Published in final edited form as:

Curr Top Med Chem. 2009 ; 9(13): 1206–1216.

The Identification, Characterization and Optimization of Small Molecule Probes of Cysteine Proteases: Experiences of the Penn Center for Molecular Discovery with Cathepsin B and Cathepsin L

Donna M. Huryn^{*} and Amos B. Smith III

Department of Chemistry, University of Pennsylvania, 231 South 34th Street, Philadelphia, PA 19104-6323, USA, Penn Center for Molecular Discovery, University of Pennsylvania, 1024 Vagelos Research Laboratories, Philadelphia, PA 19104-6383, USA

Abstract

During the pilot phase of the NIH Molecular Library Screening Network, the Penn Center for Molecular Discovery focused on a series of projects aimed at high throughput screening and the development of probes of a variety of protease targets. This review provides our medicinal chemistry experience with two such targets – cathepsin B and cathepsin L. We describe our approach for hit validation, characterization and triage that led to a critical understanding of the nature of hits from the cathepsin B project. In addition, we detail our experience at hit identification and optimization that led to the development of a novel thiocarbamate probe of cathepsin L.

Keywords

cathepsin B; cathepsin L; cysteine proteases; DTT; false positives; hit validation; hit triage; redox cycling; thiocarbamates

Introduction

As part of the Molecular Libraries Screening Center Network (MLSCN),^{i,ii} the Penn Center for Molecular Discovery (PCMD)ⁱⁱⁱ undertook a series of high throughput screening (HTS) campaigns targeting a variety of proteases, particularly members of the papain-class of cysteine proteases. Completion of the screens was followed by hit evaluation, characterization and triage, and when appropriate, medicinal chemistry optimization to improve potency and/or properties to develop small molecule probes of these proteases. This effort had several goals. First, our work would produce a sizeable data set thoroughly profiling a large compound library in a wide number of protease assays. While numerous inhibitors of proteases have been described in the literature,^{iv} the published data on these compounds typically includes the specific target activity along with selectivity against a handful of genomically-related or pharmacologically and toxicologically relevant enzymes. Second, this effort would contribute to the NIH Roadmap Initiative to annotate the NIH Small Molecule Repository (NIH SMR), by making data from high throughput screening campaigns on biomedical targets, physicochemical characterization assays, and toxicity profiling publicly available via PubChem.^v Finally, we were particularly interested in

^{*}Corresponding author: Huryn@sas.upenn.edu.

following up reports on the essential roles that cathepsin B and L play in allowing entry of SARS-coronavirus and Ebola viruses into host cells.^{vi} We sought potent, selective, cell-permeable inhibitors that would allow further elucidation of this mechanism, and might serve as starting points for drug discovery efforts. In this review, we describe the PCMD's medicinal chemistry experience in identification, characterization and triage of hits from high throughput screens of the NIH SMR for inhibitors of cathepsin B and cathepsin L, as well as the successful development of novel, potent and selective probes of cathepsin L.

Cathepsin B

Cathepsin B is a lysosomal cysteine protease that has been implicated in a number of pathological conditions^{vii} such as cancer,^{viii} immunological and inflammatory disorders,^{ix,x} Alzheimer's disease,^{xi} and infectious diseases.^{vi} As such, the identification of peptidic and small molecule inhibitors of Cathepsin B has been an active area of research, however, to date, no compounds have been clinically approved as therapeutics.^{ivc,xii}

High Throughput Screen and Triage of Cathepsin B Hits: Pyrimidotriazine-diones and Pyrazole sulfonamides

The PCMD screened approximately 63,000 NIH SMR samples for inhibition of human liver cathepsin B. The assay incorporated a modified aminomethylcoumarin dipeptide substrate, which upon cleavage by cathepsin B generated a fluorescent signal. Seventy-five (75) compounds inhibited cathepsin B by at least 20% at 10 μ M and were characterized as "primary hits."^{xiii} Upon re-testing using fresh samples, 37 of the initial 75 hits were confirmed in a dose-response assay.

