The Cancer Degradome

Dylan Edwards • Gunilla Høyer-Hansen
Francesco Blasi • Bonnie F. Sloane
Editors

The Cancer Degradome

Proteases and Cancer Biology

Springer

Editors
Dylan Edwards
University of East Anglia
School of Biological Sciences
Norwich
United Kingdom NR4 7TJ
Dylan.Edwards@uea.ac.uk

Gunilla Høyer-Hansen
Finsen Laboratory
Copenhagen Biocenter
Ole Maaløes Vej 5
DK–2200 Copenhagen N
Denmark
gunilla@finsenlab.dk

Francesco Blasi
Università Vita Salute San Raffaele
and IFOM (Fondazione Istituto
FIRC di Oncologia Molecolare)
Via Adamello 16
20139 Milan
Italy
francesco.blasi@ifom-ieo-campus.it

Bonnie F. Sloane
Wayne State University
540 E. Canfield
Detroit MI 48201
USA
bsloane@med.wayne.edu

ISBN 978-0-387-69056-8 e-ISBN 978-0-387-69057-5

DOI 10.1007/978-0-387-69057-5

Library of Congress Control Number: 2008928134

Printed on acid-free paper.

springer.com

Preface

Proteolysis is essential for life. From the breakdown of proteins in food for biosynthesis, through to antigen processing in the immune system, the blood clotting cascade, and the hormone-regulated remodelling of female reproductive tissues in adult mammals – proteolysis governs functionality, homeostasis, and fate at the levels of the cell and the entire organism. For the cancer cell, intracellular proteolysis carried out by caspases and the proteasome must be enlisted and controlled to allow it to escape apoptosis. Functioning on the cancer cell surface or in the extracellular milieu, secreted proteases (primarily metalloproteinases, serine proteases, and cathepsins) determine the interactions of cells with their environments. Once considered simply as promoting tumour cell invasion through tissue barriers, proteolysis is now known to be integral to many aspects of cancer biology, including angiogenesis, regulation of the bioavailability of growth factors, cellular adhesion, cytokine/chemokine signalling, inflammatory cell recruitment, and the mobilization of normal cells from their tissue compartments to act as accomplices in metastasis. The last decade has witnessed a revolution in our thinking concerning the role of extracellular proteolysis in cancer biology: this is the primary focus of this book.

The full repertoire of proteases and their inhibitors – collectively called the degradome – has now been revealed from the sequence analysis of several animal genomes. The first section of this book discusses our current perception of the degradome, and the "degradomic" technologies that have been developed for its study. Chapters cover such topics as the bioinformatic analysis of the human degradome, the use of different technology platforms for transcriptomic studies, substrate identification using proteomics and mass spectrometry, and finally the use of activity-based probes to image protease action in cultured cells and whole organisms.

Section II switches focus to deal with particular classes of proteases and inhibitors, discussing new insights into their roles in cancer biology, primarily derived from the study of mouse model systems. A reader looking for comprehensive coverage here will be disappointed, as we felt that there have been many outstanding recent reviews of the basic biology of protease families such as the matrix metalloproteinases (MMPs) and a disintegrin and metalloprotease (ADAMs), and our intention was therefore to highlight other enzymes that have not been covered

extensively, such as the transmembrane serine proteases, or new concepts that have emerged. Chapters also discuss model systems that have been employed in angiogenesis and tumour cell invasion. Section III carries the theme of new perspectives of protease function further, dealing in particular with the connections between proteolysis and cell signalling. Chapters discuss invadopodia as membrane regions where the cellular proteolytic and signalling machineries congregate, the role of urokinase plasminogen activator (uPAR) signalling in haematopoietic stem cell mobilization, the connections between MMPs, cytokine signalling and tumour–bone interactions, and the linkage of distinct proteolytic pathways in the "protease web" during tumour metastasis.

In Section IV the use of the degradome as a source of tumour biomarkers is highlighted. There are chapters reviewing the state of play with established markers based on the uPA system, and other valuable indicators such as cysteine cathepsins and TIMP-1. Two chapters cover information from bioinformatic analysis of transcriptomic data. This leads to the final section of the book, in which the potential for targeted cancer therapeutics based on the degradome is evaluated. As well as discussing the problems associated with clinical trials of metalloproteinase inhibitors, chapters in Section V cover the development of novel selective inhibitors based on thorough structural knowledge of specific targets. In addition, exciting new strategies for anti-cancer therapies are discussed that take advantage of tumour-associated proteases to generate cytotoxic payloads from latent prodrugs, and for improved delivery of drugs to the tumour vasculature.

The five sections share a similar arrangement of subject topics, with proteases or their inhibitors being dealt with in the order of serine proteases, cysteine cathepsins, and metalloproteinases. Sections I–III begin with chapters that the reader may find particularly useful in providing an overview to a particular area.

Following the disappointments of the clinical trials of early synthetic metalloproteinase inhibitors, the cancer protease field is now resurgent, as basic cancer biology and the pharmaceutical industry take on board the new knowledge of the multifaceted roles of proteases. Not only do some proteases antagonize tumour growth, rendering them "anti-targets" that must be spared in the design of novel, more selective agents, but their involvement in the tumour–host interplay identifies entirely new areas for intervention. Also, beyond active site-directed inhibitors, new cancer targets emerge based on knowledge of exosites, substrate cleavages, and protein–protein interactions.

