1
Matrix Proteases and the Degradome

Clara Soria-Valles, Carlos López-Otín, and Ana Gutiérrez-Fernández

1.1
Introduction

Proteases are defined as enzymes that have the ability to perform the hydrolysis of peptide bonds. Owing to this characteristic, proteases were initially described as nonspecific enzymes of protein catabolism, participating in processes such as tissue destruction or degradation of dietary proteins. More recently, a better understanding of their functions has allowed consideration of proteases as enzymes that perform highly specific reactions and take part in multiple biological processes such as DNA replication and transcription, cell proliferation, differentiation and migration, tissue morphogenesis and remodeling, heat shock and unfolded protein responses, neurogenesis, angiogenesis, ovulation, fertilization, wound repair, stem cell mobilization, coagulation, immunity, inflammation, senescence, autophagy, apoptosis, and necrosis [1]. According to the essential roles performed by proteases in all living organisms, alterations in their proteolytic activities may lead to important pathologies such as arthritis, cardiovascular alterations, neurodegenerative disorders, progeroid syndromes, and cancer [2, 3].

The biochemical reaction catalyzed by all proteases consists in the hydrolysis of a peptide bond through the nucleophilic attack at the carbonyl group. However, the way this reaction is performed differs between specific proteases. This characteristic feature has allowed the establishment of six different catalytic classes of proteases according to the group performing the nucleophilic attack: aspartic, metallo, cysteine, serine, and threonine proteases, as well as the most recently described group of glutamic proteases, which has only been found in some species of fungi and bacteria. In the case of aspartic, glutamic, and metalloproteases, a polarized water molecule located in the active center acts directly as a nucleophile, while in the other three classes, the reactive element is a hydroxyl (serine and threonine) or sulfhydryl (cysteine) group from the corresponding catalytic core [4]. Within each class, proteases can be further subdivided into different families and clans according to sequence conservation and three-dimensional structure similarities.

The diversity and complexity of proteases have made necessary the introduction of concepts and tools for their global analysis and characterization. Thus, the term

Matrix Proteases in Health and Disease, First Edition. Edited by Niels Behrendt.
© 2012 Wiley-VCH Verlag GmbH & Co. KGaA. Published 2012 by Wiley-VCH Verlag GmbH & Co. KGaA.

degradome defines the complete set of protease genes expressed by a cell, tissue, or organism at a specific moment or circumstance [5]. Likewise, the degradome of a certain protease is the complete substrate repertoire of that specific enzyme. The ability to explore the complexity of proteases in an organism has been catalyzed by the impressive advances in the sequencing and annotation of complete genomes. In fact, since the completion in 1995 of the genome of *Haemophilus influenzae* (1 830 140 bp and 1740 genes) [6], the number of available genome sequences has continuously increased at a rapid pace. Nowadays, the genome sequence for most model organisms as well as several vertebrate species and thousands of microorganisms is publicly available [7–13]. This genomic progress has made the study of degradomes more accessible to the scientific community. The information obtained from the analysis of the degradome of a certain organism constitutes an important tool for the comprehension of biological and pathological processes and could be the key to find out new ways to diagnose, treat, or prevent human pathologies.

In this chapter, we discuss available bioinformatic tools for the construction and analysis of degradomes and present an overview of characteristic features and evolutionary aspects of several degradomes of biomedical interest. We also discuss human diseases of proteolysis and, finally, introduce the different classes of proteases with ability to degrade the extracellular matrix (ECM), which is the topic of this book.

1.2 Bioinformatic Tools for the Analysis of Complex Degradomes

The development of novel molecular technologies has significantly reduced the time and cost of generating a genome sequence. However, important parts of the subsequent analysis, including the annotation of functional elements in the genome and the integration of this information into biological processes, are still a difficult task [14]. The complexity of genomic information, in which the coding sequence is interrupted by the presence of large introns, or the existence of numerous pseudogenes with high sequence identity to bona fide genes, hampers the use of straightforward bioinformatic approaches for both the reliable identification of genes and the prediction of protein structure and function. Therefore, although many genes can be directly annotated by using bioinformatic approaches, current tools have limitations in distinguishing genes from pseudogenes. Accordingly, manually supervised annotation is still the most common method to annotate complex sequences such as the human genome and degradome [15]. Another consideration when dealing with homologous sequences, as in the case of protease-coding genes, is the importance of distinguishing between different types of gene relationships. Thus, homologous genes can be classified as either orthologous genes, which have originated as a result of a speciation event derived from a single ancestral gene in the last common ancestor of two given species, or paralogous genes, which originate as a result of a duplication event within the same

genome [16]. Orthologous genes generally maintain the same ancestral function, while paralogous genes tend to evolve and acquire novel functions.

One of the main aims of genome sequencing studies is the comparison of the complement of genes between different species. Using this approach, it is possible to identify the presence of novel genes in one species, or the loss of other genes in another species, providing clues about the molecular mechanisms underlying some of the physiological differences between them. Despite the fact that genomes are sequences of nucleic acids, the comparison of the protease repertoire between two genomes is generally performed using protein sequences, as they are more conserved than DNA sequences, and, therefore, they are more informative and the searches are more sensitive [15]. Nucleotide sequences are only compared when analyzing noncoding regions or to determine evolutionary parameters using protein-coding regions.

Another issue to take into account for gene prediction when using available genome sequences is the possibility of having assembly artifacts or sequencing errors in the analyzed sequence [15]. Thus, assembly artifacts usually lead to the collapse of clusters of highly similar genes into an artificial single copy gene. Owing to the fact that about 20% of protease genes are located in clusters [17], careful examination of those regions and additional experimental approaches should be performed to correctly annotate protease clusters. On the other hand, the presence of sequencing errors could lead to the annotation of real genes as pseudogenes, although detailed examination of sequence traces or resequencing of specific regions can solve these problems. Pseudogenes are nonfunctional copies of a gene and depending on the structure and mechanism of generation we can distinguish two main types: conventional pseudogenes and processed pseudogenes. Conventional pseudogenes usually originate from a functional copy of a gene that has been inactivated by mutation and afterwards will degenerate through the accumulation of new mutations. By contrast, processed pseudogenes are derived from the mRNA copy of a gene that is retrotransposed into the genome. The creation of pseudogenes has been a major mechanism in the evolution of the mammalian degradomes [18]; therefore, their study is important to get a global picture when comparing different genomes.

Numerous databases, web pages, and programs are freely available and constitute valuable tools for the study and comparison of complete genomes. Owing to the extension of these resources, we briefly introduce some remarkable tools and databases that can be useful for the genomic study of proteases:

Degradome database (*http://degradome.uniovi.es/*): resource containing manually annotated information about all proteases and protease inhibitor genes from humans, the chimpanzee, mouse, and rat organized in catalytic classes and families. This database also provides a catalog of human hereditary diseases of proteolysis or degradomopathies and additional information about protease structures, ancillary domains present in proteases, and differences between mammalian degradomes [3, 17, 19, 20].

MEROPS (*http://merops.sanger.ac.uk*): comprehensive annotation of proteases and inhibitors in all sequenced organisms. Proteases are classified at the protein

domain level and those that are statistically and significantly similar in amino acid sequences are grouped in a family. If the families share a common ancestor, they are grouped in a clan. Protease inhibitors are classified in the same way [21, 22].

