

Proteases and their Inhibitors in Neurodegenerative Disease

Focused Meeting held at the Royal Institute of British Architects, London, U.K., 17 November 2004. Organized by and Edited by H. Nagase and J. Saklatvala (Imperial College London, U.K.).

A genomic view of the complexity of mammalian proteolytic systems

X.S. Puente, L.M. Sánchez, A. Gutiérrez-Fernández, G. Velasco and C. López-Otín¹

Departamento de Bioquímica y Biología Molecular, Facultad de Medicina, Instituto Universitario de Oncología, Universidad de Oviedo, 33006-Oviedo, Spain

Abstract

Proteolytic enzymes play an essential role in different physiological processes, including development, reproduction and host defence, as well as in numerous pathologies, like inflammatory diseases, neurological disorders or cancer. The completion of the human genome sequence allowed us to determine that more than 2% of all human genes are proteases or protease inhibitors, reflecting the importance of proteolysis in human biology. To understand better the complexity of proteases in human and other model organisms, we have used the available genome sequences of different mammalian organisms, including mouse, rat and chimpanzee, to identify and compare their degradomes, the complete set of protease genes in these species. Surprisingly, the rodent protease complement is more complex when compared with that of primates, mainly due to the expansion of protease families implicated in reproduction and host defence. Similarly, most differences between human and chimpanzee proteases are found in genes implicated in the immune system, which might explain some of the differences between both organisms. We have also found several genes implicated in reproduction, nutrition and the immune system, which are functional in rat, mouse or chimpanzee, but have been inactivated by mutations in the human lineage. These findings suggest that pseudogenization of specific protease genes has been a mechanism contributing to the evolution of the human genome. Finally, we found that proteases implicated in human hereditary diseases, and especially in neurodegenerative disorders, are highly conserved among mammals.

Proteolysis represents an important mechanism that participates in numerous biological functions, including development, apoptosis, homeostasis, reproduction and host defence. Over the last decade, proteases have generated considerable biomedical interest owing to the identification of several human pathologies in which these enzymes are implicated, including inflammatory diseases, neurodegenerative disorders and cancer [1–6]. Proteases constitute a diverse group of enzymes with the common ability to hydrolyse peptide bonds. According to their catalytic mechanism, these enzymes can be classified as aspartic-, cysteine-, serine-, threonine-proteases and metalloproteases, although the recent finding of fungi-specific glutamic-peptidases has forced

the introduction of an additional class of proteolytic enzymes in which a glutamic residue plays a central role in the catalytic mechanism [7,8]. The growing importance of proteases in human biology and pathology has made necessary the use of novel concepts for the global study of proteolysis. Thus we have introduced the term degradome to define the complete set of protease genes present in one organism [9,10]. Also in this regard, the recent completion of the genome sequence of several mammalian species has opened the opportunity to investigate the complexity of their degradomes as well as to identify the differences between these organisms from a proteolytic perspective.

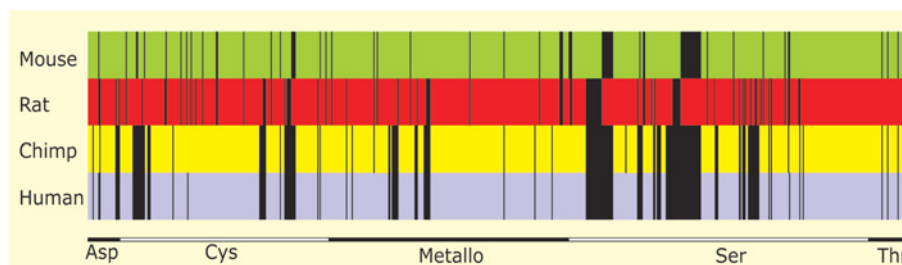
To characterize the complete repertoire of human protease genes, a bioinformatic analysis of the human genome sequence was performed, which allowed us to determine that the human degradome is composed of 561 protease and protease-related genes and more than 156 protease inhibitor

Key words: cancer, degradome, genome, immune system, mammalian proteolytic system, protease, protease inhibitor.

¹To whom correspondence should be addressed (email clo@uniovi.es).

