Molecular Dynamics and Free Energy Analyses of Cathepsin D−Inhibitor Interactions: Insight into Structure-Based Ligand Design

Shuanghong Huo,[†] Junmei Wang,[‡] Piotr Cieplak,[§] Peter A. Kollman,[‖] and Irwin D. Kuntz*

Department of Pharmaceutical Chemistry, University of California, San Francisco, 513 Parnassus Avenue, San Francisco, California 94143-0446

In this study, we compare the calculated and experimental binding free energies for a combinatorial library of inhibitors of cathepsin D (CatD), an aspartyl protease. Using a molecular dynamics (MD)-based, continuum solvent method (MM-PBSA), we are able to reproduce the experimental binding affinity for a set of seven inhibitors with an average error of ca. 1 kcal/mol and a correlation coefficient of 0.98. By comparing the dynamical conformations of the inhibitors complexed with CatD with the initial conformations generated by CombiBuild (University of California, San Francisco, CA, 1995), we have found that the docking conformation observed in an X-ray structure of one peptide inhibitor bound to CatD (*Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 6796−6800) is in good agreement with our MD simulation. However, the DOCK scoring function, based on intermolecular van der Waals and electrostatics, using a distance-dependent dielectric constant (*J. Comput. Chem.* **1992**, *13*, 505−524), poorly reproduces the trend of experimental binding affinity for these inhibitors. Finally, the use of PROFEC (*J. Comput.-Aided Mol. Des.* **1998**, *12*, 215−227) analysis allowed us to identify two possible substitutions to improve the binding of one of the better inhibitors to CatD. This study offers hope that current methods of estimating the free energy of binding are accurate enough to be used in a multistep virtual screening protocol.

Introduction

Cathepsin D (CatD) is a lysosomal aspartic proteinase. It can cleave β-amyloid precursor protein to release Alzheimer's peptide.[5,6] An elevated level of CatD is also implicated in breast cancer[7,8] and ovarian cancer.[9,10] Structure-based design coupled with combinatorial chemistry has been successfully used to design nonpeptide inhibitors of CatD.[11] On the basis of the X-ray crystallographic structure of human CatD complexed with a natural peptide inhibitor,[2] pepstatin, a structure-based design algorithm, CombiBuild,[1] was used to search for the optimal conformation of a novel scaffold[11] and then to select the side chains to attach to the scaffold. Subsequent experimental measurements demonstrated that using the three-dimensional structure information for the receptor yielded a "hit rate" at 100 nM, seven times that for a library generated by diverse selection.[11] Recent biological studies of the designed inhibitors have shown that they are able to block the production of precursors of neurofibrillary tangles.[12]

The CatD inhibitors described by Kick et al.[11] offer a significant computational challenge to quantitatively evaluate the binding affinity. Although all of the inhibitors in the combinatorial library share a common scaffold, the side chains attached to the scaffold are quite diverse. Thus, the computationally intensive

traditional free energy methods such as thermodynamic integration and the free energy perturbation method are not well-suited to evaluate the binding free energy for this system.[13] Recently, several alternative methods have been developed to estimate the binding free energy in a fast and practical way, including the molecular dynamics (MD) based approach, MM-PBSA,[14] linear interaction energy (LIE) method,[15−19] and some empirical "Ludi"-like approaches.[20,21] Even though the empirical methods are computationally efficient, they lack the physical rigor of the theoretical treatments. Furthermore, the parameters developed from the training set are often not easily transferred to other systems. LIE is a semiempirical method first proposed by Åqvist[15] and further developed by the Jorgensen group.[18] On the basis of linear−response assumptions, the binding free energy is calculated as the combination of weighted electrostatic and van der Waals (vdW) interactions between the ligand and the receptor. Again, the weighting factors of the electrostatic and vdW interaction energies appear to vary for different systems,[18,19] which creates difficulties in using LIE to evaluate a diverse combinatorial library.

In this paper, we calculate the binding of the designed inhibitors to CatD using the MM-PBSA method.[14] This method is one of a class of approaches that attempts to include free energy terms directly into a molecular simulation.[14,19] In previous studies, the MM-PBSA method has shown great promise in its ability to describe the free energy of macromolecular systems at moderate computational cost.[22,23] It combines explicit solvent MD simulations with implicit solvation models, Poisson−Boltzmann (PB) analysis,[24,25] and solvent accessible surface area-dependent nonpolar solvation free

* To whom correspondence should be addressed. Tel: (415)476-1937. Fax: (415)502-1411. E-mail: kuntz@cgl.ucsf.edu.
† Current address: Gustaf H. Carlson School of Chemistry and Biochemistry, Clark University, Worcester, MA 01610.
‡ Current address: Texas Biotechnology Corporation, 7000 Fannin, Houston, TX 77030.
§ Permanent address: Department of Chemistry, University of Warsaw, Pasteura, 1, 02-093, Warsaw, Poland.
‖ Deceased.

energy calculations[26] to estimate free energies. A set of "snapshots" along an MD trajectory for the CatD−inhibitor complex are saved as representative conformations of the macromolecule. The set of structures is then postprocessed with the explicit solvent replaced by continuum solvent models. The free energy of the macromolecular system consists of the molecular mechanics potential energy of the complex, solvation free energies, and an entropy term for the complex. The solvation free energy is composed of an electrostatic or polar portion obtained by solving the PB equation and a nonpolar solvation contribution associated with the formation of a protein-sized cavity in the solvent and vdW interactions between the complex and the solvent. The entropy of the macromolecule can be estimated by either normal mode or quasi-harmonic analyses.[27] Related free energy approaches have been described by Jayaram[28] and Vorobjev.[29] The MM-PBSA approach was successfully used to calculate the binding free energy for protein−ligand association, such as a dianionic hapten bound to the 48G7 antibody Fab fragment,[30] seven biotin analogues bound to avidin,[23] and HIV-1 RT and TIBO derivative association.[31]

We have two goals for this paper. First, we ask if the MM-PBSA method provides a quantitative evaluation of the binding of a set of compounds to CatD with no ad hoc parametrizations of the weighting factors of the electrostatic and vdW interactions. Second, we explore modifications of these compounds that might yield higher binding affinity ligands.

