29. **Shultz, L. D., F. Ishikawa, and D. L. Greiner.** 2007. Humanized mice in translational biomedical research. Nat. Rev. Immunol. **7**:118–130.
30. **Traggiai, E., L. Chicha, L. Mazzucchelli, L. Bronz, J. C. Piffaretti, A. Lanzavecchia, and M. G. Manz.** 2004. Development of a human adaptive immune system in cord blood cell-transplanted mice. Science **304**:104–107.
31. **Walker, U. A., B. Setzer, and N. Venhoff.** 2002. Increased long-term mitochondrial toxicity in combinations of nucleoside analogue reverse-transcriptase inhibitors. AIDS **16**:2165–2173.
32. **Yahata, T., K. Ando, Y. Nakamura, Y. Ueyama, K. Shimamura, N. Tamaoki, S. Kato, and T. Hotta.** 2002. Functional human T lymphocyte development from cord blood CD34+ cells in nonobese diabetic/Shi-scid, IL-2 receptor gamma null mice. J. Immunol. **169**:204–209.

Prediction of Potency of Protease Inhibitors Using Free Energy Simulations with Polarizable Quantum Mechanics-Based Ligand Charges and a Hybrid Water Model

Debananda Das,[†] Yasuhiro Koh,[‡] Yasushi Tojo,[‡] Arun K. Ghosh,[§] and Hiroaki Mitsuya*[,†,‡]

Experimental Retrovirology Section, HIV and AIDS Malignancy Branch, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892-1868, Departments of Hematology and Infectious Diseases, Kumamoto University Graduate School of Medical and Pharmaceutical Sciences, Kumamoto 860-8556, Japan, and Departments of Chemistry and Medicinal Chemistry, Purdue University, West Lafayette, Indiana 47907

Received August 26, 2009

Reliable and robust prediction of the binding affinity for drug molecules continues to be a daunting challenge. We simulated the binding interactions and free energy of binding of nine protease inhibitors (PIs) with wild-type and various mutant proteases by performing GBSA simulations in which each PI's partial charge was determined by quantum mechanics (QM) and the partial charge accounts for the polarization induced by the protease environment. We employed a hybrid solvation model that retains selected explicit water molecules in the protein with surface-generalized Born (SGB) implicit solvent. We examined the correlation of the free energy with the antiviral potency of PIs with regard to amino acid substitutions in protease. The GBSA free energy thus simulated showed strong correlations ($r > 0.75$) with antiviral IC_{50} values of PIs when amino acid substitutions were present in the protease active site. We also simulated the binding free energy of PIs with P2-bis-tetrahydrofuranylurethane (bis-THF) or related cores, utilizing a bis-THF-containing protease crystal structure as a template. The free energy showed a strong correlation ($r = 0.93$) with experimentally determined anti-HIV-1 potency. The present data suggest that the presence of selected explicit water in protein and protein polarization-induced quantum charges for the inhibitor, compared to lack of explicit water and a static force-field-based charge model, can serve as an improved lead optimization tool and warrants further exploration.

INTRODUCTION

Virtual screening has been successful in the discovery of certain novel inhibitors, and a number of these inhibitors have advanced to clinical trials.[1] When the structure of a target protein is available, virtual screening involves docking potential inhibitors against the protein and ranking the inhibitors by their predicted affinity using a scoring function. Molecular mechanics Poisson–Boltzmann surface area (MM-PBSA) or molecular mechanics generalized Born surface area (MM-GBSA) have been used in some instances in the postprocessing and reranking of results from molecular docking.[2] Of note, docking and scoring have currently been an integral part of drug discovery efforts and produced documented successes; however, there is an urgent need for improvement of the accuracy of docking and scoring results.[3] With this regard, Clark described four areas of improvement, i.e., better scoring functions, treatment of protein flexibility, treatment of water molecules, and improved technology for data analysis of virtual screening results.[1] The scoring functions fail if they do not properly account for solvation, entropy, or polarizability.[1,4]

Water molecules form polar interactions with both proteins and ligands, fill empty spaces in cavities, and serve as an important component of molecular recognition. Lu et al. analyzed water molecules present at the interfaces of 392 X-ray crystal structures of protein–ligand complexes and reported high correlations between the polar van der Waals surface area of ligands and the number of ligand-bound water molecules in the crystal structures.[5] In some instances, as many as 21 water molecules are bound to a ligand, with the average being 4.6.[5] Despite their importance, the treatment of water molecules in docking calculations have not been widespread because of methodological limitations and poor understanding of how many and which water molecules are to be included in the simulation. By sampling multiple water positions during docking, Huang and Shoichet recently assessed the ligand enrichment against 24 targets.[6] Inclusion of water molecules increased enrichment against 12 targets while remaining largely unaffected for the others.[6] Fornabaio et al. reported that waters play a significant role in the energetics of binding and performed a hydropathic analysis of HIV-1 protease complexes.[7] They reported a significant improvement of the correlation between their HINT free energy scores and experimentally determined binding constants when appropriate bridging water molecules were taken into account.[7]

Most of the studies measure the accuracy of scoring functions by their ability to correctly rank the activity of a congeneric set of ligands. The prediction of activity of a ligand against mutant proteins is equally important in light of drug resistance in several diseases including acquired immune deficiency syndrome (AIDS) and cancers. In the

* Corresponding author address: 10 Center Drive, Room 5A11-MSC 1868, Bethesda, MD 20892-1868; phone: 301-496-9238; fax: 301-402-0709; e-mail, hmitsuya@helix.nih.gov.
[†] National Institutes of Health.
[‡] Kumamoto University Graduate School of Medical and Pharmaceutical Sciences.
[§] Purdue University.

10.1021/ci900320p CCC: $40.75
Published on Web 11/24/2009

present study, we focus on the resistance mutations of HIV-1 protease. HIV-1 protease acquires amino acid substitutions under the selection pressure of protease inhibitors (PIs), rendering HIV-1 resistant to such PIs.[8] For example, an Asp30Asn (D30N) substitution causes resistance against nelfinavir. Some amino acid substitutions, while being initially selected under drug pressure against one inhibitor, confer on HIV-1 cross-resistance against other inhibitors.[8] One example of such a substitution is M46I, which is a primary indinavir-resistance-associated substitution, but M46I-containing HIV-1 is resistant to other inhibitors such as ritonavir, nelfinavir, and atazanavir.[9] Analysis of the crystal structures of interactions of PIs with mutant proteases have shown that a number of drug resistance-associated mutations, such as G48V, V82A, and I84V, occur in the catalytically active site of protease.[10−12] Analyses of crystal structures of mutant proteases have revealed that there are, in general, no major conformational changes to the backbone conformation in such proteases, and the changes in binding interactions from the wild-type may involve different polar interactions with a mutant side chain(s) or loss of favorable van der Waals contacts.[13−15] Structural interactions, which are sometimes able to provide a rational explanation of the mechanism of resistance, are not able to predict a priori, for example, whether V82A causes a higher resistance for ritonavir compared to DRV. More reliable predictions of the potency of inhibitors against protease with drug-resistant mutations would be of use in the design of novel and more potent inhibitors.

The free energy of binding of ligands to proteins can be simulated by methods such as free energy perturbation and linear interaction energy (LIE) approximation.[16,17] LIE is a semiempirical method and based on a linear approximation of polar and nonpolar free energy contributions from molecular dynamics simulation averages.[16] The LIE method has recently been used in calculating the binding free energy of *N*-sulphonyl-glutamic acid inhibitors to MurD ligase and in probing the DNA replication fidelity.[18,19] In the current study we simulated the binding free energies of nine protease inhibitors against wild-type (PRO^{WT}) and mutant proteases (PRO^{MT}) with standard and hybrid GBSA protocols. While a number of water molecules are present in the X-ray crystal structures of protease−inhibitor complexes, a water molecule that mediates hydrogen-bond interactions of the protease inhibitors with Ile50 and Ile50′ in the flap is common across several different inhibitor−protease complexes and present in the complexes for eight FDA-approved PIs. In this work, we explicitly incorporated water molecule bridging hydrogen bonds with the protease flap. For inhibitors nelfinavir and atazanavir, two additional water molecules that mediate hydrogen bonds between these inhibitors and other protease residues were also explicitly included. We compared the GBSA free energy of binding obtained from simulations with selected explicit water molecules in implicit solvation (a hybrid solvation model) with free energies that did not have the water molecule explicitly present. Furthermore, in the simulations, the inhibitor atoms had either force-field-derived fixed partial charges or quantum mechanics-based partial charges that accounted for the polarization induced by the surrounding protein environment (a hybrid charge model). We also analyzed the correlation of the GBSA free energies obtained by the simulations with antiviral potency data (IC_{50} values). Our current data suggest that selective inclusion of explicit water molecule(s) and protein polarization effects may improve the robustness of GBSA free energy simulations and aid the design of inhibitors that are potent against both wild-type and multidrug-resistant HIV-1 variants.

METHODS

Crystal Structures Used as Starting Templates. We explored various wild-type protease crystal structures from the Protein Data Bank as starting templates for docking and subsequent free energy simulations. For convenience of protein expression and crystallization, some of the structures deposited in the Protein Data Bank as wild-type structures have several mutations such as Q7K, K14R, R41K, L63P, and I64V that are distant from the inhibitor binding site.[15,20,21] These nonactive site mutations may not drastically alter the conformation of the protease and its interactions with inhibitors compared to a pristine wild-type protease of $HIV\text{-}1_{LAI}$ or $HIV\text{-}1_{NL4\text{-}3}$. While one nonactive site mutant, depending on the residue and location, may not necessarily affect the binding affinity comparisons, crystal structures with four or five nonactive site mutations are unsuitable to be used for free energy simulations, especially when comparing the simulation data with antiviral potency against wild-type HIV-1. We used 2FDE, obtained from the Protein Data Bank, as the starting template for docking against darunavir (DRV), amprenavir (APV), GRL-98065, GRL-02031, and GRL-06579. 2FDE is a cocrystal of brecanavir and $HIV\text{-}1_{LAI}$ wild-type protease, and brecanavir (BCV) has a bis-THF ligand as a core.[22] The PDB IDs of the crystal structures used for our simulations of the other inhibitors are as follows: 1HXB[23] for saquinavir (SQV); 2O4P[24] for TPV; 1OHR[25] for nelfinavir (NFV); 1MUI[26] for lopinavir (LPV), and 2AQU[27] for atazanavir (AZV). Waters were not modeled in the crystal structure of LPV[26] but were present in all other structures.

Our goal was to explore the prediction of free energy of binding once a correct binding mode was obtained. In the present study, we demonstrate that the correct binding mode was reliably obtained when a ligand was docked against a protease structure obtained with a similar core. To decrease uncertainty arising due to cross docking of ligands to different proteases, we docked ligands against the native protease crystal or against a protease structure obtained with a similar core. It is important to keep in mind that protease side chains may undergo subtle conformational changes to accommodate protease inhibitors of different shapes and sizes (the molecular weights of the PIs in the current study range from 506 to 705), and these changes might be difficult to capture by simple minimization following ligand docking to non-native crystal structures.

Docking. The interactions of protease inhibitors with wild-type HIV-1 protease were examined using computational structural modeling and molecular docking. Besides accounting for the conformational flexibility of the inhibitor, the polarization induced in the inhibitor by the protease was taken into consideration by employing polarizable quantum charges in the docking computations. The use of polarizable quantum charges has recently been shown to substantially improve the prediction of protein−ligand complex structures.[28] The QM-polarized ligand docking protocol utilizing Glide version 4.5, QSite version 4.5, Jaguar version 7.0, and

Maestro version 8.5 (Schrödinger, LLC, New York, NY 2007) was used as described below. The crystal coordinates described above were obtained from the Protein Data Bank (http://www.rcsb.org/) and used as starting templates. Hydrogens were optimized with constraints on the heavy atoms. The crystal water that mediates the interactions between protease inhibitors and the protease flap was retained, and all other crystal waters were deleted. Close interactions in the protease were annealed, and the docking grid was setup. Polarizable ligand charges were determined at the B3LYP/6-31G* level. The extra-precision mode of Glide,[29,30] which has a higher penalty for unphysical interactions, was used. For each docking simulation, up to five final poses were retained and compared with available X-ray structures to verify that the conformations were reasonable. It was particularly important that the correct ring conformations were obtained during docking. LPV produced ring conformations that were different than the conformations obtained from crystal complexes in some docking solutions, and such conformations were discarded for the subsequent GBSA simulations.

