

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

Junmei Wang,[†] Paul Morin,[‡] Wei Wang,[†] and Peter A. Kollman^{*,†}

Contribution from the Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143-0446, and Dupont Pharmaceuticals, DuPont Experimental Station, Wilmington, Delaware 19880-0353

Received October 31, 2000

Abstract: In this work, a new ansatz is presented that combines molecular dynamics simulations with MM-PBSA (Molecular Mechanics Poisson–Boltzmann/surface area) to rank the binding affinities of 12 TIBO-like HIV-1 RT inhibitors. Encouraging results have been obtained not only for the relative binding free energies, but also for the absolute ones, which have a root-mean-square deviation of 1.0 kcal/mol (the maximum error is 1.89 kcal/mol). Since the root-mean-square error is rather small, this approach can be reliably applied in ranking the ligands from the databases for this important target. Encouraged by the results, we decided to apply MM-PBSA combined with molecular docking to determine the binding mode of efavirenz SUSTIVATM another promising HIV-1 RT inhibitor for which no ligand–protein crystal structure had been published at the time of this work. To proceed, we define the following ansatz: Five hundred picosecond molecular dynamics simulations were first performed for the five binding modes suggested by DOCK 4.0, and then MM-PBSA was carried out for the collected snapshots. MM-PBSA successfully identified the correct binding mode, which has a binding free energy about 7 kcal/mol more favorable than the second best mode. Moreover, the calculated binding free energy (–13.2 kcal/mol) is in reasonable agreement with experiment (–11.6 kcal/mol). In addition, this procedure was also quite successful in modeling the complex and the structure of the last snapshot was quite close to that of the measured 2.3 Å resolution crystal (structure the root-mean-square deviation of the 54 C_α around the binding site and the inhibitor is 1.1 Å). We want to point out that this result was achieved without prior knowledge of the structure of the efavirenz/RT complex. Therefore, molecular docking combined with MD simulations followed by MM-PBSA analysis is an attractive approach for modeling protein complexes a priori.

Introduction

The main biological function of HIV-1 reverse transcriptase (RT) is to transcribe the HIV-1 RNA genome into double-stranded DNA, which is subsequently integrated into the host cell genome by an integrase enzyme. The RT is a heterodimer composed of two subunits, p66 and p51, and two binding sites are important for this target. The first, the dNTP substrate binding site, where the 3'-OH of the primer terminus contacts with the enzyme, is located in the p66 palm subdomain. Besides the dNTP site, HIV-1 RT also has an allosteric site, which is also located in the p66 palm subdomain, but distinct from the dNTP-binding site. In recent years, tens of crystal structures of HIV-1 RT have been solved including unliganded structures, complexes with DNA, and complexes with allosteric inhibitors. Given these crystal structures, people have begun to elucidate the mechanisms of how the DNA and the inhibitors interact with the enzyme, and HIV-1 RT has become an active target of rational drug development for AIDS. By now, FDA has approved several inhibitors of this target as anti-HIV drugs, which fall into two categories: nucleoside analogue RT inhibitor (NRTI, such as AZT, ddI, ddC, 3TC, and d4T)^{1–4} and nonnucleoside analogue RT inhibitor (NNRTI, such as NEVI-

RAPINE, DELAVIRDINE, and efavirenz).^{3,5–14} Both the NRTIs and NNRTIs attack HIV at the same stage (reverse transcription and viral DNA synthesis) in its life cycle, by inhibiting the HIV-1 enzyme reverse transcriptase. The NRTIs compete with the viral DNA to bind in the same site (Figure 1c), while NNRTIs bind in an allosteric site (Figure 1b), which interferes with viral RNA binding to HIV-1 RT by inducing a conformational change of the enzyme. The root-mean-square deviation

(1) Larder, B. A. *Inhibitors of HIV reverse transcriptase as antiviral agents and drug resistance*; Cold Spring Harbor Lab. Press: Cold Spring Harbor, NY, 1993; pp 205–222.

(2) Schinazi, R. F. Competitive inhibitors of human immunodeficiency virus reverse transcriptase. *Perspect. Drug Discov. Des.* **1993**, *1*, 151–180.

(3) De Clercq, E. Antiviral therapy of human immunodeficiency virus infection. *Clin. Microbiol. Rev.* **1995**, *8*, 200–239.

(4) De Clercq, E. HIV resistance to reverse transcriptase inhibitors. *Biochem. Pharmacol.* **1994**, *47*, 155–169.

(5) Young, S. D. Nonnucleoside inhibitors of HIV-1 reverse transcriptase. *Perspect. Drug Discov. Des.* **1993**, *1*, 181–192.

(6) De Clercq, E. Toward improved anti-HIV chemotherapy: therapeutic strategies for intervention with HIV infections. *J. Med. Chem.* **1995**, *38*, 2491–2517.

(7) Arnold, E.; Das, K.; Ding, J.; Yadav, P. N. S.; Hsiou, Y.; Boyer, P. L.; Hughes, S. H. Targeting HIV reverse transcriptase for anti-AIDS drug design. *Drug Des. Discov. Des.* **1996**, *13*, 29–47.

(8) Ding, J.; Das, K.; Moereels, H.; Koymans, L.; Andries, K.; Janssen, P. A. J.; Hughes, S. H.; Arnold, E. Structures of HIV-1 RT/TIBO R 86183 complex reveals similarity in the binding of diverse nonnucleoside inhibitors. *Struct. Biol.* **1995**, *2*, 407–415.

* Address correspondence to this author.

[†] University of California.

[‡] DuPont Experimental Station.

between the allosteric liganded complex and unliganded protein is more than 5.0 Å and large orientation changes occur in helix 7 (254–269), helix 8 (277–282), and helix 9 (297–311) of the p66 subunit. A large conformational difference also exists between the complexes bound to DNA and the NNRTI (the root-mean-square deviation is about 3.5 Å), especially for a three-stranded β -sheet in the p66 subunit that contains the catalytically essential amino acids for viral DNA binding (primarily a triad of aspartic acids). NNRTIs are particularly attractive drug candidates because the binding site is unique to the RT of HIV-1 and, therefore, they are less likely to cause adverse side effects by disruption of the normal DNA polymerase activity.^{15–18}

Even though the free energy perturbation (FEP) and thermodynamic integration (TI) yield rigorous and accurate free energy differences on model potential surfaces of interest (including the biological targets), they are extremely time-consuming. In addition, since the methods are implemented numerically, sufficient statistical sampling must be carried out. These kinds of problems prevent FEP and TI from being widely used in structure-based design.^{19,20} Recently, several alternative methods have been developed to estimate the binding free energy in a

(9) 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-1H-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.

(10) 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.; Lin, J. H.; Chen, I.-W.; Schleif, W. A.; Sardana, V. V.; Long, W. J.; Byrnes, V. B.; Emimi, 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.

(11) Patel, M.; Ko, S. S.; Mchugh, R. J., Jr.; Markwalder, J. A.; Srivastava, A. S.; Cordova, B. C.; Klabe, R. M.; Erickson-Viitanen, S.; Trainor, G. L.; Seitz, S. P. Synthesis and evaluation of analogs of efavirenz (SUSTIVA™) as HIV-1 reverse transcriptase inhibitors. *Bioorg., Med. Chem. Lett.* **1999**, *9*, 2805–2810.

(12) Merluzzi, V. J.; Hargrave, K. D.; Labadia, M.; Grozinger, K.; Skoog, M.; Wu, J. C.; Shih, C. K.; Eckner, K.; Hattox, S.; Adams, J.; Rosenthal, A. S.; Faanes, R.; Eckner, R. J.; Koup, R. A.; Sullivan, J. L. Inhibition of HIV-1 replication by a nonnucleoside reverse transcriptase inhibitor. *Science* **1990**, *250*, 1411–1413.

