Predicting Binding Modes from Free Energy Calculations

Martin Nervall, Peter Hanspers, Jens Carlsson, Lars Boukharta, and Johan Åqvist*

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

Received September 26, 2007

To produce reliable predictions of bioactive conformations is a major challenge in the field of structure-based inhibitor design and is a requirement for accurate binding free energy predictions with structure-based methods. A series of HIV-1 reverse transcriptase inhibitors was cross-docked using a non-native crystal structure that resulted in two distinct clusters of possible conformations. One of these clusters was compatible with an existing crystal structure, whereas the other displayed a flipped heterocyclic group. Binding free energies, using the non-native crystal structure, calculated from several scoring functions, were similar for the two clusters, and no conclusion about the binding mode could be drawn from these results. The two clusters could be separated through rescoring with the linear interaction method (LIE) in combination with molecular dynamics simulations, which leads to a binding mode prediction in line with experimental crystallographic data. Further, the LIE model produces the best correlation between experimental and calculated binding free energies among the tested scoring methods.

Introduction

One of the great challenges in computer-aided ligand design is accurate predictions of binding free energies. The collection of methods to perform such predictions falls in two main categories, nonstructure-based and structure-based, and this work will deal with the latter type of approach. The structure-based methods range from theoretically accurate, but very time-consuming, methods like free energy perturbation,[1] to more approximate and faster methods such as knowledge-based and empirical scoring functions.[2–4] Here, we will focus on the faster range of these methods, which are used in lead discovery and lead optimization efforts. All structure-based affinity prediction methods ultimately rely on a three-dimensional model of a receptor in complex with the ligand of interest, and it is thus essential to produce proper models of the bioactive complex between receptor and ligand for high accuracy in the predictions. A common strategy when the structure of the receptor is known from, e.g., X-ray crystallography or NMR, is to use automated docking to produce the necessary models of the protein−ligand complexes for binding free energy calculations.[5] In this work, the impact of using different model structures of the protein is investigated by comparing affinity predictions from different X-ray crystal structures. This is of particular interest in the drug design process when novel compounds are evaluated in a receptor with known structure but where the receptor was determined in complex with a ligand from another compound class. In most such cases, the existing structural data will show the binding of compounds that are not closely related to the novel compounds, which may cause difficulties in determining the bioactive conformations of the latter. Without accurate models of the receptor−ligand complexes, the result from structure-based affinity prediction methods are often irrelevant.

Two different structures of HIV-1 reverse transcriptase (HIV-RT[a]) are used here to assess the importance of the protein model in a docking and scoring approach. Thus, what is the impact of cross-docking on affinity prediction? Initially, a set of scoring functions are evaluated for their capacity to rank a set of ligands according to affinity. The conventional use of scoring functions is to score a single receptor−ligand complex for each compound, and in addition to that, the impact of using scoring functions in combination with conformational sampling is evaluated. Such sampling can be done by, e.g., molecular dynamics (MD) or Monte Carlo methods. The purpose is to generate an ensemble of thermodynamically accessible conformations of the ligand and from this ensemble calculate averages of the affinity estimates. The concept of scoring ensembles has been investigated previously,[6–8] and this idea is further examined to assess whether sampling can improve the relative ranking of the inhibitors. In addition, the results from the scoring functions are compared to a more rigorous technique, the linear interaction energy method (LIE).[9] The scoring functions and the LIE method are evaluated to assess whether conformational sampling can eliminate false positive conformations from a set of docking solutions with very similar scores. Thus, can the sampling help us discriminate between putative binding modes? The problem with several possible binding modes may arise in a situation where a docking program suggests two or more clusters of solutions for an inhibitor. The proposed conformations have comparable scores from the docking, and manual inspection of the binding modes gives no structural rationale to rank one ahead of the other. In an attempt to find a suitable method to discriminate between binding modes, a test case was pursued with the following properties: (i) a series of ligands with available binding affinity data, (ii) structural data from the ligand series must be available to confirm the predicted bioactive conformations of the ligands, (iii) affinities of the ligands in the series should span several magnitudes, and (iv) the ligands should be easy to dock and, at the same time, present two or more distinct binding modes with similar scores.

A system possessing all of the aforementioned features is provided by HIV type 1 reverse transcriptase. The protein has been extensively studied with an abundance of structures available[10–13] and with numerous inhibitor series with corresponding binding affinity data.[14–17] Inhibitors of HIV-RT are classified into two groups: the nucleoside and the non-nucleoside RT inhibitors (NNRTIs). The NNRTIs are generally small, rigid

* Corresponding author. Phone: +46-(0)18-471 41 09. Fax: +46-(0)18-53 69 71. E-mail: aqvist@xray.bmc.uu.se.

[a] Abbreviations: LIE, linear interaction energy; RT, reverse transcriptase; NNRTI, non-nucleoside reverse transcriptase inhibitors; GGS, GOLD Goldscore function; GCS, GOLD Chemscore function; QCS, Q Chemscore function.

10.1021/jm701218j CCC: $40.75
Published on Web 04/12/2008

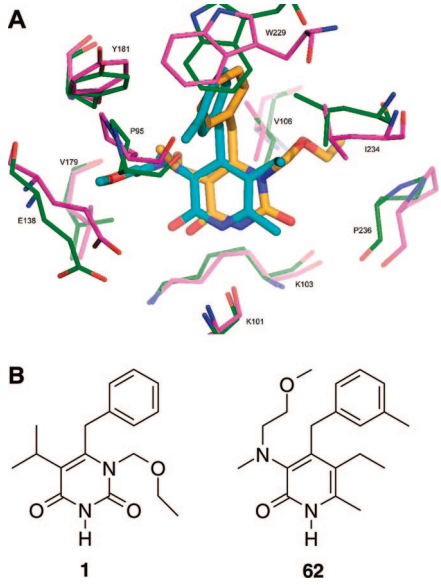


Figure 1. (A) Superposition of the NNRTI binding site in the 1RT1 and 2BAN crystal models of HIV-RT. The protein and ligand (**1**) in 1RT1 are represented by magenta and orange sticks, respectively. 2BAN and its corresponding ligand, **62**, are show in green and cyan. The major difference between the two ligands is the two ether moieties protruding from the pyridinone core in opposite directions. This difference is reflected in the two protein structures, where V106 and P236 are moved apart in the 1RT1 structure and E138 and P95 are moved apart in 2BAN. (B) Structures of the ligands from the 1RT1 and 2BAN crystal structures, respectively.

molecules that are fast to dock, and some of these may be docked in two conformations related by a 180° rotation of a ring moiety. In a study performed by Benjahad et al.,[14] the authors hypothesized two different binding modes of an inhibitor series before determining the correct conformation by X-ray crystallography (Figure 1). The X-ray structure was deposited in the Protein Data Bank (PDB) with code 2BAN, and contains HIV-RT in complex with compound **62** (see Table 1).[10] The compounds in the series studied by Benjahad et al. are chemically related to the HIV-RT inhibitor **1** (MKC-442, Emivirine),[18] whose binding conformation was determined in the crystal structure with PDB code 1RT1.[18] These two compounds bind in analogous conformations, which ensures major shape complementarity of the binding site in the two structures. Figure 1 shows the similarities of **62** and **1**, together with a superposition of the two ligands in their respective crystal structure. It is evident from the figure that the binding sites of the 1RT1 and 2BAN protein models are very similar, yet they contain a few important differences. The pocket is wider in 1RT1 on one side to accommodate the methyl-ethylether substituent on the pyridinone ring of **1**. This is accomplished by an increase in the distance between Pro236 and Val106 from 5.4 Å in 2BAN to 7.0 Å in 1RT1. On the other side of the pocket, the situation is inverted; in 2BAN, the pocket is widened compared to 1RT1 to accommodate the methyl-ethylether moiety of **62**. This is accomplished by shifts of Glu138 and Pro95, together with neighboring residues, of over 1.4 Å.

