Journal of Medicinal Chemistry Article

J. Med. Chem. **2011,** *54,* 709–718 **709**
DOI: 10.1021/jm101052g

Deconstruction of Non-Nucleoside Reverse Transcriptase Inhibitors of Human Immunodeficiency Virus Type 1 for Exploration of the Optimization Landscape of Fragments

Peter Brandt,[†] Matthis Geitmann,[†] and U. Helena Danielson*[,†,‡]

[†]*Beactica AB, Box 567, SE-751 22 Uppsala, Sweden, and* [‡]*Department of Biochemistry and Organic Chemistry, Uppsala University, Box 576, SE-751 23 Uppsala, Sweden*

Received August 13, 2010

This study has taken a closer look at the theoretical basis for protein−fragment interactions. The approach involved the deconstruction of 3 non-nucleoside inhibitors of HIV-1 reverse transcriptase and investigation of the interaction between 21 substructures and the enzyme. It focused on the concept of ligand efficiency and showed that ligand independent free energy fees (ΔG_{ind}) are crucial for the understanding of the binding affinities of fragments. A value of 7.0 kcal mol^{-1} for the ΔG_{ind} term is shown to be a lower limit for the NNRTI binding pocket of HIV-1 RT. The addition of the ΔG_{ind} term to the dissociation free energy in the calculation of a corrected ligand efficiency, in combination with the lack of an efficient ligand binding hot spot in the NNIBP, fully explains the existence of nonbinding NNRTI substructures. By applying the concept to a larger set of ligands, we could define a binding site profile that indicates the absence of an efficient fragment binding hot spot but an efficient binding of full-sized NNRTIs. The analysis explains some of the challenges in identifying fragments against flexible targets involving conformational changes and how fragments may be prioritized.

Introduction

The success of fragment-based drug discovery (FBDD[a])[1−3] and fragment library screening[4,5] relies on the ability of a target protein to bind a small compound that can be evolved into a full-sized ligand with suitable properties for a drug. This is typically possible when the target has a well-defined ligand binding pocket, a situation common for kinases, for example.[6,7] In addition, FBDD depends on the ability to detect the binding of weakly interacting fragments (often millimolar affinities) to the target protein. Much focus in FBDD has been on the use of sensitive and informative biophysical assays for detecting such weakly interacting fragments.

In an accompanying paper (DOI: 10.1021/jm1010513), we present an SPR-biosensor based approach to screen a library consisting of 1040 fragments against the non-nucleoside inhibitor binding pocket (NNIBP) of human immunodeficiency virus type 1 reverse transcriptase (HIV-1 RT).[8] The NNIBP is a rather hydrophobic binding site lacking strong determinants. This would suggest a promiscuous binding site and hence a high hit rate. However, among the 1040 fragments assayed, we could only identify 20 fragments binding to the NNIBP. These fragments all showed unexpectedly low ligand efficiencies (LE)[9] and fit qualities (FQ).[10−12]

Although successful in terms of identifying fragments both binding to the NNIBP and inhibiting the enzyme, there were some outstanding questions remaining to be answered. For example, are the low ligand efficiencies and fit qualities observed for the hits a consequence of a poor sampling of fragment-space or are they characteristics of the particular binding site? Do the unexpectedly low ligand efficiencies of the hits indicate that these are poor starting points for lead generation? In addition, the observation that fragments more similar to published non-nucleoside reverse transcriptase inhibitors (NNRTI) is not more frequent among the hits than among the inactives could indicate a difficult optimization pathway to full-sized NNRTIs. The surprisingly low hit rate from the fragment screen and the inefficient binding of the discovered hits prompted us to take a closer look at the theoretical basis for protein−fragment interactions.

Our approach was to investigate the binding of fragments of previously developed NNRTIs.[13] The deconstruction of NNRTIs was performed by using compounds previously described in the literature. The interactions between the substructures and HIV-1 RT were assayed using a surface plasmon resonance (SPR) based biosensor. It is a method we have previously used for screening and hit characterization, including for detailed interaction studies of HIV-1 RT and small molecule ligands.[14−18] It has also been used for screening of a fragment library, as described in an accompanying study.[8] The current experimental data set was complemented by data for carbonic anhydrase I, thrombin, heat shock protein 90 (Hsp90), and β-secretase (BACE-1), extracted from

*To whom correspondence should be addressed. Address: Department of Biochemistry and Organic Chemistry, Uppsala University, Box 576, SE-751 23 Uppsala, Sweden. Phone: +46-18-471 45 45. Fax: +46-18-55 84 31. E-mail: helena.danielson@biorg.uu.se.

[a]Abbreviations: AIDS, acquired immunodeficiency syndrome; BACE-1, β-site amyloid precursor protein cleaving enzyme 1; BL, binding late report point; FBDD, fragment-based drug discovery; FQ, fit quality; HIV-1, human immunodeficiency virus type 1; Hsp90, heat shock protein 90; LE, ligand efficiency; n_{HA}, number of heavy atoms; NNIBP, non-nucleoside reverse transcriptase inhibitor binding pocket; NNRTI, non-nucleoside reverse transcriptase inhibitor; RT, reverse transcriptase; SPR, surface plasmon resonance.

Published on Web 01/05/2011

BindingDB.[19,20] This enabled an analysis of the generality of the conclusions.

By investigating the binding of increasingly smaller substructures of selected NNRTIs, we could probe the NNRTI binding site for binding hot spots and possibly also identify regions less important for binding.[21] We focused our study on the concept of ligand efficiency (LE).[9] This is a parameter that has been defined as the free energy of dissociation divided by the number of heavy atoms of the compound,

$$\mathrm{LE} = -RT\ln(K_D)/n_{HA} \quad (1)$$

Ligand efficiency, initially denoted Δg, was introduced by Hopkins et al.,[9] based on the work by Kuntz et al.,[22] as a tool to guide the lead generation and optimization processes. Among the many variants of ligand efficiency and ranking metrics put forward since,[4,23−25] few, if any, have been rigorously shown to be suitable for direct prioritization of hits or leads.

Although LE is a useful concept, the approach ignores the loss of translational and rotational entropy, ΔS_{tr}, upon binding of a fragment to a protein. This loss will be rather independent of ligand size[31] and is therefore of fundamental importance in the discussion of an atomic based contribution to binding. In order to address the simplification of the ligand efficiency approach, we have returned to the early paper of Page and Jencks,[26] where entropic contributions to rates of bimolecular reactions are discussed. The authors conclude that for a simple bimolecular reaction in solution, the loss of translational and rotational entropy, ΔS_{tr}, is about 40−50 cal mol^{-1} K^{-1}, corresponding to 12−15 kcal mol^{-1} in terms of free energy at 298 K.[27] In a later paper, Andrews et al.[28] use a value of 14 kcal mol^{-1} for the term $T\Delta S_{tr}$. The authors use this energy in the estimation of functional group contributions to protein−ligand interactions. Warshel and co-workers have later claimed that the value used by Page and Jencks is an overestimation of the entropic term due to the assumption that all translational and rotational motions are lost upon protein binding.[29] On the basis of molecular dynamics simulations and free energy perturbations, they estimate the entropic contribution to the free energy of a substrate−protein association to be only 2.5 kcal mol^{-1} for the particular model system studied. This lower contribution from translational and rotational motions is stated to be a result of the flexibility of the ligand remaining after binding to the protein. In other words, there are new low frequency vibrational modes that compensate for the loss in translational and rotational entropy in the complex. Thus, the value of 2.5 kcal mol^{-1} used by Warshel and co-workers does not correspond to the loss of translational and rotational entropy, $T\Delta S_{tr}$. The value derived by Warshel and co-workers is supported by a fragment linking study performed by Borsi et al. on MMP-12 inhibitors.[30] They determined a value of a linking coefficient, E, to 2.1 × 10^{-3} M^{-1} which would correspond to 3.6 kcal mol^{-1}, a contribution that is stated to be totally entropic in nature.

