Published in final edited form as:

Nat Struct Mol Biol. ; 19(1): 9–16. doi:10.1038/nsmb.2203.

New tools for dissecting protease function: implications for inhibitor design, drug discovery and probe development

Edgar Deu¹, Martijn Verdoes¹, and Matthew Bogyo^{1,2}

¹Department of Pathology, Stanford University School of Medicine, 300 Pasteur Dr. Stanford, CA 94305-5324

²Department of Microbiology and Immunology, Stanford University School of Medicine, 300 Pasteur Dr. Stanford, CA 94305-5324

Abstract

Proteases have long been considered targets for pharmaceutical development because of our deep understanding of their enzymatic mechanism and their regulatory roles in many pathologies. However, despite our ability to develop potent inhibitors, many clinical lead compounds have failed due either to a lack of specificity or a limited understanding of the biological roles of the targeted protease. In order to successfully develop protease inhibitors as drugs, it is necessary to first understand protease function and second, to expand the platform of inhibitor development beyond active site-directed design and *in vitro* optimization. Several newly developed technologies will enable much broader assessment of drug selectivity in living cells and in animal models, allowing for lead optimization to focus on compounds that show high specificity and minimal side effects *in vivo*. In this perspective, we highlight the current advances in the development of new chemical probes, proteomic methods, and screening tools that we feel will help facilitate this paradigm shift in drug discovery methods.

INTRODUCTION

Proteases are one of the most abundant classes of enzymes in living organisms and are involved in a wide range of biological processes including cell cycle progression, protein trafficking, cell death, immune response, cell proliferation, and cell signaling. Additionally, proteases play pathological roles in many human diseases ranging from degenerative and inflammatory diseases to infectious diseases including viruses (e.g. HIV), bacteria (e.g. cholera) and parasites (e.g. malaria). Therefore, proteases are often targeted by the pharmaceutical industry, and protease inhibitor drugs are currently in use for treatment of coagulation disorders, hypertension, AIDS, cancer, and diabetes¹.

Proteases bind their substrates through hydrogen bond interactions with the substrate peptide backbone and hydrophobic and electrostatic contacts between the substrate side chains and well-defined pockets within the active site (Figure 1A). Proteases have been classified into seven distinct classes based on the amino acid or ion that catalyzes peptide bond cleavage. Cysteine, serine and threonine proteases catalyze amide cleavage through nucleophilic attack of the key side chain residue, which leads to the formation of a covalent acyl-enzyme intermediate. Hydrolysis by a water molecule liberates the peptide products (Figure 1B–D). In the case of aspartate, glutamate and metalloproteases, acid/based catalysis takes place through direct activation of a water molecule by a carboxylic acid group or metal ion (Figure 1E–F). The seventh and newest protease family, the asparagine peptide lyases, cleave

themselves using an asparagine residue as the nucleophile². The mechanism of substrate cleavage determines which type of chemical entity can be used to inhibit each family of proteases³. For cysteine, serine and threonine proteases, an electrophilic group can be used to covalently modify the catalytic residue in a reversible or irreversible manner. In the case of metalloproteases, the primary strategy is to use functional groups that coordinate the catalytic metal to achieve potent inhibition. For all protease families, potent transition state analogues can be designed based on structural and enzymatic studies.

Current advances in high throughput screening (HTS) technologies, structural biology, computational modeling and combinatorial chemistry have enabled the design of potent protease inhibitors^{1,4,5}. However, because the reaction mechanism is highly conserved among each protease class and because proteases often have many closely related family members, lead compounds often inhibit more than one target, potentially resulting in unwanted side effects *in vivo*. Achieving target specificity has been the main hurdle in developing protease inhibitors as drugs. Unfortunately, the traditional approach of hit-to-lead optimization only takes into account off-target effects fairly late in the process (Figure 2A). Drug discovery efforts usually start with HTS using recombinant or purified enzyme, followed by structural studies of lead compounds to aid in the design of second-generation inhibitors with improved affinity. Usually only a limited number of enzymes within the same family are available to assess off-target reactivity *in vitro*. Only then are compounds tested in cells or in animal models where the more significant hurdles usually arise. While advances in structural biology have revolutionized the way we design inhibitors, it is necessary to broaden the platform of inhibitor development to allow direct screening of compounds in cells and *in vivo*. This approach focuses development on compounds that show minimal side effects and that inhibit the target protease in a biologically relevant context (Figure 2B).

Although there have been a number of success stories in the development of drugs that target proteases^{1,5}, here we highlight what we believe are the two main challenges in the development of protease inhibitors as drugs: 1) Validating proteases as drug targets and obtaining a clear understanding of their biological functions. 2) Designing specific inhibitors despite the large number of proteases found in an organism. We also present recent technological advances that we believe will help address these challenges. We emphasize the use of activity-based probes (ABPs) as tools to understand the biological role of proteases, and to measure off-target effects and on-target inhibition *in vitro* and *in vivo*. We also describe recent advances in screening technologies in intact cells and in living animals, as well as recently developed proteomic approaches for the identification of protease substrates.

UNDERSTANDING THE BIOLOGICAL FUNCTION OF PROTEASE TARGETS

Dissecting how proteases perform their biological function is extremely challenging mainly due to the multiple mechanisms by which the cell regulates this class of enzymes. Direct recognition of the substrate by the active site is but one of the multiple strategies used to control the specificity and activity of proteases (Figure 3A). Proteases can be regulated by small molecules, by post-translational modifications or through interactions with other proteins^{1,5,6}. Some proteases recognize their substrates through interactions at an exosite distal to the active site⁷. On a cellular level, protease activity and specificity can be regulated by mechanisms such as zymogen activation, targeted degradation, sub-cellular localization, and by binding of endogenous inhibitors. Therefore, mRNA levels and protein abundance and localization often fail to inform when and where a protease is active.

The majority of proteases are translated as zymogens that need to be activated, usually through proteolysis or conformational alteration. This activation event is spatially and temporally controlled by distinct mechanisms. For example, cysteine cathepsins are processed in a pH-dependent manner when they reach the lysosome⁸. Other proteases, such as the Cysteine Protease Domain (CPD) of the MARTX toxins found in *Vibrio cholera*⁹, have evolved allosteric mechanisms of activation. Alternatively, activation can be regulated through the formation of large protein complexes that ensure a protease does not become activated prematurely (i.e. activation of caspase-1 upon formation of the inflammasome complex¹⁰, or by formation of receptor-triggered complexes, as exemplified by initiator caspases activation¹¹). Finally proteases such as the proteasome regulate their activity through self-compartmentalization of their proteolytic subunits.