The ligand series published by Benjahad et al. and the two crystal structures, 2BAN and 1RT1, will serve here as a test case to evaluate different methods to distinguish between putative binding modes. The ligands were previously docked and LIE scored in 2BAN with good results,[19] and the focus here will hence be on the cross-docking and scoring with the 1RT1 model. Initially, the ligands are docked in the two structures, and the results are analyzed for a consensus binding mode among the obtained solutions. Next, the docking solutions are scored with a variety of scoring functions, and the binding free energies are correlated to experimental IC_{50} values and to molecular weight.

Methods

A set of 34 HIV-RT inhibitors[14] were docked using GOLD3.0[20,21] in two crystal structures with PDB codes 2BAN and 1RT1. The studied inhibitors are a series of benzylpyridinone derivatives, which are shown in Table 1. The binding site was defined using GOLD's flood fill option in a 10 Å radius sphere around one hydrogen bound to Cδ1 in Leu100 pointing into the cavity. Each ligand was docked 20 times, and the best 15 conformations were saved. The maximum number of operations in each docking was set to 200000, and the remaining parameters were kept at their default values. Scores from the Goldscore scoring function were extracted from the dockings, as indicated by instructions on the Cambridge Crystallographic Data Centre Web site, where the strongest correlation with relative free energies of binding is achieved by scaling the contributions according to $S^{\text{hb-ext}} + 1.375 \times S^{\text{vdW-ext}}$.[22] For the Chemscore function implemented in GOLD, the estimated free energy of binding was extracted directly from the "Gold.Chemscore.DG" feature in the corresponding output file. Relative binding free energy estimates were compared to experimentally measured IC_{50} values[14] through the equation

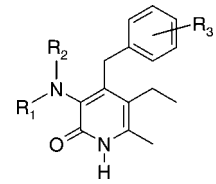
$$\Delta G_{\text{obs}} = RT\ln(IC_{50}) + c \quad (1)$$

where c is a constant[23] ($c = -RT \ln[1 + S/K_M]$) that does not affect the relative free energies because the IC_{50}'s were measured using identical assays.

Additional binding free energy estimates were calculated using three scoring functions implemented in the molecular dynamics package Q,[24] Chemscore, X-score, and PMF. The three scoring functions are described in detail elsewhere.[25–27] Goldscore, Chemscore, and X-score are empirical scoring functions that are linear functions of separate contributions (e.g., h-bonds, van der Waals). The weights on the different contributions have been determined by linear regression. PMF is a knowledge-based scoring function, which in contrast to empirical scoring functions has been derived from crystallographic data and does not rely on fitting to observed affinities.[27] In PMF, radial distribution functions were derived for atom type pairs from the crystallographic data and distance dependent interaction energy functions were calculated from the distributions.

Molecular dynamics simulations and LIE calculations were performed with the software package Q.[24] The MD simulations were conducted in a 18 Å radius sphere, centered on the backbone carbonyl of Lys101. Ionizable residues close to the ligand were charged, and residues close to the boundary edge were made neutral. The charged residues were Lys101, Lys102, Lys103, Lys104, Glu138, Asp237, Asp192, and Glu194, which left the total net charge of the sphere to be zero. The sphere was solvated with TIP3P water[28] and slowly heated to 310 K. The system was equilibrated for 150 ps at 310 K to allow full relaxation of the system before snapshot collection. The collection phase was 650 ps with a 1.5 fs time step, and snapshots of the configuration was saved every 1.5 ps. Water molecules at the sphere boundary were restrained to mimic the dipole distribution of bulk water through the SCAAS model,[29] and the SHAKE algorithm was used to constrain the geometry of the water molecules and solute bonds involving hydrogens.[30] The nonbonded cutoff was set to 10 Å for all atoms, with the exception of ligand atoms, for which no cutoff was applied. Beyond the cutoff, long-range electrostatic interactions were treated using the local reaction field (LRF) multipole expansion approximation.[31] Simulations of the ligands in the free state were performed under similar conditions in a sphere of water.[19] Partial charges were assigned according to the OPLS-AA force field,[32] with the exception of the R_1 and R_2 groups in compounds **67**, **68**, and **70**. As no OPLS-

Table 1. NNRTIs Used in This Work and Their Experimental IC_{50} Values (μM)[14a]



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

[a] The naming of the inhibitors has been adopted from ref 14.

AA charges were available for these groups, they were derived using the RESP scheme.[33]

The LIE method was used to calculate binding free energies from the MD simulations.[9,34] The method is a semiempirical scoring function based on linear free energy relations. Ligand-surrounding energies (l-s) were extracted from the collection phase MD trajectories, and binding free energies were calculated according to

$$\Delta G_{calc} = \alpha\Delta\langle V_{l\text{-}s}^{vdW}\rangle + \beta\Delta\langle V_{l\text{-}s}^{el}\rangle + \gamma \qquad (2)$$

where $\langle\ \rangle$ denotes averages of the van der Waals (vdW) and electrostatic (el) interaction energies. The Δ's represent the difference of the average interaction energies between the free and the bound states. The parameter α has been empirically determined to 0.18,[34] and β is a theoretically derived parameter that varies depending of the chemical nature of the ligand; for neutral ligands containing no hydroxyl groups, $\beta = 0.43$ was used, and for neutral ligands containing one hydroxyl group, $\beta = 0.37$ was used.[34] The constant offset γ was set to minimize the rms error between the experimental and the calculated relative binding free energies estimated for the *correct* conformations. The same set of parameters was used to calculate binding free estimates for the *correct* and *wrong* conformations.

Performances of the different scoring functions were assessed by comparing the Pearson's correlation coefficient of the experimental and calculated binding free energies. The Pearson coefficient is calculated as

$$r = \frac{\sum_i (\Delta G_i^{obs} - \Delta\overline{G}_i^{obs})(\Delta G_i^{calc} - \Delta\overline{G}_i^{calc})}{\sqrt{\sum_i (\Delta G_i^{obs} - \Delta\overline{G}_i^{obs})^2 \sum_i (\Delta G_i^{calc} - \Delta\overline{G}_i^{calc})^2}} \qquad (3)$$

where $\Delta\overline{G}^{obs}$ and $\Delta\overline{G}^{calc}$ represent the mean values of the observed and calculated binding free energies, respectively. The square of the Pearson coefficient is a measure of the fraction of the variance explained by the model relative to the total variance. In the Results Section, r values will be reported to account for negative correlations because r^2 does not provide information about the sign of the relationship. By using the Pearson coefficient as a quality measure, we allow a scoring function to predict relative values. Thus, a perfect correlation can be achieved, but with the wrong slope. In practice, this means that linear regression is performed on the calculated binding free energies to fit a line, $\Delta\overline{G}^{obs} = a\,\Delta\overline{G}_{calc} + b$. If strong correlation is achieved with $a \neq 1$, it can be viewed as a measure of the best correlation that the model could achieve if it would be reparameterized with an external variable on the given data. The scoring functions described here should give positive correlation with experimental data except Goldscore, which was designed to produce negative correlation (i.e., higher score means more negative binding free energy).

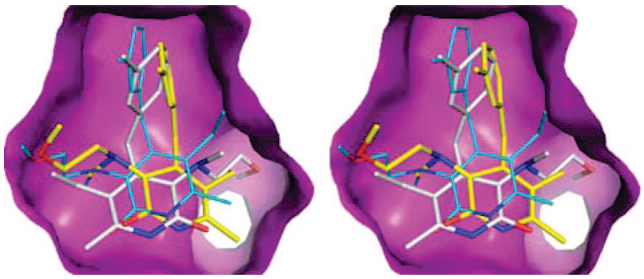


Figure 2. Figure shows the *correct* (yellow) and *wrong* (white) docked conformations of **62** together with the crystal structure conformation (cyan) of **62**, in wall-eyed stereo. In magenta, the 1RT1 X-ray structure of HIV reverse transcriptase is displayed in surface representation, with residues in the foreground left out for clarity. The *correct* and *wrong* docking solutions were collected from dockings into the 1RT1 crystal structure. Note the resemblance of the *wrong* conformation (white) with the conformation of the native ligand in 1RT1, **1** (orange molecule in Figure 1).