(13) Eriksson, M. A.; Pitera, J.; Kollman, P. A. Prediction of the binding free energies of new TIBO-like HIV-1 reverse transcriptase inhibitors using a combination of PROFEC, PBSA, CMC/MD and free energy calculations. *J. Med. Chem.* **1999**, *42*, 868–881.

(14) Smith, R. H.; Jorgensen, W. L.; Tirado-Rives, J.; Lamb, M. L.; Janssen, P. A. J.; Michejda, C. J.; Smith, B. K. Prediction of binding affinities for TIBO inhibitors of HIV-1 reverse transcriptase using Monte Carlo simulations in a linear response method. *J. Med. Chem.* **1998**, *14*, 5272–5286.

(15) Gao, H. Q.; Sarafianos, S. G.; Arnold, E.; Hughes, S. H. Similarities and differences in the RNase H activities of human immunodeficiency virus type 1 reverse transcriptase and moloney murine leukemia virus reverse transcriptase. *J. Mol. Biol.* **1999**, *294*, 1097–1113.

(16) Sarafianos, S. G.; Das, K.; Ding, J. P.; Boyer, P. L.; Hughes, S. H.; Arnold, E. Touching the heart of HIV-1 drug resistance: the fingers close down on the dNTP at the polymerase active site. *Chem. Biol.* **1999**, *6*, R137–R146.

(17) Sarafianos, S. G.; Das, K.; Clark, A. D.; Ding, J. P.; Boyer, P. L.; Hughes, S. H.; Arnold, E. Lamivudine (3TC) resistance in HIV-1 reverse transcriptase involves steric hindrance with beta-branched amino acids. *Proc. Natl. Acad. Sci.* **1999**, *96*, 10027–10032.

(18) Madrid, M.; Jacobo-Molina, A.; Ding, J.; Arnold, E. Major subdomain rearrangement in HIV-1 reverse transcriptase simulated by molecular dynamics. *Proteins-Struct. Funct., Genet.* **1999**, *35*, 332–337.

(19) Kollman, P. A. Free energy calculations: Applications to chemical and biochemical phenomena. *Chem. Rev.* **1993**, *93*, 2395–2417.

(20) Beveridge, D. L.; Dicapua, F. M. Free energy via molecular simulations: Application to chemical and biochemical system. *Annu. Rev. Biophys. Chem.* **1989**, *18*, 431–492.

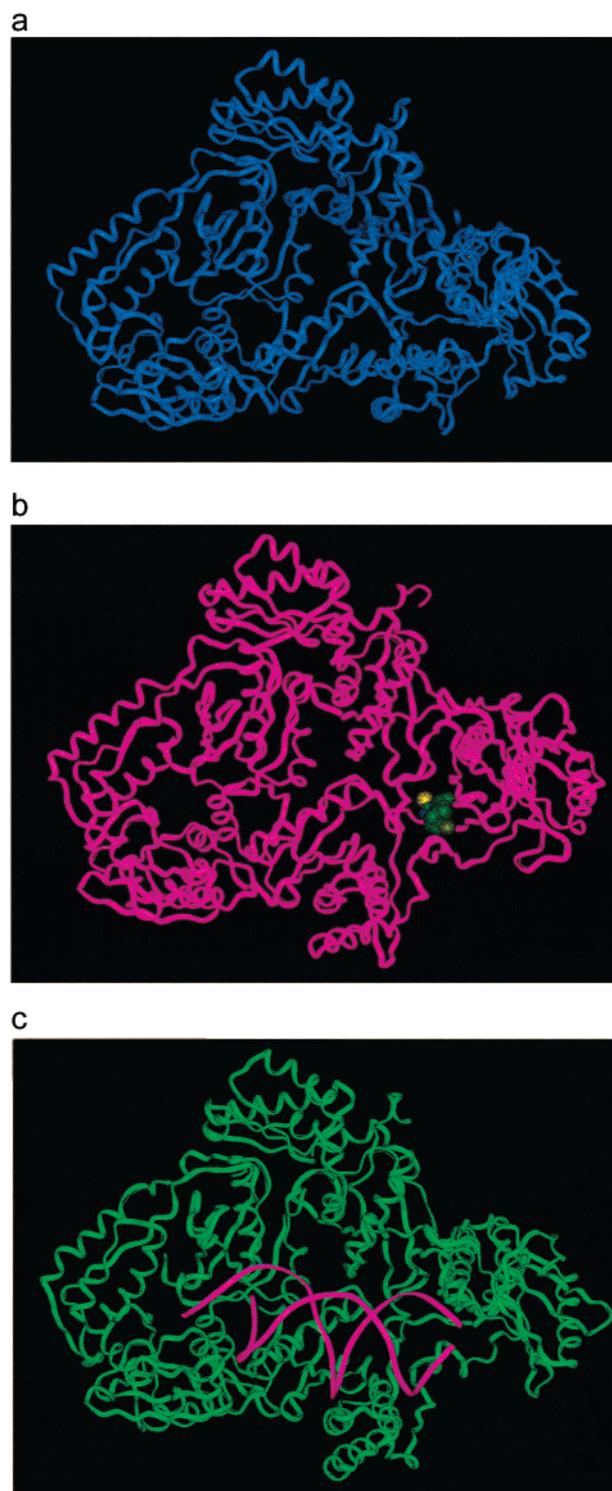
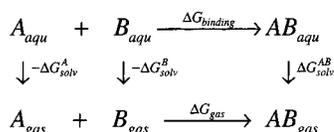


Figure 1. Ribbon representation of the crystal structures of HIV-1 reverse transcriptase: (a) unliganded HIV-1 RT (1DLO⁵⁹); (b) complex with 8Cl-TIBO at the allosteric site (1UWB³⁹), root-mean-square deviation of C α between 1UWB and 1DLO is 5.38 Å; and (c) complex with DNA (1C9R⁶⁰), root-mean-square deviations of C α between 1C9R and 1DLO, and 1C9R and 1UWB are 4.49 and 3.47 Å, respectively.

fast and practical way, including the linear interaction (LIE) method,^{21–25} CMC-MD,¹³ MM-PBSA, and some “Ludi”-like approaches.^{26,27} “Ludi”-like approaches, which are widely used in *de novo* design and database screening, are pure empirical

(21) Åqvist, J.; Medina, C.; Samuelsson, J. New method for prediction binding affinity in computer-aided drug design. *Protein Eng.* **1994**, *7*, 385–391.

methods in estimating the binding free energy. For example, the entropy is simply estimated by the number of degrees of freedom lost upon ligand binding. Although these approaches are efficient and less computer resource demanding than the other methods noted above, they are not as physically correct as, for example, molecular dynamics based approaches such as LIE and MM-PBSA. Moreover, the parameters developed from the training systems may not be transferred to others without problems. LIE, on the other hand, is a semiempirical method first proposed by Åqvist et al.^{21–24} It is based on linear-response-like assumptions that the binding free energy is the combination of the weighted electrostatic and van der Waals interactions between the ligand and the receptor. Although the coefficient of the electrostatic interaction energy is about 0.5, the van der Waals contribution is quite varied for different systems. This suggests that LIE might have difficulty in predicting the binding affinities for significantly different compounds. The LIE approach has been further extended by Jorgensen's group by including other empirical or semiempirical parameters.¹⁴ Sham et al. have recently compared a series of LIE-like approaches.²⁵ Unlike the linear interaction energy (LIE) method, MM-PBSA²⁸ applies no empirical parameters in its free energy calculations, which makes it a promising method for ranking very different compounds from database searching. In MM-PBSA, the free energy of $A + B \rightarrow AB$ is calculated using the following thermodynamic cycle:



$$\begin{aligned} \Delta G_{\text{binding}} &= \Delta G_{\text{gas}} - \Delta G_{\text{sol}}^A - \Delta G_{\text{sol}}^B + \Delta G_{\text{sol}}^{AB} \\ &= \Delta H_{\text{gas}} - T\Delta S - \Delta G_{\text{PBSA}}^A - \Delta G_{\text{PBSA}}^B + \Delta G_{\text{PBSA}}^{AB} \\ &= \Delta H_{\text{gas}} - T\Delta S + \Delta\Delta G_{\text{PB}} + \Delta\Delta G_{\text{SA}} \end{aligned} \quad (1)$$

$$\Delta H_{\text{gas}} \approx \Delta E_{\text{gas}} = \Delta E_{\text{intra}} + \Delta E_{\text{electrostatic}} + \Delta E_{\text{vdW}} \quad (2)$$

$$\Delta\Delta G_{\text{PB}} = \Delta G_{\text{PB}}^{AB} - (\Delta G_{\text{PB}}^A + \Delta G_{\text{PB}}^B) \quad (3)$$

$$\Delta\Delta G_{\text{SA}} = \Delta G_{\text{SA}}^{AB} - (\Delta G_{\text{SA}}^A + \Delta G_{\text{SA}}^B) \quad (4)$$

