Exploring Inhibitor Binding at the S′ Subsites of Cathepsin L

Shafinaz F. Chowdhury,[†,‡,§] Lissa Joseph,[†,||] S. Kumar,[||] Shenoy Rajesh Tulsidas,[||] Sathesh Bhat,[‡,⊥,#] Edmund Ziomek,[‡] Robert Ménard,[‡] J. Sivaraman,*[,||] and Enrico O. Purisima*[,‡,⊥]

Biotechnology Research Institute, National Research Council Canada, 6100 Royalmount Avenue, Montreal, Quebec H4P 2R2, Canada, Department of Biological Sciences, 14 Science Drive 4, National University of Singapore, Singapore 117543, and Department of Biochemistry, McGill University, 3655 Promenade Sir William Osler, Montreal, Quebec H3G 1Y6, Canada

Received September 21, 2007

We report a series of noncovalent, reversible inhibitors of cathepsin L that have been designed to explore additional binding interactions with the S′ subsites. The design was based on our previously reported crystal structure that suggested the possibility of engineering increased interactions with the S′ subsites (Chowdhury et al. *J. Med. Chem.* **2002**, *45*, 5321−5329). A representative of these new inhibitors has been co-crystallized with mature cathepsin L, and the structure has been solved and refined at 2.2 Å. The inhibitors described in this work extend farther into the S′ subsites of cathepsins than any inhibitors reported in the literature thus far. These interactions appear to make use of a S3′ subsite that can potentially be exploited for enhanced specificity and/or affinity.

Introduction

There are 11 human cysteine protease cathepsins, which carry out terminal protein degradation in lysosomes. Aside from their housekeeping roles, cathepsins perform many additional functions in normal cells.[1] Deregulation of the endogenous activities of cathepsins can lead to or promote disease states. For example, cathepsin L has been implicated in tumor growth and invasion.[2–4] Cathepsins and other papain-like cysteine proteases have been recognized as viable drug targets for major diseases, such as osteoporosis, arthritis, immune-related diseases, atherosclerosis, and cancer, as well as for a variety of parasitic infections.[1–6] Hence, the development of cysteine protease inhibitors has broad potential.

Papain, the archetype of the cysteine protease family to which cathepsin L belongs, has a range of identifiable binding subsites that span from S3 to S2′.[7–10] Figure 1 illustrates the locations of these subsites in the active-site groove of cathepsin L along with a putative S3′ subsite. The majority of inhibitors reported for cathepsins have traditionally used primarily the S2- and S3-binding sites.[5,11–13] These inhibitors are typically substrate analogues with a reactive warhead that covalently modifies the active-site cysteine. More recently, inhibitors with a diaminoacetone core linking S- and S′-binding moieties have been developed.[14,15] In these peptidomimetics, the groups binding in the S subsites have peptide bonds in a substrate-like orientation. Those binding in the S′ subsites have peptide bonds oriented in a reverse direction compared to the substrate orientation. The diaminoacetone core permits the linking of these two oppositely oriented peptide-like fragments.

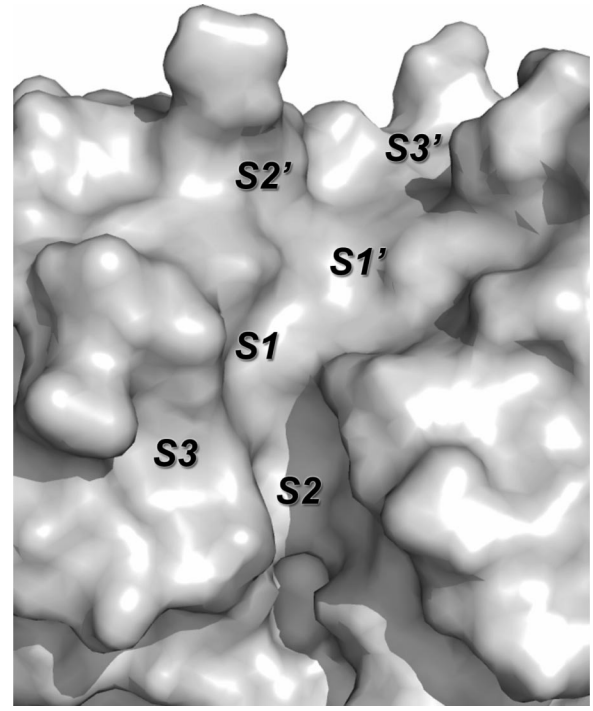


Figure 1. Location of the binding subsites in the active-site cleft of cathepsin L. (Figures 1, 2, 4, and 6−9 were generated using PyMOL).[44]

With the determination of the crystal structures of a number of cathepsin proenzymes, an alternate strategy for spanning the S- and S′-binding sites was revealed.[16,17] The autoinhibitory propeptide segment of cathepsins contains a polypeptide chain that spans the S and S′ subsites of the active site but binds in a direction reversed with respect to the normal substrate-binding mode. This effectively renders the polypeptide resistant to hydrolysis while blocking access to the active site. The challenge in taking advantage of this strategy was to reduce the size of the active-site spanning fragment while maintaining a reverse-binding mode to ensure stability against hydrolysis. We have previously reported a congeneric series of noncovalent inhibitors of cathepsin L designed to mimic the mode of autoinhibition

* To whom correspondence should be addressed. E-mail: dbsjayar@nus.edu.sg (J.S.); E-mail: enrico.purisima@nrc.ca (E.O.P.).
† These authors contributed equally to this work.
‡ National Research Council Canada.
§ Present address: Lady Davis Institute, McGill University, 3755 Côte Ste-Catherine Road, Montreal, Quebec H3T 1E2, Canada.
|| National University of Singapore.
⊥ McGill University.
Present address: Merck Frosst Canada Ltd. 16711 Trans Canada Highway, Kirkland, Quebec H9H 3L1 Canada.

[a] Abbreviations: DIPEA, *N*,*N*-diisopropylethylamine; PEG 8K, polyethylene glycol 8000; TBTU, 2-(*H*-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium tetrafluoroborate.

