REVIEW

nature structural & molecular biology

New approaches for dissecting protease functions to improve probe development and drug discovery

Edgar Deu[1], Martijn Verdoes[1] & Matthew Bogyo[1,2]

Proteases are well-established targets for pharmaceutical development because of their known enzymatic mechanism and their regulatory roles in many pathologies. However, many potent clinical lead compounds have been unsuccessful either because of a lack of specificity or because of our limited understanding of the biological roles of the targeted protease. In order to successfully develop protease inhibitors as drugs, it is necessary to understand protease functions and to expand the platform of inhibitor development beyond active site–directed design and *in vitro* optimization. Several newly developed technologies will enhance assessment of drug selectivity in living cells and animal models, allowing researchers to focus on compounds with high specificity and minimal side effects *in vivo*. In this review, we highlight advances in the development of chemical probes, proteomic methods and screening tools that we feel will help facilitate this paradigm shift in drug discovery.

Proteases are one of the most abundant classes of enzymes and are involved in a wide range of biological processes, including cell-cycle progression, cell signaling, proliferation and death, protein trafficking and immune response. Proteases are also involved in many human diseases, ranging from degenerative and inflammatory diseases to infectious diseases. Therefore, proteases are often targeted by the pharmaceutical industry, and protease inhibitor drugs are currently in use for treatment of coagulation disorders, hypertension, HIV infection, cancer and diabetes[1].

Proteases bind their substrates through hydrogen bond interactions with the substrate peptide backbone and by hydrophobic and electrostatic contacts between the substrate side chains and well-defined pockets within the active site (**Fig. 1a**). There are seven distinct classes of proteases (aspartate, cysteine, glutamate, metalloproteases, serine, threonine and the newly identified asparagine peptide lyases[2]), grouped according to the amino acid or ion that catalyzes peptide bond cleavage (**Fig. 1b–f**), and the mechanism of substrate cleavage determines which type of chemical entity can be used to inhibit each protease family[3]. For cysteine, serine and threonine proteases, an electrophilic group can covalently modify the catalytic residue in a reversible or irreversible manner. In the case of metalloproteases, functional groups that coordinate the catalytic metal can achieve potent inhibition. For all protease families, potent transition state analogs can be designed that are based on structural and enzymatic studies.

Although recent 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], developing protease inhibitors as drugs has remained a challenge, and achieving target specificity has been the main hurdle to overcome. This is because the reaction mechanism is highly conserved among each protease class and because proteases often have many closely related family members, resulting in lead compounds that often inhibit more than one target, potentially causing unwanted side effects. Unfortunately, the traditional approach of hit-to-lead optimization takes into account *in vivo* off-target effects fairly late in the process (**Fig. 2a**). Although advances in structural biology have revolutionized inhibitor design, it is necessary to broaden the platform of inhibitor development to allow for 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 (**Fig. 2b**).

There have been a number of success stories in the development of drugs that target proteases[1,5]; however, here we highlight what we believe to be the two main challenges in the development of protease inhibitors as drugs: (i) validating proteases as drug targets, including obtaining a clear understanding of their biological functions, and (ii) designing specific inhibitors despite the large number of proteases found in a given organism. We also present recent technological advances that we believe will help address these challenges, emphasizing 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*. Finally, we describe recent advances in screening technologies for intact cells and living animals, as well as recently developed proteomic approaches for the identification of protease substrates.

Understanding the biological function of proteases

Dissecting how proteases carry out their biological functions is extremely challenging, as the cell regulates this class of enzymes through several mechanisms. 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 (**Fig. 3a**). Proteases can also 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 site[7]. On a cellular level, protease activity and specificity can

[1]Department of Pathology, Stanford University School of Medicine, Stanford, California, USA. [2]Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, California, USA. Correspondence should be addressed to M.B. (mbogyo@stanford.edu).

