Proteases in bacteriophage defense systems and their potential in bioengineering

Kalogeropoulos, Konstantinos; van Beljouw, Sam P.B.; Feldmann, Dani; van den Berg, Daan F.; Brouns, Stan J.J.

DOI
10.1016/j.tibs.2025.10.006

Publication date
2026

Document Version
Final published version

Published in
Trends in Biochemical Sciences

CellPress
OPEN ACCESS

Trends in
Biochemical Sciences

Review

Proteases in bacteriophage defense systems and their potential in bioengineering

Konstantinos Kalogeropoulos [1,2,3,4,*], Sam P.B. van Beljouw [1,2,4], Dani Feldmann [1,2], Daan F. van den Berg [1,2], and Stan J.J. Brouns [1,2,*]

Novel phage defense systems featuring diverse enzymatic activities are continually being discovered. Among these, defense systems employing proteolytic enzymes have been identified, revealing a previously unrecognized enzymatic activity in phage defense. These protease-associated defense systems represent an untapped reservoir for new biotechnological tools and may serve as a springboard for the development of proteome editors. This review outlines recent advancements in the discovery and characterization of protease-containing defense systems, proposes methods for further exploration and investigation of protease activity, and considers the prospect of protease defense systems for modulating protein processing and cell fate.

Highlights

A broad spectrum of bacterial phage defense systems featuring proteases likely still awaits discovery, showcasing diverse mechanisms for sensing, effector activation, and function.

Unlike traditional DNA or RNA editing, activatable proteases present new avenues for engineering biological systems through specific and direct modification of proteins.

Engineering proteases and their substrates opens possibilities for manipulating protein processing and function.

Adaptation of protease systems to other targets and pathways offers significant potential for protein signaling pathways and cell process regulation.

A guild of protease-associated defense systems enters the frame

Bacteria and their viruses, phages, are engaged in a continuous evolutionary conflict, which has driven the development of sophisticated mechanisms of resistance and counteraction [1,2]. Over the past decade, advancements in genomic mapping of defense systems, protein comparisons, and clustering techniques have facilitated the discovery and characterization of numerous phage defense mechanisms [3,4]. These systems can sense a variety of triggers – for example, DNA damage and **phage-associated molecular patterns (PhAMPs)** (see Glossary) – and employ diverse enzymatic effectors (e.g., metabolite depleting proteins and restriction enzymes) to mediate phase defense [5,6].

Among the most recent discoveries are multiple novel defense systems that deploy **proteases** [7–9], which are enzymes that cleave proteins via peptide bond hydrolysis, regulating proteostasis and protein activity [10]. Notable examples of such bacterial **protease-associated defense systems (PADS)** are **clustered regularly interspaced short palindromic repeats (CRISPR)–CRISPR-associated protein (Cas)**, **antiviral signal transduction ATPases with numerous domains (STANDs) (Avs)**, **bacterial cyclic oligonucleotide-based anti-phage signaling system (CBASS)**, and **toxin–antitoxin (TA) systems**, many of which have been unearthed through bioinformatic approaches and subsequent experimental characterization [11,12]. However, most proteases in PADS are poorly understood and understudied.

Besides revealing novel insights into the biology of bacterial immunity, studying PADS offers unique avenues for bioengineering. The inducible nature of these protease systems, along with their adaptability as scaffolds, opens up possibilities for application in diverse biological and therapeutic contexts [13,14]. Unlike enzymes acting on DNA or RNA, proteases can act directly on proteins. In contexts where a rapid response to infection is critical, direct protein-level modulation enables faster cellular adaptation and may offer an evolutionary advantage. Therefore, direct protein processing by proteases can be advantageous in scenarios where immediate modification of cell makeup is essential, or when genome editing is undesirable [15].

[1]Department of Bionanoscience, Delft University of Technology, 2629 HZ Delft, The Netherlands
[2]Kavli Institute of Nanoscience, 2629 HZ Delft, The Netherlands
[3]Department of Biotechnology and Biomedicine, Technical University of Denmark, Søltofts Plads B221, Kgs. Lyngby, 2800, Denmark
[4]These authors contributed equally.

*Correspondence:
konka@dtu.dk (K. Kalogeropoulos) and
stanbrouns@gmail.com (S.J.J. Brouns).

After outlining the conservation of proteases, we go on to explore recent discoveries of PADS in detail. We then highlight how engineering their proteases and sensing mechanisms could enable the next generation of post-translational editors. Such systems may offer a safer alternative to nucleic acid-based genome editing, owing to the transient nature of their activity.

Protease activity is conserved across life

Proteases are indispensable enzymes that play pivotal roles in numerous cellular processes across all domains of life. In eukaryotes, proteases have evolved into highly complex systems essential for maintaining homeostasis and orchestrating organismal processes. Proteases also play critical roles in cell differentiation by regulating key signaling pathways and cell migration, and they are indispensable for higher-order processes (Box 1).

In eukaryotic immunity, caspases play a central role in eukaryotic cell death by initiating apoptosis by cleaving proteins in response to cellular damage [16]. Inflammatory caspases within **inflammasomes** also cleave proinflammatory cytokines – pro-interleukin (IL)-1β and pro-IL-18 – into their active forms and trigger **pyroptosis**, thereby amplifying the inflammatory response to infection [17]. The complement cascade likewise relies on sequential serine proteases that activate one another in a **zymogen** cascade, resulting in opsonization of pathogens and assembly of membrane attack complexes for direct lysis of microbes [18]. Intracellularly, lysosomal proteases such as cathepsins are used in the process of **xenophagy** [19]. Furthermore, the **ubiquitin–proteasome system** and similar ubiquitin-like modifier proteins target viral proteins for proteasomal degradation, limit viral replication, and regulate innate immune signaling pathways [20–22].

Bacterial proteases share several important functions with their eukaryotic counterparts, including the maintenance of cellular homeostasis through the degradation of misfolded, unnecessary, or aggregated proteins, particularly under stress [23,24]. Additionally, bacterial proteases are

Box 1. Protease classification, function, and regulation

Proteases are classified into clans based on the key catalytic residues, with further classification into subclans and families based on sequence homology [102]. Proteases are termed endopeptidases if they cleave internal peptide bonds, or exopeptidases if they target terminal bonds, with further classification into aminopeptidases or carboxypeptidases according to whether cleavage occurs at the N or the C terminus [103].

The mechanism of action typically involves a nucleophilic attack towards the scissile bond, initiated by substrate binding. Through a series of intermediate steps, two polypeptide chains, each with a neo N and C terminus, are released [104]. The substrate-binding pocket plays a critical role in determining substrate range, influenced by the residues around the cleavage site (named P1, P1′ for substrate and S1, S1′ for substrate pocket residues). Substrate specificity varies widely – from proteases with highly specific recognition sequences (e.g., SUMO proteases) to those with broad activity (e.g., **matrix metalloproteinases**), depending on specificity, localization, and cofactors – leading to diverse functions and substrate repertoires [105,106]. Auxiliary domains may assist in substrate binding, providing additional specificity and functionality in their activity [107].

In higher organisms, many proteases are synthesized as inactive zymogens with propeptide sequences that inhibit activity until removed by proteolysis, a process that is catalyzed by either the protease itself or another protease [108]. Additionally, protease activity is regulated by inhibitors that bind to the active site, preventing substrate access [109]. Proteases often require cofactors such as metal ions or proteins for activity, and activity is often further controlled by exosite binding and allosteric changes to their structural conformation. Functional activity may require the formation of higher-order structures such as dimers (caspase activation) or hexamers (proteasome), and proteases can be components of large complexes such as the proteasome, or exist as simpler entities with only the protease domain [16].

Unlike other enzymes that modify proteins post-translationally, proteases induce irreversible changes in the proteome landscape, with precise regulatory mechanisms in place to avoid detrimental activity. Proteases typically operate in biological pathways where irreversible decisions are made, often being integrated into positive feedback loops and signaling cascades controlled by multiple checkpoints [110]. Proteolysis is vital for maintaining proteome homeostasis, contributing to protein turnover and localization, as well as regulating critical processes such as tissue remodeling and cellular fate determination [111]. Dysregulated protease activity is implicated in numerous diseases, highlighting the need for stringent control of protease action [112].

Glossary

Alphafold 3: a deep-learning-based protein structure prediction tool that extends the Alphafold framework to multimodal modeling, incorporating ligands, nucleic acids, and protein complexes.
Antiviral signal transduction ATPases with numerous domains (STANDs) (Avs): proteins that detect phage components and activate downstream defense mechanisms.
CHOPPER: click chemistry enrichment with positive strategies: a chemical proteomics method for selective labeling, enrichment, and analysis of proteins or peptides with mass spectrometry.
Clustered regularly interspaced short palindromic repeats (CRISPR): together with Cas proteins CRISPR forms adaptive immune systems in bacteria and archaea.
Craspase: a CRISPR-guided caspase-like system, and the first described member of the protease-associated defense systems (PADS).
CRISPR-associated proteins (Cas): proteins that mediate target recognition and cleavage in CRISPR adaptive immune systems of bacteria and archaea.
Cyclic oligonucleotide-based anti-phage signaling system (CBASS): a system that uses cyclic nucleotides as second messengers to activate diverse defense effectors.
Cyclic triadenylate (cA_3): a small cyclic oligoadenylate second messenger produced by Type III CRISPR–Cas systems to activate downstream defense effectors.
Degron: a short amino acid sequence at a protein terminus that marks the protein for degradation through ubiquitin-dependent or ubiquitin-independent pathways.
Deubiquitinases (DUBs): a family of proteases that remove ubiquitin or ubiquitin-like molecules from target proteins by isopeptidase activity.
Inflammasome: a multiprotein complex in a host cell that detects pathogens or cellular stress and activates caspase-1, leading to cytokine maturation and pyroptosis.
JAB protease: a metalloprotease family member (JAB1/MPN/Mov34) that processes ubiquitin-like proteins by priming them for conjugation or removing them from targets, thereby enabling recycling.

essential in regulating cell division, fate determination, and cellular death [25–27]; their role also extends into transcriptional regulation, sensing, and signaling. These include the control of TA systems [28], degrading transcription repressors (e.g., in response to DNA damage, stress response and protein quality control, and gene transfer) [29], and signal processing (e.g., in quorum sensing) [30]. Proteases are essential not only for bacteria but also for viruses. Their use spans protein maturation from polyprotein proforms [31], virion assembly initiation with cleavage of structural proteins [32], and counter-defense via degradation of critical components of host immune systems [33]. This breadth of conserved functions across life underscores protease activity as a highly versatile and conserved enzymatic mechanism, and suggests an ancient evolutionary origin of proteolysis [10].

