

## Design of Noncovalent Inhibitors of Human Cathepsin L. From the 96-Residue Proregion to Optimized Tripeptides

Shafinaz F. Chowdhury, J. Sivaraman, Jing Wang,<sup>†</sup> Gopal Devanathan, Paule Lachance, Hongtao Qi, Robert Ménard, Jean Lefebvre, Yasuo Konishi, Mirosław Cygler, Traian Sulea, and Enrico O. Purisima\*

Biotechnology Research Institute, National Research Council of Canada, 6100 Royalmount Avenue, Montreal, Quebec, H4P 2R2, Canada

Received May 31, 2002

A novel series of noncovalent inhibitors of cathepsin L have been designed to mimic the mode of autoinhibition of procathepsin L. Just like the propeptide, these peptide-based inhibitors have a reverse-binding mode relative to a substrate and span both the S' and S subsites of the enzyme active site. In contrast to previous studies in which even moderate truncation of the full-length propeptide led to rapid reduction in potency, these blocked tripeptide-sized inhibitors maintain nanomolar potency. Moreover, these short peptides show higher selectivity (up to 310-fold) for inhibiting cathepsin L over K versus only 2-fold selectivity of the 96-residue propeptide of cathepsin L. A 1.9 Å X-ray crystallographic structure of the complex of cathepsin L with one of the inhibitors confirms the designed reverse-binding mode of the inhibitor as well as its noncovalent nature. Enzymatic analysis also shows the inhibitors to be resistant to hydrolysis at elevated concentrations of the enzyme. The mode of inhibition of these molecules provides a general strategy for inhibiting other cathepsins as well as other proteases.

The increasing evidence supporting the role of cathepsins, cysteine proteases of the papain superfamily, in a number of specific physiological and pathological processes has regenerated the interest for this group of enzymes as promising therapeutic targets. Originally considered to mainly play a role in general protein degradation in lysosomes,<sup>1</sup> cathepsins have been found to have important physiological roles in bone resorption and remodeling,<sup>2,3</sup> thyroid hormone liberation,<sup>4</sup> and immune response processes.<sup>5</sup> They have also been implicated in a number of degenerative processes that include osteoporosis, rheumatoid arthritis,<sup>6</sup> emphysema, and muscular dystrophy.<sup>7</sup> In addition, there is strong evidence for the participation of cathepsins in cancer, where they have been shown to act extracellularly in tumor invasion<sup>8</sup> and metastasis<sup>9,10</sup> as well as intracellularly.<sup>11</sup> These enzymes, and in particular, cathepsin L, are therefore attractive therapeutic targets for the development of anticancer agents.

Most small-molecule inhibitors of cathepsins carry a reactive group that covalently modifies the enzyme.<sup>12–15</sup> Far less common are noncovalent inhibitors. However, cathepsin L itself, in its proenzyme form, provides an example of an inhibition mode that does not rely on covalent modification of the active site.<sup>16,17</sup> Similar to other cysteine proteases, cathepsin L is synthesized as an inactive proenzyme containing an autoinhibitory 96-residue N-terminal propeptide. Inhibition is accomplished by having part of the propeptide span the active site, thereby blocking access to it. The propeptide binds to the active site with a backbone direction reverse to that of a normal peptide substrate, making it resistant to hydrolysis by the enzyme. Activation of the pro-

enzyme to mature functional cathepsin L requires removal of the propeptide. The independent propeptide fragment is a potent inhibitor of mature cathepsin L. A synthetic peptide (Phe<sup>4p</sup>-Gln<sup>90p</sup>) consisting of 87 of the 96 residues of the propeptide sequence has an inhibition constant,  $K_i$ , of 0.088 nM against cathepsin L. However, successive truncation of the propeptide sequence results in a dramatic reduction in potency.<sup>18</sup> The fragments Arg<sup>21p</sup>-Tyr<sup>95p</sup> and Gly<sup>52p</sup>-Tyr<sup>95p</sup> have  $K_i$  values of 11.5 and 2900 nM, respectively. Those truncation studies suggest that the binding affinity of these propeptides arises from the extensive contact with the enzyme as well as the maintenance of the appropriate three-dimensional structure. It is therefore a major challenge and a nontrivial task to construct low molecular weight molecules that utilize the propeptide inhibition mode.

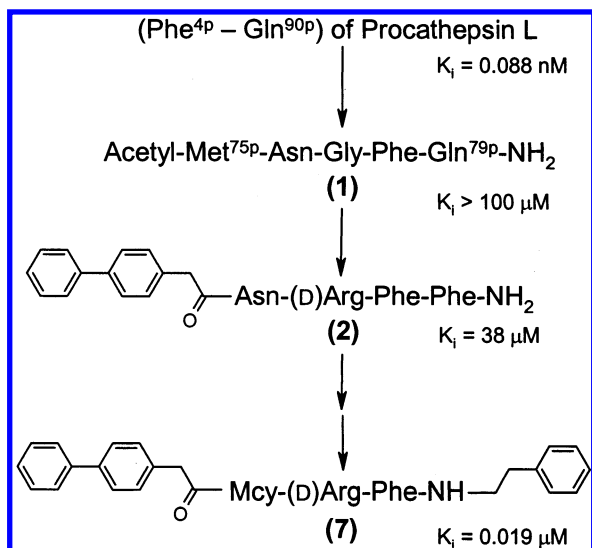
In this work, we describe the design and synthesis of short peptide-based molecules with nanomolar inhibitory potency that mimic the propeptide reverse-binding mode and span from the S2' to S3 subsites of cathepsin L. Making use of both the primed and unprimed subsites is expected to improve potency and specificity. The reverse-binding mode and its noncovalent nature were confirmed through the determination of a 1.9 Å resolution crystal structure of cathepsin L complexed with one of these inhibitors.

### Results and Discussion

**Structure-Based Design of a Starting Inhibitor for Further Optimization.** The crystal structure of the human procathepsin L was used as starting point for inhibitor design (Figure 1). We focused on the five-residue stretch of the propeptide, Met<sup>75p</sup>-Asn-Gly-Phe-Gln<sup>79p</sup>, that spans from subsite S2' to S3 of the mature enzyme in the reverse-substrate-binding mode. We synthesized the capped pentapeptide, compound **1**, based on this native sequence and found no inhibition

\* To whom correspondence should be addressed. Phone: (514) 496-6343. Fax: (514) 496-5143. E-mail: rico@bri.nrc.ca.

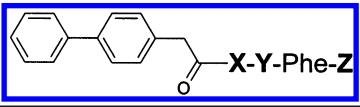
<sup>†</sup> Current address: Structural Bioinformatics, 10929 Technology Place, San Diego, CA 92127.



