Chemical Science

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: M. Poreba, W. Rut, M. Vizovisek, K. Groborz, P. Kasperkiewicz, D. Finlay, K. Vuori, D. Turk, B. Turk, G. Salvesen and M. Drag, *Chem. Sci.*, 2018, DOI: 10.1039/C7SC04303A.

ROYAL SOCIETY OF CHEMISTRY

rsc.li/chemical-science

View Article Online
DOI: 10.1039/C7SC04303A

Selective imaging of cathepsin L in breast cancer by fluorescent activity-based probes

Marcin Poreba[1,2], Wioletta Rut[1], Matej Vizovisek[3], Katarzyna Groborz[1], Paulina Kasperkiewicz[1,2], Darren Finlay[2], Kristiina Vuori[2], Dusan Turk[3], Boris Turk[3], Guy S. Salvesen[2], Marcin Drag[1]

[1]Department of Bioorganic Chemistry, Faculty of Chemistry, Wroclaw University of Technology, Wyb. Wyspianskiego 27, 50-370 Wroclaw, Poland

[2]Sanford Burnham Prebys Medical Discovery Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA

[3]Department of Biochemistry and Molecular and Structural Biology, Jožef Stefan Institute, SI-1000 Ljubljana, Slovenia

Corresponding authors: marcin.poreba@pwr.edu.pl, marcin.drag@wpr.edu.pl

View Article Online
DOI: 10.1039/C7SC04303A

ABSTRACT

Cysteine cathepsins normally function in the lysosomal degradation system where they are critical for maintenance of cellular homeostasis and MHC II immune response, and have been found to have major roles in several diseases including tumor progression. Selective visualization of individual protease activity within a complex proteome is of major importance to establish their roles in both normal and tumor cells, thereby facilitating our understanding of regulation of proteolytic networks. A generally accepted means to monitor protease activity are small molecule substrates and activity-based probes. However, there are eleven human cysteine cathepsins, with a few of them displaying overlapping substrate specificity, making the development of small molecules selectively targeting a single cathepsin very challenging. Here, we utilized HyCoSuL, a positional scanning substrate approach, to develop a highly-selective fluorogenic substrate and activity-based probe for monitoring cathepsin L activity in the breast cancer cell line MDA-MB-231. Use of this probe enabled us to distinguish activity of cathepsin L from other cathepsins, particularly cathepsin B, which is abundant and ubiquitously expressed in normal and transformed cell types. We found that cathepsin L localization in MDA-MB-231 cells greatly overlaps with that of cathepsin B, however, several cathepsin L-rich lysosomes lacked cathepsin B activity. Overall, these studies demonstrate that HyCoSuL-derived small molecule probes are valuable tools to image cathepsin L activity in living cells. This approach thus enables evaluation of cathepsin L function in tumorigenesis and is applicable to other cysteine cathepsins.

View Article Online
DOI: 10.1039/C7SC04303A

1. INTRODUCTION

Cysteine cathepsins are a group of eleven structurally related proteases in humans, normally found in lysosomes that are responsible for intracellular protein degradation, autophagy and immune response (1). Cysteine cathepsins also have major roles in a number of diseases, including cancer, arthritis and osteoporosis, and were therefore considered of major importance in drug development (2-4). Cathepsin L is ubiquitously expressed and is one of the most abundant cathepsins (3). It is synthesized as a proenzyme, which is activated in the acidic milieu of endolysosomal system autocatalytically (5) or by other proteases (3). It is an endopeptidase with the highest catalytic activity at pH 5.0- 6.0 (6, 7) and displays broad substrate specificity (3, 8). In addition to its role in protein degradation, cathepsin L was found to be critical for the MHC II-mediated antigen processing and presentation in thymus (9). Furthermore, cathepsin L was found to have a major role in the nucleus by processing the H3 histone (10) as well as the Cux-1 transcription factor (11), and has an important role in keratinocyte differentiation (12). Similarly to several other cathepsins, cathepsin L was found to have an important role in the progression of cancer. However, in contrast to cathepsins B and S that are present in both stromal cells, such as macrophages, and tumor cells, cathepsin L is primarily secreted by tumor cells (13-15).

Cathepsin L was suggested to participate in tumor cell mechanisms underlying chemo-resistance (16). Active forms of cathepsin L found in the nucleus reportedly serve as either tumor promoters (17, 18) or suppressors (19), depending on the microenvironment. Finally, secreted cathepsin L participates in degradation of extracellular proteins like the insulin receptor, matrix fibronectin, elastin, and collagen (13, 20), as well as several membrane anchored cell adhesion molecules (CAMs) and receptors (21), thereby additionally contributing to tumorigenesis.

While in mouse a single cathepsin L gene exists, in humans there are two closely related cathepsin L orthologues, cathepsin L and cathepsin V. Both cathepsins L and V have maximum activity in the same pH range and share 79.5% of amino acid identity (22, 23). Moreover, several functions originally attributed to cathepsin L based on mouse studies, were found to be taken over by cathepsin V in humans, including MHC II – mediated immune response in thymus (24). This is likely true also for the nuclear functions of cathepsin L, as human cathepsin V, but not cathepsin L, was found to bind DNA (25), however, a thorough validation is still lacking. Moreover, the two cathepsins were found to have very similar substrate specificities, as revealed by the classical Positional Scanning Synthetic Combinatorial Library (PS-SCL) (8). In addition, not only cathepsin V, but also other cathepsins have similar and largely overlapping substrate specificities to cathepsin L (8), also confirmed by several proteomic studies (26, 27). Hence, a cathepsin L selective peptide motif is still to be discovered.

View Article Online
DOI: 10.1039/C7SC04303A

In order to be able to discriminate between cathepsin L and related cathepsins and to correctly address their physiological functions, one has to understand their specificity (28). This is also of crucial importance for development of selective and potent inhibitors, as well as appropriate tools for monitoring their activity in situ. Among the latter are also selective substrates and activity-based probes (ABPs) (29-31). All early substrates aimed to target cathepsin L were unspecific and were efficiently hydrolyzed by cathepsin B and other proteases, and were largely useful for detection of cathepsin activity in biological samples rather than probes for cathepsin L (32, 33). Similarly, the cresyl violet peptidic substrates, which allow monitoring of cathepsin activities in cells, are based on nonselective peptidic motifs (34). A major step forward was application of the reverse-design principle, which is based on the conversion of a selective med-chem optimized inhibitor into a cleavable substrate by replacement of the warhead with a quencher-fluorophore pair. This lead to the development of selective cathepsin S substrates (35, 36), whereas selective substrates for other proteases are still lacking.

ABPs are mostly short peptides coupled to a reactive group (warhead) that binds to catalytic machinery, and a detectable tag (biotin, radioisotope, or fluorophore) (37, 38). Use of ABPs in protease and related enzyme studies was first proposed by Powers (39) and later exploited by others (40-43). Application of ABPs to analysis of cysteine cathepsins was pioneered by the Bogyo group, who initially designed probes based on simple broad-spectrum radiolabeled inhibitors (44), although fluorophore-labeled probes (45) and quenched activity-based probes have also been applied to living cells and whole organisms (46, 47). Some successes have been achieved in the design of selective cathepsin S probes (48, 49), but selective tools for other cathepsins remain to be developed.

Currently, the most successful approach to rapidly identify sequences with high activity and selectivity for individual proteases is HyCoSuL, which utilizes a combination of natural and unnatural amino acids and positional scanning library technology to explore extended chemical space around protease active sites (40, 50, 51). We used two library scanning procedures to obtain highly active substrates for cathepsin L. In the first, we employed a fixed P1 Arg HyCoSuL to dissect cathepsin L specificity in the P4-P2 positions. In the second, we also utilized a P1 individual substrate library to explore cathepsin L preferences in the P1 position. With these two types of libraries we were able to develop potent and highly selective fluorogenic substrates and ABPs for cathepsin L. Finally, we confirmed the utility of identified probes by defining cathepsin L activity and localization in the breast cancer line MDA-MB-231.

View Article Online
DOI: 10.1039/C7SC04303A

2. EXPERIMENTAL

2.1. Reagents. All chemicals were purchased from commercial suppliers and used without further purification. For ACC-labeled individual substrate synthesis, we used Rink amide RA resin (loading 0,48 mmol/g, Iris Biotech GmbH), Fmoc-protected amino acids (purity > 98%, Iris Biotech GmbH, Bachem, Creosalus, P3 BioSystems, QM Bio), N-hydroxybenzotriazole (HOBt monohydrate purity > 98%, Creosalus), N,N-diisopropylethylamine (DIPEA, peptide grade, Iris Biotech GmbH), diisopropylcarbodiimide (DICI, peptide grade, Iris Biotech GmbH), HATU and HBTU (peptide grade, ChemPep Inc.), 2,4,6-trimethylpyridine (2,4,6-collidine, peptide grade, Sigma-Aldrich Sp. z o.o.), N,N`-dimethylformamide (DMF, peptide grade, WITKO Sp. z o.o.), acetonitrile (ACN, HPLC pure, WITKO Sp z o.o.), piperidine (PIP, peptide grade, Iris Biotech GmbH), trifluoroacetic acid (TFA, purity 99%, Iris Biotech GmbH), triisopropylsilane (TIPS, purity 99%, Sigma-Aldrich Sp. z o.o.), methanol (MeOH, pure for analysis, POCh), dichloromethane (DCM, pure for analysis, POCh), diethyl ether (Et_2O, pure for analysis, POCh), acetic acid (AcOH, purity 98%, POCh), and phosphorus pentoxide (P_2O_5, purity 98%, POCh). Fmoc-6-ahx-OH, biotin, and 2-chlorotrityl chloride resin (100-200 mesh, 1.59 mmol/g) were purchased from Iris Biotech GmbH. TFE (2,2,2-trifluoroethanol), HBr (30% wt. in AcOH), anhydrous THF (tetrahydrofuran), NMM (4-methylmorpholine), IBCF (isobutylchloroformate) and 2,6-DMBA (2,6-dimethylbenzoic acid) were from Sigma-Aldrich. Fluorescent tags (Cyanine5 NHS/Cy5-NHS, Cyanine3 NHS/Cy3-NHS, and BODIPY-FL NHS/BDP_{FL}-NHS) were purchased from Lumiprobe. Diazomethane was generated according to the Aldrich Technical Bulletin (AL-180) protocol.

2.2. Preparation of recombinant cathepsins B, L, K, V, and S. Detailed description of expression and purification is found elsewhere (52, 53). The concentration of active cathepsins was determined via active site titration with E-64 inhibitor (Peptide Institute).

2.3. P4-P2 Hybrid Combinatorial Substrate Library (HyCoSuL) synthesis. A detailed protocol for synthesis of the P4-P2 fluorogenic substrate combinatorial library is provided elsewhere (54).

2.4. Enzymatic kinetic studies. All kinetic experiments were performed using fMax (Molecular Devices) and CLARIOStar (BMG LABTECH) spectrofluorimeter operating in kinetic mode using 96-well plates. ACC-labeled substrates were screened using 355 nm and 460 nm wavelengths (excitation and emission, respectively). Cathepsin assay buffer (for library, substrates and kinetic inhibitors) contained 100 mM sodium acetate, 100 mM sodium chloride, 10 mM DTT, 1 mM EDTA, pH 5.5. Buffer was prepared at room temperature and enzyme kinetic studies were performed at 37 °C.

View Article Online
DOI: 10.1039/C7SC04303A

2.5. Characterization of cathepsin L P4-P2 substrate specificity using HyCoSuL. P4, P3, and P2 sub-libraries were each screened at 100 μM concentration with cathepsin L in 100 μL final volume. The active enzyme concentration was in the 0.5-2.0 nM range, depending on sub-library. Assay time was 30 min, but only the linear portions of progression curves (5-15 min) were used for velocity (RFU/s) calculations. Each sub-library screening was repeated at least 3 times, and the average value was used to create the substrate specificity matrix. The best recognized amino acid at each position was set to 100%, and other amino acids were normalized accordingly.

2.6. Characterization of cathepsin L P1 substrate specificity. To determine cathepsin L preferences in the P1 position we used a fluorogenic substrate library of 139 components with the general formula Ac-Ala-Arg-Leu-P1-ACC. The library contains 19 natural and 120 unnatural amino acids. Synthesis and detailed structure of the library are found elsewhere (51). The library was screened with cathepsin L at two concentrations: 4 μM (below K_M values) and 200 μM (above K_M values). The substrate cleavage assay was carried out for ~30 min, but only linear portions of hydrolysis curves (10-15 min) were analyzed. Both screenings were repeated at least three times, and average values are presented. The Ac-Ala-Arg-Leu-Arg-ACC substrate served as a positive control, and the rate of its hydrolysis was set as 100%. Values of other substrates were normalized to the P1 Arg substrate.

2.7. Synthesis of individual optimized substrates. Each ACC-labeled cathepsin L substrate analyzed was synthesized and purified as described elsewhere (55).

2.8. Screening of potential cathepsin L-selective substrates with other cathepsins. All potentially selective cathepsin L substrates were tested for their selectivity against a panel of five different cathepsins containing cathepsins L, V, S, K and V at concentrations of 2 μM and 100 μM . To obtain reliable results, the concentration of each cathepsin was 5 nM. Substrate hydrolysis was carried out for 10 min (cathepsin L) and up to 60 min (for other cathepsins). Hydrolysis rates (RFU/s) are presented as average values from at least three screenings.

