NRC Publications Archive
Archives des publications du CNRC

A combined crystallographic and molecular dynamics study of cathepsin L retrobinding inhibitors

Shenoy, Rajesh T.; Chowdhury, Shafinaz F.; Kumar, S.; Joseph, Lissa; Purisima, Enrico O.; Sivaraman, J.

This publication could be one of several versions: author’s original, accepted manuscript or the publisher’s version. / La version de cette publication peut être l’une des suivantes : la version prépublication de l’auteur, la version acceptée du manuscrit ou la version de l’éditeur.
For the publisher’s version, please access the DOI link below./ Pour consulter la version de l’éditeur, utilisez le lien DOI ci-dessous.

Publisher’s version / Version de l'éditeur:
https://doi.org/10.1021/jm900596y
Journal of Medicinal Chemistry, 52, 20, pp. 6335-6346, 2009-09-17

NRC Publications Record / Notice d'Archives des publications de CNRC:
https://nrc-publications.canada.ca/eng/view/object/?id=e52ea8f3-7f6f-4655-aa0a-78c0dcd9a372
https://publications-cnrc.canada.ca/fra/voir/objet/?id=e52ea8f3-7f6f-4655-aa0a-78c0dcd9a372

Access and use of this website and the material on it are subject to the Terms and Conditions set forth at
https://nrc-publications.canada.ca/eng/copyright
READ THESE TERMS AND CONDITIONS CAREFULLY BEFORE USING THIS WEBSITE.

L’accès à ce site Web et l’utilisation de son contenu sont assujettis aux conditions présentées dans le site
https://publications-cnrc.canada.ca/fra/droits
LISEZ CES CONDITIONS ATTENTIVEMENT AVANT D’UTILISER CE SITE WEB.

Questions? Contact the NRC Publications Archive team at PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca. If you wish to email the authors directly, please see the first page of the publication for their contact information.

Vous avez des questions? Nous pouvons vous aider. Pour communiquer directement avec un auteur, consultez la première page de la revue dans laquelle son article a été publié afin de trouver ses coordonnées. Si vous n’arrivez pas à les repérer, communiquez avec nous à PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca.

National Research Council Canada
Conseil national de recherches Canada

Journal of
Medicinal
Chemistry
Article

J. Med. Chem. 2009, 52, 6335–6346 6335
DOI: 10.1021/jm900596y

A Combined Crystallographic and Molecular Dynamics Study of Cathepsin L Retrobinding Inhibitors[†]

Rajesh T. Shenoy,[‖,‡] Shafinaz F. Chowdhury,[§,‖] S. Kumar,[‡] Lissa Joseph,[‡] Enrico O. Purisima,[*,§] and J. Sivaraman[*,‡]

[‡]*Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, Singapore 117543, and* [§]*Biotechnology Research Institute, National Research Council Canada, 6100 Royalmount Avenue, Montreal, Quebec H4P 2R2, Canada.* [‖]*These authors contributed equally to the work.*

Received May 7, 2009

We report the crystal structures of three noncovalent retrobinding inhibitors in complex with mature cathepsin L up to resolutions of 2.5, 1.8, and 2.5 Å, respectively. These inhibitors were Bpa-(Nε-Bpa)-Lys-DArg-Tyr-Npe, Bpa-(Nε-Bpa)Lys-DArg-Phe-Npe, and Bpa-MCys-DArg-Phe-Npe, where Bpa = biphenylacetyl and Pea = *N*-phenylethyl. These were selected to clarify the binding mode of the biphenyl groups in the S′ subsites because the addition of a second biphenyl does not improve potency. Examination of the symmetry-related monomers in the crystal structures revealed inhibitor–inhibitor crystal packing interactions. Molecular dynamics simulations were then used to explore the structure and dynamical behavior of the isolated protein–ligand complexes in solution. In the simulations, the backbone biphenyl groups for all three inhibitors ended up in the same location despite having started out in different orientations in the initial crystal structure conformations. The lack of improved potency of the larger inhibitors over the smaller one is attributed to a correspondingly greater entropic cost of binding.

Introduction

The papain superfamily of cysteine proteases consists of numerous members among which cathepsins form the largest group. In humans, 11 cathepsins are known to date.[1] The main function of these cathepsins is the terminal protein degradation in the lysosome. Although mainly being scavengers in the lysosome, they play vital physiological roles like in bone resorption and remodelling, T-cell maturation, prohormone processing, and apoptosis.[2] Recent studies have shown the involvement of cathepsins in several pathological conditions like tumor invasion and metastasis, osteoporosis, arthritis, atherosclerosis, emphysema, muscular dystrophy, parasitic infections, and cancer.[3–6] Cathepsin L is an important member of this group, which has been implicated in tumor growth and invasion. Therefore cathepsin L has been recognized as a viable target for therapeutics and cysteine proteases in general are attractive targets for drug design.

Cathepsin L is biosynthesized as an autoinhibited zymogen with a 96-residue propeptide segment that spans across the entire active site cleft, completely blocking access to it. The direction of the propeptide backbone is reversed as compared to the substrate binding mode, conferring resistance to hydrolysis.

We have previously reported a series of noncovalent retrobinding inhibitors designed to span the S–S′ regions of the active site of cathepsin L that mimic the propeptide binding mode.[7,8] These inhibitors showed nanomolar potency and high selectivity for cathepsin-L. The first crystal structure of such an inhibitor (PDB code 1MHW) confirmed the reverse-binding noncovalent nature of inhibition. However, this inhibitor contained a Cys amino acid that, during cocrystallization, caused the inhibitor to dimerize through disulfide bond formation.[8] The resulting bound conformation placed the biphenyl rings from each monomer in the S′ subsites of the protein. This raised some uncertainty as to the nature of the protein–ligand interactions in the S′ subsites for the monomeric inhibitor.

In the subsequent work, we reported the second crystal structure of a member of this class of inhibitors containing an *S*-benzyl Cys derivative that precluded dimer formation (PDB code 3BC3). The binding mode of the inhibitor at the S1–S3 subsites was identical to that previously reported for the dimeric form. However, the interactions of the biphenyl groups in the S′ subsite were significantly different between the two crystal structures.

In this paper, we report the crystal structures of three more inhibitors that have been selected to help clarify and elucidate the binding mode of this class of inhibitors. In addition, we have carried out molecular dynamics (MD[a]) simulations to supplement the information provided by the crystal structure studies.

Results and Discussion

Design of Dimer-Mimetic Propeptide Inhibitors. We have previously described the design and synthesis of inhibitors **4** and **9**, which were inspired by the crystal structure of inhibitor **1** (PDB code 1MHW).[7,8] In that structure, inhibitor **1** had inadvertently dimerized through a disulfide bond and positioned two biphenyl groups in the S′ subsites of cathepsin L. Inhibitors **4** and **9** were designed to mimic the interactions seen in the inhibitor **1** crystal structure. These inhibitors

---

[†] Coordinates and structure factors have been deposited with the RCSB Protein Data Bank[26] for cathepsin L complexed with inhibitors 4, 9, and 14 with accession codes 3H89, 3H8B, and 3H8C, respectively.

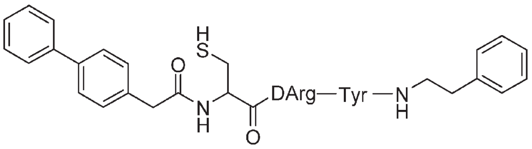
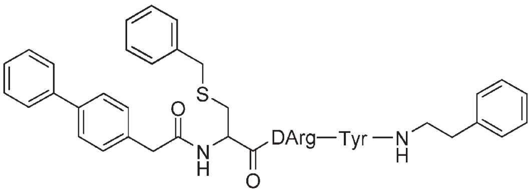
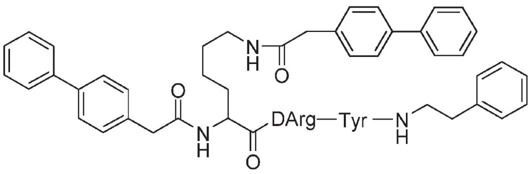
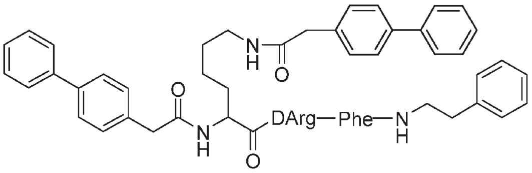
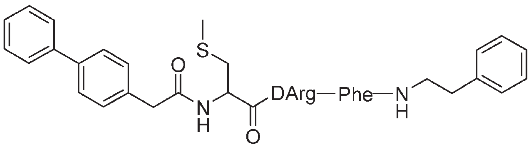
[*] To whom correspondence should be addressed. For E.O.P.: phone, (514) 496-6343; fax, (514) 496-5143; E-mail, Enrico.Purisima@nrc.ca. For J.S.: phone, + 65 6516 1163; fax, + 65 6779 5671; E-mail, dbsjayar@nus.edu.sg.

