Hierarchical Database Screenings for HIV-1 Reverse Transcriptase Using a Pharmacophore Model, Rigid Docking, Solvation Docking, and MM−PB/SA

Junmei Wang,*,† Xinshan Kang,‡,§ Irwin D. Kuntz,‡ and Peter A. Kollman‡,||

Encysive Pharmaceuticals Inc., 7000 Fannin, Houston, Texas 77030, and Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143-0446

Received May 27, 2004

In this work, an efficient strategy was presented to search drug leads for human immunodeficiency virus type 1 reverse transcriptase (HIV-1 RT) using hierarchical database screenings, which included a pharmacophore model, multiple-conformation rigid docking, solvation docking, and molecular mechanics−Poisson−Boltzmann/surface area (MM−PB/SA) sequentially. Encouraging results were achieved in searching a refined available chemical directory (ACD) database: the enrichment factor after the first three filters was estimated to be 25-fold; the hit rate for all the four filters was predicted to be 41% in a control test using 37 known HIV-1 non-nucleoside reverse transcriptase inhibitors; 10 out of 30 promising solvation-docking hits had MM−PB/SA binding free energies better than −6.8 kcal/mol and the best one, HIT15, had −17.0 kcal/mol. In conclusion, the hierarchical multiple-filter database searching strategy is an attractive strategy in drug lead exploration.

Introduction

Lead discovery is one of the most important components in rational drug design. There are two basic approaches to identify drug leads, which are de novo design and database screening. De novo design is aimed to design novel compounds that have electrostatic and hydrophobic properties complementary to those of the target. Thus, it requires 3D structures of drug targets. On the other hand, database screening applies filters to identify potential drug leads from real or virtual databases. Although de novo design is more attractive in concept, database screening is still the dominating approach in drug lead exploration. All database screening approaches fall into two categories: the query-based and the scoring-function-based approaches. Typically, the scoring-function-based approaches, but not necessarily the query-based approaches, require 3D structures of the drug targets.

Query-based approaches apply search queries to filter databases. Queries can be as simple as molecular weights or numbers of hydrogen-bond donors and acceptors or as complicated as 2D or 3D molecular structures and pharmacophore models. A well-known filter in this category is Lipinski's "rule-of-five", which states that a good drug candidate should have a molecular weight smaller than 500, a calculated log *P* smaller than 5.0, and numbers of hydrogen-bond donors and acceptors less than 5 and 10, respectively.[1] It has been estimated that roughly 10% of compounds that enter development eventually become marketed drugs and 40% of compounds fail due to poor pharmacokinetic properties.[2] Recently, more and more attention has been focused on the development of drug leads that have not only good drug potencies and selectivity but also good ADMET (absorption, distribution, metabolism, excretion, and toxicity) and PK (pharmacokinetics) properties. It is realized that a "parallel model" of lead optimization, which optimizes drug potencies and ADMET/PK properties simultaneously, is superior to the traditional serial−cyclical model of lead optimization.[3] As a result, it is important to develop reliable ADMET/PK models and apply them as queries in database screening. Query-based database screenings are typically efficient and less computer resource demanding, since they do not require the 3D structures of drug targets. However, inappropriate queries may generate too many or too few hits.

Scoring-function-based approaches apply target functions to select hits. The widely used target functions include those that calculate the free energy of an inhibitor binding to a receptor. The 3D structure of the receptor, either from X-ray, NMR, or a homology model, is the prerequisite of this kind of database search. The most rigorous and accurate methods of calculating the free energy of an inhibitor binding to a receptor are free energy perturbation (FEP) and thermodynamic integration (TI). However, the two methods are extremely time-consuming and not appropriate for database screenings. Recently, several alternative methods, which include the linear interaction energy (LIE) method,[4,5] solvation docking,[6] MM−GB/SA (generalized Born/solvent accessible surface area),[7] and MM−PB/SA,[8−11] have been developed to estimate binding free energies in fast and practical ways. LIE is based on a linear response hypothesis, which assume that the free energy of an inhibitor binding to a receptor is the combination of the weighted electrostatic and van der Waals interaction energies between the inhibitor and receptor. The solv-

* To whom correspondence should be addressed. Tel: (713) 578-6649. Fax: (713) 578-6720. E-mail: jwang@encysive.com.
† Encysive Pharmaceuticals Inc.
‡ University of California.
§ Current address: PTC Therapeutics Inc., 100 Corporate Court, South Plainfield, NJ 07080.
|| Deceased.

10.1021/jm049606e CCC: $30.25
Published on Web 02/19/2005

ation docking, which was initially proposed by Zou et al.,[6] is similar to LIE, except that the solvation energy is estimated using GB/SA. In contrast to LIE and solvation docking shown in formula I, MM−GB/SA and MM−PB/SA apply a thermodynamic cycle to calculate the binding free energy of A + B → AB.

$$\begin{array}{ccccc} A_{aqueous} & + & B_{aqueous} & \xrightarrow{\Delta G_{binding}} & AB_{aqueous} \\ \downarrow -\Delta G^{A}_{solv} & & \downarrow -\Delta G^{B}_{solv} & & \downarrow -\Delta G^{AB}_{solv} \\ A_{gas} & + & B_{gas} & \xrightarrow{\Delta G_{gas}} & AB_{gas} \end{array}$$

$$\Delta G_{\text{binding}} = \alpha \Delta G_{\text{elec}} + \beta \Delta E_{\text{vdw}} + \gamma \Delta \text{SAS} \quad \text{(I)}$$

$$= \Delta H_{\text{gas}} - T\Delta S + (\Delta G^{\text{AB}}{}_{\text{PB/GBSA}} - \Delta G^{\text{A}}{}_{\text{PB/GBSA}} - \Delta G^{\text{B}}{}_{\text{PB/GBSA}}) \quad \text{(II)}$$

$$\Delta H_{\text{gas}} \approx \Delta E_{\text{gas}} = \Delta E_{\text{internal}} + \Delta E_{\text{electrostatic}} + \Delta E_{\text{vdw}} \quad \text{(III)}$$

$$\Delta G_{\text{PBSA/GBSA}} = \Delta G_{\text{PB/GB}} + \Delta G_{\text{SA}} \quad \text{(IV)}$$

$$\Delta G_{\text{binding}} = \Delta E_{\text{electrostatic}} + \Delta E_{\text{vdw}} \quad \text{(V)}$$

ΔG_{gas} is the interaction energy between A and B in the gas phase; $\Delta G^{\text{A}}{}_{\text{solv}}$, $\Delta G^{\text{B}}{}_{\text{solv}}$, and $\Delta G^{\text{AB}}{}_{\text{solv}}$ are the solvation free energies of A, B, and AB with GB/SA or PB/SA models, respectively; $\Delta G_{\text{binding}}$, the binding free energy, is calculated with formulas II−IV. To facilitate the absolute and relative binding free energy calculations, ΔE_{inter} is typically neglected with the assumption that the intramolecular energy of the ligand does not change significantly upon binding. In contrast to the linear interaction energy method, MM−GB/SA and MM−PB/SA do not apply empirical parameters at all for free energy calculations, which makes them attractive methods for directly estimating the binding free energies. Recently, we successfully applied molecular docking combined with MM−PB/SA to determining the binding mode of HIV-1 RT/efavirenz.[9] In this blind test, not only was the calculated binding free energy in good agreement with the experiment, but the crystal structure, which was released after our paper had been submitted, was also well-predicted by the combination of molecular docking and molecular dynamic simulations (the rmsd of the non-nucleoside reverse transcriptase inhibitor and 54 α-carbons of the key residues around the binding site was 1.1 Å). If the solvation energies ($\Delta G^{\text{AB}}{}_{\text{solv}}$, $\Delta G^{\text{A}}{}_{\text{solv}}$, and $\Delta G^{\text{B}}{}_{\text{solv}}$) and entropy ($T\Delta S$) contributions are omitted, formula II becomes formula V. The scoring function described by formula V, which only considers the gas-phase intermolecular energy ($\Delta E_{\text{elec}}+\Delta E_{\text{vdw}}$) between the inhibitor and the receptor, is widely used in docking programs[12,13] because of its simplicity. In this study, one of our screening filters, rigid docking, also applied this scoring function.

However, a good scoring function is only part of the story. Conformational sampling also plays an important role in calculating the binding free energy accurately and efficiently. There are a variety of sampling protocols available, including molecular dynamics and Monte Carlo (MC) simulations, genetic algorithms, and simulated annealing. The conformational flexibilities of an inhibitor and a receptor can be taken into account at hierarchal levels, ranging from the highly expensive level, with both the inhibitor and receptor being flexible, to the less costly, which makes the inhibitor flexible but the receptor rigid, to the least expensive one, with both the inhibitor and receptor being rigid.

What is the state of the art in applying the above scoring functions and sampling protocols in binding free energy calculations? A wise and often necessary strategy is one that employs the rapid and lower level methods (such as rigid docking using the scoring function described by formula V) at the beginning and turns to the more accurate and quantitative methods (such as MM−PB/SA analysis sampled by molecular dynamics simulations) at the end. Naturally, it is wise to combine expensive sampling protocols, such as MD and MC simulations, with expensive energy calculation approaches such as free energy perturbations, TI, and MM−PB/SA.

Most docking programs apply much simpler scoring functions and only the flexibilities of inhibitors are taken into account. There are three different approaches to consider the flexibility of an inhibitor. The commonly used approach is to take the torsional angles of rotatable bonds in a ligand as variables in addition to the three translation vectors and three Euler angles in docking optimizations. The flexibility of a ligand can also be taken into account implicitly through anchor docking followed by molecular fragment growing up and assembling. In the third approach, a set of diverse conformations of a ligand is first generated and then rigid docking is performed for each conformation, and the one that gives the best docking score is taken as the "active" conformation of the ligand. One often needs to find a balance between accuracy and efficiency, because the more physical a scoring function is and the more conformational space it is sampled, the more time-consuming a database search will be.

