

# Chemical Approaches for Functionally Probing the Proteome\*<sup>§</sup>

Doron Greenbaum<sup>‡</sup>, Amos Baruch<sup>§</sup>, Linda Hayrapetian<sup>§</sup>, Zsuzsanna Darula<sup>‡¶</sup>, Alma Burlingame<sup>‡</sup>, Katlin F. Medzihradsky<sup>¶¶</sup>, and Matthew Bogoy<sup>§||</sup>

**With the availability of complete genome sequences, emphasis has shifted toward the understanding of protein function. We have developed a functional proteomic methodology that makes use of chemically reactive fluorescent probes to profile and identify enzymes in complex mixtures by virtue of their catalytic activity. This methodology allows a comparison of changes in activity of multiple enzymes under a variety of conditions using a single two-dimensional separation. The probes can also be used to localize active enzymes in intact cells using fluorescence microscopy. Furthermore, the probes enable screens for selective small molecule inhibitors of each enzyme family member within crude lysates or intact cells. Ultimately, this technology allows the rapid identification of potential drug targets and small molecule lead compounds targeted to them. *Molecular & Cellular Proteomics* 1:60–68, 2002.**

Over the past few years the complete genome sequences of multiple organisms have been determined. These efforts have been followed by the annotation of genes that code for all proteins of a given proteome. Although this information is likely to prove valuable, a great deal of effort is still required to define the function of individual gene products. Informatics techniques have been developed to assign function to individual genes by analyzing patterns of co-inheritance throughout multiple organisms (1, 2). Furthermore, analysis of genome-wide changes in transcription in response to different stimuli allows clustering of genes of similar function based on transcriptional co-regulation (3). Although these methods help to broadly classify proteins into families, the assignment of functions to specific members within a large enzyme family remains a difficult task.

Proteomic approaches address some of the gaps in genomic methodologies by profiling and identifying bulk changes in protein levels (4, 5). However, these methodol-

ogies only provide information for abundant proteins, and proteins with difficult biochemical properties (*i.e.* membrane proteins) are often excluded from analysis. Moreover, for most enzymes, their activity, and therefore their function, is regulated by a complex set of post-translational controls. Therefore, even proteomic profiles in many cases provide an incomplete picture of how enzymes are functionally regulated (6).

Classical genetic approaches are tried and true methods to assign functions to specific gene products. In many biological systems it is possible to disrupt a desired gene and assess the resulting phenotype. However, this process is often tedious, and in cases where multiple related proteins have similar functions, compensation adjustments make the results difficult to interpret.

To circumvent these problems, small molecules can be used to manipulate the activity of protein targets (7, 8). This “chemical genetic” approach makes use of libraries of small molecules to screen for compounds that perturb a given biological process. The resulting “hits” can then be used to begin to assign function to specific enzyme or protein targets. However, the utility of this process is limited by the difficult task of identifying the relevant target of the small molecule.

In the case of traditional drug discovery, small molecule libraries are screened against a single pre-defined target. Lead compounds are often identified from large chemical libraries using an *in vitro* assay. Although many of these compounds are effective against the purified target, little is usually known about their selectivity in a crude proteome. Therefore, a method that allows screening for small molecule inhibitors in cell and tissue extracts or intact cells would allow identification of lead compounds based on multiple criteria such as potency, selectivity, and cell permeability. Furthermore, compounds could be screened against entire enzyme families thereby increasing the chances of identifying useful compounds for therapeutic intervention.

We have developed chemically reactive affinity probes that can be used to (i) identify the members of a given enzyme family within a proteome, (ii) determine the relative activity levels of individual family members, (iii) localize active enzymes within a cell, and (iv) screen small molecule libraries directly in crude protein extracts for inhibitors that can ultimately be used to determine biological functions of specific target enzymes. In this study, we have chosen to focus on the

From the Departments of <sup>§</sup>Biochemistry and Biophysics and <sup>‡</sup>Pharmaceutical Chemistry, University of California, San Francisco, California 94143 and the <sup>¶</sup>Mass Spectrometry Facility, Biological Research Center, Hungarian Academy of Sciences, H-6701 Szeged, Hungary

Received, September 5, 2001

Published, MCP Papers in Press, September 11, 2001, DOI 10.1074/mcp.T100003-MCP200

<sup>§</sup>The on-line version of this article (available at <http://www.mcponline.org>) contains supplemental material.

papain family of cysteine proteases for several reasons. First, these proteases are synthesized as inactive zymogens that are activated post-translationally (9, 10). Their activity can also be regulated by interaction with macromolecular inhibitors resulting in transcription/translational profiles that provide only limited information regarding their functional regulation. Second, the papain family is composed of many closely related family members whose functions are poorly defined (11). Third, many small molecule covalent inhibitors of this class of enzyme have been developed that can be used for probe design (see Ref. 12). Finally, these enzymes have been found to play an important role in many disease conditions such as cancer (13), osteoporosis (14), asthma (11), and rheumatoid arthritis (15) making them a potential important class of enzymes for drug development.

#### MATERIALS AND METHODS

All chemicals used in the synthesis of the peptide epoxides were purchased from Advanced Chemtech (Louisville, KY) and Sigma-Aldrich Chemical Co. (St. Louis, MO). Purified human liver cathepsin B and H and purified turkey liver cathepsin C were purchased from Calbiochem (San Diego, CA). Purified, recombinant human cathepsins K, F, and V were a kind gift of Dieter Brömme (Mt. Sinai School of Medicine, New York, NY). Purified recombinant human cathepsin L was a kind gift of Vito Turk (Jozef Stefan Institute, Ljubljana, Slovenia).

#### Synthesis Protocols

*Synthesis of Ethyl (2S,3S)-oxirane-2,3-dicarboxylate and Ethyl (2R,3R)-oxirane-2,3-dicarboxylate and DCG-04*—The synthesis of (2R,3R)-oxirane-2,3-dicarboxylate is identical to that reported for the (2S,3S) isomer (18). The synthesis of DCG-04 is reported elsewhere (19).