Pyrimidotriazine-diones

Of the 37 confirmed hits, three structurally related analogs, CID 653297, CID 64750 and CID 66541/Toxoflavin (Figure 1) displayed significantly more potent activity than others. These three exhibited IC₅₀'s of 72nM, 71nM and 46nM, respectively. However, the relatively flat structure of the pyrimidotriazine-diones raised concerns, as did the lack of any structural similarity to other prototypical cysteine protease inhibitors. An analysis of the compounds' biological activity in Pubchem raised further flags. At the time of this work, according to Pubchem, these compounds had been assayed in thirteen MLSCN assays and were designated "active" in half of them. Most, but not all, of these assays targeted phosphatases; a second common factor was the presence of Dithiothreitol (DTT) in the assay medium, most probably to maintain the enzyme targets in a reduced, active state. Further literature searching led to reports describing pyrimidotriazines' indirect mechanism of inhibition of caspases^{xiv} and phosphatases.^{xv} The authors demonstrated that compounds of this class exert their activity through reaction with the reducing agent (e.g., DTT) present in the assay medium, rather than via direct binding and inhibition of the enzyme. Reaction of the putative inhibitors with DTT generated hydrogen peroxide (H₂O₂), which oxidized the active site cysteine and caused apparent inhibition.

To confirm the hypothesis that these hits were acting through a false positive mechanism, the cathepsin B assay was performed using cysteine, a less powerful reductant, in place of DTT. Under these new conditions, the pyrimidotriazine-diones, along with several structurally unrelated compounds, were inactive.^{xvi} This result confirmed that these compounds were acting through a DTT-dependent redox cycling mechanism, rather than through direct inhibition of the enzyme. While there have been sporadic reports describing redox cycling as a mechanism of false positives in enzyme assays, this promiscuous activity has only recently garnered further attention with several groups developing high throughput

assays to detect compounds which act through this mechanism.^{xvii,xviii} It is hoped that the availability of these assays, in conjunction with those to detect other undesirable or false positive mechanisms of action such as aggregation,^{xix} fluorescence interference^{xx} and chemical reactivity^{xxi} will significantly improve the efficiency of the HTS triage process and perhaps allow the development of predictive models of false positive behavior.^{xixa, xxii}

This pyrimidotriazine-dione class of compounds appears to be generally promiscuous through redox cycling, and perhaps through other, as yet unknown, mechanisms. It has also been reported to interfere with fluorescence detection methods.^{xxa} According to Pubchem, as of July 2009, CID 653297 has been designated as “active” in 81 out of the 276 assays in which it has been evaluated. While many of these assays include DDT or other reductants as part of the media, some do not, suggesting additional mechanisms of promiscuity. This collective data supports the removal of these compounds from small molecule screening libraries, or at least their careful and complete characterization when designated as “active.”

Pyrazole sulfonamides

After elimination of redox-active false positives, and other nuisance compounds, a class of pyrazole sulfonamides of general structure **1** (Figure 2) appeared initially promising as cathepsin B inhibitors. Eight analogs in the screening library exhibited inhibitory activity ranging from 250nM to 2 μ M, suggesting a relationship between structure and potency. Furthermore the compounds were deemed “drug-like” and had acceptable Lipinski parameters^{xxiii} (MW ~350–380, logP~3, HBD=2, HBA = 6–9). We postulated that if appropriately reactive, the ester carbonyl in **1** could provide an electrophilic site for reaction with the active site cysteine in cathepsin B, in analogy to most other known warhead-containing cathepsin inhibitors.^{iva} We pursued this series further based on its potency and excellent properties, as well as the potential to identify novel chemotypes that inhibit cathepsin B.

As is our practice, we synthesized a representative set of the actives in order to assure that the structure depicted in the NIH SMR was correctly assigned, and to provide sufficient material for follow-up studies. The synthesis, shown in general terms in Scheme 1,^{xxiv} was based on several literature reports^{xxv, xxvi, xxvii, xxviii} and proceeded over three steps and in moderate yields. The spectroscopic characteristics were identical to those of the original screening samples, and the structure of **1** (R₁=*p*-fluorophenyl; R₂=2-thiophene) was confirmed by x-ray crystallography. The newly prepared samples were then re-tested in the cathepsin B assay and their activity confirmed.

Further biochemical studies^{xxiv} however, indicated that these compounds were acting as competitive substrates, rather than as conventional inhibitors of the enzyme. Data to support this conclusion included biochemical experiments that demonstrated the inhibition to be competitive and reversible. This characterization differs from most known cathepsin B inhibitors, which are irreversible or only slowly reversible.^{ivc} Most compelling, however, were the results of studies that monitored the products of a stoichiometric reaction between cathepsin B and Compound **1a** (Scheme 2). After 15 minutes under those conditions, 0% of Compound **1a** remained; only the hydrolysis product, pyrazalone **4a**, was detected.^{xxiv}