In a paper more closely related to fragment-based drug discovery, Murray et al.[31] separate the dissociation free energy ΔG into ΔG_{rigid} and ΔG_{int}, where the former is the free energy change derived from loss of rigid body entropy and the latter describes the intrinsic binding affinity. The authors estimate ΔG_{rigid} for small molecules bound to proteins to be 3.6−4.8 kcal mol^{-1}. But the $T\Delta S_{tr}$ contribution to the ΔG_{rigid} term, estimated by statistical mechanical calculations, is estimated to be about 20 kcal mol^{-1} for a 200 Da fragment. Again, the lower value of ΔG_{rigid} compared to $T\Delta S_{tr}$ originates from remaining flexibility of the ligand bound to the protein.

Finally, in a paper by Chang et al.,[32] discussing "ligand configurational entropy and protein binding", amprenavir is estimated to lose 12−16 kcal mol^{-1} in free energy at 298 K because of losses in rotational and translational entropy upon binding to HIV protease. For a review of the theory of free energy and entropy in noncovalent binding, see Zhou et al.,[33] and for additional examples, see Lou et al.[34] In the following, we will discuss the contribution to the free energy from loss of translational and rotational degrees of freedom, $T\Delta S_{tr}$, separate from any gain in low vibrational degrees of freedom caused by binding of a ligand to a protein.

Although the estimation of ligand efficiencies can be improved by accounting for loss of translational and rotational entropies, the procedures described so far are all still based on the assumption of a single equilibrium, adequately described by K_D. This may be an oversimplification, as ligand interactions with a target can be composed of several independent equilibria and hence equilibrium constants. In the case of HIV-1 RT, it has been shown that ligand binding occurs by a three-step mechanism involving both a pre-equilibrium

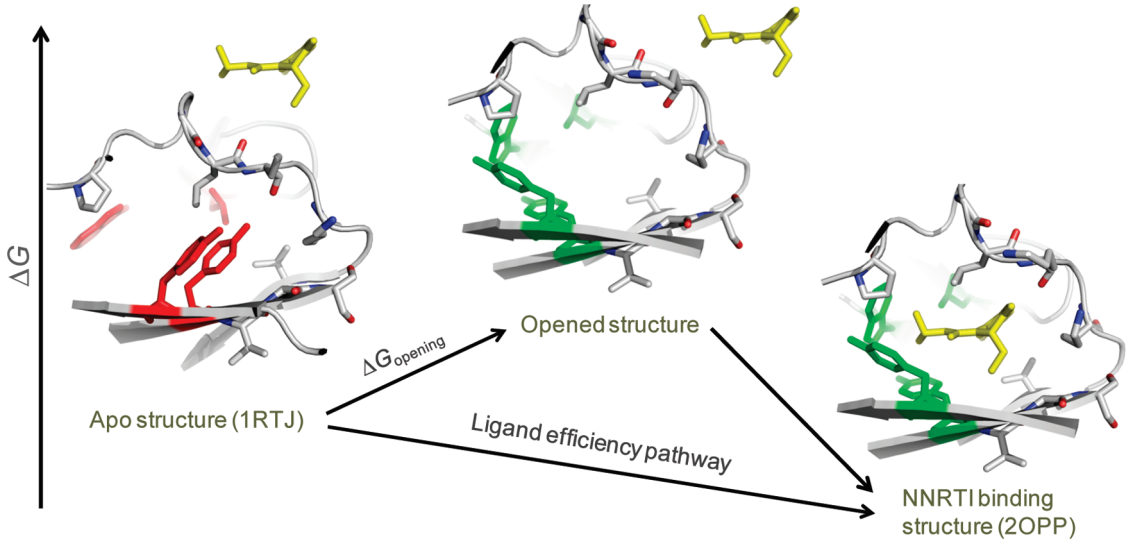

Figure 1. Energy diagram for the mechanism of ligand binding to the non-nucleoside binding pocket of HIV-1 RT.

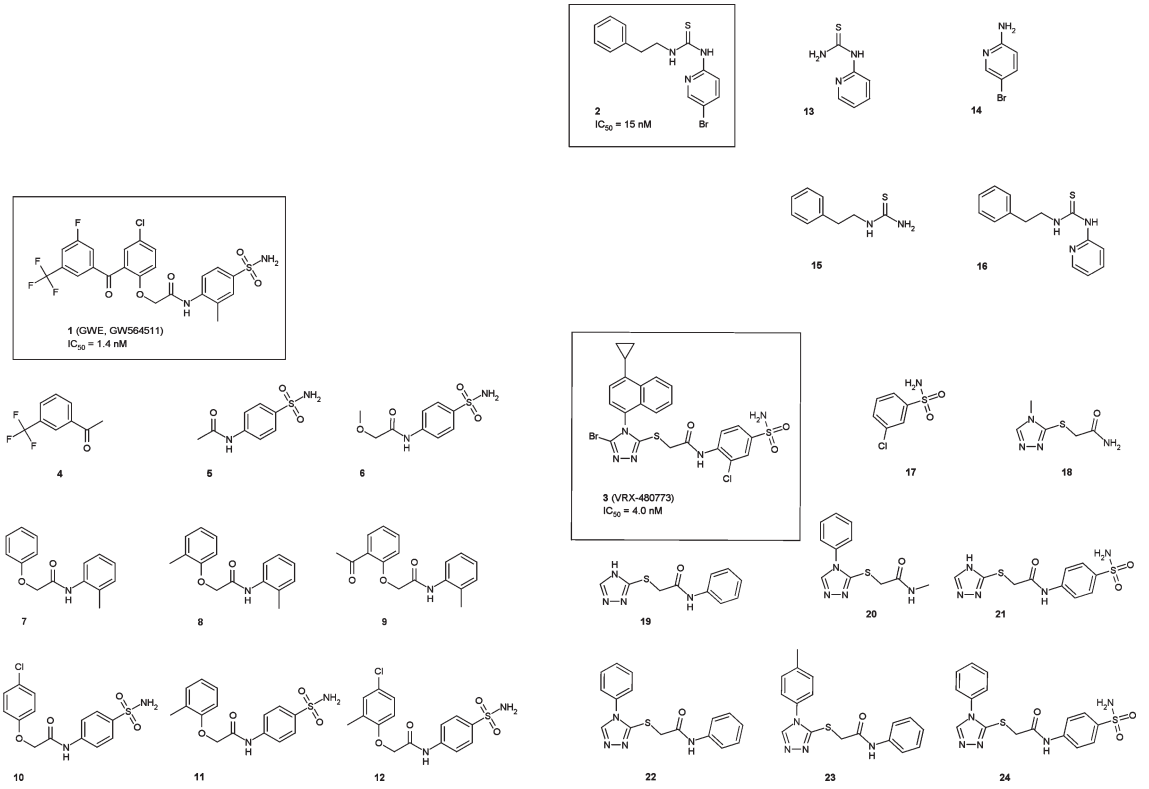

Figure 2. Three NNRTIs selected for deconstruction (**1**−**3**) and the corresponding substructures selected for analysis (**4**−**24**). Literature values of IC_{50} are given for the parent NNRTIs **1**−**3**.[35]

between at least two forms of the free enzyme (selected fit) and a ligand induced conformational change (induced fit).[14] The selected fit step implies an endergonic activation step where the NNIBP is opened for NNRTIs, followed by a ligand binding step (Figure 1). The free energy required to open the binding site could be treated as being independent of the structure and size of the NNRTI. In this article, we will use the term "ligand independent free energy fees" (ΔG_{ind}) for the free energy charged for preparing the enzyme for binding and for reducing translational and rotational degrees of freedom. The increased understanding of the energetics of interacting protein−ligand systems gained in this study has clear implications on our current application of SPR biosensor technology for screening of fragment libraries[18] and the metrics used for identification and prioritization of hits.

Results

Deconstruction of NNRTIs. Three parent NNRTIs (compounds **1**−**3**, Figure 2)[35] were selected on the basis of the availability of commercial fragments that represented substructures of the compounds. The substructures themselves were chosen on a naive prediction of expected K_D values. Thus, it was assumed that the free energy of the protein−ligand interaction would be equally distributed over the atoms in the parent NNRTIs. On the basis of the ligand efficiency of the parent NNRTIs (**1**−**3**, Figure 2),[35] 21 commercial substructures (**4**−**24**, Figure 2) estimated to have IC_{50} values below 3 mM were purchased.