GBSA Scoring Simulations. The general principle of a GBSA model has been described before. The free energy of binding, ΔG_{bind} is calculated as[2]

$$\Delta G_{bind} = \Delta E + \Delta G_{solv} + \Delta G_{SA}$$

$$\Delta E = E_{complex} - E_{protein} - E_{ligand}$$

where $E_{complex}$, $E_{protein}$, and E_{ligand} are the minimized energies of the protease−inhibitor complex, protease, and inhibitor, respectively.

$$\Delta G_{solv} = G_{solv(complex)} - G_{solv(protein)} - G_{solv(ligand)}$$

where $G_{solv(complex)}$, $G_{solv(protein)}$, and $G_{solv(ligand)}$ are the solvation free energies of the complex, protein, and inhibitor, respectively.

$$\Delta G_{SA} = G_{SA(complex)} - G_{SA(protein)} - G_{SA(ligand)}$$

where $G_{SA(complex)}$, $G_{SA(protein)}$, and $G_{SA(ligand)}$ are the surface area energies for the complex, protease, and inhibitor, respectively. The simulations were carried out using the GBSA continuum model[31] in Prime, version 2.0 (Schrödinger, LLC, New York, NY, 2008). Prime uses a surface-generalized Born (SGB) model employing a Gaussian surface instead of a van der Waals surface for better representation of the solvent-accessible surface area.[31]

GBSA simulations were carried out for the protease−ligand complex structures obtained by molecular docking. The simulations were carried out in four different scenarios. (i) No water molecules were retained in the protease and ligand atoms have fixed charges based on the OPLS force field. This is the standard MM-GBSA simulation carried out in implicit solvation. The change in free energy obtained is denoted by ΔG_{mm}, and the correlation coefficients are denoted by r_{mm}. (ii) No water molecules were retained in the protease and the ligand has protein-polarized QM charges at the B3LYP/6-31G* level. The protein-polarized charge on the ligand is determined from the docked complex and used while computing $E_{complex}$, $G_{solv(complex)}$, and $G_{SA(complex)}$ as well as for E_{ligand}, $G_{solv(ligand)}$, and $G_{SA(ligand)}$. The change in free energy obtained is denoted by ΔG_{qm}, and the correlation coefficients are denoted by r_{qm}. (iii) The bridging water molecule mediating the hydrogen-bond interactions of inhibitors DRV, GRL-98065, APV, GRL-02031, GRL-06579, NFV, SQV, and AZV with Ile50 and Ile50′ in the flap was explicitly retained. This is a hybrid solvation model since implicit GBSA solvation terms for the whole system were used. For tipranavir (TPV), GBSA with the hybrid solvation model was performed by retaining a water molecule that bridges hydrogen-bond interactions with Gly48 of one monomer of the protease. NFV and AZV were observed to have two additional bridging water molecules, and additional calculations in the presence of three explicit water molecules were performed for NFV and AZV. In the hybrid solvation model, the inhibitors either had MM charges (change in free energy and correlation coefficient denoted by $\Delta G_{mm/wat}$ and $r_{mm/wat}$, respectively) or (iv) protein-polarized QM (B3LYP/6-31G*) charges (change in free energy and correlation coefficient denoted by $\Delta G_{qm/wat}$ and $r_{qm/wat}$, respectively). In all simulations, the protease has OPLS charges. The strain energies of the ligands were taken into account.

Antiviral Agents. DRV, GRL-98065, and GRL-02031 were synthesized as described previously.[32−35] SQV and ritonavir (RTV) were kindly provided by Roche Products Ltd. (Welwyn Garden City, United Kingdom) and Abbott Laboratories (Abbott Park, IL), respectively. APV was a kind gift from Glaxo-Wellcome, Research Triangle Park, NC. NFV and indinavir (IDV) were kindly provided by Japan Energy Inc., Tokyo, Japan. LPV was synthesized by previously published methods.[36] AZV was a kind gift from Bristol-Myers Squibb (New York, NY). TPV was obtained through the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, National Institutes of Health.

Generation of Recombinant HIV-1 Clones. To generate HIV-1 clones carrying desired mutations, site-directed mutagenesis using the QuikChange Site-Directed Mutagenesis Kit (Stratagene, La Jolla, CA) was performed, and the mutation-containing genomic fragments were introduced to pHIV-1_{NLSma}, as previously described.[35,37] Determination of the nucleotide sequences of plasmids confirmed that each clone had the desired mutations but no unintended mutations. Each recombinant plasmid was transfected into 293T cells using Lipofectoamine 2000 Transfection Reagent (Invitrogen, Carlsbad, CA), and thus generated infectious virions were harvested 48 h after transfection and stored at −80 °C until use. HIV-1 carrying D30N substitution (HIV^{D30N}) was generated since residue-30 is in the active site (Figure 1) and the D30N substitution is known to cause primary drug resistance against the FDA-approved protease inhibitor NFV.[38] HIV^{I50V} was generated since Ile50 is in the flap region (Figure 1) and interacts with various PIs through a bridging water molecule.[39] The I50V mutation has been associated with resistance to APV, LPV, and RTV.[40] $HIV^{V82I/I85V}$ was also generated since Val82 is located in the active site and its substitution to Ile is associated with HIV-1 resistance to most PIs, presumably due to the expansion of the active site and loss of favorable van der Waals contact.[15,20,40,41] We recently reported the emergence of I85V as a resistant mutation against a PI, GRL-98065, and chose to study the combined effect of V82A/I85V.[32] HIV-1-infected patients who failed to respond to PI-containing regimens often have HIV-1 variants carrying both active site

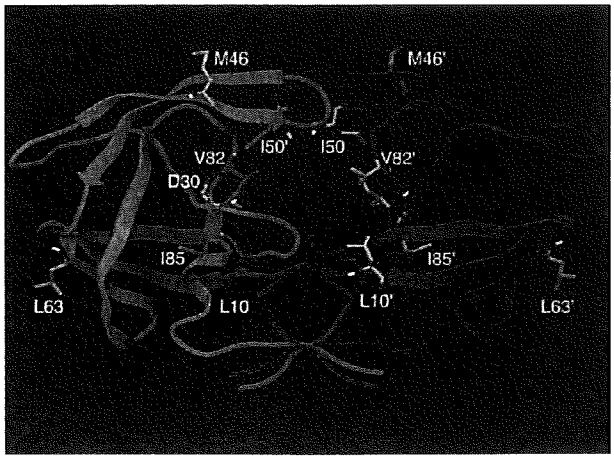


Figure 1. Structure of dimerized HIV-1 protease. The monomer subunit is shown in a red or green ribbon. The locations of the mutant residue positions are indicated. Only polar hydrogens are shown, and the following atom colors are used in this and all subsequent figures: C, gray; H, white; O, red; N, blue; S, yellow.

and nonactive site mutations in protease, and we chose to explore such clinical HIV-1 isolates, HIV^{2840} and HIV^{2841}. The former contained L10R, M46I, L63P, V82T, and I84V substitutions, while the latter contained M46I, L63P, V82T, and I84V substitutions.

Drug Susceptibility Assay. To determine the drug susceptibilities of certain laboratory HIV-1 strains, MT-4 cells were employed as target cells, as described previously,[37] with minor modifications. In brief, MT-4 cells (10^5/mL) were exposed to 100 $TCID_{50}$s of drug-resistant HIV-1 strains in the presence or absence of various concentrations of drugs and incubated at 37 °C. On day 7 of culture, the supernatants were harvested and the amounts of the p24 Gag protein were determined by using a fully automated chemiluminescent enzyme immunoassay system (Lumipulse F; Fujirebio Inc., Tokyo, Japan).[42] The drug concentrations that suppressed the production of p24 Gag protein by 50% (IC_{50}) were determined by comparison of the amount of p24 Gag protein produced in drug-treated cell cultures with the level of p24 Gag protein produced in a drug-free control cell culture. All assays were performed in duplicate or triplicate on more than three different occasions, and the data are shown as means ± 1 SD.

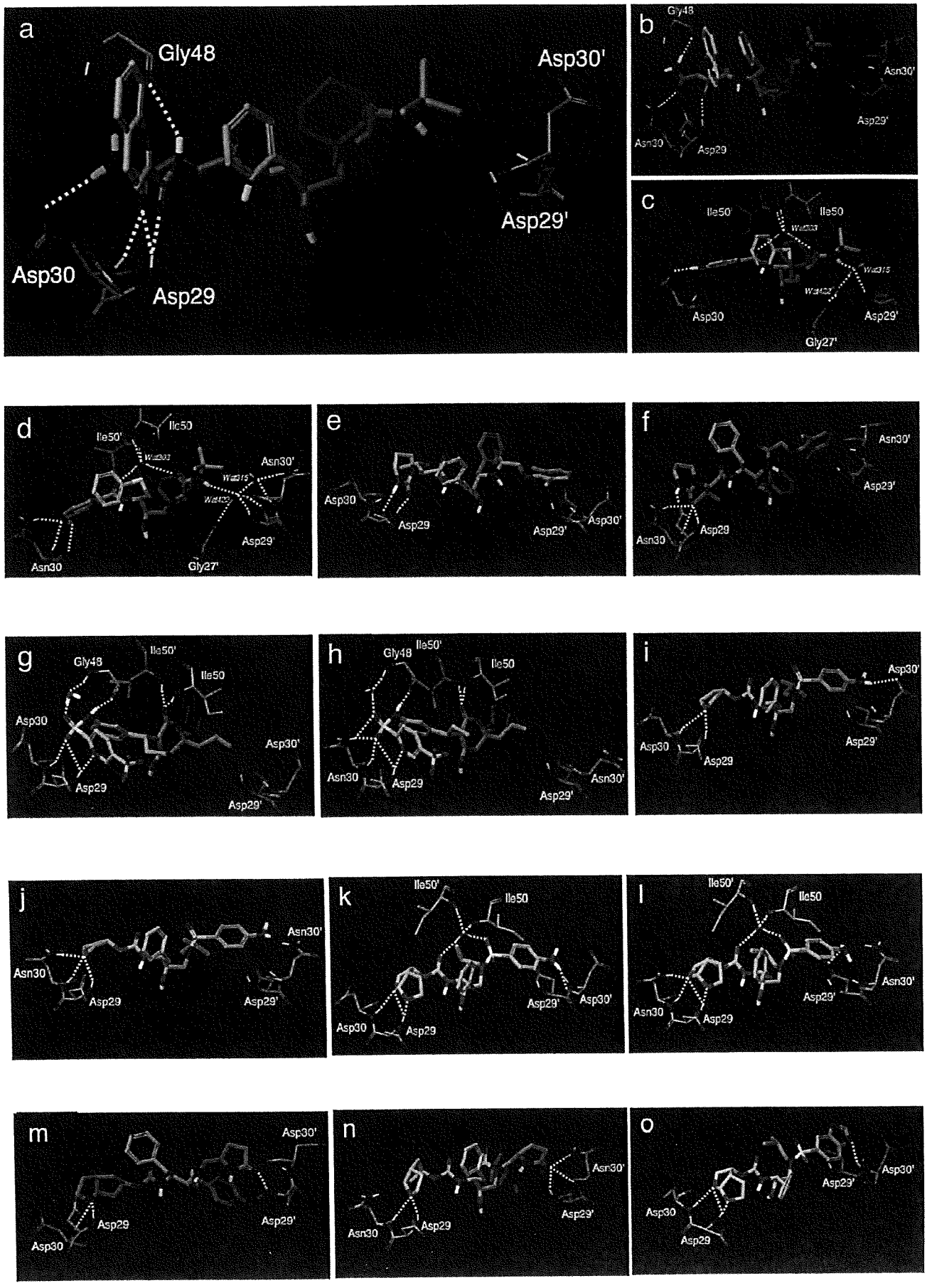
RESULTS

Analysis of Water Molecules in Protease Crystal Structures. Since the number of ligand-bound water molecules in high-resolution crystal structures varies widely,[5] we first analyzed the number of water molecules in the five protease−inhibitor crystal structures used in the current study. The total number of water molecules ranged from 51 (protease−NFV complex,[25] PDB ID 1OHR) to 124 (protease−TPV complex,[24] PDB ID 2O4P). Within 4 Å of the bound ligand, the number of water molecules present in the complexes with PDB IDs 1HXB, 2FDE, 1OHR, 2O4P, and 2AQU were 3, 5, 5, 10, and 10, respectively. We then analyzed the number of water molecules bridging hydrogen-bond interactions between the protease and the ligands. The water molecule forming tetracoordinated hydrogen-bond interactions with Ile50 and Ile50′ in the flap was the only water molecule interacting with the ligand in the protease complexes of SQV (PDB ID 1HXB), BCV (PDB ID 2FDE), and APV (PDB ID 1HPV[43]). Since this tetracoordinated water molecule is considered an important pharmacophore for protease−inhibitor interactions,[44] it was explicitly included in our docking and GBSA free energy simulations. Three water molecules formed bridging hydrogen-bond interactions in the NFV−protease (PDB ID 1OHR) and AZV−protease (PDB ID 2AQU) complexes (Figure 2c and 2q), and simulations were carried out with both one and three bridging water molecules explicitly present for these inhibitors. TPV directly formed hydrogen bonding with Ile50 and Ile50′, and a water molecule bridged hydrogen bonding with Gly48 in the flap (Figure 2g)[24,45,46] and was explicitly included in the simulations. Simulations involving LPV did not include any crystal waters since none were present in the native LPV−protease complex.[26]