Limited efforts to convert this substrate into a competitive inhibitor by eliminating the ester function were pursued. Towards that end, the pyrazaolone **4** (R₁= *p*-methoxyphenyl) was used as a pivotal intermediate to generate compounds such as ether **5** and sulfonate **6** (Figure 3). Both compounds were inactive in the cathepsin B assay, indicating that reactivity of the ester function group, and stability of the pyrazolone product, rather than binding energy was the major contributor to apparent potency.^{xxiv}

In the course of evaluating how modifications of the ester substituent (R_2) in analogs of **1** (Figure 4) affected potency and kinetic properties, we obtained a series of perplexing results. First, during biochemical time course experiments, a decrease in potency was noted with increasing pre-incubation times. Second, while a number of compounds containing heterocyclic and aromatic groups such as 2-thienyl (**7**), 2-furyl (**8**) and substituted aryl (**9**) at the R_2 position exhibited apparent potent inhibitory activity; other closely related derivatives such as the pyridine ester (**10**) were inactive. As a reactive carbonyl was presumed to be the primary determinant of activity, the inactivity of this highly electrophilic ester was puzzling. Further studies showed that the presence of DTT in the assay medium caused these confounding results. In this instance however, the nucleophilic character of DTT, rather than its reducing properties, was the culprit.^{xxiv}

A series of experiments showed that upon incubation of **7** (R_1 =phenyl) in the DTT-containing assay medium, DTT reacted with the ester carbonyl to generate pyrazolone **4a** and a DTT-adduct **11** (Scheme 3). After 2 hours incubation, 80% of **7** was converted to **4a**. This transesterification reaction occurred in all analogs of **1** tested, however the rates of reaction varied and were dependent on the nature of R_2 . For example, complete conversion (>99%) of Compound **10** (R_1 =*p*-methoxyphenyl) occurred with only a 1 hour incubation time. This result explained the apparent “inactivity” of certain compounds: only small amounts of the ester were present during the assays, as they were completely converted to the corresponding inactive pyrazolone and DTT-adduct. In this case, replacement of DTT with cysteine did not solve the instability issue as the transesterification occurred in the presence of cysteine as well.^{xxiv}

It is interesting to note that another MLSCN Center identified members of this series as inhibitors of the two component NS2B-NS3 Proteinase of West Nile Virus.^{xxix,xxx} CID 655490 (R_1 =*p*-methoxyphenyl; R_2 =phenyl) was characterized as a slow-binding, reversible, non-competitive inhibitor. Similar to our observations, instability of the compounds was observed under the conditions of this assay. However, in this case, neither DTT nor cysteine was present. We speculate that Tris in the NS2B-NS3 assay protocol could be responsible for hydrolysis of the ester, and that this series of pyrazole sulfonamides is particularly prone to hydrolysis and degradation.

Our experience with the cathepsin B assay supports the contention that a thorough analysis of high throughput screening data and significant follow up studies need to be carefully considered prior to assigning a compound as “active.” The presence of DTT, a seemingly innocuous and common additive to assay media, can often provide misleading data not only due to its potential to act as a potent reductant and H_2O_2 generator, but also through its nucleophilic reactivity. In some instances, replacement of DTT with other reductants such as β -mercaptoethanol or cysteine may attenuate the promiscuous reductant activity, and thereby serve as useful replacements. However, when the nucleophilicity of DTT is a confounding factor, these alternatives are similarly problematic. In fact, the general stability of each hit series under the assay conditions should be evaluated as part of the routine hit triage process. In this study, the characterization of the mode of action of the pyrazole sulfonamides was essential to identify issues of instability, as well as to determine the nature of the inhibition as competitive substrates. These data on pyrimidotriazine-diones as well pyrazole sulfonamides such as **1** should be considered before including these or similar compounds in screening libraries or advancing them as hits.

Cathepsin L

Like cathepsin B, cathepsin L is a cysteine protease of the papain-superfamily that plays a variety of roles in physiological function. Cathepsin L, or its inappropriate or overactivity,

has also been implicated in a number of pathophysiological conditions such as osteo- and rheumatoid arthritis, cancer, and osteoporosis,^{vii} as well as in viral infectivity.^{vi} Furthermore, cathepsin L-like (and cathepsin B-like) proteases have been isolated from and implicated in the replication of parasites such as *Plasmodium falciparum*, *Leishmania major* and *Trypanosoma brucei* and *cruzi*.^{xxxi} Known inhibitors of cathepsin L include peptide as well as non-peptide analogs.^{iv}