Interaction Analysis and Determination of K_D Values. The interactions between the selected substrucures of NNRTIs (**1**−**3**) and the parent NNRTI (**2**), and HIV-1 RT were analyzed with a biosensor-based assay. Attempts to extract K_D values from the experimental sensorgrams using steady-state responses and eq 7 were made for all compounds. However, mechanistically meaningful K_D values with submillimolar affinities could reproducibly be determined for only 3 of the 21 tested NNRTI substructures (Figures 3 and 4 and Table 1). These three compounds were **10** (30 μM), which is an analogue of GW4511 (**1**), compound **16** (6 μM), which is a very close analogue to **2**, and finally compound **23** (809 μM), which is a substructure of VRX-387902 (**3**).

An additional 6 compounds that appeared to bind to the protein were also identified. The K_D estimates (given in parentheses in Table 1) were above the highest concentration tested and were flawed by a large linear component in the equation or by a less good fit to the data. The rough K_D estimates for these compounds are included as a basis for the discrimination of compounds interacting with HIV-1 RT from noninteracting compounds. With this generous definition, only 9 out of 24 compounds were found to bind to the protein. This was an unexpectedly low fraction and prompted us to investigate the outcome further.

Only 3 of the 9 NNRTI substructures identified to bind to the enzyme were small enough to qualify as fragments when defining a fragment as a molecule with a molecular weight below 300 Da. These were compounds **9**, **15**, and **16**. The average number of heavy atoms for identified binders was 21 and the average for nonbinders was 14 with an average molecular weight of 211 Da. Thus, the nonbinders were typical fragment-sized molecules whereas the binders were larger chemical structures being more similar to the parent NNRTIs. This apparent correlation between the ability to bind and size is an important observation that is further discussed below.

Correlation between Predicted and Experimental ΔG Values. On the basis of the determined affinities, two different

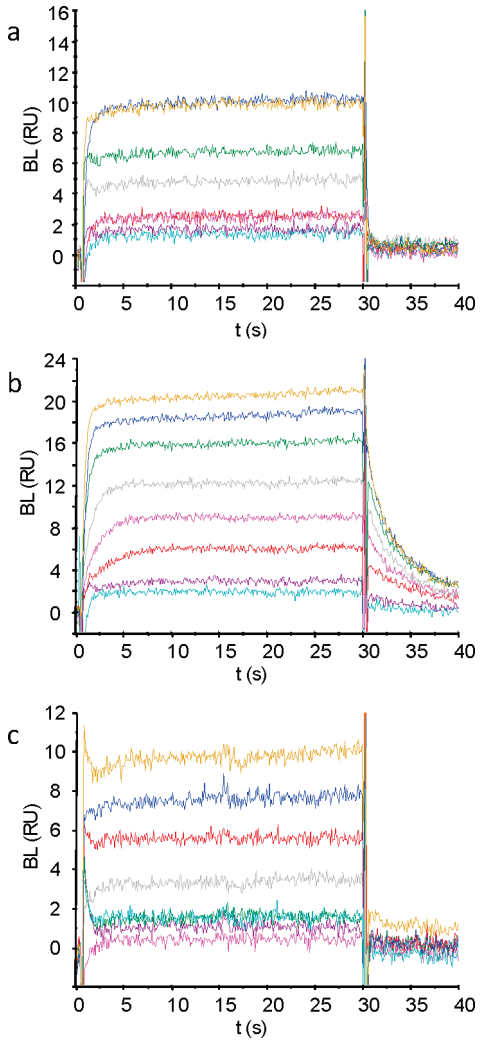

Figure 3. Sensorgrams for the interaction between the three substructures **10** (a), **16** (b), and **23** (c) and HIV-1 RT. Compounds were injected in 2-fold dilution series (50−0.39 μM for compounds **10** and **16** and 500−3.9 μM for compound **23**).

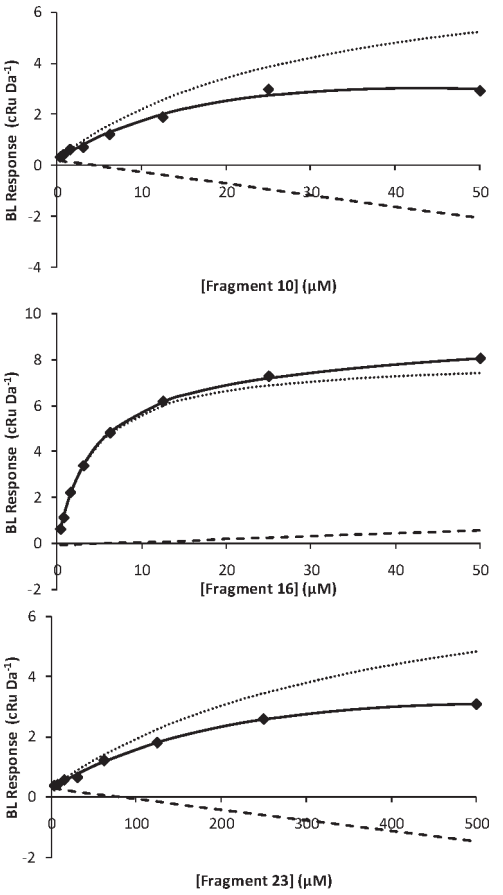

Figure 4. Concentration dependence of the signal at the end of the injection (BL report point) and the contributions from ligand binding, secondary effects, and small errors in blank corrections to the response. Experimental data points (◆) are theoretically described by eq 7 (−) where secondary signals (- - -) are subtracted from the Langmuir isotherm (· · ·). The BL response is given in cRu/Da, where c stands for centi.

estimates of ligand efficiencies were calculated (Table 1). The purpose was to compare the usefulness of the different types of ligand efficiency calculations and to illustrate the importance of ligand independent free energy fees. In this work, we therefore reintroduced the translational and rotational entropy into the discussion of ligand efficiencies. Thus, we defined a corrected ligand efficiency as

$$LE^* = [\Delta G_{opening} - T\Delta S_{tr} - RT\ln(K_D)]/n_{HA} \quad (2)$$

where $T\Delta S_{tr}$ is the estimated loss in free energy caused by a loss of translational and rotational entropy. We also included $\Delta G_{opening}$, a term accounting for the free energy required to open the NNRTI binding site. The ligand independent free energy fee is thus defined as

$$\Delta G_{ind} = \Delta G_{opening} - T\Delta S_{tr} \quad (3)$$

In the calculations, ΔG_{ind} was somewhat arbitrarily set to 7.0 kcal mol^{-1}, representing the mean of 14 kcal mol^{-1} used for the $T\Delta S_{tr}$ term by Andrews[28] and 0 kcal mol^{-1} used in eq 1.

By use of LE* for the parent NNRTIs, K_D values of fragments could be estimated simply by scaling the corresponding binding free energy according to the number of heavy atoms remaining in the fragment (eq 13). This resulted in a good coherence with experimental data, as shown in Figure 5a. In contrast, the use of LE and eq 12 results in a serious overestimation of the affinity, as seen in Figure 5b. In fact, if $\Delta G_{opening} + T\Delta S_{tr}$ is neglected, all but compounds **17** and **18** are predicted to bind with submillimolar affinities.

The consequence of this overestimation is illustrated for compound **6** (Figure 6). Without consideration of the ligand independent contributions, K_D for compound **6** is predicted to be about 100 μM using the equation

$$K_{D,substructure} = e^{-LE_{parent\ NNRTI} n_{HA,substructure}/(RT)} \quad (4)$$

Taking the ligand independent free energy fee into account, that is $\Delta G_{opening}$ - $T\Delta S_{tr}$, using instead the equation:

$$K_{D,substructure} = e^{(\Delta G_{ind} - LE^*_{parent\ NNRTI} n_{HA,substructure})/(RT)} \quad (5)$$

binding of compound **6** is predicted to be insignificant. Figure 6 also illustrates that the difference between predictions of the dissociation free energy for a fragment based on LE compared to LE* is increasing with decreasing size of the fragment relative the parent compound. The same is true for any error in the estimation of the ligand independent free energy fee for binding. On the other hand, any error in the K_D value for the parent compound will become less severe for the

Table 1. Summary of Data for the NNRTI Substructures

compd	n_{HA}	MW (Da)	K_D (μM)[a]	N[b]	LE (kcal mol^{-1})	LE*[c] (kcal mol^{-1})	FQ	FQ*[d]
4	13	188						
5	14	214						
6	16	244						
7	18	241						
8	19	255						
9	21	283	(800)	4	0.20	0.54	0.33	0.56
10	22	341	30	3	0.28	0.60	0.48	0.66
11	22	320	(1100)	2	0.18	0.50	0.31	0.55
12	23	355						
2	19	336	2.0	4	0.41	0.78	0.62	0.75
13	10	153						
14	8	173						
15	12	180	(700)	3	0.36	0.94	0.39	0.63
16	18	257	6	5	0.40	0.78	0.57	0.72
17	11	192						
18	11	172						
19	16	234						
20	17	248						
21	20	313	(1600)	2	0.19	0.54	0.30	0.55
22	22	310	(1700)	3	0.17	0.49	0.29	0.54
23	23	324	800	5	0.18	0.49	0.32	0.56
24	26	389	(1300)	4	0.15	0.42	0.30	0.53

[a] Values in parentheses are rough estimates of the magnitude of K_D values for compounds that appear to interact with weak affinity. [b] N is the number of dose-response curves used for the estimation of the corresponding K_D value. [c] LE* calculated by eq 2. [d] FQ* = LE*/LE_Scale* (see eqs 6 and 10).