Published online 5 January 2012; doi:10.1038/nsmb.2203

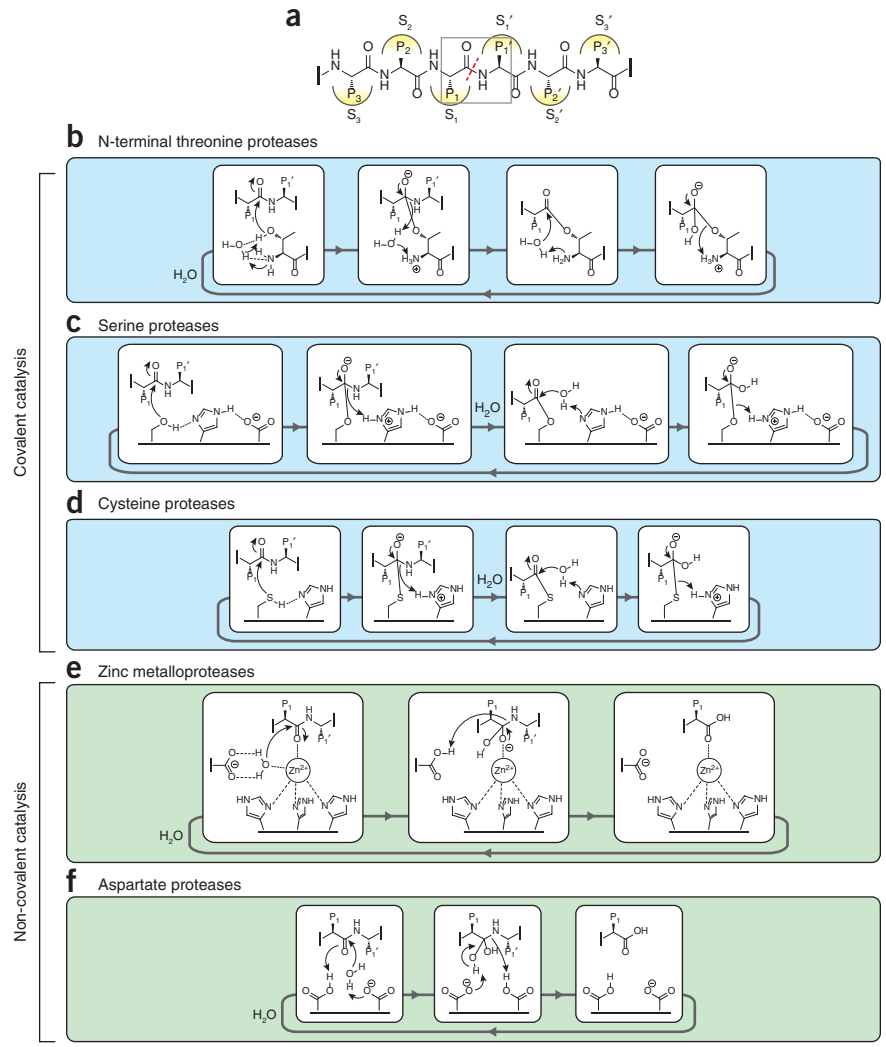

Figure 1 Mechanism of substrate hydrolysis by the primary families of proteases. (**a**) Protease substrates bind through interactions of the side chain residues (P and P′ residues) with the substrate pockets of the protease (S and S′ pockets). The red dashed line indicates a scissile bond. (**b**–**d**) The architecture of the active site and mechanism of hydrolysis for N-terminal threonine, serine and cysteine proteases that use an acyl-enzyme intermediate formed through nucelophilic attack by the catalytic side chain residue. (**e**,**f**) In the case of zinc metalloproteases (**e**), aspartate proteases (**f**) and glutamate proteases (not depicted), a carboxylic acid group or metal ion activates a water molecule, leading to acid-base catalysis. The seventh and newest protease family, the asparagine peptide lyases, cleave themselves using an asparagine residue as the nucleophile[2] (not depicted).

be regulated by multiple post-translational mechanisms. As a result, mRNA levels and protein abundance and localization often fail to reveal when and where a protease is active, so other approaches are needed to evaluate protease functions.

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 lysosome[8], and other proteases, such as the cysteine protease domain (CPD) of the multifunctional autoprocessing repeats toxin (MARTX) found in *Vibrio cholera*[9], which autoprocesses in response to inositol hexakisphosphate (Ins_6P) binding, have evolved allosteric mechanisms of activation. Alternatively, protease activation can be regulated through the formation of large protein complexes (for example, through activation of caspase-1 upon formation of the inflamasome complex[10], or through formation of receptor-triggered complexes, as exemplified by initiator caspase activation[11]) or by self-compartmentalization of their proteolytic subunits such as with the proteasome.

Proteases have multiple biological functions, often interacting with major signaling pathways[12,13]. These functions can be controlled by changes in localization of the protease or substrate in response to different stimuli. For example, substrates are targeted for degradation by the proteasome through polyubiquitination, and their access to the proteolytic cavity is regulated in an ATP-dependent manner[14]. Cathepsins are a particularly complex example of this; in the lysosome, degradation of the substrates of cathepsins is controlled by trafficking substrates into this organelle. By contrast, they have also been shown to regulate innate and adaptive immunity and tumor progression, depending on their localization[8]. Release of cathepsin B into the cytosol activates caspase 8 and triggers apoptosis[8]. Export of cathepsins to the tumor microenvironment results in extracellular matrix (ECM) degradation[8,15]. 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 inflammation[8].

This complexity of regulation and effects makes successful targeting of proteases extremely challenging, as was poignantly illustrated by the clinical inhibition of MMPs in cancer. Targeted for their involvement in ECM degradation during angiogenesis and metastasis, MMPs are also involved in maintaining homeostasis of the extracellular environment, as well as in regulating cell signaling and innate immunity. Inhibition of some MMPs has actually been shown to promote tumor growth. This duality of function, coupled with the difficulty in designing specific inhibitors, has caused many MMP inhibitors in clinical trials to fail[16]. In complex multifactorial diseases such as cancer, it is essential to know which specific protease is important in each disease subpopulation. This is particularly evident with cysteine cathepsins; although increased levels of cathepsins B and L activities correlate with tumor malignancy, an increase in cathepsins H and S activities can result in either a better or a worse prognosis, depending on the type of cancer[15].

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. 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 out[17]. However, proteases involved at the later stages of degradation, such as dipeptidyl aminopeptidase 1 (DPAP1)[18,19] and several other aminopeptidases[20],

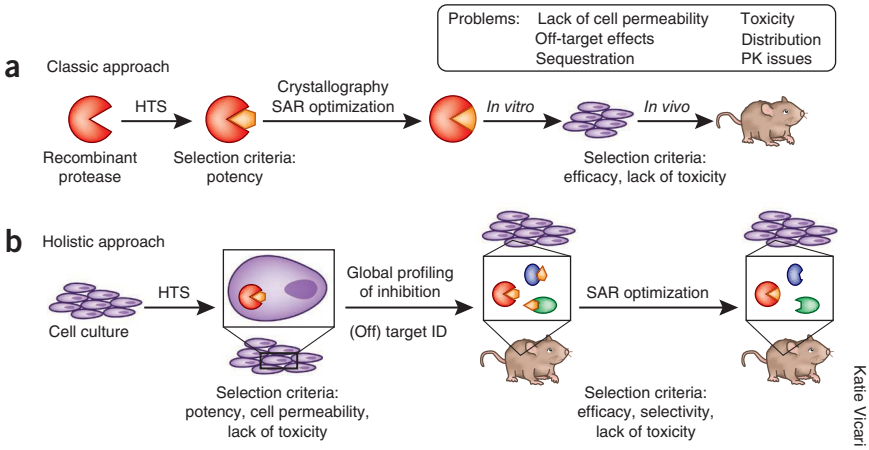

Figure 2 Schematic presentation of the hit-to-lead process. (**a**) In a classical protease drug discovery approach, the emphasis of screening and optimization is on maximizing the potency of a hit compound for a recombinant protease. Off-target effects and efficacy are usually tested after the optimization process, and problems encountered when testing the compounds in cultures and *in vivo* require either modifying the structure of the lead inhibitor to solve a particular issue or selecting a different chemotype for further optimization. (**b**) In this review, we propose a holistic approach, in which the emphasis is on identifying hits in a more complex and relevant context (intact cells), incorporating the specificity profile of hits to identify and optimize lead compounds. We believe that placing the emphasis of the hit-to-lead optimization process on selectivity instead of just on potency will help prevent off-target effects and thus increase the chances for developing protease inhibitor drugs with minimum side effects.

are essential. This underscores that it is imperative to identify those proteases that are required for a pathway to function properly.

New tools to study protease function

Given the need to further understand the complex roles of proteases, several techniques have been developed and enhanced to identify protease targets, locations and modes of regulation. These include ABPs as well as proteomics-based approaches.