[a] Abbreviations: MD, molecular dynamics; MeCys, *S*-methylcysteine; PME, particle mesh Ewald.

© 2009 American Chemical Society

Published on Web 09/17/2009

Table 1[a]

Inhibitor ID	Structure	K_i (μM)
1		0.045
2		0.155
4		0.023
9		0.511
14		0.019

[a] All K_i values were taken from Chowdhury et al.[7,8] The crystal structures of **4**, **9**, and **14** are reported in the current paper.

contain a second biphenyl group without the additional bulk of a second monomer (Table 1). This was accomplished by attaching a biphenylacetyl group to Nε of a Lys side chain in the inhibitor. However, neither **4** nor **9** exhibited improved binding affinity compared to **14**, the most potent compound in the congeneric series.[8] To understand the lack of improved potency, we solved the crystal structures of **4**, **9**, and **14**.

Overall Structure of the Inhibitor Complexes. The structure of mature human cathepsin L complexed with inhibitors **4**, **9**, and **14** was solved by using rotating anode and/or synchrotron data sets and refined up to 2.5, 1.8, and 2.5 Å resolution, respectively. The structures were determined by the molecular replacement method using mature cathepsin L coordinates taken from the procathepsin L (PDB code 1CS8). Each of the models has been refined with good stereochemical parameters (Table 2). Statistics for the Ramachandran plot from an analysis using PROCHECK for the three models gave over 83% of nonglycine residues in the most favored regions and no residues in the disallowed region. Electron density corresponding to the residues Thr175 to Gly179 of mature cathepsin L is not observed and is presumably disordered. The inhibitor molecules are well-defined in the electron density maps. The inhibitors occupy the active site in a noncovalent and reverse-binding mode of inhibition. The overall structure of mature cathepsin L resembles that of the mature part of the proenzyme. Thus we will not provide any details of the mature cathepsin L structure.

Inhibitor 4. The binding mode of inhibitor **4** in the S1−S3 subsites is similar to that seen in previous crystal structures of this class of inhibitors (Figures 1a and Supporting Information Figure S1a).[7,8] The Arg residue of the inhibitor is located in S1 with its side chain solvent-exposed and its backbone carbonyl making a hydrogen bond with the NH group of Gly68. In the S2 subsite, the inhibitor Tyr side chain has nonpolar interactions with Leu69, Met70, and Ala135 and its amide NH forms a hydrogen bond with the Gly68 carbonyl. The *N*-(2-phenylethyl)-amide group packs against the Gly68 main chain and the side chain of Leu69. Inhibitor **4** makes fairly extensive interactions with the S′ subsites. The *N*ε-biphenylacetyl-modified Lys side chain is in a fairly extended conformation in the S′ region. The distance from the Lys Cα to the carbon at the tip of the biphenyl group is 15.6 Å. The biphenyl rings of the *N*ε-biphenylacetyl-Lys side chain interact with the side chains of Glu141, Phe145, and somewhat more distantly with Leu144. The biphenylacetyl methylene group interacts with Cα of Gly139. The Lys side chain methylene groups pack against the protein surface. The amide group of the *N*ε-biphenylacetyl-Lys side chain forms a hydrogen bond with the carbonyl of Ala138. In the S2′ subsite, the 4-biphenylacetyl rings makes hydrophobic interactions with the side chain of Leu144. The carbonyl oxygen of the biphenylacetyl group makes a hydrogen bond with the indole nitrogen of Trp189. Figure 2a shows the simulated annealing $F_o - F_c$ omit map for inhibitor **4**.

Table 2. Crystallographic Data and Refinement Statistics

	inhibitor **4**	inhibitor **9**	inhibitor **14**
space group	$C2$	$C2$	$P2_12_12_1$
cell parameters (Å, deg)	a = 99.40, b = 61.08, c = 205.18, and β = 89.91	a = 61.99, b = 99.23, c = 206.46, and β = 90.05	a = 57.92, b = 59.01, and c = 132.62
resolution range (Å)	50−2.43	50−1.79	50−2.5
wavelength (Å)	1.000	1.000	1.5418
obsd hkl	51706	866755	120192
unique hkl	42927	116974	16224
completeness (%)	91.8	99.7	99.1
overall $I/\sigma I$	22.2	17.1	17.8
$^{a}R_{\text{sym}}$	0.048	0.06	0.067
Refinement and Quality of the Model			
^{b}resolution range (Å)	25−2.5	45.0 − 1.8	25.0−2.5
$^{c}R_{\text{work}}$ (no. reflections)	0.22 (34041)	0.216 (100135)	0.237 (13732)
$^{d}R_{\text{free}}$ (no. reflections)	0.28 (3002)	0.250 (8770)	0.274 (838)
Root Mean Square Deviation			
bond length (Å)	0.008	0.01	0.008
bond angle (deg)	1.3	1.3	1.13
Ramachandran Plot (%)			
favored region	84.2	84.7	82.5
allowed regions	15.8	15.3	17.5
generously allowed region	0.0	0.0	0.0
disallowed regions	0.0	0.0	0.0
^{e}Average B Factors (Å^2)			
main chain atoms	25.65	23.97	54.95
side chain atoms	30.27	30.76	56.49
overall protein atoms (no. atoms)	27.88 (9954)	26.76 (9953)	55.69 (3322)
waters (no. atoms)	26.09 (226)	32.94 (562)	60.05 (129)
ligand (no. atoms)	42.23 (426)	42.30 (420)	29.34 (106)

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

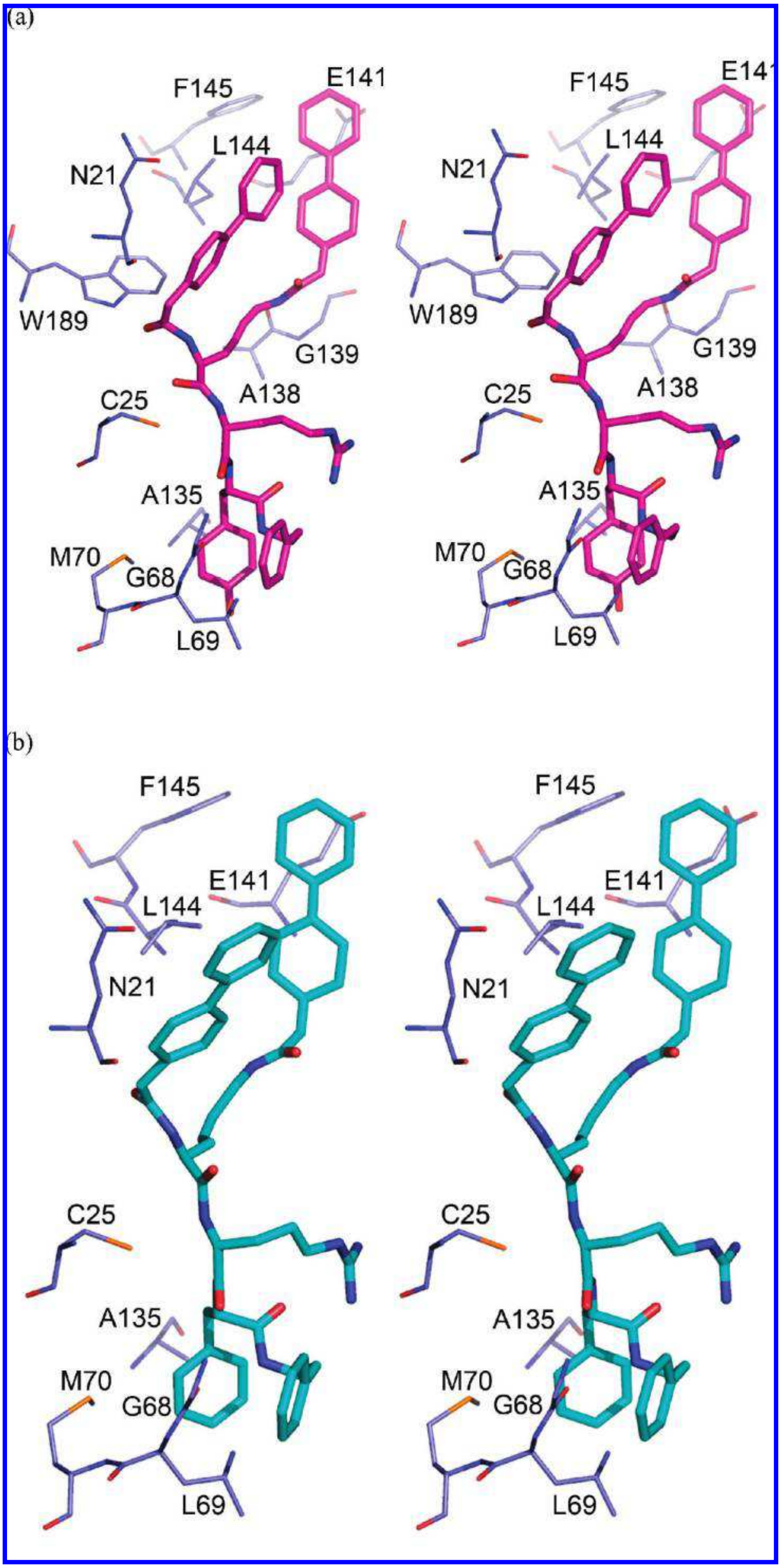
Inhibitor 9. The binding mode of inhibitor **9** in the S1−S3 subsites is essentially the same as that seen for inhibitor **4** (Figure 1b and Supporting Information Figure S1b). D-Arg is found in the S1 subsite, with the side chain guanidium group exposed to solvent and backbone carbonyl hydrogen bonded to the Gly68 amide. The Phe side chain makes nonpolar interactions with the side chains of Leu69, Met70, and Ala135 in the S2 subsite. In the S3 subsite, the NH group of the *N*-(2-phenylethyl)-amide makes a direct hydrogen bonding contact with the main chain carbonyl oxygen of Gly68. The phenyl ring of the *N*-(2-phenylethyl)-amide group packs against the backbone atoms of Gly68 and the side chain of Leu69. The orientation of the *Nε*-biphenylacetyl-Lys side chain in the S′ subsites is very similar between inhibitors **4** and **9**, but the position of the other biphenylacetyl group is significantly different in the two structures. As with inhibitor **4**, the rings of the *Nε*-biphenylacetyl-Lys side chain interacts with the side chains of Glu141, Leu144, and Phe145. The amide group of the *Nε*-biphenylacetyl-Lys makes a direct hydrogen bond with the carbonyl of Ala138. Unlike inhibitor **4**, the other biphenylacetyl group does not appear to have extensive interactions with the protein. Its main points of contact are nonpolar interactions with the backbone and side chains of Gln21. Figure 2b shows the simulated annealing $F_o - F_c$ omit map for the inhibitor **9**. The electron density of the inhibitor backbone biphenyl is somewhat weaker than that seen for inhibitor **4**.

