

MM-GB/SA Rescoring of Docking Poses in Structure-Based Lead Optimization

Cristiano R. W. Guimarães^{*,†} and Mario CardozoDepartment of Molecular Structure, Amgen Inc., 1120 Veterans Boulevard,
South San Francisco, California 94080

Received January 7, 2008

The critical issues in docking include the prediction of the correct binding pose and the accurate estimation of the corresponding binding affinity. Different docking methodologies have all been successful in reproducing the crystallographic binding modes but struggle when predicting the corresponding binding affinities. The aim of this work is to evaluate the performance of the MM-GB/SA rescoring of docking poses in structure-based lead optimization. To accomplish that, a diverse set of pharmaceutically relevant targets, including CDK2, FactorXa, Thrombin, and HIV-RT were selected. The correlation between the MM-GB/SA results and experimental data in all cases is remarkable. It even qualifies this approach as a more attractive alternative for rank-ordering than the Free Energy Perturbation and Thermodynamic Integration methodologies because, while as accurate, it can handle more structurally dissimilar ligands and provides results at a fraction of the computational cost. On the technical side, the benefit of performing a conformational analysis and having an ensemble of conformers to represent each ligand in the unbound state during the MM-GB/SA rescoring procedure was investigated. In addition, the estimation of conformational entropy penalties for the ligands upon binding, computed from the Boltzmann distribution in water, was evaluated and compared to a commonly used approach employed by many docking scoring functions.

INTRODUCTION

The computational methodologies to understand structural and energetic relationships to binding vary in speed and accuracy. The molecular dynamics (MD) and Monte Carlo (MC) simulations coupled with free-energy perturbation (FEP) or thermodynamic integration (TI) calculations are the most rigorous computational approaches currently used to estimate relative binding affinities.^{1–6} Although these methods have provided impressive results for several protein–ligand systems, they are computationally intensive and have generally been applied to study a small number of ligands in a congeneric series.

The estimation of binding affinities may be accomplished with a more approximate method such as the linear response (LR) theory, originally introduced by Åqvist.⁷ The method was generalized by Jorgensen and co-workers to include not only the interaction energies for the ligands in the solvated protein–ligand complexes and in the unbound state but also other sensible descriptors. Counts of hydrogen bonds, number of rotatable bonds for the ligands, and separate components of SASA (solvent accessible surface area), e.g., the hydrophobic, hydrophilic, and aromatic surface areas, are also considered in the multivariate fitting approach.^{8–13} The LR theory is significantly less demanding computationally than the FEP and TI alternatives because no intermediate state is required to compute binding affinities, but still considerably slow due to the sampling of the phase space.

Small-molecule docking is designed to orient and score a large number of molecules for complementarity against a

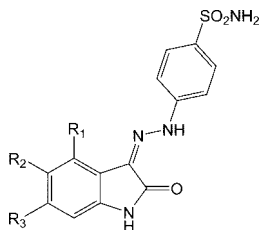
macromolecular binding site in a short period of time.^{14–20}

The critical issues in docking include the prediction of the correct binding pose and the accurate estimation of the corresponding binding affinity. Despite the enormous size of the conformational space for the ligands, different docking methodologies, e.g., force-field-based, empirical, and knowledge-based, have all been successful in reproducing the crystallographic binding modes.^{21–25} However, they still need improvement when it comes to predicting binding affinities,^{26–29} particularly the scoring function terms responsible for the estimation of the ligand desolvation, intramolecular, and conformational entropy penalties upon binding. This has been discussed recently, where the change from the unbound to the bound conformation and the loss of conformational degrees of freedom for the ligand have been collectively termed “conformer focusing”.³⁰

Since the docking algorithms provide good-quality binding poses, an energy function with a more physically reasonable description of binding contributions can be employed to rescore the docking results. MM-PB/SA calculations, pioneered by Kollman and co-workers, use a combination of molecular mechanics and continuum solvation to compute average binding energies for configurations extracted from MD simulations of the unbound and bound states.^{31a} The encouraging results obtained with this methodology, despite limitations,^{31b} have inspired several authors to use molecular-mechanics-based scoring functions with GB/SA³² as the implicit solvent model in the rescoring process. When compared to docking scoring functions, the MM-GB/SA procedure provided more accurate docking poses, improved enrichment in the virtual screening of databases, and superior correlation between calculated binding affinities and experi-

^{*} Corresponding author phone: (860) 686-2915; e-mail: cristiano.guimaraes@pfizer.com.

[†] Current address: Pfizer Global Research & Development, 558 Eastern Point Rd, B220/351A, Groton, CT 06340.

Table 1. Enzymatic Activities of Selected CDK2 Inhibitors (Core 1)


R ₁	R ₂	R ₃	IC ₅₀ (nM) ^a
H	H	H	120
CH ₂ CH ₃	H	H	7.9
CH(CH ₃) ₂	H	H	2.5
CH ₂ CH(CH ₃) ₂	H	H	1.2
OCH(CH ₃) ₂	H	H	3.4
OPh	H	H	13
NO ₂	H	H	2400
H	F	H	34
H	Cl	H	43
H	Br	H	60
H	CH ₃	H	46
H	OH	H	10
H	OCH ₃	H	12
H	NH ₂	H	74
H	SO ₂ CH ₃	H	16
H	SO ₂ NH ₂	H	43
H	CONH ₂	H	4.5
H	CON(CH ₃) ₂	H	17
H	H	Br	43
H	H	CH ₂ CH ₃	21
H	H	CH(CH ₃) ₂	75
Cl	CH ₃	H	13
Cl	OCH ₃	H	54

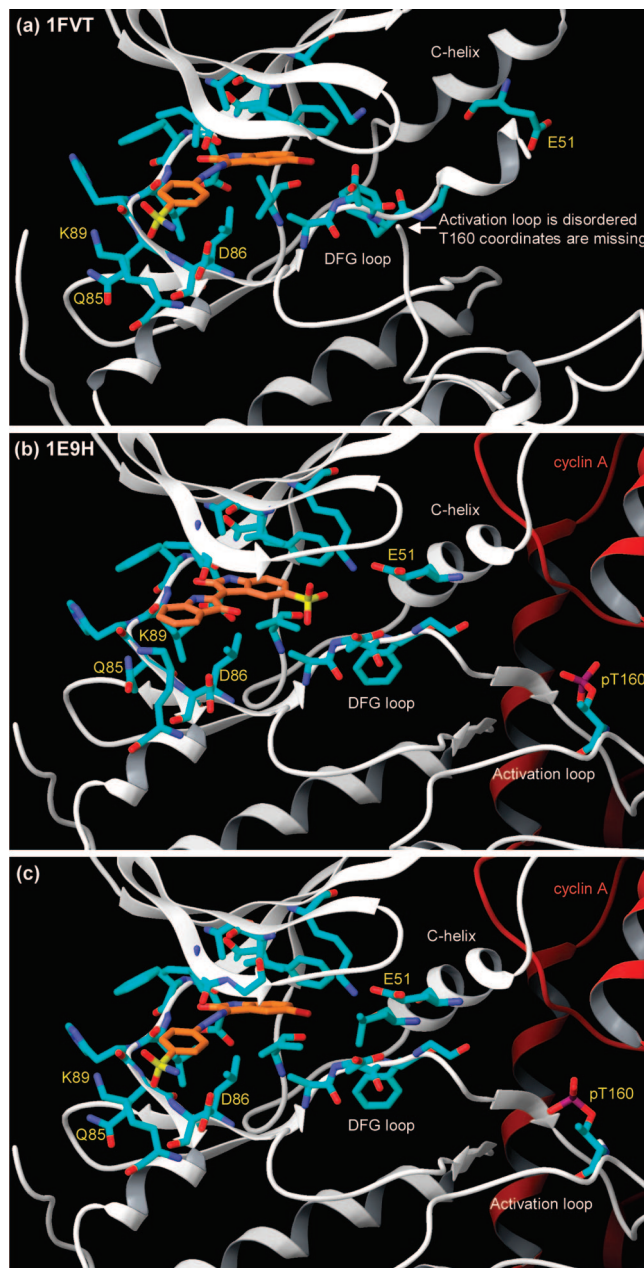
^a Ref 37.

mental data in the lead optimization of four sets of congeneric kinase inhibitors.³³

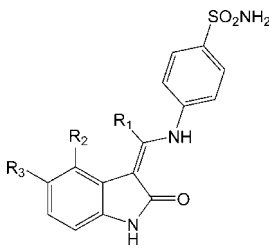
The aim of this work is to investigate the performance of MM-GB/SA rescoring in structure-based lead optimization for a more diverse set of pharmaceutically relevant targets, including CDK2, FactorXa, Thrombin, and HIV-RT. Another goal is to study the benefit of doing a conformational analysis and having an ensemble of conformers to represent each ligand in the unbound state; most of the MM-GB/SA methodologies in the literature relax the bound state conformation for the ligand to its nearest energy minimum in water when performing the rescoring. Finally, the estimation of conformational entropy penalties for the ligands upon binding, computed from the Boltzmann distribution in water, is evaluated and compared to a commonly used approach employed by many docking scoring functions, which penalizes each rotatable bond by 0.65 kcal/mol.^{30,34}

METHODS

Docking. The crystal structures for CDK2, Factor Xa, Thrombin, and HIV-RT complexed with the inhibitors Indirubin-5-sulfonate (PDB ID: 1E9H), ZK-807834 (PDB ID: 1FJS), 4-TAPAP (PDB ID: 1ETT), and UC-781 (PDB ID: 1RT4), respectively, were employed in the docking calculations performed by the Glide 4.0 XP scoring function.^{24,35} Although there is a complex between CDK2 and an inhibitor belonging to the selected congeneric series (Table 1) available in the Protein Data Bank (PDB ID:

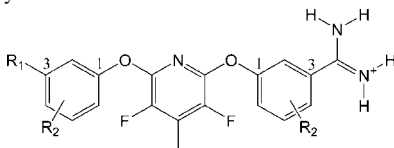
**Figure 1.** (a) 1FVT and (b) 1E9H crystal structures between CDK2 and inhibitors. (c) Modified 1E9H structure complexed to the 1FVT inhibitor used in the docking calculations.

