Distinct Specificities of Mycobacterium tuberculosis and Mammalian Proteasomes for N-Acetyl Tripeptide Substrates

Gang Lin¹*, Christopher Tsu²*, Lawrence Dick², Xi K. Zhou³ and Carl Nathan⁴*

¹Department of Microbiology and Immunology, Weill Medical College of Cornell University, New York, NY, 10065; ²Millennium Pharmaceuticals Inc., Cambridge, MA 02139; ³Department of Public Health, Weill Medical College of Cornell University, New York, NY, 10065

*Contributed equally.

To whom correspondence should be addressed at: Gang Lin, gal2005@med.cornell.edu; tel. no.: (212) 746-2985; fax no.: (212) 746-8536; Carl Nathan, cnathan@med.cornell.edu; tel. no.: (212) 746-6505; fax no.: (212) 746-8587

Running Title: Substrate specificity of Mtb vs bovine proteasomes

The proteasome of Mycobacterium tuberculosis (Mtb) is a validated and druggable target for therapeutics. To lay groundwork for developing peptide-based inhibitors with a useful degree of selectivity for Mtb’s proteasome over those of the host, we used a library of 5,920 N-acetyl-tripeptide-aminomethylcoumarins to contrast the substrate preferences of the recombinant Mtb proteasome wild type and open-gate mutant, the Rhodococcus erythropolis proteasome and the bovine proteasome with activator PA28. The Mtb proteasome was distinctive in strictly preferring P1 = tryptophan, particularly in combination with P3 = glycine, proline, lysine or arginine. Screening results were validated with Michalis-Menten kinetic analyses of 21 oligopeptide aminomethyl-coumarin substrates. Bortezomib, a proteasome inhibitor in clinical use, and 17 analogues varying only at P1 were used to examine the differential impact of inhibitors on human and Mtb proteasomes. The results with the inhibitor panel confirmed those with the substrate panel in demonstrating differential preferences of Mtb and mammalian proteasomes at the P1 amino acid. Changing P1 in bortezomib from Leu to m-CF₃-Phe led to a 220-fold increase in IC₅₀ against the human proteasome, while changing a P1 Ala to m-F-Phe decreased the IC₅₀ 400-fold against the Mtb proteasome. The change of a P1 Ala to m-CI-Phe led to an 8000-fold shift in inhibitory potency in favor of the Mtb proteasome, resulting in 8-fold selectivity. Combinations of preferred amino acids at different sites may thus improve the species selectivity of peptide-based inhibitors that target the Mtb proteasome.

Mycobacterium tuberculosis (Mtb), the single leading cause of death from bacterial infection, is growing even more dangerous with the spread of resistance to the existing chemotherapeutic agents (1). After 4 decades of efforts to find new chemophores that inhibit new targets and to shorten the 6- to 9-month treatment regimen that makes compliance problematic (2-4). Many authorities believe that this will require inclusion of agents active against non-replicating Mtb (5-8).

Proteasomes provide the major pathway of intracellular protein degradation in eukaryotic cells, where they are essential for adaptation to changing circumstances and avoidance of toxicity from oxidized or denatured proteins (9-11). Mtb is one of the few bacteria known to express a proteasome (12). The Mtb proteasome is emerging as a potential anti-tuberculosis drug target. It defends the bacterium against host nitrosative stress (12) and is essential for Mtb to persist in mice (13). The Mtb proteasome can be inhibited by small molecules both as a recombinant protein complex (14,15) and within intact mycobacteria (12). The structure of the Mtb proteasome has been solved with a peptidyl boronate (15). However, for proteasome inhibition to be a viable approach, the inhibitors must be much more active against the bacterium’s proteasome than against the host’s. Partial
inhibition of host proteasomes has been successful in the treatment of certain malignancies (16-18). Toxicity in this setting is often considered acceptable, but would be highly undesirable in the treatment of an infectious disease. The extensive conservation of proteasome structure makes it challenging to find inhibitors with species selectivity.

A standard approach to developing selective protease inhibitors is to attach a warhead to a selective substrate (19). Among proteasome inhibitors, peptidyl boronates exemplify this approach and have achieved clinical utility (16-18). Thus we have embarked on a program to develop peptidyl boronates with greater potency for the Mtb proteasome than for the β5 subunit of the human proteasome, whose inhibition is chiefly responsible for the toxicity of these compounds (16,17). As a step toward that goal, the present study sought potentially exploitable differences in the preferences of bacterial and mammalian proteasomes for tripeptide substrates.

The proteasome core, here called “20S”, is a cylindrical 28-meric protein complex with proteolytic sites shielded within a central chamber (11), access to which is regulated by additional subunits that bind and hydrolyze ATP. The overall structures of proteasomes among eukaryotes and prokaryotes (archaea and eubacterial Actinomycetes) share a barrel structure, with α subunits forming two heptameric outer rings and β subunits forming two heptameric inner rings. Proteolysis occurs at N-terminal threonines on β subunits that are exposed by autocatalytic removal of a propeptide upon assembly (20). In eukaryotes there are 7 types of α and 7 types of β subunits, with only three of the β subunits (β1, β2, and β5) displaying proteolytic activities. With oligopeptide substrates, these activities are caspase-like, trypsin-like, and chymotrypsin-like, respectively (21) (Fig. 1a). In contrast, prokaryotic proteasomes usually include only one or two types of α and β subunit, and the active sites are usually chymotrypsin-like with oligopeptide substrates. Mtb’s β subunits are of a single type. Cryo-electron microscopy, X-ray crystallography and mutation analysis suggested that the α subunits have a gating function and confirmed that the 14 β subunits each provide an active site N-terminal threonine OH (14,15) (Fig. 1a). In the present work we used a mutant form of the recombinant Mtb proteasome (Mtb20SOG) in which deletion of the N-terminal octapeptide from the α subunits mimics the presumed “open-gate” configuration and increases specific activity toward oligopeptide substrates by about an order of magnitude (14).