2.9. Determination of kinetic parameters (k_{cat}, K_M, k_{cat}/K_M) for individual substrates. Classic substrate velocity plots were analyzed according to Michaelis-Menten kinetic and a detailed protocol for determination of kinetic parameters for ACC-tagged substrates is found elsewhere (55). Because the substrates demonstrated a high degree of selectivity for cathepsin L, we had to employ higher concentrations of other cathepsins to observe substrate hydrolysis, thus the following cathepsin concentrations were used: cathepsin L, 0.1-0.5 nM; cathepsin B, 4-15 nM; cathepsin K, 5-20 nM; cathepsin V, 10-20 nM; and cathepsin S, 5-20 nM. Experiments were repeated at least three times, and results are presented as an average. S.D. were below 15%.

View Article Online
DOI: 10.1039/C7SC04303A

2.10. Synthesis of irreversible activity-based probes. Two biotin-labeled probes (**MP-cL1 and MP-cL2**) were synthesized as described by Poreba et al. (30). In brief, a biotin-6-ahx-His(Trt)-*D*Thr(*t*Bu)-Phe(F_5)-OH peptide was synthesized on solid support using 2-chlorotrityl chloride resin and used for further synthesis without purification. In parallel, two amino acyl-warheads amino acids conjugated to electrophilic binding groups. (Boc-Lys(Cbz)-AOMK and Boc-Arg(Cbz)$_2$-AOMK) were synthesized in solution followed by use of diazomethane in diethyl ether (Lys/Arg-CH_2N_2), HBr in AcOH (Lys/Arg-CH_2Br) and 2,6-DMBA in DMF (Lys/Arg-AOMK). Next, Boc-Lys(Bzl)-AOMK and Boc-Arg(Cbz)$_2$-AOMK were N-terminally de-protected in TFA/DCM (1:1 v/v with 3% of TIPS, 30 min) and coupled with the above peptide to obtain (1) biotin-6-ahx-His(Trt)-*D*Thr(*t*Bu)-Phe(F_5)-Lys(Cbz)-AOMK and (2) biotin-6-ahx-His(Trt)-*D*Thr(*t*Bu)-Phe(F_5)-Arg(Cbz)$_2$-AOMK, followed by purification on HPLC. Next, Trt/tBu protecting groups were removed in TFA/DCM and Cbz groups were removed via hydrogenolysis (H_2, Pd/C, DMF). The final compounds were purified again on HPLC. Fluorescence probes were similarly synthesized. **Cy5-labeled MP-cL3 probe.** Boc-6-ahx-His(Trt)-*D*Thr(*t*Bu)-Phe(F_5)-OH peptide was synthesized on solid support and used without further purification. Boc-Cys(Bzl)-AOMK was synthesized and N-terminally de-protected as described above. Next, the peptide and the H_2N-Cys(Bzl)-AOMK were coupled in DMF to obtain Boc-6-ahx-His(Trt)-*D*Thr(*t*Bu)-Phe(F_5)-Cys(Bzl)-AOMK. All protecting groups were then removed in TFA/DCM (1:1 v/v with 3% of TIPS, 30 min), and excess TFA was removed in a rotary evaporator under reduced pressure (the mixture was washed several times with DCM). The crude product (1eq) was dried over P_2O_5 in a desiccator for several hours. In a separate flask, 1eq of Cy5-NHS was dissolved in DMF, followed by the addition of 5eq of DIPEA. The solution was immediately mixed with crude peptide-AOMK (1eq), and the reaction was carried out at room temperature for 30 min. Reaction progress was monitored by HPLC (220 nm for peptide-AOMK or 646 nm for Cy5-NHS). The final product was purified on HPLC to obtain Cy5-6-ahx-His-*D*Thr-Phe(F_5)-Cys(Bzl)-AOMK. **Cy5/BDP_{FL}/Cy3-labeled MP-pc1/2/3 pan-cathepsin specific probes.** First, the Boc-6-ahx-Ala-Arg(Pbf)-Leu-OH peptide was synthesized on a solid support and used without further purification. In parallel, Boc-Arg(Boc)$_2$-AOMK was synthesized and N-terminally de-protected as described above. Next, the peptide and warhead were coupled in DMF to obtain Boc-6-ahx-Ala-Arg(Pbf)-Leu-Arg(Boc)$_2$-AOMK. The crude product was then purified on HPLC, de-protected with TFA/DCM, and dried over P_2O_5 to obtain H_2N-6-ahx-Ala-Arg-Leu-Arg-AOMK. Finally, the fluorescent tag (BDP_{FL}, Cy3) was attached to the N-terminus and the product was purified on HPLC to obtain BDP_{FL}-6-ahx-Ala-Arg-Leu-Arg-AOMK (MP-pc2) and Cy3-6-ahx-Ala-Arg-Leu-Arg-AOMK (MP-pc3). Cy5-Ala-Arg-Leu-Arg-AOMK (MP-pc1) was synthesized without a linker; thus first the Boc-Ala-Arg(Pbf)-Leu-OH peptide was synthesized and then coupled to the H_2N-Arg-AOMK warhead. Next, the compound was de-protected and labeled with a Cy5 tag in a manner similar to BDP_{FL} and Cy3 probes.

View Article Online
DOI: 10.1039/C7SC04303A

2.11. Determination of kinetic parameters (k_{obs}/I) for activity-based probes. For kinetic studies, five recombinant human cathepsins (B, L, V, K, and S) were expressed and purified and active site titrated using E64 inhibitor. The k_{obs}/I parameter was measured under pseudo-first order kinetic conditions, as described (40). The cathepsin of interest was mixed with various concentrations of probe (at least 5-fold excess over cathepsin) and 100 μM ACC-fluorescence substrate. Experiments were repeated at least three times, and results are presented as an average. S.D. were below 20%.

2.12. Detection of cathepsin L in HEK293T cells with biotinylated ABPs. To evaluate applicability of bMP-cL1 and bMP-cL2 probes using affinity-based enrichment assays, HEK293T cells were seeded on two 10 cm plates and cultured 2 days. Medium was changed, and cells were then transfected using the pcDNA3::ppCatL construct (prepared in-house for expression of human preprocathepsin L in mammalian cells) and Lipofectamine 2000™ based on the manufacturer's guidelines (Thermo Scientific). Forty-eight hours later, cells were harvested and whole cell lysates were prepared in a lysis buffer maintaining cathepsin activity (50 mM sodium phosphate, 100 mM NaCl, 1 mM EDTA, 1 mM DTT, 0.5% NP-40 (nonyl phenoxypolyethoxylethanol), pH 6.0) as described elsewhere (26). Non-transfected HEK293T cells served as controls. Transfection efficiency was calculated by measuring cathepsin L activity increase in HEK293T cells after transfection. For cathepsin L labeling, cell lysates were mixed with various (0 - 10 μM) concentrations of MP-cL1 and MP-cL2 probes and incubated 2 hours at 37 °C. To control for non-specific binding, 25 μM E-64d, a pan-cathepsin cell permeable inhibitor was added to lysates 1 hour prior to probe addition. After incubations, samples were mixed with 6x SDS-PAGE loading buffer and analyzed by immunoblotting. Nitrocellulose membranes were blocked after transfer with 5 ml 3% BSA (bovine serum albumin) in PBS, pH 7.4, for 1 hour before the streptavidin-HRP conjugate was added to membranes at a 1:2000 dilution, according to the manufacturer's guidelines (Alpha diagnostics). Membranes were incubated 45 minutes with gentle rocking and then washed 4 times for 15 min in 10 ml PBS, pH 7.4. Membranes were then developed using standard ECL detection reagents (GE Healthcare), and signals were detected on bioluminescence sensitive film (Kodak).

2.13. ABP uptake into cells. To evaluate probe uptake, 250,000 HEK293T cells transfected with pcDNA3:ppCatL were seeded into 6-well plates and incubated with various concentrations of bMP-cL1 and bMP-cL2 (0 - 10 μM) for up to 24 hours. In control experiments cells were preincubated with 25 μM E-64d, a pan-cathepsin inhibitor, for 2 hours prior to probe addition. After incubations, samples were collected and immunoblotted. Biotinylated proteins were detected as described above.

2.14. Evaluation of probe selectivity by immunocapture. To evaluate MP-cL1 and MP-cL2 selectivity, 250,000 HEK293T cells overexpressing cathepsin L were plated and treated with probes and inhibitors as described above. Streptavidin-bound magnetic beads were prepared in separate tubes according to the manufacturer's guidelines (Pierce™ Streptavidin Magnetic Beads, Thermo Scientific).

View Article Online
DOI: 10.1039/C7SC04303A

Briefly, 50 μL beads were washed with sample lysis buffer and then lysates were added. Tubes were incubated at room temperature on a rocking platform for 2 hours. Beads were then extensively washed with 50 mM Tris-buffered saline containing 0.1% Tween-20 and collected between each washing step. Beads were then re-suspended in 1xSDS loading buffer at a final volume of 100 μL, and the samples were placed at 95 °C and prepared for immunoblotting as described above to detect biotinylated proteins in the lysates, as well as in the unbound and eluted fractions. To confirm cathepsin L/B selectivity, we seeded HEK293T cells as described above. Probes were added at 5 μM final concentration and incubated 18 hours before cells were harvested and lysed to isolate biotinylated proteins using streptavidin magnetic beads, as described above. Lysates, plus the unbound and eluting fractions were then evaluated to detect cathepsin B (in-house mouse polyclonal Ab [51]; dilution 1:1000) and cathepsin L (Abcam ab62710).

2.15. Cathepsin reactivity with MP-cL3 and MP-pc1 ABPs based on SDS-PAGE analysis. To determine Cy5-probes potency and selectivity, each active site-titrated cathepsin was preincubated separately in the assay buffer for 15min at 37 °C at 10 nM (experiment A) or 100 nM (experiment B) concentration, and then mixed with various concentrations of either MP-cL3 or MP-pc1 probe for 30 min. In the experiment A, probe concentrations were 2 nM, 10 nM, 50 nM, and 250 nM, while in the experiment B probe concentrations were 20 nM, 100 nM, 500 nM, and 2.5 μM. Reaction volume for all samples was 200 μL. After 30 min, 100 μL of 3xSDS/DTT was added to reaction mixtures, samples were boiled for 5 min, and then 20 μL of sample was run on 4-12% Bis-Tris Plus 15-well gels for 30 min at 200 V with 2 μL of PageRuler^{TM} Prestained Protein Ladder. Gels were then directly scanned at 700 nm (red channel for Cy5 detection; excitation 685 nm) using an Odyssey fluorescence imaging system (LI-COR) and images analyzed using Image Studio software.

2.16. Detection of cathepsin L in MDA-MB-231 cells by SDS-PAGE and Western blotting. Low passage MDA-MB-231 cells were seeded on 12-well plates (200,000 cells per well per 1 mL of medium) and cultured in Dulbecco`s Modification of Eagle`s Medium (DMEM) supplemented with 10% of Fetal Bovine Serum, 2 mM of L-glutamine, 100 units/mL penicillin and 100 □g/mL streptomycin. The next day, MP-cL3 and MP-pc1 probes (1 μM) were added to the medium, and cells were incubated for 0-18 hours. To control for non-specific binding, cells were preincubated with E64d (25 μM) for 2 hours prior to probe addition. At various time points, the medium was aspirated and cells were gently washed twice with 1 mL of Dulbecco`s Phosphate-Buffered Saline (DPBS) and then trypsinized, harvested, and centrifuged for 10 min (500×g). Supernatants were removed by aspiration and pellets washed with 1 mL of DPBS and centrifuged (5 min, 500×g). Supernatants were again removed and pellets solubilized in 100 μL 1xSDS/DTT and then boiled 5 min. Each sample was then cooled to room temperature and sonicated (Branson Sonifier 250, duty cycle: 50, output control: 5, cycles: 5) and then 30 μL was loaded on 4-12% Bis-Tris Plus

View Article Online
DOI: 10.1039/C7SC04303A

gels and run for 30 min at 200V with a SDS-PAGE Standards/Broad Range ladder. Proteins were transferred to a 0.2 μm nitrocellulose membrane for 1 hour at 10 V followed by Ponceau staining to verify equal loading. Membranes were blocked with 2.5% BSA in Tris-buffered saline/0.1% v/v Tween 20 (TBS-T) for 1 hour at room temperature. Cathepsins L and B were detected by primary goat anti-hCathepsin L (R&D, AF952, 0.2mg/mL, 1:2,000) and goat anti-hCathepsin B (R&D, AF953, 0.2mg/mL, 1:1,000) antibodies, respectively, followed by incubation with secondary antibody (IRDye® 800CW, LI-COR, donkey anti-goat, 1:10,000). Membranes were incubated with primary antibodies for 2 hours, and with (fluorescent) secondary antibodies for 30 min (both in TBS-T, at room temperature). Fluorescence was read at 700nm (Cyanine5; excitation 685 nm) and 800nm (secondary antibody; excitation 785 nm) using an Odyssey fluorescence imaging system (LI-COR). Images were analyzed using Image Studio software.