Determination of the Susceptibility of Recombinant Infectious HIV-1 Clones Carrying Amino Acid Substitutions to PIs in Vitro. We determined the susceptibility of six recombinant infectious HIV-1 clones to each of the PIs that were chosen in the present study: 7 clinically available PIs (SQV, NFV, APV, TPV, DRV, and AZV) and 2 experimental PIs (GRL-02031 and GRL-98065) in the HIV-1 p24 Gag production inhibition assay, as previously described.[35,37] As illustrated in Table 1, most of the recombinant clones showed reduced susceptibility to the PIs examined by up to 16.5-fold. However, it was also noted that HIV^{D30N} and HIV^{2841} had increased susceptibility to certain PIs. The increased susceptibility of HIV_{D30N} to TPV with 33.3-fold was notable, although HIV^{D30N} was also less susceptible to SQV, LPV, and NFV (Table 1). Both HIV^{I50V} and HIV^{2840} were also less susceptible to most of the PIs (Table 1).

Binding Interactions with Wild-Type and D30N Mutant Protease. We next determined and analyzed the binding modes of nine different PIs with PRO^{WT} and a protease with an amino acid substitution at position 30 from an aspartic acid to asparagine (PRO^{D30N}). SQV has four hydrogen-bond interactions with Asp29 and Asp30 in the S2 site of the wild-type protease but has no hydrogen bonds with Asp29′ or Asp30′ in the S2′ site (Figure 2a). When protease acquires the D30N mutation, SQV loses two hydrogen bonds with Asp29 and Asp30 and does not form any new and compensating hydrogen bonds with other protease residues (Figure 2b). Comparison of the antiviral data of SQV shown in Table 1 indicates that there was a 3.9-fold decrease in antiviral potency with the D30N mutation. It is possible that the decrease of antiviral potency of SQV for D30N mutant is due to the loss of hydrogen bonds with residues 29 and 30 for the mutant. Examining the hydrogen bonds in the S2 site for NFV against PRO^{WT} and PRO^{D30N} mutant protease (Figure 2c and 2d), one observes that NFV has more hydrogen bonds with Asn30 of PRO^{D30N} compared to Asp30 of PRO^{WT}. An X-ray crystal structure has also demonstrated that NFV has a larger number of hydrogen bonds with PRO^{D30N} than with PRO^{WT}.[11]

However, D30N is a major amino acid substitution[38] associated with NFV resistance of HIV-1, and HIV^{D30N} showed a 5.3-fold increase in IC_{50} values in antiviral assays


a
Gly48
Asp30'
Asp29'
Asp30
Asp29
b
Gly48
Asn30'
Asp29'
Asn30
Asp29
c
Ile50'
Ile50
Asp30
Asp29'
Gly27'
d
Ile50'
Ile50
Asn30'
Asp29'
Asn30
Gly27'
e
Asp30
Asp29
Asp29'
Asp30'
f
Asn30'
Asp29'
Asn30
Asp29
g
Gly48
Ile50'
Ile50
Asp30
Asp30'
Asp29
Asp29'
h
Gly48
Ile50'
Ile50
Asn30
Asp29
Asn30'
Asp29'
i
Asp30'
Asp30
Asp29'
Asp29
j
Asn30'
Asn30
Asp29
Asp29'
k
Ile50'
Ile50
Asp29'
Asp30'
Asp30
Asp29
l
Ile50'
Ile50
Asn30'
Asp29'
Asn30
Asp29
m
Asp30'
Asp29
Asp29'
Asp30
n
Asn30'
Asp29'
Asn30
Asp29
o
Asp30'
Asp29'
Asp30
Asp29

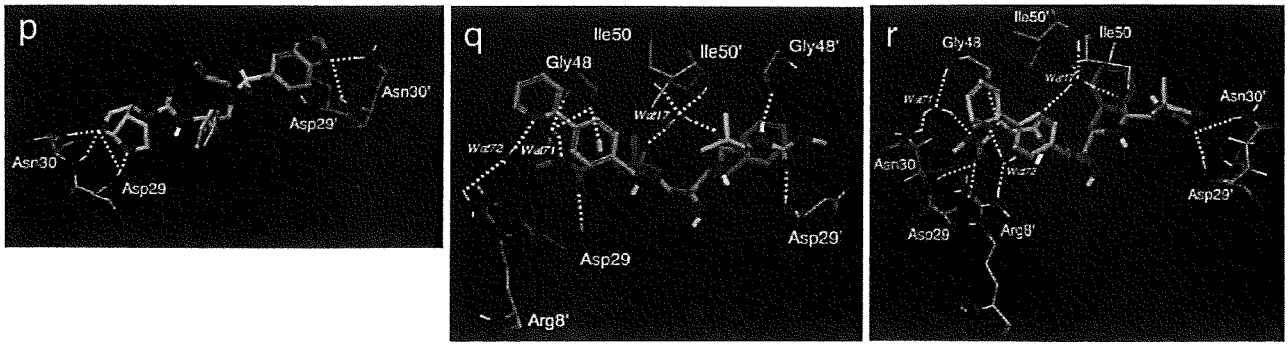


Figure 2. Interactions of protease inhibitors with wild-type and D30N mutant protease. Interactions of SQV with (a) wild-type (PRO^{WT}) and (b) D30N-containing mutant protease (PRO^{D30N}) are shown. SQV has hydrogen-bond interactions (dotted yellow lines) only in the S2 site but no such interactions in the S2′ site. The hydrogen bonds of SQV with the mediating water molecule, Asp25, Asp25′, and Gly27, are present but not shown. Hydrogen-bond interactions of NFV with (c) PRO^{WT} and (d) PRO^{D30N}. The bridging water-303 has hydrogen bonds with Ile50 and Ile50′, water-315 hydrogen bonds with Asp29′, and water-422 hydrogen bonds with Gly27′. Simulations were carried out for NFV with these three waters molecules explicitly present, with one of the bridging water molecules and without any explicit water molecule. Nefinavir has more hydrogen bonds with residues 29 and 30 in PRO^{D30N} even though it has a lower antiviral potency against $HIV\text{-}1^{D30N}$. Hydrogen bonds of LPV with (e) PRO^{WT} and (f) PRO^{D30N}. LPV has more hydrogen-bond interactions with the S2 site for the mutant than it has for the PRO^{WT}. The crystal structure for LPV 1MUI (PDB ID) does not have waters. Hydrogen-bond interaction of TPV with (g) PRO^{WT} and (h) PRO^{D30N}. Unlike other inhibitors in this study, TPV directly hydrogen bonds with Ile50 and Ile50′ but a bridging water molecule hydrogen bonds with Gly48 in the flap. Interactions of APV with (i) PRO^{WT} and (j) PRO^{D30N} are shown. In the S2 site of the protease, the THF ligand of APV has an additional hydrogen bond with Asn30 (PRO^{MT}) compared to Asp30 (PRO^{WT}). In the S2′ site, the aniline nitrogen of APV has a hydrogen bond with the side chain of Asp30 (PRO^{WT}) and the hydrogen bond is lost for the Asn30 mutant. Structural interaction of DRV with (k) PRO^{WT} and (l) PRO^{D30N} are shown. The bis-THF ligand of DRV has more hydrogen-bond interactions with PRO^{D30N} compared to the PRO^{WT}. The water molecule mediating hydrogen bonds between the inhibitor and Ile50 and Ile50′ in the flap is shown. The water molecule is present in docking and free energy simulations of APV, GRL-98065, GRL-02031, and SQV but is not shown. Hydrogen-bond interactions of GRL-02031 with (m) PRO^{WT} and (n) PRO^{D30N} are shown. (o) Hydrogen-bond interactions of GRL-98065 with wild-type protease are shown. The bis-THF ligand of GRL-98065 has hydrogen-bond interactions with backbone atoms of Asp29 and Asp30 in the S2 site and with Asp30′ in S2′. (p) Besides maintaining the backbone hydrogen bonds with Asp29 and residue-30 and -30′ of PRO^{D30N}, GRL-98065 forms additional hydrogen bonds with the side chains of Asn30 and Asn30′. Hydrogen-bond interactions of AZV with (q) PRO^{WT} and (r) PRO^{D30N}.

Table 1. Antiviral IC_{50} Values (μM) of Protease Inhibitors against Wild-Type and Mutant Protease

inhibitor	HIV^{WT} IC_{50} (μM)	HIV^{D30N} IC_{50} (μM)	HIV^{I50V} IC_{50} (μM)	$HIV^{V82I/I85V}$ IC_{50} (μM)	HIV^{2840} IC_{50} (μM)	HIV^{2841} IC_{50} (μM)
saquinavir	0.008 ± 0.003[a]	0.031 ± 0.005 (3.9)[b]	0.034 ± 0.001 (3.9)	0.023 ± 0.01 (2.9)	0.015 ± 0.009 (1.9)	0.004 ± 0.0004 (0.5)
nelfinavir	0.018 ± 0.002	0.096 ± 0.005 (5.3)	0.11 ± 0.08 (6.1)	0.083 ± 0.071 (4.6)	0.22 ± 0.05 (12.4)	0.031 ± 0.005 (1.7)
lopinavir	0.018 ± 0.001	0.055 ± 0.011 (3.1)	0.21 ± 0.14 (11.7)	0.037 ± 0.007 (2.1)	0.30 ± 0.009 (16.5)	0.048 ± 0.009 (2.7)
amprenavir	0.028 ± 0.007	0.0066 ± 0.0049 (0.2)	0.16 ± 0.13 (5.7)	0.058 ± 0.024 (2.1)	0.083 ± 0.018 (3.0)	0.038 ± 0.003 (1.4)
tipranavir	0.15 ± 0.05	0.0047 ± 0.0002 (0.03)	0.34 ± 0.03 (2.3)	0.28 ± 0.003 (1.9)	0.32 ± 0.01 (2.2)	0.12 ± 0.05 (0.8)
darunavir	0.003 ± 0.0007	0.0015 ± 0.0013 (0.5)	0.014 ± 0.012 (4.7)	0.0057 ± 0.0024 (1.9)	0.0033 ± 0.0001 (1.1)	0.0018 ± 0.0008 (0.6)
GRL-02031	0.02 ± 0.008	0.0087 ± 0.0067 (0.4)	0.014 ± 0.001 (0.7)	0.036 ± 0.002 (1.8)	0.033 ± 0.0009 (1.7)	0.012 ± 0.001 (0.6)
GRL-98065	0.0003 ± 0.00006	0.00055 ± 0.00035 (1.8)	0.0018 ± 0.001 (6.0)	0.0018 ± 0.0007 (6.0)	0.0004 ± 0.0001 (1.3)	0.00015 ± 0.00007 (0.5)
atazanavir	0.004 ± 0.0004	0.0016 ± 0.0008 (0.4)	0.0019 ± 0.001 (0.5)	0.0015 ± 0.0001 (0.4)	0.0086 ± 0.0041 (2.2)	0.003 ± 0.001 (0.8)

[a] The IC_{50} values shown represent means ± 1 SD of assays conducted in duplicate or triplicate on more than three different occasions for the study. [b] The fold change in activity of the mutants from the wild-type IC_{50} are shown within brackets.