Materials and Methods

1. Molecular Mechanics Model of the Inhibitors. We first need molecular mechanics parameters for the set of inhibitors. The parm94[32] parameter set of the AMBER program[33] does not include the atomic charges and the necessary internal and vdW parameters for these compounds. The parameters were generated as follows. AM1[34] geometry optimization was followed by RHF/6-31G* single point calculation with Gaussian98 (Gaussian, Inc., Pittsburgh, PA) to obtain the electrostatic potentials. Next, the restrained electrostatic potential (RESP) method[35] was used for charge fitting for each inhibitor. Because all of the inhibitors share a common scaffold (Figure 1a), the set of charges of the scaffold atoms for all inhibitors was assumed to be identical. The bond, angle, torsional angle, and vdW parameters not included in parm94 were transferred from the newly developed parm99 parameter set[36] and a generalized force field.[37]

2. Protocols of the MD Simulations. The starting conformations for the MD simulations of CatD−inhibitor complexes were generated by CombiBuild. The acetyl and *N*-methyl groups were patched at the N terminus and C terminus of chain A and chain B, respectively, using Sybyl (Tripos Associates, St. Louis, MO). The N-linked oligosaccharides of CatD were removed. Because the binding interface of CatD and its ligand is similar to that of plasmepsin II-pepstatin, we protonated His77 and Glu260.[38] The electronic structure calculations on the penicillopepsin−substrate complex suggested that the Asp33 outer oxygen was protonated.[39] Because of the high structural similarity at the active site among the aspartic proteinases, we chose the Asp33 outer oxygen to be protonated in CatD−inhibitor complexes. Seven ligands were chosen from the combinatorial libraries of CatD inhibitors with the K_i values ranging from nanomolar to micromolar.[11] Each complex was partially solvated with a 22 Å cap of TIP3P[40] water molecules centered at the inhibitor. Atoms beyond 16 Å of the inhibitor were frozen during the MD simulation. All-atom force field parameters of AMBER were used with a 12 Å nonbonded cutoff. All bond lengths involving hydrogen atoms

were constrained with the SHAKE algorithm.[41] The time step was 1.5 fs. The nonbonded pair list was updated every 25 steps. The temperature of the system was regulated with the Berendsen[42] coupling algorithm with a coupling constant of 0.2 ps. Each complex was simulated using the following protocol: (i) the system was minimized by 30 steps of steepest descent followed by 970 steps of conjugate gradient to release the bad contacts. (ii) After the minimization, the system was gradually heated to 300 K with a 5 ps interval per 100 K. (iii) The production MD trajectory was collected for 450 ps with each snapshot saved every 1.5 ps after a 30 ps equilibration. We have checked the adequate convergence for the measured quantities, such as potential energy as a function of time, and equilibration for each simulation. All of the simulations are well-converged in the production runs. Because of the limited space, we only show the figure of potential energy vs simulation time of the CatD−EHMA complex in the Supporting Information.

The protein conformations of the free CatD and the CatD−pepstatin complex detected by X-ray crystallography are almost identical except for the flexible flap region.[2] Furthermore, the side chains of the inhibitors are relatively rigid. Thus, we assumed that CatD and its inhibitor undergo limited conformational changes from the unbound to the bound state. As a result, the conformations of the free CatD and inhibitor can be extracted from the trajectory of the corresponding complex for the $\Delta G_{binding}$ calculation.

3. $\Delta G_{binding}$ Calculation Methods. The binding free energy is defined as

$$\Delta G_{binding} = G(complex) - [G(CatD) + G(inhibitor)] \quad (1)$$

$$G(molecule) = \langle E_{MM}\rangle + \langle G^{polar}_{solvation}\rangle + \langle G^{nonpolar}_{solvation}\rangle - TS \quad (2)$$

$$\langle E_{MM}\rangle = \langle E_{internal}\rangle + \langle E_{electrostatic}\rangle + \langle E_{vdW}\rangle \quad (3)$$

The conformational space of the active site of the CatD−inhibitor complex was sampled by the MD simulation. $\langle\ \rangle$ denotes an average over a set of snapshots along an MD trajectory. $E_{internal}$ includes the bond, angle, and torsional angle energies, while $E_{electrostatic}$ and E_{vdW} denote intramolecular electrostatic and vdW energies, respectively. In the analysis of $\Delta G_{binding}$, the water molecules were replaced with the implicit solvation models. The intramolecular electrostatic and vdW interactions were calculated using AMBER with no cutoff. The polar contribution to the solvation free energy ($\Delta G^{polar}_{solvation}$) was calculated by solving the PB equation with the Delphi program.[25] For this calculation, PARSE[43] vdW radii and parm94 charges were used. The grid spacing was set to 0.5 Å. The molecule filled 80% of the grid box. To ensure that the maximum change in potential was less than 0.001 kT/e, 500 iterations were performed. The dielectric constants inside and outside the molecule were 1.0 and 80.0, respectively. We compared the $\Delta G_{binding}$ under the conditions of ionic strength equal to 0 and 100 mM[11] for CatD−EHMA. Because the difference in $\Delta G_{binding}$ was not significant (0.25 kcal/mol), we used an ionic strength of zero for all of the PB calculations in this paper. The nonpolar solvation contribution is described as[43]

$$\Delta G^{nonpolar}_{solvation} = \gamma A + b$$

where A is the solvent accessible surface area calculated by the MSMS program,[26] and γ and b are 0.005 42 kcal/mol Å² and 0.92 kcal/mol, respectively.[43] The probe radius was 1.4 Å.