ΔG_{gas} is the interaction energy between A and B in the gas phase and ΔG_{sol}^A , ΔG_{sol}^B , and $\Delta G_{\text{sol}}^{AB}$ are the solvation free energies of A , B , and AB , which are estimated using a continuum approach (Poisson–Boltzmann/surface area), i.e., $\Delta G_{\text{sol}}^{AB} = \Delta G_{\text{PBSA}}^{AB} =$

(22) Hansson, T.; Åqvist, J. Estimation of binding free energies for HIV protease inhibitors by molecular dynamics simulations. *Protein Eng.* **1995**, *8*, 1137–1144.

(23) Åqvist, J. Calculation of absolute binding free energies for charged ligands and effects of long-range electrostatic interactions. *J. Comput. Chem.* **1996**, *17*, 1587–1597.

(24) Wang, W.; Wang, J.; Kollman, P. A. What determines the van der Waals coefficient β in the LIE (linear interaction energy) method to estimate binding free energies using molecular dynamics simulations. *Proteins: Struct. Funct., Genet.* **1999**, *34*, 395–402.

(25) Sham, Y. Y.; Chu, Z. T.; Warshel, A. Examining methods for calculations of binding free energies: LRA, LIE, PDL-LRA, and PDL/S-LRA calculations of ligands binding to an HIV protease. *Proteins: Struct., Funct., Genet.* **2000**, *39*, 393–407.

(26) Bohm, H. J. The development of a simple empirical scoring function to estimate the binding constant for a protein–ligand complex of known three-dimensional structure. *J. Comput. Aided Mol. Des.* **1994**, *8*, 243–256.

(27) Zou, X. Q.; Sun, Y.; Kuntz, I. D. Inclusion of solvation in ligand binding free energy calculations using generalized-Born model. *J. Am. Chem. Soc.* **1999**, *121*, 8033–8043.

$\Delta G_{\text{PB}}^{AB} + \Delta G_{\text{SA}}^{AB}$, etc. For a series of compounds with similar structures and binding modes, the entropy contribution can be omitted if one is only interested in the relative order of binding affinities. However, if one intends to obtain the absolute binding free energy or the compounds are very different, the contribution of entropy cannot be neglected. There are several approaches to estimate the entropy, including the normal-mode analysis,²⁹ the quasi-harmonic analysis,^{30–32} and the quasi-Gaussian approach.³³ The first two approaches, which are more practical for the biological systems, are much similar in theory except that the atomic fluctuation matrix in quasi-harmonic analysis is not from a normal mode calculation, but is obtained from the snapshots of MD simulations. Estimates of entropy from normal-mode analysis have some limitations, for example, the anharmonic contribution is not taken into account and low frequency modes that lead to large displacements are not treated accurately in the harmonic limit. On the other hand, quasi-harmonic analysis, which is an alternative to harmonic analysis, incorporates some effects due to the anharmonic nature of macromolecules. However, there are some serious problems for the quasi-harmonic analysis if the microstates are not well sampled with MD simulations. Considering that normal-mode analysis has been successfully applied in estimating the binding entropy for several biological systems,^{25–30} in this study, the conformational entropies were calculated through normal-mode analysis. By now, MM-PBSA has been successfully applied to binding free energy calculations for several systems.^{28,34–38}

Recently, two crystal structures of HIV-1 RT in complex with 8Cl-TIBO (R86183) and 9Cl-TIBO (R82913)^{8,39} have been solved at 3.0 Å resolution. More than tens of TIBO derivatives with various activities^{14,40} have also been reported. It is interesting to investigate how well the available theoretical approaches predict the binding affinities of 8Cl-TIBO and its derivatives. With promising free energy calculation methods,

(28) Srinivasan, J.; Miller, J.; Kollman, P. A.; Case, D. A. Continuum solvent studies of stability of RNA hairpin loops and helices. *J. Biol. Struct. Dyn.* **1998**, *16*, 671–682.

(29) Wilson, E. B.; Decius, J. C.; Cross, P. C. *Molecular Vibrations*; McGraw-Hill: New York, 1955.

(30) Brooks, B. R.; Janežič, D.; Karplus, M. Harmonic analysis of large system. I. methodology. *J. Comput. Chem.* **1995**, *16*, 1522–1542.

(31) Janežič, D.; Brooks, B. R. Harmonic analysis of large system. II. comparison of different protein models. *J. Comput. Chem.* **1995**, *16*, 1543–1553.

(32) Janežič, D.; Venable, R. M.; Brooks, B. R. Harmonic analysis of large system. III. comparison with molecular dynamics. *J. Comput. Chem.* **1995**, *16*, 1554–1566.

(33) Roccatano, D.; Amadei, A.; Apol, M. E. F.; Nola, A. D.; Berendsen, H. J. C. Application of the quasi-Gaussian entropy theory to molecular dynamics simulations of Lennard-Jones fluids. *J. Chem. Phys.* **1998**, *109*, 6358–6363.

(34) Chong, L. T.; Duan, Y.; Wang, L.; Massova, I.; Kollman, P. A. Molecular dynamics and free-energy calculations applied to affinity maturation in antibody 48G7. *Proc. Natl. Acad. Soc.* **1999**, *96*, 14330–14335.

(35) Massova, I.; Kollman, P. A. Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. *Perspect. Drug Discov. Des.* **2000**, *18*, 113–135.

(36) Massova, I.; Kollman, P. A. Computational alanine scanning to probe protein–protein interactions: a novel approach to evaluate binding free energies. *J. Am. Chem. Soc.* **1999**, *121*, 8133–8143.

(37) Lee, M. R.; Duan, Y.; Kollman, P. A. Use of MM-PBSA in estimating the free energies of proteins: application to native, intermediates, and unfolded villin headpiece. *Proteins: Struct. Funct., Genet.* **2000**, *39*, 309–316.

(38) Reyes, C. M.; Kollman, P. A. Investigating the binding specificity of U1A-RNA by computational mutagenesis. *J. Mol. Biol.* **2000**, *295*, 1–6.

(39) Das, K.; Ding, J.; Hsiou, Y.; Clark, A. D. J.; Moereels, H.; Koymans, L.; Andries, K.; Pauwels, R.; Janssen, P. A. J.; Boyer, P. L.; Clark, P.; Smith, R. H. J.; Kroeger-Smith, M. B.; 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.