Other resources for extracting protease information are Ensembl (*http://www.ensembl.org*), which developed a software for maintaining an automatic annotation and analysis on selected eukaryotic genomes; InterPro (*http://www.ebi.ac.uk/interpro*), which includes tools for predicting functional sites and protein domains [23] and alignment search tools such as BLAST (*http://www.ncbi.nlm.nih.gov/BLAST*) and BLAT (*http://genome.ucsc.edu*). However, in some cases, these methodologies are unable to identify distantly related protease homologs because of the high divergence of their sequences. In these situations, it is necessary to use more sensitive approaches, such as the hidden Markov models or position-specific score matrix, which are probabilistic models that collect information of specific positions from multiple sequence alignments and apply this information for the recognition of protein or DNA sequences in the genome [15]. In this regard, an important tool when working with different protease sequences is the usage of multiple sequence alignment methods developed under the principle of hierarchical clustering. Clustal (*http://www.ebi.ac.uk/clustal*) is one of the most commonly used methods for hierarchical multiple alignment [24].

1.3 Evolution of Mammalian Degradomes

The availability of complete sequences from several mammalian genomes opens the possibility of performing comparative studies of degradomes between species. This might lead to the identification of either highly conserved elements or genetic differences occurring during mammalian evolution and contribute to clarify the molecular basis of biological pathways and pathologies involving proteolytic systems [25, 26].

1.3.1 Human Degradome

Shortly after the completion of the human genome sequence draft, we performed an exhaustive bioinformatic analysis to try to find out new human protease-coding genes with sequence similarity to proteases already described in other organisms [3, 20]. By using this methodology, 570 proteases and protease-related genes, as well as 150 protease inhibitor genes, were identified, representing more than 2% of the total genes in the human genome. Interestingly, a total of 93 of these protease-related genes encode functional proteins that are catalytically inactive because of substitutions in one or more residues critical to their proteolytic activity. These nonprotease homologs may regulate the activation of other proteases or their access to substrates or inhibitors, although their precise contribution to human biology is still poorly known [1].

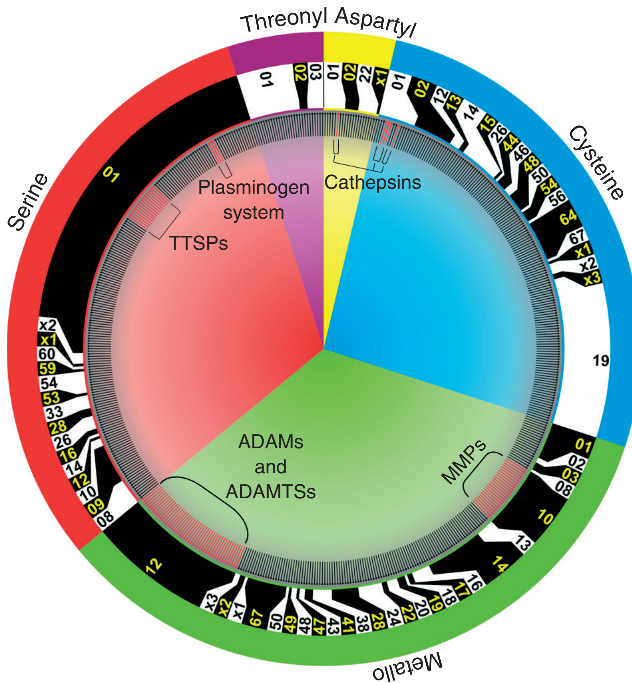

Figure 1.1 The human degradome wheel. Proteases are distributed in 5 catalytic classes and 68 different families. The code number for each protease family is indicated in the outer ring. Proteases discussed in this book are shown with red lines.

Human proteases are divided into 5 catalytic classes (aspartic, metallo, serine, cysteine, and threonine proteases) and 68 families (Figure 1.1). The most abundant classes are metalloproteases (191 members), serine proteases (178 members), and cysteine proteases (153 members). Threonine and aspartic proteases are more specialized classes and, therefore, they only have 27 and 21 members, respectively. From the 68 families of proteolytic enzymes, the most densely populated is the S01 family of serine proteases that includes enzymes of interest in the context of the present book such as plasmin, uPA (urokinase plasminogen activator), tPA (tissue-type plasminogen activator), and TTSPs (type II transmembrane serine proteases). Other representative families in the human degradome are the C01 family of cysteine proteases, mainly formed by cathepsins; the M10 metalloproteases containing more than 20 different matrix metalloproteinases (MMPs) and the M12 metalloproteases that include ADAMs (a disintegrin with metalloprotease domains) and ADAMTSs (ADAMs with thrombospondin repeats). Although the function of many of these proteases is still unknown, most of them are highly conserved between humans and other mammals, indicating that they appeared

before the mammalian expansion and their conservation is probably due to their relevance in some biological function [27].

1.3.2 Rodent Degradomes

Mouse (*Mus musculus*) and rat (*Rattus norvegicus*) are the most frequently used models for the study of human physiological and pathological processes. However, the detailed analysis of their genomes has revealed important differences from the human genome, and this information should be taken into consideration when trying to interpret and extrapolate to humans the results obtained using these animal models. In the case of proteolytic enzymes, the availability of the mouse and rat genomes has allowed us to precisely define their similarities and differences with the human degradome. Surprisingly, the complement of proteases in rodents is more complex than that in humans (656 mouse proteases and 646 rat proteases) despite their genomes are roughly 15% smaller. These proteolytic differences are mainly due to the expansion in rodents of specific protease families involved in reproduction and immunological defense, or from specific pseudogenization events in the human degradome that have led to the inactivation of some protease-coding genes [3, 15]. The expansion of specific protease families in rodents can be illustrated by the comparison of the kallikrein cluster (S01 serine proteases). In humans, this cluster is formed by 15 protease genes, while in mouse and rat it has considerably expanded and it is now composed of 26 and 23 genes, respectively. Similarly, the SENP family of sentrin-specific proteases has 14 members in the mouse but only 7 in humans [3, 15]. There are also placental cathepsins and testases that are present in rodents but absent in the human genome [17].

Conversely, the reduced number of protease genes in the human genome is mainly due to the inactivation of specific protease genes during evolution. Some representative examples of protease genes functional in rodents but pseudogenized in humans include seven members of the ADAM family of metalloproteinases (ADAM-1a, -1b, -3b, -4, -4b, -5, and -6) proposed to be involved in ovum–sperm interaction, five testis-specific serine proteases (Tessp3, Tessp6 and Tesp1, 2, and 3), and five proteases with functional roles in digestion (chymosin, intestinal Disp, trypsins Try10 and Try15, and pancreatic elastase Ela1) [15].