We hope that *The Cancer Degradome: Proteases and Cancer Biology* will convey the prevalent sense of excitement and optimism as protease research enters this new era.

Dylan Edwards
Francesco Blasi
Gunilla Høyer-Hansen
Bonnie F. Sloane

Contents

Contributors

Nesrine I. Affara
Department of Pathology, University of California, San Francisco, CA 94143, USA

Risto Ala-aho
Department of Dermatology and MediCity Research Laboratory, University of Turku, FI-20521, Turku, Finland

Kasper Almholt
Finsen Laboratory, Rigshospitalet, Copenhagen Biocenter, Ole Maaløes Vej 5, DK-2200 Copenhagen, Denmark

María Beatriz Durán Alonso
Department of Genetic Toxicology and Cancer Biology, National Institute of Biology, Večna pot 111, 1000 Ljubljana, Slovenia

Kumari L. Andarawewa
Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), CNRS UMR 7104, INSERM U596, ULP, BP 163, 67404 Illkirch Cedex, C.U. de Strasbourg, France

Lisbeth M. Andersen
Department of Molecular Biology, University of Aarhus, 10C Gustav Wied's Vej, 8000 Aarhus C, Denmark

Peter A. Andreasen
Department of Molecular Biology, University of Aarhus, 10C Gustav Wied's Vej, 8000 Aarhus C, Denmark

Vira V. Artym
Department of Oncology, Lombardi Comprehensive Cancer Center, Georgetown University Medical School, Washington, DC 20057, USA

and

Laboratory of Cell and Developmental Biology, National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, MD 20892, USA

Andrew H. Baker
British Heart Foundation Glasgow Cardiovascular Research Centre, University of Glasgow, Glasgow G12 8TA, UK

Niels Behrendt
Finsen Laboratory, Rigshospitalet, Copenhagen Biocenter, Ole Maaløes Vej 5, DK-2200 Copenhagen, Denmark

Sarah Berndt
Laboratory of Tumor and Development Biology, University of Liège, Sart-Tilman B23, B-4000 Liège, Belgium

and

Centre de Recherche en Cancérologie Expérimentale (CRCE), Groupe Interdisciplinaire de Génoprotéomique Appliquée-Recherche (GIGA-Research), University of Liège, Sart Tilman B23, B-4000 Liège, Belgium

Silvia Blacher
Laboratory of Tumor and Development Biology, Centre de Recherche en Cancérologie Expérimentale (CRCE), Groupe Interdisciplinaire de Génoprotéomique Appliquée (GIGA-R), University of Liège, Tour de Pathologie (B23), B-4000 Liège, Belgium

Francesco Blasi
Università Vita Salute San Raffaele and IFOM (Fondazione Istituto FIRC di Oncologia Molecolare), Via Adamello 16, 20139 Milan, Italy

Matthew Bogyo
Department of Pathology and Department of Microbiology and Immunology, Stanford University, Stanford, CA 94305, USA

Nils Brünner
Section of Biomedicine, Department of Veterinary Pathobiology, Faculty of Life Sciences, University of Copenhagen, DK-1870 Frederiksberg C, Denmark

Françoise Bruyère
Laboratory of Tumor and Development Biology, University of Liège, Sart-Tilman B23, B-4000 Liège, Belgium

and

Centre de Recherche en Cancérologie Expérimentale (CRCE), Groupe Interdisciplinaire de Génoprotéomique Appliquée-Recherche (GIGA-Research), University of Liège, Sart Tilman B23, B-4000 Liège, Belgium

Thomas H. Bugge
Proteases and Tissue Remodeling Unit, Oral and Pharyngeal Cancer Branch, National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, MD, USA

Valeria R. Caiolfa
Department of Molecular Biology and Functional Genomics, San Raffaele Scientific Institute, Milan, Italy

and

Italian Institute of Technology Network Research, Unit of Molecular Neuroscience, San Raffaele Scientific Institute, Milano, Italy

Ib Jarle Christensen
Finsen Laboratory, Rigshospitalet, Copenhagen Biocenter, Ole Maaløes Vej 5, DK-2200 Copenhagen, Denmark

Lisa M. Coussens
Department of Pathology, Cancer Research Institute and Comprehensive Cancer Center, University of California, San Francisco, CA 94143, USA

Jennifer H. Cox
Centre for Blood Research, Departments of Oral Biological and Medical Sciences and Biochemistry and Molecular Biology, University of British Columbia, Vancouver, BC, Canada

William Cruz-Munoz
Department of Medical Biophysics, Ontario Cancer Institute, University Health Network, Toronto, Canada

Silvia D'Alessio
Department of Molecular Biology and Functional Genomics, DIBIT, Università Vita-Salute San Raffaele, Via Olgettina 58, 20132 Milan, Italy

Bernard Degryse
Department of Molecular Biology and Functional Genomics, DIBIT, University Vita-Salute San Raffaele, Via Olgettina 58, 20132 Milan, Italy

Vincent Dive
CEA, Institut de Biologie et des Technologies de Saclay, Service d'Ingénierie Moléculaire des Proéines, 91191 Gif/Yvette Cedex, France

Dylan R. Edwards
School of Biological Sciences, University of East Anglia, Norwich, NR4 7TJ, UK

Lars H. Engelholm
Finsen Laboratory, Rigshospitalet, Copenhagen Biocenter, Ole Maaløes Vej 5, DK-2200 Copenhagen, Denmark