The emergence of the PADS repertoire of bacteria

While a plethora of immune proteins has been experimentally described (e.g., CRISPR–Cas, restriction enzymes, RecBCD) [34–36], only a few PADS have been studied (Figure 1), though

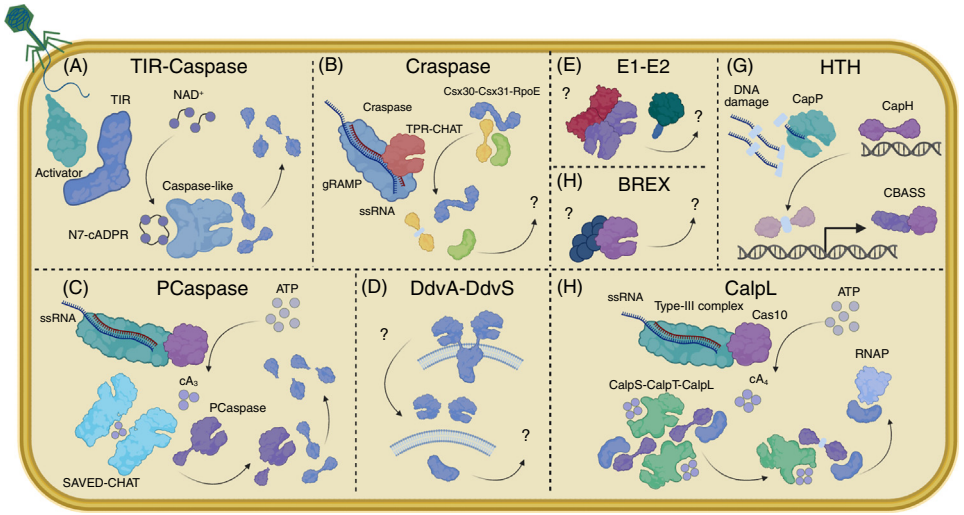

Figure 1. The increasing repertoire of characterized protease-associated defense systems (PADS). PADS employ diverse mechanisms to couple phage signals to protease activation. (A) Toll/interleukin 1 (IL-1) receptor/resistance protein (TIR)–caspase is activated by a phage component. The TIR domain uses NAD^+ to generate N7-cyclic adenosine diphosphate ribose (cADPR), which in turn activates a caspase-like protease that cleaves multiple host substrates to promote defense. (B) In the Craspase system, hybridization of guide RNA (gRNA) with its complementary RNA induces conformational changes in the complex that activate the associated caspase HetF associated with tetratricopeptide repeat (TPR) (CHAT) protease. The protease cleaves Csx30, which is in complex with Csx31 and an RpoE-like transcription factor, releasing transcriptional inhibition and inducing an abortive infection response. (C) Prokaryotic caspase (PCaspase) is also triggered by complementary RNA; the system produces cyclic tri-adenylate (cA_3). This signaling molecule activates a SMODS-associated and fused to various effector domains (SAVED)–CHAT domain, which then activates PCaspase to cleave cellular substrates broadly, driving antiviral defense. (D) The DdvA–DdvS system is proposed to sense a phage-derived periplasmic signal: DdvA self-cleaves its linker, initiating a regulated intramembrane proteolysis cascade that releases the transcription factor DdvS to induce antiviral gene expression. (E) E1–E2 conjugation systems employ cyclase or ubiquitin-like proteins to tag host or phage proteins, potentially blocking phage assembly or regulating cellular pathways. JAB proteases within these systems balance conjugation by priming and removing modifiers. (F) BREX (bacteriophage exclusion) systems are established as phage defense pathways, but the precise role of the protease component remains unknown. (G) In the cyclic oligonucleotide-based anti-phage signaling system (CBASS)-associated Cap system, DNA damage sensing relieves inhibition of CapP, a metalloprotease that degrades the transcriptional repressor CapH. This activates expression of the CBASS operon and triggers an abortive infection response. (H) The CalpL system detects complementary RNA to stimulate synthesis of cyclic tetra-adenylate (cA_4). This activates CalpL protease, which cleaves CalpT to release CalpS, a transcriptional regulator of antiviral genes. Abbreviation: RNAP, RNA polymerase. Figure created with BioRender.

Mass spectrometry (MS): an analytical method for precise mass-to-charge measurement of ions. In proteomics, especially bottom-up proteomics, MS is routinely used to study protease cleavage events (degradomics).
Matrix-assisted laser desorption/ionization (MALDI): a soft ionization method in MS often used for peptide and protein analysis.
Matrix metalloproteinases: zinc-dependent endopeptidases that degrade extracellular matrix proteins and regulate tissue remodeling, signaling, and inflammation.
Metallo-β-lactamases (MBLs): a family of metal-dependent hydrolases best known for degrading β-lactam antibiotics, but also widespread in diverse enzymatic contexts.
Phage-assisted continuous evolution (PACE): a technique that links phage replication to the activity of a target protein, enabling rapid laboratory evolution.
Phage-associated molecular patterns (PhAMPs): conserved phage-derived molecules recognized by bacterial immune systems as infection signatures.
Protease: an enzyme that catalyzes the hydrolysis of peptide bonds in polypeptide chains, thereby regulating protein turnover and function.
Protease-associated defense systems (PADS): systems that use proteolysis and signaling to protect bacteria from phage infection.
ProteinMPNN: a deep-learning-based framework for protein sequence design that generates amino acid sequences compatible with a given protein backbone structure.
Proteomic identification of protease cleavage sites (PICS): a mass spectrometry-based method for unbiased, global mapping of protease substrate specificity.
Pyroptosis: a form of inflammatory lytic cell death triggered by inflammasome signaling and executed through gasdermin pore formation.
RFdiffusion: a generative modeling method for protein design that leverages diffusion models to sample novel structures and sequences with high designability.
SMODS-associated and fused to various effector domains (SAVED): a class of bacterial defense systems centered on STAND ATPases that detect phage infection and trigger downstream effectors.

it is becoming clear that bacterial proteases also play essential roles in bacterial immune responses against phages. Here they often function as signal transductors or effectors, degrading viral components, regulating stress responses, or triggering cell dormancy or death to prevent phage propagation. We review PADS that have been experimentally described thus far, including members from the cysteine, metalloprotease, and serine protease families.

Cysteine proteases

Cysteine proteases comprise a ubiquitous class of proteolytic enzymes that have a cysteine in their active pocket, with the caspase family as the predominant signature in emerging PADS.

Importantly in the context of this review, a recent study revealed that a cysteine protease in bacteria can activate gasdermin-like cell death effectors via proteolytic cleavage upon sensing a viral invader, paralleling inflammasome-driven pyroptosis in eukaryotes [37]. In particular, caspase recruitment domain (CARD) motifs, also central to eukaryotic immune complexes, were identified at the N termini of these bacterial proteases (although these are thought to be trypsin-like proteases) that activate gasdermins [37]. The structural and functional parallels between bacterial and eukaryotic CARD protease and gasdermin modules support an ancient evolutionary origin for this immune strategy.

Another experimentally studied immune system involving a cysteine protease is **type IV Thoeris** [38]. In this PADS, infection is sensed by a Toll/IL-1 receptor/Resistance protein homology domain (TIR) protein through recognizing unknown phage-derived constituents, generating a cyclic adenosine diphosphate ribose (ADPR) (N7-cADPR) molecule from NAD^+ (Figure 1A). This N7-cADPR in turn activates a caspase-like protease, resulting in massive reduction in cellular – and potentially phage – proteins that appears to prevent the invading phage from successfully reproducing [38].

Several PADS employ a caspase HetF associated with **tetratricopeptide repeat (TPR)** (CHAT) protease domain, which has been classified as an evolutionary relative of caspases. This domain is characterized by its association with TPRs and is implicated in protein–protein interactions and pathway signaling [39]. As described here, CHAT domains are employed by several bacterial PADS. First, proteases activated by CRISPR–Cas systems have been classified as the CRISPR-controlled protease family [40]; this currently contains three members, of which two are cysteine proteases: **Craspase** and prokaryotic caspase (PCaspase). Craspase, also known as CRISPR–Cas type III-E, is a complex of an RNA-recognizing CRISPR–Cas effector (gRAMP/Cas7–11) and a caspase-like protease (TPR-CHAT/Csx29) [41,42] (Figure 1B). Upon binding of RNA complementary to the guide RNA (gRNA) of Craspase, a conformational change results in protease activation [43,44]. Craspase then cleaves the host-encoded protein Csx30, releasing the inhibition on a transcription factor (RpoE) and leading to an abortive infection phenotype [7,45]. Cleavage of the bound RNA shuts off the proteolytic activity, making Craspase an RNA-controlled protease [43].

In the second CRISPR–Cas activated protease system, CRISPR–Cas type III-B, two proteases are involved: second messenger oligonucleotide or dinucleotide synthetase **(SMODS)-associated and fused to various effector domains (SAVED)**–CHAT – and PCaspase [9]. Upon recognition of target RNA by the type III-B effector protein, **cyclic triadenylate (cA_3)**, signaling molecules are generated which bind to the SAVED domain of SAVED–CHAT. This stimulates its proteolytic action, specifically cleaving and thereby activating PCaspase (Figure 1C). Activated PCaspase cleaves various protein substrates, mounting a strong defense phenotype [9]. It could be envisioned that such systems might be employed as programmable proteases, cleaving specific proteins in response to designed RNA triggers.

