Binding Affinities for Sulfonamide Inhibitors with Human Thrombin Using Monte Carlo Simulations with a Linear Response Method

Deborah K. Jones-Hertzog[†] and William L. Jorgensen*

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

Received October 1, 1996[⊗]

The binding of sulfonamide inhibitors to human thrombin is examined to evaluate the viability of calculating free energies of binding, ΔG_b, utilizing Monte Carlo (MC) statistical mechanics with a linear response approach. Coulombic and van der Waals energy components determined from MC simulations of the bound and unbound inhibitors solvated in water plus a solvent-accessible surface area term, as an index for cavity formation, were correlated with the free energies of binding for the inhibitor MD-805 and six derivatives. The best correlations yield an average error of 0.8 kcal/mol for the seven binding affinities, which cover an observed range of 6.0 kcal/mol. The MC simulations also provided insights into the interactions occurring in the active site and the origins of variations in ΔG_b. Equatorial placement of the carboxylate group at C2 in the piperidine ring of the inhibitors causes electrostatic destabilization with the side chain of Glu-H192, while axial disposition of the C4-methyl group reduces favorable hydrophobic interactions in the P-pocket of the enzyme.

Introduction

A common strategy utilized for the calculation of free energies of binding (ΔG_b) of enzyme−inhibitor complexes is to carry out molecular dynamics (MD) or Monte Carlo (MC) simulations with free energy perturbation (FEP) methods or thermodynamic integration (TI).[1] There are limitations to these approaches,[1] so there is need for alternate procedures and approximate methods to allow for more facile treatment of large numbers of structurally diverse systems. One such semiempirical approach, the linear response (LR) method devised by Åqvist *et al.*,[2] requires only simulations at the endpoints of mutations; a linear combination of the differences in the average inhibitor−environment interaction energies between the bound and unbound states are used to determine ΔG_b. For the unbound system, these energies arise solely from inhibitor−solvent interactions, while inhibitor−solvent plus inhibitor−protein interactions result in the bound-state energies.

In the Åqvist LR expression (eq 1), the interactions are broken down into electrostatic (Coulombic) and van der Waals contributions. A van der Waals scaling

$$\Delta G_b = \beta\langle\Delta U_{elec}\rangle + \alpha\langle\Delta U_{vdw}\rangle \qquad (1)$$

factor, $\alpha = 0.161$, was derived to give the best fit to experimental binding data, and the electrostatic scaling factor, $\beta = 0.5$, follows from the quadratic dependence of free energy on solute charge, as embodied in the Born model for ion solvation. MD simulations were used to determine the required energy components. The resulting free energies of binding for several protein−ligand systems correlated well with experimental data. Furthermore, the α and β scaling parameters were found to be transferable between proteins.[2] The methodology has also been applied recently with excellent success to a larger dataset of 11 inhibitors with cytochrome P450cam.[3] A different force field was used, and an increase of α to 1.043 was found to be appropriate in this case.[3]

Extension of the LR approach to calculate free energies of hydration (ΔG_{hyd}) incorporated a third term proportional to the solute's solvent-accessible surface area (SASA), as an index for the cost of cavity formation within the solvent.[4] The latter is needed for cases with positive ΔG_{hyd} such as aklanes. It was also found that additional improvement occurred when both α and β were allowed to vary. Equation 2 gives the corresponding expression for ΔG_b.

$$\Delta G_b = \beta\langle\Delta U_{elec}\rangle + \alpha\langle\Delta U_{vdw}\rangle + \gamma\langle\Delta SASA\rangle \qquad (2)$$

In the present work, the LR expressions, eqs 1 and 2, have been applied to complexes of the serine protease thrombin and sulfonamide inhibitors. The objectives of this study are 2-fold: to evaluate the viability of LR methods for calculating free energies of binding and to obtain insights into key thrombin−inhibitor interactions that lead to variations in the binding affinities.

Due to its central role in thrombosis and haemostatis, thrombin has been implicated in many disease states,[5] such as myocardial infarction, stroke, and pulmonary embolism, and thus is currently receiving wide attention. It is a multifunctional enzyme which plays a critical role in the blood coagulation cascade mechanism,[6a] including thrombus formation through conversion of its natural substrate, fibrinogen, to fibrin and induction of platelet aggregation *via* thrombin receptor cleavage.[6b,c] Thrombus formation involves selective cleavage of the peptide bond adjacent to an arginine residue in fibrinogen, yielding fibrin monomers that polymerize forming a network of fibers. This is the major component of a blood clot that traps platelets and plasma proteins blocking blood flow at the site of injury.[6d−f] Additional stability is imparted to the fibrin structure by a thrombin-activated transamidase that forms cross-linking bonds between lysine and glutamic acid side chains.

Crystallographic investigations[7] have revealed that thrombin is composed of disulfide-linked A and B chains. It also has an anion binding exosite positioned approximately 20 Å from the active site along a groove. The active site has three binding pockets: (1) the S1

[†] Current address: Neurogen Corp., Branford, CT 06405.
[⊗] Abstract published in *Advance ACS Abstracts,* April 1, 1997.

S0022-2623(96)00684-X CCC: $14.00

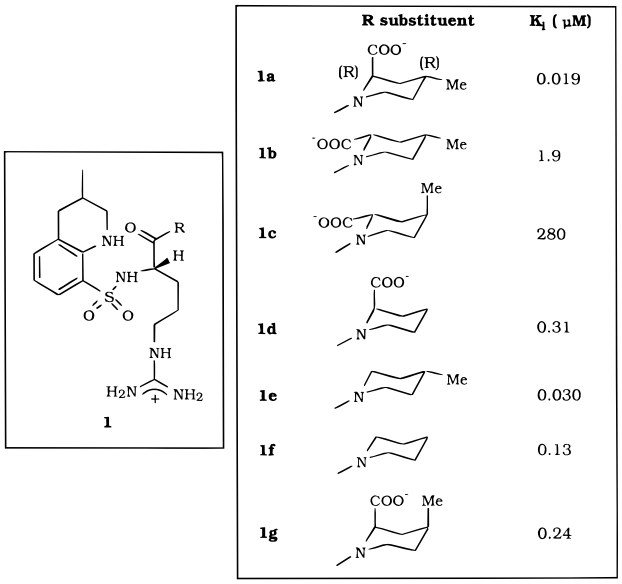

Figure 1. Structures and experimental inhibition constants (ref 7h) for sulfonamide inhibitors of human thrombin.

recognition pocket, with residue Asp189 at its base, (2) a large hydrophobic D-pocket (D designates "distal" from the catalytic site), and (3) a small hydrophobic P-pocket (P designates "proximal" to the catalytic site). The catalytic triad of Asp102, His57, and Ser195 is responsible for the proteolytic activity. The crystal structure of the complex with the sulfonamide inhibitor, (2*R*,4*R*)-4-methyl-1-[N2-[(3-methyl-1,2,3,4-tetrahydro-8-quinolinyl)sulfonyl]-L-arginyl]-2-piperidinecarboxylic acid (**1a**)[7b] (see Figure 1), also known as MD-805, Argatroban, MQPA, Novastan, and Slounon, was used as the structural starting point of the present work. The extensive inhibition constant (K_i) data available for **1a** and its analogs, **1b**−**g**,[7h] which feature relatively simple structural changes at the 2- and 4-positions of the piperidine ring (see Figure 1), made this an attractive series of compounds for computational studies.

