Proteomics Evaluation of Chemically Cleavable Activity-based Probes*

Marko Fonović‡§, Steven H. L. Verhelst‡, Mark T. Sorum‡, and Matthew Bogyo‡¶

Activity-based probes (ABPs) that specifically target subsets of related enzymatic proteins are finding increasing use in proteomics research. One of the main applications for these reagents is affinity isolation of probe-labeled targets. However, the use of cheap and efficient biotin affinity tags on ABPs can be problematic due to difficulty in release of captured proteins. Here we describe the evaluation of activity-based probes carrying a chemically cleavable linker that allows selective release of probe-labeled proteins under mild elution conditions that are compatible with mass spectrometric analysis. Specifically, we compare results from standard on-bead digestion of probe-labeled targets after affinity purification with the results obtained using chemoselective cleavage. Results are presented for multiple APBs that target both serine and cysteine proteases. These results highlight significant improvements in the quality of data obtained by using the cleavable linker system. ***Molecular & Cellular Proteomics 6:1761–1770, 2007.***

Over the last decade, activity-based proteomics has become a prominent method for profiling enzyme activities in complex protein samples (for reviews, see Refs. 1–5). This approach is dependent on small molecules termed activity-based probes (ABPs)[1] that use chemically reactive functional groups to covalently modify the active sites of a specific enzyme class. Activity-based probes have been specifically designed to target enzyme families with well established catalytic mechanisms including proteases, kinases, lipases, glycosylases, histone deacylases, and phosphatases (for a general review, see Ref. 2). Among proteases, activity-based probes have been mainly used for profiling the serine and cysteine families because both utilize catalytic nucleophiles in their active site that can be effectively targeted by a number of classes of synthetically accessible electrophiles (6–12).

One strength of ABPs is that they allow direct monitoring of the regulation of enzyme activity, thereby providing information that can be used to infer the function of targets. Perhaps even more importantly, ABPs provide an efficient way to enrich complex proteomes thereby making subsequent analytical analysis much simpler. Biotin is often used as a tag to facilitate isolation of labeled proteins by affinity chromatography on immobilized avidin. Biotin is commonly used because it is relatively cheap and provides diffusion-limited binding to avidin, thus allowing even low abundance proteins to be isolated with high efficiency. The disadvantage of this high affinity interaction with avidin is the harsh denaturing conditions required to disrupt the biotin-avidin interaction. Because these conditions are not directly compatible with mass spectrometry applications, samples must typically first be analyzed by SDS-PAGE, and subsequently proteins must be cut from a gel for identification. This additional step leads to significant loss of sample and can cause problems for proteins that do not resolve well via gel electrophoresis. Alternatively, proteins can be directly digested from the avidin resin; however, contamination by avidin and endogenously biotinylated proteins can be significant and in many cases will prevent detection of low abundance targets. Furthermore the on-bead digestion method leads to significantly more false positives as a result of nonspecific binding of proteins to the avidin resin. These false positive are particularly problematic because hit validation is often the most costly and time-consuming phase of proteomics studies.

To address some of the limitations of biotin, various cleavable linkers have been reported. These include acid-cleavable linkers (13), peptide linkers that can be cleaved by a proteolytic enzyme (14), and disulfide linkers (15). Although all of these regents have proved advantageous for specific applications, all have disadvantages that limit their general use. In particular, elution by acidic cleavage is often nonspecific, and TFA must be removed from the sample prior to MS analysis. Furthermore proteolytic elution requires introduction of an additional protease and tends to be less practical because elution efficiency is strongly dependent on protease activity. Probes that utilize a disulfide-cleavable linker are not compatible with buffers containing reducing reagents such as DTT and are prematurely cleaved under many cellular conditions.

We have recently developed a diazobenzene linker that can be cleaved to the corresponding anilines using a mild reducing agent (16). We present here the analysis of this cleavable linker system applied to ABPs that target both cysteine and

From the ‡Department of Pathology, Stanford University School of Medicine, Stanford, California 94305 and §Department of Biochemistry and Molecular and Structural Biology, Jozef Stefan Institute, Jamova 39, SI-1000 Ljubljana, Slovenia

Received, March 23, 2007, and in revised form, June 18, 2007

Published, MCP Papers in Press, July 5, 2007, DOI 10.1074/mcp.M700124-MCP200

[1] The abbreviations used are: ABP, activity-based probe; DIEA, *N*,*N*-diisopropylethylamine; PyBOP, benzotriazol-1-yloxytris(pyrrolidino)phosphonium hexafluorophosphate; Fmoc, *N*-(9-fluorenyl)methoxycarbonyl; DMF, *N*,*N*-dimethylformamide; UCH-L1, ubiquitin carboxyl-terminal hydrolase L1.

serine proteases. When compared with the on-bead digestion of labeled targets, specific chemical elution produces more reliable proteomics data that are free of background protein contamination.

EXPERIMENTAL PROCEDURES

Synthesis of SV1 and SV31—Synthesis of these probes was carried out using previously reported solid phase synthesis methods (8, 16). The cleavable linker was incorporated into the probe by coupling of 4-(5-(2-*N*-Fmoc ethyl)-2-hydroxyphenylazo)benzoic acid using standard solid phase peptide synthesis methods.

Synthesis of Cleavable Version of Fluorophosphonate DAP22C Probe (MTS-I-49)—(2′-Chloro)-chlorotrityl resin was treated with Fmoc-protected alanine under standard solid phase synthesis conditions (4 eq of DIEA in methylene chloride). The chain was extended under standard conditions using Fmoc-protected alanine (3 eq of 2-(benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate, 3 eq of 1-hydroxybenzotriazole, 6 eq of DIEA), 4-(5-(2-*N*-Fmoc ethyl)-2-hydroxyphenylazo)benzoic acid (1.5 eq of PyBOP, 3.0 eq of DIEA), and biotin (1.5 eq of PyBOP, 3.0 eq of DIEA) sequentially, gently shaking in DMF. Removal of Fmoc protecting groups was performed with 20% piperidine in DMF at room temperature. Cleavage from the resin with 2% TFA in methylene chloride (10 × 2 ml) provided the free acid, which was concentrated *in vacuo* and used directly without further purification. The above acid was treated with (1-amino-2-phenylethyl)phosphonic acid diphenyl ester (1 eq), PyBOP (1.5 eq), and DIEA (3 eq) in DMF. The solution was stirred at room temperature overnight to provide MTS-I-49, which was purified by HPLC.

Preparation of Mouse Tissue Homogenates and Protein Labeling—For tissue homogenate preparations, 60–90 mg of total protein from mouse liver, kidney, spleen, and brain tissue were used. Mouse tissue samples were homogenized using a bead beater in the presence of acetate buffer (50 mM sodium acetate, 2 mM EDTA, pH 5.5) for cysteine proteases or Tris buffer (50 mM Tris, 2 mM EDTA, pH 7.4) for serine hydrolases. The non-soluble fraction was removed by centrifugation (5 min at 13,000 × *g*), and total protein concentration was determined by the Bradford method. Total protein concentrations in homogenate preparations were in the range of 20–40 mg/ml, and 5 mg of total protein was used for a single labeling and enrichment experiment. Cysteine proteases were labeled by addition of the SV1 probe (10 μM final concentration) in the presence of DTT (2 mM final concentration). Labeling was carried out for 2 h at room temperature. For SDS-PAGE analysis the fluorescent probe SV31 (also at 10 μM) was used in place of SV1 because it allowed easier detection of labeled protein bands. Serine hydrolases were labeled by addition of MTS-I-49 (10 μM final concentration), and incubation was carried out for 2 h at room temperature. Labeled proteins were detected by avidin blot. Samples were separated by SDS-PAGE and transferred to a nitrocellulose membrane (Bio-Rad). The membrane was blocked in casein (5% solution in PBS) and incubated with streptavidin-horseradish peroxidase (dilution, 1:3500; Sigma) for 1 h at room temperature. After washing three times with 0.5% Tween in PBS, the membrane was treated with SuperSignal West Pico Chemiluminescent Substrate (Pierce) and exposed to HyBlot CL autoradiography film (Denville Scientific).

