# The Discovery of Proteases and Intramembrane Proteolysis

<table>
<thead>
<tr>
<th>Journal:</th>
<th>Biochemistry and Cell Biology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>bcb-2018-0186.R1</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>Invited Review</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>24-Jul-2018</td>
</tr>
<tr>
<td>Complete List of Authors:</td>
<td>Paschkowsky, Sandra; McGill University, Pharmacology and Therapeutics  Hsiao, Jacqueline Melissa; McGill University, Biochemistry  Young, Jason C.; McGill University  Munter, Lisa Marie; McGill University, Pharmacology and Therapeutics</td>
</tr>
<tr>
<td>Keyword:</td>
<td>Proteases, History, Intramembrane proteases</td>
</tr>
<tr>
<td>Is the invited manuscript for consideration in a Special Issue? :</td>
<td>CSMB Special Issue</td>
</tr>
</tbody>
</table>

[https://mc06.manuscriptcentral.com/bcb-pubs](https://mc06.manuscriptcentral.com/bcb-pubs)
The Discovery of Proteases and Intramembrane Proteolysis

Sandra Paschkowsky¹#, Jacqueline Melissa Hsiao²#, Jason Young², Lisa Marie Munter¹*

¹: McGill University, Department of Pharmacology & Therapeutics, Bellini Life Sciences Complex, 3649 Promenade Sir William Osler, Montreal, QC, Canada H3G 0B1
²: McGill University, Department of Biochemistry, McIntyre Building, 3655 Promenade Sir William Osler, Montreal, QC Canada, H3G 1Y6

#: These authors contributed equally to the work.

*Correspondence should be addressed to Lisa Marie Munter, McGill University, Department of Pharmacology & Therapeutics, Bellini Life Sciences Complex, 3649 Promenade Sir William Osler, Montreal, QC, Canada H3G 0B1, phone: 1-514-398 2159, lisa.munter@mcgill.ca

Abstract

Proteases carry out a wide variety of physiological functions. This review presents a brief history of protease research, starting with the original discovery of pepsin in 1836. Following the path of time, we revisit how proteases were originally classified based on their catalytic mechanism and how chemical and crystallographic studies unravelled the mechanism of serine proteases. Ongoing research on proteases addresses their biological roles, small molecule inhibitors for therapeutic uses, and protein engineering to modify their activities. The discovery of intramembrane proteases is more recent, beginning with the discovery of site-2 protease in 1997. Since then, different mechanistic classes of intramembrane proteases have been characterized, and many of these act in regulated intramembrane proteolysis in signaling pathways. Furthermore, the rhomboid intramembrane proteases were discovered by genetic and biochemical experiments in Drosophila, then in human cells. Research on the intramembrane proteases is expanding, as their biological importance is recognized.
Keywords
Proteases, History, Intramembrane Proteases

History of protease discovery

Early in his career, Theodor Schwann, who is often considered the father of modern cell biology, described the purification of an active component from gastric juices that was able to digest egg albumin. He named this compound pepsin, which is the first protease isolated in history (Mueller 1836). Nowadays, the discovery of pepsin is widely regarded as the beginning of modern protease research. However, it was not until 40 years after his discovery when trypsin, one of the best studied enzymes, was purified by physiologist Wilhelm Friedrich Kühne (Naturhistorisch-Medizinischer Verein 1877). Within the same context, Kühne further introduced the term ‘enzyme’ to the scientific literature in order to distinguish the difference between organisms and active compounds. Translated from German, Kühne defined enzymes as “[…] the unformed or unorganized ferments whose action can occur without the presence of organisms and outside of the same […]” (Naturhistorisch-Medizinischer Verein 1877).

For a long time, there has been some debate regarding the nomenclature of the originally discovered enzyme as protease. It was once suggested that proteases should be defined as enzymes that possess the ability to degrade proteins, while proteinases should be defined as endopeptidases. In recent years, this distinction has been mostly lost, and the terms ‘protease’, ‘proteinase’, and ‘peptidase’ are generally used interchangeably (Grassmann & Dyckerhoff 1928). For this review, proteases will be considered as a class of enzymes that carries out hydrolysis of peptide bonds.