**Figure 1.** Design path for the minimization of the full-length propeptide to a capped tripeptide.

( $K_i > 100 \mu\text{M}$ ) of mature cathepsin L. This result was anticipated. Without the anchoring residues that flank this fragment in the full-length propeptide, peptide **1** may not have sufficiently strong interactions with the protein to lock it into the reverse-binding mode. Furthermore, it could even adopt several substrate-like binding modes in the active site cleft and thus be readily degraded by mature cathepsin L. Therefore, a number of chemical modifications of this starting compound were modeled in the binding site of the enzyme, aimed at enhancing binding affinity in the reverse-substrate-binding mode and increasing the proteolytic stability by hindering the substrate-like binding mode. Modifications at three sites were carried out. The N-terminal acetyl-Met fragment of **1** was replaced by a 4-biphenyl-acetyl group. Molecular docking experiments show that the rigid, aromatic structure of the 4-biphenyl substituent could enhance the nonpolar interactions in the primed subsites relative to the acetyl-Met fragment. Similarly, the C-terminal Gln was replaced with Phe, which fits well into the S3 subsite of the enzyme. The third modification was the substitution of Gly in the middle of the peptide with a D-Arg. A D- rather than L-amino acid was selected for this position because the designed reverse-binding mode of our inhibitory peptide alters the orientation of the  $C\beta$  carbon, making a D-amino acid sterically favored over an L-amino acid whose  $C\beta$  carbon would clash with Gly23 of the protein. Molecular docking calculations suggested D-Arg as one of the best amino acid replacements for Gly. Moreover, incorporation of a D-amino acid at this position has the added advantage of potentially improving proteolytic stability. Docking of the peptide in the substrate-binding mode results in a collision of a large side chain of a D-amino acid with the protein atoms, regardless whether the D-amino acid occupies the S2, S1, or S1' subsite. Hence, the D-Arg substitution would make any of the three C-terminal peptide bonds resistant to cleavage. The remaining putative cleavage site before Asn would not be favored because in the substrate-like binding mode, the S2 subsite of the enzyme, the most important determinant of binding affinity and specificity, would not be occupied by the peptide. Furthermore, a biphen-

**Table 1.** Optimization of Peptide **2** for Cathepsin L Inhibition

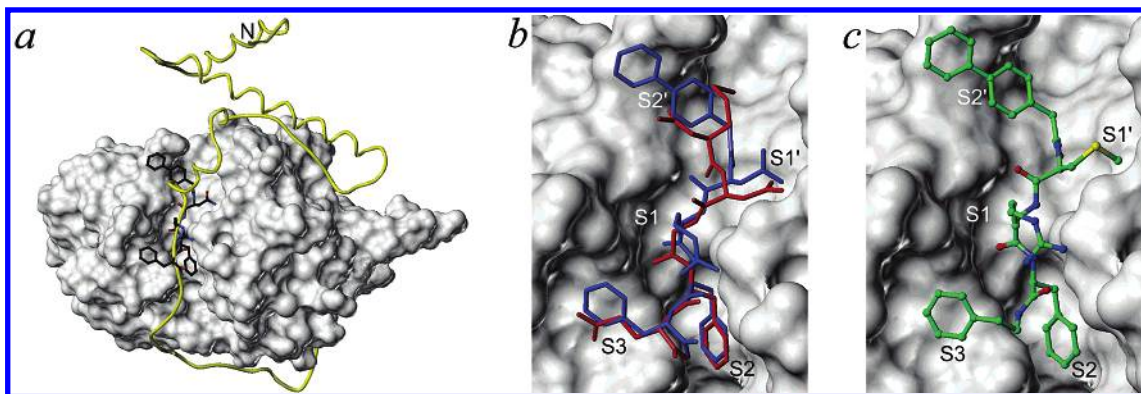
				
compd	X	Y	Z	$K_i$ ( $\mu\text{M}$ )
<b>2</b>	Asn	D-Arg	Phe-NH <sub>2</sub>	38
<b>3</b>	Ser	D-Arg	Phe-NH <sub>2</sub>	74
<b>4</b>	Cys	D-Arg	Phe-NH <sub>2</sub>	0.17
<b>5</b>	Mcy	D-Arg	Phe-NH <sub>2</sub>	0.070
<b>6</b>	Cys	D-Arg	N-(2-phenylethyl)amide	0.021
<b>7</b>	Mcy	D-Arg	N-(2-phenylethyl)amide	0.019
<b>8</b>	Mcy	D-Arg	N-(3-phenylpropyl)amide	0.21
<b>9</b>	Mcy	D-Arg	N-(benzyl)amide	6.6
<b>10</b>	Nva	D-Arg	N-(2-phenylethyl)amide	0.49
<b>11</b>	Mcy	D-Orn	N-(2-phenylethyl)amide	0.2
<b>12</b>	Mcy	Gly	Phe-NH <sub>2</sub>	0.93

ylacetyl group in S1 would not be well accommodated. Compound **2** was synthesized and inhibits cathepsin L with a  $K_i$  of  $38 \mu\text{M}$ .

**Predicted Binding Mode of the Propeptide-Based Inhibitors.** The modeled structures of **1** and **2** docked into the cathepsin L binding site are shown in parts a and b of Figure 2. Several general conclusions can be drawn from the model. (i) The side chains of these reverse-substrate-binding mode peptides are oriented to potentially utilize the canonical binding subsites of the enzyme, consistent with available data from the proenzyme structures.<sup>17</sup> (ii) The backbone conformation of the inhibitors (interacting with subsites S2 to S1') preserves the intermolecular hydrogen bonding interactions established by the template propeptide. (iii) The aliphatic portion of the side chain of a D-Arg residue at the central position of the tripeptide core is well-accommodated in the S1 subsite and has its guanidinium group largely solvent-exposed. Introduction at this position of an L-amino acid with the same backbone conformation would be sterically disfavored, since its side chain would collide with protein atoms. (iv) Good nonpolar interactions in the S3 and S2' subsites of cathepsin L can be achieved by incorporating aromatic moieties appropriately linked to the peptide core. As mentioned earlier, the modifications leading to **2** are believed to help anchor the peptide in the inhibitory, reverse-binding mode as well as to hinder its accommodation in the cleavable, substrate-binding mode.

**Inhibitor Optimization.** According to the model, Asn, D-Arg, and Phe-NH<sub>2</sub> of **2** interact with the S1', S1, and S3 subsites of cathepsin L, respectively. These residues were targeted for inhibitor optimization. Table 1 shows the inhibition constants for a series of synthesized analogues. Replacement of Asn by Ser (compound **3**) resulted in a 2-fold decrease in activity. However, replacement of the same residue by Cys led to compound **4**, with an improved  $K_i$  of  $0.17 \mu\text{M}$ . To test whether the increased inhibition by **4** is due to a possible disulfide bond formation, we synthesized compound **5**, which contains an S-methylated cysteine (Mcy). This compound proved to be an even better inhibitor of cathepsin L, with a  $K_i$  of  $0.07 \mu\text{M}$ . Thus, simply by replacing Asn in the starting inhibitor **2** by Cys or Mcy, we obtained improvements in the inhibition constant of 220- or 540-fold, respectively.