During the data analysis, different r values were estimated for varying number of data points, creating the need to account for each r value's significance separately. The significance of each r value was estimated by performing a null hypothesis t test, where the t value was calculated from

$$t = r\sqrt{\frac{n-2}{1-r^2}} \tag{4}$$

Here, n is the number of degrees of freedom, in this case the number of data points. The null hypothesis was tested at the confidence level of $\alpha_{conf} = 0.01$ for all r values. The lowest value of r at any given size of the data set can be estimated by

$$r = \sqrt{\frac{t^2}{(n-2+t^2)}} \tag{5}$$

where t is the Student's t test value at $n - 2$ degrees of freedom and n is the number of samples.

Results

Docking in Two Crystal Structures. The set of 34 chemically related HIV-RT inhibitors[14] were cross-docked into the X-ray model 1RT1,[18] with the ligand and waters removed. Two distinct clusters (binding modes) were found by the docking program, related by a 180° flip (see Figure 2). The docking program was not able to provide a consensus solution valid for all the ligands in the series regarding the binding mode. Both the Goldscore and Chemscore fitness functions implemented in GOLD3.0 were used, and neither of them was able to present a consensus solution. The two modes will hereafter be referred to as the *correct* and the *wrong* modes (see Figure 2), depending on the alignment with the ligand of the crystal structure with PDB code 2BAN. The 2BAN structure is an X-ray model of HIV-RT cocrystallized with one of the ligands in the docked set, **62**, and the assumption is made that these chemically closely related ligands will bind in a conserved mode. Hence, the mode that superimposes with the ligand in 2BAN is called *correct*, and the flipped mode is called *wrong*. The assumption about a conserved binding mode finds support in a number of published structures of HIV-RT, where a carbonyl oxygen is oriented in the direction of Glu138, just as in the *correct* mode (e.g., 2BE2, 2B5J, 1TKT, 1TKX, 1FK9, 1TKZ).[10–13] In the *wrong* mode, the pyridinone ring is flipped so that the carbonyl group is directed toward Pro236. There are a few cases where there is a carbonyl group in that position, but in all of those cases, there is a carbonyl group oriented toward Glu138 as well (e.g., 1RT1). The most convincing example of the conserved binding mode is found in the structure with PDB code 2BE2, in which the ligand is closely related to 2BAN. The pyridinone and the (di)methylphenyl groups in 2BAN and 2BE2 are identical, but they extend large substituents from the pyridinone ring in opposite directions; in 2BAN, a large substituent is directed toward Glu138, and in 2BE2, a large substituent is placed toward Pro236 in a similar manner to **1** in 1RT1 (see Figure 1). Yet, the pyridinone rings of the 2BAN and 2BE2 ligands are aligned in the same orientation. Thus, even though the ligands in 2BAN and 2BE2 have quite dissimilar substituents, the cores superimpose perfectly in the two crystal structures.

Initially the ability of GOLD to find the accurate consensus binding mode was assessed. When we compare the resulting conformations from the Goldscore and Chemscore fitness functions in the 1RT1 dockings, we notice slightly better results from the Goldscore function (Table 2). Goldscore ranks the *correct* mode ahead of the *wrong* for 50% of the ligands compared to only 20% for Chemscore. Clearly, these dockings are too ambiguous to suggest a consensus mode. When analogous dockings were performed using the 2BAN structure, a single consensus mode appeared (Table 2). The Goldscore and Chemscore fitness functions rank the *correct* mode first for 100% and 97% of the ligands, respectively. As mentioned earlier, the 2BAN structure is the structure of HIV-RT in complex with **62**, and the dockings performed with 2BAN can thus be regarded as redockings, a situation where most docking programs perform well.[35] The dockings with 1RT1, on the other hand, are cross-dockings, i.e., dockings to a structure modeled from a complex containing a chemically unrelated ligand. It is also interesting to note that redocking of **1** into the 1RT1 structure works perfectly, whereas cross-docking **1** into the protein model from 2BAN gives two different conformations as solutions (data not shown). The opposite is also true as the results above have shown, i.e., redocking of **62** in 2BAN works excellently, whereas cross-docking to 1RT1 does not. Thus, the performance in cross-docking is generally much lower than in redocking. However, cross-docking is a more realistic case study, as it more reflects a true ligand design situation where only structural information from other ligand series exists. The parallel dockings with the two protein models illustrate the effect of receptor flexibility in docking. Even though the difference between 1RT1 and 2BAN is moderate (rmsd ~1.31 Å for all heavy atoms in a 7 Å sphere around the ligand, see Figure 1), the two protein structures are dissimilar enough to affect rigid protein docking considerably. When the 1RT1 structure was used for docking, the program was not capable of finding a consensus mode common to all ligands in the series. This is a problem when it comes to ranking the ligands by affinity, as the presence of several putative binding modes makes the ranking less trustworthy.

Affinity Predictions from Scoring Functions. In addition to the analysis of the docked conformations above, affinity predictions and the relative ranking of the ligands were analyzed. Affinity scores predicted from the dockings in 2BAN by GOLD's Chemscore function (GCS) are plotted in Figure 3 for all inhibitors except **77a**, for which neither the *correct* mode nor the *wrong* was found, and **51** and **53**. Compound **51** and **53** were left out in all affinity correlation estimations due to difficulties in reproducing the experimentally observed affinity with any of the binding free energy prediction methods. These two ligands are predicted to bind more than 5 kcal/mol better than that experimentally observed for all methods evaluated, which clearly indicates they do not fit any model. These characteristics could possibly be explained by the fact that the

Table 2. Docking Results Using the 1RT1 and 2BAN Crystal Structures[a]

	1RT1				2BAN			
	Goldscore		Chemscore		Goldscore		Chemscore	
compound	correct	wrong	correct	wrong	correct	wrong	correct	wrong
36	1	-	12	1	1	-	1	6
37	1	-	1	-	1	-	1	-
38	1	-	1	12	1	14	1	-
39	1	7	10	1	1	-	1	-
40	-	1	4	1	1	13	1	-
41	3	1	-	1	1	-	1	-
42	1	13	-	1	1	-	1	-
43	11	1	-	1	1	-	1	-
44	1	3	-	1	1	-	1	-
45	-	1	-	1	1	-	1	-
46	-	1	-	1	1	-	1	-
47	1	13	4	1	1	-	1	-
48	12	1	7	3	1	-	1	-
49	11	1	-	1	1	-	1	-
50	1	-	1	-	1	4	1	-
51	1	13	1	2	1	-	1	-
52	-	1	-	1	1	-	1	-
53	-	-	1	2	1	-	1	-
54	1	-	6	1	1	-	1	-
55	10	3	2	1	1	-	1	-
56	3	1	-	1	1	-	1	-
59	1	13	1	4	1	7	1	-
60	1	3	2	1	1	-	1	-
61	7	1	-	1	1	-	1	-
62	1	15	7	1	1	-	1	-
63	-	1	-	1	1	-	1	-
64	1	-	1	-	1	-	1	-
65	1	8	5	4	1	4	1	-
66	1	3	-	1	1	-	1	-
67	-	1	-	1	1	-	1	-
68	-	-	-	1	1	-	1	-
70	-	1	-	1	1	10	1	-
77a	15	13	11	1	1	-	-	-
77b	2	6	12	2	1	13	1	-

[a] The rank is given for the best docking solution found for each binding mode in the two protein models and for each scoring function. For example, when compound **36** was docked in 1RT1, the first *correct* conformation was found in the top ranked solution with Goldscore, but only in the 12th best ranked solution with Chemscore. A dash indicates that no solution was found for that particular mode. The table clearly shows that cross-docking the ligand set with 1RT1 is much more difficult than docking with the native 2BAN model.