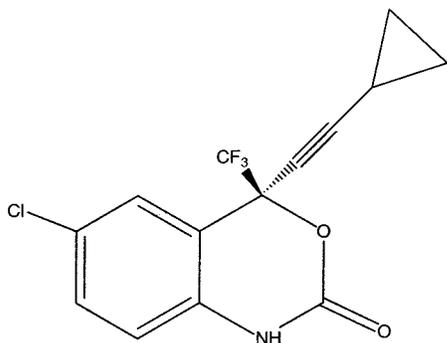


Figure 2. Sketch of efavirenz (also known as SUSTIVA™ and DMP-266).

one might further screen hits from a database search and reliably evaluate the compounds from the *de novo* design. Smith and Jorgensen recently applied the LIE method combined with Monte Carlo simulations to predict the binding free affinities for 12 TIBO-like inhibitors and obtained encouraging results.¹⁴ For their LIE method, a cavity term was also included and all of the coefficients were determined empirically. Eriksson, Pitara, and Kollman have also successfully predicted the binding free energies of several TIBO-like compounds using their chemical Monte Carlo/molecular dynamics (CMC/MD) approach.¹³ With this new method, the rank order of predicted binding affinities was in good agreement with the experimental results. Moreover, a compound suggested by PROFEC⁴¹ (a *de novo* design program) has been calculated to have higher biological activity than any other known TIBO derivatives. This encouraging discovery was further verified by free energy perturbation and thermodynamic integration calculations.¹³

efavirenz (SUSTIVA™, Figure 2)^{10,11} is a novel, highly potent NNRTI of HIV-1 RT and it is very similar to TIBO in structure: efavirenz has a benzoxazine group, while the TIBO-like compound has a similar benzazole group, and efavirenz has a hydrophobic cyclopropane, while the TIBO-like compound has a hydrophobic ethylene group. Moreover, both of them have a carbonyl (-thione) group at a similar site within the heteroatomic rings. Considering that the crystal structure of efavirenz/HIV-1 RT has not been released, it is a good case for testing how well the theoretical methods predict the binding mode of this complex.

This paper describes how we have applied MM-PBSA to predict the binding free energies for 12 TIBO-like NNRTIs shown in Table 1. We also discuss how to improve the method so that it can be applied to rank tens, even hundreds of inhibitors with a reasonable calculation time in the future. At the “Structural Biology of AIDS Targets” meeting at the NIH, June 2000, two of us, Junmei Wang and Peter Kollman (UCSF), met Paul Morin (Dupont Pharmaceuticals) and agreed to test the methodologies of structure/binding predictions in a “blind” fashion since the crystal structure of efavirenz bound to HIV-1 RT was unpublished but determined and known at Dupont Pharmaceuticals at high resolution. To predict the binding mode of efavirenz, J.W. applied the following ansatz: molecular

(40) Ho, W.; Kukla, M. J.; Breslin, H. J.; Ludovici, D. W.; Grous, P. P.; Diamond, C. J.; Miranda, M.; Rodgers, J. D.; Ho, C. Y.; De Clercq, E.; Pauwels, R.; Andries, K.; Janssen, M. A. C.; Janssen, P. A. J. Synthesis and anti-HIV-1 activity of 4,5,6,7-tetrahydro-5-methylimidazo-[4,5,1-jk]-[1,4]benzodiazepin-2(1H)-one (TIBO) derivatives. 4. *J. Med. Chem.* **1995**, *38*, 794–802.

(41) Radmer, R. J.; Kollman, P. A. The application of three approximative free energy calculation methods to structure based ligand design: Trypsin and its complex inhibitors. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 215–227.

Table 1. Select Set of TIBO Derivatives

compd no.	R ₈	R ₉	R	binding free energy (exptl, in kcal/mol) ^a
I	Cl	H	S	−11.87
II	F	H	S	−11.69
III	CH ₃	H	S	−11.16
IV	C ₂ H ₅	H	S	−10.69
V	CH ₃ O	H	S	−10.60
VI	H	Cl	S	−10.60
VII	H	H	S	−10.44
VIII	CN	H	S	−10.29
IX	C ₂ H ₅	H	O	−9.02
X	CH ₃	H	O	−8.52
XI	CN	H	O	−8.43
XII	H	H	O	−7.81

^a Experimental binding free energies are calculated from IC₅₀ (refs 14 and 40) using the following relationship: $\Delta G_{\text{binding}} = RT \ln K_{\text{dissociated}} = RT \ln(\text{IC}_{50} + 0.5C_{\text{enzyme}}) \approx RT \ln \text{IC}_{50}$, where R is ideal gas constant, T is temperature in K (298 K is used in this paper), and C_{enzyme} is the concentration of enzyme, which is a very small number after equilibration and can be omitted in most cases.

docking was performed to generate several distinct binding orientations; then a 500 ps molecular dynamics simulation was performed to further relax the complex; finally MM-PBSA was applied to evaluate the binding affinity for each binding mode. The binding mode that has the lowest binding free energy was expected to be the most favorable binding mode. We discuss the result of this prediction below.

Computational Method

Quantum Mechanics Calculations. The 12 TIBO-like ligands listed in Table 1 and efavirenz were first optimized at the HF/6-31G* level using the Gaussian 94 package⁴² and electrostatic potentials (ESP) were then generated using Merz–Singh–Kollman van der Waals parameters.^{43,44} The atomic charges used for the molecular mechanics calculations were derived from the ESP using the RESP^{45,46} program implemented in the AMBER 5 package.⁴⁷ With the aim to investigate how well our new version of the AMBER force field (Parm99)⁴⁸ describes the structures and conformations of this class of compounds,

(42) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T. A.; Peterson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. *Gaussian 94*; Gaussian, Inc.: Pittsburgh, PA, 1995.

(43) Besler, B. H.; Merz, K. M.; Kollman, P. A. Atomic charges derived from semiempirical methods. *J. Comput. Chem.* **1990**, *11*, 431.

(44) Singh, U. C.; Kollman, P. A. An approach to computing electrostatic charges for molecules. *J. Comput. Chem.* **1984**, *5*, 129.

(45) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **1993**, *97*, 10269–10280.

(46) Fox, T.; Kollman, P. A. Application of the RESP methodology in the parametrization of organic solvents. *J. Phys. Chem. B* **1998**, *102*, 8070–8079.

we also performed high-level ab initio calculations for 8-Cl TIBO and 9-Cl TIBO using six different conformers, respectively. All structures were minimized at the HF/6-31G* level, which was then followed by a single point calculation at the B3LYP/6-311G(2d,p) level. Zero point energy (ZPE) for each structure was calculated at the same level as the minimization.

Molecular Docking. For efavirenz, the binding orientations were first estimated with the docking program DOCK4.0.^{49,50} Considering the similarities of efavirenz and TIBO in structure, we used the crystal structure of HIV-1 RT/8Cl-TIBO (1uwf)³⁹ to generate the receptor site and the energetic grids for the following docking calculations. To get all the possible binding orientations, rigid and flexible docking were performed not only for efavirenz but also for its five analogues (compounds 4c, 4d, 4e, 4f, and 4m in ref 11), since they should have a similar binding mode as that of efavirenz. Five binding modes were identified after cluster analysis. The two kinds of docking were also carried out for 8Cl-TIBO and 9Cl-TIBO as controls.

Molecular Dynamics Simulations. Molecular dynamics simulations of 12 complexes were carried out using the AMBER 5.0 suite of programs⁴⁷ with a new version of AMBER force field (Parm99)⁴⁸ described by eq 5.

$$V = \sum_{\text{bonds}} K_r(r - r_{eq})^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_{eq})^2 + \sum_{\text{dihedrals}} \frac{V_n}{2}(1 + \cos(n\phi - \gamma)) + \sum_{i < j} \left\{ \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} \right] + \frac{q_i q_j}{\epsilon R_{ij}} \right\} \quad (5)$$

For each TIBO complex, a 20 Å water cap was first added near the allosteric binding site. The ligand, water molecules, and protein residues that are within 20 Å of the mass center of the ligand were flexible during the MD simulations (cyan and interior area in Figure 3). The simulations were carried out at 300 K with a time step of 2.0 fs. The nonbonded cutoff was set to 9.0 Å and SHAKE⁵¹ was applied for all the bonds involving hydrogen atoms (we found that applying a larger cutoff did not significantly change the subsequent MM-PBSA analysis). After equilibration for 100 ps, conformations were collected every 2 ps for the following 200 ps simulation. Finally, 100 snapshots were collected for postprocessing analysis.