Figure 1 | Overview of human, chimpanzee, rat and mouse degradomes

The Figure represents the complete set of protease and protease homologue genes from the indicated species. Catalytic classes are indicated at the bottom.



genes [2,11]. Taking into account that the human genome is estimated to contain < 30 000 genes [12], degradome analysis indicates that protease and protease inhibitor genes represent more than 2% of the total genes in the human genome, thus underscoring the importance of proteolysis in human biology. Human proteases can be divided into five different catalytic classes, with metalloproteases and serine proteases being the most abundant ones (187 and 175 genes respectively), whereas aspartic and threonine peptidases are composed of a limited number of genes (21 and 28 respectively) (Figure 1). Nevertheless, there are 92 protease genes that have lost key residues necessary for their proteolytic activity and have been classified as non-protease homologues [2]. Despite the lack of proteolytic activity, these inactive proteases have acquired different biological properties, and some of them might regulate the activation of other proteases or their access to substrates or inhibitors [13]. Although the function of most of these inactive proteases is not fully understood to date, many of them show a high degree of conservation between human and rodents, suggesting that they appeared before the human–rodent evolutionary divergence and have been conserved through evolution, probably because of their relevance in diverse biological functions.

On the other hand, the availability of the mouse and rat genome sequences [14,15] also represents an opportunity to characterize their degradomes and gain insights into the evolution of mammalian proteases. Surprisingly, rodent degradomes are more complex than the human degradome, with 641 genes in mouse and 626 in rat, compared with the 561 proteases present in the human degradome (Figure 1) [11]. This complexity is mainly due to the expansion in rodents of specific families of protease genes implicated in reproduction and host defence. The most significant expansions include the kallikrein cluster of serine proteases, with 26 genes in mouse and 23 in rat, and only 15 in human; the mast cell protease subfamily, implicated in host defence functions and composed of four genes in human, 17 in mouse and 28 in rat; and placental cathepsins, with eight members in mouse and ten in rat, but absent from the human genome. These results indicate that reproduction and host defence have been major forces acting during the evolution of these mammalian species. In addition, the recent sequencing of the chimpanzee genome

has allowed us to identify its degradome, which is virtually identical with the human degradome, with 559 protease genes, and with more than 99.1% identity at the amino acid level. Interestingly, most differences between human and chimpanzee proteases are found in genes implicated in immunological functions (X.S. Puente, A. Gutiérrez-Fernández, LaD.W. Hillier and C. López-Otin, unpublished work), thereby supporting the previous findings comparing human and rodent degradomes.

The comparative study of human and rodent degradomes suggests that the evolution of mammalian proteases has been mainly due to the expansion of protease genes in rodent genomes. However, there is no evidence of specific expansion of protease genes in the human genome when compared with rodents, suggesting that the evolution of human proteases has been different. In this sense, it is interesting to notice that several protease genes implicated in digestion, reproduction and host defence, which are functional in rat and mouse, show inactivating mutations in the human genome and they have become pseudogenes [2,11,15]. Among this group of human pseudogenes, there are seven members of the ADAM (a disintegrin and metalloproteinase) family of metalloproteases, including fertilins and testases that participate in reproductive functions in rodents, five members of the testis-specific serine-protease subfamily and five proteases implicated in the digestion of nutritional proteins in mouse (Table 1). It is probable that the conversion of specific genes into pseudogenes (pseudogenization) of human proteases might have been the result of the loss of function for these proteases in human processes and subsequent accumulation of mutations in the human gene. Nevertheless, it cannot be ruled out that mutations in these genes might have contributed to changes in human physiology by disrupting specific processes related to reproduction, nutrition and immunology. The analysis of the chimpanzee degradome reinforces the hypothesis of pseudogenization, as we have found evidence that some human pseudogenes are still functional proteases in chimpanzee, including caspase 12 and napsin B (X.S. Puente, A. Gutiérrez-Fernández, LaD.W. Hillier and C. López-Otin, unpublished work), whereas other proteases have been inactivated by different mutations in human and chimpanzee. Moreover, comparison of other gene families,

Table 1 | Protease pseudogenes in the human genome still functional in other species