*Synthesis of BODIPY558/568-DCG-04, BODIPY588/616-DCG-04, BODIPY530/550-DCG-04, and BODIPY493/503-DCG-04*—All fluorophores were purchased from Molecular Probes (Eugene, OR). A free amino version of DCG-04 was synthesized by replacing the terminal biotinylated lysine with lysine using the reported synthesis protocols for DCG-04 (19). Free amino DCG-04 (6 mg, 8.8  $\mu\text{mol}$ , 1.5 eq) and BODIPY558/56-OSu<sup>1</sup> (3.0 mg, 6.0  $\mu\text{mol}$ , 1.0 eq), BODIPY 588/616-OSu (1.0 eq), BODIPY530/550-OSu (1.0 eq), or BODIPY493/503-OSu (1 eq) were dissolved in 100  $\mu\text{l}$  of Me<sub>2</sub>SO. Diisopropylethylamine was then added (12.0  $\mu\text{mol}$ , 2.0 eq). The reaction was monitored by high pressure liquid chromatography (HPLC). After 2 h the product was purified on a C<sub>18</sub> reverse phase HPLC column (Delta Pak; Waters) using a linear gradient of 0–100% water-acetonitrile. Fractions were pooled and lyophilized to dryness. The identity of the product was confirmed by mass spectrometry. The electrospray mass spectrum was as follows: [M + H] calculated for BODIPY558/568-DCG-04 C<sub>49</sub>H<sub>69</sub>BF<sub>2</sub>N<sub>9</sub>O<sub>10</sub> 979.5, found 978.5; BODIPY 588/616-DCG-04 C<sub>60</sub>H<sub>76</sub>BF<sub>2</sub>N<sub>9</sub>O<sub>12</sub>S 1196.5, found 1197.0; BODIPY530/550-DCG-04 C<sub>57</sub>H<sub>69</sub>BF<sub>2</sub>N<sub>8</sub>O<sub>10</sub> 1075.5, found 1075.0; and BODIPY493/503-DCG-04 C<sub>49</sub>H<sub>63</sub>BF<sub>2</sub>N<sub>8</sub>O<sub>10</sub>S 1005.4, found 1004.5.

*Synthesis of Positional Scanning Libraries*—Synthesis of the P2 constant PSL library was performed using a 96-well manifold (Flex-

Chem; Robbins Scientific). Each library was constructed using a constant amino acid at the P2 position and an isokinetic mixture of all natural amino acids (minus cysteine and methionine plus norleucine) at the variable position. The isokinetic mixture was created using a ratio of equivalents of amino acids based on their reported coupling rates (24). The total mixture was adjusted to 10-fold excess total amino acids over resin load. For constant positions, a single amino acid was coupled using 10-fold excess. In addition to the natural amino acids, a set of 42 non-natural hydrophobic amino acids were used for the constant P2 position (see Table I in Supplemental Material). Couplings were carried out using diisopropylcarbodiimide and hydroxybenzotriazole under standard conditions for solid phase peptide synthesis. Libraries and single components were cleaved from the resin by addition of 90% trifluoroacetic acid, 5% water, 5% triisopropyl saline for 2 h. Cleavage solutions were collected, and products were precipitated by addition of cold diethyl ether. Solid products were isolated, and the crude peptides were dissolved in Me<sub>2</sub>SO (50 mM stock) based on average weights for each mixture. Libraries and single compounds were stored at –20 °C and further diluted to 10 mM stock plates for use in experiments.

*Synthesis of YQ-(R, R)Eps and YG-(R, R)Eps*—All single component peptide epoxides were synthesized on the solid support using the protocols reported for DCG-04 (19). The inhibitors were cleaved from the resin by addition of 90% trifluoroacetic acid, 5% water, 5% triisopropyl saline for 2 h. Ice-cold ether (15 ml) was used to precipitate the products. The crude products were purified on a C<sub>18</sub> reverse phase HPLC column (Waters) using a linear gradient of 0–100% water-acetonitrile. Fractions containing the product were pooled, frozen, and lyophilized to dryness. The identity of the product was confirmed by mass spectrometry. The electrospray mass spectrum was as follows: [M + H] calculated for YG-(R, R)Eps C<sub>17</sub>H<sub>21</sub>N<sub>3</sub>O<sub>7</sub> 380.1, found 380.1; YQ-(R, R)Eps C<sub>20</sub>H<sub>26</sub>N<sub>4</sub>O<sub>8</sub> 451.2, found 451.2.

#### Radiolabeling of Inhibitors

All compounds were iodinated and isolated using the protocol reported previously (18).

#### Preparation of Cell and Tissue Lysates

Tissues were Dounce-homogenized in Buffer A (50 mM Tris, pH 5.5, 1 mM DTT, 5 mM MgCl<sub>2</sub>, 250 mM sucrose), and extracts were centrifuged at 1,100 × *g* for 10 min at 4 °C, and the supernatant was centrifuged at 22,000 × *g* for 30 min at 4 °C. Cells were homogenized using glass beads in Buffer A, and supernatants were centrifuged for 15,000 × *g* for 15 min at 4 °C. The total protein concentration of the final supernatants (soluble) was determined by BCA protein quantification (Pierce).

*Labeling of Lysates and Purified Cathepsins with DCG-04, <sup>125</sup>I-DCG-04, <sup>125</sup>I-MB-074, <sup>125</sup>I-YQ-(R, R)Eps, Yellow-DCG-04, Blue-DCG-04, Green-DCG-04, or Red-DCG-04*

Lysates (100  $\mu\text{g}$  of total protein in 100  $\mu\text{l}$  of Buffer B (50 mM Tris, pH 5.5, 5 mM MgCl<sub>2</sub>, 2 mM DTT)) or purified cathepsins (0.1  $\mu\text{g}$  in Buffer B) were labeled for 1 h at 25 °C unless noted otherwise. DCG-04 was added to a final concentration of 10  $\mu\text{M}$ . Equivalent amounts of all radioactive inhibitor stock solutions (approximately 10<sup>6</sup> cpm per sample) were used for all labeling experiments. Fluorescent compounds were added to samples to a final concentration of 0.1  $\mu\text{M}$ . Samples were quenched by addition of 4× SDS sample buffer (for one-dimensional SDS-PAGE) or by addition of solid urea to a final concentration of 9.5 M (for 2D SDS-PAGE). Fluorescent samples were analyzed using an ABI 377 DNA sequencer. Standard 15% SDS-PAGE gels of 0.4-mm thickness were prepared using 15-cm plates provided by the

<sup>1</sup> The abbreviations used are: OSu, O-succinimide ester; ABP, activity-based probe; CID, collision-induced dissociation; E-64, trans-Epoxy succinyl-L-leucylamido(4-guanidino)butane; HPLC, high pressure liquid chromatography; IEF, isoelectric focusing; MS, mass spectrometry; PSL, positional scanning library; DTT, dithiothreitol; 2D, two-dimensional.

manufacturer. Samples were loaded and electrophoresed for 3–4 h at a constant current of 35 mA with voltage limited to 750 V. Gel images were created using the Gene Scan software provided by the manufacturer. In some experiments, fluorescent samples were analyzed by standard SDS-PAGE followed by scanning with a Molecular Dynamics Typhoon laser scanner.

### *In Situ Fluorescence Labeling*

Dendritic cells (DC2.4) were plated on a 24-well dish ( $10^5$  cells/well) embedded with sterile microscope coverslips, in RPMI medium containing 10% fetal bovine serum. After 16 h, cells were washed with 1 ml of TC-199 medium and incubated with 1  $\mu$ M Green-DCG-04 in TC-199 for 12 h at 37 °C. Cells were washed three times with 1 ml of TC-199 and incubated for 5 h in probe-free medium. Subsequently, cells were either lysed in Buffer A and analyzed on a 12.5% SDS-PAGE using a fluorescent scanner or viewed under a fluorescent microscope.

### *Gel Electrophoresis*

One-dimensional SDS-PAGE and two-dimensional IEF were performed as described (25).