Affinity Enrichment and Elution of Labeled Proteins—After protein labeling, free probe was removed by a PD-10 gel filtration column (Amersham Biosciences), and the sample was eluted in PBS buffer (pH 7.4). Streptavidin beads (Pierce) were washed with PBS buffer and resuspended in the eluate. The sample was incubated at room temperature with shaking for 1 h. Streptavidin beads were separated from the unbound fraction by centrifugation. The supernatant was discarded, and the beads were sequentially washed with a series of PBS buffers containing 0.05% SDS, 1 M NaCl, and 10% EtOH. Finally the beads were washed with 100 mM ammonium hydrogen carbonate. Washing was performed three times with each buffer solution. After washing, beads were either used for "on-bead" digestion or chemical elution. Chemical elution was performed by addition of elution buffer (100 mM ammonium hydrogen carbonate, 25 mM sodium hydrosulfite) followed by incubation at room temperature for 15 min.

Sample Preparation for LC-MS/MS Analysis—For on-bead digestion, streptavidin beads with bound proteins were resuspended in 100 μl of denaturing buffer (50 mM sodium hydrogen carbonate, 6 M urea). Bound proteins were reduced in the presence of 10 mM DTT for 1 h. Samples were alkylated by addition of 200 mM iodoacetamide (20 μl) and incubated for 1 h in the dark. Unreacted iodoacetamide was neutralized by addition of 200 mM DTT (20 μl). The urea concentration was reduced by addition of distilled H_2O (800 μl). Samples were incubated with trypsin overnight at 37 °C and purified on a Vivapure C_{18} spin column according to the manufacturer's instructions (Sartorius).

Proteins chemically eluted from streptavidin beads were denatured by addition of an equal volume of 12 M urea. Reduction, alkylation, and trypsinization steps were performed in the same way as described for on-bead digestion. Prior to MS analysis, samples were desalted using a Vivapure C_{18} spin column (Sartorius).

Detection of Labeled Active Site Peptides of Purified Bovine Cathepsin B—Purified bovine cathepsin B was purchased from Calbiochem. Five micrograms of the protein was labeled by 10 μM SV1 probe in the presence of labeling buffer (50 mM sodium acetate, 2 mM EDTA, 2 mM DTT, pH 5.5). Samples were incubated for 2 h at room temperature. Streptavidin beads (20 μl of bead slurry washed with acetate buffer) were added to the sample, and incubation continued for 1 h at room temperature. Unbound proteins were washed away, and on-bead digestion was performed as described above. The digested supernatant was stored, and the beads were washed three times with 100 mM sodium hydrogen carbonate solution (1 ml). Labeled cathepsin active site peptides were eluted by incubation of the beads in cleavage buffer (100 mM ammonium hydrogen carbonate, 25 mM sodium hydrosulfite) for 15 min. The cleavage buffer was removed and replaced with fresh solution two times. All three eluted supernatants were pooled and prepared for LC-MS/MS analysis as described below.

LC-MS/MS and Database Search—Samples were analyzed on an LCQ DecaXP Plus ion trap mass spectrometer (Thermo Fisher) coupled to a nano-LC liquid chromatography unit (Eksigent). Peptides were separated on a BioBasic Picofrit C_{18} capillary column (New Objective). Elution was performed with an acetonitrile gradient from 0 to 50% in a 0.1% solution of formic acid over 40 min with overall flow rate of 350 nl/min. The three most intense base peaks in each scan were analyzed by MS/MS. Dynamic exclusion was set at a repeat count of 2 with exclusion duration of 2 min. Database searches were performed using the mouse National Center for Biotechnology Information (NCBI) protein database using the Sequest algorithm (Thermo Fisher). Peptides with XCorr values over 1.5 (+1 charge), 2 (+2 charge), and 2.5 (+3 charge) and ΔCn values over 0.1 were considered for further evaluation. Protein and peptide hits were statistically reevaluated by Scaffold (Proteome Software). Peptide identifications were accepted if they could be established at greater than 95% probability as specified by the Peptide Prophet algorithm (17). Protein identifications were accepted if they could be established at greater than 99.0% probability and contained at least two identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm (18).