Terminal amine isotopic labeling of substrates (TAILS): a proteomics method for enrichment and MS analysis of protein N and C termini.
Tetratricopeptide repeat domains (TPRs): structural motifs that mediate protein–protein interactions and assembly of multiprotein complexes.
Toxin–antitoxin (TA) systems: bacterial modules composed of a stable toxin and a labile antitoxin that regulate growth, stress responses, and defense against phages.
Type IV Thoeris: a bacterial immune system that defends against phages using a TIR-domain protein to produce the cyclic signaling molecule N7-cADPR.
Ubiquitin–proteasome system: a eukaryotic protein quality control pathway in which ubiquitin tags proteins for degradation by the 26S proteasome.
Xenophagy: a selective form of autophagy in which host cells target and degrade intracellular pathogens.
Zymogen: an inactive precursor of an enzyme that requires proteolytic cleavage or conformational change to become active.

Beyond the CRISPR-controlled proteases, the CHAT-encoding DdvA–DdvS system is indirectly involved in phage defense by regulating a phage defense island [46] (Figure 1D). The DdvA protein features a dimeric structure, with a periplasmic TPR–CHAT domain that resembles caspase proteases. DdvA is proposed to sense a phage signal in the periplasm, triggering a conformational change that activates its protease activity. This leads to cleavage of its own linker region, initiating a regulated intramembrane proteolysis (RIP) cascade that releases DdvS, enabling transcription of antiviral defense genes [46]. Analogous systems could be engineered to link extracellular cues to controlled gene expression in synthetic contexts.

Lastly, the single-gene Borvo system encodes a CHAT protease that provides broad protection against Siphoviruses in *Escherichia coli* [47]. Its exact mode of action is still unclear, but one plausible model is that the Borvo CHAT protease cleaves bacterial gasdermin homologs encoded in different loci. Another possibility is that they become activated upon sensing damage and trigger dormancy through cleavage of essential proteins.

Metalloproteases

Metalloproteinases are proteases that harbor a metal ion (Zn^{2+}, Mg^{2+}, Cu^{2+}) in their active site, coordinating hydrolysis of peptide bonds [48]. Several such proteases of diverse families have been discovered to be part of bacterial PADS.

Several CBASS systems with similarity to eukaryotic immunity, in particular the cyclic GMP–AMP synthase (cGAS) and the ubiquitin system, were discovered [12]. These systems utilize E1–E2 fusion enzymes (cap2) to catalyze conjugation of a protein encoded in the system to other protein substrates likely to enhance its activity [49] (Figure 1E). **Deubiquitinase (DUB)**/JAMM-like-metalloproteases seem to play an important role in these systems, potentially through removing the conjugated proteins to balance the system, akin to deubiquitinases [50,51]. For example, the Bil system, an interferon-stimulating gene (ISG)-15-like system containing a **JAB protease**, deploys an E1–E2 enzymatic cascade to conjugate a ubiquitin-like protein to the central tail fiber of infecting phages [47,52]. This conjugation interferes with tail assembly or obstructs the tail tip, resulting in the release of tailless or non-infective phage particles. The JAB protease was shown to both prime the ubiquitin-like protein for conjugation and remove it from targets, suggesting a dual role in conjugation and recycling [53]. Another anti-phage defense system in which ubiquitin-like conjugation is central for antiviral activity is 6A-**metallo-β-lactamase (MBL)**, which features a unique fusion of E1, E2, and JAB domains. Mutations in the active sites of E1, E2, and JAB domains completely abolished the defensive phenotype, demonstrating the critical role of these components [54]. In the future, such systems could be reprogrammed for ubiquitin conjugation of selected targets, enabling conditional protein degradation in response to defined triggers.

Another PADS featuring metalloproteases is the PD-λ-2 system which prevents the infection of lambda phages and consists of three genes, including a HigA antitoxin (PD-λ-2B), a P4 phage antitoxin (PD-λ-2C), and a third gene that encodes both a HigB-like toxin domain and a Zn-peptidase domain (PD-λ-2A) that is also present in the PADS RosmerTA and IrrE/ImmA [55]. By contrast, overexpression of the proposed antitoxin PD-λ-2B was toxic and was rescued by the presence of Zn-peptidase encoding PD-λ-2A. The role of PD-λ-2C in this system is unclear, apart from being required for phage defense. These findings suggest that the peptidase of PD-λ-2B is rendered inactive upon phage infection, causing PD-λ-2B to inhibit phage replication in a self-inflicted toxic mechanism that as yet is not understood [55].

MucP, a membrane zinc metalloprotease, was discovered in *Pseudomonas aeruginosa* and shown to prevent the propagation of single-stranded RNA (ssRNA) phage PP7 by cleaving its

lysis protein [56]. Members of the Lamassu family encoding predicted peptidase effectors have also been implicated in antiviral immunity [47]. Although mutation of the putative protease active site did not impair phage protection, the peptidase may contribute to a secondary layer of defense, as observed in other multi-effector systems.

Lastly, metalloproteases have recently also been implicated in the indirect regulation of phage defense systems. For instance, capP, a metalloprotease related to CBASS systems, is triggered by ssDNA to degrade the HTH family transcriptional repressor of the cognate operon, capH [57] (Figure 1G). This results in the upregulation of the cognate CBASS system (encoded in the same operon) and abortive infection. It is conceivable that such a system could be engineered to sense DNA damage in eukaryotic cells, triggering a downstream response to that stimulus.

Serine proteases

Serine proteases were also shown to be involved in defense against bacteriophages, mainly with Lon-like and trypsin-like proteases. Lon proteases are ATP-dependent serine proteases conserved across bacteria, archaea, and eukaryotes, best known for their role in protein quality control and stress response. They degrade misfolded, damaged, or short-lived regulatory proteins, and also contribute to cellular processes [58].

Two instances of Lon-encoding PADS have been characterized, including a subset of CRISPR–Cas type III and BREX (bacteriophage exclusion). CalpL, which exists in a complex with two co-occurring proteins, CalpT and CalpS, is a Lon-like CRISPR-controlled protease from a CRISPR–Cas type III subtype [8] (Figure 1H). The Cas10 protein of the CRISPR–Cas system senses infection and produces cyclic tetra-adenylate (cA_4) to activate CalpL through oligomerization, which subsequently cuts CalpT. This is thought to expose a **degron**, targeting CalpT for degradation via the ClpX/P system, ultimately freeing the σ-factor CalpS to initiate a transcriptional response [59,60].

The BREX system is widespread in bacterial clades and also contains a Lon-like protease domain, conferring broad resistance by utilizing multiple enzymatic activities [61] (Figure 1F). Although the mechanism of action remains to be elucidated, deletion of the protease-encoding brxL gene abolishes protection, while individual expression of the same gene is highly toxic for the host [62].

In two recent studies, another PADS called HamABM was shown to exhibit a protease-controlled nuclease activation mechanism, in which the trypsin-like serine protease domain of HamA activates the proenzyme form of the nuclease HamM leading to DNA degradation [63,64]. It is proposed that, upon DNA sensing, ATP-dependent release of HamA initiates the system's cascade, allowing HamA to cleave HamM, thereby serving as the trigger and regulatory step for protease function. Interestingly, this mechanism contrasts with other characterized PADS such as PCaspase and Craspase, where protease activity occurs downstream of nuclease activation, highlighting the mechanistic diversity within PADS. Protease-gated switches such as the HamABM could inspire synthetic systems in which proteolysis serves as the trigger for nuclease or enzyme activation.

Another PADS consisting of a single component, a trypsin-like secreted protease, was identified in *Salinispora mooreana* that reduces infectivity of *Streptomyces* phages by seemingly promoting premature DNA ejection from viral particles [65]. The protease acts extracellularly, with escaper phages acquiring mutations in a tail-associated structural protein, thus establishing a novel example of a protease-mediated, secreted antiphage defense.

Lastly, bacterial TA systems that do not contain protease components are often regulated by the proteolytic regulation of Lon-like and ClpP-like serine proteases, which are able to rapidly degrade the antitoxin when necessary [66]. For example, the antitoxin of phage defense TA system MazEF is degraded by the cellular serine proteases upon detecting a phage infection of a neighboring cell [67,68]. Similarly, the Dodola system might also tap into these serine proteases of the host to convey phage protection, since it encodes the caseinolytic protease B (ClpB), a protein that is known to be part of the ClpP-regulated stress response [29].

Expanding the PADS repertoire

Bioinformatic efforts have resulted in the prediction of several potential bacterial PADS (Figure 2). In one of these exploratory studies, Shmakov and colleagues developed a pipeline for the identification of novel CRISPR loci [69]. Among these, they described several clusters containing putative proteolytic enzymes that would later be experimentally validated, including the CalpL and Craspase systems (clusters icity0089 and 6485, respectively). In addition, they described a CRISPR cluster containing aspartyl protease genes in the Sulfolobales order of thermoacidophilic archaea [70]. This cluster, called icity0034, includes operons with repeat associated mysterious proteins (RAMP). Among hypothetical proteins is an ATPase and aspartyl proteases with a fold structurally similar to the protease domain of the human protease DDI2 [71] (Figure 2, DDI2), which recognizes ubiquitinated proteins. Given its organization, this

