Combining Docking, Molecular Dynamics and the Linear Interaction Energy Method to Predict Binding Modes and Affinities for Non-nucleoside Inhibitors to HIV-1 Reverse Transcriptase

Jens Carlsson, Lars Boukharta, and Johan Åqvist*

Department of Cell and Molecular Biology, Uppsala University, Biomedical Center, Box 596, SE-751 24 Uppsala, Sweden

Received September 26, 2007

Docking, scoring, molecular dynamics (MD), and the linear interaction energy (LIE) method are used here to predict binding modes and affinities for a set of 43 non-nucleoside inhibitors to HIV-1 reverse transcriptase. Starting from a crystallographic structure, the binding modes of 43 inhibitors are predicted using automated docking. The Goldscore scoring function and the LIE method are then used to determine the relative binding free energies for the inhibitors. The Goldscore scoring function does not reproduce the relative binding affinities for the inhibitors, while the standard parametrization of the LIE method reproduces the experimental binding free energies for 39 inhibitors with an $R^2 = 0.70$ and an unsigned average error of 0.8 kcal/mol. The present calculations provide a validation of the combination of docking, MD, and LIE as a powerful tool in structure-based drug design, and the methodology is easily scalable for attaining a higher throughput of compounds.

1. Introduction

HIV-1 reverse transcriptase (RT) is one of the main targets in the development of drugs against AIDS. An interesting class of anti-HIV drugs is the non-nucleoside RT inhibitors (NNRTIs). The NNRTIs are a diverse group of compounds that bind to a cavity which is created upon formation of the RT−inhibitor complex, thereby inducing a conformational change that inhibits the function of RT. Several NNRTIs, e.g., Nevirapine and Efavirenz, are already used in AIDS therapy, but a major problem is that drug resistant mutants arise quickly after treatment is initiated. A great challenge is therefore to continuously develop new inhibitors that are effective against both the wild type and mutant forms of the enzyme.[1–3]

In the case of RT, where crystallographic data is available, computational structure-based drug design is a promising approach to identify and optimize drug candidates. Starting from a known receptor structure, the binding modes for a set of ligands can often be successfully modeled using automated docking.[4–8] In general, these methods allow only limited or no receptor flexibility and, hence, it is assumed that the chosen structure provides a reasonable representation of the protein in its complexed form. For example, docking of inhibitors to the uncomplexed state of RT would be useless considering that significant conformational changes in the protein are required to create the NNRTI binding site. Once binding modes of the ligands have been predicted, the choice of method to estimate the free energies of binding is, in practice, determined by the number of compounds to be analyzed.[9] Empirical[10,11] and knowledge-based[12,13] scoring functions estimate the binding affinity from a single structure of the protein−ligand complex and can be used to filter out drug candidates from large databases of compounds ($\sim 10^7$). Whereas the empirical scoring functions are appealing for their speed, they treat contributions to the binding free energy, e.g., entropy and solvation, in a very approximate fashion, which makes it difficult to obtain accurate predictions. For smaller sets of ligands, explicit solvent molecular dynamics (MD)[a] and Monte Carlo (MC) simulations can be used to accurately describe receptor flexibility and solvation. The most advanced approaches used with MD and MC simulations are the free energy perturbation (FEP) and thermodynamic integration (TI) methods.[14,15] FEP and TI can be used to estimate absolute or relative binding free energies, but the calculations are computationally demanding and the techniques are limited to only a few similar ligands (~ 10).[16–19] The linear interaction energy (LIE) method is a semiempirical approach that combines the advantages of FEP and TI with useful approximations that improve convergence and decrease computation time.[20] The LIE method gives the opportunity to study larger sets of ligands ($\sim 10^3$) and is suitable for lead optimization. LIE binding free energies are estimated from MD or MC simulations of the bound and the free state of a ligand, and from these two simulations, the ligand-surrounding (l-s) electrostatic and van der Waals energies are collected. The binding free energy is evaluated from

$$\Delta G_{\mathrm{bind}}^{\mathrm{LIE}} = \alpha \Delta \langle U_{\mathrm{l\text{-}s}}^{\mathrm{vdW}} \rangle + \beta \Delta \langle U_{\mathrm{l\text{-}s}}^{\mathrm{el}} \rangle + \gamma \qquad (1)$$

where the Δ's denote that the difference between the energies for the bound and free states is taken. $\langle U_{\mathrm{l\text{-}s}}^{x} \rangle$ represents MD or MC averages of intermolecular electrostatic ($x = \mathrm{el}$) and van der Waals ($x = \mathrm{vdW}$) energies for the ligand with its surroundings. In our standard parametrization of the LIE method, $\alpha = 0.18$ and the β values vary between 0.33 and 0.5 depending on the chemical nature of the ligand.[21] The parameter γ is a constant term for ligands that bind to the same receptor and only needs to be considered for calculations of absolute binding free energies.[20–22] Our standard parametrization of the LIE method has previously been applied to several different protein−ligand complexes with impressive results.[21–25]

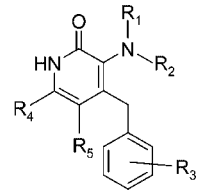
In this work, docking, scoring, MD simulations, and the LIE method are combined to predict binding modes and estimate binding free energies for 43 NNRTIs to RT. The GOLD docking program[4,5] is used to predict the binding modes for the inhibitors. The relative affinities of the inhibitors are then

* Corresponding author. Phone: +46 18 471 4109. Fax: +46 18 53 69 71. E-mail: aqvist@xray.bmc.uu.se.

[a] Abbreviations: MD, molecular dynamics; LIE, linear interaction energy; NNRTI, non-nucleoside reverse transcriptase inhibitor; RT, reverse transcriptase; MC, Monte Carlo; FEP, free energy perturbation; TI, thermodynamic integration; rms, root-mean-square.

10.1021/jm7012198 CCC: $40.75 © 2008 American Chemical Society
Published on Web 04/12/2008

Table 1. NNRTIs Used in This Work and Their Experimental IC_{50} Values (μM)[36a]



	R_1	R_2	R_3	R_4	R_5	IC_{50}
9	CH_3	CH_3	3,5-diCH_3	CH_3	C_2H_5	0.008
18b	CH_3	CH_3	3,5-diCH_3	CH_3	CH_3	0.005
18c	CH_3	CH_3	3,5-diCH_3	C_2H_5	CH_3	0.010
18d	CH_3	CH_3	3,5-diCH_3	CH_3	*i*-C_4H_9	0.050
18e	CH_3	CH_3	3,5-diCH_3	*i*-C_3H_7	CH_3	0.794
18f	CH_3	CH_3	3,5-diCH_3	CH_3	*n*-C_3H_7	0.050
18g	CH_3	CH_3	3,5-diCH_3	H	H	7.943
18h	CH_3	CH_3	3,5-diCH_3	CH_3	H	0.398
18i	CH_3	CH_3	3,5-diCH_3	$-(CH_2)_4-$	$-(CH_2)_4-$	0.010
36	H	CHO	3,5-diCH_3	CH_3	C_2H_5	0.079
37	H	CH_3	3,5-diCH_3	CH_3	C_2H_5	0.010
38	CH_3	C_2H_5	3,5-diCH_3	CH_3	C_2H_5	0.008
39	CH_3	C_3H_7	3,5-diCH_3	CH_3	C_2H_5	0.016
40	CH_3	$CH(CH_3)CH_2OCH_3$	3,5-diCH_3	CH_3	C_2H_5	0.006
41	CH_3	$(CH_2)_3SCH_3$	3,5-diCH_3	CH_3	C_2H_5	0.025
42	CH_3	$CH_2CH_2OCH_3$	3,5-diCH_3	CH_3	C_2H_5	0.002
43	CH_3	$(CH_2)_5OH$	3,5-diCH_3	CH_3	C_2H_5	0.004
44	H	$COCH_3$	3,5-diCH_3	CH_3	C_2H_5	0.398
45	H	COC_2H_5	3,5-diCH_3	CH_3	C_2H_5	3.981
46	H	COC_3H_7	3,5-diCH_3	CH_3	C_2H_5	100
47	H	C_2H_5	3,5-diCH_3	CH_3	C_2H_5	0.016
48	H	C_3H_7	3,5-diCH_3	CH_3	C_2H_5	0.020
49	H	C_4H_9	3,5-diCH_3	CH_3	C_2H_5	0.126
50	C_2H_5	C_2H_5	3,5-diCH_3	CH_3	C_2H_5	0.016
51	C_4H_9	C_4H_9	3,5-diCH_3	CH_3	C_2H_5	50.119
52	H	$CH_2C_6H_5$	3,5-diCH_3	CH_3	C_2H_5	0.251
53	$CH_2C_6H_5$	$CH_2C_6H_5$	3,5-diCH_3	CH_3	C_2H_5	100
54	$(CH_2CH_2)_2O$	$(CH_2CH_2)_2O$	3,5-diCH_3	CH_3	C_2H_5	0.158
55	$(CH_2)_5$	$(CH_2)_5$	3,5-diCH_3	CH_3	C_2H_5	0.631
56	$-CH=CH-CH=CH-$	$-CH=CH-CH=CH-$	3,5-diCH_3	CH_3	C_2H_5	0.0126
59	CH_3	$(CH_2)_2OH$	3-CH_3	CH_3	C_2H_5	0.005
60	CH_3	$(CH_2)_3OH$	3-CH_3	CH_3	C_2H_5	0.003
61	CH_3	$(CH_2)_5OH$	3-CH_3	CH_3	C_2H_5	0.010
62	CH_3	$(CH_2)_2OCH_3$	3-CH_3	CH_3	C_2H_5	0.001
63	CH_3	$(CH_2)_2OC_2H_5$	3-CH_3	CH_3	C_2H_5	0.013
64	CH_3	CH_2CN	3-CH_3	CH_3	C_2H_5	0.004
65	CH_3	$(CH_2)_2CN$	3-CH_3	CH_3	C_2H_5	0.016
66	CH_3	$(CH_2)_3CN$	3-CH_3	CH_3	C_2H_5	0.005
67	H	NH-CS-NHC_2H_5	3-CH_3	CH_3	C_2H_5	25.119
68	H	NH-CS-NHC_6H_5	3-CH_3	CH_3	C_2H_5	3.162
70	H	NH-CS-NH_2	3-CH_3	CH_3	C_2H_5	0.316
77a	CH_3	C_2H_5	3-CH=CHCN	CH_3	C_2H_5	0.001
77b	CH_3	$(CH_2)_2OCH_3$	3-CH=CHCN	CH_3	C_2H_5	0.001