Studies of substrate specificities have been reported for proteasomes of human (22,23) and another eukaryote, Trypanosoma brucei (24), using positional scanning substrate or inhibitor libraries. The positional scanning libraries were usually constructed as 20 pools per position (P1, P2, P3 and P4), each of which contained large numbers of peptides fixed at a single residue and otherwise variant. These studies provided a global analysis of substrate preference at defined positions in the peptide adjacent to the cleavage site (P1) and more distally (P2, P3, P4) and served well for identification of selective inhibitors for different β subunits (25). However, using positional scanning libraries of substrates or inhibitors is likely to miss the interactions among amino acids at different positions. It is also likely to be more challenging to identify differences in substrate preference between two highly homologous β subunits from different species that are both predominantly chymotryptic. To discern potentially subtle differences in substrate preferences, we needed to take into account the influence of the amino acid at each position on the preferences at each other position. We approached this goal by using a robotic microfluidic device to assess the specific activity of Mtb and bovine proteasomes toward a combinatorial library of 5,920 fluorescently tagged tripeptides, assaying each substrate individually. For comparison, we also tested recombinant proteins from another Actinomycete, Rhodococcus erythropolis. Finally, we validated our findings through kinetic analyses with 21 tripeptides and 18 P1 amino acid analogues of bortezomib.

**Materials and Methods**

Overexpression and purification of recombinant Mtb proteasome, Mtb PreAB-OG and Rhodococcus 20S followed the reported method (14). Bovine RBC 20S, a generous gift of Dr. George DeMartino, University of Texas Southwestern Medical Center, was purified as described (26). Human RBC 20S was purchased from Boston Biochem (Cambridge, MA). The recombinant alpha subunit of the rat PA28 activator was purified as described (27,28). The concentrations of the proteasomes were calculated...
based on their molecular weight (~700k Da); multiplicity of active sites was not taken into account. The ChemRX Protease Profiler library (29) was purchased from Discovery Partners International (South San Francisco, CA). The library was reconfigured from 96-well to 384-well format. For assay, substrate plates were prepared by mixing 1 µL of the 1 mM stock and 70 µL of microfluidics buffer (50 mM Tris, pH 7.8; 20 mM NaCl, 0.5 mM EDTA; 0.005% Triton X-100) in 384 well polypropylene plates, yielding a substrate concentration of 14 µM. On-chip dilution of substrate was 70%, for a final substrate concentration of 10 µM. Individual substrates for kinetic analysis were custom synthesized by AnaSpec (San Jose, CA). Suc-LLVY-AMC was from Bachem Biosciences (King of Prussia, PA). Z-VLR-AMC was from MD Biosciences, Inc. (St. Paul, MN). Bortezomib and its analogues were synthesized in-house (Millennium Pharmaceuticals Inc., Cambridge, MA).

**Caliper-chip assay**— A detailed description of method will be published elsewhere. In brief, the Caliper 220 Drug Discovery Systems (Caliper Technologies, Mountain View, CA) was set up according to the manufacturer’s instructions. Taking into account an on-chip dilution of 30%, the final protein concentrations of the Mtb20SWT, Mtb20SOG and Rhod20S were 500 nM, 250 nM and 24 nM, respectively. The final concentration of bov20S and PA28α were 225 nM, and 675 nM, respectively. The final concentration of each substrate was 10 µM. The screening was carried out in buffer (50 mM Tris, pH 7.8, 100 mM NaCl, 0.5 mM EDTA, 0.005% Triton X-100, 5 mM β-ME) at 26°C with 60% humidity. A FS267 or TF460 microfluidics chip was loaded into a Caliper 220 or 250 robot. The run pressure was set to make the on-chip reaction time ~90 seconds for Mtb20SWT, ~60S for Mtb20SOG and Rhod20S, and ~30 seconds for bov20S with PA28α. Relative fluorescence units (RFU) were recorded (excitation 355 nm, emission 420-480 nm). The data were collected in less than 12 hours by the Caliper LabChip HTS software. Fluorogenic activity was determined for each sample using Caliper HTS Well Analyzer software. Sipper-to-sipper variations in the raw data were normalized by using the Cy5 (Caliper Technologies, Mountain View, CA) marker (FS267) or AMC control as a reference (TF460). The background fluorescence of the unhydrolyzed substrate was detected in real-time in a parallel channel lacking enzyme (TF460: Mtb20WT, Mtb20SOG, Rhod20S; Fig. 1c) or by the subtraction of a separate substrate-only data set (FS267: Bov20S). The final annotated data were exported to Spotfire Descisionsite (Spotfire, Somerville, MA) for display and quantitative analysis.

**Single-substrate kinetic assays**— These were performed in 96-well format at final substrate concentrations of 1.56 µM to 200 µM in 1% DMSO at 37°C in 20 mM HEPES pH 7.5 and 0.5 mM EDTA. Enzyme concentrations were 13.3 nM for Mtb20S or Mtb20SOG, 0.5 nM Rhod20S, and 1.4 – 2.9 nM for human 20S and 5.5 – 11 nM PA28α. All reactions were done in 200 µL triplicate on a PolarStar Galaxy spectrofluorimeter plate reader (BMG Labtech, Durham, NC). Hydrolysis of AMC substrates was monitored fluorometrically with periodic shaking (excitation, 340 nm; emission, 460 nm). The linear ranges of the time-course curve were used to calculate the initial reaction velocities. Raw fluorescence was converted to nM/min using an AMC standard curve. The steady-state parameters $k_{cat}$ and $K_M$ were determined by non-linear regression in Kaleidagraph (Synergy Software). Occasional outliers were omitted from the analysis, but no fewer than five concentrations were used. In some cases, only $k_{cat}/K_M$ values could be obtained due to substrate inhibition and/or precipitation occurring at high concentrations. Error in the fit was less than 10% for bov20S, 15% for Mtb20SOG.