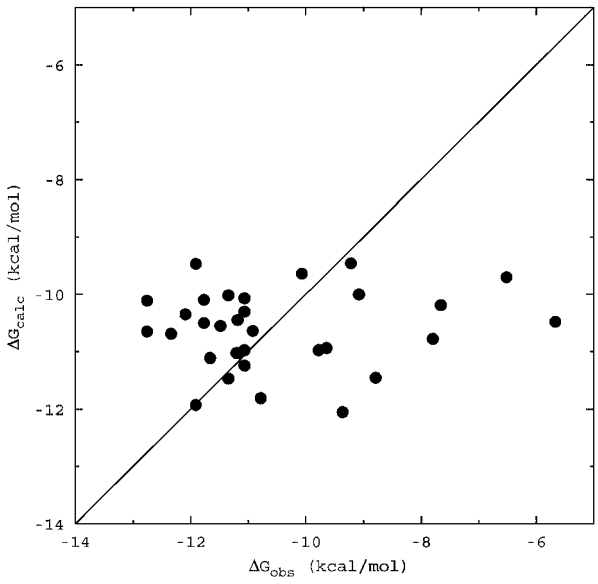


Figure 3. Plot of estimated binding free energies from the Chemscore function in GOLD vs experimental affinities. The binding free energies were estimated from docking solutions in the 2BAN X-ray model.

observed IC_{50} values were measured on cells in vitro[14] and may not reflect inhibition to RT but may be caused by solubility problems or interaction with other macromolecular species in the cells. It is obvious from the scatter plot in Figure 3 that the predicted affinities are not correlated with the experimental ones ($r = 0.10$, see Table 3). However, when the scores are plotted versus molecular weight (MW), a statistically significant trend emerges ($r = -0.73$), see Figure 4. To be statistically significant, r should be $|r| \geq 0.46$ for 31 data points at $\alpha_{conf} = 0.01$. The same trend is found for GOLD's Goldscore function (GGS), with no significant correlation between predicted and experimental binding affinity ($r = -0.26$) but with significant correlation between predicted affinity and MW ($r = -0.66$). When the experimental affinities are examined, we do not find a significant correlation with MW ($r = -0.24$), as shown in Figure 4. Thus, both GCS and GGS have a tendency to favor large ligands over small in a manner that is not supported in the experimental data. The experimental affinities do not display any correlation with MW, and ranking based on MW will therefore fail. Thus, the docking program is able to find the *correct* binding mode for all ligands when the native protein model 2BAN is used, but is not able to rank the ligands correctly according to affinity.

Ranking the ligands based on the cross-dockings performed with the 1RT1 model was found to be just as difficult for the two scoring functions as with the 2BAN dockings. When the scores from the best docked conformations for each ligand were extracted from the GCS docking, a correlation of $r = -0.04$ was achieved to the experimental values. The conformations from these dockings were, however, a mix of the *correct* and

Table 3. Computed Pearson's Correlation Coefficients for Different Combinations of Scoring Functions and Protein Models[a]

	1RT1 ($r_{calc-obs}/r_{calc-MW}$)			2BAN ($r_{calc-obs}/r_{calc-MW}$)
	all	subset[b] *correct*	subset[b] *wrong*	all
GCS	−0.04/−0.80	−0.31/−0.55		0.10/−0.73
GGS[c]	0.01/0.39	−0.55/0.24		−0.26/0.66
QCS		−0.30/−0.50	−0.28/−0.36	0.01/−0.58
PMF		−0.19/0.06	−0.23/−0.27	0.28/−0.51
X-score		−0.46/−0.22	−0.53/−0.28	−0.53/−0.43
LIE		0.79/−0.21	0.60/−0.30	

[a] The tabulated values are coefficients for the correlation of estimated binding free energies to experimental binding free energies, and estimated binding free energy to molecular weight for each combination of scoring function and protein structure. The correlation between experimental affinities and molecular weight is $r = -0.24$. [b] Different subsets were used, see text for details. [c] Note that Goldscore should produce a negative correlation coefficient, whereas all the other scoring functions should produce a positive.

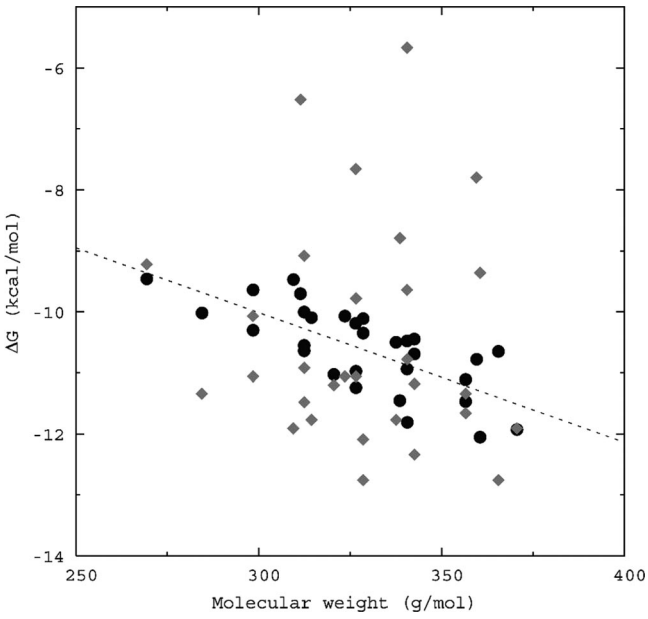


Figure 4. Plot of estimated binding free energies from the Chemscore function in GOLD vs molecular weight (circles) and experimental binding free energies vs molecular weight (diamonds). The estimated values are from dockings in the 2BAN crystal structure. The dotted line represents the least-squares fit line ($r = -0.73$) and clearly shows that larger ligands are generally ranked better. The experimental binding free energies show no evident correlation with molecular weight.

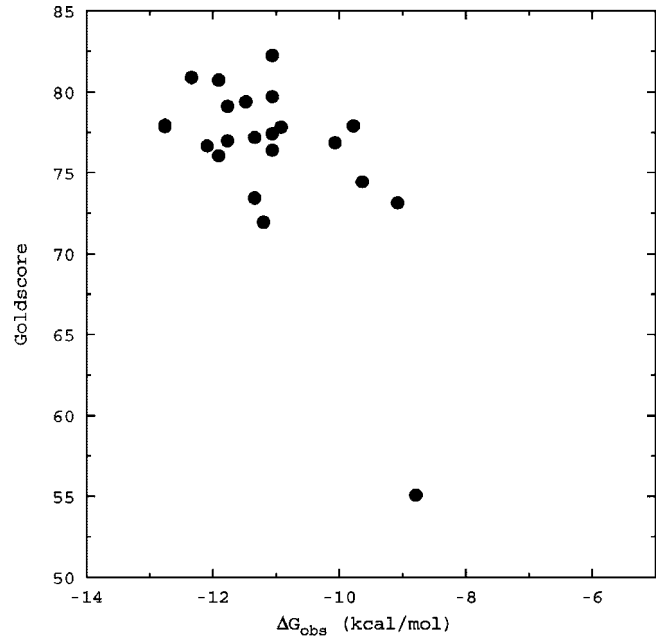


Figure 5. Scatter plot of estimated free energies of binding from the Goldscore function in GOLD from dockings in the 1RT1 model. Shown in the figure are binding free energy estimates from the best ranked *correct* conformations.

the *wrong* binding modes (see Table 2), and to be fair to the scoring function, a subset containing only the 18 ligands docked in the *correct* mode was selected. The correlation with experimental affinities for this subset was ($r = -0.31$). When the GCS score values for the *correct* conformations from 2BAN and 1RT1 are compared, a strong correlation is found, although with a slight offset. Thus, the relative ranking is conserved between the two protein models, even though the absolute values are not the same. The reason for the offset is mainly due to the lipophilic term in the scoring function which is always more favorable for the 1RT1 scoring.