Limiting or enhancing access to substrates within the cell is one of the predominant ways in which protease function is regulated. For example, turnover of substrates by the proteasome is tightly controlled by the ubiquitination pathway, and access to the proteolytic cavity is regulated in an ATP-dependent manner¹². Alternatively, protein turnover by proteases such as the cathepsins is controlled by trafficking of substrates into lysosomes. However, depending on their localization, cathepsins have also been shown to regulate innate and adaptive immunity as well as tumor progression⁸. Release of cathepsin B into the cytosol activates caspase-8 and triggers apoptosis⁸; export of cathepsins to the tumor microenvironment results in degradation of the extracellular matrix (ECM)^{8,13}, and trafficking of cathepsin C in NK cell and T lymphocyte granules results in the activation of serine proteases involved in the innate immune response and inflammation⁸.

Therefore, proteases play multiple biological functions that are dynamically controlled by changes in localization in response to different stimuli. Moreover, proteases function in networks that interact with other major signaling pathways to communicate between cell types^{14,15}. This makes the identification of protease targets extremely complicated. For example, MMPs were initially thought to be ideal anti-cancer targets because of their extracellular localization and involvement in the degradation of the ECM during angiogenesis and metastasis. However, because MMPs are also involved in maintaining homeostasis of the extracellular environment, as well as cell signaling and innate immunity regulation, inhibition of some MMPs has been shown to promote tumor growth. This duality of functions, coupled with difficulty in designing specific inhibitors, resulted in the failure of many MMP inhibitors in clinical trials¹⁶. In complex multi-factorial diseases such as cancer, it is also important to know which protease is important in each disease sub-population. While increased levels of cathepsin B and cathepsin L activities correlate with tumor malignancy, an increase in cathepsin H and cathepsin S activities can result in either better or worse prognosis depending on the type of tumor¹³.

In some biological pathways, multiple proteases work together to achieve a common goal such as ECM degradation during metastasis or hemoglobin degradation in *Plasmodium species*. In these cases, it is imperative to identify proteases that are essential for the pathway. For example, most proteases involved in the initial stages of hemoglobin degradation (falcipains and plasmepsins) have redundant roles as demonstrated by the viability of parasites in which these proteases have been knocked out¹⁷. However, proteases involved at the later stages of degradation such as dipeptidyl aminopeptidase 1 (DPAP1)^{18,19} and several aminopeptidases²⁰ are essential. Therefore, a deep understanding of the interplay between proteases is crucial to identify drug targets. In the last few years a variety of new chemical, proteomic and protein engineering tools have been developed to study the biological function and regulation of protease activity.

New tools to study protease function in intact cells and live animals

Activity-Based Probes—ABPs are small molecule reporters designed to react only with the catalytically active form of a protease (Figure 3B)^{21–25}. Most ABPs consist of three parts; a warhead that irreversibly reacts with the active site nucleophile, a spacer and/or recognition element that targets the probe to a specific protease, and a tag, usually a fluorescent dye and/or an affinity handle. Because most ABPs are designed to be highly selective for a given protease or protease family, they can be used to assess protease activity within living cells or in whole organisms^{23–25}. Furthermore, because ABPs are generally inhibitors of the target protease, they can be used with better temporal control than using conditional gene disruption or RNAi methods. For these reasons, ABPs have begun to find applications in early stage drug discovery to identify and validate new targets^{18,26–30}. For example, biotin versions of eponemycin and epoxomycin were used to identify the proteasome as the target responsible for the antitumor activity of these natural products^{31,32}, and eventually lead to the development of epoxyketone inhibitors as proteasome-specific drugs currently in clinical development (Carfilzomib)³³. Because most ABPs irreversibly modify the enzyme active site, they can also be used to directly isolate target enzymes. Libraries of covalent inhibitors can be used in forward chemical genetics screens to identify compounds that induce a specific phenotype. Hits can then be converted into ABPs to isolate targeted enzymes for identification by LC-MS/MS (Figure 4A). This method has been successfully used to identify proteases involved in *Plasmodium falciparum* egress from infected red blood cells (RBCs)²⁶, and to identify proteins involved in *Toxoplasma gondii* invasion³⁴. ABPs are also especially useful in biological systems that are not amenable to genetic manipulation or regulation by RNAi. For example, a cysteine protease ABP was used to show that *Plasmodium falciparum* activates host calpain at the RBC membrane during egress³⁵. In the tumor microenvironment, ABPs have been used to identify infiltrating macrophages as the cells that express and activate cysteine cathepsins at the tumor invasive front in response to IL-4 secreted by cancer cells³⁶. ABPs have also been used to follow the kinetics of caspases activation during apoptosis³⁷. The use of ABPs in these examples was particularly important to differentiate and localize the active forms of the proteases.

ABPs have also been used to determine the extent of target inhibition needed to observe beneficial downstream effects. For example, complete and prolonged inhibition of cathepsin C is required to decrease the level of activity of downstream serine proteases that are involved in inflammation³⁸. Similarly, because DPAP1 is highly expressed in *P. falciparum*, its activity needs to be inhibited for at least 3 hours in order to kill the malaria parasite¹⁸.

Perhaps one of the most powerful applications for ABPs is their use to visualize protease activity using modern imaging modalities. Because ABPs covalently bind only to active protease targets, it is possible to monitor not only activation but also localization of protease activity. For example, cysteine cathepsin probes have been used to non-invasively detect cathepsin activity in tumors in multiple mouse models of cancer^{11,39,40}. In a single experiment, it is possible to localize protease activity in living animals, determine the distribution of protease activity in tissues, analyze protease activity at the cellular level and finally, perform biochemical analysis of protease activation (Figure 4B). ABPs can also be used *in vivo* to monitor efficacy of drug treatment³⁹ or to measure protease activation in response to a drug^{11,40}.