1FVT), this structure is not consistent with the conditions for the biological assay; it is not complexed to cyclin A and phosphorylated on Thr160. Figures 1a and 1b illustrate the differences between the 1FVT and 1E9H CDK2 structures. The absence of cyclin A and Thr160 phosphorylation in 1FVT cause repositioning of the C-helix and a major conformational change on the activation loop that render CDK2 inactive. The inhibitor Indirubin-5-sulfonate in 1E9H, however, does not have the sulfonamide group that is sandwiched between the D86 and K89 residues, like in the 1FVT complex. The absence of this interaction causes a conformational change on Q85 and K89. Since the series of selected CDK2 inhibitors (Tables 1 and 2) displays this interaction with the enzyme, the Q85 and K89 side-chain conformations in 1E9H were changed to reflect the conformation adopted in the 1FVT crystal structure. After that, the inhibitor in 1FVT was optimally overlaid in the modified

Table 2. Enzymatic Activities of Selected CDK2 Inhibitors (Core 2)


R ₁	R ₂	R ₃	IC ₅₀ (nM) ^a
H	H	H	690
CH ₃	H	H	360
CH ₃	H	Cl	22
H	CH ₂ OH	H	54
H	H	N(CH ₃) ₂	310

^a Ref 37.

Table 3. Enzymatic Activities of Selected Factor Xa Inhibitors


R ₁	R ₂	R ₃	K _i (nM) ^a
CONH ₂	H	H	280
CONHMe	H	H	1200
CONMe ₂	H	H	80
COMe	H	H	1400
NO ₂	H	H	2500
NH ₂	H	H	3300
NMe ₂	H	H	160
NHEt	H	H	530
OMe	H	H	1350
OCF ₃	H	H	1800
F	H	H	3200
Cl	H	H	1700
OH	H	H	5000
CF ₃	H	H	1600
CONMe ₂	5-OMe	H	140
NMe ₂	2-Me	H	320
CONMe ₂	H	6-NH ₂	14
CONMe ₂	H	6-OH	1.8
NMe ₂	H	6-Me	1200
NMe ₂	H	6-NH ₂	64
NMe ₂	H	6-OH	3
NMe ₂	H	6-OMe	1400

^a Ref 38.

1E9H binding site and the generated complex used as a starting point in the CDK2 docking calculations (Figure 1c). All complexes were submitted to a series of restrained, partial minimizations using the OPLS_2005 force field.³⁶ Before the docking calculations, the selected congeneric series of CDK2,³⁷ Factor Xa,³⁸ and Thrombin³⁹ inhibitors shown in Tables 13 and Figure 2 and the non-nucleoside inhibitors of HIV-RT⁴⁰ illustrated in Figure 3 were submitted to a pre-energy minimization using the OPLS_2005 force field and the GB/SA method as the implicit water model.³² In order to accommodate for the fact that the protein structure used for docking will not in general be optimized to fit a particular ligand, the van der Waals radii for nonpolar protein atoms were scaled by a factor of 0.8, while those for the ligands were not scaled.

MM-GB/SA Rescoring. In our implementation of the MM-GB/SA rescoring (Figure 4), a conformational search for the inhibitors in the unbound state and energy minimization for the complexes using OPLS_2005 and GB/SA within MacroModel⁴¹ were performed. Applying energy minimization for the complexes rather than MD simulations greatly increases computational efficiency and provides a method with a time scale compatible with synthetic chemistry—biological test cycles. On the other hand, lack of sampling could in theory pose a significant limitation on the method since the protein would not be able to relax to accommodate different scaffolds after docking. This problem should be minimized when scoring a congeneric series. In addition, a recent study suggests that a single, relaxed structure for each complex provides superior results when compared to the standard averaging over MD trajectories.⁴² A possible explanation for this is the introduction of noise in the scoring as each complex could be visiting different regions of the phase space due to short trajectories.

The Monte Carlo multiple minimum (MCM) method implemented in MacroModel⁴¹ was used to perform the conformational analysis in the unbound state. This method is highly efficient in performing global searching, exploring close as well as distant areas of the potential energy surface. To ensure that the stochastic search was exhaustive and approached convergence, the extended protocol for the torsion sampling⁴¹ and energy minimization to a low gradient norm were employed. All conformers within 5.0 kcal/mol from the lowest-energy conformer were retained. A root-mean-square deviation (rmsd) value of 0.3 Å for heavy atoms and hydrogens connected to heteroatoms was used to obtain unique conformations. Assuming a Boltzmann distribution, the probabilities for each conformer (P_i) were calculated and the Boltzmann-averaged intramolecular energy and solvation free energy in the unbound state for every compound obtained. The conformational entropies (S_{conf}) were computed from the probabilities using eq 1, where k_B is the Boltzmann constant.

$$S_{\text{conf}} = -k_B \sum_{i=1}^n P_i \ln P_i \quad (1)$$

To better account for the protein flexibility, the best pose for each inhibitor was energy-minimized in the bound state. The conjugate gradient minimization scheme that uses the Polak–Ribiere first derivative method (PRCG), considered the best general method for energy minimization,⁴¹ was employed with a very tight convergence threshold. In the energy minimization, no constraints were applied to residues within 5 Å from the center of the system. A second shell of 3 Å around the first shell was defined, and constraints of 50 kcal/mol·Å² were applied to the residues therein. The remaining residues were held fixed. This was done with the purpose of reducing the aforementioned noise in the scoring; each complex could be driven to different local minima in a fully flexible energy minimization scheme. After the energy minimization step, the protein energy (E_{PTN}) values for all complexes were extracted. This term describes the protein deformation imposed by each ligand. Besides E_{PTN} , the energy-minimized structures for the complexes provided the intramolecular energies and solvation free energies for the ligands in the protein environment and the protein–ligand

intermolecular van der Waals (E_{VDW}) and electrostatic (E_{Elect}) interaction energies. In the bound state, it was assumed that there was only one conformation accessible to each ligand, and its conformational entropy is therefore zero. In this manner, the binding energy (ΔG_{bind}) was calculated as shown in eq 2.

$$\Delta G_{bind} = \Delta E_{intra} + \Delta G_{solv} - T\Delta S_{conf} + E_{VDW} + E_{Elect} + E_{PTN} \quad (2)$$

In eq 2, ΔE_{intra} and ΔG_{solv} are the intramolecular and desolvation penalties for each ligand upon binding. These penalties reflect how their intramolecular energies and

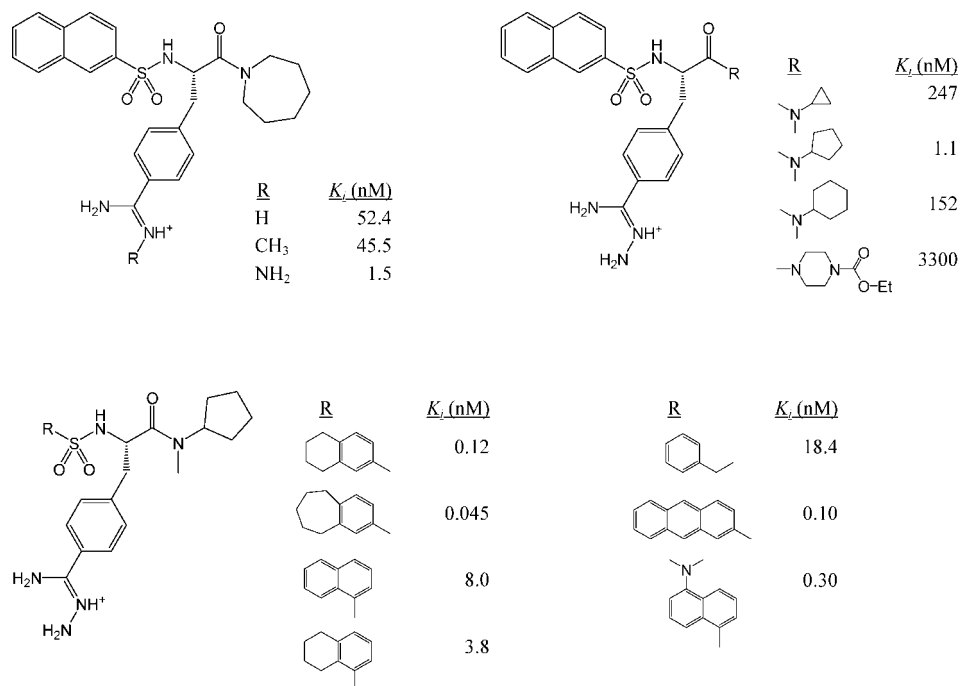


Figure 2. Enzymatic activities of selected Thrombin inhibitors. Experimental data from ref 39.

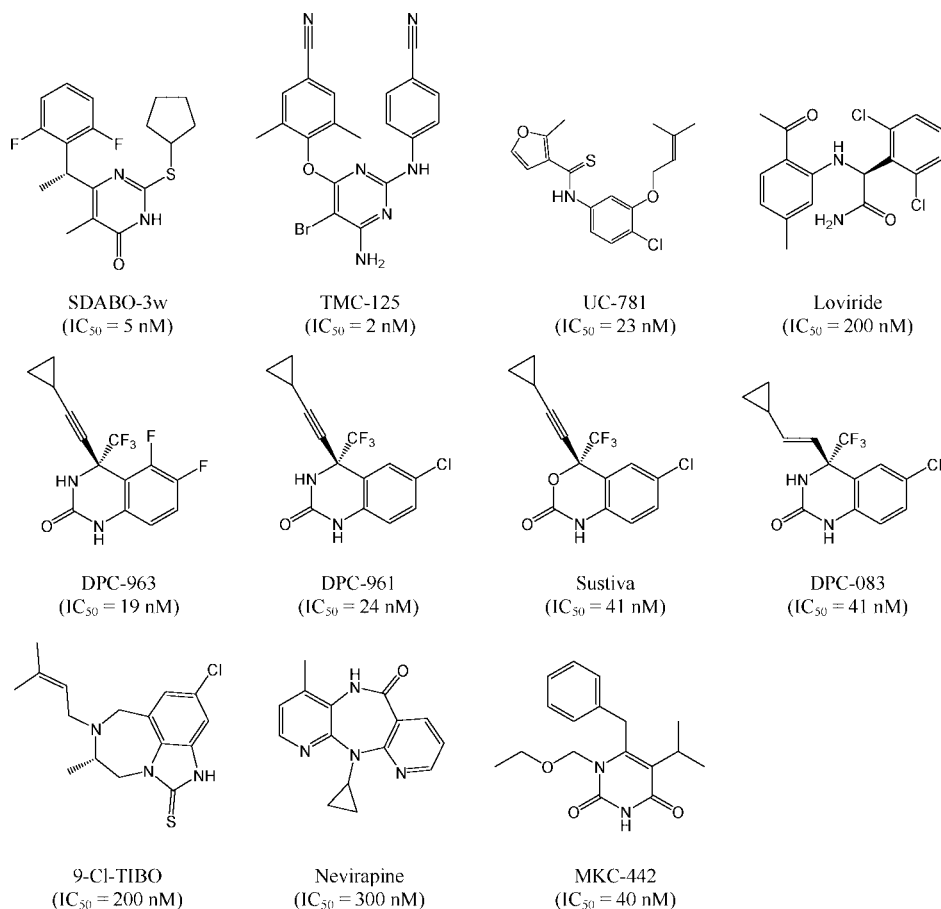


Figure 3. Enzymatic activities of selected HIV-RT inhibitors. Experimental data from ref 40.