(Table 1). Examining the structural interactions of LPV and TPV against PRO^{WT} and PRO^{D30N} mutants (Figure 2e–h), we observed that both inhibitors have more hydrogen bonds with residues 29 and 30 of PRO^{D30N} than they have for those of PRO^{WT}. Antiviral data show that while LPV has a 3.1-fold increase in IC_{50}, HIV^{D30N} was about 30 times more sensitive to TPV (Table 1). Comparison of the structural interactions of DRV, TPV, GRL-02031, and GRL-98065 with PRO^{WT} and PRO^{D30N} (Figure 2k–p) revealed that all of them have more hydrogen-bond interactions with PRO^{D30N}. APV, which is 5-fold more potent against HIV^{D30N} (Table 1), has three hydrogen bonds with Asp29, Asp30, and Asp30′ for PRO^{WT} and forms three hydrogen bonds with Asp29 and Asn30 for PRO^{D30N} in structural models (Figure 2i and 2j). Thus, it is clear that while the number of hydrogen bonds may provide an intuitive understanding of binding and/or antiviral potency, it may not always explain why certain PIs show a decrease in antiviral potency with D30N substitution while other PIs show an increase.

Free Energy Changes in Complexes with the D30N Mutation Determined by GBSA Simulations. Since the number of hydrogen bonds between PIs and protease does not always help predict the potency of PIs as discussed above, we examined the free energies under four different simulation conditions: with and without explicit water(s) and with QM or with MM charges on the inhibitor (Table 2, Supporting Information Tables S1 and S2). It was assumed that an increase in the change of free energy of binding ($\Delta\Delta G$ more positive) is to be expected for a decrease in antiviral activity and vice versa. With the D30N mutation in PRO^{D30N}, SQV showed a reduction in antiviral activity by 3.9-fold. With the bridging water and QM charges on SQV, the free energy change ($\Delta\Delta G_{qm/wat}$) of the SQV–protease complex increased by 4 kcal/mol for the D30N mutation (Table 2). HIV^{D30N}

Table 2. Change in Free Energy (kcal/mol) Associated with D30N Mutation[a]

inhibitor	$\Delta\Delta G_{qm/wat}$ (kcal/mol)	$\Delta\Delta G_{mm/wat}$ (kcal/mol)	$\Delta\Delta G_{qm}$ (kcal/mol)	$\Delta\Delta G_{mm}$ (kcal/mol)	IC_{50} fold change[b]
saquinavir	+4	+4	−1	−2	3.9
nelfinavir	+7	+9	−11	−9	5.3
lopinavir	n.d.	n.d.	+17	+21	3.1
amprenavir	−12	+12	+11	+7	0.2
tipranavir	−5	−2	−12	−9	0.03

[a] The qm subscript refers to the use of protein-polarized QM charges on the inhibitor, and the mm subscript refers to the use of fixed force-field-based charges. The use of an explicit bridging water is shown by the wat subscript. A positive $\Delta\Delta G$ indicates a decrease in binding affinity, while a negative $\Delta\Delta G$ indicates an increase. [b] IC_{50} fold change greater than 1 indicates a reduction in potency with D30N mutation, while a fold change of less than 1 indicates sensitivity.

Table 3. Correlation Coefficient (r) of the Free Energies of Binding versus Experimental IC_{50} Data[a]

	Set-1 (PRO^{WT},PRO^{D30N},PRO^{I50V})				Set-2 (PRO^{WT},PRO^{D30N},PRO^{I50V},$PRO^{V82I/I85V}$)				Set-3 (PRO^{WT} and all mutant proteases)			
	r_{mm}	r_{qm}	$r_{mm/wat}$	$r_{qm/wat}$	r_{mm}	r_{qm}	$r_{mm/wat}$	$r_{qm/wat}$	r_{mm}	r_{qm}	$r_{mm/wat}$	$r_{qm/wat}$
DRV	0.30	**0.91**	**0.93**	**0.93**	0.14	**0.65**	**0.53**	**0.81**	0.29	0.23	0.46	**0.83**
98065	0.35	0.18	**0.99**	**0.94**	0.31	0.40	0.32	**0.66**	**0.53**	0.17	0.46	0.48
APV	0.26	**0.66**	**0.57**	**0.99**	0.34	0.23	0.47	**0.88**	0.31	0.19	0.29	**0.60**
02031	0.16	**0.78**	0.24	**0.78**	0.46	0.48	0.21	0.10	0.61	**0.62**	0.02	0.19
TPV	**0.96**	**0.70**	0.49	**0.83**	**0.91**	**0.71**	0.01	**0.68**	0.14	0.04	0.04	**0.50**
NFV	0.21	**0.51**	**0.83**	**0.77**	0.26	0.35	**0.71**	**0.59**	0.11	0.31	0.41	0.39
NFV	0.21	**0.51**	**0.62**[b]	**0.97**[b]	0.26	0.35	0.48[b]	**0.92**[b]	0.11	0.31	0.18[b]	**0.58**[b]
LPV	**0.76**	**0.97**	n.d.	n.d.	0.42	**0.59**	n.d.	n.d.	0.16	**0.66**	n.d.	n.d.
SQV	0.06	0.45	**0.96**	**0.99**	0.11	0.43	**0.94**	**0.87**	**0.78**	0.22	**0.57**	**0.84**
AZV	**0.76**	**0.54**	0.06	0.16	0.13	0.14	0.35	0.16	0.24	0.17	**0.66**	**0.59**
AZV	**0.76**	**0.54**	**0.91**[b]	**0.64**[b]	0.13	0.14	0.34[b]	**0.53**[b]	0.24	0.17	**0.62**[b]	**0.58**[b]

[a] The qm subscript refers to the use of protein-polarized QM charges on the inhibitor, and the mm subscript refers to the use of fixed force-field-based charges. The use of an explicit bridging water is shown by the wat subscript. Bold font indicates correlation ($r > 0.5$). [b] Calculated with three water molecules that mediate hydrogen bonds between the protease and the inhibitor.

was resistant to NFV by 5.3-fold compared to HIV^{WT}, and $\Delta\Delta G_{qm/wat}$ showed an increase in the free energy of binding by +7 kcal/mol. HIV-1 containing PRO^{D30N} was more sensitive to APV and TPV, and $\Delta\Delta G_{qm/wat}$ for PRO^{D30N} with APV and TPV was −12 and −5 kcal/mol, respectively (Table 2). The $\Delta\Delta G_{qm/wat}$ values showed that both TPV and APV had a higher affinity for PRO^{D30N} than for PRO^{WT} (Table 2) and correlated with the increase in sensitivity to these inhibitors.

We next compared the free energy changes ($\Delta\Delta G_{mm/wat}$) simulated using the bridging water molecule but with force-field-based fixed MM charges on the inhibitors. The $\Delta\Delta G_{mm/wat}$ values of +4 and +9 kcal/mol for SQV and NFV, respectively, showed that the simulation results in an increase in the free energy of binding correlating with the reduction of antiviral activity. However, the $\Delta\Delta G_{mm/wat}$ value showed a wrong trend for APV (+12 kcal/mol) and was alike for TPV (−2 kcal/mol). The $\Delta\Delta G_{qm}$ and $\Delta\Delta G_{mm}$ values were negative for SQV and NFV, respectively, while they were positive for APV. Thus, $\Delta\Delta G_{qm}$ and $\Delta\Delta G_{mm}$ values, which did not incorporate the bridging water molecule explicitly, simulated inaccurate changes in the free energy of binding for SQV, NFV, and APV. The crystal structure for LPV (PDB ID 1MUI) did not have any water molecules present, and all simulations involving LPV were carried out with implicit water. The $\Delta\Delta G_{qm}$ and $\Delta\Delta G_{mm}$ values for LPV were +17 and +21 kcal/mol, respectively. The increase in the change in the free energy of binding of LPV with PRO^{D30N} was consistent with its decrease in antiviral potency with D30N substitution. For TPV, the negative $\Delta\Delta G_{qm}$ and $\Delta\Delta G_{mm}$ values indicated favorable free energy of binding for PRO^{D30N} compared to PRO^{WT} and correlated with the increase in antiviral potency of TPV with the D30N mutant.

In summary, the $\Delta\Delta G_{qm/wat}$ values provided consistent trends of the change in free energy of binding for PRO^{D30N}, which is resistant to SQV and NFV but is more sensitive to APV and TPV. $\Delta\Delta G_{qm}$ and $\Delta\Delta G_{mm}$ did not always provide the correct trend of change in the free energies of binding.

Correlation of Free Energy and Antiviral Potency for Active Site Mutants. The GBSA free energies were simulated under four conditions: (i) with implicit solvation terms and MM charges on both ligand and protein (ΔG_{mm}), (ii) with explicit water and implicit solvation terms (hybrid solvation model) with MM charges on both ligand and protein ($\Delta G_{mm/wat}$), (iii) with implicit solvation terms and protein-polarized QM charges on the ligand and MM charges on the protein (ΔG_{qm}), and (iv) with explicit water and implicit solvation terms (hybrid solvation model) with QM charges on the ligand and MM charges on the protein ($\Delta G_{qm/wat}$). All the free energy values determined are shown in Supporting Information Tables S1 and S2.

We analyzed the correlation of the free energies thus computed with the experimentally determined antiviral potency data (IC_{50}), and the resultant correlation coefficients are shown in Table 3, where Set-1 refers to the correlation coefficient for PRO^{WT}, PRO^{D30N}, and PRO^{I50V}, Set-2 refers to the correlation coefficient for PRO^{WT}, PRO^{D30N}, PRO^{I50V}, and $PRO^{V82I/I85V}$, and Set-3 refers to the correlation coefficients for PRO^{WT}, PRO^{D30N}, PRO^{I50V}, $PRO^{V82I/I85V}$, PRO^{2840}, and PRO^{2841}. For Set-1, r_{mm} (correlation coefficient of ΔG_{mm} vs IC_{50}) showed a strong correlation for only TPV, LPV, and AZV (Table 3). The r_{mm} value was poor for the other PIs and indicates the difficulty of obtaining reasonable

correlations between free energies and antiviral potency. The r_{qm} (correlation coefficient of ΔG_{qm} vs IC_{50}) values that represented correlation coefficients when the free energies were simulated with polarized QM charges on the ligands showed significant improvement and a strong correlation for DRV, APV, GRL-02031, and LPV. However, both r_{mm} and r_{qm} values were poor for GRL-98065 and SQV.

We next determined, for Set-1, the correlation obtained by the hybrid water model that has an explicit bridging water molecule between the inhibitor and the protease flap. The explicit water was treated as a part of the protein, and implicit solvation terms were used. The $r_{mm/wat}$ value (correlation coefficient of $\Delta G_{mm/wat}$ vs IC_{50}) represented a greater correlation than r_{mm} for all PIs except TPV and AZV. TPV directly formed hydrogen bonds with Ile50 and Ile50′, and the water molecule included in this calculation formed hydrogen bonds with Gly48 of one monomer of the protease dimer. For other PIs, the bridging water molecule formed hydrogen bonds with the flaps from both monomers. The $r_{qm/wat}$ value (correlation coefficient of $\Delta G_{qm/wat}$ vs IC_{50}) had a high degree of correlation for all PIs except AZV. Thus, the explicit inclusion of the water molecule bridging hydrogen bonds with the flap and protein-polarized QM charges for the inhibitors provided strong correlation ($r > 0.75$) for seven out of eight inhibitors. The correlation coefficient $r_{qm/wat}$ for NFV with three bridging waters was 0.97, a significant improvement over the correlation coefficient of 0.77 obtained with one bridging water molecule. The $r_{qm/wat}$ value for AZV also improved from 0.16 to 0.64 with the inclusion of three bridging water molecules.

We also determined r_{mm}, r_{qm}, $r_{mm/wat}$, and $r_{qm/wat}$ values for Set-2, which included $PRO^{V82I/I85V}$ as well (Table 3). The r_{mm} value was poor for all PIs except TPV, while the r_{qm} values showed good correlations for DRV, LPV, and TPV. The $r_{mm/wat}$ value showed strong correlation for only SQV and good correlation ($0.55 < r < 0.75$) for NFV. The $r_{qm/wat}$ value showed strong correlations for DRV, APV, and SQV and good correlations for GRL-98065 and TPV. The $r_{qm/wat}$ value for NFV jumped from 0.59 to 0.92 with the incorporation of three bridging water molecules instead of one.