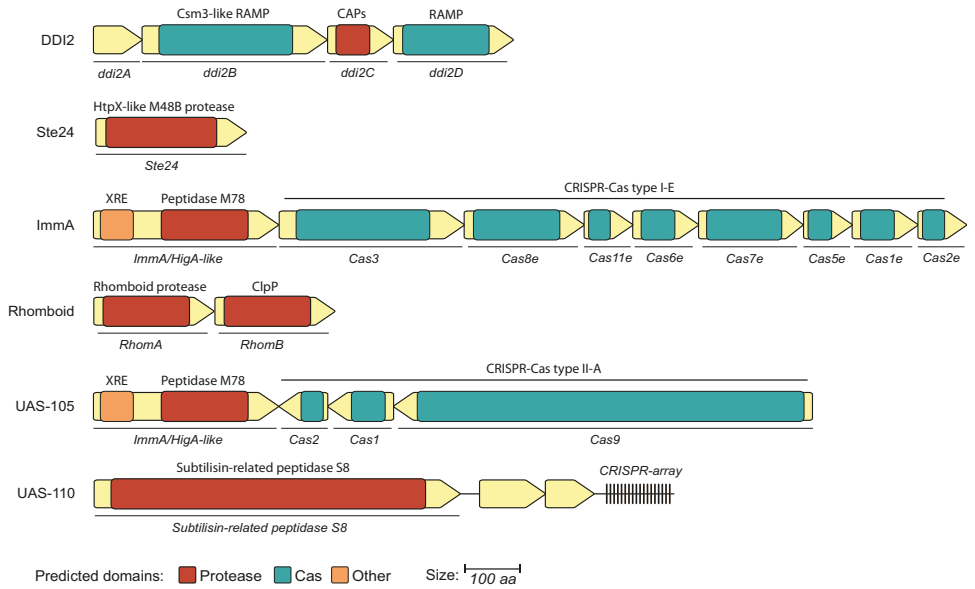

Figure 2. The expanding landscape of protease-associated defense systems (PADS). Several candidate systems have been proposed upon association with clustered regularly interspaced short palindromic repeats (CRISPR) arrays and associated effectors, or neighborhood mining. The highlighted systems in this figure show intriguing architectures with potentially novel functions, but neither they nor their homologs have yet been characterized experimentally. These potential PADS feature different classes of proteases and various domains and effectors. For instance, DDI2 could regulate stress responses and protein quality control, and Ste24 may act similarly to Borvo. The ImmA metalloprotease system could function in a similar fashion to the cyclic oligonucleotide-based anti-phage signaling system (CBASS)-associated Cap system, regulating expression of the nearby Cascade operon, although that would imply autocatalysis and regulation. ClpP proteases could be effectors in stress response pathways, similar to bacterial and eukaryotic homologs. UAS-105 could function in a manner similar to ImmA, while UAS-110 could operate in a similar fashion to the bacteriophage exclusion (BREX) system. Antiviral signal transduction ATPases with numerous domains (STANDs) (Avs) systems might resemble similar systems but employ metallo-β-lactamase (MBL) conjugation or different enzymatically active (EAD) domains to destroy or activate downstream proteins. Abbreviation: cas, CRISPR-associated protein.

system might hint at potential proteasome-like activity in these extremophiles, induced by foreign nucleic acid recognition.

In the same study [69], a cluster of transmembrane metalloproteases (icity0083) with homology to the M48 family of the MEROPS database (Figure 2, Ste24) – which comprises stress response proteases in prokaryotes and eukaryotes – was discovered [72]. Two more clusters of putative metalloproteases (clusters 4229 and 4790 according to the study numbering) resemble TA systems similar to the ImmA protease domain belonging to the M78 family (Figure 2, ImmA), which is associated with cleavage of the immunity repressor ImmR in transposon systems [73]. Finally, they also predict a serine protease cluster (5921) of the rhomboid family (Figure 2, Rhomboid), which are transmembrane proteases with various cellular functions catalyzed by intramembrane processing of substrates in higher-order organisms [74]. Shay *et al.* contemporaneously performed a similar study, searching for accessory proteins in type III CRISPR–Cas cassettes [75]. This study also describes the Lon-like protease (cluster 43 in the study numbering) and aspartyl protease clusters mentioned earlier (cluster 47).

More recently, a large genome mining and clustering study aiming to discover associated CRISPR loci was performed by Altae-Tran *et al*. [76], uncovering a large number of potential novel CRISPR systems and effectors. Among them, approximately 250 candidate systems encoding genes with protease activity are described, including previously discovered and experimentally characterized SAVED-CHAT (UAS-54 and UAS-68), TPR CHAT (UAS-5), and Lon proteases (UAS-55). Several other systems with associated protease effectors are present, either as the bait of the cluster or in locus: UAS-76 and UAS-79 (ImmA), UAS-105 (Figure 2, UAS-105), and UAS-110 (Figure 2, UAS-110).

Additionally, Lowey *et al*. investigated the prevalence of CRISPR-associated Rossmann fold (CARF)-related effectors in CBASS systems, describing CBASS operons containing proteases with predicted caspase-like or metalloprotease activity, fused to SAVED domains [77]. Gao and co-authors have also described several predicted anti-phage Avast systems that encode an effector protease with a trypsin-like fold fused to a terminase sensor TPR domain, next to an MBL or trypsin-co-occurring domain 2 (Trypco2) encoding protein [4,78]. Trypco2-encoding proteins are often found next to trypsin proteases and are believed to cooperate with the trypsin domain through an unknown mechanism [79]. Unlike Trypco2, MBL proteins can operate more independently from trypsin proteases and can be found in a multitude of different gene architectures. MBL proteins usually function as metal-dependent hydrolases and act on small molecules [80]. However, recent studies focusing on other systems provide solid evidence that MBL is a hydrolase-like nuclease [63,64]. How this function is connected to the effector activity of the Avast trypsin domain is unclear, but it might be similar to the activity of the MBL-containing JAB-encoding CBASS type II systems, which are yet to be fully characterized. Anti-phage Avast proteins have also been identified that encode an EAD8/10 domain together with an upstream caspase protein. It is therefore plausible that these domains regulate the activity of the caspase during phage infection. In an earlier study, Gao and colleagues also identified another PADS named ietAS, which seems to be a TA system encoding a serine protease of the S8 family [4].

Lastly, Millman *et al.* used clustering and enrichment near known defense genes to discover a range of novel putative defense systems [47]. This study documents 83 clusters with predicted protease genes (which contain known TA and BREX protease-containing systems, as well M78 and aspartyl proteases identified in previous studies), and characterize the protection conferred from a number of them with a phage panel against *E. coli*.

These studies highlight the diversity of putative bacterial PADS uncovered through genome mining, clustering, and functional studies. The bioinformatically identified systems often exhibit remarkable versatility, likely arising from domain duplication, fusion events, acquisition of new effector activities, and evolutionary pressures favoring adaptability [81,82]. Yet fundamental questions remain as to how these proteases become activated, what molecular cues or signals they sense, and what their substrates are. Biochemical and analytical techniques could be employed to answer these questions (Box 2).

Protease engineering and substrate expansion

Expanding the range of substrates for a protease can be accomplished by identifying existing proteins or domains that can be cleaved naturally by the protease [41,45]. This involves screening various proteins to find those susceptible to protease activity. Alternatively, the protease itself can be engineered to recognize and cleave new substrates [83]. This engineering can be achieved by altering the protease's active site, modifying its substrate-binding regions, or incorporating new auxiliary domains that confer broader or more specific substrate recognition (Figure 3).

Protease engineering

Conventional techniques such as site-directed mutagenesis and guided mutagenesis via *in silico* prediction of substrate pocket interactions can be employed to engineer a protease [83]. This mutagenesis can be focused, random, or fully conducted *in silico*. Additionally, **phage-assisted continuous evolution (PACE)** and other methods of directed evolution *in vivo* are powerful tools for refining protease activity [84]. For example, through PACE, the TEV (Tobacco Etch Virus) protease was engineered to cleave the target sequence HPLVGHIM instead of the canonical ENLYFQIS, a drastically different specificity that processes an exposed loop of the cytokine IL-23 [84]. Directed evolution can be specifically targeted to enhance cleavage efficiency, while exosite engineering may be necessary for better substrate binding [85]. Functional scanning of

Box 2. A toolkit for protease activity investigation

A variety of techniques can be employed to investigate protease activity, ranging from traditional assays and gel electrophoresis to advanced **mass spectrometry (MS)**-based methods for substrate discovery and specificity profiling (Figure I). Traditional assays utilize casein or its derivatives for monitoring unspecific protease activity [113]. Sodium dodecylsulfate–polyacrylamide gel electrophoresis (SDS–PAGE), whether native or denaturing, can be used to visualize truncated proteins or cleavage products [114], while Edman degradation can precisely determine cleavage sites via sequencing of the newly formed neo-N terminus [115]. Assays that utilize fluorescent probes or fluorophore-quencher peptide moieties can also be employed to obtain time-dependent cleavage profiles [116].

Degradomics methods share the common principles of enrichment of protein termini, with techniques such as **terminal amine isotopic labeling of substrates (TAILS)** and high-efficiency undecanal-based N-termini enrichment (HUNTER) employing negative enrichment strategies, while **CHOPPER** and subtiligase employ positive enrichment strategies [117–120] (Figure IA). These techniques can precisely identify substrates and cleavage positions in substrates *in vivo* or proteomes incubated with purified proteases. Proteome digests can also be utilized in combination with MS to probe substrate specificity in linear epitopes, aiming to profile the sequence-based cleavage specificity of promiscuous proteases. **Proteomic identification of protease cleavage sites (PICS)** uses a biotin pull-down of newly generated cleavage products, following the blocking of the natural N terminus with a different reagent [121]. Peptide libraries and peptide microarrays can also prove valuable tools for characterizing protease specificity [122] (Figure IB). The incorporation of unnatural amino acids can be employed to probe specificity, enhance cleavage efficiency for recombinant substrates, and develop potent inhibitors. Substrate phage-display libraries, when coupled with next-generation sequencing, provide an alternative method for analysis of protease specificity [87].

Beyond cleavage determination and substrate discovery, substrate validation is typically performed through SDS–PAGE and Edman degradation, or by monitoring the cleavage of peptide epitopes or full proteins via **matrix-assisted laser desorption/ionization (MALDI)** or liquid chromatography–MS (LC–MS)/MS (Figure IC). The location and activity of proteases can be investigated using active site probes (Figure ID), serving as proxies for determining the active fraction of a protease zymogen [123].