2.17. Detection of cathepsin L in MDA-MB-231 cells by immunofluorescence. Glass coverslips 15 mm in diameter were placed in 12-well plates and coated with poly-*D*-lysine (500 μL, 1 mg/mL in water, Specialty Media) for 30 minutes. Poly-*D*-lysine was aspirated and coverslips were washed twice with sterile water and air-dried. Next, low passage MDA-MB-231 cells in DMEM (100,000 cells/mL) were added to wells (1mL per well) and allowed to attach for 2 hours. Cells were then treated with Cy5-labeled probes (1 μM) for 8 or 24 hours. Control cells were preincubated with E64d for 2 hours. After various incubation times, medium was removed and cells were gently washed twice with DPBS (1mL each). Cells were then fixed with ice-cold methanol (500 μL per well) for 15 minutes at -20 °C. Methanol was gently removed and fixed cells were washed twice with DPBS, blocked and permeabilized with 10% BSA in DPBS (with 0.5% Triton-X, v/v) for 1 hour at room temperature, and washed twice with 3% BSA in DPBS (with 0.5% Triton-X, v/v). Cathepsins L and B were detected by primary goat anti-hCathepsin L (R&D, AF952, 0.2mg/mL, 1:200) and goat anti-hCathepsin B (R&D, AF953, 0.2mg/mL, 1:80), respectively, in 3% BSA in DPBS (with 0.5% Triton-X, v/v) for 2 hours at room temperature. Antibodies were removed and cells were washed twice with 3% BSA in DPBS (with 0.5% Triton-X, v/v) and labeled with secondary antibody (Alexa Fluor™ 488 donkey anti-goat IgG (H+L), A11055, Invitrogen, 1:500) in DPBS (with 0.5% Triton-X, v/v) for 1 hour at room temperature. Next, cells were washed twice with DPBS, and coverslips were mounted using Vectashield mounting medium fluorescence containing DAPI (Vector Lab. H-1000). In brief, a drop of mounting medium was placed on the Superfrost Plus microscope slide (Fisher Scientific, no. 12-550-15) and the coverslip was placed on the drop upside-down. Excess mounting medium was removed with a kimwipe, and nail polish was used to seal coverslips before subjecting slides to confocal microscope analysis (Axio Observer.A1 Inverted Microscope, Zeiss, with BD Carv II confocal unit, objective 40x, and Zeiss LSM 710 NLO Confocal Microscope, objective PlanApochromat 63x/1.4 DIC oil). DAPI was detected by the UV channel, the Cy5 probe was detected with Cy5 filter (excitation 647 nm (Axio Observer) or 633 nm (Zeiss LSM), emission 667 nm; red

View Article Online
DOI: 10.1039/C7SC04303A

channel) and cathepsins/secondary antibodies were read using FITC filters (excitation 495 nm (Axio Observer) or 488 nm (Zeiss LSM), emission 517 nm; red channel; green channel). All images were acquired in .tiff format in MetaMorph software either in .lsm format in ZEN software and analyzed using ImageJ and ZEN software. Images are representative views of cells from two coverslips.

2.18. Weighted colocalization coefficients. To quantify ABP/cathepsin colocalization we calculated weighted colocalization coefficient between MP-cL3 ABP (red) and cathepsin L (green) and between MP-cL3 ABP (red) and cathepsin B (green) using ZEN 2011 software. These calculations were performed by summing the pixels in the colocalized regions (red and green) and dividing by the sum of pixels in red channel (ABP). The value of each pixel was equal to its intensity value (from 0 to 1), thus a weighted colocalization coefficient is more accurate here than the standard colocalization coefficient where all the pixels, regardless of their intensity, have a value of 1. To eliminate the red and green staining background in these calculations, we set crosshairs according to single label controls. A graphical example of the calculated weighted colocalization coefficient is presented in **Figure S12**.

3. RESULTS & DISCUSSION

3.1. P4-P1 combined approach with unnatural amino acids.

To develop novel active and selective substrates and ABPs for cathepsin L, we undertook a combined approach employing HyCoSuL and a P1-individual substrate library containing unnatural amino acids. **Figure 1** shows an outline of this method.

View Article Online
DOI: 10.1039/C7SC04303A

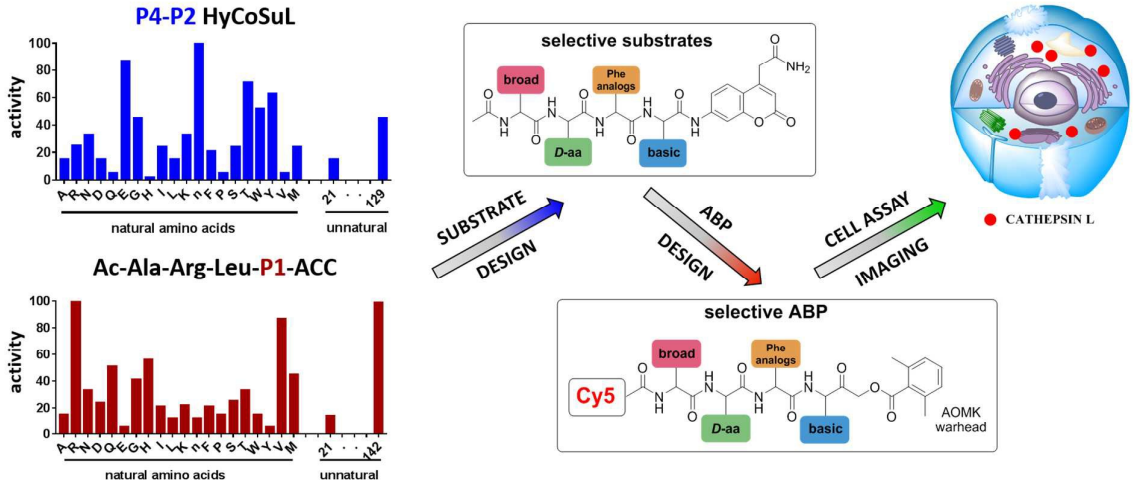


Figure 1. A general procedure to optimize cathepsin L fluorogenic substrates and ABP. P4-P2 enzyme preferences are determined using previously reported HyCoSuL technology, and the P1 position is screened using a fixed P4-P3-P2 sequence, but varying P1 (Ac-Ala-Arg-Leu-P1-ACC) allowing determination of the full substrate specificity profile (P4-P1). Next, the most selective substrate is converted into fluorescent-labeled ABP which allows for the selective imaging of cathepsin L in cells.

3.2. P4-P2 Hybrid Combinatorial Substrate Library. Cathepsin L substrate specificity was determined using a substrate combinatorial library containing Arg at P1 (54). Arg is reportedly recognized by cysteine cathepsins in the S1 pocket (8). The overall library architecture was Ac-P4-P3-P2-Arg-ACC, and the library was composed of three sub-libraries: P4, Ac-X-Mix-Mix-Arg-ACC; P3, Ac-Mix-X-Mix-Arg-ACC; and P2, Ac-Mix-Mix-X-Arg-ACC (Mix: an equimolar mixture of natural amino acids, omitting cysteine and substituting norleucine for methionine). In addition to natural amino acids, we also used 101 unnatural derivatives that can explore extended chemical space around the cathepsin L active site pockets. As a fluorescent tag in this library (and all individual substrates for cathepsin L) we selected an ACC (7-amino-4-carbamoylcoumarin) reporter. This group is widely used for generation of fluorescent substrates for proteases due to its high quantum yield and great signal-to-noise ratio. Moreover, ACC when quenched by a peptide bond displays almost non-detectable fluorescence (56, 57). The library was screened with cathepsin L and the data were used to create a broad substrate specificity matrix (**Figure 2**). Screening the enzyme with the natural amino acids library confirmed the specificity profile obtained by Choe et al. (8); however, use of the HyCoSuL approach allowed us to obtain a highly detailed picture of cathepsin L active site preferences. We found that cathepsin L has relatively narrow specificity in the S2 pocket, in agreement with proteomic studies (26, 27, 58). Moreover, unnatural Phe derivatives were well-recognized (the best being Phe(3-Cl)). Other amino acids were only poorly recognized or ignored by cathepsin L. Based on the P2 specificity matrix, we conclude that the cathepsin L S2 pocket is hydrophobic but not large, as some

View Article Online
DOI: 10.1039/C7SC04303A

Phe analogs (for example, hPhe or 1-Nal) were very weakly hydrolyzed. The P3 position appeared more tolerant than P2. Cathepsin L mostly preferred basic (Arg, Lys, Orn, Dab) and some hydrophobic (Phg, Nle(OBzl)) amino acids. Interestingly, some *D*-amino acids were better recognized by cathepsin L than their *L*-enantiomers, indicating that the S3 pocket is not stereochemically specific. Moreover, several amino acids (mostly Phe derivatives) were completely ignored by cathepsin L. Similarly to P3, P4 also displayed broad specificity with preference for basic amino acids (Dab, Dap, Orn, Agp) and ability to tolerate *D*-amino acids. Among natural amino acids, only branched Leu, Ile and Val and hydrophobic Phe and Tyr were very poorly recognized by cathepsin L. Among unnatural amino acids, only Phe derivatives were not recognized.

Finally, HyCoSuL screening represents a valuable tool in design of potent or selective enzyme substrates; however, we note that HyCoSuL (like PS-SCL (8, 59)) is a combinatorial technique, so potential cooperativity of protease subsites must be considered. To rule out this possibility in terms of cathepsin L, we synthesized three sets of substrates with the formulas Ac-P4-Lys-Phe-Arg-ACC, Ac-His-P3-Phe-Arg-ACC, and Ac-His-Lys-P2-Arg-ACC and performed kinetic studies. When we compared their k_{cat}/K_m profiles to HyCoSuL screening data, we found that cathepsin L displays almost no S4-S2 subsite cooperativity (**Figure S1**).

3.3. P1 individual substrate library. Cathepsin L reportedly cleaves substrates at the carboxylic site of Arg (P1 position); however, other amino acids are tolerated by this protease in the S1 pocket. Choe *et al.* demonstrated that at P1 cathepsin L also recognizes Lys, Gln, Thr, Nle and Met (8), in agreement with proteomic data (26, 58). To gain further insight into cathepsin L P1 preferences we utilized a 139-membered individual substrate library containing over 100 unnatural amino acids. As a P4-P2 scaffold we selected an Ala-Arg-Leu sequence, which is recognized by almost all cysteine cathepsins (8). The Ac-Ala-Arg-Leu-P1-ACC library contains diverse chemical groups at P1: basic, acidic, hydrophilic, hydrophobic, and linear, among others. The specificity matrix is presented in **Figure 3** (the reference Arg and the best amino acids are shown in red). Our analysis revealed that some unnatural amino acids, namely, Cys(Bzl), Cys(MeBzl), Cys(MeOBzl) and Nle(OBzl), are significantly better recognized at P1 than the natural residue Arg. This observation demonstrates that the cathepsin L S1 pocket has dual properties: it is negatively charged as it accepts Arg and hArg, but it can also accommodate large, hydrophobic amino acid side chains.

View Article Online
DOI: 10.1039/C7SC04303A

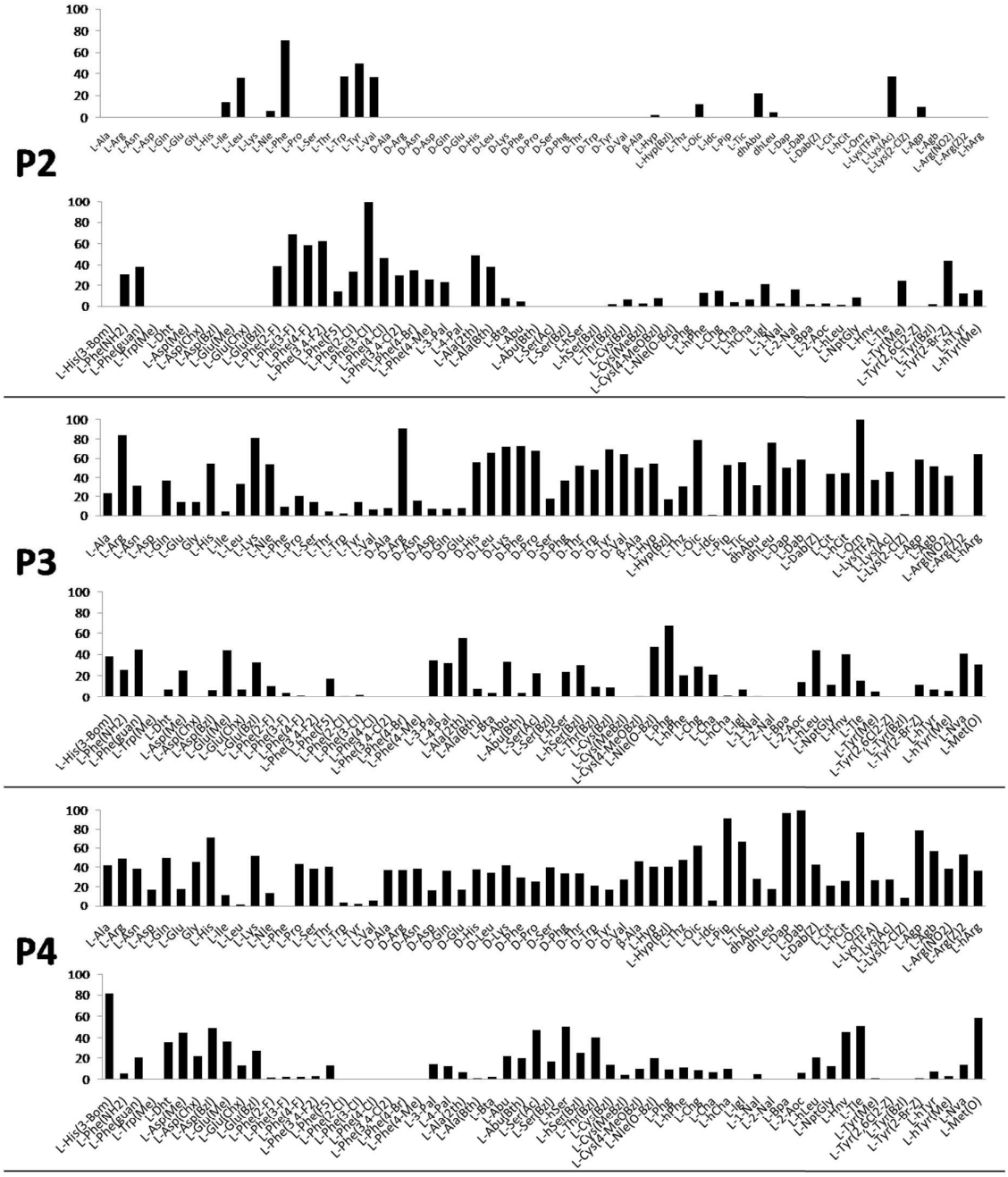


Figure 2. Human cathepsin L preferences at the P4, P3, and P2 positions. Human cathepsin L preferences were determined using three HyCoSuL sub-Libraries, Ac-P4-Mix-Mix-Arg-ACC, Ac-Mix-P3-Mix-Arg-ACC, Ac-Mix-Mix-P2-Arg-ACC, all containing natural and unnatural amino acids. The x axis shows abbreviated names of amino acids and the y axis displays relative activity presented as a percentage of the best-recognized amino acid. Standard deviations calculated from three screening were 15% of values shown in the figure.

