Epoxide electrophiles as activity-dependent cysteine protease profiling and discovery tools

Doron Greenbaum [1], Katalin F Medzihradszky [1,2], Alma Burlingame [1,2] and Matthew Bogyo [3]

Background: Analysis of global changes in gene transcription and translation by systems-based genomics and proteomics approaches provides only indirect information about protein function. In many cases, enzymatic activity fails to correlate with transcription or translation levels. Therefore, a direct method for broadly determining activities of an entire class of enzymes on a genome-wide scale would be of great utility.

Results: We have engineered chemical probes that can be used to broadly track activity of cysteine proteases. The structure of the general cysteine protease inhibitor E-64 was used as a scaffold. Analogs were synthesized by varying the core peptide recognition portion while adding affinity tags (biotin and radio-iodine) at distal sites. The resulting probes containing a P2 leucine residue (DCG-03 and DCG-04) targeted the same broad set of cysteine proteases as E-64 and were used to profile these proteases during the progression of a normal skin cell to a carcinoma. A library of DCG-04 derivatives was constructed in which the leucine residue was replaced with all natural amino acids. This library was used to obtain inhibitor activity profiles for multiple protease targets in crude cellular extracts. Finally, the affinity tag of DCG-04 allowed purification of modified proteases and identification by mass spectrometry.

Conclusions: We have created a simple and flexible method for functionally identifying cysteine proteases while simultaneously tracking their relative activity levels in crude protein mixtures. These probes were used to determine relative activities of multiple proteases throughout a defined model system for cancer progression. Furthermore, information obtained from libraries of affinity probes provides a rapid method for obtaining detailed functional information without the need for prior purification/identification of targets.

[1]Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143, USA
[2]Mass Spectrometry Facility, University of California, San Francisco, CA 94143, USA
[3]Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94143, USA

Correspondence: Matthew Bogyo
E-mail: mbogyo@biochem.ucsf.edu

Keywords: Affinity labeling; Cysteine protease; E-64; Electrophile; Epoxide; Proteomics

Received: **14 April 2000**
Accepted: **9 May 2000**

Published: **1 August 2000**

Chemistry & Biology 2000, **7**:569–581

1074-5521/00/$ – see front matter

PII: S1074-5521(00)00014-4

Introduction

New approaches for studying global cellular processes must permit the analysis of differential changes within large sets of known and unknown genes or proteins. DNA microarray techniques allow analysis of genome-wide changes in mRNA transcription for a given cellular stimulus [1,2]. Advances in two-dimensional (2D) gel electrophoresis coupled to highly sensitive mass spectrometry techniques now allow the rapid identification of proteins from whole cells or tissue extracts [3,4]. While these techniques have revolutionized the global analysis of biological processes, often information about function of enzymatic proteins can only be inferred by analysis of transcriptional/translational co-regulation of sets of genes under different stimuli. In fact, levels of transcription and translation of an enzyme, in many cases, do not correlate with its activity [5].

To assign a function to enzymatic proteins on a genome-wide scale, a method to obtain direct information about enzymatic activity is necessary. Since the simultaneous targeting of all enzyme classes with a single probe is likely to be impossible, we chose to focus our effort on proteolytic enzymes. The papain family of cysteine proteases serves as a good model system for several reasons. Firstly, most cysteine proteases are synthesized with an inhibitory propeptide that must be proteolytically removed to activate the enzyme [6,7], resulting in expression profiles that do not directly correlate with activity. Secondly, the largest set of papain-like cysteine proteases, the cathepsins, act in concert to digest a protein substrate. Thus, information regarding regulation of activity of each member relative to one another is critical for understanding their collective function. Furthermore, the cathepsins are involved in many critical biological processes, and biochemical studies of function have been limited to family members that have been cloned and expressed or purified from crude tissue. Finally, a large body of information is available regarding covalent, mechanism-based inhibitors that specifically target this family of cysteine proteases.

The papain family is classified into several major groups, most notable of which are the bleomycin hydrolases, calpains, caspases and cathepsins. To date, 16 human cathepsins have been cloned and sequenced. Several of these proteases are key players in normal physiological processes such as antigen presentation [8], bone remodeling [9] and prohormone processing [10]. In addition, several of these proteases are involved in pathological processes such as rheumatoid arthritis [11], cancer invasion and metastasis [12] and Alzheimer's disease [13,14].

The enzymatic mechanism used by the papain family of proteases has been well studied and is highly conserved. Thus, electrophilic substrate analogs that are only reactive in the context of this conserved active site can be used as general probes of function. A wide range of electrophiles has been developed as mechanism-based, cysteine protease inhibitors including diazomethyl ketones [15], fluoromethyl ketones [16], acyloxymethyl ketones [17], *O*-acylhydroxylamines [18], vinyl sulfones [19] and epoxysuccinic derivatives [20]. These inhibitors all consist of a peptide specificity determinant attached to an electrophile that becomes irreversibly alkylated when bound in close proximity to an attacking nucleophile.

Several groups have recognized the value of using irreversible mechanism-based inhibitors as affinity labels [21–25]. Addition of a reporter function, such as a radioactive iodine, to inhibitors permits the visualization of covalently modified proteases in a standard SDS–PAGE gel format. Labeling intensity provides a read-out of relative enzymatic activity. Furthermore, both known and novel proteases are targets for analysis by this methodology. Similar affinity labeling approaches have been used extensively to study or identify proteases such as the proteasome [22,26,27], caspases [28,29], cathepsins [23,25] and methionine amino peptidase [30,31]. Recently, Cravatt and coworkers have taken advantage of the broad class-specific reactivity of fluorophosphonates towards serine proteases [32]. By incorporation of a simple, extended alkyl chain capped with a biotin moiety, they have created a broad serine protease-specific probe (FP-Biotin) for functional proteomic analysis of serine proteases in crude cellular extracts.

We have developed functional proteomics tools that can be used to determine global patterns of activity for the papain family of cysteine proteases based on the broad reactivity of the natural product E-64. These tools provide functional information that can be used in concert with existing ge-

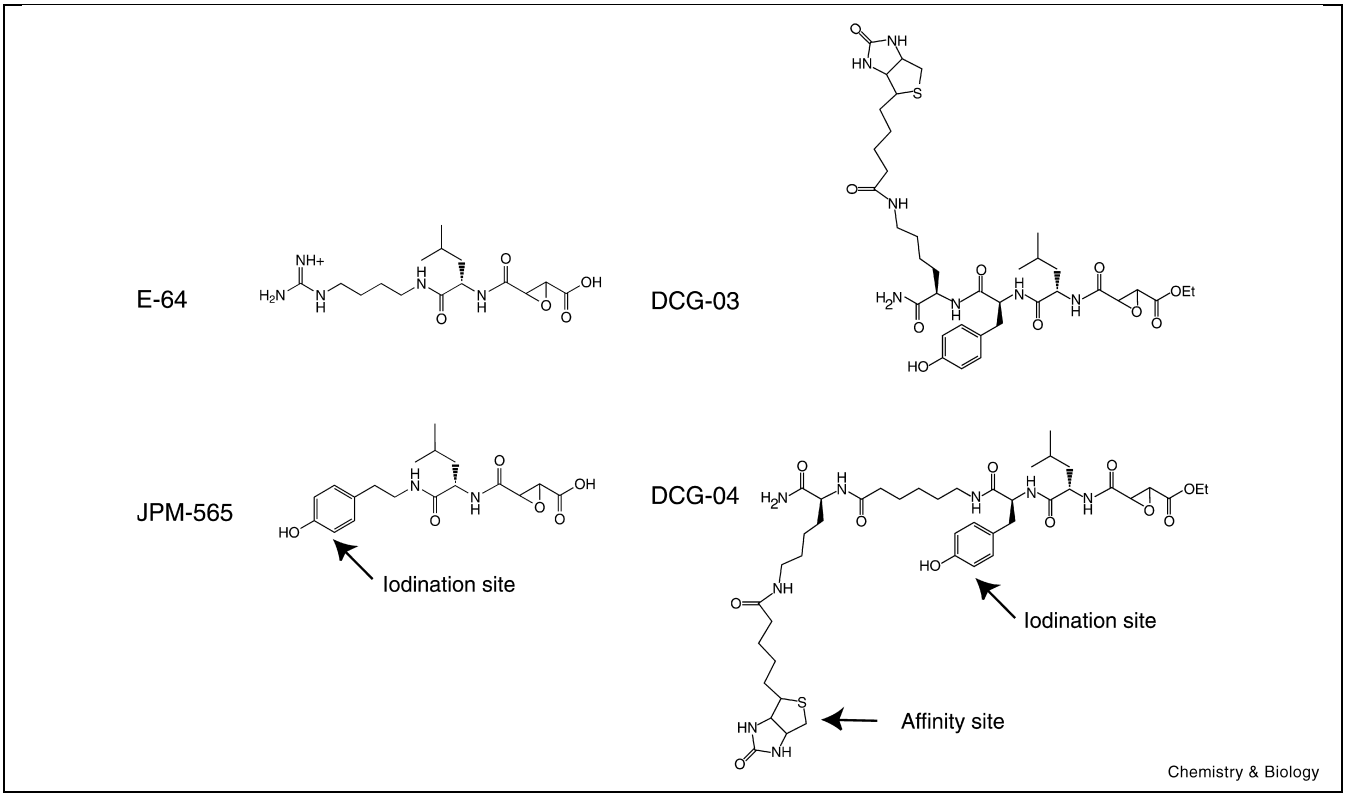

Figure 1. Structures of epoxide inhibitors and probes E-64, JPM-565, DCG-03 and DCG-04. Radiolabel attachment and affinity sites are indicated for each compound.

nomic and proteomic methods to correlate gene and protein expression profiles with enzymatic activity. Furthermore, diversification of core compounds using solid-phase combinatorial chemistry provides libraries of compounds that can be used to obtain information about inhibitor specificities of targeted protease. This information is likely to be of use in the generation of selective inhibitors without the need for prior characterization and purification of protease targets. Finally, the probes themselves can be used to rapidly identify targets after covalent modification.