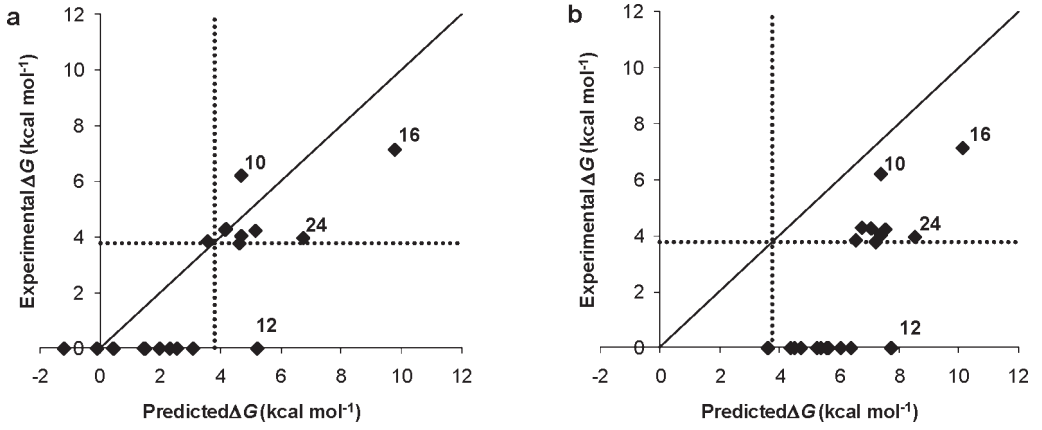

Figure 5. Correlation between predicted and experimental values of ΔG for parent NNRTIs and compounds using (a) LE* and eq 13 and (b) LE and eq 12. The dotted lines illustrate the apparent detection limit of this particular experimental setup. The compounds shown with experimental ΔG values of 0.0 are compounds for which affinities could not be determined.

prediction of dissociation free energies for smaller fragments. Thus, the rough assumption that K_D values for parent NNRTIs **1**−**3** would be in the same order of magnitude as the IC_{50} values reported in the literature will introduce less error for smaller fragments. For larger fragments, such as **16**, a mismatch between K_D and IC_{50} values would result in a larger error as seen in Figure 5 (compare K_D and IC_{50} values for NNRTI **2** in Figure 2 and Table 1).

LE, LE*, and Ranking of Compounds. By accounting for a ligand independent free energy fee, it is expected that the potential of small fragments found to interact with a target would be better appreciated (as shown by Figures 5 and 6). In the current data set, the ranking of fragments was slightly different when based on LE* compared to LE. This can be seen as smaller compounds being compensated for a relatively large effect resulting from the ligand independent free energy fee.

Smaller-sized molecules generally show higher values of LE*, as illustrated for the tested NNRTIs in Figure 7. A similar conclusion can be reached from the same analysis for all the NNRTI HIV-1 RT inhibitors reported in BindingDB (Figure 8). This trend is indicative of an uneven contribution to binding from different parts of the protein in the ligand binding pocket.

Fit Quality and LE_Scale. In an attempt to circumvent the size dependency of ligand efficiency, Reynolds et al.[9] introduced a scaling factor (LE_Scale, eq 11) to convert ligand efficiency into fit quality (FQ in eq 10). Considering the importance of ligand independent free energy terms in the estimation of ligand binding, the FQ parameter could similarly be recalculated based on LE*. The conversion factor, LE_Scale, for transforming LE into FQ is based on the maximum observed ligand efficiency for each heavy atom count. By use of 7.0 kcal mol^{-1} for ΔG_{ind} to calculate a corrected ligand efficiency LE*, the scaling could thus be converted to

$$\text{LE\_Scale}^* = RT \ln 10\, \text{LE\_Scale} - \Delta G_{ind}/n_{HA}$$
$$= 0.0975 + 17.3/n_{HA} + 35.1/n_{HA}^2 - 493/n_{HA}^3 \quad (6)$$

An FQ* value close to 1 indicates optimal binding, while a fragment with FQ* below 1 would be classified as suboptimal.

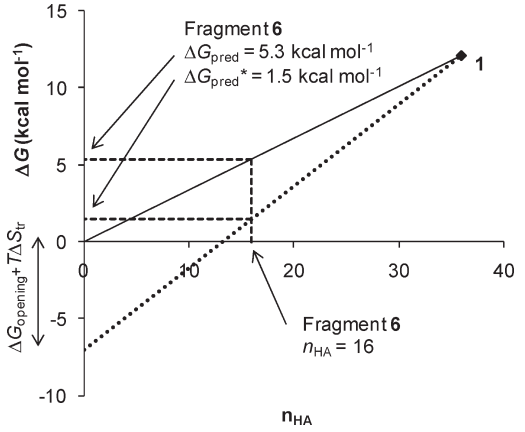

Figure 6. Effect of accounting for a ligand independent free energy fee, $\Delta G_{opening} - T\Delta S_{tr}$, in the calculation of predicted ΔG values, illustrated for compound **6**, a substructure of NNRTI **1**. Prediction of ΔG based on LE (solid line) and based on LE* (dotted line) differs by a magnitude that is dependent on the number of heavy atoms in the fragment (n_{HA}).

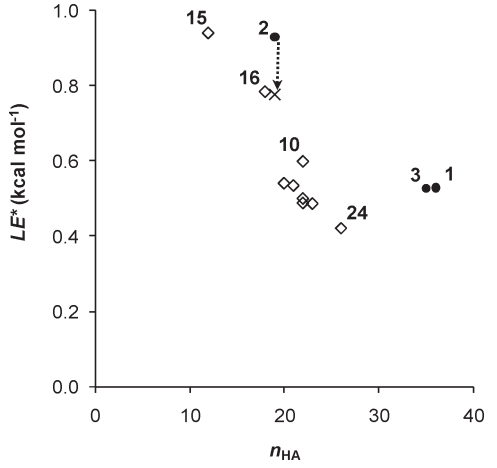

Figure 7. A more efficient use of the atoms in smaller fragments is illustrated for the studied NNRTI substructures (open diamonds). Filled circles show the parent NNRTIs. The arrow from filled circle 2 represents the shift from using a literature IC_{50} value to the K_D value of this paper. The graph corresponds to eq 2.

As an example, a ligand with a MW of 500 Da (36 heavy atoms) and K_D of 10 nM would have a LE = 0.30 kcal mol^{-1} and FQ = 0.76. The corresponding corrected ligand efficiency (LE*) would be 0.50 kcal mol^{-1} and FQ* = 0.84.

Figure 9 illustrates that all tested NNRTI substructures in this study appear suboptimal in that they neither show FQ values close to 1 nor preserve the fit quality of the parent NNRTIs (unless the difference between K_D and IC_{50} values for NNRTIs **1** and **3** would be even more different than for **2**). Performing again the same analysis using IC_{50} values for all the NNRTIs reported in BindingDB gave a similar result (Figure 10).

In an attempt to compare the ligand binding energetics of the NNIBP with ligand binding of other targets, FQ* values for four targets are plotted against n_{HA} in Figure 11. This type of graph can be interpreted as a "binding site profile", where a trend line illustrates how increasing the size of a fragment can be translated into a higher or lower fit quality.