View Article Online
DOI: 10.1039/C7SC04303A

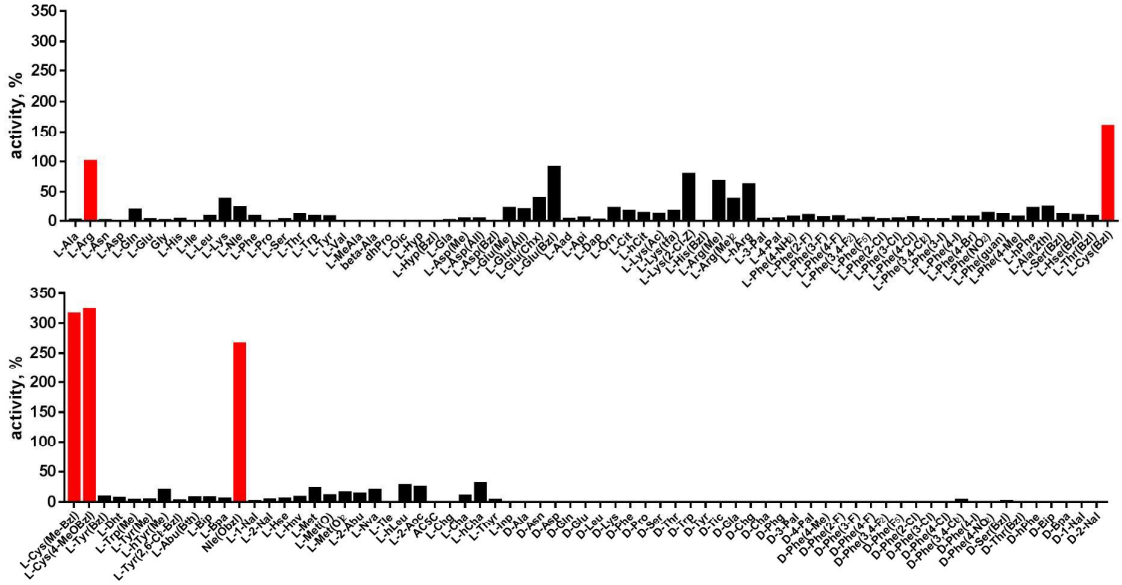
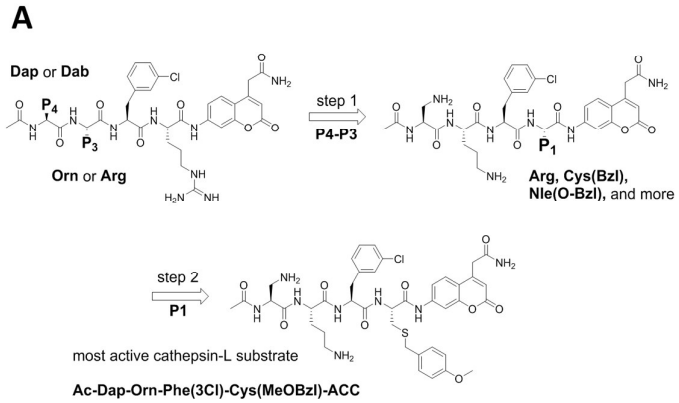


Figure 3. Human cathepsin L preferences at the P1 position. P1 human cathepsin L preferences were determined using an individual fluorogenic substrate library with the general formula Ac-Ala-Arg-Leu-P1-ACC. In this library, the P1 position is occupied by either natural or a diverse set of unnatural amino acids. The x axis shows abbreviated names of amino acids and the y axis displays relative activity presented as a percentage of the best-recognized amino acid. Standard deviations calculated from three screening were 15% of values shown in the figure. Red bars indicate the best natural (Arg) or unnatural amino acids.

3.4. Development of novel, potent cathepsin L substrates. To develop active cathepsin L substrates we combined the HyCoSuL approach with results from P1 library screening (**Figure 4**). Initially, we synthesized four potentially best substrates with various P4-P2 regions (**Table S1**) and compared the hydrolysis efficiency with the reference Ac-His-Arg-Phe-Arg-ACC substrate (**Figure 4**). Kinetic analysis revealed the best cathepsin L substrate (with fixed Arg in P1) to be Ac-Dap-Orn-Phe(3-Cl)-Arg-ACC, which was recognized almost three times better (k_{cat}/K_M) than the reference substrate. Next, we synthesized a 10-membered library of substrates in which the P4-P2 region was fixed with the best recognition sequence, while varying the P1 position (**Figure 4**, **Tables S1,S2**). k_{cat}/K_M analysis allowed us to synthesize two highly active substrates, Ac-Dap-Orn-Phe(3-Cl)-Cys(Me*O*Bzl)-ACC and Ac-Dap-Orn-Phe(3-Cl)-Nle(*O*Bzl)-ACC, which are the most potent cathepsin L tetrapeptide substrates reported to date: both are hydrolyzed almost 4-times faster by cathepsin L than is the Ac-His-Arg-Phe-Arg-ACC substrate.

View Article Online
DOI: 10.1039/C7SC04303A



B

substrate	k_{cat}/K_M, $M^{-1}s^{-1}$
Natural amino acids	
Ac-His-Arg-Phe-Arg-ACC (reference)	**13,050,000**
Unnatural aa (P4-P2 HyCoSuL)	
Ac-Dap-Orn-Phe(3-Cl)-Arg-ACC	**36,310,000**
Ac-Dap-Arg-Phe(3-Cl)-Arg-ACC	**25,400,000**
Ac-Dab-Orn-Phe(3-Cl)-Arg-ACC	**35,070,000**
Ac-Dab-Arg-Phe(3-Cl)-Arg-ACC	**31,960,000**
Unnatural aa (P1 optimization)	
Ac-Dap-Orn-Phe(3-Cl)-Arg-ACC	**36,310,000**
Ac-Dap-Orn-Phe(3-Cl)-Lys-ACC	**19,810,000**
Ac-Dap-Orn-Phe(3-Cl)-Leu-ACC	**11,290,000**
Ac-Dap-Orn-Phe(3-Cl)-hArg-ACC	**30,500,000**
Ac-Dap-Orn-Phe(3-Cl)-Glu(Bzl)-ACC	**34,270,000**
Ac-Dap-Orn-Phe(3-Cl)-Cys(Bzl)-ACC	**45,420,000**
Ac-Dap-Orn-Phe(3-Cl)-Cys(MeBzl)-ACC	**41,660,000**
Ac-Dap-Orn-Phe(3-Cl)-Cys(MeOBzl)-ACC	**48,350,000**
Ac-Dap-Orn-Phe(3-Cl)-Nle(OBzl)-ACC	**47,400,000**
Ac-Dap-Orn-Phe(3-Cl)-Lys(2-Cl-Z)-ACC	**29,280,000**

Figure 4. New, potent cathepsin L substrates. Panel A Schematic representation of the two-step procedure used to develop novel, active cathepsin L substrates. The protocol combines the HyCoSuL approach with use of the P1 individual substrate library. Using the former, we extracted the Dap-Orn-Phe(3-Cl) motif as the optimal in P4-P2 position. Then we added the best amino acids from the P1 library screening to obtain the Ac-Dap-Orn-Phe(3-Cl)-Cys(*O*MeBzl)-ACC substrate. **Panel B** The kinetic parameters (k_{cat}/K_M) for the most active cathepsin L substrates developed through combining HyCoSuL profiling with P1-optimization; aa – amino acids.

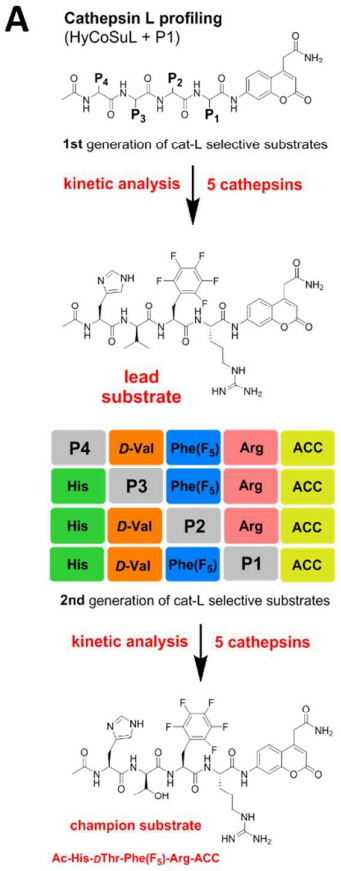
3.5. Analysis of cathepsin L-selective substrates. Although the family of Ac-Dap-Orn-Phe(3-Cl)-**P1**-ACC substrates are very potent cathepsin L substrates, they display only moderate or even poor selectivity towards other cysteine cathepsins (data not shown). Next, to develop a highly selective cathepsin L substrate, we synthesized eleven tetrapeptide substrates with various P4-P1 sequences (**Figure 5**, **Table S3**). For each position we selected the most characteristic amino acids from the cathepsin L P4-P2 profile (**Figure 2).** For the P4 position we selected a variety of amino acid side chains: Tic, His, Ala, and Met(O). In P3 we chose the best amino acids (Orn, Dab) and several *D*-amino acids, which are also well tolerated. P2 was occupied by Phe derivatives, the group preferred by the enzyme. For P1, we selected mostly basic amino acids (Arg, Orn, and Lys). We then hydrolyzed these first-generation substrates (10 μM each) with 5nM of cysteine cathepsins B, K, L, S, and V (**Table S4**) and utilized the most selective substrate (Ac-His-*D*Val-Phe(F_5)-Arg-ACC) as a candidate for further structure optimization. We next synthesized a second-generation library of substrates, replacing one position at a time (from P4 to P1) (**Figure 5**, **Table S5 and S6**). For P4 we selected His,

View Article Online
DOI: 10.1039/C7SC04303A

Arg and hArg; for P3, several *D*-amino acids; for P2, Phe analogues; and for P1, 10 amino acids best recognized by cathepsin L (Ac-Ala-Arg-Leu-P1-ACC library profiling, **Figure 3**). k_{cat}/K_M profiling of these substrates with L, V, B, K, and S cathepsins revealed several "cathepsin L selectivity requirements" (**Figure 5**). In P4, the most selective amino acid appeared to be His, as this substrate displayed a higher k_{cat}/K_M value towards cathepsin L than the P4 Arg derivatives, while maintaining similar catalytic parameters for the other four cathepsins. P3 was occupied by four different *D*-amino acids, among which the *D*-Thr displayed the highest affinity for cathepsin L and only residual activity in the presence of other enzymes (k_{cat}/K_M of 286 $M^{-1}s^{-1}$ for cathepsin V and 94 $M^{-1}s^{-1}$ for cathepsin B). Moreover, P2 was critical in terms of discriminating cathepsin L from other cysteine cathepsins. Mono- and di-substituted Phe are more active than 2,3,4,5,6-pentafluorophenylalanine (Phe(F_5)) in P2 on cathepsin L, however Phe(F_5) is poorly recognized by other cathepsins, rendering Phe(F_5) in the P2 position the most selective amino acid for cathepsin L.

Analysis of P1 specificity showed that various amino acids can be used to construct cathepsin L-selective substrates. Overall, the kinetic analysis performed on the second-generation cathepsin L substrates revealed that the most selective substrate from this library was Ac-His-*D*Thr-Phe(F_5)-Arg-ACC, which displayed almost non-detectable hydrolysis when tested with other cysteine cathepsins (**Figure 5**). Moreover, a detailed kinetic analysis of two cathepsin L selective substrates revealed that discrimination between cathepsin L and other cathepsins is mainly driven by k_{cat}, whereas the K_M parameter has almost no impact on selectivity (**Table S7**).