Activity-based probes. Activity-based probes are small-molecule reporters designed to be highly selective for the catalytically active form of a protease or protease family (**Fig. 3b**), allowing assessment of protease activity within living cells or in whole organisms[21–25]. Most ABPs consist of three parts: a warhead, a spacer and/or recognition element and a tag (**Fig. 3b**). They are also inhibitors of the target protease, allowing them to be used with better temporal control than conditional gene disruption or RNA-mediated interference (RNAi) methods. For these reasons, ABPs have begun to find applications in early stage drug discovery[18,26–32]. For example, biotin versions of eponemycin and epoxomycin identified the proteasome as the target of these natural antitumor products[18,26–32]. This emphasis on early stage drug discovery has led to the clinical development of epoxyketone inhibitors such as carfilzomib as proteasome-specific drugs[33]. As most ABPs irreversibly modify the enzyme active site, libraries of covalent inhibitors can be used in forward chemical genetics screens to identify compounds that induce a specific phenotype. Hits are then converted into ABPs to isolate targeted enzymes for identification by liquid chromatography–tandem mass spectrometry (LC-MS/MS) (**Fig. 4a**). This approach has successfully identified proteases involved in malaria parasite *Plasmodium falciparum* egress from infected red blood cells (RBCs)[28], and it has also identified proteins involved in *Toxoplasma gondii* invasion[34]. ABPs are especially useful in biological systems that are not amenable to genetic manipulation or regulation by RNAi. For example, an ABP has been used to demonstrate that *P. falciparum* activates the RBC calpain during egress[35]. There are other examples where ABPs were found to be essential for the identification and localization of the active forms of the proteases. One is a study that identified infiltrating macrophages as the cells that express and activate cysteine cathepsins at the tumor invasive front in response to IL-4 secreted by cancer cells[36]. Another study followed the kinetics of caspases activation during apoptosis[37].

A further use of ABPs *in vivo* has been to determine the extent of target inhibition needed to observe beneficial downstream effects. ABPs were used to show that complete and prolonged inhibition of cathepsin C is required to decrease the level of activity of downstream serine proteases involved in inflammation[38]. Similarly, because DPAP1 is highly expressed in *P. falciparum*, ABPs revealed that its activity needs to be inhibited for at least 3 h in order to kill the parasite[18].

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, in a single experiment, not only activation but also localization, distribution and biochemical activity of proteases in cells and *in vivo* (**Fig. 4b**). This has been shown in multiple mouse models of cancer, using cysteine cathepsin probes to noninvasively detect cathepsin activity in tumors[11,39,40]. Additionally, ABPs can be used to monitor the efficacy of drug treatment *in vivo*[39] 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, but a few success stories have emerged. In a recent study, an ABP that selectively inhibited and labeled the hepatitis C virus NS2/3 protease was designed by taking advantage of a cysteine residue that is relatively close to the active site[41]. An electrophilic group that reacts with the cysteine residue was added to a reversible inhibitor of NS2/3, creating a covalent linkage between the inhibitor and the enzyme, which improved potency and selectivity. This approach may also be broadly applicable to the study of proteases, using engineered point mutations that render the target protease sensitive to ABP binding, similarly to recent methods for selective targeting and inhibition of kinases[42].

Proteomics-based substrate identification. Although linking proteases to their substrates is essential to understand their functions, it is notoriously difficult to identify protease substrates because the end products of the reaction must be identified among the pool of cellular proteins. Furthermore, the resultant peptide fragments provide little information regarding the protease that produced them. A number of proteomic approaches have recently been developed that allow specific enrichment of the newly generated peptide fragments resulting from proteolysis, within a biologically relevant context (**Supplementary Box 1**)[43–55].These methods have been used to map proteolytic events during processes such as apoptosis[48,50], inflammation[43] or rupture of malaria-infected RBCs[45].

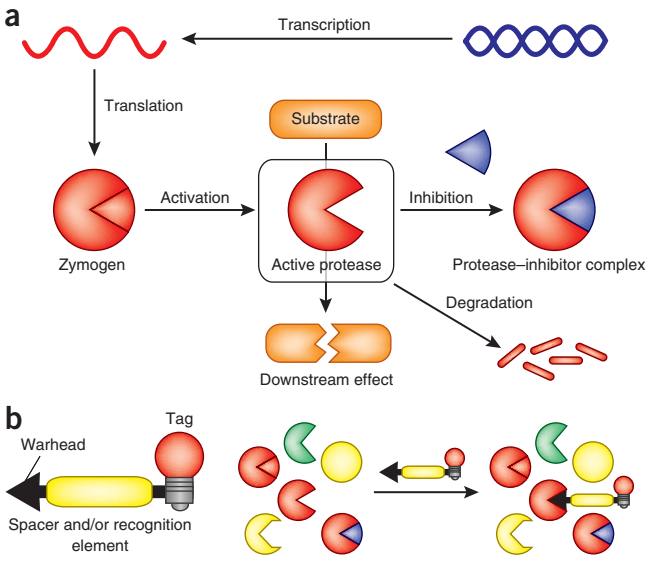

Figure 3 Activity-based probes report on tightly regulated protease activity. (**a**) Proteases are not only regulated on the transcription and translation levels but also highly regulated on the protein level. Expressed as zymogens, proteases are activated in a variety of ways and by a variety of factors, depending on the protease, including allosteric activation, environmental changes, localization, protein-protein interactions and processing by upstream proteases. Endogenous inhibitors and targeted degradation form yet another layer of regulation. (**b**) Activity-based probes are small-molecule reporters that use the active protease's own chemistry 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).

In addition, new approaches to selectively activate a protease-mediated biological process have been devised. In one such example, the proenzyme cleavage site of several apoptotic caspases was converted into a tobacco etch virus (TEV) cleavage sequence, allowing them to be activated by a small molecule–regulated TEV protease. This conferred temporal control of caspase activity in the cell, allowing identification of distinct caspase substrates during apoptosis[56]. As many proteases require proteolytic cleavage to be activated, this method could be broadly applied. Combining the use of ABPs with proteomic approaches to identify natural substrates should help determine the biological function of proteases.

Challenges for *in vitro* development 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 subfamily generally have similar substrate and inhibitor specificities (**Fig. 1**). Consequently, structural biologists have focused on finding alternate ways to inhibit proteases, such as blocking allosteric sites and exosites[1]. Recent studies suggest that some proteases cleave their substrates through an induced-fit mechanism[57,58], which could be targeted by inhibitors that stabilize the inactive conformation. Alternatively, as seen with caspases, molecules can be designed to lock an enzyme in its zymogen conformation[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 process[61].