Previous theoretical studies of thrombin have included use of molecular mechanics to design new inhibitors, correlation of binding affinities with non-bonded interaction energies obtained from molecular mechanics energy minimization for unsolvated thrombin−inhibitor complexes, FEP calculations for complexes, and active-site mapping.[8]

Table 1. Sulfonamide Nonbonded Parameters and Partial Charges

atom	σ (Å)	ϵ (kcal/mol)	q (e)
S	3.55	0.25	1.22
O	2.96	0.17	−0.61
N	3.25	0.17	−0.80
H	0.00	0.00	0.41

Computational Details

Parameterization. All computations have used the OPLS force field with an all-atom representation of the inhibitors.[9] As usual with a new project, some force-field parameters were missing. First, it was necessary to develop Lennard−Jones parameters, σ and ϵ, to represent the S and O of the sulfonamide moiety. An established protocol was employed of pure liquid simulations with iterative modification of σ and ϵ until experimental physical data, particularly, heat of vaporization and density, were reproduced. Simulations of liquid divinyl sulfone with the MC program BOSS[10] yielded the missing σ and ϵ for the S−O unit, as listed in Table 1.[11] Partial charges for the sulfonamide subunit were obtained from model compounds by fitting it to the electrostatic potential surfaces from ab initio molecular orbital calculations with the 6-31+G* basis set (see Figure 2 and Table 1).[12] The bond-stretching and angle-bending parameters for the sulfonamide group were adopted from a study by Rossi *et al.*[13] on a sulfonamide/human carbonic anhydrase complex, as summarized in Table 2.[14]

The missing torsional parameters were derived using the model sulfonamides **2**−**5** (Figure 2). Ab initio calculations with the 6-31+G* basis set yielded energy profiles for rotation about the H−N−S−C bond in **2**, H−C−N−S and C−N−S−C bonds in **3**, C−C−N−S bond in **4**, and C−C−S−N bond in **5**. Least-squares fits of these results with remaining parameters from the OPLS force field provided the torsional parameters shown in Table 3. Torsion terms for dihedrals to the oxygens on sulfur are not needed with these parameters.

Monte Carlo Simulations. A full-geometry optimization for inhibitor **1a** by itself was carried out starting from its structure in the observed complex with thrombin.[7b] This gave the initial structure for the MC simulations for the unbound inhibitors in water. In all cases, the inhibitor was surrounded by a water sphere of ca. 20 Å radius containing 1101 TIP4P[15]

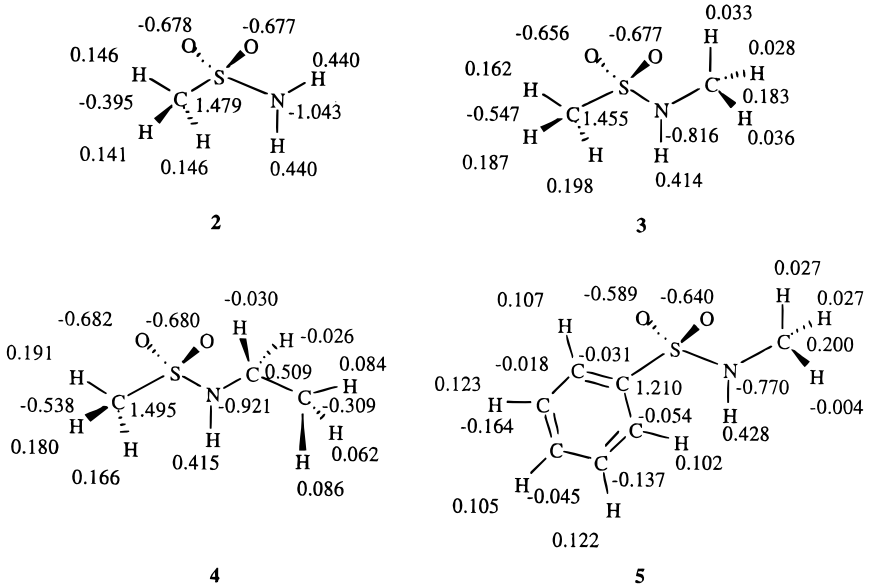

Figure 2. Model compounds used in the development of the partial charges and torsional parameters for sulfonamide groups. Partial charges from fitting to the electrostatic potential surface from ab initio 6-31+G* calculations.

Table 2. New Bond Length and Bond Angle Parameters for Inhibitors **1a**−**g**[a]

bond	K_r (kcal/mol·Å^2)	r_{eq} (Å)
	Sulfonamide Moiety	
S═O	700.0	1.44
S−N	434.0	1.67
S−C(ar)[b]	340.0	1.77
bond angle	K_ϑ (kcal/mol·rad^2)	ϑ_{eq} (deg)
	Sulfonamide Moiety	
O═S═O	104.0	119.0
S−N−C	50.0	120.0
S−N−H	55.0	111.0
O═S−C(ar)	74.0	107.2
N−S−C(ar)	100.0	103.0
S−C(ar)−C(ar)	85.0	119.4
O═S−N	120.0	107.0
	Quinoline Moiety	
C−N−H	35.0	118.4
C(ar)−C−H	35.0	109.4
C(ar)−C(ar)−C	70.0	120.0
C(ar)−C−C	63.0	114.0
N−C−C	80.0	111.2

[a] AMBER (ref 17a) bond length and bond angle parameters used otherwise. [b] ar designates an aromatic atom.

Table 3. Torsional Parameters (kcal/mol) for Sulfonamides **1a**−**g**

	V0	V1	V2	V3
	All Atom[a]			
H−N−S−C(ar)	2.836	1.671	−4.901	0.669
H−C−N−S	0.120	1.362	−1.457	0.149
C−N−S−C(ar)	−0.760	2.074	−2.966	2.473
C−C−N−S	−0.088	2.929	−2.533	0.497
C(ar)−C(ar)−S−N	0.010	1.656	−0.768	−0.117
C−C−N−H	0.000	0.000	0.000	0.000
H−C−N−H	0.000	0.000	0.000	0.000
	United Atom[b]			
H−N−S−C(ar)	2.836	1.671	−4.901	0.669
C−N−S−C(ar)	−0.152	2.536	−4.057	2.553
C−C−N−S	−0.086	2.436	−1.724	2.034
C(ar)−C(ar)−S−N	−0.063	1.876	−0.812	−0.187

[a] All-atom parameters used in MC simulations with MCPRO. [b] United-atom parameters used during minimization of the enzyme−inhibitor complex with AMBER.

water molecules. The MC simulations were carried out with MCPRO.[16] This program is a derivative of BOSS;[10] however, it includes a knowledge of residues that are used for residue-based cutoff procedures, that form the basis of lists that aid in computational efficiency and that allow Monte Carlo moves of one residue at a time. The MC simulation for unbound **1a** initially involved (1) 10×10^6 configurations of equilibration with only the water moving, (2) 3×10^6 configurations equilibration with the water and inhibitor moving, and (3) 3×10^6 configurations of averaging. The simulations were carried out with solvent−solvent, solvent−solute, and non-bonded cutoffs of 8.5, 15, and 15 Å, respectively. A residue-based cutoff procedure is implemented in MCPRO. The water−water cutoff is based on the O−O separation only; if the distance is within the cutoff, the entire water−water interaction is included. For the solute−water interactions, if any distance between an atom in a residue and the water O is within the cutoff, the entire residue−water interaction is included. The inhibitors were treated as one residue. The temperature was constant at 37 °C, as in the binding experiments.[7h] Bond lengths were fixed, while bond angles and torsion angles were sampled except for those in the quinoline moiety.

After geometry optimization and solvation as above, MC simulations of unbound inhibitors **1b**−**g** were performed in a similar manner: (1) 5×10^6 configurations of equilibration for the water only, (2) 3×10^6 configurations equilibration for the water and inhibitor, and (3) 3×10^6 configurations of averaging. In order to enhance the precision of the results, the averaging periods were then extended to 9×10^6 for inhibitors **1b**−**g** and to 12×10^6 for the parent **1a**.

For the thrombin−**1a** complex, the crystal structure was first energy-optimized using AMBER.[17,18] Then, amino acid (aa) residues more than approximately 16 Å from the active site were removed to reduce the complex to a computationally more tractable 164 residue fragment with **1a** bound in the active site. Acetyl and *N*-methylamide blocking groups were used to cap the truncation points, and overall charge neutrality was maintained for the complex.

An all-atom representation of bound **1a** was used with the same variable intramolecular degrees of freedom as in the unbound case. For the enzyme fragment, a partial united-atom model was employed with only hydrogens on saturated carbon not explicit.[9] The complex was solvated with a 20 Å cap containing 469 TIP4P water molecules centered on the inhibitor. The protein backbone and any aa residues protruding from the water cap were held fixed. Using an internal-coordinate representation, the sampling of the protein side chains involved variation of all bond angles and torsion angles. On an attempted MC move for the protein, one residue is chosen at random and all its variable angles are randomly modified. On an attempted move for the inhibitor, all its variable angles are changed, and it is totally translated and rotated as well. Since the TIP4P water molecules are rigid, their attempted moves involve just the total translations and rotations. Attempts to move a protein residue and the inhibitor were made every 10 and 24 configurations, respectively. The remaining attempted moves were for the water.