The correlation coefficients with a hybrid water model and with QM-polarized ligand charges ($r_{qm/wat}$) on the PIs were higher for most of the PIs compared to the correlation coefficients obtained without any explicit water molecule and MM charges (r_{mm}). The only exception was for TPV, which was the only nonpeptidomimetic inhibitor among the PIs examined. TPV displaces the tetracoordinated water molecule and interacts directly with Ile50 and Ile50′ in the flap.[46] The hydrogen-bond interaction of the bridging water molecule with TPV and Gly48 of one chain might not be an important contributor to its potency. Also, in general, the $r_{qm/wat}$ values provided better correlations than r_{qm}.

Correlation of Free Energy and Antiviral Potency for Active and Nonactive Site Amino Acid Substitutions. We next analyzed the correlations of the free energies with the antiviral potency (IC_{50} values) for PRO^{WT}, PRO^{D30N}, PRO^{I50V}, $PRO^{V82I/I85V}$, PRO^{2840} that contains L10R, M46I, L63P, V82T, and I84V, and PRO^{2841} that contains M46I, L63P, V82T, and I84V substitutions (Set-3 in Table 3). The analysis of PRO^{2840} and PRO^{2841} was substantially complex since both proteases contained nonactive site substitutions, but it was worth examining the ability of the

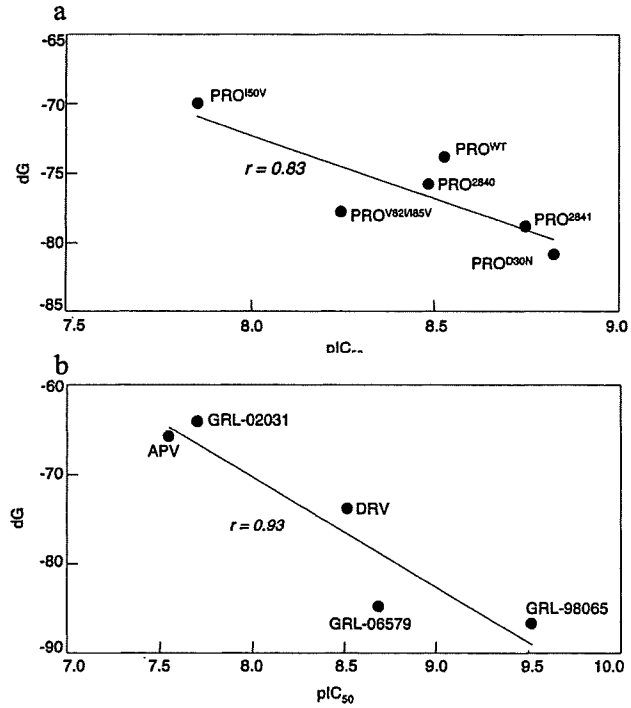


Figure 3. Correlation of free energy and antiviral activity. (a) Scatter plot of $\Delta G_{qm/wat}$ vs pIC_{50} for DRV for Set-3. (b) Scatter plot of $\Delta G_{qm/wat}$ vs pIC_{50} (PRO^{WT}) for ligands with bis-THF-related cores. The correlation coefficient is shown in both figures.

GBSA energy function to correlate with antiviral activity when substitutions distant from the inhibitor were present. In general, the correlation coefficients for Set-3 turned out to be low, indicating a lower correlation between the free energies and antiviral potencies when nonactive site mutants were present. For DRV, the r_{mm}, r_{qm}, and $r_{mm/wat}$ values indicated that the corresponding free energies had no correlation with IC_{50} values. However, the value of $r_{qm/wat}$ for DRV was 0.83, showing a strong correlation for simulations with polarized QM charges on the inhibitor with a hybrid water model (Table 3, Figure 3a). For SQV, strong correlations were obtained for both ΔG_{mm} with IC_{50} ($r_{mm} = 0.78$) and $\Delta G_{qm/wat}$ with IC_{50} ($r_{qm/wat} = 0.84$). For APV and TPV, $r_{qm/wat}$ values were 0.60 and 0.50, respectively, but there was no correlation when other GBSA protocols were used for these inhibitors. The $r_{qm/wat}$ values were greater than 0.75 for two inhibitors and greater than 0.55 for five inhibitors. The r_{qm} and $r_{mm/wat}$ values were greater than 0.55 for two inhibitors. Thus, the hybrid water model and inclusion of polarization effects, compared to the other protocols, simulated free energies with better correlation with antiviral IC_{50} for more inhibitors.

GBSA with Hybrid Water Model and Polarizable Quantum Charges on PIs as a Lead Optimization Tool. Obtaining a correct relative rank of activity for inhibitors that had potency in the nanomolar range has been a real challenge for scoring methods.[3] Scoring methods providing sufficiently high correlation with potency may serve as a lead optimization tool. In our data set, DRV and GRL-98065 have a bis-THF group as the core, GRL-06579 and GRL-02031 have a Cp-THF as the core, and APV has a THF group as the core. All these inhibitors were extremely potent against PRO^{WT} and had a narrow range of activity

Table 4. Experimental Free Energies[a] of Protease Inhibitors Against PRO^{WT}

	DRV	APV	GRL-98065	GRL-02031	GRL-06579
IC_{50} (μM)	0.003	0.028	0.0003	0.02	0.002
ΔG_{exp} (kcal/mol)	−11.6	−10.3	−13.0	−10.5	−11.9

[a] Experimental free energies determined as $\Delta G_{exp} = RT \ln(IC_{50})$, where R and T represent the universal gas constant and temperature, respectively.

ranging from 0.3 to 28 nM (Table 1). We computed ΔG_{exp} from the antiviral IC_{50} values (Table 4). Such transformations have recently enhanced the understanding of the binding of *N*-sulphonyl-glutamic acid inhibitors to MurD ligase and in understanding the efficiency of DNA catalysis.[19,47] Substitution of the THF group of APV with the bis-THF group (DRV) resulted in a −1.3 kcal/mol improvement in the free energy of binding. GRL-98065 has a 1,3-benzodioxole group as P2′ ligand compared to an aniline group in DRV, which resulted in the ΔG_{exp} of GRL-98065 being lower (i.e., better binding affinity) by −1.4 kcal/mol than DRV. Both GRL-06579 and GRL-02031 have a Cp-THF as the P2 ligand but different substituents interacting with the S1′ and S2′ locations in the protease active site. The ΔG_{exp} of GRL-06579 is lower by 1.4 kcal/mol than GRL-02031. Among these five PIs, APV and GRL-02031 have the worst ΔG_{exp} and GRL-98065 has the best ΔG_{exp}. The absolute magnitudes of our simulated GBSA free energies are larger than those measured experimentally due to the force-field parameters and the form of the energy function, but trends and correlation with experiments can be deduced. Our $\Delta G_{qm/wat}$ values simulated the correct trend by predicting that the worst binding free energy was for APV and GRL-02031, and the best binding free energy was for GRL-98065 (Supporting Information Table S1). None of the other GBSA protocols predicted the trend correctly (Supporting Information Tables S1 and S2). We further analyzed the correlation of the free energies calculated by our methods against IC_{50} values for these five PIs. The r_{mm} and $r_{mm/wat}$ values were 0.10 and 0.03, respectively, indicating that there was no correlation of GBSA free energies simulated by using MM charges with or without explicit inclusion of water. The difficulty of obtaining good correlation of free energies against antiviral activity is evident from this small but nontrivial data set. The r_{qm} value was 0.74, indicating that incorporation of protein-polarized QM charges on the inhibitors substantially improved the correlation. Including the bridging water molecule with protein-polarized QM charges on the PIs resulted in a $r_{qm/wat}$ value of 0.93 for PRO^{WT} (Figure 3b). Thus, a simulation using the hybrid water model and protein-polarized QM charges on the ligands resulted in a strong correlation, while there was no correlation for simulations with fixed MM charges.

For our docking and subsequent GBSA simulations, we used the crystal coordinates of BCV−protease complex (PDB ID 2FDE)[22] as our starting template. BCV, DRV, and GRL-98065 had a bis-THF moiety as the core ligand, and APV, GRL-06579, and GRL-02031 had cores that had a high similarity with bis-THF. The high correlation obtained with $\Delta G_{qm/wat}$-IC_{50} indicated that free energies obtained with a hybrid water model and polarized QM charges on the ligands would be a promising approach for lead optimization. Currently, we are exploring the utility of GBSA free energy simulations once a correct binding mode is obtained, and we wanted to use native or native-like crystal structure as a template as a proof-of-principle study. We have not explored docking ligands against non-native protease crystal structures in the current study. A more exhaustive examination that includes poses obtained by cross docking needs to be performed.

DISCUSSION

Progress has been made in improving the robustness of scoring functions to determine the relative affinity of inhibitors for target proteins, but there are vast scopes where significant improvement can be made.[1,3] Part of the reason for the inaccuracies in the scoring functions arise because (i) the current methodologies largely account for enthalpic changes while completely ignoring entropic changes, (ii) they do not properly treat protein flexibility, (iii) they do not properly account for solvation and desolvation effects, and (iv) they do not account for the polarization induced by the protein and the ligand on each other.[1,3,7] Most of the studies on scoring functions deal with rank ordering the activity of a congeneric set of ligands. The prediction of activity of a ligand against mutant proteins is very important in light of drug-resistance mutations that emerge in many therapeutic areas.

We explored the correlation of GBSA free energies with antiviral potency (IC_{50} values) of nine different PIs against wild-type and mutant proteases. Correlation with enzymatic K_i would result in similar conclusions since these PIs have good cell penetration and are highly protease-specific inhibitors. Furthermore, these PIs do not form aggregates and do not have the problematic properties discussed by Shoichet et al.[48,49]

The GBSA free energies were simulated by four different protocols: (i) fixed MM charges in implicit solvent (ΔG_{mm}), (ii) protein-polarized QM charges of PIs in implicit solvent (ΔG_{qm}), (iii) fixed MM charges in a hybrid solvent ($\Delta G_{mm/wat}$), and (iv) protein-polarized QM charges of PIs in a hybrid solvent ($\Delta G_{qm/wat}$). The hybrid solvent protocols have retained an explicit water molecule(s) that bridges hydrogen bonding with the protease in an otherwise implicit solvent environment. The protein-polarized ligand charges were determined during ligand docking in the protein environment at the B3LYP/6-31G* level, and these polarized charges were maintained for the ligand for the full GBSA simulation cycle. These enable us to analyze the effect of a different charge model compared to the static charges from a force field (Supporting Information Figures S3 and S4). Free energies from a successful GBSA protocol should provide a good correlation with the antiviral activity against wild-type and mutant protease. Resistance mutations in the protease active site arise primarily due to loss of favorable binding interactions. Resistance caused by nonactive site mutations is more difficult to understand and rationalize,

although some attempts to elucidate the mechanism have recently been made.[12,50–52]

We initially compared the correlation coefficients for PRO^{WT}, PRO^{D30N}, and PRO^{I50V} (Set-1). The antiviral potency of the inhibitors for wild-type and mutant proteases are shown in Table 1, and the inhibitors have potency (IC_{50} values) in the nanomolar range. The $r_{qm/wat}$ value was higher than 0.75 for 7 out of 8 PIs for Set-1, with $r_{qm/wat}$ being more than 0.90 for 4 PIs (Table 3). PRO^{D30N} was associated with SQV, NFV, and LPV resistance of HIV-1 and increased susceptibility to APV and TPV (Table 1). It is noteworthy that $r_{qm/wat}$ showed substantial correlation values of greater than 0.75 even though the fold change in antiviral activity for PRO^{D30N} is nonmonotonic. AZV is the only PI that did not show a correlation of $\Delta G_{qm/wat}$ with the IC_{50} value with one bridging water molecule. By including two additional bridging water molecules for NFV and AZV, $r_{qm/wat}$ improved to 0.97 and 0.64, respectively. Selective inclusion of explicit water molecules needs to be explored and validated. For Set-1, $r_{mm/wat}$, r_{qm}, and r_{mm} had strong correlation for 4, 3, and 3 PIs, respectively.