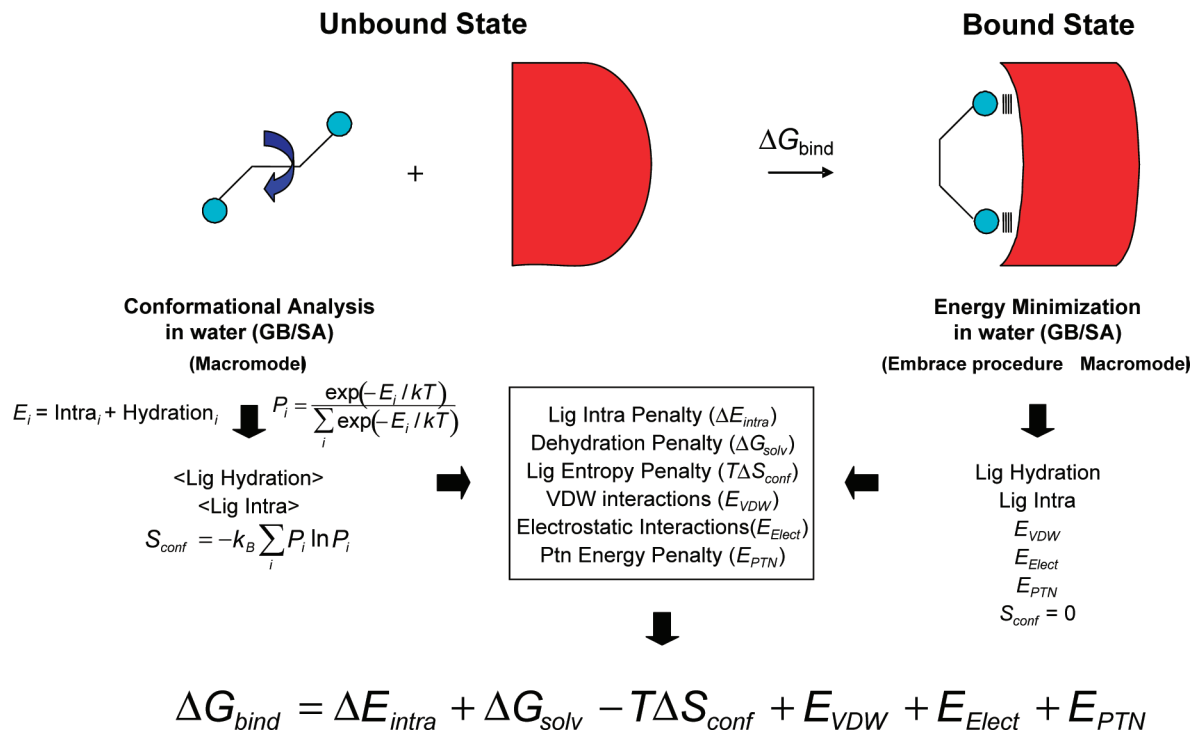


Figure 4. Schematic representation of the MM-GB/SA rescoring procedure.

solvation free energies change upon transfer from the unbound to the bound state. Similarly, $-T\Delta S_{\text{conf}}$ is the ligand conformational entropy penalty, multiplied by the temperature to convert it into free energy. The final ranking was obtained by calculating relative binding energies ($\Delta\Delta G_{\text{bind}}$) using the top-scoring inhibitor of each target as a reference.

Unbound State Representation. A common procedure employed in the literature to represent the unbound state for each ligand is to relax its bound state conformation to the nearest local energy minimum in water.³³ It is obvious that the MM-GB/SA scoring in this case would lack conformer-focusing-type contributions to protein–ligand binding.³⁰ To investigate the benefit of having an ensemble of conformers in the unbound state as an alternative, the ΔE_{intra} and ΔG_{solv} penalties in the binding of ligands to biologically relevant targets were calculated using both the single conformer and ensemble representations. To accomplish that, one hundred complexes were selected from the Protein Data Bank, including 4 Cox-2, 18 CDK2, 13 Factor Xa, 18 HIV-RT, 14 p38, 16 Thrombin, 5 GSK3, 6 HIV protease, and 6 DHFR complexes (Table 4). The correlations between the experimental data for the selected CDK2, Factor Xa, Thrombin, and HIV-RT inhibitors and the MM-GB/SA scoring using the single conformer representation were also investigated to further evaluate both unbound state descriptions. It is important to note that the MM-GB/SA scoring in this case does not include the conformational entropy penalty, as no conformational analysis is performed in the unbound state. Finally, the $-T\Delta S_{\text{conf}}$ penalties for the inhibitors of the data set shown in Table 4 were computed according to eq 1 and compared to the approach that penalizes each rotatable bond by 0.65 kcal/mol.

RESULTS AND DISCUSSION

Unbound State Analysis. Figure 5 compares the ΔG_{solv} and ΔE_{intra} penalties, calculated as defined above, for 100

conformationally diverse inhibitors upon binding to their corresponding biological targets (Table 4) using the single conformer and ensemble representations for the unbound state. While ΔG_{solv} is not dramatically affected by the description of the unbound state (Figure 5a), ΔE_{intra} may be quite different whether one conformer or a collection of conformers is used. It can be underestimated by as much as 20 kcal/mol (Figure 5b). This is obviously more drastic for the cases where the local energy minimum found in water upon relaxation of the bound conformation is not close in energy to the global energy minimum, which dominates the ensemble average. Figure 5c shows that the combined intramolecular and desolvation penalties in the single conformer representation ($\Delta E_{\text{intra}} + \Delta G_{\text{solv}}$) are underestimated with respect to the one calculated using the ensemble representation ($\langle \Delta E_{\text{intra}} \rangle + \langle \Delta G_{\text{solv}} \rangle$) by approximately the energy gap value between the local and global minima in water ($\Delta E_{\text{local} \rightarrow \text{global}}$). For example, if the energy gap in water is ca. 6 kcal/mol, so is the difference between the combined penalties using the ensemble average and using only one conformer. As shown in Figure 5d, a small energy gap occurred in ca. 50% of the cases, with the largest gaps for ligands having more rotatable bonds (Figure 5e). In the bound state, the ligands are deformed as much as possible to maximize their interactions with the protein, and those with an increased degree of flexibility are more easily deformed. In these cases, the probability of finding the global minimum in water when relaxing the bound conformation is very low.

It can also be noted in Figure 5b that some of the calculated ΔE_{intra} and $\langle \Delta E_{\text{intra}} \rangle$ penalties upon binding are very large. This problem has its origin in the force field function. Force fields are fairly accurate at providing relative energies between energy minima but tend to overestimate energy barriers. Since the bound conformations for the ligands are not energy minima in the potential energy surface, the calculated intramolecular penalties can be overestimated.

Table 4. Complexes Selected from the Protein Data Bank

PDB ID	resolution, Å	protein	PDB ID	resolution, Å	protein
1CX2	3.00	COX-2	1TKX	2.85	HIV-RT
2AYL	2.00	COX-2	1TKZ	2.81	HIV-RT
4COX	2.90	COX-2	1VRU	2.40	HIV-RT
1PXX	2.90	COX-2	1A9U	2.50	p38
1DI8	2.20	CDK2	1BL6	2.50	p38
1DM2	2.10	CDK2	1BL7	2.50	p38
1E1V	1.95	CDK2	1M7Q	2.40	p38
1E9H	2.50	CDK2	1OUK	2.50	p38
1FVT	2.20	CDK2	1OUY	2.50	p38
1FVV	2.80	CDK2	1OVE	2.10	p38
1HDV	1.90	CDK2	1W7H	2.21	p38
1JSV	1.96	CDK2	1W83	2.50	p38
1KE5	2.20	CDK2	1WBS	1.80	p38
1KE6	2.00	CDK2	1YQJ	2.00	p38
1KE7	2.00	CDK2	1ZZL	2.00	p38
1OI9	2.10	CDK2	2BAL	2.10	p38
1W0X	2.20	CDK2	2GFS	1.75	p38
2A0C	1.95	CDK2	1D4P	2.00	Thrombin
2B52	1.88	CDK2	1DWK	3.00	Thrombin
2B53	2.00	CDK2	1K22	1.93	Thrombin
2B54	1.85	CDK2	1KTS	2.40	Thrombin
2EXM	1.80	CDK2	1KTT	2.10	Thrombin
1EZQ	2.20	Factor Xa	1MU6	1.99	Thrombin
1F0R	2.10	Factor Xa	1OYT	1.67	Thrombin
1FJS	1.92	Factor Xa	1SB1	1.90	Thrombin
1IOE	2.90	Factor Xa	1T4U	2.00	Thrombin
1IQE	2.90	Factor Xa	1TA2	2.30	Thrombin
1IQH	3.00	Factor Xa	1WAY	2.02	Thrombin
1IQL	2.70	Factor Xa	1YPE	1.81	Thrombin
1IQN	2.60	Factor Xa	1YPK	1.78	Thrombin
1LQD	2.70	Factor Xa	1YPM	1.85	Thrombin
1MQ5	2.10	Factor Xa	1ZGI	2.20	Thrombin
1MQ6	2.10	Factor Xa	1ZGV	2.20	Thrombin
1NFW	2.10	Factor Xa	1Q3W	2.30	GSK3
2BMG	2.70	Factor Xa	1Q4L	2.77	GSK3
1EET	2.73	HIV-RT	1Q5K	1.94	GSK3
1EP4	2.50	HIV-RT	1R0E	2.25	GSK3
1IKW	3.00	HIV-RT	1UV5	2.80	GSK3
1IKY	3.00	HIV-RT	1QBU	1.80	HIV-1 protease
1KLM	2.65	HIV-RT	1OHR	2.10	HIV-1 protease
1LW0	2.80	HIV-RT	1HPV	1.90	HIV-1 protease
1RT1	2.55	HIV-RT	1HPX	2.00	HIV-1 protease
1RT2	2.55	HIV-RT	1TCX	2.30	HIV-1 protease
1RT4	2.90	HIV-RT	3AID	2.50	HIV-1 protease
1RT6	2.90	HIV-RT	1AOE	1.60	DHFR
1RT7	3.00	HIV-RT	1DLR	2.30	DHFR
1S6P	2.90	HIV-RT	1IA1	1.70	DHFR
1S6Q	3.00	HIV-RT	1DG5	2.00	DHFR
1S9E	2.60	HIV-RT	1DG7	1.80	DHFR
1SUQ	3.00	HIV-RT	1KMS	1.09	DHFR