Although ABPs are valuable tools to study proteases, it is difficult to make probes that are selective for only one member of a protease family. In a recent example, probes that selectively inhibit and label the hepatitis C virus NS2/3 protease were designed by taking advantage of a cysteine residue that is relatively close to the active site⁴¹. A reversible inhibitor of NS2/3 was modified by adding a reactive electrophilic group such that it reacts

with the proximal cysteine residue. The covalent linkage between the inhibitor and the enzyme increased the compound potency and improved its overall selectivity. This approach may also be broadly applicable to proteases by engineering point mutations that render the target protease sensitive to ABP binding. This would be similar to recent methods for selective targeting and inhibition of kinases⁴².

Defining protease function through identification of substrates—It is notoriously difficult to identify protease substrates because the end product of the reaction has little information regarding the protease that produced it, and it must be identified in the sea of other cellular proteins. However, linking substrates to proteases is essential to understand their function. A number of proteomic approaches have recently been developed that allow specific enrichment of the newly generated peptide fragments that result from proteolysis within a biologically relevant context (Box 1)^{43–55}. These methods have been used to map proteolytic events during processes such as apoptosis^{48,50}, inflammation⁴³, or rupture of malaria infected RBCs⁴⁵.

In addition, new approaches to selectively activate a protease-mediated biological process have been devised. In one such example, the pro-enzyme cleavage site of several apoptotic caspases was converted into a TEV cleavage sequence such that the engineered caspases would remain inactive until cleaved. Using an engineered TEV protease that is activated by the addition of a small molecule, it was possible to temporally control caspase activity in the cell and identify distinct substrates for individual caspases during apoptosis⁵⁶. Since many proteases require proteolytic cleavage to be activated, this method could be broadly applied. Therefore, the use of ABPs in combination with proteomic approaches to identify natural substrates should help dissect the biological function of proteases.

DESIGNING SPECIFIC INHIBITORS

In vitro development of protease inhibitors

New approaches to structure-based design of protease inhibitors—The design of small molecule inhibitors is often challenging because the chemistry of peptide bond cleavage overlaps within protease classes, and proteases belonging to the same sub-family generally have similar substrate and inhibitor specificities (Figure 1). Consequently, structural biologists have focused on finding alternative ways to inhibit proteases, such as blocking allosteric sites and exosites¹. Recent studies suggest that some proteases cleave their substrates through an induced fit mechanism^{57,58}. For these proteases, inhibitors could be designed to prevent the induced fit movement by stabilizing the inactive conformation. Alternatively, molecules can be designed to lock an enzyme in its zymogen conformation as has been demonstrated for caspases^{59,60}. Finally, it may also be possible to inhibit a protease by targeting the acyl-enzyme intermediate if its hydrolysis is the rate limiting step in the catalytic process⁶¹.

Challenges for in vitro approaches for the development of inhibitors

Generally, the conditions in which inhibitors are tested *in vitro* (using recombinant or purified enzymes) are different from the enzyme's environment *in vivo*. Therefore IC₅₀s obtained *in vitro* often do not reflect *in vivo* values. This discrepancy has been demonstrated for the kinase family where the effect of inhibitors in cell-based assays correlate better with IC₅₀s determined in lysates than with recombinant enzymes⁶². There are number of possible reasons why *in vitro* testing with purified targets may provide only limited information on leads. First, specific binding of the inhibitor to off-targets or non-specific association with other cellular components might reduce the effective concentration of an inhibitor in the cell. Second, assay conditions *in vitro* might not reflect *in vivo* conditions because pH, salt concentration, viscosity, or protein concentrations can affect enzyme activity, specificity and

stability. Third, proteins are dynamic macromolecules that adopt a range of different conformations, and interactions with small molecules and other proteins can change the distribution of these conformations. For example, the CPD domain of MARTX toxin was initially reported to be allosterically activated by GTP⁶⁴. However, recent work has shown that inositol hexakisphosphate (InsP₆) is the biologically relevant activator^{9,65}. Moreover, the CPD domain of *C. difficile* shows some activity even in the absence of InsP₆⁶³. Therefore, prior to the identification of the activator, IC₅₀ values of CPD inhibitors would have been dramatically underestimated. Finally, *in vitro* systems fail to take into account issues such as delivery, partitioning, cell permeability and stability. For example, probes and inhibitors that seem to be highly specific for a cytosolic target *in vitro* often show cross-reactivity with lysosomal proteases due to endocytosis when tested in intact cells^{11,66}.

New approaches to define inhibitor specificity in intact cells and living animals

Screening protease inhibitors in a biological context—Performing assays in living cells has the potential to provide a more accurate determination of inhibitor potency, efficacy, and specificity. One approach is to design a reporter substrate that is primarily cleaved by the target protease in a complex cellular extract or even in intact cells. While it is typically very difficult to make substrates with absolute specificity for a given protease, ABPs can be used to screen for selective substrates by correlating target labeling with inhibition of substrate turnover. This method was recently used to identify the (Pro-Arg)₂Rho fluorescent substrate as selective for DPAP1 in *P. falciparum* lysates and for cathepsin C in rat liver extract^{18,67}. This substrate can be used for HTS thus avoiding the need to purify or express these proteases. This method could possibly identify selective reporter substrates in intact cells using cell-permeable ABPs. In support of this concept, the cathepsin C substrate Gly-Phe-AFC was found to be cell permeable and was not cleaved in cathepsin C deficient cell lines or in bone marrow lysates obtained from cathepsin C knockout mice⁶⁸.

Performing assays in lysates or in intact cells also allows for measuring multiple activities at the same time. For example, Wakata *et al.* developed two specific fluorogenic substrates, each containing a different fluorophore, that can be used to simultaneously measure the chymotrypsin- and caspase- like proteasomal activities in yeast lysates⁶⁹. Because of the recent advances in methods to map substrate specificity of proteases⁷⁰, we believe it will become more feasible to design substrates that can be used for screening in lysates and cells.

Finally, display methods can be used to screen large libraries of peptides to identify selective substrates *in vivo*. In a recent example, a phage-display library of peptides was injected into mice bearing mammary tumors⁷¹. After extraction and lysis of the tumor, cleaved phages were isolated, amplified, and re-injected into mice for additional rounds of selection. After seven rounds, five sequences were identified and used to make fluorogenic substrates that could be used for non-invasive imaging of cancer. Although the tumor proteases responsible for cleaving these substrates were not identified, such tumor-specific reporters of proteolysis could also be used to screen for novel inhibitors in an *in vivo* setting.