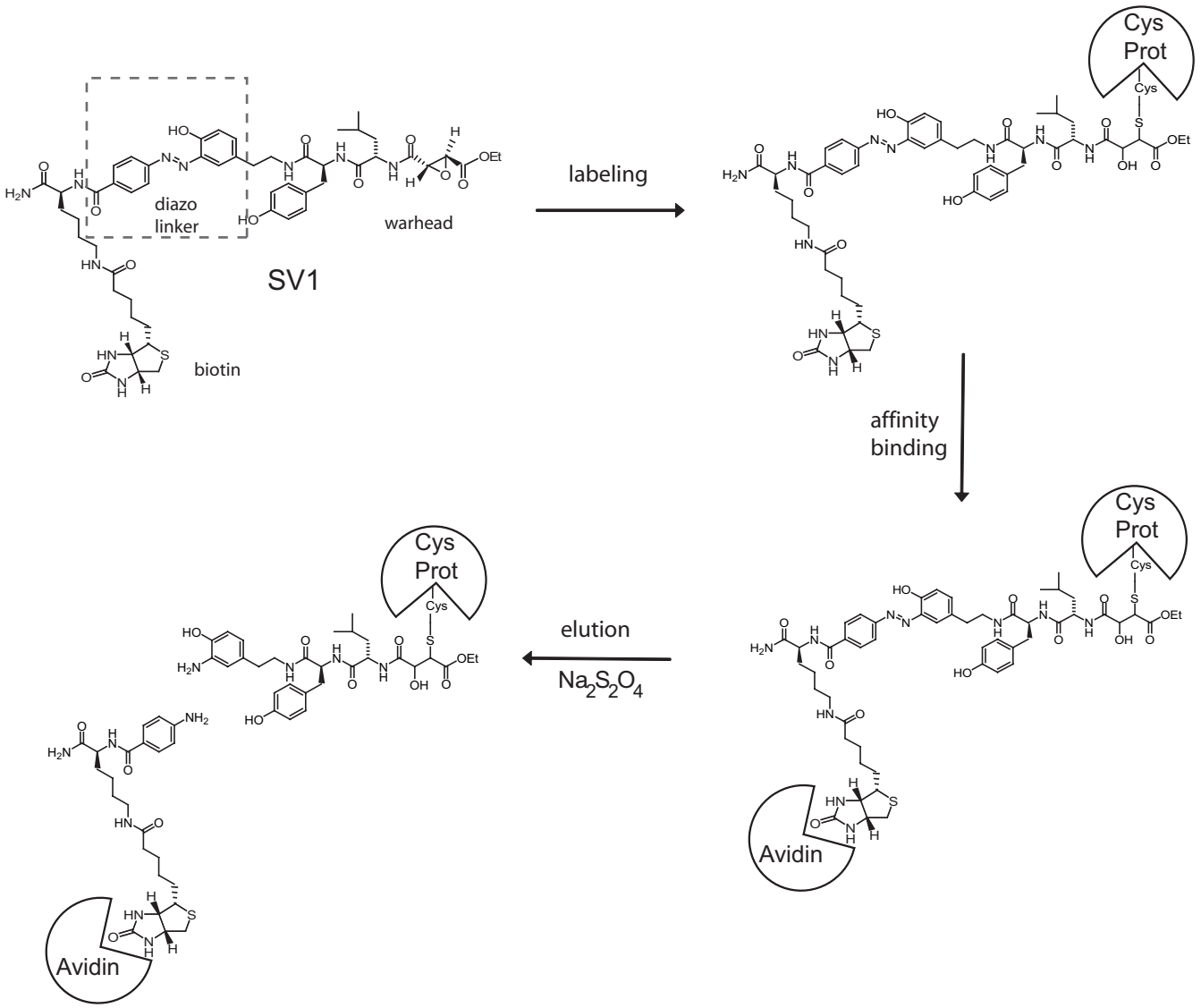
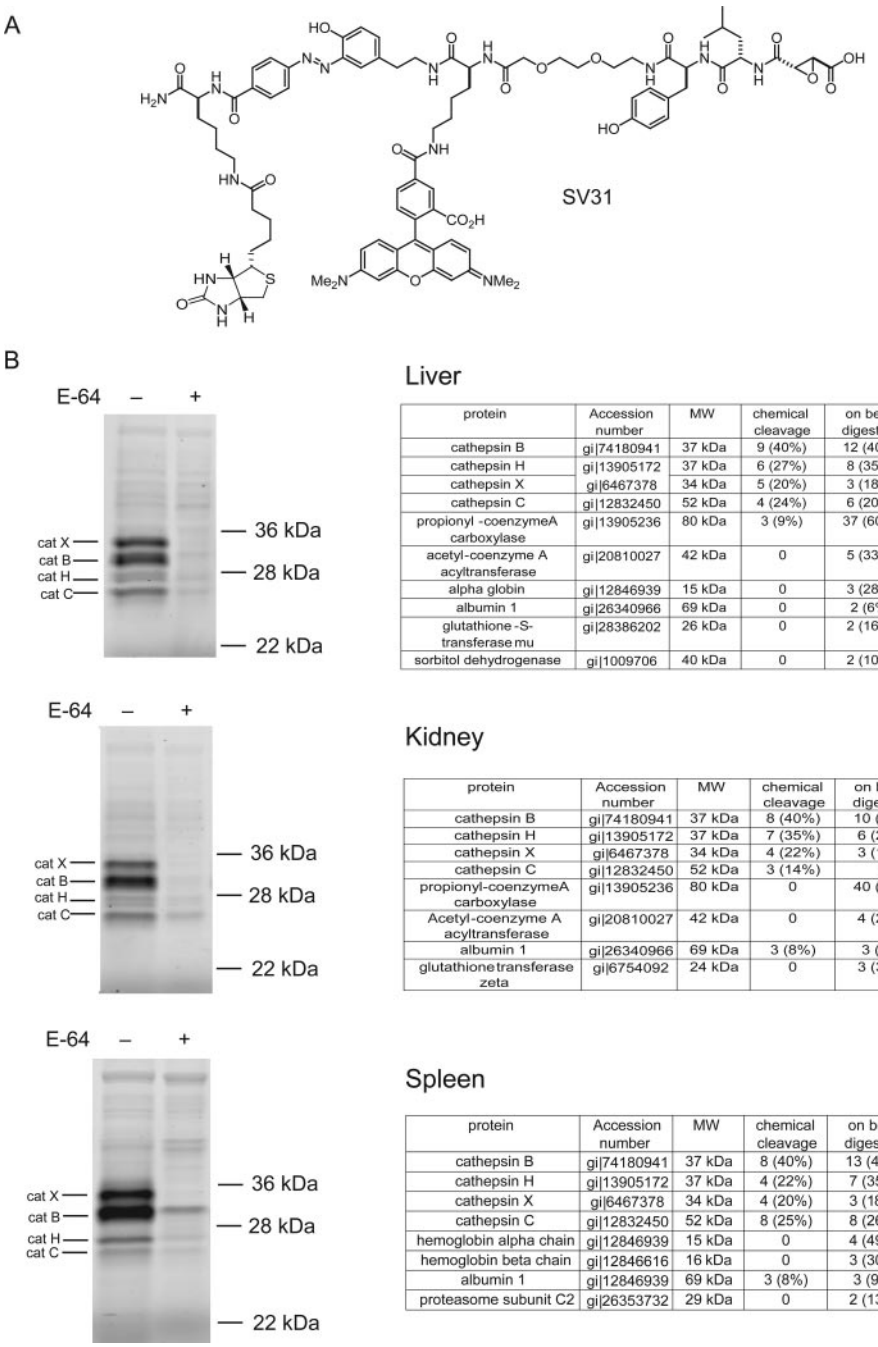


FIG. 1. **Experimental outline of sample preparation for a cleavable activity-based probe.** During the labeling step, the epoxide warhead covalently binds to the active site cysteine in the targeted protease. Labeled proteases are then bound to immobilized avidin, and unlabeled proteins are washed away. Specifically labeled proteins are released by reduction of the diazo linker in the presence of sodium hydrosulfite. Eluted proteins can then be prepared for LC-MS/MS analysis by "in-solution" digestion. *Cys Prot*, cysteine protease.

RESULTS AND DISCUSSION

We have recently reported the synthesis of the probe SV1 that specifically targets papain family cysteine proteases (16). This probe has an epoxide warhead and a biotin tag and contains the diazobenzene cleavable linker that can be specifically cleaved with greater than 90% elution efficiency using mild reducing conditions (16). Using this cleavable ABP, labeled target cysteine proteases are affinity-purified using an avidin resin and then eluted by cleavage of the diazobenzene linker using sodium hydrosulfite (Fig. 1). We therefore decided to use this general probe as a starting point to evaluate the utility of the cleavable linker system for activity-based protein profiling applications.

When using standard (non-cleavable) activity-based probes labeled with biotin, enriched proteins must be prepared for MS analysis in either of two ways. They can be eluted by denaturation by boiling in the presence of SDS, analyzed by SDS-PAGE, and digested "in gel," or they can be digested on bead. In-gel digestion is time-consuming, requires higher sample quantity, and is more prone to keratin contamination. On-bead digestion is one of the most commonly used non-gel methods for proteomics analysis of affinity-purified targets (19–22). In this approach, enriched proteins are denatured, reduced, alkylated, and digested while they are bound to affinity matrix such as immobilized avidin. However, natively biotinylated and abundant nonspecifically bound proteins can still be present during the digestion step even after stringent washing. These proteins produce false



Liver

protein	Accession number	MW	chemical cleavage	on bead digestion
cathepsin B	gi|74180941	37 kDa	9 (40%)	12 (40%)
cathepsin H	gi|13905172	37 kDa	6 (27%)	8 (35%)
cathepsin X	gi|6467378	34 kDa	5 (20%)	3 (18%)
cathepsin C	gi|12832450	52 kDa	4 (24%)	6 (20%)
propionyl -coenzymeA carboxylase	gi|13905236	80 kDa	3 (9%)	37 (60%)
acetyl-coenzyme A acyltransferase	gi|20810027	42 kDa	0	5 (33%)
alpha globin	gi|12846939	15 kDa	0	3 (28%)
albumin 1	gi|26340966	69 kDa	0	2 (6%)
glutathione -S-transferase mu	gi|28386202	26 kDa	0	2 (16%)
sorbitol dehydrogenase	gi|1009706	40 kDa	0	2 (10%)