In equation 2, S is the molecular entropy. The vibrational component of this entropy can be calculated with normal mode analysis. The solvent entropy changes were included in the polar and nonpolar solvation free energy terms. Because of the large size of CatD, we used a procedure previously reported[23,30,31] where residues lying more than 8 Å away from the inhibitor were removed before minimization and normal mode analysis. The corresponding structures of the free CatD
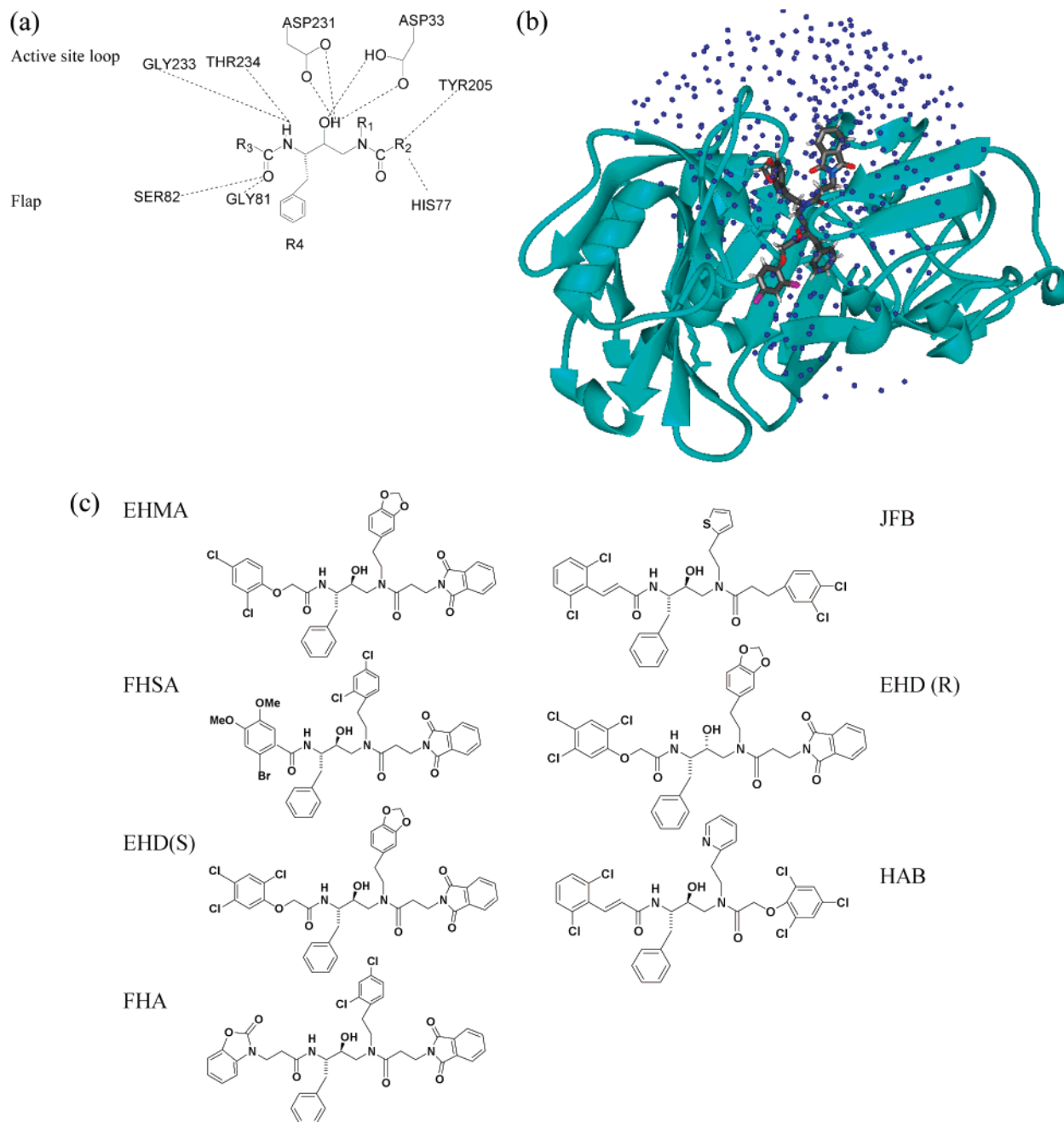
Figure 1. (a) Structure of the scaffold and hydrogen bond networks observed throughout the simulations of all of the CatD−inhibitor complexes. For all of the hydrogen bonds shown here, the fraction of time when a hydrogen bond is well-formed throughout the simulations is greater than 60%. (b) The structure of CatD−EHMA complex. The two aspartic acids at the catalytic center are explicitly shown. The water molecules are shown as dots. (c) The structures of the inhibitors.

and its inhibitor were extracted from the reduced size complex. Each of these structures was minimized with a distance-dependent dielectric constant ($\epsilon = 4r$) to include solvent screening effects.[44−46] The translational and rotational entropies of the protein, ligand, and their complex were readily calculated according to statistical mechanics.[47]

One could, in principle, run separate trajectories of the free enzyme, free ligand, and ligand enzyme complex and then evaluate G with equation 2 for each of these three trajectories. This is computationally demanding and, because the intramolecular terms do not cancel with this approach, subject to structural noise. Thus, this protocol is not practical for the MM-PBSA method.

Results and Discussion

In general, there are two key requirements for the computer-aided structure-based drug design methods:

correctly generating conformations of docked ligands and accurately predicting the binding affinity. In the following sections, we will compare the ligand conformations generated by CombiBuild and those sampled by MD simulations. The performance of the DOCK scoring function[3] will also be addressed.

1. Simulations of the CatD−Inhibitor Complexes. Unlike HIV protease, which is a dimer, CatD is a monomer in a two chain form due to a posttranslational cleavage event.[48,49] Each chain contributes one aspartic acid to the catalytic center. More specifically, D33 is located in chain A while D231 is in chain B. Figure 1b shows the initial conformation of the CatD−EHMA complex. The structures of the nonpeptide inhibitors studied in this paper are shown in Figure 1c.

Table 1. RMS Deviation (Å) of Each Simulation from Its Starting Structure

	$C\alpha^a$		flap[b]		inhibitor (all-atom)		scaffold	
CatD−inhibitor	mean	σ	mean	σ	mean	σ	mean	σ
CatD−EHMA	0.69	0.05	0.76	0.12	1.91	0.54	0.79	0.25
CatD−FHSA	0.95	0.05	1.45	0.22	1.06	0.16	0.67	0.13
CatD−(S)−EHD	1.01	0.07	1.03	0.14	1.51	0.17	0.72	0.15
CatD−FHA	0.66	0.06	0.92	0.13	1.47	0.17	0.47	0.11
CatD−JFB	1.19	0.12	1.14	0.21	1.84	0.16	0.86	0.14
CatD−(R)−EHD	1.08	0.07	1.61	0.25	1.61	0.20	1.04	0.35
CatD−HAB	0.59	0.04	0.95	0.17	2.25	0.28	0.93	0.15

[a] The rms deviation of the protein does not include the atoms in the frozen region during the simulation. [b] The flap region corresponds to residues 75−83 in CatD−pepstatin (pdb code: 1LYB).