This is particularly true when using the ensemble representation for the unbound state. $\langle \Delta E_{\text{intra}} \rangle$ is generally larger because the high intramolecular energy for the ligand bound conformation is compared to an average intramolecular energy dominated by the global minimum conformation in the unbound state. The corresponding ΔE_{intra} , calculated using the single conformer representation, tends to be smaller because the reference conformation in the unbound state is often a local minimum. Although the values are larger for $\langle \Delta E_{\text{intra}} \rangle$ and do not reflect typical intramolecular penalties for active compounds that would be obtained by more sophisticated energy functions, $\langle \Delta E_{\text{intra}} \rangle$ is still more physically meaningful than ΔE_{intra} due to the more faithful representation of the unbound state. In addition, the errors in both $\langle \Delta E_{\text{intra}} \rangle$ and ΔE_{intra} should largely cancel when computing the relative binding energies ($\Delta \Delta G_{\text{bind}}$) to obtain the final rank for congeneric series.

Figure 6 compares the estimation of the $-T\Delta S_{\text{conf}}$ penalties obtained from the Boltzmann distribution with the approach that penalizes each rotatable bond that becomes frozen upon binding by 0.65 kcal/mol. This last procedure assumes that each rotatable bond has three degenerate conformations, giving a total of 3^N conformations, all equal in energy, for a molecule with N rotatable bonds. The data set in Table 4 contains conformationally diverse compounds, with the number of rotatable bonds ranging from 0 to 16. The $-T\Delta S_{\text{conf}}$ penalties calculated for this data set using the constant penalty approximation are significantly overestimated with respect to the ones obtained from the Boltzmann distribution. This can be illustrated by the HIV-1 protease inhibitor (PDB ID: 3AID) shown in Scheme 1. The $-T\Delta S_{\text{conf}}$ value calculated by the constant penalty approximation is 9.1 kcal/mol since this compound has 14 rotatable bonds. The one calculated from the Boltzmann

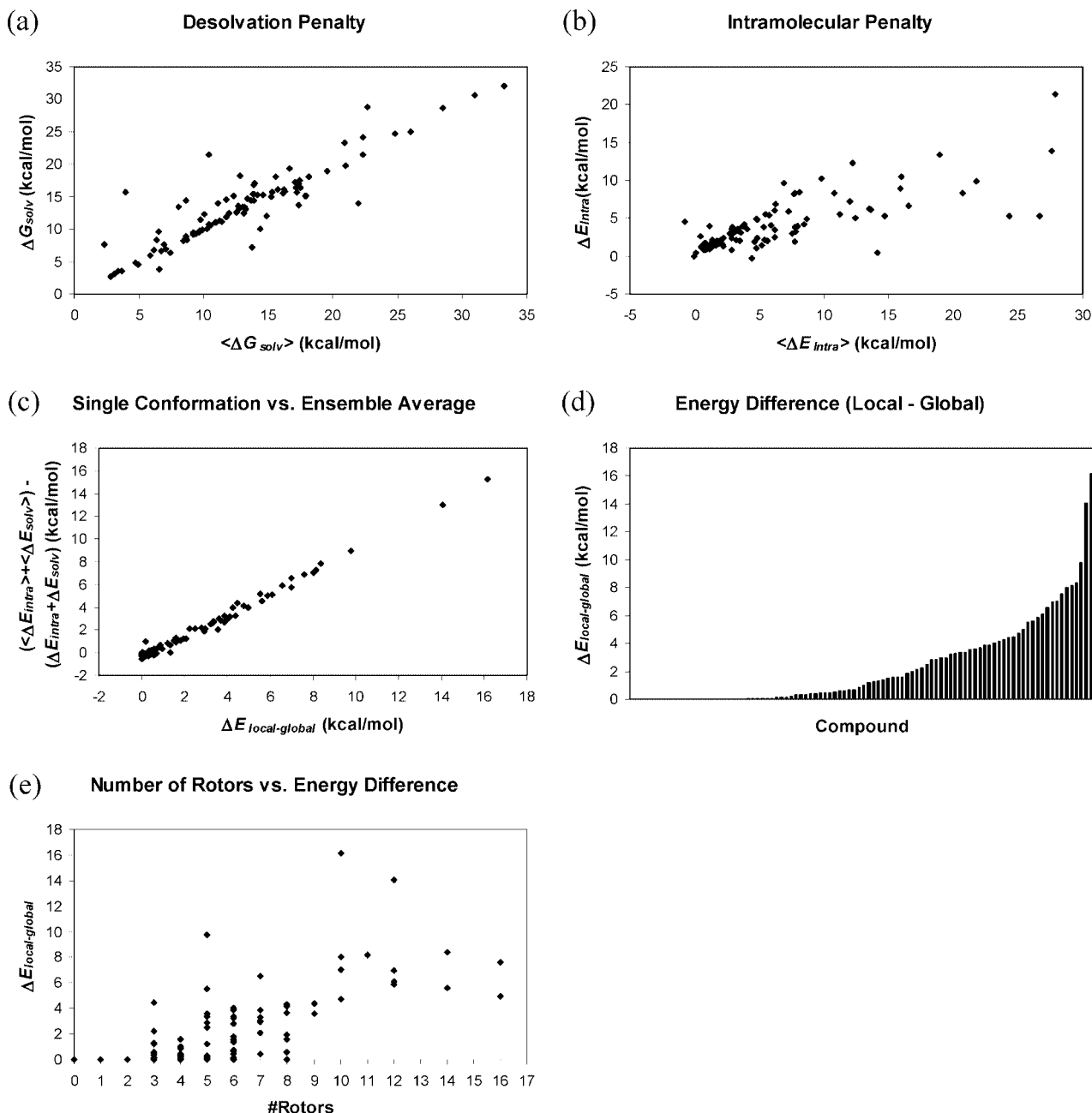


Figure 5. (a) Desolvation and (b) intramolecular penalties for 100 conformationally diverse inhibitors upon binding to their corresponding biological targets obtained using the single conformer (ΔG_{solv} and ΔE_{intra}) versus ensemble representations ($\langle \Delta G_{solv} \rangle$ and $\langle \Delta E_{intra} \rangle$). (c) Underestimation of the combined ($\Delta E_{intra} + \Delta G_{solv}$) in the single conformer representation with respect to the one obtained using the ensemble representation ($\langle \Delta E_{intra} \rangle + \langle \Delta G_{solv} \rangle$) versus energy difference between the local and global minima in water ($\Delta E_{local-global}$). (d) $\Delta E_{local-global}$ for 100 conformationally diverse inhibitors. (e) Number of rotatable bonds versus $\Delta E_{local-global}$.

distribution is only 0.7 kcal/mol. Although this compound is apparently very flexible with almost 100 conformations within 5 kcal/mol from the global energy minimum in water, the conformational entropy calculated from the Boltzmann distribution is small because the lowest and the second lowest energy conformers have probabilities of 73% and 14%, while the remaining ones have very small values. Both low-energy conformations are stabilized by an intramolecular hydrogen bond between the amide NH and the carbamate carbonyl group. Hydrophobic collapse of the aromatic rings that provides additional stabilization for these conformations was also observed. Hence, the ranking of conformationally diverse compounds by scoring functions that employ the constant penalty approximation should be adversely affected.

Interestingly, the $-T\Delta S_{conf}$ penalties calculated from the Boltzmann distribution for molecules that are so diverse in their degree of flexibility have a very small range, from 0 to ca. 3 kcal/mol (Figure 6). When analyzing each congeneric series, the typical range for this term was about only 1 kcal/mol, much narrower than the other terms in the scoring, indicating that the conformational entropy penalty should not contribute significantly for rank-ordering. However, it is important to have in mind that inaccuracies in the force field and in the GB/SA method might lead to imperfections in the weighing of the conformer distributions and, consequently, poor estimation of the $-T\Delta S_{conf}$ penalties.