The MC simulation procedure for the solvated enzyme−**1a** complex involved (1) 10×10^6 configurations of equilibration with only water moves, (2) 3×10^6 configurations of equilibration with water and complex moves, and (3) 3×10^6 configurations of averaging. Other simulation variables were the same as in the unbound calculations. The residue−water and residue−inhibitor interactions were cutoff at 15 Å. Since an experimental structure has only been reported for the thrombin/MD-805 complex, the **1a** complex from the MC simulation was modified to give the initial structures for the complexes with **1b**−**g**. These were then simulated in the same manner. An additional 3×10^6 configurations of averaging was performed subsequently for the **1a** complex.

Results and Discussion

Optimization of the LR Equations. The average electrostatic and van der Waals interaction energies between the bound and free inhibitors and their respective environments were determined during the MC simulations (see Table 4). The average SASA terms were calculated using the SAVOL2 program[19] and the radii established by Rashin.[20,21] In this program molecules are represented with united atoms. The average SASA was determined from structures saved every 2×10^5 configurations during the MC runs, as summarized in Table 5.

The LR expression (eq 1) and α (0.161) and β (0.5) parameters from Åqvist[2a−c] were ineffective at reproducing the experimental ΔG_b values for inhibitors **1a**−**g** with the energetic results from the present simulations; the rms deviation between the computed and observed values is 15 kcal/mol. Åqvist has used the GROMOS force field, and it is not surprising that the parameters are not transferable to the present study. The Carlson−Jorgensen LR equation (2) with the α, β, and γ parameters developed for ΔG_{hyd} calculations[4] gave more reasonable ΔG_b values, but the rms error of 3 kcal/mol is still unacceptable for use as a predictive tool.

Therefore, new α, β, and γ parameters were derived by fitting the present energetic and SASA results to the experimental ΔG_b values with a Simplex-based program. The results of the various fits are summarized

Table 4. Average Interaction Energies and Standard Deviations from the Bound and Unbound MC Simulations (kcal/mol)[a]

inhibitor	config × 10^6	U_{elec} (inhibitor−solvent)	U_{vdW} (inhibitor−solvent)	U_{elec} (inhibitor−enzyme)	U_{vdw} (inhibitor−enzyme)
1a, unbound	3	−268.17 (2.29)	−25.41 (0.55)		
	6	−273.62 (1.91)	−25.68 (0.45)		
	9	−271.85 (1.46)	−26.12 (0.38)		
	12	−267.20 (1.61)	−26.01 (0.33)		
1a, bound	3	−170.22 (1.11)	0.03 (0.26)	−76.22 (0.64)	−51.59 (0.32)
	6	−167.79 (0.90)	−0.47 (0.20)	−78.82 (0.70)	−50.79 (0.26)
1b, unbound	3	−261.93 (2.69)	−23.59 (0.55)		
	6	−258.83 (1.79)	−24.47 (0.52)		
	9	−259.48 (1.45)	−25.76 (0.36)		
1b, bound	3	−165.52 (1.38)	−0.97 (0.48)	−65.14 (1.27)	−52.80 (0.45)
1c, unbound	3	−278.60 (2.39)	−24.35 (0.73)		
	6	−278.95 (1.43)	−23.88 (0.46)		
	9	−280.89 (1.13)	−24.61 (0.38)		
1c, bound	3	−175.81 (1.60)	−4.49 (0.55)	−60.77 (0.46)	−49.10 (0.28)
1d, unbound	3	−258.32 (2.04)	−23.58 (0.66)		
	6	−259.30 (1.49)	−24.34 (0.43)		
	9	−256.31 (1.58)	−25.29 (0.45)		
1d, bound	3	−165.58 (1.58)	−1.18 (0.37)	−84.53 (0.81)	−47.47 (0.40)
1e, unbound	3	−196.84 (2.20)	−33.56 (0.35)		
	6	−194.94 (1.62)	−33.48 (0.37)		
	9	−191.60 (1.49)	−33.37 (0.37)		
1e, bound	3	−16.39 (0.80)	−8.41 (0.20)	−174.48 (0.72)	−49.10 (0.28)
1f, unbound	3	−192.20 (2.12)	−31.00 (0.68)		
	6	−194.72 (1.41)	−31.61 (0.47)		
	9	−195.32 (1.12)	−32.56 (0.40)		
1f, bound	3	−21.56 (0.68)	−9.61 (0.35)	−163.15 (0.32)	−47.61 (0.17)
1g, unbound	3	−273.47 (2.20)	−23.71 (0.73)		
	6	−269.59 (1.93)	−23.60 (0.44)		
	9	−266.62 (1.63)	−24.08 (0.35)		
1g, bound	3	−154.59 (1.39)	3.50 (0.34)	−91.24 (0.94)	−50.73 (0.40)

[a] Standard deviations from batch means calculations given in parentheses.

Table 5. Calculated Average Solvent-Accessible Surface Areas (Å^2)

	SASA			
	3×10^6 configs	6×10^6 configs	9×10^6 configs	12×10^6 configs
1a, unbound	722.4	721.6	718.7	716.9
1a, bound	106.6	109.3		
1b, unbound	685.3	687.3	688.0	
1b, bound	100.2			
1c, unbound	705.9	703.9	704.2	
1c, bound	107.9			
1d, unbound	667.4	670.5	677.8	
1d, bound	113.5			
1e, unbound	679.6	684.1	686.8	
1e, bound	97.5			
1f, unbound	668.0	665.2	669.4	
1f, bound	98.5			
1g, unbound	686.0	686.6	689.2	
1g, bound	110.4			

in Table 6. Initially, eq 1 was reparameterized by fixing β to 0.5 and optimizing α. The new fit resulted in substantial improvement, but the rms error was still an unacceptable 4.40 kcal/mol. However, upon refitting α and β, the error dropped to 1.34 kcal/mol with β and α reduced to 0.165 and 0.476, respectively. The largest error is 2.4 kcal/mol for **1a**. Addition of the SASA term and reparameterization of α, β, and γ for eq 2 further lowered the rms error to 0.77 kcal/mol. However, α adopted a negative value, which is physically unreasonable. A fit was then made with α fixed at 0.236, which is a value obtained from a correlation for an expanded set of free energies of hydration.[22] This led to an rms error of 1.15 kcal/mol with $\beta = 0.146$ and $\gamma = 0.010$. Finally, a two-parameter fit was tried with the same coefficient for α and β, which resulted in $\alpha = \beta = 0.131$, $\gamma = 0.014$, and an rms error of 1.02 kcal/mol. The principal discrepancy, 1.85 kcal/mol, is for inhibitor **1a**. The weak binding of **1b** and **1c** with the equatorial carboxylate groups is well-reproduced along with the strong binding for the most hydrophobic inhibitors, **1e** and **1f**. The average unsigned error for the last fit with seven inhibitors and two parameters is 0.8 kcal/mol, which can be compared with values of 0.4 and 1.8 kcal/mol reported by Åqvist and co-workers for the smaller sets of four endothiopepsin and three HIV proteinase inhibitors.[2a,c] In an expanded study for a total of 18 protein−inhibitor complexes, Hansson, Marelius, and Åqvist have now obtained an average error of 0.72 kcal/mol using eq 1 with both α and β variable and an average error of 0.54 kcal/mol with a three-parameter equation.[2e]

A few more comments on the fitting are in order. First, the reduction in SASA for these systems falls in a relatively narrow range, 564−698 Å^2 (Table 5). Multiplication times the γ of 0.014 yields an attractive term between 7.9 and 9.8 kcal/mol. In fact, a fit was made to eq 2 with $\alpha = \beta$ and with the SASA term replaced by a constant, δ; the result was $\alpha = \beta = 0.124$, $\delta = -8.23$ kcal/mol, and the rms error only increases to 1.10 kcal/mol. We have retained the SASA term owing to the improved fit and because its value is expected to more clearly emerge for a series of inhibitors that show greater variation in size. Secondly, it may be noted that the values of β obtained here are smaller than the value of 0.5 from the Born equation for the solvation of atomic ions in a uniform dielectric continuum. Values of β near 0.3 and 0.2 have been found to be appropriate for small, neutral solutes in the MD and MC studies with explicit solvent models for water and chloroform, respectively.[2d,4,22] In the present case, the combination of the large solute size, the specific protein environment, and the mix of zwitterionic inhibitors and inhibitors with a net charge of +1 has led to further damping of the