Mari Enoksson
Burnham Institute for Medical Research, La Jolla, CA 92037, USA

Gregg B. Fields
Department of Chemistry and Biochemistry, Florida Atlantic University, Boca Raton, FL 33431, USA

Barbara Fingleton
Department of Cancer Biology, Vanderbilt University School of Medicine, Nashville, TN 37232, USA

Camilla Frederiksen
Department of Surgical Gastroenterology 435, Copenhagen University Hospital Hvidovre, Kettegaard Allé 30, DK-2650 Hvidovre, Denmark

Norbert E. Fusenig
Group Tumor and Microenvironment (A101), German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany

Henrik Gårdsvoll
Finsen Laboratory, Rigshospitalet, Copenhagen Biocenter, Ole Maaløes Vej 5, DK-2200 Copenhagen, Denmark

Jelena Gavrilovic
Biomedical Research Centre, School of Biological Sciences, University of East Anglia, Norwich, NR4 7TJ, UK

Jason H. Gill
Institute of Cancer Therapeutics, University of Bradford, West Yorkshire, BD7 1DP, UK

Vasilena Gocheva
Cancer Biology and Genetics Program, Memorial Sloan Kettering Cancer Center, New York, NY 10021, USA

and

Weill Graduate School of Medical Sciences, Cornell University, New York, NY 10021, USA

Kirsty Anne Green
Finsen Laboratory, Rigshospitalet, Copenhagen Biocenter, Ole Maaløes Vej 5, DK-2200 Copenhagen, Denmark

Thore Hillig
Finsen Laboratory, Rigshospitalet, Copenhagen Biocenter, Ole Maaløes Vej 5, DK-2200 Copenhagen, Denmark

Gunilla Høyer-Hansen
Finsen Laboratory, Rigshospitalet, Copenhagen Biocenter, Ole Maaløes Vej 5, DK-2200 Copenhagen, Denmark

Kristiina Iljin
Medical Biotechnology, VTT technical Research Centre of Finland and University of Turku, Turku, Finland

Johanna Ivaska
Medical Biotechnology, VTT technical Research Centre of Finland and University of Turku, Turku, Finland

Benedikte Jacobsen
Finsen Laboratory, Rigshospitalet, Copenhagen Biocenter, Ole Maaløes Vej 5, DK-2200 Copenhagen, Denmark

Christopher Jedeszko
Department of Pharmacology and Barbara Ann Karmanos Cancer Institute, Wayne State University School of Medicine, Detroit, MI 48201, USA

Maud Jost
Laboratory of Tumor and Development Biology, Centre de Recherche en Cancérologie Expérimentale (CRCE), Groupe Interdisciplinaire de Génoprotéomique Appliquée (GIGA-R), University of Liège, Tour de Pathologie (B23), B-4000 Liège, Belgium

Johanna A. Joyce
Cancer Biology and Genetics Program, Memorial Sloan Kettering Cancer Center, New York, NY 10021, USA

Anna Juncker-Jensen
Finsen Laboratory, Rigshospitalet, Copenhagen Biocenter, Ole Maaløes Vej 5, DK-2200 Copenhagen, Denmark

Veli-Matti Kähäri
Department of Dermatology and MediCity Research Laboratory, University of Turku, FI-20521, Turku, Finland

Olli Kallioniemi
Medical Biotechnology, VTT technical Research Centre of Finland and University of Turku, Turku, Finland

Reinhild Kappelhoff
Centre for Blood Research, Departments of Oral Biological and Medical Sciences, University of British Columbia, Vancouver, BC, V6T 1Z3, Canada

Thomas Kelly
Department of Oncology, Lombardi Comprehensive Cancer Center, Georgetown University Medical School, Washington, DC 20057, USA

and

Department of Pathology, Arkansas Cancer Research Center, University of Arkansas for Medical Sciences, Little Rock, AR 72205, USA

Ulrich auf dem Keller
The UBC Centre for Blood Research, Department of Oral Biological and Medical Sciences, and Biochemistry and Molecular Biology, University of British Columbia, Vancouver, BC, Canada

Rama Khokha
Department of Medical Biophysics, Ontario Cancer Institute, University Health Network, Toronto, Canada

Sami Kilpinen
Medical Biotechnology, VTT Technical Research Centre of Finland and University of Turku, Turku, Finland

Magnus Kjaergaard
Finsen Laboratory, Rigshospitalet, Copenhagen Biocenter, Ole Maaløes Vej 5, DK-2200 Copenhagen, Denmark

Charlotte Kopitz
Klinikum rechts der Isar der Technischen Universität München, Institut für Experimentelle Onkologie und Therapieforschung

Achim Krüger
Klinikum rechts der Isar der Technischen Universität München, Institut für Experimentelle Onkologie und Therapieforschung

Tamara T. Lah
Department of Genetic Toxicology and Cancer Biology, National Institute of Biology, Večna pot 111, 1000 Ljubljana, Slovenia

Ross Laxton
Department of Clinical Pharmacology, William Harvey Research Institute, St. Barts and the London School of Medicine, John Vane Science Centre, London EC1M 6BQ, UK

Quanwen Li
Department of Pharmacology and Barbara Ann Karmanos Cancer Institute, Wayne State University School of Medicine, Detroit, MI 48201, USA

Karin List
Proteases and Tissue Remodeling Unit, Oral and Pharyngeal Cancer Branch, National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, MD, USA