### *Competition Labeling and Analysis of Data*

Rat liver lysates (100  $\mu$ g of total protein in 100  $\mu$ l of Buffer B (50 mM Tris, pH 5.5, 5 mM  $MgCl_2$ , 2 mM DTT)) or purified cathepsins (1  $\mu$ g of protein in 100  $\mu$ l of Buffer A) were pre-incubated with 10  $\mu$ M of each library member (diluted from 10 mM  $Me_2SO$  stocks) for 30 min at room temperature. Samples were then labeled by addition of  $^{125}I$ -DCG-04 to each sample followed by further incubation at room temperature for 1 h. Samples were quenched by the addition of 4 $\times$  sample buffer, resolved by SDS-PAGE, and analyzed by phosphorimaging (Molecular Dynamics). Bands corresponding to each labeled protease were quantitated. Inhibitor-treated samples were compared with an untreated control sample. Numerical values for percent competition were analyzed as described previously (26) using the programs Tree View and Cluster written by Eisen and co-workers (3). These programs can be obtained from [www.microarrays.org](http://www.microarrays.org).

### *Purification and Identification of Affinity-labeled Proteases from Rat Liver*

Protein lysates prepared in Buffer C (50 mM Acetate buffer, 5 mM DTT, 0.1% Triton X-100) were incubated with 5  $\mu$ M DCG-04 for 1.5 h at room temperature. After incubation the protein lysate was passed through a PD10 column pre-equilibrated with Buffer D (50 mM Tris-Base 7.4, 150 mM NaCl), and proteins were eluted with the same buffer. SDS was added to eluted proteins to a final concentration of 0.5%, and the solution was boiled for 10 min, diluted 2.5-fold with Buffer D (to reduce SDS concentration to 0.2%), and incubated with a 100- $\mu$ l bed volume of pre-washed streptavidin beads for 1 h at room temperature. Beads were washed five times with Buffer D, and bound proteins were eluted by boiling for 10 min in the presence of 100  $\mu$ l of SDS sample buffer. For 2D analysis, samples in SDS sample buffer were diluted 1:1 with IEF sample buffer (9.5 M, 5%  $\beta$ -mercaptoethanol, 2% Nonidet P-40, 1.6% ampholines, pH 5–7, and 0.4% ampholines, pH 3.5–10), and pure Nonidet P-40 was added (25% of volume of sample). Samples were applied to IEF tube gels and electrophoresed at 1000 V for 13 h followed by separation in the second dimension on 15% SDS-PAGE gels. The resulting gels were fixed in 12% acetic acid, 50% methanol stained with silver according to reported protocols (25). Spots were excised, digested with trypsin, and fractionated by reverse phase HPLC on an Ultimate system, equipped with a FAMOS auto-injector (LC Packings, San Francisco, CA). Experimental conditions were as follows: 1- $\mu$ l injection; 75- $\mu$ m  $\times$

150-mm PepMap column; solvent A ( $H_2O$  with 0.1% formic acid); solvent B (acetonitrile with 0.1% formic acid; gradient, 0–30% solvent B in 40 min at a flow rate of  $\sim$ 250 nl/min. Mass spectrometry detection was performed on a QSTAR quadrupole orthogonal acceleration-time-of-flight tandem mass spectrometer (Applied Biosystems/MDS Sciex, Foster City, CA) in information-dependent acquisition mode; 2-s survey acquisitions were followed by 5-s CID acquisitions, in which the most abundant ion of each survey scan was selected as the precursor. All the singly charged ions, as well as some trypsin autolysis products, were excluded from the precursor ion selection. The collision energy was optimized and adjusted automatically depending on the charge state and the  $m/z$  value of the precursor ions selected. The mass range recorded in survey acquisitions was  $m/z$  300–1400. For CID experiments the lower mass limit was changed to  $m/z$  60. All the data were measured using a two-point external calibration. The instrument affords  $\sim$ 8000 resolution and 30 ppm mass accuracy with external calibration in both MS and CID mode. Proteins were identified automatically by Mascot data base search using the MS/MS data (Matrix Science Ltd., London, UK).

## RESULTS AND DISCUSSION

*Probe Design and Application to Crude Homogenates and Intact Cells*—Several laboratories have developed small molecule electrophiles that show class-specific reactivity toward nucleophilic active site residues of several different enzyme families. These include serine (16, 17) and cysteine (18–20) hydrolases, as well as aldehyde dehydrogenases (21). In each case, electrophiles have been designed that exhibit broad irreversible reactivity for enzyme family members, while remaining relatively inert toward free-circulating nucleophiles such as thiols, hydroxyls, and amines. The resulting activity-based probes (ABPs) can be used to covalently label specific target enzymes within the complex mixture of proteins from a cell or tissue sample. Our laboratory has developed probes based on the structure of the general cysteine protease inhibitor trans-epoxysuccinyl-L-leucylamido(4-guanidino)butane (E-64) (19). These ABPs can be used to affinity-label papain family cysteine proteases. They also allow rapid purification of labeled proteases by incorporation of a biotin affinity tag. Here we have used the core peptide epoxide analog of E-64 to create four fluorescently labeled ABPs for papain family cysteine proteases (Fig. 1).

These probes incorporate four different fluorescent moieties, each with non-overlapping excitation and emission spectra, allowing for multiplexing of probes. Four BODIPY analogs were chosen based on the excitation and emission wavelengths of fluorophores commonly used in DNA sequencing protocols. We reasoned that it should be possible to visualize and quantify fluorescently labeled proteins using a standard DNA sequencing apparatus equipped with a high intensity laser. Fig. 2A shows the gel image that results from incubation of eight different purified or recombinant papain family cysteine proteases with each of the four fluorescent ABPs followed by analysis on an ABI 377 DNA sequencer (see “Materials and Methods”). Using these probes, it is possible to load all eight proteases in a single gel lane and distinguish each, based on differences in molecular weight and emission

wavelength of fluorescent labels.

The same four probes were next used to profile the repertoire of papain family proteases within a complex protein

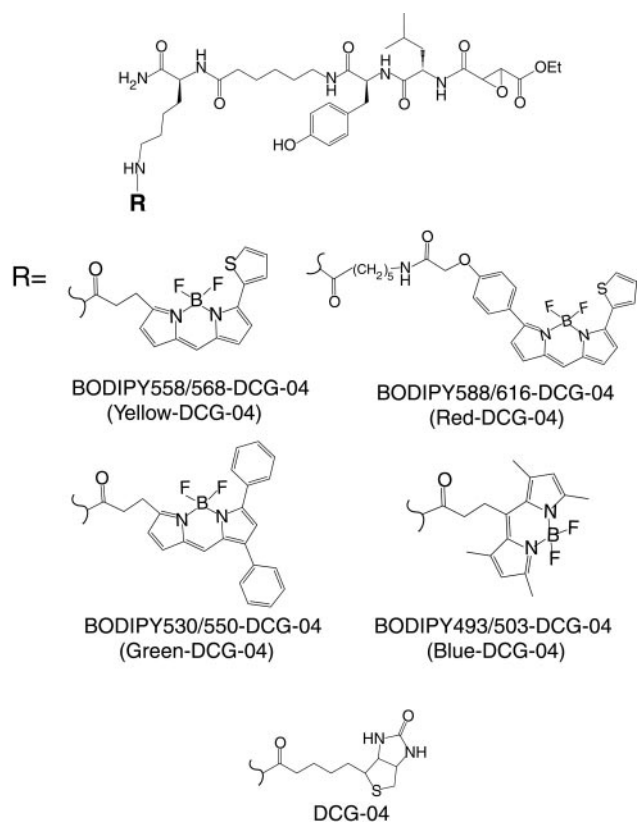


FIG. 1. Structures of the fluorescent ABPs DCG-04, Yellow-DCG-04, Red-DCG-04, Green-DCG-04, and Blue-DCG-04.