Computational methods such as co-folding or docking can be employed to predict interactions, although these results may be inaccurate since proteases function through transient interactions rather than stable binding [124]. Candidate substrates often co-localize with their proteases and may be part of the same operon [7]. To find interactors or substrates, pulldown assays coupled with MS could also be used.

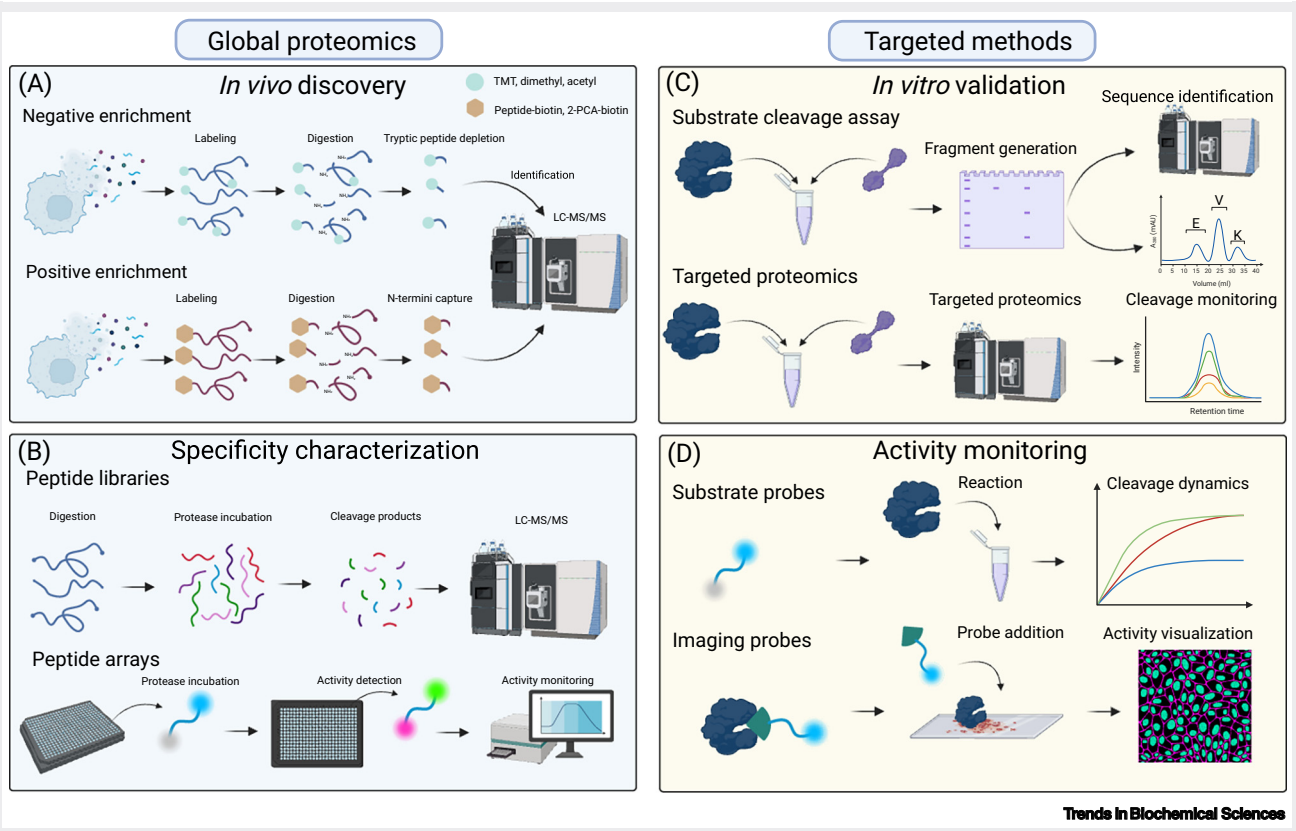

Figure I. Methods for protease characterization and substrate discovery. A range of complementary strategies enables the identification of protease substrates, cleavage motifs, and activity in diverse contexts. Global proteomics (A): system-wide discovery of protease substrates and cleavage sites can be achieved through positive enrichment approaches (labeling and capturing protein N termini) or negative enrichment strategies (blocking N termini and depleting tryptic peptides), both followed by liquid chromatography (LC)–mass spectrometry (MS)/MS analysis. Peptide libraries and arrays (B): panels of synthetic peptides provide a scalable means to screen proteases, define substrate specificity, and identify consensus cleavage motifs. Targeted methods (C): approaches such as targeted proteomics or Edman degradation enable precise mapping of cleavage positions and accurate quantification of cleavage products in known protein substrates. Substrate and imaging probes (D): chemical probes and reporter substrates allow kinetic monitoring of cleavage events and spatial visualization of protease activity in cells or tissues. Abbreviations: 2-PCA, 2-pyridinecarboxaldehyde; TMT, tandem mass tag. Figure created with BioRender.

mutational libraries can prove useful in the identification of variants with improved binding and cleavage of desired substrates. Importantly, applying such strategies to PADS proteases could allow their reprogramming toward novel triggers or targets, paving the way for programmable proteolytic circuits in biotechnology and therapeutic contexts.

Proteases must exhibit transient binding with substrates to ensure that they are released and recycled for continuous activity, which is an important consideration in their engineering [86]. Adding domains that confer substrate specificity might enhance protease functionality, but care must be taken not to disrupt its natural sensing, regulation, and activation mechanisms [14]. Engineering efforts should avoid interfering with these elements to maintain proper protease function while recognizing new substrates. Inducible activation of proteases presents a strategic approach, allowing for controlled activity and the creation of protease display libraries. These libraries can be instrumental in the engineering and discovery of protease variants with desired specificities. For example, substrate phage-display libraries coupled with next-generation sequencing (NGS) was shown to allow for sensitive protease specificity profiling [87].

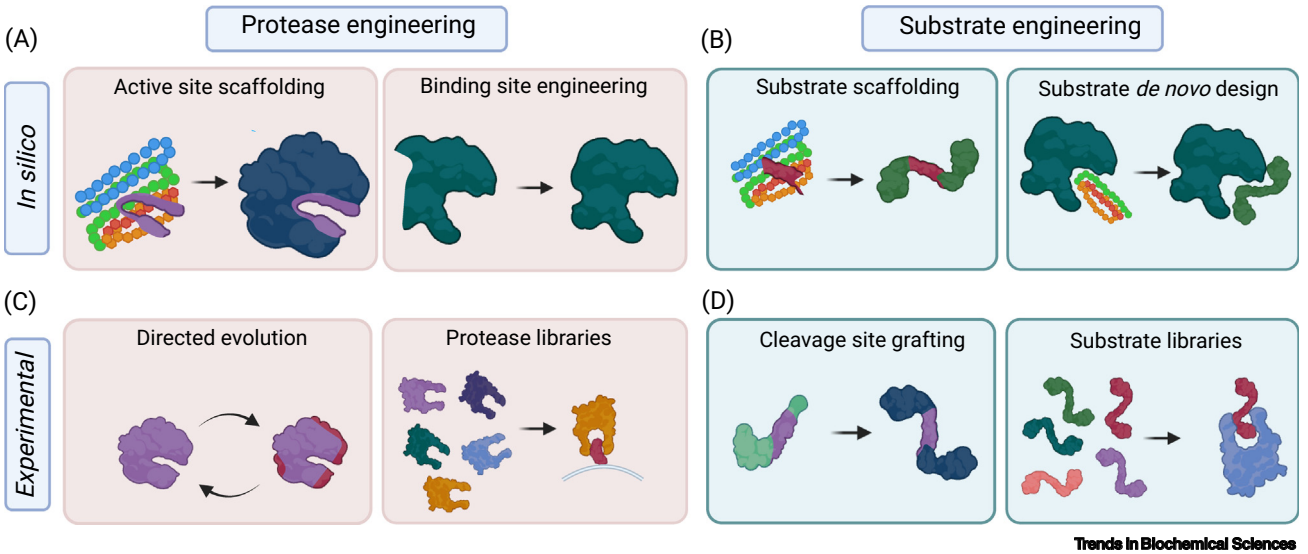

Figure 3. Methods for protease and substrate engineering. Recent advances in protein design enable both computational and experimental strategies for tailoring protease activity and substrate recognition. (A) Protease engineering (*in silico*): catalytic sites can be fixed and designed around active site scaffolding, while allosteric binding sites can be engineered for altered regulation. Artificial intelligence (AI)-guided tools such as RFdiffusion and ProteinMPNN expand the design space for protease scaffolds and binding pockets. (B) Protease engineering (experimental): protease variant libraries can be generated and screened against target substrates, while directed evolution methods such as phage-assisted continuous evolution (PACE) enable selection for novel or enhanced protease activities. (C) Substrate engineering (*in silico*): motif scaffolding around characterized cleavage sites, or fully *de novo* design of new substrates, can yield stable proteins with desired protease recognition properties. These approaches increasingly leverage deep learning frameworks originally developed for binder and inhibitor design. (D) Substrate engineering (experimental): cleavage site grafting into proteins of interest or screening of substrate libraries provides tractable routes for substrate discovery. Cleavage can be monitored using functionalized phage-display systems or mass spectrometry (MS)-based readouts. Figure created with BioRender.

Computationally guided modeling of enzymatic activity is an emerging frontier in protease engineering. Generative artificial intelligence (AI) approaches, such as **RFdiffusion** [88] and RF All-atom models [89], alongside **ProteinMPNN** [90] and ESM2 [91] for sequence generation, hold significant potential for generating sequences from backbone structures, thereby improving enzyme stability. In recent years, considerable progress has also been made on the enzyme design front, enabling the development of active protein enzymes with desired properties, including proteases [92–96]. These methods show great promise in protease engineering, with clear applications in the engineering of PADS. Advances in models that enable the design of more compact enzyme domains while retaining substrate specificity would also be particularly advantageous. Utilizing tools like Alphafold, especially with recent improvements in biomolecule incorporation and complex interaction seen in **Alphafold 3** [97], will be essential for *in silico* design. Integrating these generative and structural modeling approaches could enable the creation of proteases that are not only highly specific and stable but also controllable by defined molecular triggers, offering a blueprint for programmable defense-like systems in synthetic biology.