Table 6. Comparison of Experimental and Calculated ΔG_b (kcal/mol) after Refitting the Scaling Factors[a]

	ΔG_b					
compd	exptl	$\beta = 0.5$, $\alpha = 0.734$, $\gamma = 0.0$ (eq 1)	$\beta = 0.165$, $\alpha = 0.476$, $\gamma = 0.0$ (eq 1)	$\beta = 0.075$, $\alpha = -0.462$, $\gamma = 0.038$ (eq 2)	$\beta = 0.146$, ($\alpha = 0.236$), $\gamma = 0.010$ (eq 2)	$\beta = \alpha = 0.131$, $\gamma = 0.014$ (eq 2)
1a	−11.0	−8.2	−8.6	−9.9	−9.0	−9.1
1b	−8.1	−6.1	−8.6	−7.2	−8.3	−8.1
1c	−5.0	0.9	−6.3	−5.9	−6.3	−6.3
1d	−9.2	−14.0	−10.1	−10.2	−10.3	−10.1
1e	−10.7	−17.7	−11.6	−11.0	−11.6	−11.3
1f	−9.8	−12.8	−10.0	−9.5	−10.0	−9.8
1g	−9.4	−6.6	−7.6	−9.7	−8.2	−8.4
rms		4.40	1.34	0.77	1.15	1.02

[a] Using results from the longest Monte Carlo runs in each case.

electrostatic contribution. If the two charged inhibitors, **1e** and **1f**, are removed from the fits, then β increases to 0.20−0.22. A larger database might reveal the necessity of using multiple values of β depending on the nature of the inhibitor, as advocated now by Åqvist and co-workers.[2e]

Another notable point is that the roughly constant attractive term of ca. −8 kcal/mol is providing a base value for these inhibitors that is then modulated by the U_{elec} and U_{vdW} terms. It is important to reemphasize that these two terms only include the changes in unbound ligand−solvent vs bound ligand−solvent plus bound ligand−protein interaction energies. Other terms including the changes in protein−solvent and solvent−solvent interactions are absorbed in the α and β parameters and in the roughly constant term; the gain in solvent−solvent interactions for the released water molecules provides a particularly attractive electrostatic contribution to the ligand binding. Analysis of the energy components in Table 4 shows net positive ΔU_{elec} (1−44 kcal/mol) and net negative ΔG_{vdW} (−23 to −29 kcal/mol) terms. This implies nothing concerning the relative importance of electrostatics and van der Waals interactions for the *absolute* free energies of binding of the ligands; it does mean that the *differences* in the binding affinities for these seven inhibitors are determined more by differences in the electrostatic interactions than the van der Waals ones.

The sign of ΔU_{elec} is sensitive to the combination of the protein's structure and the details of truncation of the potential functions; inclusion of more or less of the protein and/or use of different residue−residue cutoffs could shift all of the ΔU_{elec} values. The present choices of system sizes and cutoffs have evolved through MC and MD studies by us on many systems and are consistent with the high end of choices by others; e.g., Åqvist *et al.* have used 8−15 Å cutoffs with spherical systems of 16−20 Å radius.[2] As emphasized by Åqvist, an essential point is to have the same net charge for the bound and unbound systems for an inhibitor, which is achieved here by having the protein fragment neutral.[2] Otherwise, ambiguous Born-type corrections need to be made. However, for the present protein fragment, there is a relatively high fraction of the positively charged residues near the boundary of the system and, therefore, an excess of negatively charged groups closer to the inhibitor. This is consistent with the very attractive inhibitor−enzyme ΔU_{elec} values (Table 4) for inhibitors **1e** and **1f**, which have a net charge of +1. On the other hand, the other inhibitors, which are zwitterionic, have much more attractive inhibitor−solvent interactions both in the bound and unbound states. The balance of all the effects is such that **1e** and **1f** are observed and computed to be strong binders, which would be difficult to predict *a prior*.

Clearly, it is desirable to test the methodology against still larger databases and multiple protein−ligand systems. Until this is done, many points will remain less than lucid such as the optimal terms to include in the correlation equations and the extent of the transferability of the parameters between systems. Nevertheless, the present results in conjunction with the earlier studies[2−4] are encouraging for the potential utility of such correlative methods as a short cut to predicted free energy changes for many applications. For the free energies of hydration, the database has been expanded to 35 diverse organic molecules with an rms error of 1.0 kcal/mol using eq 2.[22]

Uncertainties in the Fits and LR Components. To determine the uncertainties in the α, β, and γ parameters, a cross-validation technique was employed. Specifically, the parameters were optimized seven individual times such that the data for each inhibitor was left out once. The average values for the parameters and the standard deviations from the seven subset results are reported in Table 7. In all cases, the average values of the seven sets of parameters agree closely with the parameter values obtained from the fits with the full database. The standard deviations for the parameter values are also small.

Of course, the potential accuracy of the LR approaches depends on the precision of the calculation of the energy components and SASA values in Tables 4 and 5. The results as a function of the length of the averaging periods are listed and provide a sense of the associated uncertainties. The average uncertainty in the SASA appears to be 5 Å^2 or less, so the difference between the bound and unbound case should be no more than 10 Å^2. With the optimized γ coefficients of near 0.015 kcal/mol Å^2, the associated uncertainty in the free energy of binding is less than 0.15 kcal/mol. For the energy components in Table 4, statistical uncertainties were estimated with the batch means procedure during the Monte Carlo runs using batch sizes of 3×10^5 configurations. For the same averaging periods the uncertainties are about twice as large for the unbound calculations than for the bound ones. This is the reason for the extension of the averaging for the unbound calculations to $(9-12) \times 10^6$ configurations. The observation is consistent with our continual experience that simulations in water are always noisier than in organic solvents, and the environment of a protein is more like an organic solvent than water. The batch means procedure yielded uncertainties (1σ) of 0.2−0.6 kcal/mol

Table 7. Summary and Results of Cross-Validation for the α, β, and γ Parameters Determined by Fitting to the Experimental ΔG_b Data[a,b]

linear response method	α	β	γ (kcal/mol Å^2)	rms to expt (kcal/mol)
1/2 Coulombic and van der Waals	0.734 *0.747 ± 0.031*	(0.500)[c]		4.40
Coulombic and van der Waals	0.476 *0.474 ± 0.032*	0.165 *0.160 ± 0.019*		1.34
Coulombic, van der Waals, and SASA	−0.462 *0.462 ± 0.148*	0.075 *0.071 ± 0.020*	0.038 *0.038 ± 0.006*	0.77
Coulombic, van der Waals, and SASA	(0.236)[c]	0.146 *0.142 ± 0.027*	0.010 *0.010 ± 0.001*	1.15
Coulombic, van der Waals, and SASA	0.131[d] *0.126 ± 0.029*	0.131[d] *0.126 ± 0.029*	0.014 *0.014 ± 0.000*	1.02

[a] Cross-validated values with uncertainties given in italics. [b] Parameters from fitting results from the longest MC simulations. [c] Fixed value. [d] α set equal to β.