HIV-1 reverse transcriptase is an important target in AIDS-related drug design. The main biological role of HIV-1 RT is to transcribe the viral RNA into a double-stranded DNA. This is a crucial step in the life cycle of HIV-1. The viral nucleic acid is accommodated in a large groove through the enzyme. At one point in this groove, the dNTP site is located, formed partly by a nucleic acid template primer and partly by residues in the enzyme; this is the site of action of phosphorylated nucleoside RT inhibitors. In addition, HIV-1 RT has an allosteric site, which is colocated in the p66 palm subdomain but distinct from the dNTP binding site. The inhibitors that target the allosteric site are so-called non-nucleoside reverse transcriptase inhibitors (NNRTIs). We have reported the mechanism of allosteric binding elsewhere from a conformation point of view.[9] Recently several NNRTIs have been approved by FDA as anti-AIDS drugs, which include nevirapine, delavirdine, and efavirenz.[14−17]

In recent years, many crystal structures of HIV-1 RT in complex with NNRTIs have been solved[18−29] and tens of NNRTIs have been identified. This makes HIV-1 RT a good target to test drug lead exploration strategies. In this study, we have applied a set of hierarchical filters, from a simple query-based pharmacophore model to a much more promising free-energy-based scoring function, MM−PB/SA, to screen the ACD database for

Table 1. A List of Known HIV-1 NNRTIs

no.	compd abbrev	comp name/code	PDB code[a]	IC_{50} (μM)[b]	ΔG_{expt}[c] (kcal/mol)	ref
1	AAA	2,6-dibromo-α-APA; R95845	1HNI	0.42	−8.7	29
2	AAP	2,6-dichloro-α-APA; R90385	1VRU	0.1	−9.5	20
3	EFZ	L-743726; DMP-266; efavirenz	1FK9 1FKO 1IKW 1IKV 1JKM	0.00293	−11.6	30
4	HBY	HBY 097; quinoxaline derivative	1BQM	0.006	−11.2	31
5	HEF	HEPT; 1-[(2-hydroxyethoxy)methyl]-6-(phenylthio)thymine	1RVM 1RTI	7	−7.0	32
6	IIN	NSC119833; 2-methyl-3-amino-3-phenylisoindol-1-one	1RVN	10	−6.8	23
7	MKC	emivirine; coactinon; MKC-442; I-EBU; HEPT derivative	1RT1	0.008	−11.0	24
8	NEV	(R)-nevirapine; BI-RG-587	1VRT 3HVT	0.2	−9.1	29
9	SPP	*S*-cysteinesulfinic acid; *S*-sulfinocysteine; 1-(5-methansulfonamido-1*H*-indol-2-ylcarbonyl)-4-[methylaminopyidinyl]piperazine	1KLM	1.1	−8.1	33
10	TB9	9-chloro-TIBO; R82913	1TVR 1RVQ	0.033	−10.6	26
11	TBO	8-chloro-TIBO; R86183	1UWB 1HNV	0.003	−11.9	26
12	THI	thiazolo-3-phenylisoin dol-1-one	1RVP	0.016	−10.70	34
13	TNK	TNK-651; 6-benzyl-1-benzyloxymethyl-5-isopropyluracil	1RT2	0.006	−11.2	24
14	U05	1051U91	1RT3	0.4	−8.7	20
15	UC1	UC-781	1RT4	0.009	−11.0	28
16	UC2	UC-10; carboxanilide derivative	1RT5	0.142	−9.3	28
17	UC3	UC-38	1RT6	0.029	−10.3	28
18	UC4	UC-84; carboxanilide derivative	1RT7	0.042	−10.1	28
19	UKC	UK-129485; imidazodipyridodiazepine	1RVR	0.19	−9.23	23
20	ADA	ADAM analogue; alkenyldiarylmethane analog		16.0	−6.5	35
21	ATA	aurintricarboxylic acid; Dupont ATA; Dupont DA 639; SD-095345		5.0	−7.2	36
22	DAB	DABO; 6-benzyl-5-methyl-2-(cyclohexyloxy)pyrimidin-4-one		4.7	−7.3	37
23	D83	DPC083; DMP-266 derivative		0.0004	−12.8	d
24	D61	DPC961; DMP-266 derivative		0.00036	−12.9	d
25	EEB	EEBU; HEPT derivative; MKC-442 derivative		0.041	−10.1	39
26	HES	HEPT-S; 1-[2-hydroxyethoxymethyl]-6-phenylthiothymine		1.6	−7.9	38
27	HET	HEPT and MKC-442 derivatives		0.0027	−11.7	39
28	INO	inophyllum B		1.4	−8.0	40
29	L69	L-697639; 2-pyridinone-3-benzoxazole MeNH derivative		0.05	−10.0	41
30	PE1	trovirdine; LY-300046; PETT derivative		0.02	−10.5	42
31	PE2	PETT Cl, F derivative		0.05	−10.0	42
32	PE3	PETT derivative		0.01	−10.9	42
33	PE4	PETT derivative; LY 73497		1.3	−8.0	42
34	R14	R14458; TIBO analogue		62	−5.7	43
35	R87	R87232; α-APA derivative		0.033	−10.2	44
36	R88	R88703; phNH-phAcNH derivative; α-APA deriv		0.026	−10.3	44
37	S11	capravirine; S1153		0.0031	−11.6	45

[a] Crystal structures of HIV-1 RT complexes were available for no. 1−19. [b] IC_{50} is defined as the effective concentration that inhibits 50% of viral production, 50% of viral infectivity, or 50% of the virus-induced cytopathic effect. IC_{50} is also commonly referred to as ED_{50} or EC_{50}. If more than one IC_{50} value is available, the MT-4 cell and HIV-1 (IIIB) strain's values are first adopted. [c] Experimental binding free energies were estimated using IC_{50}.

drug leads for HIV-1 RT. This virtual screening strategy is intended to efficiently and reliably condense "real inhibitors" in a database rather than to replace the true high throughput (HTS) screening. The combination of virtual and real HTS screenings can significantly improve the successful rate of finding real inhibitors in a database with less time and lower financial cost. It is notable that this protocol can be applied to other systems only if the receptor structures are available.

To evaluate the performance of a database screening, hit rate (HR) and enrichment factor (EF) are defined in formulas VI and VII, where n is the number of compounds in the whole database that includes q known inhibitors. Suppose m hits, which include p known inhibitors, passed one or multiple screening filters; HR is then defined as the ratio of p to q (formula VI) and EF is the ratio of p_m to p_n, where p_m and p_n are probabilities of finding a real inhibitor among the hits (p/m) and in the whole database (q/n), respectively (formula VII).

$$\mathrm{HR} = p/q \tag{VI}$$

$$\mathrm{EF} = \frac{p_m}{p_n} = \mathrm{HR} \times \frac{n}{m} \tag{VII}$$

Computational Methods

Filter I—A Pharmacophore Model. Thirty-seven known HIV-1 NNRTIs binding at the allosteric site have been selected as a control set to evaluate the performance of the database screening strategy.[18,20,26,28−45] The experimental binding free energies of those NNRTIs are listed in Table 1 and the structures are shown in Figure 1. The crystal structures of HIV-1 RT in complex with 19 out of the 37 NNRTIs (Compounds **1**−**19** in Table 1) were superimposed on a crystal structure of

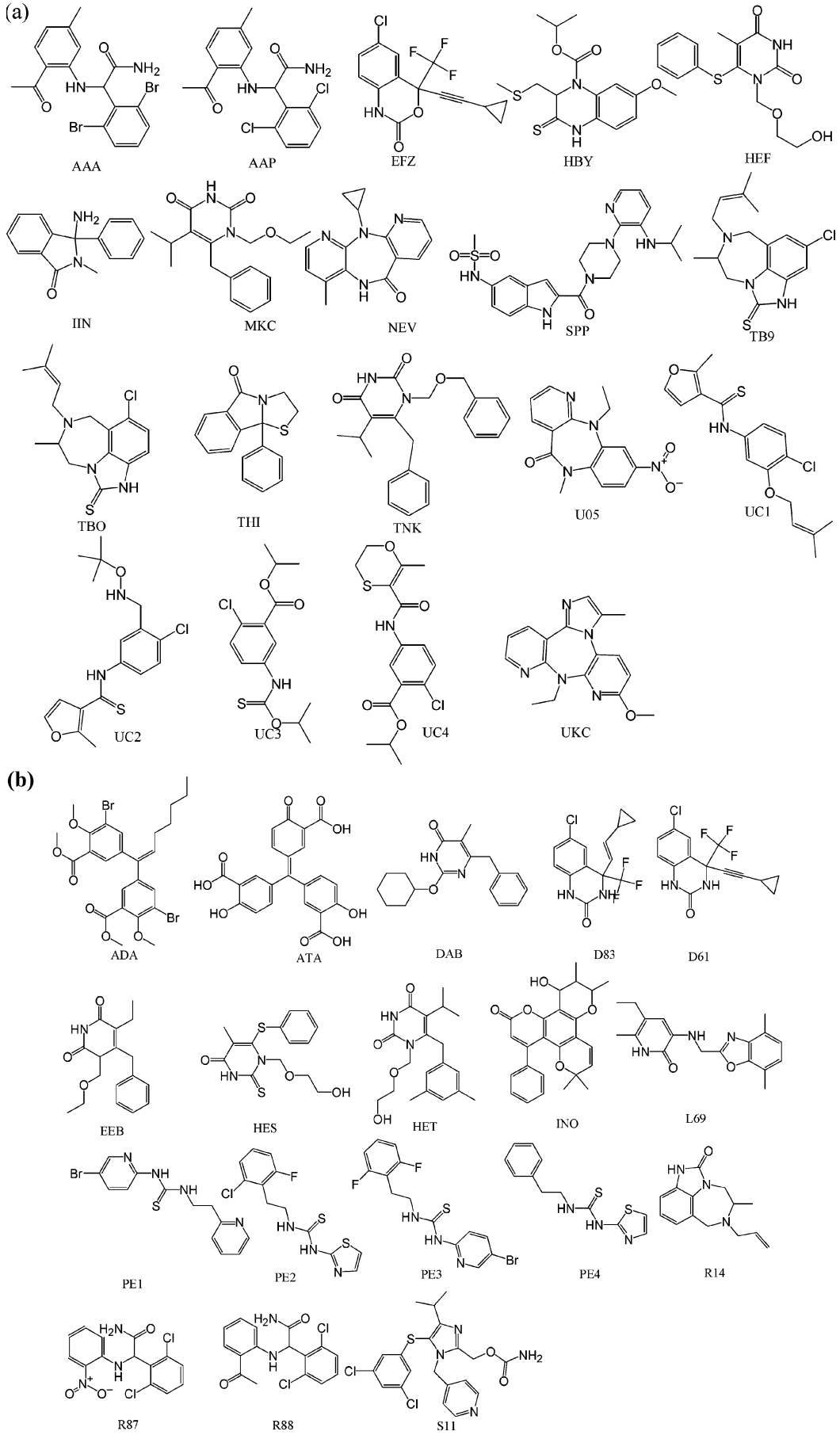

Figure 1. The molecular structures of the known NNRTIs with (a) and without (b) crystal structures in complex with HIV-1 RT.