Kidney

protein	Accession number	MW	chemical cleavage	on bead digestion
cathepsin B	gi|74180941	37 kDa	8 (40%)	10 (40%)
cathepsin H	gi|13905172	37 kDa	7 (35%)	6 (27%)
cathepsin X	gi|6467378	34 kDa	4 (22%)	3 (18%)
cathepsin C	gi|12832450	52 kDa	3 (14%)	0
propionyl-coenzymeA carboxylase	gi|13905236	80 kDa	0	40 (57%)
Acetyl-coenzyme A acyltransferase	gi|20810027	42 kDa	0	4 (27%)
albumin 1	gi|26340966	69 kDa	3 (8%)	3 (6%)
glutathione transferase zeta	gi|6754092	24 kDa	0	3 (30%)

Spleen

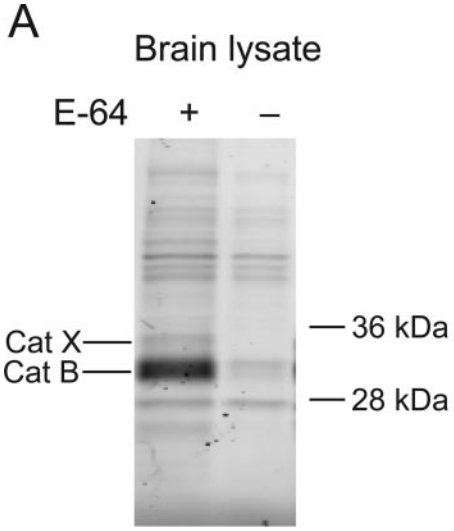
protein	Accession number	MW	chemical cleavage	on bead digestion
cathepsin B	gi|74180941	37 kDa	8 (40%)	13 (40%)
cathepsin H	gi|13905172	37 kDa	4 (22%)	7 (35%)
cathepsin X	gi|6467378	34 kDa	4 (20%)	3 (18%)
cathepsin C	gi|12832450	52 kDa	8 (25%)	8 (26%)
hemoglobin alpha chain	gi|12846939	15 kDa	0	4 (49%)
hemoglobin beta chain	gi|12846616	16 kDa	0	3 (30%)
albumin 1	gi|12846939	69 kDa	3 (8%)	3 (9%)
proteasome subunit C2	gi|26353732	29 kDa	0	2 (13%)

FIG. 2. **Comparison of target identification by on-bead digestion and chemical elution.** *A*, structure of the cleavable fluorescent probe SV31 that allows direct analysis of probe-labeled proteins by SDS-PAGE followed by scanning with a laser scanner. *B*, mouse liver, kidney, and spleen lysates were labeled with SV31 for SDS-PAGE analysis (*left panels*) and by the SV1 probe for direct mass spectrometric analysis (*right panels*). Fluorescent protein bands that are competed by pretreatment with the cysteine protease inhibitor E-64 were identified as cathepsins. After enrichment on immobilized avidin, each of the three labeled proteomes was divided into two aliquots. One was prepared for LC-MS/MS analysis by on-bead digestion, whereas the other was chemically eluted and digested in solution. Proteins identified in each tissue lysate are listed with their accession number and theoretical molecular mass. To compare both approaches, the number of identified peptides and percentage of amino acid sequence coverage are listed for each protein. Please note that molecular masses of cathepsins listed in the tables are for unprocessed zymogens and are therefore significantly larger than the masses of processed, active cathepsins observed on the SDS-PAGE gels. *cat*, cathepsin.

positives and can interfere with identification of low abundance targets.

One of the most significant advantages of the cleavable linker strategy is that it avoids background resulting from nonspecifically retained or endogenously biotinylated proteins. However, for the cleavable linker to be of real value it must allow highly efficient elution of targets. To test the sensitivity of the cleavable probe we directly compared samples of labeled proteins isolated by on-bead digestion and by chemical cleavage (Fig. 2). We labeled mouse liver, kidney, and spleen with the SV1 probe and divided each sample into two aliquots. One was digested on bead, whereas the other proteins were chemically eluted from the beads by linker cleavage and digested in solution as described under "Experimental Procedures." In addition, a small amount of each proteome was labeled with the related fluorescent probe SV31 (Fig. 2*A*). This probe allows direct analysis of labeled proteomes by SDS-PAGE followed by scanning of the gel with a laser scanner. As expected the labeling profiles were similar in all three proteomes, and the pattern of labeled proteins that were blocked by pretreatment with the general inhibitor E-64 corresponded to the patterns observed previously for labeling

A



B

protein	accession number	MW	chemical cleavage	on bead digestion
cathepsin B	gi|74180941	37 kDa	7 (23%)	11(36%)
cathepsin X	gi|11066226	34 kDa	2 (7%)	0
ubiquitin carboxy-terminal hydrolase L1	gi|25058057	25 kDa	0	5 (36%)
glyceraldehyde-3-phosphate dehydrogenase	gi|55153885	36 kDa	0	7(36%)
propionyl -coenzyme A carboxylase	gi|3905236	80 kDa	0	7(17%)
lactate dehydrogenase	gi|52880	37 kDa	0	4(20%)
tubulin beta	gi|21746161	50 kDa	0	4(20%)
tubulin alpha	gi|13435777	50 kDa	2(6%)	4(13%)
triosephosphate isomerase	gi|1864018	23 kDa	0	4(36%)
actin	gi|49868	42 kDa	0	4(15%)
alpha globin	gi|12846939	15 kDa	0	3(42%)
hemoglobin	gi|12846616	15 kDa	2(22%)	4(36%)

