

Global analysis of proteasomal substrate specificity using positional-scanning libraries of covalent inhibitors

Tamim Nazif and Matthew Bogyo*

Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94143

Communicated by Keith R. Yamamoto, University of California, San Francisco, CA, January 18, 2001 (received for review November 20, 2000)

The proteasome is a large protease complex consisting of multiple catalytic subunits that function simultaneously to digest protein substrates. This complexity has made deciphering the role each subunit plays in the generation of specific protein fragments difficult. Positional scanning libraries of peptide vinyl sulfones were generated in which the amino acid located directly at the site of hydrolysis (P1 residue) was held constant and sequences distal to that residue (P2, P3, and P4 positions) were varied across all natural amino acids (except cysteine and methionine). Binding information for each of the individual catalytic subunits was obtained for each library under a variety of different conditions. The resulting specificity profiles indicated that substrate positions distal to P1 are critical for directing substrates to active subunits in the complex. Furthermore, specificity profiles of IFN- γ -regulated subunits closely matched those of their noninducible counterparts, suggesting that subunit swapping may modulate substrate processing by a mechanism that does require a change in the primary sequence specificity of individual catalytic subunits in the complex. Finally, specificity profiles were used to design specific inhibitors of a single active site in the complex. These reagents can be used to further establish the role of each subunit in substrate processing by the proteasome.

active-site labeling | vinyl sulfones | electrophiles

The proteasome is responsible for a wide range of cellular functions ranging from controlled processing of antigens to complete destruction of long-lived and misfolded proteins (1). The proteasome is also one of the most complex catabolic systems, owing to its multiple active sites, multitude of accessory proteins that regulate its function, and the dynamic nature of the assembled protease complex that results from swapping of active subunits upon stimulation by IFN- γ .

The proteasome is a large barrel-shaped protein complex that is made up of α and β subunits. The β subunits contain the active sites where amide bond hydrolysis takes place (2, 3). Of the seven eukaryotic β subunits only three are believed to be enzymatic (1–3). These catalytic subunits create a proteolytic system that cleaves amide bonds adjacent to a variety of amino acids. The primary hydrolytic activities of the proteasome have been categorized based on the amino acid residues found directly adjacent to the scissile amide bond (4). However, several studies have now established the importance of distal amino acid residues in directing a substrate to an active site in the complex. Detailed kinetic studies using fluorogenic substrates (5) and covalent inhibitors have established the importance of both the P3 and P4 residues for subsite binding (6, 7).

IFN- γ -induced exchange of subunits in the active complex originally was shown to result in modulation of both the activity and substrate specificity of the proteasome complex (8–12). Subunit swapping thereby serves to direct processing of a peptide substrate for downstream loading onto MHC class I molecules and presentation to cytotoxic T lymphocytes. Although it is clear that IFN- γ stimulation results in the generation of peptide fragments terminating in a C-terminal aliphatic or hydrophobic amino acid, recent studies suggest that these changes in product

formation may not be directly caused by differences in primary sequence specificities of the inducible catalytic subunits (13).

Several elegant techniques have been developed to determine the substrate specificity of proteolytic enzymes in the hope that this information will aid in the identification of downstream protein substrates (14, 15). Although these techniques have proven useful, they only provide information regarding bulk product formation and therefore are unable to resolve specificities of enzymes such as the proteasome that contain multiple active sites. Analysis of substrate specificity of the proteasome has remained focused on studies of *in vitro* protein digests and catalytic mutants in yeast (16–18). Although these studies have paved the way toward a better understanding of the proteasome's catalytic mechanism, they are limited by the need for purified preparations of enzyme and therefore are unable to make claims as to the direct *in vivo* relevance of the data. They also fail to systematically address the differences in substrate specificity among the individual catalytic subunits.

We have developed a method for the global analysis of the substrate specificities of each of the proteasome's multiple active sites. This method takes advantage of peptide-based covalent inhibitors of the proteasome. These inhibitors bind to the active site of a catalytic β subunit such that they mimic a bona fide protein substrate. Thus, the amino acid residues of the inhibitor interact with the various binding pockets responsible for defining substrate specificity. These enzyme specificity pockets are identified by their position relative to the site of amide bond hydrolysis. Similarly, the amino acids of substrates and inhibitors are designated based on their position relative to each specificity pocket (Fig. 1A).

Libraries of compounds containing a complete set of amino acid residues at the P2, P3, and P4 sites were used to assess binding to each of the active sites. Furthermore, individual subunits could be resolved by SDS/PAGE, allowing analysis of IFN- γ inducible subunits without the need to obtain homogeneous populations of immunoproteasomes. The resulting specificity map was organized and displayed by using mathematical algorithms that clustered the data, allowing rapid analysis of global specificity profiles. Our results suggest that whereas IFN- γ stimulation leads to swapping of active β subunits, these newly incorporated subunits have nearly identical substrate specificities at the P2–P4 sites as the subunits that they are replacing. Furthermore, these data highlight the importance of these P2–P4 subsites for directing substrate processing. Finally, the information from the inhibitor scan provided a roadmap for the design of inhibitors that specifically targeted the Z (β 2; trypsin-like) subunit of the proteasome. This general methodology is likely to

Abbreviations: VS, vinyl sulfone; MCA, methylcoumarin amide.