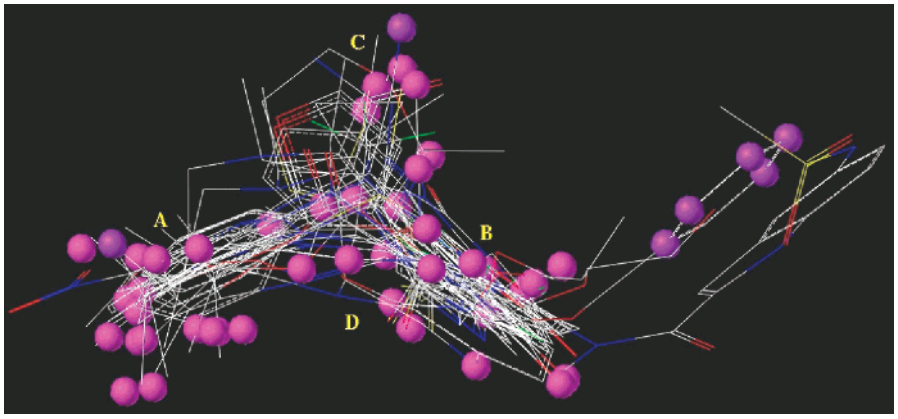

Figure 2. The 19 superimposed HIV-1 non-nucleoside reverse transcriptase inhibitors. The alignment of the inhibitors was generated by superimposing the crystal structures of HIV-1 RT/NNRTIs on that of HIV-1 RT/TBO (PDB code 1uwb). Only the main chain atoms of key residues around the allosteric binding site, namely, residues 92−110, 178−191, 227−236, 316−319 of chain A and residue 138 of chain B, were used for superimpositions. In this figure, magenta spheres, which indicate the possible positions occupied by potential ligand atoms, were generated using SPHGEN in the DOCK4.0 software package. Six purple spheres were manually added in order to get a better docking performance for TNK and SPP, which have a branch sticking out from wing B of the "butterfly".

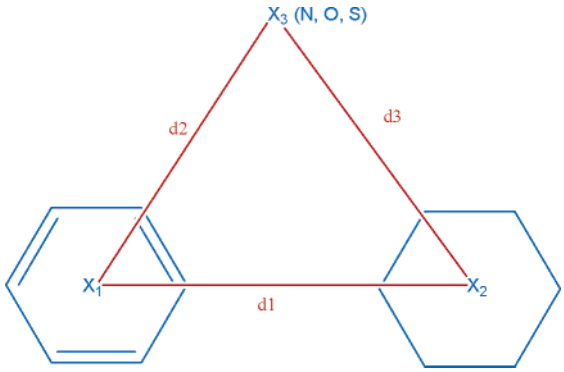

Figure 3. The first filter—a three-featured pharmacophore model derived from the 19 crystal structures of HIV-1 RT in complex with NNRTIs. In this figure, X_1 represents a five- or six-membered aromatic ring; X_2 represents a five- to seven-membered ring; and X_3 can be nitrogen, oxygen, or sulfur; d_1, d_2, and d_3 are distances between X_1 and X_2, X_1 and X_3, and X_2 and X_3, respectively. Two patterns were identified and the distance parameters are listed as the following: pattern I, d_1 (4.5−6.0 Å), d_2 (3.5−4.5 Å), and d_3 (4.5−6.5 Å); pattern II, d_1 (2.4−2.8 Å), d_2 (2.5−4.5 Å), and d_3 (4.0−5.5 Å).

HIV-1 RT/TBO (PDB code 1uwb),[18] which was applied in generating the spheres (the so-called negative images of the receptor, where the inhibitor atoms will reside) and calculating the energetic potentials in the sequential rigid and solvation docking studies. Only the main chain atoms of key residues around the allosteric binding site were considered, which are residues 92−110, 178−191, 227−236, 316−319 of chain A and residue 138 of chain B. Figure 2 shows the superimposed structures of the inhibitors. The overall shape of the inhibitors bound to the enzyme is a "butterfly" with two wings, A and B, a head C, and a tail D. Some flexible ligands, such as TNK and SPP, have branches sticking out from the wings. This "butterfly" shape reflects the overall shape of the allosteric binding site of the enzyme. On the basis of the "butterfly" shape, we designed a three-featured pharmacophore model shown in Figure 3. This pharmacophore model has two distinct binding patterns that occupy one or both of the butterfly wings, respectively.

Conformational flexible searches (CFS) were conducted for both patterns to search a refined ACD database using ISIS 2.4 of MDL Information Systems Inc. Default parameters were applied except that the rmsd cutoff was set to 1.0 Å and the van der Waals energy difference was set to 10.0 kcal/mol.

Filter II—Multiple-Conformation Rigid Docking. The second filter in this database screening strategy is a multiple-conformation rigid docking. All the docking studies were carried out using DOCK4.0.[12,13] The crystal structure of HIV-1 RT/8Cl-TBO (1uwb)[18] was chosen to generate the spheres and to calculate electrostatic and van der Waals potentials around the allosteric binding site of the receptor. The solvent accessible surface area was first generated using the DMS program in the Midas Plus package (Computer Graphic Laboratory, University of California—San Francisco) after the hydrogen atoms of the receptor were removed. The spheres, where inhibitor atoms would occupy, were then generated using the SPHGEN module in DOCK4.0.[12,13] A cluster analysis was followed and one cluster that had roughly 30−40 spheres around the binding site was selected for docking. Manually inserting and deleting some spheres may be necessary. To generate the energetic potentials, hydrogen atoms of the receptor as well as counterions were first added to make the whole system neutral. The charges and van der Waals parameters were taken from AMBER.[46] The electrostatic and van der Waals potentials were generated using the GRID module in DOCK4.0.

Conformational searches were carried out for the hits of the pharmacophore model filter using the OMEGA (optimized ensemble generation application) program from OpenEye Scientific Software. The following are the basic parameters that control the calculations: maximum number of searched conformations (GP_MAX_POP), 1000; maximum number of saved conformations (GP_NUM_OUTPUT_CONFS), 50; energy threshold (GP_ENERGY_WINDOW), 5.0 kcal/mol; and rmsd cutoff (GP_RMS_CUTOFF), 1.5 Å. We believed that this setting was adequate to generate a set of diverse low-energetic conformations with a good balance between conformation quality and computer time.

Rigid docking was then performed for all the conformations, and the key docking parameters are listed as the following: maximum orientations, 1000; minimum matching nodes, 4; maximum matching nodes, 15; no intramolecular score; and dielectric constant (ϵ), 4.0. Both the conformational searches and docking were performed with an SGI Origin (R10000).

Filter III—Solvation Docking. On the basis of our experience, the solvation docking parameters in formula I may be varied from one system to another. Thus, we set out to develop a solvation-docking model just for HIV-1 RT in order to achieve an optimal solvation docking performance. A training set of 12 known NNRTIs, which have crystal structures of HIV-1 RT complexes available, was applied to derive the model. Each molecule in the training set had an rmsd less than 3.0 Å between the docked and its own crystal structure, and the rigid docking score was better than the threshold (−24.5 kcal/mol). It is noted that the selected docked structure is not necessarily the one that has the best docking score. The α, β, and γ parameters in formula I were then optimized to reproduce the experimental binding free energies.

Solvation docking was performed for the molecules that passed filter II using a solvation docking program developed by Zou et al.[6] This program outputs the van der Waals energy (hydrophobic interaction) and the screened electrostatic energy ($\Delta G^{\text{elec}}{}_{\text{solv}} + \Delta E^{\text{elec}}{}_{\text{MM}}$, polarization interaction) as well as the polar and nonpolar solvent accessible surface areas. The electrostatic solvation energy, $\Delta G^{\text{elec}}{}_{\text{solv}}$, was calculated using a generalized Born model. Finally, the binding free energies were estimated using the derived solvation docking model.

Filter IV—MM−PB/SA. Only the known HIV-1 NNRTIs and the 30 most promising hits that passed the previous filters were screened by this filter, subject to our limited computational resources. Unlike the first three filters for which only the ligand flexibility was taken into account, this last filter applied MD simulations to sample the conformational space of both the inhibitor and the receptor. For each molecule, a 20 Å water cap was added around the allosteric binding site; MD simulations were then carried out at 300 K with a time step of 2.0 fs, the nonbonded cutoff being set to 9.0 Å. The inhibitor, the water molecules, and the receptor residues that are within 20 Å of the mass center of the inhibitor were allowed to move during the belly MD simulations. After equilibration for 50 ps, 20 snapshots were collected for postprocessing at a frequency of every 1 ps. All the MD simulations were carried out with the SANDER module in AMBER7.0[46] using the PARM99 force field[47] described by formula VIII.

$$E_{\text{pair}} = \sum_{\text{bonds}} k_r(r - r_{\text{eq}})^2 + \sum_{\text{angles}} k_\theta(\theta - \theta_{\text{eq}})^2 + = \sum_{\text{dihedrals}} \frac{v_n}{2}[1 + \cos(n\phi - \gamma)] + \sum_{i<j}\left[\frac{A_{ij}}{R_{ij}{}^{12}} - \frac{B_{ij}}{R_{ij}{}^{6}} + \frac{q_i q_j}{\epsilon R_{ij}}\right] \quad \text{(VIII)}$$

For each snapshot, MM−PB/SA analysis was performed to calculate the binding free energy. The inhibitor−receptor interaction energies ($\Delta E_{\text{electrostatic}}$, ΔE_{vdw}) were calculated with the ANAL module in the AMBER7.0 package.[46] As to the solvation energies, the electrostatic part (reaction field energies, ΔG_{PB}) was calculated with DELPHI[48,49] using the PARSE radii.[50] The nonpolar contributions (ΔG_{SA}) were estimated using a simple empirical formula:[50] $\Delta G_{\text{SA}} = \sigma A + b$, where A is the solvent-accessible surface area that was estimated using the MSMS program.[51] σ and b are empirical constants, and in this work 0.0054 and 0.92 kcal/mol were used, respectively. To speed up the solvation energy calculations, only the residues within 30 Å of the mass center of the inhibitor were used for the PBSA calculations. The entropy ($T\Delta S$) was estimated through normal-mode analysis using the NMODE module in the AMBER package. Considering that this type of calculation is extremely time-consuming for large systems, only residues within 12 Å of the mass center of the inhibitor (including the ligand, but excluding the water molecules) were used for the normal-mode analysis. A thorough minimization was first performed with a distance dependent dielectric constant ($\epsilon = 4R_{ij}$) prior to the normal-mode analysis.

