# Estimation of Binding Affinities for Selective Thrombin Inhibitors via Monte Carlo Simulations

Albert C. Pierce and William L. Jorgensen*

*Department of Chemistry, Yale University, New Haven, Connecticut 06520-8107*

*Received September 15, 2000*

Monte Carlo simulations have been performed on a series of 20 active-site-directed thrombin inhibitors to determine the interactions and energetics associated with the binding of these compounds. Physicochemical descriptors of potential value in the prediction of binding affinities were averaged during simulations of each inhibitor unbound in water and bound to thrombin. Regression equations based on 3−5 descriptors are able to reproduce the experimental binding affinities, which cover a 7 kcal/mol range, with rms errors of 1.0−1.3 kcal/mol, and yield correlation coefficients, $r^2$, of 0.7−0.8. On the basis of these results, the quantities most important in determining the binding affinities are: (1) the enhancement of van der Waals interactions in going from solution to the bound state, (2) the intramolecular strain induced in the inhibitor upon binding, (3) the number of hydrogen bonds lost in the binding process, and (4) the number of rotatable bonds in the inhibitor. The descriptors are physically reasonable and, in combination with the insights gained from analysis of the simulation structures, suggest directions for the development of improved thrombin inhibitors.

## Introduction

Thrombin is a multifunctional member of the trypsin family of serine proteases, and it is the final enzyme of the blood coagulation cascade. The activity of thrombin in this cascade is responsible for the cleavage of fibrinogen to form fibrin and the activation of platelets via the thrombin receptor.[1,2] Fibrin then polymerizes to form a network of fibers, entrapping the platelets and leading to clot formation and cessation of blood flow. Thrombin-induced clot formation is a necessary part of the wound-healing process, but it is also associated with many disease states including myocardial infarcation, pulmonary embolism, and stroke.[3] While anticoagulant therapeutics have traditionally been available in the form of heparin, which is administered by injection, and the orally active coumadin, both of these treatments have significant limitations in safety and efficacy.[4] An injectable form of the thrombin inhibitor argatroban has recently been approved by the FDA, but only for the relatively rare condition of heparin-induced thrombocytopenia.[5] The direct inhibition of thrombin by an orally available small-molecule inhibitor is therefore still a high priority in medicinal chemistry research. The inhibition of any of the closely related serine proteases in the coagulation cascade might be therapeutically effective, and eventually it may be possible to selectively inhibit these proteases for optimal treatment of specific conditions. However, at present the ease with which thrombin can be purified and studied in soluble form, along with the wealth of structural information available, has made it the most intensely studied target for anticoagulation therapy.[6]

Given the abundance of structural information, which includes numerous high-resolution crystal structures of the enzyme complexed with a variety of inhibitors, thrombin is also an ideal target for structure-based drug design. While a cursory examination of these enzyme−inhibitor complexes can give some direction in the design process, a quantitative model for activity prediction should provide a more discriminating test of the inhibitory potential of new compounds. To this end, numerous theoretical studies have been performed to estimate the activities of series of thrombin inhibitors. The methods used in these studies include correlation of binding affinity with interaction energy,[7] active site mapping,[8] comparative molecular field analysis,[9] and linear response calculations.[10] All of these analyses have their merits, but also their associated weaknesses. For example, the interaction energy approach considers only a single optimized conformation for the protein−ligand complex, and the CoMFA model does not explicitly consider the protein structure at all. Among these studies, the linear response calculations are the most thorough in their inclusion of explicit solvent and extensive sampling of protein and ligand conformations. The present work is intended to extend the linear response analysis to a larger set of inhibitors in an effort to create a more complete and accurate model for the prediction of binding affinities.

The linear response method was devised as an alternative to the free energy perturbation (FEP) approach to calculating the free energies of association ($\Delta G_b$) for molecules in solution. The FEP method is theoretically rigorous for the calculation of $\Delta G_b$, explicitly accounting for all enthalpic and entropic contributions to molecular association in the limit of complete sampling in molecular dynamics (MD) or Monte Carlo (MC) simulations. It has proven successful in reproducing experimental binding energies, and analysis of the accompanying structures has generally provided good insight into the structural basis of binding preferences.[11,12] Due to sampling issues, 10 or more simulations are required for a single FEP to determine the relative binding affinity of two similar drugs. This is quite time-consum-

* To whom correspondence should be addressed. Tel: 203-432-6278. Fax: 203-432-6144. E-mail: william.jorgenson@yale.edu.

10.1021/jm000405u CCC: $20.00 
Published on Web 03/02/2001

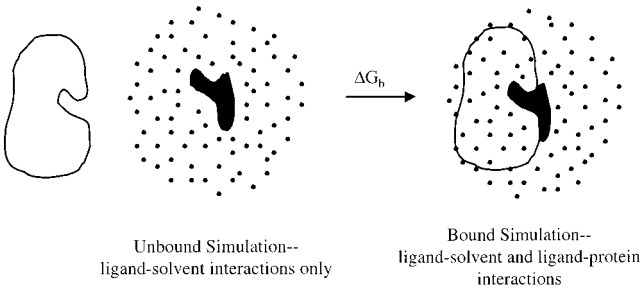


**Figure 1.** Schematic representation of the binding event and the environments of the ligand in the bound and unbound simulations.

ing, but depending upon the required accuracy, it can be worthwhile. However, computation of absolute binding affinities or of relative affinities between two dissimilar drugs requires many more simulations to provide adequate sampling and is generally impractical for protein−ligand complexes.

The 'linear response' (LR) or 'linear interaction energy' method was developed to enable more rapid calculation of absolute binding affinities of diverse sets of ligands, while maintaining the ensemble averaging, explicit solvent, and some of the theoretical rigor of the FEP methodology.[13] The LR method requires only two simulations per inhibitor: one with the inhibitor free in solution, the other with the inhibitor bound to its target protein with the entire complex solvated (see Figure 1). The $\Delta G_b$ is then calculated based on the interaction energies of the ligand with its bound environment as compared to the corresponding energies of the ligand in solution. As originally formulated by Åqvist,[13] the average van der Waals and electrostatic energies are calculated for the ligand interacting with its environment in each of these two states. The free energy of binding is then expressed as a linear combination of these two energy differences according to eq 1:

$$\Delta G_b = \alpha\langle\Delta E_{vdW}\rangle + \beta\langle\Delta E_{Coul}\rangle \quad (1)$$

The parameter $\beta$ was set to 0.5, based on the quadratic form of the free energy response of polar solutions to changes in electric field. There was no theoretical basis for a linear relationship between van der Waals (Lennard−Jones) interaction and binding energy, but it was found that by adjusting $\alpha$ to 0.161, a very reasonable fit to the experimental results was obtained for a set of five endothiapepsin inhibitors.[13]