**IC_{50} of P1 analogues of bortezomib against Mtb proteasome**— These were performed in a 384-well format assay. 1 µL of stock DMSO solutions (5 mM - 0.25 µM) of 22 inhibitors was spotted in duplicate in the wells of a black 384-well plate. For the Mtb20S proteasomes, 25 µL of 2x substrate buffer (100 µM in 20 mM HEPES, 0.5 mM EDTA, pH 7.5) were dispensed into each well using a Multidrop instrument (ThermoFisher Scientific, Waltham, MA). The plate was shaken for 3 minutes on an orbital shaker. Then, 25 µL of 2x enzyme buffer (10 nM in 20 mM HEPES, 0.5 mM EDTA, pH7.5) were dispensed into each well. The plate was then spun at 1000 rpm for 1 minute and the progress of the reactions at 37°C was monitored at 37°C in a BMG plate reader as above for the AMC substrates. IC_{50} curves were determined by non-linear regression to the Hill equation in ActivityBase (IDBS, Burlington, MA).
The human RBC 20S assays were performed as described for the Mtb20S proteasomes, with the following modifications: The 2X substrate solution was 30 µM Ac-WLA-AMC, 0.05% SDS activator in 20 mM HEPES, 0.5 mM EDTA, pH 7.5. The 2X enzyme solution was 3 nM human RBC proteasome in 20 mM HEPES, 0.5 mM EDTA, pH 7.5. Error in the fit was less than 10% for both hu20S and Mtb20SOG.

RESULTS

Ac-P3P2P1-AMC library and assay design. The design of the combinatorial Ac-P3P2P1-AMC library is depicted in Fig. 1b. Besides the 20 amino acids commonly incorporated into polypeptides, two additional natural amino acids, L-citrulline (B) and L-ornithine (O), were incorporated in the P1 position for a total of 22. 20 amino acids (omitting L-citrulline and L-cysteine) were incorporated into the P2 and P3 positions. The amino terminus of P3 was capped with an acetyl group. This has the distinct advantage of minimizing the influence of the artificial blocking group, as compared to the aromatic benzyloxy carbonyl (Z) or acidic succinyl (Suc) group, as compared to the aromatic benzyloxy carbonyl (Z) or acidic succinyl (Suc) often used in peptides. Out of the possible 8,800 tripeptides (22 P1 x 20 P2 x 20 P3), 5,920 were usable based on yield and purity. The average purity was > 90% and not less than 85% for any individual substrate (29). A robotically controlled microfluidic chip facilitated the performance of the 5,920 reactions with each of 4 proteasome preparations (Fig. 1c): Mtb20SWT, Mtb20SOG, Rhod20S and bov20S. Substrate concentrations were kept at 10 µM, equal to or a few-fold lower than the values of $K_M$ (10 – 200 µM) previously reported for tri/tetrapeptide substrates (24,30). Under these conditions, reaction rates likely reflected the specific activity ($k_{cat}/K_M$) in the absence of substrate saturation effects. Stream splitting and simultaneous fluorescence detection facilitated subtraction of the background fluorescence of each unhydrolyzed substrate (Fig. 1c).

Overall intra- and inter-species comparisons. As reported with other substrates (14), Mtb20SWT was up to 20-fold less active than Mtb20SOG. In the initial set of experiments, we used twice as much Mtb20SWT as Mtb20SOG and an increased reaction time to evaluate the correlation of specific activities for different substrates between Mtb20SWT and Mtb20SOG. This allowed us to establish that Mtb20SOG could serve as an authentic reporter of the Mtb proteasome for subsequent interspecies comparisons. Thus, a plot of the activity of Mtb20SWT and Mtb20SOG for each substrate yielded a correlation coefficient $R^2 = 0.78$ (Fig. S1a), indicating that the substrate preferences of the two Mtb proteasome preparations were highly similar.

In contrast, the correlation coefficients between results for Mtb20SOG and Rhod20S ($R^2 = 0.11$), Mtb20SOG and bov20S ($R^2 = 0.03$), and Rhod20S and bov20S ($R^2 = 0.45$) (Figs. S1b, c, d, respectively) showed that the Mtb proteasome stood apart in its tripeptide substrate preferences from the preferences exhibited by proteasomes of the other two species.

Preferences of Mtb, Rhodococcus, and bovine proteasomes for individual N-Ac-tripeptides. Our earlier biochemical characterization of the Mtb proteasome with Z-capped tripeptide and tetrapeptide substrates revealed multiple proteolytic activities with a preference for basic P1 residues (14). In contrast, with N-acetyl tripeptides, Mtb proteasomes demonstrated only chymotryptic activity ($P1 = F, L, I, W, Y$) (Fig. 2a, b), with the exception of a few tripeptides with $P1 = H$ (Table 1). These observations confirmed that constituents at one position (in this case, the NH$_2$-terminus) could markedly affect the proteasome’s preference for particular side chains at a different position (in this case, P1). Among the N-acetyl tripeptides, Mtb20SOG was strongly biased toward $P1 = W$. The most preferred substrate, WQW, allowed a 2-fold higher specific activity than RWH, the most favored tripeptide that lacked $P1 = W$. Among the 30 most preferred substrates (top 0.5 percent of the library), 28 had $P1 = W$ and 2 had $P1 = H$. At P2, Q and W were the most favored amino acids. The S3 site appeared to be more accommodating, as most amino acid residues were represented at the P3 position in the top 1% of the most preferred substrates. However, acidic residues (D and E) were disfavored at P3. This might be explained by the presence of Asp30 at the bottom of the S3 pocket (25).

Rhod20S behaved as a typical chymotryptic proteasome (Fig. 2c), with a preference for hydrophobic and aromatic amino acids at P1 (Table 1). Unlike the Mtb proteasome, there was no strong bias toward one P1 amino acid. Also in contrast to the Mtb proteasome,
Rhod20S displayed no activity for any substrates having P1 = H. Among the top 1% of the most preferred substrates for Rhod20S, most of the uncharged amino acids were found at P2 and P3.