Correlations for Set-2 included the mutant $PRO^{V82I/I85V}$. While V82I is located in the protease active site and represents one of the major resistance-associated mutations for PIs, I85V was selected as a resistance-associated mutation for GRL-98065 even though it does not form a direct van der Waals contact.[32] For Set-2, the number of PIs that showed strong correlation coefficients for $r_{qm/wat}$ and $r_{mm/wat}$ were 4 and 1, respectively, while r_{qm} and r_{mm} did not show strong correlations for any PI. Set-3 included Set-2 as well as PRO^{2840} and PRO^{2841}, which had mutations distant from the active site. We included such mutations because they are seen in drug-resistant HIV-1-harboring patients, and we wanted to test the ability of the GBSA energy function to correlate with the antiviral IC_{50} for such protease substitutions. The $r_{qm/wat}$ values showed strong correlation coefficients for only DRV and SQV for Set-3 and moderate correlation for 3 other PIs. The correlation coefficients, $r_{mm/wat}$ and r_{qm}, were from 0.55 to 0.75 for two PIs in Set-3, suggesting that those two PIs had a moderate correlation between the free energies and the IC_{50} values.

Analysis of correlation coefficients in Table 3 indicated that $r_{qm/wat}$ had strong ($r > 0.75$) and moderate ($0.55 < r < 0.75$) correlation for more PIs than $r_{mm/wat}$, r_{qm}, or r_{mm}. This suggested that the GBSA free energies simulated with a hybrid water model with protein-polarized QM charges on the inhibitors had a higher correlation with antiviral IC_{50} than the other free energy simulation protocols. Others[1,3] have suggested that improved treatment of solvation and polarizability may improve the robustness of scoring functions, and we demonstrated that our use of selected explicit water molecule(s) and protein-polarized QM partial charges on the inhibitor provided greater correlation with antiviral potency. Further improvement might be achieved by (i) improving upon the hybrid water model (ii) by accounting for the polarization induced by the inhibitors on the protein atoms and (iii) by including changes in entropy.

ACKNOWLEDGMENT

This work was supported in part by the Intramural Research Program of the Center for Cancer Research, National Cancer Institute, National Institutes of Health (D.D. and H.M.), a grant from the National Institutes of Health (GM 53386 to A.K.G.), a grant from the Kumamoto University Global Center of Excellence Program, Global Education and Research Center Aiming at the control of AIDS (H.M.) supported by the Ministry of Education, Culture, Sports, Science, and Technology (Monbu-Kagakusho), a Grant-in-Aid for Scientific Research (Priority Areas to H.M.) from the Ministry of Education, Culture, Sports, Science, and Technology (Monbu-Kagakusho) of Japan (H.M.), and a Grant for Promotion of AIDS Research from the Ministry of Health, Labor and Welfare (Kosei-Rodosho) of Japan (H.M.). The work utilized the computational resources of the Biowulf cluster at the NIH.

Supporting Information Available: Simulated GBSA free energies; charges on DRV for OPLS2005 force field (MM level) and protein-polarized charges at the B3LYP/6-31G* level. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Clark, D. E. What has virtual screening ever done for drug discovery. *Expert Opin. Drug Discovery* **2008**, *3*, 841–851.
(2) Lyne, P. D.; Lamb, M. L.; Saeh, J. C. Accurate prediction of the relative potencies of members of a series of kinase inhibitors using molecular docking and MM-GBSA scoring. *J. Med. Chem.* **2006**, *49*, 4805–4808.
(3) Leach, A. R.; Shoichet, B. K.; Peishoff, C. E. Prediction of protein-ligand interactions. Docking and scoring: Successes and gaps. *J. Med. Chem.* **2006**, *49*, 5851–5855.
(4) Rajamani, R.; Good, A. C. Ranking poses in structure-based lead discovery and optimization: Current trends in scoring function development. *Curr. Opin. Drug Discovery Dev.* **2007**, *10*, 308–315.
(5) Lu, Y. P.; Wang, R. X.; Yang, C. Y.; Wang, S. M. Analysis of ligand-bound water molecules in high-resolution crystal structures of protein-ligand complexes. *J. Chem. Inf. Model.* **2007**, *47*, 668–675.
(6) Huang, N.; Shoichet, B. K. Exploiting ordered waters in molecular docking. *J. Med. Chem.* **2008**, *51*, 4862–4865.
(7) Fornabaio, M.; Spyrakis, F.; Mozzarelli, A.; Cozzini, P.; Abraham, D. J.; Kellogg, G. E. Simple, intuitive calculations of free energy of binding for protein-ligand complexes. 3. The free energy contribution of structural water molecules in HIV-1 protease complexes. *J. Med. Chem.* **2004**, *47*, 4507–4516.
(8) Yin, P. D.; Das, D.; Mitsuya, H. Overcoming HIV drug resistance through rational drug design based on molecular, biochemical, and structural profiles of HIV resistance. *Cell. Mol. Life Sci.* **2006**, *63*, 1706–1724.
(9) Clemente, J. C.; Moose, R. E.; Hemrajani, R.; Whitford, L. R. S.; Govindasamy, L.; Reutzel, R.; McKenna, R.; Agbandje-McKenna, M.; Goodenow, M. M.; Dunn, B. M. Comparing the accumulation of active- and nonactive-site mutations in the HIV-1 protease. *Biochemistry* **2004**, *43*, 12141–12151.
(10) Hong, L.; Zhang, X. J. C.; Hartsuck, J. A.; Tang, J. Crystal structure of an in vivo HIV-1 protease mutant in complex with saquinavir: Insights into the mechanisms of drug resistance. *Protein Sci.* **2000**, *9*, 1898–1904.
(11) Kozisek, M.; Bray, J.; Rezacova, P.; Saskova, K.; Brynda, J.; Pokorna, J.; Mammano, F.; Rulisek, L.; Konvalinka, J. Molecular analysis of the HIV-1 resistance development: Enzymatic activities, crystal structures, and thermodynamics of nelfinavir-resistant HIV protease mutants. *J. Mol. Biol.* **2007**, *374*, 1005–1016.
(12) Mahalingam, B.; Wang, Y. F.; Boross, P. I.; Tozser, J.; Louis, J. M.; Harrison, R. W.; Weber, I. T. Crystal structures of HIV protease V82A and L90M mutants reveal changes in the indinavir-binding site. *Eur. J. Biochem.* **2004**, *271*, 1516–1524.
(13) Ghosh, A. K.; Sridhar, P. R.; Leshchenko, S.; Hussain, A. K.; Li, J. F.; Kovalevsky, A. Y.; Walters, D. E.; Wedekind, J. E.; Grum-Tokars, V.; Das, D.; Koh, Y.; Maeda, K.; Gatanaga, H.; Weber, I. T.; Mitsuya, H. Structure-based design of novel HIV-1 protease inhibitors to combat drug resistance. *J. Med. Chem.* **2006**, *49*, 5252–5261.
(14) Kovalevsky, A. Y.; Tie, Y. F.; Liu, F. L.; Boross, P. I.; Wang, Y. F.; Leshchenko, S.; Ghosh, A. K.; Harrison, R. W.; Weber, I. T.

Effectiveness of nonpeptide clinical inhibitor TMC-114 on HIV-1 protease with highly drug resistant mutations D30N, I50V, and L90M. *J. Med. Chem.* **2006**, *49*, 1379–1387.

(15) Tie, Y. F.; Kovalevsky, A. Y.; Boross, P.; Wang, Y. F.; Ghosh, A. K.; Tozser, J.; Harrison, R. W.; Weber, I. T. Atomic resolution crystal structures of HIV-1 protease and mutants V82A and I84V with saquinavir. *Proteins* **2007**, *67*, 232–242.

(16) Aqvist, J.; Medina, C.; Samuelsson, J. E. New Method for Predicting Binding-Affinity in Computer-Aided Drug Design. *Protein Eng.* **1994**, *7*, 385–391.

(17) Straatsma, T. P.; McCammon, J. A. Computational Alchemy. *Annu. Rev. Phys. Chem.* **1992**, *43*, 407–435.

(18) Bren, U.; Martinek, V.; Florian, J. Free energy simulations of uncatalyzed DNA replication fidelity: structure and stability of T.G and dTTP.G Terminal DNA mismatches flanked by a single dangling nucleotide. *J. Phys. Chem. B* **2006**, *110*, 10557–10566.

(19) Perdih, A.; Bren, U.; Solmajer, T. Binding free energy calculations of N-sulphonyl-glutamic acid inhibitors of MurD ligase. *J. Mol. Model.* **2009**, *15*, 983–996.

(20) Wang, Y. F.; Tie, Y. F.; Boross, P. I.; Tozser, J.; Ghosh, A. K.; Harrison, R. W.; Weber, I. T. Potent new antiviral compound shows similar inhibition and structural interactions with drug resistant mutants and wild type HIV-1 protease. *J. Med. Chem.* **2007**, *50*, 4509–4515.

(21) Vega, S.; Kang, L. W.; Velazquez-Campoy, A.; Kiso, Y.; Amzel, L. M.; Freire, E. A structural and thermodynamic escape mechanism from a drug resistant mutation of the HIV-1 protease. *Proteins* **2004**, *55*, 594–602.

(22) Miller, J. F.; Andrews, C. W.; Brieger, M.; Furfine, E. S.; Hale, M. R.; Hanlon, M. H.; Hazen, R. J.; Kaldor, I.; McLean, E. W.; Reynolds, D.; Sammond, D. M.; Spaltenstein, A.; Tung, R.; Turner, E. M.; Xu, R. X.; Sherrill, R. G. Ultra-potent P1 modified arylsulfonamide HIV protease inhibitors: The discovery of GW0385. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 1788–1794.

(23) Krohn, A.; Redshaw, S.; Ritchie, J. C.; Graves, B. J.; Hatada, M. H. Novel Binding Mode of Highly Potent HIV-Proteinase Inhibitors Incorporating the (R)-Hydroxyethylamine Isostere. *J. Med. Chem.* **1991**, *34*, 3340–3342.

(24) Muzammil, S.; Armstrong, A. A.; Kang, L. W.; Jakalian, A.; Bonneau, P. R.; Schmelmer, V.; Amzel, L. M.; Freire, E. Unique thermodynamic response of tipranavir to human immunodeficiency virus type 1 protease drug resistance mutations. *J. Virol.* **2007**, *81*, 5144–5154.

(25) Kaldor, S. W.; Kalish, V. J.; Davies, J. F.; Shetty, B. V.; Fritz, J. E.; Appelt, K.; Burgess, J. A.; Campanale, K. M.; Chirgadze, N. Y.; Clawson, D. K.; Dressman, B. A.; Hatch, S. D.; Khalil, D. A.; Kosa, M. B.; Lubbehusen, P. P.; Muesing, M. A.; Patick, A. K.; Reich, S. H.; Su, K. S.; Tatlock, J. H. Viracept (nelfinavir mesylate, AG1343): A potent, orally bioavailable inhibitor of HIV-1 protease. *J. Med. Chem.* **1997**, *40*, 3979–3985.

(26) Stoll, V.; Qin, W. Y.; Stewart, K. D.; Jakob, C.; Park, C.; Walter, K.; Simmer, R. L.; Helfrich, R.; Bussiere, D.; Kao, J.; Kempf, D.; Sham, H. L.; Norbeck, D. W. X-ray crystallographic structure of ABT-378 (lopinavir) bound to HIV-1 protease. *Bioorg. Med. Chem.* **2002**, *10*, 2803–2806.

(27) Clemente, J. C.; Coman, R. M.; Thiaville, M. M.; Janka, L. K.; Jeung, J. A.; Nukoolkarn, S.; Govindasamy, L.; Agbandje-McKenna, M.; McKenna, R.; Leelamanit, W.; Goodenow, M. M.; Dunn, B. M. Analysis of HIV-1CRF_01 A/E protease inhibitor resistance: Structural determinants for maintaining sensitivity and developing resistance to atazanavir. *Biochemistry* **2006**, *45*, 5468–5477.

(28) Cho, A. E.; Guallar, V.; Berne, B. J.; Friesner, R. Importance of accurate charges in molecular docking: Quantum mechanical/molecular mechanical (QM/MM) approach. *J. Comput. Chem.* **2005**, *26*, 915–931.

(29) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739–49.

(30) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Med. Chem.* **2006**, *49*, 6177–96.

(31) Yu, Z. Y.; Jacobson, M. P.; Friesner, R. A. What role do surfaces play in GB models? A new-generation of surface-generalized Born model based on a novel Gaussian surface for biomolecules. *J. Comput. Chem.* **2006**, *27*, 72–89.

(32) Amano, M.; Koh, Y.; Das, D.; Li, J. F.; Leschenko, S.; Wang, Y. F.; Boross, P. I.; Weber, I. T.; Ghosh, A. K.; Mitsuya, H. A novel bis-tetrahydrofuranylurethane-containing nonpeptidic protease inhibitor (PI), GRL-98065, is potent against multiple-PI-resistant human immunodeficiency virus in vitro. *Antimicrob. Agents Chemother.* **2007**, *512*143–2155.