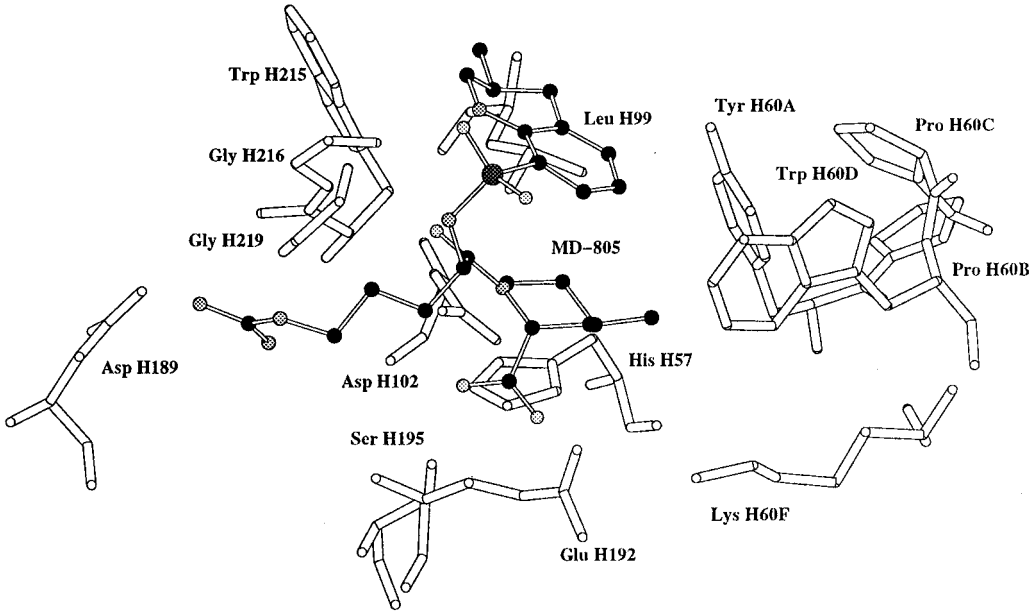

Figure 3. Residues of thrombin near the binding site for the sulfonamide inhibitors. The quinoline and piperidine rings are in the hydrophobic D- and P-pockets, respectively.

for the van der Waals energies at the ends of the runs. The oscillations of U_{vdw} in Table 4 are consistent with this, which with an α of 0.1−0.2 contribute an uncertainty of ca. 0.1 kcal/mol to the computed ΔG_b values. The greatest uncertainty stems from the electrostatic energy terms, which have standard deviations of 0.5−1.6 kcal/mol at the ends of the runs from the batch means procedure. Given the present β values of 0.1−0.2, the electrostatic terms contribute an uncertainty of ca. 0.3 kcal/mol to the free energies of binding. Thus, from this rough analysis, the total statistical uncertainty in the computed ΔG_b values from the noise in the Monte Carlo simulations is ca. 0.5 kcal/mol. The average deviation between the experimental and computed values for ΔG_b is not going to be less than this except by chance.

The issue of convergence clearly needs to be monitored in all applications of such correlative methods. The precision can be improved with ever longer simulations. For the present study, improvement in the correlation between the computed and experimental binding data was obtained as the simulations were extended. However, the gains turned out to be relatively modest. Specifically, when fits to eq 2 with $\alpha = \beta$ are made using the bound results after 3×10^6 configurations and the unbound results after 3, 6, and 9×10^6 configurations of averaging, the rms errors for ΔG_b are 1.24, 1.30, and 1.18 kcal/mol, respectively, which are not far from the final result of 1.02 kcal/mol in Table 6. Although this may not be a general finding, it suggests the qualitatively useful results can possibly be obtained from averaging runs of ca. 3×10^6 configurations for systems of the present size.

On the other hand, it should be realized that values of α, β, and γ that are larger than those found in this study will lead to proportionately greater uncertainties. Also, the choice of initial structures, as always in simulation studies, is critical; convergence to the correct structure from a poor starting point may not be possible in a simulation of tenable length. The other obvious issue affecting the generality of the present LR approach is the absence of terms for intramolecular energetics in eqs 1 and 2. This simplification can be expected to be problematic for comparison of ligands with significantly varying flexibility and for larger alterations that require varying distortion of the protein upon binding. For the present system, addition of terms to the LR equations for changes in intramolecular energetics did not lead to notable improvement in the fits.

Structures of the Complexes. The inhibitors have three subunits, which match up with the three pockets in the binding site from the crystal structure with **1a**.[7b] As reflected in Figure 3, the quinoline ring resides in the hydrophobic D-pocket, the piperidine ring is in the

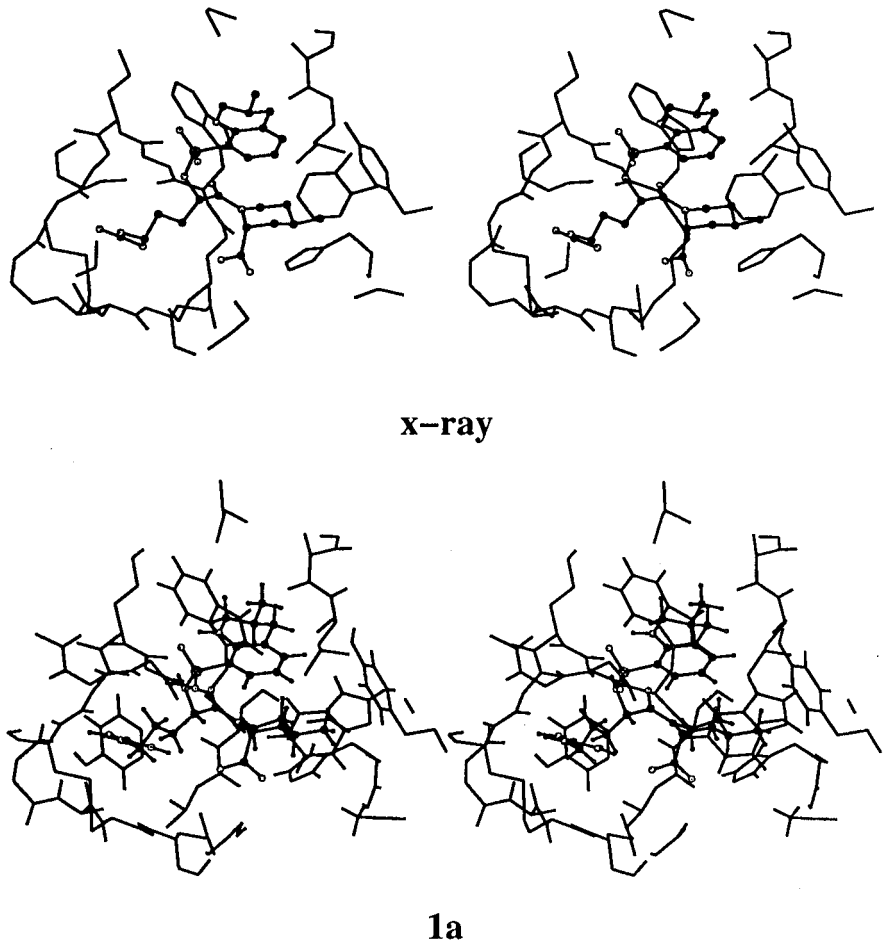

Figure 4. Stereoplots of the binding site for the complex of **1a** (MD-805) with thrombin from the X-ray study (top) and from the end of the Monte Carlo simulation (bottom). Only atoms of the protein within 6 Å of any atom of the inhibitor are shown in Figures 4−6.

hydrophobic P-pocket, and the quanidinium-containing side chain is in the arginine recognition site forming a salt bridge with Asp-H189. (The chymotrypsin numbering is used here, as in the Protein Data Bank entry, 1DWC.[7b]) The piperidine is more deeply buried in the protein and contacts the bottom face of the quinoline, while the top face of the quinoline is solvent-exposed. The oxygens of the sulfonamide group are also solvent-exposed, while the sulfonamide NH forms a hydrogen bond with the oxygen of Gly-H216. The P-pocket is formed from the side chains of His-H57, Trp-H60D, Tyr-H60A, and Leu-H99. A space-filling representation reveals a tight steric match between the P- and D-pockets and their contents, the piperidine and quinoline rings. The carboxylate group attached to the piperidine ring in **1a** forms a hydrogen bond with the hydroxyl group of Ser-H195 and with two water molecules in the crystal structure. It does not appear to be particularly well-accommodated unless there are more hydrogen bonds with water molecules and it is also close to the carboxylate group of Glu-H192; two of the O−O distances are below 6 Å. In contrast, the other group that is modified for **1a**−**g**, the piperidine C4 methyl group, snuggly fills the end of the hydrophobic P-pocket for the equatorial orientation in **1a**.

Insights on the variation in the observed binding affinities for **1a**−**g** emerge from the structures of the complexes obtained from the Monte Carlo simulations. Representative stereoplots of the inhibitors with any surrounding atoms of the protein within 6 Å of any atom in the inhibitor are shown in Figures 4−6. The illustrations are from the last of the 3×10^6 configurations in the averaging runs. The backbone of the protein is aligned in each case to reveal readily the structural changes in the vicinity of the piperidine ring. All water molecules have been removed for clarity; from the MC simulation of the **1a** complex, there are ca. 10 water molecules in the pocket between the piperidine carboxylate group, Glu-H192, His-H57, and Lys-H60F. The X-ray structure[7b] and the structure from the MC simulation of the **1a** complex are shown in Figure 4. Except for the added hydrogen atoms in the MC structure, there is little difference in the two structures. The piperidine C4-methyl group is in close contact with a face of the aryl ring of Tyr-H60A, while the C2-carboxylate is within 6 Å of the side-chain carboxylate of Glu-H192. One difference is that the hydrogen bond between the C2-carboxylate and the side chain of Ser-H195 in the X-ray structure becomes water-bridged in the MC calculation.

