutionary path of these paralogous inferred from the tree could be further confirmed by taking into account their chromosomal location (Figure III.4). In addition, these gene copies are located in the telomeric region of the chromosome 20. Telomeric regions are regarded as having a greater tolerance for the incorporation of new genetic material without affecting the organism and that in these chromosomal regions a much greater length of time is required to delete duplicated segments because of their constrains for recombination [180]. This might explain the large number of copies of this gene and their maintenance in *Pongo abelii* genome.

Duplications in the telomeric regions are quite common also in the human genome and there are several examples described in the literature. Such as the *CST1* gene in *Pongo abelii*, the human *TUBB4Q* gene is also present in the 4q telomere and in ten more regions of the human genome, but only four of these paralogous are transcribed [181]. This behavior could also be shared with the *CST1* genes from *Pongo abelii*, but further studies are necessary to determine whether all copies of this gene are equally transcribed. Moreover, the multiple copies of *CST1* gene in the telomeric region of *Pongo abelii* chromosome 20 is not unheard, since, for example, in the human genome the *OR-A* gene present multiple copies in subtelomeric regions and could also vary its chromosomal location among humans [182]. These events could create gene families by

multiple duplications and modification processes which could allow the adaptation of the organisms to the environment.

In particular, in these duplicates, there are amino acid substitutions in the two characteristic cystatin conserved domains (QXVXG and PW segments), that may involve the acquisition of novel functions by these genes in *Pongo abelii*. This observation is in agreement with the postulated idea by Ohno (1970) about gene duplication [183], which state that the gene' copies are free to accumulate mutations and could subsequently lead to loss or gain of new functions, while at least one of the copies still performs the ancestral task [184,185]. Thus, these amino acid substitutions could reveal the acquisition of novel functions while the ancestral *CST1* gene maintains its original function, which could be also seen as a mechanism for the preservation of these copies in the genome. Although all of these *CST1* genes are from protein coding sequences, it is not clear if all of them are expressed and if the observed mutations can interfere with the cysteine protease inhibitor activity typical of cystatins.

With regard to cystatin SA (*CST2*) and cystatin S (*CST4*), they emerged only in *Homo sapiens* and *Pan troglodytes* and presents a high degree of similarity among each other, suggesting a common evolution even after its divergence into distinct genes. According to these data, these genes were originated at ~6,4mya, the average time for the emergence of these two species [186].

As can be seen from the data, there is also the presence of a *Macaca mulatta* cystatin SA (*CST2*) and a *Callithrix jacchus* cystatin S (*CST4*). However, due to the other observed results, such as its position in the phylogenetic tree, these nucleotide sequences may not be correctly assigned. Moreover, the information from the chromosomal location of the gene assigned as *CST2* in *Macaca mulatta* places it in an unusual region compared to that of the *CST2* genes from human and chimpanzee. According to their position in the tree these genes are more likely representatives of the cystatin SN (*CST1*) in these species. The same misclassification seems to have occurred in the sequences of cystatin S and cystatin D from *Rattus norvegicus* and cystatin C from *Cricetulus griseus*. Probably, these genes arose in Rodents before their divergence and nowadays can only be found in rat and Chinese hamster, since they may have been deleted

from the mouse genome. However, it is very unlikely they are type-II cystatins, because they present larger genetic distances to other type-II cystatins than to chicken cystatin that has diverged from mammals ~300mya (http://www.timetree.org). Besides, considering that cystatin S originated before the divergence of *Homo sapiens* and *Pan troglodytes*, it is unlikely to have appeared independently also in the rat genome. It is also important to refer that the sequence assigned as cystatin S from *Rattus norvegicus* was formerly known as LM protein, being latter assigned to cystatin S mostly due to similarities of the amino acid sequence with other type-II cystatins and due to its ability to inhibit some cysteine proteases [187]. Indeed, these Rodent cystatins present similar functional domains to that of the type-II cystatins (http://prosite.expasy.org/). However, human cystatin S presents poor inhibitory activity but have other functions in the oral protection. No studies exist regarding the antimicrobial activity and the hydroxyapatite binding ability of these proteins in Rodents. Also, the expression pattern of these proteins differs from that of human cystatin S since human cystatin S may be found in adult saliva and these Rodent proteins can only be found in younger animals or by stimulation with isoproterenol [43].

The interpretation of the data from the proteomic approach under an evolutionary perspective allowed clarifying the distribution of C, D and S-type cystatins in the saliva's analyzed. S-type and D cystatins could not be found in dog, lamb or rabbit saliva, since they are Primate specific, and the protein assigned as cystatin D (*CST5*) found in rat saliva, according to the phylogenetic tree, is a sequence misannotated, belonging to a group of Rodents specific proteins that although presenting similar segments to the cystatin domains they are not true C, D or S-type cystatin. Although similar to cystatins, the function of this rat protein in saliva is not clear. In addition, and especially for S-type cystatins, misannotation in the databases may be due to their high similarity and since they are the most recent proteins this impairs the acquisition of characteristics that allow their clear distinction.

Finally, it remains to explain why cystatin C (*CST3*) is not found in rat, rabbit, dog or lamb saliva. The phylogenetic tree shows the presence of this cystatin in several placental mammals genome and, since it could be found in

human saliva, its presence was expected in other mammal species. However, the mechanism behind cystatin C secretion remains to be understood being crucial to explain its presence in human saliva but not in the other mammals' saliva.

CHAPTER IV – General discussion and conclusion

IV. General discussion and conclusion

Nowadays, advances in large scale DNA sequencing technology allows sequencing the whole genomes of several organisms. The exhaustive analysis of these genomes gives us the ability to search for common genes in different species, enabling the sequences to be subsequently compared [188]. Phylogeny is, in that way, the basis for comparative genomic and allows understanding the genes under its evolutionary path [189]. However, little is known about the evolution of the characteristic peptides found in human saliva.

To know if these genes are expressed in the same way in saliva of other species besides man, the evaluation of mRNA presence in saliva could be performed by DNA microarrays or other methods [188]. However, due to the characteristics of this biofluid, searching for mRNA molecules would be a difficult task mainly because RNA is quickly and highly degraded [190]. Thus, the solution involves the use of proteomics approaches for the evaluation of these peptides presence in saliva from several species. Nevertheless, the analysis of proteins can also be complex mainly due the post-translational modifications and the diverse physiochemical characteristics of the peptides [188].

Mass spectrometry is usually the most used approach in proteomics, but it is directed for a panoramic figure of the sample, allowing to identify a large number of proteins simultaneously. Thus, it could not be the appropriate method to identify specific proteins among large numbers of other proteins that could even be expressed in the sample in higher amounts [188]. However, pre-separation of proteins by gel electrophoresis or by HPLC (high performance liquid chromatography), as performed in this work, allows to reduce sample complexity and achieve a better identification.

Phylogeny could help to understand which are the genes expressed in a specific sample and to optimize the proteomic approach used to clearly identify one protein from a specific group of proteins. In this work, such approach was used to identify thymosin β4 in animal's saliva. This peptide were previously identified in human saliva and the comparative analysis of the Tβ4 amino acid sequences from different mammal species allows to notice that this peptide is

highly conserved. Thus, it seems evident that Tβ4 could also be present in saliva from other species beyond humans, and so, the proteomic approach used was directed for the identification of this particular peptide.