Table 2. Least-Squares Fittings of HIV-1 RT/TBO (1uwb) to the Other 19 HIV-1 RT/NNRTIs Crystal Structures[a]

no.	compd abbrev	PDB code	rmsd (Å)	ref	no.	compd abbrev	PDB code	rmsd (Å)	ref
1	AAA	1HNI	0.639	19	11	TBO	1HNV	0.39	18
2	AAP	1VRU	0.647	20	12	THI	1RVP	1.18	b
3	EFZ	1FK9	0.711	21	13	TNK	1RT2	0.875	24
4	HBY	1BQM	0.796	22	14	U05	1RT3	0.799	27
5	HEF	1RTI	0.854	20	15	UC1	1RT4	0.837	28
6	IIN	1RVN	1.18	*b*	16	UC2	1RT5	0.921	28
7	MKC	1RT1	1.09	24	17	UC3	1RT6	0.910	28
8	NEV	1VRT	0.641	20	18	UC4	1RT7	0.927	28
9	SPP	1KLM	0.967	25	19	UKC	1RVR	1.20	b
10	TB9	1TVR	0.798	26					

[a] Alignments were performed only for the main chain atoms around the allosteric binding site: residues 92−110, 178−191, 227−236, 316−319 of chain A and residue 138 of chain B. [b] *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 3911.

Results and Discussion

Filter I—Pharmacophore Model. Superimpositions were carried out for the crystal structures of 19 HIV-1 RT/NNRTIs to that of HIV-1 RT/TBO (1uwb). The average rmsd is 0.86 Å for the 19 NNRTIs. The aligned non-nucleoside reverse transcriptase inhibitors are shown in Figure 2 and the root-mean-square deviations are listed in Table 2. On the basis of the "butterfly" shape, a three-featured pharmacophore model was designed as shown in Figure 3. The first feature, X_1, is a five- or six-membered aromatic ring; the second feature, X_2, is a five- to seven-membered ring, which can either be aromatic or aliphatic; the third feature, and X_3, is a hydrophilic center that can be nitrogen, oxygen, or sulfur. d_1, d_2 and d_3 are distances between X_1 and X_2, X_1 and X_3, X_2 and X_3, respectively. There are two patterns of this pharmacophore model. For pattern I, d_1, d_2, and d_3 are 4.5−6.0 Å, 3.5−4.5 Å, and 4.5−6.5 Å, respectively. For pattern II, d_1, d_2, and d_3 are 2.4−2.8 Å, 3.5−4.5 Å, and 4.0−5.5 Å, respectively. All the 37 known HIV-1 NNRTIs could match at least one pattern of this pharmacophore model. To be a little bit conservative, it was assumed that the hit rate of this filter was 95%. A total of 40 000 compounds (35 000 for

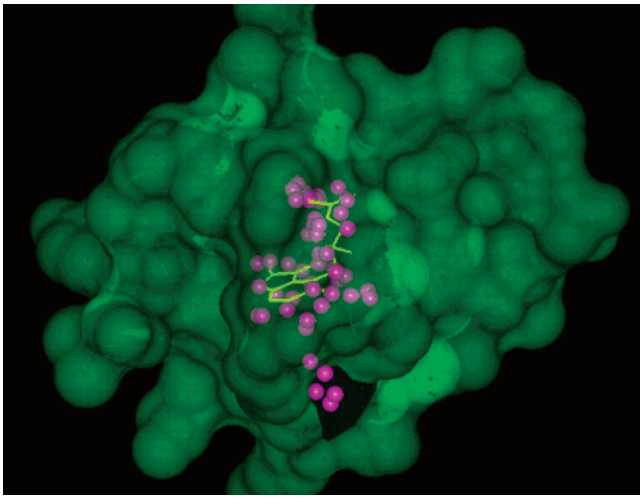

Figure 4. The transparent solvent accessible surface area of the key residues of HIV-1 RT (1uwb) around the allosteric binding site: residues 92−110, 178−191, 227−236, 316−319 of chain A and residue 138 of chain B. In this figure, the inhibitor, TBO (**11** in Table 1), is colored in yellow. The magenta spheres, which indicate the possible positions to be occupied by potential ligand atoms, were generated using the SPHGEN program in the DOCK4.0 software package.

pattern I and 5000 for pattern II) out of 150 000 passed this filter. The enrichment factor for this step was 3.56 according to formula VII.

Our three-featured pharmacophore model was conceived from the superimposed crystal structures of protein complexes. In other cases when there are much fewer or even no crystal structures available, one may build up pharmacophore models purely based on the known inhibitors. The following are some widely used software packages to generate pharmacophore models: CATALYST, of Accelrys Inc., and DISCO, DISOTECH, and GASP, of Tripos Inc. Pharmacophore database searches are typically very fast as long as the adequate databases have been preconstructed. As to our case, database searching was performed with ISIS and the CPU time was a couple of hours for each pattern.

Filter II—Multiple Conformation Rigid Docking. Conformational searches were performed with the OMEGA program for the 40 000 hits from the first filter. On average, 30 conformations were generated per molecule. More than 1 million conformations in total were subjected to rigid docking. The transparent solvent accessible surface of the key residues of HIV-1 RT (1uwb) around the allosteric binding site is shown in Figure 4. The non-nucleoside reverse transcriptase inhibitor 8Cl-TIBO (in short TBO, **11** in Table 1) is colored in yellow. The magenta balls, which were generated by SPHGEN in DOCK4.0, specify where the inhibitor atoms might reside. The rigid docking scores for the known NNRTIs and the 30 most promising hits, respectively, are listed in Tables 3 and 4. The hit rate of this step was estimated as 76% with a docking score criterion of −24.5 kcal/mol. Some 16 000 compounds out of 40 000 had at least one conformation pass this filter. The enrichment factor for this step was 1.89.

In Figure 5, a plot of the experimental binding free energies versus the docking scores is shown for 28 known HIV-1 NNRTIs that passed the rigid docking filter. The standard deviation was 3.3 kcal/mol and the

Table 3. Results of Multiple-Conformation Rigid Docking, Solvation Docking, and MM−PB/SA for the Known HIV-1 NNRTIs[a]

no.	compd abbrev	rigid docking score (kcal/mol)	solvation docking score (kcal/mol)	MM−PB/SA (kcal/mol)	experimental binding free energy (kcal/mol)
1	AAA	−29.2	−8.9	−7.2	−8.7
2	AAP	−28.7	−9.3	−7.2	−9.5
3	EFZ	−30.2	−9.8	−13.2	−11.6
4	HBY	−26.7	−10.4	−9.6	−11.2
5	HEF	−32.5	−9.8	−7.5	−7.0
6	IIN	−26.2	fail		−6.8
7	MKC	−24.7	fail		−10.7
8	NEV	−31.0	−9.9	−6.8	−9.1
9	SPP	fail			−8.1
10	TB9	−24.2	−9.6	−11.6	−10.6
11	TBO	fail		−12.0	−11.9
12	THI	−28.6	−8.9	−7.6	−10.6
13	TNK	fail		−8.2	−11.2
14	U05	−26.7	−8.8	−3.4	−8.7
15	UC1	−30.2	−9.7	−11.8	−11.0
16	UC2	−32.2	−9.8	−12.8	−9.3
17	UC3	−32.4	−10.8	−7.2	−10.3
18	UC4	−30.0	−10.0	−9.7	−10.1
19	UKC	−36.3	−10.9	−11.2	−9.2
20	ADA	fail			−6.5
21	ATA	fail			−7.2
22	DAB	−30.9	−9.4	−6.5	−7.3
23	D61	−32.5	−10.6	−14.8	−12.9
24	D83	−30.7	−9.6	−14.4	−12.8
25	EEB	−27.7	−10.0	3.0	−10.1
26	HES	−29.4	−9.1	−6.0	−7.9
27	HET	fail			−6.5
28	INO	fail			−8.0
29	L69	fail			−10.0
30	PE1	−25.6	fail		−10.5
31	PE2	−26.8	−9.1	−7.3	−10.0
32	PE3	−24.0	fail		−10.9
33	PE4	−27.5	−9.4	3.5	−8.0
34	R14	−28.2	−8.8	2.2	−5.7
35	R87	−33.4	fail		−10.2
36	R88	−25.4	fail		−10.3
37	S11	fail			−11.6

[a] Crystal structures of HIV-1 RT complexes were available for no. 1−19. Inhibitors that failed to pass the rigid docking filter (−24.5 kcal/mol), solvation docking filter (−8.8 kcal/mol), or MM−PB/SA filter (−6.8 kcal/mol) are indicated with "fail".

regression coefficient was 0, which implies that there is no statistical correlation between the rigid docking scores and the experimental binding free energies. The lack of statistical correlation is not a surprise at all given the fact that the rigid docking scoring function completely neglects the contributions of solvation energy and entropy. In addition, the receptor is being totally frozen in the docking studies.