The methodology has since been applied to a variety of systems with varying degrees of success. In general, the value of $\alpha$ varies with different systems and force fields, and in some studies it was also necessary to adjust $\beta$ in order to reproduce experimental binding energies.[14−16] The method was extended by Jorgensen et al. to include a third parameter based on the change in the solvent-accessible surface area (SASA).[10,17] Eq 2, which includes this term, provided improvement for calculations of ligand-binding free energies, presumably due to a more explicit treatment of the hydrophobic effect:

$$\Delta G_b = \alpha\langle\Delta E_{vdW}\rangle + \beta\langle\Delta E_{Coul}\rangle + \gamma\langle\Delta SASA\rangle \quad (2)$$

However, further studies in our laboratory have found that for free energies of solvation of ca. 200 organic solutes[18] and binding affinities of 40 HIV reverse transcriptase inhibitors,[19] the three terms in eq 2 are not optimal. In these cases, it was found that a general equation of the form of eq 3 was more suitable, with each $\xi_i$ representing a physicochemical descriptor and $c_i$ serving as the scaling factor:

$$\Delta G_b = \sum_i c_i\xi_i + \text{const} \quad (3)$$

In addition to the standard LR terms, some of the descriptors considered were the numbers of donated and accepted hydrogen bonds and the hydrophobic, hydrophilic, and aromatic components of SASA. These values are averaged in MC or MD simulations and taken as the difference between the bound and unbound states where appropriate.

This extended LR equation can be viewed as an empirical scoring function for ligand binding. Indeed, the original LR equation is at best a semiempirical formula with only an approximate theoretical foundation for the electrostatic term. The method also clearly ignores many potentially important contributions to ligand binding such as changes in the internal energy for both ligand and protein and entropic effects.[20] Equation 3 can include such factors, and its use can help identify the key contributors to variations in binding affinities. It is the aim of this work to use eq 3 to generate a model for predicting the binding affinities of active-site-directed thrombin inhibitors.

The generality of any model for predicting inhibitor activities will depend on the set of inhibitors upon which it is based (the training set). Therefore, a diverse set of 20 ligands (Figure 2) has been considered here with activities ranging from 5 $\mu$M to 45 pM.[21−25] As a target for small-molecule inhibitors, the thrombin active site is considered to have three different subsites.[6] The inhibitors under investigation include at least 7 different functional groups in each of these pockets. Inhibitors **1**−**6** and **20** differ for the S1 or specificity pocket, which is a deep cleft in the active site with an aspartate residue at its base to complement the basic functionality of the substrate or inhibitor. While the benzamidine-based side chains of inhibitors **1**−**3** and **4**−**6** are identical, the opposite stereochemistry and additional glycine residue of the second set of inhibitors cause them to bind in a different orientation from the first set.[26] The alternative orientations of the benzamidine side chains of the two groups of inhibitors lead to different interactions within the active site (see Figure 3) such that, from the perspective of compound diversity, the pairs **1**-**4**, **2**-**5**, and **3**-**6** may be considered to have different functional groups in the specificity pocket.

Inhibitors **3** and **6**−**12** are varied in the region of the inhibitor that occupies the P-pocket (proximal) of the binding site. This is a hydrophobic pocket located nearer to the site of cleavage than the second hydrophobic D-pocket (distal). The P-pocket moieties differ mainly in their steric bulk and shape, with the exception of compounds **11** and **12**, which introduce additional hydrogen-bonding functionality. The final set of inhibitors, **13**−**19**, vary in their D-pocket moiety. As this is a hydrophobic region of the binding site, these compounds differ almost exclusively in the size and orientation of the substituent. It should be noted that while all

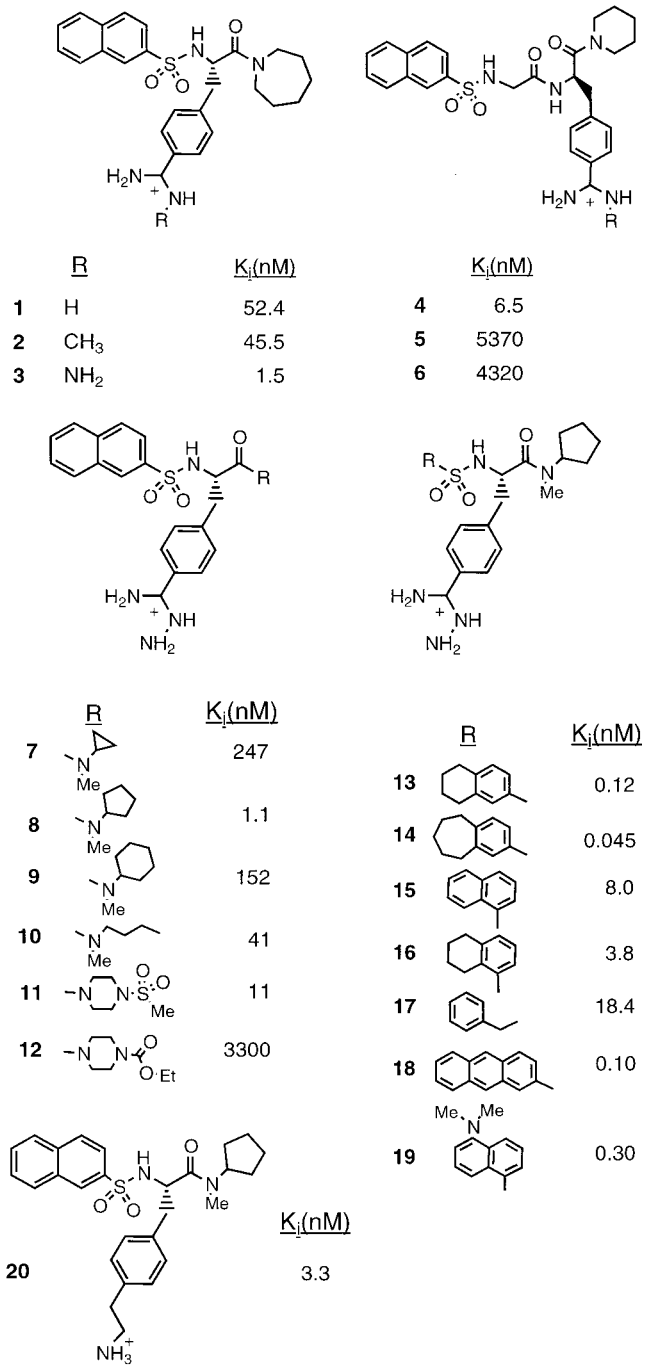


**Figure 2.** The 20 thrombin inhibitors considered and their inhibition constants.

inhibitor $K_i$ values were measured with the same assay, inhibitors **1**−**10** had their activities measured using bovine thrombin and inhibitors **11**−**20** were tested using human thrombin.[21−25] However, as these enzymes are highly homologous, with only an arginine/lysine variation in the active site, observed activity differences are small.[26]

## Computational Details

**Parameters.** All protein residues were represented with the OPLS all-atom (AA) force field. Inhibitor parameters were developed with BOSS 4.2[27] in an automated fashion as follows. Starting with an unoptimized structure of the inhibitor, a PM3 single-point calculation is performed and partial charges are calculated using the CM1P procedure.[28] Atom types from the OPLS-AA force field[29] are also assigned, and they determine the Lennard−Jones parameters for each atom, the bond-stretching parameters for each pair of bonded atoms, the angle-bending parameters for triplets, and the torsion parameters for quartets. A new **Z**-matrix (internal coordinate representation) is output with all degrees of freedom variable and all force-field parameters assigned. This structure is then subject to a molecular mechanics optimization before another PM3 single-point calculation is performed and final charges are assigned. This procedure requires no user intervention, and the resulting parameters have been shown to give good results for the calculation of solvation free energies in a variety of solvents.[18] There is no scaling of the present CM1P charges and each inhibitor has a net charge of +1.