Table 2. Hydrogen Bonds between the Inhibitors and the CatD

	hydrogen bond[a]				
	scaffold				
CatD−inhibitor	initial structure	peptide bond	hydroxyl group	R2	K_i (nM)[c]
CatD−EHMA	D33 D231 S82 G233	G81, T234	D231	H77, Y205	3.0
CatD−FHSA			D231	Y205	5.2
CatD−(S)−EHD		G81, T234	D33, D231(2)[b]		73
CatD−FHA		G81, S82	D33, D231(2)	Y205	231
CatD−JFB			D231		~10^3
CatD−(R) −EHD					>5000
CatD−HAB		G81, G233	D33, D231(2)		>10^4
CatD−EHMA (H53→ NH$_2$)		G81, S82	D33, D231(2)	Y205	
CatD−EHMA (H52 → F)		G81, T234	D231		

[a] We consider the hydrogen bond defined by distances between the heavy atoms of donor and acceptor of no more than 3.2 Å and the angles of donor and acceptor diatomic groups of no less than 120°. For the hydrogen bonds listed here, the fraction of time when a hydrogen bond is well-formed during the trajectory is greater than 60%. [b] The letter code and the Arabic number denote the residue name and residue number of CatD. D231(2) denotes that residue D231 forms two hydrogen bonds with the inhibitor. [c] The experimental inhibition constants are cited from ref 11.

Except for (R)−EHD, the inhibitors are S epimers at the hydroxyl carbon of the scaffold. The R_4 side chain is a benzyl ring for all of the inhibitors studied in this paper. The R_1, R_2, and R_3 substituents are five-member rings, six-member rings, or fused five- and six-member rings. The C_α root mean square (rms) deviation of CatD−inhibitor from the X-ray structure of CatD−pepstatin for each simulation is listed in Table 1. The magnitude of the C_α rms deviation of the CatD−inhibitor complexes is on the order of 1 Å. The rms deviation of the flap region for each simulation is of the same order of magnitude as other protein regions, but the standard deviations of the rms of the flap are significantly larger than those of other protein regions. This implies that the flap is more flexible than other protein regions. The all-atom rms deviations of the inhibitors of the complexes from the initial docking structures are in the range of 1−2 Å, which indicates that the initial conformations of the nonpeptide inhibitors generated by CombiBuild are reasonable. The rms deviation of the scaffold of each inhibitor from the CombiBuild structure is less than 1 Å except the scaffold of CatD−(R)−EHD whose R configuration is unfavorable as shown in Figure 1c. We conclude that the atomic position of the scaffold built on the basis of the X-ray structure of pepstatin complexed with CatD was a good representation of the dynamic conformation of the nonpeptide inhibitor bound to CatD.

In Table 2 are listed the hydrogen bonds between the inhibitors and the CatD. The hydrogen bond between the D231 and the hydroxyl group of the S scaffold is persistent during the simulations. For CatD−(S)−EHD, CatD−FHA, and CatD−HAB, D33 also maintains a good hydrogen bond with the hydroxyl group. However, the hydroxyl group of the R epimer is not able to form any hydrogen bond with either D33 or D231. The carbonyl oxygen of the peptide bond in the S scaffold connects with the G81 and/or S82 backbone hydrogen in the following complexes: CatD−EHMA, CatD−(S)−EHD, CatD−FHA, and CatD−HAB. The G233 backbone oxygen forms a hydrogen bond with the amino group of the CatD−HAB scaffold. In addition, the hydrogen bonds between the hydroxyl group of T234 and the amino group of the S scaffold are also observed in the CatD−EHMA and CatD−(S)−EHD complexes. Persistent hydrogen bonds involving R_1 and R_3 substituents are not found during the simulations. However, one of the oxygen atoms of the R_2 substituent of EHMA, FHSA, and FHA is seen to maintain a good hydrogen bond with the hydroxyl group of Y205, while the other oxygen atom of the R_2 substituent of EHMA interacts with the H77 backbone hydrogen. It is interesting that the (identical) scaffold of the S epimeric inhibitors forms different hydrogen bond networks with the enzyme depending on the substituents. In the initial structures of the simulations, the positions of the scaffold are identical for all of the S epimeric inhibitors. Initially, the hydrogen bond networks include inhibitor(OH)−D33, inhibitor(OH)−D231, inhibitor(NH)−G233, and inhibitor(CO)−S82. After nearly half a nanosecond simulations, the rms deviation of the scaffold of each of the S epimeric inhibitors from the initial structure is small (<1 Å), but

Table 3. Calculated Binding Free Energies vs Experimental Data

| CatD−inhibitor | ΔE_{vdW}[a] | ΔG_{ele} | | ΔG_{SA} | $-T\Delta S$[b] | ΔG_{calcd} | ΔG_{exp}[c] |
		ΔE_{ele}	ΔG_{PB}				
CatD−EHMA	−74.4	54.9		−8.9	16.9	−11.5	−11.7
		−58.3	113.2				
CatD−FHSA	−75.7	56.9		−8.6	14.5	−12.9	−11.4
		−30.8	87.7				
CatD−(S)−EHD	−81.2	63.1		−9.1	17.4	−9.8	−9.8
		−72.7	135.8				
CatD−FHA	−75.8	57.3		−9.1	18.9	−8.7	−9.1
		−46.6	103.9				
CatD−JFB	−71.3	54.2		−8.7	19.2	−6.6	∼−8.2
		−24.5	78.7				
CatD−(R)−EHD	−85.0	71.6		−8.9	16.6	−5.7	>−7.3
		−39.0	110.6				
CatD−HAB	−64.8	52.1		−8.3	15.9	−5.1	>−6.9
		−54.1	106.2				

[a] All energies are in kcal/mol. All of the symbols are explained in the text. [b] $T = 298.15$ K. [c] ΔG_{exp} is calculated from K_i (ref 11). DOCK scores of the above inhibitors (from the top to the bottom): −51.79, −36.31, −55.48, −56.44, −49.41, N/A, −50.04 kcal/mol. The DOCK scores are the raw data of ref 11 provided by A. G. Skillman. Because the R scaffold was predicted to be unfavorable to the binding at the stage of scaffold design, CatD−(R)−EHD was not evaluated in the side chain screening.

these hydrogen bond networks become divergent due to the different conformations of the substituents attached to the scaffold. Although the R_3 and R_4 substituents of JFB and HAB are identical, their scaffolds form different hydrogen bond networks with the enzyme due to the difference in R_1 and R_2 substituents of these two ligands. The effect of the side chain conformation on the scaffold position was not considered in CombiBuild except through minimization. According to our present work, it is important to consider the hydrogen bond formation dynamically. While there is no direct correlation between the number of hydrogen bonds formed by the inhibitors and CatD and the experimental inhibition constants, it is certainly unfavorable for binding if the hydroxyl group in the scaffold does not form any hydrogen bond with the active site aspartic acids, for example, CatD−(R)−EHD. Therefore, it appears that the scaffold(OH)−D33/D231 hydrogen bond formation is a necessary but not sufficient factor for high binding affinity.