For efavirenz, since the initial structure of each binding mode is from molecular docking, both the ligand and the protein have greater propensities to drift away from the initial docked-structures during the MD simulations. Therefore, a larger flexible area and a longer MD simulation were necessary. For each binding mode, 500 ps MD simulations were carried out at a time step of 2.0 fs. A 30 Å water cap was added around the ligand and all the residues within 30 Å of the mass center of the ligand were allowed to be flexible during the MD simulations. The nonbonded cutoff was set to 12.0 Å and a dual cutoff of 30.0 Å was also used, with the nonbonded list for the longer range cutoff updated every 25 MD steps. One-hundred snapshots were collected from the last 200 ps simulations for postprocessing analysis.

MM-PBSA and Normal Mode Calculations. For each snapshot collected during the simulation, ligand–receptor interaction energies ($\Delta E_{\text{electrostatic}}$, ΔE_{vdW}) were calculated with the ANAL program of

(47) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E., III; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Ferguson, D. M.; Radmer, R. J.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. *AMBER* 5; University of California: San Francisco, 1997.

(48) Wang, J.; Cieplak, P.; Kollman, P. A. How Well Does a RESP (Restrained Electrostatic Potential) Model Do in Calculating the Conformational Energies of Organic and Biological Molecules? *J. Comput. Chem.* **2000**, *21*, 1049–1074.

(49) Ewing, T. J. A.; Kuntz, I. D. Critical evaluation of search algorithms used in automated molecular docking. *J. Comput. Chem.* **1997**, *18*, 1175–1189.

(50) 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.

(51) Ryckaert, J. P.; Cicciotti, G.; Berendsen, H. J. C. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–320.

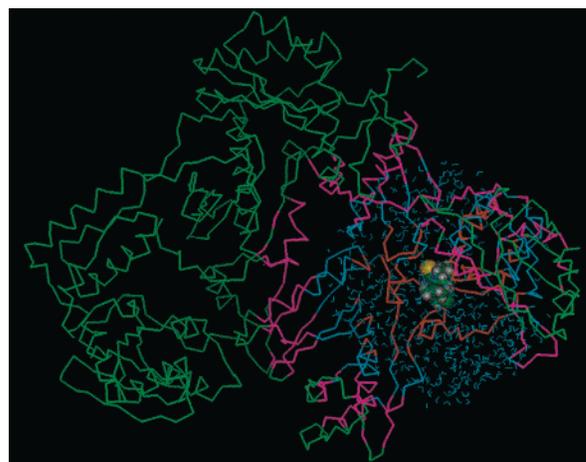


Figure 3. Residues for molecular dynamic simulations, PBSA, and normal mode calculations for HIV-1 RT/TIBO and efavirenz. All the residues in cyan and red as well as the ligand are flexible during the molecular dynamics simulations; all the residues in red (without water) including the ligand are used for normal mode calculations; all the residues in magenta and the flexible residues of receptor as well as the ligand (without water) are applied for PBSA calculations. Residue selections are different for TIBO and efavirenz (see text).

AMBER 5.0.⁴⁷ As to the solvation energies, the electrostatic parts (reaction field energies (ΔE_{PB})) were calculated using DELPHI^{52,53} with PARSE radii.⁵⁴ The nonpolar contributions ($\Delta E_{\text{nonpolar}}$) were estimated using a simple empirical relation:⁵⁴ $\Delta E_{\text{nonpolar}} = \sigma A + b$, where A is the solvent-accessible surface area that was estimated using the MSMS program.⁵⁵ σ and b are empirical constants and in this work we used 0.0054 and 0.92 kcal/mol for σ and b , respectively. To speed the solvation energy calculations, only the residues within 30 Å (for efavirenz, 40 Å used) of the mass center of the ligand were used for the PBSA calculations (including the ligand but excluding water molecules—the magenta and the interior area in Figure 3).

The normal mode calculation is extremely time-consuming for large systems; thus, only residues within 12 Å of the mass center of the ligand (including the ligand, but excluding water molecules) were used for the normal mode calculation (red and interior area in Figure 3). A distance-dependent dielectric constant ($\epsilon = 4R_{ij}$) was applied for minimizations. Considering the root-mean-square deviation of entropy is relatively small for different conformations, only 20 snapshots (every fifth snapshot of the 100 collected snapshots) were used to estimate the contribution of the entropies of association.

Results and Discussion

Inhibitor Structures. There are several important conformers of the TIBO-like compounds. The diazepine ring can adopt two different conformations: TS and TC/BS. In the TS conformation, the C5 and N6 are located on opposite sides of the plane of the ring, reminiscent of the twist-chair conformation of cyclohexane. In the TC/BS conformation, both C5 and N6 lie on the same side of the plane, somewhat analogous to the boat conformation of cyclohexane. Moreover, in solution the two chiral atoms, C5 and N6, can adopt both *R* and *S* configurations owing to the possibility of inversion. Tables 2 and 3 list the ab initio and molecular mechanical energies of six conformers of

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

(53) 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.

(54) Sitkoff, D.; Sharp, K. A.; Honig, B. Accurate calculation of hydration free energies using macroscopic solvent models. *J. Phys. Chem.* **1994**, *98*, 1978–1988.

(55) Sanner, M. F.; Olson, A. J.; Spehner, J. Reduced surface—an efficient way to compute molecular surfaces. *Biopolymers* **1996**, *38*, 305–320.

Table 2. Results of QM (B3LYP/6-311+G(2d,p) and MM (AMBER) Energies (in kcal/mol) for Six 8Cl-TIBO Conformations^a

conformation	<i>E</i> (B3LYP/ 6-311+G(2d,p))	ZPE	rel energy (ab initio)	rel energy (AMBER)
TC/BS (5 <i>S</i> , 6 <i>R</i>)	-1644.66949842	228.94350	0.0	0.0
TC/BS (5 <i>S</i> , 6 <i>R</i>)	-1644.66108854	229.51393	5.80	7.40
TC/BS (5 <i>S</i> , 6 <i>S</i>)	-1644.66807720	229.05947	1.00	0.77
TC/BS (5 <i>S</i> , 6 <i>S</i>)	-1644.66734320	228.82008	1.24	1.70
TS (5 <i>S</i> , 6 <i>R</i>)	-1644.66807237	229.22767	1.15	0.44
TS (5 <i>S</i> , 6 <i>S</i>)	-1644.66710700	229.13914	1.68	1.16

^a The average unsigned error between QM and MM is 0.57 kcal/mol.

8Cl-TIBO and 9Cl-TIBO. Overall our force field can successfully predict the structures and the conformational energies of the TIBO-like compounds. For 8Cl- and 9Cl-TIBO, the root-mean-square errors between the ab initio and molecular mechanical energies are about 0.5 kcal/mol.

When the complex with HIV-1 RT is formed, the ligand is restricted to adopt some special conformations. It is believed that at the C5 position, the *S* absolute configuration is crucial for activity against HIV-1, while at the N6 position, both the *R* and *S* are possible configurations in solution. However, in the crystal structure of the complex, only the TC/BS (5*S*, 6*R*) conformation is ever observed.^{13,56,57} The results of our 300 ps MD simulations are consistent with the above points: all the snapshots have the *S* configuration at C5 and both *R* and *S* configurations are observed for N6. The ratios of *S* isomer at N6 are 1.0%, 6.8%, 38.2%, 10.8%, 94.2%, 1.8%, 89.8%, 98.7%, 24.1%, 11.8%, 10.6%, 98.1%, 72.5%, 49.9%, and 1.6% for the 12 ligands in sequence. Interestingly, we have not observed the TS configuration during the MD simulations, although the ab initio conformational energy of TS (5*S*, 6*R*) is only marginally higher than that of TC/BS (5*S*, 6*R*). The reason is that our initial conformation is TC/BS (5*S*, 6*R*), and within the protein, molecular motions of the inhibitor are significantly restricted and, thus, the TC/BS to TS conversion appears unlikely.

