

BACKGROUND

SIGNALLING SCISSORS: NEW PERSPECTIVES ON PROTEASES

Joachim Pietzsch

167 years ago, long before anyone knew that there were such things as enzymes or what proteases were, the first protease was given its name: pepsin. This was what the founder of the cell theory in animals, Theodor Schwann, called the substance that is responsible for the breakdown of proteins in the stomach, although he could not characterize it chemically.

We now know that enzymes, in their role as biocatalysts, are the most important control points in living organisms; and the largest class of enzymes is the proteases. Proteases act as the molecular mediators of many vital processes, from embryonic development to the immune response and wound healing, and assist in the processing of cellular information. Despite this, proteases, as if they were all standing in the shadow of pepsin, are still generally thought of as aggressors, as thuggish digestive enzymes that break down proteins into their individual components for further use or for elimination.

Proteins are made up of amino acids which, following the elimination of water molecules, are linked together by peptide bonds. Proteases — which are themselves proteins and are also known as peptidases or proteinases — catalyse the cleavage of peptide bonds by activating them and exposing them to attack by water molecules — that is, they aid their hydrolysis. The catalytic centre of a protease is generally located in a groove on the surface of the enzyme. The protease holds the peptide bond of its protein substrate in this groove, as if gripped by a pair of pliers.

Meet the family

Proteases can be divided into five different groups, depending on the type of molecule in the groove that carries out the actual work of catalysis. Serine proteases attack the peptide bond of their substrate using the hydroxyl group of the side chain of the amino acid serine, which is present in their catalytic centre. Threonine proteases act in a similar way. Cysteine proteases use the

sulphur–hydrogen bond of a cysteine residue to initiate cleavage of the peptide bond. The acidic carboxyl groups of two aspartyl residues carry out this function in aspartyl proteases. Finally, metalloproteases (also known as metalloproteinases) have a tightly bound zinc atom in their catalytic centre.

Over the course of evolution, each of these groups of proteins has been specialized to perform certain functions. For example, the coagulation of blood is the responsibility of serine proteases, whereas threonine proteases are essential, in particular, for the functioning of the proteasome, a large barrel-shaped protein-degrading apparatus in the cell (see ‘Regulation through degradation’). Programmed cell death, otherwise known as apoptosis, is brought about by cysteine proteases; digestion could not occur without aspartyl proteases (which also include the HIV-1 protease from HIV); and a subgroup of metalloproteases is essential for the breakdown and rebuilding of the extracellular matrix.

Proteases do not attack their protein substrates at random. Rather, they display a high degree of specificity. This specificity is not defined by the nature of the catalytic centre alone; on either side of the catalytic centre there is a series of binding sites that favour particular amino acids in their substrates (FIG. 1). For example, in the case of the digestive enzyme trypsin; these binding sites are targeted at amino acids with positively charged side chains — they are 10,000 times more likely to bind than other side chains. So, only peptide bonds that adjoin amino acids with positively charged side chains are cleaved by trypsin to any substantial degree. To be effective, proteases must therefore interact precisely with their substrates and bind firmly to them.

As would be expected from the multiplicity of their functions, proteases come in different shapes and sizes to suit their substrates. Some proteases are so small that they are made up of little more than their catalytic centre, whereas others are huge molecules in whose

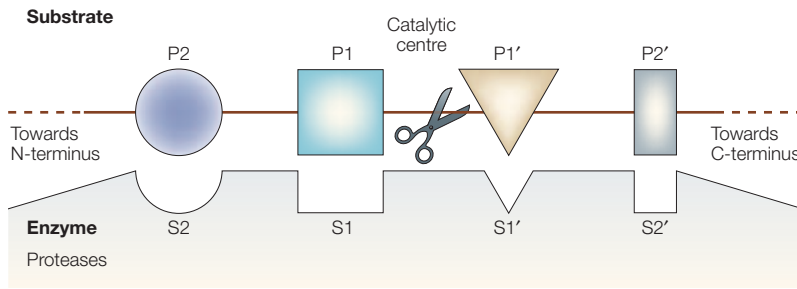


Figure 1 | **Specificity of protease binding.** The specificity of protease binding is not defined solely by the nature of the catalytic centre. On either side of the catalytic centre there is a series of binding sites (S) that favour particular amino acids (P). This level of specificity ensures that proteases carry out their task only when truly needed. Reproduced from Barrett, A. J. (December 2000) Proteases. In: *Nature Encyclopedia of Life Sciences*. London: Nature Publishing Group. <http://www.els.net/> [doi:10.1038/npg.els.0000670] © Macmillan Magazines Ltd.

entwined coils the catalytic centre seems to disappear. Yet others aggregate to form powerful complexes in which, as in proteasomes, they carry out their work together, as if in collaboration. Evolution, it seems, has been more prolific in the development of a wide variety of proteases than in any other class of enzymes.