Results and discussion

Design and synthesis of DCG-04

The natural product E-64 is a promiscuous irreversible cysteine protease inhibitor that is broadly reactive toward the papain family of cysteine proteases [20] (Figure 1). Its leucine side chain mimics the P2 amino acid of a substrate, occupying the target's S2 binding pocket while the agmatine moiety binds in the S3 position [33]. Rich et al. synthesized JPM-565 (Figure 1), a derivative in which a tyramine moiety replaces the agmatine side chain of E-64 [34,35]. This closely related compound was found to have similar class-specific reactivity for cysteine proteases as E-64. Since the P2 position of a substrate is considered to be the main specificity determinant for many cysteine proteases, we reasoned that further extension of the non-prime binding portion of JPM-565 would not significantly perturb binding affinity for a target protease. In addition, modification to the non-prime site binding element of the E-64 derivative CA-074 had little effect on binding to cathepsin B [23,36]. Elaboration of the peptide portion of E-64 allowed both incorporation of an affinity tag as well as attachment of the compound to a solid support. The resulting bi-functional compounds, DCG-03 and DCG-04, contain both the iodinatable phenol ring of JPM-565 and the additional affinity site created by incorporation of a side chain biotinylated lysine residue (Figure 1). Addition (DCG-04) or removal (DCG-03) of an amino hexanoic acid spacer between the affinity site and the electrophile was used to determine the space requirement for binding and recognition of the affinity label by support-bound avidin.

Peptide epoxides were synthesized using a combination of solution and solid-phase chemistries. The solution-phase synthesis of the epoxide acid building block (Figure 2A) starting from commercially available diethyl tartrate has been reported elsewhere [23]. Standard solid-phase peptide chemistry was used to build the peptide portion of DCG-04 and related compounds (Figure 2B). This methodology provides a flexible system with which to incorporate virtually any peptide sequence prior to attachment of the electrophilic epoxide. Surprisingly, the epoxy acid building block was stable to standard solid-phase peptide synthesis cleavage conditions (95% TFA). The use of solid-phase chemistry also allowed the synthesis of a diverse library in which the P2 leucine of DCG-04 was replaced with each of the natural amino acids (except cysteine due to reactivity with the epoxide and methionine due to oxidation). The non-natural amino acid norleucine was used as an isosteric methionine analog. The results obtained using this 19 member library of compounds are described below.

DCG-04 is an activity-dependent affinity label

Dendritic cells express relatively high levels of lysosomal cathepsins, making them a logical source of material for

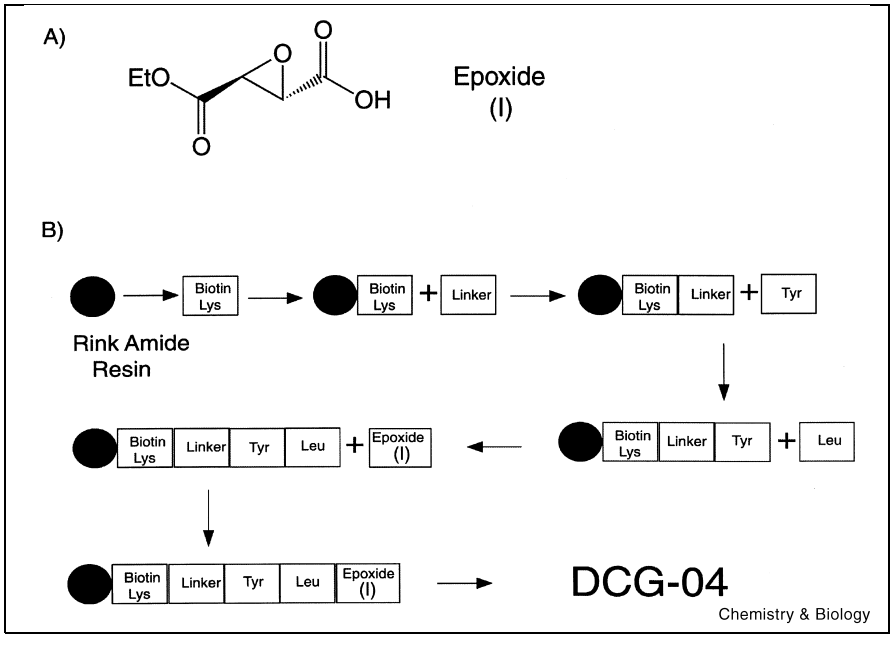

Figure 2. Synthesis of DCG-04. **(A)** Epoxy acid building block (I) and **(B)** solid-phase synthesis scheme for DCG-04. Details of the synthesis and characterization of peptide epoxides can be found in Materials and methods.

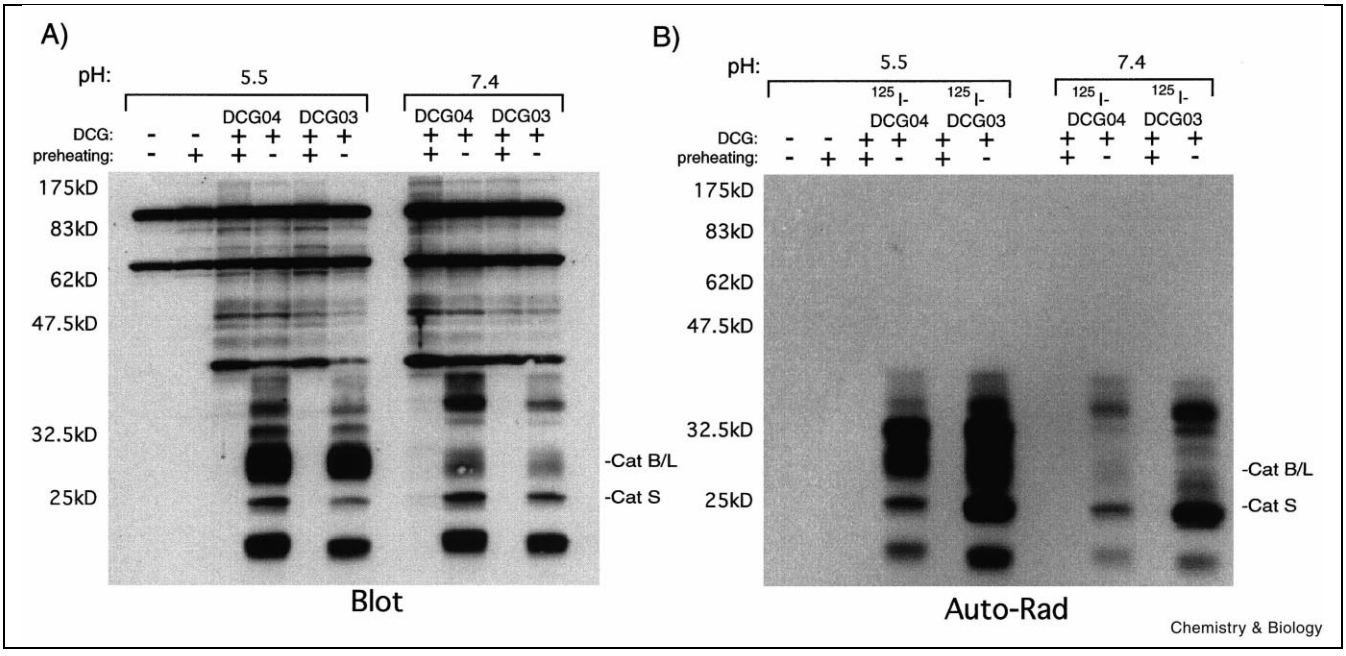

Figure 3. DCG-03 and DCG-04 label active proteases in dendritic cell extracts. **(A)** Total cell extracts from DC2.4 cells were diluted into either pH 5.5 or pH 7.4 buffer, pre-heated to 100°C for 1 min (+pre-heating) or not (−pre-heating) and labeled with 50 μM DCG-03 and DCG-04. Samples were separated by SDS–PAGE (12.5% gel) and labeled bands visualized by affinity blotting as described in Materials and methods. **(B)** Same as for **(A)** except ^{125}I-labeled versions of DCG-03 and DCG-04 were used and gels analyzed by autoradiography. The locations of cathepsins B, L and S are indicated for reference based on their known molecular weights.

establishing parameters for the use of DCG-04. Figure 3 shows the labeling profile of polypeptides modified by incubation with either DCG-03, DCG-04, ^{125}I-DCG-03 or ^{125}I-DCG-04 followed by SDS–PAGE analysis. Radio-iodinated (autoradiogram) and non-radio-iodinated (blot) DCG-03 and DCG-04 labeled multiple polypeptides in the range of 20–40 kDa.

Although the intrinsic reactivity of the epoxide electrophile portion of DCG-04 towards free thiols is quite poor, we wanted to determine if DCG-04 and its derivatives were capable of non-specific alkylation of proteins in crude cellular extracts. A pre-heating control was used to reveal non-specific labeling, with the assumption that denatured, inactive proteins modified by DCG-03 and DCG-04 represent non-specific modifications. Enzymatically active proteins were deduced by subtraction (Figure 3). Labeling of all of the major species in the 20–40 kDa size range was lost upon heat denaturation of samples prior to addition of compounds, suggesting that labeling is dependent on enzymatic activity and that these bands correspond to the major proteases in the extract. Several higher molecular weight species were observed by affinity blotting of both denaturing controls and samples in which no inhibitor was added. These species are likely to represent non-specific alkylations and endogenously biotinylated proteins.

Comparison of labeling, at neutral (pH 7.4) and at the acidic pH of the lysosome (pH 5.5), indicated that several of the modified polypeptides in the 30 kDa size range required reduced pH for activity. This result is consistent with reported findings that several lysosomal cysteine proteases either reversibly or irreversibly lose activity upon de-acidification of lysosomal compartments [37].