For efavirenz, there is only one chiral atom in the benzoxazine and the *S* configuration is necessary for anti-AIDS activity.

Reparametrization of Poisson–Boltzmann Calculations. Among the several solvation models, PBSA is regarded as an attractive approach for drug design since it works well not only for small and medium size organic compounds but also for the biological molecules such as proteins and DNA. As a model, the parametrization of atomic radii plays an important role in the performance of solvation energy calculations. In 1994, Sitkoff et al. worked out a set of van der Waals radii (PARSE) with a somewhat ad hoc set of partial charges.⁵⁴ Their model achieves an encouraging performance in predicting the solvation energies for both proteins and organic molecules that have function groups similar to those in the amino acids. By now, PARSE radii are widely used in different solvation models, sometimes with slight adjustment. In our previous applications of MM-PBSA,^{25–38} we found that the combination of the Cornell et al. charges⁵⁸ and PARSE radii worked well, presumably because of some cancellation of errors.

(56) Liaw, Y.-C.; Gao, Y.-G.; Robinson, H.; Wang, A. H.-J. Molecular structure of a potent HIV-1 inhibitor belonging to the TIBO family. *J. Am. Chem. Soc.* **1991**, *113*, 1857–1859.

(57) Boessenkool, I. K.; Boeyens, J. A. *J. Cryst. Mol. Struct.* **1980**, *10*, 11–18.

(58) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

Table 3. Results of QM (B3LYP/6-311+G(2d,p) and MM (AMBER) Energies (in kcal/mol) for Six 9Cl-TIBO Conformations^a

conformation	<i>E</i> (B3LYP/ 6-311+G(2d,p))	ZPE	rel energy (ab initio)	rel energy (AMBER)
TC/BS (5 <i>S</i> , 6 <i>R</i>)	-1644.67026316	228.86974	0.0	0.0
TC/BS (5 <i>S</i> , 6 <i>R</i>)	-1644.66303324	229.34254	4.97	7.18
TC/BS (5 <i>S</i> , 6 <i>S</i>)	-1644.66951784	229.00706	0.90	0.31
TC/BS (5 <i>S</i> , 6 <i>S</i>)	-1644.66846141	228.76786	1.04	1.94
TS (5 <i>S</i> , 6 <i>R</i>)	-1644.67031972	229.17900	0.25	0.26
TS (5 <i>S</i> , 6 <i>S</i>)	-1644.66869998	229.09679	1.19	0.10

^a The average unsigned error between QM and MM is 0.50 kcal/mol.

Table 4. Parse Parametrization for F, Cl, and CN^a

no.	compd name	<i>E</i> ^{elec}	SA	<i>E</i> ^{SA}	<i>E</i> ^{calcd}	<i>E</i> ^{exptl}	Δ <i>E</i>
1	fluoromethane	-2.45	159.4	1.78	-0.67	-0.22	-0.45
2	1,1-difluoroethane	-2.82	205.5	2.04	-0.78	-0.11	-0.67
3	2,2,2-trifluoroethanol	-7.12	231.6	2.18	-4.94	-4.31	-0.63
4	fluorobenzene	-3.27	256.4	2.31	-0.96	-0.78	-0.18
5	dichloromethane	-3.26	202.9	2.02	-1.24	-1.36	0.12
6	<i>cis</i> -1,2-dichloroethylene	-2.80	222.9	2.13	-0.67	-0.76	0.09
7	<i>o</i> -chlorotoluene	-3.43	299.4	2.54	-0.89	-1.15	0.26
8	chlorobenzene	-3.23	270.4	2.38	-0.85	-1.01	0.16
9	benzotrile	-7.99	263.0	2.34	-5.65	-4.10	-1.55
10	acetonitrile	-7.29	166.7	1.82	-5.47	-3.89	-1.58
11	propionitrile	-6.80	201.9	2.01	-4.79	-3.85	-0.94
12	butyronitrile	-6.89	233.3	2.18	-4.71	-3.64	-1.07
9 ^b	benzotrile	-6.64	263.0	2.34	-4.30	-4.10	-0.20
10 ^b	acetonitrile	-5.66	166.7	1.82	-3.84	-3.89	0.05
11 ^b	propionitrile	-5.31	201.9	2.01	-3.30	-3.85	0.55
12 ^b	butyronitrile	-5.35	233.3	2.18	-3.17	-3.64	0.47

^a After reoptimization, the average unsigned error and the root-mean-square error of 14 compounds are 0.32 and 0.38 kcal/mol, respectively. The experimental solvation energies are from ref 61. ^b PB is calculated with new parameters (the radii of carbon and nitrogen in the cyano group are set to 1.85 and 1.75 Å, respectively).

However, we should note that the PARSE radii only considered a small subset of atom types in the Cornell et al. force field. It is more consistent to reparametrize the radii based on RESP charges and such a general effort is in progress (S. Huo and J. Wang, unpublished). However, in this work we used PARSE radii, but carried out some parametrization for the new atom types not especially considered by Sitkoff et al.

Since the radius parameters of F and Cl are missing in the PARSE parameter set, we simply applied the AMBER vdW parameters (*R*^{*}). Considering that compounds **VIII** and **XI** also contain cyano groups, which do not exist in the 20 amino acids, it was necessary to verify if the PARSE parameter gives reasonable solvation energy for the cyano compounds. Table 4 lists the solvation energies of some compounds containing F, Cl, or cyano groups. For fluorides and chlorides, the extended parameter set has an unsigned average error of 0.32 kcal/mol. The PARSE parameter set uniformly overestimates the solvation energy of cyano compounds by 1.0–1.5 kcal/mol. Therefore we tried to optimize the radii of carbon and nitrogen in the cyano group and finally they were set to 1.85 and 1.75 Å, respectively. With the new parameters, the unsigned average error for four cyano compounds is reduced to 0.32 kcal/mol. Overall, the extended and revised PARSE parameter set gives very encouraging solvation energies for the 12 compounds in our test set with a root-mean-square deviation of 0.38 kcal/mol.

Binding Free Energies of TIBO-like Inhibitors. Three-hundred picosecond molecular dynamics simulations were performed for all 12 complexes. After 100 ps equilibration, 100 snapshots were collected for the following 200 ps molecular dynamics simulations. Overall, compared to the crystal structure, the root-mean-square deviations of the ligands and the main