[a] The naming of the inhibitors has been adopted from ref 36.

estimated using the Goldscore scoring function and MD simulations combined with the LIE method. Both the standard and a series of alternative LIE models, in which an additional scaling factor or intramolecular ligand energies are introduced, are investigated. The protein residues that are responsible for the difference in binding affinities between the inhibitors are also identified and these results are compared to experimental data on common mutants of RT. The present work can also be viewed as a benchmark that serves as a validation of the combined docking and LIE methodology in a realistic inhibitor design situation, and the procedure is straightforward to scale up into a higher throughput of compounds.

2. Methods

2.1. Solute Starting Structures and Docking Calculations. The studied inhibitors are a series of benzylpyridinone derivatives, which are shown in Table 1. A crystallographic structure of HIV-1 RT in complex with **62** (PDB code 2BAN[26]) was used as protein starting structure in all docking and MD simulations of the RT−inhibitor complexes. Each inhibitor was first minimized in the program Gaussian[27] using AM1[28] and then docked into RT using GOLD[4,5] with default genetic search parameters and 20 docking runs. The top three docking solutions from each docking calculation were inspected manually. If the three solutions represented a similar binding mode, these conformations were selected for MD simulations. In cases where the three top solutions had considerably different binding modes, three conformations from the largest cluster of solutions were used in the MD simulations. For **77a** and **77b**, the binding mode of a related inhibitor is available (PDB code: 2B5J[26]), and in these two cases, three conformations similar to that of this structure were chosen.

2.2. MD Simulations. MD simulations were performed using the program Q.[29] Simulations of the bound and free state for each inhibitor were carried out in an 18 Å sphere centered on the inhibitor. Except for the partial charges of the thiourea inhibitors, OPLS all-atom[30] force field parameters were used in all simulations.

For the substituents of inhibitors **67**, **68**, and **70**, 6-31G*/RESP charges were derived.[31] In the simulations, Lys103A, Glu138B, Asp192A, Asp186A, Asp237A, Lys172A, Lys101A, and Lys102A were ionized and all other residues close to the sphere edge or further away than 18 Å were set to their neutral state. Because all ligands in the present data set are neutral, no correction for distant ionizable groups is needed.[18] The protonation states of the histidines within the sphere were set by manual inspection. Each system was solvated by adding a spherical TIP3P[32] water grid of radius 18 Å and removing all water molecules closer than 2.4 Å of any solute heavy atom. For the solvated RT−inhibitor system, all protein residues within 18 Å of the center of the sphere were explicitly included in the simulations. All atoms outside the sphere were tightly restrained to their initial coordinates and excluded from nonbonded interactions. For the free inhibitors in water, a weak harmonic restraint was applied to the geometrical center of the solute to prevent it from approaching the sphere edge. The SHAKE[33] algorithm was applied to all solvent bonds and angles and the water molecules at the sphere surface were subjected to radial and polarization restraints.[29,34] The nonbonded cutoff was set to 10 Å for all atoms except for the inhibitor, to which no cutoff was applied. Long-range electrostatic interactions were treated with the local reaction field approximation.[35] The time step was set to 1 fs, and the simulations were carried out at a constant temperature of 310 K. For each simulation of the RT−inhibitor complex, the system was slowly heated while restraints on the solute coordinates to their initial positions were gradually released. This was followed by 50 ps of unrestrained equilibration and a 500 ps production phase in which ligand-surrounding energies were collected every 25 fs. Starting from slightly different initial conditions, two sets of MD simulations of the RT−inhibitor complexes were carried out for each of the three conformations extracted from the docking calculations. The reported ligand-surrounding energy for each ligand is taken as an average over these trajectories, and errors were estimated from the standard error of the mean. In water, a short equilibration was followed by a 500−600 ps production phase and three simulations, starting from different initial conformations, were carried out for each inhibitor.

2.3. Experimental and Computational Estimates of Binding Free Energies. Approximate (relative) experimental binding free energies were estimated from IC_{50} values[36] at 310 K using the relationship $\Delta G^{obs}_{bind} = RT \ln(IC_{50}) + c$, where c is a constant ($c = -RT \ln[1 + S/K_M]$).[37] Computationally, the relative affinities of the inhibitors were estimated using the Goldscore scoring function and the LIE method. The Goldscore, S^{GOLD}, is defined as described on the Cambridge Crystallographic Data Centre Web site (www.ccdc.cam.ac.uk), i.e., $S^{GOLD} = 1.375 \times S^{vdW\text{-}ext} + S^{hb\text{-}ext}$, where $S^{vdW\text{-}ext}$ and $S^{hb\text{-}ext}$ are the external van der Waals and hydrogen bond terms in the GOLD fitness function, respectively. S^{GOLD} does not explicitly represent a free energy but has been found to correlate well with experimental binding free energies.[38] For the LIE method, three different models were first compared by using different values for α and β in eq 1. The value of the β coefficient can be derived from linear response approximation, which predicts that $\beta = 0.5$.[39] However, on the basis of rigorous FEP calculations in different solvents carried out by Åqvist and Hansson,[39] the β value used for a ligand in the standard parametrization of the LIE method, β_{FEP}, is determined by its chemical groups. For the inhibitors studied here, β_{FEP} is equal to 0.43 in all cases except for **43** and **59**−**61** (for which $\beta_{FEP} = 0.37$).[21] The second term in eq 1 represents the nonpolar contribution to binding. This approximation is based on the linear relationships that are observed between solvation free energies of nonpolar molecules and the ligand-surrounding van der Waals energy.[20,40] In the standard LIE parametrization, α was determined to be 0.18 based on simulations of 18 protein−ligand complexes.[21] Because only relative binding free energies can be extracted from the experimental data, the constant γ is freely optimized to minimize the root-mean-square (rms) deviation from experiment for all studied models. The standard model was also compared to a model where the β_{FEP} values of the standard model are replaced by refined β_{FEP} values according to the scheme that was derived by Almlöf et al. (model A: for all inhibitors, $\beta_{FEP} = 0.43$, except for **37**, **43**, **47**−**49**, **52**, **59**−**61**, ($\beta_{FEP} = 0.41$), and **70** ($\beta_{FEP} = 0.42$)).[41] These models were also compared to model B, where α, β, and γ are all freely optimized by minimizing the rms deviation from experimental free energies of binding. This model thus serves as a check of the robustness of earlier parametrizations. In model C, an alternative form of eq 1 is investigated by scaling the electrostatic ligand-surrounding energies for the bound (b) and free (f) state separately[21,41]