mixture derived from a tissue homogenate. Fig. 2B shows the profiles of cysteine proteases in total rat liver homogenates obtained by labeling with the biotinylated probe DCG-04, the radiolabeled version of DCG-04, and the four fluorescent analogs of DCG-04. All ABPs labeled the same four predominant protease (bands 1–4) species with only slight differences in relative intensities observed for each probe. These results suggest that the presence of structurally diverse labeling groups at the distal affinity site of the molecules had little effect on the ability of a compound to covalently modify its targets.

Because covalent modification of target proteases by the ABPs requires modification of the active site thiol nucleophile, labeling intensities can be used as an indirect measure of enzymatic activity. Thus, unlike antibodies that can only be used to monitor bulk levels of specific proteins, these reagents allow analysis of changes in levels of enzymatic activity. In the past, our laboratory has used these reagents to follow activity of cysteine proteases during processes such as tumor progression/cell invasion and cataract formation (18, 22). These newly developed ABPs therefore provide an efficient method for monitoring changes in protease activities within a proteome.

Because the fluorescent probes are cell-permeable they make ideal tools for imaging of protease activity in intact cells or tissue sections. Fig. 3 shows the dendritic cell line DC2.4 either directly labeled *in situ* with Green-DCG-04 or pre-treated with E-64 and then labeled with the fluorescent probe. Cells directly treated with the green ABP showed a fluorescence staining pattern characteristic of lysosomal compartments. Cells that had been pre-treated with E-64 showed

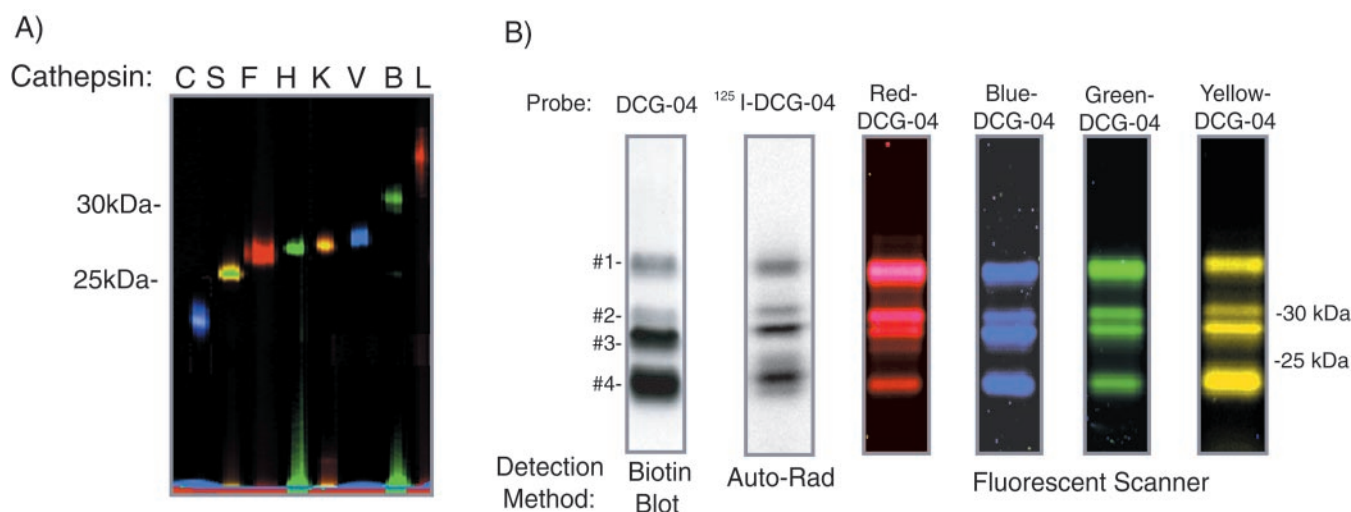


FIG. 2. Affinity labeling of papain family proteases using fluorescent ABPs. A, purified cathepsins (as indicated) were diluted into pH 5.5 buffer and labeled with 100 nM Yellow-DCG-04, Red-DCG-04, Green-DCG-04, or Blue-DCG-04 for 1 h. Samples were separated on a 15% SDS-PAGE gel, and labeled bands were visualized using an ABI 377 DNA sequencer as described under “Materials and Methods.” B, total cell extracts from rat liver were diluted into pH 5.5 buffer and labeled with 10  $\mu$ M DCG-04, <sup>125</sup>I-DCG-04 (approximately  $1 \times 10^6$  cpm), or 100 nM Red-, Blue-, Green-, and Yellow-DCG-04. Samples were separated on a 15% SDS-PAGE gel, and labeled bands were visualized (as indicated at bottom) by affinity blotting or autoradiography or using a Molecular Dynamics Typhoon laser fluorescence scanner.