In viewing the structure for the **1b** complex in Figure 5 two key contributors to the 100-fold reduction in binding relative to **1a** are evident. The equatorial placement of the C2-carboxylate forces it to be closer to the side chain of Glu-H192, an electrostatically unfavorable situation. From the MC structures, the shortest distances between the carboxylate oxygens decline from 4.6 and 5.0 Å for **1a** to 4.3 and 4.1 Å for **1b**. Mutation of Glu-H192 to, for example, a glutamine residue can be predicted to be favorable for binding **1b**. The

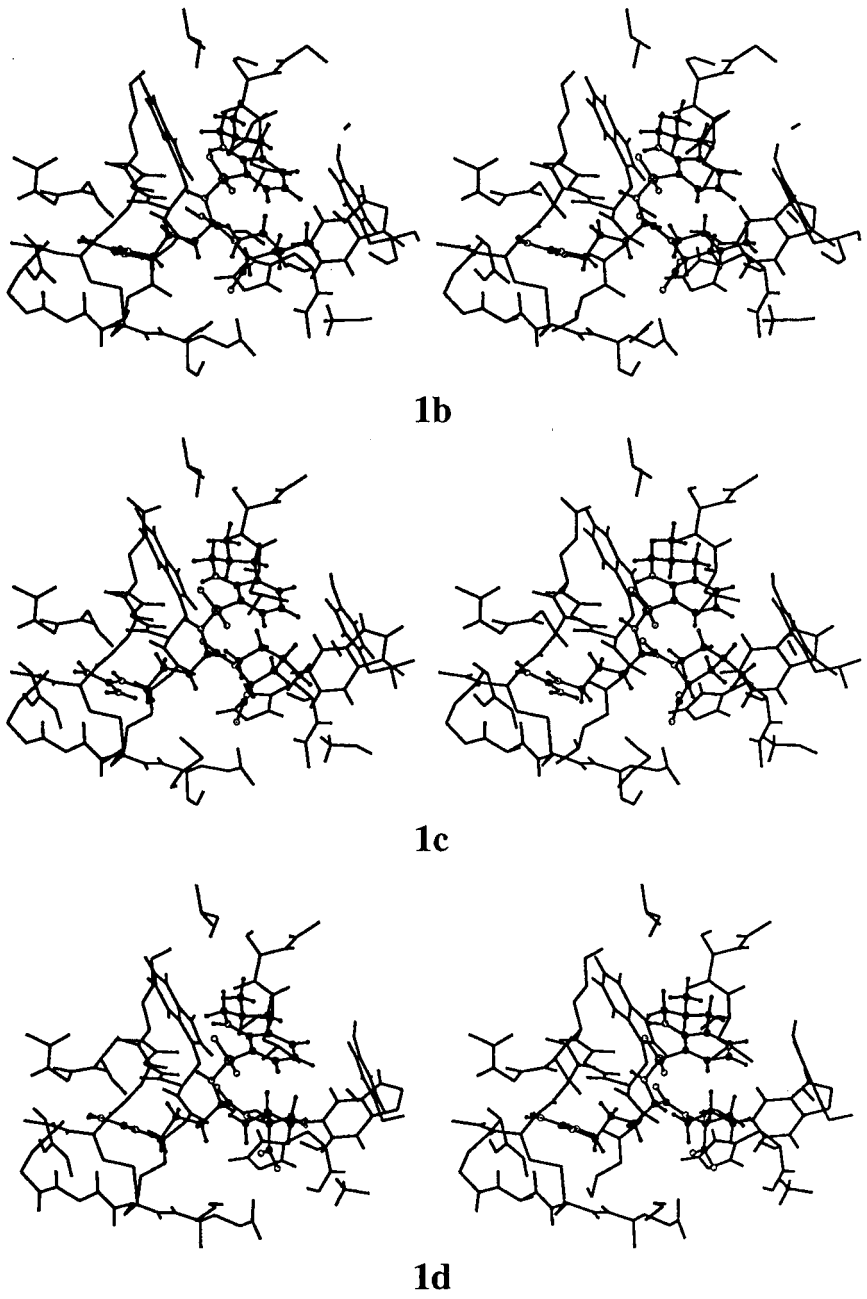

Figure 5. Stereoplots of the binding site for the complexes of inhibitors **1b**, **1c**, and **1d** with thrombin from the end of the Monte Carlo simulations.

equatorial orientation of the C2-carboxylate also brings it closer to the trimethylene spacer in **1b**, which shields it more from hydration; the number of water molecules hydrogen-bonded to the carboxylate group declines from five or six for the axial orientation to three or four for the equatorial form. The position of the piperidine C4-methyl group is unchanged for the complexes of **1a** and **1b**.

The structure for the complex of **1c** in Figure 5 shows the same problems as for the equatorial C2-carboxylate in **1b** with additional trouble for the axial C4-methyl group, which leads to a further 147-fold reduction in binding. The methyl group loses the contact with the π-face of Tyr-H60A, descends from the P-pocket and likely interferes with the hydration of His-H57 and Lys-H60F. The piperidine ring appears to try to compensate by tilting up more toward the quinoline ring to retain the methyl group in the P-pocket as much as possible.

For the complex of **1d** in Figure 5, loss of the C4-methyl group simply creates a hole between the piperidine ring and Tyr-H60A. Full display of the illustrated MC configuration shows no water molecule filling the hole; it is simply void space. The price for the emptiness is an observed 16-fold reduction in binding relative to **1a** (Figure 1). The methyl group is restored for the complex of **1e** in Figure 6, which is now missing the C2-carboxylate group. This deletion has almost no effect on the binding relative to **1a**. The MC structure does reveal a normal degree of hydration for the axial C2-carboxylate, so destabilization from the proximal side chain of Glu-H192 is apparently not significant. In going from **1e** to **1f** in Figure 6, the C4-methyl group has again been removed with an accompanying 4-fold reduction in binding. The gap between the piperidine ring and Tyr-H60A reemerges, though it is not as great as for **1d**. For the **1f** complex,