View Article Online
DOI: 10.1039/C7SC04303A



B

substrate	k_{cat}/K_M, $M^{-1}s^{-1}$ parameter				
	Cath-L	Cath-V	Cath-B	Cath-K	Cath-S
Ac-His-DVal-Phe(F_5)-Arg-ACC	747,400	539	252	< 10	< 10
P4 position					
Ac-Arg-DVal-Phe(F_5)-Arg-ACC	387,300	333	101	< 10	< 10
Ac-hArg-DVal-Phe(F_5)-Arg-ACC	454,800	647	242	96	< 10
P3 position					
Ac-His-DLeu-Phe(F_5)-Arg-ACC	405,900	353	135	< 10	< 10
Ac-His-DGln-Phe(F_5)-Arg-ACC	336,900	1,070	130	< 10	< 10
Ac-His-DThr-Phe(F_5)-Arg-ACC	729,200	286	94	< 10	< 10
Ac-His-DLys-Phe(F_5)-Arg-ACC	715,000	1,010	175	29	< 10
P2 position					
Ac-His-DVal-Phe(2-Cl)-Arg-ACC	5,689,000	11,600	1,360	13,000	2,760
Ac-His-DVal-Phe(3,4-F_2)-Arg-ACC	6,031,000	10,100	2,520	1,010	580
Ac-His-DVal-Phe(3-F)-Arg-ACC	8,432,000	12,600	6,280	5,700	814
Ac-His-DVal-Thr(Bzl)-Arg-ACC	158,520	424	480	130	< 10
Ac-His-DVal-Phe(4-F)-Arg-ACC	6,991,000	14,550	3,100	2,580	1,800
P1 position					
Ac-His-DVal-Phe(F_5)-Lys-ACC	231,000	< 10	< 10	< 10	< 10
Ac-His-DVal-Phe(F_5)-Leu-ACC	135,600	325	684	646	912
Ac-His-DVal-Phe(F_5)-hArg-ACC	317,900	175	187	< 10	< 10
Ac-His-DVal-Phe(F_5)-Glu(Bzl)-ACC	599,800	710	941	626	1,110
Ac-His-DVal-Phe(F_5)-Cys(Bzl)-ACC	732,900	1,250	1,370	190	< 10
Ac-His-DVal-Phe(F_5)-Cys(MeBzl)-ACC	1,476,000	1,740	2,290	< 10	< 10
Ac-His-DVal-Phe(F_5)-Cys(MeOBzl)-ACC	2,135,000	2,070	3,800	< 10	< 10
Ac-His-DVal-Phe(F_5)-Nle(OBzl)-ACC	2,290,000	3,820	2,500	906	508
Ac-His-DVal-Phe(F_5)-Lys(2-Cl-Z)-ACC	686,700	329	1,320	361	< 10

Figure 5 Analysis of cathepsin L-selective substrates. Panel A Outline of the two-step procedure used to develop cathepsin L-selective substrates. Initially, HyCoSuL and P1-library screening data is used to select P4-P1 cathepsin L-characteristic amino acids, which are then selected to synthesize the first generation set of substrates. After kinetic analysis of five cathepsins, the lead substrate is selected and used for second generation substrate synthesis. The second round of kinetic analysis reveals highly specific cathepsin L substrates. **Panel B** Kinetic analysis (k_{cat}/K_M) of defined second-generation cathepsin L- substrates toward five cysteine cathepsins. Experiments were triplicated and S.D. were below 15%.

3.6. Design and kinetic analysis of cathepsin ABPs. We next designed probes to track cathepsin L activity in cells by synthesizing two biotin-labeled ABPs containing the most selective cathepsin L sequence (MP-cL1 and MP-cL2, **Table S8**). Both were equipped with an electrophilic AOMK warhead. Kinetic analysis revealed that although probes are selective for cathepsin L over V, S, and K, they slightly cross-react with cathepsin B (**Figure 6**). Changing the biotin tag to cyanine-5 and the P1 Arg to Cys(Bzl) significantly improved probe activity and selectivity (MP-cL3, **Figure 6, Table S8**). We then compared compound specificity with generic (pan-specific) cathepsin probes

View Article Online
DOI: 10.1039/C7SC04303A

containing the Ala-Arg-Leu-Arg sequence by synthesizing three fluorescently-labeled probes and measuring their k_{obs}/I values in the presence of recombinant cathepsins (MP-pc1, MP-pc2 and MP-pc3, **Table S8**). All were broad spectrum probes (**Figure 6**) however, the MP-pc1 probe displayed the highest activity for all five enzymes and was thus further analyzed as a representative pan-specific cathepsin probe. Interestingly, we observed a significant difference in potency between Cy3- and Cy5-labeled probes (**Figure 6**). Since these tags have almost identical structures, we conclude that the differences are caused by the 6-ahx spacer, which may be reducing the probe's potency.

3.7. Detection of cathepsin L in cells using biotinylated probes. Biotin-tagged ABPs currently represent the gold standard in terms of biological application and have been successfully used to isolate labelled proteins by affinity enrichment/purification on appropriate beads (29, 60, 61). Previous *in vitro* kinetic assays with recombinant cathepsins show that MP-cL1 and MP-cL2 probes are potent in detecting cathepsin L and display only minor off-target activity against other cathepsins (mainly cathepsin B) (**Figure 6**). Thus we wanted to validate probe utility in living cells. Given that the biotin tag can significantly reduce permeability of a ABPs, we first investigated probes efficiency in human embryonic kidney (HEK293T) cathepsin L-transfected cells. After incubation of these probes in the 0 - 10 μM range with cells, we observed a moderately strong signal in the cathepsins MW range at probe concentrations as low as 100 nM (**Figure S2, Panel A**). The signal was increased when the probes were used at 1μM and 10μM, respectively. Moreover, there was no cathepsin labelling, when cells were preincubated with 25μM of E-64d, a cell-permeable, broad spectrum cathepsins inhibitor. Next, to assess applications of these two probes in biological studies, we tested their ability to interact with cathepsin L as a potential means to identify active proteases. To do so, we performed a pulldown assay using streptavidin-coated magnetic beads and observed strong signals corresponding to cathepsin L in the eluting fractions, signals that were completely abolished following use of the E-64d inhibitor (**Figure S2B**). Furthermore, to estimate cathepsin L/B cross-reactivity in HEK293T cells overexpressing cathepsin L, we performed a pulldown assay on streptavidin magnetic beads then performed immunoblotting of cathepsin L and B in eluting fractions. To do so, we incubated HEK293T cells overexpressing cathepsin L with both probes, prepared cell lysates, and then evaluated the enrichment of biotinylated proteins using streptavidin magnetic beads. We found that both MP-cL1 and MP-cL2 probes could be used to enrich cathepsin L, based on appearance of distinct bands in the cathepsin MW range (**Figure S2C**). Strong bands corresponding to cathepsin L were seen in eluting fractions (**Figure S2C**), but only a weak band corresponding to cathepsin B was detected in those fractions (**Figure S2D**). Based on the signal intensity, we observed only minor cross-reactivity of these probes with cathepsin B even after prolonged incubation, in agreement with the kinetic data (**Table 3**).

3.8. Cathepsin labeling by two Cy5 ABPs. Since biotin-labeled probes are not suitable for direct visualization of protease activity, we utilized the cathepsin L selective sequence (His-*D*Thr-

View Article Online
DOI: 10.1039/C7SC04303A

Phe(F_5)) to construct a fluorescently-tagged probe (MP-cL3). To confirm probe specificity (**Figure 6**) and assess activity of the MP-pc1 probe, we performed *in vitro* labeling using five human cathepsins (L, V, B, S, and K). Both probes were incubated separately at various concentrations with active site-titrated cathepsins (10 nM or 100 nM) for 30 min, followed by SDS-PAGE analysis (**Figure 6**). In both experiments, regardless of cathepsin concentration, the probe/enzyme ratio was kept constant at 0.2, 1, 5, and 25. This analysis revealed that the broad spectrum MP-pc1 probe can label cathepsins L, V, B, and S (10 nM) even at a 1:1 ratio. When cathepsins were used at the higher concentration of 100 nM this probe also labels cathepsin K. Importantly, the MP-cL3 probe displayed a very high degree of selectivity, regardless of cathepsin concentration used (**Figure 6**). Cathepsin V, the closest cathepsin L homolog, and cathepsin B are slightly labeled and only at high probe/enzyme concentrations, demonstrating that HyCoSuL screening data can be directly applied to ABP design.

A

code	sequence	Cath-L	Cath-V	Cath-B	Cath-S	Cath-K
		k_{obs}/I, $M^{-1}s^{-1}$				
MP-cL1	biotin-6-ahx-His-DThr-Phe(F_5)-Arg-AOMK	26,600	780	1,900	< 100	< 100
MP-cL2	biotin-6-ahx-His-DThr-Phe(F_5)-Lys-AOMK	10,500	360	1,250	< 100	< 100
MP-cL3	Cy5-6-ahx-His-DThr-Phe(F_5)-Cys(Bzl)-AOMK	322,000	380	780	225	< 200
MP-pc1	Cy5-Ala-Arg-Leu-Arg-AOMK	47,500	53,200	265,000	202,500	30,150
MP-pc2	$BODIPY_{FL}$-6-ahx-Ala-Arg-Leu-Arg-AOMK	8,200	9,800	23,200	18,500	350
MP-pc3	Cy3-6-ahx-Ala-Arg-Leu-Arg-AOMK	16,900	6,700	75,400	69,000	1,350

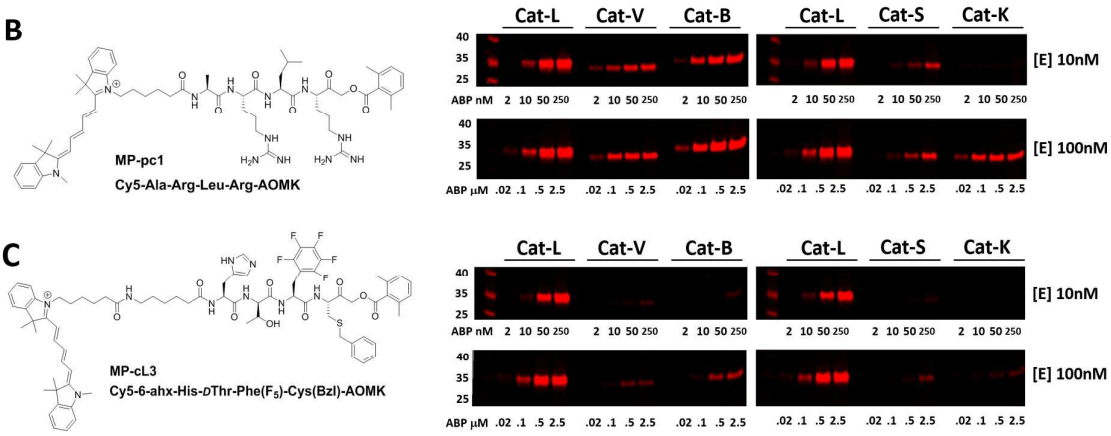


Figure 6 Broad spectrum and cathepsin L selective ABPs. Panel A Kinetic parameters (k_{obs}/I, $M^{-1}s^{-1}$) of six ABPs assessed in the presence of five recombinant human cathepsins. Experiments were triplicated and S.D. were below 20%. **Panel B,C** Labeling of recombinant cathepsins using two Cy5-labeled ABPs: a pan-cathepsin specific MP-pc1 probe (**panel B**) and the cathepsin L-specific MP-cL3 probe (**panel C**). Enzymes (10 nM or 100 nM) were incubated separately with various probe concentrations for 30 min and then subjected to SDS-PAGE analysis. Fluorescence was scanned using the 700 nm channel (LI-COR). In both experiments (enzyme at 10 nM or 100 nM) the probe/enzyme ratio was kept constant at 0.2, 1, 5, and 10.

View Article Online
DOI: 10.1039/C7SC04303A

3.9. Detection and selective labeling of cathepsin L in cells. Our ultimate goal is to develop a chemical tool capable of selective labeling active cathepsin L in cells. This task is challenging, since cathepsin B is highly expressed in numerous cells (including cancer cells) and cathepsins B and L share some substrate preferences. Our initial analyses indicate that cathepsin L can be distinguished from cathepsin B (**Figure 6**). To validate these results in a more biological setting we tested Cy5-labeled MP-pc1 and MP-cL3 using HEK293T cells, which reportedly express higher levels of cathepsin L than B [55]. To do so we incubated HEK293T cells with each probe for 0-18 hours, and in parallel we performed immunoblotting analysis. Both probes could label active cathepsins in living cells, as they were taken up within an hour of incubation into cellular lysosome/endosome compartments (**Figure S3**). Moreover, the pan-specific MP-pc1 probe labeled both cathepsins L (25kDa and 30kDa) and B (30kDa), while the HyCoSuL-derived selective MP-cL3 probe bound only cathepsin L, even after prolonged (18 hour) incubation. Interestingly, analysis using a cathepsin L antibody shows that the major form of cathepsin L in HEK293T cells is the heavy chain (25 kDa) form this was the primary form labeled by the MP-cL3 probe (**Figure S3**). The single chain (30 kDa) form, which also displays catalytic activity, was labeled by the probe (**Figure S3**).