10.1021/jm701190v CCC: $40.75
Published on Web 02/16/2008

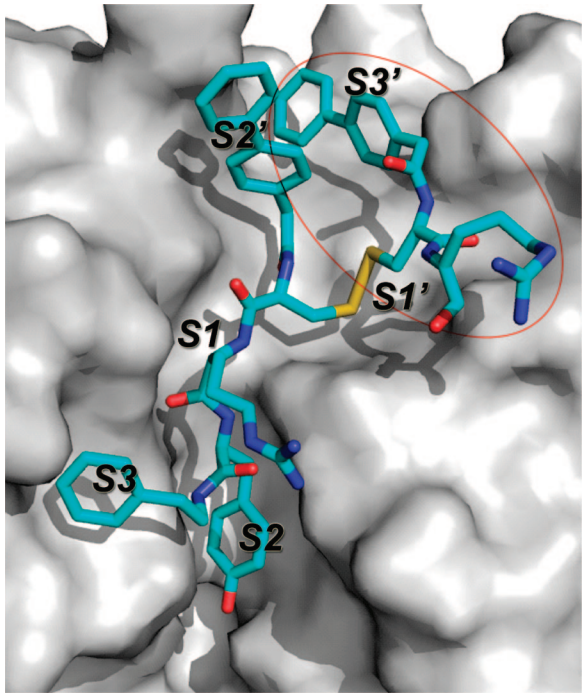


Figure 2. Crystal structure (PDB code 1MHW) of the complex of cathepsin L with the dimer of inhibitor **1**. Encircled with a thin red ellipse is the part of the second monomer (biphenylacetyl-Cys-D-Arg) that is visible in the X-ray data. The biphenyl group of the second monomer packs against the S3′ subsite of cathepsin L.

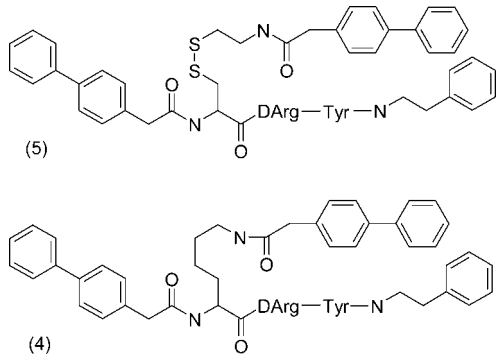


Figure 3. Dimer mimic. The structure of an inhibitor (**5**) with a hypothetical Cys derivative from a substructure extracted from the dimer of inhibitor **1**. Also shown is the synthetically more tractable congener with a *N*ε-biphenylacetyl-Lys replacing Cys (**4**).

of procathepsin L with nanomolar potency.[18] The 1.9 Å resolution crystal structure [Protein Data Bank (PDB) code 1MHW] of the complex of mature cathepsin L with one of the inhibitors confirmed the noncovalent nature and reverse-binding mode of these inhibitors.[18] However, in the crystal structure, the inhibitor that was co-crystallized had dimerized via a free cysteine residue and part of the second monomer was occupying an additional region, which we will refer to as the S3′ subsite of cathepsin L (see Figure 2). It should be pointed out, however, that the high potency of the methylcysteine derivative of the inhibitor, which precludes disulfide bond formation, demonstrates that dimerization is not essential for potency. Nevertheless, there were some lingering questions about whether the details of the binding of the monomeric inhibitor would be different. At the same time, the crystal structure of the dimeric inhibitor revealed more extensive interactions with the S′ subsites than previously seen in any inhibitor class, except for full-length propeptides. These findings motivated the present

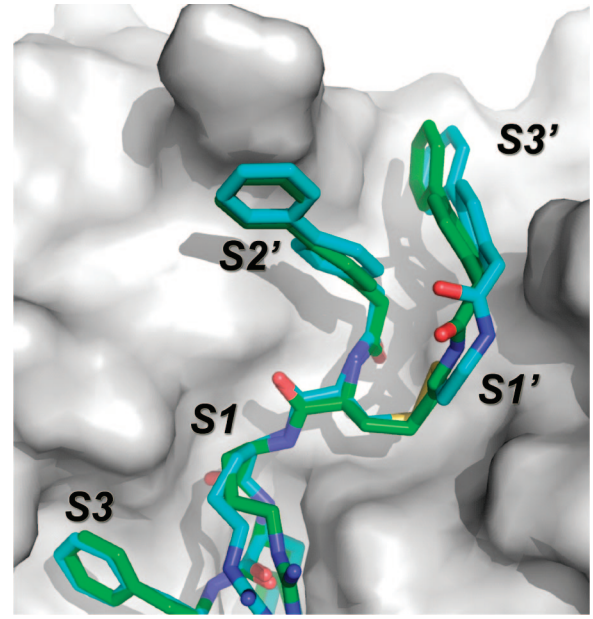


Figure 4. Model of the bound conformation of **4**. For a comparison, the crystal structure of the dimeric inhibitor is also shown with selected atoms from the second monomer stripped away for clarity.

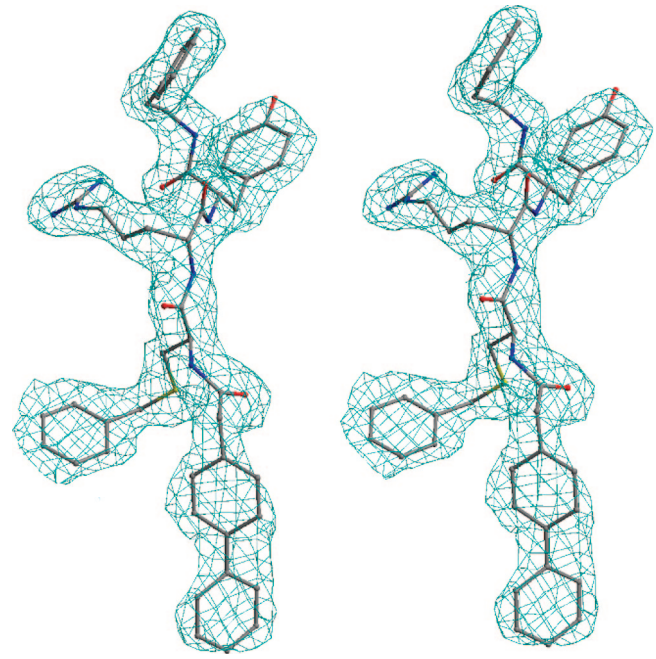


Figure 5. Simulated-annealing $F_o - F_c$ omit map in the active-site region of the cathepsin L inhibitor complex. Inhibitor molecule and all atoms within 3 Å from the inhibitor molecule were omitted prior to refinement. The map is contoured at a level of 2σ. This figure was prepared using Bobscript.[45]

study, in which we extend our previous inhibitor scaffold to use more of the S′ subsite interactions. In addition, we explored the electrostatic requirements for binding at the S1′ subsite. We also unequivocally established the binding mode of the monomeric inhibitor by solving the crystal structure of a representative monomeric structure.