Paul M. Loadman
Institute of Cancer Therapeutics, University of Bradford, West Yorkshire, BD7 1DP, UK

Anne Fog Lomholt
Department of Surgical Gastroenterology 435, Copenhagen University Hospital Hvidovre, Kettegaard Allé 30, DK-2650 Hvidovre, Denmark

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

Leif Røge Lund
Finsen Laboratory, Rigshospitalet, Copenhagen Biocenter, Ole Maaløes Vej 5, DK-2200 Copenhagen, Denmark

Conor C. Lynch
Departments of Orthopaedics and Rehabilitation and Cancer Biology, Vanderbilt University, Nashville, TN, USA

Gabriele Malengo
Department of Molecular Biology and Functional Genomics, San Raffaele Scientific Institute, Milano, Italy

and

Università Vita-Salute San Raffaele, Milano, Italy

Lynn M. Matrisian
Department of Cancer Biology, Vanderbilt University, Nashville, TN 37232, USA

Raymond R. Mattingly
Department of Pharmacology and Barbara Ann Karmanos Cancer Institute, Wayne State University School of Medicine, Detroit, MI 48201, USA

J. Oliver McIntyre
Department of Cancer Biology, Vanderbilt University, Nashville, TN 37232, USA

Kamiar Moin
Department of Pharmacology and Barbara Ann Karmanos Cancer Institute, Wayne State University School of Medicine, Detroit, MI 48201, USA

Margareta M. Mueller
Group Tumor and Microenvironment (A101), German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany

Susette C. Mueller
Department of Oncology, Lombardi Comprehensive Cancer Center, Georgetown University Medical School, Washington, DC 20057, USA

Gillian Murphy
Department of Oncology, Cambridge University, Cancer Research UK Cambridge Research Institute, Cambridge CB2 0RE, UK

Aditya Murthy
Department of Medical Biophysics, Ontario Cancer Institute, University Health Network, Toronto, Canada

Hideaki Nagase
Kennedy Institute of Rheumatology Division, Imperial College London, London W6 8LH, UK

Hans Jørgen Nielsen
Department of Surgical Gastroenterology 435, Copenhagen University Hospital Hvidovre, Kettegaard Allé 30, DK-2650 Hvidovre, Denmark

Agnès Noël
Laboratory of Tumor and Development Biology, Centre de Recherche en Cancérologie Expérimentale (CRCE), Groupe Interdisciplinaire de Génoprotéomique Appliquée (GIGA-R), University of Liège, Tour de Pathologie (B23), B-4000 Liège, Belgium

Robert K. Nuttall
School of Biological Sciences, University of East Anglia, Norwich, NR4 7TJ, UK

Nataša Obermajer
Faculty of Pharmacy, University of Ljubljana, Aškerčeva 6, 1000 Ljubljana, Slovenia

Mary B. Olive
Department of Pharmacology and Barbara Ann Karmanos Cancer Institute, Wayne State University School of Medicine, Detroit, MI 48201, USA

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

Christopher M. Overall
University of British Columbia, Centre for Blood Research, Departments of Oral Biological and Medical Sciences, Biochemistry and Molecular Biology,Vancouver, BC, V6T 1Z3, Canada

Helle Pappot
Finsen Laboratory, Rigshospitalet, Copenhagen Biocenter, and Department of Oncology, Rigshospitalet, Copenhagen, Denmark

Margot Paulick
Department of Pathology and Microbiology and Immunology, Stanford University, Stanford, CA 4061, USA

Caroline J. Pennington
School of Biological Sciences, University of East Anglia, Norwich, NR4 7TJ, UK

Christoph Peters
Institut für Molekulare Medizin und Zellforschung, Albert-Ludwigs-Universität, D-79104 Freiburg, Germany

Simon Pilgrim
School of Biological Sciences, University of East Anglia, Norwich, NR4 7TJ, UK

Michael Ploug
Finsen Laboratory, Rigshospitalet, Copenhagen Biocenter, Copenhagen, Denmark

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

Anna Prudova
The UBC Centre for Blood Research, Department of Oral Biological and Medical Sciences, and Biochemistry and Molecular Biology, University of British Columbia, Vancouver, BC, Canada,

Pia Ragno
Università di Salerno, Dipartimento di Chimica, Via Ponte don Melillo, 84084 Salerno, Italy

Thomas Reinheckel
Institut für Molekulare Medizin und Zellforschung, Albert-Ludwigs-Universität, D-79104 Freiburg, Germany

Marie-Christine Rio
Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), CNRS UMR 7104, INSERM U596, ULP, BP 163, 67404 Illkirch Cedex, C.U. de Strasbourg, France

John Rømer
Finsen Laboratory, Rigshospitalet, Copenhagen Biocenter, Ole Maaløes Vej 5, DK-2200 Copenhagen, Denmark

Guy Salvesen
Burnham Institute for Medical Research, La Jolla, CA 92037, USA

Mansoureh Sameni
Department of Pharmacology and Barbara Ann Karmanos Cancer Institute, Wayne State University School of Medicine, Detroit, MI 48201, USA

Clara Sampieri-Ramirez
School of Biological Sciences, University of East Anglia, Norwich, NR4 7TJ, UK

Donald R. Schwartz
Biodiscovery, LLC 5692 Plymouth Road Ann Arbor, MI 48105, USA

Xanthe Scott
Biomedical Research Centre, School of Biological Sciences, University of East Anglia, Norwich, NR4 7TJ, UK

