Analyses of Activity for Factor Xa Inhibitors Based on Monte Carlo Simulations

Dennis Ostrovsky, Marina Udier-Blagović, and William L. Jorgensen*

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

Received June 13, 2003

Monte Carlo/Extended Linear Response (MC/ELR) simulations have been conducted on 60 inhibitors of human factor Xa to determine the important interactions associated with their activity. A variety of physicochemical descriptors were configurationally averaged during the course of the simulations of each inhibitor bound to factor Xa and free in water. A regression equation was then derived; it reproduces the experimental inhibition data with a correlation coefficient, r^2, of 0.74, an rms error of 0.67 kcal/mol, and an average unsigned error of 0.60 kcal/mol using only two physically reasonable descriptors. The two factors that emerged as important in determining inhibitory potential are (1) favorable van der Waals interactions between protein and ligand and (2) direct hydrogen bonding between the inhibitor and protein. The conclusions were supported with structural analyses and results of MC/free energy perturbation (FEP) calculations.

Background

The blood coagulation cascade is a complex series of enzymatic reactions that generate fibrin, the fundamental building block of all blood clots.[1] Undesirable clot formation within blood vessels may trigger serious physiological reactions, such as myocardial infarction, pulmonary embolism, and stroke. These cardiovascular events are among the leading causes of mortality in the industrialized world.[2] Current antithrombotic treatments are plagued by a variety of undesirable properties, including pharmaceutical toxicity, drug−drug interactions, and requirement of careful patient monitoring. Research in this area has centered primarily on the inhibition of thrombin. However, recent efforts to identify superior anticoagulants have instead targeted the inhibition of the factor Xa enzyme.[3]

Factor Xa is a serine protease whose main function is to convert prothrombin into thrombin. Thrombin then proceeds to catalyze reactions that result in the formation of fibrin.[4] Since factor Xa links the intrinsic and extrinsic pathways of coagulation, its inhibition may prevent thrombin formation via either route. The series of reactions leading to thrombus formation involves signal amplification, where many molecules of thrombin can be generated via a single factor Xa molecule.[5] Therefore, inhibition of factor Xa should be more efficient than inhibition of thrombin.

The present study is a part of our continuing efforts to develop computational methods for the study of inhibitor binding to biologically and pharmaceutically relevant targets. We have successfully correlated computed and experimental activities for series of inhibitors bound to thrombin,[6] COX-2,[7] and HIV reverse transcriptase.[8,9] In the present work, Monte Carlo statistical mechanics simulations have been carried out for 60 factor Xa inhibitors based on two diaryloxypyridine cores (Scheme 1), for which experimental K_i data are available (Tables 1−3).[10,11] The aims of this investigation are (1) to develop a predictive computational model for factor Xa inhibition by correlating results of atomic-level simulations that include explicit solvent molecules with the experimental K_i data and (2) to better understand the model and the accompanying structural insights. This work is also part of a larger study to process several thousand protein−ligand complexes in search of general tools for the reliable prediction of protein−ligand binding affinities as the foundation for automated ligand design.

Theoretical Methods

Free energy perturbation (FEP) and thermodynamic integration (TI) are the most rigorous computational approaches currently used to calculate binding affinities (ΔG_b).[12−15] These methods typically utilize molecular dynamics (MD) or MC simulations and have yielded impressive results for a variety of protein−ligand systems.[12−15] The estimation of ΔG_b may be accomplished with a more approximate method based on linear response (LR) theory, originally introduced by Åqvist and co-workers (eq 1).[16] The LR approach is significantly less demanding computationally than the FEP and TI alternatives, because no intermediate transformation process is required to compute binding affinities.[16]
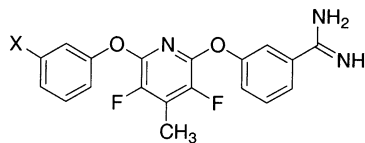
$$\Delta G_b = \alpha \langle \Delta E_{vdW} \rangle + \beta \langle \Delta E_{Coul} \rangle \tag{1}$$

In eq 1, $\langle ... \rangle$ signifies an ensemble average of the difference in interaction energies (ΔE) for the ligand in the solvated protein−ligand complex and for the unbound ligand in solution.[16] The energy terms represent differences between bound and free average van der Waals (Lennard-Jones) and electrostatic (Coulombic) contributions, respectively, which are typically calculated using a molecular mechanics force field and either MC or MD simulations. In the original report, the differences in Coulombic energy were scaled by $\beta = 0.50$, while α was determined empirically by fitting MD results to experimental binding data.[16]

The LR method was subsequently extended for calculations of free energies of solvation.[17] In this modified

* To whom correspondence should be addressed. Phone: 203 432-6278. Fax: 203 432-6299. E-mail: william.jorgensen@yale.edu.

Scheme 1

core 1   core 2
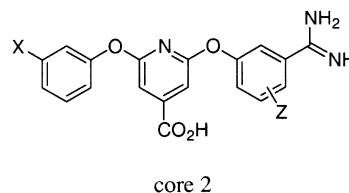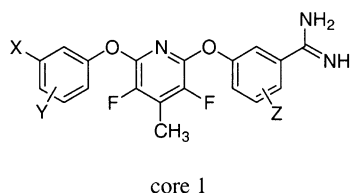
Table 1. Inhibition Data for Diaryloxypyridine Analogs with Monosubstitution on the Ring that Binds in the S4 Pocket


no.	X	K_i (nM)[a]	ΔG_{expt}[b]
1	CONH$_2$	280	−8.93
2	CO$_2$Et	5200	−7.20
3	CONHMe	1200	−8.07
4	CONMe$_2$	80	−9.67
5	CONEt$_2$	690	−8.40
6	CONMeBn	2400	−7.66
7	COPyrrolidine	220	−9.07
8	CO−morpholine	1400	−7.98
9	CO−piperidine	820	−8.29
10	CH$_2$CONMe$_2$	460	−8.64
11	(CH$_2$)$_2$CONM e$_2$	1300	−8.02
12	CH$_2$CONH$_2$	280	−8.93
13	CH$_2$COOEt	810	−8.30
14	COMe	1400	−7.98
15	COCHMe$_2$	2100	−7.71
16	NO$_2$	2500	−7.63
17	NH$_2$	3300	−7.47
18	NMe$_2$	160	−9.26
19	NHEt	530	−8.55
20	NEt$_2$	400	−8.72
21	NHSO$_2$Me	3804	−7.39
22	NHAc	5800	−7.14
23	morpholine	650	−8.43
24	OMe	1350	−8.00
25	OCF$_3$	1800	−7.83
26	OCHMe$_2$	620	−8.46
27	F	3200	−7.49
28	Cl	1700	−7.86
29	OH	5000	−7.22
30	CF$_3$	1600	−7.90
31	CH$_2$OH	1300	−8.02

[a] References 10, 11. [b] Experimental free energies from ΔG_{expt} = $RT \ln K_i$ in kcal/mol.

approach, both α and β coefficients are varied and a third term representing the solvent accessible surface area (SASA) of the solute is included and scaled by γ. The SASA term is included as a means of accounting for possible positive free energies of hydration caused by solute cavity formation in the solvent.[17] The corresponding LR equation for protein−ligand binding is then eq 2.