Substrate engineering

Recombinant substrates designed to mimic the natural substrates of proteases offer an attractive alternative to direct protease engineering. This approach has been successfully demonstrated with Craspase, where part of the target substrate Csx30 was grafted into other proteins, effectively converting them into Craspase targets [45]. While this method may be more tractable, grafting protease cleavage sites onto proteins to confer desired functions can be challenging and requires a comprehensive understanding of protease specificity [45]. Substrate libraries could also be screened to identify variants that are preferentially cleaved, providing a route to systematically probe the specificity landscape and discover new substrate sequences. These

recombinant substrates would need to be accurately delivered to their target sites, increasing the complexity of their application in practical settings. However, if efficient delivery strategies are developed and substrate engineering methods become more predictive, such approaches could enable practical and flexible deployment of engineered proteolytic systems.

Similar to emerging protease engineering methods, substrate engineering with AI-guided technologies could also be envisioned. As development of binders and inhibitors for protein targets with deep learning structural and protein language models becomes increasingly ubiquitous [91], these design principles could be adapted to design protease substrates for PADS. Designing around the sequence and structural features of the natural substrate cleavage site offers one strategy, while fully *de novo* substrate design could further expand the functional space. However, protease cleavage specificity and substrate pocket fit will likely present a challenging path. If overcome, substrate engineering could provide a powerful parallel to protease engineering, enabling programmable proteolytic circuits where both the enzyme and its synthetic substrates are rationally designed.

Concluding remarks

The discovery of PADS has fundamentally expanded our understanding of bacterial immunity. In the dynamic arms race between bacteria and phages, evolutionary pressures drive the adaptation of protease activity and inhibition.

As more PADS in bacterial immunity are identified, they are expected to yield significant insights into protease biology and engineering (see Outstanding questions). A deeper understanding of protease activation, specificity, and substrate recognition will inform strategies to predict off-target effects and optimize enzyme engineering. Phages may also encode protease inhibitors to counteract host defenses or deploy proteases to bypass bacterial immunity [98]. This interplay invites further investigation into whether bacterial and archaeal viruses, akin to eukaryotic viruses, encode proteases with moonlighting functions or immunity-suppressing roles.

Looking ahead, rational substrate design for proteases is becoming increasingly feasible, either through classical bioengineering approaches or using AI-driven technologies. Designing such substrates could provide a more controlled and predictable means of manipulating protease activity, but it also introduces additional complexities. However, substrate engineering would open a new frontier where proteolytic processing can be precisely tailored for desired molecular changes.

Promiscuity and lack of controllability have posed challenges for the biotechnological application of proteases [99]. However, the CRISPR-controlled proteases, being controlled by specific RNA molecules, have demonstrated exceptional utility in applications requiring specific recognition and activation. For example, Craspase and PCaspase have been already repurposed for RNA diagnostics, leveraging protein or peptide cleavage as a readout [45,100]. It is of note that, so far, all known CRISPR PADS are derived from type III CRISPR–Cas systems. Conventional type III CRISPR–Cas stand out due to their multifunctional effector complexes that couple RNA sensing with DNase and cyclase activities via the Cas10 subunit, triggering multistep signaling pathways [101]. Layered signal transduction is well suited to the incorporation of protease modules, whose own activity often unfolds through sequential proteolytic cascades. Integrating proteases into this architecture provides additional checkpoints, potentially enabling tighter control and lowering the risk of deleterious off-target activation, a crucial concern when highly promiscuous proteases are involved.

Outstanding questions

Prevalence of protease-containing defense systems: what is the full extent and diversity of PADS in nature, and how widely are they used in defense against phages? How did these systems evolve, and what is the full extent of similarities with defense mechanisms in other life kingdoms and multicellular eukaryotes?

Mechanisms of protease activation: what other mechanisms exist for sensing, guiding, and inducing protease activity across different systems?

Targets and specificity: what are the targets of these proteases, and are there off-target effects of these systems within the cognate proteome or in non-native organisms that would hamper applications?

Specificity of protease interfaces: what are the natural target sequences for the proteases in PADS? Which substrate properties govern the specificity of these proteases?

Target and sensing versatility: what are the limitations of targeting transcripts in eukaryotes using CRISPR-controlled proteases? Can sensor and effector domains be adapted to act on different targets?

System dynamics and kinetics: what are the dynamics and kinetics of the different components involved in PADS with multiple activities and checkpoints?

Engineering flexibility of PADS: how adaptable are the protease systems to modifications in their sensor and effector mechanisms?

Efficiency and compactness in engineered systems: when engineered, can these systems be made more compact and efficient for delivery? Is it possible to fuse domains or co-deliver substrates with the system to simplify deployment?

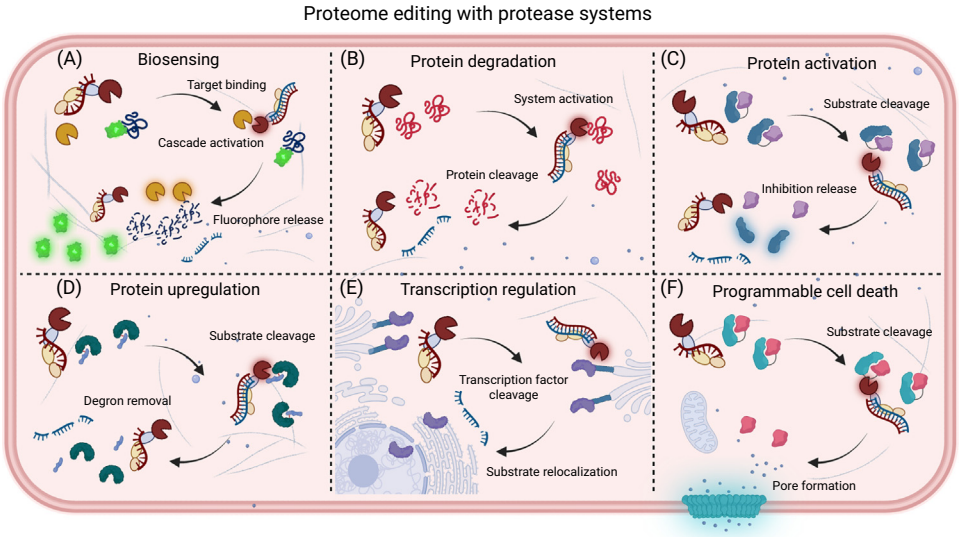

Figure 4. Potential applications of protease-associated defense systems (PADS) in bioengineering and protein editing. Craspase, the type III clustered regularly interspaced short palindromic repeats (CRISPR)-associated PADS, is shown as an example system. Here, RNA hybridization activates the protease, but the same principles could apply to protein, metabolite, or synthetic triggers. Preliminary studies have already demonstrated applications in biosensing, protein regulation, and transcriptional control [45,100]. (A) Biosensing: target recognition activates a protease that cleaves a quencher domain, releasing a fluorophore signal. (B) Protein degradation: activation of a protease leads to cleavage of a target protein, directing it for destruction. (C) Protein activation/upregulation: proteolytic processing releases a bioactive fragment or removes an inhibitory domain, stabilizing the protein and enhancing its abundance. (D) Protein upregulation via degron removal: cleavage of a degron sequence prevents degradation, increasing protein stability. (E) Transcriptional regulation: protease activation enables release or processing of a factor that translocates to the nucleus to induce gene expression. (F) Programmable cell death: cleavage of gasdermin or similar proteins allows pore formation, inducing pyroptosis specifically in cells containing the activating signal. Figure created with BioRender.

Opportunities for innovation lie in adapting these systems for biotechnological applications, where engineering proteases to sense specific signals or activate under defined conditions could lead to transformative tools in diagnostics, therapeutics, and synthetic biology (Figure 4). Nevertheless, deploying multicomponent protease systems presents challenges, particularly in delivery and functional integration within cellular environments. Accurate control over these systems within biological pathways requires advances in delivery technologies and a nuanced understanding of systems biology and cellular dynamics. Future research will determine whether PADS will live up to these expectations.

Acknowledgments

K.K. is supported by a Novo Nordisk Foundation Young Investigator Award (grant no. NNF16OC0020670) and a postdoctoral fellowship grant from the Independent Research Fund Denmark (grant no. 4257-00010B). This work was also supported by grants from the European Research Council (ERC) CoG under the European Union's Horizon 2020 research and innovation program (grant no. 101003229) to S.J.J.B.

Declaration of interests

The authors declare no conflicts of interest.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used GPT-4o in order to improve the readability and language of the manuscript text. The authors reviewed and edited the content as needed, and take full responsibility for the content of the published article.