Inhibitor 14. Inhibitor **14** is a close analogue of inhibitor **1**, the first compound in this class that was crystallized (PDB code 1MHW). The main difference is the replacement of Cys by *S*-methylCys (or MeCys, for short). This precludes the formation of a disulfide-bridged dimer as occurred in the crystallization of inhibitor **1**. This inhibitor represents the original design of this class of compounds.

The binding mode of inhibitor **14** at the S1−S3 subsites is essentially the same as that of inhibitors **4** and **9** (Figure 1c and Supporting Information Figure S1c). In the S1′ subsite, the MeCys makes nonpolar interactions with the side chain of Ala138. The 4-biphenylacetyl group is found to be in a conformation different from inhibitors **4** and **9** and is in a very extended conformation reaching toward putative S3′ subsite (Figure 1c) discussed previously.[7] The biphenyl group stacks against Trp189 and interacts with Gln21 on the other face of the rings. At the farthest tip of the biphenyl group there is some contact with the Cβ of Trp193. The simulated annealing $F_o - F_c$ omit map for inhibitor **14** shows that the position of the biphenyl rings is reasonably well-defined (Figure 2c).

Figure 3 shows a superposition of the binding modes of inhibitors **4**, **9**, and **14**. As mentioned above, the binding modes in S3−S1 is very similar. The *Nε*-biphenylacetyl-Lys side chains are also quite similar between **4** and **9**. The main difference among all three lies in the disposition of the backbone biphenyl group. Even between the highly congeneric **4** and **9**, they are oriented quite differently. The origin for the differences among the three crystal structures lies in the crystal packing interactions.

Crystal Packing Interactions. We have previously determined the crystal structure of inhibitor **2** (PDB code 3BC3), a


(a)
F145
E141
L144
N21
W189
G139
A138
C25
A135
M70
G68
L69
F145
E141
L144
N21
W189
G139
A138
C25
A135
M70
G68
L69
(b)
F145
E141
L144
N21
C25
A135
M70
G68
L69
F145
E141
L144
N21
C25
A135
M70
G68
L69

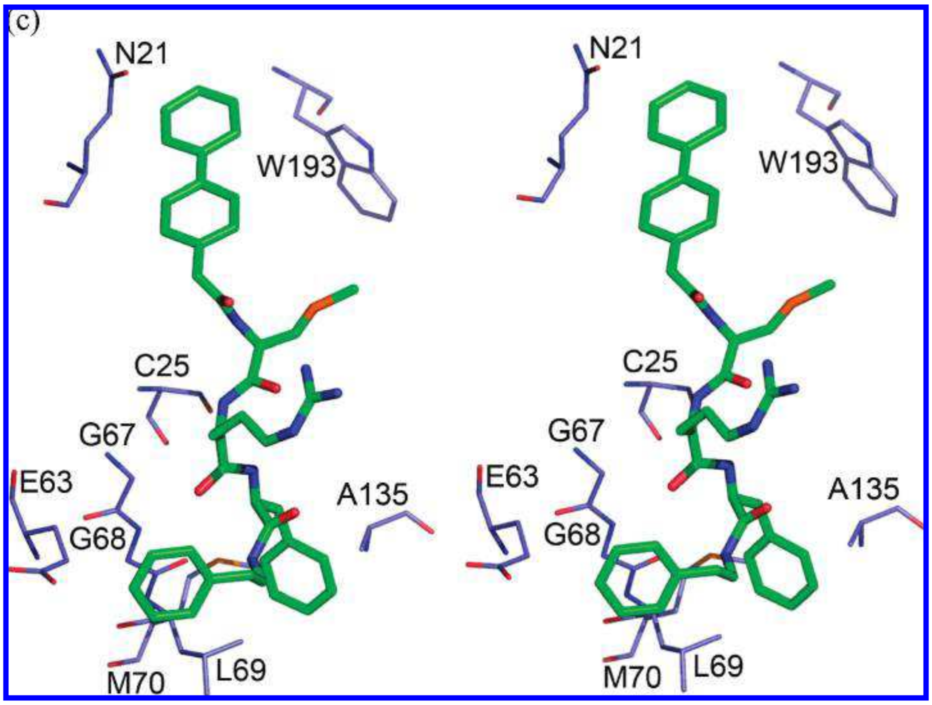


Figure 1. Stereo view of interactions between the inhibitors and cathepsin L. (a) Inhibitor **4**, (b) inhibitor **9**, and (c) inhibitor **14**. Protein atoms are shown in thin lines, whereas the inhibitors are shown in thick lines. Figures 1–4 and 8–9 were prepared using PyMOL (DeLano Scientific, Palo Alto, CA).