Another obstacle to designing small-molecule inhibitors is that the conditions in which inhibitors are tested *in vitro*, using recombinant or purified enzymes, are different from the enzyme's environment *in vivo*. This discrepancy has been demonstrated for kinases, where the effect of inhibitors in cell-based assays correlated better with IC_{50} values determined in lysates than with IC_{50} values determined with recombinant enzymes[62]. There are multiple reasons why *in vitro* testing with purified targets may provide only limited information on leads. First, specific binding of the inhibitor to off-target proteases, or other nonspecific associations, might reduce the effective concentration of an inhibitor in the cell. Second, pH, salt concentration, viscosity and protein concentrations can affect enzyme activity, specificity and stability, so assay conditions *in vitro* might not reflect those *in vivo*. 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 the MARTX toxin was initially reported to be allosterically activated by GTP[63], but recent work has shown that $InsP_6$ is the biologically relevant activator[9,64], and the CPD domain of *Clostridium difficile* shows some activity even in the absence of $InsP_6$ (ref. 65). Therefore, prior to the identification of the activator, IC_{50} values of CPD inhibitors would have been markedly underestimated. Finally, *in vitro* systems fail to take into account such issues as delivery, partitioning, cell permeability and cell stability. For example, probes and inhibitors that are highly specific for a cytosolic target *in vitro* often show cross-reactivity with lysosomal proteases when tested in cells, owing to accumulation as a result of endocytosis.[11,66].

New approaches to define inhibitor specificity *in vivo*

Given the challenges associated with developing potent and specific inhibitors using *in vitro* techniques, an alternative may be to focus on cellular and *in vivo* approaches that bypass some of the difficulties associated with *in vitro* screening while still leading to potent inhibitors.

Screening protease inhibitors in a biological context. Carrying out 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. Although 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 $(Pro\text{-}Arg)_2$Rho fluorescent substrate as selective for DPAP1 in *P. falciparum* lysates and for cathepsin C in rat liver extract[18,67]. This substrate can also be used for HTS, avoiding the need to purify or express these proteases. Using cell-permeable ABPs, correlating target labeling with inhibition of substrate turnover can also identify selective reporter substrates in intact cells, as seen with the cell-permeable substrate Gly-Phe-AFC, which was not cleaved in cell lines or in bone marrow lysates lacking cathepsin C[68].

Carrying out assays in lysates or in intact cells also allows for concurrent measurement of multiple protease activities. One example is the recent development of two specific, differently labeled fluorogenic substrates that can report chymotrypsin-like and caspase-like proteasomal activities simultaneously[69]. Because of the recent advances in methods to map substrate specificity of proteases[70], we believe that substrates designed to be used will become increasingly widespread.

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 peptide library was repeatedly injected, amplified and reinjected into mice with mammary tumors, allowing the