2. Binding Free Energies of CatD and Inhibitors. The binding free energies of the set of inhibitors to CatD are shown in Table 3. The average unsigned error between the calculated and the experimental data is 1.0 kcal/mol, and the square of the correlation coefficient is 0.96. Both the intermolecular vdW and the electrostatics are important contributions to the binding. However, the electrostatic desolvation penalty (ΔG_{PB}) offsets the favorable (negative) intermolecular electrostatics, yielding an unfavorable net electrostatic contribution to the binding. Particularly for the R epimer, CatD−(R)−EHD, the hydroxyl group in the scaffold is buried inside the catalytic center, but it is not able to form any hydrogen bond with the two aspartic acids. As a result, the net electrostatic contribution to the binding free energy of (R)−EHD and CatD is the most unfavorable among the inhibitors, consistent with the observation that R−EHD bound at worse than 5 μM.[11] The scaffold amido group of FHSA and JFB does not form any persistent hydrogen bond with the surrounding residues, and the intermolecular electrostatic interactions are accordingly weak. However, the desolvation penalties upon binding of these two inhibitors are also small. Thereby, the net electrostatic contributions

Table 4. Correlation Coefficients between the Calculated Binding Free Energy/Components and the Experimental Binding Affinity

energy term	ΔE_{vdW}	ΔE_{ele}	$\Delta E_{vdW} + \Delta E_{ele}$	$\Delta G_{MM-PBSA}$[a]	ΔG_{calcd}
r	0.16	0.16	0.20	0.87	0.98

[a] $\Delta G_{MM-PBSA} = \Delta E_{vdW} + \Delta E_{ele} + \Delta G_{solvation}^{polar} + \Delta G_{solvation}^{nonpolar}$.

to the binding for these two inhibitors are not significantly unfavorable.

We calculated the correlation coefficient between the experimental binding affinity and our calculated $\Delta G_{binding}$ as well as some components of $\Delta G_{binding}$. As seen in Table 4, the calculated $\Delta G_{binding}$ is in good agreement with the experimental results. However, the correlation between the experimental binding affinity and the sum of the intermolecular vdW and electrostatics ($\Delta E_{vdW} + \Delta E_{ele}$) is poor. These two terms form the scoring function of CombiBuild.[3] When the polar and nonpolar solvation free energy terms are added to the intermolecular interactions, the sum of intermolecular vdW, net electrostatics, and nonpolar solvation free energy changes ($\Delta G_{MM-PBSA}$) reaches good agreement with the experimental data with the correlation coefficient equal to 0.87. The fluctuation of the $\Delta G_{MM-PBSA}$ is ca. 14%. We have also noted improved scoring using a generalized Born model of electrostatics.[21] Despite the mediocre performance of the scoring function of Combi-Build in ranking the good inhibitors, it did show enrichment of the better substituents.[50] It is also encouraging that the generated ligand conformations are close to the dynamic conformation with a rms deviation in the range of 1−2 Å, Table 1.

3. Assumptions About Minimal Conformational Changes. As mentioned in the Materials and Methods section, we assume that CatD and its inhibitors undergo limited conformational changes from the unbound to the bound state. The high similarity between the X-ray structures of the free catD and the catD−pepstatin complex strongly support our assumption. To verify the limited conformational changes of the inhibitors from the unbound to the bound state, we performed an MD simulation of an EHMA ligand in a box of water with particle mesh Ewald summation[51] to treat the long-range electrostatic interaction and with a 9 Å cutoff for
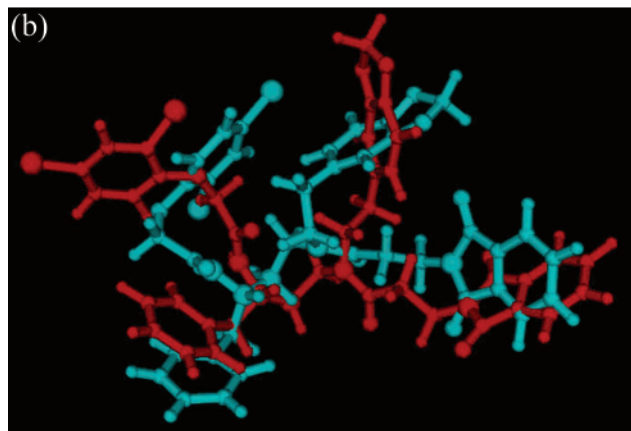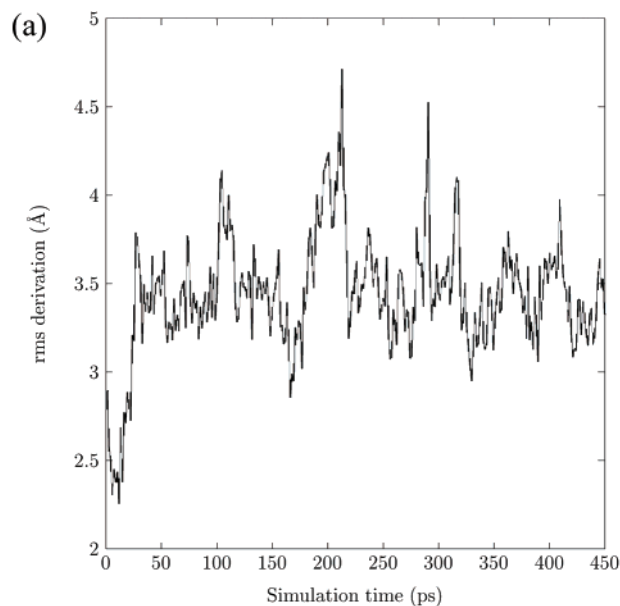
Figure 2. (a) All-atom rms deviation of the free ligand from the initial docking structure of the bound state as a function of simulation time. (b) Superposition of the average structures of EHMA (cyan) in water and complexed with catD (red).

the short-range nonbonded interaction. The production run of this simulation is 450 ps. The all-atom rms deviation of the free ligand from the initial docking structure of the bound state as a function of simulation time is shown in Figure 2a. During the simulation period, the mean of the rms deviation is 3.4 Å. The largest deviation, 4.7 Å, occurs at 142 ps. The rms deviation of the heavy atoms between the average structures of the unbound and bound state is 3.4 Å. The large deviations are found in the R_1 and R_3 side chains (Figure 2b). Therefore, we think our assumption of minimal conformational change is reasonable (at least) in this case.