References

1. Rostøl, J.T. and Marraffini, L. (2019) (Ph)ighting phages: how bacteria resist their parasites. *Cell Host Microbe* 25, 184–194
2. Hampton, H.G. *et al.* (2020) The arms race between bacteria and their phage foes. *Nature* 577, 327–336
3. Doron, S. *et al.* (2018) Systematic discovery of antiphage defense systems in the microbial pangenome. *Science* 359, eaar4120
4. Gao, L. *et al.* (2020) Diverse enzymatic activities mediate antiviral immunity in prokaryotes. *Science* 369, 1077–1084
5. Gao, Y. *et al.* (2020) Structural insights into assembly, operation and inhibition of a type I restriction–modification system. *Nat. Microbiol.* 5, 1107–1118
6. Sundaram, B. *et al.* (2024) NLRC5 senses NAD+ depletion, forming a PANoptosome and driving PANoptosis and inflammation. *Cell* 187, 4061–4077.e17
7. Hu, C. *et al.* (2022) Craspase is a CRISPR RNA-guided, RNA-activated protease. *Science* 377, 1278–1285
8. Rouillon, C. *et al.* (2023) Antiviral signalling by a cyclic nucleotide activated CRISPR protease. *Nature* 614, 168–174
9. Steens, J.A. *et al.* (2024) Type III-B CRISPR-Cas cascade of proteolytic cleavages. *Science* 383, 512–519
10. López-Otín, C. and Bond, J.S. (2008) Proteases: multifunctional enzymes in life and disease. *J. Biol. Chem.* 283, 30433–30437
11. Makarova, K.S. *et al.* (2020) Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.* 18, 67–83
12. Millman, A. *et al.* (2020) Diversity and classification of cyclic-oligonucleotide-based anti-phage signalling systems. *Nat. Microbiol.* 5, 1608–1615
13. Chung, H.K. and Lin, M.Z. (2020) On the cutting edge: protease-based methods for sensing and controlling cell biology. *Nat. Methods* 17, 885–896
14. Dyer, R.P. and Weiss, G.A. (2022) Making the cut with protease engineering. *Cell Chem. Biol.* 29, 177–190
15. Holt, B.A. and Kwong, G.A. (2020) Protease circuits for processing biological information. *Nat. Commun.* 11, 5021
16. Julien, O. and Wells, J.A. (2017) Caspases and their substrates. *Cell Death Differ.* 24, 1380–1389
17. Man, S.M. and Kanneganti, T.-D. (2016) Converging roles of caspases in inflammasome activation, cell death and innate immunity. *Nat. Rev. Immunol.* 16, 7–21
18. Dunkelberger, J.R. and Song, W.-C. (2010) Complement and its role in innate and adaptive immune responses. *Cell Res.* 20, 34–50
19. Szulc-Dąbrowska, L. *et al.* (2020) Cathepsins in bacteria-macrophage interaction: defenders or victims of circumstance? *Front. Cell. Infect. Microbiol.* 10, 601072
20. Harty, R.N. *et al.* (2009) Antiviral activity of innate immune protein ISG15. *J. Innate Immun.* 1, 397–404
21. Goldberg, K. *et al.* (2025) Cell-autonomous innate immunity by proteasome-derived defence peptides. *Nature* 639, 1032–1041
22. Liu, Y.-C. *et al.* (2005) Immunity by ubiquitylation: a reversible process of modification. *Nat. Rev. Immunol.* 5, 941–952
23. Dalbey, R.E. *et al.* (2012) Membrane proteases in the bacterial protein secretion and quality control pathway. *Microbiol. Mol. Biol. Rev.* 76, 311–330
24. Liu, G. *et al.* (2020) Bacterial rhomboid proteases mediate quality control of orphan membrane proteins. *EMBO J.* 39, e102922
25. Bieniossek, C. *et al.* (2006) The molecular architecture of the metalloprotease FtsH. *Proc. Natl. Acad. Sci.* 103, 3066–3071
26. Alqarzaee, A.A. *et al.* (2021) Staphylococcal ClpXP protease targets the cellular antioxidant system to eliminate fitness-compromised cells in stationary phase. *Proc. Natl. Acad. Sci.* 118, e2109671118
27. Johnson, A.G. *et al.* (2024) Structure and assembly of a bacterial gasdermin pore. *Nature* 628, 657–663
28. Bordes, P. and Genevaux, P. (2021) Control of toxin-antitoxin systems by proteases in *Mycobacterium tuberculosis*. *Front. Mol. Biosci.* 8, 691399
29. Rizzolo, K. *et al.* (2021) Functional cooperativity between the trigger factor chaperone and the ClpXP proteolytic complex. *Nat. Commun.* 12, 281
30. Swift, S. *et al.* (1999) Quorum sensing-dependent regulation and blockade of exoprotease production in *Aeromonas hydrophila*. *Infect. Immun.* 67, 5192–5199
31. Martín, A.C. *et al.* (1998) Pneumococcal bacteriophage Cp-1 encodes its own protease essential for phage maturation. *J. Virol.* 72, 3491–3494
32. Fokine, A. and Rossmann, M.G. (2016) Common evolutionary origin of procapsid proteases, phage tail tubes, and tubes of bacterial type VI secretion systems. *Structure* 24, 1928–1935
33. Hill, M.E. *et al.* (2018) The unique cofactor region of Zika virus NS2B-NS3 protease facilitates cleavage of key host proteins. *ACS Chem. Biol.* 13, 2398–2405
34. Mali, P. *et al.* (2013) RNA-guided human genome engineering via Cas9. *Science* 339, 823–826
35. Vasu, K. and Nagaraja, V. (2013) Diverse functions of restriction-modification systems in addition to cellular defense. *Microbiol. Mol. Biol. Rev.* 77, 53–72
36. Wilkinson, M. *et al.* (2022) Structures of RecBCD in complex with phage-encoded inhibitor proteins reveal distinctive strategies for evasion of a bacterial immunity hub. *eLife* 11, e83409
37. Wein, T. *et al.* (2025) CARD domains mediate anti-phage defence in bacterial gasdermin systems. *Nature* 639, 727–734
38. Rousset, F. *et al.* (2025) TIR signaling activates caspase-like immunity in bacteria. *Science* 387, 510–516
39. Aravind, L. and Koonin, E.V. (2002) Classification of the caspase–hemoglobinase fold: detection of new families and implications for the origin of the eukaryotic separins. *Proteins Struct. Funct. Bioinforma.* 46, 355–367
40. van Beljouw, S.P.B. and Brouns, S.J.J. (2024) CRISPR-controlled proteases. *Biochem. Soc. Trans.* 52, 441–453
41. Kato, K. *et al.* (2022) Structure and engineering of the type III-E CRISPR-Cas7-11 effector complex. *Cell* 185, 2324–2337.e16
42. van Beljouw, S.P.B. *et al.* (2021) The gRAMP CRISPR-Cas effector is an RNA endonuclease complexed with a caspase-like peptidase. *Science* 373, 1349–1353
43. Cui, N. *et al.* (2022) Structural basis for the non-self RNA-activated protease activity of the type III-E CRISPR nuclease-protease Craspase. *Nat. Commun.* 13, 7549
44. Ekundayo, B. *et al.* (2022) Structural insights into the regulation of Cas7-11 by TPR-CHAT. *Nat. Struct. Mol. Biol.* 30, 135–139
45. Strecker, J. *et al.* (2022) RNA-activated protein cleavage with a CRISPR-associated endopeptidase. *Science* 378, 874–881
46. Bernal-Bernal, D. *et al.* (2024) Structural basis for regulation of a CBASS-CRISPR-Cas defense island by a transmembrane anti-σ factor and its ECF σ partner. *Sci. Adv.* 10, eadp1053
47. Millman, A. *et al.* (2022) An expanded arsenal of immune systems that protect bacteria from phages. *Cell Host Microbe* 30, 1556–1569.e5
48. Grossman, M. *et al.* (2011) Correlated structural kinetics and retarded solvent dynamics at the metalloprotease active site. *Nat. Struct. Mol. Biol.* 18, 1102–1108
49. Yan, Y. *et al.* (2024) Phage defence system CBASS is regulated by a prokaryotic E2 enzyme that imitates the ubiquitin pathway. *Nat. Microbiol.* 9, 1566–1578
50. Ledvina, H.E. *et al.* (2023) An E1-E2 fusion protein primes antiviral immune signalling in bacteria. *Nature* 616, 319–325
51. Krüger, L. *et al.* (2024) Reversible conjugation of a CBASS nucleotide cyclase regulates bacterial immune response to phage infection. *Nat. Microbiol.* 9, 1579–1592
52. Hör, J. *et al.* (2024) Bacteria conjugate ubiquitin-like proteins to interfere with phage assembly. *Nature* 631, 850–856
53. Chambers, L.R. *et al.* (2024) A eukaryotic-like ubiquitination system in bacterial antiviral defence. *Nature* 631, 843–849
54. van den Berg, D.F. *et al.* (2024) Bacterial homologs of innate eukaryotic antiviral defenses with anti-phage activity highlight shared evolutionary roots of viral defenses. *Cell Host Microbe* 32, 1427–1443.e8
55. Vassallo, C.N. *et al.* (2022) A functional selection reveals previously undetected anti-phage defence systems in the *E. coli* pangenome. *Nat. Microbiol.* 7, 1568–1579
56. Bae, H.-W. *et al.* (2025) *Pseudomonas aeruginosa* MucP contributes to RNA phage resistance by targeting phage lysis.