1.3.3 Chimpanzee Degradome

The chimpanzee (*Pan troglodytes*) belongs to the hominid family together with humans, the orangutan, bonobo, and gorilla, and constitutes our closest relative. Humans and chimpanzees shared their last common ancestor 5–6 million years ago and have a 99.1% of identity in their genome sequences [28]. The chimpanzee degradome is virtually identical to that of humans, with 568 protease-coding genes, including protease-related genes. However, and despite the high conservation between chimpanzee and human orthologs, there are examples of differential

proteases, or proteases showing a high degree of divergence between the two species. Most differences found in comparative genomic analysis of chimpanzees and humans are the consequence of a random genetic drift with a high number of neutral mutations. However, there are a few important functional changes that might be responsible for the phenotypic differences between these closely related species [29, 30]. Similar to the situation with rodents, most changes between human and chimpanzee proteases involve genes implicated in immune response or reproduction. Among them, there are some genes present in chimpanzees that are absent or have been pseudogenized in humans, such as *NAPSB*, *CASP12*, and *HPP*. Other protease genes including *GGTLA1*, *MMP23A*, *HPR*, and *PRSS33* are absent in the chimpanzee genome but are functional in humans [15]. An interesting case is *PRSS33*, which encodes a macrophage-specific serine protease conserved in most mammalian species [31], but is lost in the chimpanzee because of an Alu-mediated recombination [32]. Despite the identification of differential genes between the chimpanzee and humans, a number of proteases are identical between both species, including components of the proteasome or proteases implicated in neurological processes such as *PSEN1*, *BACE*, and *IMMP2L* [15].

1.3.4
Duck-Billed Platypus Degradome

The duck-billed platypus (*Ornithorhynchus anatinus*) is a monotrema that shared a common ancestor with humans more than 166 million years ago. This fascinating organism is an ideal source to better understand the evolution of mammalian genomes, as it shows some characteristics of reptiles and birds. The platypus genome has 536 protease genes, 90% of them being one-to-one orthologs with human genes [18, 33]. Although the number of proteases of humans and the platypus is similar, the phylogenetic position of the platypus has made it possible to identify some proteases of vertebrates conserved in the platypus but absent in eutherians and implicated in important biological processes such as apoptosis, immune response, tooth formation, reproduction, or digestion. As an example, the family alpha-aspartyl dipeptidases (AADs), represented in the platypus by a unique gene, appears in all the organisms except in eutherians, where it was lost by pseudogenization. The high expression of this gene in eggs, ovaries, and testes of Drosophila, zebrafish, frog, or chicken, suggests that it might have implications in reproduction. Something similar happens with nothepsin, an aspartyl protease conserved in oviparous animals such as fish, amphibians, reptiles, birds, and platypus, but whose gene has been pseudogenized in metatherians and eutherians [33]. It is believed that this protein might be related to the development of platypus eggs. The analysis of the platypus degradome has also revealed that similar to primates, its evolution has been mainly shaped by the loss of specific protease-coding genes. However, while in primates the inactivation of proteolytic genes is mainly due to pseudogenization caused by the accumulation of inactivating mutations, in the case of platypus most inactivated genes are completely lost due to gene deletion. In this regard, a remarkable situation is the absence in the platypus

of important digestive proteases such as pepsinogens A, B, and C, conserved in other vertebrate species and responsible for the processing of dietary proteins [18].

1.3.5
Other Degradomes

The growing list of known degradomes also includes other important species such as orangutans (*Pongo pygmaeus* and *Pongo abelii*) and zebra finch (*Taeniopygia guttata*), which complement the information derived from the study of the above-described organisms. The genome sequencing of other evolutionary distant primates such as orangutan provides useful information for genomic comparison. In addition, the sequencing of the zebra finch genome might help understand its biological peculiarities and allow the comparison between birds and mammalian degradomes [34, 35].

As expected, the orangutan degradome is highly similar to that of the chimpanzee and humans, although there are several interesting differences that mainly affect the immune and reproductive systems [36]. In some cases, protease genes have been lost in the orangutan by pseudogenization. This is the case of *PRSS33*, which is a functional protease of the immune system in humans, but has been lost in the chimpanzee and pseudogenized in the orangutan because of the existence of a premature stop codon [36], or *HTRA4*, which acquired a stop codon in its first exon and seems to have consequences for the reproductive processes [36]. In other cases, some functional genes in orangutans have been pseudogenized in humans, including *CASP12*, a cysteine protease implicated in cytokine processing and involved in sepsis both in the chimpanzee and orangutan [19, 37]. In orangutans, CASP12 constitutes a fully functional protease, while its human ortholog is a pseudogene or encodes an inactive protease [38].

The degradome of the zebra finch contains about 460 proteases and differs from mammals in several protease-coding genes such as caspases, granzymes, acrosins, metalloproteases, and pepsinogens. These changes might be related to differences in apoptosis, immune system, bone development and reproduction between birds and mammals. For example, several members of the ADAM family, such as ADAM1–7 and ADAM30, which, in mammals, participate in fertilization are absent in the genomes of the zebra finch and chicken [34], while the gene encoding caspase-3, which has an important role in song-response habituation, has been specifically duplicated in the zebra finch genome [34, 39].

In summary, the completion of all these genome projects has provided a good overview of the proteolytic systems of different organisms. However, many proteases are still largely unknown and their identification and characterization through comparative genomics and bioinformatic approaches represent important future challenges in this field. In addition, the identification or prediction of substrates and inhibitors with this kind of bioinformatic tools, along with experimental investigation, will be necessary to facilitate a better understanding of normal and pathological functions of proteases and the discovery of new therapeutic targets [15].

1.4
Human Diseases of Proteolysis

The essential role of proteolytic enzymes in the control of cell behavior predicts that abnormal, either insufficient or excessive, protease activity may lead to important pathological processes. These diseases can be caused by direct alterations in protease genes or be due to alterations in other components of proteolytic systems. The first group can be further subdivided into two major groups: genetic disorders caused by mutation of protease genes or diseases caused by alterations in the spatiotemporal pattern of expression of proteases [3]. In relation to diseases of proteolysis caused by alterations in other components of proteolytic systems, we can mention that changes in substrates, inhibitors, regulatory factors, and transport systems of proteases generate important pathologies. For instance, mutations in the gene encoding APP enhance the processing of this protein by beta- or gamma secretases causing Alzheimer's disease [40]; similarly, removal of the protease processing site in a substrate can have dramatic consequences, generating life-threatening syndromes such as Hutchinson-Gilford progeria due to removal of the proteolytic processing site in prelamin A [41–43]. Likewise, different serpinopathies are caused by mutation in serine protease inhibitors [44], while hemophilia A, the most common form of hemophilia, is caused by mutations in the *F8* gene, a cofactor of the coagulation protease factor IX [45]. Finally, alterations in the subcellular localization of proteases due to changes in transport systems can also produce important pathologies. Thus, the transport of lysosomal proteases such as cathepsin Z is altered by mutations in *ERGIC53* that cause hematological diseases [46].

Among the best characterized diseases of proteolysis are those caused by mutations in protease-coding genes. The number of hereditary diseases of proteolysis has continuously grown during the past decade, especially since the completion of the human genome sequence and the introduction of novel tools for genomic analysis [15, 47]. To date, we have annotated a total of 89 human hereditary diseases caused by mutations in protease genes, what implies that more than 15% of the genes coding for human proteases are implicated in some form of hereditary pathology [20] (*http://www.uniovi.es/degradome*). Depending on the effect of the mutation on protease function, these pathologies can be classified as loss-of-function or gain-of-function mutations, being more frequent than those caused by loss-of-protease function [3].