Analysis of the labeling of DC2.4 lysates by both affinity blot and autoradiography techniques resulted in similar profiles of modified polypeptides, highlighting the utility of both techniques. However, the autoradiogram showed labeling of only enzymatically active polypeptides by radio-labeled forms of DCG-03 and DCG-04. Addition of the rather bulky iodine atom to DCG-03 and DCG-04 had only a modest effect on target modification yet resulted in compounds with dramatically reduced background labeling and increased sensitivity. Ultimately, the ability to use both autoradiography as well as blot techniques enhances the flexibility of these protease detection reagents and further highlights the utility of bi-functional inhibitors.

DCG-04 targets cysteine proteases inhibited by E-64 and JPM-565

Both direct labeling and indirect competition experiments were performed to confirm that DCG-04 reacts with a sim-

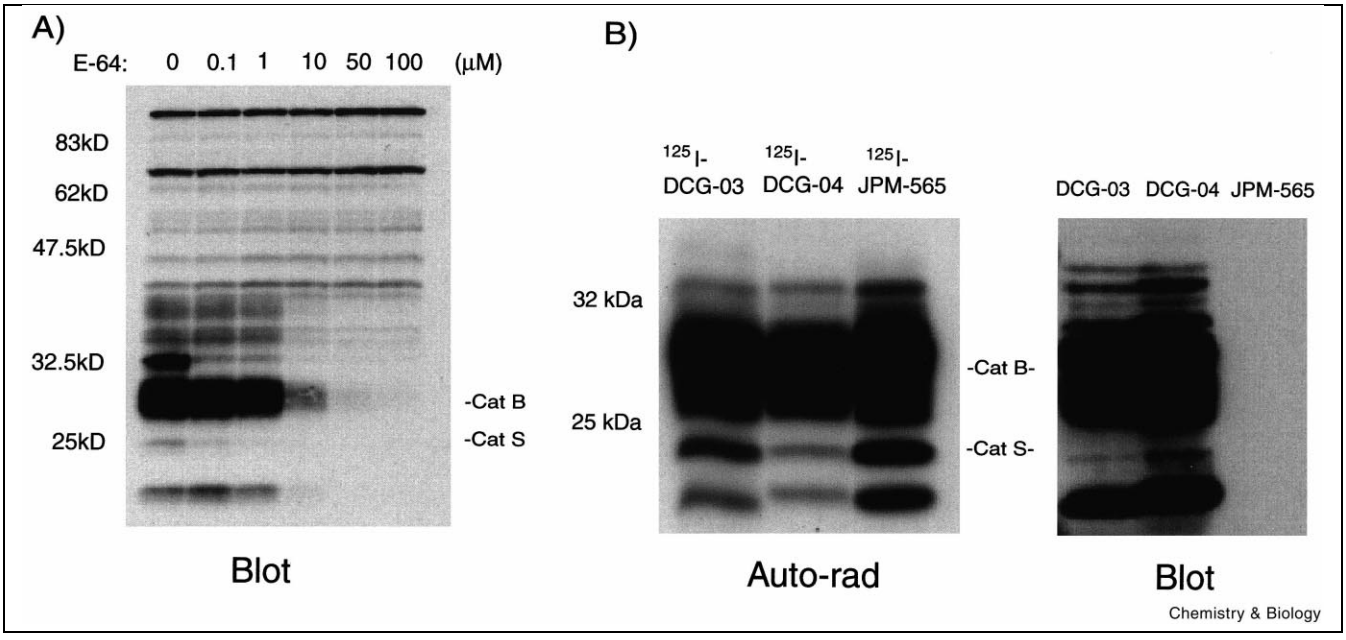

Figure 4. DCG-03 and DCG-04 target the same polypeptides as parent compounds E-64 and JPM-565. **(A)** Total cellular extracts from DC2.4 cells were incubated with increasing concentrations of E-64 as indicated for 30 min at 25°C followed by addition of 50 μM DCG-04 and further incubation for 1 h. Samples were resolved by SDS–PAGE (12.5%) and labeled bands visualized by affinity blotting. **(B)** Total cellular extracts were labeled with either ^{125}I-labeled forms (auto-rad) or with non-labeled forms (blot) of DCG-03, DCG-04 and JPM-565 followed by separation by SDS–PAGE (12.5%) and analysis as indicated. The locations of cathepsin B and S are indicated for reference based on their known molecular weights.

ilar subset of proteases to the parent compounds E-64 and JPM-565. An indirect competition experiment was required to determine the polypeptides modified by E-64 since it lacks an affinity labeling site. Extracts from the dendritic cell line DC2.4 were preincubated with an increasing concentration of E-64 followed by labeling with DCG-04. Final labeling intensity was used to indirectly monitor the extent of polypeptide modification by E-64. The competition revealed that all polypeptides labeled by DCG-04 are effectively competed by E-64, indicating that the two compounds target the same subset of proteases (Figure 4A). A similar competition experiment was performed using the cathepsin B-specific inhibitor MB-074 [23]. These results positively identified the diffuse 30 kDa polypeptide (labeled cat B in Figure 4A) as cathepsin B (data not shown).

Comparison of the specificity of DCG-03, DCG-04 and JPM-565 was accomplished using direct labeling of DC2.4 cell lysates. Labeling profiles obtained for ^{125}I-DCG-03, ^{125}I-DCG-04 and ^{125}I-JPM-565 were identical for all three probes and indicated that each targeted polypeptides in the 20–40 kDa size range (Figure 4B). Analysis of non-radiolabeled DCG-03-, DCG-04- and JPM-565-treated extracts again showed the similarity of the blotting and autoradiography detection systems. As expected, JPM-565, which lacks a biotin label, showed no labeling as detected by affinity blotting. Together these results establish that modifications to the extended binding portion of the E-64 family of compounds have little effect on selectivity or potency. However, this region of the inhibitor may still play an important role in establishing specificity of binding when equipped with the proper recognition sequence. Future work is aimed at exploring the use of extended peptide recognition motifs to fine tune selectivity of the DCG family of inhibitors for specific protease targets.

Profiling applications

The aforementioned methods established the initial parameters for use of the general cysteine protease labels DCG-03 and DCG-04. We next wanted to apply these techniques to profile the activity and specificity of cysteine proteases in several different model systems. The broadly reactive probe DCG-04 was used to generate activity profiles of multiple protease targets both in a model for disease progression and throughout multiple tissue types. Similarly, activity profiles were generated using the cathepsin B-specific probe MB-074 to provide complementary information for a single, well-defined cysteine protease target. This information was also used to positively establish the identity of cathepsin B in the DCG-04 labeling pro-

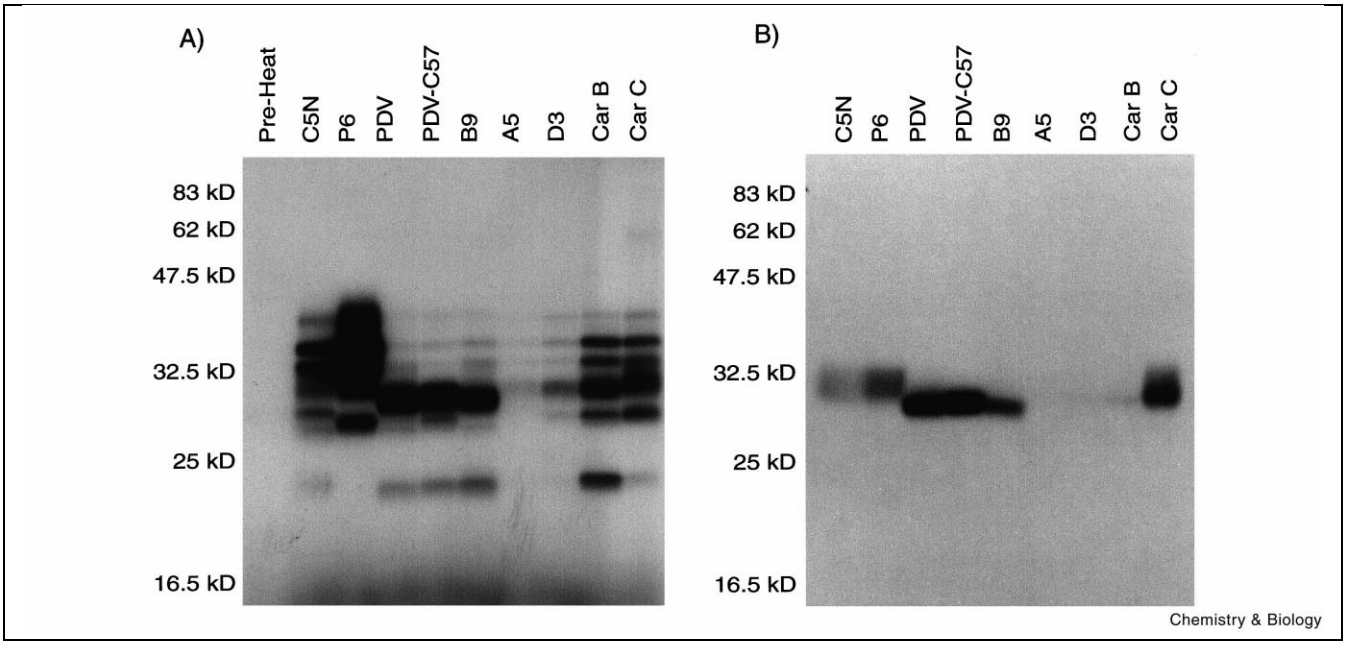

Figure 5. Activity profiling across a disease progression. Tissue culture cells were isolated from carcinomas generated by application of a chemical mutagen to the skin of mice (see Materials and methods). Progression begins at the left with the non-invasive benign cells (C5N and P6) and progresses to the right through papilloma cell lines (PDV and PDV-C57), squamous cell carcinomas (B9, A5 and D3) and finally highly invasive spindle cell carcinomas (Car B and Car C). Total cellular lysates were normalized with respect to protein concentration and labeled with **(A)** ^{125}I-DCG-04 and **(B)** the cathepsin B-specific probe ^{125}I-MB-074. A pre-heat control from the C5N lysate was included in **(A)** to show background labeling.

files. To obtain more detailed functional information for DCG-04-modified proteases, inhibitor specificity profiles were generated using a library of DCG-04 analogs in total cellular extracts. The same libraries were also used in conjunction with the cathepsin B-specific probe, ^{125}I-MB-074, as well as with purified cathepsin H to determine specificity profiles for individual target proteases. These results are described below.