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

Given the success of previous MD/LR[16,18−22] and MC/LR[23−25] binding studies, we sought to treat significantly larger data sets to determine if correlations with experimental data are maintained upon increasing the ratio of data points to parameters. Simultaneously, results from aqueous MC simulations were correlated with solvation properties for over 200 diverse organic compounds.[26] In the latter study the list of descriptors

Table 2. Inhibition Data for Polysubstituted Diaryloxypyridine Analogs
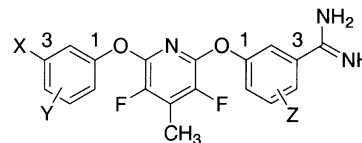
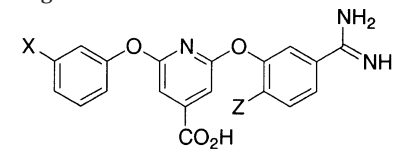| no.	X	Y	Z	K_i (nM)[a]	ΔG_{expt}[b]
32	CONMe$_2$	2-OH	H	4500	−7.29
33	CONMe$_2$	2-OMe	H	2400	−7.66
34	CONMe$_2$	4-OH	H	2000	−7.77
35	CONMe$_2$	4-OMe	H	1400	−7.98
36	CONMe$_2$	5-CONMe$_2$	H	160	−9.26
37	CONMe$_2$	5-OMe	H	140	−9.34
38	NMe$_2$	2-Me	H	320	−8.85
39	NMe$_2$	4-Cl	H	500	−8.59
40	CONMe$_2$	H	5,6-OH	23	−10.43
41	CONMe$_2$	H	5,6-OMe	180	−9.19
42	CONMe$_2$	H	5-OH, 6-OMe	160	−9.26
43	CONMe$_2$	H	5-OMe	2700	−7.59
44	CONMe$_2$	H	6-NH$_2$	14	−10.70
45	CONMe$_2$	H	6-OH	1.8	−11.92
46	CONMe$_2$	H	6-OMe	720	−8.37
47	NMe$_2$	H	6-Me	1200	−8.07
48	NMe$_2$	H	6-NH$_2$	64	−9.80
49	NMe$_2$	H	6-OH	3	−11.62
50	NMe$_2$	H	6-OMe	1400	−7.98

[a] References 10, 11. [b] Experimental free energies from ΔG_{expt} = $RT \ln K_i$ in kcal/mol.

was expanded from those in eq 2 to include, for example, counts of hydrogen bonds, and separate components of SASA, e.g., the hydrophobic, hydrophilic, and aromatic surface areas. A multivariate fitting approach was used, and eq 3 is the related general expression for estimating free energies of binding.

$$\Delta G_b = \sum_n c_n \xi_n + \text{constant} \quad (3)$$

In this equation c_n is an optimizable coefficient for its associated descriptor, ξ_n. In principle, any physically meaningful quantity could be used as a descriptor in this extended linear response (ELR) approach, although descriptors with understandable physical interpretations are likely to have greater predictive utility. For example, an ELR regression with a variety of descriptors was considered[8] and ultimately validated for a set of over 200 nonnucleoside HIV reverse transcriptase (RT) inhibitors.[9] ELR regressions have also been reported for 20 thrombin inhibitors[6] and 45 selective COX-2 inhibitors.[7] Consensus is emerging from these studies on the importance of a few key descriptors, especially, the protein−ligand van der Waals interaction and the change in number of hydrogen bonds for the ligand upon binding.

It should be noted that that LR and ELR studies to-date have all used experimental data in conjunction with descriptors obtained from computer simulations to

Table 3. Inhibition Data for Diaryloxypyridine Carboxylic Acid Analogs

no.	X	Z	K_i (nM)[a]	ΔG_{expt} [b]
51	Amidine	H	59	−9.80
52	Amidine	OH	6	−11.21
53	(imidazol-1-yl)	OH	26	−10.34
54	(2-methylimidazol-1-yl)	OH	8	−11.03
55	(4-methylimidazol-1-yl)	OH	13	−10.75
56	(1-methylimidazol-2-yl)	OH	9	−10.97
57	(1,4-dimethylimidazol-2-yl)	OH	8.4	−11.01
58	(4,5-dihydro-2-methylimidazol-2-yl)	OH	85	−9.64
59	(1,4,5-... 2-methylimidazolinyl)	OH	2.4	−11.75
60	(4,4-dimethyl-2-methylimidazolinyl)	OH	810	−8.30

[a] References 10, 11. [b] Experimental free energies from $\Delta G_{expt} = RT \ln K_i$ in kcal/mol.

derive a regression expression. Once a reasonable, cross-validated equation is derived, no additional experimental data are necessary for activity predictions of novel compounds; only simulations of bound and unbound states are required to make predictions for new compounds. Successful predictions of binding affinities for 27 nonnucleoside inhibitors of HIV-1 reverse transcriptase not included in the development of the scoring function provided strong support for the potential utility of the MC/ELR method as a useful predictive tool in drug design.[9] The same regression equation was also successful ($q^2 = 0.55$) for subsequent predictions of the activities of 47 inhibitors with the K103N mutant of HIV RT.

Computational Details

System Setup. Consistent with previous ELR protocols, a binding site model was constructed using the residues closest to the active site.[8] The initial coordinates were taken from a structure containing diphenoxypyridine inhibitor ZK-807834 bound to factor Xa (pdb entry 1fjs).[28] A representative model was built using CHOP[29] and included only residues with any atom within ca. 16.5 Å of the geometric center of ZK-807834 (Figure 1). During the pruning of the protein, residues were added or removed as necessary to minimize the number of
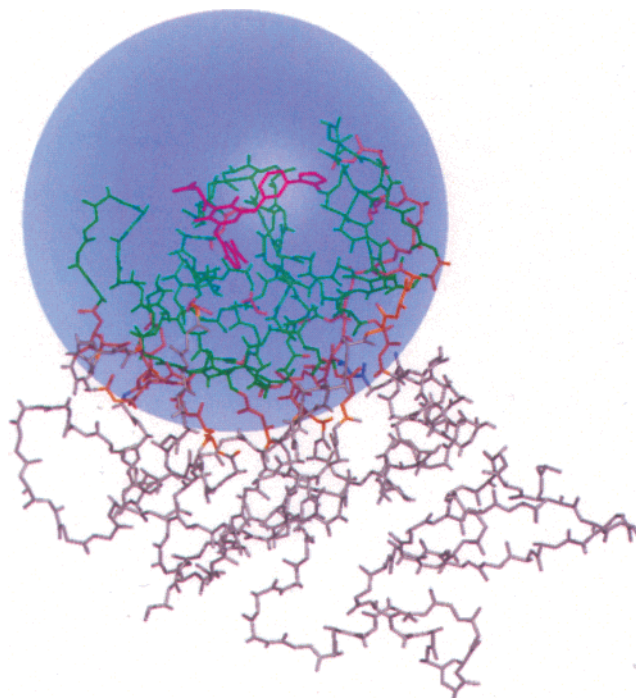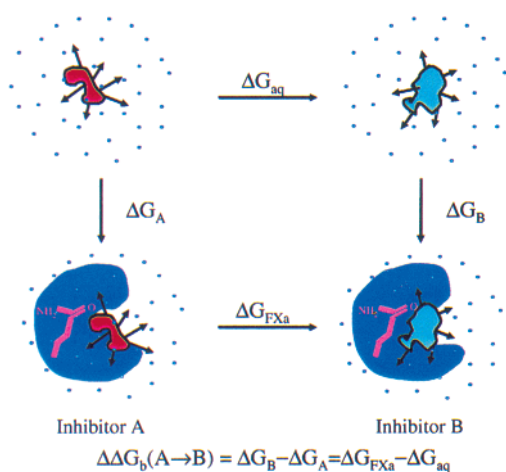