On the other hand, an interpretation of the obtained data by mass spectrometry under a phylogenetic perspective could help to understand the distribution of a protein group in a specific kind of sample. This was useful to understand the data regarding C, D and S-type cystatins, which were found in human saliva but not all of them were found in saliva from the other species. Phylogenetic analysis allows concluding that D and S-type cystatins couldn't be identified in saliva from the analyzed species because they are Primate specific. Phylogenetic analysis performed in this work also enable to determine wrongly annotated nucleotide sequences, namely the sequences assigned as cystatin S and cystatin D from rat, that besides some similarity with other cystatins are obviously wrongly annotated in the databases. Therefore, it becomes evident the importance of the link between proteomics and phylogenetic approaches, which allows to target more effectively the study and understand under an evolutionary perspective the distribution of specific proteins in different organisms. The same has already been suggested by metagenomics and may explain, for example, if a given protein found for the first time in an organism represents the emergence of a new protein family or if it is the result of the divergence of known protein families [191]. Thereby, annotation errors would be avoided as those sometimes found in databases.

The use of these two aspects in the characterization of different mammals' saliva allowed to draw some conclusions, but also left some doubts:

- Thymosin β4 was in fact identified in other saliva beyond human. This peptide was found in dog saliva, but was not found in the other analyzed species, remaining to understand the mechanisms that lead to its expression in saliva.
- Type-I cystatins were identified in dog and lamb saliva, but not in the other species, and other type-II cystatins (beyond C, D and S-type cystatins) were also reported in rat and rabbit saliva. Not being saliva

specific, the presence of these cystatins over salivary cystatins is not clear.

- Cystatin C, although present in the genome of several mammals, was not identified in any saliva analyzed in this work. However, it is not a cystatin specific to saliva, but remains to understand why it could be found in human saliva and not in the others analyzed saliva's.
- Cystatin D and S-type cystatins are Primate specific and have probably emerged after the great mammalian radiation at 80-90 million years ago.
- Proteins wrongly annotated as cystatins S and D in rat remains incorrectly characterized, being its real function not known in the oral cavity of these mammals.
- Although samples from *Pongo abelii* were not analyzed in this work, it would be interesting to see if all genes of cystatin SN found by the phylogenetic analyses are expressed in this mammal saliva, and if their function has evoluted.

By all these reasons, the study of saliva from different mammal species using these two analytical strands, phylogeny and proteomics, may provide answers to this questions and may represent a way for attaining new knowledge about saliva.

CHAPTER V – References

V. References

[1] J.E. Lambert, Primate digestion: Interactions among anatomy, physiology, and feeding ecology, Evolutionary Anthropology 7 (1998) 8-20.
[2] C.K.J. Huq N. L., Ung M., Myroforidis H., Veith P. D., Chen D., Stanton D., He H., Ward B. R., Reynolds E. C., A review of the salivary proteome and peptidome and saliva-derived peptide therapeutics, International Journal of Peptide Research and Therapeutics 13 (2007) 547-564.
[3] R.G. Schipper, E. Silletti, M.H. Vingerhoeds, Saliva as research material: biochemical, physicochemical and practical aspects, Archives of Oral Biology 52 (2007) 1114-1135.
[4] L.M.J. Amado F., Domingues P.,Duarte J. A., Vitorino R., Salivary peptidomics, Expert Review of Proteomics 7 (2010) 709-721.
[5] C.T. Castagnola M., Vitali A.,Sanna M. T.,Messana I., Biotechnological implications of the salivary proteome, Trends in Biotechnology 29 (2011) 409-418.
[6] A.V. Amerongen, E.C. Veerman, Saliva--the defender of the oral cavity, Oral Dis 8 (2002) 12-22.
[7] S.U. Gorr, M. Abdolhosseini, Antimicrobial peptides and periodontal disease, J Clin Periodontol 38 Suppl 11 (2011) 126-141.
[8] C.N. Yeoman C. J., Yildirim S., Miller M., Kent A.,, L.S.R. Stumpf R., Nelson K. E., White B. A., Wilson B. A., Towards an Evolutionary Model of Animal-Associated Microbiomes, Entropy 13 (2011) 570-594.
[9] C. Llena-Puy, The role of saliva in maintaining oral health and as an aid to diagnosis, Med Oral Patol Oral Cir Bucal 11 (2006) E449-455.
[10] M. Castagnola, I. Messana, R. Inzitari, C. Fanali, T. Cabras, A. Morelli, A.M. Pecoraro, G. Neri, M.G. Torrioli, F. Gurrieri, Hypo-phosphorylation of salivary peptidome as a clue to the molecular pathogenesis of autism spectrum disorders, J Proteome Res 7 (2008) 5327-5332.
[11] G. Peluso, M. De Santis, R. Inzitari, C. Fanali, T. Cabras, I. Messana, M. Castagnola, G.F. Ferraccioli, Proteomic study of salivary peptides and proteins in patients with Sjogren's syndrome before and after pilocarpine treatment, Arthritis Rheum 56 (2007) 2216-2222.
[12] A.M. Contucci, R. Inzitari, S. Agostino, A. Vitali, A. Fiorita, T. Cabras, E. Scarano, I. Messana, Statherin levels in saliva of patients with precancerous and cancerous lesions of the oral cavity: a preliminary report, Oral Dis 11 (2005) 95-99.
[13] L.C. Schenkels, E.C. Veerman, A.V. Nieuw Amerongen, Biochemical composition of human saliva in relation to other mucosal fluids, Critical Reviews in Oral Biology & Medicine 6 (1995) 161-175.
[14] S.A. Rayment, B. Liu, G.D. Offner, F.G. Oppenheim, R.F. Troxler, Immunoquantification of human salivary mucins MG1 and MG2 in stimulated whole saliva: factors influencing mucin levels, J Dent Res 79 (2000) 1765-1772.
[15] A. Zalewska, K. Zwierz, K. Zolkowski, A. Gindzienski, Structure and biosynthesis of human salivary mucins, Acta Biochim Pol 47 (2000) 1067-1079.
[16] E.N. Baker, H.M. Baker, Molecular structure, binding properties and dynamics of lactoferrin, Cell Mol Life Sci 62 (2005) 2531-2539.
[17] Y.R. Yamada N., Suzuki E., The current status and future prospects of the salivary proteome, Journal of Health Science (2009) 682-688.
[18] Y.M. Henskens, E.C. Veerman, M.S. Mantel, U. van der Velden, A.V. Nieuw Amerongen, Cystatins S and C in human whole saliva and in glandular salivas in periodontal health and disease, J Dent Res 73 (1994) 1606-1614.