**MC Simulations.** The energy-minimized structures of the inhibitors served as the starting point for the unbound simulations. Each inhibitor was solvated with a 24-Å cap of approximately 1915 TIP4P water[30] molecules centered on a central inhibitor atom. The simulations were carried out with the MC program MCPRO,[31] which incorporates residue-based cutoffs and moves. Each inhibitor was treated as a single residue with all bonds, angles, and dihedrals varied. The majority of attempted moves were translations and rotations of solvent, with a solute move, consisting of translation, rotation, and internal coordinate variations, carried out every 50 configurations.

Calculations performed on HIV-1 reverse transcriptase inhibitors have shown that the averaged Coulombic interaction of the ligand with solvent in the unbound state can be dependent on the starting structure of the ligand.[19] In that study a simulated annealing protocol was developed to allow increased sampling of the ligand and better convergence of the solute−solvent electrostatic interaction. As all of the inhibitors considered in this study carry a net charge, any problems with the convergence of Coulombic interactions are likely to be exacerbated. Consequently, five rounds of the following annealing protocol were applied for the simulation of every unbound inhibitor. In each round, $10^6$ configurations are covered at a temperature of 1000 K with only translational, rotational, and torsional degrees of freedom sampled for the inhibitor. The system is then equilibrated at 298 K for $5 \times 10^6$ configurations sampling all degrees of freedom, followed by $10 \times 10^6$ configurations of averaging. After this period of averaging the cycle is begun again with another high-temperature run. The final results are the averages of the results for the five averaging periods.

Two crystal structures served as the starting points for the protein−ligand complexes. The simulations of inhibitors **4**−**6** were based on the 2.3-Å crystal structure of bovine thrombin with NAPAP (compound **4**), PDB entry 1ETS.[26] All other inhibitors were modeled into the 2.5-Å crystal structure of bovine thrombin with 4-TAPAP, PDB entry 1ETT.[26] 4-TAPAP is similar to inhibitor **1**, but with a *p*-tolyl group in the D-pocket and a piperidinyl group in the P-pocket. For each complex, hydrogen atoms were added to the protein, the inhibitor was modeled into the binding site, and the complete system was conjugate-gradient minimized for 250 steps with a dielectric constant of 4*r*. If the inhibitor conformation was not clearly defined by the bound structure of NAPAP or 4-TAPAP, minimizations were carried out for each reasonable conformation of the inhibitor, and the conformation with the lowest energy was used in the subsequent simulations. Next, all protein residues beyond 16 Å from the inhibitor were removed to leave a more manageable system of 171 protein residues. Clipped N- and C-terminal ends of the protein chain were then capped with acetyl and methylamine groups, respectively. Several basic residues distant from the active site were also neutralized in order for the inhibitor to experience a charge-neutral environment in both the bound and unbound states.

Each of these systems was then solvated with a 24-Å cap of approximately 1030 TIP4P water molecules. All protein residues beyond 12 Å from the inhibitor were frozen, as were all protein bond lengths and backbone degrees of freedom. Residues with at least one atom within 12 Å of the inhibitor had

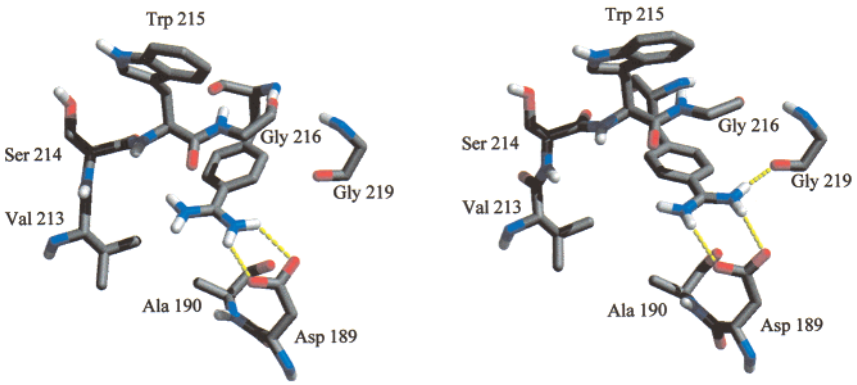


**Figure 3.** Binding site orientation of the benzamidine side chains of inhibitor **1** (left) and inhibitor **4** (right).

their angles and torsions varied, while all degrees of freedom for the inhibitor were varied. The solvent−solvent, solute−solvent, intersolute, and intrasolute nonbonded cutoffs were 9, 12, 12, and 9 Å, respectively, for all simulations. An inhibitor−protein residue interaction list was calculated during the first block of the simulation and held constant throughout the calculation. All charged protein residues were included in this list to ensure that all inhibitors experienced a neutral environment for the entirety of the bound simulation.[13,14] Finally, the systems were equilibrated at 298 K for $10 \times 10^6$ configurations before the averaging of all quantities was carried out over $10 \times 10^6$ configurations.

The quantities that were averaged and considered for the fit of eq 3 to the inhibition data were the differences in Coulombic, Lennard−Jones (van der Waals), and internal energy ($\Delta E_{Coul}$, $\Delta E_{LJ}$, and $\Delta E_{int}$), hydrophobic, hydrophilic, aromatic, and total surface areas (ΔFOSA, ΔFISA, ΔARSA, and ΔSASA), hydrogen bonds donated, hydrogen bonds accepted, and total hydrogen bonds (ΔHBDN, ΔHBAC, and ΔHB), and the number of rotatable bonds (#RB) for the ligand. The SASA calculations used a 1.4-Å solvent probe with solute atomic radii based on the OPLS-AA Lennard−Jones radius; the same probe radius has been used in all of our related studies, though other values could be tried. For the hydrogen bond counts, a hydrogen bond is defined as present when a heteroatom−hydrogen distance is less than 2.5 Å.

These quantities were fit to eq 3 with multiple linear regression in the statistical package JMP.[32] Descriptors were chosen to maximize the value of the squared correlation coefficient ($r^2$) while keeping the total number of descriptors to a minimum. The cross-validated $r^2$, or $q^2$, was calculated by generating a series of 20 fits to the binding data, in each case leaving out one inhibitor. The equation for each fit is then used to 'predict' the value of the inhibitor that was not included in the fit, and the correlation of these predictions with the actual values gives $q^2$.