High Throughput Screen, Triage and Identification of Cathepsin L Hits

Our efforts began with screening of approximately 58,000 NIH SMR compounds^{xxxii} for inhibitors of human liver cathepsin L. After the primary screen, 100 compounds were designated “active” according to the criteria of >45% inhibition at 10 μ M; approximately half of those compounds were confirmed in dose-response testing in the presence of cysteine, rather than DTT.^{xxxiii} A series of 5 oxadiazoles of general structure **12** appeared most promising based on their potent activity under the cysteine-assay conditions (IC_{50} ~200–600nM), excellent selectivity (no other activity noted in PubChem assays) and good predicted physical properties.^{xxxiv}

Completion of an analysis of the purity and integrity of the original library and re-supplied samples indicated the presence of some impurities. A synthesis of oxadiazoles (e.g., **12**: $R_1=H$; $R_2=o$ -ethylphenyl)^{xxxiv} was developed based on procedures in the literature^{xxxv} and provided pure materials as determined by LC/MS. However, when this pure sample of **12** ($R_1=H$; $R_2=o$ -ethylphenyl) was tested in the cathepsin L assay, no inhibition was observed. Further studies revealed that the impurities present in the original samples were responsible for the biological activity. Those activities were attributed to thiocarbazate **13**, the product of hydrolysis of the oxadiazole moiety.^{xxxiv} This chemotype – “CO-NH-NH-CO-S” - had not been previously described in the literature, however it bears resemblance to azapeptides, some of which have been described as papain inhibitors.^{xxxvi}

A general synthesis of thiocarbazates (Scheme 4) was quickly developed, and required three chemical steps starting from an amino acid derived hydrazone (**14**). Treatment of the hydrazone with thiocarbonyl gas,^{xxxvii} followed by alkylation generated thiocarbazate **15**. The N-Boc group was then removed using acid treatment to furnish the free amine **13**. This sequence was efficient and proceeded in moderate to good overall yield. When tested, **13** ($R_1=H$, $R_2=o$ -ethylphenyl) inhibited cathepsin L with an IC_{50} of 133nM. Some instability of **13** was noted due to diketopiperazine formation (Scheme 5); thus subsequent analogs retained the N-Boc group. The Boc-protected analogs such as **15** ($R_1=H$, $R_2=o$ -ethylphenyl) were more potent ($IC_{50}=56nM$) than the corresponding free amines.^{xxiv}

Characterization of Thiocarbazate's Inhibition

Typical cysteine protease inhibitors contain an electrophilic warhead that provides a site for reaction with the active site cysteine. In the case of the thiocarbazates, the reactivity of the C2 carbonyl as evidenced by diketopiperazine formation (Figure 7) suggested this functionality as the “warhead.” Further evidence to support this hypothesis was based on several pieces of data. First, we designed, synthesized and evaluated analogs in which the reactivity of the C2 carbonyl was modified. The carbon analog (**17**) proved inactive, the nitrogen analog (**18**) modestly active and the oxygen analog (**19**) was as potent as the thiocarbazate.^{xxxviii} This modulation in potency reflected the changes in the reactivity of the C2 carbonyl. Second, docking studies of **15a** ($R_1=H$, $R_2=o$ -ethylphenyl) placed the C2 carbonyl in close proximity to the active site cysteine.^{xxxix} The kinetic characterization of **15a** revealed a slowly reversible inhibition.^{xl} Additional studies provided further insight into the mode of action of these inhibitors. We conducted experiments in which stoichiometric amounts of enzyme were incubated with **15a**, and monitored the reaction by LC/MS for the

appearance of products of hydrolysis **20** and **21** (independently synthesized). No products were detected. Furthermore, a similar reaction in the presence of cysteine produced no detectable products either (Scheme 7). Taken together, these results suggested that the active site cysteine reacts with the C2 carbonyl of the thiocarbazate to form a stable tetrahedral intermediate **22** (Scheme 8). Unlike substrates of the enzyme, this intermediate does not collapse further as evidenced by the lack of hydrolysis products detected. In time, however, intermediate **22** may covert back to the thiocarbazate and active enzyme.