In 1929, John Northrop provided substantial evidence that proteases are proteins, which was determined from the purification and crystallization of pepsin (J. H. Northrop 1929). He was later awarded the Nobel Prize in Chemistry in 1946 for the isolation and crystallization of several other proteases in their pure forms, including chymotrypsin, trypsin, and pepsinogen, which were published in his award-winning book, Crystalline Enzymes: The Chemistry of Pepsin, Trypsin, and Bacteriophage (J.H. Northrop 1939).

Brian Hartley proposed the classification of proteases with regard to their observed mechanisms of action in 1960, which was a new approach at the time. As opposed to biological function or specificity, he proposed four main classes of enzymes: serine, thiol (cysteine),
(aspartic) acid, and metal (metalloproteinase) proteases (Hartley 1960). Despite the abundance of proteases that have been identified since 1960, the simple concept of a mechanism-based classification of proteases remains applicable today. However, proteases that use other catalytic mechanisms have been discovered. For instance, the proteasome degrades ubiquitinated proteins in the cytosol by using threonine amino acid residues for catalysis (Ciechanover, Hod & Hershko 1978; Collins & Goldberg 2017; Lowe et al. 1995). Recently, other threonine proteases have been discovered, as well as glutamic acid proteases that have been identified only in fungi and bacteria thus far (Fujinaga, Cherney, Oyama, Oda & James 2004; Jensen, Ostergaard, Wilting & Lassen 2010; Manolaridis et al. 2013; Sims, Dunn-Coleman, Robson & Oliver 2004).

**Catalytic mechanism of serine proteases**

One of the best-studied examples of enzyme mechanics include the mechanism of serine proteases, which involve the use of a catalytic triad consisting of aspartate, histidine, and serine (Polgar 2005). Because natural amino acids are not usually sufficiently strong nucleophiles at physiological pH, the reaction mechanism involves a specific coordination of the triad residues. The mechanism includes a stabilizing hydrogen bond between the aspartate and histidine to orient the histidine residue. Additionally, this hydrogen bond increases the pKa of histidine, which thereby induces an initial deprotonation of the serine residue by histidine. The attack of the peptide bond leads to the formation of an acyl-enzyme intermediate state and the stabilization of negative charges by an oxyanion hole formed with backbone amine groups (Berg, Tymoczko & Stryer 2002).

In 1956, the catalytic importance of serine and histidine within the catalytic triad for hydrolysis was demonstrated by chemical experiments (Gutfreund & Sturtevant 1956; Smillie & Hartley 1966; J. H. Wang & Parker 1967). The crystal structure of chymotrypsin further elucidated this mechanism, establishing that the distance between serine and histidine was in fact within the anticipated hydrogen bond length (Blow, Birktoft & Hartley 1969). However, the catalytic triad is not unique to serine proteases; for example, the mechanisms behind transacylation and phosphorylation use the same fundamental principles (Chen et al. 2008; Milkowski & Strack 2004). In fact, it has been suggested that this well-understood mechanism has convergently evolved up to 25 times (Buller & Townsend 2013).
Other proteases, such as the rhomboid proteases, lack the specific aspartate, and carry out hydrolysis with the use of a catalytic dyad of histidine and serine, rather than the classical catalytic triad (Arutyunova et al. 2016; Bondar, del Val & White 2009; Xue & Ha 2012). The histidine is stabilized by other interactions in the active site, and thus, these enzymes are still considered serine proteases (Lemberg et al. 2005; Lemieux, Fischer, Cherney, Bateman & James 2007; Y. Wang, Zhang & Ha 2006). The majority of thiol proteases use a similar mechanism involving a triad, except the serine is replaced by a cysteine (Nandi et al. 2012). Conversely, a few thiol proteases, such as caspases, and several other enzymes carry out hydrolysis with a catalytic dyad (Paetzel & Dalbey 1997). Similarly, the other protease classes have more divergent mechanisms. For example, aspartate and metal proteases activate water for nucleophilic attack and do not form acyl-enzyme intermediates (Dunn 2002; Feng et al. 2007).

