CHEMISTRY
AN ASIAN JOURNAL

DOI: 10.1002/asia.201100523

Applying Small Molecule Microarrays and Resulting Affinity Probe Cocktails for Proteome Profiling of Mammalian Cell Lysates

Haibin Shi,[a, c] Mahesh Uttamchandani,[a, d] and Shao Q. Yao*[a, b, c]

On the occasion of the 10th anniversary of click chemistry

Abstract: Small molecule microarrays (SMMs) are proving to be increasingly important tools for assessing protein–ligand interactions, as well as in screening for enzyme substrates and inhibitors, in a high-throughput manner. We previously described an SMM-facilitated screening strategy for the rapid identification of probes against γ-secretase, an aspartic protease. In this article, we extend upon this work with an expanded library of hydroxyethylamine-derived inhibitors which non-exclusively target aspartic proteases. Our library is diversified across P_2, P_1, P_1', and P_2' positions. Accordingly, 86 new inhibitors are synthesized using a combinatorial, solid-phase synthetic approach, bringing the total library size to 284-biotinylated compounds, which were arrayed onto avidin slides. In order to elucidate enzymatic activity and profiles within complex biological samples, screening is performed using fluorescently-labeled mammalian cell lysates. This yielded reproducible profiles or binding fingerprints that correspond with interactions from aspartic proteases or accessory proteins as well as other interacting targets that were present in the sample. The brightest microarray hits were converted to affinity-based probes (A*f*BPs) using convenient, 1-step "click" chemistry with benzophenone from the relevant building blocks. Pull-down/mass spectrometric analysis with these probes (individuals or cocktail) yielded putative protein targets that include well-known aspartic proteases, such as cathepsin D which is a clear marker for breast cancer cell lines, T47D. Many other hits were also identified, which may be secondary or tertiary interactors of aspartic proteases, or yet unreported off-targets of the hydroxyethylamine pharmacophore. Our work herein thus provides a candidate list of biomarkers for further investigations. Taken together, this SMM-facilitated strategy for the discovery of new A*f*BPs should provide a useful tool for high-throughput development of novel small molecule probes and the identification of new aspartic proteases as well as related biomarkers in the future.

Keywords: click chemistry · fluorescence · microarrays · proteins · proteomics

Introduction

Approaches in functional proteomics have classically included two-dimensional gel electrophoresis (2DE),[1] liquid chromatography-mass spectrometry (LC-MS), and DNA microarray technology.[2,3] These methods have been applied to the large-scale characterization of cellular, subcellular, or circulating proteins and biomarkers.[4] However, the majority of such technologies are geared towards detecting changes in protein abundance rather than actual functional activity.[5] To address this gap, activity-based protein profiling (ABPP) has emerged as a powerful chemical proteomic alternative for assessing the differing functional states of enzymes and other protein sub-classes within a complex proteome.[6] To date, small molecule probes (the so-called activity-based probes, or ABPs) needed for ABPP studies have

[a] H. Shi, M. Uttamchandani, Prof. Dr. S. Q. Yao
Department of Chemistry
National University of Singapore
3 Science Drive 3, Singapore 117543
Fax: (+65) 6779-1691
E-mail: chmyaosq@nus.edu.sg

[b] Prof. Dr. S. Q. Yao
Department of Biological Sciences
National University of Singapore

[c] H. Shi, Prof. Dr. S. Q. Yao
NUS MedChem Program of the Life Sciences Institute
National University of Singapore

[d] M. Uttamchandani
Uttamchandani Mahesh. Defence Medical and Environmental Research Institute
DSO National Laboratories
27 Medical Drive, Singapore 117510

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/asia.201100523.

been successfully developed for a wide range of mechanistically distinct enzymes, such as serine hydrolases, cysteine proteases, aspartic proteases, metalloproteases, phosphatases, kinases, and others.[7–13] However, gel-based separation of complex samples labeled with ABPs is highly limited, in that only a few probes, or a narrow pool of samples, may be screened at any given time. Small molecule microarrays (SMMs), on the other hand, offer high scalability in the assessment of ligand-protein interactions.[14] We and others previously described SMM-facilitated screenings, using combinatorial small molecule libraries, for the screening of purified enzymes and proteins.[15] This included a recent example for the high-throughput identification of affinity-based probes (A*f*BPs) against γ-secretase,[9c] an aspartic protease implicated in the progression of Alzheimer's disease.[16] More recently, we have successfully demonstrated the first example of SMMs for large-scale functional profiling of cysteine proteases present in native apoptotic biological samples, including lysates from mammalian cells and parasite-infected red blood cells (RBCs).[17] Herein, we extend this SMM approach to the screening of a variety of mammalian cellular lysates from numerous tumor cell lines using small molecule probes derived from hydroxyethylamine-based pharmacophores. With the type of throughput provided by SMMs in the initial screening, we were able to quickly identify selected A*f*BPs suitable for the large-scale profiling of aspartic proteases and other potential biomarkers in mammalian proteomes. Additionally, unknown cellular targets identified from screening with such probes might provide opportunities for future applications of hydroxyethylamine-based inhibitors in novel therapeutics (Figure 1).

Results and Discussion

Rationale of Our Strategy

Aspartic proteases are the smallest of all protease classes in the human genome, with only 15 members documented thus far.[18] However, these enzymes have received considerable attention as potential therapeutic targets since many have been shown to play important roles in the physiological and pathological processes.[19] For example, HIV-1 protease is an important target in the fight against AIDS.[19a,b] Renin is a target in the treatment of hypertension.[19c] Cathepsin D is implicated in breast cancer metastasis.[19d] Plasmepsins are a key mediator in malaria pathogenesis.[19e] Aspartic proteases are characterized by two catalytic aspartic acid residues located in their active sites and, in most cases, by the conserved Asp-Thr-Gly (DTG) sequence in the primary structure. Although the catalytic mechanism of aspartic proteases is poorly understood, it has been generally accepted that the aspartic residues in the enzyme active site bind a molecule of water through extensive H-bonding. Of the numerous known aspartic protease inhibitors, hydroxyethylamine-containing compounds are amongst the most potent and successful, some of which are already FDA-approved drugs.[20] For example, Amprenavir and Nelfinavir are just two examples of hydroxyethylamine-based drugs currently used in the treatment of AIDS (Figure 2a). By mimicking the transition state formed during the proteolytic reaction of aspartic proteases, the hydroxyl group on the hydroxyethylamine isostere binds tightly to the enzyme active site through H-bonding with the catalytic aspartic residues, thereby replacing the active site-bound water molecule. We previously showed that this pharmacophore could be strategically and conveniently diversified to generate inhibitor libraries that vary structurally at the corresponding P_2, P_1, P_1', and P_2' positions (Figure 2b).[9c]

In the current study, with the aim to develop new A*f*BPs, which may be used to gain further insight into potential hy-

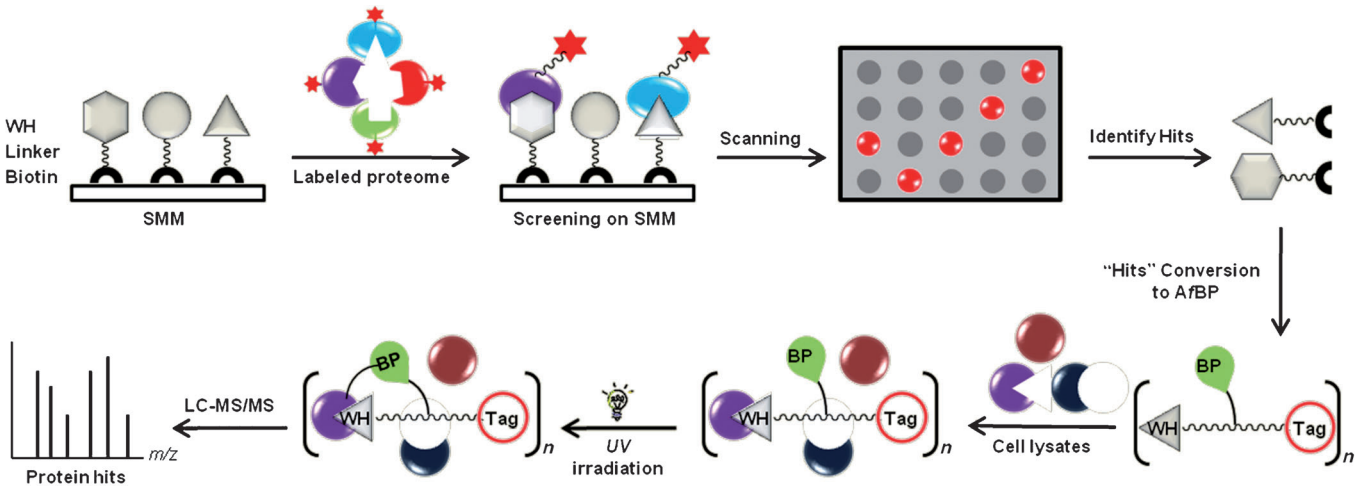


Figure 1. Overall strategy for small molecule microarray (SMM)-facilitated development of A*f*BPs for proteome profiling of aspartic proteases in mammalian cell lysates.