$$\Delta G_{bind} = \beta_b \langle U^{el}_{l\text{-}s} \rangle_b - \beta_f \langle U^{el}_{l\text{-}s} \rangle_f + \alpha \Delta \langle U^{vdW}_{l\text{-}s} \rangle + \gamma \quad (2)$$

where β_b, β_f, and γ are optimized. Finally, model D is based on eq 3 and includes both intra- (l-l) and intermolecular (l-s) ligand electrostatic energies

$$\Delta G_{bind} = \beta(\Delta \langle U^{el}_{l\text{-}s} \rangle + \Delta \langle U^{el}_{l\text{-}l} \rangle) + \alpha \Delta \langle U^{vdW}_{l\text{-}s} \rangle + \gamma \quad (3)$$

where the values of α, β, and γ are freely optimized. To compare the different models, several statistical figures of merit were computed. The coefficient of multiple determination measures the overall fit of a model and is calculated from

$$R^2 = 1 - \frac{SSE}{SST} \quad (4)$$

where $SSE = \sum_i (\Delta G_i^{calc} - \Delta G_i^{obs})^2$ and $SST = \sum_i (\Delta G_i^{obs} - \langle \Delta G^{obs} \rangle)^2$. ΔG_i^{calc} and ΔG_i^{obs} are the calculated and observed binding free energies for ligand i, respectively. An LIE model that reproduces all experimental free energies has an $R^2 = 1.0$. The leave-one-out cross-validated coefficient, Q^2_{loo}, assesses the predictivity of a model and is calculated from

$$Q^2_{loo} = 1 - \frac{PRESS}{SST} \quad (5)$$

where PRESS is the predictive residual sum of squares, $PRESS = \sum_j (\Delta G_j^{calc} - \Delta G_j^{obs})^2$. In this case, ΔG_j^{calc} is calculated from an LIE model optimized on all data points except ligand j.

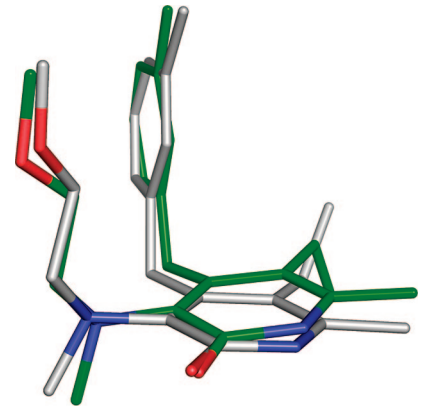


Figure 1. Experimental and docked structure of inhibitor **62**. The experimental and docked structures are shown with white and green carbon atoms, respectively.

3. Results and Discussion

3.1. Docking and Scoring. The docked conformation of inhibitor **62** closely reproduces the crystallographic structure of the complex (rmsd = 0.6 Å, Figure 1). The main difference between the two structures is the orientation of the 5-ethyl group, which also could not be resolved from the experimental electron density.[26] Inhibitor **62** binds to RT in a conformation similar to the NNRTIs 1-[(2-hydroxyethoxy)methyl]-6-(phenylthio)thymine (HEPT[42]) and 6-benzyl-1-(ethoxymethyl)-5-isopropyluracil (MKC-442[43]).[44,45]Furthermore, the obtained docking solutions for a majority of the other 42 inhibitors are similar to the binding mode of inhibitor **62**. The chosen binding modes for four of the inhibitors

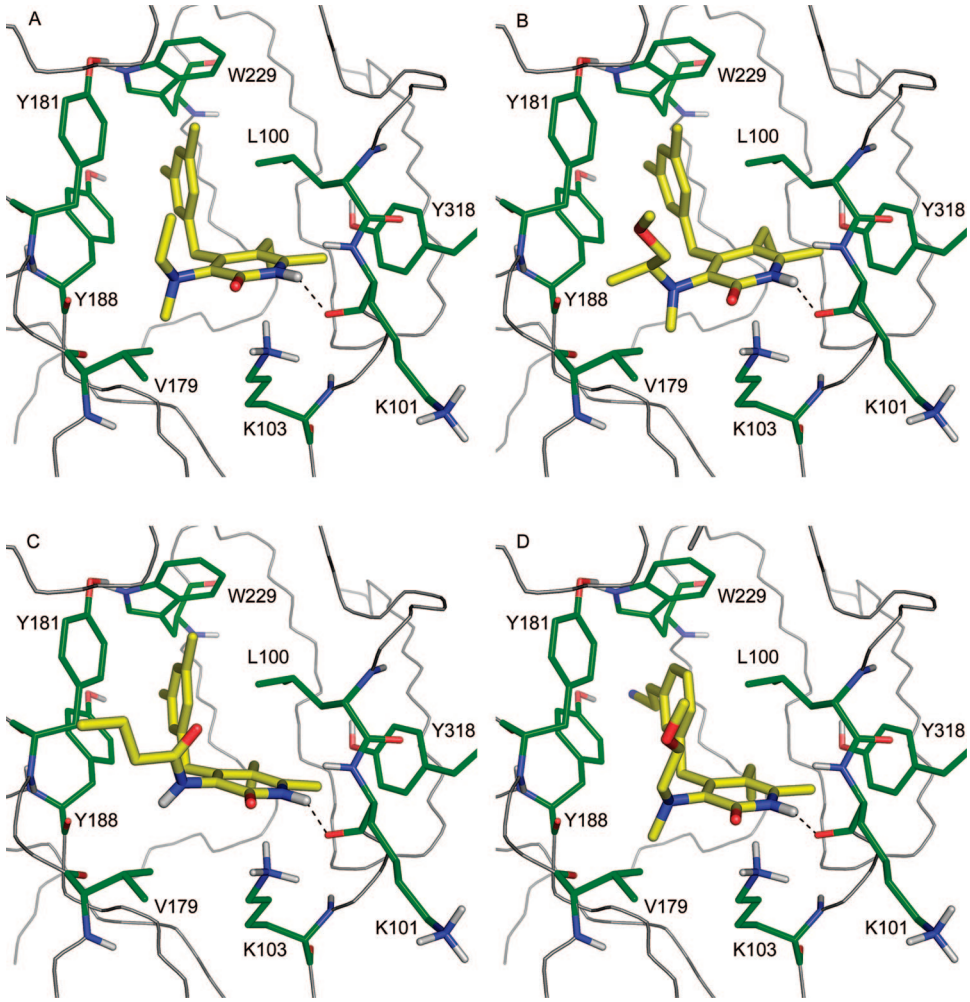


Figure 2. Docking solutions for inhibitors (A) **38**, (B) **40**, (C) **46**, and (D) **77b**.

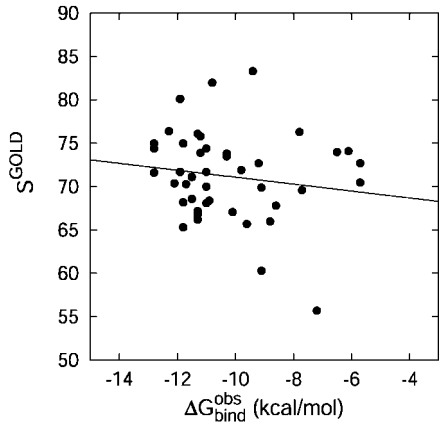


Figure 3. Goldscore (S^{GOLD}) vs experimental binding free energies (ΔG_{bind}^{obs} in kcal/mol). The solid line represents the best fit between S^{GOLD} and ΔG_{bind}^{obs}($r^2 = 0.02$).