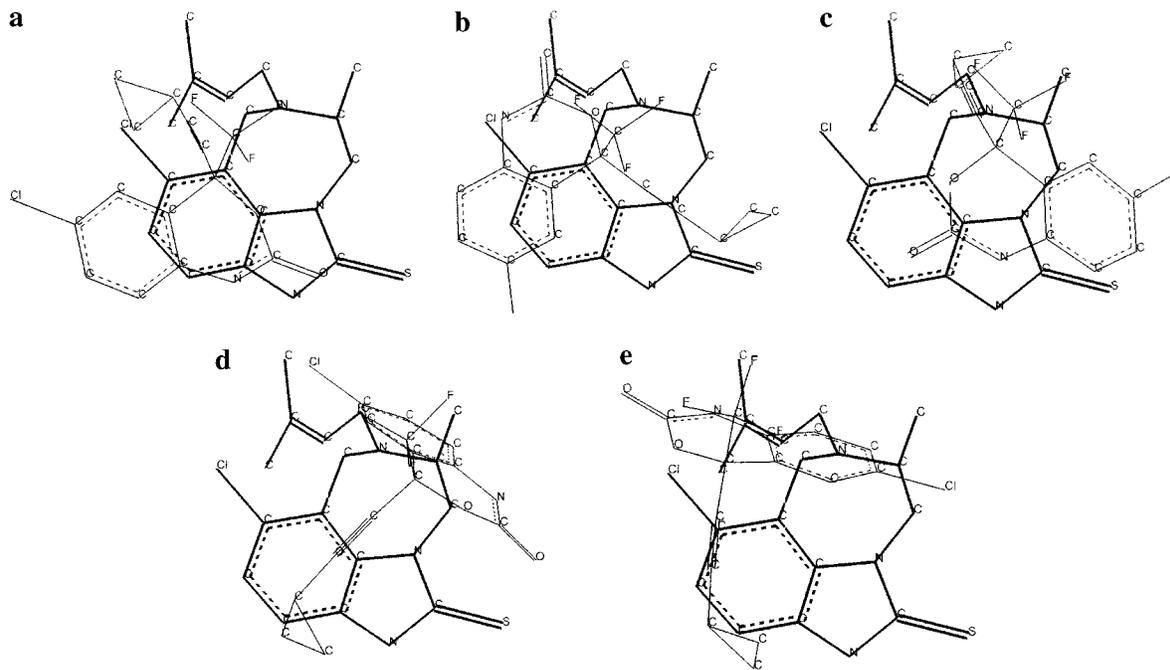


Figure 6. Relative positions and orientations of efavirenz after 500 ps simulation to 8Cl-TIBO: (a) binding mode I; (b) binding mode II; (c) binding mode III; (d) binding mode IV; and (e) binding mode V.

Table 6. Binding Free Energies of the Five Binding Modes of Efavirenz^a

no.	ΔE^{intra}	ΔE^{vdw}	ΔE^{elec}	ΔE^{inter}	$\Delta\Delta G^{PB}$	$\Delta\Delta G^{SA}$	ΔG^{MMPBSA}	$T\Delta S$	ΔG^{calc}	ΔG^{corr}
I	131.61	-43.36	-11.52	-54.88	32.66	-4.59	-26.81	13.60	-13.21	-13.21
II	133.71	-45.11	-0.93	-46.04	37.73	-4.50	-12.81	13.70	0.89	2.99
III	133.44	-41.64	-9.09	-50.72	33.53	-4.72	-21.92	13.72	-8.20	-6.37
IV	132.71	-41.97	-4.28	-46.25	30.29	-4.73	-20.68	13.37	-7.31	-6.21
V	133.25	-41.98	-4.12	-46.11	27.95	-4.98	-23.14	15.24	-7.90	-6.26

^a All energies are in kcal/mol. The experimental binding free energy (-11.63 kcal/mol) is calculated using K_i (ref 10). ^b The corrected ΔG is calculated by adding the relative intramolecular energy of the ligand to the ΔG^{calc} .

rigid and flexible docking, respectively. If hydrogens are included in the ligand, the rigid docking failed to find the lowest energy orientation with the correct binding mode, which has a root-mean-square deviation of 11.25 Å to the crystal structure. On the other hand, the correct binding orientation, which has an root-mean-square deviation of 0.87 Å, is 2.2 kcal/mol less favorable according to the docking score function. However, for the flexible docking the lowest energetic conformation is very close to the crystal structure and the root-mean-square deviation is 1.24 Å. For 9Cl-TIBO, both the rigid and flexible docking successfully docked the ligand back to the binding site with a root-mean-square deviation of 0.25 and 1.40 Å, respectively. However, the energy score for the rigid docking is actually unfavorable. Apparently, some torsional flexibility in the ligand is more crucial if one includes hydrogens in docking these ligands into their crystal structures to achieve favorable docking scores.

It is necessary to point out that molecular docking may give very different lowest energy orientations, even for very similar compounds that should have the same binding mode. The reason lies in the fact that the present docking program does not consider the flexibility of the protein. Therefore, to not miss favorable orientations, both rigid and flexible docking were then performed for efavirenz and its five derivatives (compounds 4c, 4d, 4e, 4f, and 4m in ref 11). Finally, after cluster analysis, a total of five binding modes were recognized.

Five-hundred picosecond molecular dynamics simulations were then performed for all five binding modes. Although for most of modes the systems are well equilibrated after 100–

200 ps MD simulations, the 100 snapshots for further processing analysis were selected from only the last 200 ps of the trajectories. After equilibration, the MD simulations are very stable and the root-mean-square deviations of the main chain are smaller than 0.10 Å and those of the ligand are smaller than 0.40 Å. Figure 6 shows the positions and orientations of efavirenz after 500 ps simulation relative to the X-ray position of 8Cl-TIBO for the five binding modes. Overall these orientations and positions are significantly different from each other.

Table 6 lists the components of molecular mechanics and solvation energies. The binding mode I has the most favorable binding free energy, which is -13.21 kcal/mol, about 5 kcal/mol more negative than the second best binding mode (binding mode III, -8.20 kcal/mol). Moreover, the ligand in binding mode I has the lowest intramolecular energy, which is about 2 kcal/mol more negative than those of the other binding modes. For the five binding modes, the van der Waals energies are not very different, which means they all have good hydrophobic contacts. However, the electrostatic energies ($\Delta E^{elec} + \Delta\Delta G^{PB}$) are significantly different: 21.1, 36.8, 24.4, 26.0, and 23.8 kcal/mol for the five binding modes in sequence. We conclude that in this case the electrostatic term is the predominant factor for determining the different binding orientations. This term not only includes ligand–protein interactions but also both ligand and protein desolvation upon binding.

Figure 7 shows the root-mean-square deviation fluctuations of the main chain and the ligand along the simulation time for the first binding mode. Figure 8 shows the alignment of the last snapshot of the first binding mode (in yellow) and the 2.3

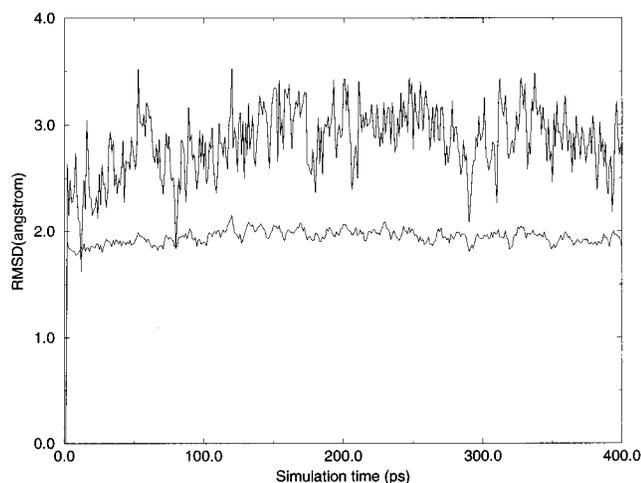


Figure 7. Root-mean-square deviations (Å) of efavirenz (upper line) and the α -carbons of HIV-1 RT (lower line, only the C_{α} of the flexible residues in MD simulations are considered) for 500 ps simulation compared with the docked-structure. The snapshots of the first 100 ps MD simulation, which is intended for equilibration, are omitted.

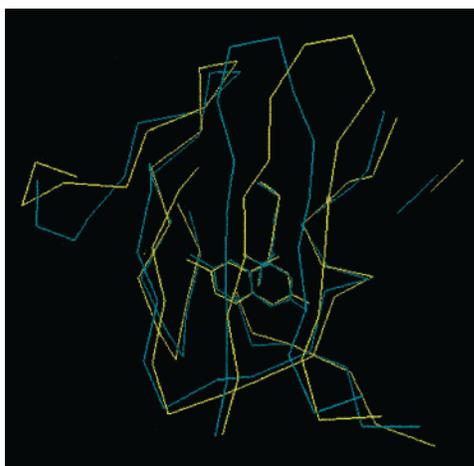


Figure 8. Alignment of the model structure (yellow, binding mode I) and the crystal structure (cyan) for HIV-1 RT/efavirenz. The least-squares fit was only performed for the C_{α} of residues 94–109, 226–240, 176–192 (loop), and 316–321 and the inhibitor. The root-mean-square deviation is 1.1 Å.