*To whom reprint requests should be addressed at: Department of Biochemistry and Biophysics, University of California, 513 Parnassus Avenue, San Francisco, CA 94143-0448. E-mail: mbogyo@biochem.ucsf.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

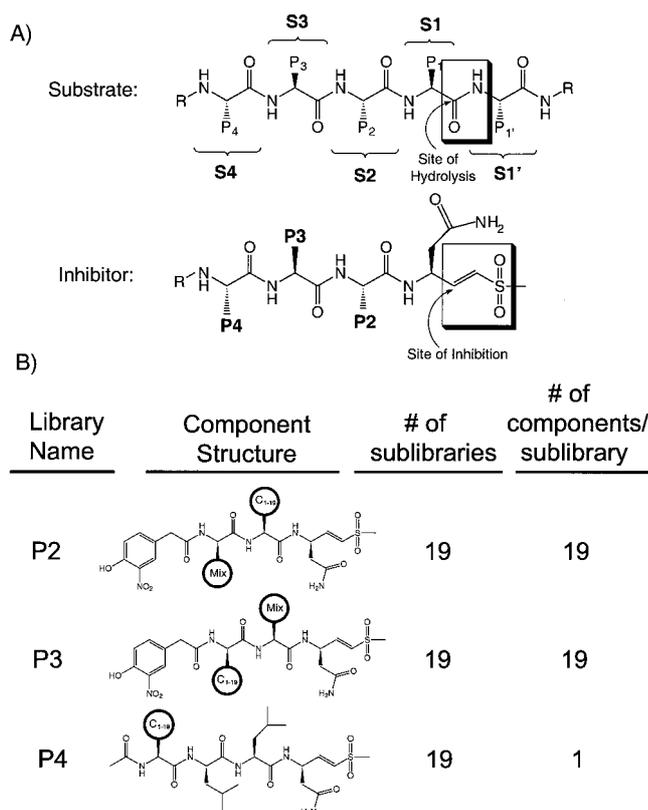


Fig. 1. Inhibitor nomenclature and structure/composition of libraries of P1 aspartyl peptide vinyl sulfones. (A) P notation used to define positions of interaction between inhibitor/substrate and enzyme active site. Sites labeled with S designate binding pockets of the enzyme. (B) Diagram of P2 and P3 positional scanning libraries and the P4 single component library. Constant positions made up of all possible natural amino acids (minus cysteine and using norleucine instead of methionine) are indicated by C₁₋₁₉. Variable positions that contain an isokinetic mixture of amino acids (the same 19 amino acids as constant positions) are indicated as Mix. Note that the P2 and P3 scanning libraries are composed of 19 sublibraries each containing a mixture of 19 members (one Mix position) whereas the P4 library is made up of 19 single compounds (no Mix position).

be of value as a rapid method for optimizing inhibitors against protein targets in crude cellular homogenates.

Experimental Methods

Synthesis of Peptide Vinyl Sulfone. The synthesis of fluorenylmethoxycarbonyl-Asp-VS (IV) was accomplished essentially as described for other amino acid vinyl sulfones (6). Details of the synthesis and Scheme 1 are published as supplemental material on the PNAS web site, www.pnas.org.

Synthesis of Positional Scanning and Single Component Peptide Libraries. Fluorenylmethoxycarbonyl-Asp-VS-loaded resin (20 mg/well 0.7 mmol per g) was weighed into 96 position filter blocks (FlexChem system; Robbins Scientific, Mountain View, CA). Each library consisted of 19 sublibraries in which each of the natural amino acids (minus cysteine and methionine plus norleucine) was used at a designated constant position and an isokinetic mixture of those same 19 amino acids was coupled in the variable position. The isokinetic mixture was created by using a ratio of equivalents of amino acids based on their reported coupling rates (22). The total mixture was adjusted to 10-fold excess total amino acids over resin load. For constant positions a single amino acid was coupled by using 10-fold excess. Couplings were carried out with diisopropylcarbodiimide

and hydroxybenzotriazole under standard conditions. The mixture libraries (P2 and P3) were amino terminally capped with 4-hydroxy-3-nitrophenylacetic acid and the single component P4 libraries were capped by using acetic anhydride. Libraries and single components were cleaved from the resin by addition of 90% trifluoroacetic acid, 5% water, and 5% triisopropyl silane for 2 h. Cleavage solutions were collected and products were precipitated by addition of cold diethyl ether. Solid products were isolated by centrifugation followed by lyophilization to yield the crude peptides that were dissolved in DMSO (50 mM stock) based on average weights for each mixture. Libraries and single compounds were stored at -20°C and further diluted to 5 mM stock plates for use in experiments.

Synthesis of P1 Leucine Peptide Vinyl Sulfones. The peptide vinyl sulfone Ac-YVRLV-VS was synthesized exactly according to the procedures for the synthesis of a related vinyl sulfone reported previously (23).

Radiolabeling of Inhibitors. All compounds were iodinated and isolated as reported (19).

Cell Culture and IFN- γ Stimulation. All cells (EL-4 and NIH 3T3) were maintained in RPMI media with 10% FCS. Cells were stimulated with IFN- γ by addition of 4 μl of a stock solution (50 units/ml) to 50 ml of media. Cells were harvested after 2 days of stimulation.

Preparation of Cell Lysates. Cells were lysed with glass beads (<104 microns; Sigma) in buffer A (50 mM Tris, pH 5.5/1 mM DTT/5 mM MgCl₂/250 mM sucrose) and supernatants were centrifuged for 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).