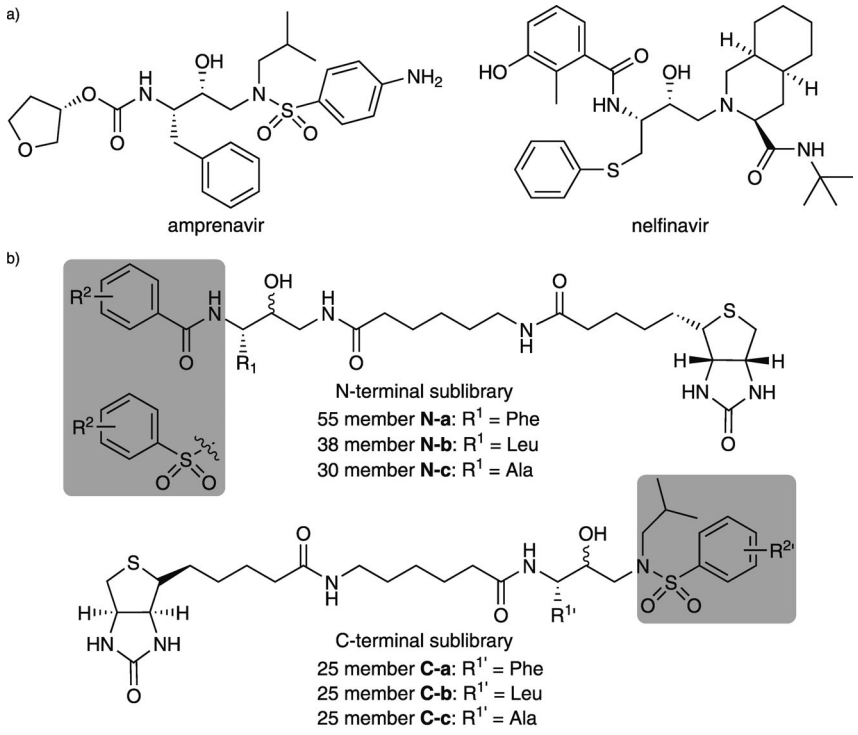
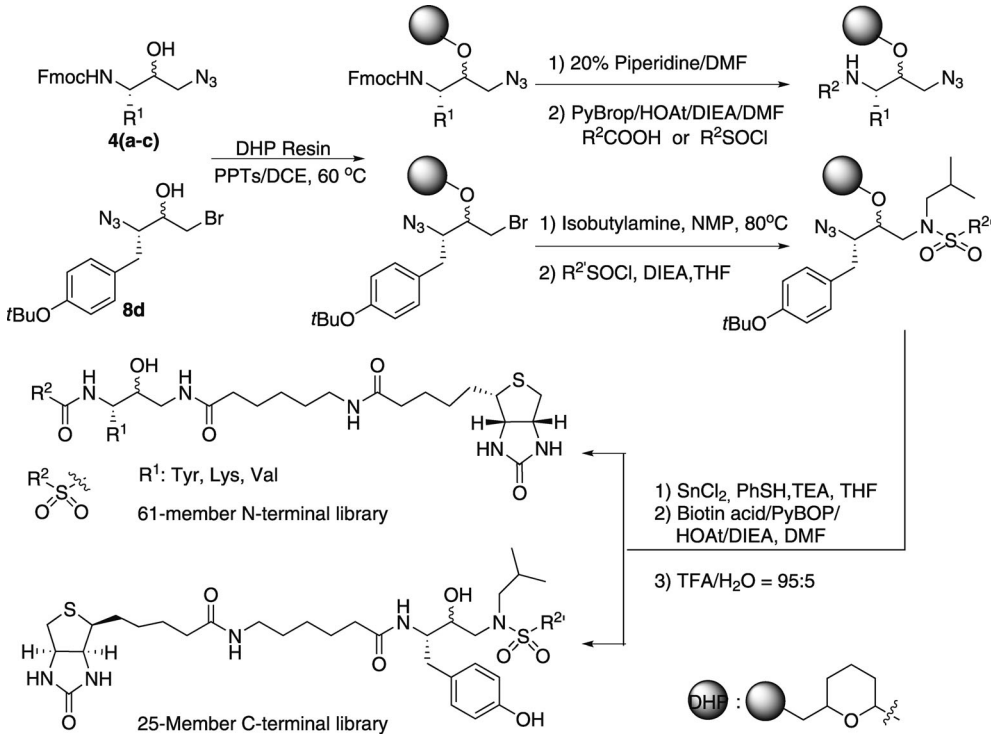


Figure 2. a) Representative structures of two commercially available drugs against HIV protease, an aspartic protease. b) The 184-member library against aspartic proteases based on the hydroxyethylamine pharmacophore.[9c] In both N- and C-terminal sublibraries, the P_1/P_1' positions were diversified by building blocks derived from three amino acid residues (Phe, Leu, and Ala; in red). The P_2/P_2' positions (highlighted in gray boxes), were diversified by various carboxylic acids and sulfonyl chloride building blocks.

droxyethylamine-interacting proteins in various mammalian proteomes and to discovery new cancer biomarkers, we conceived a multi-step approach, as depicted in Figure 1; in the first stage, an SMM platform was established, on which a structurally diverse library of hydroxyethylamine inhibitors, each containing a biotin tag (and a connecting linker) for SMM immobilization, was used to probe the binding profiles of endogenous aspartic proteases and other proteins from complex proteome samples. We had recently shown this is feasible with fluorescently-labeled crude proteome lysates.[17] Subsequently, positive hits (e.g., bright spots) obtained from the SMM screening were identified, and the corresponding compounds were conveniently converted to affinity-based probes (A*f*BPs) by the Cu^I-catalyzed 1,3-dipolar cycloaddition (click chemistry or CuAAC).[21] This step was made possible owing to the modular design of our compound libraries, and the introduction of a benzophenone (Bp) as well as a reporter tag to the vicinity of these hydroxyethylamine pharmacophores enabled the covalent capture of any potential interacting proteins. In the next step, the resulting A*f*BPs, either as individual probes or a mixture (that is, "cocktail") of probes, are used for gel-based proteome profiling against mammalian cell lysates. Finally, large-scale pull-down experiments followed by mass spectrometric analysis with selected probes allows positive identification of hydroxyethylamine-binding protein targets that may be further developed into potential cancer biomarkers.

Design and Synthesis of Hydroxyethylamine-Containing Inhibitor Library

Hydroxyethylamine isosteres are transition-state mimics that were introduced as the backbone replacement of the amide bond located between the P_1/P_1' position. We previously developed a solid-phase synthetic strategy capable of diversifying all four positions around the hydroxyethylamine core, that is, P_1/P_1' and P_2/P_2' residues (Figure 2b).[9c] From that study, two sublibraries (N-terminal and C-terminal as shown in Figure 2b) with a total of 198 different compounds were already available. To further expand the coverage of our library of compounds to be spotted on the SMM, an additional 86 new compounds were prepared in the current study (Scheme 1). In the new N-terminal sublibrary, the R_1 position, corresponding to the P_1 residue upon binding to aspartic proteases, was varied with three amino acids (Tyr, Lys, and Val). The R_2 position, corresponding to the P_2 residue, was diversified with 24 aromatic/aliphatic moieties (Figure S3 in the Supporting Information). In the new C-terminal sublibrary, the P_1' residue (denoted as R_1' in the inhibitor library) was fixed with Tyr for its general importance at this position amongst many known aspartic protease substrates/inhibitors. The P_2' residue (denoted as R_2') was similarly varied with a variety of commercially available sulfonyl chlorides (see Supporting Information). As shown in Scheme 1, the synthesis of these compounds was carried out by following our previously reported solid-phase approaches on different pre-loaded 3,4-dihydro-2H-pyran (DHP) resins.[9c] Starting from scaffolds **4d–f** and **8d**, which were synthesized from the corresponding commercially available L-amino acids following previously reported procedures with some modifications,[22] compounds **4d–f** and **8d** were immobilized at their −OH groups onto the DHP resin under acidic conditions, giving preloaded resins for N- and C-terminal sublibraries, respectively. Next, for the N-terminal sublibrary, Fmoc deprotection of the resin was carried out, followed by standard acylation with 24 different acid building blocks (**1–24** in Figure S3 in the Supporting Information). Subsequent azide reduction, which yielded the corresponding amines, attachment of a biotin linker, and TFA cleavage yielded the 61-member N-terminal compounds. For the synthesis of C-terminal compounds, the corresponding preloaded DHP resin was first al-



Scheme 1. Solid-phase synthesis of the N- and C-terminal sublibraries using preloaded DHP resins. A total of 86 new hydroxyethylamine-derived compounds were synthesized. Fmoc = fluorenylmethyloxycarbonyl, DCE = 1,2-dichloroethane, DMF = *N*,*N*-dimethylformamide, PyBrop = bromo-tris-pyrrolidino phosphoniumhexafluorophosphate, HOAt = 1-hydroxy-7-azabenzotriazole, DIEA = ethyldiisopropylamine, NMP = 1-methyl-2-pyrrolidone, PyBOP = (benzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate, THF = tetrahydrofuran.

kylated with *iso*-butylamine, then acylated with commercially available sulfonyl chloride building blocks (**1–25** in Figure S4 in the Supporting Information). Subsequent azide reduction with $SnCl_2$/PhSH provided the corresponding amines, which, upon coupling with the biotin linker and TFA cleavage, afforded the 25-member C-terminal compounds. The resulting 86 new compounds were analyzed by LC-MS and shown to be of sufficient purity (>80%) for subsequent microarray immobilization and screening (Table S1 in the Supporting Information).

Proteome Profiling of Mammalian Cell Lysates on SMMs

The 86 new compounds were combined with the 198 previously reported compounds, giving a total of 284 hydroxyethylamines with diversities at all $P_2/P_1/P_1'/P_2'$ positions. They were robotically spotted onto avidin-functionalized slides in duplicate to generate the corresponding small molecule microarray (see Supporting Information for spotting pattern). Eight different human cell lines, seven of which are derived from tumor cells (e.g., human normal kidney cell line HEK293T, Ovary cancer cells OVCAR-3, erythromyeloblastoid leukemia cells K562, colon cancer cells HCT116, hepatocellular liver carcinoma cells HepG2, and three breast cancer cell lines MDA-MB-435, MCF7, and T47D), were selected for SMM screening (Figure 3). Briefly, cellular proteome lysates from these cell lines were minimally labeled by an amino-reactive Cy3 dye and directly applied onto the microarray, based on previously developed procedures.[17] To ensure that all fluorescence signals were indeed attributable to activity-dependent interactions between the hydroxyethylamine inhibitors and their targeted proteins, the heat-denatured form of representative cellular lysates (that is, T47D) was routinely applied to the same array as a negative control experiment. As shown in Figure 3a, all eight mammalian cell lysates produced highly distinctive and similar SMM binding profiles, with most of the brightest spots originating from selected members of the C-terminal sublibrary (Figure 3b). We speculated that this arises from the presence of some commonly targeted proteins in both normal and cancer cells. Colored heat maps were generated from the corresponding SMM images in order to better visualize and compare different binding fingerprints obtained from the eight different cell lines (Figure 3b); each mammalian cell line displayed characteristic patterns, and closer examination further revealed that spots containing hydroxyethylamine compounds having Tyr and Ala at the P_1' position provided the strongest binders. However, the effect of different substituent groups at the P_2' position did not have much of an effect on the binding signals observed. Taken together, the SMM screening provided a rapid tool to quickly delineate the protein–ligand interaction and identify potential "hit" compounds for subsequent follow-up studies (that is, conversion to A*f*BPs and subsequent proteome profiling experiments). Additionally, the unique binding fingerprints obtained from these SMM experiments may also provide