Figure 1. Factor Xa binding site model solvated by a 22-Å cap of water. Green residues sampled in the MC simulations, orange residues rigid, gray residues not used. Compound ZK-807834 in magenta shows the location of the binding site. Crystal structure coordinates (pdb entry 1fjs) from ref 28.

protein chains and to maximize the individual chain lengths. Clipped residues were capped by acetyl or *N*-methylamine groups, and hydrogen atoms were added using the PEPZ program.[30] Residues with all atoms outside a 12 Å sphere from the center of ZK-807834 were kept rigid during MC simulations. The final system contained 143 protein residues plus the inhibitor. Flexible residues consisted of 17−19A, 30A, 32−33A, 40−44A, 54−61A, 94−103A, 136−151A, 156−163A, 168A, 171−177A, 180−199A, and 211−229A. Rigid residues were 20−21A, 28−29A, 31A, 34−39A, 53A, 62−63A, 91−93A, 104−105A, 132−135A, 152−155A, 164−167A, 169−170A, 178−179A, 200A, 210A, and 230−233A. The overall charge on the system was neutralized as previously described.[8] Each inhibitor has a charge of +1; for compounds **1−50**, the amidine group is protonated, and for compounds **51−60**, a combination of two positively charged groups (amidine and imidazolyl) and a negatively charged carboxylate also resulted in a net +1 charge. Initial structures were generated using the GenMol program.[31] GenMol performs a conformational search and optimization for a ligand in a binding site, after which the lowest energy complex is saved for subsequent MC calculations. Partial atomic charges for the ligands were obtained using the CM1A procedure.[32] All energy evaluations used the OPLS-AA force field[33] with the exception of the inclusion of the CM1A charges for the ligand. Experimental K_i values were converted to a free-energy scale via $\Delta G_{expt} = RT \ln K_i$.

MC Simulations. To relax the crystal structure prior to the MC simulations, 50 steps of conjugate gradient energy minimization were performed on each protein−ligand complex using a distance-dependent dielectric constant of 4 ($\epsilon = 4r$). The MC simulations used a 22 Å TIP4P[34] water cap centered on the ligand containing 1470 and 853 water molecules for the bound and unbound simulations, respectively. The side chains of the protein residues with any atom within 12 Å of the ligand were sampled, the protein backbone was fixed after the conjugate gradient minimizations, and the ligands were fully flexible throughout the simulations. All bond lengths for the protein remained fixed after the initial energy minimizations. A protein residue−ligand list was determined for each complex during the initial solvent equilibration stage of the simulation and was subsequently kept constant. A random MC

Scheme 2

Figure 2. Computed activities (ΔG_{calc}) using eq 4 vs experimental activities (ΔG_{expt}) for the 60 diaryloxypyridine inhibitors with factor Xa.

move for a side chain was attempted every 10 configurations, while a ligand move was attempted every 56 configurations; the remaining moves were for solvent molecules. Solvent–solvent neighbor lists were used, and the maximum number of internal coordinates varied in an attempted move was limited to 30. All MC simulations and energy calculations were performed using the MCPRO program.[35]

Other standard procedures were followed.[8,9] Residue-based cutoffs were set to 9 Å for solvent–solvent, solute–solute, and intersolute nonbonded interactions. Protein–inhibitor simulations in the bound state consisted of 10 million configurations of solvent-only equilibration, 10 million configurations of full equilibration, and 10 million configurations of averaging. Unbound simulations utilized an annealing protocol in order to lessen the impact of starting conditions on the simulation results.[8] This protocol entails 10 million configurations of solvent-only equilibration and a further 5 million configurations during which only ligand translation, rotation, and dihedral angles are sampled in addition to the solvent molecules. To increase the acceptance rates for this latter stage, the system temperature is raised to 727 °C (1000 K) for each attempted ligand move. After this high-temperature phase, the system is equilibrated for 5 million configurations at the normal simulation temperature (25 °C), followed by 10 million configurations of averaging. The latter three stages are repeated for a total of five cycles.[8]

FEP. To obtain additional understanding of the large difference in activity for compounds **49** and **50** (Table 2) and compounds **58** and **59** (Table 3), rigorous free energy perturbation (FEP) calculations were performed for these pairs using well-established methods.[12] Following protocols described in more detail previously,[36,37] one inhibitor was converted to the other unbound in water and bound to the protein, and the relative free energy of binding was obtained using the thermodynamic cycle shown in Scheme 2.

FEP calculations were performed by converting inhibitor A to inhibitor B in 20 windows, using the 22 Å TIP4P water cap with the 1470 and 853 water molecules for the unbound and bound simulations, respectively. Each window consisted of 10 million configurations of solvent only equilibration, 10 million configurations of full equilibration, and 10 million and 20 million configurations of averaging for bound and unbound states, respectively.

Results and Discussion

Regression Equation. All statistical analyses were performed using the JMP program.[38] Roughly 40 descriptors were evaluated during the simulations and included protein–ligand and ligand–solvent interaction energies, counts of hydrogen bonds, surface area changes, and predicted properties for the ligand such as the octanol/water log P and polarizability. The resultant
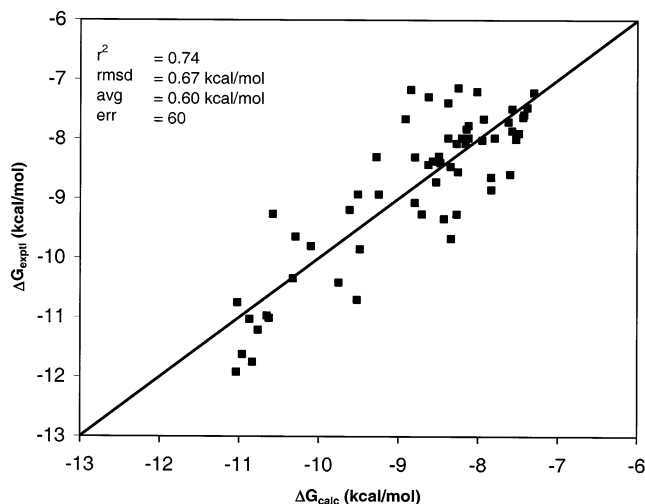
simplest regression equation containing just two physically reasonable descriptors correlated the experimental activities with the computed properties according to eq 4.