Direct Labeling of the Proteasome with Iodinated Peptide Vinyl Sulfones. Equivalent amounts of radioactive inhibitor stock solutions (approximately 10⁶ cpm per sample) were used for all labeling experiments. Samples of lysates (100 μg of total protein in 100 μl of buffer; 50 mM Tris, pH 7.4/5 mM MgCl₂/2 mM DTT) were labeled for 1 h at 25°C unless noted otherwise. Samples were quenched by addition of SDS sample buffer.

Library Competitions and Quantitation of Data from Crude Extract and Purified 20S Proteasomes. In a typical experiment NIH 3T3 or EL-4 lysates were prepared as described above and diluted to 1 mg/ml in reaction buffer (RB) (50 mM Tris, pH 7.4/5 mM MgCl₂/2 mM DTT). Lysates (100 μg total protein) or purified 20S proteasomes (1 μg per sample in RB with 0.01% SDS) were incubated with 1 μl from a 5 mM library stock (final concentration of 50 μM) for 30 min at room temperature. Iodinated inhibitors (diluted 1:10 or 1:5 in RB) were added to each reaction and incubation continued for an additional 90 min at room temperature. The reactions were quenched by addition of SDS/PAGE sample buffer followed by boiling for 5 min. All samples were separated on 12.5% SDS/PAGE gels, and data were obtained by exposure of the gels to PhosphorImager screens (Molecular Dynamics). Bands of activity were quantitated with IMAGEQUANT software (Molecular Dynamics), and ratios of each band's intensity relative to the corresponding control untreated sample were obtained.

Analysis of Library Scanning Data Using Genomic Software and Hierarchical Clustering. To prepare data for display and clustering, ratios of signals in inhibitor-treated samples relative to non-treated samples were expressed in decimal format. The ratios were multiplied by 2 and log base 2-transformed to convert the data to values in a range from -1 to 1. Using this approach,

compounds that showed 100% competition were assigned a value of -1 and compounds that showed no inhibition were assigned a value of 1 . The log-transformed data then were converted to color format and clustered by using the program CLUSTER designed by Eisen and coworkers (21). Data then were displayed in a tree diagram after clustering using the program TREE VIEW. Both programs can be obtained at <http://www.microarrays.org>.

Kinetic Analysis of Selectivity of Proteasome Inhibitors. In a typical assay, $0.5 \mu\text{g}$ ($0.3 \mu\text{g}$ for Z-LLE-MCA; MCA indicates methylcoumarin amide) of purified RBC 20S proteasome was diluted to a final volume of $500 \mu\text{l}$ in buffer (20 mM Hepes, pH 8.0/0.5 mM EDTA/0.01% SDS). Fluorogenic substrates were then added to a final concentration of $10 \mu\text{M}$. Progress curves were measured by monitoring aminomethylcoumarin fluorescence at 37°C . The initial rate was measured for 5–10 min. Vinyl sulfone inhibitors were then added from DMSO stocks to the desired concentration and the reaction was allowed to proceed for a total time of 1 h. Substrate hydrolysis was measured independently for each inhibitor at 4–8 different concentrations. Fluorescence data were collected as a function of time and processed by using the SCIENTIST scientific graphing program (MicroMath Scientific Software, Salt Lake City). k_{obs} values were obtained by a nonlinear least-squares fit of the data to the equation for time-dependent or slow-binding inhibition, ($\text{fluorescence} = v_s t + [(v_o - v_s)/k_{\text{obs}}][1 - \exp(-k_{\text{obs}}t)]$), where v_o is the initial velocity that decays over time to a final velocity, v_s , with a rate constant, k_{obs} . $k_{\text{obs}}/[I]$ values are reported as an average of three independent experiments obtained with different inhibitor concentrations.

Results and Discussion

Design and Synthesis of Positional Scanned Libraries of Proteasome Inhibitors. Although originally designed to target cysteine proteases, peptide vinyl sulfones have been demonstrated to act as irreversible covalent inhibitors of the proteasome (6, 19). Because these reagents function by mimicking a peptide substrate, we reasoned that they would provide structure/potency profiles for the proteasome that would directly correlate with substrate specificity. Furthermore, because these reagents covalently bind to the active site threonine of the proteasome, it is possible to trace and resolve binding to individual subunits. We generated libraries of peptide vinyl sulfones in which the P1 amino acid was held constant so that the role of secondary sequence determinants could be assessed. The amino acid asparagine was chosen for the P1 residue because it allowed attachment to a solid support via a side-chain amide bond linkage. Commercially available fluorenylmethoxycarbonyl-protected, *t*-butyl side chain-protected aspartic acid was converted to the corresponding C-terminal vinyl sulfone. Removal of the side chain-protecting group resulted in a free carboxylic acid moiety that was used as a site of attachment to a Rink polystyrene resin. Cleavage from the resin resulted in conversion of this residue to asparagine. This method allowed rapid synthesis of virtually any peptide sequence terminating in a C-terminal asparagine vinyl sulfone.