As expected, bov20S displayed three protease-like activities (Fig. 2d). A preference at P1 for hydrophobic and aromatic amino acid residues reflected chymotryptic activity; for R and K, tryptic activity; and for D and E, caspase-like activity. For chymotryptic activity, the preferred tripeptide sequence was P1 = L/F/Y, P2 = Q/S/L, P3 = Y/F/W (Table 1). However, the P1 = L and P3 = Y were the most favored, whereas the hydrophilic P1 = N and P3 = O/N were the least favored (Fig. S2a). For trypsin-like activity, the preferred amino acids were P1 = R and P3 = R and K, with acidic residues strongly disfavored at P3. This differed from the profile of human 20S determined in a tetrapeptidyl 7-amino-4-carbamoylmethylcoumarin positional scanning library when P1 was fixed as R, where the preferred P3 was E and K, whereas P3 = R was disfavored (22,31). For caspase-like activity of bov20S, the P2 amino acids appeared to be more determinant than the P3 amino acids, as the preferred sequences were Ac-Xaa-F/W/Y-D/E, whereas the most disfavored were Ac-E/K/R-Xaa-D/E-AMC (Fig. S2b). The preference of bov20S for basic residues at P3 of tryp tic substrates (Fig. S2c) may reflect the presence of D28 of the β2 subunit of the proteasome at the bottom of the S3 pocket, which can form hydrogen bonds or salt bridges with protonated guanidino or amino groups (25). P2 amino acids were less of a determinant than P1 and P3.

Comparison of preferences of Mtb and bovine proteasomes for shared substrates. To aid in comparison of large data sets, we performed hierarchical clustering analysis of substrate specificity for proteasomes from the three species, as illustrated in Fig. 3 for the comparison of Mtb20SOG and bov20S at P1 and P3. The clustergram revealed 12 subgroups with different pairs of residues at P1 and P3 that contain more preferred substrates for Mtb20SOG than for bov20S: W/A, W/G, W/K, W/O, W/P, W/R, T/E, T/Q, R/P, R/D, O/H and O/N.

Because Mtb20SOG was only substantially active on substrates with P1 = W, we next focused on this subgroup of substrates in comparing preferences of Mtb20SOG and bov20S. The correlation between their activities yielded $R^2$ = 0.33 (not shown), indicative of a partial overlap in substrate preferences. However, two subgroups were much more favored by Mtb20SOG than by bov20S: those with sequences Ac-G-Xaa-W-AMC and Ac-P-Xaa-W-AMC (Fig. 4a, 4b). The activity of Mtb20SOG on Ac-G-Xaa-W-AMC appeared to depend on bulkiness and neutral hydrophilicity (Q/N) of P2: the bulkier, the more preferable the substrate was for the Mtb proteasome. Although the effect of P2 amino acids on Ac-P-Xaa-W-AMC was variable, Ac-P-W-W-AMC was a selective substrate for Mtb20SOG.

In a tetrapeptide inhibitor with a vinyl sulfone warhead, Ac-PRLN-VS, inclusion of P3 = R in combination with P1 = N conferred selectivity for human and yeast 20S β2 subunits over their β5 subunits (23). This was attributable to interaction between the guanidino group of Arg and the D28 of the β2 subunit of the proteasome of both species at the bottom of the S3 pocket (23,25), given that D28 is not conserved in β5 of human or yeast 20S. However, the homologous residue, D30, is conserved in the β subunits of the Mtb proteasome (15). As shown in Fig. 4c, 4d, comparison of the two subgroups of substrates with formulas Ac-R/K-Xaa-W-AMC for Mtb20SOG and bov20S did not reveal a sharp difference in preferences. However, hydrophobic or aromatic P2 amino acids were favored by Mtb20SOG and basic P2 amino acids by bov20S.

Kinetics of hydrolysis of individual substrates. For a better understanding of the kinetic behavior of these proteasomes, we undertook a Michaelis-Menten steady state analysis of specificity with a group of 19 N-acetyl-tripeptidyl-AMC substrates, along with Z-VLR-AMC and suc-LLVY-AMC. Of this set of 21 substrates, 16 would be conventionally considered chymotryptic, 2 tryptic, and 3 caspase-like substrates (Table 2). Although Z-VLR-AMC was a good substrate for Mtb proteasomes (14), activity of N-acetyl-VLR-AMC was undetectable. Thus, Z appears to contribute significantly to binding affinity.

In general, the steady state kinetic analysis of individual, resynthesized substrates agreed well with the large-scale substrate specificity assay conducted at a fixed concentration and time. With few exceptions, the Mtb proteasome was active only on chymotryptic substrates with P1 = W, whereas bov20S was active on all types of substrates. The $K_M$ values of all substrates for which activity was detected were between 10 - 180...
µM for bov20S and similarly, 10 – 100 µM for Mtb proteasomes (Table 2). No $k_{\text{cat}}$ and $K_M$ could be estimated for a few substrates due to either precipitation or substrate inhibition at high concentrations before saturation was achieved. The $k_{\text{cat}}/K_M$ values for these substrates were estimated as the slopes of the linear plots of each set of data. Almost all the preferred substrates had similar $K_M$ values for Mtb20SWT as for Mtb20SOG, but with 20-30 fold reduction of $k_{\text{cat}}$ (Mtb20SWT data not shown). The correlation of $k_{\text{cat}}/K_M$ values of Mtb20SOG and Mtb20SWT yielded an $R^2$ 0.82 (Fig. S3), matched well with the R2 0.78 of the correlation of results from high throughput screening of substrates between Mtb20SOG and Mtb20SWT.