Identification of CID16725315 (**15a**) as a Probe of Cathepsin L

The potent activity and selectivity of **15a** warranted the designation of this compound as a MLSCN probe.^{xxx} Table 1 summarizes its activity, kinetic properties, selectivity and toxicity. As cathepsin inhibitors have been reported to be efficacious in cellular models of *Leishmania major* and *Plasmodium falciparum* infection, **15a** was evaluated and exhibited IC₅₀'s of 12.5μM and 15.4μM respectively.^{xl}

Docking Studies with Compound **15a**

Docking experiments were performed using an x-ray structure of papain bound to a diketo-epoxide inhibitor^{xli} as a model of cathepsin L (Figure 9).^{xxxix} These studies find that thiocarbazate **15a** binds to the active site with similar energies as the diketo-epoxide inhibitor, and makes a number of identical interactions. Three sub-sites are fully occupied: the S1' sub-site is occupied by the 2-ethylphenyl group; the indole group occupies the S2 sub-site, and the tert-butoxy group sits in the S3 sub-site. Table 1 records the types of interactions **15a** makes with specific residues in papain/cathepsin L as suggested by the docking studies. This model was useful to rationalize the activity of analogs, as well as to help optimize them.

Design and Synthesis of Analogs of **15a**

A series of analogs in which modifications to the amino acid side chain (A) and sulfur substituent (B) were designed and prepared. For example, the N-Boc tryptophan moiety was replaced with the corresponding phenylalanine, leucine, alanine and glycine units to afford **23**, **24**, **25** and **26** respectively.^{xlii} All exhibited reduced activity, with decreased potency correlating with decreased size. This result is consistent with the docking model that positions the indole group of **15a** in the relatively large S2 site, and predicts that smaller substituents would be less active. Deletion of the NH-Boc group (**27**)^{xxxviii,xxxix} reduced potency by nearly 400-fold, presumably due to loss of the H-bond to Asp158/162 and lack of occupancy of the S3 site.

Modification at position B revealed trends that also could be rationalized by the docking studies. Replacement of the anilide with constrained analogs such as tetrahydroquinoline (**28**) or tetrahydroisoquinoline (**29**) resulted in comparable or improved potency. However, removal of the carbonyl function through incorporation of an aniline (**30**) generated an inactive compound.^{xlii} This result provided support for the importance of the hydrogen bond between the anilide carbonyl and Trp177/189. An extensive library of compounds was prepared that incorporated a wide diversity of additional modifications at each of these positions. The compounds were evaluated in cathepsin L, as well as cathepsin B and cathepsin S assays, and SAR trends were delineated.^{xlii}

Identification of CID 23631927 (**19**) as an Optimized Probe of Cathepsin L

One of the most potent analogs identified was Compound **19** (CID 23631927). We designed this molecule by combining the improved potency of oxacarbazates with ring constraints at

the anilide moiety. This compound exhibited an IC₅₀ against cathepsin L at 7nM (0.4nM with a 4 hour pre-incubation) and exhibited exquisite selectivity compared to cathepsin B (>700 fold). Furthermore, in a zebrafish whole organism assay, no toxicity was detected up to 100µM. In cellular models, analog **19** prevented the entry of SARS-CoV and Ebola virus into cells at sub-micromolar concentrations.^{xxx} This result confirms the role of cathepsin L in viral entry, and suggests compounds such as **19** may find utility as starting points in drug discovery efforts.

Conclusions

Our experience with identification, characterization and triage of hits from screening of the NIH SMR for inhibitors of cysteine proteases enforced several important lessons. First, as demonstrated in our work on cathepsin B, the mechanism of action of hits should be evaluated prior to significant assignment of resources. False positives plague all high throughput screens, and rapid elimination of compounds that act through such mechanisms such as redox cycling assures that poor hits and leads are not pursued. Second, the stability of hits in assay media should always be evaluated to confirm that the integrity of the structure is intact throughout the biological assay. The availability of the public database, PubChem, if correctly annotated, should provide a wealth of information about compounds and substructures with these liabilities. Third, the mechanism of action of hits should be considered and evaluated prior to hit and lead optimization. Finally, careful analysis of library samples for purity and integrity is essential, as is re-synthesis and re-assay of hit samples. This process ensures that the activity is ascribed to the correct structure, and provides high quality data for generating valid structure activity relationships. Through studies such as these, we were able to quickly eliminate false positives and nuisance compounds identified in primary HTS assays for cathepsin B. We were also able to identify potential liabilities, such as instability, to distinguish alternate substrates from inhibitors, and to focus medicinal chemistry efforts accordingly. In the cathepsin L project, analysis of purity and integrity lead to the identification of a novel chemotype, a thiocarbazate, that acted as a potent inhibitor of cathepsins. Optimization of this moiety led to novel, potent and highly selective probes of cathepsin L that have considerable potential in preventing viral entry to cells.