Positional scanning libraries (20) were constructed in which tripeptides were varied through the P2 and P3 positions (Fig. 1B). The nitrophenol group was included at the N terminus because it has been shown to increase the potency of tripeptide vinyl sulfones for the proteasome (19). Libraries were synthesized by coupling a single amino acid in the constant position (C_{1-19}) and an isokinetic mixture of this same set of amino acids in a variable position (mix). All natural amino acids were used except cysteine due to its reactivity with the vinyl sulfone. Norleucine was used instead of methionine to prevent heterogeneity resulting from oxidation. This method produced sets of 19 sublibraries each containing 19 components. To address the

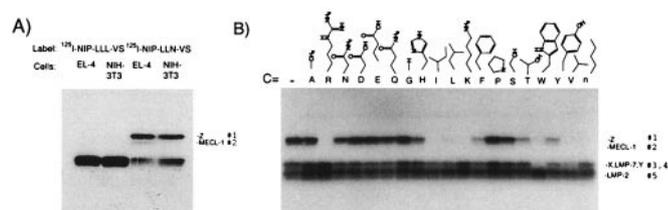


Fig. 2. Application of the P1 asparaginyl peptide vinyl sulfones to the proteasome from crude cellular extracts. (A) Direct labeling of the proteasome's multiple active sites in NIH 3T3 and EL-4 cells using two labels that differ only in the P1 residue. Change from P1 leucine to P1 asparagine leads to an increased labeling of the Z and MECL-1 subunits. (B) A typical competition profile obtained from incubation of crude EL-4 cell extracts with the complete set of P2 scanning sublibraries followed by labeling with ^{125}I -NIP-LLN-VS. The control, nontreated lane is indicated at the far left, and the identity of the constant amino acid for each sublibrary is shown along with its side-chain structure at the top of the gel.

specificity at the P4 position, libraries of acetyl-capped tetrapeptides also were constructed. These libraries consisted of single compounds in which the P2 and P3 residues were held constant as leucine, previously shown to be potent for inhibition of the proteasome (6). The single component methodology was chosen to avoid complications associated with multiple variable positions and sublibraries of increased complexity.

Application of Inhibitor Libraries to the Proteasome. Changing the P1 leucine residue of the previously reported peptide vinyl sulfone NLVS (NIP-LLL-VS) to asparagine resulted in a new covalent inhibitor (NIP-LLN-VS; Fig. 2A), which targets all of the active sites of the proteasome. Libraries of P1 asparagine vinyl sulfones therefore would serve as a good starting point for specificity analysis due to their propensity to target all of the proteasome's active sites.

Initial library studies were performed by using crude homogenates from two different cell lines. The mouse lymphoblastoid cell line EL-4 was chosen because it constitutively expresses the LMP-2, LMP-7, and MECL-1 IFN- γ -inducible subunits of the proteasome. The mouse fibroblast line NIH 3T3 was used because it expresses inducible subunits only upon stimulation by IFN- γ . Analysis of inhibitor binding was performed by using a competition assay in which libraries or single compounds were added to crude homogenates followed by an incubation/binding period. Modification of each subunit then was visualized by addition of a general radiolabeled inhibitor. Three different radioactive affinity labels of the proteasome (NLVS, YL₃-VS, and NP-LLN-VS; see refs. 6 and 19) were used to ensure that specificity profiles were not affected by the potency and selectivity of the label used for analysis. Fig. 2B shows a typical SDS/PAGE profile of a library competition assay. Compounds that efficiently targeted an active site reduced labeling by the radioactive probe. Quantitation of samples relative to the control labeling provided a percent competition value for each subunit. Assays were performed by adding each library to each cell line that had or had not been stimulated by IFN- γ . Overall, this method provided data for all populations of proteasomes, including 26S, 20S, and other regulator-associated complexes regardless of their abundance in the cell; however, minor populations of proteasome complexes with different substrate specificities may exist that would be poorly represented in our bulk data.

Data sets from competition experiments rapidly became overwhelmingly large due to the large number of libraries, extract conditions, and active sites. Hundreds of data points were generated, and a simple and efficient method for analyzing the data was required. We turned our attention to informatics efforts