In all discussions of thrombin structures, residues are referred to according to the chymotrypsin numbering scheme.[26]

## Results

**Regression Equation.** The experimental free energies of binding show the largest-in-magnitude correlation coefficients ($r$) with $\Delta E_{int}$ (0.71), ΔHB (−0.53), $\Delta E_{LJ}$ (0.42), #RB (0.26), ΔSASA (−0.23), and $\Delta E_{Coul}$ (−0.23). The stongest correlations between these descriptors are −0.43 for $\Delta E_{LJ}/\Delta E_{Coul}$, 0.41 for $\Delta E_{LJ}/$ΔSASA, 0.38 for $\Delta E_{int}/$#RB, and 0.37 for ΔHB/ΔSASA. If one takes the first 4 descriptors and thereby avoids the two strongest cross-correlations, the fit in eq 4 yields a correlation coefficient of 0.83, $r^2$ of 0.69, cross-validated $q^2$ of 0.62, rms error of 1.28 kcal/mol, and mean unsigned error of 1.00 kcal/mol for the 20 data points that range over 7.0 kcal/mol in $\Delta G_b$ (Figure 4):

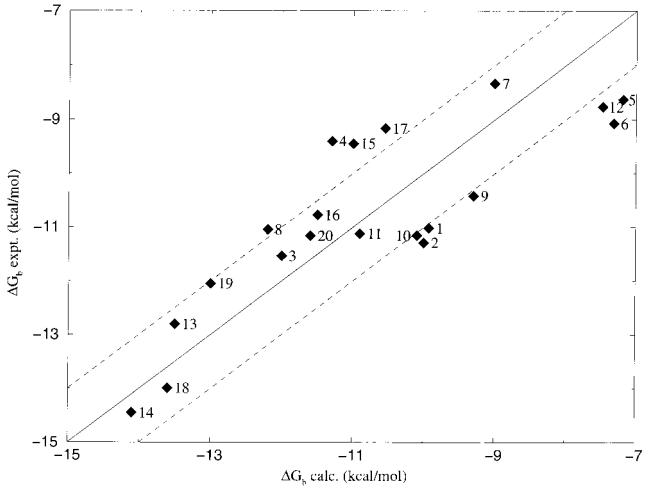


**Figure 4.** Plot of the predicted binding affinities ($\Delta G_b$ calc.) from eq 4 vs the experimental binding energies ($\Delta G_b$ expt.).

$$\Delta G_b = 0.142\langle\Delta E_{int}\rangle - 0.740\langle\Delta HB\rangle + 0.147\langle\Delta E_{LJ}\rangle + 0.441\langle\#RB\rangle - 13.30 \quad (4)$$

If the #RB is dropped, the fit is about the same with an $r^2$ of 0.67 and rms error of 1.29 kcal/mol. There are many alternative fits with 3−5 descriptors that yield regressions with $r^2$ values in the 0.73−0.83 range and only slightly lower $q^2$ values. With $\Delta E_{LJ}$, ΔSASA, and $\Delta E_{int}$, the $r^2$ is 0.73 and rms error is 1.15 kcal/mol, while with those terms plus $\Delta E_{Coul}$ and #RB, eq 5 yields an $r^2$ of 0.82, rms error of 1.00 kcal/mol, and mean unsigned error of 0.67 kcal/mol:

$$\Delta G_b = 0.167\langle\Delta E_{int}\rangle - 0.052\langle\Delta E_{Coul}\rangle + 0.217\langle\Delta E_{LJ}\rangle + 0.636\langle\#RB\rangle - 0.031\langle\Delta SASA\rangle - 32.07 \quad (5)$$

However, we suspect that at this point the data are being overfitted; the ratio of data points to descriptors is only 4, the intercept is large, and the negative coefficients for the $\Delta E_{Coul}$ and ΔSASA terms are not clearly physically reasonable. It may also be noted that eq 2 yields a poor fit for the present data set with an $r^2$ of 0.38 and rms error of 1.75 kcal/mol.

Equation 4 appears more sound because of use of the 4 least-correlated descriptors and because the coefficients of the terms are all physically reasonable. In our study of HIV-RT with 40 inhibitors, the most important descriptors were also found to reflect the protein−ligand Lennard−Jones interaction energy and the change in number of hydrogen bonds, ΔHB.[19] The

HIV-RT inhibitors are comparatively nonflexible, so the $\Delta E_{int}$ and #RB terms are not important in that case. For eq 4, the first term, based on the internal energy change for the ligand, accounts for the fact that an inhibitor is less likely to bind to the protein if conformational strain is induced upon binding. A descriptor for strain energy is intuitively appealing and has proven useful in a number of studies.[33,34] The second term, ΔHB, accounts for the reduction in inhibitor potency associated with the loss of hydrogen bonds on transfer from water to the protein active site. Due to the ideal hydrogen-bonding environment provided by liquid water, all inhibitors form fewer hydrogen bonds in the active site of the protein, and the penalty for binding free energies is 0.74 kcal/mol per hydrogen bond lost. This is in reasonable accord with the results for 40 HIV-1 reverse transcriptase inhibitors, where each hydrogen bond lost resulted in a 0.94 kcal/mol penalty.[19] The third term, $\Delta E_{LJ}$, is the same descriptor used in the original LR equation to describe binding due to enhanced van der Waals contacts in the protein environment, with the coefficient falling quite close to Åqvist's value of 0.161.[13] The last independent variable in the fit is the number of rotatable bonds, which should be correlated with the entropy loss due to conformational restrictions imposed on the ligand in the bound state. In an empirical scoring function based on 45 protein−ligand systems studied by Bohm et al.,[35] this was one of four terms found to be necessary to reproduce binding affinities. In that case, the coefficient for this quantity was 0.34, which is reasonably close to the value of 0.44 found here. Finally, the constant term in the fit includes the effects of binding forces that are not explicitly accounted for in eq 4 and that are presumably similar for all 20 ligands.