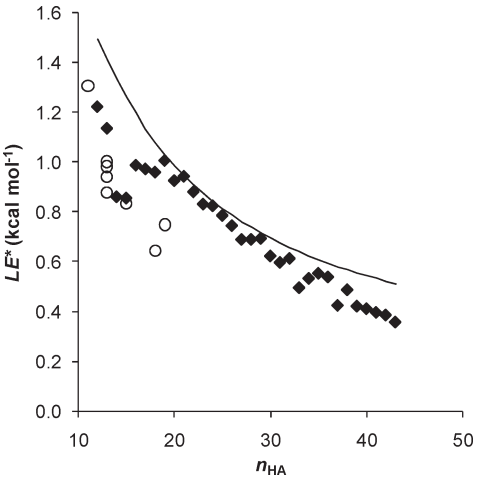

Figure 8. Illustration of the more efficient binding (higher LE*) of smaller NNRTIs, using data for the most efficient NNRTIs according to BindingDB (filled diamonds) combined with data from a fragment screen[8] (open circles). Shown is LE* (based on IC_{50}) vs number of heavy atoms (n_{HA}) for the most potent inhibitor for each molecular size. The line represents LE_Scale* (eq 6).

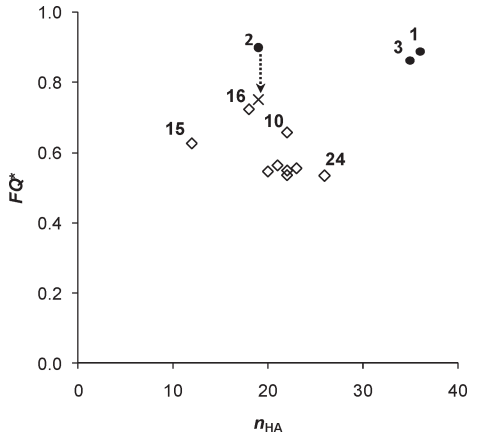

Figure 9. Suboptimal binding of NNRTI substructures illustrated by FQ* < 1 for the HIV-1 RT binding NNRTI substructures (open diamonds) and parent NNRTIs (filled circles). The arrow from filled circle 2 represents the shift from using a reported IC_{50} to the measured K_D value of this paper.

Discussion

This deconstruction study has highlighted two aspects of interactions that are typically neglected in the popular ligand efficiency metrics used for estimating the quality of fragments and ranking hits. The first concerns losses in translational and rotational degrees of freedom and the second concerns the complexities of protein−ligand interactions, both aspects that are largely independent of the characteristics of the fragments. The basis for the study is described in an accompanying paper (DOI: 10.1021/jm1010513), involving screening of a fragment library against HIV RT.[8]

HIV-1 RT is a highly flexible protein[36] with multiple functions and binding sites.[37] One particular class of HIV-1 RT inhibitors, the NNRTIs, bind to an allosteric site. This non-nucleoside inhibitor binding pocket (NNIBP) is absent in X-ray structures of the apo form of HIV-1 RT but must be transiently available for the NNRTIs to bind. Considering the absence of the NNIBP in the apo structure, the conformational

change opening the NNIBP is most likely endergonic. Binding of NNRTIs to HIV-1 RT results in considerable restrictions to the conformational flexibility of the enzyme. The inhibition of the catalytic function of HIV-1 RT by NNRTIs is considered to be an effect of an induced distortion of the polymerase active site in combination with the imposed conformational restrictions.[38,39] The endergonic nature of binding site opening in combination with imposed restrictions to the enzyme by NNRTI binding makes HIV-1 RT an excellent model system for the study of fragment screening against proteins with flexible ligand binding sites.

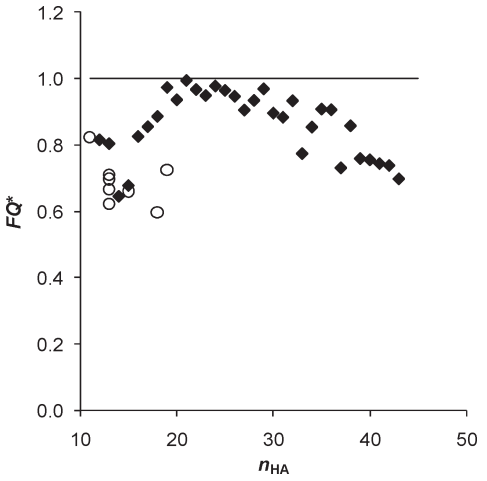

Figure 10. NNRTI binding site profile described by FQ* based on IC_{50} vs number of heavy atoms for the most potent HIV-1 RT inhibitor in each molecular size. Shown are data from BindingDB (filled diamonds) combined with data from a fragment screen (open circles).[8]

Binding of NNRTI Substructures. This study was designed and analyzed from a ligand efficiency perspective. On the basis of the simplistic assumption that the binding free energy is partitioned equally over all atoms in the parent NNRTIs, ignoring any presence of binding hot spots, most of the 21 substructures were expected to bind (Figure 5b). The result that only nine substructures were found to bind to the enzyme could thus be seen as surprising.

However, recognizing that NNRTI binding to HIV-1 RT is in fact a two step process, the ligand efficiency should be calculated differently. Accounting for ligand independent free energy fees included in the free energy of dissociation, calculated from the dissociation constant (eq 9), will result in ligand efficiency metrics more focused on the actual protein–ligand interaction. As a result, the discovery of nine binding substructures is about what could be expected from this set of substructures (Figure 5a). In addition, the nine interacting substructures are identical to those predicted to bind, with the exception of compound **12** that was found inactive. This somewhat more elaborate model again makes the

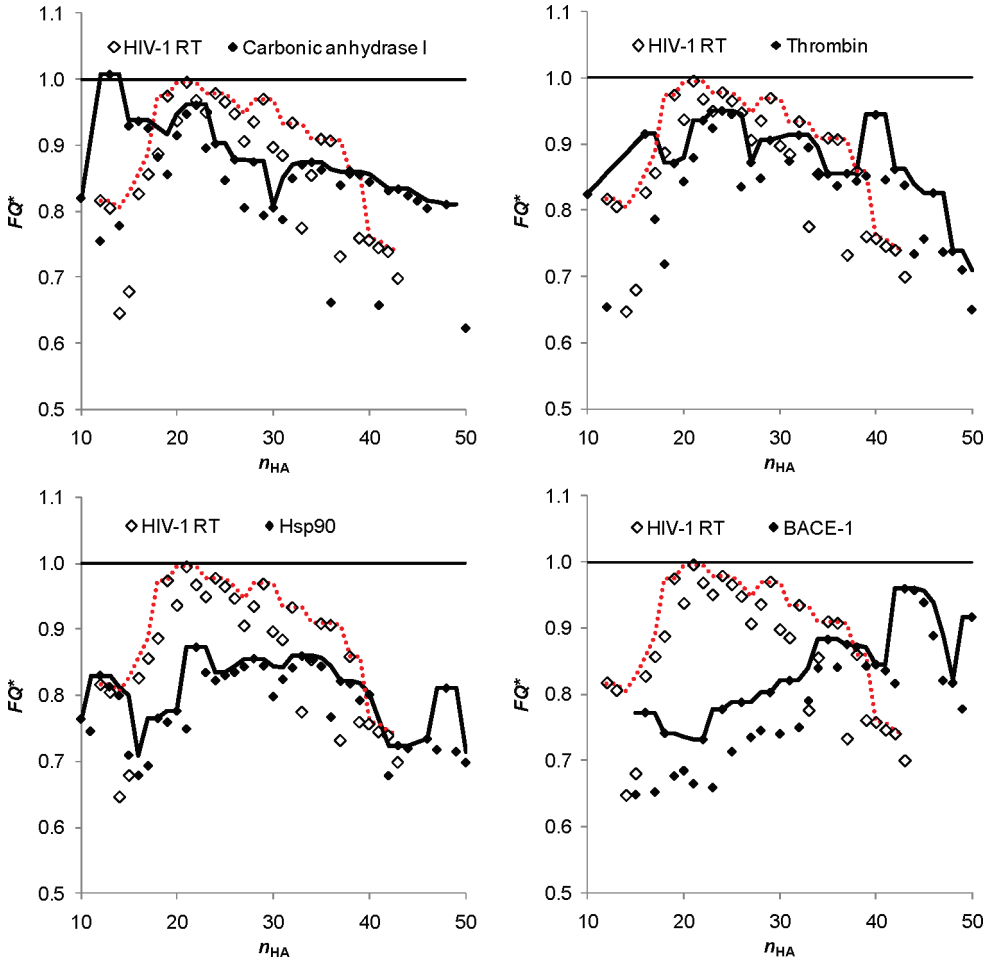

Figure 11. Comparison between the NNRTI binding site profile described by FQ* with binding sites of other targets. To make the differences more apparent, a trend line is added for each target. This line represents a moving maximum over three points. All data are from the BindingDB.