C

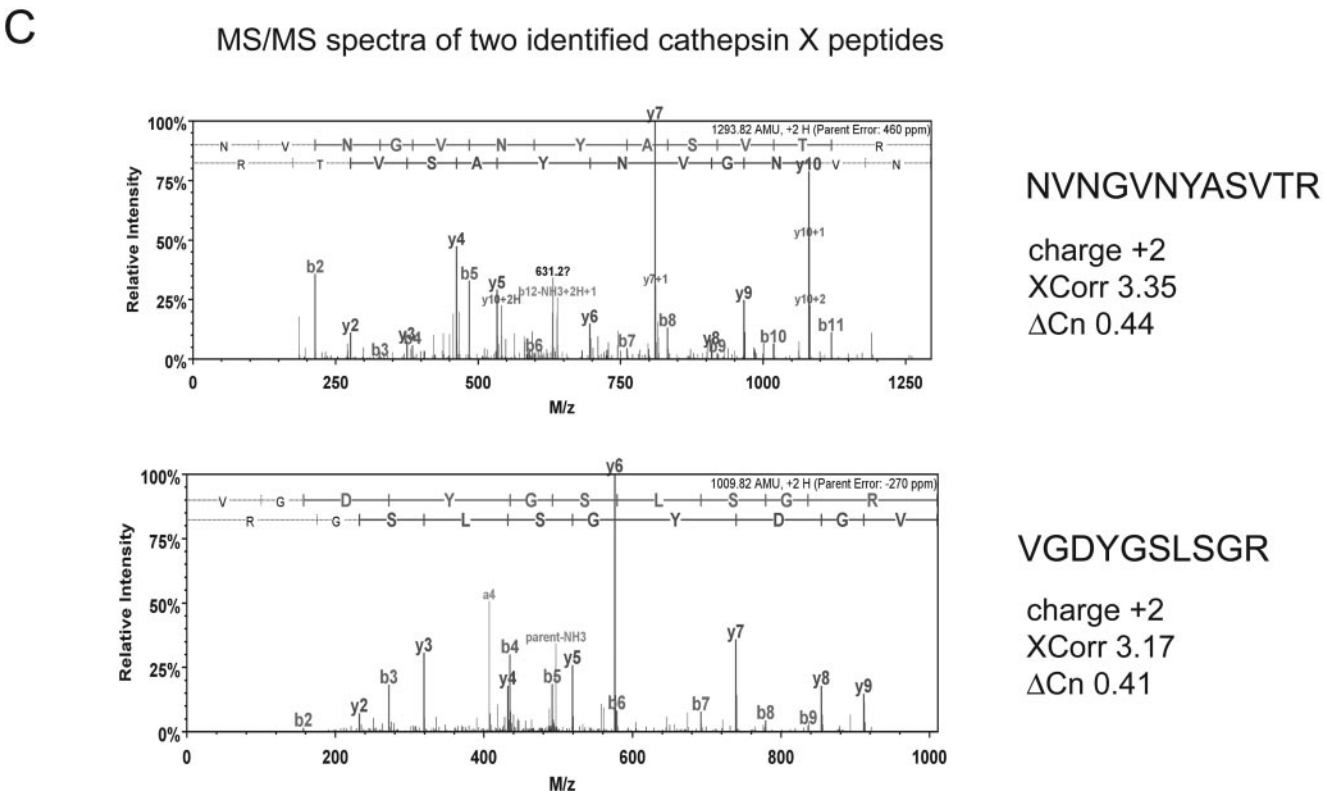


FIG. 3. **Determination of SV1 probe targets in mouse brain lysate.** Mouse brain lysate was labeled with SV31 for SDS-PAGE analysis (*A*) and with SV1 for direct mass spectrometric analysis (*B*). Targets were enriched by affinity binding to immobilized avidin and eluted by chemical cleavage or digested on beads. All identified proteins determined with chemical cleavage and on-bead digestion are listed with their number of identified peptides and amino acid sequence coverage (*B*). *C*, MS/MS spectra for both cathepsin X peptides determined in the sample prepared by chemical elution are shown together with their charge and Sequest XCorr and ΔCn values. *Cat*, cathepsin.

of cathepsins B, X, H, and C by the related probe DCG-04 (23). Consistent with these reported patterns of labeling, cathepsin X (also called cathepsin Z) was resolved above cathepsins B and H on the gel even though its molecular weight based on sequence alone suggests it should be smaller than both cathepsins B and H. This may be due to differences in secondary structure or post-translational modifications, or effects based on detergent denaturation. In all three proteomes on-bead digestion yielded samples contaminated with background proteins. The main contaminant was propi-

onyl-coenzyme A carboxylase, which is a natively biotinylated protein. Other identified background contaminants were abundant tissue proteins such as hemoglobin, albumin, and transferases (Fig. 2*B*). The highest contamination was observed in liver and kidney proteomes, where over 60% of all identified peptides corresponded to background proteins. Contamination was not as high in spleen lysate due to the absence of carboxylases in this tissue. Samples obtained from chemical cleavage on the other hand showed a minimal number of background proteins while retaining good coverage of previously identified cysteine proteases (23). Importantly the overall sequence coverage of the identified cathepsin targets was virtually identical for both methods suggesting that elution is nearly quantitative as reported previously for this probe (16). As a further indication of the selectivity of the linker system, on-bead digestion of the kidney proteome failed to identify one of the primary cathepsin target proteins (cathepsin C), whereas chemical cleavage provided confident identification of cathepsin C with 18% sequence coverage. Taken together these results suggest that chemical cleavage of the diazobenzene linker is highly efficient and results in significant reduction in background signals, thus reducing overall false negatives produced by abundant background proteins.

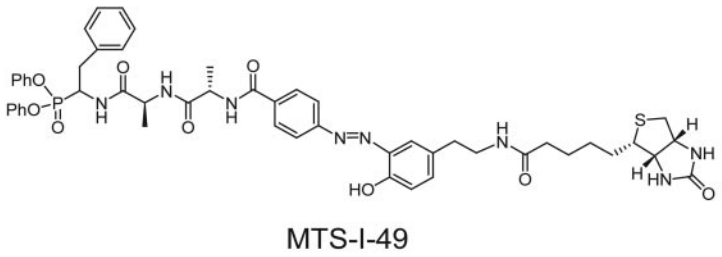
We also applied the cleavable probe SV1 to brain lysate to determine whether the reduced background and high sensitivity observed in liver, spleen, and kidney samples could be observed in other tissues that have typically produced high background in our activity-based protein profiling experiments (Fig. 3). SDS-PAGE analysis of brain lysate incubated with the fluorescent probe SV31 showed labeling of two cysteine protease activities that could be inhibited by E-64 (Fig. 3*A*). Based on their molecular weight and inhibitor sensitivity, these two proteases were predicted to be cathepsins X and B. Direct on-bead digestion of enriched brain lysate identified a total 11 proteins, but only two of them were positively identified as cysteine proteases (cathepsin B and ubiquitin carboxyl-terminal hydrolase L1; Fig. 3*B*). In contrast, MS analysis of the chemically cleaved sample identified four proteins, and two of these were identified as cysteine proteases (cathepsins B and X; Fig 3*B*). As cathepsin B, cathepsin X, and ubiquitin carboxyl-terminal hydrolase L1 (UCH-L1) are all members of the papain family of cysteine proteases, they all represent potential targets of the SV1 probe. However, UCH-L1 is a highly abundant protein that makes up over 1% of total brain protein content (24) making it a potential nonspecific background protein. Because SDS-PAGE did not show any fluorescently labeled band at the predicted size of UCH-L1 and the SV1 probe also failed to label recombinant UCH-L1,[2] we concluded that UCH-L1 is in fact a false positive background protein that was only identified in the on-bead sample. Furthermore the weakly labeled 32-kDa protein that can be blocked by E-64 pretreatment (Fig. 3*A*) was identified as cathepsin X in the chemical cleavage sample but was missed in the on-bead digestion sample. Thus, the chemical cleavage method both prevented identification of a false positive protein that seemed to be a legitimate target and helped to identify a false negative target that was missed by on-bead digestion due to high background contamination. In addition, chemical cleavage produced clean MS/MS spectra that allowed confident identification of probe-labeled cathepsin X (Fig. 3*C*), suggesting that the enhanced selectivity does not come at the cost of reduced spectral quality in the raw MS data.

To show applicability of the cleavable linker approach for profiling of other catalytic classes of proteases in the context of other probe scaffolds, we incorporated the diazobenzene linker into the peptide backbone of the commercially available probe DAP22C to make the probe MTS-I-49 (Enzyme Systems Products; Fig. 4*A*). DAP22C is a peptidyl aminoalkanephosphonate that has been reported to target the serine proteases cathepsin G and chymotrypsin *in vitro* (25). However, it has not been used for any proteomics applications. We used MTS-I-49 to label mouse liver extracts and analyze the labeled sample by both SDS-PAGE (Fig. 4*B*) and by direct chemical cleavage or on-bead digestion (Fig. 4*C*). Although the on-bead digest again identified a large number of background and endogenously biotinylated proteins, the chemical cleavage revealed predominant recovery of the S9 family clan SC serine hydrolase, carboxylesterase 4. Overall the level of background peptides was decreased by over 90% when chemical elution was used for the sample preparation (Fig. 4*C*). Although the number of carboxylesterase peptides identified was reduced compared with the on-bead sample we could still make a confident identification with nearly 40% coverage of peptide sequence (Fig. 4*D*). Carboxylesterase 4 has not been characterized on a functional level and has never been confirmed to be catalytically active (26). It is not clear why we were only able to identify this one esterase, but the lack of other reasonable targets in the on-bead digestion sample suggests that the probe may simply be inefficient or highly selective for targets that are not active in mouse liver extracts.