Some LR studies have found that it is possible to replace the SASA term with a constant,[14,36] as this quantity varies little in many analogue series. In the present case the constant term may also include the electrostatic attraction, which is attenuated by the ΔHB term. While it would be preferable to explicitly include the term for $\Delta E_{Coul}$, the handling of this quantity has generally been problematic for charged inhibitors,[15,37] and its addition to eq 4 has negligible effect on the present fit. An additional advantage of the ΔHB term is that it converges much faster than $\Delta E_{Coul}$ and is not influenced by long-range effects. Consequently, smaller systems could be studied with significantly shorter periods of configurational averaging. This would reduce the computational cost while still producing an effective model for predicting binding affinities. A fit based on the current simulations after only 1 million configurations of averaging of the bound and unbound systems gives an $r^2$ of 0.65 with the same terms used in eq 4. With some reasonable reduction in the equilibration periods as well, the current processing time of 1−2 days/inhibitor on a 750-MHz PentiumIII processor could be reduced to a few hours. Of course, much faster approaches are available for estimating protein−ligand binding energies that do not involve thermal averaging and explicit representation of the solvent.[20,35,38]

**Analysis of Binding Trends.** Based on eq 4, in combination with the data in Table 1, it is straightforward to assess the importance of the different contributions to binding affinity. The free energies associated with the change in van der Waals (Lennard−Jones) interaction cover a range of 2.4 kcal/mol. For the changes in internal energy, hydrogen bonding, and rotatable bonds the ranges are 4.2, 2.2, and 1.3 kcal/mol, respectively. Thus, each term makes a significant contribution to the overall binding affinities. Also, due to the fact that $\Delta E_{int}$, ΔHB, and #RB all penalize the binding of most inhibitors, the final constant term is necessarily negative. This is in accord with the argument that the constant term is accounting for large and favorable contributions from the burial of surface area and electrostatic attraction, including the salt bridge with Asp189.

**Table 1.** Contributions from Individual Terms in Eq 4 to the Free Energies of Binding[a]

| | contributions from | | | | | |
|---|---|---|---|---|---|---|
| inhibitor | $\Delta E_{LJ}$ | $\Delta E_{int}$ | ΔHB | #RB | $\Delta G_b$ calcd | $\Delta G_b$ expt |
| **1** | −3.76 | 0.97 | 1.55 | 3.53 | −11.02 | −9.93[b] |
| **2** | −3.42 | 0.97 | 0.93 | 3.53 | −11.29 | −10.0[b] |
| **3** | −4.25 | 1.02 | 1.47 | 3.53 | −11.53 | −12.0[b] |
| **4** | −4.32 | 1.29 | 2.52 | 4.41 | −9.40 | −11.3[b] |
| **5** | −3.02 | 1.15 | 2.12 | 4.41 | −8.63 | −7.19[b] |
| **6** | −3.82 | 2.05 | 1.58 | 4.41 | −9.08 | −7.32[b] |
| **7** | −3.66 | 2.13 | 2.52 | 3.97 | −8.33 | −9.01[c] |
| **8** | −4.22 | 0.67 | 1.84 | 3.97 | −11.04 | −12.2[c] |
| **9** | −4.35 | 1.97 | 1.30 | 3.97 | −10.41 | −9.30[c] |
| **10** | −4.53 | 1.21 | 0.61 | 4.86 | −11.16 | −10.1[c] |
| **11** | −4.40 | 0.53 | 2.07 | 3.97 | −11.12 | −10.9[d] |
| **12** | −5.37 | 3.34 | 2.16 | 4.41 | −8.76 | −7.48[d] |
| **13** | −4.70 | 0.22 | 1.01 | 3.97 | −12.80 | −13.5[e] |
| **14** | −4.82 | −0.82 | 0.53 | 3.97 | −14.45 | −14.1[e] |
| **15** | −3.83 | 2.54 | 1.18 | 3.97 | −9.45 | −11.0[e] |
| **16** | −3.85 | 1.14 | 1.28 | 3.97 | −10.77 | −11.5[e] |
| **17** | −3.36 | 2.76 | 0.33 | 4.41 | −9.16 | −10.5[e] |
| **18** | −4.88 | −0.11 | 0.33 | 3.97 | −13.99 | −13.6[e] |
| **19** | −4.76 | 0.83 | 0.77 | 4.41 | −12.04 | −13.0[e] |
| **20** | −3.96 | 0.91 | 0.78 | 4.41 | −11.16 | −11.6[f] |

[a] $\Delta G_b$ calcd from eq 4. $\Delta G_b$ expt = RT ln($K_i$). All quantities in kcal/mol. [b] Ref 21. [c] Ref 22. [d] Ref 23. [e] Ref 24. [f] Ref 25.

For specific binding trends in this set of inhibitors, it is most useful to consider the three sets of inhibitors that vary in their specificity pocket, P-pocket, and D-pocket binding moieties. Inhibitors **1**−**6** and **20** all vary in the basic group that forms the ion pair with Asp189 in the specificity pocket. The benzamidine side chain of compound **1** ($K_i$ = 52.4 nM) mimics the arginine side chain of the preferred substrate and served as the starting point for the experimental exploration of this binding pocket.[21] As illustrated in Figure 3, this inhibitor forms two hydrogen bonds in the S1 pocket, leaving two potential hydrogen bond donors unsatisfied. The loss of two hydrogen bonds is clearly a liability for this inhibitor, and its potency could likely be enhanced by the elimination of one or more of these unpaired hydrogen bond donors. The replacement of the benzamidine side chain with benzylamine accomplishes this goal in compound **20** ($K_i$ = 3.3 nM), while the same end is achieved for inhibitor **2** ($K_i$ = 45.5 nM) with the substitution of a methyl group for one of the amidinium protons not involved in a hydrogen bond. However, the methyl group that replaces the amidinium proton in compound **2** is too bulky for the binding pocket (see $\Delta E_{LJ}$ in Table 1) and there is no gain in potency. In contrast to inhibitors **2** and **20**, compound **3** ($K_i$ = 1.5 nM) has its hydrogen bonding optimized through the introduction of an additional hydrogen-bonding amine function-

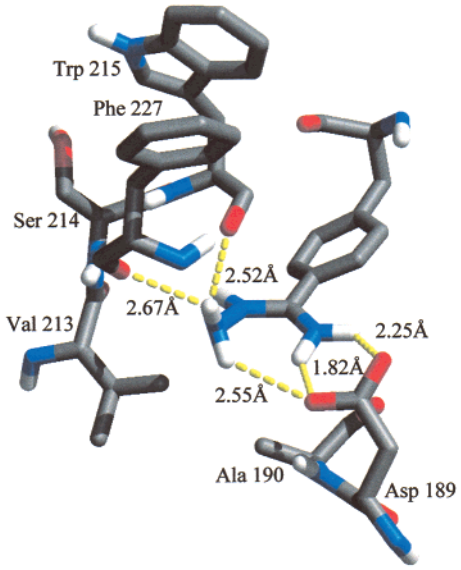


**Figure 5.** Hydrogen-bonding pattern of the benzamidrazone side chain of inhibitors **3** and **7**−**19**.

ality in the S1 pocket. The amine substitution leads to the burial of two additional hydrogen-bonding atoms relative to compound **1**, but due to the orientation of this amine group (see Figure 5) additional hydrogen bonds can be formed with Asp189 and the backbone carbonyl groups of Trp215 and Phe227. It is this net gain of one hydrogen bond with no accompanying steric penalty that leads to enhanced activity for compounds with this side chain. As this side chain also confers selectivity over the closely related serine protease trypsin, it was used as the S1 moiety for the majority of the subsequently developed inhibitors in this series.[21] Due to the different binding mode of the NAPAP-based inhibitors (**4**−**6**), substitutions at the benzamidine side chain lead to steric and strain penalties, which reduce the potency of **5** and **6**.