The contribution of the C-terminal amide group to the binding affinity of the two best inhibitors identified thus



**Figure 2.** Modeled structures for the molecules in Figure 1: (a) **2** in the active site of mature cathepsin L displayed as a Connolly surface, where the full-length propeptide is depicted as a yellow tube; (b) expanded view of **1** (red) and **2** (blue) in the active site; (c) binding mode of the best inhibitor, **7**.

far, **4** and **5**, was determined by replacing Phe-NH<sub>2</sub> with *N*-(2-phenylethyl)amide. The tripeptides **6** and **7**, which lack the C-terminal CONH<sub>2</sub> group, inhibited cathepsin L with *K<sub>i</sub>* values of 0.021 and 0.019 μM, respectively, which are 8- and 4-fold lower than those of the corresponding inhibitors **4** and **5**. Changing the aliphatic carbon linkers between the phenyl ring and the amide group of the C-terminal blocking group (three-carbon linker in **8** and one-carbon linker in **9**) had an adverse effect on the inhibition constant, suggesting that a two-carbon linker has an optimal length for binding in the S3 pocket. The modeled complex of inhibitor **7** depicting the interactions with subsites S3 and S2' is shown in Figure 2c. This theoretical structural prediction was later confirmed by the crystal structure of cathepsin L complexed with a potent inhibitor analogue (see below).

Comparison of the activities of **3** and **4** shows that the sulfur atom, predicted to interact with the S1' subsite of cathepsin L, contributes significantly to the potency of the inhibitor. To further verify this, we synthesized compound **10**, an analogue of **7** containing norvaline (Nva) in place of the Mcy, thus replacing the sulfur atom by a methylene group. This led to a 26-fold increase in inhibition constant, confirming the importance of introducing a sulfur atom at this position of the inhibitor.

The contribution of the D-Arg side chain to the binding affinity was also tested. An analogue of **7** was made in which D-Arg was replaced by d-Orn (**11**). Although both side chains are positively charged in the assay conditions (pH 5.5), the D-Orn-containing compound **11** inhibited cathepsin L with a *K<sub>i</sub>* of 0.2 μM, which is 11-fold higher than that of the corresponding D-Arg-containing inhibitor **7**. Interestingly, complete removal of the D-Arg side chain has a similar effect, a 13-fold increase in *K<sub>i</sub>* (compare **12** and **5**).

The two best inhibitors, **6** and **7**, both contain a phenylalanine side chain predicted to interact with the hydrophobic S2 subsite of cathepsin L in the reverse-substrate-binding mode. We tested a series of analogues of **6** and **7** in which Phe was substituted by various natural and unnatural residues. The results are presented in Table 2. Although none of the analogues displayed better inhibition of cathepsin L compared to the parent compounds, these data provide an informative structure–activity profile. Replacement of Phe in **6** by Tyr (**13**) or Trp (**14**) resulted in only a 2- or 3-fold increase in *K<sub>i</sub>*, respectively. However, the use of Tyr

**Table 2.** Structure–Activity Relationship at Phe Position of the Tripeptide Inhibitors **6** and **7** of Cathepsin L

compd	X	Y	<i>K<sub>i</sub></i> (μM)
<b>6</b>	Cys	Phe	0.021
<b>7</b>	Mcy	Phe	0.019
<b>13</b>	Cys	Tyr	0.045
<b>14</b>	Cys	Trp	0.067
<b>15</b>	Cys	Cha	0.16
<b>16</b>	Cys	Npa	0.16
<b>17</b>	Cys	Bpa	0.27
<b>18</b>	Cys	Abu	0.21
<b>19</b>	Cys	Arg	3.9
<b>20</b>	Mcy	Leu	0.27
<b>21</b>	Mcy	Met	0.24

instead of Phe improves the solubility of the peptide. It was **13** that subsequently led to successful cocrystallization with cathepsin L for X-ray structure determination. Replacement of Phe by a variety of other nonpolar groups (compounds **15**–**18** and **20**–**21**) gave a somewhat larger change in *K<sub>i</sub>* but still within a narrow range from 0.16 to 0.27 μM (an 8- to 14-fold increase). In contrast, analogue **19**, which contains a charged Arg side chain instead of Phe in **6**, has a *K<sub>i</sub>* of 3.9 μM (190-fold increase relative to **6**). These results indicate a requirement for a hydrophobic residue at this position of the inhibitor but with no strong preference beyond that. This is in agreement with the predicted binding of this residue in the hydrophobic S2 pocket of cathepsin L.

Finally, the contribution of the moieties located at the ends of the scaffold were tested using one of the best inhibitors, compound **6**. Removal of one of the two aromatic rings of the 4-biphenylacetyl N-terminal blocking group (**22**) resulted in a significant loss of cathepsin L inhibition (680-fold increase in *K<sub>i</sub>*). This confirms our predictions based on the inhibitor docking simulations, which motivated the incorporation of the 4-biphenylacetyl blocking group. Deletion of the Phe residue resulted in the dipeptide **23**, which inhibited cathepsin L weakly with a *K<sub>i</sub>* of only 81 μM. The decrease in activity can be explained by the loss of interactions with the S3 subsite as well as the loss of two hydrogen bonds with the protein.

**Proteolytic Stability and Selectivity.** Three of the best inhibitors (**7**, **13**, and **14**) were incubated with cathepsin L (10 nM) for 4 h at 28 °C, and the mixtures



**Table 3.** Specificity of Cathepsin L, K, and B Inhibition by Compounds **7**, **13**, and **14**<sup>a</sup>

compd	<i>K<sub>i</sub></i> (μM)		
	cathepsin L	cathepsin K	cathepsin B
<b>7</b>	0.019	5.9 (310)	4.1 (210)
<b>13</b>	0.045	2.9 (64)	15 (330)
<b>14</b>	0.067	1.0 (15)	0.56 (9)

<sup>a</sup> In parentheses for cathepsin K and B is the ratio of *K<sub>i</sub>* relative to that of cathepsin L for a given inhibitor.

were analyzed by reverse-phase HPLC for the stability of the inhibitor. The inhibitors were completely resistant to proteolysis by cathepsin L, since there was no sign of their cleavage under these standard assay conditions (see Supporting Information). This confirms that the designed peptides do not bind to cathepsin L in a substrate-like binding mode at least for those compounds that show good potency.

The selectivity of these three inhibitors was measured against cathepsins B, K, and L. The results are summarized in Table 3. Compound **7** shows a 310- and 210-fold selectivity for cathepsin L over cathepsins K and B, respectively. This selectivity for cathepsin L over K is remarkable considering that by comparison the full-length propeptide of cathepsin L shows only 2-fold selectivity for cathepsin L over K.<sup>19</sup> Thus, our low molecular weight peptides exhibit both potency and selectivity.

**Crystal Structure of Inhibitor 13 Complexed with Cathepsin L.** We determined the crystal structure of mature human cathepsin L in complex with **13** at 1.9 Å resolution. The inhibitor **13**, 4-biphenylacetyl-Cys-D-Arg-Tyr-*N*-(2-phenylethyl)amide, is a more water-soluble analogue of our best inhibitor **7** and exhibits a cathepsin L inhibition constant of 45 nM, which is only 2-fold higher than that of **7** (Table 2). The asymmetric unit of the crystal contains two cathepsin L–**13** complexes that adopt very similar conformations.