Results and Discussion

Design of Inhibitors with Extended S′ Interactions. Inhibitor **1** (Table 1) is the monomeric component of the previously solved crystal structure (PDB code 1MHW).[18] In the crystal

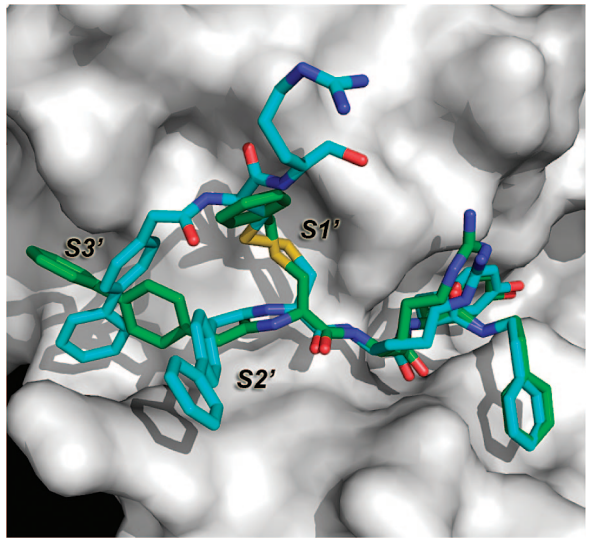


Figure 6. Overlay of the 1MHW crystal structure (cyan carbon chain) and the crystal structure of **2** (green carbon chain).

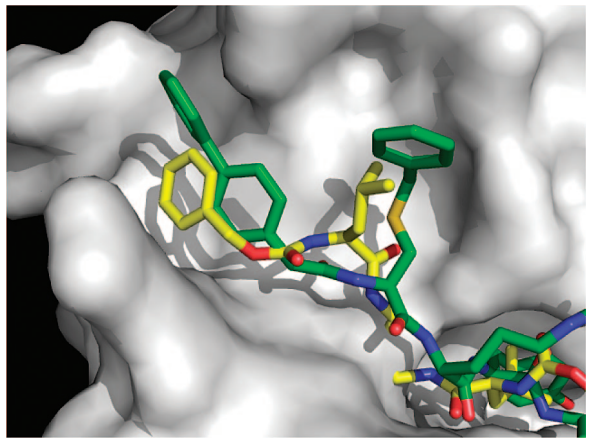


Figure 7. Overlay of **2** with a diaminoketone inhibitor (PDB code 1AU0).

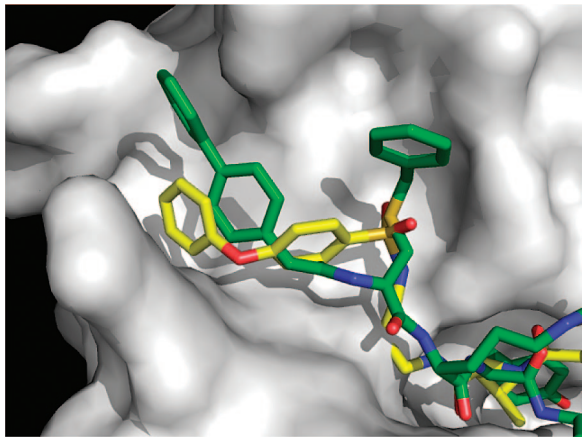


Figure 8. Overlay of **2** with a vinylsulfone inhibitor (PDB code 1AU2).

structure, a disulfide-bonded dimeric form of the inhibitor was observed. As seen in Figure 2, one monomer is completely visible, while only the biphenylacetyl-Cys-Arg fragment of the second monomer is well-defined. Presumably, the rest of the second monomer is disordered with no clear electron density. The Cys and biphenyl group of the second monomer occupy

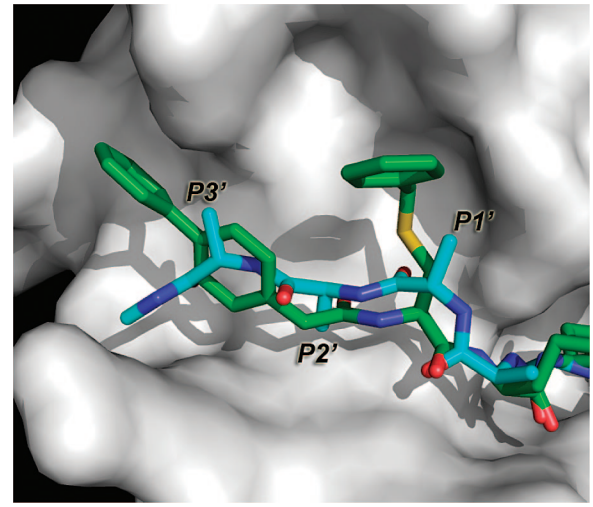


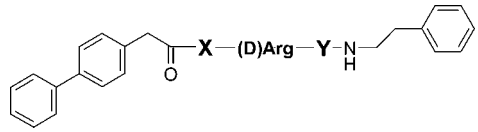
Figure 9. Overlay of **2** with a hypothetical polyalanine substrate.