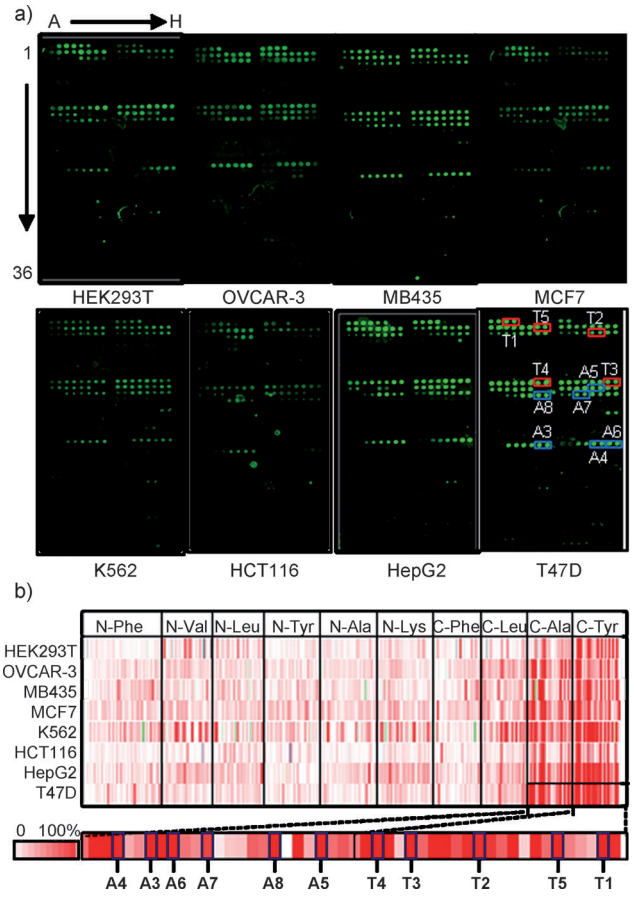


Figure 3. SMM profiles obtained by screening with eight different mammalian cell lysates. a) The binding profiles of Cy3-labeled mammalian cell lysates obtained from screening against the 284-member hydroxyethylamine SMM. All 284 compounds were spotted in duplicate (see Supporting Information for spotting patterns/IDs). Positions of 11 brightest inhibitor spots are highlighted on the SMM. b) Color heat maps displaying binding of the 284-member compound library against eight Cy3-labeled mammalian cell lysates. The zoomed heat map shows inhibitor potencies of the C-terminal sublibrary presenting Ala and Tyr at the P_1' position. The scale bar represents the relative potency ratio.

useful clues for future design and diversification of hydroxyethylamine pharmacophores as potential drug candidates.[23]

Click Assembly of Select A*f*BPs

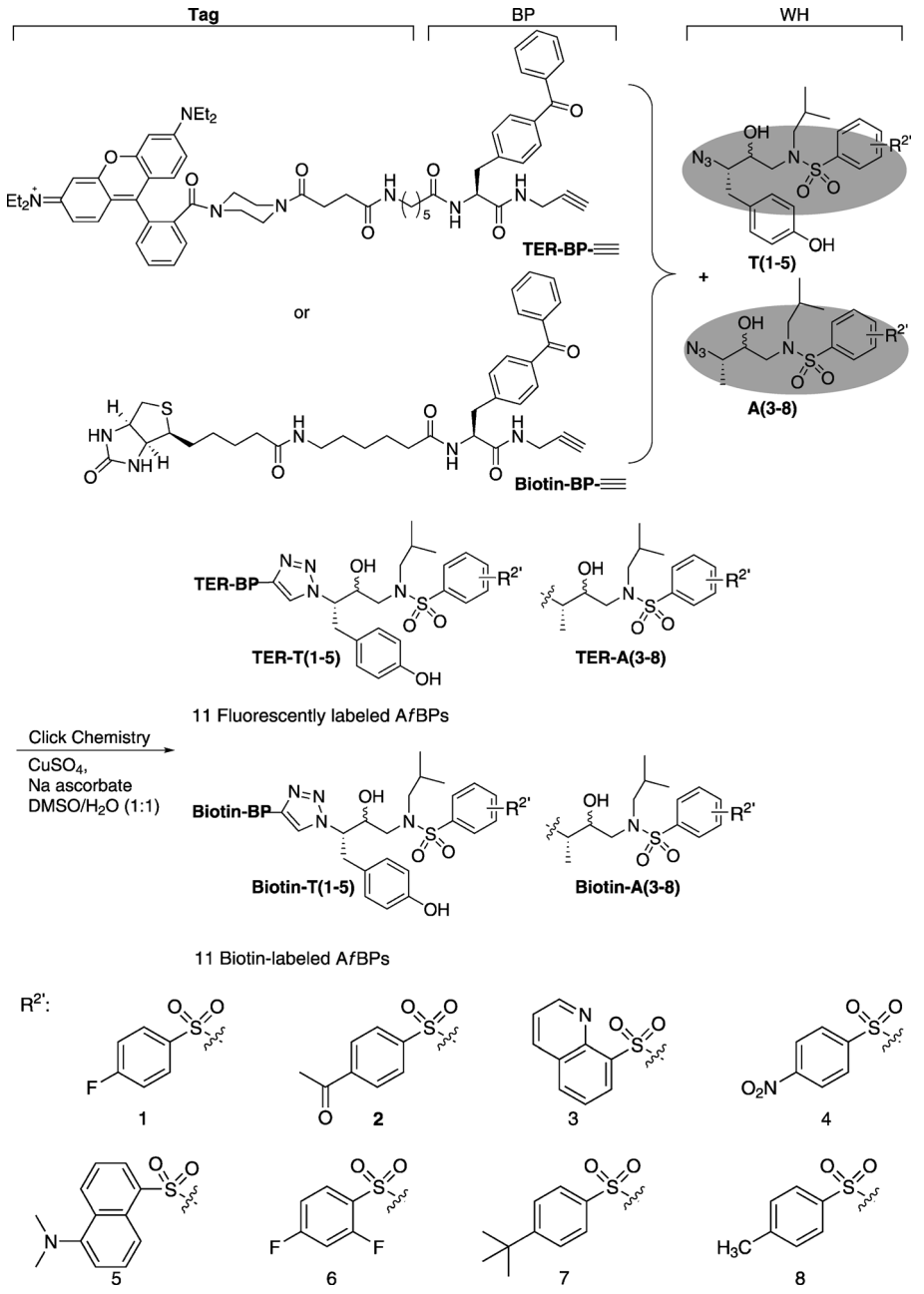
To further examine the positive cellular protein targets that exhibited strong binding to members of our hydroxyethylamine library, the eleven brightest spots (boxed in Figure 3a) were chosen, and converted into the corresponding affinity-based probes (A*f*BPs). As was mentioned earlier, this step was made possible owing to the modular design of our compound libraries, and the implementation of click chemistry into our design principle. The highly efficient and modular Cu^I-catalyzed cycloaddition reaction between alkynes and azides (that is, click chemistry) had previously been successfully exploited to facilitate the synthesis of other classes of activity-based probes.[21] As shown in Scheme 2, selected azide-containing **WH**, **T(1–5)**, and **A(3–8)**, which were actually key intermediates in the solid-phase synthesis of our hydroxyethylamine libraries (Scheme 1), were "click" assembled with one of the two alkyne-containing photo-cross-linkers, TER-Bp-≡ and Biotin-Bp-≡, giving 11 fluorescently labeled A*f*BPs (for subsequent in-gel fluorescence-based activity-based profiling) and 11 biotin-labeled A*f*BPs (for subsequent pull-down/target identification). Introduction of the photoreactive moiety benzophenone (Bp) enabled us to covalently capture potential protein binders from within the proteome and perform MS analysis for de novo target identification. A variety of "click" conditions were investigated; eventually, conditions (0.1 eq $CuSO_4$/0.4 eq NaAsc with DMSO/H_2O as co-solvents) were found that afforded the desired 22 A*f*BPs with excellent yields and purities, within 12 h.

Activity-Based Profiling of the A*f*BP Library with Recombinant Cathepsin D

To demonstrate the ability of our SMM-derived A*f*BPs for UV-initiated, activity-based profiling of aspartic proteases, we first carried out in-gel fluorescence scanning of recombinant cathepsin D labeled with each of the eleven tetraethylrhodamine (TER) A*f*BPs. Cathepsin D is a well-known lysosomal aspartic protease endogenously expressed in mammalian cells and has been reported to play an important role in protein degradation and tumor progression. Following previously published protocols with some modifications,[9c] individual probes were incubated with cathepsin D for 30 min before being exposed to UV irradiation (at ~350 nm) for 25 min. This was followed by SDS-PAGE separation and in-gel fluorescence scanning. As shown in Figure 4a, one distinct labeled band, which corresponded to cathepsin D, was highly visible across all eleven probes, with varying labeling intensities. It thus indicated that all selected probes could positively label recombinant cathepsin D, albeit with different degrees of reactivity. The labeling intensity of each probe also indicated the relative binding preference of the enzyme to the corresponding hydroxyethylamine pharmacophore. In order to investigate the detection limit of the labeling reaction, varied amounts of recombinant cathepsin D were labeled by a cocktail of the eleven probes (that is, a mixture of an equal amount of each probe; Figure S7 in the Supporting Information); results showed proportional increases in fluorescence intensity of the labeled cathepsin D with increased amounts of the enzyme. We estimated that by using this method, as little as 1.5 pmol of cathepsin D could be readily detected.

In vitro Proteome Profiling

To further investigate potential cellular binding targets of the probes within complex mammalian proteomes, we next carried out in vitro proteome profiling experiments. Four TER probes (**T1**, **T3**, **A3**, and **A6**) were used to label the eight aforementioned mammalian cell lysates. Briefly, each



Scheme 2. Hit-to-probe conversion by click chemistry and structures of the twenty-two A*f*BPs. DMSO = dimethyl sulfate.

probe (1 μM final probe concentration; an equal amount of lysates was used in each labeling reaction) was treated with each of the eight mammalian cellular lysates for 30 min followed by 25 min UV irradiation. Subsequent SDS-PAGE separation and in-gel fluorescence scanning revealed the sub-proteome binding profiles of the mammalian lysates as a result of positive interactions between endogenous proteins and the hydroxyethylamine-derived probe. As shown in Figure 4b, different human cell lines were each distinctly labeled despite obvious similarities. For example, in T47D and HepG2 cancer cell lines, many more positively labeled bands were observed, indicating that significantly more positively interacting proteins were present in the total cellular proteome of these cell lines. On the contrary, far fewer fluorescently labeled bands were observed for MDA-MB-435 and HCT116 cell lines (which was consistent with their earlier comparatively weaker fluorescence binding signals on the SMM; Figure 3). It is also interesting to note that, while the four TER probes did generate mostly similar proteome profiles within each mammalian cell line as expected, relative fluorescence intensities amongst the labeled bands varied from one probe to another.

Next, all eleven probes were used to label T47D cell lysates in order to obtain comprehensive proteome "fingerprinting" profiles as a result of hydroxyethylamine/protein binding interaction (Figure 4c); each probe again showed highly similar yet distinctive proteome labeling profiles, indicating that the both R_1/R_2' residues exerted a considerable influence over probe interactions with cellular protein targets. The experiments also revealed that each probe had different reactivity profiles. For example, a 90 kDa protein showed extremely strong labeling with probes **T4**, **A3**, **A4**, and **A6** (labeled with an arrow in Figure 4c), but not with other probes. However, a 55 kDa-protein was positively labeled by most of the probes except probe **A4** and **A5** (labeled with an asterisk in Figure 4c). To validate whether this 55 kDa labeled band corresponded to endogenous cathepsin D, which is widely over-expressed in human breast cancer cells, we performed pull-down experiments using the biotinylated probe of **T2** which had the best labeling performance for T47D cell lysates based on the labeling profile. Upon labeling with probe, T47D cell lysates were subjected to affinity enrichment and subsequently analyzed by immunoblotting with cathepsin D antibody (right gel in Figure 4c); results clearly showed that cathepsin D was successfully labeled and isolated by the biotinylated probe of **T2**.