How well did the multiple-conformation rigid docking perform in predicting the structures of the HIV-1 RT/NNRTI complex? Sixteen out of the 19 NNRTIs for which crystal structures of HIV-1 RT complexes are available successfully passed the rigid docking criterion. For each inhibitor that survived the filter, its own crystal structure was superimposed on that of HIV-1 RT/TBO (only the key residues around the binding site were considered). Then the rmsd of the best scored docking conformation to the crystal structure was calculated directly without fitting. The average and root-mean-square deviation of the rmsd for the 16 known HIV-1 NNRTIs were 3.2 and 3.4 Å, respectively (Table 5). It is still a considerable challenge to accurately model a protein complex (rmsd of less than 2.0 Å) using a

Table 4. Results of Multiple-Conformation Rigid Docking, Solvation Docking, and MM−PB/SA for the Top 30 Solvation Docking Hits[a]

no.	compd abbrev	rigid docking score (kcal/mol)	solvation docking score (kcal/mol)	MM−PB/SA (kcal/mol)
1	HIT01	−33.1	−8.9	−9.6
2	HIT02	−30.1	−9.0	0.4
3	HIT03	−31.8	−9.3	−0.8
4	HIT04	−28.7	−9.5	−2.5
5	HIT05	−35.7	−8.9	−5.4
6	HIT06	−30.0	−9.8	0.3
7	HIT07	−31.9	−9.2	−3.3
8	HIT08	−31.8	−9.8	−8.6
9	HIT09	−37.8	−9.6	−0.3
10	HIT10	−32.6	−9.8	−0.7
11	HIT11	−34.2	−9.2	5.3
12	HIT12	−34.8	−9.2	−6.6
13	HIT13	−35.2	−9.2	−9.5
14	HIT14	−34.5	−10.0	−10.8
15	HIT15	−25.7	−10.2	−17.7
16	HIT16	−30.6	−9.9	−4.0
17	HIT17	−31.6	−10.0	−3.6
18	HIT18	−33.8	−9.9	−0.3
19	HIT19	−35.6	−10.6	−5.1
20	HIT20	−27.2	−9.6	−3.1
21	HIT21	−32.6	−10.8	−2.2
22	HIT22	−31.8	−10.7	−7.1
23	HIT23	−37.6	−9.6	−8.1
24	HIT24	−29.3	−9.6	−1.9
25	HIT25	−35.2	−11.1	−1.8
26	HIT26	−28.5	−10.0	−7.9
27	HIT27	−31.4	−9.9	−6.8
28	HIT28	−30.9	−11.0	−8.1
29	HIT29	−27.5	−8.8	2.7
30	HIT30	−36.4	−9.5	1.4

[a] Promising solvation docking hits that passed the MM−PB/SA filter are shown in bold.

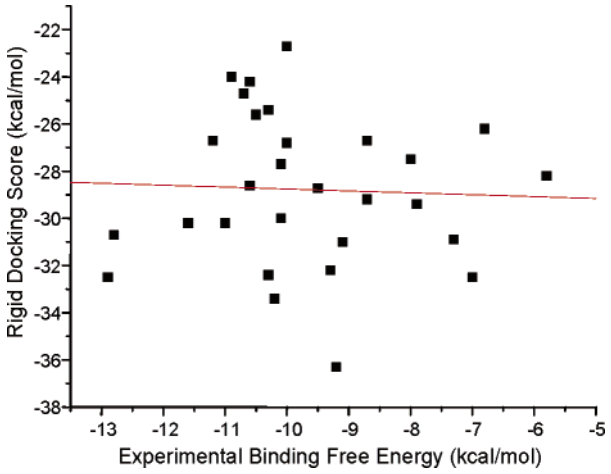

Figure 5. Plot of the experimental binding free energies versus docking scores for the 29 known NNRTIs that passed the rigid docking filter. The regression coefficient was 0.0 and the standard deviation was 3.3 kcal/mol.

receptor structure other than the one that forms the complex, even though it is a crystal structure.

What is the key to achieving a good docking performance? First of all, a set of good spheres that represent the active site is crucial. Those spheres should be evenly (not exactly) distributed around the binding site and some of them should occupy the atomic positions of the cocrystallized inhibitor. One may manually insert or delete some spheres if needed. Second, a high-quality conformational set is important in multiple-conformation rigid docking studies. A good conformational set

Table 5. Root Mean Square Deviations of the Rigid Docking, Solvation Docking, and Molecular Dynamic Simulations Structures of 16 NNRTIs Compared to the Crystal Structures[a]

no.	compd abbrev	rigid docking (Å)	solvation docking (Å)	MD (Å)
1	AAA	4.2	4.4	4.4
2	AAP	4.0	4.2	3.7
3	HBY	1.5	1.5	1.4
4	HEF	3.6	3.6	2.9
5	IIN	3.1		
6	MKC	5.8		
7	NEV	1.3	1.3	2.8
8	EFZ	3.0	3.0	1.1
9	TB9	2.1	2.1	1.4
10	THI	4.6	5.6	5.2
11	U05	3.5	3.6	3.9
12	UC1	1.6	1.1	0.8
13	UC2	1.2	1.4	1.2
14	UC3	5.3	1.7	2.4
15	UC4	3.9	2.9	2.8
16	UKC	1.7	1.7	1.2
	no. of data	16	14	14
	av deviations	3.2	2.7	2.5
	rms deviations	3.4	3.0	2.8

[a] TBO, TNK, and SPP were omitted since they did not meet the rigid docking criterion.

Table 6. A List of the Three Energy Terms of Solvation Docking for 12 HIV-1 NNRTIs That Were Applied To Derive the Solvation Docking Model[a]

no.	compd abbrev	rmsd (Å)	ΔG_{elec} (kcal/mol)	ΔE_{vdw} (kcal/mol)	ΔSAS (Å^2)	ΔG_{calc} (kcal/mol)	ΔG_{expt} (kcal/mol)
1	EFZ	3.0	1.9	−33.7	−890	−9.8	−11.6
2	HBY	1.5	3.6	−39.3	−882	−10.4	−11.2
3	HEF	2.7	14.2	−39.4	−863	−8.5	−7.0
4	IIN	2.7	7.6	−29.2	−914	−8.1	−6.8
5	NEV	1.3	1.4	−35.0	−852	−9.9	−9.1
6	TB9	2.1	5.3	−38.9	−793	−9.6	−10.6
7	U05	3.5	11.6	−36.6	−795	−8.1	−8.7
8	UC1	2.4	9.2	−35.3	−905	−8.9	−11.0
9	UC2	1.2	9.2	−42.7	−811	−9.7	−9.3
10	UC3	1.6	−3.0	−29.7	−867	−9.9	−10.3
11	UC4	1.9	4.2	−34.7	−892	−9.6	−10.1
12	UKC	1.8	1.8	−35.7	−881	−10.2	−9.2

[a] All 12 NNRTIs passed the rigid docking filter (−24.5 kca/mol), and the rmsd of the docked structures are within 3.0 Å to their individual crystal structures. Coefficients of electrostatic (α), van der Waals (β), and solvent accessible surface area (γ) energy terms are 0.1736, 0.1709, and 0.0049, respectively. The average unsigned error and root-mean-square error between ΔG_{calc} and ΔG_{expt} are 1.03 and 1.16 kcal/mol, respectively.

should be sufficiently diverse to cover the whole conformational space and the structures energetically favorable. In addition, the intramolecular energies ($\Delta E_{internal}$) of an inhibitor, which were not considered in this work, may be important, and it will be a good idea to add this term to the scores calculated with formula V in the future.

Rigid docking with DOCK4.0 is not very time demanding, and the typical process time for each conformation is within 10 s on a single-process SGI Origin (R10000).

Filter III—Solvation Docking. In Table 6, the three energetic terms (ΔG_{elec}, ΔE_{vdw} and ΔSAS) in formula I are listed for 12 known NNRTIs that have crystal structures for HIV-1 RT complexes. All 12 inhibitors have not only good docking scores (<−24.5 kcal/mol), but also high-quality docked structures (rmsd < 3.0 Å). The coefficients, α, β, and γ in formula I, were deter-

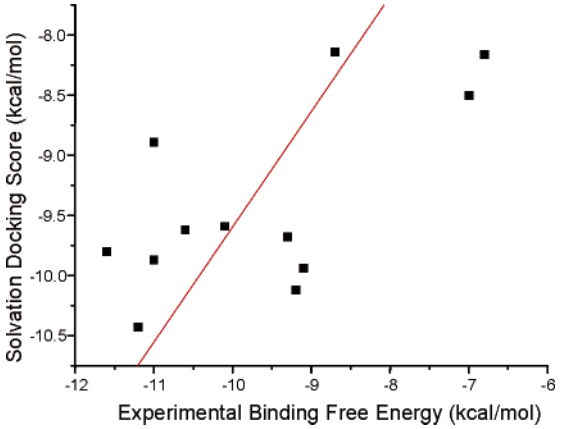

Figure 6. The performance of a solvation docking model derived from the 12 known NNRTIs that achieved both good docked structures (rmsd < 3.0 Å) and good docking scores (<−24.5 kcal/mol). The slope was 0.96 with the intercept being set to 0.0. The regression coefficient and standard deviation were 0.67 and 1.2 kcal/ mol, respectively.

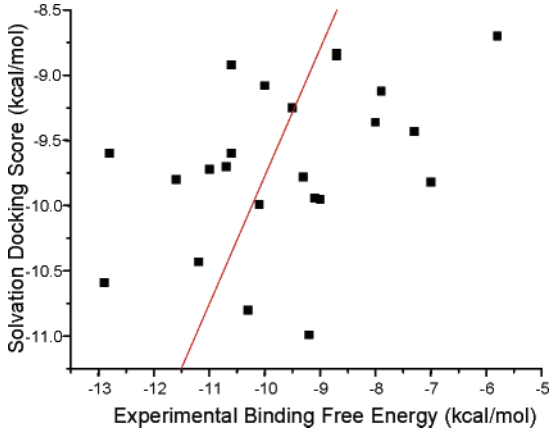

Figure 7. Plot of the experimental binding free energies versus solvation docking scores for the 22 compounds that passed the solvation docking filter. The slope was 0.98 with the intercept being set to 0.0. The regression coefficient and standard deviation were 0.42 and 1.6 kcal/mol, respectively.

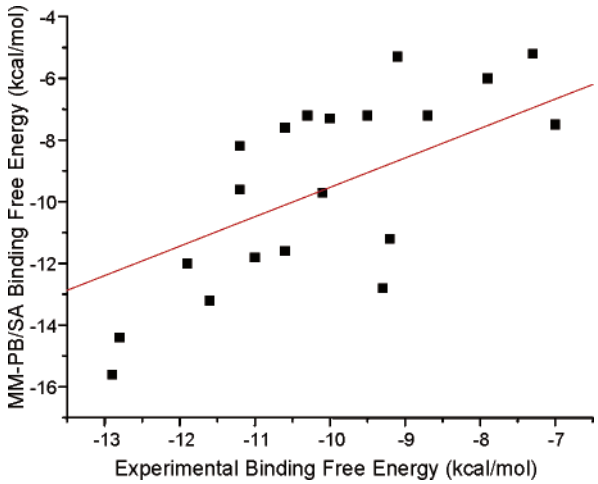

Figure 8. Plot of the experimental versus MM−PB/SA binding free energies for the 20 compounds that had their MM−PB/SA scores better than −6.0 kcal/mol. The slope was 0.95 with the intercept being set to 0.0. The regression coefficient and standard deviation were 0.74 and 2.2 kcal/mol, respectively.

mined by a least-squares fitting in order to reproduce the experimental binding free energies. The solvation-docking model ($\alpha = 0.1736$, $\beta = 0.1709$, $\gamma = 0.0049$) achieved average unsigned and rms errors of 1.03 and 1.16 kcal/mol between ΔG_{calc} and ΔG_{expt}, respectively. Figure 6 demonstrates how well the solvation-docking model predicts the binding free energies for 12 molecules in the training set. The slope was 0.96 with the intercept being set to 0.0 (since we intended to make the solvation docking scores reproduce the experimental binding free energies). The regression coefficient and standard deviation were 0.67 and 1.2 kcal/mol, respectively.