Bonnie F. Sloane
Department of Pharmacology and Barbara Ann Karmanos Cancer Institute, Wayne State University School of Medicine, Detroit, MI 48201, USA

Helene Solberg
Finsen Laboratory, Rigshospitalet, Copenhagen Biocenter, Ole Maaløes Vej 5, DK-2200 Copenhagen, Denmark

Paul N. Span
Department of Chemical Endocrinology, Radboud University Nijmegen Medical Centre, The Netherlands

M. Patrizia Stoppelli
Institute of Genetics and Biophysics, National Research Council, Via Castellino, 111-80131 Naples, Italy

Fred C. Sweep
Department of Chemical Endocrinology, Radboud University Nijmegen Medical Centre, The Netherlands

Silvia Vosseler
Group Tumor and Microenvironment (A101), German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany

Giuseppina Votta
Institute of Genetics and Biophysics, National Research Council, Via Castellino, 111-80131 Naples, Italy

Matthew Wallard
School of Biological Sciences, University of East Anglia, Norwich, NR4 7TJ, UK

Ekkehard Weber
Institute of Physiological Chemistry, Martin Luther University, Halle-Wittenber, 06114 Halle, Germany

Claire H. Wilson
Centre for Blood Research, Department of Biochemistry and Molecular Biology, University of British Columbia, Vancouver, BC, V6T 1Z3, Canada

Shu Ye
Department of Clinical Pharmacology, William Harvey Research Institute, St. Barts and the London School of Medicine, John Vane Science Centre, London EC1M 6BQ, UK

Athanasios Yiotakis
Department of Chemistry, Laboratory of Organic Chemistry, University of Athens, Panepistimiopolis Zografou 15771, Athens, Greece

Moreno Zamai
Department of Molecular Biology and Functional Genomics, San Raffaele Scientific Institute, Milano, Italy

and

Italian Institute of Technology Network Research, Unit of Molecular Neuroscience, San Raffaele Scientific Institute, Milano, Italy

Wenhong Zhu
Burnham Institute for Medical Research, La Jolla, CA 92037, USA

Section I
The Degradome and Its Analysis

Chapter 1
Protease Genomics and the Cancer Degradome

Xose S. Puente, Gonzalo R. Ordóñez, and Carlos López-Otín

Abstract Proteases comprise a large group of enzymes involved in multiple physiological and pathological processes, which has made necessary the introduction of global concepts for their study. Thus, the human degradome has been defined as the complete set of proteolytic genes encoded by the human genome. Likewise, the term cancer degradome defines the set of protease genes expressed by a tumour at a specific time. Detailed genomic analyses have revealed that the human degradome is composed of 569 protease-coding genes, whereas mouse and rat degradomes are even more complex, containing 649 and 634 genes, respectively. The precise knowledge of these differences is essential to understand the utility and limitations of these animal models to investigate human diseases, including cancer. In this regard, recent studies with genetically modified mice have shown that proteases contribute to all stages of tumour progression and not only to the later stages as was originally proposed. These studies have also revealed the existence of proteolytic enzymes with tumour-suppressive functions. Accordingly, any attempt to understand the biological and pathological relevance of proteases in cancer must take into account the large structural and functional diversity of proteolytic systems operating in all stages of the disease. Hopefully, the novel information derived from protease genomics may finally lead to the validation of some of these enzymes as important components of future strategies for cancer treatment.

Introduction

Proteases constitute a group of enzymes with the ability to hydrolyze peptide bonds. The irreversibility of this type of reaction makes it suitable for multiple cellular processes, which has contributed to a widespread use of this mechanism in different

C. López-Otín
Departamento de Bioquímica y Biología Molecular, Facultad de Medicina, Universidad de Oviedo, 33006 Oviedo-Spain, e-mail: clo@uniovi.es

D. Edwards et al. (eds.), *The Cancer Degradome*.

biological contexts, including development, apoptosis, homeostasis, reproduction or host defense (López-Otín and Overall 2002). Since their initial discovery, this group of enzymes has attracted the interest of numerous researchers because of their participation in important physiological processes such as food digestion and blood coagulation. This fact has contributed to clarify important features of this type of enzymes, including the biochemical mechanisms implicated in their catalysis, the structural determinants which define their substrate specificity and the different mechanisms by which their activity is regulated, either by specific activation through limited proteolysis of an inactive precursor or by the action of endogenous protease inhibitors (Rawlings et al. 2004). As novel proteases were identified and additional features discovered, the interest on proteolytic enzymes has grown accordingly. In fact, over the last two decades, proteases have acquired great biomedical interest due to the identification of numerous human pathologies in which proteolytic enzymes are implicated. These protease-associated diseases include inflammatory conditions, cardiovascular alterations, neurodegenerative disorders and cancer (Coussens et al. 2000, Esler and Wolfe 2001, Mohammed et al. 2004, Overall and López-Otín 2002, Puente et al. 2003). In most of these cases, the diseases are linked to an increased proteolytic activity, resulting in enhanced protein degradation and finally leading to tissue damage and destruction. That is the case for inflammatory diseases such as rheumatoid arthritis, in which an excessive protease activity results in cartilage degradation and impaired joint function, or cancer, in which proteases acting at the leading edge are responsible for basal membrane degradation, facilitating the invasion of tumour cells and the further development of metastasis (Zucker et al. 2003). These studies together with the introduction of improved cloning technologies have resulted in the identification of numerous novel human proteases and their association with specific pathologies, which has led to the consideration of these enzymes as promising targets to treat different human diseases (Turk 2006). In this chapter, we will discuss our current knowledge on human proteolytic enzymes and the utility of comparative genomic analysis to understand their evolutionary history and to evaluate experimental data on proteases obtained in animal models. Finally, we will specifically discuss the relevance for cancer of this genomic analysis of proteolytic systems.