Profiling across disease progression using DCG-04 and MB-074

The mouse skin model of multi-stage carcinogenesis has been well studied genotypically and phenotypically, has discrete steps in the progression, but lacks information on cysteine protease involvement [38,39]. The role of cathepsins in tumor biology has mostly focused on cathepsins B and L. Upregulated levels of both cathepsins B and L have been shown to correlate with an invasive phenotype [12,40]. Furthermore, cathepsins B and L are secreted by many types of tumorigenic cells and treatment of invasive cells with the cysteine protease inhibitor E-64 results in a block in cellular invasion into a synthetic matrix [41,42]. These data indicate that cathepsins are likely to play an important role in the metastatic process.

Ten cell lines representing various steps in the progression from benign skin cell (C5N) to highly invasive spindle cell carcinomas (Car B and Car C) were used to analyze global changes in activity of cathepsins throughout this multi-stage carcinogenesis model. The carcinoma progression also includes benign papilloma cell lines P6, PDV and PDV-C57, and more invasive squamous cell carcinoma cell lines B9, A5 and D3. Equal amounts of protein from each cell lysate were labeled with both the broadly reactive probe, ^{125}I-DCG-04, as well as the cathepsin B-specific probe, ^{125}I-MB-074 at pH 5.5 (Figure 5). The results show that several protease activities, including cathepsin B, dramatically fluctuate across the panel of cell lines.

The broadly reactive probe ^{125}I-DCG-04 highlights the activity of several proteases in the lysosomal cysteine protease size range in each of the cell types (Figure 5A). The benign cell lines C5N and P6 both contain multiple labeled polypeptides between 28 and 45 kDa; however, the labeling intensity observed for the P6 line is dramatically increased for all polypeptides in this range. Interestingly, the major difference between these cell lines is an activating mutation in the *ras* gene [43]. It has previously been shown that various classes of proteases, including the cathepsins, are upregulated downstream of Ras; however, these studies were limited to analysis of expression levels of cathepsin B and H [44].

The papilloma cell lines PDV and PDV-C57 show nearly identical patterns of labeling (Figure 5A). However, these profiles are dramatically different than the profile observed for C5N and P6 lysates. A predominant 30 kDa polypeptide (cathepsin B; see below) is labeled along with a less intensely labeled 21 kDa polypeptide. The squamous cell carcinoma cell lines B9, A5 and D3 result in similar profiles of modified polypeptides. While all three lines are nearly identical cancer cell types, only B9 shows appreciable labeling of the major 30 kDa and 21 kDa polypeptides. Similarly, the two highly invasive spindle cell carcinomas Car B and Car C show similar, but not identical, labeling profiles. The 21 kDa species, in particular, shows differential labeling in the two cell types. These findings illustrate that cells isolated from different tumor sources have different protease activities. This signature of protease activity may in fact be unique to each cell and/or tumor, much the same way genomics studies by Browne and colleagues have shown that individual tumor cells have unique global gene expression profiles [45].

The cathepsin B-specific label ^{125}I-MB-074 was used to directly examine the profile of cathepsin B activity in the same collection of cells described (Figure 5B). This probe has been found to label cathepsin B in a highly specific manner [23]. Labeling of cathepsin B dramatically changed across the profile of cell types with the greatest activity observed for the PDV and PDV-C57 lines. Furthermore, the apparent molecular weight, as well as the sharpness of the cathepsin B, band differed for the benign and spindle cell carcinomas, suggesting that this enzyme is modified differently in these cell types. This change in migration for cathepsin B may be due to changes in glycosylation or other post-translational modifications. Cathepsin B has been found to exist as different isoforms with differing p*I*s in various tumor cells as a result of changes in glycosylation and trafficking [46]. Changes in the post-translational modification of cathepsin B are likely to affect the localization of active forms of the enzyme and therefore may play an important role in the control of cathepsin B activity in tumors [46]. Overall, the results obtained from labeling with ^{125}I-MB-074 further highlight the variability

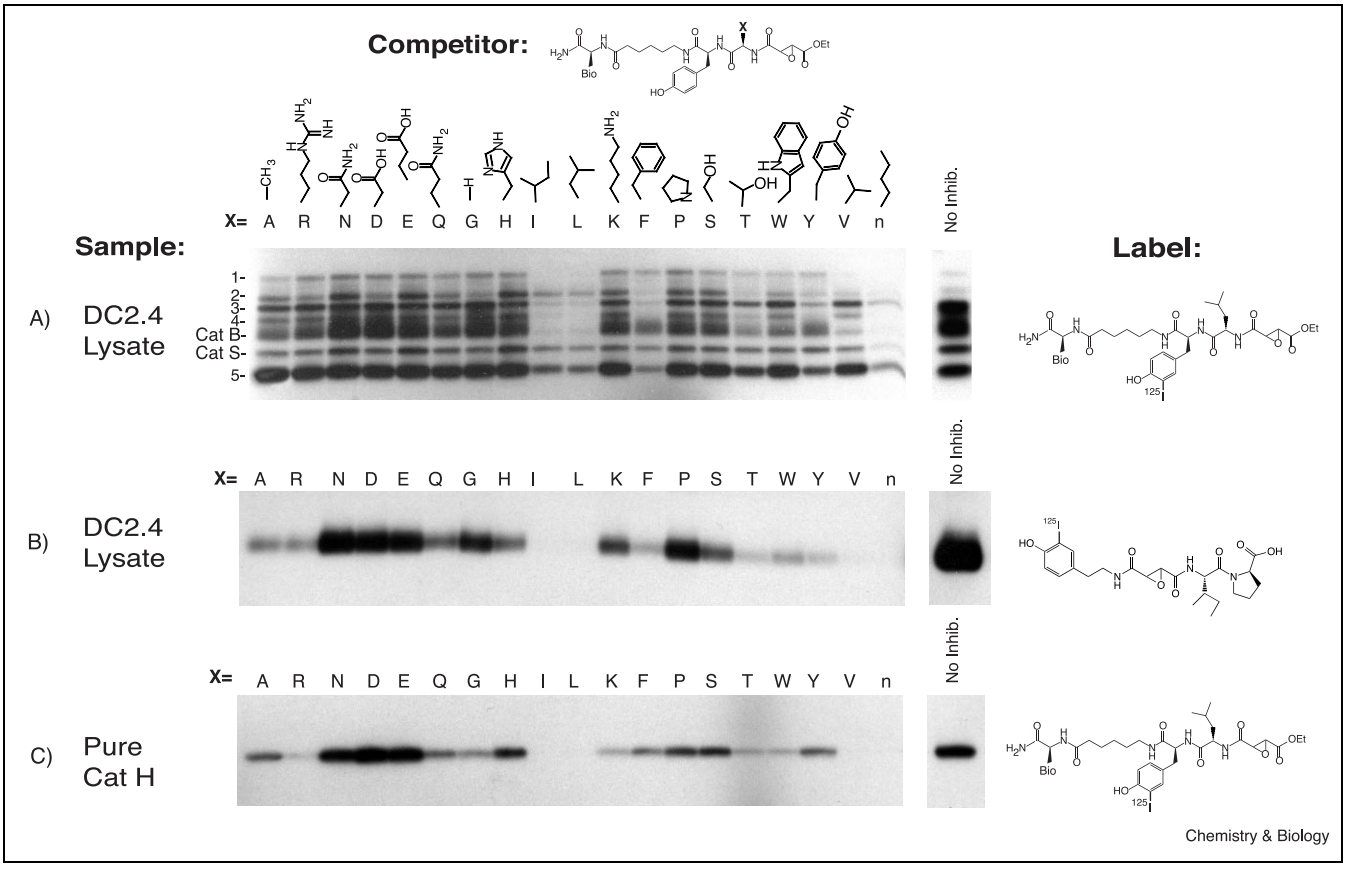

Figure 6. Profiling protease inhibitor specificity. Lysates from the dendritic cell line DC2.4 **(A,B)** or purified cathepsin H **(C)** were preincubated with 50 μM of each of the 19 derivatives of DCG-04 and then labeled with ^{125}I-DCG-04 **(A,C)** or ^{125}I-MB-074 **(B)** as indicated. The general structures of the inhibitors are shown with the variable amino acid side chain indicated as an X (competitor; top). The predominant labeled polypeptides in **(A)** are labeled with numbers, and positions of cathepsins B and S are indicated for reference.

of cathepsin B activity found in different types of tumor cells as well as in nearly identical cell lines derived from different sources.

Profiling protease specificity using a library of inhibitors

To take advantage of the flexibility and ease of synthesis of the DCG-04 family of compounds, we created a small library of compounds in which the peptide recognition portion of the molecule is modified. It has been proposed that the main specificity regions within the active binding site of the cathepsins are S2, S1, S1′ and S2′, with S2 containing the main binding pocket [47]. Since the leucine residue of E-64 binds in the critical S2 pocket of many proteases [33], changes to this residue are likely to have the greatest effect on specificity of our inhibitor for a given target. A complete scanning library consisting of 18 natural amino acids and the isosteric methionine analog norleucine was constructed. This library of inhibitors was used to create profiles of inhibitor specificity for proteases targeted by DCG-04 and MB-074 (Figure 6).

Competition analysis was used to determine the potency of each member of the P2 scanning library towards multiple protease targets. Lysates from DC2.4 cells were preincubated with 50 μM of each of the 19 DCG library members and residual activity measured for multiple proteases using ^{125}I-DCG-04 (Figure 6A). In general, residues containing non-charged aliphatic side chains [isoleucine (I), leucine (L; DCG-04) and norleucine (n)] show the highest activity and the lowest amount of specificity across the profile of polypeptides. More interesting was the apparent selectivity of several DCG family compounds for a subset of labeled polypeptides. For example, the valine containing compound competed for polypeptides 1, 2 and cathepsin B but had little effect on the remaining species. In contrast, both the phenylalanine and tyrosine containing compounds showed specificity for polypeptides 2, 3, 4 and 5. Furthermore, while the aspartic acid and glycine containing compounds showed relatively poor activity overall, they showed some degree of specificity against polypeptide 2. Using these data to simultaneously score inhibitors for potency and selectivity will be valuable for the development of specific inhibitors.