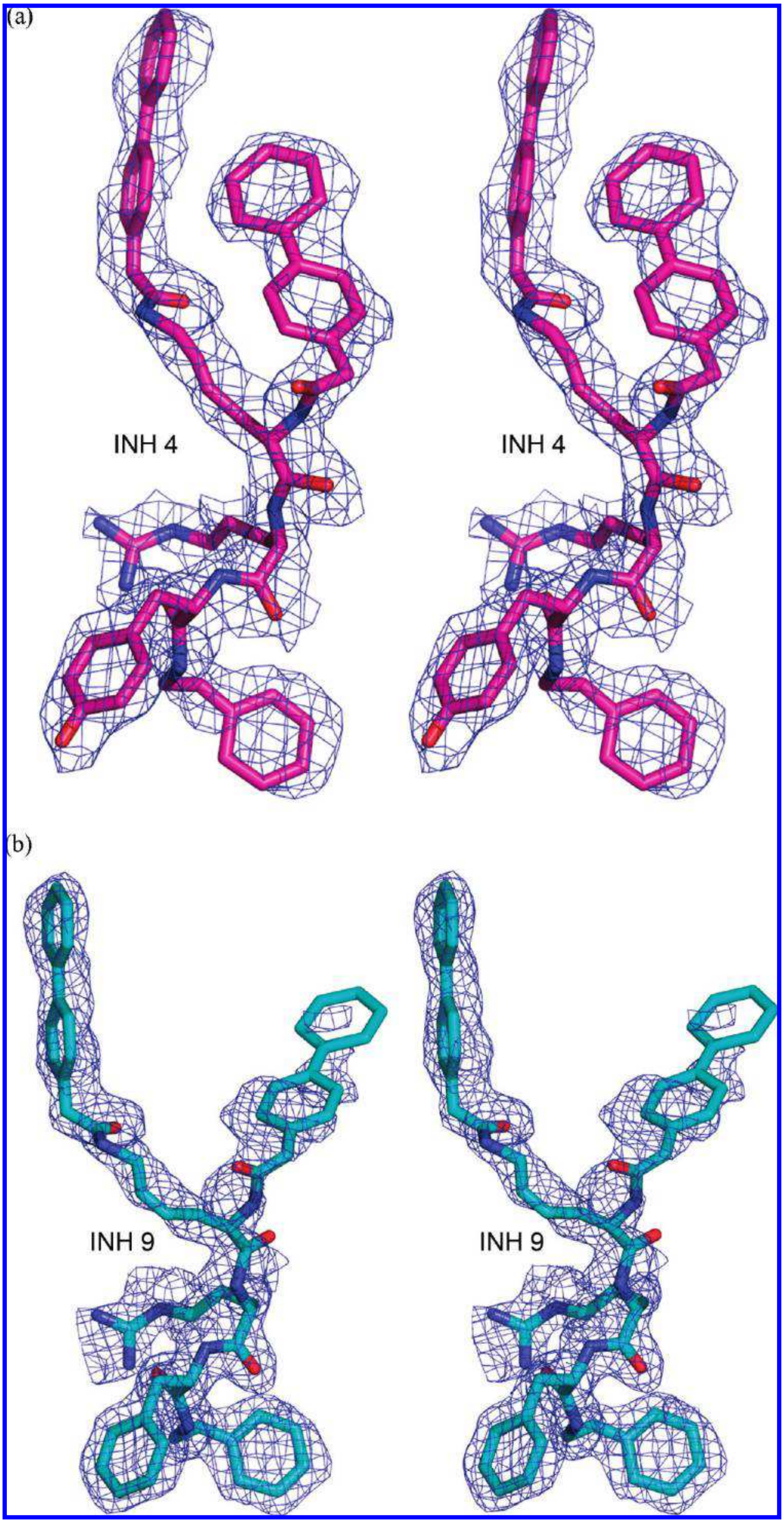
close congener of **14**.[7] A comparison of the bound conformations of the two inhibitors shows a similar binding mode, with the biphenyl rings extended toward the putative S3′ binding site[7] and the biphenyl rings stacked against the Trp189 rings. However, in the crystal structure of **2**, the biphenyl ring interacts more with the Trp189 indole ring and comes close to the Phe 145 side chain. In the crystal structure of **14**, a symmetry-related interaction between side chains of corresponding Leu144 residues of the protein occludes the region between Trp189 and Phe145, altering the binding mode of the biphenyl group relative to that seen in the crystal structure of **2**.

In the crystal structures of inhibitors **4** and **9**, there are even more significant symmetry-related interactions. Examination of the crystal packing contacts of inhibitor **4** revealed that the S′ regions of two adjacent protein molecules face each other and create a shared binding cavity at the interface that subsequently gets occupied by parts of the inhibitors bound to each protein unit. The biphenyl rings of the inhibitor bound to one protein interact with the biphenyl rings of the other inhibitor bound to the other protein (Figure 4a). The *Nε*-biphenyl of one protein is packed between the backbone biphenyl and the Gln21 side chain of the other protein. As in inhibitor **4**, examination of the symmetry related contacts of **9** revealed that the S′ regions of two adjacent proteins face each other. The backbone biphenyl from one inhibitor stacks against the side chain biphenyl from the other inhibitor (Figure 4b). For both **4** and **9**, it is thus possible that the bound conformation of the inhibitor in the S′ subsites in this crystal structure may not be representative of that found in an isolated protein–ligand complex in solution. Binding in the S subsites, however, does not seem to be affected by the crystal packing.

To explore how the binding modes might change in the absence of crystal packing interactions, molecular dynamics simulations were run as described in the next section.

Molecular Dynamics. Molecular dynamics simulations of cathepsin L protein complexed with **4**, **9**, and **14** were carried out in order to further clarify the binding mode of the inhibitors. Trajectories of 12 ns were generated for each of the inhibitors. Duplicate runs using different initial random velocities were carried out to confirm the reproducibility of the equilibrium structures. Figure 5 shows plots of the time series of selected distances describing the positions of S′-binding groups of **4**. The S′-binding moieties of **4** were highly mobile during the first few nanoseconds of the simulation before finally settling down to a reasonably stable conformation. The time series of the distance between C4 of the backbone biphenyl (see Supporting Information Figure S1a for ligand atom numbering) and the CG atom of side chain of Trp184 is shown in Figure 5a. The backbone biphenyl group quickly ends up packing against the Trp184 indole ring. There is some increased fluctuation in position from about 2–4 ns and a much larger brief excursion at about 9.5 ns, but the contact with Trp184 seems to be quite stable. In contrast, the side chain *Nε*-biphenyl is quite mobile during the first 7.5 ns. It eventually starts to stay close to Leu144 and Glu141 at around 5 ns and stays relatively localized in that region from 8 ns onward (Figure 5b). The lysine side chain N*ε* of **4** breaks its initial hydrogen bond to the Ala138 carbonyl O but reforms it after 1.5 ns and remains mostly formed for the rest of the simulation (Figure 5c).

Figure 6 shows the corresponding time series for selected distances describing the positions of S′-binding groups of inhibitor **9**. The evolution of the distance between the C4 of the backbone biphenyl (see Supporting Information Figure S1b for ligand atom numbering) and the CG atom of Trp184 is shown in Figure 6a. The backbone biphenyl group starts quite distant (more than 7 Å away) from Trp184 but finds its way toward it in about 1 ns. It fluctuates somewhat around that region but settles down from 5 ns onward, packing against the Trp184 indole ring. The side


(a)
INH 4
INH 4
(b)
INH 9
INH 9

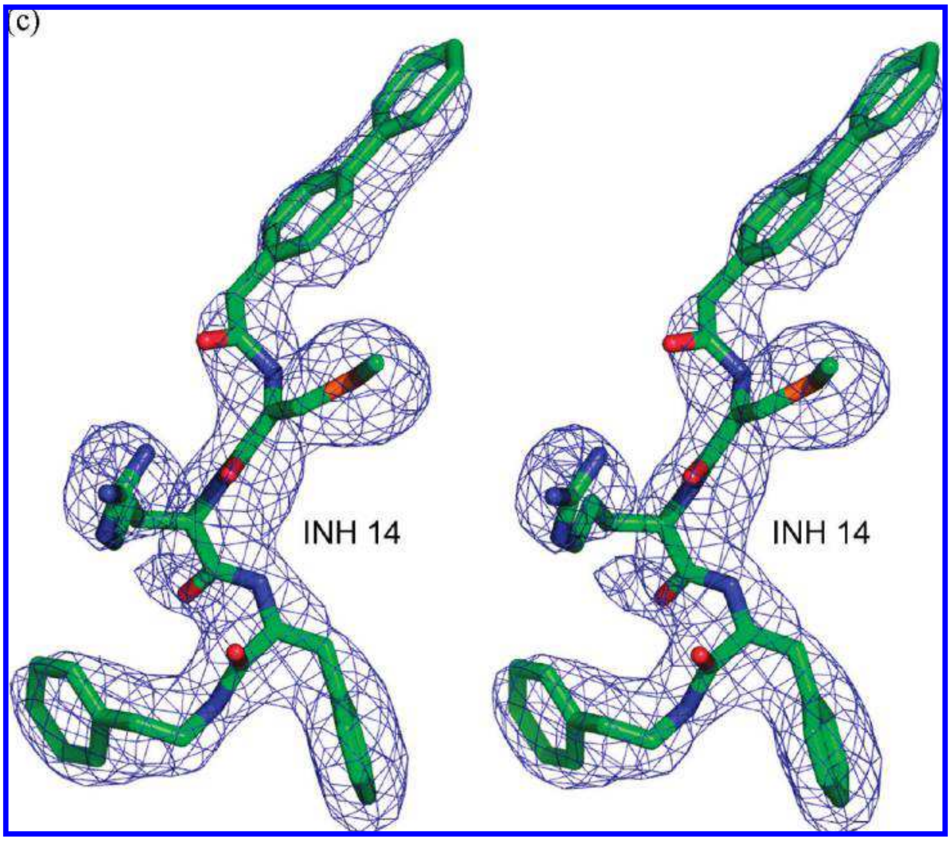


Figure 2. Stereo view of the simulated annealing $F_o - F_c$ omit map in the active site region of cathepsin L. The bound inhibitor and all atoms within 3 Å of the inhibitor molecule were omitted prior to refinement. The map contoured at a level of 2σ for inhibitor **4** and **14** and 2.5σ for inhibitor **9**. These figures were prepared by using PyMOL (DeLano Scientific, Palo Alto, CA).