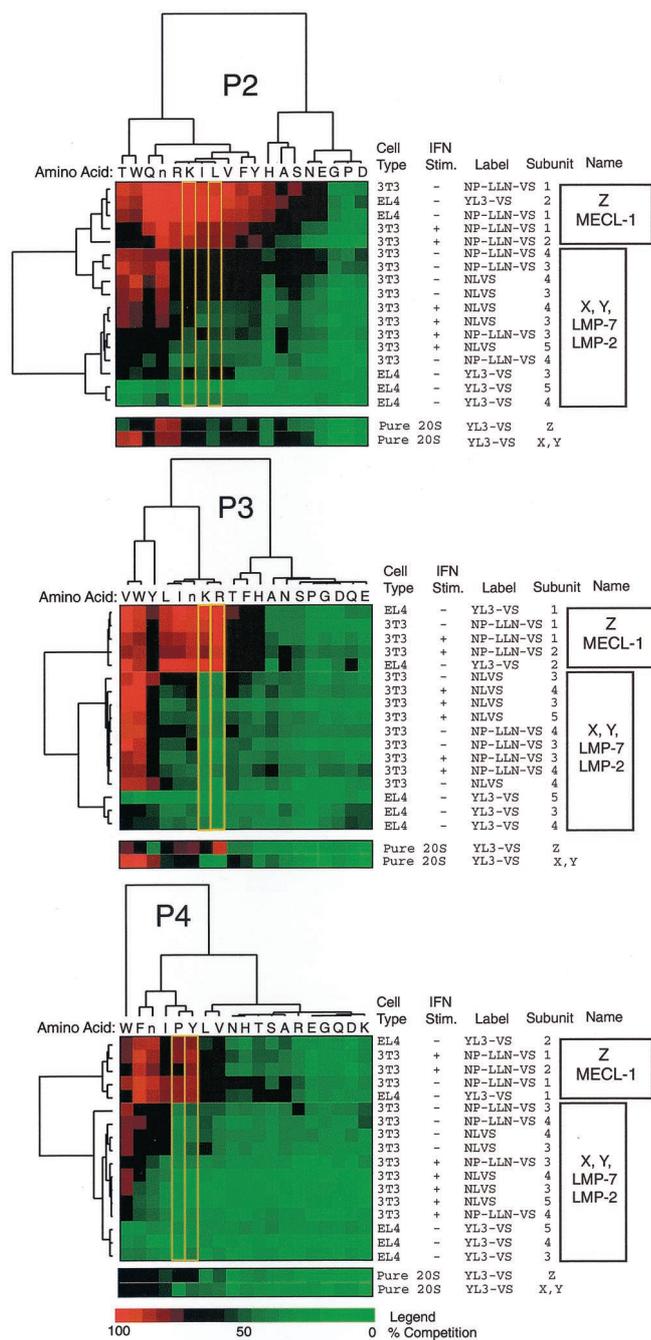


Fig. 3. Cluster analysis of the complete data set obtained for each positional scanning library. Percent competition data obtained from gels as shown in Fig. 2B was converted to color format as described in *Experimental Methods*. Multiple cell lines, conditions of stimulation, and labels were used for analysis in crude cell extracts and are indicated on the right. The tree structure at the top and left of the diagram was obtained by hierarchical clustering and indicates the degree of similarity as a function of the height of the lines connecting profiles. Subunits are numbered according to the bands labeled in Fig. 2B, and the boxed regions to the far right indicate groups of high similarity across the series of constant amino acids. Note the similarity and clustering of the Z and MECL-1 subunit profiles (subunits 1 and 2). Yellow boxes highlight amino acid residues at each position that show the greatest degree of selectivity for the Z and MECL-1 subunits over the X, Y, LMP-2, and LMP-7 subunits. Data from addition of scanning libraries to purified preparations of 20S proteasomes using the methods used for analysis in crude extracts are displayed at the bottom of the clustergrams for comparison.

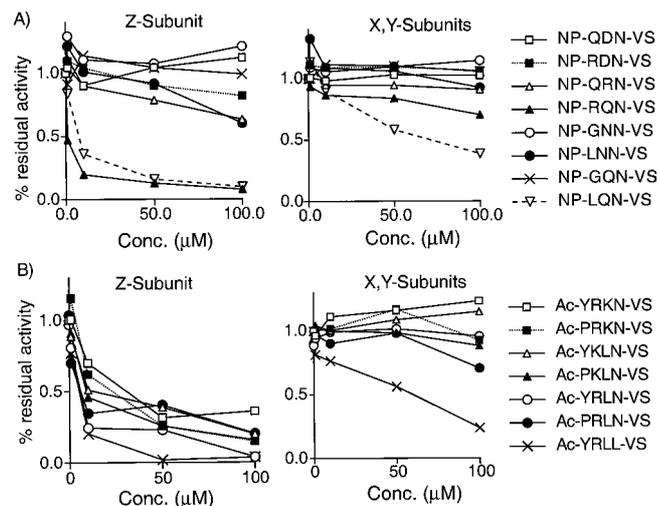


Fig. 4. Library validation and testing of Z-specific inhibitors designed from scanning data. (A) Competition analysis using single components selected based on data from the P2 and P3 positional scanning libraries. Data were obtained by addition of increasing concentrations of the indicated compounds to NIH 3T3 crude cell extracts. Intensity of labeled bands was measured for each concentration and expressed as a ratio to the control, nontreated lane. Residues for this study were chosen based on their overall activity against the multiple active sites of the proteasome. These include strong binding (R, L), weak binding (G, N, D), or positionally dependent strong binding (Q) amino acids. (B) Competition analysis as in A for the Z-subunit selective inhibitors designed based on the optimized residues outlined in Fig. 3. The non-P1 arginine compound Ac-YRLL-VS also is shown. Note the loss of selectivity upon conversion of the P1 amino acid to leucine.

developed for genome-wide transcription profiles. Programs by Eisen and coworkers (21) assign color to data points based on a numerical value for fold induction of genes in the genome. Large data sets can be mathematically clustered to group genes based on similar patterns of expression across a given set of conditions. This software was used for analysis of our specificity data, with red assigned to compounds that showed the greatest competition (hot binders), green assigned to compounds that showed little or no competition (cold binders), and black to compounds that showed 50% competition (Fig. 3). The complete data sets obtained for each library (P2–P4) were clustered individually. The resulting three clustergrams of data grouped subunits with similar specificity profiles across the set of variable amino acids together along the y axis and amino acids with similar reactivity for the multiple proteasomal active sites together along the x axis.

The clustergrams identified amino acids at each variable position (P2–P4) that behave most similarly with respect to subunit modification throughout the various experimental conditions. Not surprisingly, amino acids with similar structure, charge, or polarity such as leucine and isoleucine, lysine and arginine, tyrosine and phenylalanine showed nearly identical profiles of specificity for the proteasome's multiple active sites. Interestingly, some amino acids of nearly identical side-chain structure such as asparagine and glutamine showed dramatically different specificity profiles at distinct positions on the inhibitor, suggesting that the proteasome has well-defined selectivity at non-P1 binding sites. The clustered profiles also show that each β subunit's specificity across each library was identical regardless of the cell type and label used, indicating that the results were highly reproducible. However, the profiles for each library were distinct, suggesting that substrate selectivity is variable and in fact is more narrowly defined at sites further from the P1 position.