Acknowledgments

Financial support for this work was provided by the NIH (5U54HG003915-03). The authors would like to thank our PCMD colleagues Scott Diamond, Mary Pat Beavers, Michael Myers, Zhuqing Liu, Phillip Beneditti, Barry Cooperman, Parag Shah, Andrew Napper and Barry Cooperman along with Paul Bates. DMH is grateful to Jay Kostman for editorial assistance.

References

- i. Austin CP, Brady LS, Insel TR, Collins FS. *Science* 2004;306:1138–1139. [PubMed: 15542455]
- ii. Huryn DM, Cosford NDP. *Ann Reports Med Chem* 2007;42:401–416.
- iii. <http://www.seas.upenn.edu/~pcmd/>
- iv. For several reviews of this area see: (a) McGrath ME. *Annu Rev Biophys Biomol Struct* 1999;28:181–204. [PubMed: 10410800] (b) Hernandez AA, Roush WR. *Curr Opin Chem Biol* 2002;6:459–465. [PubMed: 12133721] (c) Schirmeister T, Kaeppler U. *Mini-Rev Med Chem* 2003;3:361–373. [PubMed: 12678829]
- v. <http://pubchem.ncbi.nlm.nih.gov/>
- vi. (a) Simmons G, Gosalia DN, Rennekamp AJ, Reeves JD, Diamond SL, Bates P. *Proc Natl Acad Sci USA* 2005;102:11876–11881. [PubMed: 16081529] (b) Kaletsky RL, Simmons G, Bates P. *J Virol* 2007;13378–13384. [PubMed: 17928356] (c) Chandran K, Sullivan NJ, Felbor U, Whelan SP, Cunningham JM. *Science* 2005;308:1643–1645. [PubMed: 15831716]

- vii. Vasiljeva O, Reinheckel T, Peters C, Turk D, Turk V, Turk B. *Curr Pharm Des* 2007;13:385–401.
- viii. Michaud S, Gour J. *Exp Opin Therap Patents* 1998;8:645–672.
- ix. Guicciardi ME, Miyoshi H, Bronk SF, Gores GJ. *Amer J Path* 2001;159:2045–2054. [PubMed: 11733355]
- x. Tohyama T, Kiyochi H, Narumoto K, Honda K, Kobayashi N, Kanoh M, Saibara T, Katunuma N. *Transplant Proceed* 2000;32:944–945.
- xi. Hook VYH, Kindy M, Hook G. *J Biol Chem* 2008;388:7745–7753. and references cited therein. [PubMed: 18184658]
- xii. For a recent review of this area see: Frlan R, Gobec S. *Curr Med Chem* 2006;13:2309–2327. [PubMed: 16918357]
- xiii. PubChem web address for PCMD cathepsin B assay results.
<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pcassay&term=453,523,820>
- xiv. Smith GK, Barrett DG, Blackburn K, Cory M, Dallas WS, Davis R, Hassler D, McConnell R, Moyer M, Weaver K. *Arch Biochem Biophys* 2002;399:195–205. [PubMed: 11888206]
- xv. Tjernberg A, Hallén D, Schultz J, James S, Benkestock K, Byström S, Weigelt J. *Bioorg Med Chem Lett* 2004;14:891. [PubMed: 15012988]
- xvi. Motlekar N, Diamond SL, Napper AD. *Assay Drug Dev Technol* 2008;6:395–405.
- xvii. Lor LA, Schneck J, McNulty DE, Diaz E, Brandt M, Thrall SH, Schwartz B. *J Biomol Screen* 2007;12:881–890. [PubMed: 17579124]
- xviii. Johnston PA, Soares KM, Shinde SN, Foster CA, Shun TY, Takyi HK, Wipf P, Lazo JS. *Assay Drug Dev Technol* 2008;6:505–518.
- xix. (a) Feng BY, Shelat A, Doman TN, Guy RK, Shoichet BK. *Nat Chem Biol* 2005;1:146–148. [PubMed: 16408018] (b) Feng BY, Simeonov A, Jadhav A, Babaoglu K, Inglese J, Shoichet BK, Austin CP. *J Med Chem* 2007;50:2385–2390. [PubMed: 17447748] (c) Dalvit C, Caronni D, Mongelli N, Veronesi M, Vulpetti A. *Curr Drug Discov Technol* 2006;3:115–124. [PubMed: 16925519]
- xx. (a) Simeonov A, Jadhav A, Thomas CJ, Wang Y, Huang R, Southall NT, Shinn P, Smith J, Austin CP, Auld DS, Inglese J. *J Med Chem* 2008;51:2363–2371. [PubMed: 18363325] (b) Turek-Etienne TC, Small EC, Soh SC, Xin TA, Gaitonde PV, Barrabee EB, Hart RF, Bryant RW. *J Biomol Screen* 2003;8:176–184. [PubMed: 12844438]
- xxi. Huth JR, Mendoza R, Olejniczak ET, Johnston RW, Cothron DA, Liu Y, Lerner CG, Hajduk PJ. *J Am Chem Soc* 2005;127:217–224. [PubMed: 15631471]
- xxii. Babaoglu K, Simeonov A, Irwin JJ, Nelson ME, Feng B, Thomas CL, Cancian L, Costi MP, Maltby DA, Jadhav A, Inglese J, Austin CP, Shoichet BK. *J Med Chem* 2008;51:2502–2511. [PubMed: 18333608]
- xxiii. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. *Adv Drug Deliv Rev* 2001;46:3–26. [PubMed: 11259830]
- xxiv. Myers MC, Napper AD, Motlekar N, Shah PP, Chiu CH, Beavers MP, Diamond SL, Huryn DM, Smith AB. *Bioorg Med Chem Lett* 2007;17:4761–4766. [PubMed: 17656088]
- xxv. Elgemeie GH, Metwally NH. *J Chem Res* 1999;6:384–385.
- xxvi. Elgemeie GH, Elghandour AH, Elzanate AM, Masoud WA. *Phosphorus, Sulfur and Silicon* 2000;163:91.
- xxvii. Drummond JT, Johnson G. *J Heterocycl Chem* 1988;25:1123.
- xxviii. Vostrova LM, Grenad'orova MV, Klad'ko LG. *Ukr Khim Zh* 2004;9–10:115.
- xxix. Johnston PA, Phillips J, Shun TY, Shinde S, Lazo JS, Huryn DM, Myers MC, Ratnikov BI, Smith JW, Su W, Dahl R, Cosford NDP, Shiryayev SA, Strongin AY. *Assay Drug Dev Technol* 2007;5:737–750. [PubMed: 18181690]
- xxx. <https://mli.nih.gov/mli/mlp-probes/>
- xxxi. McKerrow JH, Engel JC, Caffrey CR. *Bioorg Med Chem* 1999;7:639–644. [PubMed: 10353643]
- xxxii. PubChem web address for PCMD cathepsin L primary screen assay and hits.
<http://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?aid=460>
- xxxiii. PubChem web address for PCMD cathepsin L dose response testing in the presence of cysteine.
<http://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?aid=1627>