We next challenged the selectivity of the MP-cL3 probe in MDA-MB-231 cells, similarly to HEK293T cells, since these cells contain more cathepsin B. We found that MP-cL3 is rapidly taken up by cells and detectable after 1 hour and demonstrated no labeling of cathepsin B up to 8 hours (**Figure 7, Figure S4**). Thereafter, the fluorescence signal appeared saturated, suggesting that 8 hours is sufficient to inhibit (and detect) cathepsin L activity. Prolonged (18 hour) incubation of the probe with cells resulted in non-specific but faint labeling of cathepsin B (**Figure 7A**). As all three bands were assigned to cathepsin L, or cathepsin B, this analysis also demonstrated that MP-cL3 probe does not label cathepsin V (L2), a close cathepsin L homolog. Interestingly, MDA-MB-231 cells may process cathepsin L differently than HEK293T cells, as the former do not display clear differences in levels of cathepsin L single and heavy chains. Moreover, as expected, the broad spectrum MP-pc1 probe bound mostly to cathepsin B, since cathepsin L labeling was almost negligible (**Figure 7B**).

View Article Online
DOI: 10.1039/C7SC04303A

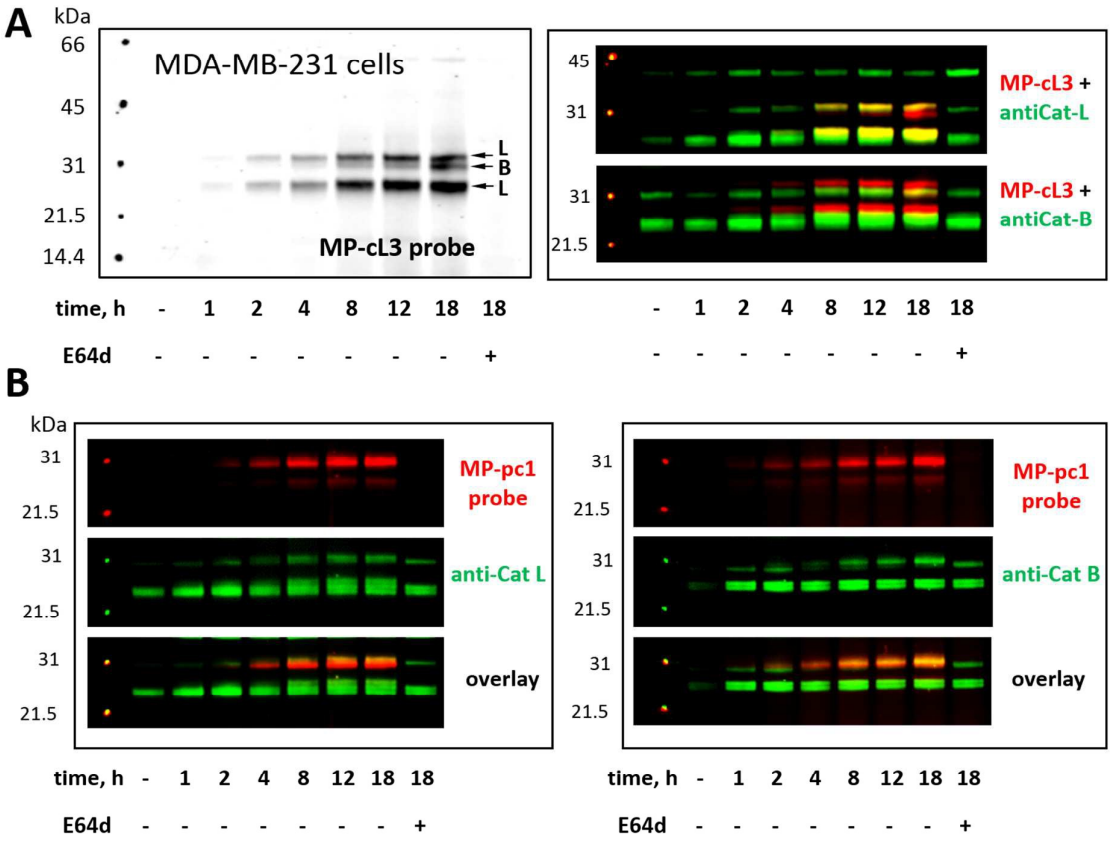


Figure 7 Cathepsin labeling in MDA-MB-231 cells using Cy5-labeled ABPs. **Panel A.** MP-cL3 begins to label active cathepsin L in living cells after 1 hour and remains selective for cathepsin L up to 8 hours (when the fluorescence signal becomes almost saturated). After 8 hours the MP-cL3 probe labels also cathepsin B; however, even after 18 hours this labeling is poor. As indicated by antibody analysis, the MP-cL3 probe labels both single chain (30kDa) and heavy chain (25kDa) forms of cathepsin L. Pre-incubation of cells with 25 μM of E64d blocked all cysteine cathepsin activity (panel A, last lane). **Panel B.** The cathepsins generic probe, MP-c1, labels mostly cathepsin B, as indicated by antibody analysis.

3.10. Detection of active cathepsin L by fluorescence microscopy. Our data indicate that we could achieve high selectivity for cathepsin L by incubating MP-cL3 in cells for up to 8 hours (**Figure 7A**). We utilized this information to visualize cathepsin L activity in MDA-MB-231 cells using confocal fluorescence microscopy. We incubated 1μM of MP-cL3 probe with MDA-MB-231 cells for 8 hours, and then methanol-fixed cells were labeled with either cathepsin L or cathepsin B antibodies for imaging (**Figure 8, Figures S5-S8**). Since the MP-cL3 is not a quenched probe (qABP), it has to be stressed here, that all non-bound probe must be washed from fixed cells prior to imaging in order to avoid false-positive signals. We found that probe signals (red) greatly overlapped with those of cathepsin L antibodies (green), and importantly, the MP-cL3 probe labeled only active cathepsin L

View Article Online
DOI: 10.1039/C7SC04303A

(**Figure 8A, Figures S5, S6**). Our Western blot analysis demonstrates that there is also a portion of inactive (pro)cathepsin L, however, we could not detect this on fluorescence microscopy as it seems that active and inactive cathepsin L share the same localization (**Figure S10**). More crucially, we detected very few red spots alone in that overlay image, strongly suggesting that there is negligible off-target binding by the MP-cL3 probe. In parallel experiments, we stained cells with cathepsin B antibody and merged images with MP-cL3 probe localization (**Figure 8B, Figures S7, S8**). The presence of overlapping signals strongly suggests that cathepsins L and B reside in the same lysosomes/endosomes; however, we observed some organelles in which only cathepsin B, but not cathepsin L, is active (**Figure 8B, Figures S7, S8**). Additional analysis confirmed that prolonged incubation of the MP-cL3 probe with MDA-MB-231 cells (24 hours) results in additional labeling likely due to other cathepsins, confirming the MP-cL3 probe becomes non-selective after 8 hours (**Figures S9, S10).** Thus, the probe concentration and time of incubation with cells are both critical to achieve selectivity in cells. Finally, we also incubated MDA-MB-231 cells with the broad spectrum MP-pc1 probe (1μM, 8 hours) and found that it labeled primarily cathepsin B, since the degree of probe/enzyme overlap was higher than for cathepsin L enzyme (**Figure S11**).

View Article Online
DOI: 10.1039/C7SC04303A

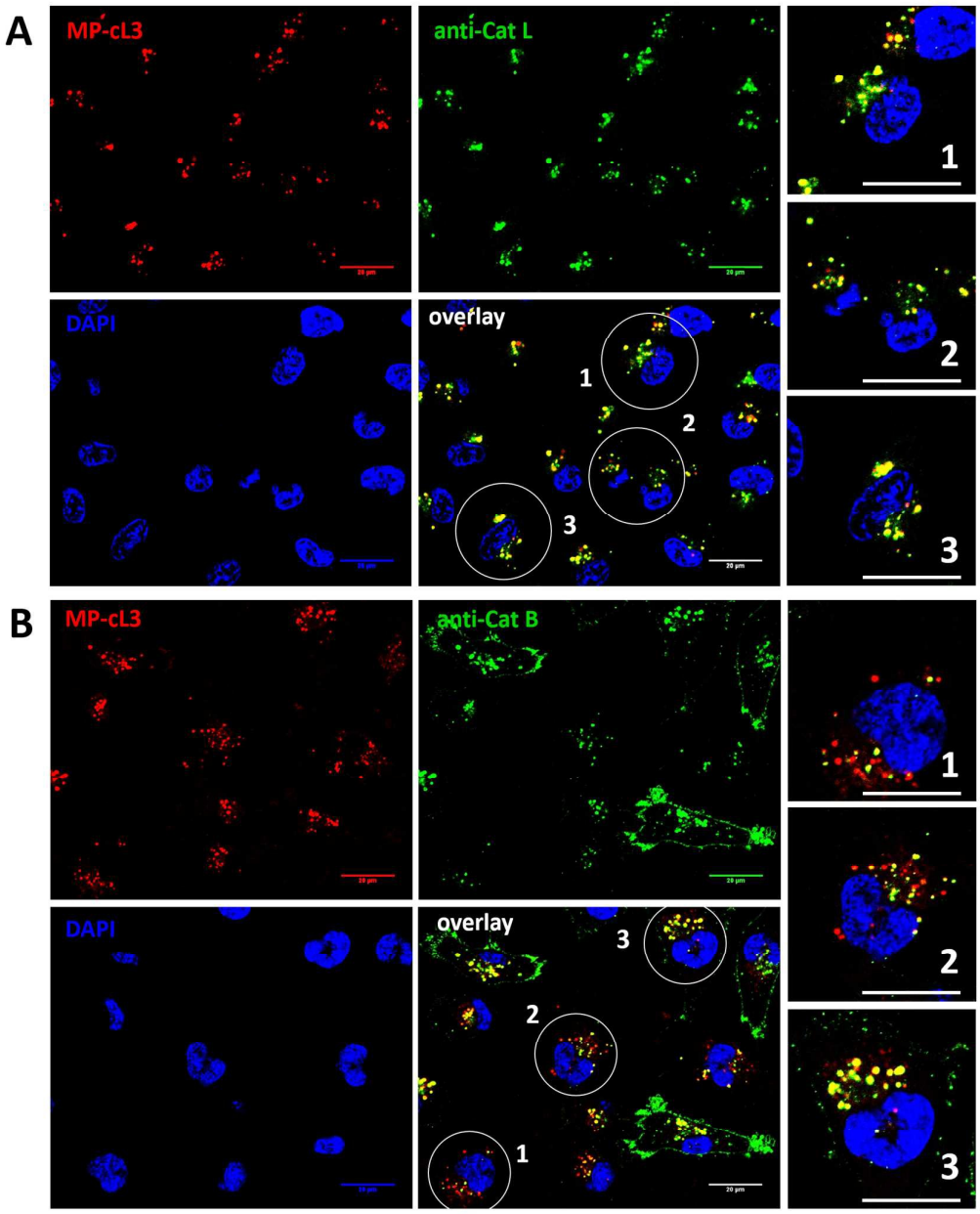


Figure 8 Cathepsin L labeling of MDA-MB-231 cells. **Panel A.** The cathepsin L selective probe MP-cL3 (1 μM) was incubated with MDA-MB-231 cells for 8 hours and cells were then subjected for confocal fluorescence microscopy. Cathepsin L antibody detects both the proenzyme and active enzyme; however, Cy5 detects only active cathepsin L. In the overlay, green spots indicate inactive enzyme, as yellow show only active cathepsin L. The lack of red spots in the overlay image strongly indicates high selectivity of the MP-cL3 probe. **Panel B.** The same experiment as in Panel A, but here localization of the MP-cL3 probe was paired with staining with a cathepsin B antibody. The results demonstrate that cathepsin B is not distributed equally with cathepsin L, as there are lysosomes/endosomes where both enzymes are active, and others in which only active cathepsin B is present. Both experiments (Panel A and B) were performed three times and representative pictures are presented. Scale bar 20 μm.

View Article Online
DOI: 10.1039/C7SC04303A

3.11. Colocalization between cathepsin L, cathepsin B and MP-cL3 ABP. To investigate the colocalization of MP-cL3 with cathepsin L protein in MDA-MB-231 cells we performed a quantitative pixel-by-pixel analysis of multiple fluorescence microscopy images. A reliable parameter for such analysis is a weighted colocalization coefficient (wcc), which provides information not only about single pixels overlapping, but also correlates their intensities, making the analysis more accurate (62, 63). For this analysis we used: (1) MP-cL3 probe (selective labeling – 8 hours) + anti-cathepsin L (8 images, 98 single cells analyzed), (2) MP-cL3 probe (non-selective labeling – 24 hours) + anti-cathepsin L (8 images, 59 single cells analyzed) and (3) MP-cL3 probe (selective labeling – 8 hours) + anti-cathepsin B (7 images, 68 single cells analyzed) (**Figure 9, Figure S12**). We demonstrated that the highest weighted colocalization coefficient was obtained when MP-cL3 probe was incubated with MDA-MB-231 cells for 8 hours (wwc = 0.937 and 0.928) (**Figure S5**). Prolonged incubation of this probe with cells (24 hours) resulted in a decline of this parameter (wwc = 0.725 and 0.695) (**Figure S9**). Colocalization of MP-cL3 probe with cathepsin B antibody resulted in the lowest wwc value (0.646 and 0.635) (**Figure S7**). To further demonstrate the usefulness of the wwc parameter we constructed histograms from randomly selected lines from the microscope images, and overlaid probe signal (red) with cathepsin L either cathepsin B antibody signal (green) (**Figures S6 and S8**). Histograms for cathepsin L antibody (green) and MP-cL3 probe (red) overlapped very well (**Figure S6B**), whereas histograms for cathepsin B antibody and MP-cL3 matched poorly (**Figure S8B**).