[19] E. Lamy, G. da Costa, R. Santos, E.S.F. Capela, J. Potes, A. Pereira, A.V. Coelho, E. Sales Baptista, Sheep and goat saliva proteome analysis: a useful tool for ingestive behavior research?, Physiol Behav 98 (2009) 393-401.
[20] E.I. McDougall, Studies on ruminant saliva. 1. The composition and output of sheep's saliva, Biochem J 43 (1948) 99-109.
[21] V. de Almeida Pdel, A.M. Gregio, M.A. Machado, A.A. de Lima, L.R. Azevedo, Saliva composition and functions: a comprehensive review, J Contemp Dent Pract 9 (2008) 72-80.
[22] E. Scarano, A. Fiorita, P.M. Picciotti, G.C. Passali, L. Calo, T. Cabras, R. Inzitari, C. Fanali, I. Messana, M. Castagnola, G. Paludetti, Proteomics of saliva: personal experience, Acta Otorhinolaryngol Ital 30 (2010) 125-130.
[23] I.R. Messana I., Fanali C., Cabras T., Castagnola M., Facts and artifacts in proteomics of body fluids. What proteomics of saliva is telling us?, Journal of Separation Science 31 (2008) 1948-1963.
[24] S. Bandhakavi, M.D. Stone, G. Onsongo, S.K. Van Riper, T.J. Griffin, A dynamic range compression and three-dimensional peptide fractionation analysis platform expands proteome coverage and the diagnostic potential of whole saliva, J Proteome Res 8 (2009) 5590-5600.
[25] T. Cabras, E. Pisano, R. Boi, A. Olianas, B. Manconi, R. Inzitari, C. Fanali, B. Giardina, M. Castagnola, I. Messana, Age-dependent modifications of the human salivary secretory protein complex, J Proteome Res 8 (2009) 4126-4134.
[26] R. Vitorino, M.J. Lobo, A.J. Ferrer-Correira, J.R. Dubin, K.B. Tomer, P.M. Domingues, F.M. Amado, Identification of human whole saliva protein components using proteomics, Proteomics 4 (2004) 1109-1115.
[27] Y. Yao, E.A. Berg, C.E. Costello, R.F. Troxler, F.G. Oppenheim, Identification of protein components in human acquired enamel pellicle and whole saliva using novel proteomics approaches, J Biol Chem 278 (2003) 5300-5308.
[28] S. Hu, Y. Xie, P. Ramachandran, R.R. Ogorzalek Loo, Y. Li, J.A. Loo, D.T. Wong, Large-scale identification of proteins in human salivary proteome by liquid chromatography/mass spectrometry and two-dimensional gel electrophoresis-mass spectrometry, Proteomics 5 (2005) 1714-1728.
[29] M. Hardt, L.R. Thomas, S.E. Dixon, G. Newport, N. Agabian, A. Prakobphol, S.C. Hall, H.E. Witkowska, S.J. Fisher, Toward defining the human parotid gland salivary proteome and peptidome: identification and characterization using 2D SDS-PAGE, ultrafiltration, HPLC, and mass spectrometry, Biochemistry 44 (2005) 2885-2899.
[30] E.J. Helmerhorst, F.G. Oppenheim, Saliva: a dynamic proteome, J Dent Res 86 (2007) 680-693.
[31] P.O. Soder, Proteolytic activity in the oral cavity: proteolytic enzymes from human saliva and dental plaque material, J Dent Res 51 (1972) 389-393.
[32] R. Inzitari, T. Cabras, G. Onnis, C. Olmi, A. Mastinu, M.T. Sanna, M.G. Pellegrini, M. Castagnola, I. Messana, Different isoforms and post-translational modifications of human salivary acidic proline-rich proteins, Proteomics 5 (2005) 805-815.
[33] M. Castagnola, R. Inzitari, D.V. Rossetti, C. Olmi, T. Cabras, V. Piras, P. Nicolussi, M.T. Sanna, M. Pellegrini, B. Giardina, I. Messana, A cascade of 24 histatins (histatin 3 fragments) in human saliva. Suggestions for a pre-secretory sequential cleavage pathway, J Biol Chem 279 (2004) 41436-41443.
[34] X. Sun, E. Salih, F.G. Oppenheim, E.J. Helmerhorst, Activity-based mass spectrometric characterization of proteases and inhibitors in human saliva, Proteomics Clin Appl 3 (2009) 810-820.
[35] G.E.W. Tandler B., Nagato T., Phillips, C. J., Secretion by striated ducts of mammalian major salivary glands: review from an ultrastructural, functional, and evolutionary perspective, Anatomical record 264 (2001) 121-145.

[36] D. Brawand, M. Soumillon, A. Necsulea, P. Julien, G. Csardi, P. Harrigan, M. Weier, A. Liechti, A. Aximu-Petri, M. Kircher, F.W. Albert, U. Zeller, P. Khaitovich, F. Grutzner, S. Bergmann, R. Nielsen, S. Paabo, H. Kaessmann, The evolution of gene expression levels in mammalian organs, Nature 478 (2011) 343-348.
[37] C.A.M. Castle J. D., Sorting and secretion of salivary proteins, Critical Reviews in Oral Biology & Medicine 4 (1993) 393-398.
[38] C.T. Castagnola M., Denotti G., Fadda M. B., Gambarini G., Lupi A., Manca I., Onnis G., Piras V., Soro V., Tambaro S., Messana I., Circadian rhythms of histatin 1, histatin 3, histatin 5, statherin and uric acid in whole human saliva secretion, Biological Rhythm Research 33 (2002) 213-222.
[39] B.C.A. Ferguson D. B., Circadian variations in the flow rate and composition of whole saliva stimulated by mastication, Archives of Oral Biology 24 (1979) 877-881.
[40] E.M. Ghezzi, L.A. Lange, J.A. Ship, Determination of variation of stimulated salivary flow rates, J Dent Res 79 (2000) 1874-1878.
[41] Z. Grzonka, E. Jankowska, F. Kasprzykowski, R. Kasprzykowska, L. Lankiewicz, W. Wiczk, E. Wieczerzak, J. Ciarkowski, P. Drabik, R. Janowski, M. Kozak, M. Jaskolski, A. Grubb, Structural studies of cysteine proteases and their inhibitors, Acta Biochim Pol 48 (2001) 1-20.
[42] M. Morzel, O. Palicki, C. Chabanet, G. Lucchi, P. Ducoroy, C. Chambon, S. Nicklaus, Saliva electrophoretic protein profiles in infants: changes with age and impact of teeth eruption and diet transition, Archives of Oral Biology 56 (2011) 634-642.
[43] D.P. Dickinson, Salivary (SD-type) cystatins: over one billion years in the making--but to what purpose?, Crit Rev Oral Biol Med 13 (2002) 485-508.
[44] J. Ochieng, G. Chaudhuri, Cystatin superfamily, J Health Care Poor Underserved 21 (2010) 51-70.
[45] M. Alvarez-Fernandez, Y.H. Liang, M. Abrahamson, X.D. Su, Crystal structure of human cystatin D, a cysteine peptidase inhibitor with restricted inhibition profile, J Biol Chem 280 (2005) 18221-18228.
[46] M.T. Stubbs, B. Laber, W. Bode, R. Huber, R. Jerala, B. Lenarcic, V. Turk, The refined 2.4 A X-ray crystal structure of recombinant human stefin B in complex with the cysteine proteinase papain: a novel type of proteinase inhibitor interaction, EMBO J 9 (1990) 1939-1947.
[47] D. Kordis, V. Turk, Phylogenomic analysis of the cystatin superfamily in eukaryotes and prokaryotes, BMC Evol Biol 9 (2009) 266.
[48] D.P. Dickinson, Cysteine peptidases of mammals: their biological roles and potential effects in the oral cavity and other tissues in health and disease, Crit Rev Oral Biol Med 13 (2002) 238-275.
[49] A. Hall, K. Hakansson, R.W. Mason, A. Grubb, M. Abrahamson, Structural basis for the biological specificity of cystatin C. Identification of leucine 9 in the N-terminal binding region as a selectivity-conferring residue in the inhibition of mammalian cysteine peptidases, J Biol Chem 270 (1995) 5115-5121.
[50] M.S. Lamkin, F.G. Oppenheim, Structural features of salivary function, Crit Rev Oral Biol Med 4 (1993) 251-259.
[51] L. Bjorck, Proteinase inhibition, immunoglobulin-binding proteins and a novel antimicrobial principle, Mol Microbiol 4 (1990) 1439-1442.
[52] M.F.J. Blankenvoorde, W. van't Hof, E. Walgreen-Weterings, T.J.M. van Steenbergen, H.S. Brand, E.C.I. Veerman, A.V.N. Amerongen, Cystatin and cystatin-derived peptides have antibacterial activity against the pathogen Porphyromonas gingivalis, Biological Chemistry 379 (1998) 1371-1375.
[53] D. Keppler, Towards novel anti-cancer strategies based on cystatin function, Cancer Lett 235 (2006) 159-176.