Å crystal structure (cyan) for HIV-1 RT/efavirenz. Least-squares fit was only performed for the C_{α} of residues 94–109, 226–240, 176–192 (loop), and 316–321 and the inhibitor. From this figure, one can see that the position and orientation of the inhibitor was well predicted and overall the root-mean-square deviation is only 1.1 Å. Figure 9 shows the last snapshot of the first binding mode. Several favorable interactions between the ligand and the enzyme are clearly shown in this figure: one hydrogen bond formed between the ligand and the receptor and another one formed between the ligand and water 2845; the favorable electrostatic interaction between LYS 101, LYS 103, and the nitrogen atoms in the ring; and the favorable hydrophobic interaction between PRO 95, PHE 227, LEU 234, VAL 106, TYR 188, and the hydrophobic cyclopropyl of the inhibitor.

Further Efficiencies in MM-PBSA Calculations. MM/PBSA is a newly developed free energy calculation method. Although it does not have as solid a theoretical basis as FEP and TI, it has some advantages compared to other less computationally demanding free energy prediction methods including LIE. The advantage of this method lies in the fact that it is more computationally efficient than FEP and TI and it

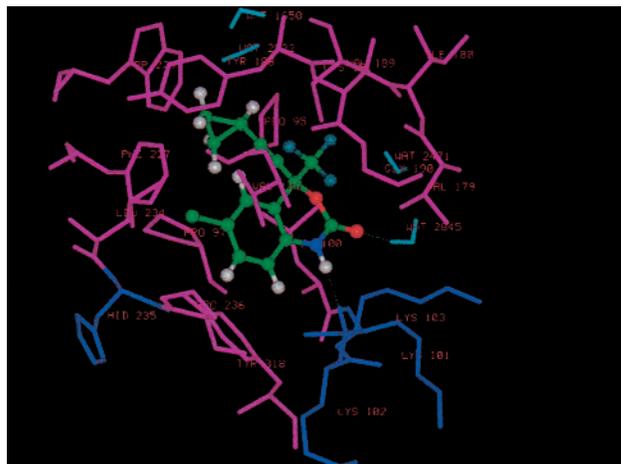


Figure 9. The final snapshot of the 500 ps MD simulation of HIV-1 RT/efavirenz (binding mode I). All residues (nonpolar amino acid in magenta, polar in blue, and water in cyan) within 4.0 Å from the mass center of efavirenz are shown. Hydrogen bonds formed by the ligand and the surroundings are indicated as green dotted lines.

does not need a training set to derive the empirical parameters as does LIE. Therefore, combining this method with molecular docking and molecular dynamics simulations, one can hope to reliably model a protein and DNA complex *a priori*.

As a new method of calculating the binding free energy, there are still things to be done to make MM-PBSA more efficient and reliable. There are several sources of error in MM-PBSA, which include the force field, MD sampling, and solvation free energy by PBSA and entropy estimated by normal-mode analysis. Moreover, it seems that the MM-PBSA approach does not work as well in calculating the absolute binding free energy for charged ligands as it does for neutral ones. We are now working on the following aspects to improve this method: (1) developing a new way to sample the conformational space for tens of compounds in only one simulation (Jed Pitera, Bernd Kuhn, and Shuanghong Huo, unpublished); (2) reparametrizing the atomic radii for PBSA calculation with an extended training set including all kinds of common organic compounds to reduce the errors from the solvation free energy calculations (Junmei Wang and Shuanghong Huo, unpublished); (3) developing a fast PB algorithm to allow the molecular dynamics simulations to be carried out rapidly in continuum solvent, which should also considerably speed up the MM-PBSA analysis (Ray Luo, unpublished); (4) improving the normal-mode analysis and making it more efficient and faster; and (5) improving the current force fields to calculate the molecular structure and energy more accurately. With increased computer power, we hope that an improved MM-PBSA approach can be used as a reliable and practical approach to predict the binding affinities for tens and even hundreds of compounds in a reasonable time.

Conclusions

Binding affinities of 12 TIBO-like inhibitors of HIV-1 RT have been predicted using a novel ansatz that combines MM-PBSA and normal mode calculations. The average unsigned error of the absolute binding free energy is about 1.0 kcal/mol, and a reasonable ranking order was also achieved. Although more work is still needed to improve this method, it is hopeful that MM-PBSA can be used as a fast and reliable method to predict and rank the binding affinities of hundreds of database hits and *de novo* designed compounds.

MM-PBSA, combined with normal mode calculation, has also been successfully applied in the prediction of the binding

geometry of HIV-1 RT/efavirenz, which was not known to us when this work was carried out. Among the five binding geometries suggested by docking, the binding free energy of the most favorable one (the correct one) is about 7 kcal/mol lower than that of the runner up. For the correct binding mode (mode I), after 100 to 200 ps MD equilibration, the MD simulation is very stable (root-mean-square deviations of main chain and ligand are only 0.06 and 0.31 Å, respectively), and the conformations sampled by MD simulation are very similar to the crystal one. For the last snapshot, the root-mean-square deviation of the 54 C α around the binding site is only 1.1 Å, which is particularly impressive given that the C α of the protein were initially \sim 2 Å root-mean-square deviation away from the correct structure (Figure 7). Thus, we have shown that MM-PBSA can be successfully used both to rank different ligands binding in a given site and to rank different binding modes of a single ligand.

In contrast to other binding free energy calculation methods such as the linear interaction energy (LIE) method, MM-PBSA contains no empirical parameters and, thus, it is more likely to be useful in determining the relative free energies of binding of quite different compounds and systems for which there is more limited experimental data, although a protein structure of the target is, of course, required. With some improvement in conformational sampling and solvation energy calculations, this method can be used to further rank tens and maybe even hundreds of potential compounds from docking database searching and *de novo* design in a reasonable length of computer and human time.

After this work had been submitted for publication, we learned of a related study using docking, Monte Carlo simulations to infer the binding modes of efavirenz to HIV-1 RT. That study used the known resistance profiles of efavirenz to validate their model.⁶²

Acknowledgment. P.A.K. is grateful to acknowledge research support from the NIH (GM-29072 and GM-56609 (E. Arnold, P. I.)). J.M.W. is grateful to acknowledge computer time at NCSA (MCB000013N (J. Wang, P. I.)), supported by the NSF. We are grateful to C. H. Chang, K. Hardman, X. Wen and L. Gonville at Dupont Pharmaceuticals for their work to solve the structure of HIV-1 RT bound to efavirenz.

JA003834Q

(59) Hsiou, Y.; Ding, J.; Das, K.; Clark, A. D., Jr.; Hughes, S. H.; Arnold, E. Structure of unliganded human immunodeficiency virus type 1 reverse transcriptase at 2.7 angstroms resolution: implications of conformational changes for polymerization and inhibition mechanisms. *Structure* **1996**, *4*, 853–860.

(60) Ding, J.; Das, K.; Hsiou, Y.; Sarafianos, S. G.; Clark, A. D., Jr.; Jacobo-molina, A.; Tantillo, C.; Hughes, S. H.; Arnold, E. Structure and functional implications of the polymerase active site region in a complex of HIV-1 RT with double-stranded DNA and antibody FAB fragment at 2.8 angstroms resolution. *J. Mol. Biol.* **1998**, *284*, 1095.

(61) Li, J.; Zhu, T.; Hawkins, G. D.; Winget, P.; Liotard, D. A.; Cramer, C. J. Extension of the platform of applicability of the SM5.42R universal solvation model. *Theor. Chem. Acc.* **1999**, *103*, 9–63.

(62) Rizzo, R. C.; Wang, D. P.; Tirado-Rives, J.; Jorgensen, W. L. Validation of a model for the complex of HIV-1 reverse transcriptase with sustiva through computation of resistance profiles. *J. Am. Chem. Soc.* **2000**, *122*, 12898–12900.