Protease gene	Status in other mammals	Process
ADAM-1, -3, -4, -4b, -5, -6, -25	Functional in rodents	Reproduction
Caspase 12	Functional in chimpanzee and rodents	Host defence (?)
Chymosin	Functional in rodents	Digestion
Distal intestinal serine protease	Functional in rodents	Digestion
Airway trypsin-like-2, -3	Functional in rodents	Host defence (?)
Implantation serine protease-2, -2L	Functional in rodents	Reproduction
Napsin B	Functional in chimpanzee	Host defence (?)
Pancreatic elastase	Functional in rodents	Digestion
Testis serine proteases Tesp-2, -3, Tessp-3, -6	Functional in rodents	Reproduction
Trypsin-10, -15	Functional in rodents	Digestion

including olfactory receptors and members of the cytochrome P450 family, implicated in chemosensation and detoxification, has revealed that their evolution has been similar to protease genes. Thus, these gene families also exhibit a large expansion in rodents, and specific pseudogenization in the human genome [15–17], suggesting that this might have constituted a general mechanism during evolution.

On the other hand, the expansion of proteases in rodent genomes might result in an increased proteolytic activity in rodent tissues. Therefore it is tempting to speculate that mechanisms aimed at controlling the proteolytic activity in rodent tissues might also be expanded in these organisms. To evaluate this hypothesis, the protease inhibitor complement of mammalian species was identified and compared. As expected, the rodent protease inhibitor complement is more complex than the human protease inhibitor, with 199 genes in mouse and 183 in rat and only 156 in human. Interestingly, a detailed analysis of the observed expansion events suggests that there is a parallel evolution between proteases and protease inhibitor genes, as there is a correlation between the expansion of some clusters of protease inhibitor genes and the expansion of protease families they inhibit. These results suggest that the increased number of protease inhibitor genes has been a general mechanism during evolution to compensate for the increase in proteolytic activity due to the expansion of protease genes.

Despite the mentioned differences in mammalian degradomes, there is strong conservation of many protease genes between human and rodent. In addition to proteases implicated in housekeeping processes, such as the proteasome components, the highest degree of conservation can be found in those proteases that participate in neurological processes (Table 2). In this group, most proteases show more than 95% identity at the protein level between human and rodents, whereas most chimpanzee genes code for proteins that are identical with their human counterparts, including the amyloid precursor protein secretases PSEN1 and BACE implicated in Alzheimer's disease [5,18], or IMMP2L, associated with Gilles de la Tourette syndrome [19]. These results suggest that chimpanzees or rodents can suffer the same neurological disorders as humans, although it is possible that

Table 2 | Conservation of protease genes implicated in neurological disorders

Protease gene	Identity with chimpanzee gene (%)	Identity with mouse gene (%)	Disorder
BACE	100	96.4	Alzheimer's disease
IMMP2L	100	90.9	Gilles de la Tourette syndrome
PSEN1	100	92.7	Alzheimer's disease
PSEN2	99.8	95.5	Alzheimer's disease
RELN	99.8	94.8	Lissencephaly syndrome
UCHL1	100	95.1	Parkinson's disease
USP14	99.8	96.8	Ataxia mice (ax)

other differences, either in gene regulation or substrate sequence, might be responsible for the different susceptibility to neuropathologies in these species. Nevertheless, the high degree of amino acid identity between human and rodent proteases implicated in neurological disorders supports the use of these model organisms for the study of human pathologies, and to test the effect of specific-protease inhibitors to treat these devastating diseases. Finally, we have found that all protease genes implicated in human hereditary diseases are conserved in mouse and rat, with the exception of caspase 10. Mutations in this gene or the closely related protease caspase 8 cause autoimmune lymphoproliferative syndromes in humans, whereas disruption of the Casp8 gene in mice results in embryonic lethality [20]. This is probably the result of a compensation mechanism due to the presence of two related activities in human, i.e. caspase 8 and caspase 10. Nevertheless, rodents appear to be excellent models to study most human diseases, although caution should be taken when studying protease families that have been expanded in mouse or rat.