[54] W. Muller-Esterl, H. Fritz, J. Kellermann, F. Lottspeich, W. Machleidt, V. Turk, Genealogy of mammalian cysteine proteinase inhibitors. Common evolutionary origin of stefins, cystatins and kininogens, FEBS Lett 191 (1985) 221-226.
[55] W.M. Brown, K.M. Dziegielewska, Friends and relations of the cystatin superfamily - New members and their evolution, Protein Science 6 (1997) 5-12.
[56] N.D. Rawlings, A.J. Barrett, Evolution of Proteins of the Cystatin Superfamily, Journal of Molecular Evolution 30 (1990) 60-71.
[57] M. Edgerton, S.E. Koshlukova, Salivary histatin 5 and its similarities to the other antimicrobial proteins in human saliva, Adv Dent Res 14 (2000) 16-21.
[58] I.B. Holbrook, P.C. Molan, A further study of the factors enhancing glycolysis in human saliva, Archives of Oral Biology 18 (1973) 1275-1282.
[59] K. De Smet, R. Contreras, Human antimicrobial peptides: defensins, cathelicidins and histatins, Biotechnol Lett 27 (2005) 1337-1347.
[60] J.J. Pollock, L. Denepitiya, B.J. MacKay, V.J. Iacono, Fungistatic and fungicidal activity of human parotid salivary histidine-rich polypeptides on Candida albicans, Infect Immun 44 (1984) 702-707.
[61] B.J. MacKay, L. Denepitiya, V.J. Iacono, S.B. Krost, J.J. Pollock, Growth-inhibitory and bactericidal effects of human parotid salivary histidine-rich polypeptides on Streptococcus mutans, Infect Immun 44 (1984) 695-701.
[62] L.M. Sabatini, T. Ota, E.A. Azen, Nucleotide sequence analysis of the human salivary protein genes HIS1 and HIS2, and evolution of the STATH/HIS gene family, Mol Biol Evol 10 (1993) 497-511.
[63] N.M. Harvey, G.H. Carpenter, G.B. Proctor, J. Klein, Normal and frictional interactions of purified human statherin adsorbed on molecularly-smooth solid substrata, Biofouling 27 (2011) 823-835.
[64] L.M. Sabatini, E.A. Azen, Histatins, a family of salivary histidine-rich proteins, are encoded by at least two loci (HIS1 and HIS2), Biochem Biophys Res Commun 160 (1989) 495-502.
[65] F.G. Oppenheim, T. Xu, F.M. McMillian, S.M. Levitz, R.D. Diamond, G.D. Offner, R.F. Troxler, Histatins, a novel family of histidine-rich proteins in human parotid secretion. Isolation, characterization, primary structure, and fungistatic effects on Candida albicans, J Biol Chem 263 (1988) 7472-7477.
[66] R.F. Troxler, G.D. Offner, T. Xu, J.C. Vanderspek, F.G. Oppenheim, Structural relationship between human salivary histatins, J Dent Res 69 (1990) 2-6.
[67] T. Cabras, C. Fanali, J.A. Monteiro, F. Amado, R. Inzitari, C. Desiderio, E. Scarano, B. Giardina, M. Castagnola, I. Messana, Tyrosine polysulfation of human salivary histatin 1. A post-translational modification specific of the submandibular gland, J Proteome Res 6 (2007) 2472-2480.
[68] X. Sun, E. Salih, F.G. Oppenheim, E.J. Helmerhorst, Kinetics of histatin proteolysis in whole saliva and the effect on bioactive domains with metal-binding, antifungal, and wound-healing properties, FASEB J 23 (2009) 2691-2701.
[69] J. Wiesner, A. Vilcinskas, Antimicrobial peptides: the ancient arm of the human immune system, Virulence 1 (2010) 440-464.
[70] M. Rijnkels, L. Elnitski, W. Miller, J.M. Rosen, Multispecies comparative analysis of a mammalian-specific genomic domain encoding secretory proteins, Genomics 82 (2003) 417-432.
[71] T. Xu, E. Telser, R.F. Troxler, F.G. Oppenheim, Primary structure and anticandidal activity of the major histatin from parotid secretion of the subhuman primate, Macaca fascicularis, J Dent Res 69 (1990) 1717-1723.
[72] M.J. Oudhoff, J.G. Bolscher, K. Nazmi, H. Kalay, W. van 't Hof, A.V. Amerongen, E.C. Veerman, Histatins are the major wound-closure stimulating factors in human saliva as identified in a cell culture assay, FASEB J 22 (2008) 3805-3812.