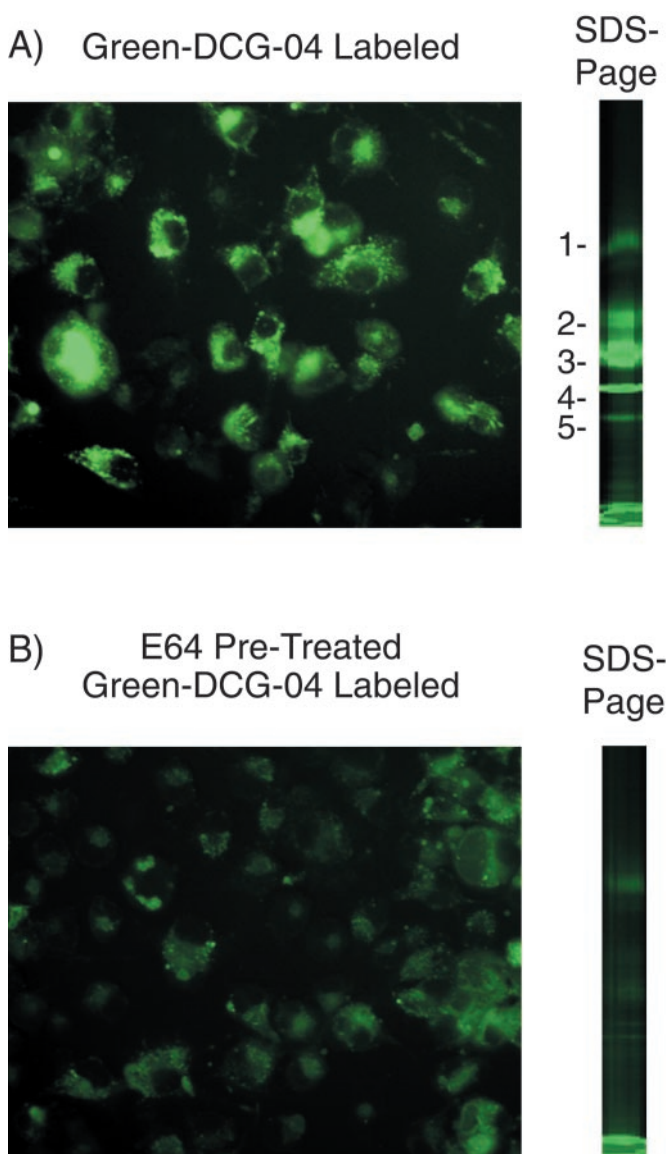


FIG. 3. **Localization of protease activity *in situ*.** DC2.4 cells were grown in culture in serum-free media and treated overnight with Green-DCG-04 (1  $\mu\text{M}$  final concentration) (A) or (B) pre-treated with 10  $\mu\text{M}$  E-64 for 1 h and then labeled with 1  $\mu\text{M}$  Green-DCG-04. Fresh medium was added, and cells were incubated for 5 h to remove excess probe. Cells were visualized by fluorescence microscopy (left panels) and then collected, lysed in SDS sample buffer, and analyzed by SDS-PAGE on an ABI 377 DNA sequencer (right panels). Labeled proteases in the untreated cells are indicated with numbers. Note the complete competition of all protease species by E-64 pre-treatment.

diffuse fluorescence throughout the cytosol, likely because of residual free probe that failed to be washed away. The cells were collected after imaging, lysed, and analyzed by SDS-PAGE and fluorescence detection. The resulting profiles indicated that multiple protease species were labeled by the fluorescent probe and that these proteases were completely inhibited by pre-treatment of cells with E-64. Thus the fluorescent staining observed in the non-pretreated cells repre-

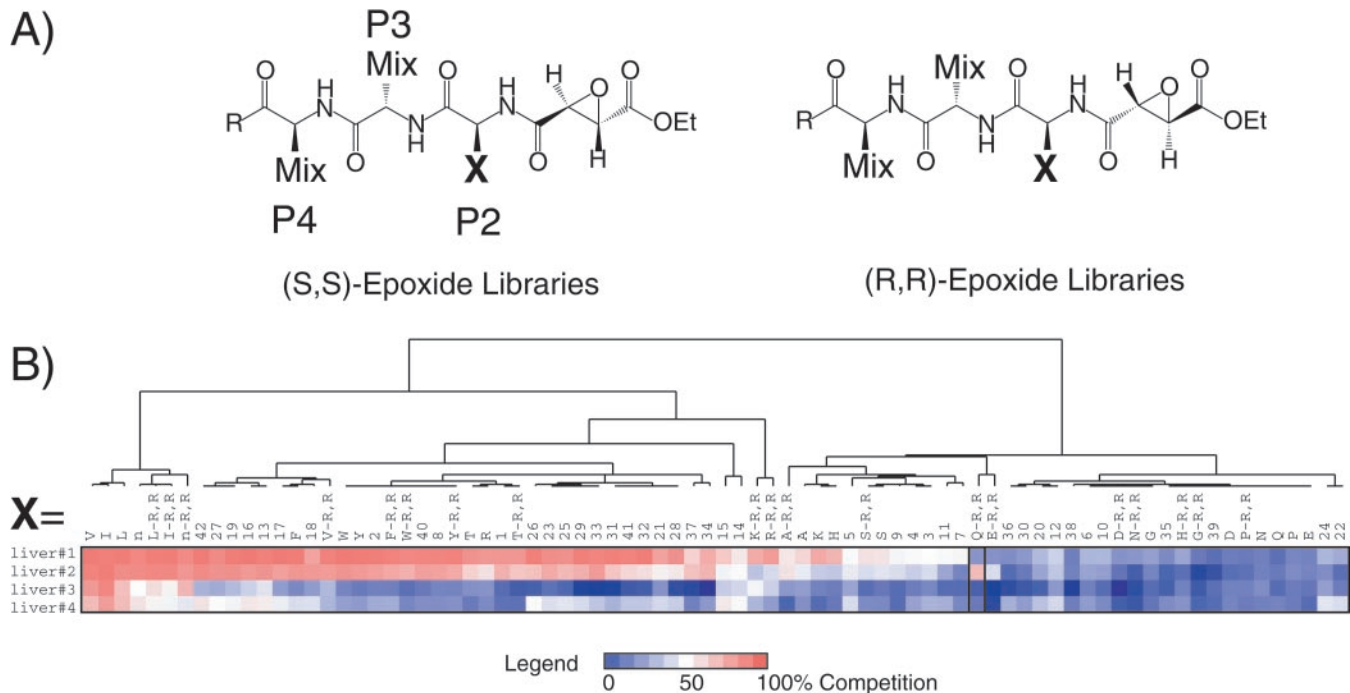
sents the localization of active papain family cysteine proteases. This method is likely to be applicable to tissue samples and may serve as a convenient way to image protease activities in tissues derived from important clinical samples such as solid tumors.

*Using ABPs to Screen for Selective Inhibitors of Papain Family Cysteine Proteases in Crude Tissue Extracts*—Perhaps the most powerful attribute of ABPs is their ability to facilitate screening of small molecule inhibitors against complete enzyme families without the need to first identify, clone, and express individual targets. Furthermore, the data that is obtained from the screening process provides information not only regarding potency of the potential lead compounds but also regarding selectivity of the compounds in a physiologically relevant sample that contains many closely related family members.

To demonstrate the utility of this approach a series of small molecule inhibitor libraries were designed based on a core peptide backbone coupled to the epoxide electrophile contained in the DCG-04 probes (Fig. 4A). Initially, PSLs were synthesized in which a single amino acid position was scanned through a series of natural and non-natural amino acids, whereas the remaining two positions were coupled with a mixture of all possible natural amino acids (minus cysteine and methionine and including norleucine). The resulting sublibraries were composed of 361 members each. Scanning of constant amino acids at the P3 and P4 positions through all natural amino acids indicated that these elements did not significantly contribute to selectivity of inhibitor binding to protease targets (data not shown). Therefore, only data compiled for scanning of the constant P2 position are presented. To increase the diversity of the small molecules in the PSLs we included 42 hydrophobic non-natural amino acids as building blocks (see Table I in Supplemental Material). In addition, each of the natural amino acids was coupled to the mirror-image enantiomeric form of the epoxide (2R, 3R versus 2S, 3S). Previous work indicates that this change in stereochemistry favors binding of the inhibitors on the prime side of the active site resulting in more diversity in our libraries (23).