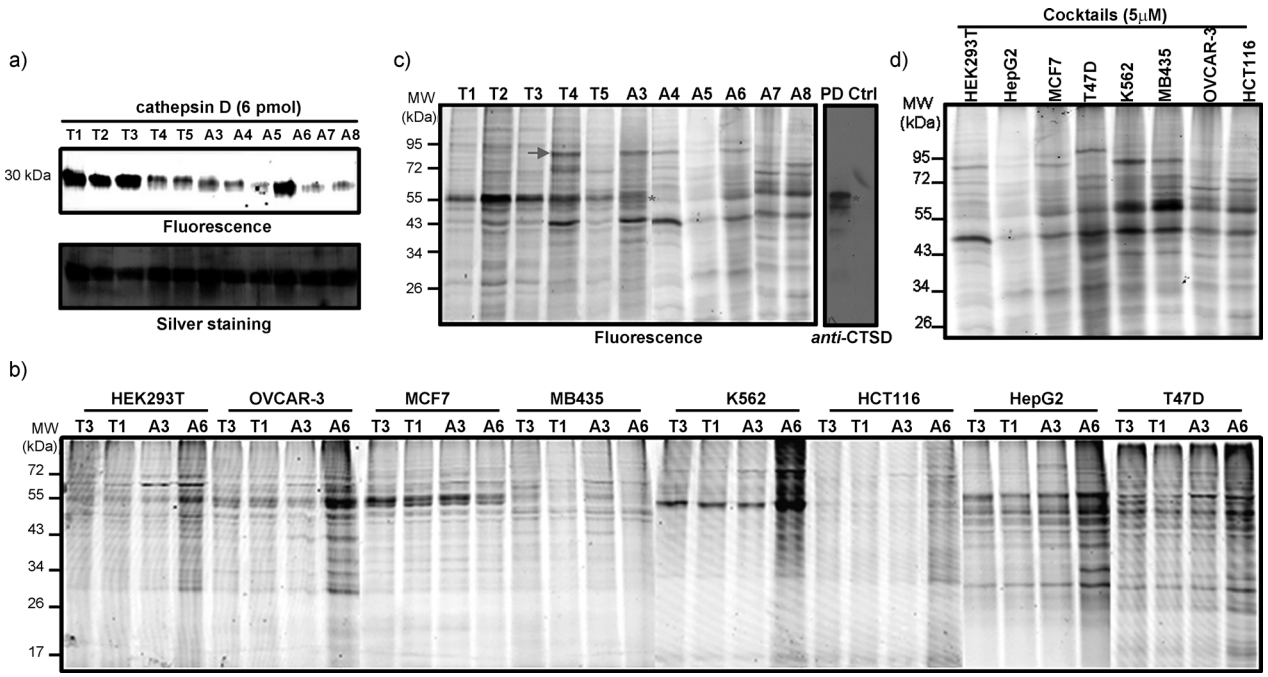


Figure 4. In-gel fluorescence images of recombinantly purified cathepsin D and mammalian cell lysates labeled with different A*f*BPs (individual probe or/and mixture cocktails). a) Fluorescence labeling profiles of recombinant cathepsin D with 11 individual fluorescently labeled TER probes. b) Fluorescence labeling profiles of eight mammalian cell lysates with four representative TER probes (**T3**, **T1**, **A3**, and **A6**; structures are shown in Scheme 2). The final concentration of each probe used in the labeling reaction was 1 μM. c) In vitro proteome profiling of T47D cell lysates with 11 individual TER probes. The asterisk indicates the protein band corresponding to cathepsin D, which was confirmed by pull-down/immunoblotting with *anti*-cathepsin D antibody (right gel). The arrow indicates the distinct 90 kDa protein band that appeared only in selected cell lines. d) In vitro proteome profiles of eight different mammalian cell lysates with a mixture cocktail of 11 TER probes. The final concentration of each probe within the cocktail was 0.5 μM. Abbreviations: PD = positive pull-down assay; Ctrl = negative pull-down assay (avidin beads without probe).

This experiment thus demonstrates the feasibility of our affinity-based probe for profiling of endogenous cathepsin D in cellular lysates. At this stage, attempts were not made to determine whether some other fluorescently labeled bands in our proteome profiling experiments corresponded to other endogenous aspartic proteases (vide infra).

While the detailed proteome profiling analysis of T47D cell lysates with the 11 individual probes provided useful information about the interaction between cellular proteins and the hydroxyethylamine pharmacophore, the experiment itself was laborious, severely limiting these probes' potential application in large-scale proteome profiling against multiple mammalian cell proteomes. Previously, a cocktail-based approach, in which a mixture of several activity-based probes instead of individual probes, had been successfully applied in the profiling of metalloprotease activities in complex proteomes.[24] We speculated that a similar approach may be adopted herein to increase the throughput of our profiling experiments, therefore enabling much more comprehensive proteome profiles of multiple mammalian cell lysates to be readily obtained within a reasonable time frame. We therefore applied an equal amount of all eleven A*f*BPs as a single mixture cocktail in our profiling experiments. By applying the same labeling protocol, the cocktail (comprising 0.5 μM of each probe, giving a final probe concentration of ~5 μM) was applied to each of the eight mammalian cellular lysates (Figure 4d); in general, cancer cell line K562 and three human breast cancer cell lines, namely MCF-7, MDA-MB-435, and T47D, showed similar proteome labeling profiles, with the exception of a 100 kDa band which appeared only in the T47D cell line. Similar proteome labeling profiles were also observed for ovary cancer cell line OVCAR-3 and colon cancer cell line HCT116. Comparatively weak proteome labeling profiles were seen for HEK293T and HepG2 cell lines. Collectively, the success of this cocktail approach indicates it may be more widely adopted in future for large-scale proteome profiling experiments, and, in some cases, to differentiate different types of mammalian cell lysates.

Pull-Down Experiments and Target Identification

Next, in order to identify potential cellular targets of the hydroxyethylamine pharmacophore, we performed affinity pull-down experiments with the biotinylated version of the probe cocktail. The mixture of 11 biotinylated probes (0.5 μM of each probe) was incubated with T47D cell lysates for 30 min, followed by 25 min UV irradiation to initiate covalent cross-linking of interacting protein-probe pairs. Upon precipitation with acetone and washing with methanol, the resulting biotinylated proteins were reconstituted with SDS buffer, then affinity-enriched with Neutravidin beads. Fol-

lowing SDS-PAGE gel separation and silver staining (Figure 5), the entire gel lane from the pull-down sample was then cut into slices and each slice was independently processed for in-gel trypsin digestion as described in the Experimental Section. As a negative pull-down control, the entire experiment was repeated with the same cell lysates treated with 5% DMSO instead of the actual probe cocktail.

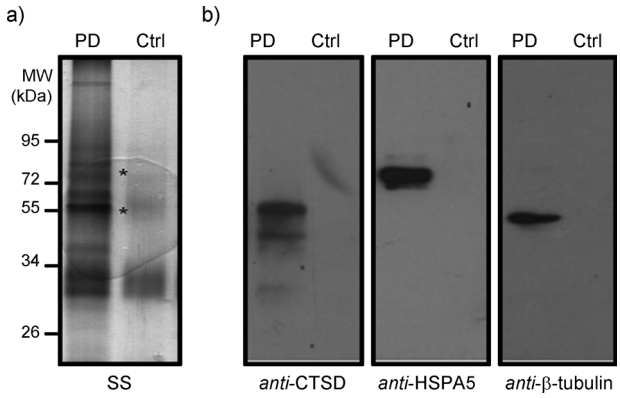


Figure 5. Pull-down experiments using the biotinylated probe cocktail and subsequent target validation of selected "hits" by Western blotting analysis. a) Silver-stained gel of the samples from pull-down experiment. Asterisks show the expected locations of cathepsin D, HSP5A, and β-tubulin. b) Western blotting analysis of pulled-down fractions of T47D cell lysates treated with the biotinylated probe cocktail (or negative controls without the cocktail; right lanes) and target validation of selected hits with their respective antibodies. Abbreviations: SS, Silver stain; PD, positive pull-down assay; Ctrl, Neutravidin beads without probes (5% DMSO instead).

Following LC-MS/MS analysis of the trypsin-digested samples, all proteins were identified by setting a minimum MOWSE (molecular weight search) score of 40 as the cut-off threshold. In addition, proteins that appeared in the negative control pull-down experiment (beads only) were excluded as potential false positives. In order to further minimize false positives in our protein list and obtain more reliable putative hits, we also performed competitive pull-down experiments in the presence of pepstatin A which is a well-known general aspartic protease inhibitor. Briefly, the entire positive pull-down/LC-MS experiment was repeated, with the introduction of a pre-incubation step of T47D lysates with pepstatin A (100 μM) for 30 min, prior to the addition of the probe cocktail. By further subtracting the protein list obtained from the pepstatin A-competitive pull-down experiment from that of the positive pull-down experiment, a total of 39 proteins were finally identified as highly confident putative proteins hits (Table 1). It should be noted that all MS-based results obtained herein (including all the putative protein hits listed in Table 1) should only be used as preliminary data. In our current study, deliberate efforts were made to improve our pull-down/LC-MS results. Nevertheless, owing to the highly complex cellular environment and the intrinsic limitation of affinity pull-down assay and mass spectrometry,[25,26] false positives/non-specific protein binders could be minimized but not eliminated entirely from our results. Consequently, proper follow-up studies and validation experiments will be needed before any biological conclusion can be made to some of these protein hits. Of the 39 identified proteins, cathepsin D was one of them. This result is well expected as our previous immunoblotting results had independently verified this (Figure 4c). It is interesting to note that missing from the list are other known mammalian aspartic proteases, which might have been caused by their relatively lower expression level in the T47D cell line. Among other proteins identified, both HSP5A and β-tubulin are two of the well-known interactors of cathepsin D.[27,28] Their positive interaction with our A*f*BPs were further confirmed by pull-down/western blotting with their respective antibodies (Figure 5b). Another two proteins known to be regulated by cathepsin D, namely superoxide dismutase (SOD) and eukaryotic translation elongation factor 1 delta isoform 4 (eEF1),[27] were also identified. Notably, a known pepstatin A-binding protein, disulfide-isomerase (PDI),[29] was also identified with a high MOWSE score. Additionally, a number of other targeted proteins were pulled down including one well-known serine protease (lon protease homolog or LONP1), several ribosomal proteins (60S ribosomal protein L5 or RPL5 and isoform 1 of plectin-1 or PLEC1), RNA recognition proteins (isoform 1 of heterogeneous nuclear ribonucleoprotein M or HNRNPM and Isoform 1 of heterogeneous nuclear ribonucleoprotein Q SYNCRIP), leucyl-tRNA synthetase, cytoplasmic (or LARS), and DNA-directed RNA polymerase II subunit RPB2 (or POLR2B). At this stage, we do not have further evidence to establish whether these hits directly interacted with our probes as well as hydroxyethylamine pharmacophores in general, or were captured indirectly through interactions with other binding proteins. Another plausible explanation is that these hydroxyethylamine moieties may also potentially interact/target other protein classes besides aspartic proteases, and some of the hits identified herein may be considered potential off-targets of hydroxyethylamine-derived drugs in cellular environments.[13c] Collectively, it demonstrates that our SMM-facilitated A*f*BP strategy should potentially be useful in the application of large inhibitor libraries for comparative proteomic profiling, with the potential for novel and disease-related biomarker identification.