The Human Degradome

The importance of proteolysis for life is underscored by the fact that all living organisms contain proteases which are required for normal development or growth (Barrett et al. 2004). Although proteases perform the same catalytic reaction, the hydrolysis of a peptide bond, this activity has evolved independently several times leading to the emergence of numerous enzymes with different mechanisms capable of performing this type of reaction. As a result, proteolytic enzymes can be classified in six different classes according to their catalytic mechanism, including aspartic-, cysteine-, serine-, threonine-, and metalloproteases, as well as the recently identified fungi-specific class of glutamic-peptidases (Fujinaga et al. 2004,

Rawlings et al. 2004). The large number of identified proteases and their importance 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 or the repertoire of proteases expressed by a certain tissue (López-Otín and Overall 2002). Likewise, the term cancer degradome has been rapidly coined to define the set of protease genes expressed by a tumour at a specific time (Overall and López-Otín 2002). Although our current understanding of the role of proteases in tumour development is still limited, it is generally accepted that a detailed knowledge of the proteases expressed by a tumour at a certain stage will be extremely useful for early detection and prognosis evaluation of the disease as well as for designing specific treatments based on the degradome of the tumour.

The availability of the human genome sequence opened the possibility to characterize the complete repertoire of human protease genes. To this aim, we first performed a bioinformatic analysis of the human genome to classify all previously known protease-coding genes and to identify novel genes encoding proteins with sequence similarity to previously known proteases from human or other organisms. This allowed us to determine that the human degradome is composed of 569 protease and protease-related genes (López-Otín et al. 2004, Puente et al. 2003). Taking into account that the human genome is estimated to contain less than 25,000 genes (Collins et al. 2004, Hubbard et al. 2007), the analysis of the human degradome indicates that proteases represent more than 2% of the total genes in the human genome, underscoring the importance of proteolysis in human biology.

Human proteases can be divided into five different catalytic classes, with metalloproteases, serine and cysteine proteases being the most abundant ones (194, 176 and 150 genes in the human genome, respectively), while aspartic and threonine peptidases are composed of a limited number of members (21 and 28, respectively) (Fig. 1.1). Nevertheless, it is worthwhile mentioning that among the 569 human proteases, 92 have lost key residues necessary for their proteolytic activity and have been classified as non-protease homologues (Puente et al. 2003). 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 (Boatright et al. 2004). Although the function for most of these catalytically inactive proteases is not fully understood to date, many of them show a high degree of conservation between human and other mammals, suggesting that they appeared before the mammalian expansion and have been conserved through evolution probably because of their relevance in diverse biological functions. Another interesting characteristic of proteolytic enzymes is the presence in most of them of one or several auxiliary domains. These domains lack proteolytic activity and in most cases can be also found in other types of proteins. The presence of these ancillary regions confers novel biological functions to proteolytic enzymes, facilitating their interaction with specific substrates, activators or inhibitors or their localization on specific cellular compartments (Overall 2002).