Solvation docking was performed for all the hits that survived the multiple-conformation rigid docking, and then the binding free energies were predicted with the above solvation docking model. A total of 3360 compounds successfully passed this filter with a threshold of −8.8 kcal/mol. As for the known NNRTIs in the control test, 22 out of the 28 rigid docking hits successfully passed the solvation docking filter. The HR and EF were estimated as 79% and 3.74, respectively. The solvation docking scores for all the known NNRTIs as well as the 30 selected promising hits are listed in Tables 3 and 4. A plot of the experimental binding free energies versus solvation docking scores is shown in Figure 7 for the 22 known NNRTIs that passed the solvation docking filter. The standard deviation was 1.6 kcal/mol and the regression coefficient was 0.4. The slope was 0.9778 with the intercept being set to 0.0. The regression coefficient, although significantly better than that of rigid docking, is still very low, as the flexibility of the receptor was not considered in this filter.

Fourteen out of the 19 known NNRTIs for which crystal structures of HIV-1 RT complexes exist passed the solvation docking filter. The rmsd values of the compounds docked to the crystal structures are listed in Table 5. The average and root-mean-square deviation of the rmsd for the 14 inhibitors were 2.7 and 3.0 Å, respectively. This performance is slightly better than that of rigid docking.

Compared to MM−GB/SA, MM−PB/SA, and LIE, solvation docking is a much faster approach, due to its simpler sampling protocol (the flexibility of the receptor is neglected). To apply this filter, one needs to find or build up a proper model. Unfortunately, just like LIE, most solvation docking models suffer from poor transferability. Solvation docking consumes a little more time than rigid docking, and the typical process time for one conformation was 20 s on a single-process SGI Origin (R10000) in this work.

Filter IV—MM−PB/SA. The last filter, MM−PB/SA, is the most time and resource demanding step in our lead discovery strategy. We thus only performed MM−PB/SA for the 22 known NNRTIs in the control set and the 30 most promising hits that passed the solvation docking filter. However, we do believe that it is practical to perform MM−PB/SA for up to 5000 compounds within a reasonable time frame for a pharmaceutical company. In Tables 3 and 4, the MM−PB/SA binding free energies for the known NNRTIs and the 30 promising compounds, respectively, are listed. Sixteen out of the 22 known NNRTIs have MM−PB/SA binding affinities better than −6.8 kcal/mol, which corresponds to an IC_{50} value of 10 μM. Figure 8 is a plot of MM−PB/SA versus experimental binding free energies for the 20 NNRTIs that have MM−PB/SA scores

Table 7. A List of the Individual Energy Terms of MM−PB/SA for the Known HIV-1 NNRTIs That Passed the Rigid Docking and Solvation Docking Filters[a]

		ΔE_{gas}		$\Delta G_{solvation}$ (PB/SA)					
no.	compd abbrev	ΔE_{vdw}	ΔE_{elec}	ΔG_{polar}	$\Delta G_{nonpolar}$	$\Delta G_{MM-PB/SA}$	$T\Delta S$	ΔG_{calc}	ΔG_{expt}
1	AAA	−48.6	−7.3	38.8	−5.6	−22.7	15.5	−7.2	−8.7
2	AAP	−48.4	−15.6	42.6	−5.4	−26.7	19.5	−7.2	−9.5
3	EFZ	−43.4	−11.5	32.7	−4.6	−26.8	13.6	−13.2	−11.6
4	HBY	−48.5	−12.6	39.0	−5.3	−27.4	17.8	−9.6	−11.2
5	HEF	−47.1	−16.3	43.6	−4.6	−24.5	17.0	−7.5	−7.0
6	NEV	−45.7	−5.1	33.4	−4.3	−21.7	14.9	−6.8	−9.1
7	TB9	−48.4	−24.1	50.2	−5.0	−27.3	15.7	−11.6	−10.6
8	THI	−44.5	−4.4	31.7	−3.5	−20.8	13.1	−7.6	−10.6
9[b]	U05	−46.9	−4.6	37.6	−4.6	−13.9	15.1	−3.4	−8.7
10	UC1	−55.1	−25.3	56.9	−5.6	−29.1	17.2	−11.8	−11.0
11	UC2	−54.0	−8.4	38.6	−5.6	−29.4	17.5	−11.2	−9.3
12	UC3	−42.9	−16.5	41.2	−5.3	−23.6	16.3	−7.2	−10.3
13	UC4	−45.5	−10.4	36.0	−5.3	−25.2	15.5	−9.7	−10.1
14	UKC	−48.8	−10.4	35.8	−5.2	−28.7	17.5	−11.2	−9.2
15[b]	DAB	−50.0	−8.5	41.3	−5.2	−22.4	15.9	−6.5	−7.3
16	D61	−45.7	−16.3	36.9	−4.5	−29.6	14.8	−14.8	−12.9
17	D83	−45.7	−13.7	35.2	−4.4	−28.6	14.2	−14.4	−12.8
18[b]	EEB	−46.9	−4.5	40.4	−5.1	−16.1	19.2	3.0	−10.1
19[b]	HES	−45.4	−16.3	45.1	−4.2	−20.8	14.7	−6.0	−7.9
20	PE2	−47.5	−2.3	34.5	−4.9	−20.2	12.8	−7.3	−10.0
21[b]	PE4	−41.1	−3.5	36.9	−3.6	−11.2	14.8	3.5	−8.0
22[b]	R14	−43.1	−10.3	41.6	−4.3	−16.1	18.2	2.2	−5.7

[a] All energy terms are in kcal/mol. [b] Compounds that failed to pass the MM−PB/SA filter (−6.8 kcal/mol).

better than −6.0 kcal/mol. The slope was 0.95 with an intercept of 0.0. The regression coefficient and standard deviation were 0.74 and 2.2 kcal/mol, respectively. The regression coefficient was significantly better than those of the rigid docking and solvation docking filters. The individual energetic terms (ΔE_{vdw}, ΔE_{elec}, ΔG_{polar}, and ΔG_{SASr}) of MM−PB/SA are listed in Table 7 for the known NNRTIs that passed the rigid docking and solvation docking filters. Ten promising hits out of 30 had MM−PB/SA scores better than −6.8 kcal/mol. HIT15, the best one, had a binding free energy of −17.7 kcal/mol, which has a good chance to be a real HIV-1 non-nucleoside reverse transcriptase inhibitor.

In summary, 16 out of the 37 known NNRTIs survived all the filters, and the overall hit rate of the whole procedure was 41%, assuming that the hit rate of the first filter was 95%. The hit rate and the enrichment factor of the first three filters were 56% and 25, respectively. An enrichment factor of 25 implies that the probability of finding a real inhibitor randomly from the hits of the first three filters is 25-fold higher than from the whole database.

Future Development. A cartoon representation of our hierarchical multiple-filter database screening strategy is shown in Figure 9. In this figure, hearts stand for compounds, with active compounds colored in red and inactive colored in blue. From the top to the bottom, the active compounds are gradually condensed. It is notable that the numbers of red and blue hearts do not directly correspond to the real numbers of active and inactive compounds in a database. It should also be pointed out that the above protocol can certainly be modified to meet an individual's special needs. For example, one may first apply a set of ADMET/PK models to screen the whole database to eliminate compounds with bad ADMET/PK properties. For the last filter, MM−PB/SA may be replaced with a less computer resource demanding GB/SA model. Recently, we developed a solvation model[52] based on weighted

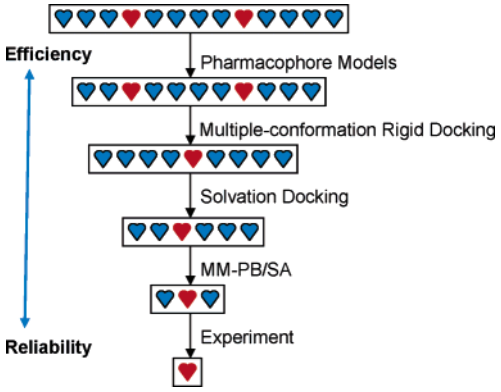

Figure 9. Representation of the hierarchical database screening strategy using multiple filters. From the most efficient to the most reliable, these filters are pharmacophore models, multiple-conformation rigid docking, solvation docking, and MM−PB/SA. The hits that survived all the four filters may be subjected to testing their biological activities. In this figure, hearts stand for compounds, and active compounds are colored red and inactive blue. From the top to the bottom, the active compounds are condensed gradually. The numbers of red and blue hearts do not directly correspond to the real numbers of active and inactive compounds in practical database searches.

solvent accessible surface area (WSAS), which is a much more efficient method compared to MM−PB/SA or MM−GB/SA. This solvation model has also demonstrated relatively good performance in free energy calculations. One may substitute MM−WSAS for the solvation docking and MM−PB/SA in some situations.

As pointed out previously,[9] although MM−PB/SA does not have as solid a theoretical basis as FEP and TI, it is computationally more efficient than FEP and TI. Moreover, compared to other free energy calculation methods including LIE and solvation docking, MM−PB/SA is more promising under some circumstances, because it does not require a training set to derive empirical parameters in the first place, while LIE and solvation docking do. Therefore, when combining

MM−PB/SA with molecular docking and molecular dynamics simulations, one could reliably model a protein and DNA complex a priori. This is why MM−PB/SA was applied as the last filter in our lead exploring strategy. In this work, for each inhibitor, the MD simulation required 4 h of CPU time, and the subsequent MM−PB/SA analysis and entropy calculations needed another 2 h on the SGI Origin (R10000). If one has to process 3500 molecules, it would take roughly 1 month with a 30-processer SGI Origin (R10000). It is shown that this step still requires massive amounts of computational resources. Fortunately, most new generation computers are much faster than the one we used. It is expected that one may conduct MM−PB/SA analysis for 3500 molecules within 1 week.