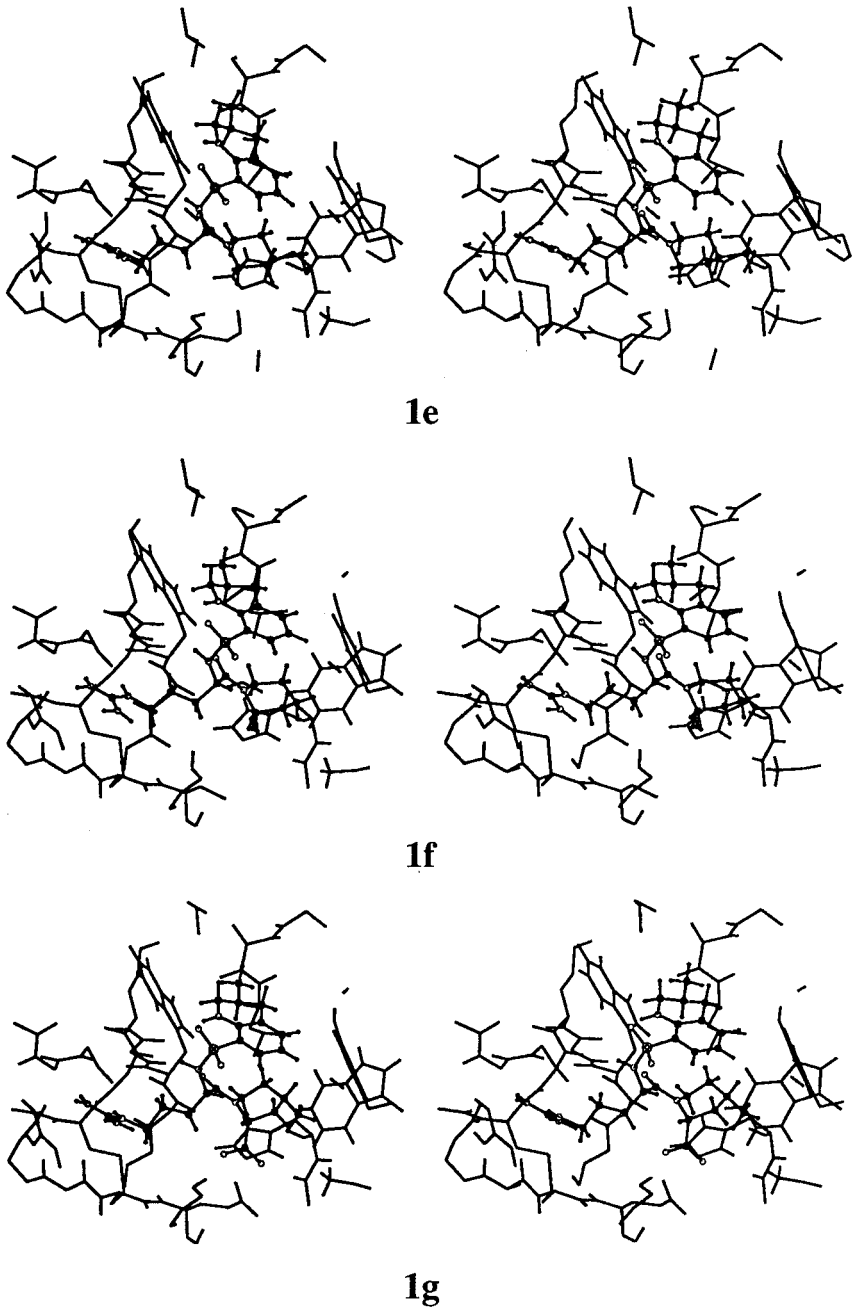

Figure 6. Stereoplots of the binding site for the complexes of inhibitors **1e**, **1f**, and **1g** with thrombin from the end of the Monte Carlo simulations.

the inhibitor has translated about 1 Å toward the P-pocket to help fill the void. This modification is likely inhibited for **1d** by positioning the C2-carboxylate for optimal hydration.

In Figure 6, **1g** is seen to reintroduce the problems with an axial C4-methyl group as in **1c**. However, it is not clear why the observed ratio of inhibition constants for **1c**/**1b** (147) is greater than for **1g**/**1a** (12).[7h] It is possible that the difference has an entropic component; simultaneous placement of the unfavorable equatorial carboxylate and axial methyl group may cause restricted motional freedom for inhibitor **1c**. It is notable that the tilting up of the piperidine ring in **1g** moves the axial carboxylate close enough to Ser195 that a direct hydrogen bond is observed, as in the X-ray structure for **1a** (bottom of the stereoplots in Figures 4 and 6). For **1a**−**d**, water molecules bridge the piperidine carboxylate and the side chains of both Ser-H195 and Glu-H192. A water molecule is also found in the MC structures to bridge the axial carboxylate of **1a**, **1d**, **1g**, and Lys-H60F.

Finally, turning to the arginine recognition pocket, examination of crystal structures that contain arginine−carboxylate interactions has revealed 10 distinct hydrogen-bonding arrangements.[23] In the crystal structure for the **1a** complex, type 6 geometry (single NH1−twin O) is observed for the guanidinium group and the carboxylate of Asp-H189. The MC simulations also revealed that type 6 hydrogen-bonding arrangements are dominant for all the inhibitors (leftmost region in Figures 4−6). An additional hydrogen bond is found in this pocket between the oxygen of Gly-H219 and the guanidinium unit. There are also two or three water molecules in the pocket that form hydrogen bonds with the guanidinium hydrogens.

Conclusion

Correlative linear response approaches in conjunction with Monte Carlo simulations were used to compute free energies of binding for seven inhibitors of thrombin. Equation 2 in a two-parameter form with $\alpha = \beta$ yielded more accurate results than the alternative eq 1 also with two parameters. The rms deviation of 1.0 kcal/mol and average unsigned error of 0.8 kcal/mol for the observed range of 6.0 kcal/mol in binding affinities show that the procedure has useful predictive value. The nature of the structural differences between the inhibitors involving epimerizations and complete removal of a charged group would be a difficult challenge for free-energy perturbation procedures. Moreover, it was interesting to observe that the correlations obtained after only 3×10^6 configurations of averaging for the unbound inhibitors also gave reasonable relative and absolute ΔG_b values. The rms deviations were only a few tenths of a kcal/mol higher than from the results of the longest simulations. The structural findings from the Monte Carlo simulations also provided insights into the variations in observed binding affinities. Equatorial placement of the carboxylate group at C2 in the piperidine ring of the inhibitors causes electrostatic destabilization with the side chain of Glu-H192, while axial disposition of the C4-methyl group reduces favorable hydrophobic interactions in the P-pocket of the enzyme. Though further work is needed with larger sets of ligands and multiple proteins to evaluate the methodology fully, the illustrated energetic and structural characterization of enzyme inhibition indicates that the Monte Carlo/linear response approach has much promise as a tool for ligand design.

Acknowledgment. Gratitude is expressed to Drs. Dongchul Lim and Julian Tirado-Rives for computational assistance and to the National Institutes of Health, the Office of Naval Research, and Bayer Pharmaceutical Division for support of this research. Receipt of a preprint of ref 2e from Prof. Åqvist is also appreciated.

Supporting Information Available: Structural results for the model compounds used in the potential function development, force field parameters and thermodynamic results for liquid divinyl sulfone, and nonbonded parameters for inhibitors **1a**−**g** (14 pages). Ordering information is given on any current masthead page.

References

(1) For reviews, see: (a) Kollman, P. A. Free Energy Calculations: Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* **1993**, *93*, 2395−2417. (b) van Gunsteren, W. F. Methods for Calculation of Free Energies and Binding Constants: Successes and Problems. In *Computer Simulation of Biomolecular Systems*; van Gunsteren, W. F., Weiner, P. K., Eds.; ESCOM: Leiden, 1989; pp 27−59.

(2) (a) 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. (b) Åqvist, J.; Mowbray, S. M. Sugar Recognition by a Glucose/Galactose Receptor. *J. Biol. Chem.* **1995**, *270*, 9978−9981. (c) Hansson, T.; Åqvist, J. Estimation of Binding Free Energies for HIV Proteinase Inhibitors by Molecular Dynamics Simulations. *Protein Eng.* **1995**, *8*, 1137−1144. (d) Åqvist, J.; Hansson, T. *J. Phys. Chem.* **1996**, *100*, 9512−9521. (e) Hansson, T.; Marelius, J.; Åqvist, J. Ligand Binding Affinity Prediction by Linear Interaction Energy Methods. Submitted for publication.

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

(4) Carlson, H. A.; Jorgensen, W. L. An Extended Linear Response Method for Determining Free Energies of Hydration. *J. Phys. Chem.* **1995**, *99*, 10667−10673.

(5) 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.

(6) (a) Fenton, J. W. Regulation of thrombin generation and functions. *Semin. Thromb. Hemostasis* **1988**, *14*, 234−240. (b) Vu, T. K. H.; 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. (c) Liu, L. W.; Vu, T. K. H.; Esmon, C. T.; Coughlin, S. R. The region of the thrombin receptor resembling hirudin bonds to thrombin and alters enzyme specificity. *J. Biol. Chem.* **1991**, *266*, 16977−16980. (d) Doolittle, R. F. Fibrinogen to fibrin transformation. *Annu. Rev. Biochem.* **1984**, *53*, 195−229. (e) Blomback, B. Fibrinogen to fibrin transformation. In *Blood Clotting Enzymology*; Seegers, W. H., Eds.; Academic Press: New York, 1967; pp 143−215. (f) Chen, R.; Doolittle, R. F. Isolation, characterization, and location of a donor-acceptor unit from cross-linked fibrin. *Proc. Natl. Acad. Sci. U.S.A.* **1970**, *66*, 472−479.