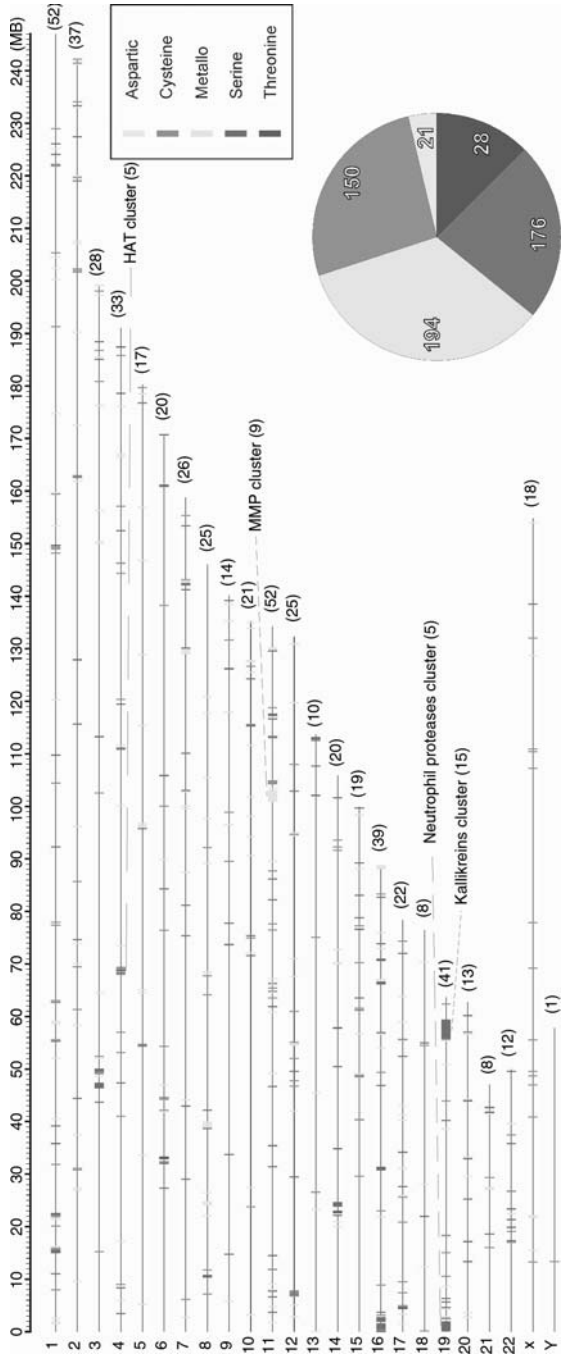

Fig. 1.1 Genomic view of the human degradome. Distribution of protease genes along individual chromosomes showing genes as boxes coloured according to the catalytic class to which they belong. The location of specific protease gene clusters in the human genome is indicated, as well as the number of genes present in each cluster. The number of proteases in each human chromosome is shown on the right side. The circle represents the distribution of human protease genes per catalytic class. (*See also* Color Insert I)

The completion of the human genome sequence has been a starting point to define the complexity of the human degradome. In this regard, the number of proteases and protease homologues currently annotated in the degradome database (*http://www.uniovi.es/degradome*) must be viewed as a current estimate of human protease-coding genes and not as a definitive number of human proteases. Accordingly, the number of human proteases has grown slightly during the last few years as novel structural designs and catalytic mechanisms have been unveiled and the corresponding human orthologs identified (Diaz-Perales et al. 2005). However, it must be taken into account that about one third of all human proteins cannot be classified into any of the protein families currently present in the protein family (Pfam) database (Finn et al. 2006). Therefore, it is expected that the experimental analysis of these orphan proteins could result in the identification of as yet unknown novel catalytic mechanisms which might contribute to expand the dimensions of the human degradome.

Tools for Degradome Research

The growing interest in proteolytic enzymes during the last decade was mainly due to the observation that proteolytic activity was associated with the progression of numerous human diseases, and especially cancer. The ability to determine the set of protease genes expressed by a tumour, or cancer degradome, can be extremely useful to understand the invasive potential of the tumour and to decide personalized treatments based on the use of specific protease inhibitors. In this regard, the definition of the human protease repertoire has opened the possibility to understand the complexity of the human degradome and has allowed the design of novel tools to study the implication of proteases in physiological and pathological processes. Thus, the knowledge of the coding sequences for all human proteases and inhibitors has been first used to develop a cDNA microarray, the CLIP-CHIP, for the detection of all proteases and protease inhibitors in human and mouse samples (Overall et al. 2004). Similar approaches based on oligonucleotide microarrays have also been used by different groups to analyze the expression of proteases and their inhibitors in malignant tumours (Schwartz et al. 2007). More recently, low-density arrays based on specific TaqMan probes have been developed to allow the detection and quantification of even low-expressing protease genes (G.R.O, X.S.P. and C.L.O., unpublished). The recent availability of these discovery tools opens the possibility to analyze in more detail the complexity of biological or pathological processes in terms of proteases, and to understand the molecular mechanisms underlying different human diseases, including cancer.

Increased Complexity of Rodent Degradomes

Although many proteases have been identified due to their expression in different human physiological or pathological conditions, the contribution of a certain

protease to processes such as inflammation or tumour growth cannot be inferred from its expression pattern, making necessary the introduction of other approaches to address this question. In this regard, different animal models including mouse, rat, macaque and chicken have proved useful to understand the molecular mechanisms underlying several human diseases. In fact, rat is widely used to study pathologies such as cardiovascular or neurodegenerative diseases, as well as to carry out pharmacological studies of specific drugs. However, the ability to easily manipulate the mouse genome to either mutate or overexpress specific genes has made this species one of the most valuable tools to understand the molecular mechanisms underlying certain human diseases such as cancer (Rosenthal and Brown 2007). In any case, the broad use of animal models to study these processes and to test novel protease inhibitors implies the need to fully define the complexity of their proteolytic systems.

The recent availability of the mouse and rat genome sequences (Gibbs et al. 2004, Waterston et al. 2002) has provided an excellent opportunity to characterize their degradomes and to gain insights into the evolution of mammalian proteases. Surprisingly, despite their smaller genomes, rodent degradomes are more complex than the human degradome, with 649 genes in mouse and 634 in rat, compared to the 569 proteases present in the human degradome (López-Otín and Matrisian 2007, Puente and López-Otín 2004). These evolutionary differences between human and rodent degradomes can be explained by two different mechanisms. The first one proposes that the increased number of proteases in rodents could be due to an expansion of protease-coding genes in their genomes after the rodent and human lineages diverged about 75 million years ago. Alternatively, the reduced number of proteases in the human degradome could be a consequence of the loss or inactivation of protease-coding genes in this lineage. Comparison of the degradomes of these mammalian species has shown that both mechanisms have been acting during evolution yielding the current differences in the complexity of mammalian degradomes.