[73] M.J. Oudhoff, P.A. van den Keijbus, K.L. Kroeze, K. Nazmi, S. Gibbs, J.G. Bolscher, E.C. Veerman, Histatins enhance wound closure with oral and non-oral cells, J Dent Res 88 (2009) 846-850.
[74] D.M. Rothstein, P. Spacciapoli, L.T. Tran, T. Xu, F.D. Roberts, M. Dalla Serra, D.K. Buxton, F.G. Oppenheim, P. Friden, Anticandida activity is retained in P-113, a 12-amino-acid fragment of histatin 5, Antimicrobial Agents and Chemotherapy 45 (2001) 1367-1373.
[75] P.A. Raj, E. Marcus, D.K. Sukumaran, Structure of human salivary histatin 5 in aqueous and nonaqueous solutions, Biopolymers 45 (1998) 51-67.
[76] E.J. Helmerhorst, P. Breeuwer, W. van't Hof, E. Walgreen-Weterings, L.C. Oomen, E.C. Veerman, A.V. Amerongen, T. Abee, The cellular target of histatin 5 on Candida albicans is the energized mitochondrion, J Biol Chem 274 (1999) 7286-7291.
[77] B. Manconi, T. Cabras, A. Vitali, C. Fanali, A. Fiorita, R. Inzitari, M. Castagnola, I. Messana, M.T. Sanna, Expression, purification, phosphorylation and characterization of recombinant human statherin, Protein Expr Purif 69 (2010) 219-225.
[78] J.L. Jensen, M.S. Lamkin, R.F. Troxler, F.G. Oppenheim, Multiple forms of statherin in human salivary secretions, Archives of Oral Biology 36 (1991) 529-534.
[79] P.A. Raj, M. Johnsson, M.J. Levine, G.H. Nancollas, Salivary statherin. Dependence on sequence, charge, hydrogen bonding potency, and helical conformation for adsorption to hydroxyapatite and inhibition of mineralization, J Biol Chem 267 (1992) 5968-5976.
[80] N. Ramasubbu, L.M. Thomas, K.K. Bhandary, M.J. Levine, Structural characteristics of human salivary statherin: a model for boundary lubrication at the enamel surface, Crit Rev Oral Biol Med 4 (1993) 363-370.
[81] K. Makrodimitris, D.L. Masica, E.T. Kim, J.J. Gray, Structure prediction of protein-solid surface interactions reveals a molecular recognition motif of statherin for hydroxyapatite, J Am Chem Soc 129 (2007) 13713-13722.
[82] G. Goobes, R. Goobes, W.J. Shaw, J.M. Gibson, J.R. Long, V. Raghunathan, O. Schueler-Furman, J.M. Popham, D. Baker, C.T. Campbell, P.S. Stayton, G.P. Drobny, The structure, dynamics, and energetics of protein adsorption-lessons learned from adsorption of statherin to hydroxyapatite, Magn Reson Chem 45 (2008) S32-S47.
[83] E.J. Helmerhorst, G. Traboulsi, E. Salih, F.G. Oppenheim, Mass spectrometric identification of key proteolytic cleavage sites in statherin affecting mineral homeostasis and bacterial binding domains, J Proteome Res 9 (2010) 5413-5421.
[84] J.T. Leito, A.J. Ligtenberg, K. Nazmi, E.C. Veerman, Identification of salivary components that induce transition of hyphae to yeast in Candida albicans, FEMS Yeast Res 9 (2009) 1102-1110.
[85] W.H. Li, M. Tanimura, The molecular clock runs more slowly in man than in apes and monkeys, Nature 326 (1987) 93-96.
[86] H.S. Kim, K.M. Lyons, E. Saitoh, E.A. Azen, O. Smithies, N. Maeda, The structure and evolution of the human salivary proline-rich protein gene family, Mamm Genome 4 (1993) 3-14.
[87] R. Vitorino, R. Alves, A. Barros, A. Caseiro, R. Ferreira, M.C. Lobo, A. Bastos, J. Duarte, D. Carvalho, L.L. Santos, F.L. Amado, Finding new posttranslational modifications in salivary proline-rich proteins, Proteomics 10 (2010) 3732-3742.
[88] S.S. Molloy, P.A. Bresnahan, S.H. Leppla, K.R. Klimpel, G. Thomas, Human furin is a calcium-dependent serine endoprotease that recognizes the sequence Arg-X-X-Arg and efficiently cleaves anthrax toxin protective antigen, J Biol Chem 267 (1992) 16396-16402.

[89] M. Shintani, K. Minaguchi, K. Suzuki, K.A. Lim, Allelic variants of acidic proline-rich proteins observed in Japanese, Chinese, and Malays, Biochem Genet 28 (1990) 173-184.
[90] R.J. Gibbons, D.I. Hay, D.H. Schlesinger, Delineation of a segment of adsorbed salivary acidic proline-rich proteins which promotes adhesion of Streptococcus gordonii to apatitic surfaces, Infect Immun 59 (1991) 2948-2954.
[91] I. Messana, T. Cabras, E. Pisano, M.T. Sanna, A. Olianas, B. Manconi, M. Pellegrini, G. Paludetti, E. Scarano, A. Fiorita, S. Agostino, A.M. Contucci, L. Calo, P.M. Picciotti, A. Manni, A. Bennick, A. Vitali, C. Fanali, R. Inzitari, M. Castagnola, Trafficking and postsecretory events responsible for the formation of secreted human salivary peptides: a proteomics approach, Mol Cell Proteomics 7 (2008) 911-926.
[92] S. Isemura, E. Saitoh, K. Sanada, Fractionation and characterization of basic proline-rich peptides of human parotid saliva and the amino acid sequence of proline-rich peptide P-E, J Biochem 91 (1982) 2067-2075.
[93] E. Saitoh, S. Isemura, K. Sanada, Further fractionation of basic proline-rich peptides from human parotid saliva and complete amino acid sequence of basic proline-rich peptide P-H, J Biochem 94 (1983) 1991-1999.
[94] E. Saitoh, S. Isemura, K. Sanada, Complete amino acid sequence of a basic proline-rich peptide, P-F, from human parotid saliva, J Biochem 93 (1983) 883-888.
[95] E. Saitoh, S. Isemura, K. Sanada, Complete amino acid sequence of a basic proline-rich peptide, P-D, from human parotid saliva, J Biochem 93 (1983) 495-502.
[96] Y. Lu, A. Bennick, Interaction of tannin with human salivary proline-rich proteins, Archives of Oral Biology 43 (1998) 717-728.
[97] H.S. Kim, O. Smithies, N. Maeda, A physical map of the human salivary proline-rich protein gene cluster covers over 700 kbp of DNA, Genomics 6 (1990) 260-267.
[98] B.A. Dale, Periodontal epithelium: a newly recognized role in health and disease, Periodontol 2000 30 (2002) 70-78.
[99] B.A. Dale, L.P. Fredericks, Antimicrobial peptides in the oral environment: expression and function in health and disease, Curr Issues Mol Biol 7 (2005) 119-133.
[100] T. Ganz, Defensins: antimicrobial peptides of innate immunity, Nat Rev Immunol 3 (2003) 710-720.
[101] C. Goebel, L.G. Mackay, E.R. Vickers, L.E. Mather, Determination of defensin HNP-1, HNP-2, and HNP-3 in human saliva by using LC/MS, Peptides 21 (2000) 757-765.
[102] S. Crovella, N. Antcheva, I. Zelezetsky, M. Boniotto, S. Pacor, M.V. Verga Falzacappa, A. Tossi, Primate beta-defensins--structure, function and evolution, Curr Protein Pept Sci 6 (2005) 7-21.
[103] A.L. Hughes, Evolutionary diversification of the mammalian defensins, Cell Mol Life Sci 56 (1999) 94-103.
[104] M. Puklo, A. Guentsch, P.S. Hiemstra, S. Eick, J. Potempa, Analysis of neutrophil-derived antimicrobial peptides in gingival crevicular fluid suggests importance of cathelicidin LL-37 in the innate immune response against periodontogenic bacteria, Oral Microbiol Immunol 23 (2008) 328-335.
[105] G. Diamond, N. Beckloff, L.K. Ryan, Host defense peptides in the oral cavity and the lung: similarities and differences, J Dent Res 87 (2008) 915-927.
[106] M. Zanetti, Cathelicidins, multifunctional peptides of the innate immunity, J Leukoc Biol 75 (2004) 39-48.
[107] M. Groschl, O. Wendler, H.G. Topf, J. Bohlender, H. Kohler, Significance of salivary adrenomedullin in the maintenance of oral health: stimulation of oral cell proliferation and antibacterial properties, Regul Pept 154 (2009) 16-22.