4. PROFEC Analysis. The PROFEC free energy estimation software[4] was used to suggest how the inhibitors could be modified to improve the binding affinity. We chose the best inhibitor, EHMA, as a template. By postprocessing dynamics trajectories of the inhibitor in solution and complexed with CatD, the PROFEC estimated the free energy cost of adding a test particle to the inhibitor. The analysis of the region around the R_1 side chain revealed no modifications. The first interesting suggestion from PROFEC was that the oxygen linker of R_3 (Figure 3) should be substituted with
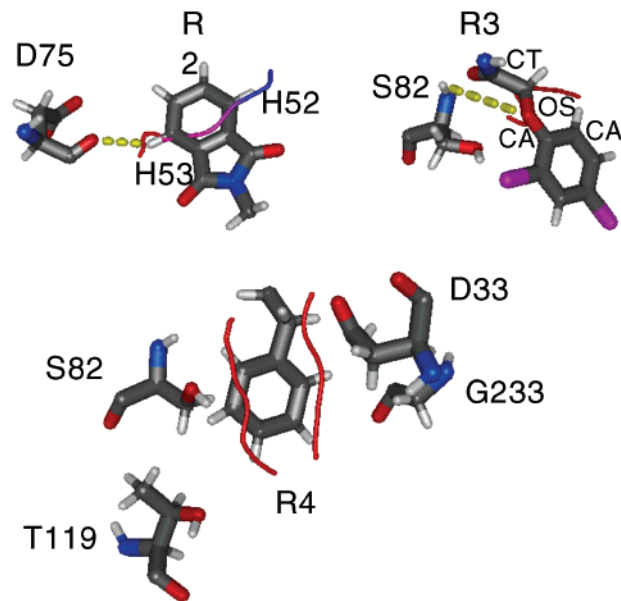

Figure 3. PROFEC contour map of zero vdW interaction for inhibitor EHMA. For clarity, only the slice of contour map through the plane of the ring is shown. The color of the contour line indicates the sign of charge that the test particle should have. Blue indicates a negative charge, while red denotes a positive charge. The dash line denotes potential hydrogen bonds if the ligand is modified as PROFEC suggested. For the designed compound, EHMA (H53 → NH₂), the amino group contacts with Asp75 (O) with the average distance of NH···O is 3.55 (±0.32) Å throughout the simulation.

a positively charged atom, but the conformation of the dichloro-phenyl ring is constrained by the oxygen linker. On the basis of the previous ab initio calculation of the gas-phase torsional barrier of anisole, the conformers with the methoxy substitute coplaner with the phenyl ring corresponded to free energy minima.[52] When the chloro substitution occurred ortho, the torsional angle, CT−OS−CA−CA (connecting to H), is restrained at 0°. Because of the large conformational change to be caused by replacing the oxygen linker, which was not considered in PROFEC, we rejected this suggestion. The next area considered by PROFEC is the replacement of H with large groups for the R_4 side chain and the addition of positively charged substituents to the phenyl ring. Although PROFEC detects little room for larger groups at the R_4 position, some new potent CatD inhibitors synthesized by the Ellman group do have large side chains at the R_4 position,[53] presumably tolerated by rearrangements in the protein that PROFEC does not explore. The most practical modification suggested by the PROFEC is in the R_2 position as shown in Figure 3. As suggested by the PROFEC, H53 should be changed to a positively charged group and H52 should be replaced with a negatively charged group. As seen in Figure 3, D75 is a potential hydrogen bond acceptor; therefore, we designed a compound with H53 replaced with an amino group. Another new compound is a fluorine substituent at the H52 position. With the RESP charge fitting scheme,[35] we kept the charges of the scaffold and R_1, R_3, and R_4 unchanged. Then, we performed new MD simulations of these two designed compounds complexed with CatD following the same protocol as mentioned in the Materials and Methods

Table 5. Calculated Binding Free Energy for the Designed Compounds

CatD–inhibitor	ΔE_{vdW}	ΔG_{ele}		ΔG_{SA}	$-T\Delta S$	ΔG_{calc}
		ΔE_{ele}	ΔG_{PB}			
CatD–EHMA	−83.0	62.3		−9.0	16.9	−13.1
(H53 → NH$_2$)		−63.1	125.4			
CatD–EHMA	−82.6	63.1		−8.9	16.9	−11.5
(H52 → F)		−48.7	111.8			

section. The binding free energies for the designed compounds are listed in Table 5. When H53 is replaced with an amino group, the intermolecular vdW and electrostatic interactions become more favorable, while more desolvation penalty is paid. The binding free energy is improved by 1.3 kcal/mol. Because the charge distribution of the R$_2$ substituent changed completely after introducing the amino group to the ring, the conformation of the whole ligand changed accordingly. As a result, the hydrogen bond network between the ligand and the CatD is different from the original with one more hydrogen bond formed, Table 2. However, the increase in favorable intermolecular electrostatics (4.8 kcal/mol) was offset by the high cost of desolvation (12.2 kcal/mol) even though the amino group was partially exposed to the solvent. On the other hand, the favorable intermolecular vdW interaction increased by 8.6 kcal/mol. However, the fluoro substitution does not improve the binding. Even though relative to EHMA this substituent improves the intermolecular vdW interaction by 8.2 kcal/mol, the intermolecular electrostatics becomes 9.6 kcal/mol less favorable than that of EHMA while the desolvation cost is slightly less than that of EHMA.

In summary, the use of PROFEC analysis allowed us to identify two possible substitutions to improve the binding of EHMA and CatD. Subsequent MD simulations of the modified ligands complexed with CatD allow for the rearrangement of the binding pocket in response to the substitutions. As a result of this process, it appears that the amino substitution at the R$_2$ position ought to result in a more tightly binding ligand.