Field of proteolysis research

Although proteases presumably evolved as enzymes purely associated with protein catabolism, it is now clear that proteases have important roles beyond nonspecific degradation. Since the discovery of the first protease, a variety of physiological functions have been identified, including the regulation of protein interactions, protein activity, and protein anabolism (Lopez-Otin & Bond 2008). Specific functions of proteases have been identified in the regulation of various pathways, such as cell division, cell differentiation, stem cell mobilization, the blood-clotting cascade, cellular senescence, and apoptosis (Lopez-Otin & Bond 2008). It is therefore unsurprising that the disruption of protease mechanisms are implicated in many human diseases, including cancer, neurodegenerative, inflammatory, and cardiovascular diseases (Chakraborti, Chakraborti & Dhalla 2017).

As a result, there has been extensive research on small molecule protease inhibitors as potential treatments for these diseases. However, since the discovery of the first protease, it took almost 150 years before the first marketed protease inhibitor, captopril, which is an inhibitor of the angiotensin-converting enzyme, was approved by the Food and Drug Administration (FDA) (Smith & Vane 2003). One of the most successful stories in protease research is the development of protease inhibitors engineered as anti-viral drugs for the treatment of HIV infections (Richman 2001). Mechanistically, the drug has the ability to inhibit a specific viral protease, leading to viral maturation defects and subsequently decreasing the reproduction of the virus.
Likewise, protease inhibitors are successful pharmacological strategies to treat hepatitis C infections (Opar 2010). Other drugs that inhibit proteases are used in the treatment of hypertension (angiotensin-converting enzyme inhibitor), cancer (proteasome inhibitor), and blood coagulation (thrombin and factor Xa inhibitors) (Drag & Salvesen 2010).

Protein engineering of proteases has modified their mechanisms to alter their specificity or allow regulation. For example, a ‘metal-switch’ was engineered into trypsin: when a metal-binding element is introduced, the enzyme is inhibited by increasing concentrations of Cu$^{2+}$ (Higaki, Fletterick & Craik 1992). On the other hand, engineering proteases that are resistant to inhibitors is becoming widespread for specific therapeutic applications, such as for proteases used in wound healing and clot dissolving that may then possess higher efficacy and may not be inactivated by natural inhibitors present in the tissue (Drag & Salvesen 2010).

**Discovery of intramembrane proteolysis**

The proteases that were discovered early in protease research are predominantly water-soluble, with some proteases that are membrane anchored, such as the family of ‘A disintegrin and metalloproteases’ (ADAMs) (Reiss & Bhakdi 2017). However, cellular membranes constitute a significant proportion of the cell (Almen, Nordstrom, Fredriksson & Schioth 2009). The idea that there exists a mechanism to cleave membrane spanning proteins was first introduced in the early 1990s from observations within the Alzheimer’s Disease research field, which suggested that the amyloid precursor protein (APP) is cleaved within its transmembrane sequence (TMS) (Selkoe 1994). However, it was not until 1997 when the first intramembrane protease was discovered, namely, the human site-2 protease (S2P) (Rawson et al. 1997). This delay was partially due to controversies regarding the idea of hydrolysis in an environment where water was presumed not to be present (M. K. Lemberg & M. Freeman 2007). The crystallographic structure of the *Escherichia coli* protease GlpG provided unambiguous evidence that the active center of the enzyme lies within the plane of the membrane, but is still accessible to water (Y. Wang et al. 2006). This notion was confirmed for various other proteases (Feng et al. 2007; Lemieux et al. 2007; Takagi, Tominaga, Sato, Tomita & Iwatsubo 2010; Tomita 2014). Intramembrane proteases cleave within TMSs of their α-helical substrates. It is believed that the steric hindrance of the amino acid side chains in this secondary structure may prevent access of the protease to the substrate peptide bond (Wolfe 2009). The protease would therefore need to
create a hydrophilic environment, while altering the conformation of the substrate to allow hydrolysis. Although intramembrane proteases have minimal sequence similarities to water-soluble proteases, the overall mechanism of hydrolysis is fairly similar (Wolfe 2009).