Mutations in protease-coding genes may have multiple and diverse outcomes, generating hematological, neurological, immunological, digestive, reproductive, and bone pathologies. For further definition of the genes and gene products referred to in the following, the reader is referred to: *http://degradome.uniovi.es/diseases.html*. The hematological diseases caused by protease gene mutations mainly involve proteases associated with the coagulation system or implicated in blood homeostasis. Thus, hemophilia B, and factor VIIa, Xa, XIa, or XIIa deficiencies are caused by mutations in *F9*, *F7*, *F10*, *F11*, and *F12*, respectively [48–52], while thrombophilia is caused by defects in *PLG* or *PROC* [53]. Hyper- or hypoprothrombinemia can be originated by either gain or loss of *F2* expression [54], while mutations

in *ADAMTS13* are responsible for thrombotic thrombocytopenic purpura [55], and cyclic hematopoiesis is caused by gain-of-function mutations in *ELA2* [56]. In addition, a wide range of neurological pathologies, from paraplegias and ataxias to degenerative disorders, are caused by mutations in proteases. Thus, gain-of-function mutations in *PSEN1* and *PSEN2* cause familial Alzheimer's disease [57, 58], while loss-of-function mutations in *SPG7* or *CLN2* originate spastic paraplegia or neuronal ceroid lipofuscinosis, respectively [59, 60]. Other neurological syndromes caused by mutations in protease genes include Gilles de la Tourette syndrome, Parkinson disease type V or nonsyndromic mental retardation, because of alterations in *IMMP2L*, *UCHL1*, or *PRSS12* [61–63]. Likewise, there are immunological diseases of proteolysis, including autoimmune lymphoproliferative syndromes (I and II), caused by mutation of either *CASP8* or *CASP10* [64, 65]. This group also includes pathologies associated with mutations in components of the complement system such as *C1R*, *C1S*, *C2*, *DF*, and *IF*.

In addition, some diseases of the digestive system are caused by loss-of-function mutations, such as *PRSS7* deficiency that results in failure to convert inactive trypsinogen into active trypsin [66], or by gain-of-function mutations such as hereditary pancreatitis caused by mutations in *PRSS1* [67]. Alterations in some proteases can also affect reproduction. Thus, mutations in *USP9Y* cause azoospermia [68], while gonadal dysgenesis is caused by mutations in *DHH* [69]. Finally, bone diseases can be developed by loss-of-function mutations in *MMP2*, which cause multicentric osteolysis with nodulosis and arthropathy [70]; or in *MMP9* and *MMP13*, causing recessive metaphyseal anadysplasia [71]; while activating mutations in *MMP13* cause spondyloepimetaphyseal dysplasia [72]. Moreover, alterations in *ADAMTS2*, *ADAMTS10*, or *ADAMTS17*, cause Ehlers-Danlos syndrome type VIIC [73] or Weill-Marchesani syndrome [74].

Besides this variety of hereditary pathologies, proteases are key players in cancer through their ability to participate in all steps of tumor progression and invasion [27]. Remarkably, other pathologies such as arthritis or cardiovascular diseases are also associated with important alterations in the spatiotemporal expression of protease genes. All these characteristics make proteolytic enzymes suitable as therapeutic targets for the development of drugs and as important prognosis biomarkers. Genomic, transcriptomic, proteomic, and degradomic studies, together with well-characterized animal models, may have the key to understand the origin and progression of pathological processes in which proteases are implicated.

1.5
Matrix Proteases and Their Inhibitors

As a prelude to the forthcoming chapters, we briefly describe in the context of the entire human degradome (Figure 1.1), the different ECM-degrading proteases and their inhibitors whose detailed analysis is addressed in this book. Among the numerous proteolytic enzymes with ECM-degrading capacity, attention was primarily focused on MMPs and the plasminogen system. Later, other metalloproteases,

belonging to both ADAM and ADAMTS families, raised considerable interest because of their putative structural and functional links with MMPs. In addition, other proteases such as cathepsins acquired importance for their vital role in mammalian cellular turnover. More recently, new families of extra- or pericellular enzymes such as TTSPs have occupied a relevant place among the enzymes with ECM-degrading properties.

MMPs, ADAMs, and ADAMTSs belong to the catalytic class of metalloproteases, a large group of enzymes that can be classified according to their catalytic mechanism, substrates and structural homology [4]. The most characteristic feature of metalloproteases is that they use a metal ion to polarize a water molecule and perform the hydrolysis. Depending on the way each metalloprotease coordinates the metal ion and accommodates the substrate in a polypeptide scaffold, these enzymes can also be classified in different groups termed *clans* or *families*. The sequence HExxH is the most conserved zinc-binding motif in metalloproteases. Here, the two histidines coordinate the zinc ion and the glutamate acts as a general base in the catalytic reaction. This motif is characteristic of the MA clan, which contains two subclasses and 40 families of peptidases, including MMPs, ADAMs, and ADAMTSs [4].

MMPs are a group of more than 20 enzymes that are either secreted or membrane anchored. Their structural design is similar to reprolysins, mainly constituted by a peptidase motif in their catalytic domain, an N-terminal signal peptide, and a prodomain. In addition, in their archetypal domain organization there is a C-terminal hinge region followed by one hemopexin domain except for MMP-7, MMP-23A and -23B, and MMP-26 that lack this domain [75]. Some of these enzymes, such as stromelysin-3 or epilysin, have a furin-like target sequence inserted in their prodomains, and constitute the group of secreted MMPs activatable by furin convertases. Other MMPs are membrane anchored by a glycosylphosphatidylinositol or by type I or type II transmembrane segments, constituting a group called membrane-type matrix metalloproteinases (MT-MMPs). For many years, MMPs have been considered as proteases with the ability to degrade all major protein components of ECM and basement membranes, playing an important role in several physiological processes such as bone development, tissue resorption, or wound healing [76], and also in different pathological processes in which ECM remodeling was necessary, such as arthritis or tumor progression. Beyond this classic function of MMPs, further studies have shown that these enzymes target a large number of nonmatrix substrates, including growth factors, peptidase inhibitors, cytokines, receptors, adhesion molecules, clotting factors, and other proteases [77]. This variety of protease substrates allows these enzymes to regulate many other processes such as migration, apoptosis, proliferation, angiogenesis, and inflammation [78].

The expression of most MMPs is regulated by hormones, growth factors, cytokines, and tissue inhibitors of metalloproteinases (TIMPs). The mammalian TIMPs are a family of protease inhibitors constituted by four members: TIMP-1, TIMP-2, TIMP-3, and TIMP-4. They have two domains, an N-terminal domain of about 125 amino acids and a C-terminal domain of around 65 residues, and each

one is stabilized by 3 disulfide bonds [79]. Expression of TIMPs and MMPs must be balanced for maintaining ECM metabolism and controlling the remodeling and physiological conditions of the tissues. In addition, TIMP-2 has been shown to be necessary for the binding and activation of proMMP2 by MT1-MMP, illustrating the complex relationships between these molecules.

The ADAM subfamily of reprolysins shares a structural domain organization based on a peptidase unit followed by a disintegrin domain, an epidermal growth factor (EGF)-like module, a transmembrane region, and a cytoplasmic tail. Most ADAMs are localized on the cell surface or are secreted, although some of them are situated at the Golgi. Their main function is the shedding of several transmembrane proteins such as cytokines and growth factor receptors from the cell surface [4]. In mammals, there are 38 different ADAMs; the mouse genome encodes 37 family members and the human genome 21. Interestingly, some ADAMs that are functional in rodents have become pseudogenes in humans, such as *ADAM1*, *3*, *4*, and *25* [3]. Conversely, *ADAM20* is a human-specific gene that is absent in rodents.