Being a bottleneck of this lead exploring strategy, it is important for us to make MM−PB/SA more efficient. Actions may be taken in the following two aspects. First of all, one may run MD simulations using implicit water models, such as in GB/SA and PB/SA, instead of an explicit water model. GB-MD has been available in AMBER since its sixth version. Although GB-MD may not sample conformational space as well as MD with explicit solvents, it can save a lot of CPU hours not only on sampling but also on postanalysis. Second, a new algorithm is being developed to calculate the entropy accurately and efficiently. With the proposed new improvements, we believe our hierarchical multiple-filter drug lead exploring strategy will be more efficient and will have broader applications in drug discovery.

Conclusions

In this work, a hierarchical multiple-filter database screening strategy has been successfully applied to explore the drug leads for the HIV-1 reverse transcriptase. Starting from a refined ACD database that had 150 000 compounds, the first filter, a pharmacophore model, efficiently reduced the compound number to 40 000. The HR and EF are estimated at 95% and 3.56, respectively. In the subsequent rigid docking filter, the 40 000 hits were further reduced to 16 000 with HR and EF of 76% and 1.89, respectively. The third filter, solvation docking, reduced the compound number from 16 000 to 3360 with HR and EF of 79% and 3.74, respectively. In the last step, molecular dynamics simulations combined with MM−PB/SA were performed for the 30 most promising solvation docking hits. Ten out of the 30 molecules achieved MM−PB/SA binding free energies better than −6.8 kcal/mol, and HIT15, the best one, had a binding free energy of −17.0 kcal/mol. The EF for the first three steps and the HR of the whole procedure were 25 and 41%, respectively. Although MM−PB/SA calculations were not calculated for all the 3360 hits of the solvation docking filter, it has been demonstrated that it is possible for pharmaceutical companies, equipped with better computational resources, to perform these calculations in a reasonable time frame.

For the 19 known HIV-1 NNRTIs that have crystal structures of HIV-1 RT complexes, the docked and MD structures were superimposed on the crystal ones and the root-mean-square deviations (rmsd) were 3.2, 2.7, and 2.5 Å for the rigid docking, solvation docking, and MD simulations, respectively. This was an encouraging performance, given that all the docking studies were based on one single-crystal structure of HIV-1 RT/TBO (1uwb), rather than the more desirable individual crystal complexes.

In conclusion, our hierarchical multiple-filter database searching strategy achieved not only high efficiency but also high reliability. It is an attractive strategy in drug lead exploration.

Acknowledgment. This paper is dedicated to Peter A. Kollman, who directed this project prior to his death in May 2001. We are grateful to acknowledge financial support from NIH (GM-29072, CA-25644 and GM-56609) and the computer time at NCSA (MCB000013 (J. Wang, P. I.)). The structures of promising hits in Table 4 are provided upon request.

References

(1) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.* **2001**, *46*, 3−26.
(2) Blake, J. F. Chemoinformatics—Predicting the physicochemical properties of 'drug-like' molecules. *Curr. Opin. Biotechnol.* **2000**, *11*, 104−107.
(3) Beresford, A. P.; Selick, H. E.; Tarbit, M. H. The emerging importance of predictive ADME simulation in drug discovery. *Drug Discovery Today* **2002**, *7*, 109−116.
(4) Åqvist, J.; Medina, C.; Samuelsson, J. E. New method for prediction binding affinity in computer-aided drug design. *Protein Eng.* **1994**, *7*, 385−391.
(5) Hansson, T.; Åqvist, J. Estimation of binding free energies for HIV proteinase inhibitors by molecular dynamics simulations. *Protein Eng.* **1995**, *8*, 1137−1144.
(6) Zou, X. Q.; Sun, Y. X.; Kuntz, I. D. Inclusion of solvation in ligand binding free energy calculations using the generalized-Born model. *J. Am. Chem. Soc.* **1999**, *121*, 8033−8043.
(7) Guvench, O.; Shenkin, P.; Kolossváry, I.; Still, W. C. Application of the frozen atom approximation to the GB/SA continuum model for solvation free energy. *J. Comput. Chem.* **2002**, *23*, 214−221.
(8) Srinivasan, J.; Miller, J.; Kollman, P. A.; Case, D. A. Continuum solvent studies of the stability of RNA hairpin loops and helices. *J. Biol. Struct. Dyn.* **1998**, *16*, 671−682.
(9) Wang, J. M.; Morin, P.; Wang, W.; Kollman, P. A. Use of MM−PBSA in reproducing the binding free energies to HIV-1 RT of TIBO derivatives and predicting the binding mode to HIV-1 RT of efavirenz by docking and MM−PBSA. *J. Am. Chem. Soc.* **2001**, *123*, 5221−5230.
(10) Huo, S. H.; Wang, J. M.; Cieplak, P.; Kollman, P. A.; Kuntz, I. D. Molecular dynamics and free energy analyses of cathepsin D-inhibitor interactions: Insight into structure-based ligand design. *J. Med. Chem.* **2002**, *45*, 1412−1419.
(11) Wang, W.; Lim, W. A.; Jakalian, A.; Wang, J. M.; Luo, R.; Bayly, C. T.; Kollman, P. A. An analysis of the interactions between the sem-5 SH3 domain and its ligands using molecular dynamics, free energy calculations and sequence analysis. *J. Am. Chem. Soc.* **2001**, *123*, 3986−3994.
(12) Ewing, T. J. A.; Kuntz, I. D. Critical evaluation of search algorithms for automated molecular docking and database screening, *J. Comput. Chem.* **1997**, *18*, 1175−1189.
(13) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* **1982**, *161*, 269−288.
(14) De Clercq, E. Antiviral therapy of human immunodeficiency virus infections. *Clin. Microbiol. Rev.* **1995**, *8*, 200−239.
(15) Young, S. D. Nonnucleoside inhibitors of HIV-1 reverse transcriptase. *Perspect. Drug Discovery Des.* **1993**, *1*, 181−192.
(16) De Clercq, E. Toward improved anti-HIV chemotherapy: Therapeutic strategies for intervention with HIV infections. *J. Med. Chem.* **1995**, *38*, 2491−2517.
(17) Esnouf, R.; Ren, J. S.; Ross, C.; Jones, Y.; Stammers, D.; Stuart, D. Mechanism of inhibition of HIV-1 reverse transcriptase by non-nucleoside inhibitors. *Nat. Struct. Biol.* **1995**, *2*, 303−308.
(18) Ding, J. P.; Das, K.; Moereels, H.; Koymans, L.; Andries, K.; Janssen, P. A. J.; Hughes, S. H.; Arnold, E. Structures of HIV-1 RT/TIBO R86183 complex reveals similarity in the binding of diverse nonnucleoside inhibitors. *Nat. Struct. Biol.* **1995**, *2*, 407−415.
(19) Ding, J. P.; Das, K.; Tantillo, C.; Zhang, W.; Clark, A. D., Jr.; Jessen, S.; Lu, X.; Hsiou, Y.; Jacobo-Molina, A.; Andries, K.;

Pauwels, R.; Moereels, H.; Koymans, L.; Janssen, P. A. J.; Smith, R. H., Jr.; Koepke, M. K.; Michejda, C. J.; Hughes, S. H.; Arnold, E. Structure of HIV-1 reverse transcriptase in a complex with the non-nucleoside inhibitor α-APA R95845 at 2.8 Å resolution. *Structure* **1995**, *3*, 365−379.

(20) Ren, J. S.; Esnouf, R.; Garman, E.; Somers, D.; Ross, C.; Kirby, I.; Keeling, J.; Darby, G.; Jones, Y.; Stuart, D.; Stammers, D. High-resolution structures of HIV-1 RT from four RT−inhibitor complexes. *Nat. Struct. Biol.* **1995**, *2*, 293−302.

(21) Ren, J. S.; 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 resistant mutations in HIV-1 reverse transcriptase. *Structure* **2000**, *8*, 1089−1094.

(22) Hsiou, Y.; Das, K. Y.; Ding, J. P.; Clark, A. D., Jr.; Kleim, J. P.; Rosner, M.; Winkler, I.; Riess, G.; Hughes, S. H.; Arnold, E. Structures of Tyr188Leu mutant and wild-type HIV-1 reverse transcriptase complexed with the non-nucleoside inhibitor HBY 097: Inhibitor flexibility is a useful design feature for reducing drug resistance. *J. Mol. Biol.* **1998**, *284*, 313−323.

(23) Gussio, R.; Pattabiraman, N.; Zaharevitz, D. W.; Kellogg, G. E.; Topol, I. A.; Rice, W. G.; Schaeffer, C. A.; Erickson, J. W.; Burt, S. K. All-atom models for the non-nucleoside binding site of HIV-1 reverse transcriptase complexed with inhibitors: A 3D QSAR approach. *J. Med. Chem.* **1996**, *39*, 1645−1650.

(24) 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.

(25) Esnouf, R. M.; Ren, J. S.; Hopkins, A. L.; Ross, C. K.; Jones, E. Y.; Stammers, D. K.; Stuart, D. I. Unique features in the structure of the complex between HIV-1 reverse transcriptase and the bis(heteroaryl)piperazine (BHAP) U-90152 explain resistance mutations for this nonnucleoside inhibitor. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 3984−3989.

(26) Das, K.; Ding, J. P.; Hsiou, Y.; Clark, A. D.; Moereels, H.; Koymans, L.; Andries, K.; Pauwels, R.; Janssen, P. A. J.; Boyer, P. L.; Clark, P.; Smith, R. H.; Smith, M. B. K.; Michedja, C. J.; Hughes, S. H.; Arnold, E. Crystal structures of 8-Cl and 9-Cl TIBO complexed with wild-type HIV-1 RT and 8-Cl TIBO complexed with the Tyr181Cys HIV-1 RT drug resistance mutant. *J. Mol. Biol.* **1996**, *264*, 1085−1100.