Compounds **7**−**12** are varied in the region of the inhibitor that binds in the hydrophobic P-pocket, which is defined by residues His57, Tyr60A, Trp60D, and Ile174. Among this set of inhibitors relative binding affinities are determined largely by the steric fit of the inhibitors in the binding pocket and intramolecular strain induced in achieving this fit. Also, in the case of compounds **11** and **12**, the burial of two potential hydrogen bond acceptors (Figure 6) leads to reduced binding affinity. Despite this hydrogen bond loss, compound **11** has a $K_i$ of 11 nM, suggesting that a more hydrophobic isostere would be a very potent inhibitor. Analysis of the interactions of compound **12** with thrombin illustrates the complex interplay among the different forces affecting ligand binding. This inhibitor has the most favorable $\Delta E_{LJ}$ of all the compounds considered in this study, yet it also has one of lowest activities, with a $K_i$ of 3.3 $\mu$M. The reason for this is that compound **12** also has the least favorable contribution from $\Delta E_{int}$ of any inhibitor considered here. The gas-phase minimum energy conformation of this molecule, and the conformation adopted in solution, has the carbamate N−C(=O)−O−C torsion at or near 180°, as in a Z-ester. In binding to thrombin this torsion must take a value of approximately 60° (see Figure 6). The gas-phase molecular mechanics energy of this conformation is 10.5 kcal/mol higher than the energy of the global minimum conformation. The torsional strain in the bound conformation leads to the poor activity of this inhibitor. In this case, substituting an ethylamide for the carbamate could significantly improve inhibitor activity due to the relatively small (0.5 kcal/mol) energy difference between the 60° and 180° conformations of the amide N−C(=O)−C−C torsion.

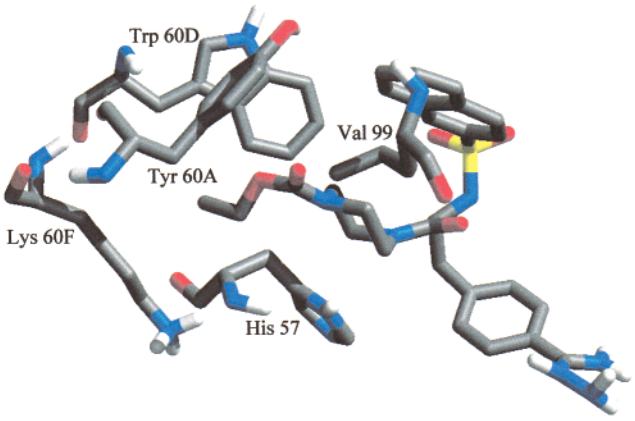


**Figure 6.** Binding site orientation of inhibitor **12** in the P-pocket of thrombin. The P-pocket sulfonamide S of inhibitor **11** occupies approximately the same position as the carbamate carbon depicted here. Neither of these groups is accessible to hydrogen bond donors.

Compounds **8** and **13**−**19** vary in the region of the inhibitor that binds in the D-pocket, a hydrophobic pocket formed by residues Tyr60A, Leu99, Ile174, and Trp215. The relative binding affinities among these inhibitors are determined primarily by the quality of the steric fit of their D-pocket moieties. The activities of inhibitors **13**−**19** were determined against human thrombin, so for comparison with these compounds the activity of inhibitor **8** versus human thrombin ($K_i$ = 0.38 nM)[24] should be used. It is clear from the relative binding affinities of compounds **13**, **14**, and **16** that the cycloalkyl groups of these inhibitors consistently form more favorable interactions in the D-pocket than the corresponding naphthyl substituent of inhibitors **8** and **15**. It can also be seen in Figure 7 that there is a recess behind the dimethylamino group of inhibitor **19**. The D-pocket moieties of compounds **14** and **19** are able to bind in this niche, enhancing the van der Waals interactions and potency of these inhibitors relative to inhibitors **8** and **15**. This is a fairly subtle effect that is nevertheless revealed in the calculated $\Delta E_{LJ}$ terms. The explanation for the relatively poor activity of compound **17** ($K_i$ = 18.4) compared to inhibitor **8** can be deduced from its binding mode, depicted in Figure 8, and its $\Delta E_{LJ}$ and $\Delta E_{int}$ terms, reported in Table 1. The reduced affinity seems to be more a result of the methylene linker introduced in this inhibitor rather than its smaller aryl group. This additional $sp^3$-hybridized carbon does not allow the aryl group to lay flat in the binding pocket, so the only D-pocket interactions possible are at the sides of this binding pocket and even these contacts require some degree of intramolecular strain to achieve.

## Conclusions

Regression equations have been developed for the estimation of binding affinities of active-site-directed,

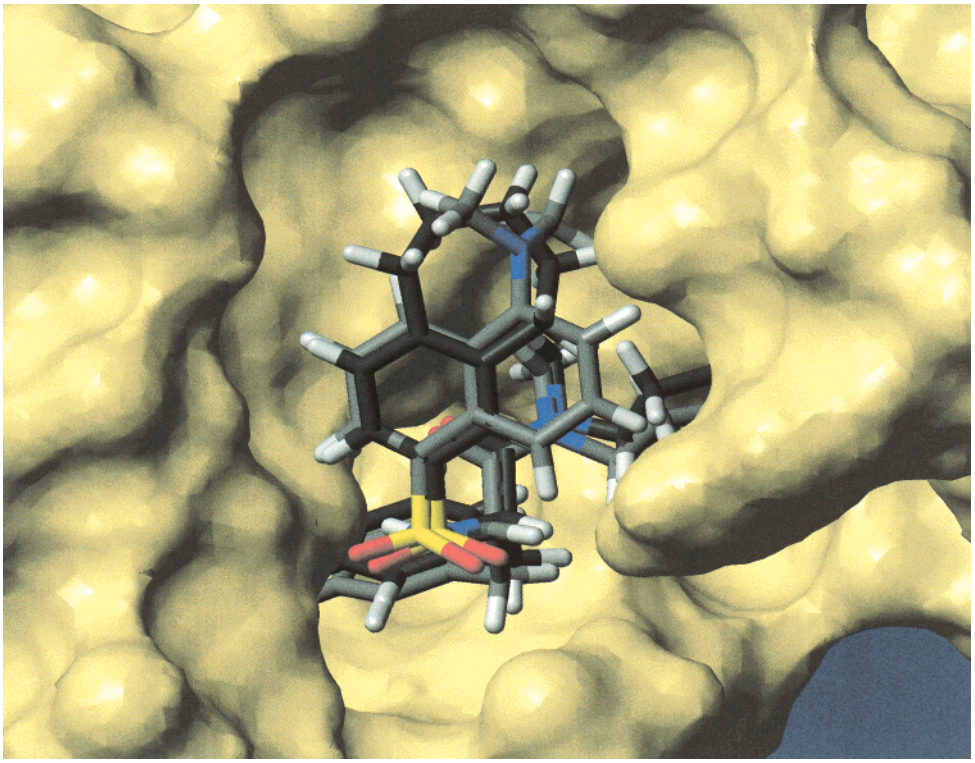

**Figure 7.** Inhibitors **14** (dark) and **19** (light) bound to thrombin. Both of these inhibitors take advantage of a niche in the upper right of the hydrophobic D-pocket.