(**38**, **40**, **46**, and **77b**) are shown in Figure 2. From the docking calculations, a Goldscore for each inhibitor can be extracted and compared to experimental binding data. The results, based on the average score of the three chosen conformations for each inhibitor, are shown in Figure 3. There is no correlation between Goldscore and the experimental data and these results agree with recently published studies, which indicate that docking algorithms in many cases can identify the correct binding mode but they are not able to rank different inhibitors by affinity.[46]

3.2. Molecular Dynamics Simulations and LIE Binding Free Energies. Calculated ligand-surrounding energies for the inhibitors are shown in Table 2 and are based on averages over all the MD simulations carried out for each inhibitor. For all the investigated LIE models, it was found that excluding inhibitors **36**, **37**, **51**, and **53** resulted in significantly better correlation with experiment, and in all the results presented below, these inhibitors have been excluded from the analysis. The inaccurate predictions made for these four inhibitors may partially be explained by the estimation of experimental free energies of binding from IC_{50} values, which are based on cell-assay experiments and might not only reflect inhibition of RT, but also solubility problems. Another explanation, is that it may be difficult to predict the correct binding modes for inhibitors that are significantly larger (**51**, **53**) or smaller (**36**, **37**) compared to **62**. It can be speculated that the protein structure in complex with some of the studied inhibitors might deviate significantly from the RT−inhibitor **62** complex. For example, Tyr181 has been found to have two completely different conformations for inhibitors similar to those studied here.[26,45] Because protein flexibility is not explicitly taken into account in the docking calculations, this may lead to the prediction of incorrect binding modes for some inhibitors, resulting in difficulties to predict accurate binding free energies.

The results for all the investigated LIE models are summarized in Table 3. The standard parametrization of the LIE method ($\beta = \beta_{FEP}$ and $\alpha = 0.18$) reproduces the experimental results with an average unsigned error of 0.8 kcal/mol ($R^2 =$

Table 2. Ligand-Surrounding Electrostatic ($\langle U_{l\text{-}s}^{el}\rangle$) and van der Waals ($\langle U_{l\text{-}s}^{vdW}\rangle$) Energies (kcal/mol) for the Bound (b) and Free State (f) of the Inhibitors[a]

inhibitor	$\langle U_{l\text{-}s}^{vdW}\rangle_b$	$\langle U_{l\text{-}s}^{vdW}\rangle_f$	$\langle U_{l\text{-}s}^{el}\rangle_b$	$\langle U_{l\text{-}s}^{el}\rangle_f$	ΔG_{bind}^{obs}	ΔG_{bind}^{calc}
9	−51.5 ± 0.2	−28.1 ± 0.0	−18.2 ± 0.3	−27.9 ± 0.1	−11.5	−10.2 ± 0.2
18b	−49.8 ± 0.4	−26.9 ± 0.0	−17.7 ± 0.2	−27.5 ± 0.1	−11.8	−10.1 ± 0.2
18c	−51.7 ± 0.4	−28.4 ± 0.1	−19.4 ± 0.5	−27.0 ± 0.1	−11.3	−11.1 ± 0.4
18d	−55.6 ± 0.1	−30.4 ± 0.1	−18.4 ± 0.3	−26.9 ± 0.1	−10.3	−11.0 ± 0.2
18e	−53.8 ± 0.3	−29.7 ± 0.0	−16.0 ± 0.5	−26.2 ± 0.1	−8.6	−10.1 ± 0.3
18f	−53.5 ± 0.3	−29.5 ± 0.1	−18.0 ± 0.2	−27.3 ± 0.2	−10.3	−10.5 ± 0.2
18g	−45.0 ± 1.0	−24.3 ± 0.1	−12.4 ± 1.4	−26.8 ± 0.1	−7.2	−7.7 ± 0.8
18h	−47.7 ± 0.1	−26.0 ± 0.0	−19.4 ± 0.4	−26.9 ± 0.1	−9.1	−10.9 ± 0.2
18i	−54.3 ± 0.4	−29.2 ± 0.1	−18.7 ± 0.4	−27.5 ± 0.1	−11.3	−10.9 ± 0.3
38	−53.7 ± 0.4	−29.5 ± 0.0	−17.7 ± 0.2	−26.0 ± 0.1	−11.5	−11.0 ± 0.2
39	−55.9 ± 0.3	−30.7 ± 0.0	−18.2 ± 0.2	−26.5 ± 0.2	−11.0	−11.2 ± 0.3
40	−57.9 ± 0.6	−32.0 ± 0.1	−21.7 ± 0.4	−28.8 ± 0.3	−11.7	−11.8 ± 0.4
41	−61.1 ± 0.2	−33.1 ± 0.1	−23.5 ± 0.3	−32.6 ± 0.0	−10.8	−11.3 ± 0.2
42	−56.8 ± 0.4	−31.0 ± 0.2	−22.8 ± 0.3	−30.3 ± 0.1	−12.3	−11.6 ± 0.3
43	−59.7 ± 0.1	−31.8 ± 0.2	−33.4 ± 0.7	−41.6 ± 0.5	−11.9	−12.2 ± 0.5
44	−52.2 ± 0.2	−28.1 ± 0.1	−31.8 ± 0.5	−46.2 ± 0.2	−9.1	−8.3 ± 0.4
45	−55.4 ± 0.3	−29.3 ± 0.1	−27.9 ± 1.3	−46.5 ± 0.3	−7.7	−6.9 ± 0.7
46	−57.6 ± 0.3	−30.9 ± 0.1	−27.5 ± 0.5	−45.6 ± 0.2	−5.7	−7.2 ± 0.4
47	−52.2 ± 0.6	−28.1 ± 0.0	−17.3 ± 0.9	−27.6 ± 0.1	−11.0	−10.1 ± 0.5
48	−55.4 ± 0.2	−29.5 ± 0.2	−16.2 ± 0.2	−27.3 ± 0.1	−10.9	−10.1 ± 0.2
49	−57.2 ± 0.2	−31.9 ± 0.1	−16.1 ± 0.3	−28.7 ± 0.1	−9.8	−9.3 ± 0.2
50	−54.0 ± 0.2	−30.5 ± 0.1	−18.8 ± 0.3	−24.9 ± 0.4	−11.0	−11.8 ± 0.3
52	−59.7 ± 0.1	−33.4 ± 0.3	−19.2 ± 0.4	−30.5 ± 0.4	−9.4	−10.1 ± 0.4
54	−54.5 ± 0.4	−30.5 ± 0.1	−23.4 ± 0.6	−33.2 ± 0.1	−9.6	−10.3 ± 0.4
55	−55.9 ± 0.1	−31.7 ± 0.0	−18.8 ± 0.7	−29.9 ± 0.1	−8.8	−9.8 ± 0.4
56	−55.5 ± 0.2	−30.8 ± 0.0	−21.4 ± 0.3	−30.4 ± 0.1	−11.2	−10.8 ± 0.2
59	−50.2 ± 0.1	−26.2 ± 0.1	−35.1 ± 0.3	−38.7 ± 0.1	−11.8	−13.2 ± 0.2
60	−51.6 ± 0.2	−27.6 ± 0.1	−35.3 ± 0.6	−40.2 ± 0.4	−12.1	−12.7 ± 0.4
61	−57.0 ± 0.2	−30.5 ± 0.1	−33.3 ± 0.2	−40.3 ± 0.5	−11.3	−12.4 ± 0.3
62	−54.4 ± 0.2	−29.7 ± 0.1	−23.0 ± 1.1	−29.0 ± 0.1	−12.8	−12.0 ± 0.6
63	−57.0 ± 0.2	−31.1 ± 0.1	−22.5 ± 0.4	−29.1 ± 0.0	−11.2	−12.0 ± 0.3
64	−50.5 ± 0.3	−27.1 ± 0.0	−24.1 ± 0.4	−33.3 ± 0.1	−11.9	−10.5 ± 0.3
65	−52.3 ± 0.3	−28.2 ± 0.1	−26.9 ± 0.3	−34.8 ± 0.4	−11.0	−11.1 ± 0.4
66	−55.5 ± 0.4	−29.2 ± 0.1	−27.1 ± 1.1	−36.4 ± 0.2	−11.8	−11.0 ± 0.7
67	−55.4 ± 0.2	−29.8 ± 0.1	−32.9 ± 0.2	−49.8 ± 0.2	−6.5	−7.5 ± 0.2
68	−61.1 ± 0.3	−33.6 ± 0.2	−32.4 ± 0.3	−50.6 ± 0.5	−7.8	−7.3 ± 0.5
70	−48.1 ± 0.3	−25.5 ± 0.0	−44.4 ± 2.1	−54.7 ± 0.1	−9.2	−9.8 ± 1.0
77a	−54.5 ± 0.3	−30.5 ± 0.1	−21.3 ± 0.8	−30.2 ± 0.3	−12.8	−10.7 ± 0.6
77b	−58.0 ± 0.2	−32.3 ± 0.2	−26.1 ± 0.3	−34.4 ± 0.3	−12.8	−11.3 ± 0.4

[a] Experimental and calculated LIE (standard parameterization) binding free energies (kcal/mol) are also shown. The naming of the inhibitors has been adopted from ref 36.

