1

Proteases: a primer

Nigel M. Hooper[1]

Proteolysis Research Group, School of Biochemistry and Molecular Biology, University of Leeds, Leeds LS2 9JT, U.K.

Abstract

A protease can be defined as an enzyme that hydrolyses peptide bonds. Proteases can be divided into endopeptidases, which cleave internal peptide bonds in substrates, and exopeptidases, which cleave the terminal peptide bonds. Exopeptidases can be further subdivided into aminopeptidases and carboxypeptidases. The Schechter and Berger nomenclature provides a model for describing the interactions between the peptide substrate and the active site of a protease. Proteases can also be classified as aspartic proteases, cysteine proteases, metalloproteases, serine proteases and threonine proteases, depending on the nature of the active site. Different inhibitors can be used experimentally to distinguish between these classes of protease. The MEROPs database groups proteases into families on the basis of similarities in sequence and structure. Protease activity can be regulated *in vivo* by endogenous inhibitors, by the activation of zymogens and by altering the rate of their synthesis and degradation.

Introduction

For many researchers, proteases are often considered to be unwanted biological pests. Many do their utmost to inactivate proteases in order to prevent the breakdown of their particular protein of interest. For others, proteases are tools to be used to selectively destroy or chop up a protein prior to its further analysis, e.g. the digestion of a protein with trypsin prior to sequencing the smaller tryptic fragments. However, more and more

[1]*E-mail: n.m.hooper@leeds.ac.uk*

researchers are recognizing that proteases are often key players in a wide range of biological processes; for example, in regulating the cell cycle, cell growth and differentiation, antigen processing and angiogenesis. In addition, it is becoming apparent that the aberrant functioning of certain proteases may be involved in several disease states, including Alzheimer's disease, in cancer metastasis and in inflammation. An understanding of the role played by proteases in these processes may provide the opportunity for therapeutic intervention, and inhibitors of certain proteases have already proved to be effective therapeutic agents in hypertension and heart failure, some forms of cancer, and against certain viruses. The aim of this volume is to highlight some of the more recent developments in this area and to provide an insight into the future of protease research. When one considers that almost 2% of the human genome codes for proteases [1], it is clear that there is a lot still to be learned. The remainder of this chapter is devoted to a brief introduction to the basic terminology used in protease biology.

Definition of proteases

A protease is defined as an enzyme that hydrolyses one or more peptide bonds (Figure 1) in a protein or peptide [2]. Thus, proteases can, potentially, degrade anything containing a peptide bond, from a dipeptide up to a large protein containing thousands of amino acids. However, many proteases have a preference for protein substrates, while others will only cleave short peptides or even just dipeptides. As these enzymes hydrolyse peptide bonds, some have argued that they all should be termed 'peptidases', and that the term 'protease' be restricted to those peptidases that hydrolyse proteins. Other commonly found terms in the literature include 'proteinase' and 'proteolytic enzyme'.

Cleavage-site specificity

The terminology used in describing the cleavage-site specificity of proteases is based on a model proposed by Schechter and Berger [3]. In this model, the catalytic site is considered to be flanked on one or both sides by specificity subsites, each of which is able to accommodate the side chain of a single amino acid residue (Figure 2). By convention, the substrate amino-acid residues are called P (for peptide) and the subsites on the protease that interact with the substrate are called S (for subsite). The subsites are numbered outwards from the catalytic site, S1, S2, S3, etc. towards the N-terminus of the substrate, and S1′, S2′, S3′, etc. towards the C-terminus (Figure 2). The side chains of the amino-acid residues in the substrate that these sites accommodate are numbered P1, P2, etc. and P1′, P2′, etc., outwards from the scissile peptide bond (see Figure 2). The residues are usually not numbered beyond P6 on either side of the scissile bond. Different proteases have different requirements for subsite interactions to determine the specificity of cleavage. For example, the S1 subsite of trypsin has a marked preference for the binding of basic