[108] S. Kapas, K. Pahal, A.T. Cruchley, E. Hagi-Pavli, J.P. Hinson, Expression of adrenomedullin and its receptors in human salivary tissue, J Dent Res 83 (2004) 333-337.
[109] T. Huff, C.S. Muller, A.M. Otto, R. Netzker, E. Hannappel, beta-Thymosins, small acidic peptides with multiple functions, Int J Biochem Cell Biol 33 (2001) 205-220.
[110] D. Crockford, N. Turjman, C. Allan, J. Angel, Thymosin beta4: structure, function, and biological properties supporting current and future clinical applications, Ann N Y Acad Sci 1194 (2010) 179-189.
[111] G. Sosne, P. Qiu, A.L. Goldstein, M. Wheater, Biological activities of thymosin beta4 defined by active sites in short peptide sequences, FASEB J 24 (2010) 2144-2151.
[112] P. Qiu, M. Kurpakus-Wheater, G. Sosne, Matrix metalloproteinase activity is necessary for thymosin beta 4 promotion of epithelial cell migration, Journal of Cellular Physiology 212 (2007) 165-173.
[113] D. Fanni, C. Gerosa, S. Nemolato, A. Locci, V. Marinelli, T. Cabras, I. Messana, V. Fanos, M. Castagnola, G. Faa, Thymosin beta 10 expression in developing human salivary glands, Early Hum Dev 87 (2011) 779-783.
[114] S. Nemolato, I. Messana, T. Cabras, B. Manconi, R. Inzitari, C. Fanali, G. Vento, C. Tirone, C. Romagnoli, A. Riva, D. Fanni, E. Di Felice, G. Faa, M. Castagnola, Thymosin beta(4) and beta(10) levels in pre-term newborn oral cavity and foetal salivary glands evidence a switch of secretion during foetal development, PLoS One 4 (2009) e5109.
[115] R. Inzitari, T. Cabras, E. Pisano, C. Fanali, B. Manconi, E. Scarano, A. Fiorita, G. Paludetti, A. Manni, S. Nemolato, G. Faa, M. Castagnola, I. Messana, HPLC-ESI-MS analysis of oral human fluids reveals that gingival crevicular fluid is the main source of oral thymosins beta(4) and beta(10), Journal of Separation Science 32 (2009) 57-63.
[116] M. Akhter, I. Kobayashi, T. Kiyoshima, K. Matsuo, H. Yamaza, H. Wada, J.Y. Honda, X. Ming, H. Sakai, Possible functional involvement of thymosin beta 4 in developing tooth germ of mouse lower first molar, Histochem Cell Biol 124 (2005) 207-213.
[117] R. Reti, E. Kwon, P. Qiu, M. Wheater, G. Sosne, Thymosin beta4 is cytoprotective in human gingival fibroblasts, European Journal of Oral Sciences 116 (2008) 424-430.
[118] M.F. Carlier, M. Hertzog, D. Didry, L. Renault, F.X. Cantrelle, C. van Heijenoort, M. Knossow, E. Guittet, Structure, function, and evolution of the beta-thymosin/WH2 (WASP-Homology2) actin-binding module, Ann N Y Acad Sci 1112 (2007) 67-75.
[119] I. Striz, I. Trebichavsky, Calprotectin - a pleiotropic molecule in acute and chronic inflammation, Physiol Res 53 (2004) 245-253.
[120] C.W. Heizmann, The multifunctional S100 protein family, Methods Mol Biol 172 (2002) 69-80.
[121] C. Kerkhoff, M. Klempt, C. Sorg, Novel insights into structure and function of MRP8 (S100A8) and MRP14 (S100A9), Biochimica Et Biophysica Acta-Molecular Cell Research 1448 (1998) 200-211.
[122] M.S. Springer, W.J. Murphy, Mammalian evolution and biomedicine: new views from phylogeny, Biol Rev Camb Philos Soc 82 (2007) 375-392.
[123] S. Kumar, S.B. Hedges, A molecular timescale for vertebrate evolution, Nature 392 (1998) 917-920.
[124] L. Bromham, Why do species vary in their rate of molecular evolution?, Biol Lett 5 (2009) 401-404.
[125] R. Lanfear, J.J. Welch, L. Bromham, Watching the clock: studying variation in rates of molecular evolution between species, Trends Ecol Evol 25 (2010) 495-503.
[126] S. Whelan, P. Lio, N. Goldman, Molecular phylogenetics: state-of-the-art methods for looking into the past, Trends Genet 17 (2001) 262-272.

[127] R.D. Sleator, Phylogenetics, Arch Microbiol 193 (2011) 235-239.
[128] L.N. Kinch, N.V. Grishin, Evolution of protein structures and functions, Curr Opin Struct Biol 12 (2002) 400-408.
[129] G.H. Perry, N.J. Dominy, K.G. Claw, A.S. Lee, H. Fiegler, R. Redon, J. Werner, F.A. Villanea, J.L. Mountain, R. Misra, N.P. Carter, C. Lee, A.C. Stone, Diet and the evolution of human amylase gene copy number variation, Nat Genet 39 (2007) 1256-1260.
[130] C.B. Bailey, Saliva secretion and its relation to feeding in cattle. 4. The relationship between the concentrations of sodium, potassium, chloride and inorganic phosphate in mixed saliva and rumen fluid, Br J Nutr 15 (1961) 489-498.
[131] N.L. Huq, K.J. Cross, M. Ung, E.C. Reynolds, A review of protein structure and gene organisation for proteins associated with mineralised tissue and calcium phosphate stabilisation encoded on human chromosome 4, Archives of Oral Biology 50 (2005) 599-609.
[132] C.G. Elsik, R.L. Tellam, K.C. Worley, R.A. Gibbs, D.M. Muzny, G.M. Weinstock, D.L. Adelson, E.E. Eichler, L. Elnitski, R. Guigo, D.L. Hamernik, S.M. Kappes, H.A. Lewin, D.J. Lynn, F.W. Nicholas, A. Reymond, M. Rijnkels, L.C. Skow, E.M. Zdobnov, L. Schook, J. Womack, T. Alioto, S.E. Antonarakis, A. Astashyn, C.E. Chapple, H.C. Chen, J. Chrast, F. Camara, O. Ermolaeva, C.N. Henrichsen, W. Hlavina, Y. Kapustin, B. Kiryutin, P. Kitts, F. Kokocinski, M. Landrum, D. Maglott, K. Pruitt, V. Sapojnikov, S.M. Searle, V. Solovyev, A. Souvorov, C. Ucla, C. Wyss, J.M. Anzola, D. Gerlach, E. Elhaik, D. Graur, J.T. Reese, R.C. Edgar, J.C. McEwan, G.M. Payne, J.M. Raison, T. Junier, E.V. Kriventseva, E. Eyras, M. Plass, R. Donthu, D.M. Larkin, J. Reecy, M.Q. Yang, L. Chen, Z. Cheng, C.G. Chitko-McKown, G.E. Liu, L.K. Matukumalli, J. Song, B. Zhu, D.G. Bradley, F.S. Brinkman, L.P. Lau, M.D. Whiteside, A. Walker, T.T. Wheeler, T. Casey, J.B. German, D.G. Lemay, N.J. Maqbool, A.J. Molenaar, S. Seo, P. Stothard, C.L. Baldwin, R. Baxter, C.L. Brinkmeyer-Langford, W.C. Brown, C.P. Childers, T. Connelley, S.A. Ellis, K. Fritz, E.J. Glass, C.T. Herzig, A. Iivanainen, K.K. Lahmers, A.K. Bennett, C.M. Dickens, J.G. Gilbert, D.E. Hagen, H. Salih, J. Aerts, A.R. Caetano, et al., The genome sequence of taurine cattle: a window to ruminant biology and evolution, Science 324 (2009) 522-528.
[133] G.M. Molenaar A., Davis S., Kim K., McCracken J., Callaghan M., Harris P., Rijnkels M., Maqbool N., Oden K., Good L., Singh K., Haigh B., Wieliczko R., Stelwagen K, The histatherin gene - a chimera of histatin and statherin in cattle, identified through targeted screening of an EST database, 5TH international symposium. Milk genomics and human health, Sydney, Australia, 2008.
[134] B. Manconi, C. Fanali, T. Cabras, R. Inzitari, M. Patamia, E. Scarano, A. Fiorita, A. Vitali, M. Castagnola, I. Messana, M.T. Sanna, Structural characterization of a new statherin from pig parotid granules, J Pept Sci 16 (2010) 269-275.
[135] D.M. Carlson, Proline-rich proteins and glycoproteins: expression of salivary gland multigene families, Biochimie 70 (1988) 1689-1695.
[136] J.L. Cox, P.A. Shaw, Structure, organization and regulation of a rat cysteine proteinase inhibitor-encoding gene, Gene 110 (1992) 175-180.
[137] L.M. Sabatini, Y.Z. He, E.A. Azen, Structure and sequence determination of the gene encoding human salivary statherin, Gene 89 (1990) 245-251.
[138] J.P. Freije, M. Abrahamson, I. Olafsson, G. Velasco, A. Grubb, C. Lopez-Otin, Structure and expression of the gene encoding cystatin D, a novel human cysteine proteinase inhibitor, J Biol Chem 266 (1991) 20538-20543.
[139] C. Huh, J.W. Nagle, C.A. Kozak, M. Abrahamson, S. Karlsson, Structural organization, expression and chromosomal mapping of the mouse cystatin-C-encoding gene (Cst3), Gene 152 (1995) 221-226.