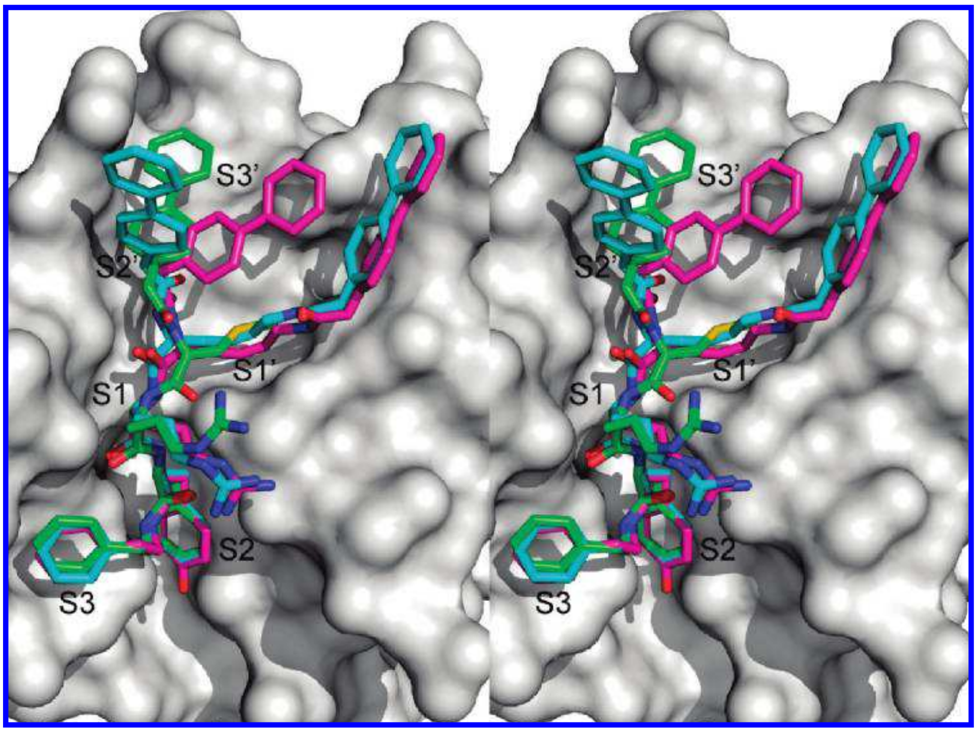


Figure 3. Stereo view of superposed binding modes of inhibitors **4**, **9**, and **14**. The three crystal structures were superposed, and the inhibitors were extracted and overlaid onto a molecular surface representation of cathepsin L taken from the crystal structure of **9**. To distinguish the three inhibitors, the carbon atoms of **4**, **9**, and **14** are in magenta, cyan, and green, respectively. The binding subsites S3−S3′ are labeled. Note the disparate positions of the biphenyl groups in the S2′ and S3′ positions.

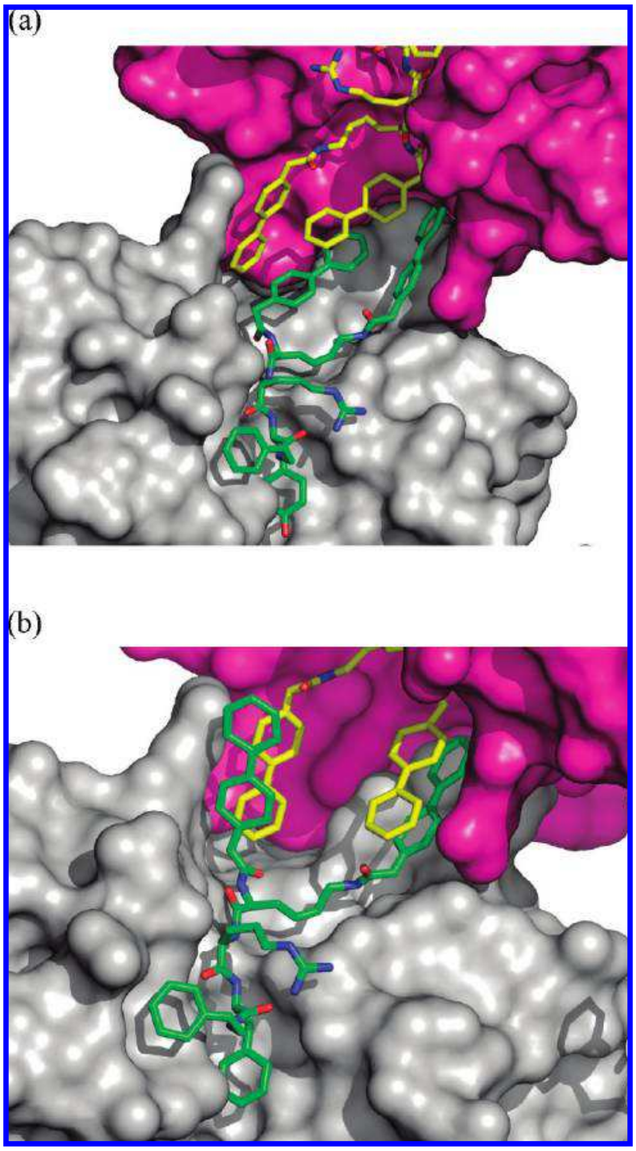


Figure 4. Crystal packing interactions. Shown are two adjacent protein−ligand complexes in the unit cell. (a) Inhibitor **4** and cathepsin L, (b) inhibitor **9** and cathepsin L. The biphenyl rings in one inhibitor interact with those of another inhibitor in a symmetry related monomer. This is particularly pronounced with inhibitor **9**.

chain *N*ε-biphenyl group is much more mobile than its counterpart in **4**. At the end of the 12 ns simulation, it has not really settled down to a defined binding site. This is consistent with the lower binding affinity observed for **9** as compared to **4** and **14**. The side chain biphenyl does preferentially fluctuate around and interacts with the side chains of Leu144 and Glu141 but does not stay localized for long periods of time. However, the lysine side chain Nε of **9** manages to maintain a hydrogen bond to the Ala138 carbonyl O for more of the 12 ns trajectory (Figure 6c).

The binding mode of **14** was quite stable throughout the 12 ns simulation. Figure 7 shows time series of the distance between the C4 carbon of the backbone biphenyl and the CG atom of Trp184. It remains stable at around 3.8 Å throughout the 12 ns trajectory. The position of the *S*-methylcysteine (MeCys) side chain also remains stable throughout (data not shown).

Figure 8 shows a superposition of snapshots for each of the inhibitors taken at 8.8 ns into the trajectory. This is a relatively "quiet" section of the trajectory in terms of the fluctuations reported in Figures 5−7. Visual inspection of the structures in the trajectory confirm that these selected