While cell-based screening methods, when coupled with effective tools for monitoring target inhibition, have the potential to be an effective way to identify lead compounds, it should also be noted that compounds that alter expression levels of protease targets would appear as active-site directed inhibitors. However, these types of hits might be valuable and could also be separated from the true active site inhibitors using counter screening assays or by assessment of protease levels using Western blots.

Design of non-peptidic inhibitors—One approach to generate specific protease inhibitors and substrates is to move away from the natural peptide scaffold by incorporating

non-natural amino acids. This approach has been used extensively to identify potent and selective inhibitors of multiple cysteine proteases^{18,26,72–75}. While peptidic inhibitors generally have poor pharmacological properties, the structure of a protease bound to a peptidic inhibitor has often served as the initial template for structure-based design of non-peptidic inhibitors and transition state analogues^{1,5}. Non-peptidic scaffolds for protease inhibitors are usually identified through screening methods such as HTS of small drug-like molecules^{1,5}, and more recently through *in silico* screening and fragment-based screening⁴. Once a hit has been identified, a combination of structural (NMR, crystallography), computational (docking), and SAR (structure-activity relationship) methods are used to optimize initial hits into potent drug leads^{1,4,5}.

Another promising method to identify new inhibitor scaffolds is substrate activity screening (SAS)^{76,77}. In this approach, diverse, non-peptidic molecules containing a coumarin fluorophore are screened against a protease of interest. The advantage of this method is that scaffolds with poor binding affinity can still be identified because turnover of these substrates results in accumulation of fluorescent signal over time. These hits can then be optimized and converted into potent inhibitors by replacing the coumarin group with an appropriate pharmacophore. This approach has been successfully used to identify novel non-peptidic inhibitor scaffolds with improved stability in mouse serum and better efficacy *in vivo*^{18,78,79}.

Global profiling of compound specificity—While selective ABPs are very useful tools to study the biology of a specific target, broad-spectrum ABPs that target a family of enzymes are also valuable to globally profile the specificity of an inhibitor against all members of the family (Figure 4C–D). Typically, these assays are performed such that the members of the targeted family can be resolved based on migration in an SDS-PAGE gel. Inhibition results in a diminished level of labeling compared to the control sample and can be readily quantified (Figure 4C)⁸⁰. This method can also be performed *in vivo* by pre-treating a whole animal with inhibitor and then labeling tissue extracts with an ABP^{28,81} or directly *in vivo* by intravenous injection of the probe⁸². This type of assay also provides information regarding the partitioning of the drug within specific tissues. For example, a cell-permeable broad-spectrum proteasome probe was used to show that bortezomib, a proteasome inhibitor currently used for the treatment of multiple myeloma, not only inhibits the $\beta 5$ and $\beta 5i$ subunit, but also targets $\beta 1$ and $\beta 1i$ ⁸³. Additionally, a broad-spectrum ABP was used to identify several serine proteases as off-targets of bortezomib⁸⁴, which can explain some of the side effects of this drug. In yet another example, an ABP-based SAR study showed that inhibitors of cathepsin K that failed in phase II clinical development for the treatment of osteoporosis (due to the formation of skin lesions in some patients)⁸⁵ accumulate in lysosomes and inhibit off-target cathepsins *in vivo*. This leads to an accumulation of intracellular collagen, which could explain the observed side effects^{23,86}. The lysosomotropism is presumably due to the presence of a basic amine group that becomes charged in the acidic lysosomes. In agreement with this assessment, the inhibitor Odanacatib which does not contain a basic amine, did not show this effect and is currently in phase III clinical trial.

All the ABPs mentioned above contain an electrophilic reactive group to irreversibly modify the cysteine, serine or threonine catalytic residues of target proteases. This strategy is not applicable to aspartate and metal proteases because hydrolysis of the peptide bond is mediated by a water molecule. Nonetheless, ABPs have been designed for metalloproteases by incorporating a moiety that coordinates the metal ion inside the active site. This strong interaction can then be used to pull down proteases for MS identification and quantification⁸⁷ or for *in vivo* imaging applications⁴⁴. Another approach is to introduce a photo-crosslinking group in addition to the chelating moiety to achieve covalent

modification of metalloproteases⁸⁸. This strategy has been used to develop aminopeptidase-specific probes⁸⁹ as well as broad-spectrum metalloprotease ABPs⁹⁰. These probes have been used to study the biological function of these proteases, select HTS hits based on their specificity profile⁹¹, or identify active metalloproteases in tissue extracts⁸⁸. This photo-crosslinking approach has also been used to target the presenilin aspartate proteases involved in processing of amyloid proteins⁹². While ABPs can be valuable to assess target selectivity, this method depends on having probes that can broadly target all the members of a protease family. Recently, a method was developed for global profiling the reactivity of all cysteines in a proteome by using an isotopically-labeled probe containing iodoacetamide as a general alkylating agent⁹³ (Figure 4D). Because cysteine proteases contain a highly reactive cysteine, this method could be used to monitor the effects of a given compound on the entire protease class. In principle, this method should also be translatable to serine protease by using a general serine reactive reporter.

In summary, by performing HTS in intact cells, it is possible to select for compounds that are cell permeable and able to inhibit the target in its biological context. The specificity of these hits can then be quickly evaluated using an activity-based global profiling method, which also provides the SAR information necessary to design more specific inhibitors. Such an approach was recently used to identify triazole ureas as potent serine hydrolase inhibitors and to develop potent inhibitors with highly selective inhibition profiles *in vivo*³⁰.

CONCLUSION

In summary, we feel that the main challenges facing protease drug development is our lack of understanding of the complex mechanisms by which protease activities are regulated, and of the multiple biological roles they play in diverse biological pathways. In addition, one of the main hurdles not only to develop safe protease inhibitor drugs, but also to understand the biological function of proteases is the difficulty in developing specific inhibitors in an *in vivo* context. While the classical development of inhibitors *in vitro* yields potent inhibitors, often these compounds do not translate into specific and/or effective inhibitors *in vivo* due to stability, cell permeability or partitioning issues. To overcome some of these challenges, we believe that compounds should be screened in intact cells and in animal models at an earlier stage in the drug development process. This would allow focusing of the hit-to-lead optimization process on those compounds that are able to reach their target within a biologically relevant context. Then, activity-based protein profiling methods can be used to determine the specificity of lead inhibitors against all members of a protease family and thus facilitate the design of highly specific inhibitors with reduced off-target effects.