ADAMTSs are a group of secreted reprolysins that share some domains with ADAMs but lack the transmembrane and cytoplasmic tail and contain additional domains, being the most characteristic in a central thrombospondin (TS) type-1 and a series of C-terminal TS repeats [80, 81]. ADAMTS-7 and -12 also have a mucin domain, ADAMTS-20 and ADAMTS-9 a GON domain, ADAMTS-13 two CUB domains, and several other ADAMTSs a PLAC domain [82]. ADAMTSs are expressed in a wide range of adult and fetal tissues, and perform multiple physiological functions. As an example, ADAMTS-1, -4, -5, -8, -9, and -15 are aggrecanases with ability to degrade aggrecan, brevican, and versican, whereas ADAMTS-2, -3, and -14 are associated with the release of N-terminal propeptides of collagen.

Cathepsins were originally described as proteases functioning in acidic cellular compartments, just as lysosomes and endosomes, but they are present in the extracellular space as well as in the cytosol and nucleus [83]. Cathepsins can be classified based on their structure and catalytic type into aspartic (cathepsins D and E), serine (cathepsins A and G), and cysteine cathepsins (cathepsins B, C, F, H, K, L, S, O, L2, W, and Z) [84]. Among all these enzymes, cathepsins D, H, K, L, and S are capable of efficiently cleaving ECM components and other specific substrates and contribute to several physiological processes such as collagen turnover in bone and cartilage [85, 86], antigen presentation in the immune system [87], and neuropeptide and hormone processing [88, 89]. The main inhibitors of cathepsins are the proteins called *cystatins*, *stefins*, *thyropins*, and *serpins*, which bind to the target enzyme preventing substrate hydrolysis [90].

In addition to these matrix-degrading metalloproteases and cysteine proteases, there are several serine proteases implicated in the ECM metabolism including components of the plasminogen system and several members of the TTSPs. The plasminogen system is involved in numerous physiological and pathological processes because of its ability to degrade fibrin and several ECM proteins and to activate MMPs [91]. Plg is converted to the serine protease plasmin by one of the two plasminogen activators, tPA or uPA. The uPA receptor (uPAR) is a regulator of the plasminogen activator system. This glycosylphosphatidylinositol

(GPI)-bound protein acts as a membrane receptor for uPA and its zymogen form, pro-uPA [92]. When uPA is activated, it cleaves plasminogen, generating the protease plasmin, which, in a feedback process, cleaves and activates pro-uPA. Cell-associated plasminogen activation by uPA-uPAR can facilitate cell migration through ECM by pericellular proteolysis [92]. In turn, plasmin also cleaves ECM components, is essential for the development of fibrinolysis and activates MMPs, such as MMP3, MMP9, MMP12, and MMP13 [93]. The plasminogen activation system can be regulated by plasminogen activator inhibitor-1 (PAI-1; also known as *SERPINE1*) and -2 (PAI-2; also known as *SERPINB2*), that inhibit uPA and tPA preventing the activation of plasminogen into plasmin. Furthermore, the activity of plasmin can be also inhibited by α2-antiplasmin (SERPINF2).

The TTSPs are S01 serine proteases with a transmembrane domain close to the amino terminus of the protein that separates a short intracellular domain from a larger extracellular portion of the molecule containing a carboxy-terminal chymotrypsin domain (S1) and a variable internal stem region [94]. Members of this class include enteropeptidase, hepsin, spinesin, corin, matriptase, matriptase-2, TMPRSS2, TMPRSS3, and TMPRSS4. These enzymes have diverse roles in vertebrate physiology, being involved in the development and homeostasis of mammalian tissues such as the skin, heart, inner ear, placenta, and digestive tract [95]. The TTSPs are regulated by endogenous protease inhibitors, specifically Kunitz-domain-containing inhibitors and serpins. The transmembrane Kunitz-type inhibitor hepatocyte growth factor activator inhibitor-1 (HAI-1) is implicated in the inhibition of matriptase [96], hepsin [97], and prostasin [98]; and HAI-2 inhibits hepsin and matriptase [97, 99].

In summary, genomic and functional analyses have revealed the diversity and complexity of matrix proteases. Their roles in basic biological process as well as in several pathologies suggest the importance of these enzymes. This dual implication of proteases in life and disease should always be taken into consideration before dealing with clinical applications, but hopefully, the increasing knowledge of these enzymes will contribute to understand their diverse functional roles *in vivo* and to identify new therapeutic targets.

Acknowledgments

Our work is supported by grants from the Ministerio de Ciencia e Innovación-Spain, Fundación "M. Botín", and European Union (FP7 MicroEnviMet). The Instituto Universitario de Oncología is supported by Obra Social Cajastur and Acción Transversal del Cáncer-RTICC.

References

1. Lopez-Otin, C. and Bond, J.S. (2008) Proteases: multifunctional enzymes in life and disease. *J. Biol. Chem.*, **283**, 30433–30437.
2. Lopez-Otin, C. and Hunter, T. (2010) The regulatory crosstalk between kinases and proteases in cancer. *Nat. Rev. Cancer*, **10**, 278–292.

3. Puente, X.S., Sanchez, L.M., Overall, C.M., and Lopez-Otin, C. (2003) Human and mouse proteases: a comparative genomic approach. *Nat. Rev. Genet.*, **4**, 544–558.
4. Ugalde, A.P., Ordonez, G.R., Quiros, P.M., Puente, X.S., and Lopez-Otin, C. (2010) Metalloproteases and the degradome. *Methods Mol. Biol.*, **622**, 3–29.
5. Lopez-Otin, C. and Overall, C.M. (2002) Protease degradomics: a new challenge for proteomics. *Nat. Rev. Mol. Cell Biol.*, **3**, 509–519.
6. Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.F., Dougherty, B.A., Merrick, J.M. *et al.* (1995) Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science*, **269**, 496–512.
7. Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F. *et al.* (2000) The genome sequence of Drosophila melanogaster. *Science*, **287**, 2185–2195.
8. C.E.S.C. (1998) Genome sequence of the nematode C. elegans: a platform for investigating biology. *Science*, **282**, 2012–2018.
9. Gibbs, R.A., Weinstock, G.M., Metzker, M.L., Muzny, D.M., Sodergren, E.J., Scherer, S., Scott, G., Steffen, D., Worley, K.C., Burch, P.E. *et al.* (2004) Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature*, **428**, 493–521.
10. I.C.G.S.C. (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*, **432**, 695–716.
11. Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.
12. Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A. *et al.* (2001) The sequence of the human genome. *Science*, **291**, 1304–1351.
13. Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P. *et al.* (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature*, **420**, 520–562.
14. Lewis, S., Ashburner, M., and Reese, M.G. (2000) Annotating eukaryote genomes. *Curr. Opin. Struct. Biol.*, **10**, 349–354.
15. Ordonez, G.R., Puente, X.S., Quesada, V., and Lopez-Otin, C. (2009) Proteolytic systems: constructing degradomes. *Methods Mol. Biol.*, **539**, 33–47.
16. Fitch, W.M. (2000) Homology a personal view on some of the problems. *Trends Genet.*, **16**, 227–231.
17. Puente, X.S. and Lopez-Otin, C. (2004) A genomic analysis of rat proteases and protease inhibitors. *Genome Res.*, **14**, 609–622.
18. Ordonez, G.R., Hillier, L.W., Warren, W.C., Grutzner, F., Lopez-Otin, C., and Puente, X.S. (2008) Loss of genes implicated in gastric function during platypus evolution. *Genome Biol.*, **9**, R81.
19. Puente, X.S., Gutierrez-Fernandez, A., Ordonez, G.R., Hillier, L.W., and Lopez-Otin, C. (2005a) Comparative genomic analysis of human and chimpanzee proteases. *Genomics*, **86**, 638–647.
20. Quesada, V., Ordonez, G.R., Sanchez, L.M., Puente, X.S., and Lopez-Otin, C. (2009) The Degradome database: mammalian proteases and diseases of proteolysis. *Nucleic Acids Res.*, **37**, D239–D243.
21. Rawlings, N.D., Morton, F.R., and Barrett, A.J. (2006) MEROPS: the peptidase database. *Nucleic Acids Res.*, **34**, D270–D272.
22. Rawlings, N.D., Tolle, D.P., and Barrett, A.J. (2004) MEROPS: the peptidase database. *Nucleic Acids Res.*, **32**, D160–D164.
23. Mulder, N.J., Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Binns, D., Bork, P., Buillard, V., Cerutti, L., Copley, R. *et al.* (2007) New developments in the InterPro database. *Nucleic Acids Res.*, **35**, D224–D228.

24. Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994) 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**, 4673–4680.
25. Page, M.J. and Di Cera, E. (2008) Evolution of peptidase diversity. *J. Biol. Chem.*, **283**, 30010–30014.
26. Puente, X.S., Sanchez, L.M., Gutierrez-Fernandez, A., Velasco, G., and Lopez-Otin, C. (2005b) A genomic view of the complexity of mammalian proteolytic systems. *Biochem. Soc. Trans.*, **33**, 331–334.
27. Edwards, D.R. *et al.* (eds) (2008) *The Cancer Degradome: Proteases and Cancer Biology*, Springer, New York.
28. C.S.A.C. (2005) Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature*, **437**, 69–87.
29. Donaldson, I.J. and Gottgens, B. (2006) Evolution of candidate transcriptional regulatory motifs since the human-chimpanzee divergence. *Genome Biol.*, **7**, R52.
30. Varki, A. and Altheide, T.K. (2005) Comparing the human and chimpanzee genomes: searching for needles in a haystack. *Genome Res.*, **15**, 1746–1758.
31. Chen, C., Darrow, A.L., Qi, J.S., D'Andrea, M.R., and Andrade-Gordon, P. (2003) A novel serine protease predominately expressed in macrophages. *Biochem. J.*, **374**, 97–107.
32. Johnson, M.E., Cheng, Z., Morrison, V.A., Scherer, S., Ventura, M., Gibbs, R.A., Green, E.D., and Eichler, E.E. (2006) Recurrent duplication-driven transposition of DNA during hominoid evolution. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 17626–17631.
33. Warren, W.C., Hillier, L.W., Marshall Graves, J.A., Birney, E., Ponting, C.P., Grutzner, F., Belov, K., Miller, W., Clarke, L., Chinwalla, A.T. *et al.* (2008) Genome analysis of the platypus reveals unique signatures of evolution. *Nature*, **453**, 175–183.
34. Quesada, V., Velasco, G., Puente, X.S., Warren, W.C., and Lopez-Otin, C. (2010) Comparative genomic analysis of the zebra finch degradome provides new insights into evolution of proteases in birds and mammals. *BMC Genomics*, **11**, 220.
35. Warren, W.C., Clayton, D.F., Ellegren, H., Arnold, A.P., Hillier, L.W., Kunstner, A., Searle, S., White, S., Vilella, A.J., Fairley, S. *et al.* (2010) The genome of a songbird. *Nature*, **464**, 757–762.
36. Locke, D.P., Hillier, L.W., Warren, W.C., Worley, K.C., Nazareth, L.V., Muzny, D.M., Yang, S.P., Wang, Z., Chinwalla, A.T., Minx, P. *et al.* (2011) Comparative and demographic analysis of orang-utan genomes. *Nature*, **469**, 529–533.
37. Roy, S., Sharom, J.R., Houde, C., Loisel, T.P., Vaillancourt, J.P., Shao, W., Saleh, M., and Nicholson, D.W. (2008) Confinement of caspase-12 proteolytic activity to autoprocessing. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 4133–4138.
38. Xue, Y., Daly, A., Yngvadottir, B., Liu, M., Coop, G., Kim, Y., Sabeti, P., Chen, Y., Stalker, J., Huckle, E. *et al.* (2006) Spread of an inactive form of caspase-12 in humans is due to recent positive selection. *Am. J. Hum. Genet.*, **78**, 659–670.
39. Huesmann, G.R. and Clayton, D.F. (2006) Dynamic role of postsynaptic caspase-3 and BIRC4 in zebra finch song-response habituation. *Neuron*, **52**, 1061–1072.
40. Bertram, L., Lill, C.M., and Tanzi, R.E. (2010) The genetics of Alzheimer disease: back to the future. *Neuron*, **68**, 270–281.
41. De Sandre-Giovannoli, A., Bernard, R., Cau, P., Navarro, C., Amiel, J., Boccaccio, I., Lyonnet, S., Stewart, C.L., Munnich, A., Le Merrer, M. *et al.* (2003) Lamin a truncation in Hutchinson-Gilford progeria. *Science*, **300**, 2055.
42. Eriksson, M., Brown, W.T., Gordon, L.B., Glynn, M.W., Singer, J., Scott, L., Erdos, M.R., Robbins, C.M., Moses, T.Y., Berglund, P. *et al.* (2003) Recurrent de novo point mutations in lamin A cause Hutchinson-Gilford progeria syndrome. *Nature*, **423**, 293–298.
43. Varela, I., Pereira, S., Ugalde, A.P., Navarro, C.L., Suarez, M.F., Cau, P.,