Conclusions

We have synthesized a new 86-member library of hydroxyethylamine-derived pharmacophores in this study. By combining with a previously published compound library, we have, for the first time, developed a small molecule microarray (SMM) platform for facilitated screening and identification of potential affinity-based probes (A*f*BPs) suitable to be used in large-scale profiling of multiple mammalian cell lysates. Aided by our SMM screening results, a panel of

CHEMISTRY
AN ASIAN JOURNAL

Table 1. Proteins identified by pull-down and mass spectrometry.

#	ID	Protein Description	MS	Score	Peptides matched	Family
1	IPI00011229	Cathepsin D	45.0	163	4	Aspartic protease
2	IPI00023598	Tubulin beta-4 chain	50	1497	74	Tubulin family
3	IPI00786995	protein kinase, DNA-activated, catalytic polypeptide	47	801	38	Protein kinase
4	IPI00299402	Pyruvate carboxylase, mitochondrial	130.3	531	21	Biotin binding
5	IPI00646779	TUBB6 protein	50.5	531	23	Tubulin family
6	IPI00017454	Putative tubulin-like protein alpha-4B	27.8	510	18	Tubulin family
7	IPI00024993	Enoyl-CoA hydratase, mitochondrial	31.8	412	14	Enoyl-CoA hydratase
8	IPI00171903	Isoform 1 of Heterogeneous nuclear ribonucleoprotein M	77.7	376	22	RNA recognition domain
9	IPI00020984	highly similar to Calnexin	71.9	289	13	Calreticulin family
10	IPI00641706	TUBB6 46 kDa protein	46.3	259	22	Tubulin family
11	IPI00000494	60S ribosomal protein L5	34.6	244	10	Ribosomal protein
12	IPI00014898	Isoform 1 of Plectin-1	53.3	188	19	Ribosomal protein L18P
13	IPI00037070	HSPA8 54 kDa protein	53.7	187	10	Plakin or cytolinker
14	IPI00010796	Protein disulfide-isomerase	57.5	181	6	Heat shock protein
15	IPI00420014	Isoform 1 of U5 small nuclear ribonucleoprotein 200 kDa helicase	24.6	180	6	Helicase family.
16	IPI00292496	Tubulin beta-8 chain	50.3	177	17	Tubulin
17	IPI00003362	HSPA5 protein	72.5	175	10	Heat shock protein
18	IPI00003925	Isoform 1 of Pyruvate dehydrogenase E1 component subunit beta,	39.6	172	5	–
19	IPI00022891	ADP/ATP translocase 1	33.3	167	9	Mitochondrial carrier
20	IPI00029737	Isoform Long of Long-chain-fatty-acid–CoA ligase 4	80.2	156	10	ATP-dependent AMP-binding enzyme family.
21	IPI00005040	Medium-chain specific acyl-CoA dehydrogenase, mitochondrial	47	128	4	Acyl-CoA dehydrogenase
22	IPI00219682	Erythrocyte band 7 integral membrane protein	31.9	125	5	Bad 7/mec-2 family
23	IPI00943207	Ribonucleoside-diphosphate reductase	81.7	125	9	–
24	IPI00103994	Leucyl-tRNA synthetase, cytoplasmic	135.6	123	5	Class-I aminoacyl-tRNA synthetase family.
25	IPI00003348	Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-2	38.05	107	5	WD repeat G protein
26	IPI00027808	DNA-directed RNA polymerase II subunit RPB2	135.2	107	4	RNA polymerase beta chain family
27	IPI00022314	Superoxide dismutase [Mn], mitochondrial	24.9	99	6	iron/manganese superoxide dismutase
28	IPI00009634	Sulfide:quinone oxidoreductase, mitochondrial	50.2	94	3	SQRD family.
29	IPI00031517	DNA replication licensing factor MCM6	93.8	92	3	MCM family
30	IPI00413641	Aldose reductase	36.2	89	10	Aldo/keto reductase
31	IPI00465233	Eukaryotic translation initiation factor 3, subunit E interacting protein	71.1	86	2	–
32	IPI00005158	Lon protease homolog, mitochondrial	106.9	81	4	Serine protease
33	IPI00021048	Isoform 1 of Myoferlin	23.6	81	2	Ferlin family
34	IPI00009253	Alpha-soluble NSF attachment protein	33.7	80	4	SNAP family
35	IPI00018140	Isoform 1 of Heterogeneous nuclear ribonucleoprotein Q	69.8	80	3	RNA recognition
36	IPI00290462	Carbonyl reductase [NADPH] 3	31.2	80	4	Dehydrogenases
37	IPI00064086	Eukaryotic translation elongation factor 1 delta isoform 4	28.7	78	2	EF-1-beta family
38	IPI00169383	Phosphoglycerate kinase 1	44.9	73	4	Kinase
39	IPI00008986	Large neutral amino acids transporter small subunit 1	55.7	72	1	L-type amino acid transporter

A*f*BPs were rapidly synthesized by click chemistry. Subsequent in vitro proteome profiling experiments confirmed these probes are useful tools for interrogation of endogenous aspartic proteases present in a variety of mammalian cell lines. A 'cocktail' screening approach was further developed to facilitate rapid and simultaneous profiling of multiple cellular lysates. Large-scale affinity pull-down/LC-MS identifications of putative protein hits enabled us to identify both direct and indirect binding targets of the hydroxyethylamine pharmacophore. Our SMM-facilitated A*f*BP discovery strategy thus might provide a useful tool for future discovery of new aspartic proteases and other potential cancer biomarkers.

Experimental Section

Chemicals and Antibodies

Pepstatin A and plain glass slides were purchased from Sigma Aldrich (USA), and modified to generate the corresponding avidin-coated surface as previously described.[30] Recombinant cathepsin D was purchased from Invitrogen, and monoclonal cathepsin D antibody was purchased from Santa Cruz.

Cell Culture Conditions

HEK293T, HepG2, OVCAR-3, T47D, MDA-MB-435, HCT116, and MCF7 cells were grown in DMEM (Invitrogen, Carlsbad, CA) containing 10% heat-inactivated fetal bovine serum (FBS; Gibco Invitrogen), 100 U mL^{-1} penicillin, and 100 g mL^{-1} streptomycin (Thermo Scientific, Rockford, IL) and maintained in a humidified 37°C incubator with 5%

CO_2. K562 cells were grown in RPMI 1640 medium (Invitrogen Corporation) supplemented with 10% heat-inactivated fetal bovine serum, 100 U mL^{-1} penicillin, and 100 g mL^{-1} streptomycin and maintained in a humidified 37 °C incubator with 5% CO_2.

Preparation of Cellular Lysates

Mammalian lysates were prepared as follows: fresh cell pellets were thawed on ice by suspension in 100 μL of lysis buffer (50 mM HEPES, *p*H 5.4, 1 mM EDTA, 100 mM NaCl, 10% Glycerol, 0.1% NP-40), incubated for 15 min on ice, and then spun for 20 min at 13000 rpm at 4 °C. The cleared supernatant was decanted to an Eppendorf tube and kept at 4 °C. Protein concentration was quantified by Bradford assay (Bio-Rad, USA), aliquoted and stored in −80 °C, and used for subsequent screening on array.

Small Molecule Microarray Preparation and Screening

Procedures for the construction of the SMM were based on previously published protocols,[30] and are described in detail in the Supporting Information. Cellular lysates were minimally labeled with the Cy3 dye for 1 h on ice following the manufacturer's protocols. The excessive dye was quenched with a 10-fold molar excess of hydroxylamine for a further 1 h, then removed by buffer exchange with a Microcon centrifugal filter (Millipore, USA) or extensive dialysis at 4 °C overnight (Amersham, GE Healthcare, USA). The labeled lysate was reconstituted in a final buffer volume of 30 μL of HEPES (*p*H 5.4) containing 1% bovine serum albumin. In a standard microarray experiment, the labeled lysate (4 μg in 30 μL of HEPES) was applied under a cover slip to the array. The samples were then incubated with the array in a humidified chamber for 1 h at RT before repeated rinses with PBST (PBS containing 0.05% of Tween 20) with gentle shaking. Slides were scanned using an ArrayWoRx microarray scanner installed with the relevant filters (Cy3, λex/em = 548/595 nm).

Probe Labeling and In-gel Fluorescence Scanning

For labeling of individual probes, lysates from mammalian cells (~50 μg of total proteins as determined by Bradford assay) were incubated with the TER probe (1 μM final concentration; 5% DMSO) in HEPES buffer (50 mM HEPES, *p*H 5.4, 1 mM EDTA, 100 mM NaCl) for 30 min at RT, then irradiated on ice for 25 min using a B100 A lamp (UVP) at a distance of 5 cm. The reactions were then resuspended in 50 μL 1 × SDS-loading buffer followed by heating for 10 min at 95 °C; around 20 μg of protein was loaded per gel lane for separation by SDS-PAGE (12% SDS gel), then visualized by in-gel fluorescence scanning using a Typhoon 9410 Variable Mode Imager scanner. For cocktail labeling, lysates from mammalian cells (~50 μg of total proteins) were incubated with a cocktail of 11 TER probes (0.5 μM of each probe, 5% DMSO) in HEPES buffer (50 mM HEPES, *p*H 5.4, 1 mM EDTA, 100 mM NaCl) for 30 min at RT, then irradiated on ice for 25 min using a B100 A lamp (UVP) at a distance of 5 cm. The reaction was quenched by addition of pre-chilled acetone (0.5 mL), then placed at −20 °C for 30 min. Upon centrifugation at 13,000 rpm for 10 min at 4 °C to precipitate proteins, the supernatant was discarded and the resulting pellet was washed two times with 200 μL of pre-chilled methanol, then resuspended in 50 μL 1 × SDS-loading buffer followed by heating for 10 min at 95 °C; around 20 μg of protein was loaded per gel lane for separation by SDS-PAGE (12% SDS gel), then visualized by in-gel fluorescence.

Affinity Pull Down and Mass Spectrometric Target Identification

Two sets of pull-down assays were carried out in this study. For pull down with an individual probe, biotinylated probe **T2** (10 μM final concentration) was incubated with fresh T47D cell lysates (~5 mg) for 0.5 h at 4 °C in an acidic buffer (50 mM HEPES, *p*H 5.4, 1 mM EDTA, 100 mM NaCl). Meanwhile, the same volume of DMSO was incubated with T47D cell lysates as the negative control experiment. Subsequently, the samples were irradiated on ice for 25 min under UV lamp. The resulting mixtures were incubated with Neutravidin agarose beads for further 2 h at 4 °C. After that, the supernatant was removed by centrifugation, and the beads were washed with 0.1% SDS in PBS once, and washed four times with PBS. After washing, the beads were boiled in the elution buffer (200 mM Tris at *p*H 6.8 containing 400 mM DTT and 4% SDS) for 15 min. Eluted proteins were separated on 12% SDS-PAGE gel (together with negative pull-down control), then visualized by coomassie blue/silver staining.