In the case of GGS, no correlation to experimental affinities was found when the scores from the best cross-docked conformations in were analyzed ($r = 0.01$). In analogy to the GCS analysis, a subset containing only the best *correct* docked conformations were constructed. This subset contained 24 ligands and shows better correlation with experimental affinities ($r = -0.55$) (Figure 5). Note that GGS should give a negative r in contrast to GCS. Although this seems like an improvement, the r value is largely dependent on one influence point, compound **55**, and without it the r value drops to $r = -0.33$. It is interesting to note that the affinity predictions from GGS in 1RT1 and 2BAN do not show the strong correlation as the corresponding values from GCS do, which implies that the Goldscore fitness function is more sensitive to the conformation of the protein than Chemscore.

It is evident from the cross-docking and ranking that GGS and GCS are not able to separate between the *correct* and the *wrong* binding modes in the 1RT1 protein model. The solutions from the dockings were a mix of the two modes, and no consensus binding orientation could be found. To find a better procedure to rank the docked ligands and to find a way to discriminate between the two modes in the cross-docking case, additional affinity prediction methods were investigated for both the *correct* and the *wrong* binding modes. Three different scoring functions were used for this purpose; Chemscore,[25] X-Score,[26] and PMF,[27] as implemented in the molecular dynamics package Q.[24] From the 2BAN dockings, the 32 *correct* conformations were selected for affinity predictions with the scoring functions. A subset of 14 ligands was extracted from the 1RT1 dockings where at least one *correct* and one *wrong* conformation were found as solutions by the docking program. Thus, the best ranked *correct* and *wrong* pose were selected for the compounds where such docking solutions were available. The subset enables comparisons of the computed scores for the *correct* and *wrong* conformation for each ligand to assess if any consensus binding mode could be found. Figure 6 shows scatter plots of the results from applying the scoring functions

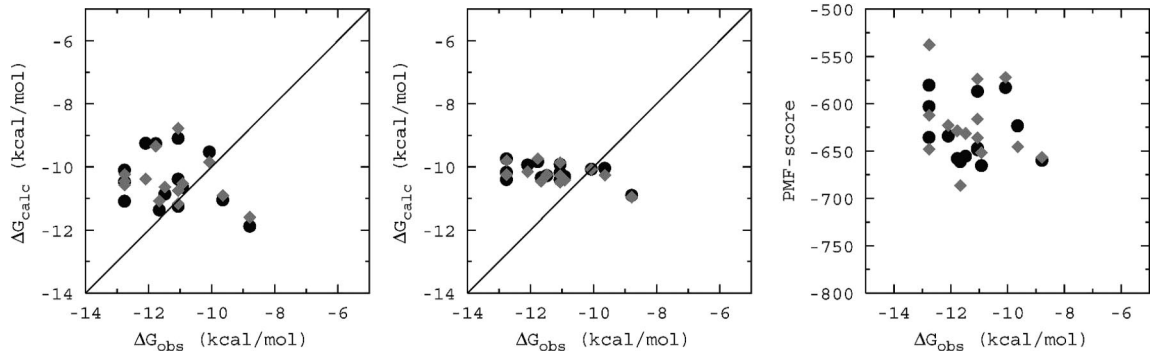


Figure 6. Scatter plots of binding free energy estimates from Chemscore, X-score, and PMF vs experimental binding free energies, from left to right. The circles and the diamonds represent scores from the top ranked *correct* and *wrong* pose, respectively. The estimates were calculated from single, docked conformations in the 1RT1 crystal structure.

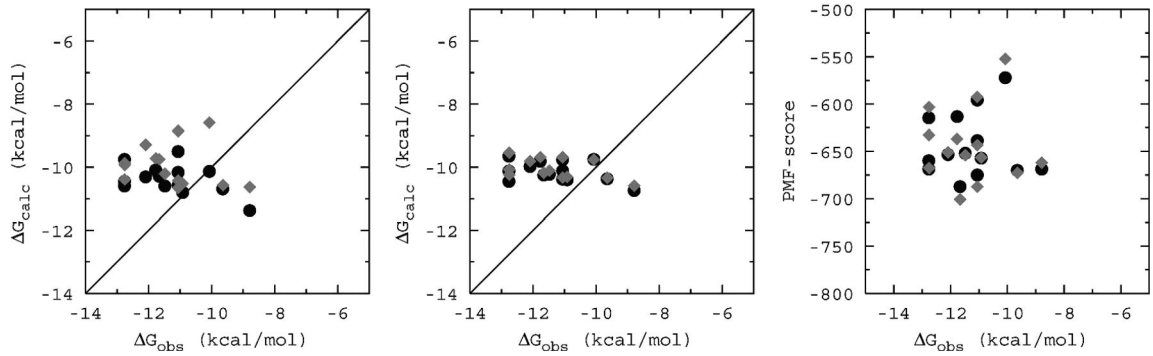


Figure 7. Scatter plots of binding free energy estimates from Chemscore, X-score, and PMF vs experimental binding free energies, from left to right. The circles and the diamonds represent averaged scores from MD snapshots for the *correct* and *wrong* modes, respectively, in the 1RT1 crystal structure.

on the docked poses from 1RT1, and the results are summarized below in text.

The implementation of the Chemscore function in Q (QCS) is analogous to the implementation in GOLD, and the affinity predictions are hence very similar. In the case of the 2BAN docked conformations, QCS displays no correlation with experimental data ($r = 0.01$), and just as with GCS, the correlation with MW is high ($r = -0.58$) and significant at $\alpha_{conf} = 0.01$. The 1RT1 subset of 14 data points gives a correlation with experimental affinities of $r = -0.30$. When the X-score function was applied to the 2BAN conformations, a correlation of $r = -0.53$ was observed between calculated and experimental binding free energies, and the corresponding value for the 1RT1 subset values is -0.46. The PMF function produced a correlation of $r = 0.28$ for the 2BAN conformations and $r = -0.19$ for the 1RT1 subset. Thus, for the three scoring functions, QCS, X-score, and PMF, no significant correlation was achieved with experimental data. In fact, for the 1RT1 subset, all three scoring functions produce negative correlation, indicating that good binders are more likely to be ranked as bad and vice versa.

Further, the impact of using the three scoring functions in consensus was investigated. The idea to combine scoring functions to receive more robust scoring has been proposed earlier with excellent results.[36–38] However, we could not find any combination of the three considered scoring functions that gives a significantly improved correlation with the experimental values. The best correlation was found for the combination QCS+PMF, but it still gives a negative r value of -0.30. To achieve any effect of combining different scoring functions, these functions must perform reasonably well individually. In this case, we cannot expect to attain acceptable results due to the bad performance of the scoring functions when applied separately. Any combination of the three scoring functions in a linear model results in at least one negative coefficient, and the predictivity of such a model must be considered very low considering that the scoring functions were developed separately and are each intended to have positive correlation with experimental data. If the coefficients in the linear model are restrained to be positive, the best model that can be achieved is to use the offset (set to the mean value) as the only predictor, with all linear coefficients set to zero.

Next, we tried to improve the scoring by sampling different system configurations through MD simulations. The concept of scoring a number of Boltzmann weighted snapshots was introduced by Marelius et al.[6] The idea is to generate a number of thermally accessible configurations, score them, and use the arithmetic mean. Furthermore, the MD simulations allows for a relaxation of the protein−ligand complex, which might be strained due to imperfect docking or implicit flexibility in the docking algorithm. Only poses from dockings with the Chemscore fitness function were selected for MD simulations. The Goldscore fitness function has a very soft repulsion potential to account for implicit flexibility and hence produces steric clashes in the cross-docking case that are not suitable as starting points for MD simulations. The docked complexes were simulated for 450 ps, and a snapshot of the system was saved every 1.5 ps. The snapshots were then scored and averaged. Figure 7 show plots of the averaged scores, and the correlation coefficients are shown in Table 3. From these results it appears that none of the scoring functions are sufficiently close to the correct values to benefit from sampling and averaging.