(7) (a) Bode, W.; Mayr, I.; Baumann, U.; Huber, R.; Stone, S. R.; Hofsteenge, J. The refined 1.9 Å crystal structure of human α-thrombin: interaction with D-Phe-Pro-Arg chloromethylketone and significance of the Tyr-Pro-Pro-Trp insertion segment. *EMBO J.* **1989**, *8*, 3467−3475. (b) Banner, D. W.; Hadvary, P. Crystallographic analysis at 3.0-Å resolution of the binding of human thrombin of four active site-directed inhibitors. *J. Biol. Chem.* **1991**, *266*, 20085−20093. (c) Hilpert, K.; Ackermann, J.; Banner, D. W.; Gast, A.; Gubernator, K.; Hadvary, P.; L. Labler, Muller, K.; Schmid, G.; Tschopp, T. B.; van de Waterbeemd, H. Design and Synthesis of Potent and Highly Selective Thrombin Inhibitors. *J. Med. Chem.* **1994**, *37*, 3889−3901. (d) Bode, W. T.; Turk, D.; Sturzebecher, J. Geometry of binding of the benzamidine- and arginine-based inhibitors NAPAP and MQPA to human α-thrombin. X-ray crystallographic determination of the NAPAP-trypsin complex and modeling of NAPAP-thrombin and MQPA-thrombin. *Eur. J. Biochem.* **1990**, *193*, 175−182. (e) Brandstetter, H.; Turk, D.; Hoeffken, W.; Grosse, D.; Sturzebecher, J.; Martin, P. D.; Edwards, B. F. P.; Bode, W. Refined 2.3Å X-ray Crystal Structure of Bovine Thrombin Complexes Formed with the Benzamidine and Arginine-based Thrombin Inhibitors NAPAP, 4-TAPAP and MQPA. *J. Mol. Biol.* **1992**, *226*, 1085−1099. (f) Mathews, I. I.; Tulinsky, A. Active-Site Mimetic Inhibition of Thrombin. *Acta Crystallogr.* **1995**, *D51*, 550−559. (g) Zdanov, A.; Wu, S.; DiMaio, J.; Konishi, Y.; Li, Y.; Wu, X.; Edwards, B. F. P.; Martin, P. D.; Cygler, M. Crystal structure of the complex of human α-thrombin and nonhydrolyzable bifunctional inhibitors, hirutonin-2 and hirutonin-6. *Proteins: Struct., Funct., Genet.* **1993**, *17*, 252−65. (h) Kikumoto, R.; Tamao, Y.; Teauka, T.; Tonomura, S.; Hara, H.; Ninomiya, K. Selective Inhibition of thrombin by (2R,4R)-4-methyl-1-(N2-[(3-methyl-1,2,3,4-tetrahydro-8-quinolinyl)suphonyl)-L-arginyl)]-2-piperidine-carboxylic acid. *Biochemistry* **1984**, *23*, 85−90. (i) Bauer, M.; Brandsetter, H.; Turk, D.; Sturzebecher, J.; Bode, W. Crystallographic Determination of Thrombin Complexes With Small Synthetic Inhibitors as a Starting Point for the Receptor-Based Design of Antithrombotics. *Semin. Thromb. Hemostasis* **1993**, *19*, 352−360.

(8) (a) Obst, U.; Gramlich, V.; Diederich, F.; Weber, L.; Banner, D. W. Design of Novel, Nonpeptidic Thrombin Inhibitors and Structure of a Thrombin-Inhibitor Complex. *Angew. Chem., Int. Ed. Engl.* **1995**, *34*, 1739−1742. (b) Grootenhuis, P. D. J.; van Galen, P. J. M. Correlation of Binding Affinities with Nonbonded Interaction Energies of Thrombin-Inhibitor Complexes. *Acta Crystallogr.* **1995**, *D51*, 560−566. (c) Tsuda, Y.; Szewczuk, Z.; Wang, J.; Yue, S. Y.; Purisima, E.; Konishi, Y. Interactions of Hirudin-Based Inhibitor with Thrombin: Critical Role of the Ile^{H59} Side Chain of the Inhibitor. *Biochemistry* **1995**, *34*, 8708−8714. (d) Wang, J.; Szewczuk, Z.; Yue, S.-Y.; Tsuda, Y.; Konishi, Y.; Purisima, E. O. Calculation of Relative Binding Free Energies and Configurational Entropies: A Structural and Thermodynamic Analysis of the Nature of Non-polar binding of Thrombin Inhibitors Based on $Hirudin^{55-65}$. *J. Mol. Biol.* **1995**, *253*, 473−492. (e) Lau, W. F.; Tabernero, L.; Scak, J. S.; Iwanowicz, E. J. Molecular Modeling Studies of Novel Retro-binding Tripeptide Active-site Inhibitors of Thrombin. *Bioorg. Med. Chem.* **1995**, *3*, 1039. (f) 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) (a) Jorgensen, W. L.; Tirado-Rives, J. The OPLS Potential Functions for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin. *J. Am. Chem. Soc.* **1988**, *110*, 1657−1666. (b) Jorgensen, W. L.; Severance, D. L. Aromatic-Aromatic Interactions: Free Energy Profiles for the Benzene Dimer in Water, Chloroform and Liquid Benzene. *J. Am. Chem. Soc.* **1990**, *112*, 4768−4774. (c) Jorgensen, W. L.; Nguyen, T. B. Monte Carlo simulations of the hydration of substituted benzenes with OPLS potential functions *J. Comput. Chem.* **1993**, *14*, 195−205. (d) Kaminski, G.; Duffy, E. M.; Matsui, T.; Jorgensen, W. L. Free Energies of Hydration and Pure Liquid

Properties of Hydrocarbons from the OPLS All-Atom Model. *J. Phys. Chem.* **1994**, *98*, 13077−13082. (e) 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.* **1997**, *119*, 11225−11236.

(10) Jorgensen, W. L. *BOSS*, Version 3.6; Yale University: New Haven, CT, 1995.

(11) See the Supporting Information for a more detailed presentation on the determination of these nonbonded parameters and results for liquid divinyl sulfone.

(12) Frisch, M. L.; Trucks, G. W.; Head-Gordon, M.; Gill, P. M. W.; Wong, M. W.; Foresman, J. B.; Johnson, B. G.; Schlegel, H. B.; Robb, M. A.; Replogle, E. S.; Gomperts, R.; Andres, J. L.; Raghavachari, K.; Binkley, J. S.; Gonzalez, C.; Martin, R. L.; Fox, D. J.; Defrees, D. J.; Baker, J.; Stewart, J. J. P.; Pople, J. A. *Gaussian94*, Revision A; Gaussian, Inc.: Pittsburgh, PA, 1994.

(13) Rossi, K. A.; Merz, K. M.; Smith, G. M.; Baldwin, J. J. A Test of The Free Energy Perturbation Method on Human Carbonic Anhydrase II Inhibitors. *J. Med. Chem.* **1995**, *38*, 2061−2069.

(14) See the Supporting Information for a complete listing of σ, ϵ, and q parameters for inhibitors **1a**−**g**.

(15) 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.

(16) Jorgensen, W. L. *MCPRO*, Version 1.3; Yale University: New Haven, CT, 1995.

(17) (a) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. A. A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins. *J. Am. Chem. Soc.* **1984**, *106*, 765−784. (b) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. An All Atom Force Field for Simulations of Proteins and Nucleic Acids. *J. Comput. Chem.* **1986**, *7*, 230−252.

(18) Pearlman, D. A.; Case, D. A.; Caldwell, J. C.; Seibel, G. L.; Singh, C.; Weiner, P.; Kollman, P. A. (1991), AMBER 4.0, University of California, San Francisco.

(19) Skell, J. M.; Pearlman, R. S. SAVOL2, University of Texas, Austin.

(20) (a) Rashin, A. A. Buried Surface Area, Conformational Entropy, and Protein Stability. *Biopolymers* **1984**, *23*, 1605−1620. (b) Rashin, A. A.; Iofin, M.; Honig, B. Internal cavities and buried waters in globular proteins. *Biochemistry* **1986**, *25*, 3619−3625.

(21) For a full listing of atomic radii values used to determine the SASAs, see the Supplementary Material.

(22) McDonald, N. A.; Carlson, H. A.; Jorgensen, W. L. *J. Phys. Org. Chem.*, in press.

(23) Singh, J.; Thornton, J. M.; Snarey, M.; Campbell, S. F. The geometries of interacting arginine-carboxyls in proteins. *FEBS Lett.* **1987**, *224*, 161−171.

JM960684E