[140] J. Frygelius, M. Oscarson, K. Nordqvist, A. Wedell, V. Tohonen, The reproductive tissue specific cystatin subgroup of genes: expression during gonadal development in wildtype and testatin knockout animals, Sex Dev 1 (2007) 363-372.
[141] F. Esnard, A. Esnard, D. Faucher, J.P. Capony, J. Derancourt, M. Brillard, F. Gauthier, Rat cystatin C: the complete amino acid sequence reveals a site for N-glycosylation, Biol Chem Hoppe Seyler 371 Suppl (1990) 161-166.
[142] E.A. Azen, E. Amberger, S. Fisher, A. Prakobphol, R.L. Niece, PRB1, PRB2, and PRB4 coded polymorphisms among human salivary concanavalin-A binding, II-1, and Po proline-rich proteins, Am J Hum Genet 58 (1996) 143-153.
[143] F.G. Oppenheim, G.D. Offner, R.F. Troxler, Phosphoproteins in the parotid saliva from the subhuman primate Macaca fascicularis. Isolation and characterization of a proline-rich phosphoglycoprotein and the complete covalent structure of a proline-rich phosphopeptide, J Biol Chem 257 (1982) 9271-9282.
[144] E. Saitoh, H.S. Kim, O. Smithies, N. Maeda, Human cysteine-proteinase inhibitors: nucleotide sequence analysis of three members of the cystatin gene family, Gene 61 (1987) 329-338.
[145] E. Saitoh, L.M. Sabatini, R.L. Eddy, T.B. Shows, E.A. Azen, S. Isemura, K. Sanada, The human cystatin C gene (CST3) is a member of the cystatin gene family which is localized on chromosome 20, Biochem Biophys Res Commun 162 (1989) 1324-1331.
[146] L. Padovan, Segat,L., Tossi,A., Pontillo,A., Antcheva,N. and Crovella,S., Histatins in non-human primates: gene variation and functional effects.
[147] X. Xu, H. Nagarajan, N.E. Lewis, S. Pan, Z. Cai, X. Liu, W. Chen, M. Xie, W. Wang, S. Hammond, M.R. Andersen, N. Neff, B. Passarelli, W. Koh, H.C. Fan, J. Wang, Y. Gui, K.H. Lee, M.J. Betenbaugh, S.R. Quake, I. Famili, B.O. Palsson, The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line, Nat Biotechnol 29 (2011) 735-741.
[148] G. Yan, G. Zhang, X. Fang, Y. Zhang, C. Li, F. Ling, D.N. Cooper, Q. Li, Y. Li, A.J. van Gool, H. Du, J. Chen, R. Chen, P. Zhang, Z. Huang, J.R. Thompson, Y. Meng, Y. Bai, J. Wang, M. Zhuo, T. Wang, Y. Huang, L. Wei, J. Li, Z. Wang, H. Hu, P. Yang, L. Le, P.D. Stenson, B. Li, X. Liu, E.V. Ball, N. An, Q. Huang, W. Fan, X. Zhang, W. Wang, M.G. Katze, B. Su, R. Nielsen, H. Yang, X. Wang, Genome sequencing and comparison of two nonhuman primate animal models, the cynomolgus and Chinese rhesus macaques, Nat Biotechnol 29 (2011) 1019-1023.
[149] L.H. Wei, L.C. Walker, E. Levy, Cystatin C. Icelandic-like mutation in an animal model of cerebrovascular beta-amyloidosis, Stroke 27 (1996) 2080-2085.
[150] R.O. Lopez-Solis, U. Kemmerling, Codominant expression of genes coding for different sets of inducible salivary polypeptides associated with parotid hypertrophy in two inbred mouse strains, J Cell Biochem 95 (2005) 99-107.
[151] Q. Zhang, A.A. Szalay, J.M. Tieche, E. Kyeyune-Nyombi, J.F. Sands, K.C. Oberg, J. Leonora, Cloning and functional study of porcine parotid hormone, a novel proline-rich protein, J Biol Chem 280 (2005) 22233-22244.
[152] A.L. Goldstein, F.D. Slater, A. White, Preparation, assay, and partial purification of a thymic lymphocytopoietic factor (thymosin), Proc Natl Acad Sci U S A 56 (1966) 1010-1017.
[153] M. Badamchian, A.A. Damavandy, H. Damavandy, S.D. Wadhwa, B. Katz, A.L. Goldstein, Identification and quantification of thymosin beta4 in human saliva and tears, Ann N Y Acad Sci 1112 (2007) 458-465.
[154] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J Mol Biol 215 (1990) 403-410.