In relation to the first mechanism, we have reported that the increased complexity of rodent degradomes is mainly due to the expansion of specific families of protease genes, most of them implicated in reproduction and host defense (Puente et al. 2005b). Several examples illustrate the existence of major differences in the functioning of proteolytic systems associated with these important physiological processes. Thus, the comparison of human and rodent proteases has revealed the presence of genes encoding placental cathepsins, testins and testases in mouse and rat, while no human orthologs could be identified for this group of proteases implicated in placental development and fertilization (Deussing et al. 2002, Puente and López-Otín 2004, Puente et al. 2003). However, the increased complexity of rodent degradomes is also due to the expansion of certain protease families which are also present in humans. This is the case for the mast cell protease subfamily of trypsin-like serine proteases, which are implicated in host defense functions. In humans, this group of proteases is composed of just four genes—cathepsin G, chymase and granzymes B and H—which are clustered in a small region of chromosome 14q11. Interestingly, a closer look at the syntenic regions in mouse

chromosome 14C1 and rat chromosome 15p13 has shown that this cluster of proteolytic genes has undergone a large expansion, and now contains up to 17 protease genes in mouse and 28 in the rat (Puente and López-Otín 2004). Similarly, the genes encoding kallikreins are also located in a cluster with 26 genes in mouse, 23 in rat and only 15 in human. As reflected by these examples, expansion of proteolytic genes has contributed to the increased complexity of rodent degradomes. Remarkably, these differences correspond almost exclusively to genes implicated in reproductive or immunological functions, indicating that these processes have been major forces acting during mammalian evolution. The precise knowledge of these differential genes will be extremely useful when interpreting experimental data obtained using animal models.

The occurrence of rodent-specific protease subfamilies constitutes an extreme case of differences between human and rodent degradomes. Despite the existence of numerous examples that illustrate the importance of gene expansion events during the evolution of rodent degradomes, we have been unable to identify similar expansions in the human genome, with the single exception of a recent duplication involving the gene encoding MMP23, which has originated two almost identical copies of this gene (MMP23A and MMP23B) (Gururajan et al. 1998, Puente et al. 2005a). Based on these data, it appears that gene expansion could be sufficient to explain the larger number of genes present in rodent degradomes when compared to that of humans. However, the loss of protease-coding genes in the human genome has also contributed to the observed differences with rodents. In fact, if for each mouse and rat protease gene absent in humans we analyze the corresponding region in the human genome, we will be able to identify a human sequence with high similarity to the murine counterpart in about 30 cases. However, a detailed analysis of this sequence will rapidly reveal the presence of premature stop codons, frameshifts or partial gene deletions which have contributed to the inactivation of this protease gene in the human genome. Therefore, pseudogenization has also been an important mechanism contributing to the increased degradome complexity of rodents when compared to humans. Interestingly, a detailed comparative analysis has revealed that most of the pseudogenized proteases which are still functional in rodents are involved in reproductive processes or in immunological functions, reinforcing the importance of these processes during mammalian evolution.

Complexity of the Protease Inhibitor Repertoire

The overall picture emerging from this comparison of human and rodent degradomes suggests that the larger number of rodent proteases would result in an increased proteolytic activity in rodent tissues. Therefore, it is tempting to speculate that additional mechanisms could have evolved to compensate this increase in proteolysis. As an initial approach to address this question, it should be of interest to investigate whether the genes encoding protease inhibitors are different between these species, as this group of proteins is responsible for the inhibition of specific

proteases under physiological conditions. In this regard, determination of the protease inhibitor complement in the genomes of human, mouse and rat has shown that changes in this group of genes might compensate the increased proteolytic potential of rodent tissues. Thus, the repertoire of protease inhibitor genes present in the human genome consists of more than 156 members, while mouse and rat show a higher complexity, with 199 and 183 members, respectively (Puente and López-Otín 2004).

Similar to the case of protease-coding genes, the increased complexity of protease inhibitor genes in rodents is mainly due to the expansion of gene clusters in these species. In fact, a detailed genomic analysis has revealed that a series of protease inhibitor genes expanded in the rodent genomes belong to groups which specifically inhibit some of the protease families which were also expanded in these species. An interesting case is that of a group of serine protease inhibitors of the serine proteinase inhibitor B (SERPINB) family. This group of inhibitors is located in human chromosome 6p25 and is composed of three different genes (SERPINB-1, -6 and -9), while the syntenic regions in mouse chromosome 13A4 or rat 17p12 have undergone a gene expansion process resulting in the presence of eight functional genes encoding SERPINB inhibitors in rat and fifteen in mouse. Similarly, the cystatin gene family, encoding a group of protease inhibitors with high specificity for cathepsins, has also been expanded in rodents. Together, these data suggest that the expansion of protease inhibitor genes in the rodent genomes might constitute a general mechanism to compensate the increased proteolytic activity in rodent tissues which might result from the expansion of protease-coding genes.

Applications of the Comparative Analysis of Mammalian Degradomes

One of the main conclusions that can be raised from the genomic comparison between the human degradome and those of mouse and rat is the increased complexity of the protease complement in these species. Taking into account that these rodents are the most widely used animal models to investigate human diseases, the characterization of their degradomes constitutes a valuable resource to evaluate experimental data obtained with these animals. In this regard, the presence in rodents of large protease families with high sequence identity and similar substrate specificity among their members might generate compensatory mechanisms which complicate the analysis of animals deficient in specific protease genes. Moreover, this increased protease complexity should be taken into consideration when studying the efficacy of novel protease inhibitors. In fact, it is possible that the existence of other family members not affected by the inhibitor might be able to compensate the biological function performed by the target protease, leading to discouraging results with compounds which otherwise could be useful in humans or if experimented in other animal models. The understanding of the different complexity of human and mouse degradomes will be helpful to anticipate