Perhaps more importantly, the IFN- γ -inducible subunits (MCL-1, LMP-2, and LMP-7) showed nearly identical P2–P4

Table 1. Kinetic inhibition of the three primary protease activities of the 20S proteasome by Z-specific inhibitors

Inhibitor	$k_{obs}/[I]$ (1/M)		
	Suc-LLVY-MCA	Boc-LRR-MCA	Z-LLE-MCA
Ac-PRLN-VS	ND	1,572 ± 3 (0.2–2 μM)	ND
Ac-YRLN-VS	ND	1,530 ± 1 (0.2–5 μM)	ND
Ac-PKLN-VS	ND	584 ± 3 (0.5–10 μM)	ND
Ac-YKLN-VS	ND	456 ± 2 (0.5–10 μM)	ND
Ac-PRKN-VS	ND	2,149 ± 4 (0.5–5 μM)	ND
Ac-YRKN-VS	5.4 ± 0.4 (100–500 μM)	1,442 ± 7 (1–5 μM)	13.9 ± 0.5 (100–500 μM)

ND indicates no inhibition of substrate hydrolysis was observed at concentrations as high as 100 μM.

specificity profiles as their noninducible counterparts (Fig. 3). On first inspection these data seem to be in conflict with initial studies of the effects of subunit swapping to the substrate specificity of the proteasome. However, these earlier studies relied on fluorogenic substrates that provide only information regarding bulk substrate processing by the complex and fail to indicate specificity changes of individual subunits. Therefore one cannot rule out the possibility that changes in processing of specific substrates upon IFN stimulation is due to an increase in the activity and not the specificity of one or more active sites in the complex as supported by our findings. Our results are also consistent with the recent findings that suggest that inducible β subunits play a role in modulation of structural integrity of the proteasome complex and may have little effect on direct primary sequence recognition (13).

We also compared the profiles observed for proteasome complexes in crude homogenates with those observed for purified preparations of 20S proteasomes that had been artificially activated by addition of low concentrations of SDS (Fig. 3). In general the profiles are similar; however, the P2 and P4 data from purified 20S proteasomes, unlike the data from homogenates, showed nearly identical profiles for the Z and X/Y subunits. The Z subunit of the purified 20S complex also was significantly less active and had a significantly less defined substrate specificity than the crude complexes. These differences may reflect differences in activation of the complex with SDS compared with a physiologically relevant regulator such as the 19S or 11S caps. These data also serve to highlight the differences between activity of purified, homogeneous populations of proteasomes and those that exist within the cell.

Because covalent inhibitors in some ways may fail to accurately mimic a genuine cleavable peptide substrate, we compared our results using the inhibitor libraries to results obtained using positionally scanned libraries of fluorogenic substrates (J. Harris and C. Craik, personal communication). Although these substrate libraries provide only bulk information regarding the specificity of the complex, we observed similar profiles with our inhibitor library scan, suggesting that the peptide vinyl sulfones bind in a manner similar to a substrate and therefore provide information that correlates with global patterns of substrate specificity.

Design of Subunit-Specific Inhibitors Based on Specificity Profiles.

Compiling and sorting the data into clustergrams allowed rapid identification of optimal residues for the design of selective inhibitors to target a single active site of the proteasome. The synthesis of individual compounds based on scanning data also served as a method for validating the library approach. A major shortcoming of positional-scanning libraries is their inability to determine cooperative effects of multiple side chains on substrate or inhibitor binding. By creating mixtures of compounds these effects are lost through averaging of interactions at the variable positions. We chose to synthesize a number of individual nitro phenol-capped tripeptides containing either strong binding (R, L), weak binding (G, N, D), or positionally dependent strong binding (Q; binds only in the P2 position) amino acids. Aspar-

agine and glutamine residues were chosen for their similarity of structure but difference in potency toward the Z subunit of the proteasome when placed in the P2 position.

Compounds were analyzed for potency against each of the active sites of the proteasome by competition analysis in crude extracts from NIH 3T3 cells (Fig. 4A). Residues with poor reactivity in the library scan lead to weak inhibitors upon incorporation into individual compounds. Furthermore, the effects of weak binders outweighed those of strong binding amino acids, leading to poor inhibitors for all compounds containing weak binding amino acids at any position. More subtle effects observed in the library scan also were validated by using this approach. Arginine in the P3 position was accepted by the Z subunit and was excluded by the X and Y subunits in the library scans. Leucine, on the other hand, behaved very similarly to arginine with respect to the Z subunit but bound weakly to the X and Y subunits. This slight difference in potency toward the X/Y subunits was reflected in the single compounds. NP-RQN-VS showed inhibition of only the Z subunit whereas NP-LQN-VS showed inhibition of the Z subunit and also weak activity against the X and Y subunits. Thus, the library approach provided both qualitative and semiquantitative information that could be directly translated into the design of selective inhibitors.