Similar competition experiments were performed with the library of DCG analogs to obtain profiles of single proteases. DC2.4 lysates were preincubated with P2 library and then labeled with the cathepsin B-specific compound ^{125}I-MB-074 (Figure 6B). This method allowed analysis of cathepsin B specificity in crude extracts. As found in the ^{125}I-DCG-04 labeling (Figure 6B), isoleucine, leucine, valine and norleucine analogs showed the highest activity followed by the aromatic amino acids (W, Y, F) containing compounds. In order to explore specificity profiles for additional cysteine proteases that could not be specifically labeled in crude extracts, we performed the same competition labeling experiment described above using a purified enzyme. Preincubation of purified cathepsin H with the library of compounds followed by ^{125}I-DCG-04 labeling resulted in a specificity profile that was remarkably similar to the profile observed for cathepsin B in crude extracts (Figure 6C). While these two proteases are quite different in their biological functions, it is clear from these data that the two have similar inhibitor specificity in the S2 pocket.

Since it is unlikely that two distinct proteases will exhibit identical reactivity across a diverse set of inhibitors, it may be possible to use this information from inhibitor libraries to generate 'specificity fingerprints' for a series of well-characterized proteases. Establishment of a database of protease inhibitor profiles could potentially be used to establish target identification by labeling of crude protein mixtures in the presence of compound libraries. Furthermore, extension of this methodology to longer, more diverse peptide substrate analogs may further accentuate the specificity differences of closely related protease species.

Profiling across tissue types

Having determined that both DCG-03 and DCG-04 were capable of covalently modifying multiple papain family

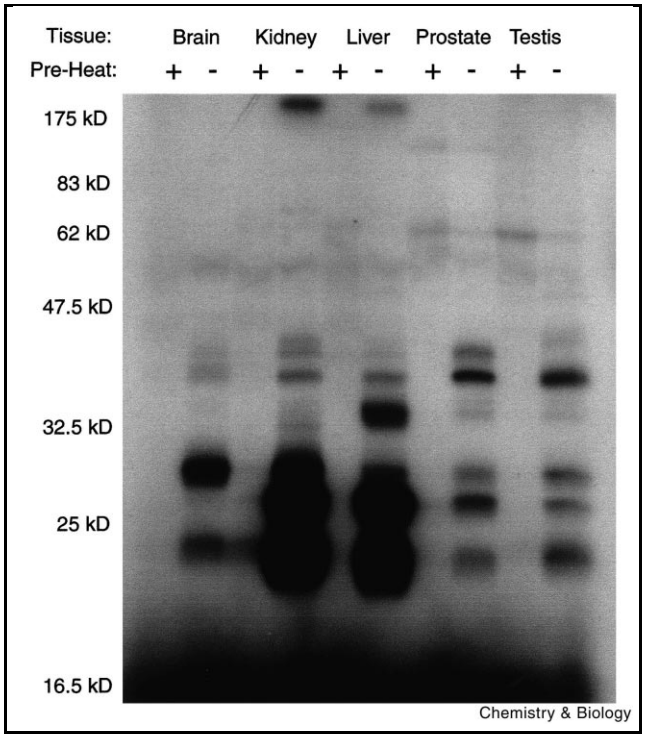

Figure 7. Activity profiling of cysteine proteases across tissue types. Labeling of total cellular extracts (100 μg protein/lane) from rat brain, kidney, liver, prostate and testis with ^{125}I-DCG-04 at pH 5.5. Samples were analyzed by SDS–PAGE followed by autoradiography. A pre-heating control was included for each tissue type to indicated background labeling.

proteases in extracts generated from several cell lines, we wanted to test the utility of these reagents for profiling protease activity patterns in various tissues. In this way, a crude map of protease activities can be created for each tissue and ultimately the identity of these major species can be determined by virtue of their reactivity towards the DCG-04 affinity probe.

Samples of rat brain, kidney, liver, prostate and testis tissue were used to make crude homogenates at the reduced pH of the lysosome (pH 5.5). Samples were labeled with ^{125}I-DCG-04 and analyzed by SDS–PAGE/autoradiography (Figure 7). The most intense labeling in the 20–30 kDa size range was observed for kidney and liver tissue, consistent with the known protein processing functions of these organs. Comparison of the labeling profiles across tissue samples indicated that while some of the modified polypeptides were observed in multiple tissues at nearly identical intensities, several polypeptides showed increased or specific activity in a given tissue type. These data are consistent with the findings that cathepsin expression patterns and activities are differentially regulated across tissue types [48]. In addition, the major species labeled by ^{125}I-DCG-04 were in the 20–30 kDa size range and are likely to be lysosomal cathepsins such cathepsins B, H and L. To confirm this hypothesis, we chose rat kidney as a starting material for the affinity purification of targeted cysteine protease using DCG-04 as an affinity tag. The results of this purification are described below.

Identification of DCG-04-modified proteins in rat kidney by affinity chromatography

Perhaps the greatest attribute of a functional proteomics tool is its ability to aid in the identification of targeted proteins. As shown above, rat kidney contains several polypeptides that were efficiently targeted by DCG-04 (Figure 7). Three prominently labeled species of 23 kDa, 28 kDa and 30 kDa were identified in total kidney extract (Figure 8A). When subjected to anion exchange chromatography, these polypeptides partitioned over a wide range of the elution gradient as determined by DCG-04 labeling of column fractions (Figure 8B). Two pools of fractions were

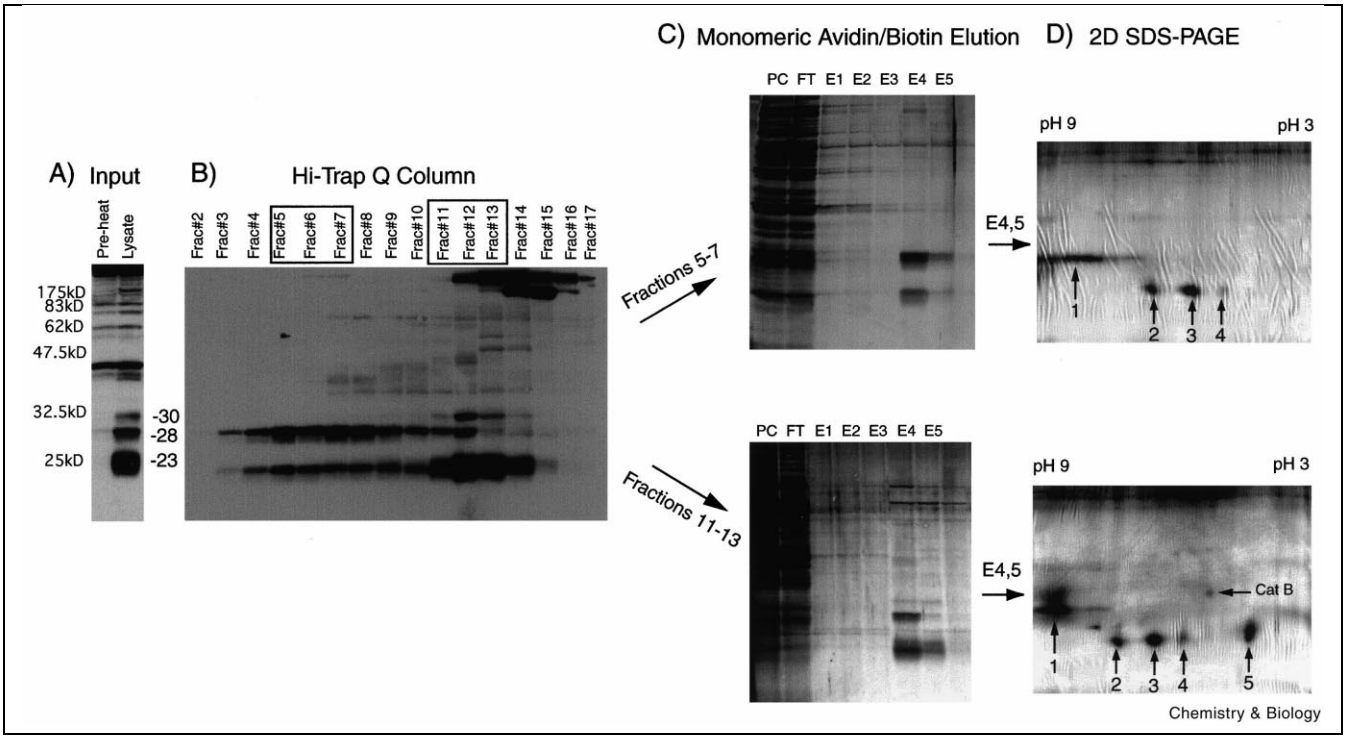

Figure 8. Affinity purification of DCG-04-targeted proteases from rat kidney. **(A)** Labeling of total cellular extracts (100 μg protein/lane) from rat kidney with 50 μM DCG-04 at pH 5.5. Samples were analyzed by SDS–PAGE followed by affinity blot. **(B)** Anion exchange chromatography of rat kidney lysate using a gradient from 0.05 to 1 M NaCl, pH 9.0. Fractions were analyzed by addition of DCG-04 (50 μM) followed by SDS–PAGE and affinity blotting. Fractions containing DCG-04-labeled proteins were pooled (fractions 5–7 and fractions 11–13). **(C)** Pooled fractions were labeled with DCG-04 (50 μM), and DCG-04-modified proteins bound to a monomeric avidin column, washed with 1 M NaCl, and eluted using 2 mM biotin. A sample of material from pools prior to application to the affinity column (PC) along with column flow through (FT) and biotin elution fractions (E1–E5) were analyzed by SDS–PAGE followed by silver staining. **(D)** Elutions containing labeled proteins were pooled, volumes reduced and analyzed by 2D IEF electrophoresis followed by silver staining. Spots labeled with numbers were excised and used for sequencing.