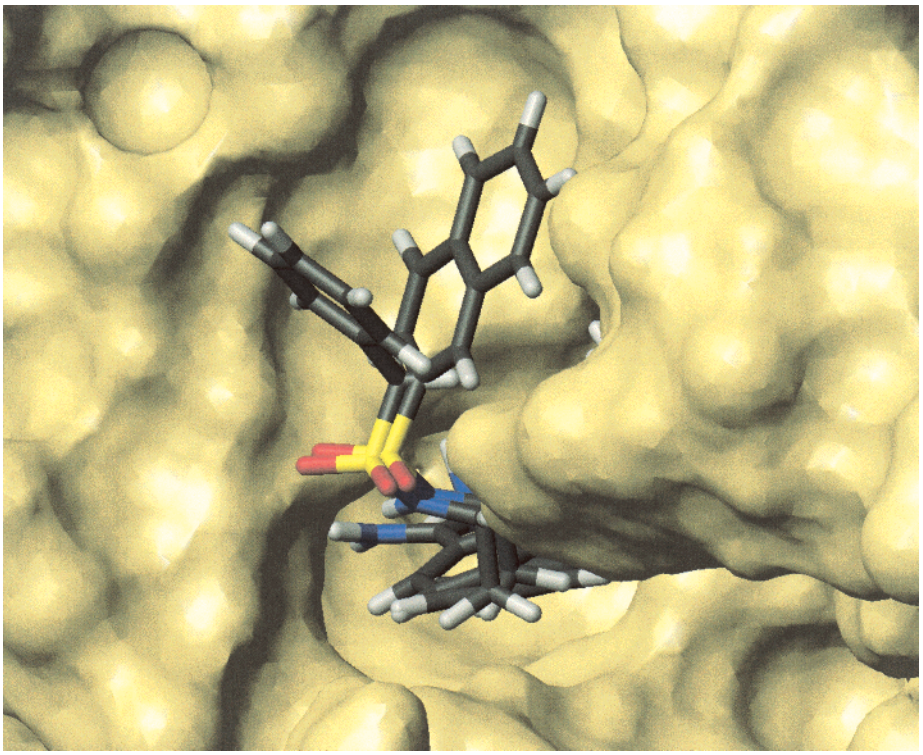

**Figure 8.** Naphthyl moiety of inhibitor **8** lays flat in the D-pocket relative to the more solvent-exposed benzyl group of inhibitor **17**.

small-molecule thrombin inhibitors. As in the original LR method, this technique calculates the binding energy as a function of the interactions of the inhibitor with its environment in the bound and unbound states. On the basis of eq 4, the four quantities found to be important for the prediction of binding affinities are: (1) the enhancement of van der Waals interactions upon binding, (2) the number of hydrogen bonds lost in going from solution to the bound state, (3) the intramolecular strain induced in the inhibitor upon binding, and (4) the number of rotatable bonds in the inhibitor. These quantities are all physically reasonable as descriptors for the calculation of $\Delta G_b$, and this model for activity prediction, along with the structural insights gained in its generation, could aid in the development of novel thrombin inhibitors. Expansion of the data set is also

desirable to obtain greater statistical confidence in the choice of optimal descriptors and to better gauge the predictive abilities. The seemingly better results with eq 5 indicate that caution is warranted in such studies to avoid potential overfitting of the data that includes significantly cross-correlated descriptors and coefficients for descriptors that seem to have the wrong sign on a physical basis.

While the calculations performed here are virtually identical to those used in the standard LR method, the flexibility of eq 3 allows for consideration of factors other than the van der Waals and electrostatic interactions in the estimation of binding affinities. In particular, inhibitor strain and entropy loss upon binding, two factors that cannot be accounted for with the standard LR equation, are found to be important for the reproduction of experimental activities. Another advantage of the terms in eq 4 is that since none of the descriptors involve long-range interactions or slowly converging quantities, significantly diminished computational effort may be possible in future projects. By considering a wider variety of contributions to protein−ligand binding in calculations that can be executed in a shorter period of time, the suggested approach represents a significant improvement over the standard LR method.

**Acknowledgment.** Gratitude is expressed to Dr. Julian Tirado-Rives and Robert Rizzo for helpful discussions throughout the course of this work and to the National Institutes of Health (Grant GM32136) for support.