In summary, the study of mammalian genomes has helped to clarify the complexity of proteolytic enzymes and the importance of proteolysis for human biology. Current efforts, aimed at identifying the substrate specificity and biological function for most protease genes, together with the developing of novel protease inhibitors with increased specificity

and reduced side effects, will greatly benefit the treatment of numerous human diseases.

This work was supported by grants from CICYT (Spain), Fundación la Caixa and European Union (FP5 and Cancer Degradome-FP6). The Instituto Universitario de Oncología is supported by Obra Social Cajastur-Asturias, Spain and Red de Centros de Cancer, Instituto Carlos III, Spain. X.S.P. and L.M.S. belong to the Ramón y Cajal Program from the MCyT (Spain).

References

- 1 Mohammed, F.F., Smookler, D.S., Taylor, S.E., Fingleton, B., Kassiri, Z., Sanchez, O.H., English, J.L., Matrisian, L.M., Au, B., Yeh, W.C. et al. (2004) *Nat. Genet.* **36**, 969–977
- 2 Puente, X.S., Sánchez, L.M., Overall, C.M. and López-Otín, C. (2003) *Nat. Rev. Genet.* **4**, 544–558
- 3 Coussens, L.M., Tinkle, C.L., Hanahan, D. and Werb, Z. (2000) *Cell (Cambridge, Mass.)* **103**, 481–490
- 4 Balbín, M., Fueyo, A., Tester, A.M., Pendás, A.M., Pitiot, A.S., Astudillo, A., Overall, C.M., Shapiro, S.D. and López-Otín, C. (2003) *Nat. Genet.* **35**, 252–257
- 5 Esler, W.P. and Wolfe, M.S. (2001) *Science* **293**, 1449–1454
- 6 Overall, C.M. and López-Otín, C. (2002) *Nat. Rev. Cancer* **2**, 657–672
- 7 Rawlings, N.D., Tolle, D.P. and Barrett, A.J. (2004) *Nucleic Acids Res.* **32**, D160–D164 (database issue)
- 8 Fujinaga, M., Cherney, M.M., Oyama, H., Oda, K. and James, M.N. (2004) *Proc. Natl. Acad. Sci. U.S.A.* **101**, 3364–3369
- 9 López-Otín, C. and Overall, C.M. (2002) *Nat. Rev. Mol. Cell. Biol.* **3**, 509–519
- 10 Overall, C.M., Tam, E.M., Kappelhoff, R., Connor, A., Ewart, T., Morrison, C.J., Puente, X.S., López-Otín, C. and Seth, A. (2004) *Biol. Chem.* **385**, 493–504
- 11 Puente, X.S. and López-Otín, C. (2004) *Genome Res.* **14**, 609–622
- 12 Consortium, I.H.G.S. (2004) *Nature (London)* **431**, 931–945
- 13 Boatright, K.M., Deis, C., Denault, J.B., Sutherlin, D.P. and Salvesen, G.S. (2004) *Biochem. J.* **382**, 651–657
- 14 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) *Nature (London)* **420**, 520–562
- 15 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) *Nature (London)* **428**, 493–521
- 16 Nelson, D.R., Zeldin, D.C., Hoffman, S.M., Maltais, L.J., Wain, H.M. and Nebert, D.W. (2004) *Pharmacogenetics* **14**, 1–18
- 17 Godfrey, P.A., Malnic, B. and Buck, L.B. (2004) *Proc. Natl. Acad. Sci. U.S.A.* **101**, 2156–2161
- 18 Dingwall, C. (2001) *J. Clin. Invest.* **108**, 1243–1246
- 19 Petek, E., Windpassinger, C., Vincent, J.B., Cheung, J., Boright, A.P., Scherer, S.W., Kroisel, P.M. and Wagner, K. (2001) *Am. J. Hum. Genet.* **68**, 848–858
- 20 Varfolomeev, E.E., Schuchmann, M., Luria, V., Chiannikulchai, N., Beckmann, J.S., Mett, I.L., Rebrikov, D., Brodianski, V.M., Kemper, O.C., Kollet, O. et al. (1998) *Immunity* **9**, 267–276

Received 10 January 2005