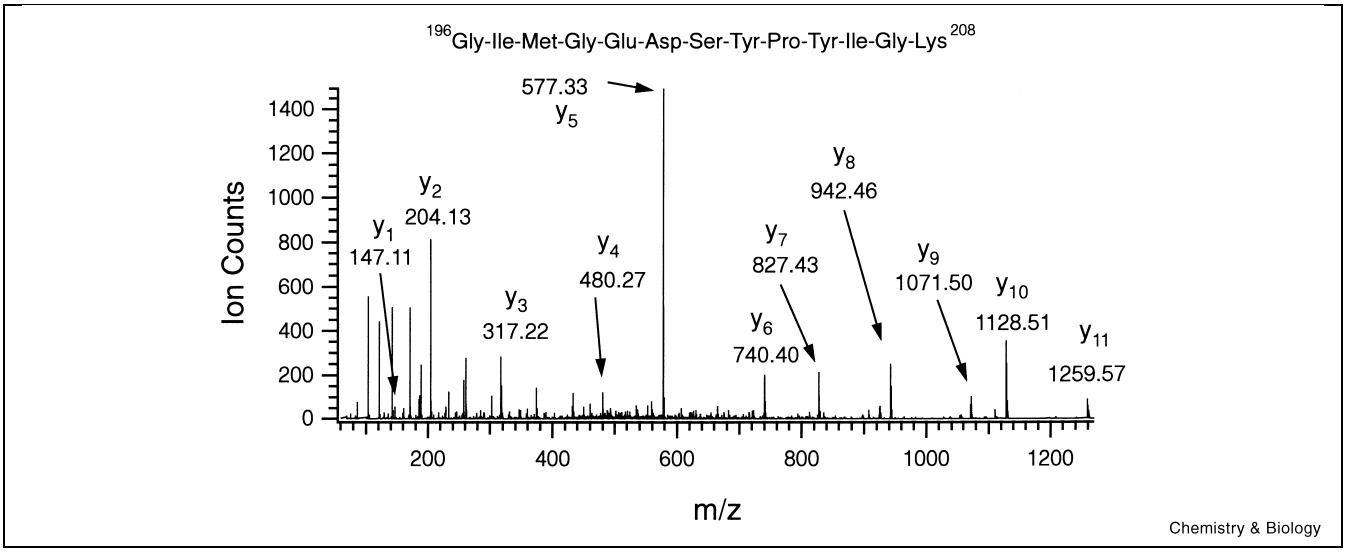

Figure 9. Low energy CID spectrum of tryptic peptides with $MH^+ = 1429.7$. The doubly charged ion at *m/z* 715.35 was selected as a precursor ion. Only the C-terminal fragment ions used for sequence determination are labeled.

chosen based on differences in labeled protein composition. Fractions 7–9 contained predominantly the 23 and 28 kDa species, and fractions 11–13 contained the 23, 28 and 30 kDa species. Modified proteins were affinity-purified using a monomeric avidin column that has a reduced binding affinity for biotin and thus the bound proteins could be competitively eluted with high concentrations of biotin (2 mM). The affinity column purified all DCG-04-modified polypeptides in both pools as visualized by SDS–PAGE and silver staining of eluted fractions (Figure 8C). To further resolve DCG-04-modified polypeptides, peak fractions were concentrated, separated by 2D SDS–PAGE and visualized by silver staining (Figure 8D).

The 30 kDa polypeptide (cat B) yielded a single spot near the acidic end of the gel, while the 28 kDa polypeptide (spot #1) resolved into a streak near the basic end of the gel. The 23 kDa band yielded three distinct spots ranging in p*I* from acidic to basic (spots #2–5). All spots were excised from the gel and subjected to in-gel trypsin digestion, followed by peptide extraction and analysis by mass spectrometry. The protein amount in the 30 kDa spot was not sufficient for unambiguous identification based on MS data alone. Thus its identity was confirmed as cathepsin B by labeling of anion exchange column fractions with the cathepsin B-specific label ^{125}I-MB-074 [23] (data not shown).

The tryptic mass fingerprint obtained for the 28 kDa band as well as two of the three 23 kDa spots (#2, #3) indicated the presence of cathepsin H. Furthermore, all three digests contained a MH^+ 1429.7 peptide that was sequenced by low energy dissociation analysis (CID; Figure 9). The resulting sequence, MGEDSYPYL/IGK, unequivocally matched cathepsin H. The amino-terminus of cathepsin H is heterogeneous, explaining the presence of multiple cathepsin H isoforms at similar molecular weights [49]. In addition, cathepsin H exists as both single chain and two-chain isoforms differing by about 5 kDa [49]. Thus, spot #1 is likely to be the single chain form of cat H while spots #2 and #3 may represent heavy chain versions of the two-chain isoform.

The remaining 23 kDa spots (#4, #5) did not yield sequence data; however, spot #5 was identified as cathepsin L based on the tryptic peptides observed in its digest, its size and p*I*. Thus, DCG-04 successfully identified the predominant active cysteine proteases in rat kidney as cathepsins B, H and L in agreement with previous studies [48].

Significance

Functional proteomics methods are becoming more important as genomics efforts complete the sequences of various organisms. Cravatt and co-workers have established the utility of a functional proteomics tool specific for the serine hydrolase family of proteases [32]. We show here that a general affinity label, DCG-04, and its radiolabeled counterpart, ^{125}I-DCG-04, can be used to profile cysteine protease activities in crude extracts from cells and tissues, as well as throughout multiple stages of a physiological process. Diversification of the peptide portion of the inhibitor using solid-phase synthesis established the utility of small libraries of compounds for determining profiles of inhibitor specificity for both characterized and potentially novel enzymes. The information obtained from these libraries provides a starting point for the development of protease-spe-

cific inhibitors and also provides functional information about a protease target that may serve as a method for rapid identification of targets in crude protein mixtures. Furthermore, DCG-04 can be used as an affinity purification reagent to aid in the identification of proteases selected by virtue of their reactivity towards our electrophilic probes. Target identification of proteases from crude extracts based on activity profiles will assist in the assignment of protein function as well as potentially identify new players in processes such as carcinogenesis. Finally, further diversification of these reagents is likely to extend their utility for the study of additional physiological processes that are regulated by proteolysis.

Materials and methods

Synthesis of DCG-04, DCG-03 and P2 diverse library

Solution-phase synthesis of ethyl (2S,3S)-oxirane-2,3-dicarboxylate. The synthesis of this compound has been reported elsewhere [23].

Solid-phase synthesis of DCG-04 and DCG-03. The details of the solid-phase synthesis are shown in Figure 2. All resins and reagents were purchased from Advanced Chemtech (Louisville, KY, USA). Dry Fmoc-Rink amide resin (0.7 mmol/g) was weighed into 1×10 cm columns (Waters). The columns were fitted with Teflon stopcocks and connected to a 20 port vacuum manifold (Waters) that was used to drain solvents and reagents from the columns. The resin was swelled using DMF. The Fmoc protecting group was removed (deprotected) by treatment with a 20% piperidine solution in DMF for 15 min. The resin was washed with 3×3 ml of DMF and 3×3 ml of CH_2Cl_2.

Fmoc-Lys(biotin)-OH (100 mg, 70 μmol, 1 eq), DIC (11.4 μl, 112 μmol, 1.5 eq), HOBT (15.1 mg, 112 μmol, 1.5 eq) were dissolved in 2 ml of DMF, added to the resin and the reaction was agitated for 1 h. The resin was washed, and the N-terminal Fmoc group was deprotected. Fmoc-6-aminohexanoic acid (74.2 mg, 210 μmol, 3 eq), DIC (21.4 μl, 210 μmol, 3 eq) and HOBT (28.4 mg, 210 μmol, 3 eq) were dissolved in 2 ml DMF, and agitated with the resin for 1 h, followed by washing and deprotection of the N-terminal Fmoc group (synthesis of DCG-03 leaves this step out). Fmoc-Tyr(But)-OH (160.8 mg, 350 μmol, 5 eq), DIC (35.6 μl, 350 μmol, 5 eq) and HOBT (47.2 mg, 350 μmol, 5 eq) were dissolved in 2 ml DMF, and the reaction agitated for 1 h followed by washing and N-terminal Fmoc group deprotection. Fmoc-leucine (61.8 mg, 350 μmol, 5 eq), DIC (35.6 μl, 350 μmol, 5 eq) and HOBT (47.2 mg, 350 μmol, 5 eq) were dissolved in 2 ml DMF, and the reaction agitated for 1 h. The resin was washed followed by deprotection of the N-terminal Fmoc group. Ethyl (2*S*,3*S*)-oxirane-2,3-dicarboxylate (22.4 mg, 140 μmol, 2 eq), DIC (14.2 μl, 140 μmol, 2 eq) and HOBT (18.9 mg, 140 μmol, 2 eq) were dissolved in 2 ml DMF, and the reaction agitated for 1 h. The resin was washed with 3×3 ml of DMF and 3×3 ml of CH_2Cl_2.

The inhibitors were cleaved from the resin using 1 ml of cleavage cocktail (95% TFA, 2.5% water, 2.5% triisopropylsilane). The mix was collected, and the resin washed with 0.5 ml of fresh cleavage cocktail. Ice cold ether (15 ml) was used to precipitate the product. The solid was collected and dissolved in a minimal amount of DMSO. The product was purified on a C18 reverse phase high performance liquid chromatography (HPLC) column (Waters, Delta-Pak) 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. Electrospray mass spectrum: [M+H] calculated for DCG-03 $C_{37}H_{55}N_7O_{10}S$ 791.0, found 791.0; calculated for DCG-04 $C_{43}H_{66}N_8O_{11}S$ 903.1, found 903.7.