(33) Ghosh, A. K.; Leshchenko, S.; Noetzel, M. Stereoselective photochemical 1,3-dioxolane addition to 5-alkoxymethyl-2(5H)-furanone: Synthesis of bis-tetrahydrofuranyl ligand for HIV protease inhibitor UIC-94017 (TMC-114). *J. Org. Chem.* **2004**, *69*, 7822–7829.

(34) Ghosh, A. K.; Leshchenko-Yashchuk, S.; Anderson, D. D.; Baldridge, A.; Noetzel, M.; Miller, H. B.; Tie, Y. F.; Wang, Y. F.; Koh, Y.; Weber, I. T.; Mitsuya, H. Design of HIV-1 Protease Inhibitors with pyrrolidinones and oxazolidinones as novel P1′-Ligands To Enhance Backbone-Binding Interactions with Protease: Synthesis, Biological Evaluation, and Protein-Ligand X-ray Studies. *J. Med. Chem.* **2009**, *52*, 3902–3914.

(35) Koh, Y.; Das, D.; Leschenko, S.; Nakata, H.; Ogata-Aoki, H.; Amano, M.; Nakayama, M.; Ghosh, A. K.; Mitsuya, H. GRL-02031, a novel nonpeptidic protease inhibitor (PI) containing a stereochemically defined fused cyclopentanyltetrahydrofuran potent against multi-PI-resistant human immunodeficiency virus type 1 In Vitro. *Antimicrob. Agents Chemother.* **2009**, *53*, 997–1006.

(36) Yoshimura, K.; Kato, R.; Yusa, K.; Kavlick, M. F.; Maroun, V.; Nguyen, A.; Mimoto, T.; Ueno, T.; Shintani, M.; Falloon, J.; Masur, H.; Hayashi, H.; Erickson, J.; Mitsuya, H. JE-2147: A dipeptide protease inhibitor (PI) that potently inhibits multi-PI-resistant HIV-1. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 8675–8680.

(37) Koh, Y.; Nakata, H.; Maeda, K.; Ogata, H.; Bilcer, G.; Devasamudram, T.; Kincaid, J. F.; Boross, P.; Wang, Y. F.; Ties, Y. F.; Volarath, P.; Gaddis, L.; Harrison, R. W.; Weber, I. T.; Ghosh, A. K.; Mitsuya, H. Novel bis-tetrahydrofuranylurethane-containing nonpeptidic protease inhibitor (PI) UIC-94017 (TMC114) with potent activity against multi-PI-resistant human immunodeficiency virus in vitro. *Antimicrob. Agents Chemother.* **2003**, *47*, 3123–3129.

(38) Markowitz, M.; Conant, M.; Hurley, A.; Schluger, R.; Duran, M.; Peterkin, J.; Chapman, S.; Patick, A.; Hendricks, A.; Yuen, G. J.; Hoskins, W.; Clendeninn, N.; Ho, D. D. *A preliminary evaluation of nelfinavir mesylate, an inhibitor of human immunodeficiency virus (HIV)-1 protease, to treat HIV infection*; University of Chicago Press: Chicago, 1998; pp 1533−1540.

(39) Wlodawer, A.; Erickson, J. W. Structure-Based Inhibitors of HIV-1 Protease. *Annu. Rev. Biochem.* **1993**, *62*, 543–585.

(40) Johnson, V. A.; Brun-Vezinet, F.; Clotet, B.; Conway, B.; Kuritzkes, D. R.; Pillay, D.; Schapiro, J.; Telenti, A.; Richman, D. Update of the drug resistance mutations in HIV-1. *Top. HIV Med. 2005* **2005**, *13*, 51–57.

(41) Baldwin, E. T.; Bhat, T. N.; Liu, B. S.; Pattabiraman, N.; Erickson, J. W. Structural basis of drug-resistance for the V82A mutant of HIV-1 proteinase. *Nat. Struct. Biol.* **1995**, *2*, 244–249.

(42) Maeda, K.; Yoshimura, K.; Shibayama, S.; Habashita, H.; Tada, H.; Sagawa, K.; Miyakawa, T.; Aoki, M.; Fukushima, D.; Mitsuya, H. Novel low molecular weight spirodiketopiperazine derivatives potently inhibit R5 HIV-1 infection through their antagonistic effects on CCR5. *J. Biol. Chem.* **2001**, *276*, 35194–35200.

(43) Kim, E. E.; Baker, C. T.; Dwyer, M. D.; Murcko, M. A.; Rao, B. G.; Tung, R. D.; Navia, M. A. Crystal-Structure of HIV-1 Protease in Complex with VX-478, a Potent and Orally Bioavailable Inhibitor of the Enzyme. *J. Am. Chem. Soc.* **1995**, *117*, 1181–1182.

(44) Wlodawer, A.; Vondrasek, J. Inhibitors of HIV-1 protease: A major success of structure-assisted drug design. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 249–284.

(45) Thaisrivongs, S.; Skulnick, H. I.; Turner, S. R.; Strohbach, J. W.; Tommasi, R. A.; Johnson, P. D.; Aristoff, P. A.; Judge, T. M.; Gammill, R. B.; Morris, J. K.; Romines, K. R.; Chrusciel, R. A.; Hinshaw, R. R.; Chong, K. T.; Tarpley, W. G.; Poppe, S. M.; Slade, D. E.; Lynn, J. C.; Horng, M. M.; Tomich, P. K.; Seest, E. P.; Dolak, L. A.; Howe, W. J.; Howard, G. M.; Schwende, F. J.; Toth, L. N.; Padbury, G. E.; Wilson, G. J.; Shiou, L. H.; Zipp, G. L.; Wilkinson, K. F.; Rush, B. D.; Ruwart, M. J.; Koeplinger, K. A.; Zhao, Z. Y.; Cole, S.; Zaya, R. M.; Kakuk, T. J.; Janakiraman, M. N.; Watenpaugh, K. D. Structure-based design of HIV protease inhibitors: Sulfonamide-containing 5,6-dihydro-4-hydroxy-2-pyrones as non-peptidic inhibitors. *J. Med. Chem.* **1996**, *39*, 4349–4353.

(46) Turner, S. R.; Strohbach, J. W.; Tommasi, R. A.; Aristoff, P. A.; Johnson, P. D.; Skulnick, H. I.; Dolak, L. A.; Seest, E. P.; Tomich, P. K.; Bohanan, M. J.; Horng, M. M.; Lynn, J. C.; Chong, K. T.; Hinshaw, R. R.; Watenpaugh, K. D.; Janakiraman, M. N.; Thaisrivongs, S. Tipranavir (PNU-140690): A potent, orally bioavailable nonpeptidic HIV protease inhibitor of the 5,6-dihydro-4-hydroxy-2-pyrone sulfonamide class. *J. Med. Chem.* **1998**, *41*, 3467–3476.

(47) Brown, K. L.; Bren, U.; Stone, M. P.; Guengerich, F. P. Inherent stereospecificity in the reaction of aflatoxin B1 8,9-epoxide with deoxyguanosine and efficiency of DNA catalysis. *Chem. Res. Toxicol.* **2009**, *22*, 913–917.

(48) McGovern, S. L.; Caselli, E.; Grigorieff, N.; Shoichet, B. K. A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening. *J. Med. Chem.* **2002**, *45*, 1712–1722.
(49) Shoichet, B. K. Screening in a spirit haunted world. *Drug Discovery Today* **2006**, *11*, 607–615.
(50) Ode, H.; Matsuyama, S.; Hata, M.; Hoshino, T.; Kakizawa, J.; Sugiura, W. Mechanism of drug resistance due to N88S in CRF01_AE HIV-1 protease, analyzed by molecular dynamics simulations. *J. Med. Chem.* **2007**, *50*, 1768–1777.
(51) Ode, H.; Matsuyama, S.; Hata, M.; Neya, S.; Kakizawa, J.; Sugiura, W.; Hoshino, T. Computational characterization of structural role of the non-active site mutation M36I of human immunodeficiency virus type 1 protease. *J. Mol. Biol.* **2007**, *370*, 598–607.
(52) Ode, H.; Neya, S.; Hata, M.; Sugiura, W.; Hoshino, T. Computational simulations of HIV-1 proteases-multi-drug resistance due to nonactive site mutation L90M. *J. Am. Chem. Soc.* **2006**, *128*, 7887–7895.

CI900320P

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 284, NO. 51, pp. 35681–35691, December 18, 2009
Printed in the U.S.A.

Mechanism of Inhibition of HIV-1 Reverse Transcriptase by 4′-Ethynyl-2-fluoro-2′-deoxyadenosine Triphosphate, a Translocation-defective Reverse Transcriptase Inhibitor*

Received for publication, June 23, 2009, and in revised form, October 14, 2009 Published, JBC Papers in Press, October 16, 2009, DOI 10.1074/jbc.M109.036616

Eleftherios Michailidis[‡], Bruno Marchand[‡1], Eiichi N. Kodama[§2], Kamlendra Singh[‡], Masao Matsuoka[§], Karen A. Kirby[‡], Emily M. Ryan[‡], Ali M. Sawani[‡], Eva Nagy[¶], Noriyuki Ashida[‖], Hiroaki Mitsuya[‡‡], Michael A. Parniak[¶], and Stefan G. Sarafianos[‡3]**

From the [‡]Christopher Bond Life Sciences Center, Department of Molecular Microbiology and Immunology, University of Missouri, Columbia, Missouri 65211, the [§]Institute for Virus Research, Kyoto University, Kyoto 606-8507, Japan, the [¶]Department of Molecular Genetics and Biochemistry, University of Pittsburgh, Pittsburgh, Pennsylvania 15261, [‖]Yamasa Corporation, Chiba 288-0056, Japan, the [**]Department of Hematology and Infectious Diseases, Kumamoto University, Kumamoto 860-8556, Japan, and the [‡‡]Experimental Retrovirology Section, HIV/AIDS Malignancy Branch, National Institutes of Health, Bethesda, Maryland 20892

Nucleoside reverse transcriptase inhibitors (NRTIs) are employed in first line therapies for the treatment of human immunodeficiency virus (HIV) infection. They generally lack a 3′-hydroxyl group, and thus when incorporated into the nascent DNA they prevent further elongation. In this report we show that 4′-ethynyl-2-fluoro-2′-deoxyadenosine (EFdA), a nucleoside analog that retains a 3′-hydroxyl moiety, inhibited HIV-1 replication in activated peripheral blood mononuclear cells with an EC_{50} of 0.05 nM, a potency several orders of magnitude better than any of the current clinically used NRTIs. This exceptional antiviral activity stems in part from a mechanism of action that is different from approved NRTIs. Reverse transcriptase (RT) can use EFdA-5′-triphosphate (EFdA-TP) as a substrate more efficiently than the natural substrate, dATP. Importantly, despite the presence of a 3′-hydroxyl, the incorporated EFdA monophosphate (EFdA-MP) acted mainly as a *de facto* terminator of further RT-catalyzed DNA synthesis because of the difficulty of RT translocation on the nucleic acid primer possessing 3′-terminal EFdA-MP. EFdA-TP is thus a translocation-defective RT inhibitor (TDRTI). This diminished translocation kept the primer 3′-terminal EFdA-MP ideally located to undergo phosphorolytic excision. However, net phosphorolysis was not substantially increased, because of the apparently facile reincorporation of the newly excised EFdA-TP. Our molecular modeling studies suggest that the 4′-ethynyl fits into a hydrophobic pocket defined by RT residues Ala-114, Tyr-115, Phe-160, and Met-184 and the aliphatic chain of Asp-185. These interactions, which contribute to both enhanced RT utilization of EFdA-TP and difficulty in the translocation of 3′-terminal EFdA-MP primers, underlie the mechanism of action of this potent antiviral nucleoside.

Nucleoside reverse transcriptase inhibitors (NRTIs)[4] are central components of first line regimens for treatment of HIV infections (1–6). Currently, there are eight clinically approved NRTIs: AZT, 3TC, FTC, ABC, ddI, ddC, d4T, and the nucleotide tenofovir (TFV; reviewed in Refs. 7 and 8). A structural hallmark of these NRTIs is the lack of a 3′-OH; it has long been considered that the absence of the 3′-OH is essential for antiviral activity. However, the absence of the 3′-OH in NRTIs also imparts detrimental properties to the inhibitor, including reduced affinity for RT compared with the analogous dNTP substrate, as well as reduced intracellular conversion to the active nucleoside triphosphate (9).