This favouritism towards proteases in biological systems might also be due to the fact that they always act in an uncompromising manner. The hydrolysis of a peptide bond is irreversible, partly because the equilibrium of the reaction in which two amino acids join together lies on the side of the hydrolysis product, and partly because proteases do not have a direct counterpart that can reverse the cleavage of a peptide bond. That is because the synthesis of proteins from messenger RNA only takes place on ribosomes. The repair of proteins that have been cleaved is not part of the business of life.

So, once a protease has cleaved a protein, there is no turning back. Because of this, active proteases must only be deployed in biological systems when they are really needed; for example, to stop bleeding after an injury. For this reason, proteases, whether inside or outside the cell, usually remain in standby mode. They are expressed in the form of inactive precursors, so-called zymogens, and become activated by proteolytic cleavage only when they are needed. This activation of proteases in living organisms is often organized as a multi-stage process, arranged in complex cascades that serve to ensure both better control and the targeted amplification of the desired biological effect. These cascades are made up almost exclusively of proteases, one activating the next by cleavage of the zymogen precursor. The complexity of these cascades is so great that, even today, it has not been possible to clarify the mode of action of even one in its entirety.

Proteases and blood clotting

This lack of complete knowledge even applies to the best-understood protease cascade, blood coagulation. The classical model of blood coagulation has coagulation factor X — a serine protease — at its centre (FIG. 2). Two different cascades activate this factor to produce factor Xa. In the common final stages of blood coagulation, factor Xa converts the zymogen prothrombin

(factor II) into thrombin (factor IIa). Finally, thrombin cleaves fibrinogen to give fibrin monomers which, with the aid of factor XIIIa — a blood proenzyme but not a protease — bind together activated blood platelets to form the solid matrix of a blood clot.

The two cascades that lead to the activation of factor X are called the intrinsic and extrinsic pathways of coagulation. The first cascade is called intrinsic because all of its components are present in the blood. It is triggered by reduced blood flow or by contact with pathologically altered vessel walls, and leads, via factors XII, XI and IX, and VIII, to the activation of factor X. A by-product of the conversion of factor XII into factor XIIa is bradykinin, a nine-amino-acid peptide that is important in other metabolic pathways. The extrinsic pathway — which is activated following injury — is triggered by contact between blood and a protein called tissue factor that is normally found outside the vessels. That leads to the activation of coagulation factor VII, which acts directly on factor X, and also, higher up in the cascade, potentiates the activation of factor IX.

To maintain the equilibrium of the blood flow and circulation, the blood coagulation cascade, which is, after all, only intended as a protective mechanism in emergency situations, requires a counterweight. This is provided by fibrinolysis, the process of dissolving the tangled fibrin mass, which is also carried out by proteases. A wealth of endogenous inhibitors, chiefly serpins and statins, control the activity of the coagulation proteases and ensure that their effects are limited both in time and in space. Haemostasis is regulated so finely and with such amazing precision in higher organisms that we are only gradually beginning to grasp the complexity of the cascades and the variety of interdependencies involved in its regulation.

Admittedly, the more deeply the mechanisms of blood coagulation are investigated, the more fragile the classical text-book model of the intrinsic and extrinsic pathways seems. The actual significance of the intrinsic pathway and the processes that lead to the activation of factor XII have become questionable. It remains unclear whether tissue factor gives a decisive push towards coagulation, and at what point and how it does so. A large part of the process seems to take place in an interwoven fashion on the surface of activated blood platelets. But even today no-one knows with certainty what molecular command initiates the cascade that leads to the coagulation of blood.

Signalling death

To date, we know much less about apoptosis than about the coagulation of blood. It too occurs via protease cascades. Apoptosis is programmed cell death, without which life would be impossible. It is the means by which our bodies control the targeted removal of cells that have completed their job, have become surplus to requirements in the course of embryonic development, or display errors in their genetic information. Life persists and adapts through apoptosis: it is estimated that about 100 billion cells in the body of an adult human die every day, to be replaced by new ones, and most of them die by apoptosis.