[155] T.A. Hall, BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT, Nucleic Acids Symposium Series 41 (1999) 95-98.
[156] E. Hannappel, G.J. Xu, J. Morgan, J. Hempstead, B.L. Horecker, Thymosin beta 4: a ubiquitous peptide in rat and mouse tissues, Proc Natl Acad Sci U S A 79 (1982) 2172-2175.
[157] S. Erickson-Viitanen, S. Ruggieri, P. Natalini, B.L. Horecker, Distribution of thymosin beta 4 in vertebrate classes, Arch Biochem Biophys 221 (1983) 570-576.
[158] J.O. Jo, Y.J. Kang, M.S. Ock, H.K. Kleinman, H.K. Chang, H.J. Cha, Thymosin beta 4 Expression in Human Tissues and in Tumors Using Tissue Microarrays, Applied Immunohistochemistry & Molecular Morphology 19 (2011) 160-167.
[159] B.T. Lahn, D.C. Page, Functional coherence of the human Y chromosome, Science 278 (1997) 675-680.
[160] I.M. Clauss, M.G. Wathelet, J. Szpirer, M.Q. Islam, G. Levan, C. Szpirer, G.A. Huez, Human Thymosin-Beta-4/6-26 Gene Is Part of a Multigene Family Composed of 7 Members Located on 7 Different Chromosomes, Genomics 9 (1991) 174-180.
[161] S.P. Yang, H.J. Lee, Y. Su, Molecular cloning and structural characterization of the functional human thymosin beta4 gene, Mol Cell Biochem 272 (2005) 97-105.
[162] M.P. Hoffman, B.L. Kidder, Z.L. Steinberg, S. Lakhani, S. Ho, H.K. Kleinman, M. Larsen, Gene expression profiles of mouse submandibular gland development: FGFR1 regulates branching morphogenesis in vitro through BMP- and FGF-dependent mechanisms, Development 129 (2002) 5767-5778.
[163] H. Yamaza, K. Matsuo, T. Kiyoshima, N. Shigemura, I. Kobayashi, H. Wada, A. Akamime, H. Sakai, Detection of differentially expressed genes in the early developmental stage of the mouse mandible, Int J Dev Biol 45 (2001) 675-680.
[164] E. Hannappel, Thymosin beta4 and its posttranslational modifications, Ann N Y Acad Sci 1194 (2010) 27-35.
[165] D. Safer, M. Elzinga, V.T. Nachmias, Thymosin beta 4 and Fx, an actin-sequestering peptide, are indistinguishable, J Biol Chem 266 (1991) 4029-4032.
[166] T. Huff, C.S. Muller, E. Hannappel, C-terminal truncation of thymosin beta10 by an intracellular protease and its influence on the interaction with G-actin studied by ultrafiltration, FEBS Lett 414 (1997) 39-44.
[167] K. Asif, S.V. Kothiwale, Phagocytic activity of peripheral blood and crevicular phagocytes in health and periodontal disease, J Indian Soc Periodontol 14 (2010) 8-11.
[168] M.F. Blankenvoorde, Y.M. Henskens, W. van't Hof, E.C. Veerman, A.V. Nieuw Amerongen, Inhibition of the growth and cysteine proteinase activity of Porphyromonas gingivalis by human salivary cystatin S and chicken cystatin, Biol Chem 377 (1996) 847-850.
[169] U.K. Laemmli, Cleavage of structural proteins during the assembly of the head of bacteriophage T4, Nature 227 (1970) 680-685.
[170] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nucleic Acids Res 22 (1994) 4673-4680.
[171] K. Tamura, J. Dudley, M. Nei, S. Kumar, MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0, Mol Biol Evol 24 (2007) 1596-1599.
[172] N. Saitou, M. Nei, The neighbor-joining method: a new method for reconstructing phylogenetic trees, Mol Biol Evol 4 (1987) 406-425.
[173] V. Turk, V. Stoka, D. Turk, Cystatins: biochemical and structural properties, and medical relevance, Front Biosci 13 (2008) 5406-5420.
[174] P.L. Zeeuwen, T. Cheng, J. Schalkwijk, The biology of cystatin M/E and its cognate target proteases, J Invest Dermatol 129 (2009) 1327-1338.

[175] T. Nashida, R. Sato, A. Imai, H. Shimomura, Gene expression profiles of the three major salivary glands in rats, Biomed Res 31 (2010) 387-399.
[176] Y. Koshizuka, T. Yamada, K. Hoshi, T. Ogasawara, U. Chung, H. Kawano, Y. Nakamura, K. Nakamura, S. Ikegawa, H. Kawaguchi, Cystatin 10, a novel chondrocyte-specific protein, may promote the last steps of the chondrocyte differentiation pathway, Journal of Biological Chemistry 278 (2003) 48259-48266.
[177] T.M. Mattioli, S. Silva, A.M. Gregio, M.A. Machado, A.A. Lima, L.R. Alanis, The effects of antidepressants and pilocarpine on rat parotid glands: an immunohistochemical study, Clinics (Sao Paulo) 66 (2011) 1605-1610.
[178] F. Agha-Hosseini, I. Mirzaii-Dizgah, L. Ghavamzadeh, A. Ghavamzadeh, Z. Tohidast-Acrad, Effect of pilocarpine hydrochloride on unstimulated whole saliva flow rate and composition in patients with chronic graft-versus-host disease (cGVHD), Bone Marrow Transplant 39 (2007) 431-434.
[179] P. Perelman, W.E. Johnson, C. Roos, H.N. Seuanez, J.E. Horvath, M.A. Moreira, B. Kessing, J. Pontius, M. Roelke, Y. Rumpler, M.P. Schneider, A. Silva, S.J. O'Brien, J. Pecon-Slattery, A molecular phylogeny of living primates, PLoS Genet 7 (2011) e1001342.
[180] R.V. Samonte, E.E. Eichler, Segmental duplications and the evolution of the primate genome, Nat Rev Genet 3 (2002) 65-72.
[181] M. van Geel, M.C. Dickson, A.F. Beck, D.J. Bolland, R.R. Frants, S.M. van der Maarel, P.J. de Jong, J.E. Hewitt, Genomic analysis of human chromosome 10q and 4q telomeres suggests a common origin, Genomics 79 (2002) 210-217.
[182] E. Linardopoulou, H.C. Mefford, O. Nguyen, C. Friedman, G. van den Engh, D.G. Farwell, M. Coltrera, B.J. Trask, Transcriptional activity of multiple copies of a subtelomerically located olfactory receptor gene that is polymorphic in number and location, Hum Mol Genet 10 (2001) 2373-2383.
[183] S. Ohno, Evolution by gene duplication, Springer-Verlag, Berlin, New York,, 1970.
[184] C. Roth, S. Rastogi, L. Arvestad, K. Dittmar, S. Light, D. Ekman, D.A. Liberles, Evolution after gene duplication: models, mechanisms, sequences, systems, and organisms, J Exp Zool B Mol Dev Evol 308 (2007) 58-73.
[185] A. Wong, E.J. Vallender, K. Heretis, Y. Ilkin, B.T. Lahn, C.L. Martin, D.H. Ledbetter, Diverse fates of paralogs following segmental duplication of telomeric genes, Genomics 84 (2004) 239-247.
[186] A. Hobolth, O.F. Christensen, T. Mailund, M.H. Schierup, Genomic relationships and speciation times of human, chimpanzee, and gorilla inferred from a coalescent hidden Markov model, PLoS Genet 3 (2007) e7.
[187] P.A. Shaw, J.L. Cox, T. Barka, Y. Naito, Cloning and sequencing of cDNA encoding a rat salivary cysteine proteinase inhibitor inducible by beta-adrenergic agonists, J Biol Chem 263 (1988) 18133-18137.
[188] R.J. Hause, H.D. Kim, K.K. Leung, R.B. Jones, Targeted protein-omic methods are bridging the gap between proteomic and hypothesis-driven protein analysis approaches, Expert Rev Proteomics 8 (2011) 565-575.
[189] D.E. Soltis, P.S. Soltis, The role of phylogenetics in comparative genetics, Plant Physiol 132 (2003) 1790-1800.
[190] T.K. Fabian, P. Fejerdy, P. Csermely, Salivary Genomics, Transcriptomics and Proteomics: The Emerging Concept of the Oral Ecosystem and their Use in the Early Diagnosis of Cancer and other Diseases, Curr Genomics 9 (2008) 11-21.
[191] A. Godzik, Metagenomics and the protein universe, Curr Opin Struct Biol 21 (2011) 398-403.