## References

(1) Davie, E. W.; Fujikawa, K.; Kisiel, W. The Coagulation Cascade: Initiation, Maintenance and Regulation. *Biochemistry* **1991**, *30*, 10363−10370.
(2) Vu, T. K.; Hung, D. T.; Wheaton, V. I.; Coughlin, S. R. Molecular Cloning of a Functional Thrombin Receptor Reveals a Novel Proteolytic Mechanism of Receptor Activation. *Cell* **1991**, *64*, 1057−1068.
(3) Fenton, J. W.; Ofosu, F. A.; Moon, D. G.; Maraganore, J. M. Thrombin Structure and Function: Why Thrombin is the Primary Target for Antithrombotics. *Blood Coagulation Fibrinolysis* **1991**, *2*, 69−75.
(4) Raskob, G. E.; George, J. N. Thrombotic Complications of Antithrombotic Therapy: A Paradox with Implications for Clinical Practice. *Ann. Int. Med.* **1997**, *127*, 839−841.
(5) Thomas, D. J. ACOVA approval letter: http://www.fda.gov/cder/foi/appletter/2000/20883ltr.pdf; accessed 7/1/2000.
(6) Banner, D.; Ackermann, J.; Gast, A.; Gubernator, K.; Hadvary, P.; Hilpert, K.; Labler, L.; Muller, K.; Schmid, G.; Tschopp, T.; Waterbeemd, H. v. d.; Wirz, B. In *Serine Proteases: 3D Structures, Mechanisms of Action and Inhibitors*; Testa, B., Kyburz, E., Fuhrer, W., Giger, R., Eds.; John Wiley & Sons: Basel, 1993.
(7) Grootenhuis, P. D. J.; Galen, P. J. M. Correlation of Binding Affinities with Nonbonded Interaction Energies of Thrombin-Inhibitor Complexes. *Acta Crystallogr. D* **1995**, *51*, 560−566.
(8) Grootenhuis, P. D. J.; Karplus, M. Functionality Map Analysis of the Active Site Cleft of Human Thrombin. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 1−10.
(9) Böhm, M.; Sturzebecher, J.; Klebe, G. Three-Dimensional Quantitative Structure−Activity Relationship Analyses Using Comparative Molecular Field Analysis and Comparative Molecular Similarity Indices Analysis To Elucidate Selectivity Differences of Inhibitors Binding to Trypsin, Thrombin, and Factor Xa. *J. Med. Chem.* **1998**, *42*, 458−477.
(10) Jones-Hertzog, D. K.; Jorgensen, W. L. Binding Affinities for Sulfonamide Inhibitors with Human Thrombin Using Monte Carlo Simulations with a Linear Response Methodology. *J. Med. Chem.* **1997**, *40*, 1539−1549.
(11) Pierce, A. C.; Jorgensen, W. L. Computational Binding Studies of Orthogonal Cyclosporin- Cyclophilin Pairs. *Angew. Chem., Int. Ed. Engl.* **1997**, *36*, 1466−1469.
(12) Rao, B. G.; Kim, E. E.; Murcko, M. A. Calculation of solvation and binding free energy differences between VX-478 and its analogues by free energy perturbation and AMSOL methods. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 23−30.
(13) Åqvist, J.; Medina, C.; Samuelsson, J.-E. A New Method For Predicting Binding Affinity in Computer-Aided Drug Design. *Protein Eng.* **1994**, *7*, 385−391.
(14) Hansson, T.; Marelius, J.; Åqvist, J. Ligand Binding Affinity Prediction by Linear Interaction Energy Methods. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 27−35.
(15) Wang, J.; Dixon, R.; Kollman, P. A. Ranking ligand binding affinities with avidin: A molecular dynamics-based interaction energy study. *Proteins* **1999**, *34*, 69−81.
(16) Wall, I. D.; Leach, A. R.; Salt, D. W.; Ford, M. G.; Essex, J. W. Binding Constants of Neuraminidase Inhibitors: An Investigation of the Linear Interaction Energy Methodology. *J. Med. Chem.* **1999**, *42*, 5142−5152.
(17) Carlson, H. A.; Jorgensen, W. L. An Extended Linear Response Method For Determining Free Energies of Hydration. *J. Phys. Chem.* **1995**, *99*, 10667−10673.
(18) Duffy, E. M.; Jorgensen, W. L. Prediction of Properties from Simulations: Free Energies of Solvation in Hexadecane, Octanol, and Water. *J. Am. Chem. Soc.* **2000**, *122*, 2878−2888.
(19) Rizzo, R. C.; Tirado-Rives, J.; Jorgensen, W. L. A Computational Monte Carlo/QSAR Study of HIV-1 Reverse Transcriptase Non-Nucleoside Inhibitors: Estimation of Binding Affinities for HEPT and Nevirapine Analogues. *J. Med. Chem.* **2001**, *44*, 145−154.
(20) Ajay; Murcko, M. A. Computational Methods to Predict Binding Free Energy in Ligand−Receptor Complexes. *J. Med. Chem.* **1995**, *38*, 4953−4967.
(21) Kim, S.; Hwang Sang, Y.; Kim Young, K.; Yun, M.; Oh Yeong, S. Rational Design of Selective Thrombin Inhibitors. *Bioorg. Med. Chem. Lett.* **1997**, *7*, 769−774.
(22) Oh, Y. S.; Yun, M.; Hwang, S. Y.; Hong, S.; Shin, Y.; Lee, K.; Yoon, K. H.; Yoo, Y. J.; Kim, D. S.; Lee, S. H.; Lee, Y. H.; Park, H. D.; Lee, C. H.; Lee, S. K.; Kim, S. Discovery of LB30057, a Benzamidrazone-Based Selective Oral Thrombin Inhibitor. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 631−634.
(23) Oh, Y. S.; Kim, S. S.; Hwang, S. Y.; Yun, M. K.; Hwang, S. R.; Hong, S. W.; Lee, Y. H.; Jeong, Y. N.; Lee, K.; Shin, Y. S. European Patent Application EP 0739886, 1996.
(24) Lee, K.; Hwang, S. Y.; Hong, S.; Hong, C. Y.; Lee, C.-S.; Shin, Y.; Kim, S.; Yun, M.; Yoo, Y. J.; Kang, M.; Oh, Y. S. Structural Modification of an Orally Active Thrombin Inhibitor, LB30057: Replacement of the D-Pocket-Binding Naphthyl Moiety. *Bioorg. Med. Chem.* **1998**, *6*, 869−876.
(25) Lee, K.; Jung, W.-H.; Park Cheol, W.; Hong Chang, Y.; Kim In, C.; Kim, S.; Oh Yeong, S.; Kwon, O. H.; Lee, S.-H.; Park Hee, D.; Kim Sang, W.; Lee Yong, H.; Yoo Yung, J. Benzylamine-Based Selective and Orally Bioavailable Inhibitors of Thrombin. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 2563−2568.
(26) Brandstetter, H.; Turk, D.; Hoeffken, H. W.; Grosse, D.; Stürzebecher, J.; Martin, P. D.; Edwards, B. F. P.; Bode, W. Refined 2.3 Å Crystal Structure of Bovine Thrombin Complexes Formed with the Benzamide and Arginine-based Thrombin Inhibitors NAPAP, 4-TAPAP and MQPA − A Starting Point for Improving Antithrombotics. *J. Mol. Biol.* **1992**, *226*, 1085−1099.
(27) Jorgensen, W. L. *BOSS*, version 4.2; Yale University: New Haven, CT, 2000.
(28) Storer, J. W.; Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. Class-IV charge models − a new semiempirical approach in quantum chemistry. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 87−110.
(29) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.
(30) 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.
(31) Jorgensen, W. L. *MCPRO*, version 1.65; Yale University: New Haven, CT, 2000.
(32) *JMP*, version 3; SAS Institute, Inc.: Cary, NC, 1995.
(33) Vajda, S.; Weng, Z.; Rosenfeld, R.; DeLisi, C. Effect of Conformational Flexibility and Solvation on Receptor−Ligand Binding Free Energies. *Biochemistry* **1994**, *33*, 13977−13988.
(34) Dominy, B. N.; Brooks, C. L. Methodology for Protein−Ligand Binding Studies: Application to a Model for Drug Resistance, the HIV/FIV Protease System. *Proteins* **1999**, *36*, 318−331.
(35) Böhm, H.-J.; Klebe, G. What Can We Learn from Molecular Recognition in Protein−Ligand Complexes for the Design of New Drugs? *Angew. Chem., Int. Ed. Engl.* **1996**, *35*, 2588−2614.
(36) Lamb, M. L.; Tirado-Rives, J.; Jorgensen, W. L. Estimation of the Binding Affinities of FKBP12 Inhibitors using a Linear Response Methodology. *Bioorg. Med. Chem.* **1999**, *7*, 851−860.
(37) Åqvist, J. Calculation of Absolute Binding Free Energies for Charged Ligands and Effects of Long-Range Electrostatic Interactions. *J. Comput. Chem.* **1996**, *17*, 1587−1597.
(38) Muegge, I.; Martin, Y. C. A General and Fast Scoring Function for Protein−Ligand Interactions: A Simplified Potential Approach. *J. Med. Chem.* **1999**, *42*, 791−804.

JM000405U