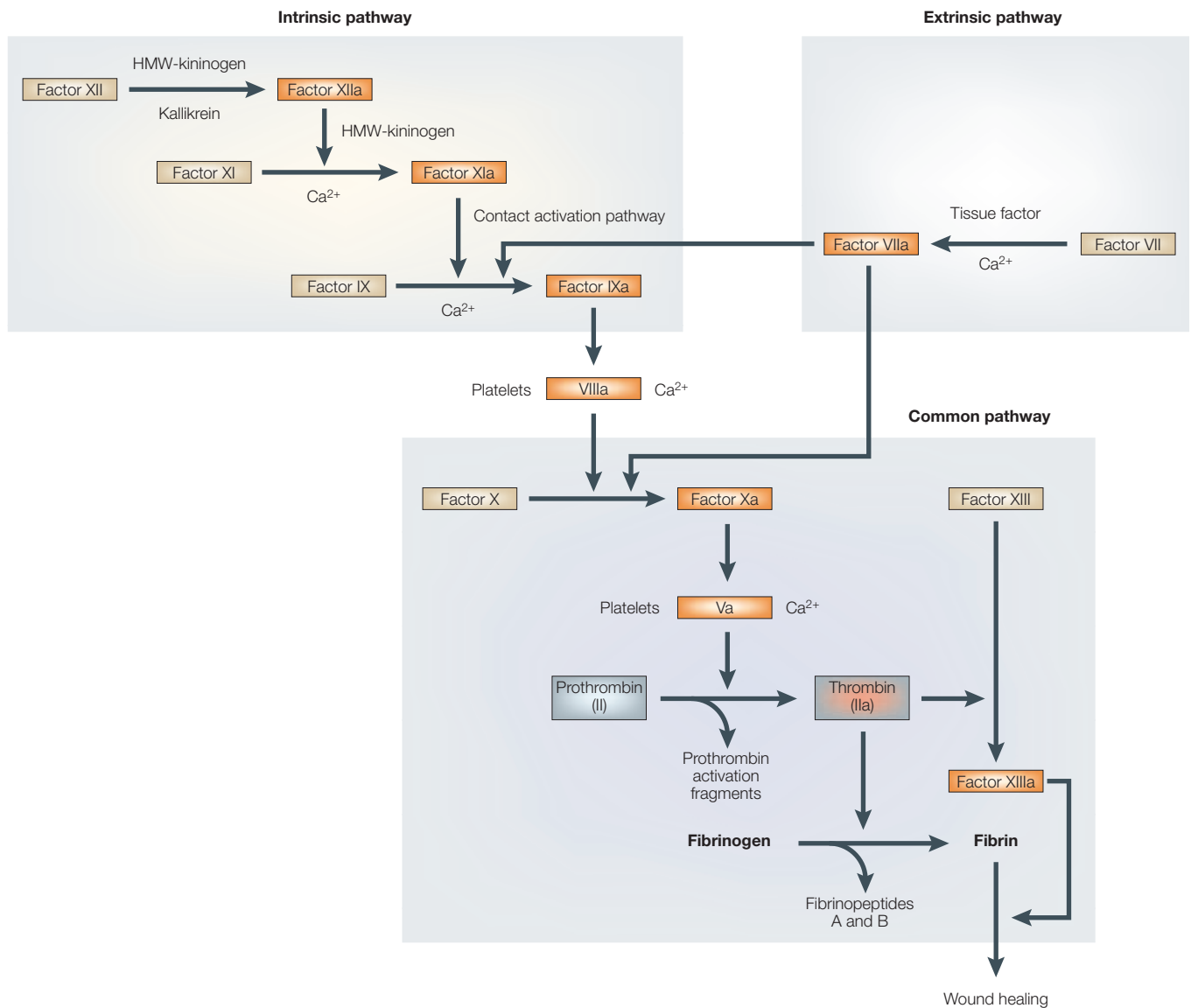


Figure 2 | **Proteases and blood coagulation.** The generation of blood clots involves a complex series of activation events, in which inactive proteins are converted to active forms after cleavage by proteases. This is initiated by the ‘initiating’ or ‘extrinsic’ pathway which activates factor X, a serine protease. Each activation results in the generation of a new protease which acts on a subsequent protein, to ultimately produce fibrin, which forms the blood clot. Modified from Van Lint, J. & Vandenheede, J. R. (March 2000) Regulatory Cascade. In: *Nature Encyclopedia of Life Sciences*. London: Nature Publishing Group. <http://www.els.net/> [doi:10.1038/npg.els.0000867] © Macmillan Magazines Ltd.