A similar protocol was used to synthesize the P2 diverse library except that synthesis was performed using a 96 well manifold (Robbins Scientific). Synthesis was carried out on 20 mg of Rink resin per well, and all coupling conditions were identical to those described above. Each of 18 natural amino acids (except cysteine and methionine) and including norleucine were coupled after addition of the amino hexanoic acid spacer group. All subsequent steps were performed as described above except peptides were used without HPLC purification due to the fact that products were found to be pure by HPLC analysis. Identity of products was confirmed by mass spectrometry. Electrospray mass spectrum: X = Ala calculated [M+H] for $C_{40}H_{60}N_8O_{11}S$ 862.0, found 861.9; Arg $C_{42}H_{66}N_{12}O_{11}S$ 946.5, found 946.7; Asn $C_{41}H_{61}N_9O_{12}S$ 905.0, found 904.9; Asp $C_{41}H_{60}N_8O_{13}S$ 906.0, found 905.9; Glu $C_{42}H_{62}N_8O_{13}S$ 920.0, found 919.8; Gln $C_{42}H_{63}N_9O_{12}S$ 919.0, found 918.9; Gly $C_{39}H_{58}N_8O_{11}S$ 848.0, found 847.7; His $C_{43}H_{62}N_{10}O_{11}S$ 928.0, found 927.7; Ile $C_{43}H_{66}N_8O_{11}S$ 904.1, found 904.0; Leu $C_{43}H_{66}N_8O_{11}S$ 904.1, found 904.0; Lys $C_{43}H_{67}N_9O_{11}S$ 919.0, found 919.0; $C_{46}H_{64}N_8O_{11}S$ 938.0, found 937.8; Pro $C_{42}H_{62}N_8O_{11}S$ 888.0, found 877.8; Ser $C_{40}H_{60}N_8O_{12}S$ 878.0, found 877.8; Thr $C_{41}H_{62}N_8O_{12}S$ 892.0, found 892.0; Trp $C_{48}H_{65}N_9O_{11}S$ 977.1, found 976.7; Tyr $C_{46}H_{64}N_8O_{12}S$ 954.1, found 953.8; Val $C_{42}H_{64}N_8O_{11}S$ 890.0, found 890.0; Nle $C_{43}H_{66}N_8O_{11}S$ 904.1, found 903.9.

Radiolabeling of inhibitors

All compounds were iodinated and isolated using the previously reported protocol [23].

Preparation of cell and tissue lysates

Tissues were Dounce-homogenized in buffer A (50 mM Tris pH 5.5, 1 mM dithiothreitol (DTT), 5 mM $MgCl_2$, 250 mM sucrose) and extracts centrifuged at $1100\times g$ for 10 min at 4°C. The resulting supernatant was centrifuged at $22\,000\times g$ for 30 min at 4°C and final supernatants used for all labeling experiments. Cells were lysed using glass beads (<104 μm) in buffer A and supernatants centrifuged at $15\,000\times g$ 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 with ^{125}I-DCG-04, ^{125}I-DCG-03 and ^{125}I-MB-074

Equivalent amounts of radioactive inhibitor stock solutions (approximately 10^6 cpm per sample) were used for all labeling experiments. Samples of lysates (100 μg total protein in 100 μl buffer; 50 mM Tris pH 5.5, 5 mM $MgCl_2$, 2 mM DTT) were labeled for 1 h at 25°C unless noted otherwise. Samples were quenched by dilution of $4\times$ SDS sample buffer to $1\times$ (for one-dimensional (1D) SDS–PAGE) or by dissolving urea to a final concentration of 9.5 M (for 2D SDS–PAGE).

Gel electrophoresis

1D SDS–PAGE, 2D IEF gels were performed as described [22].

SDS–PAGE–Western blotting detection of and autoradiography of DCG-04-modified proteins

Quenched DCG-04-labeled samples were separated by SDS–PAGE (100 μg/lane) and transferred to nitrocellulose using a semi-dry apparatus. Membranes were blocked using phosphate-buffered saline (PBS) and 5% (w/v) dry milk for 30 min at 25°C. Blots were washed briefly with PBS/0.2% Tween (PBS–Tween) and treated with avidin-horseradish peroxidase conjugate (VectaStain) in PBS–Tween for 30 min at 25°C. Blots were washed three times with PBS–Tween, treated with ECL reagents (Amersham) and exposed to film.

Competition labeling experiments

Lysates from the dendritic cell line DC2.4 were prepared at pH 5.5 as described above. Purified cathepsin H was purchased from Calbiochem (San Diego, CA, USA). Samples of lysates (100 μg total protein in 100

μl buffer B; 50 mM Tris pH 5.5, 5 mM $MgCl_2$, 2 mM DTT) or purified cathepsin H (1 μg protein in 100 μl buffer A) were preincubated with 50 μM of each library member (diluted from 5 mM DMSO stocks) for 2 h at room temperature. Samples were then labeled by addition of either ^{125}I-DCG-04 or ^{125}I-MB-074 to each sample followed by further incubation at room temperature for 1 h. Samples were quenched by the addition of 4× sample buffer to 1× followed by boiling for 5 min. Samples were analyzed by SDS–PAGE followed by autoradiography.

Preparation of mouse carcinoma cell lines

Mouse melanoma cell lines were prepared by a single topical application of 25 μg of the chemical mutagen dimethylbenzanthracene to the skin of mice followed by biweekly application of 100 μM of the tumor promoter, TPA, over an extended period of time essentially as described [50–52].

Protein identification of DCG-04-modified proteins

A soluble fraction of rat kidney lysate (80 mg total protein) was diluted into anion exchange starting buffer (50 mM Tris, 50 mM NaCl, pH 9.0). The lysate was applied to a HitrapQ anion exchange column (Amersham Pharmacia Biotech) and eluted using a linear gradient of 0.05–1 M NaCl, pH 9. An aliquot from each fraction (50 μl) was incubated with 50 μM DCG-04 at 25°C for 1 h and analyzed on a 12.5% SDS–PAGE gel followed by affinity blotting as described above.

The fractions containing peak labeling of the 25–30 kDa bands were pooled, and DCG-04 was added to a final concentration of 50 μM. Pools were incubated at 25°C for 2 h and then 12 h at 4°C. Unbound inhibitor was removed and buffer was exchanged with PBS using a PD-10 column (Pharmacia). Samples were applied to a monomeric avidin column (1 ml bed volume; Pierce), and the column was washed with 6×1 ml fractions of 1 M NaCl. Bound proteins were eluted with 0.5 ml fractions of 2 mM biotin/100 mM NH_4HCO_3 buffer. All wash and eluent fractions were analyzed by SDS–PAGE and silver staining. The fractions containing the labeled 25–30 kDa bands were pooled, the volume reduced by lyophilization and solid urea added to 9.5 M along with BME to 5%, NP-40 to 2%, pH 5–7 ampholytes to 1.6% and pH 3.5–10 ampholytes to 0.4%. Samples were applied to IEF tube gels and electrophoresed at 1000 V for 13 h followed by separation in the second dimension on 12.5% SDS–PAGE gels.

The resulting gels were fixed in 12% acetic acid/50% methanol stained with silver according to reported protocols [22]. Spots were excised, digested with trypsin and the peptide molecular weight measurements were carried out by matrix-assisted laser desorption ionization-mass spectrometry (PE Voyager DESTR). Sequence determination was performed on a quadrupole time-of-flight hybrid tandem mass spectrometer (PE QSTAR) equipped with a Protome nanospray source. This instrument affords high resolution and accuracy for mass measurement and the CID data obtained allowed unambiguous sequence determination. Database searches were performed using the Protein Prospector software package (http://prospector.ucsf.edu/).

Acknowledgements

We thank A. Balmain and B. Hahn for supplying cell lines for the tumor cell profiling experiments. We thank B. Cravatt for supplying rat tissues and for critical evaluation of the manuscript. We thank K. Williams for assistance with in-gel digestion protocols and S. Fisher for advice on sample preparation. We thank Wendell Lim for critical evaluation of the manuscript. This work was supported by funding from the Sandler Foundation (D.G. and M.B.). QSTAR shared instrumentation Grant NIH NCRR BRTP RR01614.