Conclusion

In this study, we have investigated the binding free energies between a set of inhibitors synthesized by combinatorial chemistry to CatD. In brief, we have shown that the original model-built inhibitor structures, based on the crystal structure of a peptidic inhibitor complexed to CatD, are consistent with the geometries generated during solvated MD simulations. While this result does not replace further crystallographic studies, it re-enforces the conclusion of Kick et al.[11] that structure-based starting points can lead to useful library design.

Second, we have shown that an improved scoring protocol that retains most of the underlying contributions to the free energy of binding yields a quite accurate correlation with the experimental data over a chemically diverse set of substituents: an average error of 1 kcal/mol and a correlation coefficient of 0.98. Analyses of these calculations and of the original DOCK scoring functions[3] show that no single component of the free energy, nor the interaction energy, correlated well with the experimental results, speaking to the need for quantitative treatment of entropic contribution. The

accuracy of the MM-PBSA method comes at moderate computational cost (150 CPU time/ligand for a single processor on Origin 2000) as compared to the full scale thermodynamic integration and free energy perturbation methods. No ad hoc weighting factors of the electrostatic and vdW interactions were needed for this study.

The success of the MM-PBSA approach suggests a multistep strategy for overall ligand discovery/design/optimization protocol. The first step is the use of a crude but rapid method such as DOCK or CombiBuild to scan large databases or construct virtual libraries. Reduction of the candidates from the 10^5-10^9 range to the 10−1000 range then permits the use of a more accurate estimate of the free energy of binding. These studies can be coupled with additional tools such as PROFEC to suggest ligand modifications.

The most significant challenges ahead are to reduce the time requirements of the MM-PBSA step, to improve the calculation of the conformational entropy, and to sample more completely the configurational space of the ligand–receptor complex, with the ultimate aim being an accurate virtual high-throughput screen.

Supporting Information Available: Structures and atomic charges for each inhibitor, force field parameters that are not included in parm94, and potential energy as a function of simulation time for the CatD–EHMA complex. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Roe, D. C. Application and development tools for structure-based drug design. Ph.D. Thesis, Department of Pharmaceutical Chemistry, University of California, San Francisco, 1995.
(2) Baldwin, E. T.; Bhat, T. N.; Gulnik, S.; Hosur, M. V.; Sowder, R. C., II; Cachau, R. E.; Collins, J.; Silva, A. M.; Erickson, J. W. Crystal structures of native and inhibited forms of human cathepsin D: implications for lysosomal targeting and drug design. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 6796−6800.
(3) Meng, E. C.; Shoichet, B. K.; Kuntz, I. D. Automated docking with grid-based energy evaluation. *J. Comput. Chem.* **1992**, *13*, 505−524.
(4) Radmer, R. J.; Kollman, P. A. The application of three approximate free energy calculations methods to structure based ligand design: Trypsin and its complex with inhibitors. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 215−227.
(5) Ladror, U. S.; Snyder, S. W.; Wang, G. T.; Holzman, T. F.; Krafft, G. A. Cleavage at the amino and carboxyl termini of Alzheimer's amyloid-beta by cathepsin D. *J. Biol. Chem.* **1994**, *269*, 18422−18428.
(6) Dreyer, R. N.; Bausch, K. M.; Fracasso, P.; Hammond, L. J.; Wunderlich, D.; Wirak, D. O.; Davis, G.; Brini, C. M.; Buckholz, T. M.; König, G.; et al. Processing of the pre-beta-amyloid protein by cathepsin D is enhanced by a familial Alzheimer's disease mutation. *Eur. J. Biochem.* **1994**, *224*, 265−271.
(7) Rochefort, H.; Liaudet, E.; Garcia, M. Alterations and Role of Human Cathepsin D in Cancer Metastasis. *Enzyme Protein* **1996**, *49*, 106−116.

(8) Westley, B. R.; May, F. E. B. Cathepsin D and Breast Cancer. *Eur. J. Cancer* **1996**, *32A*, 15−24.

(9) Garcia, M.; Platet, N.; Liaudet, E.; Laurent, V.; Derocq, D.; Brouillet, J. P.; Rochefort, H. Biological and Clinical Significance of Cathepsin D in Breast Cancer Metastasis. *Stem Cells* **1996**, *14*, 642−650.

(10) Leto, G.; Gebbia, N.; Rausa, L.; Tumminello, F. M. Cathepsin-D in the Malignant Progression of Neoplastic Diseases (Review). *Anticancer Res.* **1992**, *12*, 235−240.

(11) Kick, E. K.; Roe, D. C.; Skillman, A. G.; Liu, G. C.; Ewing, T. J. A.; Sun, Y. X.; Kuntz, I. D.; Ellman, J. A. Structure-based design and combinatorial chemistry yield low nanomolar inhibitors of cathepsin D. *Chem. Biol.* **1997**, *4*, 297−307.

(12) Bi, X. N.; Haque, T. S.; Zhou, J.; Skillman, A. G.; Lin, B.; Lee, C. E.; Kuntz, I. D.; Ellman, J. A.; Lynch, G. Novel cathepsin D inhibitors block the formation of hyperphosphorylated tau fragments in hippocampus. *J. Neurochem.* **2000**, *74*, 1469−1477.

(13) Kollman, P. A. Free energy calculations: Application to chemical and biochemical phenomena. *Chem. Rev.* **1993**, *93*, 2395−2417.

(14) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E., III. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* **2000**, *33*, 889−897.

(15) Aqvist, J.; Medina, C.; Samuelsson, J. E. A new method for predicting binding affinity in computer-aided drug design. *Protein Eng.* **1994**, *7*, 385−391.

(16) Hansson, T.; Aqvist, J. Estimation of binding free energies for HIV proteinase inhibitors by molecular dynamics simulations. *Protein Eng.* **1995**, *8*, 1137−1144.

(17) Aqvist, J. Calculation of absolute binding free energies for charged ligands and effects of long-range electrostatic interactions. *J. Comput. Chem.* **1996**, *17*, 1587−1597.

(18) Smith, R. H., Jr.; Jorgensen, W. L.; Tirado-Rives, J.; Lamb, M. L.; Janssen, P. A.; Michejda, C. J.; Kroeger Smith, M. B. Prediction of binding affinities for TIBO inhibitors of HIV-1 reverse transcriptase using Monte Carlo simulations in a linear response method. *J. Med. Chem.* **1998**, *41*, 5272−5286.