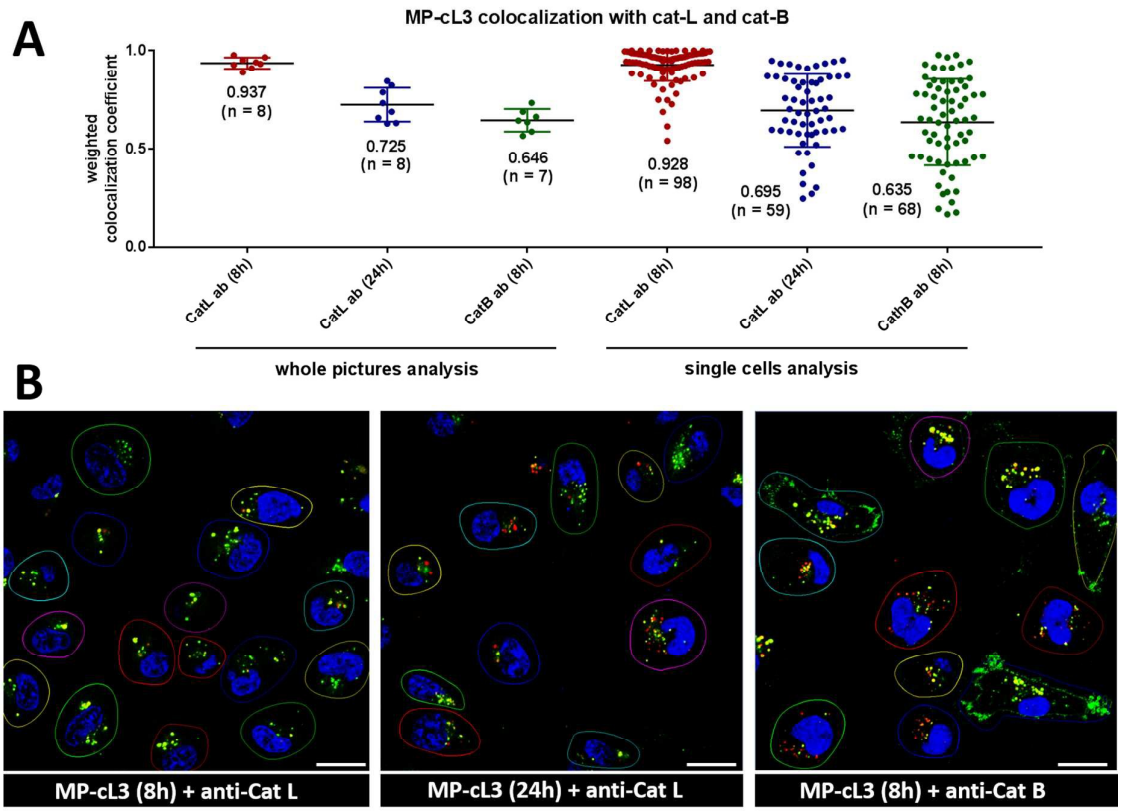


Figure 9 Colocalization of MP-cL3 probe with catepsin L and cathepsin B in MDA-MB-231 cancer cells. **Panel A** Weighted colocalization coefficients calculated for ABP/cat-L and ABP/cat-B based on (1) whole picture analysis, or (2) single cell analysis. Points on the graph indicate single pictures (left) or single cells (right); with "n" defining the number of pictures/single cells taken for analysis. Numbers in brackets on the x-axis indicate the incubation time of MP-cL3 with MDA-MB-231 cells. **Panel B** Representative images used for the

View Article Online
DOI: 10.1039/C7SC04303A

calculation of weighted colocalization coefficients between MP-cL3 probe and cathepsin L (or B) antibody. Red is ABP, green is cathepsin antibody, and blue is DAPI. Scale bar is 20 μm.

4. CONCLUSIONS

Development of protease-specific substrates offering direct readout (fluorogenic or chromogenic) is of a great significance, since such tools could be extremely useful in biochemical and biological studies of proteases. Starting from broad spectrum chromogenic substrates developed in the early 1940`s and 1950`s, the field has progressed through individual internally quenched fluorescence substrate libraries, and combinatorial mixture libraries with natural amino acids (8, 64-67). However, most of these tools are insufficient to develop selective substrates to distinguish closely-related proteases that display overlapping substrate specificities (among them, caspases, matrix metalloproteases, or cysteine cathepsins). To address this challenge, we applied HyCoSuL technology (40, 50, 54), because the use of such diverse chemical structures in a library allowed discovery of highly selective cathepsin L substrates and probes. Combined analysis of enzyme preferences in the P4-P1 region allowed us to design an extremely selective tetrapeptide substrate (Ac-His-*D*Thr-Phe(F_5)-Arg-ACC). Notably, given that we found that selectivity towards cathepsin L lies in the P4-P2 sequence, the Arg at position P1 could be exchanged for non-charged amino acids to (1) make the substrate and inhibitor more active in the presence of cathepsin L and (2) to avoid potential cross-reactivity with the numerous proteases that prefer Arg at P1. We then synthesized three small molecule probes, with MP-cL3 (Cy5-6-ahx-His-*D*Thr-Phe(F_5)-Cys(Bzl)-AOMK) being the most potent and selective toward cathepsin L. This probe bound exclusively to cathepsin L in an *in vitro* experiment testing five recombinant cysteine cathepsins. To demonstrate that the enzymes were active, we also used a broad spectrum cathepsin probe, MP-pc1 (Cy5-Ala-Arg-Leu-Arg-AOMK). More importantly, we showed that the MP-cL3 probe selectively labels cathepsin L in HEK293T and MDA-MB-231 cell lines, both of which express cathepsins L and B at different levels. MP-cL3 probe selectivity over cathepsin B is of special importance, as cathepsin B is overexpressed in many types of cancers and masks cathepsin L proteolytic activity, making its analysis more complex. Finally, using the MP-cL3 probe, we not only showed specific labeling of cathepsin L in cells but performed quantitative pixel-by-pixel localization studies using fluorescence microscopy. That analysis indicated that cathepsin L co-localizes with cathepsin B in MDA-MB-231 cancer cells and that some cathepsin B-rich lysosomes lack cathepsin L. We also prepared two biotinylated cathepsin L probes that have only minor cross-reactivity with cathepsin B and could serve as tools in a diverse biological experiments. In summary, we have developed the first selective cathepsin L ABP, one that allows accurate detection of cathepsin L activity in cancer cells. Given that this probe does not react with

View Article Online
DOI: 10.1039/C7SC04303A

other cathepsins, it has potential to facilitate understanding of cathepsin L function and provide new information about the biological framework of cathepsin L-dependent or -related cancer progression.

CONFLICT OF INTEREST

There are no conflicts to declare

ACKNOWLEDGMENTS

This work was supported by the National Science Center in Poland (grant Preludium, UMO-2013/09/N/ST5/02448 to MP), National Institutes of Health in the USA (grant GM99040 to GSS). This project has received founding from the European Union`s Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 661187. B.T. and D.T were supported by the grants P1-0140 and P1-0048 by the Slovene Research Agency. The Drag laboratory is supported by the Foundation for Polish Science.

AUTHOR CONTRIBUTIONS

M.P., G.S.S., and M.D. designed research; M.P., W.R., M.V., K.G., and P.K. performed research and collected data; D.F., K.V., D.T., and B.T contributed new reagents and enzymes; M.P., G.S.S., and M.D. analysed and interpreted the data, and wrote the manuscript; W.R., M.V., K.G., P.K., D.F., K.V., D.T, and B.T. critically revised the manuscript.

REFERENCES

1. Rawlings ND, Barrett AJ, Finn R. Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Res. 2016;44(D1):D343-50.
2. Kramer L, Turk D, Turk B. The Future of Cysteine Cathepsins in Disease Management. Trends Pharmacol Sci. 2017.
3. Turk V, Stoka V, Vasiljeva O, Renko M, Sun T, Turk B, et al. Cysteine cathepsins: from structure, function and regulation to new frontiers. Biochim Biophys Acta. 2012;1824(1):68-88.
4. Vasiljeva O, Reinheckel T, Peters C, Turk D, Turk V, Turk B. Emerging roles of cysteine cathepsins in disease and their potential as drug targets. Curr Pharm Des. 2007;13(4):387-403.
5. Menard R, Carmona E, Takebe S, Dufour E, Plouffe C, Mason P, et al. Autocatalytic processing of recombinant human procathepsin L. Contribution of both intermolecular and unimolecular events in the processing of procathepsin L in vitro. J Biol Chem. 1998;273(8):4478-84.
6. Turk B, Dolenc I, Lenarcic B, Krizaj I, Turk V, Bieth JG, et al. Acidic pH as a physiological regulator of human cathepsin L activity. Eur J Biochem. 1999;259(3):926-32.
7. Turk B, Dolenc I, Turk V, Bieth JG. Kinetics of the pH-induced inactivation of human cathepsin L. Biochemistry. 1993;32(1):375-80.

View Article Online
DOI: 10.1039/C7SC04303A

8. Choe Y, Leonetti F, Greenbaum DC, Lecaille F, Bogyo M, Bromme D, et al. Substrate profiling of cysteine proteases using a combinatorial peptide library identifies functionally unique specificities. J Biol Chem. 2006;281(18):12824-32.
9. Nakagawa T, Roth W, Wong P, Nelson A, Farr A, Deussing J, et al. Cathepsin L: critical role in Ii degradation and CD4 T cell selection in the thymus. Science. 1998;280(5362):450-3.
10. Duncan EM, Muratore-Schroeder TL, Cook RG, Garcia BA, Shabanowitz J, Hunt DF, et al. Cathepsin L proteolytically processes histone H3 during mouse embryonic stem cell differentiation. Cell. 2008;135(2):284-94.
11. Goulet B, Baruch A, Moon NS, Poirier M, Sansregret LL, Erickson A, et al. A cathepsin L isoform that is devoid of a signal peptide localizes to the nucleus in S phase and processes the CDP/Cux transcription factor. Mol Cell. 2004;14(2):207-19.
12. Reinheckel T, Hagemann S, Dollwet-Mack S, Martinez E, Lohmuller T, Zlatkovic G, et al. The lysosomal cysteine protease cathepsin L regulates keratinocyte proliferation by control of growth factor recycling. J Cell Sci. 2005;118(Pt 15):3387-95.
13. Olson OC, Joyce JA. Cysteine cathepsin proteases: regulators of cancer progression and therapeutic response. Nat Rev Cancer. 2015;15(12):712-29.
14. Reiser J, Adair B, Reinheckel T. Specialized roles for cysteine cathepsins in health and disease. J Clin Invest. 2010;120(10):3421-31.
15. Mohamed MM, Sloane BF. Cysteine cathepsins: multifunctional enzymes in cancer. Nat Rev Cancer. 2006;6(10):764-75.
16. Zheng X, Chu F, Chou PM, Gallati C, Dier U, Mirkin BL, et al. Cathepsin L inhibition suppresses drug resistance in vitro and in vivo: a putative mechanism. Am J Physiol Cell Physiol. 2009;296(1):C65-74.
17. Frade R, Rousselet N, Jean D. Intratumoral gene delivery of anti-cathepsin L single-chain variable fragment by lentiviral vector inhibits tumor progression induced by human melanoma cells. Cancer Gene Ther. 2008;15(9):591-604.
18. Yagel S, Warner AH, Nellans HN, Lala PK, Waghorne C, Denhardt DT. Suppression by cathepsin L inhibitors of the invasion of amnion membranes by murine cancer cells. Cancer Res. 1989;49(13):3553-7.
19. Dennemarker J, Lohmuller T, Mayerle J, Tacke M, Lerch MM, Coussens LM, et al. Deficiency for the cysteine protease cathepsin L promotes tumor progression in mouse epidermis. Oncogene. 2010;29(11):1611-21.
20. Fonovic M, Turk B. Cysteine cathepsins and extracellular matrix degradation. Biochim Biophys Acta. 2014;1840(8):2560-70.
21. Sobotic B, Vizovisek M, Vidmar R, Van Damme P, Gocheva V, Joyce JA, et al. Proteomic Identification of Cysteine Cathepsin Substrates Shed from the Surface of Cancer Cells. Mol Cell Proteomics. 2015;14(8):2213-28.
22. Adachi W, Kawamoto S, Ohno I, Nishida K, Kinoshita S, Matsubara K, et al. Isolation and characterization of human cathepsin V: a major proteinase in corneal epithelium. Invest Ophthalmol Vis Sci. 1998;39(10):1789-96.
23. Santamaria I, Velasco G, Cazorla M, Fueyo A, Campo E, Lopez-Otin C. Cathepsin L2, a novel human cysteine proteinase produced by breast and colorectal carcinomas. Cancer Res. 1998;58(8):1624-30.
24. Tolosa E, Li W, Yasuda Y, Wienhold W, Denzin LK, Lautwein A, et al. Cathepsin V is involved in the degradation of invariant chain in human thymus and is overexpressed in myasthenia gravis. J Clin Invest. 2003;112(4):517-26.
25. Ong PC, McGowan S, Pearce MC, Irving JA, Kan WT, Grigoryev SA, et al. DNA accelerates the inhibition of human cathepsin V by serpins. J Biol Chem. 2007;282(51):36980-6.
26. Vizovisek M, Vidmar R, Van Quickelberghe E, Impens F, Andjelkovic U, Sobotic B, et al. Fast profiling of protease specificity reveals similar substrate specificities for cathepsins K, L and S. Proteomics. 2015;15(14):2479-90.