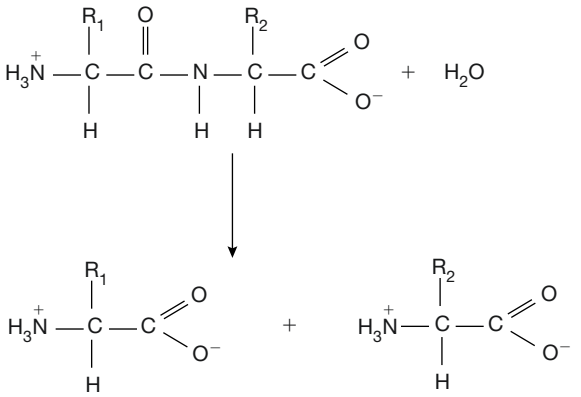

Figure 1. Peptide bond hydrolysis by a protease

amino acid residues (arginine and lysine), while interactions with several of the subsites further away from the scissile bond are critical for substrate binding to renin, the protease involved in the renin–angiotensin system (see Chapter 10), and to the caspases, proteases involved in apoptosis (see Chapter 2).

Classification of proteases

Why classify proteases? First, classification aids researchers and students in understanding the terminology in this large, and often confusing, field of research. Secondly, the grouping together of enzymes in families on the basis of sequence and structural information aids in the elucidation of common catalytic, biosynthetic processing and regulatory mechanisms. Finally, such classification is invaluable in elucidating the function of newly identified proteases. This is particularly relevant in the context of proteases that are

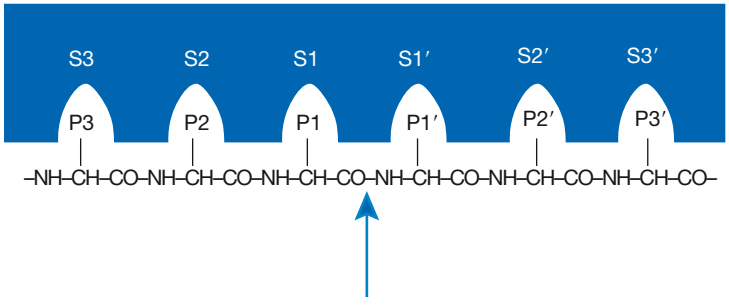

Figure 2. The Schechter and Berger [3] nomenclature for binding of a peptide substrate to a protease

The protease is represented as the blue shaded area. P1, P1′, etc. are the side chains of the six amino acids surrounding the peptide bond to be cleaved (indicated by the arrow) in the substrate. S1, S1′, etc. are the corresponding subsites on the protease.

initially identified on the basis of sequence similarity from screening genome databases [1], rather than in the more traditional way of isolating an activity that cleaves a particular protein or peptide substrate, followed by its purification and experimental characterization. Mining of genome databases for novel proteases is dealt with in more detail in Chapter 14.

Proteases can be classified on the basis of the position within a peptide of the peptide bond that is cleaved. Thus, endopeptidases cleave internal peptide bonds, while exopeptidases cleave the terminal bonds (Figure 3). Exopeptidases can be further subdivided into aminopeptidases or carboxypeptidases, depending on whether they cleave the N-terminal or C-terminal peptide bond respectively (Figure 3). Proteases are also classified on the basis of the catalytic mechanism, that is, the nature of the amino acid residue or cofactor at the active site that is involved in the hydrolytic reaction. Thus, aspartic proteases, such as the HIV protease (see Chapter 9) and renin (see Chapter 10), have a critical aspartate residue that is involved in catalysis. Metalloproteases have a bivalent metal ion, usually zinc but sometimes cobalt, iron or manganese, at the active site. Examples of metalloproteases include the matrix metalloproteases (see Chapter 3), methionine aminopeptidases (see Chapter 6), angiotensin-converting enzyme and neprilysin (see Chapter 10), and the ADAMs (a disintegrin and metalloproteinase domain) family of proteases (see Chapter 11). In the aspartic and metalloproteases, the nucleophile that attacks the peptide bond of the substrate is an activated water molecule, whereas in the other protease groups the nucleophile is part of an amino acid at the catalytic site of the protease.