assumption that the protein−ligand interaction energy is smoothly distributed over the entire parent ligand−protein interface and that there are no hot spots that contribute larger fractions of the overall energy. Such an energy distribution would result in a constant maximum ligand efficiency for differently sized ideal fragments. As seen by the varying ligand efficiencies of the binding compounds in Figures 7 and 8, this hypothesis is not likely to hold true. Although correct in the prediction of the number of binders, the approach ignores any presence of binding hot spots in the ligand binding pocket. This neglect of binding hot spots leads to an underestimation of the potential maximum affinity of fragments.

Definition of Binding Site Profile and Ligand Binding Hot Spots. The fact that substructures predicted to bind to HIV-1 RT were found to bind and that all these substructures are relatively large suggests that the hypothesis of an equally distributed interaction energy in this case is not unreasonable. At least it may be concluded that none of the smaller tested substructures address any efficient binding hot spot of the NNIBP.

Reynolds et al. interprets the drop in maximum ligand efficiency with increasing molecular size in terms of structural compromises, loss of rotational entropy in the bound state, and a factor related to the smaller surface area available for interaction.[11] On the basis of the appreciation in this work of ligand independent contributions to the dissociation free energy, we suggest that the observed drop in ligand efficiency for larger molecules has been underestimated in the literature. Thus, we conclude that protein−ligand interaction energies generally are more unequally distributed over the participating atoms. Therefore, the maximum ligand efficiency will increase with decreasing size of the ligand. For the fragments tested in this study, this is exactly what has been found (Figure 7). Our results are in line with the conclusions drawn by Hajduk, stating that "certain protein subsites are simply 'hotter' than others".[40]

An examination of the four additional targets (Figure 11) illustrates that different targets display different binding site profiles. This originates from differences in how the LE* decreases with increasing molecular size, a fact that in turn is determined by differences in the distribution of binding energy over the specific protein−ligand interactions. For example, the decrease in LE* for the NNIBP (Figure 8) is not as steep as for carbonic anhydrase I. We interpret this as a higher homogeneity of the HIV-1 RT ligand binding site. The binding site of carbonic anhydrase I, on the other hand, shows a more focused, or energetically more differentiated, binding hot spot. As a result of this analysis, we suggest that the newly introduced concept of fit quality is in fact a less promising approach for scoring of fragment hits, since it will give a score that is based on average protein−ligand interactions instead of giving a score that is tailored for a specific target protein. However, the use of fit quality as a function of the number of heavy atoms to describe the binding site of a target seems very promising.

Estimation of the Ligand Independent Free Energy Fee. In this work, we have used $\Delta G_{ind} = 7.0$ kcal mol^{-1} for our conclusions. By use of this value, the estimation of which NNRTI substructures that are likely to bind to the NNIBP corresponded with the outcome of the experiments. However, any value between 6 and 10 kcal mol^{-1} works for rationalizing the outcome of this deconstruction study. In view of the increasing values of LE* for smaller NNRTIs, the assumption that the ligand binding pocket lacks a hot spot for fragments is not entirely correct. In fact, the more efficient binding of smaller molecules suggests that there is a preferred site for binding, although this hot spot is not as efficient as for many other targets. Therefore, $\Delta G_{ind} = 7.0$ kcal mol^{-1} could be seen as a lower limit of the ligand independent free energy fee for this system. Increasing this value in combination with the introduction of a moderately efficient ligand binding hot spot could also give estimations compatible with the experimental result.

Joining two low potency substructures into one larger ligand could be very beneficial in terms of increased potency. The term often used to describe this phenomenon is superadditivity. The value of ΔG_{ind} selected in this work is accurate enough to show that the reverse of "superadditivity", that is, "supersubtractivity", is responsible for the low affinities of substructures compared to parent compounds.

LE*, Fit Quality, and Consequences for Ranking of Hits. Evident from literature, the maximum ligand efficiency generally decreases with an increase in the number of heavy atoms in the ligand.[22] This effect is not an artifact of the method used to calculate the ligand efficiency but is rather a true effect of an increased efficiency of the interactions between an ideal ligand and a protein as the size of the ligand decreases. The substructure with the best LE* will define the binding hot spot of the ligand binding pocket. Once this interaction is occupied, larger molecules need also to interact with other parts of the ligand binding pocket. Hence, a decrease in ligand efficiency will be observed for larger molecules. As a consequence, the value of the ligand efficiency should never be used directly as a ranking measure for prioritizing fragment hits, unless they are all of identical size.

As a consequence of the introduction of the ΔG_{ind} term in the calculation of a corrected ligand efficiency, the ranking of hits from a fragment-based screen will be even more in favor of smaller fragments. Therefore, LE* will not provide a solution to the size dependence but rather increase the importance of ligand size. Although useful in the context of a deconstruction study, LE* is not a more efficient measure for prioritizing compounds in a screen-to-hit or lead generation program than is LE.

By using the method of Reynolds et al. to compensate for the fact that the maximum ligand efficiency is not independent of ligand size,[11] ligand optimization will be less likely to get stuck with small fragments being too difficult to decorate without losing ligand efficiency. Instead, medicinal chemists will get a better guidance on when to select a larger and more potent compound. However, using a scaling of the ligand efficiency for hits against a target based on maximum ligand efficiencies for other targets will result in an erroneous picture of the optimization landscape. This is indisputably shown in Figure 11 for a number of different targets. A better approach could be to use data for a similar ligand binding pocket to calculate a target class specific LE_Scale. If reference ligands are available, a deconstruction of those would give a first view of the ligand binding signature as illustrated in Figure 9.

Fragment Screening against HIV-1 RT and Other Flexible Ligand Binding Pockets. The NNRTI substructures tested in this work appear suboptimal with respect to the fit quality. This could be a result of the two-step mechanism for binding of NNRTIs, where an endergonic opening of the non-nucleoside binding pocket is required before binding. In fact, comparing

calculated ligand efficiencies based on IC_{50} or apparent K_D values for multistep processes with one-step mechanisms is unfair.

From the BindingDB, 946 structures of NNRTIs were selected. The size of these inhibitors ranged from 12 to 43 heavy atoms and IC_{50} values of 0.6 nM to 300 μM. In the plot of the LE* of these compounds against the number of heavy atoms (Figure 8), it is obvious that the size dependence of LE* is not in line with the LE_Scale* which was derived from a much larger collection of targets. The discrepancy in the high end of the number of heavy atoms could be explained by the restricted volume of the ligand binding site. More interesting from a fragment screening perspective is the deviation between LE* and LE_Scale* for the smaller NNRTIs. By transformation of the LE* into FQ*, the trend becomes even more apparent (Figure 10). Below 20 heavy atoms, the FQ* begins a dramatic drop, indicating that this target is not ideal for fragment screening. The consequence of this poorer fit quality for smaller fragments is that the screening concentration required for the discovery of all relevant fragment binders needs to be rather high, or else the hits identified in a fragment-based screen against this type of protein targets will have to be relatively large.

Conclusions

Deconstruction of known inhibitors has been shown to be an efficient way to assess how amenable a binding site is for fragment-based lead generation. Here, the deconstruction data could be used to construct a preliminary picture of a ligand binding site, defining the optimization landscape. It is expected that this type of analysis can be useful for assessing the suitability of targets for fragment-based approaches and to evaluate the quality of fragments identified upon screening. It will also be useful for experimental design, for example, the selection of a suitable fragment screening concentration, which is ultimately dependent on the behavior of the ligand binding pocket.