Table 3. Comparison of Statistical Figures of Merit for the Five Analyzed LIE Models[a]

model	α	β	⟨lerrorl⟩	rms	R^2	Q_{loo}^2
standard[b]	0.18	β_{FEP}	0.82	0.95	0.70	0.68
A[c]	0.18	β'_{FEP}	0.78	0.93	0.71	0.70
B[d]	**0.20**	**0.43**	0.79	0.93	0.71	0.66
C[e]	0.18	**0.41/0.43**[g]	0.78	0.92	0.72	0.67
D[f]	**0.20**	**0.45**	0.88	1.10	0.60	0.54

[a] The values of the optimized scaling factors are shown in bold. Inhibitors **36**, **37**, **51**, and **53** have been excluded from all models. [b] γ in eq 1 is optimized, $\gamma = -10.2$ kcal/mol. β_{FEP} values are set according to ref 21. [c] γ in eq 1 is optimized, $\gamma = -10.2$ kcal/mol. β'_{FEP} values are set according to ref 41. For the calculations of β'_{FEP} values for inhibitor **67**, **68**, and **70**, it has been assumed that thioureas and amides have the same β'_{FEP} values. [d] α, β, and γ in eq 1 are optimized, $\gamma = -9.7$ kcal/mol. [e] β_b, β_f, and γ in eq 2 are optimized $\gamma = -10.6$ kcal/mol. [f] α, β, and γ in eq 3 are optimized, $\gamma = -9.7$ kcal/mol. [g] β_b/β_f.

0.70). Using the refined β_{FEP} values of Almlöf et al.[41] (model A), where a weighting scheme is used to derive a specific β value for each inhibitor, the correlation with experiment is slightly improved ($R^2 = 0.71$). However, this refined model is likely to become superior when dealing with larger compounds with more functional groups.[41] A similar correlation was found for model B, and the calculated values of α and β in this model, where both coefficients in eq 1 are optimized, are very close to the standard parametrization (α = 0.20, β = 0.43). This result again confirms the robustness of the earlier parametrizations. It should be noted that the relative binding free energies of the studied inhibitors are largely determined by the change in ligand-surrounding electrostatic energies. The nonpolar contribution, calculated from the ligand-surrounding van der Waals energies, is rather similar for the inhibitors and, hence, does not contribute significantly to the relative binding free energies. The finding that the electrostatic term in the LIE method determines the relative binding free energies for this data set also provides a possible explanation for the poor results obtained using Goldscore. The only term representing polar interactions in this scoring function is a simple hydrogen bond term, which may not be a sufficiently accurate description. In conclusion, the standard parametrization of eq 1 can predict the binding free energies of the 39 inhibitors very well, and the results for this model are shown in Figure 4. For comparison, the inhibitors that were excluded from the LIE analysis (**36**, **37**, **51**, and **53**) were also excluded for Goldscore, but this did not improve the results significantly. Previously, Jorgensen and co-workers have carried out an impressive amount of binding free energy calculations on NNRTIs using the extended linear response (ELR) method in conjunction with MC simulations.[47,48] In contrast to the LIE method, a large number (∼15) of different descriptors, e.g., surface area and the ligand's dipole moment, are used in this approach, and in order to make accurate predictions, the ELR equation has to be reparameterized for each receptor.

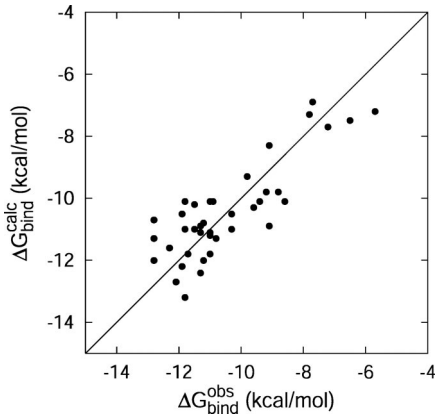


Figure 4. Calculated LIE (ΔG_{bind}^{calc}) vs experimental (ΔG_{bind}^{obs}) binding free energies (kcal/mol). The solid line represents perfect agreement between ΔG_{bind}^{calc} and ΔG_{bind}^{obs}.

To investigate if the results could be further improved by modifying the LIE equation, two alternative forms of eq 1 were analyzed. In the LIE model based on eq 2, different scaling factors for the protein and water environment are used for the electrostatic contribution to binding. Åqvist and co-workers have shown that the approximation for the polar contribution in LIE is excellent for small molecules in aqueous phase.[39,41] In eq 1, the same scaling factors are assumed to be valid for the protein binding site, but it can be argued that the β value for the protein environment (β_b) should be larger than that in aqueous phase (β_f) in many cases, e.g., when the protein is preorganized for ligand binding.[22,49] However, in the standard parametrization of the LIE method, which is based on 18 protein−ligand complexes, separate β values for the protein and water environment did not improve agreement with experiment.[21] Using this alternative, LIE formulation (model C) with $\alpha = 0.18$ and parametrizing β_b and β_f yields scaling factors that are very close to the standard parametrization ($\beta_b = 0.41$ and $\beta_f = 0.43$) and it does not improve agreement with experimental results significantly. Hence, in agreement with the conclusion of Hansson et al.,[21] introducing the complexity of having an extra parameter does not seem to be justified in this case. In eq 3, the use of intramolecular energies in the LIE method is also investigated. The differences in energies taken in eq 1 to estimate binding free energies are based only on the ligand's interactions with its surroundings. The free energy contributions arising from solvent−solvent and intramolecular receptor and ligand interactions are not explicitly included in the LIE method. Instead, given that the linear response approximation holds, the contributions from electrostatic protein−protein and solvent−solvent interactions are taken into account by the scaling factor β.[20,39,50] Although the same argument could be used for intramolecular ligand energies,[50] a recent study of the accuracy of the linear response approximation in water gave some support for also including intramolecular ligand energies in predictions of the electrostatic contribution to the free energy of hydration.[41] However, including intramolecular electrostatic energies (eq 3) and parametrizing α, β, and γ (model D) yields an $R^2 = 0.60$ and an rms deviation of 1.1 kcal/mol, which is significantly worse compared to the original LIE model (eq 1). From the results of models C and D, including intramolecular energies or an additional scaling factor for the electrostatic ligand-surrounding energies, does not improve the accuracy of standard form of the LIE method.

3.3. Which RT Residues are Important for Binding? From the MD simulations, the RT−inhibitor electrostatic and van der Waals interaction energies can be analyzed to identify residues that are important for binding. The average electrostatic interaction energies, over all inhibitors, for the RT residues that contribute significantly to the binding free energy are shown in Figure 5A. Lys101, Lys103, and Glu138 make the largest contributions to the electrostatic component of the average ligand−residue interaction energies and are equal to −6.8, −9.5, and 2.7 kcal/mol, respectively. Furthermore, the standard deviations in ligand−residue electrostatic energies, again over all inhibitors, are the largest for the same three residues. The difference in ligand−water electrostatic interaction energies between the bound and free state also vary widely among the inhibitors. As described in the previous section, the ligand-surrounding electrostatic energy is also the most relevant descriptor of the relative binding free energies. Remarkably, the sum of the electrostatic contributions from Lys101, Lys103, and Glu138, together with the desolvation term, also correlates well with the experimental binding affinities (Figure 6). This implies that these four energetic contributions largely determine the difference in binding affinity between the inhibitors. Structurally, these energetic contributions can be traced back to the strength of the hydrogen bond to the Lys101 backbone carbonyl, the inhibitors' interactions with the side chains of Glu138 and Lys103, and the degree of desolvation upon binding. The obtained correlation between activity and only a few interaction energies suggest that efficient LIE models may be based on only monitoring a subset of the ligand−residue interactions. By introducing empirical scaling factors for each interaction energy, multivariate analysis can be used to identify the energetic contributions from the binding site that reproduce experimental data. The parametrized model can then be used in scoring and lead optimization of inhibitors of unknown activity. Application of this approach, including all inhibitors in Table 1, yields a strong correlation ($R^2 = 0.89$) between calculated and experimental binding affinities and will be presented elsewhere (Carlsson et al., in preparation). The nonpolar free energy contribution, calculated from the change in ligand-surrounding van der Waals energy, is rather similar for all inhibitors and was not found to correlate with the experimental affinities. The average van der Waals interaction energies for the residues that contribute most to the binding free energy are shown in Figure 5B. Leu100, Tyr181, and Tyr188 make the largest contributions to the ligand−residue interaction energies with averages equal to −5.8, −5.0, and −4.7 kcal/mol, respectively. This shows that binding is clearly favored by van der Waals interactions with these residues. However, it should be noted that the van der Waals contribution is not as easily interpreted as the polar term in the LIE method because it does not only take into account the van der Waals interactions but also size-dependent terms such as the hydrophobic effect.[20,51]