In order to identify potential cellular targets of the hydroxyethylamine pharmacophore, pull-down experiments with a cocktail of 11 biotinylated probes were also carried out. Fresh T47D cell lysates (~5 mg) were incubated with the cocktail of 11 biotinylated probes (0.5 μM for each probe) for 0.5 h at 4 °C in an acidic buffer (50 mM HEPES, *p*H 5.4, 1 mM EDTA, 100 mM NaCl). Similarly, the same volume of DMSO was incubated with T47D lysates as a negative control. Subsequently the reactions were irradiated on ice for 25 min under UV lamp and quenched by pre-chilled acetone (10 mL) before being placed at −20 °C for 30 min. The mixtures were then centrifuged at 13,000 rpm for 10 min at 4 °C to precipitate the proteins. The supernatant was discarded and the pellet was washed two times with 10 mL of pre-chilled methanol. The protein pellets were resolubilized with 0.1% SDS-PBS buffer. Next, the solutions were incubated with Neutravidin agarose beads for 2 h at 4 °C. The supernatant was removed by centrifugation, and the beads were washed with 0.1% SDS in PBS once, and washed four times with PBS. After washing, the beads were boiled in the elution buffer (200 mM Tris at *p*H 6.8 containing 400 mM DTT and 4% SDS) for 15 min. Eluted proteins were separated on 12% SDS-PAGE gel (together with negative pull-down control), then visualized by coomassie blue/silver staining.

Each gel lane was cut into multiple slices. Subsequently, trypsin digestion (using In-Gel Trypsin Digestion Kit, Pierce) and peptide extraction (with 50% acetonitrile and 1% formic acid) were carried out. The samples were then dried in vacuo and stored at −20 °C. LC-MS/MS analysis was performed using an LTQ-FT ultra mass spectrometer (Thermo Electron, Germany) coupled with an online Shimadzu UFLC system utilizing nanospray ionization. Peptides were first enriched with a Zorbax 300SB C18 column (5 mm × 0.3 mm, Agilent Technologies), followed by elution into an integrated nanopore column (75 μm × 100 mm) packed with C18 material (5 μm particle size, 300 Å pore size). Mobile phase A (0.1% formic acid in H_2O) and mobile phase B (0.1% formic acid in acetonitrile) were used to establish the 90 min gradient, comprising 3 min of 0–5% B, then 52 min of 5–30% B, followed by 12 min of 30–60% B; maintained at 80% B for 8 min before re-equilibrating at 5% B for 15 min. Sample was injected into the MS with an electrospray potential of 1.8 kV without sheath and auxiliary gas flow, ion transfer tube temperature of 180 °C, and collision gas pressure of 0.85 mTorr. A full-survey MS scan (350–2000 *m*/*z* range) was acquired in the 7-T FT-ICR cell at a resolution of 100000 and a maximum ion accumulation time of 1000 ms. Precursor ion charge-state screening was activated. The linear ion trap was used to collect peptides where the 10 most intense ions were selected for collision-induced dissociation (CID) in MS2, performed concurrently with a maximum ion accumulation time of 200 ms. Dynamic exclusion was activated for this process, with a repeat count of one and exclusion duration of 30 s. For CID, the activation Q was set at 0.25, isolation width (*m*/*z*) 2.0, activation time 30 ms, and normalized collision energy 35%. The Extract-Msn (version 4.0) program found in Bioworks Browser 3.3 (Thermo Electron, Germany) was used to extract tandem MS spectra in the data format from the raw data of the LTQ-FT ultra. These data files were then converted into the MASCOT generic file format using an in-house program. Intensity values and fragment ion *m*/*z* ratios were not manipulated. These data were used to obtain protein identities by searching against the corresponding database by means of an in-house MASCOT server (version 2.2.03, Matrix Science, Boston, MA). The search was limited to a maximum of two missed trypsin cleavages, #13C of 2, mass tolerances of 10 ppm for peptide precursors and 0.8 Da for fragment ions. Only proteins with a MOWSE score higher than 40, corresponding to $p < 0.05$, were considered significant. The peptide/protein lists obtained were exported as an html file.

For immunoblotting experiments, samples from pull-down experiments were resolved by SDS-PAGE gel and transferred to polyvinylidene fluoride membranes. Membranes were then blocked with 3% BSA in TBST (0.05% Tween in TBS) for 1 h at RT. After blocking, membranes were incubated with cathepsin D primary antibody (1:500), HSP5A primary

antibody (1:500), and β-tubulin (1:500) for 1 h at RT, respectively. The membranes were then washed four times with TBST (10 min/time), and incubated with the respective secondary antibody. After 1 h incubation at RT, blots were washed again with TBST before the development with SuperSignal West Dura Kit (Thermo Scientific).

Chemical Synthesis

All chemicals were purchased at the highest grade available from commercial vendors and used without further purification, unless otherwise noted. All reactions were carried out under N_2 atmosphere with HPLC-grade solvents, unless otherwise stated. Reaction progress was monitored by TLC on pre-coated silica plates (Merck 60 F254, 250 μm thickness) and spots were visualized by ceric ammonium molybdate, basic $KMnO_4$, or UV light. ^{1}H NMR and ^{13}C NMR spectra were recorded on a Bruker Avance 300 MHz, or Bruker model DPX-500 MHz NMR spectrometer. Chemical shifts are reported in parts per million relative to internal standard tetramethylsilane ($Si(CH_3)_4 = 0.00$ ppm) or residual solvent peaks ($CHCl_3 = 7.26$ ppm). Mass spectra were obtained at the Center for Chemical Characterization and Analysis at NUS. Analytical HPLC was carried out on Shimadzu LC-IT-TOF or Shimadzu LC-ESI system equipped with an autosampler, using reverse-phase Phenomenex Luna 5 μm C18 100 Å 50×3.0 mm columns. TFA/H_2O (0.1%) and TFA/acetonitrile (0.1%) were used as eluents for all HPLC experiments.

Synthesis of α-Azido alcohols 4(d–f)

α-Azido ketones **3** (5 mmol) were dissolved in 20 mL of THF followed by addition of sodium borohydride (6 mmol) in 2 mL H_2O at 0 °C. The reaction mixture was then stirred for 1 h and neutralized with aqueous 1 N HCl. After extraction, the organic layers were washed with brine and dried over anhydrous Na_2SO_4, and concentrated under reduced pressure. Purification by silica-gel chromatography with 85:15 hexane/EtOAc afforded α-azido alcohol **4(d–f)** in 50–70%.

N-α-Fmoc-isopropylalanylazidoalcohol (4d)

Yield = 79%. ^{1}H NMR (500 MHz, $CDCl_3$): $\delta = 0.72$ (d, $J = 9.55$ Hz, 6H), 2.58 (d, $J = 5.65$ Hz, 1H), 3.19–3.41 (m, 2H), 3.53–3.58 (m, 1H), 3.67 (s, 1H), 4.49 (t, $J = 5.98$ Hz, 1H), 4.59–4.64 (m, 2H), 7.33 (t, $J = 7.57$ Hz, 2H), 7.40 (t, $J = 3.78$ Hz, 2H), 7.59 (t, $J = 5.98$ Hz, 2H), 7.77 ppm (d, $J = 7.55$ Hz, 2H); ^{13}C NMR (125 MHz, $CDCl_3$): $\delta = 157.1$, 143.7, 141.5, 127.7, 127.1, 124.9, 120.0, 71.2, 66.3, 58.6, 54.4, 47.5, 28.0, 20.0 ppm; LC-MS: m/z $[M+Na]^+ = 403.154$.

N-α-Fmoc-4-tert-butoxy-phenylalanylazidoalcohol (4e)

Yield = 90%. ^{1}H NMR (500 MHz, $CDCl_3$): $\delta = 1.30$ (s, 9H), 2.78–2.84 (m, 2H), 3.25–3.32 (m, 2H), 3.72–3.78 (m, 2H), 4.11–4.17 (m, 1H), 4.39–4.45 (m, 2H), 4.79 (d, $J = 6.95$ Hz, 1H), 5.08 (d, $J = 8.85$ Hz, 1H), 6.88 (d, $J = 7.55$ Hz, 2H), 7.02–7.07 (m, 2H), 7.30 (t, $J = 7.23$ Hz, 2H), 7.39 (t, $J = 7.25$ Hz, 2H), 7.51 ppm (t, $J = 7.55$ Hz, 2H), ^{13}C NMR (125 MHz, $CDCl_3$): $\delta = 156.4$, 154.2, 143.6, 141.4, 132.2–124.8(m), 124.2, 119.9, 78.4, 72.1, 66.5, 54.9, 54.2, 47.3, 35.2, 28.8 ppm; LC-MS: m/z $[M+Na]^+ = 523.207$.

N-α-Fmoc-4-Boc-butylalanylazidoalcohol (4f)

Yield = 88%. ^{1}H NMR (500 MHz, $CDCl_3$): $\delta = 1.23$–1.28 (m, 4H), 1.45 (s, 9H), 3.00–3.09 (m, 2H), 3.24–3.28 (m, 2H), 3.48–3.75 (m, 2H), 4.19 (t, $J = 6.3$ Hz, 1H), 4.42–4.61 (m, 2H), 5.01 (s, 1H), 5.13 (d, $J = 8.85$ Hz, 1H), 7.31 (t, $J = 7.55$ Hz, 2H), 7.40 (t, $J = 7.55$ Hz, 2H), 7.58 (d, $J = 6.95$ Hz, 2H), 7.76 ppm (d, $J = 7.55$ Hz, 2H); ^{13}C NMR (125 MHz, $CDCl_3$): $\delta = 156.5$, 143.7, 141.3, 127.7, 127.0, 124.9, 119.9, 79.3, 73.1, 66.4, 60.4, 53.9, 47.4, 39.5, 31.3, 29.5, 22.4, 20.9 ppm; LC-MS: m/z $[M+H]^+ = 410.205$.

*α-Azido bromomethyl-Tyr(t*Bu)-alcohol (8d)

To the solution of α-azido bromomethyl ketones **7** (8 mmol, 1 eq) in 20 mL of 95:5 THF/H_2O, sodium borohydride (12 mmol. 1.5 eq) was added gradually at 0 °C. The reaction mixture was stirred for 1 h and then neutralized with aqueous 1 N HCl. After extraction with EtOAc, the organic extracts were washed with brine, dried with Na_2SO_4, and concentrated under reduced pressure. Purification by silica-gel chromatography with 80:20 hexanes/EtOAc afforded compound **8d** as white solid. Yield = 77%. ^{1}H NMR (300 MHz, $CDCl_3$): $\delta = 1.33$ (s, 9H), 2.51 (dd, $J = 5.07$ Hz, 1H), 2.79 (d, $J = 5.66$ Hz, 1H), 2.90–3.14 (m, 2H), 3.52–3.59 (m, 1H), 3.63–3.75 (m, 2H), 6.96 (d, $J = 8.40$ Hz, 2H), 7.17 ppm (d, $J = 6.60$ Hz, 1H); ^{13}C NMR (75 MHz, $CDCl_3$): $\delta = 154.3$, 146.9, 131.8–131.5(m), 129.8, 124.5, 72.2, 71.9, 69.2, 65.0, 37.3, 36.4, 28.8 ppm. LC-MS: m/z $[M+H]^+ = 327.350$.