$$\Delta G_{\text{calc}} = 0.116\langle \text{EXX}_{\text{LJ}}\rangle - 1.12\langle \text{DtoPr}\rangle + 0.33 \quad (4)$$

EXX_{LJ} is the average protein–ligand Lennard-Jones interaction energy, while DtoPr represents the average number of hydrogen bonds donated to the protein by the ligand. A hydrogen bond was defined by an interatomic distance of less than 2.5 Å between a hydrogen attached to a heteroatom and any N, O, or S atom. The correlation coefficient (r^2) for the 60 compounds is 0.74, reflecting good agreement between theory and experiment (Figure 2). The robustness of the regression equation was probed by computing the cross-validated q^2 using the leave-one-out procedure.[6,9] In the present case, the computed q^2 of 0.63 indicates that eq 4 also has good predictive value. The overall rms error for the computed free energies is 0.67 kcal/mol and the mean unsigned error is 0.61 kcal/mol. These errors are somewhat smaller than in previous ELR studies,[8,9] possibly due to higher quality for the experimental data (K_i vs IC_{50}/EC_{50}) stemming from the use of an isolated enzyme assay. The statistical significance of the two descriptors was evaluated by computing probability $> F$ ratios. For each of the descriptors, the probability $> F$ ratio was <0.0001, implying that the likelihood of a randomly occurring significant descriptor is low. The cross-correlation between the two descriptors was also small, $r = 0.12$. Equation 4 emerges readily using the stepwise regression option with the *JMP* program; it is also obtained as optimal for two descriptors using an in-house program, xlfit, that performs exhaustive linear regression analyses for all combinations of a specified number of descriptors. Addition of more descriptors does not improve the fit.

Both descriptors in eq 4 make intuitive physical sense. EXX_{LJ} is a measure of the goodness of the steric fit between the ligand and protein, taking into account both favorable and unfavorable interactions. This term has consistently emerged as significant in previous ELR studies and its coefficient is normally near 0.1.[7–9] The
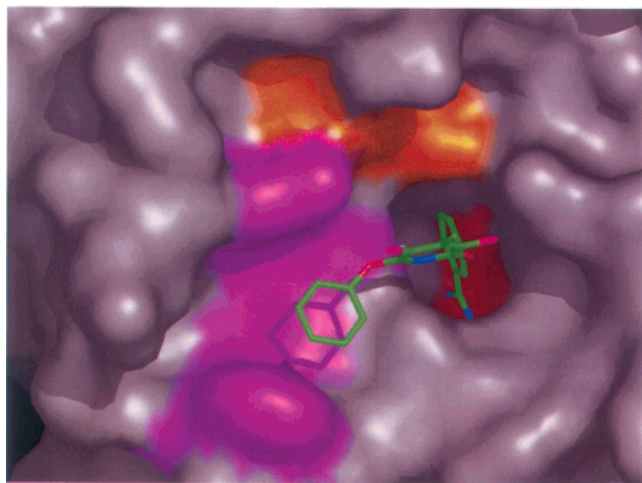
Figure 3. Factor Xa binding site with core 1 bound. The S4 binding pocket featuring residues Tyr99A, Phe174A, and Trp215A is shown in purple. Residues Asp189A and Ala190A are colored red in the S1 binding pocket. The catalytic triad (Asp102A, His57A, Ser195A) is colored orange.

DtoPr term implies that an increase in the number of protein−ligand hydrogen bonds favors binding. Although this exact term has not appeared in previous work, other hydrogen-bonding descriptors have been found to be significant.[8,9] The magnitude of the coefficient in front of hydrogen-bonding terms has varied from ca. 0.7 to 1.3 kcal/mol, thus the present value of 1.12 kcal/mol is normal.[9] The number of DtoPr hydrogen bonds varies from three to five for the present inhibitors and includes two hydrogen bonds for the protonated amidine with the Asp189A carboxylate group.

In other ELR studies,[6−9,27] a consensus is emerging for the importance of the total protein−ligand interaction energy ($EXX = EXX_{LJ} + EXX_{Coulomb}$), the change in number of hydrogen bonds for the ligand upon transfer from water to the protein binding site, ΔHB, and the number of rotatable bonds for the ligand. The lack of these terms in eq 4 likely stems from the general structural similarity of the ligands, resultant small ranges in the descriptors, and the small range in observed inhibitory potential for compounds **1**−**39** (2.5 kcal/mol). The result is that only two differentiating elements were revealed. Expansion of the database to include many more compounds, more chemotypes, and many proteins is needed to develop general models for protein−ligand binding.

Analysis of Binding Trends. The two descriptor in eq 4 contribute an approximately equal amount to the computed activities, and each descriptor largely reflects interactions in a binding subsite. The compounds in this data set have two primary sites of interaction with factor Xa,[39] and these will be discussed individually (Figure 3). One of the aryl rings is positioned in the S4 binding pocket (Figure 3, in purple), structurally defined by aromatic residues Tyr99A, Phe174A, and Trp215A. The interactions in this region are nonbonding in nature and consist primarily of van der Waals contacts with the π-systems of these aromatic residues. Favorable and unfavorable interactions are reflected in EXX_{LJ}. Despite the variety of substituents, the range of activities for compounds **1**−**39** (2.5 kcal/mol) is similar to the range in the EXX_{LJ} component of eq 4 (1.6 kcal/mol) (Table

4). Small substituents are disfavored in this pocket, e.g., compounds **27** (X = F) and **29** (X = OH), which both have EXX_{LJ} contributions of −4.2 kcal/mol (Table 4). These small groups have poor contacts with the aromatic rings in the S4 binding pocket. Compound **18** (X = NMe₂) is predicted to be a better inhibitor than the unmethylated analogue **17** (X = NH₂), in agreement with the experimental K_i trend. The less polar NMe₂ group is better matched with the aromatic pocket and suffers less of a desolvation penalty. It was also speculated that the observed difference was associated with the bulkier NMe₂ group adopting a nonplanar conformation relative to the phenyl ring in comparison to NH₂.[11] The average computed H−N−C−C and C−N−C−C dihedral angles were 22° and 85° for bound **17** and **18**, respectively, supporting the latter structural hypothesis. A similar increase in activity was observed upon NMe₂ substitution in a structurally analogous set of factor Xa inhibitors.[40] However, disubstitution on this ring generally leads to significant loss of activity (Table 2).[11] With two positions substituted, it is difficult to avoid some steric clashes with protein side chains in the S4 site. Interestingly, a smaller penalty is observed for compound **36**, which has CONMe₂ substituents in both the 3- and 5-positions. The difference in activities ($\Delta\Delta G_{expt}$) for compound **36** and its monosubstituted counterpart **4** is only 0.4 kcal/mol. For **4**, we find that the CONMe₂ group is orthogonal to the phenyl ring and is stacked between the aryl side chains of Phe174A and Tyr99A. For **36** with the C_2 symmetric nature of the 3,5-di-CONMe₂ substitution, either amide group can occupy the sandwiched position, while the other is mostly solvent exposed. Compound **37** also likely benefits in the same manner with population of both arrangements of the substituents.