The six RT residues that make the three largest van der Waals and electrostatic contributions to binding of the inhibitors are shown in Figure 7. These results agree very well with experiments carried out on three common mutants that are resistant to many NNRTIs: Lys103Asn, Tyr181Cys, and Tyr188Leu.[36] Lys103, Tyr181, and Tyr188 were here found to have either highly favorable electrostatic or van der Waals interactions with the inhibitors, indicating their importance for tight binding and inhibition. Experimentally, the Lys103Asn mutation reduces the binding free energy for the studied inhibitors by approximately 1−3 kcal/mol.[36] This strong charge-mediated hydrogen bond with Lys103 is lost in the mutation to Asn, which leads to a loss of binding affinity. Structures of RT with the Lys103Asn mutation without a bound NNRTI also indicate that Asn103

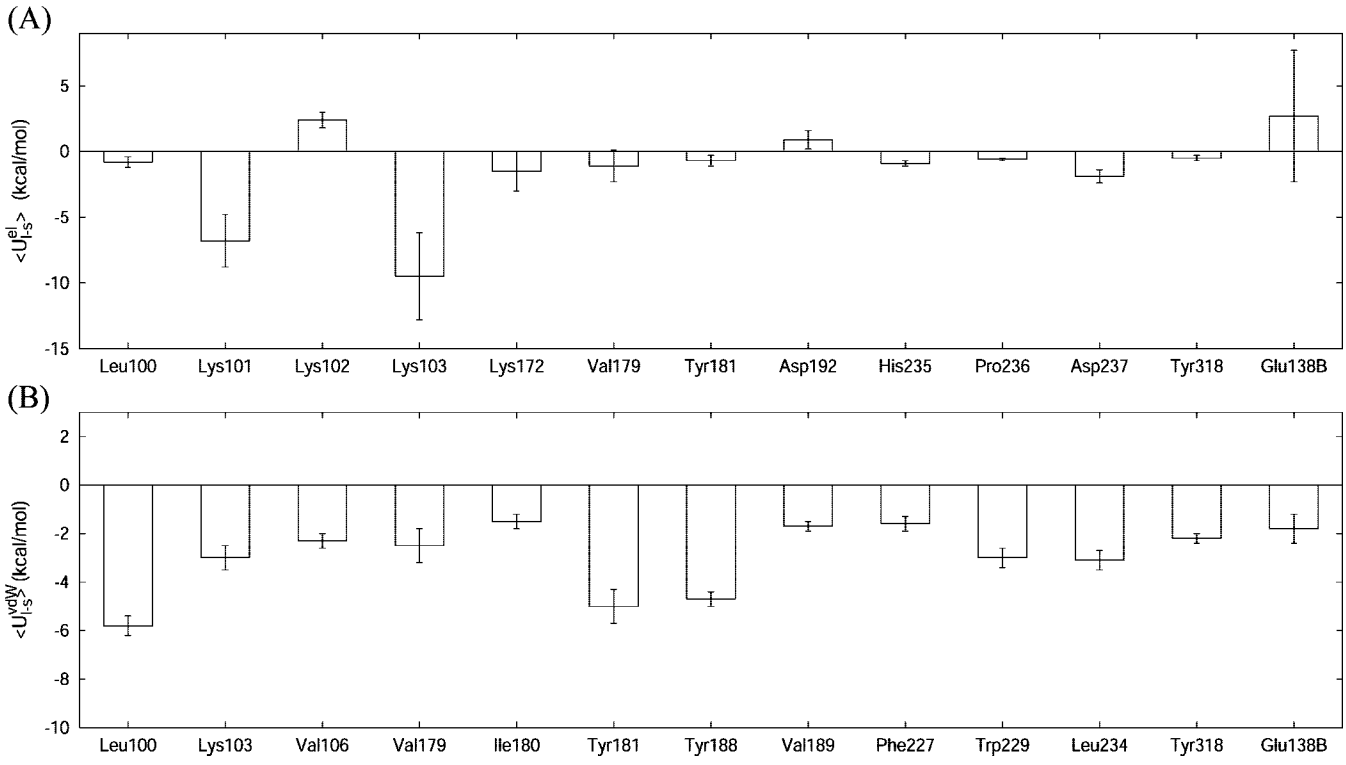


Figure 5. Average ligand−residue interaction energies ($\langle U_{l\text{-}s}^{el/vdW} \rangle$ in kcal/mol) over all inhibitors used in the LIE calculations for the residues that contribute most to the ligand-surrounding (A) electrostatic and (B) van der Waals energies. The average interaction energies are shown using bars, and the corresponding standard deviations are shown using error bars.

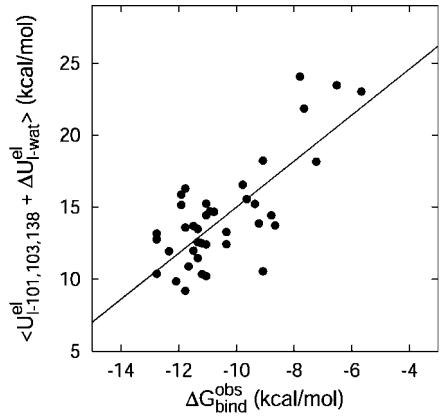


Figure 6. The sum of electrostatic ligand−residue interaction energies from Lys101, Lys103, Glu138, and the change in solvent interactions ($\langle U_{l\text{-}101,103,138}^{el} + \Delta U_{l\text{-}wat}^{el} \rangle$) vs experimental binding free energies (ΔG_{bind}^{obs} in kcal/mol). The solid line represents the best fit between $\langle U_{l\text{-}101,103,138}^{el} + \Delta U_{l\text{-}wat}^{el} \rangle$ and ΔG_{bind}^{obs} ($r^2 = 0.57$).

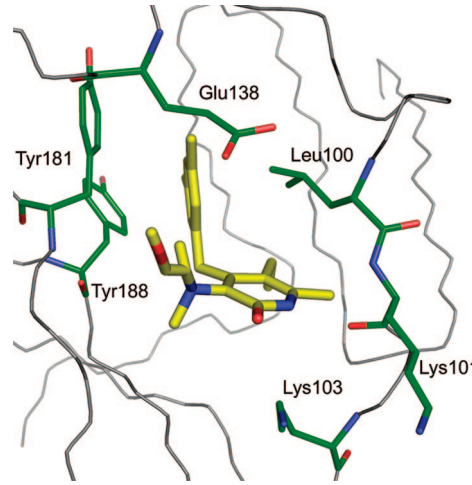


Figure 7. Six RT residues that make the three largest van der Waals and electrostatic contributions to binding for the inhibitors. The binding mode of inhibitor **40** is also shown.

stabilizes the unbound state of RT by forming a hydrogen bond to Tyr188, which could further decrease the binding free energy.[52] The Tyr181Cys and Tyr188Leu mutations make binding of the inhibitors about 1−5 kcal/mol worse compared to wild type RT.[36] Computationally, Tyr181 and Tyr188 give large contributions to the ligand−protein van der Waals interactions. Substituting Tyr for Cys or Leu decreases the favorable aromatic ring-stacking interactions and leads to a reduced binding affinity. Reduced NNRTI binding affinities were also found for mutants of the other three residues in Figure 7.[36] Possible strategies to improve the effectiveness of NNRTIs against resistant strains are to introduce inhibitor flexibility, which gives the possibility to adapt to these mutations, or to target backbone interactions and conserved residues, e.g., Trp229.[53,54]