ABPs are valuable imaging agents that can report on protease activity at the protein, cellular and whole organism levels. Therefore, they are ideal tools not only to study the biological function of proteases, but also to assess the *in vivo* efficacy of protease inhibitor treatments in real time. Moreover, identification of natural substrates using new proteomic methods will help define how proteases perform a specific biological function. This will also provide biomarkers to help evaluate whether inhibition of a protease target results in downstream biological effects. Overall, we believe that the advent of new tools such as ABPs, global protein profiling, and novel proteomic approaches will allow a shift in the focus of protease inhibitor development from that of potency *in vitro* to that of specificity and efficacy *in vivo*.

Acknowledgments

The authors thank Dr. Christopher Kirk for his comments and suggestions. This work was funded by NIH grants R01 EB005011 and R01 AI078947 (to M.B.) and a Rubicon grant from The Netherlands Organization for Scientific Research (NWO) (to M.V.).

References

1. Drag M, Salvesen GS. Emerging principles in protease-based drug discovery. *Nat Rev Drug Discov.* 2010; 9:690–701. [PubMed: 20811381]
2. Rawlings ND, Barrett AJ, Bateman A. Asparagine peptide lyases: a seventh catalytic type of proteolytic enzymes. *J Biol Chem.* 2011
3. Powers JC, Asgian JL, Ekici OD, James KE. Irreversible inhibitors of serine, cysteine, and threonine proteases. *Chem Rev.* 2002; 102:4639–4750. [PubMed: 12475205]
4. Johnson SL, Pellecchia M. Structure- and fragment-based approaches to protease inhibition. *Curr Top Med Chem.* 2006; 6:317–329. [PubMed: 16611145]
5. Turk B. Targeting proteases: successes, failures and future prospects. *Nat Rev Drug Discov.* 2006; 5:785–799. [PubMed: 16955069]
6. Shen A. Allosteric regulation of protease activity by small molecules. *Mol Biosyst.* 2010; 6:1431–1443. [PubMed: 20539873]
7. Bock PE, Panizzi P, Verhamme IM. Exosites in the substrate specificity of blood coagulation reactions. *J Thromb Haemost.* 2007; 5 (Suppl 1):81–94. [PubMed: 17635714]
8. Conus S, Simon HU. Cathepsins and their involvement in immune responses. *Swiss Med Wkly.* 2010; 140:w13042. [PubMed: 20648403]
9. Lupardus PJ, Shen A, Bogyo M, Garcia KC. Small molecule-induced allosteric activation of the *Vibrio cholerae* RTX cysteine protease domain. *Science.* 2008; 322:265–268. [PubMed: 18845756]
10. Bauernfeind F, et al. Inflammasomes: current understanding and open questions. *Cell Mol Life Sci.* 2011; 68:765–783. [PubMed: 21072676]
11. Edgington LE, et al. Noninvasive optical imaging of apoptosis by caspase-targeted activity-based probes. *Nat Med.* 2009; 15:967–973. [PubMed: 19597506]
12. Smith DM, Fraga H, Reis C, Kafri G, Goldberg AL. ATP binds to proteasomal ATPases in pairs with distinct functional effects, implying an ordered reaction cycle. *Cell.* 2011; 144:526–538. [PubMed: 21335235]
13. Palermo C, Joyce JA. Cysteine cathepsin proteases as pharmacological targets in cancer. *Trends Pharmacol Sci.* 2008; 29:22–28. [PubMed: 18037508]
14. Mason SD, Joyce JA. Proteolytic networks in cancer. *Trends Cell Biol.* 2011; 21:228–237. [PubMed: 21232958]
15. Lopez-Otin C, Hunter T. The regulatory crosstalk between kinases and proteases in cancer. *Nat Rev Cancer.* 2010; 10:278–292. [PubMed: 20300104]
16. Overall CM, Kleifeld O. Tumour microenvironment - opinion: validating matrix metalloproteinases as drug targets and anti-targets for cancer therapy. *Nat Rev Cancer.* 2006; 6:227–239. [PubMed: 16498445]
17. Omara-Opyene AL, et al. Genetic disruption of the *Plasmodium falciparum* digestive vacuole plasmepsins demonstrates their functional redundancy. *J Biol Chem.* 2004; 279:54088–54096. [PubMed: 15491999]
18. Deu E, et al. Functional studies of *Plasmodium falciparum* dipeptidyl aminopeptidase I using small molecule inhibitors and active site probes. *Chem Biol.* 2010; 17:808–819. [PubMed: 20797610]
19. Klemba M, Gluzman I, Goldberg DE. A *Plasmodium falciparum* dipeptidyl aminopeptidase I participates in vacuolar hemoglobin degradation. *J Biol Chem.* 2004; 279:43000–43007. [PubMed: 15304495]
20. Dalal S, Klemba M. Roles for two aminopeptidases in vacuolar hemoglobin catabolism in *Plasmodium falciparum*. *J Biol Chem.* 2007; 282:35978–35987. [PubMed: 17895246]
21. Sadaghiani AM, Verhelst SH, Bogyo M. Tagging and detection strategies for activity-based proteomics. *Curr Opin Chem Biol.* 2007; 11:20–28. [PubMed: 17174138]
22. Cravatt BF, Wright AT, Kozarich JW. Activity-based protein profiling: from enzyme chemistry to proteomic chemistry. *Annu Rev Biochem.* 2008; 77:383–414. [PubMed: 18366325]
23. Desmarais S, Masse F, Percival MD. Pharmacological inhibitors to identify roles of cathepsin K in cell-based studies: a comparison of available tools. *Biol Chem.* 2009; 390:941–948. [PubMed: 19453281]