Experimental Section

Enzyme. Recombinant HIV-1 reverse transcriptase (BH10 isolate) was expressed in *Escherichia coli*, strain BL21 (DE3), and purified as described by Elinder et al.[17] The enzyme had an E478Q substitution in order to extinguish the RNase H activity.

NNRTI Substructures. Compounds **2**, **4**, **5**, **9**, **13**, **15**, **17**, **18**, **20**, **22**, **23**, and **24** were purchased from Enamine (Kiev, Ukraine). Compounds **6**, **7**, **8**, **10**, **11**, **12**, **14**, **16**, **19**, and **21** were purchased from Vitas-M Laboratory (Moscow, Russia). All purchased compounds were used without further purification. Purity and identity of compounds were assessed by suppliers using LCMS and/or 1D ^{1}H NMR. Purities of Enamine compounds range from 90% to 95%, whereas compounds from Vitas-M Laboratory are more than 95% pure.

Interaction Analysis Setup. Experiments were performed with a Biacore S51 instrument (GE Healthcare) at 25 °C. Immobilization of HIV-1 reverse transcriptase was achieved essentially as described previously.[14−16] Surface densities between 10 and 15 kRU of HIV-1 RT, prepared by amine coupling to CM5 sensor chips (GE Healthcare), were used for the compound characterizations.

K_D Determination. The compounds were injected in 2-fold dilution series (500−3.9 μM) for 30 s at a flow rate of 90 μL/min. Nonspecific signals were removed by subtraction of signals from a reference channel. Corrections to compensate for differences in DMSO concentrations between running buffer and samples (bulk refractive index calibration) were also performed. Finally, data were blank subtracted and thereafter normalized with respect to the molecular weight of the injected compound. Compound **16** (50 μM) was used to assess the binding capacity of the immobilized enzyme, i.e., to determine how the R_{max} decreased with increasing number of injections. Affinities (K_D) were determined by fitting the modified Langmuir isotherm to the responses in the report point "binding late", BL, for all compounds:

$$BL = \frac{R_{max}[L]}{[L] + K_D} + u[L] + m \quad (7)$$

The report point "binding late" corresponds to the response 4 s before the end of the injection, i.e., 26 s after injection. R_{max} is the maximal response at saturation, determined by fitting the parameters of eq 7 to the data of the well behaved compound **16**. The variable [L] is the concentration of the compound. The constant u compensates for any linear low affinity component and/or signals due to secondary effects, including concentration dependent differences in bulk refractive index not taken full care of by double referencing or by bulk refractive index calibration. The constant m is an offset for each concentration series used for compensating for small errors in blank response corrections. The K_D value is thus estimated on the dose-response data after correction for experimental disturbances. The data analysis was performed using the Sprint software (Beactica AB, Uppsala, Sweden).

Calculation of Ligand Efficiency. Ligand efficiency was calculated by dividing the dissociation free energy by the number of heavy atoms:

$$LE = \Delta G / n_{HA} \quad (8)$$

where n_{HA} is the number of heavy atoms. The dissociation free energy was calculated by

$$\Delta G = -RT \ln K_D \quad (9)$$

using a temperature of 298 K. For compounds **1**−**3**, K_D values are not available in the literature. Therefore, we based the calculations on reported IC_{50} values, although strictly, this is inappropriate. Nevertheless, the order of magnitude is expected to be reasonable for this initial approach.

The corrected ligand efficiency LE* was earlier defined as

$$LE^* = [\Delta G_{opening} - T\Delta S_{tr} - RT \ln(K_D)]/n_{HA} \quad (2)$$

where $\Delta G_{opening}$ is the free energy required to open the NNRTI binding site and $-T\Delta S_{tr}$ is the estimated change in free energy caused by a loss of translational and rotational entropy by a ligand when bound to a protein. The sum of these two terms is denoted the "ligand independent free energy fee" (ΔG_{ind}):

$$\Delta G_{ind} = \Delta G_{opening} - T\Delta S_{tr} \quad (3)$$

Calculation of Fit Quality. The fit quality (FQ), introduced by Reynolds et al.[10−12] was calculated as

$$FQ = LE/LE\_Scale \quad (10)$$

where

$$LE\_Scale = 0.0975 + 10.3/n_{HA} + 35.1/n_{HA}^2 - 493/n_{HA}^3 \quad (11)$$

Estimation of ΔG Values for Compounds. In the initial naive estimations of affinities for purchased NNRTI substructures, the ligand efficiency (LE) calculated for the parent NNRTI was multiplied by the number of heavy atoms in the compound to get a rough estimate of the dissociation energy:

$$\Delta G_{substructure} = LE_{parent\ NNRTI} n_{HA,substructure} \quad (12)$$

From this, it was assumed that the reported IC_{50} values were of the same order of magnitude as the corresponding K_D values. In the estimation of ΔG for the compounds, taking the ligand

independent free energy fees into account, ΔG was calculated using LE* of the parent NNRTI multiplied by the number of heavy atoms of the compound subtracted by ΔG_{ind}:

$$\Delta G_{substructure} = LE^*_{parent\ NNRTI} n_{HA,substructure} - \Delta G_{ind} \quad (13)$$

External Data Set. A data set consisting of IC_{50} and K_i data for four different targets in addition to HIV-1 RT was assembled from the binding database (BindingDB, http://www.bindingdb.org).[41] The data extracted from the database were K_i values for carbonic anhydrase I, and IC_{50} values for thrombin, heat shock protein 90 (Hsp90), and β-secretase (BACE-1). K_i and IC_{50} values together with structures were exported from the Web site as an sd file. This was imported into Instant JChem (ChemAxon Ltd.), and the number of non-hydrogen atoms, n_{HA}, was calculated to filter out the most potent ligands for each molecular size.

Acknowledgment. The authors acknowledge Malin Elinder for production of protein and for assistance in the preparation of this manuscript.