Those proteases in which the nucleophile is the sulphydryl group of a cysteine residue are termed cysteine proteases, typified by the caspases that are involved in programmed cell death (see Chapter 2). In serine proteases the catalytic mechanism depends upon the hydroxy group of a serine residue acting as the nucleophile that attacks the peptide bond. Examples of serine proteases include chymotrypsin and trypsin (digestive enzymes of the intestine), and the proteases involved in the blood clotting cascade (see Chapter 8). In a small number of proteases the catalytic mechanism depends on the hydroxy group of a threonine residue, the so-called threonine proteases. These are exemplified by

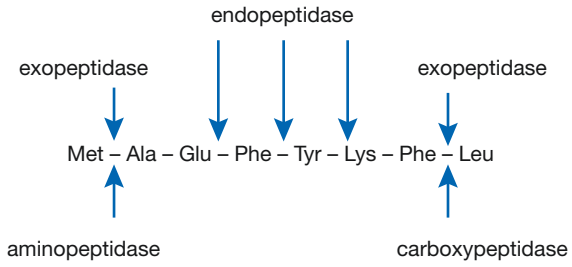

Figure 3. Cleavage site specificity of proteases

the catalytic subunit of the proteasome (see Chapter 5). At present, 3% of the proteases in the human genome have been identified as aspartic proteases, 23% as cysteine proteases, 36% as metalloproteases and 32% as serine proteases [1].

Over the past decade, Alan Barrett and his colleagues in Cambridge, U.K., have developed a more detailed classification system for proteases, the MEROPS database [4–6]. This is available online (www.merops.ac.uk) or in hard copy as the *Handbook of Proteolytic Enzymes* [7]. In the MEROPS database, proteases are classified by structural similarities in the parts of the molecules that are responsible for their enzymic activity. They are grouped into families on the basis of amino-acid sequence homology, and the families are assembled into clans based on evidence, usually similarities in tertiary structure, that they share a common ancestry (Figure 4). This classification forms a framework around which a wealth of supplementary information about the proteases is organized, including images of three-dimensional structures, amino-acid sequence alignments, comments on biomedical relevance, and literature references. A set of online searches provides access to information about the location of proteases on human chromosomes and their substrate specificity. As the MEROPS database is updated regularly, it provides an extremely valuable resource for protease researchers (see also Chapter 14).

Inhibition of proteases

The four major classes of proteases (aspartic, cysteine, metallo and serine) can be distinguished experimentally using class-specific inhibitors (Table 1). For example, chelators such as EDTA or 1,10-phenanthroline remove the critical metal ion from the catalytic site of metalloproteases, thereby inactivating them.

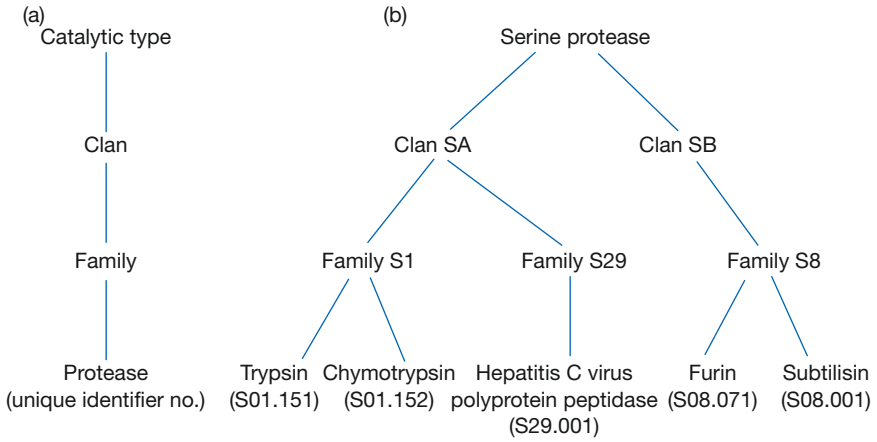

Figure 4. Overview of the MEROPS protease classification system for proteases
(a) Summary showing the relationship between catalytic type (aspartic, cysteine, metallo or serine), clan, family and individual protease within the MEROPS database. (b) Schematic showing the relationship of five serine proteases within the MEROPS database. For more information see [7].

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.