Our health depends on the correct amount of apoptosis occurring in the body. Too much apoptosis can lead to the degeneration of vital structures: diseases such as rheumatoid arthritis and Alzheimer’s disease are among the consequences. Too little apoptosis can lead to uncontrolled growth of cells, and therefore to cancer. Specific cysteine proteases, called caspases, linked into a proteolytic cascade, are the central controlling elements of apoptosis (FIG. 3). In response to endogenous or exogenous signals, they take on both the initiation and the execution of the cell-death programme. Once again, endogenous protease inhibitors look after the fine regulation of the process. The apoptosis cascade has been

clarified in outline in the course of the past decade: we know the key elements and how they relate to each other. We still do not know much about the activation of the first proteases in the cascade; that is, about how the process starts. We are also still largely in the dark about the interactions with other signal cascades in the cell — in particular with respect to the differing cell specificity of apoptotic programmes.

Therapeutic targets

The key role of proteases in many physiological and pathophysiological processes makes them attractive targets for pharmaceutical research. Protease inhibitors are

potent drugs. Their best-known representatives include angiotensin-converting enzyme (ACE) inhibitors and HIV protease inhibitors.

ACE is a metalloprotease that catalyses the conversion of angiotensin I into angiotensin II within the renin-angiotensin system, which leads to vasoconstriction. ACE inhibitors, introduced in the 1980s, were originally used as antihypertensives, but have significantly improved the treatment of other cardiovascular diseases, and are now used to treat heart failure and even to prevent heart attacks in at-risk patients, such as people with diabetes.

HIV-1 protease is used by HIV to 'tailor' its viral coat after replication in the host cell. The introduction of HIV protease inhibitors in the mid-1990s — in combination with reverse transcriptase inhibitors — led to a dramatic improvement in AIDS treatment and to a decidedly better prognosis for those infected with HIV.

However, the success of ACE inhibitors and HIV protease inhibitors should not hide the fact that proteases are extremely demanding target molecules for pharmaceutical research. In the case of many other proteases, therapeutic hopes have already been dashed (see 'The trouble with inhibitors'). This is partly because of the chemical nature of proteases: effective inhibitors must be able to satisfy the substrate specificity mentioned above but, for reasons of bioavailability, must not be proteins or peptides. Another factor is the complexity of the protease cascades: finding the right target molecule which can be inhibited to produce a therapeutic effect, not just *in vitro* but also *in vivo* is a Herculean task.

It is more the rule than the exception that protease inhibitors that look good in preclinical tests fail in clinical trials. The interdependencies and redundancies in such cascades are apparently too complex for us to predict the effects of a given therapeutic intervention. Our knowledge about what proteases regulate and how proteases are regulated is too fragmented. Because the signal is amplified enormously in the course of a cascade, with the number of molecules often increasing many thousands of times, it would be best to have an inhibitor that intervenes early on in a cascade. But even in the case of the well-understood blood-coagulation cascade it has not been possible to develop, for example, a factor VII inhibitor for clinical use. Even at an early stage in the cascades, the tangle of the proteases involved, their endogenous inhibitors, cofactors and other signal molecules is too confusing to enable us to form an accurate picture of the process that we wish to influence with treatments.

The plot thickens

Paradoxically, our overall picture of proteases becomes fuzzier the more we find out about them. Research has brought more and more details to light: there are even active proteases in the cell membrane; that is, in a hydrophobic environment (see 'RIPping and folding: intramembrane proteolysis'); the logistics of the busy protein transport within cells is controlled by proteases; the immune system is supplied with information from proteases via antigens; several receptors on important signal transmission pathways are activated by proteases (see 'Signalling from the outside in'). It is not yet possible to piece together these parts of the protease mosaic to give a conclusive picture.

So, protease researchers are turning to technologies developed in genomic and proteomic research (see 'The great substrate hunt'). This promises a new perspective in the near future: an overview of all the proteases that exist and their substrates, and a panoramic view of the proteolytic networks and their biological significance. It is estimated that up to 1,200 human genes (4.5% of our genome) encode proteases, with the majority of these genes remaining to be discovered. The current version (22 September 2003) of the MEROPS database, which specializes in proteases, lists 475 known and putative proteases and 103 homologues to known proteases in humans.