The complete P2–P4 scanning data were used to generate compounds with absolute specificity for the Z subunit of the proteasome. Two amino acids with selectivity for the Z subunit were chosen at each of the positions (Fig. 3; see yellow boxes) and used to generate six compounds containing combinations of these amino acids. Compounds were tested by competition analysis in crude extracts (Fig. 4B). All inhibitors were potent inhibitors of the Z subunit with no measurable activity toward the X and Y subunits. To further test the selectivity of these inhibitors we performed kinetic analysis on purified 20S proteasomes by using the three commonly used fluorogenic substrates (Table 1). All six compounds potentially inhibited hydrolysis of the trypsin-like substrate Boc-LRR-MCA and failed to have any measurable effect on hydrolysis of either the chymotrypsin-like (Suc-LLVY-MCA) or glutamyl peptide hydrolyzing activity (PGPH) (Z-LLE-MCA) substrates (except Ac-YRKN-VS whose effect on the PGPH and chymotrypsin substrates was extremely poor but measurable). This kinetic data provides direct proof that the Z subunit alone is responsible for the trypsin-like activity of the proteasome as measured by the fluorogenic substrate Boc-LRR-MCA.

To determine the effect of the P1 residue in the context of a P2–P4 Z-optimized inhibitor, one of the Z-specific tetrapeptides was synthesized with the P1 amino acid leucine (X/Y subunit directing) in place of asparagine (Z subunit directing). This peptide Ac-YRLL-VS, like its closely related parent compound Ac-YRLN-VS, was a potent inhibitor of the Z subunit but also showed moderate activity against the X/Y subunits (Fig. 4B). Thus, whereas the P2–P4 positions of a substrate clearly play an important role in defining the site of hydrolysis, a combination of interactions at multiple sites leads to absolute discrimination.

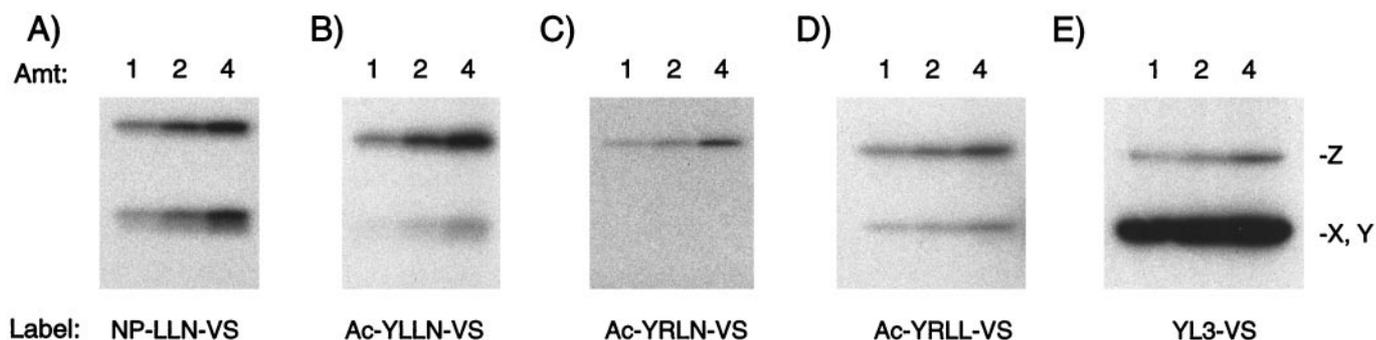


Fig. 5. Tuning the selectivity of affinity labels for the proteasome's multiple active sites. Direct labeling of proteasomes in crude NIH 3T3 lysates using a series of peptide vinyl sulfones. Beginning on the right (A) with the broadly reactive compound NP-LLN-VS and replacing the nitrophenol cap with tyrosine residue (Ac-YLLN-VS; B) increases labeling of the Z subunit. Changing the P3 leucine residue to arginine (Ac-YRLN-VS; C) yields a Z-subunit selective inhibitor. Changing the P1 to leucine (Ac-YRLL-VS; D) restores modest labeling of the X and Y subunits. Finally, returning the P3 residue to leucine (YL₃-VS; E) results in predominant labeling of the X and Y subunits of the proteasome.

Tuning Selectivity of Inhibitors Using Scanning Data. As a final determination of the selectivity of the Z-specific inhibitors we radiolabeled one of the tyrosine containing compounds (Ac-YRLN-VS). We then compared the selectivity of labeling of this compound to other compounds with various positions optimized for binding to either the X/Y subunits or the Z subunit of the proteasome. Beginning with NP-LLL-VS, a selective inhibitor of the X/Y subunits, the P1 residue was changed to asparagine (NP-LLN-VS), resulting in a dramatic increase in the labeling of the Z subunit (see Fig. 2A). Changing the P4 residue from the NP moiety to tyrosine (optimal for Z subunit binding) led to an increase in labeling of the Z subunit relative to the X/Y subunits (Fig. 5A and B). Replacement of the P3 leucine residue with arginine (Ac-YRLN-VS) resulted in the Z-specific compound, which showed labeling of only the Z subunit of the proteasome (Fig. 5C). Returning the P1 residue to the X/Y directing leucine (Ac-YRLL-VS; Fig. 5D) restores moderate labeling of the X/Y subunits and finally returning the P3 to leucine (YL₃-VS; Fig. 5E) results in predominant labeling of the X/Y subunits. Thus, using information from the positional scanning data it was possible to fine-tune the selectivity of inhibitors toward individual subunits of the proteasome.