For the Mtb proteasome, the change of P1 = W to P1 = L in Ac-YQW-AMC resulted in a 66-fold decrease in $k_{\text{cat}}/K_M$, from 0.66 µM⁻¹min⁻¹ to 0.01 µM⁻¹min⁻¹, underscoring the importance of a bulky, hydrophobic residue at P1. For the bovine proteasome β5 subunit, hydrophobic and aromatic amino acid residues at P1 contributed to activity as substrates. To a lesser degree, activity scaled with increasing size of the residue at P2. For example, changing P2 from Ac-RAW-AMC to Ac-RFW-AMC resulted in a 2.4-fold increase in the $k_{\text{cat}}/K_M$, from 0.26 µM⁻¹min⁻¹ to 0.63 µM⁻¹min⁻¹. The corresponding impact on the Mtb proteasome was a 1.8-fold increase, from 0.79 µM⁻¹min⁻¹ to 1.4 µM⁻¹min⁻¹. For the bovine proteasome, we confirmed that changing P3 from Y to W to yield Ac-WQW-AMC resulted in a 1.4-fold decrease in $k_{\text{cat}}/K_M$, from 5.12 µM⁻¹min⁻¹ to 3.7 µM⁻¹min⁻¹. The result was opposite in the Mtb proteasome: a 2-fold increase, from 0.66 µM⁻¹min⁻¹ to 1.25 µM⁻¹min⁻¹.

The most selective of the 21 substrates whose kinetics were tested was Ac-LWW-AMC, similar to Ac-GWW, which ranked as the most selective substrate in the larger library (Fig. 4a). Ac-LWW-AMC showed a 6-fold difference in $k_{\text{cat}}/K_M$ (0.09 µM⁻¹min⁻¹ vs 0.55 µM⁻¹min⁻¹) in favor of the Mtb proteasome versus bov20S β5. The Km of Ac-LWW-AMC was 9.1 µM against the Mtb proteasome, a 10-fold increase in binding affinity compared to other substrates with measurable Km values for Mtb proteasome.

We have previously reported that the Mtb20S possesses a broad specificity, primarily tryptic in nature, as judged by activity on Z-VLR-AMC, Suc-LLVY-AMC and other canonical proteasome substrates [13]. This finding is supported here, with $k_{\text{cat}}/K_M$ for Z-VLR-AMC being 5-fold greater than for Suc-LLVY-AMC. However, the P1-W mutant of bortezomib (I), which inhibits the human proteasome β5 subunit with high potency (IC₅₀ 0.016 µM) and selectivity. The bortezomib-analogue library was divided into two groups: compounds containing aliphatic, branched side chains 2 – 9 and those containing Phe and substituted benzyl analogues 10 – 18 (Table 3). We performed experiments with both Mtb proteasome preparations (wild type and open-gate mutant) in comparison to human proteasomes purified from red blood cells and activated with PA28. We used both a chymotryptic substrate (Ac-RFW-AMC) and a tryptic substrate (Z-VLR-AMC) for the Mtb proteasomes, and varied the concentration of inhibitors over the range 5 nM to 100 µM. IC₅₀ values were obtained by fitting the data to the Hill equation. The IC₅₀ values of Mtb20SWT and Mtb20SOG were almost identical; only those of Mtb20SOG are shown. Moreover, the IC₅₀/S of the inhibitors were generally identical with each substrate.

Among the inhibitors with aliphatic side chains, that with R = isobutyl (bortezomib) 1 was the most potent, which was 3-fold more potent than the next, which had R = 2-butyl (isoleucine side chain) 8. The n-propyl 9, n-butyl 7, or isopentyl 4 side chains further increased IC₅₀S to 1.11 – 1.88 µM. The continuous reduction of the size of side chain to n-ethyl 3 and methyl 5 led to the IC₅₀s rising up to 8 – 44 µM, a 154-fold decrease of potency from bortezomib 1. The neopentyl side chain 2 eliminated inhibitory potency. In sum, substitution of P1 with a 4-carbon branched side chain improved the
inhibitory potency the most, whereas extremely bulky substitution reduced it.

Consistent with P1 = W being a preferred substrate among the tripeptides for the Mtb proteasome, the aromatic side chain analogues of bortezomib were generally better inhibitors than those with aliphatic side chains (Table 3). The side chain R = phenyl of 10 appeared to be the least potent in the group, followed by m-CF$_3$-benzyl 11. The replacement of m-CF$_3$ 11 by m-CH$_3$ 13 improved potency by 3-fold, whereas p-CH$_3$ 12 further improved potency by 1.8-fold. Replacement of p-CH$_3$ by p-Cl 14 reduced potency by 2-fold. By contrast, the o- or m- substitution of Cl, or F, respectively, i.e. o-Cl 15, m-Cl 16, o-F 17, m-F 18, afforded no further improvement in inhibitory potency. Thus, bortezomib analogs with P1 aromatic side chains generally had enhanced inhibitory potency, and those with m- or o- substitution with electron-withdrawing halogens (F, Cl) were the most potent Mtb proteasome inhibitors. Since electron-donating p-CH$_3$ 12 also showed identical potency as 15 - 18, it seemed that the electronic property of the substituents, and the position of the substitution on the phenyl, might be inter-dependent in their effects on potency.

**DISCUSSION**

The Mtb proteasome is a non-redundant pathway by which Mtb protects itself against nitrosative stress and metabolic stringency in vitro (12) and it is essential for Mtb to survive in mice, even if they are immunodeficient (13). This information, combined with the ability of small, drug-like compounds to enter Mtb, inhibit its proteasome and kill the bacterium (12), focuses interest on developing selective proteasome inhibitors as potential leads for chemotherapeutics. Such compounds might offer a range of features currently prized in potential new agents for treatment of tuberculosis: that they be new chemophores active against new targets and effective against non-replicating Mtb. However, it remains a major hurdle to attain selectivity between β subunits that are predominantly chymotryptic in both pathogen and host.

Using a microfluidic assay, we analyzed the substrate specificity of Mtb, *Rhodococcus*, and bovine proteasomes with a library of 5920 tripeptidyl-AMC substrates tested individually. When each was given a preferred substrate, the bovine proteasome was approximately 3-fold more active than the *Rhodococcus* proteasome and the Mtb proteasome open-gate mutant. Proteasomes of Mtb and cow demonstrated similar $K_m$ values for most of the preferred substrates, in the range from 10 – 200 µM, whereas the major difference in terms of enzymatic efficiency lay in $k_{cat}$.