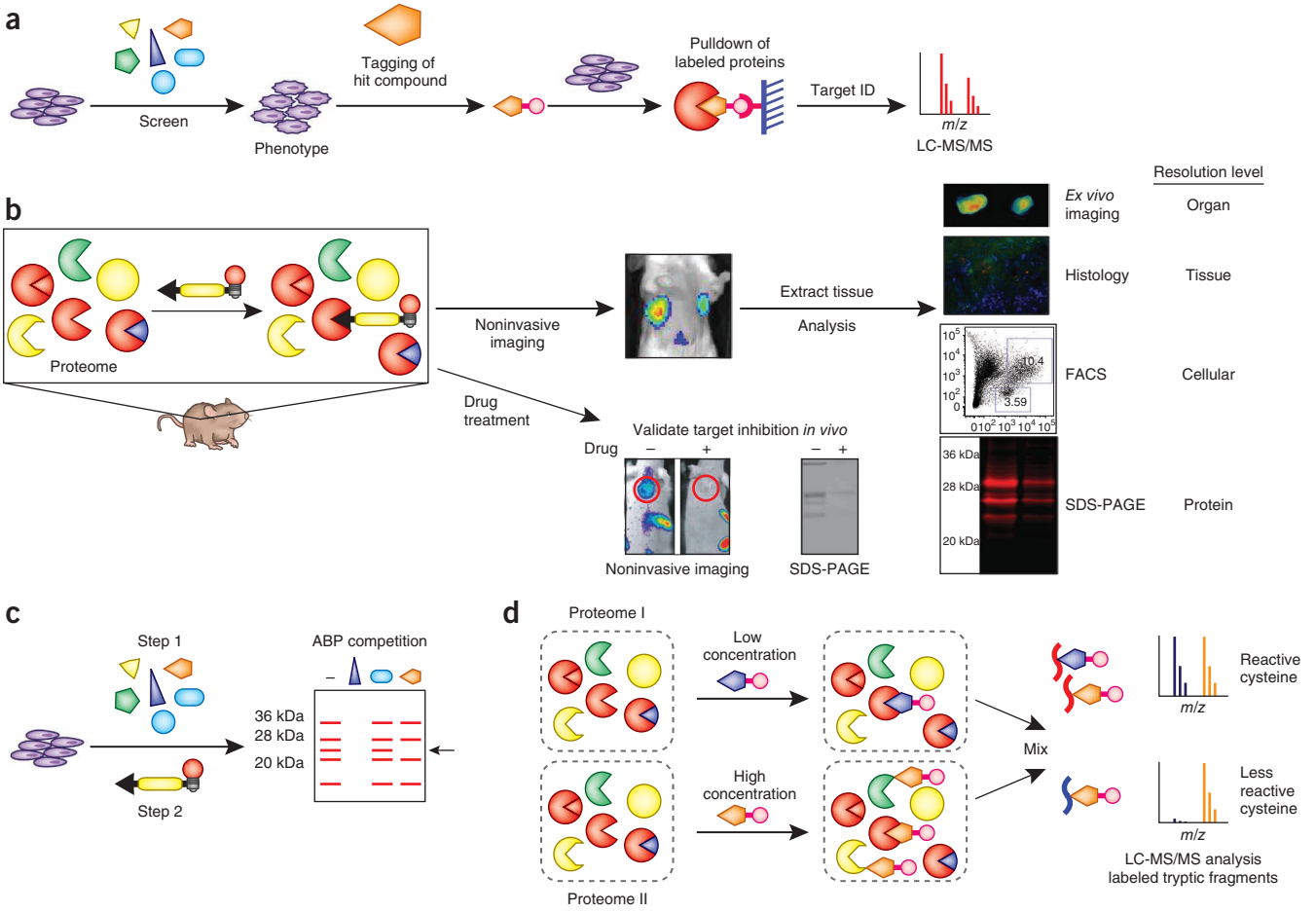

Figure 4 Chemical tools to study protease function and to measure target inhibition. (**a**) Forward chemical genetics allows for target identification through the introduction of an affinity tag to the hit compound. (**b**) Near-infrared fluorescently labeled ABPs can be applied to top-down characterization of a target protease. Whole-animal noninvasive imaging techniques allow the visualization of target distribution, and extracted tissue can be analyzed *ex vivo*. Histology shows target distribution on a microscopic level, FACS analysis identifies the types of cells that contain active protease and biochemical analysis allows characterization at the protein level. Treatment with a lead compound before labeling provides information on target inhibition. Mouse images are from our previous publication[39]. (**c**) Broad-spectrum protease probes enable a readout of the inhibition profile of a lead compound for an entire protease family in a proteome. Members of the targeted family can be resolved on a gel, and inhibition results in diminished labeling of individual proteases. (**d**) Global profiling of all reactive cysteines in a proteome; iodoacetamide-based reporter molecules will react primarily with the more reactive cysteines. Using isotopically labeled reporter molecules, this method can be used to predict functional cysteines in proteomes as well as to identify targets. When the methods described here are used to evaluate the specificity profile of a reversible inhibitor, the labeling conditions should be adjusted so that the covalent probe does not outcompete the inhibitor. Because these methods have a good dynamic range, this can be accomplished by lowering the probe concentration or decreasing the labeling times.

identification of sequences that were used to make fluorogenic substrates for noninvasive imaging of cancer[71]. 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 new inhibitors *in vivo*.

Although cell-based screening methods, when coupled with effective tools for monitoring target inhibition, have the potential to be a useful way to identify lead compounds, it should be noted that compounds that alter expression levels of protease targets could also appear to be active site–directed inhibitors. However, such occurrences might provide valuable information, and these compounds could be separated from the true active site inhibitors by using counter screening assays or by using western blots to assess protease levels.

Design of nonpeptidic inhibitors. One approach used to generate specific protease inhibitors and substrates is to move away from the natural peptide scaffold by incorporating non-natural amino acids. This has been extensively done with cysteine proteases[18,28,72–75]. Although 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 nonpeptidic inhibitors and transition-state analogs[1,5]. Nonpeptidic scaffolds for protease inhibitors are usually identified through screening methods such as HTS of small drug-like molecules[1,5], *in silico* screening or fragment-based screening[4]. Once such a scaffold has been successfully identified, a combination of structural, computational and structure-activity relationship (SAR) methods can be used to optimize initial hits into potent drug leads[1,4,5].

Alternatively, substrate-activity screening (SAS)[76,77] can be used to compare diverse, nonpeptidic substrates containing a coumarin fluorophore, against a protease of interest. The advantage of this method is that scaffolds with poor binding affinity will not be overlooked, as turnover of these substrates results in accumulation of a fluorescent signal over time. Hits are then converted into potent

inhibitors by replacing the coumarin group with an appropriate pharmacophore. This approach has been successfully used to identify new nonpeptidic inhibitor scaffolds with improved stability in mouse serum and better efficacy *in vivo*[18,78,79].

Global profiling of compound specificity. Although selective ABPs are very useful tools to study the biology of a specific target, broad-spectrum ABPs are valuable for globally profiling the effect of an inhibitor against all members of a protease family (**Fig. 4c**,**d**)[80]. This type of assay can provide information regarding the partitioning of the drug within specific tissues, as it can be done *in vivo* by pretreating an entire animal with an inhibitor and then labeling residual activity with an ABP directly *in vivo* or in tissue extracts[30,81,82]. In one application of this method, a cell-permeable broad-spectrum proteasome probe showed that the clinical proteasome inhibitor bortezomib (Velcade)—currently used for the treatment of multiple myeloma—targets unexpected proteasomal subunits[83]. In a second application of this method, another broad-spectrum ABP identified several serine proteases as additional off-target proteases[84]. In an ABP-based structure-activity relationship study, cathepsin K inhibitors that failed in phase II clinical development for the treatment of osteoporosis[85] were shown to accumulate in lysosomes *in vivo*, inhibiting off-target cathepsins and leading to an accumulation of intracellular collagen[23,86]. This lysosomotropism is presumably due to the presence of a basic amine group that becomes charged in the acidic lysosomes. A new inhibitor, odanacatib, that lacks a basic amine did not have this effect, allowing it to enter phase III clinical trials. Thus, these kinds of assays can potentially explain some of the side effects of such drugs and inform future inhibitor development.

All the ABPs mentioned above contain an electrophilic reactive group that irreversibly modifies the cysteine, serine or threonine catalytic residues of target proteases. However, this strategy is not applicable to aspartate and metal proteases because hydrolysis of the peptide bond is mediated by a water molecule. Nonetheless, metalloprotease ABPs that can be used to pull down proteases for MS identification and quantification[87] or for *in vivo* imaging applications[44] have been designed by incorporating a moiety that coordinates the metal ion inside the active site. Another approach is to introduce a photo–cross-linking group in addition to the chelating moiety to achieve covalent modification of metalloproteases[88–90]. Specific probes using these design principles have been used to study the biological function of aminopeptidases in *P. falciparum*[89], and broad-spectrum probes have been used to select HTS hits based on their specificity profile[91], or to identify active metalloproteases in tissue extracts[88]. The photo–cross-linking approach has also been used to target the presenilin aspartate proteases involved in the processing of amyloid proteins[92].

Although ABPs are 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 globally profiling the reactivity of all cysteines in a proteome by using an isotopically labeled probe containing iodoacetamide as a general alkylating agent[93] (**Fig. 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. This method should also, in principle, be applicable to studying serine protease by using a general serine reactive reporter.

In summary, by carrying out 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 compounds can then be quickly evaluated using an activity-based global profiling method, which also provides the SAR information necessary to design more specific inhibitors. The validity of this approach was recently demonstrated with the identification of triazole ureas as potent serine hydrolase inhibitors that were then developed into highly selective inhibitors *in vivo*[32].