Immobilization of 4(d–f) onto DHP Resin

According to previously published procedures,[9c] HM DHP resin (40 mg) was first swelled in DCE (5 mL) for 2 h. The solvent was then removed and **4(d–f)** (4.0 eq) and a catalytic amount of PPTS (pyridinium *p*-toluenesulfonate) (1.5 eq) in DCE (10 mL) were added at RT. The reaction was then stirred for 12 h at 60 °C. The resulting resin was washed with NMP (3×), THF (3×), CH_2Cl_2 (3×), and Et_2O (3×) and dried in vacuo, then stored at −20 °C.

Deprotection of Fmoc Group

The resulting resin was suspended in a solution of 20% piperidine in NMP and shaken for 2 h at RT. Upon solvent removal by filtration, the resin was washed with NMP (3×), THF (3×), CH_2Cl_2 (3×), and Et_2O (3×), then dried under reduced pressure. The progress of the reaction was monitored by the ninhydrin test. Blue beads indicate the presence of primary amine.

Coupling Reactions for Acids and Sulfonyl Chlorides

The resulting resin firstly was swelled in dry THF for 2 h, then the solvent was removed. To the resin was added a preactived solution containing the corresponding acid (4.0 eq), PyBOP (4.0 eq), HOAt (4.0 eq), and DIPEA (8.0 eq) in THF or a solution of the corresponding sulfonyl chloride (4.0 eq) and DIEA (8.0 eq) in dry THF. The reaction mixtures were shaken overnight at RT and the resin was finally washed with NMP (3×), THF (3×), CH_2Cl_2 (3×), and Et_2O (3×) before being dried under reduced pressure. The completeness of the reaction was monitored by the ninhydrin test. A negative test indicated the absence of primary amine and the completeness of the reaction.

Immobilization of 8d onto DHP Resin

HM DHP resin (40 mg) was swelled in DCE (5 mL) for 2 h. The solvent was then removed and **8d** (4.0 eq) and a catalytic amount of PPTS (1.5 eq) in DCE (10 mL) were added at RT. The reaction was then stirred for 12 h at 60 °C. The resulting resin was washed with NMP (3×), THF (3×), CH_2Cl_2 (3×), and Et_2O (3×) and dried in vacuo, then stored at −20 °C.

Amine Alkylation and Coupling Reactions for C-Terminal Library

Support-bound bromomethyl alcohols were added to a vial, followed by addition of a solution of *iso*-butylamine (10 eq) in NMP (10 mL). The vial was then sealed and the reaction mixture was heated at 80 °C for 36 h. The resin was subsequently washed with NMP (3×), THF (3×), CH_2Cl_2 (3×), and ether (3×), then dried in vacuo. Acylation with the sulfonyl chloride (4.0 eq) and DIEA (8.0 eq) in dry THF was carried out overnight. The resin was then washed with NMP (3×), THF (3×), CH_2Cl_2 (3×), and ether (3×), then dried under reduced pressure.

Reduction of Azido Group to Primary Amine

Reduction of the azide was accomplished using 0.2 M $SnCl_2$, 0.8 M PhSH, and 1.0 M Et_3N in THF (1 mL) for 4 h. The resin was then washed with 50% (vol) of aqueous THF solution (3×), THF (3×), CH_2Cl_2 (3×), and ether (3×), then dried in vacuo. The reaction was monitored by using the ninhydrin test. If primary amine is present, the beads will become blue after treatment with ninhydrin reagents.

Acylation of Primary Amine with Long-Chain Biotin Acids

The above resulting support-bound amine was then acylated in a preactived solution of biotin acid (4.0 eq), PyBOP (4.0 eq), HOAt (4.0 eq), and DIPEA (8.0 eq) in THF. The reaction mixture was shaken overnight

at RT, and the resins were washed with NMP (3×), THF (3×), CH_2Cl_2 (3×), and Et_2O (3×) and dried under vacuum. The completeness of the reaction was monitored by the ninhydrin test. If the coupling reaction is completed, the beads will exhibit no color change after treatment with ninhydrin reagents.

Cleavage of Small Molecules from the Solid Support

Each microreactor containing a member of the N- and C-terminal sublibraries was cleaved with a 3-mL cleavage solution containing TFA (95%) and H_2O (5%) for 20 min at RT. Following concentration in vacuo until >90% of the cleavage solution was removed, cold ether (chilled to −20°C) was added to the liquid residue to precipitate the compounds. The compounds were allowed to precipitate at −20°C overnight. The ether layer was discarded and the precipitates were dried thoroughly in vacuo. The solids were dissolved in 0.5 mL of DMSO and stored at −20°C for future use. LC-MS was performed to ensure the compounds were of correct mass and sufficient purity. Mass spectra were recorded on a Finnigan LCQ mass spectrometer or a Shimadzu LC-IT-TOF spectrometer. Analytical HPLC was carried out on a Shimadzu LC-IT-TOF system equipped with an autosampler, using reverse-phase Phenomenex Luna 5 μm C18(2) 100 Å 150×3.0 mm (for peptides) or 50×3.0 mm (for all other samples) columns. 0.1% TFA/H_2O and 0.1% TFA/acetonitrile were used as eluents. The flow rate was 0.6 mL min^{-1}. Each samples were run for 20 min under 10–100% ACN condition.

*Assembly of Affinity-Based Probes (Af*BPs) by Click Chemistry

Synthesis of the probes by click chemistry followed previously published procedures,[31] with minor modifications as indicated below: the alkyne (1.2 eq) and the azide (1.0 eq; final concentration: 10.0 mM) were dissolved in a minimal amount of DMSO. A mixture of DMSO/H_2O solution (1:1; 2 mL) was subsequently added, and the reaction was shaken for a few minutes to obtain a clear solution. The click chemistry was initiated by sequential addition of catalytic amounts of sodium ascorbate (0.4 eq) and $CuSO_4$ (0.1 eq). The reaction was incubated further for another 12 h with shaking at RT. The reaction product was then directly analyzed by LC-MS; results indicated the complete consumption of the azide and quantitative formation of the triazole final product in all cases. The final probes were subsequently purified by prep-HPLC and confirmed by LC-MS. Some probes were further characterized by NMR.

Biotin-T1

^{1}H NMR (300 MHz, [D$_6$]DMSO): δ=0.54 (d, J=7.02 Hz, 6H), 0.66 (d, J=6.15 Hz, 2H), 1.04–1.19 (m, 4H), 1.23–1.41 (m, 10H), 1.86–2.03 (m, 6H), 2.24 (d, J=11.40 Hz, 2H), 2.68–3.02 (m, 9H), 3.59–3.70 (m, 2H), 4.27 (s, 2H), 4.53–4.64 (m, 3H), 5.39 (s, 1H), 6.34 (s, 2H), 7.23–7.36 (m, 1H), 7.50–7.78 (m, 10H), 8.06 (s, 1H), 8.24 (d, J=8.19 Hz, 1H), 8.29–8.49 (m, 2H), 8.98 ppm (s, 1H); LC-MS: m/z $[M+H]^+$=1082.100.

Biotin-T3

^{1}H NMR (300 MHz, [D$_6$]DMSO): δ=0.51–0.65 (m, 6H), 1.19–1.23 (t, J=4.68 Hz, 10H), 1.96–2.05 (m, 4H), 2.20–2.42 (m, 4H), 2.60–2.75 (m, 3H), 2.82–3.26 (m, 10H), 4.05 (m, 1H), 4.21–4.35 (m, 4H), 6.49 (s, 2H), 6.75–6.81 (m, 2H), 6.89 (s, 2H), 7.06 (s, 2H), 7.24 (s, 2H), 7.26–7.36 (m, 3H), 7.55–7.85 (m, 10H), 8.22–8.35 (m, 2H), 8.36–8.49 ppm (m, 3H); LC-MS: m/z $[M+H]^+$=1115.150.

Acknowledgements

This work was supported by the Competitive Research Program (R143-000-218-281; NRF-G-CRP 2007-04) from the National Research Foundation (NRF) of Singapore, the Agency for Science, Technology and Research (R-143-000-391-305) and the Ministry of Education (R-143-000-394-112). We thank Jigang Wang for technical assistance and Dr. Siu Kwan Sze (Nanyang Technological University, Singapore) for advice and support for the LC-MS experiments.