Cadinanos, J., Osorio, F.G., Foray, N., Cobo, J. *et al.* (2008) Combined treatment with statins and aminobisphosphonates extends longevity in a mouse model of human premature aging. *Nat. Med.*, **14**, 767–772.
44. Ekeowa, U.I., Freeke, J., Miranda, E., Gooptu, B., Bush, M.F., Perez, J., Teckman, J., Robinson, C.V., and Lomas, D.A. (2010) Defining the mechanism of polymerization in the serpinopathies. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 17146–17151.
45. Summers, R.J., Meeks, S.L., Healey, J.F., Brown, H.C., Parker, E.T., Kempton, C.L., Doering, C.B., and Lollar, P. (2011) Factor VIII A3 domain substitution N1922S results in hemophilia A due to domain-specific misfolding and hyposecretion of functional protein. *Blood*, **117** (11), 3190–3198.
46. Hauri, H.P., Kappeler, F., Andersson, H., and Appenzeller, C. (2000) ERGIC-53 and traffic in the secretory pathway. *J. Cell Sci.*, **113** (Pt 4), 587–596.
47. I.H.G.S.C. (2004) Finishing the euchromatic sequence of the human genome. *Nature*, **431**, 931–945.
48. Bowen, D.J. (2002) Haemophilia A and haemophilia B: molecular insights. *Mol. Pathol.*, **55**, 127–144.
49. Kravtsov, D.V., Wu, W., Meijers, J.C., Sun, M.F., Blinder, M.A., Dang, T.P., Wang, H., and Gailani, D. (2004) Dominant factor XI deficiency caused by mutations in the factor XI catalytic domain. *Blood*, **104**, 128–134.
50. Millar, D.S., Elliston, L., Deex, P., Krawczak, M., Wacey, A.I., Reynaud, J., Nieuwenhuis, H.K., Bolton-Maggs, P., Mannucci, P.M., Reverter, J.C. *et al.* (2000) Molecular analysis of the genotype-phenotype relationship in factor X deficiency. *Hum. Genet.*, **106**, 249–257.
51. Perry, D.J. (2002) Factor VII deficiency. *Br. J. Haematol.*, **118**, 689–700.
52. Soria, J.M., Almasy, L., Souto, J.C., Bacq, D., Buil, A., Faure, A., Martinez-Marchan, E., Mateo, J., Borrell, M., Stone, W. *et al.* (2002) A quantitative-trait locus in the human factor XII gene influences both plasma factor XII levels and susceptibility to thrombotic disease. *Am. J. Hum. Genet.*, **70**, 567–574.
53. Gehring, N.H., Frede, U., Neu-Yilik, G., Hundsdoerfer, P., Vetter, B., Hentze, M.W., and Kulozik, A.E. (2001) Increased efficiency of mRNA 3' end formation: a new genetic mechanism contributing to hereditary thrombophilia. *Nat. Genet.*, **28**, 389–392.
54. Meeks, S.L. and Abshire, T.C. (2008) Abnormalities of prothrombin: a review of the pathophysiology, diagnosis, and treatment. *Haemophilia*, **14**, 1159–1163.
55. Levy, G.G., Nichols, W.C., Lian, E.C., Foroud, T., McClintick, J.N., McGee, B.M., Yang, A.Y., Siemieniak, D.R., Stark, K.R., Gruppo, R. *et al.* (2001) Mutations in a member of the ADAMTS gene family cause thrombotic thrombocytopenic purpura. *Nature*, **413**, 488–494.
56. Horwitz, M., Benson, K.F., Person, R.E., Aprikyan, A.G., and Dale, D.C. (1999) Mutations in ELA2, encoding neutrophil elastase, define a 21-day biological clock in cyclic haematopoiesis. *Nat. Genet.*, **23**, 433–436.
57. Citron, M., Westaway, D., Xia, W., Carlson, G., Diehl, T., Levesque, G., Johnson-Wood, K., Lee, M., Seubert, P., Davis, A. *et al.* (1997) Mutant presenilins of Alzheimer's disease increase production of 42-residue amyloid beta-protein in both transfected cells and transgenic mice. *Nat. Med.*, **3**, 67–72.
58. Esler, W.P. and Wolfe, M.S. (2001) A portrait of Alzheimer secretases--new features and familiar faces. *Science*, **293**, 1449–1454.
59. Casari, G., De Fusco, M., Ciarmatori, S., Zeviani, M., Mora, M., Fernandez, P., De Michele, G., Filla, A., Cocozza, S., Marconi, R. *et al.* (1998) Spastic paraplegia and OXPHOS impairment caused by mutations in paraplegin, a nuclear-encoded mitochondrial metalloprotease. *Cell*, **93**, 973–983.
60. Sleat, D.E., Donnelly, R.J., Lackland, H., Liu, C.G., Sohar, I., Pullarkat, R.K., and Lobel, P. (1997) Association of mutations in a lysosomal protein with classical late-infantile neuronal ceroid lipofuscinosis. *Science*, **277**, 1802–1805.

61. Leroy, E., Boyer, R., Auburger, G., Leube, B., Ulm, G., Mezey, E., Harta, G., Brownstein, M.J., Jonnalagada, S., Chernova, T. *et al.* (1998) The ubiquitin pathway in Parkinson's disease. *Nature*, **395**, 451–452.
62. Mitsui, S., Yamaguchi, N., Osako, Y., and Yuri, K. (2007) Enzymatic properties and localization of motopsin (PRSS12), a protease whose absence causes mental retardation. *Brain Res.*, **1136**, 1–12.
63. Petek, E., Windpassinger, C., Vincent, J.B., Cheung, J., Boright, A.P., Scherer, S.W., Kroisel, P.M., and Wagner, K. (2001) Disruption of a novel gene (IMMP2L) by a breakpoint in 7q31 associated with Tourette syndrome. *Am. J. Hum. Genet.*, **68**, 848–858.
64. Chun, H.J., Zheng, L., Ahmad, M., Wang, J., Speirs, C.K., Siegel, R.M., Dale, J.K., Puck, J., Davis, J., Hall, C.G. *et al.* (2002) Pleiotropic defects in lymphocyte activation caused by caspase-8 mutations lead to human immunodeficiency. *Nature*, **419**, 395–399.
65. Wang, J., Zheng, L., Lobito, A., Chan, F.K., Dale, J., Sneller, M., Yao, X., Puck, J.M., Straus, S.E., and Lenardo, M.J. (1999) Inherited human Caspase 10 mutations underlie defective lymphocyte and dendritic cell apoptosis in autoimmune lymphoproliferative syndrome type II. *Cell*, **98**, 47–58.
66. Holzinger, A., Maier, E.M., Buck, C., Mayerhofer, P.U., Kappler, M., Haworth, J.C., Moroz, S.P., Hadorn, H.B., Sadler, J.E., and Roscher, A.A. (2002) Mutations in the proenteropeptidase gene are the molecular cause of congenital enteropeptidase deficiency. *Am. J. Hum. Genet.*, **70**, 20–25.
67. Kereszturi, E., Szmola, R., Kukor, Z., Simon, P., Weiss, F.U., Lerch, M.M., and Sahin-Toth, M. (2009) Hereditary pancreatitis caused by mutation-induced misfolding of human cationic trypsinogen: a novel disease mechanism. *Hum. Mutat.*, **30**, 575–582.
68. Sun, C., Skaletsky, H., Birren, B., Devon, K., Tang, Z., Silber, S., Oates, R., and Page, D.C. (1999) An azoospermic man with a de novo point mutation in the Y-chromosomal gene USP9Y. *Nat. Genet.*, **23**, 429–432.
69. Canto, P., Soderlund, D., Reyes, E., and Mendez, J.P. (2004) Mutations in the desert hedgehog (DHH) gene in patients with 46,XY complete pure gonadal dysgenesis. *J. Clin. Endocrinol. Metab.*, **89**, 4480–4483.
70. Tuysuz, B., Mosig, R., Altun, G., Sancak, S., Glucksman, M.J., and Martignetti, J.A. (2009) A novel matrix metalloproteinase 2 (MMP2) terminal hemopexin domain mutation in a family with multicentric osteolysis with nodulosis and arthritis with cardiac defects. *Eur. J. Hum. Genet.*, **17**, 565–572.
71. Lausch, E., Keppler, R., Hilbert, K., Cormier-Daire, V., Nikkel, S., Nishimura, G., Unger, S., Spranger, J., Superti-Furga, A., and Zabel, B. (2009) Mutations in MMP9 and MMP13 determine the mode of inheritance and the clinical spectrum of metaphyseal anadysplasia. *Am. J. Hum. Genet.*, **85**, 168–178.
72. Kennedy, A.M., Inada, M., Krane, S.M., Christie, P.T., Harding, B., Lopez-Otin, C., Sanchez, L.M., Pannett, A.A., Dearlove, A., Hartley, C. *et al.* (2005) MMP13 mutation causes spondyloepimetaphyseal dysplasia, Missouri type (SEMD(MO). *J. Clin. Invest.*, **115**, 2832–2842.
73. Colige, A., Nuytinck, L., Hausser, I., van Essen, A.J., Thiry, M., Herens, C., Ades, L.C., Malfait, F., Paepe, A.D., Franck, P. *et al.* (2004) Novel types of mutation responsible for the dermatosparactic type of Ehlers-Danlos syndrome (Type VIIC) and common polymorphisms in the ADAMTS2 gene. *J. Invest. Dermatol.*, **123**, 656–663.
74. Morales, J., Al-Sharif, L., Khalil, D.S., Shinwari, J.M., Bavi, P., Al-Mahrouqi, R.A., Al-Rajhi, A., Alkuraya, F.S., Meyer, B.F., and Al Tassan, N. (2009) Homozygous mutations in ADAMTS10 and ADAMTS17 cause lenticular myopia, ectopia lentis, glaucoma, spherophakia, and short stature. *Am. J. Hum. Genet.*, **85**, 558–568.
75. Folgueras, A.R., Pendas, A.M., Sanchez, L.M., and Lopez-Otin, C. (2004) Matrix metalloproteinases in cancer: from new functions to improved inhibition strategies. *Int. J. Dev. Biol.*, **48**, 411–424.