(19) Wang, W.; Wang, J.; Kollman, P. A. What determines the van der Waals coefficient beta in the LIE (linear interaction energy) method to estimate binding free energies using molecular dynamics simulations? *Proteins* **1999**, *34*, 395−402.

(20) Bohm, H. J. The development of a simple empirical scoring function to estimate the binding constant for a protein−ligand complex of known three-dimensional structure. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 243−256.

(21) Zou, X.; Sun, Y.; Kuntz, I. D. Inclusion of solvation in ligand binding free energy calculations using the Generalized-Born model. *J. Am. Chem. Soc.* **1999**, *121*, 8033−8043.

(22) Srinivasan, J.; Cheatham, T. E., III; Cieplak, P.; Kollman, P. A.; Case, D. A. Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate-DNA helices. *J. Am. Chem. Soc.* **1998**, *120*, 9401−9409.

(23) Kuhn, B.; Kollman, P. A. Binding of a diverse set of ligands to avidin and streptavidin: an accurate quantitative prediction of their relative affinities by a combination of molecular mechanics and continuum solvent models. *J. Med. Chem.* **2000**, *43*, 3786−3791.

(24) Gilson, M. K.; Honig, B. Calculation of the total electrostatic energy of a macromolecular system: solvation energies, binding energies, and conformational analysis. *Proteins* **1988**, *4*, 7−18.

(25) Honig, B.; Nicholls, A. Classical Electrostatics in Biology and Chemistry. *Science* **1995**, *268*, 1144−1149.

(26) Sanner, M. F.; Olson, A. J.; Spehner, J. C. Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers* **1996**, *38*, 305−320.

(27) Brooks, B. R.; Janežič, D.; Karplus, M. Harmonic analysis of large systems. *J. Comput. Chem.* **1995**, *16*, 1522−1553.

(28) Jayaram, B.; Young, M. A.; Beveridge, D. L. Free energy analysis of the conformational preferences of A and B forms of DNA in solution. *J. Am. Chem. Soc.* **1998**, *120*, 10629−10633.

(29) Vorobjev, Y.; Almagro, J. C.; Hermans, J. Discrimination between native and intentionally misfolded conformations of proteins: ES/IS, a new method for calculating conformational free energy that uses both dynamics simulations with an explicit solvent and implicit solvent continuum model. *Proteins* **1998**, *32*, 399−413.

(30) Chong, L. T.; Duan, Y.; Wang, L.; Massova, I.; Kollman, P. A. Molecular dynamics and free-energy calculations applied to affinity maturation in antibody 48G7. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 14330−14335.

(31) Wang, J.; Morin, P.; Wang, W.; Kollman, P. A. Use of MM-PBSA in reproducing the binding free energies to HIV-1 RT of TIBO derivatives and predicting the binding mode to HIV-1 RT of efavirenz by Docking and MM-PBSA. *J. Am. Chem. Soc.* **2001**, *123*, 5221−5230.

(32) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, J. K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acid, and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.

(33) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E., III; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. A. AMBER, A package of computer-programs for applying molecular mechanics, normal-mode analysis, molecular-dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.* **1995**, *91*, 1−41.

(34) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. The development and use of quantum-mechanical molecular-models. 76.AM1- A new general-purpose quantum-mechanical molecular-model. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.

(35) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **1993**, *97*, 10269−10280.

(36) Wang, J.; Cieplak, P.; Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **2000**, *21*, 1049−1074.

(37) Wang, J.; Kollman, P. A. Unpublished results.

(38) Xie, D.; Gulnik, S.; Collins, L.; Gustchina, E.; Suvorov, L.; Erickson, J. W. Dissection of the pH dependence of inhibitor binding energetics for an aspartic protease: direct measurement of the protonation states of the catalytic aspartic acid residues. *Biochemistry* **1997**, *36*, 16166−16172.

(39) Oldziej, S.; Ciarkowski, J. Mechanism of action of aspartic proteinases: Application of transition-state analogue theory. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 583−588.

(40) Jorgensen, W. L. Revised TIPS for simulations of liquid water and aqueous solutions. *J. Chem. Phys.* **1982**, *77*, 4156−4163.

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

(42) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.

(43) Sitkoff, D.; Sharp, K. A.; Honig, B. Accurate calculation of hydration free-energies using microscopic solvent models. *J. Phys. Chem.* **1994**, *98*, 1978−1988.

(44) Pickersgill, R. W. A rapid method of calculating charge charge interaction energies in proteins. *Protein Eng.* **1988**, *2*, 247−248.

(45) Mehler, E. L.; Solmajer, T. Electrostatic effects in proteins: comparison of dielectric and charge models. *Protein Eng.* **1991**, *4*, 903−910.

(46) Solmajer, T.; Mehler, E. L. Electrostatic screening in molecular dynamics simulations. *Protein Eng.* **1991**, *4*, 911−917.

(47) McQuarrie, D. A. A. *Statistical Mechanics*; Harper & Row: New York, 1976.

(48) Barrett, A. J. *Proteinases in Mammalian Cells and Tissue*; Elsevier: New York, 1977.

(49) Gulnik, S.; Baldwin, E. T.; Tarasova, N.; Erickson, J. Human liver cathepsin D. Purification, crystallization and preliminary X-ray diffraction analysis of a lysosomal enzyme. *J. Mol. Biol.* **1992**, *227*, 265−270.

(50) Skillman, A. G. J. Structure-based design of combinatorial library. In *Pharmaceutical Chemistry*; University of California, San Francisco: San Francisco, 1999; p 389.

(51) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: A N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.

(52) Spellmeyer, D. C.; Grootenhuis, P. D. J.; Miller, M. D.; Kuyper, L. F.; Kollman, P. A. Theoretical investigations of the rotational barrier in anisole: An ab initio and molecular dynamics study. *J. Phys. Chem.* **1990**, *94*, 4483−4491.

(53) Lee, C. E.; Kick, E. K.; Ellman, J. A. General solid-phase synthesis approach to prepare mechanism-based aspartyl protease inhibitor libraries. Identification of potent cathepsin D inhibitors. *J. Am. Chem. Soc.* **1998**, *120*, 9735−9747.