Finally, we were interested to see whether the active site peptides that are modified with the probe could be isolated and identified by MS/MS analysis. This is particularly important for applications of the cleavable linker system to ICAT where probe-labeled peptides are exclusively recovered whereas non-labeled peptides are washed away. Because the active site peptides of all of the cysteine cathepsins are difficult to identify due to their large size as trypsin-digested fragments, we tested the SV1 probe on purified bovine spleen cathepsin B. After on-bead digestion of enriched labeled cathepsins, active site peptides were eluted by chemical cleavage, and active site peptides were identified from the MS/MS fragmentation pattern. Surprisingly, not only cathepsin B but also cathepsin S was identified in this sample. The molecular mass of the precursor ion and its MS/MS fragmentation pattern showed the expected mass addition of 571 Da to the

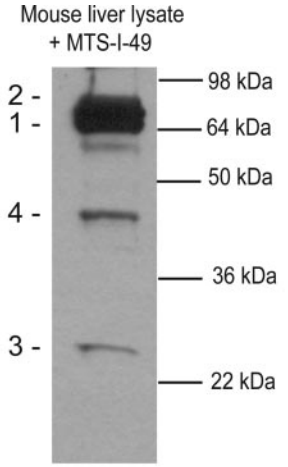
[2] M. Fonović, unpublished results.

A



MTS-I-49

B



C

	protein	Accession number	MW	chemical cleavage	on bead digestion
1 -	liver carboxylesterase4	gi|15488664	62 kDa	14(38%)	20(51%)
2 -	propionyl-Coenzyme A carboxylase	gi|13905236	80 kDa	3(8%)	22(45%)
3 -	glutathione-S-transferase	gi|2832492	25 kDa	2(12%)	8(44%)
4 -	acetyl-Coenzyme acyltransferase 2	gi|20810027	42 kDa	2(15%)	5(26%)
	carbonmonoxy hemoglobin	gi|18655689	16 kDa	0	7(68%)
	betaine-homocysteine methyltransferase	gi|22477957	45 kDa	0	4(19%)
	hemoglobin alpha 1	gi|12846939	15 kDa	0	3(25%)
	acetyl-Coenzyme A dehydrogenase	gi|26348293	19 kDa	0	3(34%)
	albumin 1	gi|26340966	69 kDa	0	2(6%)
	alcohol dehydrogenase 1	gi|32449839	40 kDa	0	3(20%)
	glyceraldehyde-3-phosphate dehydrogenase	gi|55153885	36 kDa	0	2(10%)
	3-hydroxy-3-methylglutaryl-coenzyme A-synthase	gi|12836439	57 kDa	0	2(6%)
	carbamoyl-phosphate synthase 1	gi|82879179	165 kDa	0	2(2%)
	fructose-bisphosphate aldolase	gi|15723268	40 kDa	0	2(6%)
	methionine adenosyltransferase 1	gi|74150290	44 kDa	0	2(9%)
	peroxisomal acyl-CoA oxidase	gi|2253380	75 kDa	0	2(4%)
	glycine N-methyltransferase	gi|15679953	33 kDa	0	2(10%)

D

gi 15488664 (100%), 61612.7 Da Hypothetical protein MGC18894 [Mus musculus],
gi 21450339 ref NP_659179 hypotetical protein LOC234564 [Mus musculus]
14 unique peptides, 16 unique spectra, 16 total spectra, 213/561 amino acids (38% coverage)

MFLSTLFLVS LATCVICGNP SSPPVVDTAH GKVLGKHVNV EGFSQPVAVF
LGIPFAKPPL GSLRFAPPQP AEPWSSVKNA TTYPPMCSQD AARGQAVNDL
ITNRKEKIHL EFSEDCLYLN IYTPADFSKN SRLPVMVWIH GGGLKLGGAS
SFDGRALSAY ENVVVVAIQY RLSIWGFFST GDEHSRGNWG HLDQVAALHW
VQDNIANFGG DPGSVTIFGE SAGGYSVSIL ILSPLSKNLF HRAISESGVA
FIPGMFTKDV RPITEQIAVT AGCKTTTSAV IVHCMRQKTE EELLEIMHKL
NLYKLSLQGD TKNSDQFVTS VLDGVVLPKD PKEILAEKNF NTVPYIVGIN
KQECGWLLPT MTGFLPADVK LDKKKAIALL EQFASMTGIP EDIIPVAVEK
YTKGSDDPDQ IREGVLDAMG DVAFGVPSVI VSRGHRDTGA PTYMYEYQYY
PSFSSPQRPK NVVGDHADDV YSVFGAPILR EGASEEEINL SKMVMKFWAN
FARNGNPNGK GLPHWPKYDQ KEGYLHIGGT TQQAQRLKEE EVTFWTQSLA
KKQPQPYHNE L

FIG. 4. **Use of the cleavable linker probe for identification of serine hydrolases in mouse liver.** *A*, structure of the phenoxy probe MTS-I-49 that contains the cleavable linker and is similar in structure to the commercially available serine protease probe DAP22C. *B*, mouse liver extracts were labeled with MTS-I-49 and analyzed by direct SDS-PAGE followed by biotin blotting. *C*, list of proteins identified by LC/MS/MS analysis of on bead-digested or chemically eluted proteins from mouse liver extracts labeled with MTS-I-49. A hypothetical protein was identified as the primary probe target after chemical cleavage. BLAST (Basic Local Alignment Search Tool) database searching identified it as liver carboxylesterase 4 (*D*). A number of additional endogenously biotinylated proteins were identified in the on bead-digested sample. The four most intensely labeled proteins from the gel image in *B* (*labeled 1–4*) likely represent the top four hits in the MS analysis in *C* based on predicted molecular weights. The proposed assignments are indicated using numbers corresponding to the number of the hits in the *table* in *C*. *PhO*, phosphonate.

active site cysteine corresponding to the cleaved probe fragment (Fig. 5). These data confirm that the probe labeled active site cysteine residues and that it could be used to specifically isolate and sequence the active site peptides labeled in the target proteases. In all of our labeling and enrichment experiments, we never identified any peptides derived from the pro regions of the zymogen forms of any proteases (Fig. 6). These data fit with the finding that the probe only showed modification of proteins in the size rage of the mature cathepsins when the samples were analyzed by SDS-PAGE and confirm that our probe labels only catalytically active proteases.

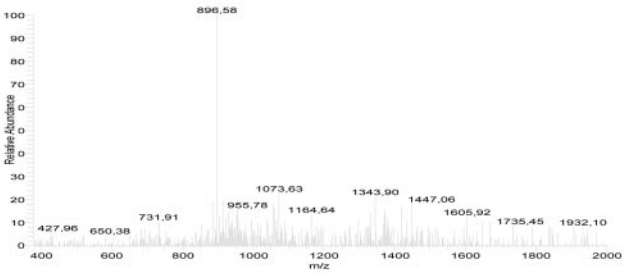
In conclusion, our results using multiple activity-based probes that contain the chemically cleavable diazobenzene linker suggest that these probes will be valuable tools for direct proteomics profiling of active serine and cysteine proteases. The major advantage of this cleavable linker system is its ability to reduce background signals resulting in reduction in the number of both false positives and false negatives. We