MM-GB/SA Rescoring. The first analyzed target was CDK2. The docking method provided binding poses for all

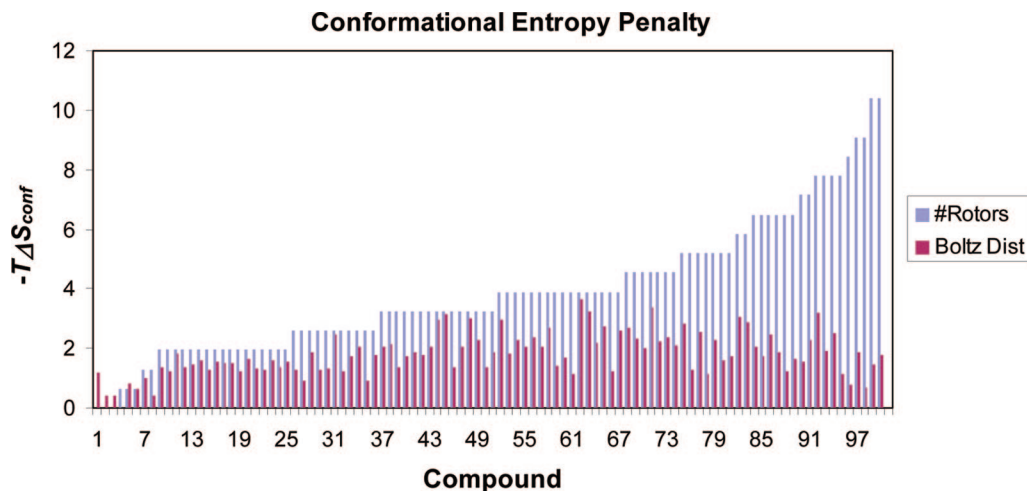
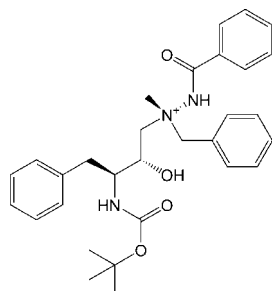


Figure 6. Estimation of the $-T\Delta S_{\text{conf}}$ penalties for 100 conformationally diverse inhibitors upon binding to their corresponding biological targets obtained from the Boltzmann distribution and using the constant penalty approximation (0.65 kcal/mol per rotatable bond).

Scheme 1. HIV-1 Protease Inhibitor (PDB ID: 3AID)



ligands in the 1E9H modified binding site (Figure 1c) that are consistent with the one displayed by the inactivated 1FVT experimental structure (Figure 1a). In spite of that, there is no correlation between the GlideXP score and the experimental data (Figure 7a). Rescoring with MM-GB/SA greatly increases the correlation, with superior results obtained using the ensemble representation for the unbound state (Figures 7b) when compared to the single conformer representation (Figure 7c). This is not caused by the lack of conformational entropic contributions in the latter. As discussed, the calculated $-T\Delta S_{\text{conf}}$ penalties do not contribute significantly for rank-ordering because they are very similar for all compounds. Rather, the lower correlation is caused by a single compound. In almost all cases, the local energy minimum found when relaxing the bound state conformation is the global energy minimum since these molecules are very rigid. The only exception is for the compound where R_2 is SO_2NH_2 , shown in Table 1 and highlighted in Figure 7c. The local energy minimum found in water in this case is 2.5 kcal/mol higher in energy than the global minimum, causing an underestimation of the combined desolvation and intramolecular penalties by approximately the same amount. Consequently, this compound is predicted to bind more favorably than it actually does, which is enough to negatively affect the correlation with experiment when using the single conformer representation.

Next, a series of Factor Xa inhibitors that have an increased degree of flexibility with 5–6 rotatable bonds were analyzed. As mentioned above, more flexible molecules have larger energy gaps between the local energy minimum found in water upon relaxation of the bound conformation and the corresponding global minimum. This can adversely affect

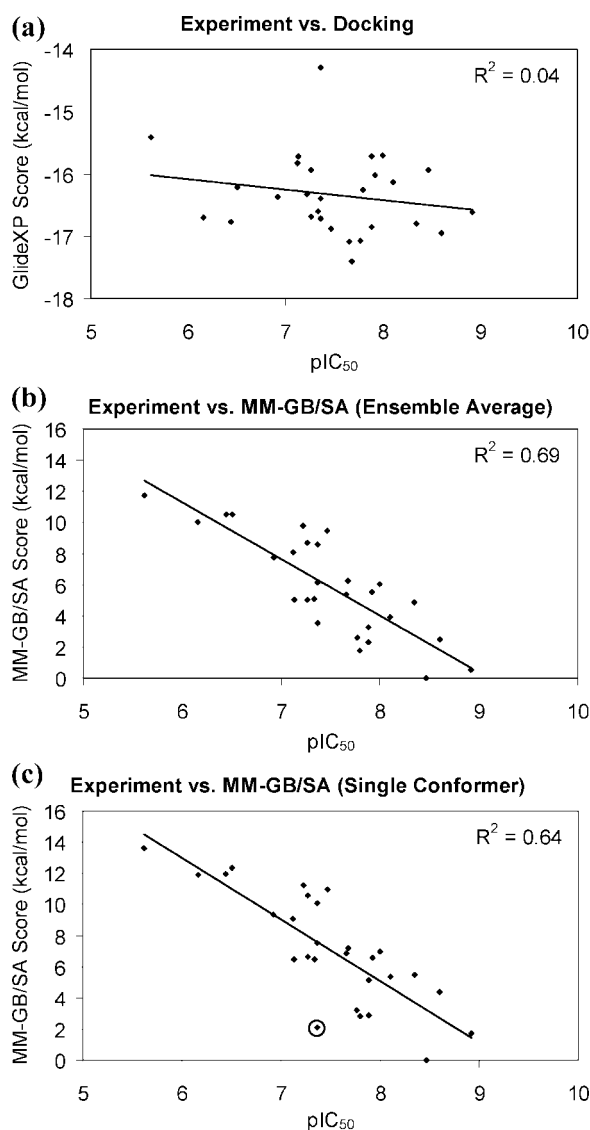


Figure 7. Correlation between experimental IC_{50} values for CDK2 inhibitors and (a) GlideXP scoring, (b) MM-GB/SA scoring using ensemble representation for the unbound state, and (c) MM-GB/SA scoring using single conformer representation.

the estimation of the combined intramolecular and desolvation penalties in the single conformer representation and have

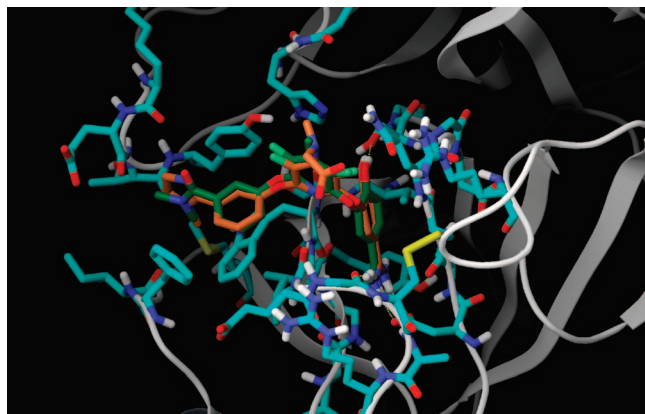


Figure 8. Comparison between the observed conformation for the inhibitor in the 1FJS crystal structure (brown) and the predicted binding mode for a closely related analogue (green) where R_1 , R_2 , and R_3 are CONMe_2 , H, and 6-OH, respectively (see Table 3).

a negative impact on the correlation with experiment. As demonstrated by the good agreement between the experimental binding mode for the inhibitor in the 1FJS crystal structure and the predicted conformation for a closely related analogue (Figure 8), accurate binding poses for the selected Factor Xa congeneric series are obtained by Glide. Rescoring with the MM-GB/SA method significantly improves correlation with the experimental K_i 's over docking, with no significant difference between the unbound state representations (Figure 9). In this case, the local energy minima in water for all compounds lie 1–3 kcal/mol above the global energy minima. The relatively small range for the energy gaps is caused by the similar degree of flexibility for the inhibitors in the congeneric series. Therefore, the corresponding underestimation of the combined intramolecular and desolvation penalties in the single conformer representation is not enough to deteriorate the results, especially because of the large dynamic range for the MM-GB/SA score (Figure 9).

The next system investigated consists of a series of Thrombin inhibitors. Again, Glide provides accurate binding poses (Figure 10) but struggles when scoring this congeneric series (Figure 11a). Excellent results are obtained with the MM-GB/SA method, with superior correlation using the ensemble representation for the unbound state (Figures 11b and 11c). Since this series not only is more flexible than the Factor Xa inhibitors but also has an increased range of flexibility with the number of rotatable bonds varying from 8 to 11, the local energy minima found in water with the single conformer representation are 3–7 kcal/mol higher in energy than the corresponding global minima. In this case, the magnitude of the noise in the combined intramolecular and desolvation penalties with respect to the more compressed dynamic range for the MM-GB/SA score causes a more significant deterioration of the correlation obtained with single conformer representation.

Despite the good results, the MM-GB/SA methodology is not perfect. For example, Figures 11b and 11c show an outlier compound, predicted to bind less favorably than observed experimentally. Analysis of the MM-GB/SA terms revealed that the outlier compound, where R is CH_3 (Figure 2), has a ΔG_{solv} penalty of 14.7 kcal/mol, while a pair of closely related analogues where R is H and NH_2 have ΔG_{solv} penalties of 12.9 and 13.0 kcal/mol, respectively. The

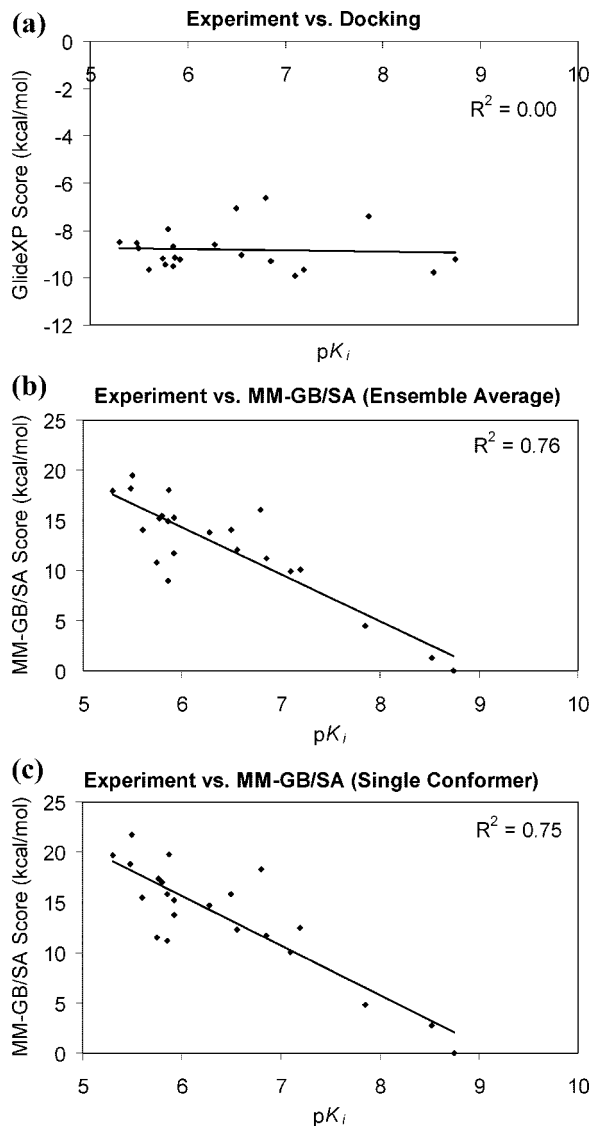


Figure 9. Correlation between experimental K_i values for Factor Xa inhibitors and (a) GlideXP scoring, (b) MM-GB/SA scoring using ensemble representation for the unbound state, and (c) MM-GB/SA scoring using single conformer representation.