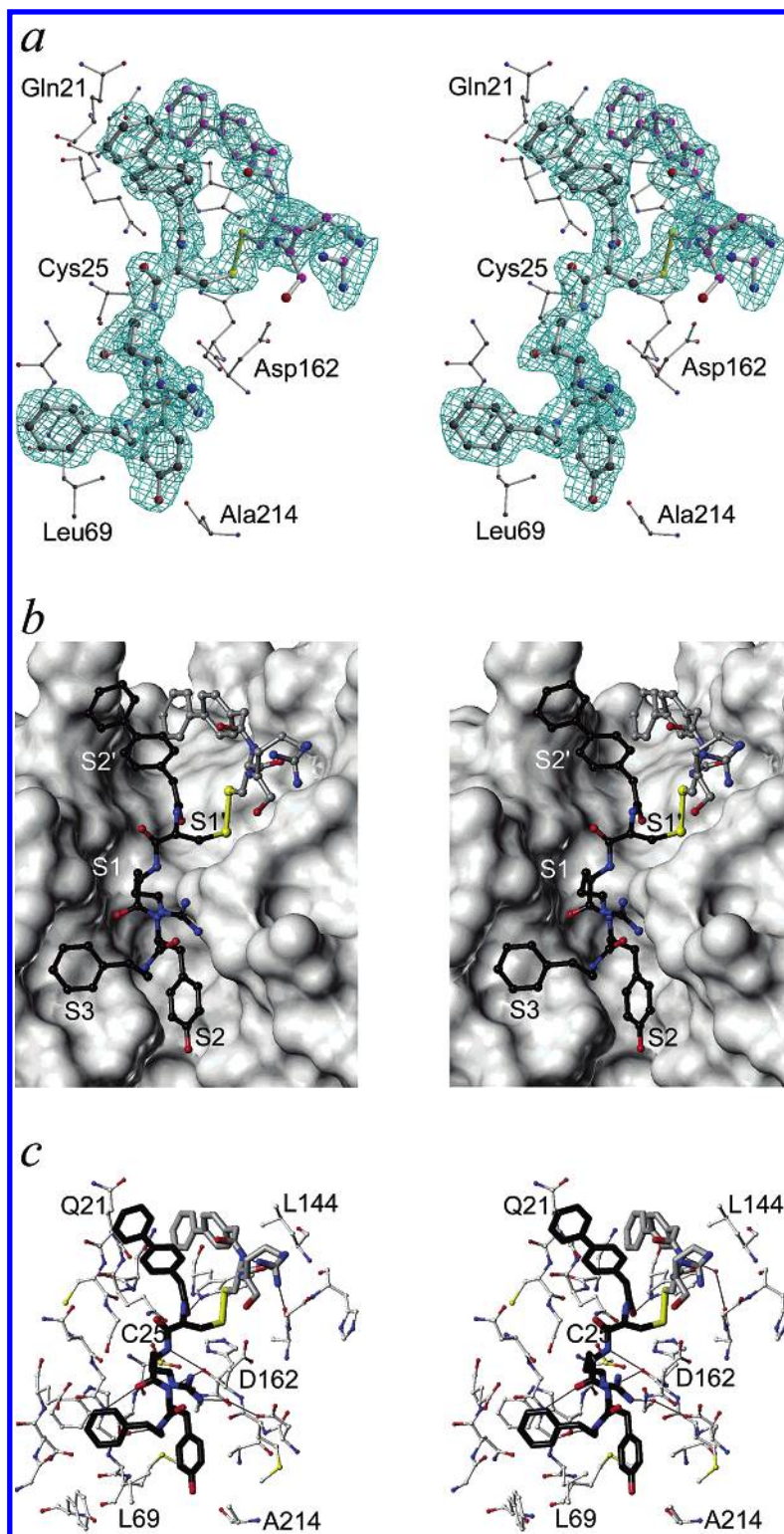
In the present complex, the structure of cathepsin L is very similar to that of the mature enzyme part as observed in the procathepsin L crystal structure at 1.8 Å resolution (PDB code 1cs8) with a root-mean-square deviation (rmsd) between equivalent Cα atoms of 0.45 Å. Thus, the procathepsin L coordinates used as a basis for our molecular modeling calculations provided a good approximation of the mature cathepsin L structure.

Figure 3a shows the  $F_o - F_c$  simulated annealing omit map for the active site region of the complex between cathepsin L and **13**. The crystal structure demonstrates that **13** inhibits human cathepsin L by binding directly into the substrate-binding site of the mature enzyme, in a binding mode that matches remarkably the one predicted by molecular modeling (see below). That is, the present experimental structure shows unambiguously that the designed blocked tripeptide **13** occupies both primed and nonprimed subsites of cathepsin L in the reverse-substrate-binding mode. Also, there is no evidence for covalent bond formation between the enzyme and inhibitor molecules. The electron density around the active site Cys is consistent with its sulfur oxidized to sulfinate (–SO<sub>2</sub><sup>–</sup>). It is worth noting that the catalytic Cys in the procathepsin L structure was also oxidized to sulfonate (–SO<sub>3</sub><sup>–</sup>) in that case.

The binding mode of **13** in cathepsin L is seen more clearly in Figure 3b, which shows how it fits into the

active site groove. The D-Arg and Phe residues and the *N*-(2-phenylethyl)amide blocking group of the inhibitor bind to the enzyme subsites S1, S2, and S3, respectively, whereas the Cys residue and the 4-biphenylacetyl blocking group of **13** are accommodated into S1' and S2' subsites, respectively. The backbone of the inhibitor establishes five direct intermolecular hydrogen bonds with the enzyme (Figure 3c). These include two hydrogen bonds between Gly68 and the *N*-(2-phenylethyl)-amide NH group and D-Arg backbone carbonyl of the inhibitor, two hydrogen bonds between the Asp162 main chain carbonyl and the inhibitor's D-Arg and Tyr backbone NH groups, and one hydrogen bond between the Trp189 indole NH group and the 4-biphenylacetyl-carbonyl of the inhibitor. The phenyl ring of the inhibitor's *N*-(2-phenylethyl)amide blocking group fits well into the relatively shallow S3 subsite and interacts with the side chains of Glu63, Leu69, Tyr72, with the CH<sub>2</sub> of Gly68 and the carbonyl of Gly61. It appears that the two-carbon aliphatic linker optimally positions the aromatic ring of the blocking group into the S3 subsite, while the linker itself does not make contact with the protein. The Tyr side chain of the inhibitor interacts extensively with the deep, hydrophobic S2 pocket of cathepsin L, making contact with the side chains of Leu69, Met70, Ala135, and Ala214. The D-Arg side chain of **13** is situated in the less-delineated S1 pocket with the guanidinium group being largely solvent-exposed. The inhibitor's Cys side chain interacts with the side chains of the S1' residues Ala138, Asp162, His163, and Trp189. Its sulfur atom is engaged in a disulfide bond with a second inhibitor molecule only half of which has clear electron density (see below). The 4-biphenylacetyl rings of the inhibitor pack against the S2' subsite, and more specifically, they interact with Gln21, Gly23, and the main chain atoms of Cys22. This conformation and the binding mode are made possible by the flexibility afforded by the methylene group between the biphenyl moiety and the carbonyl group.