bioRxiv Published online April 21, 2025. doi.org/10.1101/2024.12.09.627492
57. Lau, R.K. *et al.* (2022) A conserved signaling pathway activates bacterial CBASS immune signaling in response to DNA damage. *EMBO J.* 41, e111540
58. Bezawork-Geleta, A. *et al.* (2015) LON is the master protease that protects against protein aggregation in human mitochondria through direct degradation of misfolded proteins. *Sci. Rep.* 5, 17397
59. Binder, S.C. *et al.* (2024) The SAVED domain of the type III CRISPR protease CalpL is a ring nuclease. *Nucleic Acids Res.* 52, 10520–10532
60. Smalakyte, D. *et al.* (2024) Filament formation activates protease and ring nuclease activities of CRISPR Lon-SAVED. *Mol. Cell* 84, 4239–4255.e8
61. Goldfarb, T. *et al.* (2015) BREX is a novel phage resistance system widespread in microbial genomes. *EMBO J.* 34, 169–183
62. Gordeeva, J. *et al.* (2019) BREX system of *Escherichia coli* distinguishes self from non-self by methylation of a specific DNA site. *Nucleic Acids Res.* 47, 253–265
63. Tuck, O.T. *et al.* (2025) Recurrent acquisition of nuclease-protease pairs in antiviral immunity. *bioRxiv* Published online July 28, 2025. doi.org/10.1101/2025.07.28.667249
64. Huang, P. *et al.* (2025) The anti-phage mechanism of a widespread trypsin-MBL module. *bioRxiv* Published online September 3, 2025. doi.org/10.1101/2025.09.03.673762
65. Herbst, E. *et al.* (2025) Extracellular activity of a bacterial protease associated with reduced phage infectivity. *bioRxiv* Published online September 3, 2025. doi.org/10.1101/2025.09.03.674096
66. Donegan, N.P. *et al.* (2010) Proteolytic regulation of toxin-antitoxin systems by ClpPC in *Staphylococcus aureus*. *J. Bacteriol.* 192, 1416–1422
67. Cui, Y. *et al.* (2022) Bacterial MazF/MazE toxin-antitoxin suppresses lytic propagation of arbitrium-containing phages. *Cell Rep.* 41, 111752
68. Nikolic, N. *et al.* (2025) A bacterial toxin–antitoxin system as a native defence element against RNA phages. *Biol. Lett.* 21, 20250080
69. Shmakov, S.A. *et al.* (2018) Systematic prediction of genes functionally linked to CRISPR-Cas systems by gene neighborhood analysis. *Proc. Natl. Acad. Sci. USA* 115, E5307–E5316
70. Liu, L.-J. *et al.* (2021) Physiology, taxonomy, and sulfur metabolism of the Sulfolobales, an order of thermoacidophilic archaea. *Front. Microbiol.* 12, 768283
71. Sivá, M. *et al.* (2016) Human DNA-damage-inducible 2 protein is structurally and functionally distinct from its yeast ortholog. *Sci. Rep.* 6, 30443
72. Shimohata, N. *et al.* (2002) The Cpx stress response system of *Escherichia coli* senses plasma membrane proteins and controls HtpX, a membrane protease with a cytosolic active site. *Genes Cells Devoted Mol. Cell. Mech.* 7, 653–662
73. Bose, B. *et al.* (2008) A conserved anti-repressor controls horizontal gene transfer by proteolysis. *Mol. Microbiol.* 70, 570–582
74. Bergbold, N. and Lemberg, M.K. (2013) Emerging role of rhomboid family proteins in mammalian biology and disease. *Biochim. Biophys. Acta* 1828, 2840–2848
75. Shah, S.A. *et al.* (2019) Comprehensive search for accessory proteins encoded with archaeal and bacterial type III CRISPR-cas gene cassettes reveals 39 new cas gene families. *RNA Biol.* 16, 530–542
76. Altae-Tran, H. *et al.* (2023) Uncovering the functional diversity of rare CRISPR-Cas systems with deep terascale clustering. *Science* 382, eadi1910
77. Lowey, B. *et al.* (2020) CBASS immunity uses CARF-related effectors to sense 3′-5′- and 2′-5′-linked cyclic oligonucleotide signals and protect bacteria from phage infection. *Cell* 182, 38–49.e17
78. Gao, L.A. *et al.* (2022) Prokaryotic innate immunity through pattern recognition of conserved viral proteins. *Science* 377, eabm4096
79. Kaur, G. *et al.* (2020) Highly regulated, diversifying NTP-dependent biological conflict systems with implications for the emergence of multicellularity. *eLife* 9, e52696
80. Palzkill, T. (2013) Metallo-β-lactamase structure and function. *Ann. N. Y. Acad. Sci.* 1277, 91–104
81. Makarova, K.S. *et al.* (2011) Defense islands in bacterial and archaeal genomes and prediction of novel defense systems. *J. Bacteriol.* 193, 6039–6056
82. Koonin, E.V. *et al.* (2017) Evolutionary genomics of defense systems in Archaea and Bacteria. *Ann. Rev. Microbiol.* 71, 233–261
83. Hedstrom, L. *et al.* (1992) Converting trypsin to chymotrypsin: the role of surface loops. *Science* 255, 1249–1253
84. Packer, M.S. *et al.* (2017) Phage-assisted continuous evolution of proteases with altered substrate specificity. *Nat. Commun.* 8, 956
85. Sanchez, M.I. and Ting, A.Y. (2020) Directed evolution improves the catalytic efficiency of TEV protease. *Nat. Methods* 17, 167–174
86. Ferrall-Fairbanks, M.C. *et al.* (2020) Reassessing enzyme kinetics: considering protease-as-substrate interactions in proteolytic networks. *Proc. Natl. Acad. Sci.* 117, 3307–3318
87. Zhou, J. *et al.* (2020) Deep profiling of protease substrate specificity enabled by dual random and scanned human proteome substrate phage libraries. *Proc. Natl. Acad. Sci.* 117, 25464–25475
88. Watson, J.L. *et al.* (2023) *De novo* design of protein structure and function with RFdiffusion. *Nature* 620, 1089–1100
89. Krishna, R. *et al.* (2024) Generalized biomolecular modeling and design with RoseTTAFold All-Atom. *Science* 384, eadl2528
90. Dauparas, J. *et al.* (2022) Robust deep learning–based protein sequence design using ProteinMPNN. *Science* 378, 49–56
91. Lin, Z. *et al.* (2023) Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* 379, 1123–1130
92. Anishchenko, I. *et al.* (2025) Modeling protein-small molecule conformational ensembles with ChemNet. *bioRxiv* Published online September 27, 2025. doi.org/10.1101/2024.09.25.614868
93. Ahern, W. *et al.* (2025) Atom level enzyme active site scaffolding using RFdiffusion2. *bioRxiv* Published online April 10, 2025. doi.org/10.1101/2025.04.09.648075
94. Braun, M. *et al.* (2025) Computational design of highly active de novo enzymes. *bioRxiv* Published online June 11, 2025. doi.org/10.1101/2024.08.02.606416
95. Kim, D. *et al.* (2025) Computational design of metallohydrolases. *bioRxiv* Published online April 11, 2025. doi.org/10.1101/2024.11.13.623507
96. Lauko, A. *et al.* (2025) Computational design of serine hydrolases. *Science* 388, eadu2454
97. Abramson, J. *et al.* (2024) Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 630, 493–500
98. Hilliard, J.J. *et al.* (1998) PinA inhibits ATP hydrolysis and energy-dependent protein degradation by Lon protease. *J. Biol. Chem.* 273, 524–527
99. Martinusen, S.G. *et al.* (2025) Protease engineering: approaches, tools, and emerging trends. *Biotechnol. Adv.* 82, 108602
100. Kato, K. *et al.* (2022) RNA-triggered protein cleavage and cell growth arrest by the type III-E CRISPR nuclease-protease. *Science* 378, 882–889
101. Niewoehner, O. *et al.* (2017) Type III CRISPR–Cas systems produce cyclic oligoadenylate second messengers. *Nature* 548, 543–548
102. Rawlings, N.D. *et al.* (2017) The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res.* 46, D624–D632
103. Mótyán, J.A. *et al.* (2013) Research applications of proteolytic enzymes in molecular biology. *Biomolecules* 3, 923–942
104. Rawlings, N.D. (2020) Twenty-five years of nomenclature and classification of proteolytic enzymes. *Biochim. Biophys. Acta BBA Proteins Proteomics* 1868, 140345
105. Hickey, C.M. *et al.* (2012) Function and regulation of SUMO proteases. *Nat. Rev. Mol. Cell Biol.* 13, 755–766
106. Eckhard, U. *et al.* (2016) Active site specificity profiling of the matrix metalloproteinase family: proteomic identification of

4300 cleavage sites by nine MMPs explored with structural and synthetic peptide cleavage analyses. *Matrix Biol. J. Int. Soc. Matrix Biol.* 49, 37–60
107. Sakr, M. *et al.* (2018) Tracking the Cartoon mouse phenotype: hemopexin domain-dependent regulation of MT1-MMP pericellular collagenolytic activity. *J. Biol. Chem.* 293, 8113–8127
108. Ra, H.-J. and Parks, W.C. (2007) Control of matrix metalloproteinase catalytic activity. *Matrix Biol. J. Int. Soc. Matrix Biol.* 26, 587–596
109. Rudzińska, M. *et al.* (2021) Current status and perspectives of protease inhibitors and their combination with nanosized drug delivery systems for targeted cancer therapy. *Drug Des. Devel. Ther.* 15, 9–20
110. Turk, B. *et al.* (2012) Protease signalling: the cutting edge. *EMBO J.* 31, 1630–1643
111. Wilk, S. (2005) Proteases and signaling. *Sci. STKE* 2005, tr15
112. Turk, B. (2006) Targeting proteases: successes, failures and future prospects. *Nat. Rev. Drug Discov.* 5, 785–799
113. Cupp-Enyard, C. (2008) Sigma's non-specific protease activity assay – casein as a substrate. *J. Vis. Exp. JoVE* 17, 899
114. LAEMMLI, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680–685
115. Hewick, R.M. *et al.* (1981) A gas–liquid solid phase peptide and protein sequenator. *J. Biol. Chem.* 256, 7990–7997
116. Voss, E.W., Jr. *et al.* (1996) Detection of protease activity using a fluorescence-enhancement globular substrate. *BioTechniques* 20, 286–291
117. Kleifeld, O. *et al.* (2010) Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products. *Nat. Biotechnol.* 28, 281–288
118. Weng, S.S.H. *et al.* (2019) Sensitive determination of proteolytic proteoforms in limited microscale proteome samples. *Mol. Cell. Proteomics MCP* 18, 2335–2347
119. Bridge, H.N. *et al.* (2023) An N terminomics toolbox combining 2-pyridinecarboxaldehyde probes and click chemistry for profiling protease specificity. *Cell Chem. Biol.* 31, 534–549.e8
120. Weeks, A.M. and Wells, J.A. (2020) N-terminal modification of proteins with subtiligase specificity variants. *Curr. Protoc. Chem. Biol.* 12, e79
121. Schilling, O. and Overall, C.M. (2008) Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites. *Nat. Biotechnol.* 26, 685–694
122. Kalogeropoulos, K. *et al.* (2019) Protease activity profiling of snake venoms using high-throughput peptide screening. *Toxins* 11, 170
123. Blum, G. *et al.* (2005) Dynamic imaging of protease activity with fluorescently quenched activity-based probes. *Nat. Chem. Biol.* 1, 203–209
124. Mirdita, M. *et al.* (2022) ColabFold: making protein folding accessible to all. *Nat. Methods* 19, 679–682