Substituents on the aryloxy ring in the S1 pocket interact with the protein primarily through hydrogen bonding. Specifically, the amidine forms hydrogen bonds with Asp189A and with the backbone carbonyl of Gly218A, as illustrated in Figure 4. These three protein−ligand hydrogen bonds and the hydrogen bond with the water molecule in Figure 4 are essentially invariant for the present ligand series. This homogeneity of interactions in the S1 pocket, coupled with the lack of significant differentiation between the strengths of van der Waals interactions in the S4 pocket, explain the fairly narrow range of activies for compounds **1**−**39**.

The key structural change in this series of compounds that dramatically increases their inhibition of factor Xa is the addition of a hydrogen-bond-donating group in the 6-position on the amidine-containing aryloxy ring. For example, compound **49** (Z = OH) is experimentally observed to be a better inhibitor than compound **50** (Z = OMe) by 3.6 kcal/mol (Table 2). Compound **50** lacks the hydroxyl group, which has been observed to form a hydrogen bond directly to O$^\gamma$ of Ser195A in the crystal structure for another inhibitor, ZK-807834, with factor Xa.[28] The predicted relative binding affinity of compounds **49** and **50** is 3.2 kcal/mol in Table 4. Indeed, the present MC simulations find that the 6-OH in **49** forms a hydrogen bond with Ser195A and it also participates in an additional water-bridged hydrogen bond to His57A (Figure 5). The latter water molecule corresponds to crystallographic water molecule 606 in

Table 4. Contributions of the Separate Terms in Eq 4 to the Computed Free Energies (ΔG_{calc})

	contribution from					contribution from			
no.	EXX_{LJ}	DtoPr	ΔG_{calc}	ΔG_{expt}[a]	no.	EXX_{LJ}	DtoPr	ΔG_{calc}	ΔG_{expt}[a]
1	−5.08	−4.51	−9.25	−8.93	31	−4.84	−3.45	−7.95	−8.02
2	−4.98	−3.38	−8.01	−7.20	32	−5.18	−3.79	−8.62	−7.29
3	−5.20	−3.41	−8.27	−8.07	33	−4.90	−3.38	−7.93	−7.66
4	−5.31	−3.38	−8.34	−9.67	34	−5.07	−3.40	−8.12	−7.77
5	−5.44	−3.38	−8.47	−8.40	35	−5.17	−3.38	−8.21	−7.98
6	−5.88	−3.38	−8.91	−7.66	36	−5.64	−3.41	−8.70	−9.26
7	−5.76	−3.38	−8.79	−9.07	37	−5.36	−3.41	−8.43	−9.34
8	−5.34	−3.38	−8.37	−7.98	38	−4.79	−3.39	−7.83	−8.85
9	−5.46	−3.38	−8.49	−8.29	39	−4.57	−3.38	−7.60	−8.59
10	−4.80	−3.38	−7.83	−8.64	40	−6.46	−3.63	−9.75	−10.41
11	−5.14	−3.38	−8.17	−8.02	41	−6.18	−3.77	−9.61	−9.19
12	−5.24	−4.61	−9.51	−8.93	42	−5.76	−5.16	−10.58	−9.26
13	−5.76	−3.38	−8.79	−8.30	43	−3.76	−4.01	−7.43	−7.59
14	−5.08	−3.39	−8.12	−7.98	44	−6.36	−3.50	−9.52	−10.70
15	−4.51	−3.45	−7.62	−7.71	45	−6.80	−4.58	−11.04	−11.92
16	−4.36	−3.42	−7.43	−7.63	46	−4.90	−4.02	−8.57	−8.37
17	−4.34	−3.39	−7.38	−7.47	47	−4.14	−4.37	−8.16	−8.07
18	−5.18	−3.43	−8.27	−9.26	48	−6.24	−4.20	−10.10	−9.80
19	−5.21	−3.39	−8.25	−8.55	49	−6.50	−4.80	−10.96	−11.62
20	−5.48	−3.39	−8.52	−8.72	50	−4.46	−3.68	−7.79	−7.98
21	−5.31	−3.41	−8.38	−7.39	51	−4.72	−5.11	−9.49	−9.85
22	−5.17	−3.42	−8.25	−7.14	52	−5.06	−6.05	−10.77	−11.21
23	−5.55	−3.42	−8.62	−8.43	53	−5.26	−5.41	−10.32	−10.34
24	−4.42	−3.44	−7.52	−8.00	54	−5.50	−5.72	−10.87	−11.03
25	−5.06	−3.43	−8.15	−7.83	55	−5.37	−6.00	−11.02	−10.75
26	−5.31	−3.38	−8.35	−8.46	56	−5.61	−5.39	−10.65	−10.97
27	−4.23	−3.68	−7.57	−7.49	57	−5.53	−5.43	−10.63	−11.01
28	−4.45	−3.47	−7.57	−7.86	58	−5.45	−5.19	−10.29	−9.64
29	−4.24	−3.40	−7.30	−7.22	59	−5.60	−5.58	−10.83	−11.75
30	−4.42	−3.41	−7.49	−7.90	60	−4.11	−5.51	−9.28	−8.30

[a] Experimental free energies from $\Delta G_{expt} = RT \ln K_i$ in kcal/mol.
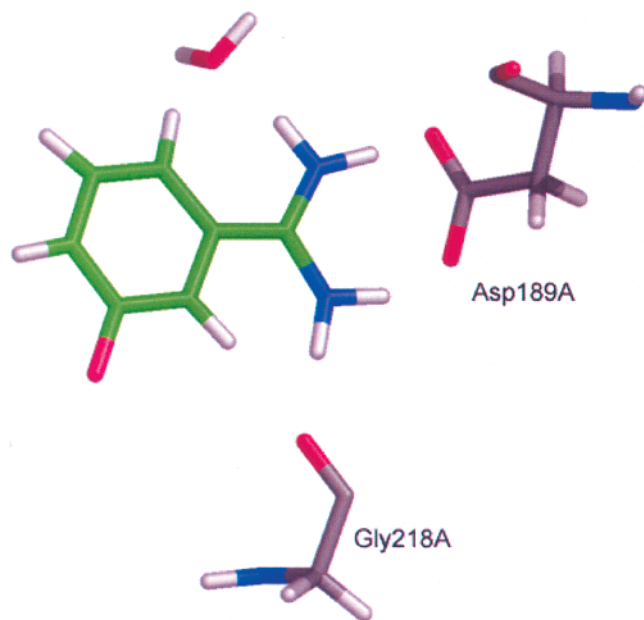

Figure 4. Key interactions in the S1 binding pocket between factor Xa and the benzamidine fragment of the diaryloxypyridine inhibitors. The illustrated water molecule, which is hydrogen bonded to the amidine group, is present throughout all the MC simulations.

pdb entry 1fjs.[28] Water-mediated hydrogen-bond interactions of this type have also been reported for inhibitors of related serine proteases.[41] A side-by-side comparison of the simulated structure of **49** with the crystal structure for ZK-807834 shows good agreement between the positions of inhibitor atoms involved in hydrogen bonding and the His57A and Ser195A residues of the catalytic triad (Figure 6). For compound **50**, the methoxy

group displaces the bridging water molecule and no hydrogen-bonding network is formed (Figure 5, right). **45** and **46** are the tertiary amide analogues and also show experimentally a 3.6 kcal/mol reduction in inhibitory ability upon replacement of the 6-hydroxy group with methoxy. The computed difference in Table 4 of 2.5 kcal/mol is still reasonable, and it is smaller than that for the **49**, **50** pair; it reflects a diminished loss for the hydrogen-bonding term, 0.56 vs 1.12 kcal/mol. The approximately half a hydrogen bond difference between the counts from two MC simulations does not appear to be statistically significant based on the variations that are obtained from multiple runs for one complex.