Figure 10. Comparison between the observed conformation for the inhibitor in the 1ETT crystal structure (brown) and the predicted binding mode for a closely related analogue (green) with a K_i value of 52.4 nM (see Figure 2).

corresponding ΔG_{solv} penalties obtained using the single conformer representation are 12.6, 9.7, and 10.7 kcal/mol. The results suggest that the ΔG_{solv} penalties for the outlier

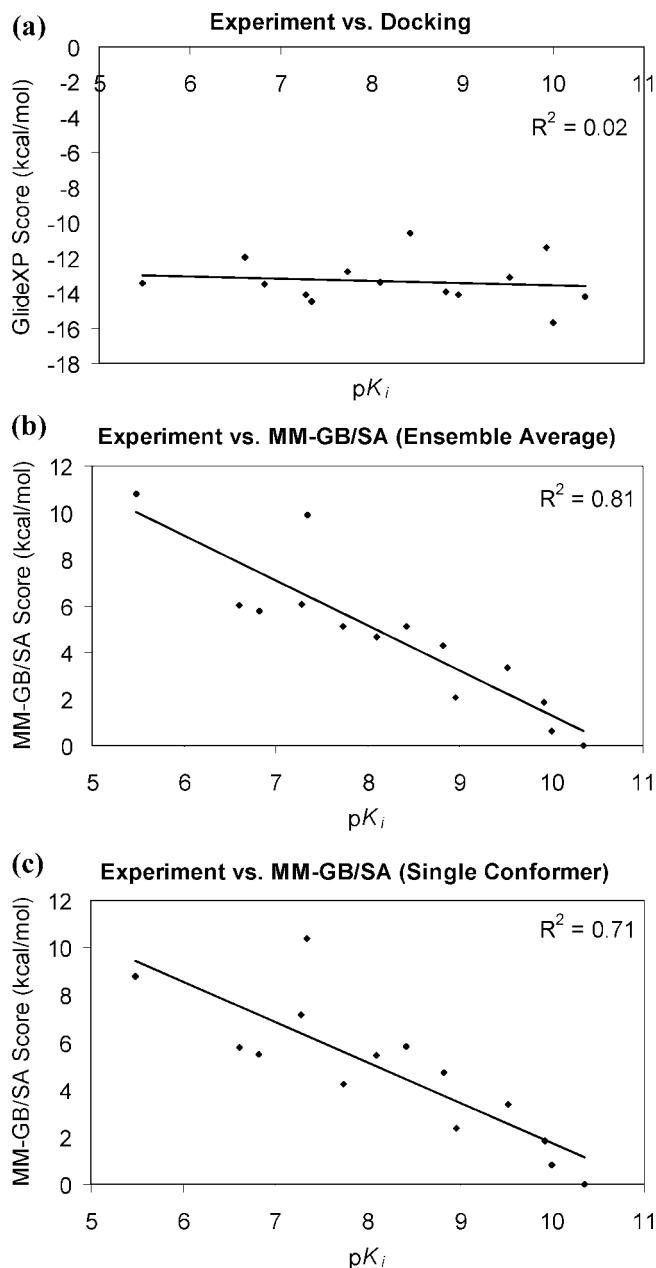


Figure 11. Correlation between experimental K_i values for Thrombin inhibitors and (a) GlideXP scoring, (b) MM-GB/SA scoring using ensemble representation for the unbound state, and (c) MM-GB/SA scoring using single conformer representation.

compound were overestimated in both cases. The methylation of the benzamidine group should make the desolvation process more favorable due to the reduction of the number of hydrogen bonds with the solvent. This is clearly a result of inaccuracies in the GB/SA method.

Finally, a set of eleven HIV-RT inhibitors containing seven different scaffolds were analyzed. This more diverse set should pose an extra challenge since inaccuracies in the force field and the GB/SA method that would affect the MM-GB/SA score are less likely to cancel. In addition, different scaffolds could potentially induce unique conformational changes in the protein that a rigid receptor docking followed by a constrained energy minimization would not be able to capture, especially in the case of a plastic enzyme like HIV-RT. In spite of its plasticity, the protein conformations of six complexes for which crystal structures are available are

very similar: their PDB IDs are 1IKW, 1RT1, 1VRU, 1TVR, 1VRT, and 1RT4. Consequently, the binding poses obtained from cross-docking of the respective inhibitors Sustiva, MKC-442, Loviride, 9-Cl-TIBO, and Nevirapine into the 1RT4 crystal structure, the complex between HIV-RT and UC-781, agree very well with the corresponding observed X-ray conformations. The rmsd between the predicted and the observed structures for five of the inhibitors is below 1.0 Å (Figure 12). The only inhibitor that presented an rmsd value above 1.0 Å was 9-Cl TIBO. This larger rmsd value can be attributed to the different position for the methyl group attached to the seven-membered ring and to the orientation of the flexible 3,3-dimethyl-allyl group.

In the case of HIV-RT, the correlation between the GlideXP score and the experimental data is far superior than for the other systems studied (Figure 13a). A similar correlation is obtained with the MM-GB/SA rescoring with no difference between the single conformer and ensemble representations (Figures 13b and 13c). This should be expected since this diverse set is composed of fairly rigid compounds. The high degree of rigidity could also be responsible for the good results obtained with the GlideXP score. The limitations of the docking scoring function associated with inaccuracies in the estimation of intramolecular, desolvation, and conformational entropy penalties should be minimized in this case.

The fact that the MM-GB/SA method succeeded for this diverse set of inhibitors indicates that potential problems with the force field and the GB/SA method and the ligand/protein induced fit effects did not play an important role. Obviously, this will not always be the case. The first problem is hard to foresee, and its only solution is reparameterization. The second could in theory be addressed using induced fit docking.⁴³ However, even if the method correctly reproduces the different protein conformations induced by each scaffold, it is unclear whether the force field would accurately compute the large variations in E_{PTN} (Figure 4), one of the MM-GB/SA scoring terms, between each conformation adopted by the protein. Investigation of more complicated cases than HIV-RT is warranted.

Finally, Figures 7, 9, 11, and 13 show a very large dynamic range for the MM-GB/SA scores when compared to the experimental range. It is possible that this has its origin in the application of a protein dielectric constant of 1 in a model where protein motions and polarization are not taken into account. In this case, electrostatic interactions are not shielded enough, and protein–ligand intermolecular electrostatic attractions and repulsions are overestimated, causing the large separation of potent and weak compounds. Another possible explanation for the large dynamic range is associated with the lack of a term in the MM-GB/SA scoring that describes enthalpy–entropy compensation. Since translational, rotational, and vibrational entropy changes upon binding are ignored, the ligands that interact the most favorably with the protein are not entropically penalized due to restriction of their vibrational modes in the binding site. Although not essential for congeneric series, as demonstrated by the good correlation obtained with experimental data, the aforementioned effects are likely very important when ranking different chemical classes and will be the subject of future research.