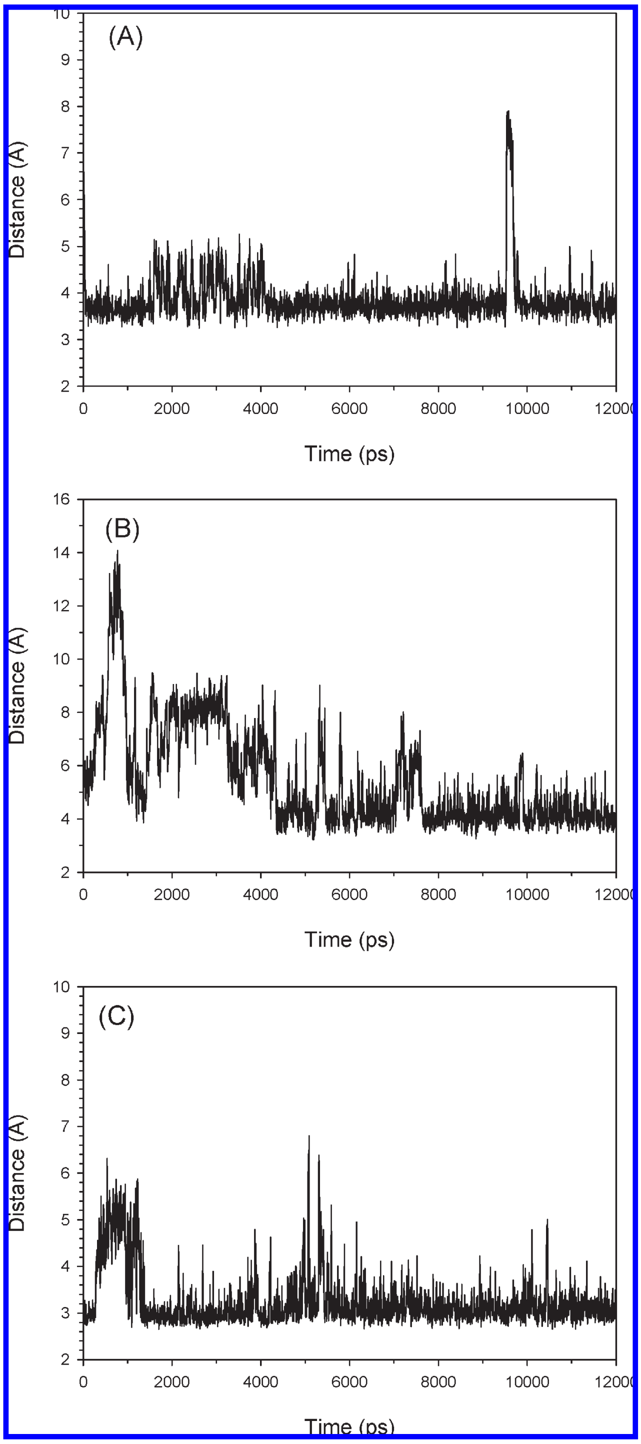


Figure 5. Interatomic distance fluctuations between cathepsin L and inhibitor **4**. (A) Trp184 CG−inhibitor C4. This tracks the interaction of the backbone biphenyl with the Trp184 indole ring. (B) Leu144 CD2−inhibitor C28. This tracks the interaction of the side chain biphenyl with a nonpolar binding site represented by the Leu144 side chain. (C) Ala138 carbonyl O−inhibitor Lys Nε. This tracks the locking of the inhibitor Lys side chain by formation of a hydrogen bond. See Supporting Information Figure S1a for atom numbering of inhibitor **4**. Together these distances provide an indication of the disposition and mobility of the biphenyl groups.

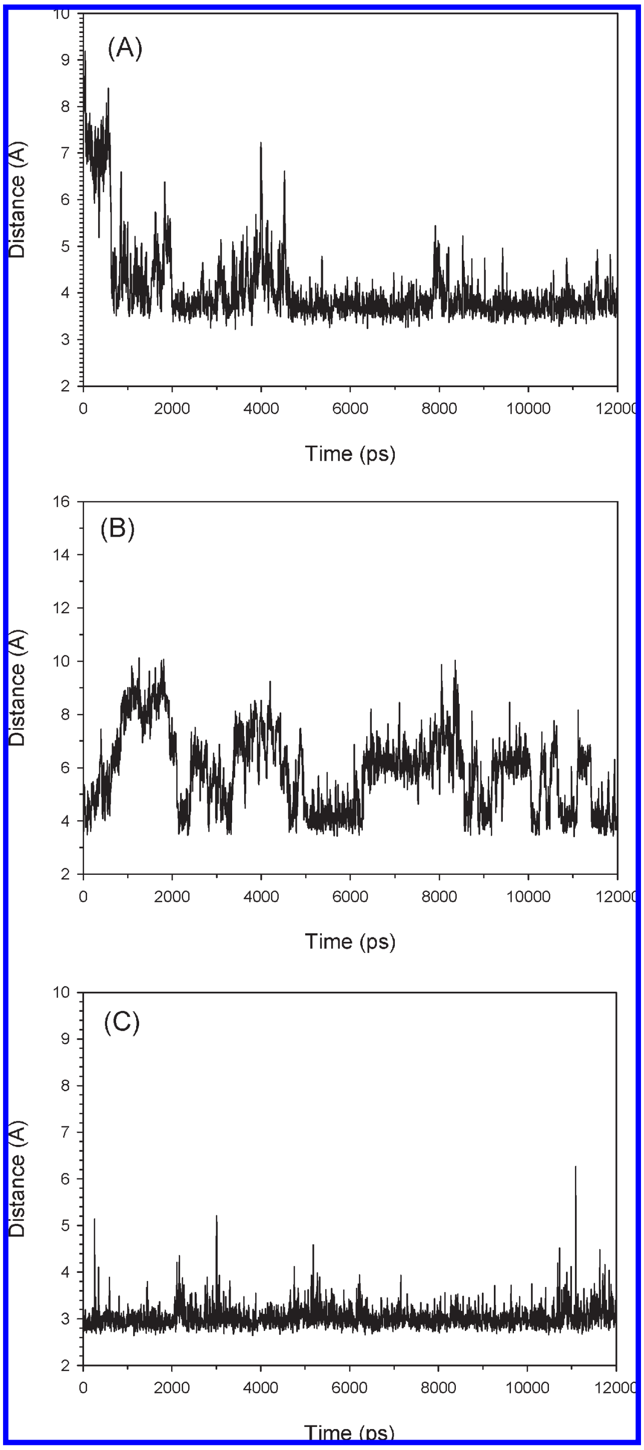


Figure 6. Interatomic distance fluctuations between cathepsin L and inhibitor **9**. (a) Trp184 CG–inhibitor C4, (b) Leu144 CD2–inhibitor C28, (c) Ala138 carbonyl O–inhibitor Lys Nε. See Supporting Information Figure S1b for atom numbering of inhibitor **4**. The significance of the various tracked distances is as in Figure 5.

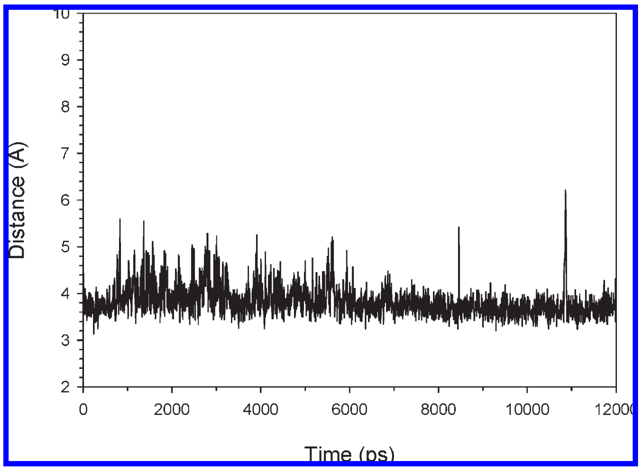


Figure 7. Interatomic distance fluctuations between cathepsin L and inhibitor **14**. Trp184 CG–inhibitor C4. See Supporting Information Figure S1c for atom numbering of inhibitor **14**. This tracks the interaction of the backbone biphenyl with the Trp184 indole ring.

snapshots are representative of the structures obtained toward the later part of the MD simulation. We see that in all three structures the backbone biphenyl groups have all converged into the same binding site. This binding site was not accessible to inhibitors **4** and **9** in the crystal structures due to symmetry-related interactions. Figure 9 shows an overlay of the MD snapshot of **14** in the crystal structure conformation. In the MD snapshot, the biphenyl ring extends farther into the S3′ binding site, reaching as far as the biphenyl ring in the crystal structure of **2**. Moreover, the peptide unit of the biphenylacetyl group has reoriented with its carbonyl facing the protein instead of pointing outward to solvent as in the crystal structure of **14**. The MD-refined orientation of the peptide plane overlays almost perfectly with that seen the crystal structure of **2**. These observations suggest that the binding mode for the backbone biphenyl groups seen in the MD simulations is probably more representative of the solution structure than that observed in the crystal structures of **4**, **9**, and **14**, which may have some perturbing effects from symmetry-related interactions in the crystalline forms. It is worth noting the synergy of X-ray crystallography and MD simulations in this case. The crystallographic structures supplied essential structural information and a reliable starting point for MD simulations, while the dynamics studies allowed further exploration of solution structures beyond what was available from the static models.

Binding Affinity. Inhibitors **4** and **9** are larger than **14**, have more van der Waals interactions with cat L, and yet are not more potent than **14**. Part of the explanation probably lies in the greater entropic cost of binding with **4** and **9**. In particular, we expect the reduction in translational, rotational, and torsional entropy to be greater for **4** and **9** versus **14**. The translational and rotational entropy change can be estimated using the Sackur–Tetrode equation.[9,10] This approximates the entropic change by the loss of rotational and translational entropy of the ligand going from the ideal gas to the bound state. At 300 K, the translational and rotational $-T\Delta S$ for **4**, **9**, and **14** are calculated to be 23.1, 23.1, and 22.4 kcal/mol, respectively. Moreover, inhibitors **4** and **9** have five more rotatable bonds than **14** because of the long derivatized lysine side chain, leading to an expected greater loss of torsional entropy upon binding. Typical empirical values for the entropic cost of freezing a rotatable bond range from about 0.3 to 0.5 kcal/mol.[11-14] This adds an additional 1.5–2.5 kcal/mol entropic cost for binding inhibitors **4** and **9** versus **14**. Hence, the lack of enhanced potency for inhibitors **4** and **9** can be rationalized in terms of entropic costs.