FEP calculations were also performed for the **49** → **50** conversion in water and in the solvated complex in order to compute an accurate relative free energy of binding for this pair. The computed $\Delta\Delta G_b$ of 2.24 ± 0.33 kcal/mol is about 1 kcal/mol less than the difference from the K_i data. As with the MC/ELR calculations, a direct hydrogen bond with the side chain of Ser195A and an additional water-mediated interaction with His57A are present in the FEP simulations (Figure 5). The preservation of these interactions lends further support to the importance of the DtoPr descriptor in the regression model (eq 4).

Furthermore, FEP calculations were performed for the conversion **58** → **59**, where the substitution of a hydrogen by a methyl group in the S1 binding pocket leads to a 35-fold increase in potency. The amidine groups are protonated in all cases. The FEP result, $\Delta\Delta G_b = -1.96 \pm 0.25$ kcal/mol, is in close accord with the experimental inhibition difference, −2.11 kcal/mol (Table 3). Further analysis of energy components then led to a more detailed understanding of the origin of
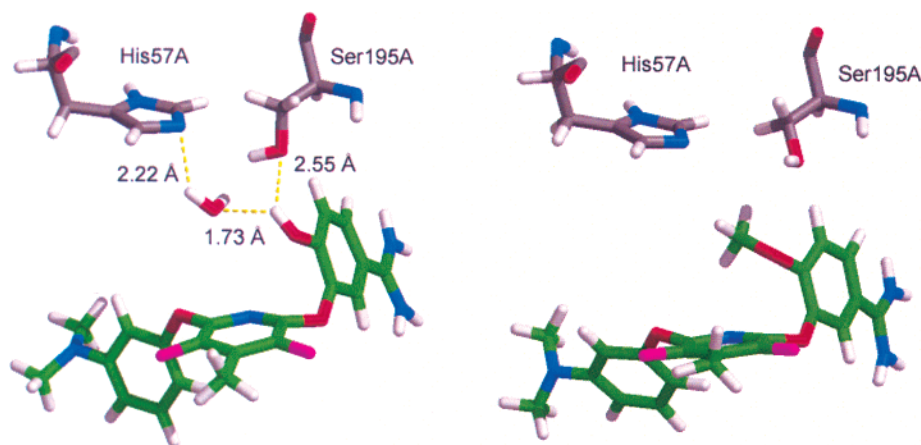
Figure 5. Comparison of computed structures for bound inhibitors **49** (left) and **50** (right). The methoxy group in **50** is incapable of hydrogen bonding to Ser195A, while the hydroxy group in **49** donates a hydrogen bond to the hydroxyl group of Ser195A. An additional water-bridged hydrogen bond to His57A is present in the complex with **49** but absent in **50** due to the steric bulk of the methoxy group.
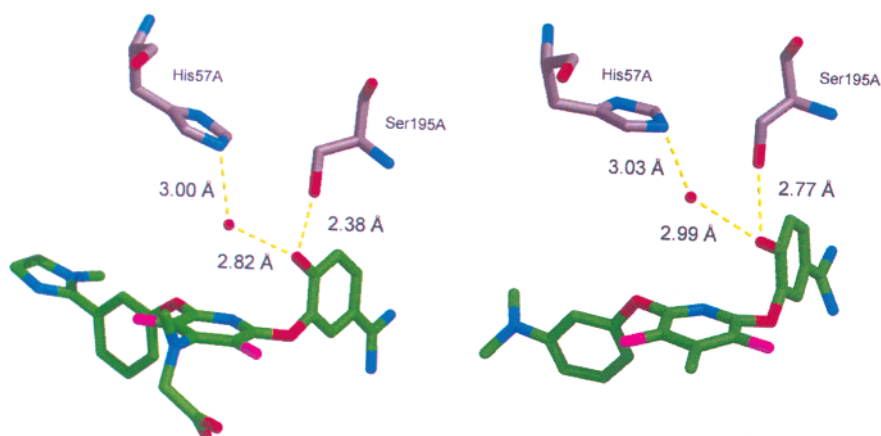

Figure 6. Interactions in the active site for inhibitors ZK-807834 (left, coordinates from ref 28) and **49** (right, after 20 million MC steps) with factor Xa. Hydrogens in the computed structure have been removed for clarity.
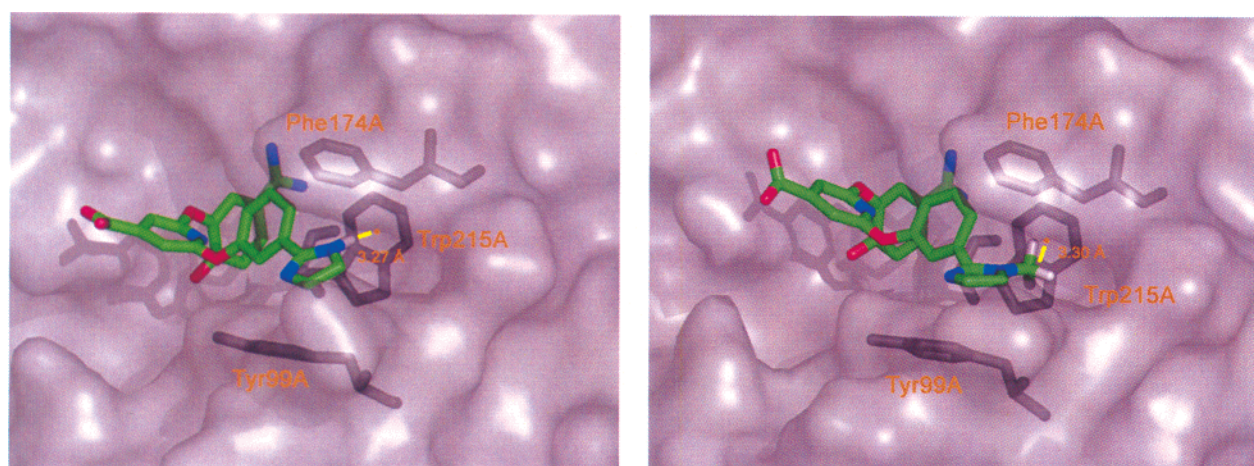

Figure 7. Interactions in the active site for inhibitors **58** (left) and **59** (right) with factor Xa. Computed structures are from the end of MC simulations covering 20 million configurations. Only hydrogens of the ligand that are proximal to the indole side chain of Trp215A are shown.