(27) Ren, J. S.; Esnouf, R. M.; Hopkins, A. L.; Jones, E. Y.; Kirby, I.; Keeling, J.; Ross, C. K.; Larder, B. A.; Stuart, D. I.; Stammers D. K. 3′-azido-3′-deoxythymidine drug resistance mutations in HIV-1 RT can induce long range conformational changes. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 9518−9523.

(28) Ren, J. S.; Esnouf, R. M.; Hopkins, A. L.; Warren, J.; Balzarini, J.; Stuart, D. I.; Stammers, D. K., Crystal structures of HIV-1 Reverse transcriptase in complex with carboxanilide derivatives. *Biochemistry* **1998**, *37*, 14394−14403.

(29) Smith, M. B. K.; Rouzer, C. A.; Taneyhill, L. A.; Smith, N. A.; Hughes, S. H.; Boyer, P. L.; Janssen, P. A. J.; Moereels, H.; Koymans, L.; Arnold, E.; Ding, J. P.; Das, K.; Zhang, W. Y.; Michejda, C. J.; Smith, R. H., Jr. Molecular modeling studies of HIV-1 reverse transcriptase nonnucleoside inhibitors: Total energy of complexation as a predictor of drug placement and activity. *Protein Sci.* **1995**, *4*, 2203−2222.

(30) Young, S. D.; Britcher, S. F.; Tran, L. O.; Payne, L. S.; Lumma, W. C.; Lyle, T. A.; Huff, J. R.; Anderson, P. S.; Olsen, D. B.; Carroll, S. S.; Pettibone, D. J.; O'brien, J. A.; Ball, R. G.; Balani, S. K.; Lin, J. H.; Chen, I. W.; Schleif, W. A.; Sardana, V. V.; Long, W. J.; Byrnes, V. W.; Emini, E. A. L-743,726 (DMP-266): A novel, highly potent nonnucleoside inhibitor of the human immunodeficiency virus type 1 reverse transcriptase. *Antimicrob. Agents Chemother.* **1995**, *39*, 2602−2605.

(31) Kleim, J. P.; Bender, R.; Kirsch, R.; Meichsner, C.; Paessens, A.; Rosner, M.; Rubsamen-waigmann, H.; Kaiser, R.; Wichers, M.; Schneweis, K. E.; Winkler, I.; Riess, G. Preclinical evaluation of hby 097, a new nonnucleoside reverse transcriptase inhibitor of human immunodeficiency virus type 1 replication. *Antimicrob. Agents Chemother.* **1995**, *39*, 2253−2257.

(32) Baba, M.; Tanaka, H.; Declercq, E.; Pauwels, R.; Balzarini, J.; Schols, D.; Nakashima, H.; Perno, C. F.; Walker, R. T.; Miyasaka, T. Highly specific inhibition of human immunodeficiency virus type 1 by a novel 6-substituted acyclouridine derivative. *Biochem. Biophys. Res. Commun.* **1989**, *165*, 1375−1381.

(33) Romero, D. L.; Morge, R. A.; Genin, M. J.; Biles, C.; Busso, M.; Resnick, L.; Althaus, I. W.; Reusser, F.; Thomas, R. C.; Tarpley, W. G. Bis(heteroaryl) piperazine (BHAP) reverse transcriptase inhibitors: Structure−activity relationships of novel substituted indole analogues and the identification of 1-[5-methanesulfonamido-1*H*-indol-2-yl]-carbonyl]-4-[3-[(1-methylethyl)amino]-pyridinyl]piperazine monomethanesulfonate (U-90152S), a second-generation clinical candidate. *J. Med. Chem.* **1993**, *36*, 1505−1508.

(34) Ren, J. S.; Esnouf, R. M.; Hopkins, A. L.; Stuart, D. I.; Stammers, D. K. Crystallographic analysis of the binding modes of thiazoloisoindolinone non-nucleoside inhibitors to HIV-1 reverse transcriptase and comparison with modeling studies. *J. Med. Chem.* **1999**, *42*, 3845−3851.

(35) Cushman, M.; Golebiewski, M.; Buckheit, R. W., Jr.; Graham, L.; Rice, W. G. Synthesis and biological evaluation of an alkenyldiarylmethane (ADAM) which acts as a novel non-nucleoside HIV-1 reverse transcriptase inhibitor. *Bioorg. Med. Chem. Lett.* **1995**, *5*, 2713−2716.

(36) Baba, M.; Schols, D.; Pauwels, R.; Balzarini, J.; De Clercq, E. Fuchsin acid selectively inhibits human immunodeficiency virus (HIV) replication in vitro. *Biochem. Biophys. Res. Commun.* **1988**, *155*, 1404−1411.

(37) Artico, M.; Massa, S.; Mai, A.; Marongiu, M. E.; Piras, G.; Tramontano, E.; La Colla, P. 3,4-dihydro-2-alkoxy-6-benzyl-4-oxopyrimidines (DABOs): A new class of specific inhibitors of human immunodeficiency virus type 1. *Antiviral Chem. Chemother.* **1993**, *4*, 361−368.

(38) Baba, M.; De Clercq, E.; Iida, S.; Tanaka, H.; Nitta, I.; Ubasawa, M.; Takashima, H.; Sekiya, K.; Umezu, K.; Nakashima, H.; Shigeta, S.; Walker, R. T.; Miyasaka, T. Anti-Human immunodeficiency virus type 1 activities and pharmacokinetics of novel 6-substituted acyclouridine derivatives. *Antimicrob. Agents Chemother.* **1990**, *34*, 2358−2363.

(39) Tanaka, H.; Takashima, H.; Ubasawa, M.; Sekiya, K.; Inouye, N.; Baba, M.; Shigeta, S.; Walker, R. T.; De Clercq, E.; Miyasaka, T. Synthesis and antiviral activity of 6-benzyl analogues of 1-[(2-hydroxyethoxy)methyl]-6-(phenylthio)thymine (HEPT) as potent and selective anti-HIV-1 agents. *J. Med. Chem.* **1995**, *38*, 2860−2865.

(40) Patil, A. D.; Freyer, A. J.; Eggleston, D. S.; Haltiwanger, R. C.; Bean, M. F.; Taylor, P. B.; Caranfa, M. J.; Breen, A. L.; Bartus, H. R.; Johnson, R. K.; Hertzberg, R. P.; Westley, J. W. The inophyllums,novel inhibitors of HIV-1 reverse transcriptase isolated from the malaysian tree, calophyllum inophyllum linn. *J. Med. Chem.* **1993**, *36*, 4131−4138.

(41) Goldman, M. E.; Nunberg, J. H.; O'brien, J. A.; Quintero, J. C.; Schleif, W. A.; Freund, K. F.; Gaul, S. L.; Saari, W. S.; Wai, J. S.; Hoffman, J. M.; Anderson, P. S.; Hupe, D. J.; Emini, E. A.; Stern, A. M. Pyridinone derivatives: Specific human immunodeficiency virus type I reverse transcriptase inhibitors with antiviral activity. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 6863−6867.

(42) Bell, F. W.; Cantrell, A. S.; Hogberg, M.; Jaskunas, S. R.; Johansson, N. G.; Jordon, C. L.; Kinnick, M. D.; Lind, P.; Morin, J. M.; Noreen, R.; Oberg, B.; Palkowitz, J. A.; Parrish, C. A.; Pranc, P.; Sahlberg, C.; Ternansky, R. J.; Vasileff, R. T.; Vrang, L.; West, S. J.; Zhang, H.; Zhou, X. X. Phenethylthiazolethiourea (PETT) compounds, a new class of HIV-1 reverse transcriptase inhibitors. 1. Synthesis and basic structure−activity relationship studies of PETT analogues. *J. Med. Chem.* **1995**, *38*, 4929−4936.

(43) Pauwels, R.; Andries, K.; Desmyter, J.; Schols, D.; Kukla, M. J.; Breslin, H. J.; Raeymaeckers, A.; Van Gelder, J.; Woestenborghs, R.; Heykants, J.; Schellekens, K.; Janssen, M. A. C.; De Clercq, E.; Janssen, P. A. J. Potent and selective inhibition of HIV-1 replication in vitro by a novel series of tibo derivatives. *Nature* **1990**, *343*, 470−474.

(44) Pauwels, R.; Andries, K.; Debyser, Z.; Van Daele, P.; Schols, D.; Stoffels, P.; De Vreese, K. D.; Woestenborghs, R.; Vandamme, A. M.; Janssen, C. G. M.; Anne, J.; Cauwenbergh, G.; Desmyter, J.; Heykants, J.; Janssen, M. A. C.; De Clercq, E.; Janssen, P. A. J. Potent and highly selective human immunodeficiency virus type 1 (HIV-1) inhibition by a series of α-anilinophenylacetamide derivatives targeted at HIV-1 reverse transcriptase. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 1711−1715.

(45) Fujiwara, T.; Sato, A.; El-Farrash, M.; Miki, S.; Abe, K.; Isaka, Y.; Kodama, M.; Wu, Y. M.; Chen, L. B.; Harada, H.; Sugimoto, H.; Hatanaka, M.; Hinuma, Y. S-1153 inhibits replication of known drug-resistant strains of human immunodeficiency virus type 1. *Antimicrob. Agents Chemother.* **1998**, *42*, 1340−1345.

(46) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E., III; Wang, J.; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Vincent, M.; Crowley, V.; Tsui, V.; Gohlke, H.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. *AMBER7*, University of California, San Francisco, 2002.

(47) Wang, J. M.; Cieplak, P.; Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **2000**, *21*, 1049−1074.

(48) Honig, B.; Nicholls, A. Classical electrostatics in biology and chemistry. *Science* **1995**, *268*, 1144−1149.

(49) Gilson, M. K.; Honig, B. Calculation of the total electrostatic energy of a macromolecular system: Solvation energies, binding energies, and conformational analysis. *Proteins* **1998**, *4*, 7−18.

(50) Sitkoff, D.; Sharp, K. A.; Honig, B. Accurate calculation of hydration free energies using macroscopic solvent models. *J. Phys. Chem.* **1994**, *98*, 1978−1988.
(51) Sanner, M. F.; Olson, A. J.; Spehner, J. C. Reduced surface—An efficient way to compute molecular surfaces. *Biopolymers* **1996**, *38*, 305−320.
(52) Wang, J. M.; Wang W.; Huo, S. H.; Lee, M.; Kollman, P. A. A solvation model based on weighted solvent accessible surface area. *J. Phys. Chem. B.* **2001**, *105*, 5055−5067.

JM049606E