76. Fanjul-Fernandez, M., Folgueras, A.R., Cabrera, S., and Lopez-Otin, C. (2010) Matrix metalloproteinases: evolution, gene regulation and functional analysis in mouse models. *Biochim. Biophys. Acta*, **1803**, 3–19.
77. Rodriguez, D., Morrison, C.J., and Overall, C.M. (2010) Matrix metalloproteinases: what do they not do? New substrates and biological roles identified by murine models and proteomics. *Biochim. Biophys. Acta*, **1803**, 39–54.
78. Kessenbrock, K., Plaks, V., and Werb, Z. (2010) Matrix metalloproteinases: regulators of the tumor microenvironment. *Cell*, **141**, 52–67.
79. Brew, K., Dinakarpandian, D., and Nagase, H. (2000) Tissue inhibitors of metalloproteinases: evolution, structure and function. *Biochim. Biophys. Acta*, **1477**, 267–283.
80. Cal, S., Obaya, A.J., Llamazares, M., Garabaya, C., Quesada, V., and Lopez-Otin, C. (2002) Cloning, expression analysis, and structural characterization of seven novel human ADAMTSs, a family of metalloproteinases with disintegrin and thrombospondin-1 domains. *Gene*, **283**, 49–62.
81. Llamazares, M., Cal, S., Quesada, V., and Lopez-Otin, C. (2003) Identification and characterization of ADAMTS-20 defines a novel subfamily of metalloproteinases-disintegrins with multiple thrombospondin-1 repeats and a unique GON domain. *J. Biol. Chem.*, **278**, 13382–13389.
82. Porter, S., Clark, I.M., Kevorkian, L., and Edwards, D.R. (2005) The ADAMTS metalloproteinases. *Biochem. J.*, **386**, 15–27.
83. Reiser, J., Adair, B., and Reinheckel, T. (2010) Specialized roles for cysteine cathepsins in health and disease. *J. Clin. Invest.*, **120**, 3421–3431.
84. Santamaria, I., Velasco, G., Pendas, A.M., Fueyo, A., and Lopez-Otin, C. (1998) Cathepsin Z, a novel human cysteine proteinase with a short propeptide domain and a unique chromosomal location. *J. Biol. Chem.*, **273**, 16816–16823.
85. Deal, C. (2009) Potential new drug targets for osteoporosis. *Nat. Clin. Pract. Rheumatol.*, **5**, 20–27.
86. Stoch, S.A. and Wagner, J.A. (2008) Cathepsin K inhibitors: a novel target for osteoporosis therapy. *Clin. Pharmacol. Ther.*, **83**, 172–176.
87. Honey, K. and Rudensky, A.Y. (2003) Lysosomal cysteine proteases regulate antigen presentation. *Nat. Rev. Immunol.*, **3**, 472–482.
88. Friedrichs, B., Tepel, C., Reinheckel, T., Deussing, J., von Figura, K., Herzog, V., Peters, C., Saftig, P., and Brix, K. (2003) Thyroid functions of mouse cathepsins B, K, and L. *J. Clin. Invest.*, **111**, 1733–1745.
89. Funkelstein, L., Toneff, T., Mosier, C., Hwang, S.R., Beuschlein, F., Lichtenauer, U.D., Reinheckel, T., Peters, C., and Hook, V. (2008) Major role of cathepsin L for producing the peptide hormones ACTH, beta-endorphin, and alpha-MSH, illustrated by protease gene knockout and expression. *J. Biol. Chem.*, **283**, 35652–35659.
90. Turk, B., Turk, D., and Salvesen, G.S. (2002) Regulating cysteine protease activity: essential role of protease inhibitors as guardians and regulators. *Curr. Pharm. Des.*, **8**, 1623–1637.
91. Behrendt, N. (2004) The urokinase receptor (uPAR) and the uPAR-associated protein (uPARAP/Endo180): membrane proteins engaged in matrix turnover during tissue remodeling. *Biol. Chem.*, **385**, 103–136.
92. Smith, H.W. and Marshall, C.J. (2010) Regulation of cell signalling by uPAR. *Nat. Rev. Mol. Cell Biol.*, **11**, 23–36.
93. Carmeliet, P., Moons, L., Lijnen, R., Baes, M., Lemaitre, V., Tipping, P., Drew, A., Eeckhout, Y., Shapiro, S., Lupu, F. *et al.* (1997) Urokinase-generated plasmin activates matrix metalloproteinases during aneurysm formation. *Nat. Genet.*, **17**, 439–444.
94. Bugge, T.H., Antalis, T.M., and Wu, Q. (2009) Type II transmembrane serine proteases. *J. Biol. Chem.*, **284**, 23177–23181.

95. Szabo, R. and Bugge, T.H. (2008) Type II transmembrane serine proteases in development and disease. *Int. J. Biochem. Cell Biol.*, **40**, 1297–1316.
96. Lin, C.Y., Anders, J., Johnson, M., and Dickson, R.B. (1999) Purification and characterization of a complex containing matriptase and a Kunitz-type serine protease inhibitor from human milk. *J. Biol. Chem.*, **274**, 18237–18242.
97. Kirchhofer, D., Peek, M., Lipari, M.T., Billeci, K., Fan, B., and Moran, P. (2005) Hepsin activates pro-hepatocyte growth factor and is inhibited by hepatocyte growth factor activator inhibitor-1B (HAI-1B) and HAI-2. *FEBS Lett.*, **579**, 1945–1950.
98. Fan, B., Wu, T.D., Li, W., and Kirchhofer, D. (2005) Identification of hepatocyte growth factor activator inhibitor-1B as a potential physiological inhibitor of prostasin. *J. Biol. Chem.*, **280**, 34513–34520.
99. Szabo, R., Hobson, J.P., List, K., Molinolo, A., Lin, C.Y., and Bugge, T.H. (2008) Potent inhibition and global co-localization implicate the transmembrane Kunitz-type serine protease inhibitor hepatocyte growth factor activator inhibitor-2 in the regulation of epithelial matriptase activity. *J. Biol. Chem.*, **283**, 29495–29504.