References

1. Schena, M., Heller, R.A., Theriault, T.P., Konrad, K., Lachenmeier, E. & Davis, R.W. (1998). Microarrays: biotechnology's discovery platform for functional genomics. *Trends Biotechnol.* **16**, 301–306.
2. DeRisi, J.L. & Iyer, V.R. (1999). Genomics and array technology. *Curr. Opin. Oncol.* **11**, 76–79.
3. Jungblut, P.R., et al., & Stöffler, G. (1999). Proteomics in human disease: cancer, heart and infectious diseases. *Electrophoresis* **20**, 2100–2110.
4. Celis, J.E., Ostergaard, M., Jensen, N.A., Gromova, I., Rasmussen, H.H. & Gromov, P. (1998). Human and mouse proteomic databases: novel resources in the protein universe. *FEBS Lett.* **430**, 64–72.
5. Gygi, S.P., Rochon, Y., Franza, B.R. & Aebersold, R. (1999). Correlation between protein and mRNA abundance in yeast. *Mol. Cell. Biol.* **19**, 1720–1730.
6. Cygler, M., Sivaraman, J., Grochulski, P., Coulombe, R., Storer, A.C. & Mort, J.S. (1996). Structure of rat procathepsin B: model for inhibition of cysteine protease activity by the proregion. *Structure* **4**, 405–416.
7. Coulombe, R., Grochulski, P., Sivaraman, J., Ménard, R., Mort, J.S. & Cygler, M. (1996). Structure of human procathepsin L reveals the molecular basis of inhibition by the prosegment. *EMBO J.* **15**, 5492–5503.
8. Villadangos, J.A., et al., & Ploegh, H.L. (1999). Proteases involved in MHC class II antigen presentation. *Immunol. Rev.* **172**, 109–120.
9. Gelb, B.D., Shi, G.P., Chapman, H.A. & Desnick, R.J. (1996). Pycnodysostosis, a lysosomal disease caused by cathepsin K deficiency. *Science* **273**, 1236–1238.
10. Beinfeld, M.C. (1998). Prohormone and proneuropeptide processing. Recent progress and future challenges. *Endocrine* **8**, 1–5.
11. Iwata, Y., Mort, J.S., Tateishi, H. & Lee, E.R. (1997). Macrophage cathepsin L, a factor in the erosion of subchondral bone in rheumatoid arthritis. *Arthritis Rheum.* **40**, 499–509.
12. Yan, S., Sameni, M. & Sloane, B.F. (1998). Cathepsin B and human tumor progression. *Biol. Chem.* **379**, 113–123.
13. Golde, T.E., Estus, S., Younkin, L.H., Selkoe, D.J. & Younkin, S.G. (1992). Processing of the amyloid protein precursor to potentially amyloidogenic derivatives (see comments). *Science* **255**, 728–730.
14. Munger, J.S., et al., & Chapman, H.A. (1995). Lysosomal processing of amyloid precursor protein to Aβ peptides: a distinct role for cathepsin S. *Biochem. J.* **311**, 299–305.
15. Shaw, E. (1994). Peptidyl diazomethanes as inhibitors of cysteine and serine proteinases. *Methods Enzymol.* **244**, 649–656.
16. Shaw, E., Angliker, H., Rauber, P., Walker, B. & Wikstrom, P. (1986). Peptidyl fluoromethyl ketones as thiol protease inhibitors. *Biomed. Biochim. Acta* **45**, 1397–1403.
17. Pliura, D.H., Bonaventura, B.J., Smith, R.A., Coles, P.J. & Krantz, A. (1992). Comparative behaviour of calpain and cathepsin B toward peptidyl acyloxymethyl ketones, sulphonium methyl ketones and other potential inhibitors of cysteine proteinases. *Biochem. J.* **288**, 759–762.
18. Brömme, D., et al., & Demuth, H.U. (1989). Potent and selective inactivation of cysteine proteinases with *N*-peptidyl-*O*-acyl hydroxylamines. *Biochem. J.* **263**, 861–866.
19. Palmer, J.T., Rasnick, D., Klaus, J.L. & Brömme, D. (1995). Vinyl sulfones as mechanism-based cysteine protease inhibitors. *J. Med. Chem.* **38**, 3193–3196.
20. Barrett, A.J., et al., & Hanada, K. (1982). L-*trans*-Epoxysuccinyl-leucylamido(4-guanidino)butane (E-64) and its analogues as inhibitors of cysteine proteinases including cathepsins B, H and L. *Biochem. J.* **201**, 189–198.
21. Rauber, P., Wikstrom, P. & Shaw, E. (1988). Iodination of peptidyl chloromethyl ketones for protease affinity labels. *Anal. Biochem.* **168**, 259–264.
22. Bogyo, M., Shin, S., McMaster, J.S. & Ploegh, H.L. (1998). Substrate binding and sequence preference of the proteasome revealed by active-site-directed affinity probes. *Chem. Biol.* **5**, 307–320.
23. Bogyo, M., Verhelst, S., Bellingard-Dubouchaud, V., Toba, S. & Greenbaum, D. (2000). Selective targeting of lysosomal cysteine proteases with radiolabeled electrophilic substrate analogs. *Chem. Biol.* **7**, 27–38.
24. Mason, R.W., Wilcox, D., Wilkstrom, P. & Shaw, E.N. (1989). The identification of active forms of cysteine proteinases in Kirsten-virus-transformed mouse fibroblasts by use of a specific radiolabeled inhibitor. *Biochem. J.* **257**, 125–129.
25. Mason, R.W., Bartholomew, L.T. & Hardwick, B.S. (1989). The use of benzyloxycarbonyl[^{125}I]iodotyrosylalanyldiazomethane as a probe

for active cysteine proteinases in human tissues. *Biochem. J.* **263**, 945–949.
26. Bogyo, M., McMaster, J.S., Gaczynska, M., Tortorella, D., Goldberg, A.L. & Ploegh, H. (1997). Covalent modification of the active site threonine of proteasomal β subunits and the *Escherichia coli* homolog HslV by a new class of inhibitors. *Proc. Natl. Acad. Sci. USA* **94**, 6629–6634.
27. Meng, L., Mohan, R., Kwok, B.H., Elofsson, M., Sin, N. & Crews, C.M. (1999). Epoxomicin, a potent and selective proteasome inhibitor, exhibits in vivo anti-inflammatory activity. *Proc. Natl. Acad. Sci. USA* **96**, 10403–10408.
28. Faleiro, L., Kobayashi, R., Fearnhead, H. & Lazebnik, Y. (1997). Multiple species of CPP32 and Mch2 are the major active caspases present in apoptotic cells. *EMBO J.* **16**, 2271–2281.
29. Nicholson, D.W., et al., & Lazebnik, Y.A. et al. (1995). Identification and inhibition of the ICE/CED-3 protease necessary for mammalian apoptosis. *Nature* **376**, 37–43.
30. Griffith, E.C., et al., & Liu, J.O. (1997). Methionine aminopeptidase (type 2) is the common target for angiogenesis inhibitors AGM-1470 and ovalicin. *Chem. Biol.* **4**, 461–471.
31. Sin, N., Meng, L., Wang, M.Q., Wen, J.J., Bornmann, W.G. & Crews, C.M. (1997). The anti-angiogenic agent fumagillin covalently binds and inhibits the methionine aminopeptidase, MetAP-2. *Proc. Natl. Acad. Sci. USA* **94**, 6099–6103.
32. Liu, Y., Patricelli, M. & Cravatt, B. (1999). Activity-based protein profiling: The serine hydrolases. *Proc. Natl. Acad. Sci. USA* **96**, 14694–14699.
33. Matsumoto, K., Mizoue, K., Kitamura, K., Tse, W.C., Huber, C.P. & Ishida, T. (1999). Structural basis of inhibition of cysteine proteases by E-64 and its derivatives. *Biopolymers* **51**, 99–107.
34. Meara, J.P. & Rich, D.H. (1996). Mechanistic studies on the inactivation of papain by epoxysuccinyl inhibitors. *J. Med. Chem.* **39**, 3357–3366.
35. Shi, G.-P., Munger, J.S., Meara, J.P., Rich, D.H. & Chapman, H.A. (1992). Molecular cloning and expression of human aveolar macrophage cathepsin S, an elastinolytic cysteine protease. *J. Biol. Chem.* **267**, 7258–7262.
36. Towatari, T., et al., & Katunuma, N. (1991). Novel epoxysuccinyl peptides. A selective inhibitor of cathepsin B, in vivo. *FEBS Lett.* **280**, 311–315.
37. Barrett, A., Rawlings, N. & Woessner, J. (1998). *Handbook of Proteolytic Enzymes*, Academic Press, San Diego, CA.
38. Kemp, C.J., Burns, P.A., Brown, K. & Balmain, A. (1994). Transgenic approaches to the analysis of ras and p53 function in multistage carcinogenesis. *Cold Spring Harbor Symp. Quant. Biol.* **59**, 427–434.
39. Yuspa, S.H., et al., & Weinberg, W.C. (1994). Role of oncogenes and tumor suppressor genes in multistage carcinogenesis. *J. Invest. Dermatol.* **103**, 90S–95S.
40. Baricos, W.H., Zhou, Y., Mason, R.W. & Barrett, A.J. (1988). Human kidney cathepsins B and L. Characterization and potential role in degradation of glomerular basement membrane. *Biochem. J.* **252**, 301–304.
41. Linebaugh, B.E., Sameni, M., Day, N.A., Sloane, B.F. & Keppler, D. (1999). Exocytosis of active cathepsin B enzyme activity at pH 7.0, inhibition and molecular mass. *Eur. J. Biochem.* **264**, 100–109.
42. Mason, R.W., Gal, S. & Gottesman, M.M. (1987). The identification of the major excreted protein (MEP) from a transformed mouse fibroblast cell line as a catalytically active precursor form of cathepsin L. *Biochem. J.* **248**, 449–454.
43. Quintanilla, M., Haddow, S., Jonas, D., Jaffe, D., Bowden, G.T. & Balmain, A. (1991). Comparison of ras activation during epidermal carcinogenesis in vitro and in vivo. *Carcinogenesis* **12**, 1875–1881.
44. Kim, K., Cai, J., Shuja, S., Kuo, T. & Murnane, M.J. (1998). Presence of activated ras correlates with increased cysteine proteinase activities in human colorectal carcinomas. *Int. J. Cancer* **79**, 324–333.
45. Alizadeh, A.A., et al., & Staudt, L.M. (2000). Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling (see comments). *Nature* **403**, 503–511.
46. Moin, K., Cao, L., Day, N.A., Koblinski, J.E. & Sloane, B.F. (1998). Tumor cell membrane cathepsin B. *Biol. Chem.* **379**, 1093–1099.
47. Turk, D., Guncar, G., Podobnik, M. & Turk, B. (1998). Revised definition of substrate binding sites of papain-like cysteine proteases. *Biol. Chem.* **379**, 137–147.
48. Kominami, E., Tsukahara, T., Bando, Y. & Katunuma, N. (1985). Distribution of cathepsins B and H in rat tissues and peripheral blood cells. *J. Biochem.* **98**, 87–93.
49. Ishidoh, K., et al., & Kominami, E. (1998). Multiple processing of procathepsin L to cathepsin L in vivo. *Biochem. Biophys. Res. Commun.* **252**, 202–207.
50. Bremner, R. & Balmain, A. (1990). Genetic changes in skin tumor progression: correlation between presence of a mutant ras gene and loss of heterozygosity on mouse chromosome 7. *Cell* **61**, 407–417.
51. Burns, P.A., Kemp, C.J., Gannon, J.V., Lane, D.P., Bremner, R. & Balmain, A. (1991). Loss of heterozygosity and mutational alterations of the p53 gene in skin tumours of interspecific hybrid mice. *Oncogene* **6**, 2363–2369.
52. Haddow, S., Fowlis, D.J., Parkinson, K., Akhurst, R.J. & Balmain, A. (1991) Loss of growth control by TGF-β occurs at a late stage of mouse skin carcinogenesis and is independent of ras gene activation. *Oncogene* **6**, 1465–1470 [Erratum. *Oncogene* **6**, 2377–2378].