Conclusions

We have developed a rapid method for the analysis of substrate and inhibitor selectivity of the proteasome. Positional scanning libraries coupled with highly selective affinity labels allowed

analysis of binding to individual active sites of the proteasome in bulk populations of proteasomes. This scanning approach has yielded global specificity profiles of individual catalytic subunits and also has led to the design of inhibitors with “tunable” selectivity toward individual active sites in the proteasome complex. Subunit selective inhibitors were used in this study to identify the Z subunit as the sole catalytic center responsible for the trypsin-like activity of the mammalian proteasome. Future work with cell-permeable analogs of these selective inhibitors is likely to further define the roles of individual subunits in specific processing events catalyzed by the proteasome. Finally, specificity profiles of the active sites analyzed under a variety of physiological conditions will provide information as to how the proteasome is able to control the processing of substrates. Surprisingly, our data suggest that the overall P2–P4 substrate specificity of subunits is static throughout conditions such as IFN stimulation, suggesting that regulation of substrate processing is likely to be controlled by factors other than modulation of the primary sequence specificity of the proteasome complex.

We thank Marty Rechsteiner, University of Utah, for the generous gift of purified 20S proteasomes and critical evaluation of the manuscript. We thank Joe DeRisi for assistance with data analysis using TREE VIEW and CLUSTER. We thank Charles Craik and Jennifer Harris for information regarding bulk substrate specificity of the proteasome obtained from fluorogenic substrate libraries. We thank Jon Ellman for experimental advice and critical evaluation of the manuscript. This work was supported by funding from the Sandler Foundation.

- Coux, O., Tanaka, K. & Goldberg, A. L. (1996) *Annu. Rev. Biochem.* **65**, 801–847.
- Lupas, A., Zwickl, P., Wenzel, T., Seemüller, E. & Baumeister, W. (1995) *Cold Spring Harbor Symp. Quant. Biol.* **LX**, 515–524.
- Bochtler, M., Ditzel, L., Groll, M., Hartmann, C. & Huber, R. (1999) *Annu. Rev. Biophys. Biomol. Struct.* **28**, 295–317.
- Wilk, S. & Orłowski, M. (1983) *J. Neurochem.* **40**, 842–849.
- Cardozo, C., Vinitzky, A., Michaud, C. & Orłowski, M. (1994) *Biochemistry* **33**, 6483–6489.
- Bogyo, M., Shin, S., McMaster, J. S. & Ploegh, H. (1998) *Chem. Biol.* **5**, 307–320.
- Elofsson, M., Splittgerber, U., Myung, J., Mohan, R. & Crews, C. M. (1999) *Chem. Biol.* **6**, 811–822.
- Fruh, K., Gossen, M., Wang, K., Bujard, H., Peterson, P. A. & Yang, Y. (1994) *EMBO J.* **13**, 3236–3244.
- Boes, B., Hengel, H., Ruppert, T., Multhaup, G., Koszinowski, U. H. & Kloetzel, P. M. (1994) *J. Exp. Med.* **179**, 901–909.
- Driscoll, J., Brown, M. G., Finley, D. & Monaco, J. J. (1993) *Nature (London)* **365**, 262–264.
- Gaczynska, M., Rock, K. L. & Goldberg, A. L. (1993) *Nature (London)* **365**, 264–267.
- Eleuteri, A. M., Kohanski, R. A., Cardozo, C. & Orłowski, M. (1997) *J. Biol. Chem.* **272**, 11824–11831.
- Sijts, A. J., Ruppert, T., Rehmann, B., Schmidt, M., Koszinowski, U. & Kloetzel, P. M. (2000) *J. Exp. Med.* **191**, 503–514.
- Matthews, D. J. & Wells, J. A. (1993) *Science* **260**, 1113–1117.
- Harris, J. L., Backes, B. J., Leonetti, F., Mahrus, S., Ellman, J. A. & Craik, C. S. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 7754–7759. (First Published June 27, 2000; 10.1073/pnas.140132697)
- Dick, T. P., Nussbaum, A. K., Deeg, M., Heinemeyer, W., Groll, M., Schirle, M., Keilholz, W., Stevanovic, S., Wolf, D. H., Huber, R., et al. (1998) *J. Biol. Chem.* **273**, 25637–25646.
- Nussbaum, A. K., Dick, T. P., Keilholz, W., Schirle, M., Stevanovic, S., Dietz, K., Heinemeyer, W., Groll, M., Wolf, D. H., Huber, R., et al. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 12504–12509.
- Holzhtuter, H. G., Frommel, C. & Kloetzel, P. M. (1999) *J. Mol. Biol.* **286**, 1251–1265.
- Bogyo, M., McMaster, J. S., Gaczynska, M., Tortorella, D., Goldberg, A. L. & Ploegh, H. L. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 6629–6634.
- Houghten, R. A., Pinilla, C., Blondelle, S. E., Appel, J. R., Dooley, C. T. & Cuervo, J. H. (1991) *Nature (London)* **354**, 84–86.
- Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 14863–14868.
- Ostresh, J. M., Winkle, J. H., Hamashin, V. T. & Houghten, R. A. (1994) *Biopolymers* **34**, 1681–1689.
- Overkleeft, H. S., Bos, P. R., Hekking, B. G., Gordon, E. J., Ploegh, H. L. & Kessler, B. M. (2000) *Tetrahedron Lett.* **41**, 6005–6009.