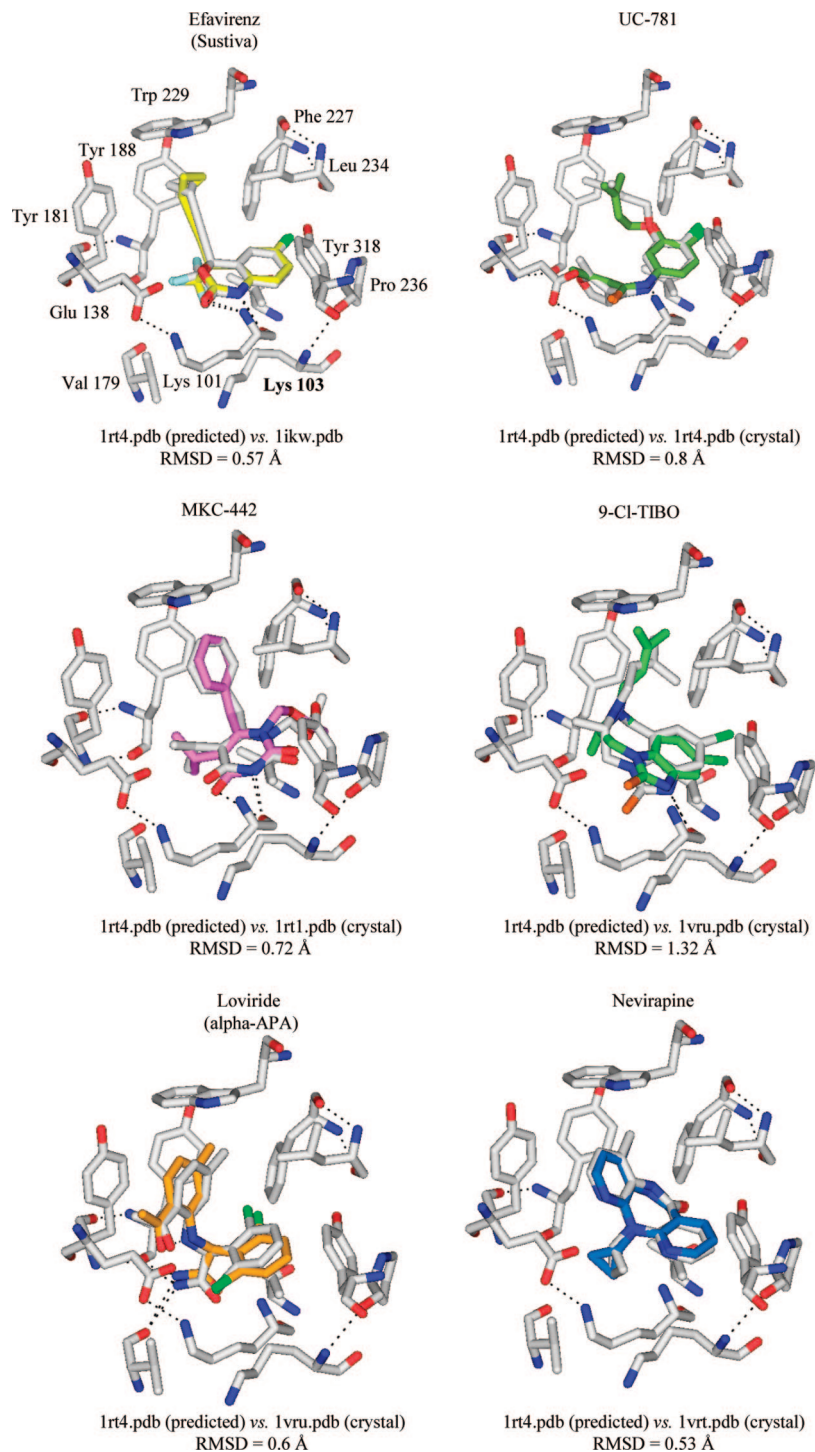


Figure 12. Comparison between the predicted and experimental X-ray binding modes for six HIV-RT inhibitors. The root-mean-square deviation (rmsd) values are shown. The predicted structures for the inhibitors are in light gray. The 1RT4 protein structure used in the docking calculations is also in light gray.

CONCLUSIONS

The performance of the MM-GB/SA rescoring of docking poses in structure-based lead optimization was investigated in this work. Overall, the correlation with experiment obtained with the physics-based scoring is far superior to docking. The remarkable results for all systems even qualify the MM-GB/SA approach as a more attractive alternative to the FEP and TI methodologies for rank-ordering. It can be as accurate, handle more structurally dissimilar ligands, and provide results at a fraction of the computational cost.

Regarding the unbound state representation, since flexible molecules are more easily deformed by the protein to maximize the intermolecular interactions, the relaxation of the bound state conformation in water, a procedure adopted in the single conformer representation, will more likely not find the global minimum or a conformation that is close in energy. As a consequence, the corresponding underestimation and noise in the combined intramolecular and desolvation penalties in the single conformer representation deteriorate the results for congeneric series containing more flexible ligands, especially for series that have a large range of

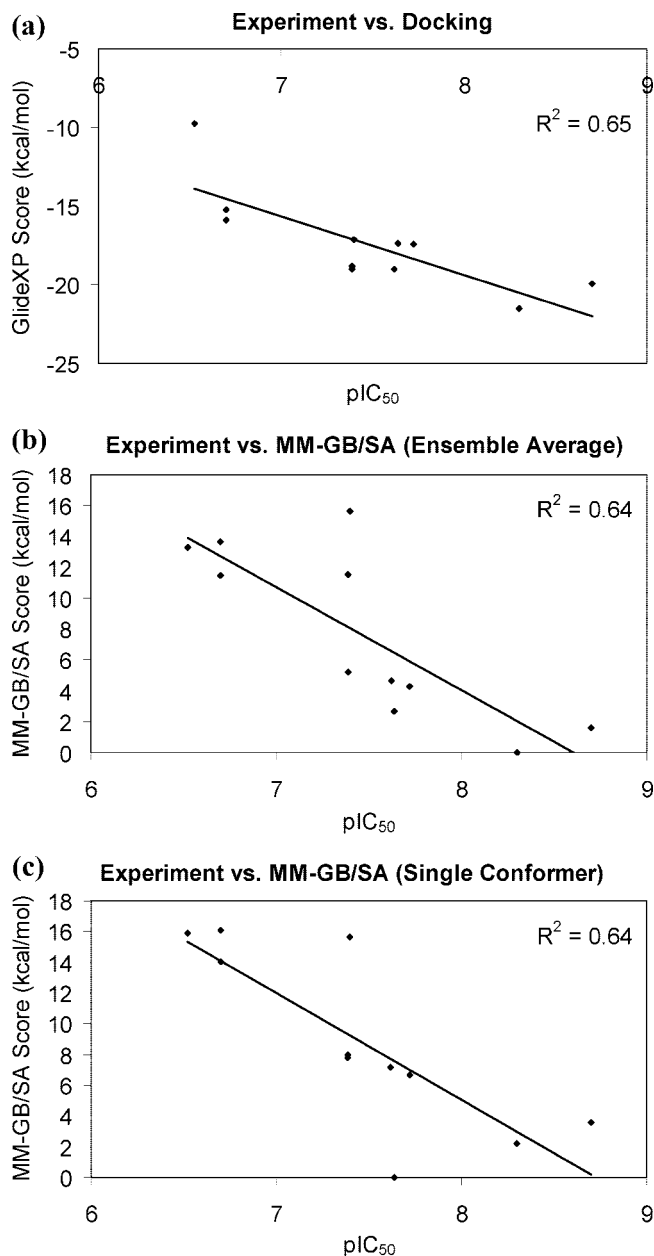


Figure 13. Correlation between experimental IC_{50} values for HIV-RT inhibitors and (a) GlideXP scoring, (b) MM-GB/SA scoring using ensemble representation for the unbound state, and (c) MM-GB/SA scoring using single conformer representation.

flexibility. The unbound state representation, however, is not a factor when dealing with more rigid compounds.

It has also been demonstrated that a penalty of 0.65 kcal/mol per rotatable bond in the ligand that becomes frozen upon binding significantly overestimates the conformational entropy penalty with respect to the one obtained using a Boltzmann distribution. Overall, the conformational entropy penalty term in the latter is very small and similar for all compounds, even when comparing ligands that are so diverse in their degree of flexibility. This suggests that this contribution should not be important for rank-ordering, especially when studying a congeneric series. However, it should be noted that inaccuracies in the force field and in the GB/SA method may lead to imperfections in the weighing of the conformer distributions and, consequently, to poor estimation of the conformational entropy penalty term.

Finally, the fact that the MM-GB/SA rescoring succeeded for the diverse set of HIV-RT inhibitors must not be generalized. Inaccuracies in the force field and the GB/SA method, and the ligand/protein induced fit effects, are still very likely to play important roles when dealing with different scaffolds. While there is not much one can do for the former, except for reparameterization, the latter can be attenuated using induced fit docking. Nevertheless, even if the different protein conformations induced by each scaffold are correctly reproduced, it is unclear whether the force field would accurately compute the large changes in the protein energy, one of the MM-GB/SA scoring terms. Investigation of these effects for other systems is definitely required.

ACKNOWLEDGMENT

Gratitude is expressed to John Eksterowicz and Nigel Walker from Amgen, Inc., and Woody Sherman at Schrodinger, Inc., for helpful discussions.

REFERENCES AND NOTES

- (1) 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.
- (2) Kollman, P. A. Free Energy Calculations: Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* **1993**, *93*, 2395–2417.
- (3) Jorgensen, W. L. Free Energy Calculations: A Breakthrough for Modeling Organic Chemistry in Solution. *Acc. Chem. Res.* **1989**, *22*, 184–189.
- (4) Simonson, T.; Georgios, A.; Karplus, M. Free Energy Simulations Come of Age: Protein–Ligand Recognition. *Acc. Chem. Res.* **2002**, *35*, 430–437.
- (5) Pearlman, D. A.; Charifson, P. S. Are Free Energy Calculations Useful in Practice? A Comparison with Rapid Scoring Functions for the p38 MAP Kinase Protein System. *J. Med. Chem.* **2001**, *44*, 3417–3423.
- (6) Guimarães, C. R. W.; Boger, D. L.; Jorgensen, W. L. Elucidation of Fatty Acid Amide Hydrolase Inhibition by Potent α -Ketoheterocycle Derivatives from Monte Carlo Simulations. *J. Am. Chem. Soc.* **2005**, *127*, 17377–17384.
- (7) 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.
- (8) Rizzo, R. C.; Tirado-Rives, J.; Jorgensen, W. L. Estimation of the 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-Blagovic, 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*, 2970–2987.
- (10) Pierce, A. C.; Jorgensen, W. L. Estimation of Binding Affinities for Selective Thrombin Inhibitors via Monte Carlo Simulations. *J. Med. Chem.* **2001**, *44*, 1043–1050.
- (11) 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*, 267–270.
- (12) Ostrovsky, D.; Udier-Blagovic, M.; Jorgensen, W. L. Analyses for Activity of Factor Xa Inhibitors Based on Monte Carlo Simulations. *J. Med. Chem.* **2003**, *46*, 5691–5699.
- (13) Tomimaga, Y.; Jorgensen, W. L. General Model for Estimation of the Inhibition of Protein Kinases using Monte Carlo Simulations. *J. Med. Chem.* **2004**, *47*, 2534–2549.
- (14) Taylor, R. D.; Jewsbury, P. J.; Essex, J. W. A Review of Protein–Small Molecule Docking Methods. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 151–166.
- (15) Shoichet, B. K.; McGovern, S. L.; Wei, B.; Irwin, J. J. Lead Discovery using Molecular Docking. *Curr. Opin. Chem. Biol.* **2002**, *6*, 439–446.
- (16) Walters, W. P.; Stahl, M. T.; Murcko, M. A. Virtual Screening - An Overview. *Drug Discovery Today* **1998**, *3*, 160–178.
- (17) Shoichet, B. K. Virtual Screening of Chemical Libraries. *Nature* **2004**, *432*, 862–865.
- (18) Powers, R. A.; Morandi, F.; Shoichet, B. K. Structure-Based Discovery of a Novel, Noncovalent Inhibitor of AmpC Beta-Lactamase. *Structure* **2002**, *10*, 1013–1023.