3.4. Analysis of Inhibitor−RT Interactions. The relative affinities for the different classes of inhibitors in the studied data set are quite well reproduced by the LIE method. For the R_1, R_2-dialkyl substituted inhibitors with an ethyl and methyl group in R_4 and R_5, respectively (**38**, **39**, **50**, **51**, and **55**), binding affinity decreases with increasing chain length. Although the low affinities of inhibitors **51** and **53** were not reproduced by Goldscore or the LIE method, it can be speculated that these bulky substituents affect the strong hydrogen bond to Lys101 and the favorable interactions with the Lys103 side chain amine group. The same trend is also observed for the inhibitors with

secondary amine substituents (**37**, **47**−**49**, and **52**), for which binding affinity decreases with increasing chain length of R_2. The inhibitors with R_1═R_2═CH_3 (**9** and **18b**−**18i**) all have alkyl substituents in R_4 and R_5. On average, these inhibitors become less desolvated upon binding than the other inhibitors and two of the molecules in this series, **9** and **18b**, were experimentally found to be relatively good binders. Inhibitors with polar substituents in R_1 and R_2 lose favorable solvent interactions upon binding. For inhibitors **43** and **59**−**61**, which were found to be relatively active against RT, the desolvation can be counterbalanced by the formation of a hydrogen bond between the carboxylate group of Glu138 and the hydroxyl of the substituents. In other cases, e.g., for the inhibitors with nitrile substituents (**64**−**66**), the substituents are partially in contact with the solvent in the bound state. The amide and thiourea substituents, on the other hand, are more buried in the RT binding site. These inhibitors (**44**−**46**, **67**, **68**, and **70**) lose a considerable amount of solvent interactions upon binding and consequently have very low activities against RT. The experimentally most potent inhibitors, **62** and **77b**, are characterized by strong electrostatic interactions with Lys103 and a small desolvation energy. These inhibitors are also active against the three mutated forms described above: Lys103Asn, Tyr181Cys, and Tyr188Leu. However, because of strong interactions with these residues, the inhibitors are still several kcal/mol less efficient toward the mutants compared to the wild type.[36]

4. Conclusions

The present study investigates the accuracy of docking, scoring, and MD simulations in combination with the LIE method on a set of NNRTIs. The standard form and parametrization of the LIE method accurately reproduces the activities for a majority of the inhibitors, while the results obtained from Goldscore did not correlate with experiment. Furthermore, no significant improvement of the results was obtained when intramolecular electrostatic ligand energies or an additional scaling factor was included in the LIE model. The protein−inhibitor interactions were analyzed in detail, and it was found that the binding free energies of the studied set of NNRTIs can be described using only a few specific inhibitor interactions. The results of this analysis also agree very well with experiments on mutants of RT, which show that identifying the principal interactions for a class of inhibitors can give important directions for further optimization of affinity and for the design of inhibitors that are less susceptible to mutations. The calculations illustrate that, although MD simulations are computationally expensive, the LIE method can provide a useful approach for obtaining accurate predictions of protein−ligand binding free energies also for relatively large data sets. There are also many techniques that can be used to reduce the required computation time for each ligand. In this project, the same amount of computation time was used for each inhibitor, which resulted in several nanoseconds of simulation time for each calculated binding free energy. For a majority of the ligands, the calculations are converged using a small fraction of this computation time and long simulations were actually only required in a few cases. In studies of larger sets of ligands, e.g., in virtual screening applications, computation time can be significantly reduced by monitoring the energy convergence and extending simulations only in cases where it is necessary. Another effective approach to improve convergence is to perform sampling using explicit solvent MD simulations and estimate the electrostatic contributions to binding by postprocessing a series of snapshots of the complex using Poisson−Boltzmann or Langevin dipoles methods.[55,56] With automated generation of force field parameters and MD simulations, the LIE method can now be used to estimate binding affinities for thousands of potential inhibitors[57,58] and, recently, we have applied this approach to rank 1000 molecules against 1-deoxy-D-xylulose-5-phosphate reductoisomerase, which is a potential drug target in the development of drugs against tuberculosis (Carlsson et al., unpublished). These examples clearly show that force field based methods to estimate binding free energies have become powerful tools that successfully can be used in computed-aided drug design.

Acknowledgment. Support from the Swedish Research Council (VR) and the Swedish Foundation for Strategic Research (SSF/Rapid) is gratefully acknowledged.

References

(1) Esnouf, R.; Ren, J. S.; Ross, C.; Jones, Y.; Stammers, D.; Stuart, D. Mechanism of Inhibition of HIV-1 Reverse-Transcriptase by Non-nucleoside Inhibitors. *Nat. Struct. Biol.* **1995**, *2*, 303–308.
(2) De Clercq, E. The role of non-nucleoside reverse transcriptase inhibitors (NNRTIs) in the therapy of HIV-1 infection. *Antiviral Res.* **1998**, *38*, 153–179.
(3) De Clercq, E. New approaches toward anti-HIV chemotherapy. *J. Med. Chem.* **2005**, *48*, 1297–1313.
(4) Jones, G.; Willett, P.; Glen, R. C. Molecular Recognition of Receptor-Sites Using a Genetic Algorithm with a Description of Desolvation. *J. Mol. Biol.* **1995**, *245*, 43–53.
(5) Jones, G.; Willett, 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.
(6) Ewing, T. J. A.; Makino, S.; Skillman, A. G.; Kuntz, I. D. DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 411–428.
(7) 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.
(8) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **2004**, *47*, 1750–1759.
(9) Gohlke, H.; Klebe, G. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angew. Chem., Int. Ed.* **2002**, *41*, 2645–2676.
(10) Böhm, H. J. The Development of a Simple Empirical Scoring Function to Estimate the Binding Constant for a Protein−Ligand Complex of Known 3-Dimensional Structure. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 243–256.
(11) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. Empirical scoring functions.1. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 425–445.
(12) 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.
(13) Gohlke, H.; Hendlich, M.; Klebe, G. Knowledge-based scoring function to predict protein−ligand interactions. *J. Mol. Biol.* **2000**, *295*, 337–356.
(14) Beveridge, D. L.; DiCapua, F. M. Free Energy via Molecular Simulation: Applications to Chemical and Biomolecular Systems. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431–492.
(15) Kollman, P. Free-Energy Calculations: Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* **1993**, *93*, 2395–2417.
(16) Tembe, B. L.; McCammon, J. A. Ligand−Receptor Interactions. *Comput. Chem. (Oxford)* **1984**, *8*, 281–283.
(17) Hansson, T.; Åqvist, J. Estimation of binding free energies for HIV proteinase inhibitors by molecular dynamics simulations. *Protein Eng.* **1995**, *8*, 1137–1144.
(18) Åqvist, J. Calculation of absolute binding free energies for charged ligands and effects of long-range electrostatic interactions. *J. Comput. Chem.* **1996**, *17*, 1587–1597.
(19) Lamb, M. L.; Jorgensen, W. L. Computational approaches to molecular recognition. *Curr. Opin. Chem. Biol.* **1997**, *1*, 449–457.
(20) Åqvist, J.; Medina, C.; Samuelsson, J. E. New Method for Predicting Binding Affinity in Computer-Aided Drug Design. *Protein Eng.* **1994**, *7*, 385–391.
(21) Hansson, T.; Marelius, J.; Åqvist, J. Ligand binding affinity prediction by linear interaction energy methods. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 27–35.