The bovine proteasome demonstrated three distinct protease activities, as anticipated based on studies of all other eukaryotic proteasomes to date. The *Rhodococcus* proteasome demonstrated only chymotryptic activity, as reported for all prokaryotic proteasomes to date except Mtb’s (14), and consistent with our earlier results using Z-capped tripeptide or tetrapeptide substrates (14). Moreover, comparing large numbers of individual N-acetyl tripeptides having hydrophobic P1 residues, the chymotryptic substrate preferences of the bovine and the *Rhodococcus* proteasomes were similar ($R^2 = 0.49$).

In contrast, the Mtb proteasome stood apart from the other two in several respects. We previously reported that the Mtb proteasome displays multiple proteolytic activities with Z-capped tripeptide or tetrapeptide substrates (14). However, with acetyl tripeptide substrates, the Mtb proteasome only showed chymotryptic activity. Compared to the chymotryptic activity of bovine and *Rhodococcus* proteasomes, the chymotryptic activity of the Mtb proteasome was strikingly restricted to substrates with P1 = W. Moreover, P1 and P3 amino acids appeared to be the major determinants of substrate preference for the bovine and *Rhodococcus* proteasomes, whereas for the Mtb proteasome, only P1 had a major effect. The Mtb proteasome’s strong preference for P1 = W was reflected in the potency of bortezomib P1-variant analogues. As P1 amino acid analogues changed from Ala to m-Cl-Phe, the ratio of IC$_{50}$ values between the human proteasome and the Mtb proteasome changed from 0.001 (5) to 8.0 (16), a remarkable 8000-fold shift in favor of the Mtb proteasome. However, the modest 8-fold selectivity for compound 16 in favor of Mtb proteasome over human proteasome suggests much work need to be done to improve the selectivity. Incorporating the selective preference by Mtb proteasome for side chains at S1 and S3 sites, together with other proven proteasome or protease warheads may further improve the species selectivity toward the Mtb proteasome. Incorporating non-natural amino acids into the inhibitor design would also provide more avenues
for improving species selectivity of proteasome inhibitors.

In sum, we discovered differences in substrate preferences, along with the differential influence of substituents at different positions in the substrate, through the use of a ~6,000 acetyl tripeptide AMC substrate library, suggesting that the primary and extended sites at the active center of the Mtb proteasome are very different than those of mammalian proteasomes. A limitation of this study is that substrate screening was performed with cow rather than human proteasomes. However, the studies with bortezomib analogs involved human proteasomes and were confirmatory. Thus, the findings reported here may provide opportunities to design peptide-based inhibitors with a useful degree of selectivity for the Mtb proteasome over that of the human host.

REFERENCES

1. Raviglione, M. C., and Smith, I. M. (2007) N Engl J Med 356, 656-659
2. Duncan, K., and Barry, C. E., 3rd. (2004) Curr Opin Microbiol 7, 460-465
3. Ginsberg, A. M., and Spigelman, M. (2007) Nat Med 13, 290-294
4. Spigelman, M. K. (2007) J Infect Dis 196 Suppl 1, S28-34
5. McCune, R. M., Feldmann, F. M., Lambert, H. P., and McDermott, W. (1966) J Exp Med 123, 445-468
6. McCune, R. M., Feldmann, F. M., and McDermott, W. (1966) J Exp Med 123, 469-486
7. Levin, B. R., and Rozen, D. E. (2006) Nat Rev Microbiol 4, 556-562
8. Warner, D. F., and Mizrahi, V. (2006) Clin Microbiol Rev 19, 558-570
9. Goldberg, A. L., Gaczynska, M., Grant, E., Michalek, M., and Rock, K. L. (1995) Cold Spring Harb Symp Quant Biol 60, 479-490
10. Goldberg, A. L. (2007) Biochem Soc Trans 35, 12-17
11. Baumeister, W., Walz, J., Zuhl, F., and Seemuller, E. (1998) Cell 92, 367-380
12. Darwin, K. H., Ehrt, S., Gutierrez-Ramos, J. C., Weich, N., and Nathan, C. F. (2003) Science 302, 1963-1966
13. Gandotra, S., Schnappinger, D., Monteleone, M., Hillen, W., and Ehrt, S. (2007) Nat Med 13, 1515-1520
14. Lin, G., Hu, G., Tsu, C., Kunes, Y. Z., Li, H., Dick, L., Parsons, T., Li, P., Chen, Z., Zwickl, P., Weich, N., and Nathan, C. (2006) Mol Microbiol 59, 1405-1416
15. Hu, G., Lin, G., Wang, M., Dick, L., Xu, R. M., Nathan, C., and Li, H. (2006) Mol Microbiol 59, 1417-1428
16. Adams, J., Behnke, M., Chen, S., Cruickshank, A. A., Dick, L. R., Grenier, L., Klunder, J. M., Ma, Y. T., Plamondon, L., and Stein, R. L. (1998) Bioorg Med Chem Lett 8, 333-338
17. Adams, J., Palombella, V. J., Sausville, E. A., Johnson, J., Destree, A., Lazarus, D. D., Maas, J., Pien, C. S., Prakash, S., and Elliott, P. J. (1999) Cancer Res 59, 2615-2622
18. Kropff, M., Bisping, G., Schuck, E., Liebisch, P., Lang, N., Hentrlich, M., Dechow, T., Kroger, N., Salwender, H., Metzner, B., Sezer, O., Engelhardt, M., Wolf, H. H., Einsele, H., Volpert, S., Heinecke, A., Berdel, W. E., and Kienast, J. (2007) Br J Haematol 138, 330-337
19. Powers, J. C., Asgian, J. L., Ekici, O. D., and James, K. E. (2002) Chem Rev 102, 4639-4750
20. Lowe, J., Stock, D., Jap, B., Zwickl, P., Baumeister, W., and Huber, R. (1995) Science 268, 533-539
21. Orlowski, M., Cardozo, C., and Michaud, C. (1993) Biochemistry 32, 1563-1572
22. Harris, J. L., Alper, P. B., Li, J., Rechsteiner, M., and Backes, B. J. (2001) Chem Biol 8, 1131-1141
23. Nazif, T., and Bogyo, M. (2001) Proc Natl Acad Sci U S A 98, 2967-2972
24. Wang, C. C., Bozdech, Z., Liu, C. L., Shipway, A., Backes, B. J., Harris, J. L., and Bogyo, M. (2003) J Biol Chem 278, 15800-15808
25. Groll, M., Nazif, T., Huber, R., and Bogyo, M. (2002) Chem Biol 9, 655-662
26. McGuire, M. J., and DeMartino, G. N. (1989) Biochem Biophys Res Commun 160, 911-916
27. Song, X., Mott, J. D., von Kampen, J., Pramanik, B., Tanaka, K., Slaughter, C. A., and DeMartino, G. N. (1996) J Biol Chem 271, 26410-26417
28. Song, X., von Kampen, J., Slaughter, C. A., and DeMartino, G. N. (1997) J Biol Chem 272, 27994-28000
29. Shepeck, J. E., 2nd, Kar, H., Gosink, L., Wheatley, J. B., Gjerstad, E., Loftus, S. M., Zubiria, A. R., and Janc, J. W. (2000) Bioorg Med Chem Lett 10, 2639-2642
30. Stein, R. L., Melandri, F., and Dick, L. (1996) Biochemistry 35, 3899-3908
31. Harris, J. L., Backes, B. J., Leonetti, F., Mahrus, S., Ellman, J. A., and Craik, C. S. (2000) Proc Natl Acad Sci U S A 97, 7754-7759