The electron density allows fitting of part of a second inhibitor molecule into the primed region of the cathepsin L binding site (Figure 3a,b). This second inhibitor molecule is covalently attached through a disulfide bond to the first inhibitor molecule described above. The C-terminal part of the second inhibitor molecule comprising the Tyr residue and the *N*-(2-phenylethyl)amide blocking group could not be built in the observed electron density map. From the visible part, the Cys and D-Arg residues have the highest temperature factors when compared to the atoms of the protein and the first inhibitor molecule. This is consistent with the limited number of intermolecular contacts established by the second inhibitor molecule, with only the biphenyl moiety packing well between the side chains of Asn18, Ile144, Trp189 on one side and Gly20 and the biphenyl moiety of the first inhibitor molecule on the other side. Hence, the present structural data indicate a high degree of flexibility in the second inhibitor molecule as opposed to the extensive, specific interactions with the enzyme established by the first inhibitor molecule. Therefore, we propose that the formation of the inhibitor homodimer is not critical to achieve potent inhibition of cathepsin L and that only one inhibitor molecule bound in the designed propeptide-like binding mode is suf-

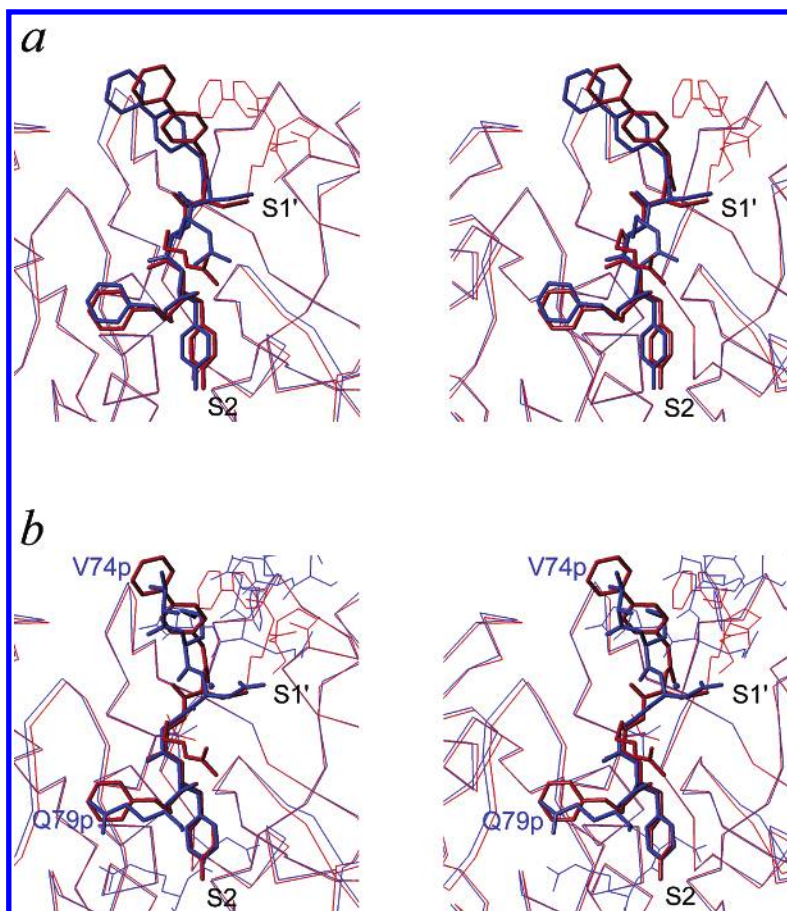


**Figure 3.** Crystal structure of **13** in complex with cathepsin L (stereoviews). (a) Simulated annealing  $F_o - F_c$  omit map shows electron density for the bound inhibitor. All atoms within 3 Å were omitted from refinement and map calculation. The map is contoured at  $2.0\sigma$ . Select cathepsin L residues are labeled to help orient the viewer. (b) Steric fit of the inhibitor in the active site groove of the enzyme which is displayed as a Connolly surface. Carbon atoms of the inhibitor dimer are shown in black for the first molecule and in gray for the second. (c) Detailed intermolecular interactions are shown. The inhibitor dimer is represented as thick capped sticks with the atom color scheme as indicated in Figure 3b. Cathepsin L residues 4.5 Å around the inhibitor dimer are displayed. Hydrogen bonds are indicated by thin black lines.

ficient to produce the observed cathepsin L inhibition. Most importantly, this hypothesis is fully supported by our biochemical data demonstrating equipotent inhibition obtained upon methylation of the inhibitor Cys residue, which precludes the formation of the disulfide-

bonded inhibitor dimer (e.g., compare analogues **4** and **5** and analogues **6** and **7**). It should be noted that mass spectral analysis of the inhibitors taken after their synthesis and also after the proteolytic stability assays did not show the presence of dimers at that stage.





**Figure 4.** Comparison of the crystallographic cathepsin L–**13** complex (red) with (a) the modeled cathepsin L–**13** complex (blue) and (b) the crystal structure of procatepsin L (blue, PDB code 1cs8) after alignment of the corresponding C $\alpha$  atoms of the proteins (stereoviews). The protein part is shown as C $\alpha$  trace. The first molecule of the cocrystallized inhibitor dimer and the active-site-spanning segment of procatepsin L propeptide (residues Val<sup>74p</sup> to Gln<sup>79p</sup>) are displayed as capped sticks. The second molecule of the cocrystallized inhibitor dimer and the propeptide residues flanking the active-site-spanning segment are shown as thin lines.

Nevertheless, the intermolecular interactions observed for the biphenyl moiety of the second inhibitor molecule, when combined with those established by the first inhibitor molecule in the primed region, constitute an interesting and certainly novel avenue for further structure-based ligand design strategies.

**Comparison of the Predicted and Crystal Structures.** The crystal structure of cathepsin L complexed with inhibitor **13** matches the theoretical predictions on the binding mode of our designed inhibitors. Figure 4a shows the cathepsin L–**13** complexes from X-ray crystallography and computational docking simulations after their superposition onto the equivalent C $\alpha$  atoms of the enzyme (rmsd of 0.46 Å). There is good agreement between the bound geometries of the cocrystallized and modeled inhibitor, with rmsd (all comparisons exclude hydrogen atoms) of 1.31 Å for all atoms and 0.40 Å for the backbone atoms alone. The largest deviations are due to the biphenyl ring atoms and D-Arg side chain with rmsd of 2.12 and 2.04 Å, respectively. The remaining inhibitor atoms overlay with an rmsd of only 0.50 Å. The crystal structure geometry of the bound inhibitor **13** can also be compared with that of the cathepsin L propeptide segment spanning the active site groove of the mature enzyme. The backbone conformations of the inhibitors and propeptide are very similar in the non-primed region, while they differ somewhat in the primed region (see Figure 4b). This is at least partly due to different oxidation states observed for the catalytic Cys

in the two structures, sulfonate in the procatepsin L crystal structure (PDB code 1cs8) and sulfinate in the present structure. In addition, the side chains and blocking groups of the tripeptide core backbone of **13** overlay well with the corresponding side chains of the cathepsin L propeptide, with the Tyr and Cys side chains of **13** occupying the positions of Phe<sup>78p</sup> and Asn<sup>76p</sup>. The phenylethyl moiety of the inhibitor superimposes onto Gln<sup>79p</sup> side chain, and its 4-biphenylmethyl moiety partially overlaps with the Val<sup>74p</sup> and Met<sup>75p</sup> side chains.