References

(1) Erlanson, D. A.; McDowell, R. S.; O'Brien, T. Fragment-based drug discovery. *J. Med. Chem.* **2004**, *47*, 3463–3482.
(2) Congreve, M.; Chessari, G.; Tisi, D.; Woodhead, A. J. Recent developments in fragment-based drug discovery. *J. Med. Chem.* **2008**, *51*, 3661–3680.
(3) de Kloe, G. E.; Bailey, D.; Leurs, R.; de Esch, I. J. P. Transforming fragments into candidates: small becomes big in medicinal chemistry. *Drug Discovery Today* **2009**, *14*, 630–646.
(4) Leach, A. R.; Hann, M. M.; Burrows, J. N.; Griffin, E. J. Fragment screening: an introduction. *Mol. BioSyst.* **2006**, *2*, 429–446.
(5) Hubbard, R. E.; Chen, I.; Davis, B. Informatics and modelling challenges in fragment-based drug discovery. *Curr. Opin. Drug Discovery Dev.* **2007**, *10*, 289–297.
(6) Everts, S. Piece by piece. *Chem. Eng. News* **2009**, *86*, 15–23.
(7) Gill, A. New lead generation strategies for protein kinase inhibitors: fragment based screening approaches. *Mini-Rev. Med. Chem.* **2004**, *3*, 301–311.
(8) Geitmann, M.; Elinder, M.; Seeger, C.; Brandt, P.; de Esch, I. J. P.; Danielson, U. H. Identification of a novel scaffold for allosteric inhibition of wild type and drug resistant HIV-1 reverse transcriptase by fragment library screening. *J. Med. Chem.* DOI: 10.1021/jm1010513.
(9) Hopkins, A. L.; Groom, C. R.; Alex, A. Ligand efficiency: a useful metric for lead selection. *Drug Discovery Today* **2004**, *9*, 430–431.
(10) Reynolds, C. H.; Bembenek, S. D.; Tounge, B. A. The role of molecular size in ligand efficiency. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 4258–4261.
(11) Reynolds, C. H.; Tounge, B. A.; Bembenek, S. D. Ligand binding efficiency: trends, physical basis, and implications. *J. Med. Chem.* **2008**, *51*, 2432–2438.
(12) Bembenek, S. D.; Tounge, B. A.; Reynolds, C. H. Ligand efficiency and fragment-based drug discovery. *Drug Discovery Today* **2009**, *14*, 278–283.
(13) For a recent paper on fragment-based deconstruction of inhibitors, see the following: Barelier, S.; Pons, J.; Marcillat, O.; Lancelin, J.-M.; Krimm, I. Fragment-based deconstruction of Bcl-xL Inhibitors. *J. Med. Chem.* **2010**, *53*, 2577–2588.
(14) Geitmann, M.; Unge, T; Danielson, U. H. Biosensor-based kinetic characterization of the interaction between HIV-1 reverse transcriptase and non-nucleoside inhibitors. *J. Med. Chem.* **2006**, *49*, 2367–2374.
(15) Geitmann, M.; Unge, T; Danielson, U. H. Interaction kinetic characterization of HIV-1 reverse transcriptase non-nucleoside inhibitor resistance. *J. Med. Chem.* **2006**, *49*, 2375–2387.
(16) Geitmann, M.; Danielson, U. H. Additional level of information about complex interaction between non-nucleoside inhibitor and HIV-1 reverse transcriptase using biosensor-based thermodynamic analysis. *Bioorg. Med. Chem.* **2007**, *15*, 7344–7354.
(17) Elinder, M.; Nordström, H.; Geitmann, M.; Hämäläinen, M.; Vrang, L.; Öberg, B.; Danielson, U. H. Screening for NNRTIs with slow dissociation and high affinity for a panel of HIV-1 RT variants. *J. Biomol. Screening* **2009**, *14*, 395–403.
(18) Danielson, U. H. Fragment library screening and lead characterization using SPR biosensors. *Curr. Top. Med. Chem.* **2009**, *9*, 1725–1735.
(19) Liu, T.; Lin, Y.; Wen, X.; Jorrisen, R. N.; Gilson, M. K. BindingDB: a web-accessible database of experimentally determined protein−ligand binding affinities. *Nucleic Acids Res.* **2007**, *35*, D198–D201.
(20) Chen, X.; Lin, Y.; Liu, M.; Gilson, M. K. The binding database: data management and interface design. *Bioinformatics* **2002**, *18*, 130–139.
(21) de Kloe, G. E.; Retra, K.; Geitmann, M.; Källblad, P.; Smit, A. B.; van Muijlwijk-Koezen, J. E.; Leurs, R.; Irth, H.; Danielson, U. H.; de Esch, I. J. P. Surface plasmon resonance biosensor based fragment screening using acetylcholine binding protein identifies ligand efficiency hot spots (LE hot spots) by deconstruction of nicotinic acetylcholine receptor alpha7 ligands. *J. Med. Chem.* **2010**, *53*, 7192–7201.
(22) Kuntz, I. D.; Chen, K.; Sharp, K. A.; Kollman, P. A. The maximal affinity of ligands. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9997–10002.
(23) Definition of ligand efficiency by Peter Kenny: $(pIC_{50} - \log P)/n_{HA}$ described at http://fbdd-lit.blogspot.com.
(24) Leeson, P. D.; Springthorpe, B. The influence of drug-like concepts on decision-making in medicinal chemistry. *Nat. Rev. Drug Discovery* **2007**, *6*, 881–890.
(25) Keserü, G. M.; Makara, G. M. The influence of lead discovery strategies on the properties of drug candidates. *Nat. Rev. Drug Discovery* **2009**, *8*, 203–212.
(26) Page, M. I.; Jencks, W. P. Entropic contributions to rate accelerations in enzymic and intramolecular reactions and the chelate effect. *Proc. Nat. Acad. Sci. U.S.A.* **1971**, *68*, 1678–1683.
(27) Jencks, W. P. On the attribution and additivity of binding energies. *Proc. Nat. Acad. Sci. U.S.A.* **1981**, *78*, 4046–4050.
(28) Andrews, P. R.; Craik, D. J.; Martin, J. L. Functional group contributions to drug−receptor interactions. *J. Med. Chem.* **1984**, *27*, 1648–1657.
(29) Štrajbl, J. V.; Glennon, T. M.; Sham, Y. Y.; Chu, Z. T.; Warshel, A. How important are entropic contributions to enzyme catalysis? *Proc. Nat. Acad. Sci. U.S.A.* **2000**, *97*, 11899–11904.
(30) Borsi, V.; Calderone, V.; Fragai, M.; Luchinat, C.; Sarti, N. Entropic contribution to the linking coefficient in fragment based drug design: a case study. *J. Med. Chem.* **2010**, *53*, 4285–4289.
(31) Murray, C. W.; Verdonk, M. L. The consequences of translational and rotational entropy lost by small molecules on binding to proteins. *J. Comput.-Aided Mol. Des.* **2003**, *16*, 741–753.
(32) Chang, C. A.; Chen, W.; Gilson, M. K. Ligand configurational entropy and protein binding. *Proc. Nat. Acad. Sci. U.S.A.* **2008**, *104*, 1534–1539.
(33) Zhou, H.-X.; Gilson, M. K. Theory of free energy and entropy in noncovalent binding. *Chem. Rev.* **2009**, *109*, 4092–4107.
(34) Lou, H.; Sharp, K. On the calculation of absolute macromolecular binding free energies. *Proc. Nat. Acad. Sci. U.S.A.* **2002**, *99*, 10399–10404.
(35) NNRTI **1**: Chan, J. H.; Freeman, G. A.; Tidwell, J. H.; Romines, K. R.; Schaller, L. T.; Cowan, J. R.; Gonzales, S. S.; Lowell, G. S.; Andrews, C. W., 3rd; Reynolds, D. J.; St Clair, M.; Hazen, R. J.; Ferris, R. G.; Creech, K. L.; Roberts, G. B.; Short, S. A.; Weaver, K.; Koszalka, G. W.; Boone, L. R. Novel benzophenones as non-nucleoside reverse transcriptase inhibitors of HIV-1. *J. Med. Chem.* **2004**, *47*, 1175–1182. NNRTI 2: Bell, F. W.; Cantrell, A. S.; Hoegberg, M.; Jaskunas, S. R.; Johansson, N. G.; Jordan, C. L.; Kinnick, M. D.; Lind, P.; Morin, J. M., Jr. Phenethylthiazolethiourea (PETT) compounds, a new class of HIV-1 reverse transcriptase inhibitors. 1. Synthesis and basic structure−activity relationship studies of PETT analogs. *J. Med. Chem.* **1995**, *38*, 4929–4936. NNRTI 3: Zhang, Z.; Xu, W.; Koh, Y. H.; Shim, J. H.; Girardet, J. L.; Yeh, L. T.; Hamatake, R. K.; Hong, Z. A novel nonnucleoside analogue that inhibits human immunodeficiency virus type 1 isolates resistant to current nonnucleoside reverse transcriptase inhibitors. *Antimicrob. Agents Chemother.* **2007**, *51*, 429–437.
(36) Paris, K. A.; Haq, O.; Felts, A. K.; Das, K.; Arnold, E.; Levy, R. M. Conformational landscape of the human immunodeficiency virus type 1 reverse transcriptase non-nucleoside inhibitor binding pocket: lessons for inhibitor design from a cluster analysis of many crystal structures. *J. Med. Chem.* **2009**, *52*, 6413–6420.
(37) Sarafianos, S. G.; Marchand, B.; Das, K.; Himmel, D. M.; Parniak, M. A.; Hughes, S. H.; Arnold, E. Structure and function of HIV-1 reverse transcriptase: molecular mechanisms of polymerization and inhibition. *J. Mol. Biol.* **2009**, *385*, 693–713.
(38) Ren, J.; Esnouf, R.; Hopkins, A.; Ross, C.; Jones, Y.; Stammers, D.; Stuart, D. The structure of HIV-1 reverse transcriptase complexed with 9-chloro-TIBO: lessons for inhibitor design. *Structure* **1995**, *3*, 915–926.
(39) Kohlstaedt, L. A.; Wang, J.; Friedman, J. M.; Rice, P. A.; Steitz, T. A. Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science* **1992**, *256*, 1783–1790.
(40) Hajduk, P. J. Fragment-based drug design: how big is too big? *J. Med. Chem.* **2006**, *49*, 6972–6976.
(41) Liu, T.; Lin, Y.; Wen, X.; Jorrisen, R. N.; Gilson, M. K. BindingDB: a web-accessible database of experimentally determined protein−ligand binding affinities. *Nucleic Acids Res.* **2007**, *35*, D198–D201.