[1] W. F. Patton, B. Schulenberg, T. H. Steinberg, *Curr. Opin. Biotechnol.* **2002**, *13*, 321–328.
[2] a) R. Aebersold, D. R. Goodlett, *Chem. Rev.* **2001**, *101*, 269–295; b) M. Gstaiger, R. Aebersold, *Nat. Rev. Genet.* **2009**, *10*, 617–627; c) B. F. Cravatt; G. M. Simon, J. R. Yates, *Nature* **2007**, *450*, 991–1000; G. M. Simon, J. R. Yates, *Nature* **2007**, *450*, 991–1000.
[3] a) S. Miura, K. Miura, H. Masuzaki, N. Miyake, K. Yoshiura, N. Sosonkina, N. Harada, O. Shimokawa, D. Nakayama, S. Yoshimura, N. Matsumoto, N. Niikawa, T. Ishimaru, *J. Hum. Genet.* **2006**, *51*, 412–417; b) P. P. Brown, D. Botstein, *Nat. Genet.* **1999**, *21*, 33–37.
[4] a) R. Kramer, D. Cohen, *Nat. Rev. Drug Discovery* **2004**, *3*, 965–972; b) S. D. Patterson, R. Aebersold, *Nat. Genet.* **2003**, *33*, 311–323.
[5] a) B. Kobe, B. E. Kemp, *Nature* **1999**, *402*, 373–376; b) B. F. Cravatt, A. T. Wright, J. W. Kozarich, *Annu. Rev. Biochem.* **2008**, *77*, 383–414; c) M. Uttamchandani, C. H. S. Lu, S. Q. Yao, *Acc. Chem. Res.* **2009**, *42*, 1183–1192.
[6] a) M. J. Evans, B. F. Cravatt, *Chem. Rev.* **2006**, *106*, 3279–3301; b) A. M. Sadaghiani, S. H. L. Verhelst, M. Bogyo, *Curr. Opin. Chem. Biol.* **2007**, *11*, 20–28; c) M. Uttamchandani, J. Li, H. Sun, S. Q. Yao, *ChemBioChem* **2008**, *9*, 667–675.
[7] Y. Liu, M. P. Patricelli and B. F. Cravatt, *Proc Natl Acad. Sci. USA* **1999**, *96*, 14694–14699.
[8] a) A. Borodovsky, H. Ovaa, N. Kolli, T. Gan-Erdene, K. D. Wilkinson, H. L. Ploegh, B. M. Kessler, *Chem. Biol.* **2002**, *9*, 1149–1159; b) D. Greenbaum, K. F. Medzihradszky, A. Burlingame, M. Bogyo, *Chem. Biol.* **2000**, *7*, 569–581; c) M. Uttamchandani, K. Liu, R. C. Panicker, S. Q. Yao, *Chem. Commun.* **2007**, 1518–1520.
[9] a) H. Fuwa, Y. Takahashi, Y. Konno, N. Watanabe, H. Miyashita, M. Sasaki, H. Natsugari, T. Kan, T. Fukuyama, T. Tomita, T. Iwatsubo, *ACS Chem. Biol.* **2007**, *2*, 408–418; b) K. Liu, H. Shi, H. G. Xiao, A. G. L. Chong, X. Z. Bi, Y. T. Chang, K. S. W. Tan, R. Y. Yada, S. Q. Yao, *Angew. Chem.* **2009**, *121*, 8443–8447; *Angew. Chem. Int. Ed.* **2009**, *48*, 8293–8297; c) H. Shi, K. Liu, A. Xu, S. Q. Yao, *Chem. Commun.* **2009**, 5030–5032.
[10] a) A. Saghatelian, N. Jessani, A. Joseph, M. Humphrey, B. F. Cravatt, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 10000–10005; b) E. W. S. Chan, S. Chattopadhaya, R. C. Panicker, X. Huang, S. Q. Yao, *J. Am. Chem. Soc.* **2004**, *126*, 14435–14446.
[11] a) S. Kumar, B. Zhou, F. Liang, W. Q. Wang, Z. Huang, Z. Y. Zhang, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 7943–7948; b) Q. Zhu, X. Huang, G. Y. J. Chen, S. Q. Yao, *Tetrahedron Lett.* **2003**, *44*, 2669–2672.
[12] a) M. S. Cohen, H. Hadjivassiliou, J. Taunton, *Nat. Chem. Biol.* **2007**, *3*, 156–160; b) M. S. Cohen, C. Zhang, K. M. Shokat, J. Taunton, *Science* **2005**, *308*, 1318–1321; c) K. A. Kalesh, S. B. D. Sim, J. Wang, K. Liu, Q. Lin, S. Q. Yao, *Chem. Commun.* **2010**, *46*, 1118–1120.
[13] a) R. Srinivasan, X. Huang, S. L. Ng, S. Q. Yao, *ChemBioChem* **2006**, *7*, 32–36; b) D. R. Blais, M. Brûlotte, Y. Qian, S. Bélanger, S. Q. Yao, J. P. Pezacki, *J. Proteome Res.* **2010**, *9*, 912–923; c) P.-Y. Yang, K. Liu, M. H. Ngai, M. J. Lear, M. Wenk, S. Q. Yao, *J. Am. Chem. Soc.* **2010**, *132*, 656–666.
[14] a) M. Uttamchandani, S. Q. Yao, *Curr. Pharm. Des.* **2008**, *14*, 2428–2438; b) K. Y. Tomizaki, K. Usui, H. Mihara, *ChemBioChem* **2005**, *6*, 783–799; c) T. Kodadek, *Trends Biochem. Sci.* **2002**, *27*, 295–300; d) J. L. Duffner, P. A. Clemons, A. N. Koehler, *Curr. Opin. Chem. Biol.* **2007**, *11*, 74–82.
[15] a) H. Sun, L. P. Tan, L. Gao, S. Q. Yao, *Chem. Commun.* **2009**, 677–679; b) C. H. S. Lu, H. Sun, F. B. A. Bakar, M. Uttamchandani, W. Zhou, Y.-C. Liou, S. Q. Yao, *Angew. Chem.* **2008**, *120*, 7548–7551; *Angew. Chem. Int. Ed.* **2008**, *47*, 7438–7441; c) H. Sun, C. H. S. Lu, M. Uttamchandani, Y. Xia, Y.-C. Liou, S. Q. Yao, *Angew. Chem.* **2008**, *120*, 1722–1726; *Angew. Chem. Int. Ed.* **2008**, *47*, 1698–1702; d) M. Uttamchandani, W. L. Lee, J. Wang, S. Q. Yao, *J. Am. Chem. Soc.* **2007**, *129*, 13110–13117; e) C. M. Salisbury, D. J. Maly, J. A. Ellman, *J. Am. Chem. Soc.* **2002**, *124*, 14868–14870; f) M. Köhn, M. Gutierrez-Rodriguez, P. Jonkheijm, S. Wetzel, R. Wacker, H. Schroeder, H. Prinz, C. M. Niemeyer, R. Breinbauer, S. E. Szedlac-

sek, H. Waldmann, *Angew. Chem.* **2007**, *119*, 7844–7847; *Angew. Chem. Int. Ed.* **2007**, *46*, 7700–7703; g) N. Winssinger, R. Damoiseaux, D. C. Tully, B. H. Geierstanger, K. Burdick, J. L. Harris, *Chem. Biol.* **2004**, *11*, 1351–1360; h) M. M. Reddy, T. Kodadek, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 12672–12677; i) R. L. W. Roska, T. G. S. Lama, J. P. Henne, R. E. Carlson, *J. Am. Chem. Soc.* **2009**, *131*, 16660–16662; j) K. Usui, K. Tomizaki, H. Mihara, *Mol. BioSyst.* **2006**, *2*, 417–420.

[16] J. Hardy, D. J. Selkoe, *Science* **2002**, *297*, 353–356.

[17] H. Wu, J. Ge, P.-Y. Yang, J. Wang, M. Uttamchandani, S. Q. Yao, *J. Am. Chem. Soc.* **2011**, *133*, 1946–1954.

[18] http://merops.sanger.ac.uk/.

[19] a) R. C. Gallo, P. S . Sarin, E. P. Gelmann, M. Robert-Guroff, E. Richardson, V. S. Kalyanaraman, *Science* **1983**, *220*, 865–867; b) F. Barre-Sinoussi, J. C. Chermann, F. Rey, M. T. Nugeyre, S. Chamaret, J. Gruest, *Science* **1983**, *220*, 868–871; c) R. Tigerstedt, P. G. Bergman, Kreislauf. N, A. Skand, *Physiologie* **1898**, *8*, 223–271; d) H. Rochefort, F. Capony, M. Garcia, V. Cavailles, G. Freiss, M. Chambon, *J. Cell. Biochem.* **1987**, *35*, 17–29; e) C. Boss, S. Richard-Bildstein, T. W. Fischli, S. Meyer, C. Binkert, *Curr. Med. Chem.* **2003**, *10*, 883–907.

[20] a) E. E. Kim, C. T. Baker, M. D. Dwyer, M. A. Murcko, B. G. Rao, R. D. Tung, M. A. Navia, *J. Am. Chem. Soc.* **1995**, *117*, 1181–1182; b) S. W. Kaldor, M. Hammond, B. A. Dressman, J. E. Fritz, T. A. Crowell, R. A. Hermann, *Bioorg. Med. Chem. Lett.* **1994**, *4*, 1385–1390.

[21] a) S. K. Mamidyala, M. G. Finn, *Chem. Soc. Rev.* **2010**, *39*, 1252–1261; b) E. M. Sletten, C. R. Bertozzi, *Angew. Chem.* **2009**, *121*, 7108–7133; *Angew. Chem. Int. Ed.* **2009**, *48*, 6974–6998; c) M. Meldal, C. W. Tornøe, *Chem. Rev.* **2008**, *108*, 2952–3015; d) K. A. Kalesh, H. Shi, J. Ge, S. Q. Yao, *Org. Biomol. Chem.* **2010**, *8*, 1749–1762.

[22] a) M. Chino, M. Wakao, J. A. Ellman, *Tetrahedron* **2002**, *58*, 6305–6310; b) A. Brik, J. Muldoon, Y.-C. Lin, J. H. Elder, D. S. Goodsell, A. J. Olson, V. V. Fokin, K. B. Sharpless, C.-H. Wong, *ChemBioChem* **2003**, *4*, 1246–1248.

[23] M. Uttamchandani, J. Wang, J. Li, M. Hu, H. Sun, K. Y.-T. Chen, K. Liu, S. Q. Yao, *J. Am. Chem. Soc.* **2007**, *129*, 7848–7858.

[24] S. A. Sieber, S. Niessen, H. S. Hoover, B. F. Cravatt, *Nat. Chem. Biol.* **2006**, *2*, 274–281.

[25] a) S. E. Ong, M. Mann, *Nat. Chem. Biol.* **2005**, *1*, 252–262; b) M. Bantscheff, M. Schirle, G. Sweetman, J. Rick, B. Kuster, *Anal. Bioanal. Chem.* **2007**, *389*, 1017–1031.

[26] a) W. Zhu, J. W. Smith, C. M. Huang, *J. Biomed. Biotechnol.* **2010**, *2010*, 840518–840523; b) H. Liu, R. G. Sadygov, J. R. Yates, 3rd, *Anal. Chem.* **2004**, *76*, 4193–4201; c) Y. Ishihama, Y. Oda, T. Tabata, T. Sato, T. Nagasu, J. Rappsilber, M. Mann, *Mol. Cell. Proteomics* **2005**, *4*, 1265–1272; d) K. Shinoda, M. Tomita, Y. Ishihama, *Bioinformatics* **2010**, *26*, 576–577.

[27] M. A. Bewley, T. K. Pham, H. M. Marriott, J. Noirel, H.-P. Chu, S. Y. Ow, A. G. Ryazanov, R. C. Read, M. K. B. Whyte, B. Chain, P. C. Wright, D. H. Dockrell, *Mol. Cell. Proteomics* **2011**, *10*, DOI: 10.1074/mcp.M111.008193.

[28] a) F. Bracco, M. Banay-Schwartz, T. Deguzman, A. Lajtha, *Neurochem. Int.* **1982**, *4*, 541–549; b) M. Koike, H. Nakanishi, P. Saftig, J. Ezaki, K. Isahara, Y. Ohsawa, W. Schulz-Schaeffer, T. Watanabe, S. Waguri, S. Kametaka, M. Shibata, K. Yamamoto, E. Kominami, C. Peters, K. V. Figura, Y. Uchiyama, *J. Neurosci.* **2000**, *20*, 6898–6906.

[29] a) A. Naguleswarana, F. Alaeddinea, C. Guionaudb, N. Vonlaufena, S. Sondac, P. Jenoed, M. Mevissenb, A. Hemphilla, *Int. J. Parasitol.* **2005**, *35*, 1459–1472; b) F. Delom, B. Mallet, P. Carayon, P. J. Lejeune, *J. Biol. Chem.* **2001**, *276*, 21337–21342; c) H. J. Ryser, E. M. Levy, R. Mandel, G. J. DiSciullo, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 4559–4563.

[30] H. Wu, J. Ge, S. Q. Yao, *Angew. Chem.* **2010**, *122*, 6678–6682; *Angew. Chem. Int. Ed.* **2010**, *49*, 6528–6532.

[31] R. Srinivasan, M. Uttamchandani, S. Q. Yao, *Org. Lett.* **2006**, *8*, 713–716.

Received: June 8, 2011
Published online: September 2, 2011