**Rational Design versus Truncation and Combinatorial Approaches.** As mentioned in the Introduction, truncation of the full-length propeptide of cathepsin L leads to rapid reduction in potency.<sup>18</sup> This is also observed in propeptide truncation studies applied to other cathepsins. For example, in a study of 10 overlapping 15-mer peptides derived from procatepsin B, the most potent sequence had a  $K_i$  of 4.6  $\mu$ M<sup>20</sup> compared to 0.4 nM for a 56-residue propeptide fragment.<sup>21</sup> Chen et al. carried out systematic truncation from either the N- or C-terminal ends of the same procatepsin B 56-residue peptide. They reported that removal of the first 25 residues or the last 15 residues resulted in  $K_i$  values of 9.7 and 83  $\mu$ M, respectively.<sup>22</sup> In another study on procongoain, again with overlapping 15-mer peptides, the most potent inhibitor had a  $K_i$  of 5.9  $\mu$ M against mature congoain.<sup>23</sup> Further reduction to a heptamer containing the core pentamer sequence blocking the

active site resulted in a  $K_i$  of 225  $\mu\text{M}$ .<sup>23</sup> It is clear from these studies that the native active site spanning sequences have only a weak binding affinity to the mature enzyme. Hence, truncation studies by themselves are unable to produce potent inhibitors. Here, by optimizing protein–ligand binding interactions, we have been able to produce potent low molecular weight peptides that mimic the propeptide mode of inhibition.

An alternative method of discovering short peptide inhibitors is through combinatorial peptide libraries. Brinker et al. synthesized an extensive pentapeptide library of more than 3 million peptides that systematically sampled substitutions at each position. The SAR developed from the study allowed them to define a set of inhibitors with submicromolar  $\text{IC}_{50}$  values.<sup>24</sup> The most potent of these had an  $\text{IC}_{50}$  of 0.5  $\mu\text{M}$  against cathepsin L. Meldal et al. constructed an octapeptide library with a D-amino acid in the fourth position from the N terminus.<sup>25</sup> The D-amino acid was incorporated in order to improve stability to hydrolysis by the enzyme. They tested the peptides against cruzipain, cathepsin B, and cathepsin L. The best peptides showed similar potencies among these three enzymes. The best inhibitor of cathepsin L had a  $K_i$  of 0.10  $\mu\text{M}$  with no selectivity for cathepsin L over K.

It is interesting to note that by using a more directed structure-based approach, the two dozen compounds designed and synthesized in this work yielded more potent and selective inhibitors of cathepsin L than did the combinatorial chemistry techniques. This success highlights the value of having and utilizing detailed structural information about the target enzyme and the binding mode of the inhibitor. The strategy used in this work should have broad applicability to the design of propeptide-based inhibitors for other enzymes.

## Experimental Section

**Synthesis.** All Fmoc-protected amino acids were purchased from Novobiochem (La Jolla, CA). 4-Biphenylacetic acid, 2-phenylethylamine, 3-phenylpropylamine, and benzylamine were obtained from Sigma-Aldrich (St. Louis, MO). Peptides were synthesized by Fmoc solid-phase chemistry using manual coupling (Fmoc-amino acid, 4 equiv; 2-(*H*-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium tetrafluoroborate (TBTU), 4 equiv; *N,N*-diisopropylethylamine (DIPEA), 6 equiv) in *N*-methylpyrrolidone. Completion of the reaction was verified by colorimetric ninhydrin assay (Kaiser test). The N terminus was deprotected, and peptides were blocked with corresponding acids by manual coupling (acids, 8 equiv; TBTU, 8 equiv; DIPEA, 12 equiv). For peptides with a C-terminal amide, both deprotection and cleavage from Rink amide resin was done by a cleavage cocktail (90% trifluoroacetic acid (TFA), 5% water, 2.5% 1,2-ethanedithiol, and 2.5% triisopropylsilane). For C-terminal blocked peptides, cleavage from Wang resin was done by incubation with the respective amines for 3–5 days and subsequent deprotection of the cleaved product by the same cleavage cocktail as described above. The peptides were purified by reverse-phase HPLC on a semipreparative Vydac C18 (1 cm  $\times$  25 cm) column using a 60 min linear gradient of 10–80% acetonitrile (containing 0.1% TFA) on a Waters Delta Prep 4000 (Waters Ltd., Mississauga, Ontario, Canada). Purity was evaluated by analytical HPLC. The molecular mass of the final products was verified using a SCIEX API III mass spectrometer (PE SCIEX, Thornhill, Ontario, Canada). HPLC retention times and mass spectrometry molecular masses are listed in the Supporting Information.

**Enzyme Inhibition Assays.** The substrate Cbz-Phe-Arg-MCA and the irreversible inhibitor E-64 were purchased from Bachem (King of Prussia, PA) and Peptides International

(Louisville, KY), respectively. Human cathepsins B, K, and L were prepared as described previously.<sup>18,26,27</sup> All recombinant enzymes were expressed in the yeast *Pichia pastoris* as a prepro- $\alpha$ -factor fusion construct using the culture conditions recommended by Invitrogen Corp. (San Diego, CA). The secreted proenzymes were autocatalytically activated, purified, and stored at either 4 °C (cathepsins B and L) or –80 °C (cathepsin K) in inhibited form by MMTS or  $\text{HgCl}_2$ .<sup>18,26,27</sup>

Kinetic experiments were performed as previously described.<sup>26</sup> Fluorescence was monitored on a SPEX Fluorolog-2 spectrofluorometer with the excitation and emission wavelengths set at 380 and 440 nm, respectively. The enzymes, stored in inhibited form, were preactivated by incubation with 2 mM DTT in the same buffer as the reaction mixture. The concentration of active enzyme was determined by titration with E-64.<sup>28</sup> All kinetic measurements were carried out at 25 °C in the presence of 2 mM DTT, 0.2 M NaCl, and 3% DMSO. The reactions were carried out at pH 6.0 for cathepsins B and K (50 mM sodium phosphate, 5 mM EDTA) and at pH 5.5 for cathepsin L (50 mM sodium citrate, 1 mM EDTA). When classical (i.e., linear) kinetics were observed, the  $K_i$  values were obtained from a graph of  $1/v_s$  vs  $[I]$  by measuring the initial rate of substrate hydrolysis ( $v_s$ ) in the presence of varying concentrations of inhibitor and at substrate concentrations kept well below  $K_M$ .<sup>29</sup> However, in most cases, nonlinearity in the initial portion of the progress curves indicated the presence of a “slow inhibition” process and the data were analyzed as described previously.<sup>21</sup>