Scoring with the Linear Interaction Energy Method. The HIV-RT inhibitors reported by Benjahad et al. were previously docked into 2BAN by Carlsson et al. and subsequently ranked

Table 4. Experimental and Computed Binding Free Energies of 14 Ligands[a]

			ligand-surrounding interactions (kcal/mol)[d]			
compound	$\Delta G_{bind,obs}$ (kcal/mol)[b]	$\Delta G_{bind,calc}$ (kcal/mol)[c]	$\langle V_{l\text{-}s}^{vdw}\rangle_p$	$\langle V_{l\text{-}s}^{el}\rangle_p$	$\langle V_{l\text{-}s}^{vdw}\rangle_w$	$\langle V_{l\text{-}s}^{el}\rangle_w$
36$_{correct}$	−10.1	−10.9 ± 0.2	−44.8 ± 0.1	−38.0 ± 0.5	−26.63	−45.09
36$_{wrong}$		−3.5 ± 0.6	−50.4 ± 0.1	−18.2 ± 1.3		
38$_{correct}$	−11.5	−10.9 ± 0.0	−55.7 ± 0.1	−15.8 ± 0.0	−29.48	−26.37
38$_{wrong}$		−9.0 ± 0.4	−54.1 ± 0.0	−12.0 ± 1.0		
39$_{correct}$	−11.1	−11.3 ± 0.1	−58.3 ± 0.1	−15.8 ± 0.3	−31.02	−25.98
39$_{wrong}$		−9.3 ± 0.3	−56.5 ± 0.3	−11.9 ± 0.6		
40$_{correct}$	−11.7	−11.8 ± 0.3	−58.6 ± 0.4	−21.6 ± 0.6	−32.05	−30.19
40$_{wrong}$		−7.3 ± 0.3	−63.0 ± 0.3	−9.2 ± 0.6		
47$_{correct}$	−11.1	−11.2 ± 0.9	−51.1 ± 0.9	−19.3 ± 1.8	−28.13	−27.81
47$_{wrong}$		−8.2 ± 0.2	−54.2 ± 0.2	−11.1 ± 0.3		
48$_{correct}$	−10.9	−10.3 ± 0.1	−57.5 ± 0.1	−15.7 ± 0.2	−29.57	−28.4
48$_{wrong}$		−7.6 ± 0.4	−56.7 ± 0.5	−9.8 ± 0.8		
54$_{correct}$	−9.6	−9.6 ± 0.1	−57.5 ± 0.0	−18.8 ± 0.1	−30.47	−32.76
54$_{wrong}$		−7.1 ± 0.4	−56.8 ± 0.0	−13.4 ± 1.0		
55$_{correct}$	−8.8	−9.9 ± 0.1	−57.5 ± 0.6	−18.4 ± 0.1	−32.05	−30.86
55$_{wrong}$		−7.9 ± 0.2	−56.9 ± 0.4	−13.8 ± 0.3		
59$_{correct}$	−11.8	−13.0 ± 0.1	−51.7 ± 0.0	−32.1 ± 0.3	−26.08	−38.4
59$_{wrong}$		−8.7 ± 0.1	−53.7 ± 0.1	−19.4 ± 0.2		
60$_{correct}$	−12.1	−12.1 ± 0.1	−54.5 ± 0.3	−31.5 ± 0.1	−28.04	−40.52
60$_{wrong}$		−8.6 ± 0.1	−54.6 ± 0.4	−21.8 ± 0.0		
62$_{correct}$	−12.8	−11.6 ± 0.0	−57.0 ± 0.1	−19.5 ± 0.0	−29.62	−29.03
62$_{wrong}$		−9.6 ± 0.0	−55.0 ± 0.0	−15.7 ± 0.0		
65$_{correct}$	−11.1	−10.0 ± 0.2	−54.7 ± 0.1	−21.8 ± 0.4	−28	−34.67
65$_{wrong}$		−8.8 ± 0.0	−52.6 ± 0.0	−19.8 ± 0.1		
77a$_{correct}$	−12.8	−12.2 ± 0.2	−57.7 ± 0.6	−20.3 ± 0.2	−30.57	−28.15
77a$_{wrong}$		−10.9 ± 1.0	−54.6 ± 1.9	−18.6 ± 1.5		
77b$_{correct}$	−12.8	−12.6 ± 0.0	−59.3 ± 0.1	−26.2 ± 0.0	−32.27	−33.06
77b$_{wrong}$		−8.8 ± 0.1	−61.3 ± 0.3	−16.5 ± 0.1		

[a] The computed values are from LIE calculations using the 1RT1 crystal structure. [b] Experimental relative binding free energies were calculated from experimentally determined IC_{50}'s using $\Delta G_{bind,obs} = RT \ln (IC_{50})$. [c] The values were calculated using a constant offset (γ) of −10.7 kcal/mol. [d] The calculated average electrostatic $\langle V_{el}\rangle$ and nonpolar $\langle V_{vdW}\rangle$ energies for ligand-surrounding (l-s) interactions. The subscripts p and w denote simulations of the ligand in complex with the protein and free in water, respectively.

using the LIE method (the calculated binding free energies for the complete data set are shown in Figure 4 of the accompanying paper).[19] Carlsson and co-workers clearly showed that the LIE method can be applied to this ligand set with excellent results in the protein model cocrystallized with compound **62** (2BAN). Here, the LIE approach was used to predict binding free energies for the 14 ligands that possessed docking solutions in both the *correct* and *wrong* modes in the crystal structure 1RT1. The results from the MD/LIE calculations are displayed in Table 4 and show a significant difference in binding free energy between the two modes (see Figure 8). The predictions for the *correct* mode correlates well with experimental values ($r = 0.79$), with a mean unsigned error of 0.57 kcal/mol (rms error 0.71 kcal/mol) and an R^2 coefficient of 0.62 (Figure 8). The achieved correlation is satisfactory considering that no reparameterization of the general LIE model has been carried out and the data used can therefore be viewed as an independent test set. The observed correlation between the affinity prediction in the *wrong* mode and experimental data is slightly worse ($r = 0.60$) but, more importantly, the LIE method predicts each of the ligands to bind in the *correct* mode. The latter is true also for the compound **51**, which was left out in the correlation analysis (data not shown). For the other compound left out from the correlation analysis, **53**, the MD simulations fail in the *correct* mode due to steric clashes, indicating that the binding pocket is too small. On the basis of these LIE results, the prediction would be made that the ligands bind in the *correct* binding mode.

By measuring the interaction potential energy from one residue at the time to the ligand in both the *correct* and *wrong* mode, the most prominent features determining the mode could be elucidated. Interaction energies from the ligand to each amino acid were extracted from the simulations and partitioned into van der Waals and electrostatic contributions. The difference for each residue between the *correct* and *wrong* modes were calculated and resulted in a fingerprint for each ligand containing information on what interactions are important for determining the binding mode. When these fingerprints were analyzed, the most striking feature found was the similarity in the van der Waals interactions between the two modes. Thus, when two conformations fit the binding site equally well the binding mode is determined by electrostatics. This is probably a key as to why the empirical and knowledge-based scoring functions fail to differentiate between the modes; they lack proper electrostatic descriptors. As a consequence of the importance of the electrostatics, the interacting residues that determine the binding mode are all situated around the pyridinone ring and not around

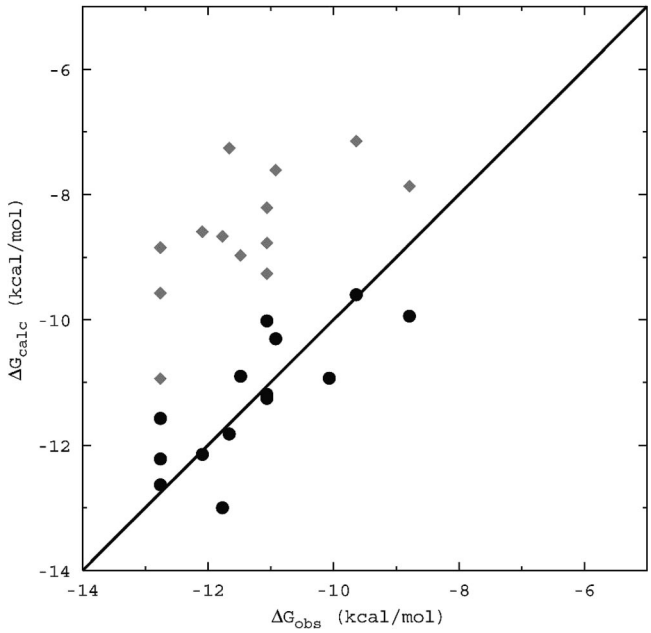


Figure 8. Scatter plot of binding free energy estimates from LIE calculations for the *correct* (circles) and *wrong* (diamonds) conformations. Docking solutions from cross-docking in the 1RT1 crystal structure were used as starting structures in the MD simulations. Note the excellent separation between the two modes in the estimates.