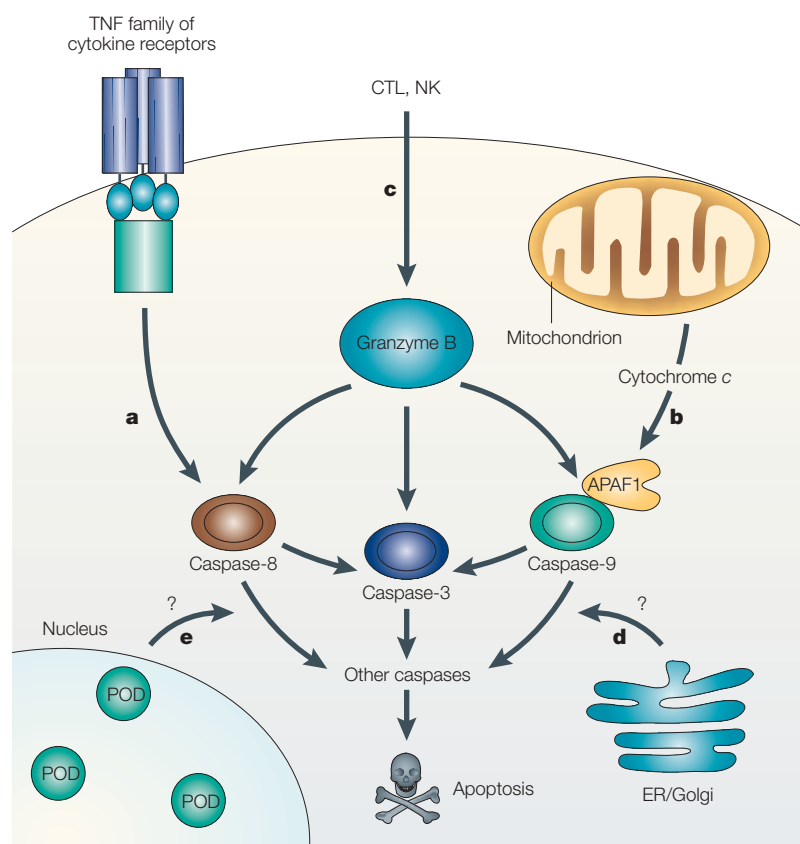


Figure 3 | **Simplified pathways of caspase activation of apoptosis.** **a** | Members of the tumour-necrosis factor (TNF) family of cytokine receptors bind to caspases to form a complex that triggers caspase activation and leads to apoptosis. **b** | Cytochrome c is released from mitochondria, binds to and activates apoptotic protease-activating factor 1 (APAF1), allowing it to then bind and activate caspase-9. This so-called apoptosome causes cell death. **c** | Immune cells, such as cytotoxic T lymphocytes (CTLs) and natural killer (NK) cells, inject apoptosis-inducing proteases, particularly granzyme B, into target cells. **d** | A caspase activation pathway that is linked to endoplasmic reticulum (ER)/Golgi body stress has been proposed in the mouse, for which caspase-12 might have a key role. **e** | The nucleus might also have a regulatory pathway. Several proteins that can promote or induce apoptosis have been localized to discrete nuclear organelles, called PML oncogenic domains (PODs) or nuclear bodies. Modified from Reed, J. C. Apoptosis-based therapies. *Nature Rev. Drug Discov.* **1**, 1111–1121 (2002) © Macmillian Magazines Ltd.

Proteases will certainly soon have shrugged off their traditional image as destructive aggressors. Their signals influence many essential processes of human physiology. Their cascades act as elaborately spun and defining threads in the web of life. That reminds us of the three fates in Greek mythology, Clotho, Lachesis and Atropos, whose job is to create the destiny for each living being — to weave the web of life, to measure and cut. Proteases could be their most important tools. More than any other molecules, they seem to manifest the dynamics of life, the eternal rhythm of life and death.

Further reading

- Barrett, A. J. (December 2000) Proteases. In: *Nature Encyclopedia of Life Sciences*. London: Nature Publishing Group. <http://www.els.net/> [doi:10.1038/npg.els.0000670]
- Essays in Biochemistry volume 38 (ed. N.M. Hooper). Proteases in biology and medicine (Portland, London, 2002).
- López-Otin, C. & Overall, C. M. Protease degradomics: a new challenge for proteomics. *Nature Rev. Mol. Cell Biol.* **3**, 509–519 (2002).
- Leung, D. *et al.* Protease inhibitors: current status and future prospects. *J. Med. Chem.* **43**, 305–341.
- Van Lint, J. & Vandenheede, J. R. (March 2000) Regulatory Cascade. In: *Nature Encyclopedia of Life Sciences*. London: Nature Publishing Group. <http://www.els.net/> [doi:10.1038/npg.els.0000867]