the greater activity of **59** than **58**. The difference in the computed protein–ligand interaction energies is 3.9 kcal/mol in favor of **58**. This value is justified, since the smaller **58** is better accommodated in the pocket and makes more favorable N–H⋯π interactions with the indole ring of Trp215A (Figure 7). Solvation of both compounds in the bound state is similar, as reflected

in the computed difference in the ligand–solvent energies of only 0.8 kcal/mol for the **58** → **59** conversion. In contrast, the computed difference in interaction energies with the solvent for **58** and **59** unbound in water is 12 kcal/mol. This clearly reflects the enhanced solvation of the dihydroimidazolium group in **58** relative to the *N*-methyl analogue in **59**. Thus, **58** pays a significantly

Table 5. Results for a Test Set of Inactive Diaryloxypyridine Inhibitors
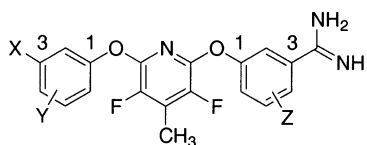
no.	X	Y	Z	ΔG_{expt}[a,b]	ΔG_{calc}
61	H	H	H	>-7.2	-7.30
62	CO−piperazineMe	H	H	>-7.2	-8.58
63	CONHPh	H	H	>-7.2	-8.32
64	CONMe$_2$	5,6-OMe	H	>-7.2	-8.17
65	CONMe$_2$	5-OH	H	>-7.2	-8.00
66	CONMe$_2$	6-Me	H	>-7.2	-7.88
67	CONMe$_2$	6-OH	H	>-7.2	-7.70
68	CONMe$_2$	6-OMe	H	>-7.2	-8.40
69	NMe$_2$	6-COOEt	H	>-7.2	-7.95
70	NMe$_2$	6-Me	H	>-7.2	-8.09
71	NMe$_2$	6-OMe	H	>-7.2	-8.36
72	CONMe$_2$	H	2-OH	>-7.2	-8.58
73	CONMe$_2$	H	2-OH, 5-OMe	>-7.2	-7.86
74	NMe$_2$	H	2,6-OMe	>-7.2	-8.51

^{a} References 10, 11. ^{b} Experimental free energies from $\Delta G_{expt} = RT \ln K_i$ in kcal/mol; all K_i values are >5000 nM.

higher desolvation penalty upon binding, and this dominates its more favorable protein−ligand interaction.

Finally, an additional test for regression equation 4 was carried out in the spirit of discriminating between potential leads and poor inhibitors. Specifically, MC/ELR calculations were performed for a set of compounds not used in deriving eq 4 and that were all inactive in the experimental assay with observed K_is of >5 μM (Table 5). Using eq 4, the activities for these compounds were consistently predicted to be in the low range of all computed and observed results (Tables 1−4, Figure 2).

Conclusion

A simple regression equation has been developed from the results of Monte Carlo simulations for the estimation of activities for a large set of factor Xa inhibitors. The derived regression, eq 4, has two physically meaningful parameters: (1) the protein−ligand Lennard-Jones interaction energy and (2) the number of direct protein−ligand hydrogen bonds. The equation displays good agreement between computed and experimental activities with an r^2 of 0.74, a cross-validated q^2 of 0.63, rms error of 0.67 kcal/mol, and an average unsigned error of 0.60 kcal/mol (Figure 2). The data set employed in this study (60 compounds) lends additional support to the utility of the ELR method as a tool for elucidating key structural and energetic information of importance in protein−ligand binding. Furthermore, inhibitors that had low activities and were not included in the training set were successfully predicted to be poor inhibitors, confirming the usefulness of the derived regression in predicting activities of new compounds. These results also extend our database of ELR results for protein−ligand complexes that is forming the basis for the development of a general predictive tool for enzyme inhibition.[6−9]

Furthermore, these computations provide insight through eq 4 as to what structural features should be beneficial in improving the activity of the compounds in this series. Further optimization of the molecular fragment in the S4 pocket to enhance van der Waals interactions with the surrounding aromatic protein resi-dues is suggested. In addition, for the S1 pocket, substituents on the aryloxy ring already participate in the amidine/Asp189A and 6-OH/Ser195A hydrogen bonds. However, the presence of the bridging water molecule between the 6-OH and His57A suggests that elaboration of the ligand to replace the water molecule and hydrogen bond directly to His57A would also be beneficial.

References

(1) Davie, E. W.; Fujikawa, K.; Kisiel, W. The coagulation cascade: Initiation, maintenance, and regulation. *Biochemistry* **1991**, *30*, 10363−70.

(2) Vlasuk, G. P. Direct Factor Xa Inhibitors. In *New Therapeutic Agents in Thrombosis and Thrombolysis*, 1st ed.; Sasahara, A. A., Loscalzo, J., Ed.; Marcel Dekker: New York, 1997.

(3) Kaiser, B. Factor Xa−A promising target for drug development. *Cell. Mol. Life Sci.* **2002**, *59*, 189−192.

(4) Mann, K. G.; Nesheim, M. E.; Church, W. R.; Haley, P.; Krishnaswamy, S. Surface Dependent Reactions of the Vitamin K-Dependent Enzyme Complexes. *Blood* **1990**, *76*, 1−16.

(5) Prasa, D.; Svendsen, L.; Sturzebecher, J. Inhibition of thrombin generation in plasma by inhibitors of factor Xa. *Thromb. Haemost.* **1997**, *78*, 1215−1220.

(6) Pierce, A. C.; Jorgensen, W. L. Estimation of Binding Affinities for Selective Thrombin Inhibitors via Monte Carlo Simulations. *J. Med. Chem.* **2001**, *44*, 1043−50.

(7) Wesolowski, S. S.; Jorgensen, W. L. Estimation of binding affinities for celecoxib analogues with COX-2 via Monte Carlo-extended linear response. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 267−70.

(8) Rizzo, R. C.; Tirado-Rives, J.; Jorgensen, W. L. Estimation of Binding Affinities for HEPT and Nevirapine Analogues with HIV-1 Reverse Transcriptase via Monte Carlo Simulations. *J. Med. Chem.* **2001**, *44*, 145−154.

(9) Rizzo, R. C.; Udier-Blagović, M.; Wang, D.; Watkins, E. K.; Smith, M. B. K.; Smith, R. H. J.; Tirado-Rives, J.; Jorgensen, W. L. Prediction of Activity for Nonnucleoside Inhibitors with HIV-1 Reverse Transcriptase Based on Monte Carlo Simulations. *J. Med. Chem.* **2002**, *45*, 2970−87.

(10) Phillips, G.; Davey, D. D.; Eagen, K. A.; Koovakkat, S. K.; Liang, A.; Ng, H. P.; Pinkerton, M.; Trinh, L.; Whitlow, M.; Beatty, A. M.; Morrissey, M. M. Design, synthesis, and activity of 2,6-diphenoxypyridine-derived factor Xa inhibitors. *J. Med. Chem.* **1999**, *42*, 1749−1756.