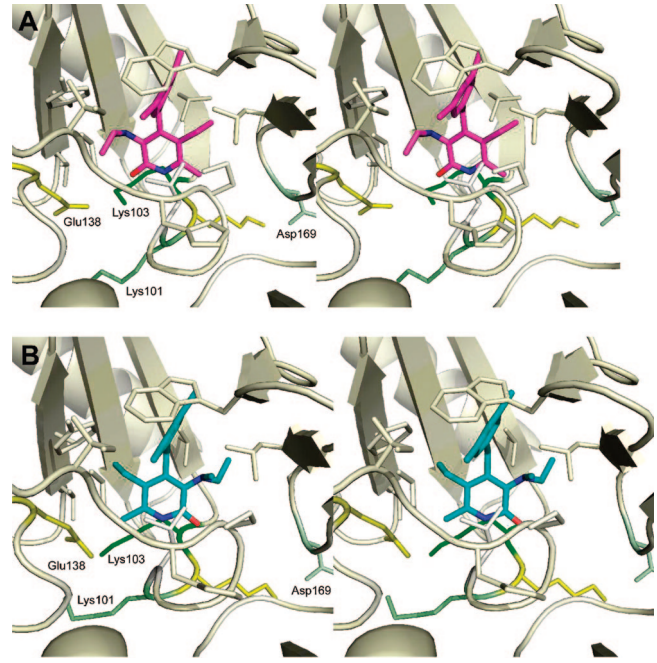


Figure 9. Shown here are two stereo figures of **38** in (A) the *correct* conformation (magenta) and (B) the *wrong* conformation (cyan) in complex with the 1RT1 model of HIV-RT. The two figures are representative MD snapshots where the protein is color-coded in green−white−yellow to show the computed difference in the electrostatic contribution the *correct* and *wrong* modes. Green residues favor the *correct* mode and yellow favor the *wrong* mode.

the hydrophobic benzyl (or dibenzyl) moiety. The electrostatic fingerprints are very similar for all ligands, and compound **38** will here serve as a representative example to explain the differences in the two modes. Ligand **38** is a particularly good example, as two of the substitutions on the pyridinone are of equal size which makes the shape of the ligand very similar in the two modes.

The fingerprint analysis showed that the single most important residue for discriminating between the two modes is Lys103 (see Figure 9). In the *correct* mode, the carbonyl of the pyridinone is oriented toward Lys103 and makes strong electrostatic interaction to the charged amine. In the *wrong* mode, the carbonyl is replaced by a methyl group and the favorable interaction is lost. This leads to a total difference in average electrostatic interaction energy of 14.5 kcal throughout the simulations in favor of the *correct* mode. The amine nitrogen of Lys103 maintains a distance of 2.8−4.5 Å to this carbonyl. Much of the effect from Lys103 is counterbalanced by Glu138 which is situated 5.5−6.0 Å from the carbonyl during the simulation (measured from Cδ in Glu138), but it is not as strong as the interaction with Lys103. The total effect from these interactions accounts for the major part of the difference in binding free energy between the *correct* and *wrong* modes. The solvation term is always more favorable for the *wrong* mode, which indicates that the ligands have stronger interactions with water molecules in the wrong mode, but these interactions are compensated by protein interactions in the *correct* mode.

Discussion

Scoring functions and other fast methods to predict binding affinities are well suited for the purpose of virtual screening where new lead hits are sought after. However, when it comes to lead optimization, these methods are too coarse-grained and more accurate methods must be used. We have demonstrated here that docking a homologous ligand series into an X-ray structure created from a protein−ligand complex containing a ligand from another series (cross-docking) is not an easy task. In some cases, several docking solutions with equal scores will emerge as possible candidates of the bioactive conformation. To draw conclusions on how the ligand series binds to the receptor affinity, predictions with increased accuracy must be performed, and when the binding mode of the series has been established, structural information can be extracted from the calculations. This structural information can be used to improve the lead compound, both in terms of increasing affinity or in terms of changing properties of the ligand to improve the ADME/PK profile, which can be done by removing or replacing groups that are found not to be important for ligand binding.

One of the problems with finding the correct bioactive mode of a ligand series is the low complexity of the scoring functions used in docking programs. The scoring functions often lack sufficient negative contributions from polar mismatches and unfavorable desolvation, and sometimes even from steric clashes, which lead them to be strongly correlated with molecular weight. Bigger is not always better in the case of homologous ligands in a lead optimization series, and to overcome this problem, a higher level of theory must be used, e.g., the LIE method. We find that the docking program GOLD is very reliable in generating sufficient conformations. In the case where a crystal structure cocrystallized with **62** (2BAN) was used, GOLD was able to generate the *correct* conformation and rank it in first place for practically all ligands. When a crystal structure belonging to another ligand series (1RT1) was used, the docking program was able to generate suitable conformations but not rank them, resulting in an ambiguous binding mode for the ligand series. This result is mainly due to the subtle shape differences in the binding site of the crystal structures; as mentioned above, the scoring functions are primarily based on shape complementarity and the 1RT1 crystal structure is modeled from a complex with a ligand that has larger substituents on one side of the central pyridinone group and

smaller substituents on the other. Thus, the binding pocket in 1RT1 is widened on one side and reduced on the other compared to 2BAN, and these shape differences have substantial impact on the docking. The results from the dockings are in line with Warren et al.,[35] who found that docking programs generally are capable of generating poses close to crystallographically determined conformations but have difficulties in ranking the correct conformation in the top position. In this situation, the question of a consensus binding mode arises. When it comes to ranking the ligands relative to each other according to binding affinity, the results found by Warren et al. were discouraging. They found that none of the tested scoring functions were able to rank ligands by affinity in a useful way. We evaluated four scoring functions to rank the ligands and came to a similar conclusion.

It was clear that we needed additional complexity in our model to succeed in ranking the ligands. One way of doing this is to use conformational sampling on the protein−ligand complex and score different configurations of the system. The sampling was intended to allow steric clashes from the docking to relax and to average out peaks and valleys in the scoring energy landscape. Improvement by conformational sampling has been observed earlier, and it has the advantage of faster convergence than force field energy based methods like the LIE method. In contrast to the work presented here, Gutiérrez-de-Terán et al. showed that conformational sampling contributed a substantial improvement to the relative ranking of nine malaria protease inhibitors.[8] The sampling increased the correlation coefficient, r, from 0.66 to 0.82 with X-score. In this case, there was a correlation between MW and observed binding affinity ($r = -0.58$), but the achieved correlation between predicted affinities from X-score and MW was close to a perfect fit ($r = -0.97$) when a single conformation was scored per inhibitor. The sampling improved the quality of the predictions from X-score both by increasing the correlation with experimental values and at the same time by decreasing the correlation with MW ($r = -0.90$). Still, the correlation to MW was higher than the correlation to observed affinities, which shows that the scoring function is strongly biased to favor large ligands. Affinity predictions calculated from the same MD simulations using the LIE method had a correlation with MW in the same region as the experimental values and at the same time good correlation with experimental affinities ($r = 0.89$). In contrast, conformational sampling did not improve the scoring of the present HIV-RT ligands as it did with the malarial protease inhibitors. The reason for this might be the small structural changes between the ligands and the strong shape complementarity to the protein for all ligands.