It has been estimated that more than a quarter of all human proteins possess α-helical transmembrane domains (Almen et al. 2009), raising the possibility that intramembrane proteases may have a role in maintaining membrane protein homeostasis. An increasing number of functionally important membrane proteins are found to be regulated by intramembrane proteases, which have been shown to play crucial roles in certain cellular functions, such as apoptosis and lipid metabolism (McCarthy, Coleman-Vaughan & McCarthy 2017). Four classes of intramembrane proteases have been discovered thus far: the S2P-type metalloproteases; the intramembrane aspartyl proteases, which include γ-secretase, signal peptide peptidase (SPP) and SPP-like (SPPL) peptidases; the superfamily of rhomboid serine proteases; and the glutamyl protease, Rce1 (Manolaridis et al. 2013; Rawson et al. 1997; Ray et al. 1999; Wolfe, Xia, Moore et al. 1999; Wolfe, Xia, Ostaszewski et al. 1999).

The first intramembrane protease discovered, S2P, was shown to be of significant importance for endogenous cholesterol metabolism and the regulation of other processes by cleaving and activating transcription factors, such as SREBP1, ATF6, and CREB (Brown, Ye, Rawson & Goldstein, 2000; Goldstein & Brown 2015; Rawson et al. 1997; Y. Wang et al. 2006). The identification of γ-secretase, an aspartyl intramembrane protease, marked the beginning of another extensive line of research, mainly due to its significant role in Alzheimer’s disease (Rawson et al. 1997; Wolfe, Xia, Moore et al. 1999; Wolfe, Xia, Ostaszewski et al. 1999). Interestingly, over 100 substrates of γ-secretase have been identified thus far, suggesting that γ-secretase has low substrate specificity, and it has been accordingly referred to as the ‘proteasome of the membrane’ (Kopan & Ilagan 2004). However, for the majority of intramembrane proteases, only a few substrates have been described, and this may either indicate a more specific function of intramembrane proteases or demonstrates the difficulty to identify substrates. SPP and SPPLs were shown to play major roles in other biological processes, such as protein degradation, the immune response, and protein glycosylation (Mentrup, Fluhrer & Schroder 2017). The superfamily of rhomboid proteases are linked to Parkinson’s Disease, diabetes, and cancer, with specific roles in the regulation of mitochondrial morphology, receptor signaling, and protein degradation at the endoplasmic reticulum (ER) (Dusterhoft, Kunzel & Freeman 2017).
Regulated intramembrane proteolysis (RIP)

In 2000, Goldstein and Brown introduced the term “Regulated intramembrane proteolysis” (RIP) upon identification of the SREBP transcription machinery for the regulation of lipid and cholesterol homeostasis (Brown et al. 2000). In general, RIP mediates intracellular signaling, as well as signalling between cells and the extracellular environment through the regulation of membrane proteins. RIP is involved in many processes, which include cellular metabolism, cellular adhesion, and proliferation (Marambaud et al. 2002; Rawson et al. 1997; Yoshida et al. 2013). This mechanism generally consists of two sequential cleavage events by two different proteases on a single TMS substrate. The first cleavage that occurs is described as ‘shedding’, which generates an extracellular or luminal proteolytic fragment. Subsequently, the resulting membrane-bound fragment is then cleaved in its TMS, producing an intracellular fragment that in some cases can act as a transcription factor in the nucleus (Brown et al. 2000). For the well characterized SREBP, low cholesterol triggers its traffic from the ER to the Golgi, where its luminal domain is cleaved by serine protease site 1 protease (S1P) that is anchored to the membrane and the cytosolic domain released by intramembrane S2P. The cytosolic fragment activates transcription of cholesterol synthesis genes (Brown et al. 2000). Many intramembrane proteases, although not all, are involved in RIP signaling. It should be noted that intramembrane proteases are also known as intramembrane-cleaving proteases (I-CLiPs).