ACKNOWLEDGEMENT

We thank the MPI Chemistry Department for bortezomib analogs and P. Zwickl (Max Planck Institute, Martinsreid, Germany) for an expression plasmid for Rhod20S. Supported by NIH PO1-AI06293 and the Milstein Program in Chemical Biology of Infectious Diseases. Dr. Zhou's contribution to this project is supported in part by NIH Clinical and Translational Science Award to WCMC (UL1-RR024996). The Department of Microbiology and Immunology is supported by the William Randolph Hearst Foundation. Dr. Christopher Tsu, and Dr. Lawrence Dick were employees of Millennium Pharmaceuticals Inc.

FOOTNOTE

The abbreviations used: 7-Amino-4-methylcoumarin, AMC; proteasome, 20S; bovine 20S, bov20S; IC₅₀, concentration required for 50% inhibition; Mtb PrcBA wild type, Mtb20SWT; Mtb PrcBA open-gate mutant, Mtb20SOG; Rhodococcus erythropolis 20S, Rhod20S; benzyloxycarbony, Z; bortezomib, pyrazinylcarbonyl-Phe-Leu-boronic acid. Single-letter codes are used for amino acids, including B for L-citrulline and O for L-ornithine.

LEGENDS

Fig. 1. Schematic illustration of enzymes, substrates, and assay system. (a) Different composition of β subunits in eubacterial and eukaryotic proteasomes. The Mtb proteasome has 7 identical β subunits, while eukaryotic proteasomes have 7 different β subunits. Only three β subunits of eukaryotic proteasomes are active once their Thr1 active sites are exposed by autocatalytic removal of propeptide: β1 – caspase-like,
β2 – trypsin-like, β5 – chymotrypsin-like. Eubacterial proteasomes are generally considered to be chymotrypsin-like when assayed with small peptide substrates. (b) Structure of the acetyl-P3-P2-P1-AMC substrate library. R1, R2, R3 refer to the side chains of P1, P2, P3 amino acids, respectively; and S1, S2, S3 refer to the binding pockets of the proteasome for P1, P2, P3 amino acid side chains, respectively. (c) Schematic illustration of microfluidic TF460 assay system. Stream splitting and simultaneous fluorescence detection facilitated subtraction of the background fluorescence of each unhydrolyzed substrate; or After substrate in buffer was pulled into the channel through vacuum, the stream was split 50:50 into parallel channels, and mixed with enzyme and buffer, respectively. The final annotated data were obtained by subtracting background fluorescence of the unhydrolyzed substrate from that of the enzymatic reaction.

**Fig. 2.** Activities of different species’ proteasomes for individual peptides organized according to P1 (alphabetical order; color-coded) and within each set with the same P1, by P2. Each dot represents one substrate that is color-coded by its P1 amino acid. (a) Mtb20SWT; (b) Mtb20SOG; (c) Rhod20S; (d) bov20S; B, citruline; O, ornithine. The labeled substrates are all in the order (from left to right) of P3-P2-P1 in sequence, and color-coded as green with black border.

**Table 1** Preferred amino acids at given positions for Mtb20SOG, Rhod20S, and bov20S.

**Table 2** Kinetic constants for bov20S, Mtb20SWT, Mtb20SOG, on individual tripeptide AMC substrates

| Substrate | kcat (s⁻¹) | Km (mM) | IC50 (μM) |
|-----------|------------|---------|-----------|
| a LLVY, Suc-LLVY-AMC; b Z-VLR, Z-VLR-AMC; c Kcat and kcat values could not be estimated as saturation not reached due either to substrate precipitation or substrate inhibition at high concentration. Their Kcat/kcat values were estimated from the linear plot of the data; d data from Ref. 14. ND, no detectable activities; NA, not available. | | | |

**Table 3** IC50 values for bortezomib P1 amino acid analogues against Mtb20SOG and human 20S.
Fig. 1

(a) 

- **bacterial**
- **αβ**
- **β1 caspase-like**
- **β2 trypsin-like**
- **β5 chymotrypsin-like**
- **β chymotrypsin-like**

(b) 

- **S2**
- **R1**
- **R2**
- **R3**
- **P1**
- **P2**
- **P3**
- **S1**
- **S3**
- **AMC**

(c) 

- **Buffer control**
- **Substrate**
- **Proteasome + buffer**
- **Detection window**
- **λ_{ex} 420-480 nm**
- **λ_{em} 355 nm**
- **Waste**
Figure 2a
Figure 2b
Figure 2c
Figure 2d