- xxxiv. Myers MC, Shah PS, Diamond SL, Huryn DM, Smith AB. *Bioorg Med Chem Lett* 2008;18:214–218.
- xxxv. Confalone PN, Woodward RB. *J Am Chem Soc* 1983;105:902–906. Verardo G, Geatti P, Lesa B. *Synthesis* 2005:559–564. von Geldern TW, Tasker AS, Sorensen BK, Winn M, Szczepankiewicz BG, Dixon DB, Chiou WJ, Wang L, Wessale JL, Adler A, Marsh KC, Nguyen B, Opgenorth TJ. *J Med Chem* 1999;42:3668–3678. [PubMed: 10479298]
- xxxvi. For a representative example see: Magrath J, Abeles RH. *J Med Chem* 1992;35:4279–4283. [PubMed: 1447732]
- xxxvii. Chande MS, Singh-Jathar K. *Indian J Chem* 1998;37B:352.
- xxxviii. Myers MC, Shah PP, Beavers MP, Napper AD, Diamond SL, Smith AB, Huryn DM. *Bioorg Med Chem Lett* 2008;18:3646–3651. [PubMed: 18499453]
- xxxix. Beavers MP, Myers MC, Shah PP, Purvis JE, Diamond SL, Cooperman BS, Huryn DM, Smith AB. *J Chem Info & Model* 2008;48:1464–1472.
- xl. Shah PP, Myers MC, Beavers MP, Purvis JE, Jing H, Grieser HJ, Sharlow ER, Napper AD, Huryn DM, Cooperman BS, Smith AB, Diamond SL. *Molec Pharm* 2008;74:34–41.
- xli. Tsuge H, Nishimura T, Tada Y, Asao T, Turk D, Turk V, Katunuma N. *Biochem Biophys Res Commun* 1999;266:411–416. [PubMed: 10600517]
- xlii. Liu Z, Myers MC, Shah PP, Beavers MP, Benedetti PA, Diamond SL, Smith AB, Huryn DM. *J Combi Chem HTS*. in press.