Concluding remarks

In summary, we feel that the main challenges facing protease drug development is a lack of understanding of the complex mechanisms by which protease activities are regulated and of the multiple functions they have in diverse biological pathways. In addition, one of the main hurdles to developing safe protease inhibitor drugs and understanding the biological function of proteases is the difficulty in developing specific inhibitors *in vivo*. Although the classical techniques for development of inhibitors *in vitro* yield potent inhibitors, these compounds often do not translate into specific and/or effective inhibitors *in vivo*, owing to stability, cell permeability or partitioning issues. Screening compounds in intact cells and animal models at an earlier stage in the drug-development process should help overcome some of these challenges. This would allow researchers to focus hit-to-lead optimization on those compounds that are able to reach their target within a biologically relevant context. Activity-based protein profiling methods could then 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. They are therefore 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 specific biological functions and provide biomarkers to 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 new proteomic approaches will allow the focus of protease inhibitor development to shift from potency *in vitro* to specificity and efficacy *in vivo*.

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

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.).

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/nsmb/.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Drag, M. & Salvesen, G.S. Emerging principles in protease-based drug discovery. *Nat. Rev. Drug Discov.* **9**, 690–701 (2010).
2. Rawlings, N.D., Barrett, A.J. & Bateman, A. Asparagine peptide lyases: a seventh catalytic type of proteolytic enzymes. *J. Biol. Chem.* **286**, 38321–38328 (2011).
3. Powers, J.C., Asgian, J.L., Ekici, O.D. & James, K.E. Irreversible inhibitors of serine, cysteine, and threonine proteases. *Chem. Rev.* **102**, 4639–4750 (2002).
4. Johnson, S.L. & Pellecchia, M. Structure- and fragment-based approaches to protease inhibition. *Curr. Top. Med. Chem.* **6**, 317–329 (2006).
5. Turk, B. Targeting proteases: successes, failures and future prospects. *Nat. Rev. Drug Discov.* **5**, 785–799 (2006).
6. Shen, A. Allosteric regulation of protease activity by small molecules. *Mol. Biosyst.* **6**, 1431–1443 (2010).