24. Paulick MG, Bogyo M. Application of activity-based probes to the study of enzymes involved in cancer progression. *Curr Opin Genet Dev.* 2008; 18:97–106. [PubMed: 18294838]
25. Blum G. Use of fluorescent imaging to investigate pathological protease activity. *Curr Opin Drug Discov Devel.* 2008; 11:708–716.
26. Arastu-Kapur S, et al. Identification of proteases that regulate erythrocyte rupture by the malaria parasite *Plasmodium falciparum*. *Nat Chem Biol.* 2008; 4:203–213. [PubMed: 18246061]
27. Chang JW, Nomura DK, Cravatt BF. A Potent and Selective Inhibitor of KIAA1363/AADACL1 that Impairs Prostate Cancer Pathogenesis. *Chem Biol.* 2011; 18:476–484. [PubMed: 21513884]
28. Ahn K, et al. Discovery and characterization of a highly selective FAAH inhibitor that reduces inflammatory pain. *Chem Biol.* 2009; 16:411–420. [PubMed: 19389627]
29. Staub I, Sieber SA. Beta-lactam probes as selective chemical-proteomic tools for the identification and functional characterization of resistance associated enzymes in MRSA. *J Am Chem Soc.* 2009; 131:6271–6276. [PubMed: 19354235]
30. Adibekian A, et al. Click-generated triazole ureas as ultrapotent in vivo-active serine hydrolase inhibitors. *Nat Chem Biol.* 2011; 7:469–478. [PubMed: 21572424]
31. Meng L, Kwok BH, Sin N, Crews CM. Eponemycin exerts its antitumor effect through the inhibition of proteasome function. *Cancer Res.* 1999; 59:2798–2801. [PubMed: 10383134]
32. Sin N, et al. Total synthesis of the potent proteasome inhibitor epoxomicin: a useful tool for understanding proteasome biology. *Bioorg Med Chem Lett.* 1999; 9:2283–2288. [PubMed: 10465562]
33. Jain S, Diefenbach C, Zain J, O'Connor OA. Emerging role of carfilzomib in treatment of relapsed and refractory lymphoid neoplasms and multiple myeloma. *Core Evid.* 2011; 6:43–57. [PubMed: 21654882]
34. Hall CI, et al. Chemical genetic screen identifies *Toxoplasma* DJ-1 as a regulator of parasite secretion, attachment, and invasion. *Proc Natl Acad Sci U S A.* 2011; 108:10568–10573. [PubMed: 21670272]
35. Chandramohanadas R, et al. Apicomplexan parasites co-opt host calpains to facilitate their escape from infected cells. *Science.* 2009; 324:794–797. [PubMed: 19342550]
36. Gocheva V, et al. IL-4 induces cathepsin protease activity in tumor-associated macrophages to promote cancer growth and invasion. *Genes Dev.* 2010; 24:241–255. [PubMed: 20080943]
37. Berger AB, et al. Identification of early intermediates of caspase activation using selective inhibitors and activity-based probes. *Mol Cell.* 2006; 23:509–521. [PubMed: 16916639]
38. Methot N, et al. In vivo inhibition of serine protease processing requires a high fractional inhibition of cathepsin C. *Mol Pharmacol.* 2008; 73:1857–1865. [PubMed: 18326050]
39. Blum G, von Degenfeld G, Merchant MJ, Blau HM, Bogyo M. Noninvasive optical imaging of cysteine protease activity using fluorescently quenched activity-based probes. *Nat Chem Biol.* 2007; 3:668–677. [PubMed: 17828252]
40. Blum G, Weimer RM, Edgington LE, Adams W, Bogyo M. Comparative assessment of substrates and activity based probes as tools for non-invasive optical imaging of cysteine protease activity. *PLoS One.* 2009; 4:e6374. [PubMed: 19636372]
41. Hagel M, et al. Selective irreversible inhibition of a protease by targeting a noncatalytic cysteine. *Nat Chem Biol.* 2011; 7:22–24. [PubMed: 21113170]
42. Blair JA, et al. Structure-guided development of affinity probes for tyrosine kinases using chemical genetics. *Nat Chem Biol.* 2007; 3:229–238. [PubMed: 17334377]
43. Agard NJ, Maltby D, Wells JA. Inflammatory stimuli regulate caspase substrate profiles. *Mol Cell Proteomics.* 2010; 9:880–893. [PubMed: 20173201]
44. auf dem Keller U, et al. Novel matrix metalloproteinase inhibitor [18F]marimastat-aryltrifluoroborate as a probe for in vivo positron emission tomography imaging in cancer. *Cancer Res.* 2010; 70:7562–7569. [PubMed: 20729277]
45. Bowyer PW, Simon GM, Cravatt BF, Bogyo M. Global profiling of proteolysis during rupture of *P. falciparum* from the host erythrocyte. *Mol Cell Proteomics.* 2010
46. Enoksson M, et al. Identification of proteolytic cleavage sites by quantitative proteomics. *J Proteome Res.* 2007; 6:2850–2858. [PubMed: 17547438]