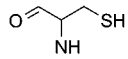
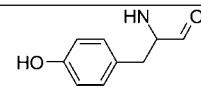
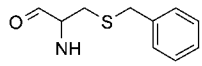
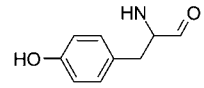
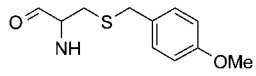
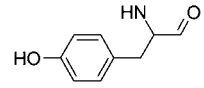
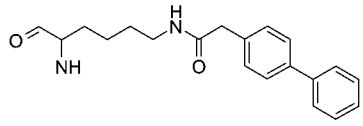
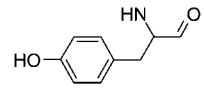
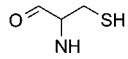
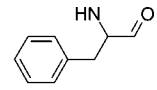
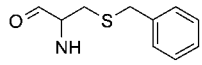
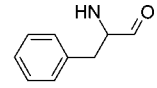
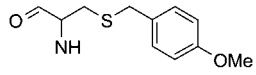
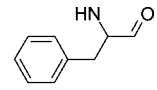
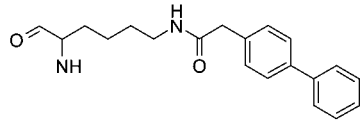
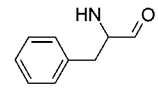
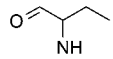
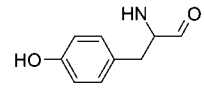
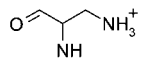
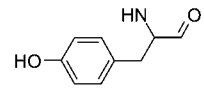
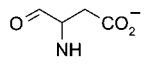
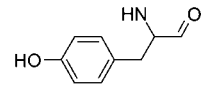
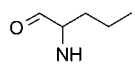
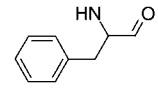
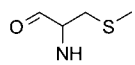
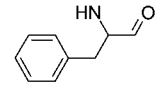
the S′ subsites. In this work, we sought to mimic the additional S′ interactions of the second monomer by modifying the Cys side chain in **1**. The first modification was the replacement of Cys with *S*-benzyl Cys (compound **2**). Modeling of the docked conformation of **2** suggests that the benzyl group has a number of nonpolar interactions at the junction of the S1′ and S3′ sites, notably with the side chains of Trp189 and Leu144 and the backbone of Gly139. It essentially mimics the interactions of the second Cys in the dimeric form of **1**. However, we do not expect this extension to be long enough to reach the region occupied by the biphenyl group of the second monomer seen in Figure 2. The K_i for **2** is 0.155 μM, around 3-fold weaker in affinity than **1**. To establish the binding mode, the crystal structure of **2** bound to cathepsin L was solved (see below).

A slightly larger extension was 4-methoxy-substituted *S*-benzyl cysteine (**3**). The methoxy group was expected to provide additional interactions with the S′ subsites. Molecular modeling of the bound conformation suggests that the benzyl fragments of **2** and **3** bind in a similar way at the junction of S1′ and S3′. Although the methoxy group makes additional contacts with the side chains of Leu144, Glu141, and the backbone of His140, the K_i for **3** is 0.112 μM, essentially the same as **2**.

Figure 3 shows compound **5** with a hypothetical Cys derivative obtained from the inhibitor dimer in the crystal structure (Figure 2) by following the bond connectivity from the second biphenyl to Cys in the first monomer. Unfortunately, this hypothetical derivative is not straightforward to synthesize. Much more tractable synthetically is **4** with *N*ϵ-biphenyl-Lys, which closely mimics the hypothetical molecule (Figure 3). The major differences are the replacement of the disulfide sulfur atoms with methylene groups and the overall shortening of the extension by one bond. The methylene groups are isosteric with the sulfur atoms and are expected to make similar contacts with the S_1' subsite. The aliphatic linker, unlike a disulfide bond, has the advantage of not being labile under reducing conditions. Figure 4 shows the modeled bound conformation of **4**, which contains the Nϵ-biphenyl-Lys replacement of Cys. For a comparison, the structure of the hypothetical Cys derivative is shown using the crystal structure coordinates from the dimeric inhibitor. The shortening by one bond slightly repositions the peptide group of the biphenylacteyl moiety compared to the inhibitor dimer crystal structure. However, the peptide group does not appear to form specific interactions with cathepsin L in either case. The K_i for **4** is 0.024 μM. Despite the predicted more extensive interactions with cathepsin L, it is only about

Table 1. Inhibitor Structures and Activity against Cathepsin L



Compound ID	X	Y	K_i (μM)
1			0.045
2			0.155
3			0.112
4			0.023
6			0.021
7			0.024
8			0.464
9			0.511
10			6.0
11			10.7
12			57.5
13			0.46
14			0.019

2-fold better than **1**. Part of reason for the lack of improvement in potency may be the entropic cost of freezing a long flexible linker upon complex formation. Another possible factor is that the methylene group may form weaker intermolecular van der Waals interactions than the sulfur of Cys in **1**.

In our previous study,[18] replacement of Tyr with Phe in **1** resulted in a K_i of 0.021 μM, a 2-fold improvement in potency. Hence, the analogues of **2**, **3**, and **4** with Tyr replaced by Phe were synthesized and tested. The results are summarized in Table 1. Compound **7** has a K_i of 0.024 μM, a 6-fold

Table 2. Crystallographic Statistics

data collection	
resolution range (Å)	50.0–2.2
wavelength (Å)	1.5418
observed *hkl*	86 866
unique *hkl*	21 015
completeness (%)	94.5
overall $I/\sigma I$	21.1
R_{sym}[a] (%)	0.048
refinement and quality of the model	
resolution range[b] (Å)	30.0−2.2
R_{work}[c] (%) (number of reflections)	18.34 (18 333)
R_{free}[d] (%) (number of reflections)	23.41 (1558)
root-mean-square deviation	
bond length (Å)	0.006
bond angle (deg)	1.044
Ramachadran plot	
favored region (%)	84.5
allowed regions (%)	15.5
generously allowed region (%)	0.0
disallowed regions (%)	0.0
average *B* factors[e] (Å^2)	
main-chain atoms	28.41
side-chain atoms	30.50
overall protein atoms (number of atoms)	29.42 (3322)
waters (number of atoms)	40.3 (395)
ligand (number of atoms)	27.90 (12)

[a] $R_{sym} = \sum|I_i - \langle I\rangle|/\sum|I_i|$, where I_i is the intensity of the *i*th measurement and $\langle I\rangle$ is the mean intensity for that reflection. [b] Reflections greater than $I > \sigma I$ were used in the refinement. [c] $R_{work} = \sum|F_{obs} - F_{calc}|/\sum|F_{obs}|$, where F_{calc} and F_{obs} are the calculated and observed structure factor amplitudes, respectively. [d] R_{free} = the same as for R_{work} but for 8.5% of the total reflections chosen at random and omitted from refinement. [e] Individual *B*-factor refinement were carried out.

improvement over **2**, consistent with our expectations. However, compounds **8** and **9** show 4- and 22-fold reductions in potency, respectively, over their Tyr counterparts, **3** and **4**. This is a somewhat puzzling result, given the potency of **6**. Molecular modeling of **7**, **8**, and **9** does not suggest any significant differences in the binding modes of these compounds over their Tyr analogues that could explain this behavior.