(22) Almlöf, M.; Brandsdal, B. O.; Åqvist, J. Binding affinity prediction with different force fields: Examination of the linear interaction energy method. *J. Comput. Chem.* **2004**, *25*, 1242–1254.
(23) Marelius, J.; Graffner-Nordberg, M.; Hansson, T.; Hallberg, A.; Åqvist, J. Computation of affinity and selectivity: Binding of 2,4-diaminopteridine and 2,4-diaminoquinazoline inhibitors to dihydrofolate reductases. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 119–131.
(24) Ljungberg, K. B.; Marelius, J.; Musil, D.; Svensson, P.; Norden, B.; Åqvist, J. Computational modelling of inhibitor binding to human thrombin. *Eur. J. Pharm. Sci.* **2001**, *12*, 441–446.
(25) Ersmark, K.; Feierberg, I.; Bjelic, S.; Hamelink, E.; Hackett, F.; Blackman, M. J.; Hulten, J.; Samuelsson, B.; Åqvist, J.; Hallberg, A. Potent inhibitors of the *Plasmodium falciparum* enzymes plasmepsin I and II devoid of cathepsin D inhibitory activity. *J. Med. Chem.* **2004**, *47*, 110–122.
(26) Himmel, D. M.; Das, K.; Clark, A. D.; Hughes, S. H.; Benjahad, A.; Oumouch, S.; Guillemont, J.; Coupa, S.; Poncelet, A.; Csoka, I.; Meyer, C.; Andries, K.; Nguyen, C. H.; Grierson, D. S.; Arnold, E. Crystal structures for HIV-1 reverse transcriptase in complexes with three pyridinone derivatives: A new class of non-nucleoside inhibitors effective against a broad range of drug-resistant strains. *J. Med. Chem.* **2005**, *48*, 7582–7591.
(27) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision C.02; Gaussian, Inc.: Wallingford, CT, 2004.
(28) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.
(29) Marelius, J.; Kolmodin, K.; Feierberg, I.; Åqvist, J. Q. A molecular dynamics program for free energy calculations and empirical valence bond simulations in biomolecular systems. *J. Mol. Graphics Modell.* **1998**, *16*, 213–225.
(30) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
(31) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: the RESP Model. *J. Phys. Chem.* **1993**, *97*, 10269–10280.
(32) Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R. W.; Klein, M. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.
(33) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of *n*-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.
(34) King, G.; Warshel, A. A Surface Constrained All-Atom Solvent Model for Effective Simulations of Polar Solutions. *J. Chem. Phys.* **1989**, *91*, 3647–3661.
(35) Lee, F. S.; Warshel, A. A Local Reaction Field Method for Fast Evaluation of Long-Range Electrostatic Interactions in Molecular Simulations. *J. Chem. Phys.* **1992**, *97*, 3100–3107.
(36) Benjahad, A.; Croisy, M.; Monneret, C.; Bisagni, E.; Mabire, D.; Coupa, S.; Poncelet, A.; Csoka, I.; Guillemont, J.; Meyer, C.; Andries, K.; Pauwels, R.; de Bethune, M. P.; Himmel, D. M.; Das, K.; Arnold, E.; Nguyen, C. H.; Grierson, D. S. 4-benzyl- and 4-benzoyl-3-dimethylaminopyridin-2 (1H)-ones: In vitro evaluation of new c-3-amino-substituted and c-5,6-alkyl-substituted analogues against clinically important HIV mutant strains. *J. Med. Chem.* **2005**, *48*, 1948–1964.
(37) Cheng, Y.; Prusoff, W. H. Relationship between Inhibition Constant (K1) and Concentration of Inhibitor Which Causes 50% Inhibition (I50) of an Enzymatic Reaction. *Biochem. Pharmacol.* **1973**, *22*, 3099–3108.
(38) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Improved protein−ligand docking using GOLD. *Proteins: Struct., Funct., Genet.* **2003**, *52*, 609–623.
(39) Åqvist, J.; Hansson, T. On the validity of electrostatic linear response in polar solvents. *J. Phys. Chem.* **1996**, *100*, 9512–9521.
(40) Ben-Naim, A.; Marcus, Y. Solvation thermodynamics of nonionic solutes. *J. Chem. Phys.* **1984**, *81*, 2016–2027.
(41) Almlöf, M.; Carlsson, J.; Åqvist, J. Improving the Accuracy of the Linear Interaction Energy Method for Solvation Free Energies. *J. Chem. Theory Comput.* **2007**, *3*, 2162–2175.
(42) Baba, M.; Tanaka, H.; Declercq, E.; Pauwels, R.; Balzarini, J.; Schols, D.; Nakashima, H.; Perno, C. F.; Walker, R. T.; Miyasaka, T. Highly Specific-Inhibition of Human Immunodeficiency Virus Type-1 by a Novel 6-Substituted Acyclouridine Derivative. *Biochem. Biophys. Res. Commun.* **1989**, *165*, 1375–1381.
(43) Yuasa, S.; Sadakata, Y.; Takashima, H.; Sekiya, K.; Inouye, N.; Ubasawa, M.; Baba, M. Selective and Synergistic Inhibition of Human Immunodeficiency Virus Type-1 Reverse-Transcriptase by a Non-nucleoside Inhibitor, MKC-442. *Mol. Pharmacol.* **1993**, *44*, 895–900.
(44) Ren, J. S.; Esnouf, R.; Garman, E.; Somers, D.; Ross, C.; Kirby, I.; Keeling, J.; Darby, G.; Jones, Y.; Stuart, D.; Stammers, D. High-Resolution Structures of HIV-1 RT from 4 RT-Inhibitor Complexes. *Nat. Struct. Biol.* **1995**, *2*, 293–302.
(45) Hopkins, A. L.; Ren, J. S.; Esnouf, R. M.; Willcox, B. E.; Jones, E. Y.; Ross, C.; Miyasaka, T.; Walker, R. T.; Tanaka, H.; Stammers, D. K.; Stuart, D. I. Complexes of HIV-1 reverse transcriptase with inhibitors of the HEPT series reveal conformational changes relevant to the design of potent non-nucleoside inhibitors. *J. Med. Chem.* **1996**, *39*, 1589–1600.
(46) 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.
(47) Rizzo, R. C.; Tirado-Rives, J.; Jorgensen, W. L. Estimation of binding affinities for HEPT and nevirapine analogues with HTV-1 reverse transcriptase via Monte Carlo simulations. *J. Med. Chem.* **2001**, *44*, 145–154.
(48) Rizzo, R. C.; Udier-Blagovic, M.; Wang, D. P.; Watkins, E. K.; Smith, M. B. K.; Smith, R. H.; 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.
(49) Sham, Y. Y.; Chu, Z. T.; Tao, H.; Warshel, A. Examining methods for calculations of binding free energies: LRA, LIE, PDLD-LRA, and PDLD/S-LRA calculations of ligands binding to an HIV protease. *Proteins: Struct., Funct., Genet.* **2000**, *39*, 393–407.
(50) Åqvist, J.; Marelius, J. The linear interaction energy method for predicting ligand binding free energies. *Comb. Chem. High Throughput Screening* **2001**, *4*, 613–626.
(51) Carlsson, J.; Åqvist, J. Calculations of solute and solvent entropies from molecular dynamics simulations. *Phys. Chem. Chem. Phys.* **2006**, *8*, 5385–5395.
(52) Hsiou, Y.; Ding, J. P.; Das, K.; Clark, A. D.; Boyer, P. L.; Lewi, P.; Janssen, P. A. J.; Kleim, J. P.; Rosner, M.; Hughes, S. H.; Arnold, E. The Lys103Asn mutation of HIV-1 RT: A novel mechanism of drug resistance. *J. Mol. Biol.* **2001**, *309*, 437–445.
(53) Ohtaka, H.; Velazquez-Campoy, A.; Xie, D.; Freire, E. Overcoming drug resistance in HIV-1 chemotherapy: the binding thermodynamics of Amprenavir and TMC-126 to wild-type and drug-resistant mutants of the HIV-1 protease. *Protein Sci.* **2002**, *11*, 1908–1916.
(54) Das, K.; Clark, A. D.; Lewi, P. J.; Heeres, J.; de Jonge, M. R.; Koymans, L. M. H.; Vinkers, H. M.; Daeyaert, F.; Ludovici, D. W.; Kukla, M. J.; De Corte, B.; Kavash, R. W.; Ho, C. Y.; Ye, H.; Lichtenstein, M. A.; Andries, K.; Pauwels, R.; de Bethune, M. P.; Boyer, P. L.; Clark, P.; Hughes, S. H.; Janssen, P. A. J.; Arnold, E. Roles of conformational and positional adaptability in structure-based design of TMC125-R165335 (etravirine) and related non-nucleoside reverse transcriptase inhibitors that are highly potent and effective against wild-type and drug-resistant HIV-1 variants. *J. Med. Chem.* **2004**, *47*, 2550–2560.
(55) Carlsson, J.; Andér, M.; Nervall, M.; Åqvist, J. Continuum solvation models in the linear interaction energy method. *J. Phys. Chem. B* **2006**, *110*, 12034–12041.
(56) Xiang, Y.; Oelschlaeger, P.; Florian, J.; Goodman, M. F.; Warshel, A. Simulating the effect of DNA polymerase mutations on transition-state energetics and fidelity: Evaluating amino acid group contribution and allosteric coupling for ionized residues in human pol beta. *Biochemistry* **2006**, *45*, 7036–7048.
(57) Huang, D. Z.; Luthi, U.; Kolb, P.; Cecchini, M.; Barberis, A.; Caflisch, A. In silico discovery of beta-secretase inhibitors. *J. Am. Chem. Soc.* **2006**, *128*, 5436–5443.
(58) Stjernschantz, E.; Marelius, J.; Medina, C.; Jacobsson, M.; Vermeulen, N. P. E.; Oostenbrink, C. Are automated molecular dynamics simulations and binding free energy calculations realistic tools in lead optimization? An evaluation of the linear interaction energy (LIE) method. *J. Chem. Inf. Model.* **2006**, *46*, 1972–1983.

JM7012198