**Molecular Modeling.** The crystal structure of procatepsin L (PDB code 1cs8) was used as the starting point for the modeling study. Structure manipulation and visualization were done in SYBYL 6.6 (Tripos, Inc., St. Louis, MO). The proregion residues were removed except for the fragment Met<sup>75p</sup>-Asn-Gly-Phe-Gln<sup>79p</sup>. Crystallographic water molecules buried in the mature enzyme were retained. Both N and C termini of the protein were modeled in the ionized state. The peptide fragment was capped at the N- and C-terminal ends with acetyl and  $\text{NH}_2$  groups, respectively. All histidine residues were protonated, and the catalytic cysteine was modeled as a thiolate. Hydrogen atoms were added, and the complex was subjected to conjugate gradient energy minimization using a distance-dependent dielectric function ( $\epsilon = 4r$ ) and an 8 Å nonbonded cutoff up to an rms gradient of 0.01 kcal mol<sup>–1</sup> Å<sup>–1</sup> using the AMBER force field.<sup>30</sup> This complex served as the template for the initial docking of our inhibitors.

Bound conformations of the inhibitors were modeled by docking them according to the template followed by a conformational search using a Monte Carlo minimization (MCM) procedure.<sup>31–34</sup> Starting structures for each cycle of minimization were obtained by randomly perturbing one or more dihedral angles in the inhibitor. The perturbations involved random changes in the side chain dihedral angles as well as crankshaft rotations of peptide units. Selected residues of the protein around the active site were allowed to relax during the minimization. The set of mobile protein residues defined as extending 8 Å around the docked template ligand was used for all the inhibitors. A total of 1000 MCM cycles was carried out for each inhibitor. The AMBER force field had to be supplemented with parameters for unnatural amino acids and blocking groups (see Supporting Information).

**Crystallization and Data Collection.** The cathepsin L–13 complex was prepared by incubating the protein with the inhibitor in the presence of 2 mM DTT. The protein was kept in a buffer containing 50 mM sodium acetate, pH 6.0, and 100 mM NaCl. Owing to the limited solubility of the inhibitor, 0.01 mM concentrations of protein and inhibitor were used to prepare the initial mixture that was then concentrated up to 8.7 mg/mL with a final ratio of 1:1.2 M (protein/inhibitor). The crystal was grown by using the hanging drop vapor diffusion method with a reservoir solution of 18% (w/v) poly(ethylene glycol) 8000, 200 mM sodium citrate, pH 4.2, 200 mM  $\text{Li}_2\text{SO}_4$ , and 8% 2-propanol at 18 °C. The crystallization drop contained 2  $\mu\text{L}$  of inhibitor complex and 2  $\mu\text{L}$  of the reservoir solution. After microseeding, rod-shaped crystals



appeared and were grown for up to 2 weeks. Diffraction data were collected on an Raxis IIC area detector mounted on a RU300 rotating anode generator with the reservoir solution supplemented with 18% glycerol as a cryoprotectant. Data evaluation performed with Denzo and Scalepack<sup>35</sup> identified a  $P2_12_12_1$  space group having cell parameters  $a = 51.52$  Å,  $b = 58.63$  Å,  $c = 151.45$  Å and two molecules per asymmetric unit. A total of 134 326 reflections observed were merged to 34 166 unique reflections with  $R_{\text{sym}} = 0.049$  and completeness of 92.1% to 1.9 Å resolution.

**Structure Determination and Refinement.** The mature cathepsin L molecule taken from the procathepsin L structure (PDB code 1cjl) was used as a starting model for molecular replacement solution using Amore.<sup>36</sup> After rotation and translation, the rigid body refinement of the cathepsin L model in Amore results in a correlation factor of 66.3 and  $R_{\text{cryst}}$  of 33.7. Further minimization in CNS<sup>37</sup> reduced the  $R$  factor to 0.31. At this stage, the calculated difference map clearly showed the presence of the inhibitor as well as part of a second molecule. The inhibitor was modeled into the map. Model building and refinement were done in O<sup>38</sup> and CNS, respectively. Appropriate entries were added to the dictionaries of both programs to accommodate the nonstandard groups of the inhibitor. After several cycles of map fitting and refinement, we obtained an  $R$  factor of 0.185 ( $R_{\text{free}} = 0.23$ ) for all the reflections within 30.0–1.9 Å resolution. The rms deviations from standard geometries of the bond lengths and angles were 0.009 Å and 1.3°, respectively. The backbone dihedral angles ( $\phi$  and  $\psi$ ) of the complex were in the allowed region of the Ramachandran plot.<sup>39</sup> Structural coordinates have been deposited in the Protein Data Bank (www.rcsb.org), accession number 1mhw.

**Acknowledgment.** S.C. thanks Dr. Alicja Kluczyk for useful discussions regarding the synthetic work. This is NRCC publication number 44856.

## Appendix

**Abbreviations.** Abu, 2-aminobutyric acid; Bpa, 4-biphenylalanine; Cha, cyclohexylalanine; Mcy, *S*-methylcysteine; MMTS, methyl methanethiolsulfonate; Npa, 2-naphthylalanine; Nva, norvaline; Orn, ornithine.

**Supporting Information Available:** Additional AMBER force field parameters for unnatural amino acids used in the molecular modeling calculations, HPLC chromatograms of the proteolytic stability assays on selected inhibitors, and HPLC retention times and molecular masses from MS. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## References