Crystal Structure of the Complex of Inhibitor 2 and Cathepsin L. The crystal structure of mature human cathepsin L complexed with inhibitor **2** was determined at 2.2 Å resolution. Compound **2** inhibits mature cathepsin L with a K_i of 0.155 μM. The asymmetric unit consists of two inhibitor complex molecules. The structure was solved by the molecular replacement method using mature cathepsin L coordinates derived from procathepsin L (PDB code 1CS8). The model has been refined with good stereochemical parameters. Statistics for the Ramachandran plot from an analysis using PROCHECK[19] for the inhibitor complex model gave 84.5% of nonglycine residues in the most favored region. Electron density of the mature cathepsin L residues from Thr175 to Gly179 is not observed and is presumably disordered. The data collection and refinement statistics are given in Table 2.

Figure 5 shows the simulated annealing $F_o - F_c$ omit map for the substrate-binding region of the cathepsin L inhibitor **2** complex. The inhibitor molecule is well-defined in the electron-density map. As predicted, the crystal structure agrees well with the molecular model of this inhibitor complex with noncovalent and reverse substrate-binding mode of inhibition. The electron density around the active-site Cys25 side chain indicates the possibility of oxidized sulfur. There are 10 direct hydrogen-bonding contacts (<3.2 Å) between the bound inhibitor and the mature cathepsin L of the complex. The overall binding mode of inhibitor **2** is comparable to that of the dimeric form of inhibitor **1** with D-Arg anchored at the S1 subsite and the Tyr and phenylethyl groups at the S2 and S3 subsites, respectively, in essentially identical orientations.[18] The major differences lie in the orientation of the biphenyl rings (Figure 6). In the current structure, the biphenyl group starts out in S2′ but projects into S3′ in an orientation perpendicular to the biphenyl in the dimeric part of inhibitor **1** (PDB code 1MHW). The orientation of the *S*-benzyl group follows the chain of the disulfide bond in 1MHW and positions the center of the benzene ring near the Cα of Cys in the second monomer. These results suggest that the positioning in S2′ of the biphenyl rings of the first half of the inhibitor dimer in 1MHW was induced by the presence of the dimer and that the preferred position for the monomer is at S3′. It will be interesting to see if the crystal structure of inhibitor **4**, which is a close mimic of the dimer, will once again displace the first biphenyl toward the S2′-binding site (work in progress).

The diaminoketone inhibitors of cathepsin K also have substantial interactions with the S′-binding sites.[14,15] In Figure 7, we aligned the cathepsin K structure (PDB code 1AU0) with the cathepsin L structure in this work and display and overlay of the cathepsin K inhibitor and molecule **2** in the cathepsin L active site. The diaminoketone inhibitor has Cbz-Leu occupying the S′ subsites. We see that the Cbz group roughly overlaps with the first half of the biphenyl group of **2** and that Leu side chain occupies part of the region used by the *S*-benzyl group. Figure 8 shows a similar alignment for a vinylsulfone inhibitor (PDB code 1AU2).[15] The phenyl rings of the inhibitor overlap with part of the biphenylacetyl group of **2**, and the sulfone S atom is positioned similarly to the *S*-benzyl sulfur of **2**. However, neither the diaminoketone nor vinylsulfone inhibitors reach as far as the second ring of the biphenyl group of **2**. These molecules represent three different classes of inhibitors that make significant interactions with the S′ subsites of cathepsins. Of these classes, the inhibitors in this work make the most extensive use of the putative S3′ subsite.

S3′ Subsite. Many years ago, Schechter and Berger suggested the existence of seven subsites in papain spanning S4−S3′ based on hydrolysis data on a series of polyalanine peptides.[8] However, there is no direct structural evidence for the existence of all these subsites. More recently, Turk and co-workers have revised the definition of substrate-binding sites in cysteine proteases to S3−S2′ based on the available crystal structures of inhibitors and substrate analogues. In their analysis, they see a S1′- and S2′-binding site but no evidence for S3′. The crystal structure in the present work reveals a binding site on cathepsin L that extends far beyond the S2′ subsite. Of course, this is a crystal structure of an inhibitor, and there is no evidence that a substrate would actually use the extended binding site. However, the proximity of the binding site to S2′ makes it a reasonable binding site for a P3′ side chain. Figure 9 shows a manually generated polyalanine peptide overlaid onto the crystal structure of inhibitor **2**. We see that there is the potential for the P3′ side chain to make use of the extended subsite. Hence, the region around the second ring in the biphenyl of **2** plausibly suggests a putative S3′ subsite.

Electrostatics of the S1′ Subsite. In the vicinity of the S1′ binding is a negatively charged Asp162. In particular, the distance between the sulfur atom of *S*-benzylcysteine and one of the carboxylate oxygens of Asp162 is 4.4 Å. This raised the possibility of designing an inhibitor with a positively charged moiety occupying S1′ to interact with the carboxylate of Asp162. To explore the electrostatic preferences at the S1′ subsite, three inhibitors (**10**, **11**, and **12**) containing neutral, positive, and negative groups for binding at S1′ were synthesized and tested. In these inhibitors, the Cys in **1** was replaced by aminobutyric,

Table 3. Relative Electrostatic Binding Free Energies of Inhibitors **1**, **10**, **11**, and **12** against Mature Cathepsin L[a]

inhibitors	E_{coul}	ΔG_{rf}	electrostatic $\Delta\Delta G_{calc}$	$\Delta\Delta G_{exp}$
1	−102.0	118.0	0.0	−3.0
10	−99.5	115.5	0.0	0.0
11	−183.6	201.0	1.4	0.3
12	−31.4	51.1	3.7	1.3

[a] Inhibitor **10** is the reference. All quantities are in kcal/mol.

diaminopropionic, and aspartic acids to provide side chains with neutral, positive, and negative charges, respectively. Compounds **10** and **11** are isosteric, while **12** is somewhat larger. We found that, despite the presence of the nearby Asp162, the neutral methyl is preferred (albeit only slightly) over the positive ammonium, which in turn is preferred over the negative carboxylate (Table 1). Notably, the differences in binding affinities among all three are small, implying that the subsite has no strong electrostatic preference.