The discovery of rhomboid proteases

Rhomboid proteases are a distinct and conserved family of intramembrane proteases. The first rhomboid protease was described years before the idea of intramembrane cleavage appeared in the literature in the early 1990s (Selkoe 1994). More specifically, Nusslein-Vollhard and Wieschaus were the first to describe the rhomboid gene in Drosophila when they screened for embryonic lethal mutations in the early 1980s (Jurgens, Wieschaus, Nusslein-Volhard & Kluding 1984; Nusslein-Volhard, Wieschaus & Jurgens 1984). The mutation identified was in the rhomboid (now called rhomboid-1) gene on chromosome 3, and characterized by the phenotype of a rhombus-shaped head of the embryo and the lack of all median denticle bands (Jurgens et al. 1984). In 1990, rhomboid-1 was also shown to play a crucial role in the development of the peripheral nervous system and the establishment
of the dorsoventral axis in *Drosophila* (Bier, Jan & Jan 1990). Upon cloning *rhomboid-1*, sequence analysis suggested that the predicted protein was likely to be an integral membrane protein with three to seven TMSs (Bier et al. 1990). It was identified in 1996 by Golembo *et al.* that the EGF receptor pathway in *Drosophila* is activated when its ligand, *spitz*, is processed and secreted in order to refine cell fate (Golembo, Raz & Shilo 1996). This activation of *spitz* was thought to involve *rhomboid-1* and another gene, namely *star*; however, the mechanism remained elusive (Golembo et al. 1996). The regulation of the EGF receptor pathway was further corroborated when Wasserman *et al.* identified *Drosophila rhomboid-3* that cooperates with *rhomboid-1* specifically in the eye (Wasserman, Urban & Freeman 2000). Moreover, Wasserman and colleagues identified five novel rhomboid-like genes in *Drosophila*, suggesting an evolutionary conservation (Wasserman et al. 2000). In 2001, the molecular mechanism of the activation of *spitz* was elucidated from cell culture experiments (Lee, Urban, Garvey & Freeman 2001). It was demonstrated that *star* was necessary for the export of transmembrane *spitz* from the ER to the Golgi, where *rhomboid-1* induced the cleavage of *spitz* to release its soluble active form. By using *Drosophila* embryos, it was suggested that *spitz* cleavage occurred in response to *rhomboid-1* expression (Lee et al. 2001), contrary to previous evidence suggesting a lack of protease motifs upon sequence analysis (Bier et al. 1990).

It was not until 2001 when *rhomboid-1* was described as a novel member of the serine protease family (Urban, Lee & Freeman 2001). At the time, mutational analysis revealed that the catalytic residues of rhomboid proteases consisted of serine, histidine, and arguably asparagine, which are distinct from the classical catalytic triad (Urban et al. 2001). Later research further indicated the importance of serine and histidine, and that asparagine in fact does not participate in the catalysis (Lemberg et al. 2005; Maegawa, Ito & Akiyama 2005; Y. Wang et al. 2006). When the human homolog RHBDL2 was found to cleave *spitz* as well, it suggested that rhomboid proteolytic activity was conserved from *Drosophila* to humans (Pascall & Brown 1998; Urban et al. 2001). When genome sequencing became widely accessible, it provided confirmation that rhomboid orthologs are found in all branches of life, revealing that rhomboid proteases are indeed evolutionarily conserved (M. Lemberg & M. Freeman 2007; Lemberg et al. 2005).