PSLs were first screened against the primary protease targets of DCG-04 in rat liver homogenates (bands 1–4 from Fig. 2B; see Fig. 4B). Potency was assessed by pretreatment of total cell extracts with each library followed by labeling with  $^{125}\text{I}$ -DCG-04 and analysis by SDS-PAGE and autoradiography. The ability of each library to block active site labeling by DCG-04 was measured as a percentage competition relative to an untreated control. The resulting values were visualized using software developed by Eisen and co-workers (3) designed to analyze data generated from microarray analysis. This software assigns a color to numerical competition values and allows clustering of profiles based on similarities across diversity positions (x axis) and enzyme family members (y axis). The resulting “clustergram” is shown in Fig. 4B.

Clustering data throughout the constant amino acid resi-



**Fig. 4. Screening of peptide epoxide PSLs.** *A*, structures of the general PSL scaffolds containing either (S, S) or (R, R) epoxides. PSLs contain a fixed P2 position (X) and P3 and P4 positions composed of an isokinetic mixture of 19 natural amino acids (all natural amino acids minus cysteine and methionine, plus norleucine; *Mix*). *B*, colorimetric cluster display of inhibition data. PSLs were used to profile proteases in rat liver extracts by pretreatment of samples with individual constant P2 libraries followed by labeling with  $^{125}\text{I}$ -DCG-04. Labeling intensity of each target relative to the control untreated sample was used to generate percent competition values. These resulting data were clustered and visualized using programs designed for analysis of microarray data (see “Materials and Methods”). *Numbers* along the *top* correspond to individual non-natural amino acids listed in Table I in Supplemental Material. *Single letters* represent the single letter codes for each of the natural amino acids. *n* corresponds to norleucine that was used in place of methionine to avoid problems with side-chain oxidation. The *tree structures* at the *top* and *left* of the *diagrams* were obtained by hierarchical clustering and indicate the degree of similarity as a function of the height of the lines connecting profiles. Unknown protease bands in rat liver are numbered 1–4 and correspond to the bands shown in Fig. 2B. The *color key* is shown at the *bottom*.

dues grouped the data such that residues that showed overall poor binding to all targets were positioned to the right, and residues that showed universal strong binding were positioned to the left. The remaining residues in the middle of the cluster-gram showed some degree of selectivity for individual enzymes. The results from the clustering indicate that the non-natural amino acids and natural amino acids linked to the (R, R) enantiomer of the epoxide provided the greatest target selectivity.

Specificity profiles for each of the major protease species labeled by DCG-04 also identified several residues that clustered to the center of the profile that confer unique specificity for an individual protease species in the extract. Therefore, this method yielded interesting lead compounds using a relatively small number of libraries (~80) with limited structural diversity. A similar screen of a larger, more structurally diverse small molecule library is likely to provide a greater number of inhibitor leads. Given the relative ease of screening and the abundance of the protein extracts, such a large-scale screen is clearly accessible using this methodology.

*Profiling Changes in Protease Activities upon Addition of Selective Small Molecule Inhibitors*—Analysis of the library

data from screening of liver extracts indicated that several PSLs showed selective binding to a single protease. We chose to focus on the constant P2 glutamine (R, R) epoxide library because of its high degree of selectivity for protease 2 in the extract. Liver extracts were either directly labeled with the Red-DCG-04 probe or treated with the library and then labeled with the Blue-DCG-04 probe. The samples were then combined and subjected to a first dimension of isoelectric focusing followed by analysis by SDS-PAGE in the second dimension using the DNA sequencer (Fig. 5A). This method allowed analysis of multiple channels of data in a single gel that could be merged to determine changes in activity of each protease species in the presence of the inhibitor library. The resulting 2D profile unambiguously demonstrated that the glutamine (R, R) library specifically binds to the active site of a single protease (spot 2) as indicated by loss of labeling in the blue channel.

To determine the identity of the protease selectively targeted by the small molecule library, we used the biotin-tagged DCG-04 to perform a single-step affinity purification of all labeled proteases from liver extracts. The resulting silver-

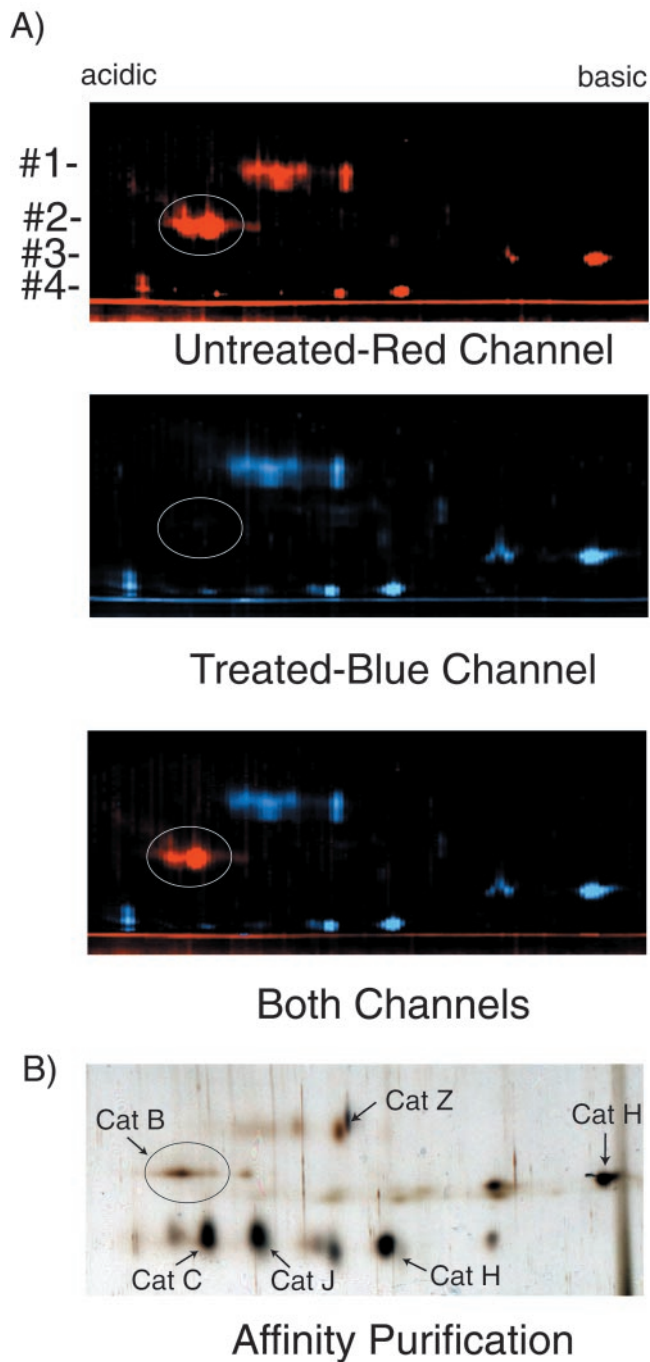


FIG. 5. **Profiling changes in protease activity upon inhibitor treatment.** A, liver extracts (100  $\mu\text{g}$  of total protein) were treated with 100 nM Red-DCG-04 or with 10  $\mu\text{M}$  Ac-XX-Q-(R, R)Eps library for 30 min and then with 100 nM Blue-DCG-04. Reactions were quenched with IEF sample buffer, and equal amounts of each reaction were co-loaded on a single IEF tube gel. Labeled proteins were separated on a 15% SDS-PAGE and analyzed using an ABI 377 DNA sequencer. The lower panel shows the red and blue channels overlaid on a single image whereas the upper and middle panels show the individual fluorescence channels. Note the loss of activity of the circled protease upon inhibitor treatment. B, active proteases in the liver extract were purified by a single step affinity purification of DCG-04-labeled liver extract. Silver-stained spots were excised and sequenced by liquid