These results are in agreement with a previous study of the substrate specificity of the S1′ subsite of cathepsin L.[20] Using intramolecularly quenched fluorogenic substrates, it was shown that cathepsin L favors amino acids with small (Ala and Ser) or long but nonbranched (Asn, Gln, and Lys) side chains in P1′ of the substrate. Asp at P1′ was about 8-fold worse than Ser.

To understand the lack of electrostatic preference at S1′, we calculated electrostatic binding free energies for inhibitors **10**, **11**, and **12**. These energies are the sum of the intermolecular Coulomb interaction plus the change in reaction field energy upon complex formation. The results are summarized in Table 3. We see that, indeed, the introduction of a positive moiety in **11** leads to a much stronger Coulomb interaction energy than in **10**. However, this is compensated by a much higher desolvation penalty upon binding, leading to a net decrease in binding affinity compared to **10**. Conversely, for **12**, the carboxylate has much poorer Coulomb interaction that is partially compensated by a reduced desolvation penalty upon binding. The calculated electrostatic binding free energies agree with the ranking and the modest differences in experimentally determined binding affinities.

Inhibitor **1**, similar to **10**, places a neutral SH group at the S1′ site but is found experimentally to be a significantly better binder. It has a higher binding affinity than would be expected from the calculated electrostatic contributions alone. A possible explanation is that sulfur in its lower oxidation states is significantly more polarizable than typical hydrocarbon moieties, such as the CH_3 group of Abu.[21] This would result in stronger van der Waals interactions with the binding site. A comparison of the binding affinities of the norvaline (**13**) and methylcysteine (**14**) derivatives parallels the significant differences in binding seen in **10** versus **1** (Table 1).

To further analyze the charge preference for ligand groups using S1′, we carried out a charge optimization calculation on the sulfur atom center of **1**. Charge optimization is an emerging tool for computer-aided molecular design.[22–25] Briefly, the method determines the optimal partial charges at selected ligand atom centers that maximize binding affinity. We find an optimal charge of $-0.047e$ at the sulfur atom coordinates, suggesting a preference for a neutral group. For a comparison, if we calculate ESP-fit charges for a cysteine residue using a united atom S to represent the SH group, we obtain a partial charge of $-0.041e$ for the sulfur atom. This suggests that Cys already has a near optimal partial charge at that position. However, upon calculating the charge selectivity, as given by the cost of altering the charge by $1e$ from its optimal value, we find only a modest penalty of 6.01 kcal/mol. In comparison to selectivity values reported in other charge selectivity calculations in the literature, this value is relatively small.[26,27] This suggests that although the S1′ subsite prefers a neutral group it does not have a strong preference for one.

Conclusion

In this work, we have established the binding conformation of the monomeric form of our retro-binding peptidomimetics. The conformation at the S1−S3 subsites was essentially identical to that previously reported for the dimeric form. However, the disposition of the biphenyl group at the S′ subsites was significantly altered. We explored extending the inhibitor to increase interactions with the S′ subsites. Although these extensions appear to be well-accommodated, they do not confer enhanced binding affinity, perhaps because of increased entropic costs upon binding. The electrostatic preferences at the S1′ subsite were also systematically studied using chemical synthesis and theoretical methods. A neutral group is slightly preferred at S1′. The interplay of coulomb interactions and desolvation costs was found to be important in modulating the electrostatic interactions at S1′.

The inhibitors described in this work have the most extensive interactions with the S′ subsites of cathepsins reported in the literature thus far. These interactions are highly suggestive of the existence of a S3′ subsite that can potentially be used for enhanced specificity and affinity.

Experimental Section

Enzyme 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 cathepsin L was prepared as described previously.[28–30] All recombinant enzymes were expressed in the yeast *Pichia pastoris* as a prepro-α-factor fusion construct using the culture conditions recommended by Invitrogen Corp. (San Diego, CA). The secreted proenzymes were autocatalytically activated, purified, and stored at 4 °C, inhibited by MMTS or $HgCl_2$.[28–30]

Kinetic experiments were performed as previously described.[28] 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 pre-activated by incubation with 2 mM dithiothreitol (DTT) in the same buffer as the reaction mixture. The concentration of active enzyme was determined by titration with E-64.[31] All kinetic measurements were carried out at 25 °C in the presence of 2 mM DTT, 0.2 M NaCl, and 3% dimethylsulfoxide (DMSO). The reactions were carried out at pH 5.5 [50 mM sodium citrate and 1 mM ethylenediaminetetraacetic acid (EDTA)]. When classical (i.e., linear) kinetics were observed, the K_i values were obtained from a graph of $1/v_s$ versus $[I]$ 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.[32] 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.[33]

Peptide Synthesis. All Fmoc-protected amino acids were purchased from Novobiochem (La Jolla, CA). 4-Biphenylacetic acid and 2-phenylethylamine 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),[a] 4 equiv; *N*,*N*-diisopropylethylamine (DIPEA), 6 equiv] in *N*-methylpyrilidone. Completion of the reaction was verified by colorimetric ninhydrin assay (Kaiser test). For compounds **4** and **9**, lysine with both the backbone and side-chain amino groups Fmoc-protected was used. This allowed simultaneous deprotection of the two amino groups for coupling with biphenylacetic acid. The N terminus was deprotected, and peptides were blocked with

corresponding acids by manual coupling (acids, 8 equiv; TBTU, 8 equiv; DIPEA, 12 equiv). Cleavage from Wang resin was performed by incubation with 2-phenylethylamine for 3−5 days and subsequent deprotection of the cleaved product by a cleavage cocktail [90% trifluoroacetic acid (TFA), 5% water, 2.5% 1,2-ethanedithiol, and 2.5% triisopropylsilane]. The peptides were purified by reverse-phase high-performance liquid chromatography (HPLC) on a semipreparative Vydac C18 (1 × 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).