CID 653297

CID647501

CID66541

Figure 1.
Structures of CID653297, CID647501 and CID66541

Figure 2.
General structure of pyrazole sulfonamides (**1**)

5 ($\text{R}_2=2\text{-thiophene}$)

6 ($\text{R}_2=2\text{-thiophene}$)

Figure 3.
Non-ester analogs of pyrazole sulfonate **1**

Figure 4.
Ester analogs

Figure 5.
CID 655490

Figure 6.
Reported structure of oxadiazole cathepsin L hits **12** and true active species **13**.

Figure 7.
Analogues with modified reactivity at C2 carbonyl

Figure 8.
Compound **15a** docked to papain

Figure 9.
Structure of cathepsin L probe CID 23631927

Scheme 1.
General synthesis of pyrazole sulfonamides **1**

1a (R_1 =Ph; R_2 =2-thiophene)

4a(R_1 =Ph)

Scheme 2.
Stoichiometric reaction of **1a** with cathepsin B

Scheme 3.
Reaction of DTT with pyrazole sulfonamides

Scheme 4.
General synthesis of thiocarbazates

Scheme 5.
Diketopiperazine formation

Scheme 7.

Proposed mechanism of activity of **15a** with cathepsin L.

Scheme 8.
Reaction of thiocarbazate **15a** with cathepsin L or cysteine.

Table 1

Biological and kinetic data on CID16725315

IC ₅₀	56nM
IC ₅₀ (4h preincubation)	1nM
k _{on}	24,000M ⁻¹ s ⁻¹
k _{off}	2.2 × 10 ⁻⁵ s ⁻¹
Ki	0.89nM
Selectivity vs cathepsin B	45-fold
Selectivity vs. cathepsins G, K, V, S	11 – >150-fold
Toxicity (zebrafish)	>100μM
Toxicity (bovine aortic endothelial cells)	16μM

Table 2Binding Interactions of **15b** docked to papain

Papain Residue	Analogous cathepsin L residue	Observed Binding Interaction
Gln14	Gln 19	H-bonding
Cys25	Cys 25	within bonding distance; H-bonding
Gly66	Gly 68	H-bonding
Asp158	Asp 162	H-bonding
Trp177	Trp 189	Hydrophobic/Aromatic; H-bonding
Ser205	Ala214	Hydrophobic/Aromatic

Table 3

Analogues of 15a and cathepsin L inhibitory activity

	<p style="text-align: center;">A</p>
Compound	A	
15a

	<p data-bbox="1031 226 1096 304">A</p> <p>Chemical structure of compound A: A Boc-protected amino acid derivative. The structure features a BocHN group attached to a chiral center (indicated by a dashed bond). This chiral center is also attached to an indole ring system. The BocHN group is further connected to a carbonyl group, which is linked to a hydrazide group (NH-NH-C=O) and a thioether group (S-CH2-CH2-).</p>	
Compound	A	
24	<p>Chemical structure of compound 24: A Boc-protected amino acid derivative. The structure features a BocHN group attached to a chiral center (indicated by a dashed bond). This chiral center is also attached to a cyclohexane ring system. The BocHN group is further connected to a carbonyl group, which is linked to a hydrazide group (NH-NH-C=O) and a thioether group (S-CH2-CH2-).</p>	

<p>A</p>		
Compound	A	
25		

<div data-bbox="902 233 1624 873"><p>A</p> </div>		
Compound	A	
26	<div data-bbox="370 1041 1507 1566"> </div>	

<p>A</p>
Compound	A	
27

	<p data-bbox="1031 220 1096 304">A</p> <p>Chemical structure of compound A: A Boc-protected amino acid derivative. The structure features a BocHN group attached to a chiral center (indicated by a dashed bond). This chiral center is also attached to an indole ring system. The BocHN group is connected via a carbonyl group to a hydrazide linkage (NH-NH-C=O), which is further connected to a thioether group (S-CH2-CH2-).</p>	
Compound	A	
28	<p>Chemical structure of compound 28: A complex molecule featuring a bicyclic system (likely a decalin derivative) with a hydrazide linkage (NH-NH-C=O) and a thioether group (S-CH2-CH2-). The structure is shown in a 3D representation with dashed and wedged bonds.</p>	

<div>A </div>		
Compound	A	
29		
30		

<div data-bbox="906 233 1624 873"><p>A</p></div>		
Compound	A	
31	<div data-bbox="878 1020 1008 1178"><p>A</p></div>	