- (19) Schapira, M.; Abagyan, R.; Totrov, M. Nuclear Hormone Receptor Targeted Virtual Screening. *J. Med. Chem.* **2003**, *46*, 3045–3059.
- (20) Alvarez, J. C. High-Throughput Docking as a Source of Novel Drug Leads. *Curr. Opin. Chem. Biol.* **2004**, *8*, 1–6.
- (21) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. A Geometric Approach to Macromolecule-Ligand Interactions. *J. Mol. Biol.* **1982**, *161*, 269–288.
- (22) Jones, G.; Willet, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and Validation of a Genetic Algorithm for Flexible Docking. *J. Mol. Biol.* **1997**, *267*, 727–748.
- (23) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A Fast Flexible Docking Method using an Incremental Construction Algorithm. *J. Mol. Biol.* **1996**, *261*, 470–489.
- (24) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47*, 1739–1749.
- (25) Muegge, I.; Martin, Y. C. A General and Fast Scoring Function for Protein-Ligand Interactions: A Simplified Potential Approach. *J. Med. Chem.* **1999**, *42*, 791–804.
- (26) Charifson, P. S.; Corkey, J. J.; Murcko, M. A.; Walters, W. P. Consensus Scoring: A Method for Obtaining Improved Hit Rates from Docking Databases of Three-Dimensional Structures into Proteins. *J. Med. Chem.* **1999**, *42*, 5100–5109.
- (27) Perola, E.; Walters, W. P.; Charifson, P. S. A Detailed Comparison of Current Docking and Scoring Methods on Systems of Pharmaceutical Relevance. *Proteins* **2004**, *56*, 235–249.
- (28) Stahl, M.; Rarey, M. Detailed Analysis of Scoring Functions for Virtual Screening. *J. Med. Chem.* **2001**, *44*, 1035–1042.
- (29) Warren, G. L.; Andrews, C. W.; Capelli, A.-M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. A Critical Assessment of Docking Programs and Scoring Functions. *J. Med. Chem.* **2006**, *49*, 5912–5931.
- (30) Tirado-Rives, J.; Jorgensen, W. L. Contribution of Conformer Focusing to the Uncertainty in Predicting Free Energies for Protein-Ligand Binding. *J. Med. Chem.* **2006**, *49*, 5880–5884.
- (31) (a) Kuhn, B.; Kollman, P. A. Binding of a Diverse Set of Ligands to Avidin and Streptavidin: An Accurate Quantitative Prediction of their Relative Affinities by a Combination of Molecular Mechanics and Continuum Solvent Models. *J. Med. Chem.* **2000**, *43* (31), 3786–3791. (b) Barril, X.; Gelpí, J. L.; López, J. M.; Orozco, M.; Luque, F. J. How Accurate Can Molecular Dynamics/Linear Response and Poisson-Boltzmann/Solvent Accessible Surface Calculations Be for Predicting Relative Binding Affinities? Acetylcholinesterase as Huprine Inhibitors as a Test Case. *Theor. Chem. Acc.* **2001**, *106*, 2–9.
- (32) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semi-analytical Treatment of Solvation for Molecular Mechanics and Dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.
- (33) (a) Bernacki, K.; Kalyanaraman, C.; Jacobson, M. P. Virtual Ligand Screening against *Escherichia coli* Dihydrofolate Reductase: Improving Docking Enrichment Physics-Based Methods. *J. Biomol. Screening* **2005**, *10*, 675–681. (b) Huang, N.; Kalyanaraman, C.; Irwin, J. J.; Jacobson, M. P. Physics-Based Scoring of Protein-Ligand Complexes: Enrichment of Known Inhibitors in Large-Scale Virtual Screening. *J. Chem. Inf. Model.* **2006**, *46*, 243–253. (c) Huang, N.; Kalyanaraman, C.; Bernacki, K.; Jacobson, M. P. Molecular Mechanics Methods for Predicting Protein-Ligand Binding. *Phys. Chem. Chem. Phys.* **2006**, *8*, 5166–5177. (d) Lyne, P. D.; Lamb, M. L.; Saeh, J. C. Accurate Prediction of the Relative Potencies of Members of a Series of Kinase Inhibitors using Molecular Docking and MM-GBSA Scoring. *J. Med. Chem.* **2006**, *49*, 4805–4808. (e) Lee, M. R.; Sun, Y. Improving Docking Accuracy through Molecular Mechanics Generalized Born Optimization and Scoring. *J. Chem. Theory Comput.* **2007**, *3*, 1106–1119. (f) Huang, N.; Jacobson, M. P. Physics-Based Methods for Studying Protein-Ligand Interactions. *Curr. Opin. Drug Discovery Dev.* **2007**, *10*, 325–331. (g) Foloppe, N.; Hubbard, R. Towards Predictive Ligand Design with Free-Energy Based Computational Methods. *Curr. Med. Chem.* **2006**, *13*, 3583–3608.
- (34) Salaniwal, S.; Manas, E. S.; Alvarez, J. C.; Unwalla, R. J. Critical Evaluation of Methods to Incorporate Entropy Loss upon Binding in High-Throughput Docking. *Proteins* **2007**, *66*, 422–435.
- (35) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. *J. Med. Chem.* **2006**, *49*, 6177–6196.
- (36) (a) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11235. (b) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. J. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *J. Phys. Chem. B* **2001**, *105*, 6474–6487.
- (37) Bramson, H. N.; Corona, J.; Davis, S. T.; Dickerson, S. H.; Edelstein, M.; Frye, S. V.; Gampe, R. T.; Harris, J. P. A.; Hassell, A.; Holmes, W. D.; Hunter, R. N.; Lackey, K. E.; Lovejoy, B.; Luzzio, M. J.; Montana, V.; Rocque, W. J.; Rusnak, D.; Shewchuk, L.; Veal, J. M.; Walker, D. H.; Kuyper, L. F. Oxindole-Based Inhibitors of Cyclin-Dependent Kinase 2 (CDK2): Design, Synthesis, Enzymatic activities, and X-ray Crystallographic Analysis. *J. Med. Chem.* **2001**, *44*, 4339–4358.
- (38) (a) 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. (b) 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.
- (39) (a) Kim, S.; Hwang, S. Y.; Kim Young, K.; Yun, M.; Oh, Y. S. Rational Design of Selective Thrombin Inhibitors. *Bioorg. Med. Chem. Lett.* **1997**, *7*, 769–774. (b) Oh, Y. S.; Yun, M.; Hwang, S. Y.; Hong, S.; Shin, Y.; Lee, K.; Yoon, K. H.; Yoo, Y. J.; Kim, D. S.; Lee, S. H.; Lee, Y. H.; Park, H. D.; Lee, C. H.; Lee, S. K.; Kim, S. Discovery of LB30057, a Benzamidrazone-based Selective Oral Thrombin Inhibitor. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 631–634. (c) Lee, K.; Jung, W.-H.; Park Cheol, W.; Hong Chang, Y.; Kim In, C.; Kim, S.; Oh Yeong, S.; Kwon, O. H.; Lee, S.-H.; Park Hee, D.; Kim Sang, W.; Lee Yong, H.; Yoo Yung, J. Benzylamine-Based Selective and Orally Bioavailable Inhibitors of Thrombin. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 2563–2568.
- (40) (a) King, R. W.; Klabe, R. M.; Reid, C. D.; Erickson-Viitanen, S. K. Potency of Nonnucleoside Reverse Transcriptase Inhibitors (NNRTIs) used in Combination with Other Human Immunodeficiency Virus NNRTIs, NRTIs, or Protease Inhibitors. *Antimicrob. Agents Chemother.* **2002**, *46*, 1640–1646. (b) Mai, A.; Sbardella, G.; Artico, M.; Ragno, R.; Massa, S.; Novellino, E.; Greco, G.; Lavecchia, A.; Musiu, C.; La Colla, M.; Murgioni, C.; La Colla, P.; Loddio, R. Structure-Based Design, Synthesis, and Biological Evaluation of Conformationally Restricted Novel 2-Alkylthio-6-[1-(2,6-difluorophenyl)alkyl]-3,4-dihydro-5-alkylpyrimidin-4(3H)-ones as Non-Nucleoside Inhibitors of HIV-1 Reverse Transcriptase. *J. Med. Chem.* **2001**, *44*, 2544–2554. (c) Sahlberg, C.; Norén, R.; Engelhardt, P.; Högborg, M.; Kangas-metsä, J.; Vrang, L.; Zhang, H. Synthesis and Anti-HIV Activities of Urea-pETT Analogs Belonging to a New Class of Potent Non-Nucleoside HIV-1 Reverse Transcriptase Inhibitors. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 1511–1516. (d) Herrewewe, Y. V.; Michiels, J.; Van Roey, J.; Fransen, K.; Kestens, L.; Balzarini, J.; Lewi, P.; Vanham, G.; Janssen, P. In Vitro Evaluation of Nonnucleoside Reverse Transcriptase Inhibitors UC-781 and TMC120-R147681 as Human Immunodeficiency Virus Microbicides. *Antimicrob. Agents Chemother.* **2004**, *48*, 337–339. (e) De Clercq, E. Perspectives of Non-Nucleoside Reverse Transcriptase Inhibitors (NNRTIs) in the Therapy of HIV-1 Infection. *Farmacology* **1999**, *54*, 26–45.
- (41) MacroModel, version 9.0; Schrödinger, LLC: New York, NY, 2005.
- (42) Kuhn, B.; Gerber, P.; Schulz-Gasch, T.; Stahl, M. Validation and Use of the MM-PBSA Approach for Drug Discovery. *J. Med. Chem.* **2005**, *48*, 4040–4048.
- (43) (a) Sherman, W.; Day, T.; Jacobson, M. P.; Friesner, R. A.; Farid, R. Novel Procedure for Modeling Ligand/Receptor Induced Fit Effects. *J. Med. Chem.* **2006**, *49*, 534–553. (b) Sherman, W.; Beard, H. S.; Farid, R. Use of Induced Fit Receptor Structure in Virtual Screening. *Chem. Biol. Drug Des.* **2006**, *67*, 8384.

CI800004W