- McGrath, M. E. The Lysosomal Cysteine Proteases. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 181–204.
- Kakegawa, H.; Nikawa, T.; Tagami, K.; Kamioka, H.; Sumitani, K.; Kawata, T.; Drobnic-Kosorok, M.; Lenarcic, B.; Turk, V.; Katunuma, N. Participation of cathepsin L on bone resorption. *FEBS Lett.* **1993**, *321*, 247–250.
- Drake, F. H.; Dodds, R. A.; James, I. E.; Connor, J. R.; Debouck, C.; Richardson, S.; Lee-Rykaczewski, E.; Coleman, L.; Rieman, D.; Barthlow, R.; Hastings, G.; Gowen, M. Cathepsin K, but not cathepsins B, L, or S, is abundantly expressed in human osteoclasts. *J. Biol. Chem.* **1996**, *271*, 12511–12516.
- Brix, K.; Lemansky, P.; Herzog, V. Evidence for extracellularly acting cathepsins mediating thyroid hormone liberation in thyroid epithelial cells. *Endocrinology* **1996**, *137*, 1963–1974.
- Chapman, H. A. Endosomal proteolysis and MHC class II function. *Curr. Opin. Immunol.* **1998**, *10*, 93–102.
- Esser, R. E.; Angelo, R. A.; Murphey, M. D.; Watts, L. M.; Thornburg, L. P.; Palmer, J. T.; Talhouk, J. W.; Smith, R. E. Cysteine proteinase inhibitors decrease articular cartilage and bone destruction in chronic inflammatory arthritis. *Arthritis Rheum.* **1994**, *37*, 236–247.
- Katunuma, N.; Kominami, E. Abnormal expression of lysosomal cysteine proteinases in muscle wasting diseases. *Rev. Physiol. Biochem. Pharmacol.* **1987**, *108*, 1–20.
- MullerLadner, U.; Gay, R. E.; Gay, S. Cysteine proteases in arthritis and inflammation. *Perspect. Drug Discovery Des.* **1996**, *6*, 87–98.
- Sheahan, K.; Shuja, S.; Murnane, M. J. Cysteine protease activities and tumor development in human colorectal carcinoma. *Cancer Res.* **1989**, *49*, 3809–3814.
- Elliot, E.; Sloane, B. F. The cysteine protease cathepsin B in cancer. *Perspect. Drug Discovery Des.* **1996**, *6*, 12–32.
- Navab, R.; Chevet, E.; Authier, F.; Di Guglielmo, G. M.; Bergeron, J. J. M.; Brodt, P. Inhibition of endosomal insulin-like growth factor-I processing by cysteine proteinase inhibitors blocks receptor-mediated functions. *J. Biol. Chem.* **2001**, *276*, 13644–13649.
- Demuth, H.-U. Recent Developments in Inhibiting Cysteine and Serine Proteases. *J. Enzyme Inhib.* **1990**, *3*, 249–278.
- Rasnick, D. Small synthetic inhibitors of cysteine proteases. *Perspect. Drug Discovery Des.* **1996**, *6*, 47–63.
- Otto, H.-H.; Schirmeister, T. Cysteine Proteases and Their Inhibitors. *Chem. Rev.* **1997**, *97*, 133–171.
- Leung, D.; Abbenante, G.; Fairlie, D. P. Protease Inhibitors: Current Status and Future Prospects. *J. Med. Chem.* **2000**, *43*, 305–341.
- Coulombe, R.; Grochulski, P.; Sivaraman, J.; Ménard, R.; Mort, J. S.; Cygler, M. Structure of human procathepsin L reveals the molecular basis of inhibition by the prosegment. *EMBO J.* **1996**, *15*, 5492–5503.
- Cygler, M.; Mort, J. S. Proregion structure of members of the papain superfamily. Mode of inhibition of enzymatic activity. *Biochimie* **1997**, *79*, 645–652.
- Carmona, E.; Dufour, E.; Plouffe, C.; Takebe, S.; Mason, P.; Mort, J. S.; Ménard, R. Potency and selectivity of the cathepsin L propeptide as an inhibitor of cysteine proteases. *Biochemistry* **1996**, *35*, 8149–8157.
- Guay, J.; Falgoutret, J.-P.; Ducret, A.; Percival, M. D.; Mancini, J. A. Potency and selectivity of inhibition of cathepsin K, L and S by their respective propeptides. *Eur. J. Biochem.* **2000**, *267*, 6311–6318.
- Chagas, J. R.; Ferrer-Di Martino, M.; Gauthier, F.; Lalmanach, G. Inhibition of cathepsin B by its propeptide: Use of overlapping peptides to identify a critical segment. *FEBS Lett.* **1996**, *392*, 233–236.
- Fox, T.; de Miguel, E.; Mort, J. S.; Storer, A. C. Potent Slow-Binding Inhibition of Cathepsin B by Its Propeptide. *Biochemistry* **1992**, *31*, 12571–12576.
- Chen, Y.; Plouffe, C.; Ménard, R.; Storer, A. C. Delineating functionally important regions and residues in the cathepsin B propeptide for inhibitory activity. *FEBS Lett.* **1996**, *393*, 24–26.
- Lalmanach, G.; Lecaille, F.; Chagas, J. R.; Authier, E.; Scharfstein, J.; Juliano, M. A.; Gauthier, F. Inhibition of trypanosomal cysteine proteinases by their propeptides. *J. Biol. Chem.* **1998**, *273*, 25112–25116.
- Brinker, A.; Weber, E.; Stoll, D.; Voigt, J.; Müller, A.; Sewald, N.; Jung, G.; Wiesmüller, K.-H.; Bohley, P. Highly potent inhibitors of human cathepsin L identified by screening combinatorial pentapeptide amide collections. *Eur. J. Biochem.* **2000**, *267*, 5085–5092.
- Meldal, M.; Svendsen, I.; Juliano, L.; Juliano, M. A.; del Nery, E.; Scharfstein, J. Inhibition of Cruzipain Visualized in a Fluorescence Quenched Solid-Phase Inhibitor Library Assay, D-Amino Acid Inhibitors for Cruzipain, Cathepsin B and Cathepsin L. *J. Pept. Sci.* **1998**, *4*, 83–91.
- Nägler, D. K.; Storer, A. C.; Portaro, F. C.; Carmona, E.; Juliano, L.; Ménard, R. Major increase in endopeptidase activity of human cathepsin B upon removal of occluding loop contacts. *Biochemistry* **1997**, *36*, 12608–12615.
- Nägler, D. K.; Tam, W.; Storer, A. C.; Krupa, J. C.; Mort, J. S.; Ménard, R. Interdependency of Sequence and Positional Specificities for Cysteine Proteases of the Papain Family. *Biochemistry* **1999**, *38*, 4868–4874.
- Barrett, A. J.; Kumbhavi, A. A.; Brown, M. A.; Kirschke, H.; Knight, C. G.; Tamai, M.; Hanada, K. 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.* **1982**, *201*, 189–198.
- Hammes, G. G. *Enzyme Catalysis and Regulation*; Academic Press: New York, 1982.
- Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
- Li, Z.; Scheraga, H. A. Monte Carlo minimization approach to the multiple-minima problem in protein folding. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 6611–6615.



- (32) Caffisch, A.; Fischer, S.; Karplus, M. Docking by Monte Carlo Minimization with a Solvation Correction: Application to an FKBP–Substrate Complex. *J. Comput. Chem.* **1997**, *18*, 723–743.
- (33) Nägler, D. K.; Zhang, R.; Tam, W.; Sulea, T.; Purisima, E. O.; Ménard, R. Human Cathepsin X: A Cysteine Protease with Unique Carboxypeptidase Activity. *Biochemistry* **1999**, *38*, 12648–12654.
- (34) Lin, L. Y.-C.; Sulea, T.; Szittner, R.; Vassilyev, V.; Purisima, E. O.; Meighen, E. A. Modeling of the Bacterial Luciferase–Flavin Mononucleotide Complex. Combining Flexible Docking with Structure–Activity Data. *Protein Sci.* **2001**, *10*, 1563–1571.
- (35) Otwinowski, Z.; Minor, W. Twinning. *Methods Enzymol.* **1997**, *276*, 307–326.
- (36) Collaborative Computational Project Number 4. The CCP4 Suite: Programs for Protein Crystallography. *Acta Crystallogr., Sect. D* **1994**, *50*, 760–763.
- (37) Brünger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; GrosseKunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L. Crystallography and NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **1998**, *54*, 905–921.
- (38) Jones, T. A.; Zou, J.-Y.; Cowan, S. W.; Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr., Sect. A* **1991**, *47*, 110–119.
- (39) Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **1993**, *26*, 283–291.

JM020238T