(11) Phillips, G.; Guilford, W. J.; Buckman, B. O.; Davey, D. D.; Eagen, K. A.; Koovakkat, S.; Liang, A.; McCarrick, M.; Mohan, R.; Ng, H. P.; Pinkerton, M.; Subramanyam, B.; Ho, E.; Trinh, L.; Whitlow, M.; Wu, S.; Xu, W.; Morrissey, M. M. Design, synthesis, and activity of a novel series of factor Xa inhibitors: Optimization of arylamidine groups. *J. Med. Chem.* **2002**, *45*, 2484−2493.

(12) Jorgensen, W. L. Free Energy Changes in Solution. In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Ed.; Wiley: New York, 1998; Vol. 2, pp 1061−1070.

(13) Lamb, M. L.; Jorgensen, W. L. Computational approaches to molecular recognition. *Curr. Opin. Chem. Biol.* **1997**, *1*, 449−457.

(14) Kollman, P. Free Energy Calculations: Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* **1993**, *93*, 2395−2417.

(15) Jorgensen, W. L. Free Energy Calculations: A Breakthrough For Modeling Organic Chemistry in Solution. *Acc. Chem. Res.* **1989**, *22*, 184−189.

(16) 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.

(17) (a) Carlson, H. A.; Jorgensen, W. L. An Extended Linear Response Method For Determining Free Energies of Hydration. *J. Phys. Chem.* **1995**, *99*, 10667−10673. (b) McDonald, N. A.; Carlson, H. A.; Jorgensen, W. L. Free energies of solvation in chloroform and water from a linear response approach. *J. Phys. Org. Chem.* **1997**, *10*, 563−576.

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

(19) Paulsen, M. D.; Ornstein, R. L. Binding free energy calculations for P450cam-substrate complexes. *Protein Eng.* **1996**, *9*, 567−571.

(20) Hulten, J.; Bonham, N. M.; Nillroth, U.; Hansson, T.; Zuccarello, G.; Bouzide, A.; Aqvist, J.; Classon, B.; Danielson, U. H.; Karlen, A.; Kvarnstrom, I.; Samuelsson, B.; Hallberg, A. Cyclic HIV-1 Protease Inhibitors Derived from Mannitol: Synthesis, Inhibitory Potencies, and Computational Predictions of Binding Affinities. *J. Med. Chem.* **1997**, *40*, 885−897.

(21) Hansson, T.; Marelius, J.; Aqvist, J. Ligand binding affinity prediction by linear interaction energy methods. *J. Comput.-Aided Mol. Des.* **1998,** *12,* 27−35.

(22) 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.

(23) Jones-Hertzog, D. K.; Jorgensen, W. L. Binding affinities for Sulfonamide Inhibitors with human Thrombin Using Monte Carlo Simulations with a Linear Response Methodol. *J. Med. Chem.* **1997**, *40*, 1539−49.

(24) Smith, R. H.; Jorgensen, W. L.; Tirado-Rives, J.; Lamb, M. L.; Janssen, P. A. J.; Michejda, C. J.; Smith, M. B. K. Prediction of Binding Affinities for TIBO Inhibitors of HIV-1 Reverse Transcriptase Using Monte Carlo Simulations in a Linear Response Methodol. *J. Med. Chem.* **1998**, *41*, 1, 5272−5286.

(25) Lamb, M. L.; Tirado-Rives, J.; Jorgensen, W. L. Estimation of the binding affinities of FKBP12 inhibitors using a linear response method. *Bioorg. Med. Chem.* **1999**, *7*, 851−860.

(26) 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.

(27) Udier-Blagović, M.; Watkins, E. K.; Tirado-Rives, J.; Jorgensen, W. L. Activity Predictions for Efavirenz Analogues with the K103N Mutant of HIV Reverse Transcriptase. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 3337-3340.

(28) Adler, M.; Davey, D. D.; Phillips, G. B.; Kim, S. H.; Jancarik, J.; Rumennik, G.; Light, D. R.; Whitlow, M. Preparation, characterization, and the crystal structure of the inhibitor ZK-807834 (CI-1031) complexed with factor Xa. *Biochemistry* **2000**, *39*, 12534−12542.

(29) Tirado-Rives, J. *CHOP,* Version 1.0; Yale University: New Haven, CT, 2001.

(30) Tirado-Rives, J. *PEPZ,* Version 1.0; Yale University: New Haven, CT, 1997.

(31) Jorgensen, W. L. *GenMol,* Version 2.2; Yale University: New Haven, CT, 2003.

(32) 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.

(33) 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.

(34) 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.

(35) Jorgensen, W. L. *MCPRO,* Version 1.67; Yale University: New Haven, Connecticut, 2001.

(36) Rizzo, R. C.; Wang, D.; Tirado-Rives, J.; Jorgensen, W. L. Validation of a Model for the Complex of HIV-1 Reverse Transcriptase with Sustiva through Computation of Resistance Profiles. *J. Am. Chem. Soc.* **2000**, *122*, 12898−12900.

(37) Udier-Blagović, M.; Tirado-Rives, J.; Jorgensen, W. L. Validation of a Model for the Complex of HIV-1 Reverse Transcriptase with Nonnucleoside Inhibitor TMC125. *J. Am. Chem. Soc.* **2003**, *125*, 6016−6017.

(38) *JMP,* Version 4; SAS Institute Inc., Cary, NC, 2000.

(39) Padmanabhan, K.; Padmanabhan, K. P.; Tulinsky, A.; Park, C. H.; Bode, W.; Huber, R.; Blankenship, D. T.; Cardin, A. D.; Kisiel, W. Structure of human des(1−45) factor Xa at 2.2 A resolution. *J. Mol. Biol.* **1993**, *232*, 947−966.

(40) Herron, D. K.; Goodson, T., Jr.; Wiley, M. R.; Weir, L. C.; Kyle, J. A.; Yee, Y. K.; Tebbe, A. L.; Tinsley, J. M.; Mendel, D.; Masters, J. J.; Franciskovich, J. B.; Sawyer, J. S.; Beight, D. W.; Ratz, A. M.; Milot, G.; Hall, S. E.; Klimkowski, V. J.; Wikel, J. H.; Eastwood, B. J.; Towner, R. D.; Gifford-Moore, D. S.; Craft, T. J.; Smith, G. F. 1,2-Dibenzamidobenzene inhibitors of human factor Xa. *J. Med. Chem.* **2000**, *43*, 859−872.

(41) Verner, E.; Katz, B. A.; Spencer, J. R.; Allen, D.; Hataye, J.; Hruzewicz, W.; Hui, H. C.; Kolesnikov, A.; Li, Y.; Luong, C.; Martelli, A.; Radika, K.; Rai, R.; She, M.; Shrader, W.; Sprengeler, P. A.; Trapp, S.; Wang, J.; Young, W. B.; Mackman, R. L. Development of serine protease inhibitors displaying a multicentered short (<2.3 Å) hydrogen bond binding mode: Inhibitors of urokinase-type plasminogen activator and factor Xa. *J. Med. Chem.* **2001**, *44*, 2753−2771.