47. Kleifeld O, et al. Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products. *Nat Biotechnol.* 2010; 28:281–288. [PubMed: 20208520]
48. Mahrus S, et al. Global sequencing of proteolytic cleavage sites in apoptosis by specific labeling of protein N termini. *Cell.* 2008; 134:866–876. [PubMed: 18722006]
49. auf dem Keller U, Prudova A, Gioia M, Butler GS, Overall CM. A statistics-based platform for quantitative N-terminome analysis and identification of protease cleavage products. *Mol Cell Proteomics.* 2010; 9:912–927. [PubMed: 20305283]
50. Simon GM, Dix MM, Cravatt BF. Comparative assessment of large-scale proteomic studies of apoptotic proteolysis. *ACS Chem Biol.* 2009; 4:401–408. [PubMed: 19415908]
51. Wildes D, Wells JA. Sampling the N-terminal proteome of human blood. *Proc Natl Acad Sci U S A.* 2010; 107:4561–4566. [PubMed: 20173099]
52. Staes A, et al. Selecting protein N-terminal peptides by combined fractional diagonal chromatography. *Nat Protoc.* 2011; 6:1130–1141. [PubMed: 21799483]
53. Prudova A, auf dem Keller U, Butler GS, Overall CM. Multiplex N-terminome analysis of MMP-2 and MMP-9 substrate degradomes by iTRAQ-TAILS quantitative proteomics. *Mol Cell Proteomics.* 2010; 9:894–911. [PubMed: 20305284]
54. Schilling O, Barre O, Huesgen PF, Overall CM. Proteome-wide analysis of protein carboxy termini: C terminomics. *Nat Methods.* 2010; 7:508–511. [PubMed: 20526347]
55. Timmer JC, et al. Profiling constitutive proteolytic events in vivo. *Biochem J.* 2007; 407:41–48. [PubMed: 17650073]
56. Gray DC, Mahrus S, Wells JA. Activation of specific apoptotic caspases with an engineered small-molecule-activated protease. *Cell.* 2010; 142:637–646. [PubMed: 20723762]
57. Truebestein L, et al. Substrate-induced remodeling of the active site regulates human HTRA1 activity. *Nat Struct Mol Biol.* 2011; 18:386–388. [PubMed: 21297635]
58. Vaidya S, Velazquez-Delgado EM, Abbruzzese G, Hardy JA. Substrate-induced conformational changes occur in all cleaved forms of caspase-6. *J Mol Biol.* 2011; 406:75–91. [PubMed: 21111746]
59. Hardy JA, Wells JA. Dissecting an allosteric switch in caspase-7 using chemical and mutational probes. *J Biol Chem.* 2009; 284:26063–26069. [PubMed: 19581639]
60. Scheer JM, Romanowski MJ, Wells JA. A common allosteric site and mechanism in caspases. *Proc Natl Acad Sci U S A.* 2006; 103:7595–7600. [PubMed: 16682620]
61. Schneck JL, et al. Chemical mechanism of a cysteine protease, cathepsin C, as revealed by integration of both steady-state and pre-steady-state solvent kinetic isotope effects. *Biochemistry.* 2008; 47:8697–8710. [PubMed: 18656960]
62. Patricelli MP, et al. In situ kinase profiling reveals functionally relevant properties of native kinases. *Chem Biol.* 2011; 18:699–710. [PubMed: 21700206]
63. Puri AW, et al. Rational design of inhibitors and activity-based probes targeting *Clostridium difficile* virulence factor TcdB. *Chem Biol.* 2010; 17:1201–1211. [PubMed: 21095570]
64. Sheahan KL, Cordero CL, Satchell KJ. Autoprocessing of the *Vibrio cholerae* RTX toxin by the cysteine protease domain. *EMBO J.* 2007; 26:2552–2561. [PubMed: 17464284]
65. Shen A, et al. Mechanistic and structural insights into the proteolytic activation of *Vibrio cholerae* MARTX toxin. *Nat Chem Biol.* 2009; 5:469–478. [PubMed: 19465933]
66. Falgout JP, et al. Lysosomotropism of basic cathepsin K inhibitors contributes to increased cellular potencies against off-target cathepsins and reduced functional selectivity. *J Med Chem.* 2005; 48:7535–7543. [PubMed: 16302795]
67. Deu E, Yang Z, Wang F, Klemba M, Bogoy M. Use of activity-based probes to develop high throughput screening assays that can be performed in complex cell extracts. *PLoS One.* 2010; 5:e11985. [PubMed: 20700487]
68. Thong B, Pilling J, Ainscow E, Beri R, Unitt J. Development and Validation of a SimpleCell-Based Fluorescence Assay for Dipeptidyl Peptidase 1 (DPP1) Activity. *J Biomol Screen.* 2010
69. Wakata A, et al. Simultaneous fluorescent monitoring of proteasomal subunit catalysis. *J Am Chem Soc.* 2010; 132:1578–1582. [PubMed: 20078037]

70. Poreba M, Drag M. Current strategies for probing substrate specificity of proteases. *Curr Med Chem*. 2010; 17:3968–3995. [PubMed: 20939826]
71. Whitney M, et al. Parallel in vivo and in vitro selection using phage display identifies protease-dependent tumor-targeting peptides. *J Biol Chem*. 2010; 285:22532–22541. [PubMed: 20460372]
72. Greenbaum DC, et al. A role for the protease falcipain 1 in host cell invasion by the human malaria parasite. *Science*. 2002; 298:2002–2006. [PubMed: 12471262]
73. Sadaghiani AM, et al. Design, synthesis, and evaluation of in vivo potency and selectivity of epoxysuccinyl-based inhibitors of papain-family cysteine proteases. *Chem Biol*. 2007; 14:499–511. [PubMed: 17524981]
74. Albrow VE, et al. Development of Small Molecule Inhibitors and Probes of Human SUMO Deconjugating Proteases. *Chem Biol*. 2011; 18:722–732. [PubMed: 21700208]
75. Ponder EL, et al. Functional Characterization of a SUMO Deconjugating Protease of *Plasmodium falciparum* Using Newly Identified Small Molecule Inhibitors. *Chem Biol*. 2011; 18:711–721. [PubMed: 21700207]
76. Patterson AW, Wood WJ, Ellman JA. Substrate activity screening (SAS): a general procedure for the preparation and screening of a fragment-based non-peptidic protease substrate library for inhibitor discovery. *Nat Protoc*. 2007; 2:424–433. [PubMed: 17406604]
77. Patterson AW, et al. Identification of selective, nonpeptidic nitrile inhibitors of cathepsin s using the substrate activity screening method. *J Med Chem*. 2006; 49:6298–6307. [PubMed: 17034136]
78. Brak K, Doyle PS, McKerrow JH, Ellman JA. Identification of a new class of nonpeptidic inhibitors of cruzain. *J Am Chem Soc*. 2008; 130:6404–6410. [PubMed: 18435536]
79. Brak K, et al. Nonpeptidic tetrafluorophenoxymethyl ketone cruzain inhibitors as promising new leads for Chagas disease chemotherapy. *J Med Chem*. 2010; 53:1763–1773. [PubMed: 20088534]
80. Bachovchin DA, et al. Superfamily-wide portrait of serine hydrolase inhibition achieved by library-versus-library screening. *Proc Natl Acad Sci U S A*. 2010; 107:20941–20946. [PubMed: 21084632]
81. Ahn K, et al. Novel mechanistic class of fatty acid amide hydrolase inhibitors with remarkable selectivity. *Biochemistry*. 2007; 46:13019–13030. [PubMed: 17949010]
82. Speers AE, Cravatt BF. Profiling enzyme activities in vivo using click chemistry methods. *Chem Biol*. 2004; 11:535–546. [PubMed: 15123248]
83. Berkers CR, et al. Activity probe for in vivo profiling of the specificity of proteasome inhibitor bortezomib. *Nat Methods*. 2005; 2:357–362. [PubMed: 15846363]
84. Arastu-Kapur S, et al. Nonproteasomal Targets of the Proteasome Inhibitors Bortezomib and Carfilzomib: a Link to Clinical Adverse Events. *Clin Cancer Res*. 2011
85. Desmarais S, et al. Effect of cathepsin k inhibitor basicity on in vivo off-target activities. *Mol Pharmacol*. 2008; 73:147–156. [PubMed: 17940194]
86. Gauthier JY, et al. The discovery of odanacatib (MK-0822), a selective inhibitor of cathepsin K. *Bioorg Med Chem Lett*. 2008; 18:923–928. [PubMed: 18226527]
87. Geurink P, et al. A peptide hydroxamate library for enrichment of metalloproteinases: towards an affinity-based metalloproteinase profiling protocol. *Org Biomol Chem*. 2008; 6:1244–1250. [PubMed: 18362965]
88. Saghatelian A, Jessani N, Joseph A, Humphrey M, Cravatt BF. Activity-based probes for the proteomic profiling of metalloproteases. *Proc Natl Acad Sci U S A*. 2004; 101:10000–10005. [PubMed: 15220480]
89. Harbut MB, et al. Bestatin-based chemical biology strategy reveals distinct roles for malaria M1- and M17-family aminopeptidases. *Proc Natl Acad Sci U S A*. 2011; 108:E526–534. [PubMed: 21844374]
90. Sieber SA, Niessen S, Hoover HS, Cravatt BF. Proteomic profiling of metalloprotease activities with cocktails of active-site probes. *Nat Chem Biol*. 2006; 2:274–281. [PubMed: 16565715]
91. Nakai R, Salisbury CM, Rosen H, Cravatt BF. Ranking the selectivity of PubChem screening hits by activity-based protein profiling: MMP13 as a case study. *Bioorg Med Chem*. 2009; 17:1101–1108. [PubMed: 18364257]