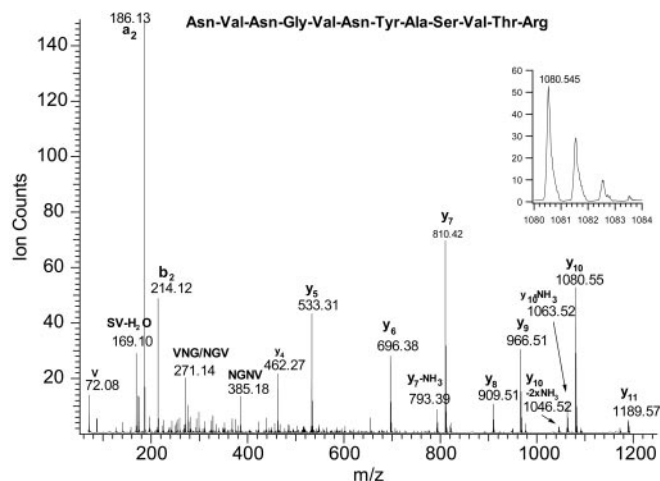
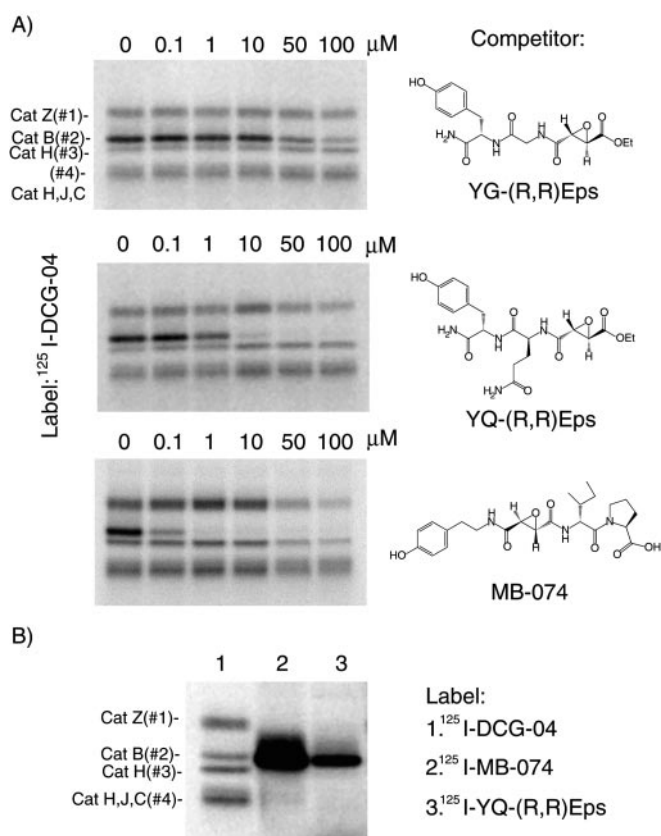


FIG. 6. **Typical CID spectrum of a tryptic peptide generated from in-gel digestion of an affinity-purified cysteine protease.** The precursor ion was  $m/z$  647.34 ( $2^+$ ). The spectrum was acquired on a quadrupole orthogonal acceleration-time-of-flight tandem mass spectrometer (Pulsar; MDS Sciex), in information-dependent acquisition mode during the liquid chromatography/MS analysis of the tryptic digest of the protein purified by 2D electrophoresis (See Fig. 5B). A data base search by Mascot identified the protein as cathepsin Z using these and other CID data. The inset shows the resolution afforded by this instrument.

stained 2D profile shows that all fluorescently labeled protease could be rapidly purified from the crude extract and correlated with the labeling profiles (Fig. 5B). The silver-stained spot corresponding to band 2 was excised and identified as cathepsin B by liquid chromatography-MS-time-of-flight CID sequencing (Fig. 6). Furthermore, several other cathepsin family members including cathepsins Z, H, C, and J were identified by this method.

*Design of Selective Inhibitors Based on Library Screening Data*—Using information from the scanning of our PSLs, we synthesized several individual compounds designed to validate the library approach. In all cases a P3 tyrosine was included as a site for radioiodination, and the P2 residue was chosen based on target selectivity. P2 glutamine attached to the (R, R) epoxide inhibitor (YQ-(R, R)-Eps) was chosen because of its selectivity for cathepsin B in the extract, and P2 glycine was chosen as a negative control. The cathepsin B-specific ABP MB-074 (18) was used as a control for comparison with YQ-(R, R)-Eps. Compounds were added to extracts over a wide concentration range, and activity for each target was assessed by labeling with  $^{125}\text{I}$ -DCG-04 (Fig. 7A). As expected, YQ-(R, R)-Eps and MB-074 selectively blocked labeling of the cathepsin B band (number 2) whereas YG-(R, R)-Eps showed little or no inhibition of all of the proteases.

chromatography-MS-time-of-flight CID. The silver-stained spot corresponding to the labeled protease inhibited by Ac-XX-Q-(R, R)-Eps library was identified as cathepsin (Cat) B. Other papain family protease were also identified and are labeled with arrows.



**FIG. 7. Evaluation of specific protease inhibitors selected from library screening.** Competition analysis of a negative control compound (YG-(R, R)Eps), a cathepsin B-specific compound identified from the library screening (YQ-(R, R)Eps), and a previously described cathepsin B-specific inhibitor (MB-074) is shown. Several concentrations of each compound were incubated with 100  $\mu\text{g}$  of total liver extract for 30 min followed by labeling with  $^{125}\text{I}$ -DCG-04 for 1 h. *Cat*, cathepsin. *A*, inhibition dose response profiles for each compound. *B*, direct labeling of 100  $\mu\text{g}$  of total liver extract with radiolabeled versions of DCG-04, MB-074, and YQ-(R, R)Eps. Note the specificity of MB-074 and YQ-(R, R)Eps for cathepsin B.

The newly developed cathepsin B inhibitor was also radioiodinated and used to label liver homogenates (Fig. 7B). The labeling profile was compared with the profiles for the cathepsin B-specific probe  $^{125}\text{I}$ -MB-074 and the generally reactive probe  $^{125}\text{I}$ -DCG-04. YQ-(R, R)-Eps, like MB-074, showed selective labeling of the band identified as cathepsin B.