View Article Online
DOI: 10.1039/C7SC04303A

27. Vidmar R, Vizovisek M, Turk D, Turk B, Fonovic M. Protease cleavage site fingerprinting by label-free in-gel degradomics reveals pH-dependent specificity switch of legumain. EMBO J. 2017;36(16):2455-65.
28. Turk B, Turk D, Turk V. Protease signalling: the cutting edge. EMBO J. 2012;31(7):1630-43.
29. Sanman LE, Bogyo M. Activity-based profiling of proteases. Annu Rev Biochem. 2014;83:249-73.
30. Poreba M, Solberg R, Rut W, Lunde NN, Kasperkiewicz P, Snipas SJ, et al. Counter Selection Substrate Library Strategy for Developing Specific Protease Substrates and Probes. Cell Chem Biol. 2016;23(8):1023-35.
31. Zhou S, Lin J, Du W, Zhang Z, Luo Q, Liu BF, et al. Characterization of proteinase activation dynamics by capillary electrophoresis conjugating with fluorescent protein-based probe. J Chromatogr B Analyt Technol Biomed Life Sci. 2006;844(1):158-62.
32. Tchoupe JR, Moreau T, Gauthier F, Bieth JG. Photometric or fluorometric assay of cathepsin B, L and H and papain using substrates with an aminotrifluoromethylcoumarin leaving group. Biochim Biophys Acta. 1991;1076(1):149-51.
33. Higashi T, Hashimoto M, Watanabe M, Yamauchi Y, Fujiwara M, Nakatsukasa H, et al. Assay procedures for cathepsin B, H and L activities in rat tissue homogenates. Acta Med Okayama. 1986;40(1):27-32.
34. Creasy BM, Hartmann CB, White FK, McCoy KL. New assay using fluorogenic substrates and immunofluorescence staining to measure cysteine cathepsin activity in live cell subpopulations. Cytometry A. 2007;71(2):114-23.
35. Watzke A, Kosec G, Kindermann M, Jeske V, Nestler HP, Turk V, et al. Selective activity-based probes for cysteine cathepsins. Angew Chem Int Ed Engl. 2008;47(2):406-9.
36. Hu HY, Vats D, Vizovisek M, Kramer L, Germanier C, Wendt KU, et al. In vivo imaging of mouse tumors by a lipidated cathepsin S substrate. Angew Chem Int Ed Engl. 2014;53(29):7669-73.
37. Fonovic M, Bogyo M. Activity based probes for proteases: applications to biomarker discovery, molecular imaging and drug screening. Curr Pharm Des. 2007;13(3):253-61.
38. Sadaghiani AM, Verhelst SH, Bogyo M. Tagging and detection strategies for activity-based proteomics. Curr Opin Chem Biol. 2007;11(1):20-8.
39. Kam CM, Abuelyaman AS, Li Z, Hudig D, Powers JC. Biotinylated isocoumarins, new inhibitors and reagents for detection, localization, and isolation of serine proteases. Bioconjug Chem. 1993;4(6):560-7.
40. Kasperkiewicz P, Poreba M, Snipas SJ, Parker H, Winterbourn CC, Salvesen GS, et al. Design of ultrasensitive probes for human neutrophil elastase through hybrid combinatorial substrate library profiling. Proc Natl Acad Sci U S A. 2014;111(7):2518-23.
41. Liu Y, Patricelli MP, Cravatt BF. Activity-based protein profiling: the serine hydrolases. Proc Natl Acad Sci U S A. 1999;96(26):14694-9.
42. 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(10):668-77.
43. Tully SE, Cravatt BF. Activity-based probes that target functional subclasses of phospholipases in proteomes. J Am Chem Soc. 2010;132(10):3264-5.
44. Bogyo M, Verhelst S, Bellingard-Dubouchaud V, Toba S, Greenbaum D. Selective targeting of lysosomal cysteine proteases with radiolabeled electrophilic substrate analogs. Chem Biol. 2000;7(1):27-38.
45. Paulick MG, Bogyo M. Development of activity-based probes for cathepsin X. ACS Chem Biol. 2011;6(6):563-72.
46. Edgington-Mitchell LE, Bogyo M, Verdoes M. Live Cell Imaging and Profiling of Cysteine Cathepsin Activity Using a Quenched Activity-Based Probe. Methods Mol Biol. 2017;1491:145-59.
47. Ofori LO, Withana NP, Prestwood TR, Verdoes M, Brady JJ, Winslow MM, et al. Design of Protease Activated Optical Contrast Agents That Exploit a Latent Lysosomotropic Effect for Use in Fluorescence-Guided Surgery. ACS Chem Biol. 2015;10(9):1977-88.

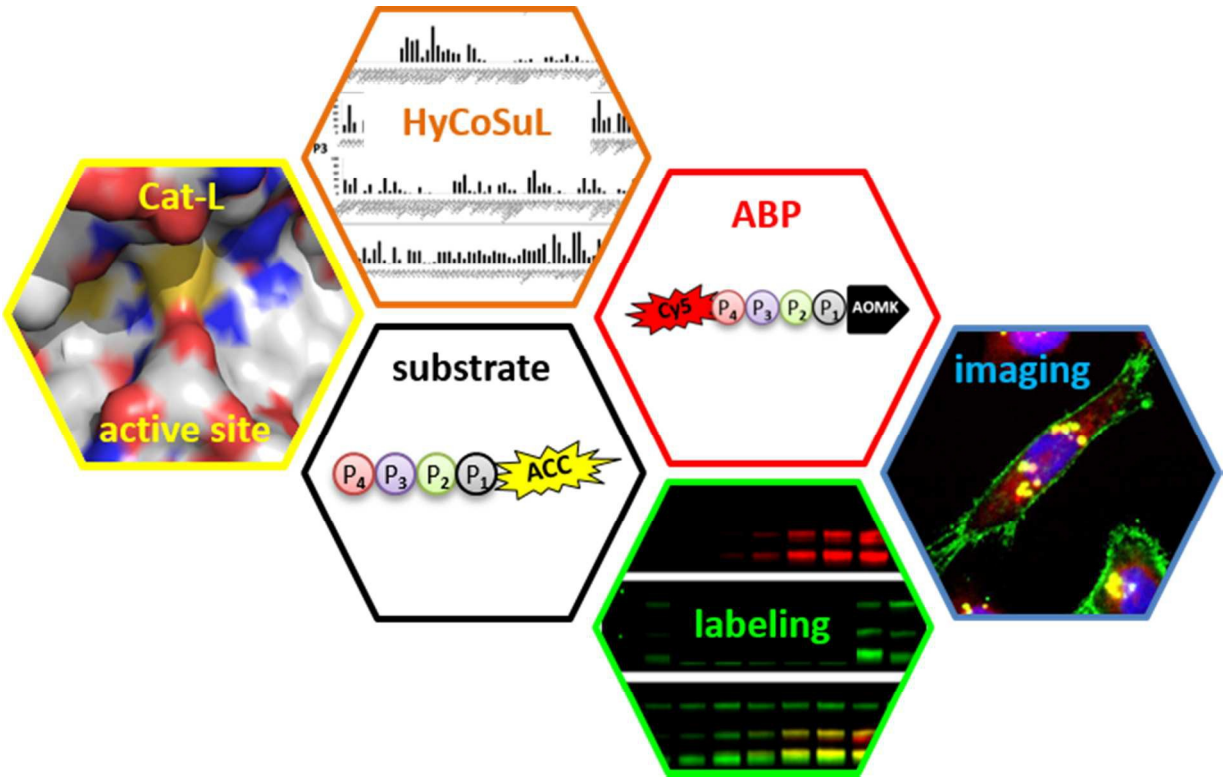
View Article Online
DOI: 10.1039/C7SC04303A

48. Oresic Bender K, Ofori L, van der Linden WA, Mock ED, Datta GK, Chowdhury S, et al. Design of a highly selective quenched activity-based probe and its application in dual color imaging studies of cathepsin S activity localization. J Am Chem Soc. 2015;137(14):4771-7.
49. Verdoes M, Edgington LE, Scheeren FA, Leyva M, Blum G, Weiskopf K, et al. A nonpeptidic cathepsin S activity-based probe for noninvasive optical imaging of tumor-associated macrophages. Chem Biol. 2012;19(5):619-28.
50. Poreba M, Kasperkiewicz P, Snipas SJ, Fasci D, Salvesen GS, Drag M. Unnatural amino acids increase sensitivity and provide for the design of highly selective caspase substrates. Cell Death Differ. 2014;21(9):1482-92.
51. Rut W, Zhang L, Kasperkiewicz P, Poreba M, Hilgenfeld R, Drag M. Extended substrate specificity and first potent irreversible inhibitor/activity-based probe design for Zika virus NS2B-NS3 protease. Antiviral Res. 2017;139:88-94.
52. Mihelic M, Dobersek A, Guncar G, Turk D. Inhibitory fragment from the p41 form of invariant chain can regulate activity of cysteine cathepsins in antigen presentation. J Biol Chem. 2008;283(21):14453-60.
53. Bromme D, Nallaseth FS, Turk B. Production and activation of recombinant papain-like cysteine proteases. Methods. 2004;32(2):199-206.
54. Kasperkiewicz P, Poreba M, Snipas SJ, Lin SJ, Kirchhofer D, Salvesen GS, et al. Design of a Selective Substrate and Activity Based Probe for Human Neutrophil Serine Protease 4. PLoS One. 2015;10(7):e0132818.
55. Poreba M, Szalek A, Kasperkiewicz P, Drag M. Positional scanning substrate combinatorial library (PS-SCL) approach to define caspase substrate specificity. Methods Mol Biol. 2014;1133:41-59.
56. Poreba M, Szalek A, Rut W, Kasperkiewicz P, Rutkowska-Wlodarczyk I, Snipas SJ, et al. Highly sensitive and adaptable fluorescence-quenched pair discloses the substrate specificity profiles in diverse protease families. Sci Rep. 2017;7:43135.
57. Harris JL, Backes BJ, Leonetti F, Mahrus S, Ellman JA, Craik CS. Rapid and general profiling of protease specificity by using combinatorial fluorogenic substrate libraries. Proc Natl Acad Sci U S A. 2000;97(14):7754-9.
58. Biniossek ML, Nagler DK, Becker-Pauly C, Schilling O. Proteomic identification of protease cleavage sites characterizes prime and non-prime specificity of cysteine cathepsins B, L, and S. J Proteome Res. 2011;10(12):5363-73.
59. Thornberry NA, Rano TA, Peterson EP, Rasper DM, Timkey T, Garcia-Calvo M, et al. A combinatorial approach defines specificities of members of the caspase family and granzyme B. Functional relationships established for key mediators of apoptosis. J Biol Chem. 1997;272(29):17907-11.
60. Wright MH, Sieber SA. Chemical proteomics approaches for identifying the cellular targets of natural products. Nat Prod Rep. 2016;33(5):681-708.
61. Speers AE, Cravatt BF. Activity-Based Protein Profiling (ABPP) and Click Chemistry (CC)-ABPP by MudPIT Mass Spectrometry. Curr Protoc Chem Biol. 2009;1:29-41.
62. Manders EMM, Verbeek FJ, Aten JA. Measurement of co-localization of objects in dual-colour confocal images. Journal of Microscopy. 1993;169(3):375-82.
63. Adler J, Parmryd I. Quantifying colocalization by correlation: the Pearson correlation coefficient is superior to the Mander's overlap coefficient. Cytometry A. 2010;77(8):733-42.
64. Katunuma N, Towatari T, Tamai M, Hanada K. Use of new synthetic substrates for assays of cathepsin L and cathepsin B. J Biochem. 1983;93(4):1129-35.
65. Portaro FC, Santos AB, Cezari MH, Juliano MA, Juliano L, Carmona E. Probing the specificity of cysteine proteinases at subsites remote from the active site: analysis of P4, P3, P2' and P3' variations in extended substrates. Biochem J. 2000;347 Pt 1:123-9.
66. Legowska M, Wysocka M, Burster T, Pikula M, Rolka K, Lesner A. Ultrasensitive internally quenched substrates of human cathepsin L. Anal Biochem. 2014;466:30-7.

View Article Online
DOI: 10.1039/C7SC04303A

67. Puzer L, Cotrin SS, Alves MF, Egborge T, Araujo MS, Juliano MA, et al. Comparative substrate specificity analysis of recombinant human cathepsin V and cathepsin L. Arch Biochem Biophys. 2004;430(2):274-83.

View Article Online
DOI: 10.1039/C7SC04303A



Highly-selective fluorogenic substrate and activity-based probe for monitoring cathepsin L activity in the breast cancer cell line MDA-MB-231