92. Li YM, et al. Photoactivated gamma-secretase inhibitors directed to the active site covalently label presenilin 1. *Nature*. 2000; 405:689–694. [PubMed: 10864326]
93. Weerapana E, et al. Quantitative reactivity profiling predicts functional cysteines in proteomes. *Nature*. 2010; 468:790–795. [PubMed: 21085121]

Figure 1. Mechanism of substrate hydrolysis by the primary families of proteases

A) Protease substrates bind through interactions of the side chain residues (P, P' residues) with the substrate pockets of the protease (S, S' pockets). The red dashed line indicates scissile bond. B-D) The architecture of the active site and mechanism of the hydrolysis for the three classes of proteases employing a covalent catalysis mechanism, B) N-terminal threonine, C) serine and D) cysteine protease is depicted. E-F) A non-covalent hydrolysis mechanism is employed by E) zinc metalloproteases, F) aspartate proteases and glutamate proteases (not depicted).

Figure 2. Schematic presentation of the hit-to-lead process

A) In a classical protease drug discovery approach the emphasis of the screen and optimization lays on the maximizing the potency of a hit compound for a recombinant protease. Off-target effects and efficacy are usually tested after the optimization process, and problems encountered when testing the compounds in cultures and *in vivo* implies either modifying the structure of the lead inhibitor to solve a particular issue or select a different chemotype for further optimization. B) In this perspective we like to propose a more holistic approach in which the emphasis lays on identifying hits in a more complex and relevant context (intact cells) and incorporating the specificity profile of hits to select hits and optimize lead compound. By putting the emphasis of the hit to lead optimization process in selectivity instead of just potency we believe would help us prevent off-target effects and thus increase the chances of developing protease inhibitor drugs with minimum side effects.

Figure 3. Protease activity is highly regulated. Activity-based probes report on protease activity
 A) Besides regulation on the transcription and translation level, proteases are highly regulated on the protein level. Being expressed as zymogens, proteases are activated in a variety of ways depending on the protease: allosteric activation, environmental changes, localization, protein-protein interactions, processing by upstream proteases, etc. Endogenous inhibitors and targeted degradation form yet another layer of regulation. B) Activity-based probes (ABPs) are small molecule reporter molecules that turn the chemistry of the active protease against itself to distinguish it from its zymogen or inhibited form. Most ABPs consist of three parts; a warhead, an electrophilic moiety that reacts with the active site nucleophile to result in a covalent and irreversible adduct, a spacer and/or recognition element that targets the probe to a specific target protease, and a tag, usually a fluorescent dye and/or an affinity handle, like biotin.

Figure 4. Chemical toolbox to study protease function and target inhibition

A) Forward Chemical Genetics allows for target identification through the introduction of an affinity tag to the hit compound. B) Near-Infrared fluorescently labeled ABPs can be applied to top-down characterization of a target protease. Whole animal non-invasive imaging techniques allow for visualization of target distribution. Extracted tissue can be analyzed *ex vivo*, histology can be performed to show target distribution on a microscopic level. FACS analysis can report on which types of cells contain active protease, whereas biochemical analysis of the fluorescently labeled proteases allows for characterization at the protein level. Treatment with a lead compound prior to labeling informs on target inhibition. C) The use of broad spectrum protease probes enables a readout for the inhibition profile of a lead compound for an entire protease family. In short, first a proteome is exposed to a compound. Inhibition of a target protease prevents probe binding in the second step allowing for target identification. D) Global profiling of all reactive cysteines in a proteome; at low concentrations the iodoacetamide based reporter molecule will react primarily with the more reactive cysteines. At higher concentrations, less reactive cysteines will also be modified. Using isotopically labeled reporter molecules, this method can be used to predict functional cysteines in proteomes as well as target identification. Most of the profiling methods described here use probes that covalently modify the active site of the targeted proteases. When used to evaluate the specificity profile of reversible inhibitors it is important to adjust the labeling conditions such that the probe will not out-compete the inhibitor. Because most

of these methods have a good dynamic range, this can be accomplished by lowering probe concentration or decreasing labeling times.