We conclude that it is possible to rapidly identify a structurally distinct class of cathepsin B-selective inhibitors by screening of libraries of limited complexity. The resulting lead compound, although not excessively potent, now serves as a template for the design of optimized inhibitors that are distinct from the CA-074 class of cell-impermeable cathepsin B inhibitors. No doubt this approach could also be used to selectively target other cathepsin family members through a more extensive library screening effort.

In summary, we have developed tools to identify families of related enzymes within a complex proteome. These tools can

be used to determine relative activity levels of these enzymes and to visualize their localization in live cells. These tools also allow rapid design and screening of small molecule inhibitors for select targets. In the current study we successfully identified a new cathepsin B-selective inhibitor by screening of a small set of libraries in crude liver extracts. Furthermore, we have developed a general method for rapid analysis of large data sets generated from library screening of multiple targets in crude cell extracts. This approach allows rapid comparison of inhibitors, as well as targets based on similarities in structure-function relationships. This general functional proteomic method, although applied here to papain family proteases, can also be used for a wide range of enzyme families through design and synthesis of new families of class-specific affinity probes.

**Acknowledgments**—We thank Dieter Brömme (Mt. Sinai School of Medicine) and Vito Turk (Jozef Stefan Institute) for the kind gift of purified cathepsins. We thank David Ginzinger for assistance with operating the ABI DNA sequencer and for troubleshooting with data analysis.

\* This work was supported in part by National Institutes of Health Grants NCRR 01614 and RR12961 (to the MS Facility Director, A. L. Burlingame, and to K. M.), by the Eotvos Scholarship of the Hungarian Scholarship Board (to Z. D.), and by funding from the Sandler Program in Basic Sciences (to D. G., L. H., A. B., and M. B.). 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: University of California, San Francisco Campus, Box 0448, 513 Parnassus Ave., San Francisco, CA 94143–0448. Tel.: 415-502-8142; Fax: 415-502-4315; E-mail: mbogyo@biochem.ucsf.edu.

## REFERENCES

- Marcotte, E. M., Pellegrini, M., Thompson, M. J., Yeates, T. O., and Eisenberg, D. (1999) A combined algorithm for genome-wide prediction of protein function. *Nature* **402**, 83–86
- Eisenberg, D., Marcotte, E. M., Xenarios, I., and Yeates, T. O. (2000) Protein function in the post-genomic era. *Nature* **405**, 823–826
- Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. (1998) Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 14863–14868
- Dove, A. (1999) Proteomics: translating genomics into products? *Nat. Biotechnol.* **17**, 233–236
- Pandey, A., and Mann, M. (2000) Proteomics to study genes and genomes. *Nature* **405**, 837–846
- Gygi, S. P., Rochon, Y., Franza, B. R., and Aebersold, R. (1999) Correlation between protein and mRNA abundance in yeast. *Mol. Cell. Biol.* **19**, 1720–1730
- Stockwell, B. R. (2000) Frontiers in chemical genetics. *Trends Biotechnol.* **18**, 449–455
- Schreiber, S. L. (1998) Chemical genetics resulting from a passion for synthetic organic chemistry. *Bioorg. Med. Chem.* **6**, 1127–1152
- Cyglar, M., Sivaraman, J., Grochulski, P., Coulombe, R., Storer, A. C., and Mort, J. S. (1996) Structure of rat procathepsin B: model for inhibition of cysteine protease activity by the prosegment. *Structure* **4**, 405–416
- Coulombe, R., Grochulski, P., Sivaraman, J., Ménard, R., Mort, J. S., and Cyglar, M. (1996) Structure of human procathepsin L reveals the molecular basis of inhibition by the prosegment. *EMBO J.* **15**, 5492–5503
- Chapman, H. A., Reese, R. J., and Shi, G.-P. (1997) Emerging roles for cysteine proteases in human biology. *Annu. Rev. Physiol.* **59**, 63–88
- Shaw, E. (1994) Peptidyl diazomethanes as inhibitors of cysteine and serine



- proteinases. *Methods Enzymol.* **244**, 649–656
13. Yan, S., Sameni, M., and Sloane, B. F. (1998) Cathepsin B and human tumor progression. *Biol. Chem.* **379**, 113–123
  14. Gelb, B. D., Shi, G. P., Chapman, H. A., and Desnick, R. J. (1996) Pycnodysostosis, a lysosomal disease caused by cathepsin K deficiency. *Science* **273**, 1236–1238
  15. Iwata, Y., Mort, J. S., Tateishi, H., and Lee, E. R. (1997) Macrophage cathepsin L, a factor in the erosion of subchondral bone in rheumatoid arthritis. *Arthritis Rheum.* **40**, 499–509
  16. Liu, Y., Patricelli, M., and Cravatt, B. (1999) Activity-based protein profiling: the serine hydrolases. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 14694–14699
  17. Kidd, D., Liu, Y., and Cravatt, B. F. (2001) Profiling serine hydrolase activities in complex proteomes. *Biochemistry* **40**, 4005–4015
  18. 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
  19. Greenbaum, D., Medzihradzky, K. F., Burlingame, A. L., and Bogyo, M. (2000) Epoxide electrophiles as activity-dependent cysteine protease profiling and discovery tools. *Chem. Biol.* **7**, 569–581
  20. Faleiro, L., Kobayashi, R., Fearnhead, H., and Lazebnik, Y. (1997) Multiple species of CPP32 and Mch2 are the major active caspases present in apoptotic cells. *EMBO J.* **16**, 2271–2281
  21. Adam, G. C., Cravatt, B. F., and Sorensen, E. J. (2001) Profiling the specific reactivity of the proteome with non-directed activity-based probes. *Chem. Biol.* **8**, 81–95
  22. Baruch, A., Greenbaum, D., Levy, E. T., Nielsen, P. A., Gilula, N. B., Kumar, N. M., and Bogyo, M. (2001) Defining a link between gap junction communication, proteolysis, and cataract formation. *J. Biol. Chem.* **276**, 28999–29006
  23. Schaschke, N., Assfalg-Machleidt, I., Machleidt, W., Turk, D., and Moroder, L. (1997) E-64 analogues as inhibitors of cathepsin B on the role of the absolute configuration of the epoxysuccinyl group. *Bioorg. Med. Chem.* **5**, 1789–1797
  24. Ostresh, J. M., Winkle, J. H., Hamashin, V. T., and Houghten, R. A. (1994) Peptide libraries: determination of relative reaction rates of protected amino acids in competitive couplings. *Biopolymers* **34**, 1681–1689
  25. Bogyo, M., Shin, S., McMaster, J. S., and Ploegh, H. L. (1998) Substrate binding and sequence preference of the proteasome revealed by active site-directed affinity probes. *Chem. Biol.* **5**, 307–320
  26. Nazif, T., and Bogyo, M. (2001) Global analysis of proteasomal substrate specificity using positional-scanning libraries of covalent inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 2967–2972