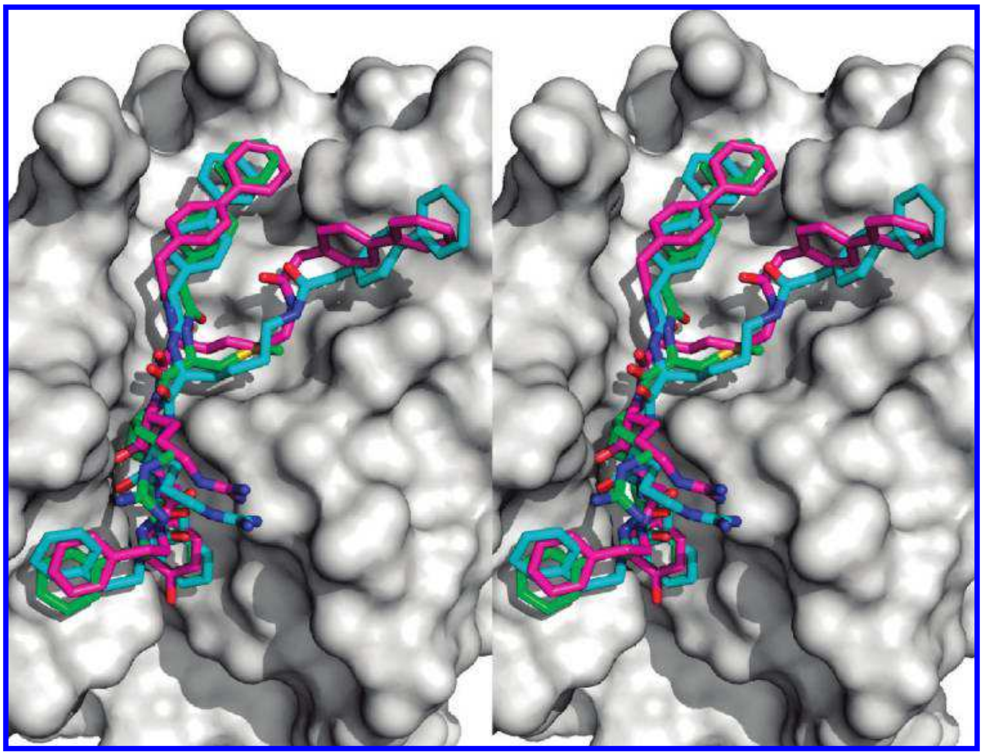


Figure 8. Stereo view of superposition of MD snapshots of inhibitors **4**, **9**, and **14**. The structures are overlaid onto a molecular surface representation of cathepsin L. The carbon atoms of **4**, **9**, and **14** are in magenta, cyan, and green, respectively. The binding subsites S3−S3′ are labeled. Note how the biphenyl groups in the S2′ and S3′ positions have converged into similar binding modes in contrast to the initial crystal structure poses seen in Figure 3.

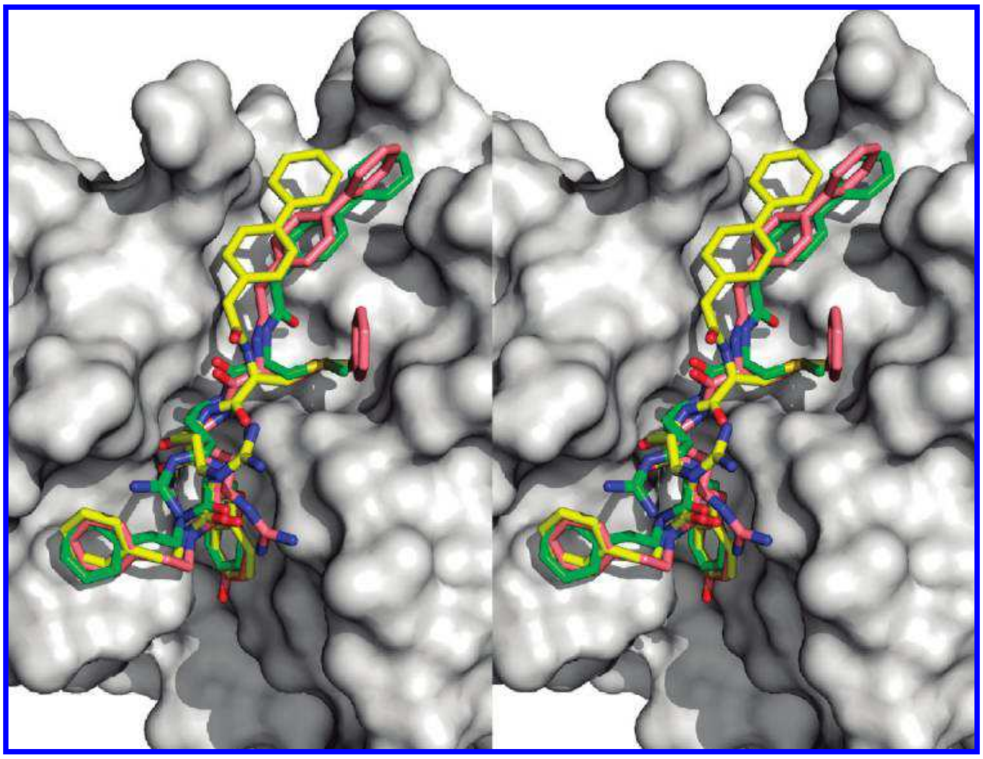


Figure 9. Stereo view of the superposition of an MD snapshot of inhibitor **14** and the crystal structures of inhibitors **2** (PDB code 3BC3) and **14**. The MD snapshot, and crystal structures of **2** and **14** have carbon chains colored green, pink, and yellow, respectively. The MD-refined binding mode of the biphenylacetyl group of **14** agrees closely with that of **2**, a previously solved closely related compound.

More difficult to understand is the difference in potency and molecular dynamics behavior between **4** and **9**, which differ by a single hydroxyl group (a Tyr to Phe change in the S2 subsite). Examination of the crystal structures and snapshots along the molecular dynamics trajectories of these two inhibitors did not reveal any special interactions made by Tyr hydroxyl group. In fact, the binding mode and dynamical behavior at the S2 and S3 subsites is quite similar between **4** and **9**, and yet the distal *Nε*-biphenylacetyl-lysyl side chain of **9** is observed to be more floppy than in **4** along the MD trajectory. Experimentally, **4** is observed to be more potent than **9**. This remains an unresolved puzzle for us. It should be noted, as reported previously,[7] that a Tyr side chain binding in S2 is not always superior to Phe. Clearly, more work has to be done to better understand the subtle influence that a Tyr/Phe in S2 has on binding.

Conclusion

We have solved the crystal structures of three inhibitors in complex with mature cathepsin L up to resolutions of 2.5, 1.8, and 2.5 Å, respectively. These inhibitors were selected to help clarify and elucidate the binding mode of this class of inhibitors. Of particular interest was the disposition of the biphenyl groups in the S′ subsites of the enzyme because the addition of a second biphenyl group to the inhibitor does not improve potency. Because of some question about the effects of symmetry-related interactions in the crystal, we have also carried out molecular dynamics simulations to supplement the information provided by the crystal structure studies and to explore the dynamical behavior of these inhibitors in the active site. After the MD simulations, all three inhibitors have one biphenyl group packing against the Trp189 rings. In inhibitor **4**, the second biphenyl packs against Leu144 and Glu141. In inhibitor **9**, the second biphenyl group also interacts with Leu144 and Glu141 but is much more mobile. The lack of improved affinity for inhibitors **4** and **9** relative to **14** may be due to the larger entropic cost of restricting the many rotatable bonds of the inhibitor lysine side chain. Perhaps forming a macrocyclic structure instead of two independent biphenyls would reduce the entropic cost and result in improved potency. The combined use of X-ray crystallography and molecular dynamics simulations has provided a better understanding of the binding mode of this class of inhibitors.

Experimental Section

Inhibitor Synthesis and Enzyme Assay. 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). The inhibitors were synthesized by standard Fmoc solid phase chemistry using manual coupling as described previously.[7,8] Purity was evaluated by analytical HPLC and was greater than 99%. The molecular mass of the final products was verified using a SCIEX API III mass spectrometer (PE SCIEX, Thornhill, Ontario, Canada). Binding affinities were taken from published data.[7,8]

Cocrystallization and Data Collection. Inhibitor 4. The complex of cathepsin L with inhibitor **4** was prepared by incubating the protein with inhibitor in the presence of 2 mM DTT at room temperature for 3 h. Protein was kept in 20 mM sodium acetate pH 5.7, 100 mM NaCl, and 1 mM EDTA. Because of the limited solubility of inhibitor, 0.01 M concentration of protein and inhibitor, which is dissolved in 25% DMSO, were used to prepare the initial mixture with a final ratio of 1:4 M (protein: inhibitor). After incubation, the complex was subsequently concentrated up to 9 mg/mL. Crystals of the complex were grown after 10 days by the hanging drop vapor diffusion method at room temperature with a reservoir solution of 19%(w/v) polyethylene glycol 8000 and 200 mM Ammonium sulfate.