RFU (Bov20S PA28α)
Figure 3
Figure 4

(a) RFU vs. YVNTSQRONLMKHGFEDAXAG for Ac-G-Xaa-W-AMC

(b) RFU vs. WYVTSRQONLKHGFEDAXA for Ac-P-Xaa-W-AMC

(c) RFU vs. YVTSPQRONMNHGFEDA for Ac-R-Xaa-W-AMC

(d) RFU vs. WYVTSRQONLKHGFEDA for Ac-K-Xaa-W-AMC
|          | P3                                      | P2                                      | P1                                      |
|----------|-----------------------------------------|-----------------------------------------|-----------------------------------------|
| Mtb20SOG | aromatic ≈ hydrophobic ≈ neutral ≈ basic | Q ≈ W » O ≈ T ≈ S ≈ F                  | W » H > Y > F > R                      |
| Rhod20S  | Y > L ≈ W ≈ F                           | non-defined                             | L > F ≈ I > W                          |
| Bov20S   | Y > W > F > L > R                       | non-aromatic                            | L » R ≈ F ≈ H > M                      |
Table 2

| Substrate | Ac-(P3P2P1)-AMC | Bov20S | Mtb20SOG | Ratio |
|-----------|----------------|--------|-----------|-------|
|           | $k_{cat}$ (min)$^{-1}$ | $K_M$ (mM) | $k_{cat}/K_M$ (min/mM) | $k_{cat}$ (min)$^{-1}$ | $K_M$ (mM) | $k_{cat}/K_M$ (min/mM) | $k_{cat}/K_M$(Mtb20SOG) | $k_{cat}/K_M$(Bov20S) |
| KQL       | 135             | 21     | 6.43      | ND    | ND    | ND    | 0.01$^c$ | 0.001 |
| TFL       | 130             | 53     | 2.45      | ND    | ND    | ND    | ND    | NA   |
| NYL       | 180             | 26     | 6.92      | ND    | ND    | ND    | ND    | NA   |
| YQL       | 490             | 47     | 10.43     | -     | -     | 0.01$^c$ | 0.04$^c$ | 0.05  |
| WLA       | 150             | 28     | 5.36      | ND    | ND    | ND    | ND    | NA   |
| WAV       | 51              | 110    | 0.46      | ND    | ND    | ND    | ND    | NA   |
| YWI       | -               | -      | 0.78$^c$  | -     | -     | 0.04$^c$ | 0.04$^c$ | 0.05  |
| KQY       | 75              | 37     | 2.03      | ND    | ND    | ND    | ND    | NA   |
| LLVY$^a$  | 160             | 30     | 5.33      | 6.5   | 80    | 0.08  | 0.015$^d$ |
| YGF       | 470             | 180    | 2.61      | ND    | ND    | ND    | ND    | NA   |
| LWW       | -               | -      | 0.09$^c$  | 5     | 9.1   | 0.55  | 6.1    |
| YQW       | 220             | 43     | 5.12      | -     | -     | 0.66$^c$ | 0.13  |
| NTW       | 57              | 37     | 1.54      | 61    | 82    | 0.74  | 0.48   |
| RAW       | -               | -      | 0.26$^c$  | 77    | 98    | 0.79  | 3.04   |
| RFW       | 10.6            | 16.7   | 0.63      | 115   | 80    | 1.40  | 2.22   |
| WQW       | -               | -      | 3.70$^c$  | -     | -     | 1.25$^c$ | 0.34  |
| Z-VLR$^b$ | 99              | 29     | 3.41      | -     | -     | 0.42$^c,d$ | 0.12  |
| RQR       | 280             | 19     | 14.74     | ND    | ND    | ND    | ND    | NA   |
| OWE       | 38              | 85     | 0.45      | ND    | ND    | ND    | ND    | NA   |
| AWE       | 86              | 50     | 1.72      | ND    | ND    | ND    | ND    | NA   |
| LLE       | 110             | 39     | 2.82      | ND    | ND    | ND    | ND    | NA   |
| R         | IC\textsubscript{50} (µM) | Ratio |  |
|-----------|-------------------|-------|---|
|           | Mtb20SOG (Ac-RFW-AMC) | Mtb-20SOG (Z-VLR-AMC) | Human 20S (Ac-WLA-AMC) | IC\textsubscript{50}(h20S) / IC\textsubscript{50}(Mtb20SOG) |
| Bortezomib | 0.29              | 0.28  | 0.016 | 0.055 |
| 2         | > 100             | > 100 | > 100 | N/A   |
| 3         | 8.12              | 10.24 | 0.024 | 0.003 |
| 4         | 1.88              | 2.24  | 0.023 | 0.012 |
| 5         | 44.60             | 28.20 | 0.048 | 0.001 |
| 6         | 3.56              | 4.88  | 0.027 | 0.008 |
| 7         | 1.11              | 1.66  | 0.026 | 0.023 |
| 8         | 0.80              | 1.04  | 0.049 | 0.061 |
| 9         | 1.33              | 2.08  | 0.023 | 0.017 |
| 10        | 1.47              | 2.11  | 0.16  | 0.11  |
| 11        | 0.67              | 0.51  | 3.5   | 5.22  |
| 12        | 0.13              | 0.13  | 0.48  | 3.69  |
| 13        | 0.24              | 0.15  | 0.95  | 3.96  |
| 14        | 0.27              | 0.25  | 0.61  | 2.26  |
| 15        | 0.12              | 0.14  | 0.40  | 3.33  |
| 16        | 0.15              | 0.13  | 1.2   | 8.00  |
| 17        | 0.14              | 0.14  | 0.93  | 6.64  |
| 18        | 0.11              | 0.11  | 0.73  | 6.64  |
Distinct specificities of mycobacterium tuberculosis and mammalian proteasomes for N-acetyl tripeptide substrates
Gang Lin, Christopher Tsu, Lawrence Dick, Xi K. Zhou and Carl Nathan

J. Biol. Chem. published online October 1, 2008

Access the most updated version of this article at doi: 10.1074/jbc.M805324200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC’s e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2008/10/02/M805324200.DC1