Molecular Modeling. The crystal structure of mature cathepsin L in complex with a dimer of inhibitor **1** (PDB code 1MHW) was used as a template for constructing models of the bound conformations of the various inhibitors discussed in this paper. Structure manipulation and visualization were performed on SYBYL6.6 (Tripos, Inc., St. Louis, MO). The structure of the inhibitor was edited by removing extraneous atoms in the second monomer of the dimer to yield a starting model of the bound conformation of the hypothetical inhibitor **5**. Preparation of the structure for molecular modeling calculations was the same as described previously.[18]

Bound conformations of the other inhibitors were modeled by mutating the derivatized Cys in **5** to the necessary moiety. Each of the enzyme–inhibitor complexes were minimized followed by a conformational search using a Monte Carlo with energy minimization (MCM) procedure.[34,35]

Charge optimization provides the ideal charge at selected atom centers that optimize electrostatic binding affinity for a given binding configuration.[26,36–39] The optimal charge and selectivity were calculated as described previously.[39]

Complex Formation, Crystallization, and Data Collection. The mature cathepsin L was inhibited with inhibitor **2** by incubating the protein and inhibitor in the presence of 2 mM DTT at room temperature for 3 h. Protein was stored in a buffer of 20 mM sodium acetate at pH 5.7, 100 mM NaCl, and 1 mM EDTA. Because of the limited solubility of the inhibitor, approximately 0.01 mM concentration of protein and inhibitor were initially mixed to obtain a final ratio of 1:3 M (protein/inhibitor). After incubation, the complex was gradually concentrated up to 14 mg/mL. Crystals of the complex were grown by the hanging drop vapor diffusion method at room temperature with a reservoir solution of 17% (w/v) polyethylene glycol 8000 (PEG 8K) and 0.2 M ammonium sulfate. The drop was composed of 1 μL of reservoir solution and 1 μL of the protein inhibitor complex. Diffraction data were collected on a Raxis IV^{+2} area detector mounted on a RU300 rotating anode detector with 25% glycerol supplemented as the cryoprotectant. Diffraction data were processed and scaled with HKL2000.[40] The complex crystallized in the $C222_1$ space group, having cell parameters $a = 74.77$ Å, $b = 90.93$ Å, and $c = 126.03$ Å, with two complex molecules per asymmetric unit.

Structure Determination and Refinement. Initial phases were obtained by the molecular replacement method using Molrep,[41] with the mature part of the procathepsins L as a search model (PDB code 1CS8). The rotation and translation results in a correlation factor of 64.1 and R_{cryst} of 32.9. Further minimization in CNS reduced the R factor to 0.30. At this stage, the calculated difference map clearly showed the presence of the inhibitor. The inhibitor was modeled into the map. Model building and refinement were carried out in O[42] and CNS,[43] respectively. Appropriate entries were added to the dictionaries of both programs to accommodate the nonstandard groups of the inhibitor. The noncrystallographic symmetry (NCS) restraints were applied in the initial cycles of refinement but were removed at the final stage. After several cycles of map fitting and refinement, we obtained a R factor of 0.183 (R_{free} = 0.234) for reflections $I > \sigma I$ within 30.0–2.2 Å resolution. The crystallographic statistics are given in Table 2.

Coordinates of cathepsin L inhibitor complex **2** have been deposited in the Protein Data Bank (http://www.pdb.org) (Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. (2000) The Protein Data Bank. *Nucleic Acids Res. 28*, 235–242), accession number 3BC3.

Acknowledgment. J.S. acknowledges the research support from the Academic Research Fund (R-154-000-254-112), National University of Singapore (NUS). E.O.P. thanks Dr. Traian Sulea for useful discussions and comments.