When the level of theory was raised another step to include MD/LIE calculations, we were able to achieve consistent results regarding the binding mode. The predicted binding free energies from the LIE method were consistently in favor of the *correct* binding mode, and the obtained relative free energies were in close agreement with experimental data. Furthermore, the MD simulations were also able to produce a sound rationale as to why LIE predicts the *correct* binding mode to be the most favorable. According to the simulations, Lys103 plays a crucial role in the alignment of the pyridinone ring of the ligands and Lys103 is also found to be an important factor in the affinity of the ligands to the protein. The importance of Lys103 is also illustrated by the fact that problems with resistance to pyridinone-based compounds may arise from Lys103 mutations.[39]

The procedure employed here can be used to predict binding modes of novel classes of compounds when the outcome of a docking procedure is ambiguous. Benjahad et al. attempted to predict the binding conformation of **62** by molecular modeling and a minimization protocol, but they were not successful in finding a trustworthy solution. Instead, they decided to solve the problem by X-ray crystallography resulting in the 2BAN model.[10] The method presented herein could thus have been used to correctly predict the binding mode of the compound series without the need of a new crystal structure, and the combination of docking and affinity predictions with the LIE method can thus save valuable resources in lead optimization projects.

Acknowledgment. Financial support from the Swedish Foundation for Strategic Research (SSF/Rapid) and the Swedish Research Council (VR) is greatly acknowledged.

References

(1) Straatsma, T. P.; McCammon, J. A. Computational Alchemy. *Annu. Rev. Phys. Chem.* **1992**, *43*, 407–435.
(2) Gohlke, H.; Klebe, G. Statistical potentials and scoring functions applied to protein−ligand binding. *Curr. Opin. Struct. Biol.* **2001**, *11*, 231–235.
(3) 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.
(4) Jorgensen, W. L. The many roles of computation in drug discovery. *Science* **2004**, *303*, 1813–1818.
(5) 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.
(6) Marelius, J.; Ljungberg, K. B.; Aqvist, J. Sensitivity of an empirical affinity scoring function to changes in receptor−ligand complex conformations. *Eur. J. Pharm. Sci.* **2001**, *14*, 87–95.
(7) Ersmark, K.; Feierberg, I.; Bjelic, S.; Hulten, J.; Samuelsson, B.; Aqvist, J.; Hallberg, A. C-2-symmetric inhibitors of *Plasmodium falciparum* plasmepsin II: Synthesis and theoretical predictions. *Bioorg. Med. Chem.* **2003**, *11*, 3723–3733.
(8) Gutiérrez-de-Terán, H.; Nervall, M.; Ersmark, K.; Dunn, B. M.; Hallberg, A.; Åqvist, J. Inhibitor binding to the plasmepsin IV aspartic protease from *Plasmodium falciparum*. *Biochemistry* **2006**, *45*, 10529–10541.
(9) Åqvist, J.; Medina, C.; Samuelsson, J. E. New method for predicting binding affinity in computer-aided drug design. *Protein Eng.* **1994**, *7*, 385–391.
(10) 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.
(11) Hopkins, A. L.; Ren, J. S.; Milton, J.; Hazen, R. J.; Chan, J. H.; Stuart, D. I.; Stammers, D. K. Design of non-nucleoside inhibitors of HIV-1 reverse transcriptase with improved drug resistance properties. 1. *J. Med. Chem.* **2004**, *47*, 5912–5922.
(12) Freeman, G. A.; Andrews, C. W.; Hopkins, A. L.; Lowell, G. S.; Schaller, L. T.; Cowan, J. R.; Gonzales, S. S.; Koszalka, G. W.; Hazen, R. J.; Boone, L. R.; Ferris, R. G.; Creech, K. L.; Roberts, G. B.; Short, S. A.; Weaver, K.; Reynolds, D. J.; Milton, J.; Ren, J. S.; Stuart, D. I.; Stammers, D. K.; Chan, J. H. Design of non-nucleoside inhibitors of HIV-1 reverse transcriptase with improved drug resistance properties. 2. *J. Med. Chem.* **2004**, *47*, 5923–5936.
(13) Ren, J.; Milton, J.; Weaver, K. L.; Short, S. A.; Stuart, D. I.; Stammers, D. K. Structural basis for the resilience of efavirenz (DMP-266) to drug resistance mutations in HIV-1 reverse transcriptase. *Structure* **2000**, *8*, 1089–1094.
(14) 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.
(15) Aly, Y. L.; Pedersen, E. B.; La Colla, P.; Loddo, R. Synthesis and anti-HIV-1 activity of novel MKC-442 analogues containing alkenyl chains or reactive functionalities in the 6-benzyl group. *Monatsh. Chem.* **2006**, *137*, 1557–1570.

(16) Ellis, D.; Kuhen, K. L.; Anaclerio, B.; Wu, B.; Wolff, K.; Yin, H.; Bursulaya, B.; Caldwell, J.; Karanewsky, D.; He, Y. Design, synthesis, and biological evaluations of novel quinolones as HIV-1 non-nucleoside reverse transcriptase inhibitors. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 4246–4251.

(17) Jiang, T.; Kuhen, K. L.; Wolff, K.; Yin, H.; Bieza, K.; Caldwell, J.; Bursulaya, B.; Wu, T. Y. H.; He, Y. Design, synthesis and biological evaluations of novel oxindoles as HIV-1 non-nucleoside reverse transcriptase inhibitors. Part I. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 2105–2108.

(18) 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.

(19) Carlsson, J.; Boukharta, L.; Aqvist, J. 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. **2008**, *51*, 2648–2656.

(20) 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.

(21) 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.

(22) Cambridge Crystallographic Data Centre. http://www.ccdc.cam.ac.uk/products/life_sciences/gold/faqs/scientific_faq.php#bindingaffinity (accessed Sept 3, 2007).

(23) Cheng, Y.; Prusoff, W. H. Relationship between inhibition constant (K_I) and concentration of inhibitor which causes 50% inhibition (I_{50}) of an enzymatic reaction. *Biochem. Pharmacol.* **1973**, *22*, 3099–3108.

(24) 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. Graph.* **1998**, *16*, 213–225.

(25) 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.

(26) Wang, R. X.; Lai, L. H.; Wang, S. M. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 11–26.

(27) 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.

(28) Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Rw, I.; Klein, M. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.

(29) King, G.; Warshel, A. A surface constrained all-atom solvent model for effective simulations of polar solutions. *J. Chem. Phys.* **1989**, *91*, 3647–3661.

(30) 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.

(31) 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.

(32) Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, 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.

(33) 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.

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

(35) 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.

(36) Charifson, P. S.; Corkery, 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.

(37) Oda, A.; Tsuchida, K.; Takakura, T.; Yamaotsu, N.; Hirono, S. Comparison of consensus scoring strategies for evaluating computational models of protein−ligand complexes. *J. Chem. Inf. Model.* **2006**, *46*, 380–391.

(38) Baber, J. C.; William, A. S.; Gao, Y. H.; Feher, M. The use of consensus scoring in ligand-based virtual screening. *J. Chem. Inf. Model.* **2006**, *46*, 277–288.

(39) Nunberg, J. H.; Schleif, W. A.; Boots, E. J.; Obrien, J. A.; Quintero, J. C.; Hoffman, J. M.; Emini, E. A.; Goldman, M. E. Viral Resistance To Human Immunodeficiency Virus Type-1-Specific Pyridinone Reverse-Transcriptase Inhibitors. *J. Virol.* **1991**, *65*, 4887–4892.

JM701218J