Inhibitor 9. The complex of cathepsin L with inhibitor **9** was prepared by following the above-mentioned procedure to a ratio of 1:5 M (protein:inhibitor) and concentrated up to 10 mg/mL. Crystals of the complex were grown after 3 weeks by the hanging drop vapor diffusion method at room temperature with a reservoir solution of 25% (w/v) polyethylene glycol 8000 and 200 mM ammonium sulfate.

Inhibitor 14. The complex of cathepsin L with inhibitor **14** was prepared by following the above-mentioned procedure to a ratio of 1:5 M (protein:inhibitor) and concentrated up to 15 mg/mL. Crystals appeared after 2−3 weeks by the hanging drop vapor diffusion method at room temperature with a reservoir solution of 30% (w/v) polyethylene glycol 8000 and 200 mM ammonium sulfate.

All the cocrystallization drops were composed of 1 μL of reservoir solution and 1 μL of the complex. We have also soaked all the crystals to ensure the inhibition of cathepsin L in the crystals by the addition of 1.5 μM inhibitor solution into the drop. Diffraction data sets were collected on an R-axis IV++ area detector mounted on RU300 rotating anode detector and/or on a synchrotron beamline, BNL, with 25% glycerol as cryoprotectant. The data collection and refinement statistics are provided in the Table 1.

Structure Solution and Refinement. Initial phases for cathepsin L inhibitor complexes were obtained by molecular replacement method with Molrep[15] and using the mature cathepsin L coordinated taken from procathespin L (PDB code 1CS8). The initial *R* factor was 42% for all the complexes and subsequent refinement was carried out with CNS.[16] The resulting model with the electron density map was examined and the model was fitted with the O program.[17] Omit maps were calculated for positioning the inhibitor molecules. All three model buildings and refinements were carried out using O[17] and CNS[16] programs with appropriate entries in their respective dictionaries. Overall geometry of final models were analyzed by PROCHECK.[18] Refinement statistics are given in Table 1.

Molecular Modeling. The starting structures for MD simulations of inhibitors **4**, **9**, and **14** were based on their solved crystal structures using the AMBER 9 suite of programs.[19] The AMBER ff03 force field[20] for the proteins supplemented with generalized amber force field (GAFF)[21] atom types for inhibitors generated using the antechamber module of AMBER was used. Partial charges of the inhibitors were calculated by the AM1-BCC method.[22]

Each enzyme−inhibitor complex conformation was solvated in a truncated octahedron TIP3P water box.[23] The initial distance between the wall of the box and the closest atom of the solute was 12.0 Å, and the closest distance between the solute and solvent atoms was 0.8 Å. Counterions (Na^+) were added to neutralize the −9 net charge of the complex. Energy minimization, applying harmonic restraints with force constants of 10 kcal/mol/$Å^2$ to all solute atoms, was carried out, followed by heating from 100 to 300 K over 25 ps in the canonical ensemble (NVT). The solvent density was then adjusted by equilibrating over 25 ps in the isothermal isobaric ensemble (NPT) to a pressure of 1 atm. The harmonic restraints were then gradually reduced to zero with four rounds of 25 ps NPT simulations. After additional 25 ps simulation, 12 ns production NPT run was obtained with snapshots collected every 1 ps. For all simulations, a 2 fs time-step and 9 Å nonbonded cutoff were used. The particle mesh Ewald (PME) method[24] was used to treat long-range electrostatic interactions. Bond lengths involving hydrogen atoms were constrained by SHAKE.[25]

Acknowledgment. We acknowledge the X12C and X29 beamlines, National Synchrotron Light Source, Brookhaven National Laboratory, for the data collection, and thank Anand Saxena for the help during the data collection. J.S. acknowledges the research support from the Academic Research Fund (grant no. R154-000-254-112), National University of Singapore (NUS). Shenoy Rajesh Tulsidas is a graduate scholar in receipt of a research scholarship from

NUS. This is National Research Council of Canada publication number 50668.

Supporting Information Available: Schematic view of cathepsin L—inhibitor interactions, stereo view of superposition of an MD snapshot of inhibitors **4** and **9** versus their crystal structure binding modes, equivalent plots of Figures 5—7 for a duplicate MD run. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) McGrath, M. E. The Lysosomal Cysteine Proteases. *Annu. Rev. Biophys. Biomol. Struct*. **1999**, *28*, 181–204.

(2) Li, W.; Kornmark, L.; Jonasson, L.; Forssell, C.; Yuan, X.-M. Cathepsin L is significantly associated with apoptosis and plaque destabilization in human atherosclerosis. *Atherosclerosis* **2009**, *202*, 92–102.

(3) Sheahan, K.; Shuja, S.; Murnane, M. J. Cysteine protease activities and tumor development in human colorectal carcinoma. *Cancer Res*. **1989**, *49*, 3809–3814.

(4) 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.

(5) Turk, V.; Turk, B.; Turk, D. New EMBO Members' Review: Lysosomal cysteine proteases: facts and opportunities. *EMBO J.* **2001**, *20*, 4629–4633.

(6) 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.

(7) Chowdhury, S. F.; Joseph, L.; Kumar, S.; Tulsidas, S. R.; Bhat, S.; Ziomek, E.; Ménard, R.; Sivaraman, J.; Purisima, E. O. Exploring inhibitor binding at the S′ subsites of cathepsin L. *J. Med. Chem.* **2008**, *51*, 1361–1368.

(8) 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 Noncovalent Inhibitors of Human Cathepsin L. From the 96-Residue Proregion to Optimized Tripeptides. *J. Med. Chem*. **2002**, *45*, 5321–5329.

(9) Steinberg, I. Z.; Scheraga, H. A. Entropy Changes Accompanying Association Reactions of Proteins. *J. Biol. Chem*. **1963**, *238*, 172–181.

(10) Schwarzl, S. M.; Tschopp, T. B.; Smith, J. C.; Fischer, S. Can the Calculation of Ligand Binding Free Energies Be Improved with Continuum Solvent Electrostatics and an Ideal-Gas Entropy Correction? *J. Comput. Chem*. **2002**, *23*, 1143–1149.

(11) Pickett, S. D.; Sternberg, M. J. E. Empirical Scale of Side-Chain Conformational Entropy in Protein Folding. *J. Mol. Biol*. **1993**, *231*, 825–839.

(12) Wang, J.; Purisima, E. O. Analysis of Thermodynamic Determinants of Helix Propensities in Nonpolar Amino Acids through a Novel Free Energy Calculation. *J. Am. Chem. Soc*. **1996**, *118*, 995–1001.

(13) Wang, J.; Szewczuk, Z.; Yue, S.-Y.; Tsuda, Y.; Konishi, Y.; Purisima, E. O. Calculation of Relative Binding Free Energies and Configurational Entropies. A Structural and Thermodynamic Analysis of the Nature of Nonpolar Binding of Thrombin Inhibitors Based on Hirudin$^{55-65}$. *J. Mol. Biol*. **1995**, *253*, 473–492.

(14) Lazaridis, T.; Masunov, A.; Gandolfo, F. Contributions to the binding free energy of ligands to avidin and streptavidin. *Proteins* **2002**, *47*, 194–208.

(15) Vagin, A.; Teplyakov, A. MOLREP: An Automated Program for Molecular Replacement. *J. Appl. Crystallogr*. **1997**, *30*, 1022–1025.

(16) Brunger, 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.

(17) 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.

(18) 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.

(19) Case, D. A.; Cheatham, T. E., III; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber Biomolecular Simulation Programs. *J. Comput. Chem*. **2005**, *26*, 1668–1688.

(20) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem*. **2003**, *24*, 1999–2012.

(21) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem*. **2004**, *25*, 1157–1174.

(22) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation. *J. Comput. Chem*. **2002**, *23*, 1623–1641.

(23) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys*. **1983**, *79*, 926–935.

(24) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An $N\cdot\log(N)$ method for Ewald sums in large systems. *J. Chem. Phys*. **1993**, *98*, 10089–10092.

(25) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *J. Comput. Phys*. **1977**, *23*, 327–341.

(26) Berman, H. M.; Westbrook, J.; Zukang, F.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shidyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res*. **2000**, *28*, 235–242.