Supporting Information Available: Analytical characterization of the synthesized peptide (Table S1). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Turk, V.; Turk, B.; Turk, D. New EMBO members' review: Lysosomal cysteine proteases: Facts and opportunities. *EMBO J.* **2001**, *20*, 4629–4633.
(2) Gocheva, V.; Zeng, W.; Ke, D.; Klimstra, D.; Reinheckel, T.; Peters, C.; Hanahan, D.; Joyce, J. A. Distinct roles for cysteine cathepsin genes in multistage tumorigenesis. *Genes Dev.* **2006**, *20*, 543–556.
(3) Turk, V.; Kos, J.; Turk, B. Cysteine cathepsins (proteases)—On the main stage of cancer. *Cancer Cell* **2004**, *5*, 409–410.
(4) Joyce, J. A.; Baruch, A.; Chehade, K.; Meyer-Morse, N.; Giraudo, E.; Tsai, F. Y.; Greenbaum, D. C.; Hager, J. H.; Bogyo, M.; Hanahan, D. Cathepsin cysteine proteases are effectors of invasive growth and angiogenesis during multistage tumorigenesis. *Cancer Cell* **2004**, *5*, 443–453.
(5) Lecaille, F.; Kaleta, J.; Brömme, D. Human and parasitic papain-like cysteine proteases: Their role in physiology and pathology and recent developments in inhibitor design. *Chem. Rev.* **2002**, *102*, 4459–4488.
(6) Sajid, M.; McKerrow, J. H. Cysteine proteases of parasitic organisms. *Mol. Biochem. Parasitol.* **2002**, *120*, 1–21.
(7) Berger, A.; Schechter, I. Mapping the active site of papain with the aid of peptide substrates and inhibitors. *Philos. Trans. R. Soc. London, Ser. B* **1970**, *257*, 249–264.
(8) Schechter, I.; Berger, A. On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **1967**, *27*, 157–162.
(9) Turk, D.; Guncar, G.; Podobnik, M.; Turk, B. Revised definition of substrate binding sites of papain-like proteases. *Biol. Chem.* **1998**, *379*, 137–147.
(10) McGrath, M. E. The lysosomal cysteine proteases. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 181–204.
(11) Demuth, H.-U. Recent developments in inhibiting cysteine and serine proteases. *J. Enzyme Inhib.* **1990**, *3*, 249–278.
(12) Rasnick, D. Small synthetic inhibitors of cysteine proteases. *Perspect. Drug Discovery Des.* **1996**, *6*, 47–63.
(13) Otto, H.-H.; Schirmeister, T. Cysteine proteases and their inhibitors. *Chem. Rev.* **1997**, *97*, 133–171.
(14) Thompson, S. K.; Halbert, S. M.; Bossard, M. J.; Tomaszek, T. A.; Levy, M. A.; Zhao, B.; Smith, W. W.; Abdel-Meguid, S. S.; Janson, C. A.; D'Alessio, K. J.; McQueney, M. S.; Amegadzie, B. Y.; Hanning, C. R.; DesJarlais, R. L.; Briand, J.; Sarkar, S. K.; Huddleston, M. J.; Ijames, C. F.; Carr, S. A.; Garnes, K. T.; Shu, A.; Heys, J. R.; Bradbeer, J.; Zembryki, D.; Veber, D. F. Design of potent and selective human cathepsin K inhibitors that span the active site. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 14249–14254.
(15) Yamashita, D. S.; Smith, W. W.; Zhao, B.; Janson, C. A.; Tomaszek, T. A.; Bossard, M. J.; Levy, M. A.; Oh, H. J.; Carr, T. J.; Thompson, S. K.; Ijames, C. F.; Carr, S. A.; McQueney, M.; D'Alessio, K. J.; Amegadzie, B. Y.; Hanning, C. R.; bdel-Meguid, S.; DesJarlais, R. L.; Gleason, J. G.; Veber, D. F. Structure and design of potent and selective cathepsin K inhibitors. *J. Am. Chem. Soc.* **1997**, *119*, 11351–11352.
(16) 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.
(17) Cygler, M.; Mort, J. S. Proregion structure of members of the papain superfamily. Mode of inhibition of enzymatic activity. *Biochimie* **1997**, *79*, 645–652.
(18) Chowdhury, S. F.; Sivaraman, J.; Wang, J.; Devanathan, G.; Lachance, P.; Qi, H.; Ménard, R.; Lefebvre, J.; Konishi, Y.; Cygler, M.; Sulea, T.; Purisima, E. O. Design of non-covalent inhibitors of human cathepsin L. From the 96-residue proregion to optimized tripeptides. *J. Med. Chem.* **2002**, *45*, 5321–5329.
(19) 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.
(20) Ménard, R.; Carmona, E.; Plouffe, C.; Bromme, D.; Konishi, Y.; Lefebvre, J.; Storer, A. C. The specificity of the S1′ subsite of cysteine proteases. *FEBS Lett.* **1993**, *328*, 107–110.

(21) Gellman, S. H. On the role of methionine residues in the sequence-independent recognition of nonpolar protein surfaces. *Biochemistry* **1991**, *30*, 6633–6636.

(22) Kangas, E.; Tidor, B. Electrostatic complementarity at ligand binding sites: Application to chorismate mutase. *J. Phys. Chem. B* **2001**, *105*, 880–888.

(23) Mandal, A.; Hilvert, D. Charge optimization increases the potency and selectivity of a chorismate mutase inhibitor. *J. Am. Chem. Soc.* **2003**, *125*, 5598–5599.

(24) Sims, P. A.; Wong, C. F.; McCammon, J. A. Charge optimization of the interface between protein kinases and their ligands. *J. Comput. Chem.* **2004**, *25*, 1416–1429.

(25) Bhat, S.; Sulea, T.; Purisima, E. O. Coupled atomic charge selectivity for ligand optimal charge distributions at protein binding sites. *J. Comput. Chem.* **2006**, *27*, 1899–1907.

(26) Lee, L.-P.; Tidor, B. Optimization of binding electrostatics: Charge complementarity in the barnase−barstar protein complex. *Protein Sci.* **2001**, *10*, 362–377.

(27) Sulea, T.; Purisima, E. O. Profiling charge complementarity and selectivity for binding at the protein surface. *Biophys. J.* **2003**, *84*, 2283–2896.

(28) 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.

(29) 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.

(30) 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.

(31) Barrett, A. J.; Kembhavi, 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.

(32) Dixon, M. The effect of pH on the affinities of enzymes for substrates and inhibitors. *Biochem. J.* **1953**, *55*, 161–170.

(33) 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.

(34) 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.

(35) 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.

(36) Lee, L.-P.; Tidor, B. Optimization of electrostatic binding free energy. *J. Chem. Phys.* **1997**, *106*, 8681–8690.

(37) Chong, L. T.; Dempster, S. E.; Hendsch, Z. S.; Lee, L.-P.; Tidor, B. Computation of electrostatic complements to proteins: A case of charge stabilized binding. *Protein Sci.* **1998**, *7*, 206–210.

(38) Kangas, E.; Tidor, B. Optimizing electrostatic affinity in ligand-receptor binding: Theory, computation, and ligand properties. *J. Chem. Phys.* **1998**, *109*, 7522–7545.

(39) Sulea, T.; Purisima, E. O. Optimizing ligand charges for maximum binding affinity. A solvated interaction energy approach. *J. Phys. Chem. B* **2001**, *105*, 889–899.

(40) Otwinowski, Z.; Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **1997**, *276*, 307–326.

(41) Vagin, A.; Teplyakov, A. MOLREP: An automated program for molecular replacement. *J. Appl. Crystallogr.* **1997**, *30*, 1022–1025.

(42) Jones, T. A.; Zou, J.-Y.; Cowan, S. W.; Kjelgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr., Sect. A: Found. Crystallogr.* **1991**, *47*, 110–119.

(43) 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., Sect. D: Biol. Crystallogr.* **1998**, *54*, 905–921.

(44) The PyMOL Molecular Graphics System. DeLano Scientific, Palo Alto, CA, 2002.

(45) Esnouf, R. Further additions to MolScript version 1.4, including reading and contouring of electron-density maps. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1999**, *55*, 938–940.

JM701190V