7. Bock, P.E., Panizzi, P. & Verhamme, I.M. Exosites in the substrate specificity of blood coagulation reactions. *J. Thromb. Haemost.* **5** (suppl. 1), 81–94 (2007).
8. Conus, S. & Simon, H.U. Cathepsins and their involvement in immune responses. *Swiss Med. Wkly.* **140**, w13042 (2010).
9. Lupardus, P.J., Shen, A., Bogyo, M. & Garcia, K.C. Small molecule-induced allosteric activation of the *Vibrio cholerae* RTX cysteine protease domain. *Science* **322**, 265–268 (2008).
10. Bauernfeind, F. *et al.* Inflammasomes: current understanding and open questions. *Cell Mol. Life Sci.* **68**, 765–783 (2011).
11. Edgington, L.E. *et al.* Noninvasive optical imaging of apoptosis by caspase-targeted activity-based probes. *Nat. Med.* **15**, 967–973 (2009).
12. Mason, S.D. & Joyce, J.A. Proteolytic networks in cancer. *Trends Cell Biol.* **21**, 228–237 (2011).
13. López-Otín, C. & Hunter, T. The regulatory crosstalk between kinases and proteases in cancer. *Nat. Rev. Cancer* **10**, 278–292 (2010).
14. Smith, D.M., Fraga, H., Reis, C., Kafri, G. & Goldberg, A.L. ATP binds to proteasomal ATPases in pairs with distinct functional effects, implying an ordered reaction cycle. *Cell* **144**, 526–538 (2011).
15. Palermo, C. & Joyce, J.A. Cysteine cathepsin proteases as pharmacological targets in cancer. *Trends Pharmacol. Sci.* **29**, 22–28 (2008).
16. Overall, C.M. & Kleifeld, O. Tumour microenvironment - opinion: validating matrix metalloproteinases as drug targets and anti-targets for cancer therapy. *Nat. Rev. Cancer* **6**, 227–239 (2006).
17. Omara-Opyene, A.L. *et al.* Genetic disruption of the Plasmodium falciparum digestive vacuole plasmepsins demonstrates their functional redundancy. *J. Biol. Chem.* **279**, 54088–54096 (2004).
18. Deu, E. *et al.* Functional studies of *Plasmodium falciparum* dipeptidyl aminopeptidase I using small molecule inhibitors and active site probes. *Chem. Biol.* **17**, 808–819 (2010).
19. Klemba, M., Gluzman, I. & Goldberg, D.E. A *Plasmodium falciparum* dipeptidyl aminopeptidase I participates in vacuolar hemoglobin degradation. *J. Biol. Chem.* **279**, 43000–43007 (2004).
20. Dalal, S. & Klemba, M. Roles for two aminopeptidases in vacuolar hemoglobin catabolism in *Plasmodium falciparum*. *J. Biol. Chem.* **282**, 35978–35987 (2007).
21. Sadaghiani, A.M., Verhelst, S.H. & Bogyo, M. Tagging and detection strategies for activity-based proteomics. *Curr. Opin. Chem. Biol.* **11**, 20–28 (2007).
22. Cravatt, B.F., Wright, A.T. & Kozarich, J.W. Activity-based protein profiling: from enzyme chemistry to proteomic chemistry. *Annu. Rev. Biochem.* **77**, 383–414 (2008).
23. Desmarais, S., Masse, F. & Percival, M.D. Pharmacological inhibitors to identify roles of cathepsin K in cell-based studies: a comparison of available tools. *Biol. Chem.* **390**, 941–948 (2009).
24. Paulick, M.G. & Bogyo, M. Application of activity-based probes to the study of enzymes involved in cancer progression. *Curr. Opin. Genet. Dev.* **18**, 97–106 (2008).
25. Blum, G. Use of fluorescent imaging to investigate pathological protease activity. *Curr. Opin. Drug Discov. Devel.* **11**, 708–716 (2008).
26. Meng, L., Kwok, B.H., Sin, N. & Crews, C.M. Eponemycin exerts its antitumor effect through the inhibition of proteasome function. *Cancer Res.* **59**, 2798–2801 (1999).
27. Sin, N. *et al.* Total synthesis of the potent proteasome inhibitor epoxomicin: a useful tool for understanding proteasome biology. *Bioorg. Med. Chem. Lett.* **9**, 2283–2288 (1999).
28. Arastu-Kapur, S. *et al.* Identification of proteases that regulate erythrocyte rupture by the malaria parasite *Plasmodium falciparum*. *Nat. Chem. Biol.* **4**, 203–213 (2008).
29. Chang, J.W., Nomura, D.K. & Cravatt, B.F. A potent and selective inhibitor of KIAA1363/AADACL1 that impairs prostate cancer pathogenesis. *Chem. Biol.* **18**, 476–484 (2011).
30. Ahn, K. *et al.* Discovery and characterization of a highly selective FAAH inhibitor that reduces inflammatory pain. *Chem. Biol.* **16**, 411–420 (2009).
31. Staub, I. & Sieber, S.A. Beta-lactam probes as selective chemical-proteomic tools for the identification and functional characterization of resistance associated enzymes in MRSA. *J. Am. Chem. Soc.* **131**, 6271–6276 (2009).
32. Adibekian, A. *et al.* Click-generated triazole ureas as ultrapotent *in vivo*-active serine hydrolase inhibitors. *Nat. Chem. Biol.* **7**, 469–478 (2011).
33. Jain, S., Diefenbach, C., Zain, J. & O'Connor, O.A. Emerging role of carfilzomib in treatment of relapsed and refractory lymphoid neoplasms and multiple myeloma. *Core Evid.* **6**, 43–57 (2011).
34. Hall, C.I. *et al.* Chemical genetic screen identifies Toxoplasma DJ-1 as a regulator of parasite secretion, attachment, and invasion. *Proc. Natl. Acad. Sci. USA* **108**, 10568–10573 (2011).
35. Chandramohanadas, R. *et al.* Apicomplexan parasites co-opt host calpains to facilitate their escape from infected cells. *Science* **324**, 794–797 (2009).
36. Gocheva, V. *et al.* IL-4 induces cathepsin protease activity in tumor-associated macrophages to promote cancer growth and invasion. *Genes Dev.* **24**, 241–255 (2010).
37. Berger, A.B. *et al.* Identification of early intermediates of caspase activation using selective inhibitors and activity-based probes. *Mol. Cell* **23**, 509–521 (2006).
38. Méthot, N. *et al. In vivo* inhibition of serine protease processing requires a high fractional inhibition of cathepsin C. *Mol. Pharmacol.* **73**, 1857–1865 (2008).
39. Blum, G., von Degenfeld, G., Merchant, M.J., Blau, H.M. & Bogyo, M. Noninvasive optical imaging of cysteine protease activity using fluorescently quenched activity-based probes. *Nat. Chem. Biol.* **3**, 668–677 (2007).
40. Blum, G., Weimer, R.M., Edgington, L.E., 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* **4**, e6374 (2009).
41. Hagel, M. *et al.* Selective irreversible inhibition of a protease by targeting a noncatalytic cysteine. *Nat. Chem. Biol.* **7**, 22–24 (2011).
42. Blair, J.A. *et al.* Structure-guided development of affinity probes for tyrosine kinases using chemical genetics. *Nat. Chem. Biol.* **3**, 229–238 (2007).
43. Agard, N.J., Maltby, D. & Wells, J.A. Inflammatory stimuli regulate caspase substrate profiles. *Mol. Cell. Proteomics* **9**, 880–893 (2010).
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.* **70**, 7562–7569 (2010).
45. Bowyer, P.W., Simon, G.M., Cravatt, B.F. & Bogyo, M. Global profiling of proteolysis during rupture of *Plasmodium falciparum* from the host erythrocyte. *Mol. Cell Proteomics* **10**, M110.001636 (2011).
46. Enoksson, M. *et al.* Identification of proteolytic cleavage sites by quantitative proteomics. *J. Proteome Res.* **6**, 2850–2858 (2007).
47. Kleifeld, O. *et al.* Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products. *Nat. Biotechnol.* **28**, 281–288 (2010).
48. Mahrus, S. *et al.* Global sequencing of proteolytic cleavage sites in apoptosis by specific labeling of protein N termini. *Cell* **134**, 866–876 (2008).
49. auf dem Keller, U., Prudova, A., Gioia, M., Butler, G.S. & Overall, C.M. A statistics-based platform for quantitative N-terminome analysis and identification of protease cleavage products. *Mol. Cell Proteomics* **9**, 912–927 (2010).
50. Simon, G.M., Dix, M.M. & Cravatt, B.F. Comparative assessment of large-scale proteomic studies of apoptotic proteolysis. *ACS Chem. Biol.* **4**, 401–408 (2009).
51. Wildes, D. & Wells, J.A. Sampling the N-terminal proteome of human blood. *Proc. Natl. Acad. Sci. USA* **107**, 4561–4566 (2010).
52. Staes, A. *et al.* Selecting protein N-terminal peptides by combined fractional diagonal chromatography. *Nat. Protoc.* **6**, 1130–1141 (2011).
53. Prudova, A., auf dem Keller, U., Butler, G.S. & Overall, C.M. Multiplex N-terminome analysis of MMP-2 and MMP-9 substrate degradomes by iTRAQ-TAILS quantitative proteomics. *Mol. Cell Proteomics* **9**, 894–911 (2010).
54. Schilling, O., Barre, O., Huesgen, P.F. & Overall, C.M. Proteome-wide analysis of protein carboxy termini: C terminomics. *Nat. Methods* **7**, 508–511 (2010).
55. Timmer, J.C. *et al.* Profiling constitutive proteolytic events *in vivo*. *Biochem. J.* **407**, 41–48 (2007).
56. Gray, D.C., Mahrus, S. & Wells, J.A. Activation of specific apoptotic caspases with an engineered small-molecule-activated protease. *Cell* **142**, 637–646 (2010).
57. Truebestein, L. *et al.* Substrate-induced remodeling of the active site regulates human HTRA1 activity. *Nat. Struct. Mol. Biol.* **18**, 386–388 (2011).
58. Vaidya, S., Velazquez-Delgado, E.M., Abbruzzese, G. & Hardy, J.A. Substrate-induced conformational changes occur in all cleaved forms of caspase-6. *J. Mol. Biol.* **406**, 75–91 (2011).
59. Hardy, J.A. & Wells, J.A. Dissecting an allosteric switch in caspase-7 using chemical and mutational probes. *J. Biol. Chem.* **284**, 26063–26069 (2009).
60. Scheer, J.M., Romanowski, M.J. & Wells, J.A. A common allosteric site and mechanism in caspases. *Proc. Natl. Acad. Sci. USA* **103**, 7595–7600 (2006).
61. Schneck, J.L. *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* **47**, 8697–8710 (2008).
62. Patricelli, M.P. *et al. In situ* kinase profiling reveals functionally relevant properties of native kinases. *Chem. Biol.* **18**, 699–710 (2011).
63. Sheahan, K.L., Cordero, C.L. & Satchell, K.J. Autoprocessing of the Vibrio cholerae RTX toxin by the cysteine protease domain. *EMBO J.* **26**, 2552–2561 (2007).
64. Shen, A. *et al.* Mechanistic and structural insights into the proteolytic activation of *Vibrio cholerae* MARTX toxin. *Nat. Chem. Biol.* **5**, 469–478 (2009).
65. Puri, A.W. *et al.* Rational design of inhibitors and activity-based probes targeting *Clostridium difficile* virulence factor TcdB. *Chem. Biol.* **17**, 1201–1211 (2010).
66. Falgueyret, J.P. *et al.* Lysosomotropism of basic cathepsin K inhibitors contributes to increased cellular potencies against off-target cathepsins and reduced functional selectivity. *J. Med. Chem.* **48**, 7535–7543 (2005).
67. Deu, E., Yang, Z., Wang, F., Klemba, M. & Bogyo, M. Use of activity-based probes to develop high throughput screening assays that can be performed in complex cell extracts. *PLoS ONE* **5**, e11985 (2010).
68. Thong, B., Pilling, J., Ainscow, E., Beri, R. & Unitt, J. Development and validation of a simple cell-based fluorescence assay for dipeptidyl peptidase 1 (DPP1) activity. *J. Biomol. Screen.* **16**, 36–43 (2011).
69. Wakata, A. *et al.* Simultaneous fluorescent monitoring of proteasomal subunit catalysis. *J. Am. Chem. Soc.* **132**, 1578–1582 (2010).
70. Poreba, M. & Drag, M. Current strategies for probing substrate specificity of proteases. *Curr. Med. Chem.* **17**, 3968–3995 (2010).
71. Whitney, M. *et al.* Parallel *in vivo* and *in vitro* selection using phage display identifies protease-dependent tumor-targeting peptides. *J. Biol. Chem.* **285**, 22532–22541 (2010).
72. Greenbaum, D.C. *et al.* A role for the protease falcipain 1 in host cell invasion by the human malaria parasite. *Science* **298**, 2002–2006 (2002).
73. Sadaghiani, A.M. *et al.* Design, synthesis, and evaluation of *in vivo* potency and selectivity of epoxysuccinyl-based inhibitors of papain-family cysteine proteases. *Chem. Biol.* **14**, 499–511 (2007).