A

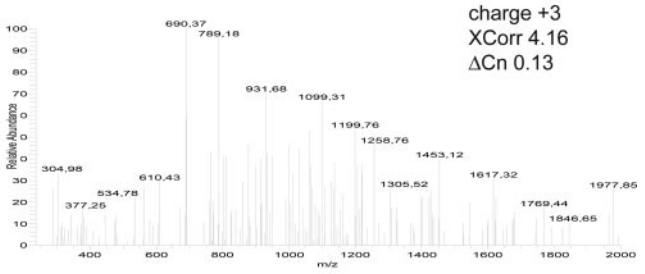
Cathepsin B

amino acid	B ions	Y ions
D	116.1	-
Q	244.23	2573.13
G	301.28 +	2445.00
S	388.35 +	2387.94
C+57	548.52	2300.87
G	605.57 +	2140.7
S	692.65 +	2083.65
C+571	1366.79 +	1996.57 +
W	1553.00	1322.43 +
A	1624.08 +	1136.22 +
F	1771.26 +	1065.14 +
G	1828.31	917.96 +
A	1899.39 +	860.91 +
V	1998.52	789.83 +
E	2127.63	690.70 +
A	2198.71	561.59 +
I	2311.87	490.51 +
S	2398.95	377.35 +
D	2514.04	290.27 +
R	-	175.19

Base peak



MS/MS spectrum

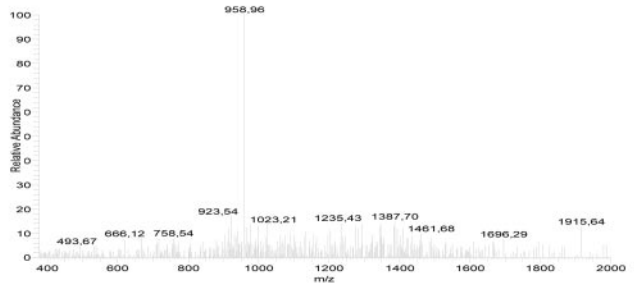


B

Cathepsin S

amino acid	B ions	Y ions
Y	164.18	-
Q	292.31	2713.36
G	349.36	2585.23
A	420.44 +	2528.18
C+57	580.61 +	2457.1
G	637.66	2296.94
S	724.74	2239.89
C+571	1399.88 +	2152.81
W	1586.09	1477.67 +
A	1657.17	1291.46 +
F	1804.34	1220.38 +
S	1891.42	1073.20 +
A	1962.50 +	986.12 +
V	2061.63	915.05 +
G	2118.68	815.92 +
A	2189.76	758.86 +
L	2302.92	687.79 +
E	2432.03	574.63 +
A	2503.11	445.51 +
Q	2631.24	374.43 +
V	2730.37	246.30
K	-	147.17

Base peak



MS/MS spectrum

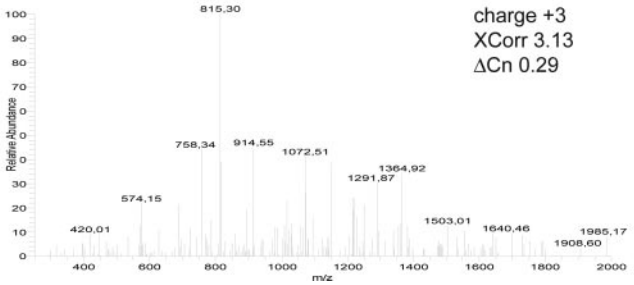


FIG. 5. **Detection of labeled active site peptides of cathepsins B and S.** Purified bovine cathepsin B (Calbiochem) was labeled with SV1 (10 μM) and bound to immobilized streptavidin. The sample was digested on bead followed by washing to remove non-active site labeled peptides. The labeled active site peptides remained bound to the beads and were specifically eluted by reduction cleavage with sodium hydrosulfite. MS/MS analysis revealed the presence of active site peptides of both cathepsins B (*A*) and S (*B*) suggesting that the commercial sample was contaminated with cathepsin S. Due to the large molecular mass of the active site peptides, only triply charged species were detected. The tables at *left* show assignments of b and y ion series (detected ions are marked) in the MS/MS scans of precursor ions; the base peaks and MS/MS spectra are shown at *right*.

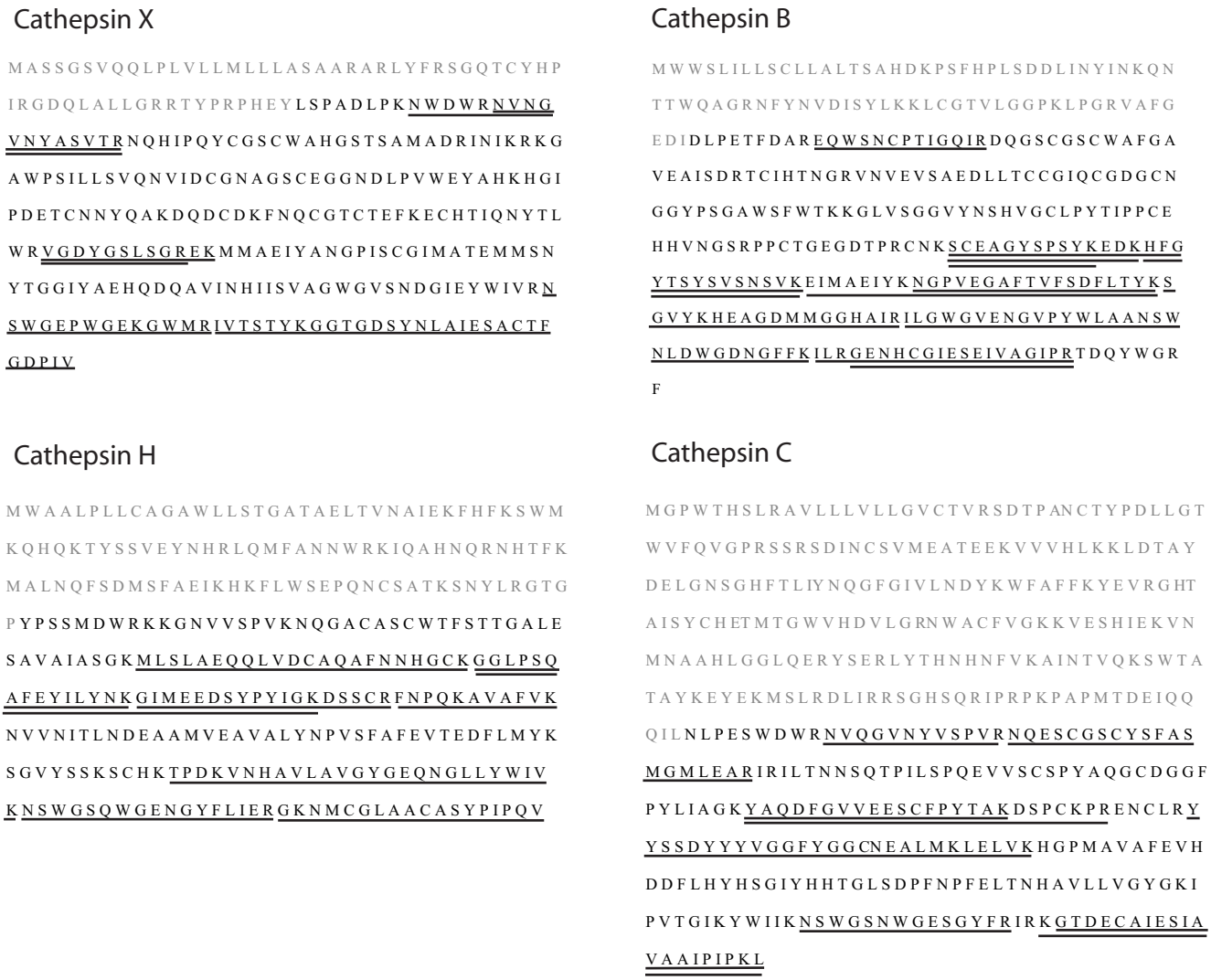


FIG. 6. **Amino acid coverage of identified cathepsins.** Positions of peptides identified in all labeling experiments (on bead-digested and chemically cleaved) are shown mapped onto the amino acid sequences of each target protease (*lines below* sequences). Cathepsin pro regions are shown in *gray*. None of the peptides mapped to the pro regions of any of the targets indicating that zymogen forms of the proteases were not labeled by the probes.

believe this linker system will have broad value for all types of small molecule proteomics probes that rely on affinity purification methods as it will help to reduce the time-consuming and expensive process of validating proteomics hits. In the examples presented here we had the benefit of significant knowledge of the types of targets we expected to identify with our probes. In cases where general probes are used and potential targets are unknown, these chemically cleavable linkers will be particularly valuable.