To date, 14 members of the mammalian rhomboid family have been identified (Bergbold & Lemberg 2013). Of these, five are catalytically active, namely RHBDL1-4 and presenilin-
associated rhomboid-like protease (PARL), while the other nine are pseudo-proteases, due to the lack of key catalytic residues (Bergbold & Lemberg, 2013). Some of the pseudo-proteases possess the ability to bind proteins in a chaperone-like manner and regulate the degradation or trafficking of these proteins (Adrain, Zettl, Christova, Taylor & Freeman 2012; Grieve et al. 2017).

Rhomboid proteases are unique amongst the intramembrane proteases, as their substrates do not require initial shedding by other proteases. Instead, rhomboid proteases cleave substrates directly in their transmembrane sequence or conduct ectodomain cleavages independently from other proteases (Maegawa, Koide, Ito & Akiyama 2007; Strisovsky, Sharpe & Freeman 2009). For example, RHBDL2 cleaves and activates thrombomodulin at the plasma membrane by cleaving within the TMS (Lohi, Urban & Freeman 2004), and RHBDL4 cleaves the amyloid precursor protein multiple times in the ectodomain region (Paschkowsky, Hamze, Oestereich & Munter 2016). In this regard, eukaryotic rhomboid proteases have been associated with a variety of human diseases, including neurodegenerative and metabolic diseases, as well as cancer (Adrain et al. 2012; Civitarese et al. 2010; Johnson et al. 2017; Paschkowsky et al. 2016; Shen, Buguliskis, Lee & Sibley 2014; Shi et al. 2011; Song et al. 2015; Stewart & Tonkin 2015; Walder et al. 2005). In addition, parasitic rhomboid proteases are involved in invasion of host cells, adhesion protein cleavage, and parasitic cell growth (Shen et al. 2014; Stewart & Tonkin 2015). Those health-related pathways demand for a deeper understanding of rhomboid proteases. Future research will further elucidate the structure and function of rhomboid proteases and continue to identify their physiologically relevant substrates. It will become interesting to develop potent rhomboid inhibitors and to determine if rhomboid proteases may be useful as novel therapeutic targets (Ticha et al. 2017; Tong et al. 2017).

**Conclusion**

Theodor Schwann’s discovery of the first protease almost 200 years ago is not what he is best remembered for. Indeed, his work on cell theory and his description of Schwann cells, which are specialized cells within the peripheral nervous system, are often regarded as his most noteworthy contribution. Nonetheless, his discovery of pepsin started a field of research that is widespread, exciting, and expanding. This brief collection of important discoveries in protease
history hopefully illustrates how proteases have adapted to different environments without changing basic underlying mechanisms.

**Funding:** JS holds a Canada Research Chair in Molecular Chaperones, Tier II and receives funding through the Canadian Institute of Health Research (CIHR), MOP-130332. LMM holds a Fonds de recherche Santé Québec (FRQS) chercheur-boursier junior 1 salary award. This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant RGPIN-2015-04645, the Canada Foundation of Innovation Leaders Opportunity Fund Grant 32565, the Scottish Rite Charitable Foundation 16112, and the Weston Brain Institute RR172187.

**References**

Adrain, C., Zettl, M., Christova, Y., Taylor, N., & Freeman, M. 2012. Tumor necrosis factor signaling requires iRhom2 to promote trafficking and activation of TACE. *Science, 335*(6065), 225-228. doi:10.1126/science.1214400


Chakraborti, S., Chakraborti, T., & Dhalla, N. S. 2017. Proteases in Human Diseases: Springer.


Naturhistorisch-Medizinischer Verein, H. 1877. *Verhandlungen*.


Song, W., Liu, W., Zhao, H., Li, S., Guan, X., Ying, J., . . . Wang, L. 2015. Rhomboid domain containing 1 promotes colorectal cancer growth through activation of the EGFR signalling pathway. *Nat Commun, 6*, 8022. doi:10.1038/ncomms9022