74. Albrow, V.E. *et al.* Development of small molecule inhibitors and probes of human SUMO deconjugating proteases. *Chem. Biol.* **18**, 722–732 (2011).
75. Ponder, E.L. *et al.* Functional characterization of a SUMO deconjugating protease of *Plasmodium falciparum* using newly identified small molecule inhibitors. *Chem. Biol.* **18**, 711–721 (2011).
76. Patterson, A.W., Wood, W.J. & Ellman, J.A. 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.* **2**, 424–433 (2007).
77. Patterson, A.W. *et al.* Identification of selective, nonpeptidic nitrile inhibitors of cathepsin s using the substrate activity screening method. *J. Med. Chem.* **49**, 6298–6307 (2006).
78. Brak, K., Doyle, P.S., McKerrow, J.H. & Ellman, J.A. Identification of a new class of nonpeptidic inhibitors of cruzain. *J. Am. Chem. Soc.* **130**, 6404–6410 (2008).
79. Brak, K. *et al.* Nonpeptidic tetrafluorophenoxymethyl ketone cruzain inhibitors as promising new leads for Chagas disease chemotherapy. *J. Med. Chem.* **53**, 1763–1773 (2010).
80. Bachovchin, D.A. *et al.* Superfamily-wide portrait of serine hydrolase inhibition achieved by library-versus-library screening. *Proc. Natl. Acad. Sci. USA* **107**, 20941–20946 (2010).
81. Ahn, K. *et al.* Novel mechanistic class of fatty acid amide hydrolase inhibitors with remarkable selectivity. *Biochemistry* **46**, 13019–13030 (2007).
82. Speers, A.E. & Cravatt, B.F. Profiling enzyme activities *in vivo* using click chemistry methods. *Chem. Biol.* **11**, 535–546 (2004).
83. Berkers, C.R. *et al.* Activity probe for *in vivo* profiling of the specificity of proteasome inhibitor bortezomib. *Nat. Methods* **2**, 357–362 (2005).
84. Arastu-Kapur, S. *et al.* Nonproteasomal targets of the proteasome inhibitors bortezomib and carfilzomib: a link to clinical adverse events. *Clin. Cancer Res.* **17**, 2734–2743 (2011).
85. Desmarais, S. *et al.* Effect of cathepsin k inhibitor basicity on *in vivo* off-target activities. *Mol. Pharmacol.* **73**, 147–156 (2008).
86. Gauthier, J.Y. *et al.* The discovery of odanacatib (MK-0822), a selective inhibitor of cathepsin K. *Bioorg. Med. Chem. Lett.* **18**, 923–928 (2008).
87. Geurink, P. *et al.* A peptide hydroxamate library for enrichment of metalloproteinases: towards an affinity-based metalloproteinase profiling protocol. *Org. Biomol. Chem.* **6**, 1244–1250 (2008).
88. Saghatelian, A., Jessani, N., Joseph, A., Humphrey, M. & Cravatt, B.F. Activity-based probes for the proteomic profiling of metalloproteases. *Proc. Natl. Acad. Sci. USA* **101**, 10000–10005 (2004).
89. Harbut, M.B. *et al.* Bestatin-based chemical biology strategy reveals distinct roles for malaria M1- and M17-family aminopeptidases. *Proc. Natl. Acad. Sci. USA* **108**, E526–E534 (2011).
90. Sieber, S.A., Niessen, S., Hoover, H.S. & Cravatt, B.F. Proteomic profiling of metalloprotease activities with cocktails of active-site probes. *Nat. Chem. Biol.* **2**, 274–281 (2006).
91. Nakai, R., Salisbury, C.M., Rosen, H. & Cravatt, B.F. Ranking the selectivity of PubChem screening hits by activity-based protein profiling: MMP13 as a case study. *Bioorg. Med. Chem.* **17**, 1101–1108 (2009).
92. Li, Y.M. *et al.* Photoactivated gamma-secretase inhibitors directed to the active site covalently label presenilin 1. *Nature* **405**, 689–694 (2000).
93. Weerapana, E. *et al.* Quantitative reactivity profiling predicts functional cysteines in proteomes. *Nature* **468**, 790–795 (2010).