Acknowledgment—We thank Andrew Guzzetta (Stanford University) for extensive technical help and advice.

* This work was supported by National Technology Center for Networks and Pathways Grant U54 RR020843 (to M. B.), National Institutes of Health Grant R01-EB005011 (to M. B.), and a postdoctoral fellowship from the Netherlands Organization for Scientific Research (to S. V.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¶ To whom correspondence should be addressed: Dept. of Pathology, Stanford University School of Medicine, 300 Pasteur Dr., Stanford, CA 94305. Tel.: 650-725-4132; Fax: 650-725-7424; E-mail: mbogyo@stanford.edu.

REFERENCES

1. Berger, A. B., Vitorino, P. M., and Bogyo, M. (2004) Activity-based protein profiling: applications to biomarker discovery, in vivo imaging and drug discovery. *Am. J. Pharmacogenomics* **4,** 371–381
2. Evans, M. J., and Cravatt, B. F. (2006) Mechanism-based profiling of enzyme families. *Chem. Rev.* **106,** 3279–3301
3. Fonović, M., and Bogyo, M. (2007) Activity based probes for proteases: applications to biomarker discovery, molecular imaging and drug

screening. *Curr. Pharm. Des.* **13,** 253–261
4. Jeffery, D. A., and Bogyo, M. (2003) Chemical proteomics and its application to drug discovery. *Curr. Opin. Biotechnol.* **14,** 87–95
5. Sadaghiani, A. M., Verhelst, S. H., and Bogyo, M. (2007) Tagging and detection strategies for activity-based proteomics. *Curr. Opin. Chem. Biol.* **11,** 20–28
6. Bogyo, M., Verhelst, S., Bellingard-Dubouchaud, V., Toba, S., and Greenbaum, D. (2000) Selective targeting of lysosomal cysteine proteases with radiolabeled electrophilic substrate analogs. *Chem. Biol.* **7,** 27–38
7. Borodovsky, A., Ovaa, H., Kolli, N., Gan-Erdene, T., Wilkinson, K. D., Ploegh, H. L., and Kessler, B. M. (2002) Chemistry-based functional proteomics reveals novel members of the deubiquitinating enzyme family. *Chem. Biol.* **9,** 1149–1159
8. Greenbaum, D., Medzihradszky, K. F., Burlingame, A., and Bogyo, M. (2000) Epoxide electrophiles as activity-dependent cysteine protease profiling and discovery tools. *Chem. Biol.* **7,** 569–581
9. Kato, D., Boatright, K. M., Berger, A. B., Nazif, T., Blum, G., Ryan, C., Chehade, K. A., Salvesen, G. S., and Bogyo, M. (2005) Activity-based probes that target diverse cysteine protease families. *Nat. Chem. Biol.* **1,** 33–38
10. Liu, Y., Patricelli, M. P., and Cravatt, B. F. (1999) Activity-based protein profiling: the serine hydrolases. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 14694–14699
11. Mahrus, S., and Craik, C. S. (2005) Selective chemical functional probes of granzymes A and B reveal granzyme B is a major effector of natural killer cell-mediated lysis of target cells. *Chem. Biol.* **12,** 567–577
12. Pan, Z., Jeffery, D. A., Chehade, K., Beltman, J., Clark, J. M., Grothaus, P., Bogyo, M., and Baruch, A. (2006) Development of activity-based probes for trypsin-family serine proteases. *Bioorg. Med. Chem. Lett.* **16,** 2882–2885
13. van der Veken, P., Dirksen, E. H., Ruijter, E., Elgersma, R. C., Heck, A. J., Rijkers, D. T., Slijper, M., and Liskamp, R. M. (2005) Development of a novel chemical probe for the selective enrichment of phosphorylated serine- and threonine-containing peptides. *Chembiochem* **6,** 2271–2280
14. Speers, A. E., and Cravatt, B. F. (2005) A tandem orthogonal proteolysis strategy for high-content chemical proteomics. *J. Am. Chem. Soc.* **127,** 10018–10019
15. Shimkus, M., Levy, J., and Herman, T. (1985) A chemically cleavable biotinylated nucleotide: usefulness in the recovery of protein-DNA complexes from avidin affinity columns. *Proc. Natl. Acad. Sci. U. S. A.* **82,** 2593–2597
16. Verhelst, S. H., Fonović, M., and Bogyo, M. (2007) A mild chemically cleavable linker system for functional proteomic applications. *Angew. Chem. Int. Ed. Engl.* **46,** 1284–1286
17. Keller, A., Nesvizhskii, A. I., Kolker, E., and Aebersold, R. (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* **74,** 5383–5392
18. Nesvizhskii, A. I., Keller, A., Kolker, E., and Aebersold, R. (2003) A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **75,** 4646–4658
19. Fu, Z., Larson, K. A., Chitta, R. K., Parker, S. A., Turk, B. E., Lawrence, M. W., Kaldis, P., Galaktionov, K., Cohn, S. M., Shabanowitz, J., Hunt, D. F., and Sturgill, T. W. (2006) Identification of yin-yang regulators and a phosphorylation consensus for male germ cell-associated kinase (MAK)-related kinase. *Mol. Cell. Biol.* **26,** 8639–8654
20. Jessani, N., Niessen, S., Wei, B. Q., Nicolau, M., Humphrey, M., Ji, Y., Han, W., Noh, D. Y., Yates, J. R., III, Jeffrey, S. S., and Cravatt, B. F. (2005) A streamlined platform for high-content functional proteomics of primary human specimens. *Nat. Methods* **2,** 691–697
21. Labugger, R., Simpson, J. A., Quick, M., Brown, H. A., Collier, C. E., Neverova, I., and Van Eyk, J. E. (2003) Strategy for analysis of cardiac troponins in biological samples with a combination of affinity chromatography and mass spectrometry. *Clin. Chem.* **49,** 873–879
22. Tagwerker, C., Flick, K., Cui, M., Guerrero, C., Dou, Y., Auer, B., Baldi, P., Huang, L., and Kaiser, P. (2006) A tandem affinity tag for two-step purification under fully denaturing conditions: application in ubiquitin profiling and protein complex identification combined with in vivo cross-linking. *Mol. Cell. Proteomics* **5,** 737–748
23. Greenbaum, D., Baruch, A., Hayrapetian, L., Darula, Z., Burlingame, A., Medzihradszky, K. F., and Bogyo, M. (2002) Chemical approaches for functionally probing the proteome. *Mol. Cell. Proteomics* **1,** 60–68
24. Wilkinson, K. D., Lee, K. M., Deshpande, S., Duerksen-Hughes, P., Boss, J. M., and Pohl, J. (1989) The neuron-specific protein PGP 9.5 is a ubiquitin carboxyl-terminal hydrolase. *Science* **246,** 670–673
25. Oleksyszyn, J., and Powers, J. C. (1991) Irreversible inhibition of serine proteases by peptide derivatives of (α-aminoalkyl)phosphonate diphenyl esters. *Biochemistry* **30,** 485–493
26. Rawlings, N. D., Morton, F. R., and Barrett, A. J. (2006) MEROPS: the peptidase database. *Nucleic Acids Res.* **34,** D270–D272