Previously we described a series of 4′-substituted NRTIs (10) that retain the 3′-OH group and have excellent antiviral properties and significantly improved selectivity indices (CC_{50}/EC_{50}) compared with the approved NRTIs. Furthermore, these NRTIs efficiently suppress various NRTI-resistant HIV. The most potent of these 4′-substituted NRTIs are the adenosine analogs that have an ethynyl group at the 4′ position of the ribose ring. Despite their high anti-HIV activity, 4′-substituted compounds are susceptible to degradation by adenosine deaminase (11), a property that limits the plasma and intracellular half-life of the drugs. To overcome the adenosine deaminase sensitivity of these 4′-ethynyl NRTIs, we developed a second generation of analogs substituted at the 2-position of the adenine ring (12). We recently reported that the 2-halogenated, 4′-ethynyl compounds have remarkably improved potency and selectivity indices (CC_{50}/EC_{50}) compared with the nonhalogenated analogs and significantly better ones compared with

* This work was supported, in whole or in part, by National Institutes of Health Grants AI076119, AI074389, AI076119-S1, and AI076119-02S1 (to S. G. S.) and AI079801 (to M. A. P.).

[1] Recipient of the amfAR Mathilde Krim Fellowship.

[2] Present address: Div. of Emerging Infectious Diseases, Tohoku University School of Medicine, Sendai 980-8575, Japan.

[3] To whom correspondence should be addressed: 471d Christopher S. Bond Life Sciences Ctr., 1201 Rollins St., Columbia, MO 65211. Tel.: 573-882-4338; E-mail: sarafianoss@missouri.edu.

[4] The abbreviations used are: NRTI, nucleoside reverse transcriptase inhibitor; TDRTI, translocation-defective RT inhibitor; RT, reverse transcriptase; HIV, human immunodeficiency virus; EFdA, 4′-ethynyl-2-fluoro-2′-deoxyadenosine; MP, monophosphate; TP, triphosphate; AZT, azidothymidine; EdA, 4′-ethynyl-2′-deoxyadenosine; EFddA, 4′-ethynyl-2-fluoro-2′,3′-dideoxyadenosine; EFd4A, 4′-ethynyl-2-fluoro-2′,3′-dihydro-2′,3′-dideoxyadenosine; Ed4T, 4′-ethynyl-2′,3′-dihydro-3′-deoxythymidine; TFV, tenofovir; PBMC, peripheral blood mononuclear cell; T/P, template/primer; $T/P_{EFdA\text{-}MP}$ or $T/P_{ddA\text{-}MP}$, template/primer possessing either EFdA-MP or ddAMP at the 3′-primer terminus (or T/P chain terminated by EFdA or ddA); N-site, nucleotide-binding site; P-site, primer site; PDB, Protein Data Bank; d4T, stavudine.

TABLE 1

DNA and RNA sequences used in this study

The primers were fluorescently labeled at the 5′-end except for the footprinting experiments, in which the template was fluorescently labeled at the 5′-end.

Polymerization experiments	
T_{d100}	5′-TAG TGT GTG CCC GTC TGT TGT GTG ACT CTG GTA ACT AGA GAT CCC TCA GAC CCT TTT AGT
	CAG TGT GGA AAA TCT CTA GCA GTG GCG CCC GAA CAG GGA C
P_{d18}	5′-Cy3 GTC CCT GTT CGG GCG CCA
T_{d31}	5′-CCA TAG CTA GCA TTG GTG CTC GAA CAG TGA C
T_{r31}	5′-CCA UAG CUA GCA UUG GUG CUC GAA CAG UGA C
P_{d18}	5′-Cy3 GTC ACT GTT CGA GCA CCA
Gel shift experiments	
T_{d43}	5′-AAT CAG TGT AGA CAA TCC CTA GCA TTG GTG CTC GAA CAG TGA C
P_{d18}	5′-Cy3 GTC CCT GTT CGG GCG CCA
Footprinting experiments	
P_{d20}	5′-TTG TCA CTG TTC GAG CAC CA
T_{d43}	5′-Cy3 CCA TAG CTA GCA TTG GTG CTC GAA CAG TGA CAA TCA GTG TAGA

other approved NRTIs. These compounds are resistant to degradation by adenosine deamination (13). The most potent of these compounds is EFdA (Fig. 1*A*), which was recently shown not to inhibit human DNA polymerases α and β or mitochondrial DNA polymerase γ (12). Notably, clinically important drug-resistant HIVs (14, 15) are sensitive or hypersensitive to this compound (13).

Despite its remarkable antiviral potency, the molecular mechanism by which EFdA and related compounds inhibit HIV is unknown. To elucidate this mechanism we carried out biochemical experiments that systematically decipher the effect of EFdA on each of the mechanistic steps of DNA synthesis by HIV RT. On the basis of these experiments we propose that EFdA-5′-triphosphate (EFdA-TP) inhibits RT by first being incorporated at the 3′-primer terminus, and after its incorporation it prevents further addition of nucleotides by blocking the translocation of the primer strand on the viral polymerase. We therefore termed EFdA a "translocation-defective reverse transcriptase inhibitor (TDRTI)." By understanding the molecular details of RT inhibition by a highly potent NRTI, we hope to gain insights into the design of even more efficacious inhibitors that may act via same or similar mechanisms.

EXPERIMENTAL PROCEDURES

Enzymes and Nucleic Acids

The RT genes coding for p66 and p51 subunits of BH10 HIV-1 were cloned in the pETDuet-1 vector (Novagen) using restriction sites NcoI and SacI for the p51 subunit and SacII and AvrII for the p66 subunit. The sequences coding for a hexahistidine tag and the 3C protease recognition sequence were added at the N terminus of the p51 subunit. RT was expressed in BL21 (Invitrogen) and purified by nickel affinity chromatography and MonoQ anion exchange chromatography (16). Oligonucleotides used in this study were synthesized chemically and purchased from Integrated DNA Technologies (Coralville, IA). Sequences of the DNA/RNA substrates are shown in Table 1. Deoxynucleotide triphosphates and dideoxynucleotide triphosphates were purchased from Fermentas (Glen Burnie, MD). EFdA was synthesized by Yamasa Corp. (Chiba, Japan) as described previously (12). Using EFdA as the starting material, the triphosphate form, EFdA-TP, was synthesized by TriLink BioTechnologies (San Diego, CA). Concentrations of nucleotides and EFdA-TP were calculated spectrophotometrically on the basis of absorption at 260 nm and their extinction coefficients. All nucleotides were treated with inorganic pyrophosphatase (Roche Diagnostics) as described previously (17) to remove traces of PP_i contamination that might interfere with the rescue assay.

Cell-based HIV-1 Replication Assays

Peripheral blood mononuclear cells (PBMCs) were isolated from healthy donor buffy coats (purchased from the Central Blood Bank, Pittsburgh, PA) using Ficoll-Hypaque (Histopaque, Sigma-Aldrich) gradient centrifugation as described previously (18). PBMCs were stimulated with 5 μg/ml phytohemagglutinin (Sigma) in RPMI 1640 containing 10% fetal bovine serum for 48 h prior to exposure to drug and virus. After washing, the activated cells were resuspended in RPMI 1640/fetal bovine serum containing interleukin-2 (10 units/ml) and varying concentrations of the NRTIs and then were infected with HIV-1_{NL4-3} at a multiplicity of infection of 0.01. HIV-1 infection was assessed by measuring HIV-1 p24 antigen in cell-free culture supernatants obtained 7 days post-infection using an HIV-1 p24 antigen capture assay kit (SAIC, Frederick, MD).

Primer Extension Assays

Characterization of EFdA-TP as a Chain Terminator—DNA or RNA template was annealed to a 5′-Cy3-labeled DNA primer (3:1 molar ratio). To monitor the primer extension, the DNA/DNA or RNA/DNA hybrid (20 nM) was incubated at 37 °C with HIV-1 RT (20 nM) in a buffer containing 50 mM Tris (pH 7.8) and 50 mM NaCl (RT buffer). Subsequently, varying amounts of EFdA-TP or ddATP were added, and the reactions were initiated by the addition of 6 mM $MgCl_2$ to a final volume of 20 μl. All dNTPs were present at a final concentration of 1 μM. The reactions were terminated after 15 min by adding an equal volume of 100% formamide containing traces of bromphenol blue. The products were resolved on a 15% polyacrylamide 7 M urea gel. In this and subsequent assays, the gels were scanned with a phosphorimaging device (FLA 5000, FujiFilm). The bands for fully extended product were quantified using Multi Gauge software (FujiFilm), and the results were plotted as percent full extension using GraphPad Prism 4 to determine the IC_{50} for EFdA-TP and other nucleotide analogs.

Steady State Kinetics—Steady state kinetic parameters, K_m and k_{cat}, for incorporation of EFdA-MP or dAMP were determined using single nucleotide incorporation in gel-based assays under saturating substrate conditions. The reactions were car-

ried out in RT buffer with 6 mM $MgCl_2$, 100 nM T_{d31}/P_{d18} or T_{r31}/P_{d18}, and 2.5 nM RT in a final volume of 20 μl and stopped at the indicated reaction times. The products were resolved and quantified as described above. K_m and k_{cat} were determined graphically using the Michaelis-Menten equation.

Incorporation of dNTP to the Template/Primer (T/P) Possessing Either EFdA-MP (T/$P_{EFdA\text{-}MP}$) or ddAMP (T/P_{ddAMP}) at the 3′-Primer Terminus—$T/P_{EFdA\text{-}MP}$ and T/P_{ddAMP} were prepared by incubating 500 nM T_{d31}/P_{d18} with 1 μM HIV-1 RT in RT buffer and 6 mM $MgCl_2$. EFdA-TP (1 μM) or ddATP (5 μM) was added into the reaction and the mixture was incubated at 37 °C for 1 h. After incorporation of the nucleotide analogs, the T/P_{analog} was purified using the QIAquick nucleotide removal kit (Qiagen, Valencia, CA). Under these conditions, the extension of T/P to $T/P_{EFdA\text{-}MP}$ or T/P_{ddAMP} was complete. Purified $T/P_{EFdA\text{-}MP}$ (5 nM) was incubated with 20 nM HIV-1 RT in RT buffer and 6 mM $MgCl_2$. The first incoming nucleotide was added at different concentrations (0–100 μM) in the presence of the other dNTPs (1 μM). The reactions were incubated at 37 °C for 15 or 60 min.

Gel Mobility Shift Assays

Formation of RT·DNA Binary Complex—$T/P_{EFdA\text{-}MP}$ and T/P_{ddAMP} were prepared using T_{d43}/P_{d18} as described above. Purified $T/P_{EFdA\text{-}MP}$ or T/P_{ddAMP} (20 nM) was incubated at room temperature for 10 min with different concentrations of HIV-1 RT in RT buffer and 6 mM $MgCl_2$. RT was used at different concentrations to obtain RT/DNA ratios that ranged from 0.25 to 7.5. Four μl of 20% sucrose was added to each mixture in a final volume of 20 μl. The complexes were subsequently resolved on a native 6% polyacrylamide Tris borate gel and visualized as described above.

Formation of RT·$DNA_{EFdA\text{-}MP}$·dTTP Ternary Complex—Purified $T/P_{EFdA\text{-}MP}$ or T/P_{ddAMP} (9 nM) was incubated at room temperature for 10 min with 100 nM HIV-1 RT, varying amounts of the next nucleotide (1–5000 μM) in RT buffer, and 6 mM $MgCl_2$. Prior to the addition of sucrose, 150 ng/μl heparin was added, and finally the products were resolved on native 6% polyacrylamide Tris borate gels and visualized as described above.

Site-specific Fe^{2+} Footprinting Assay

Site-specific Fe^{2+} footprints were monitored on 5′-Cy3-labeled DNA templates. T_{d43}/P_{d20} (100 nM) was incubated with HIV-1 RT (600 nM) in a buffer containing 120 mM sodium cacodylate (pH 7), 20 mM NaCl, 6 mM $MgCl_2$, and EFdA-TP (1 μM) to allow quantitative chain termination. Prior to treatment with Fe^{2+}, complexes were preincubated for 7 min with increasing concentrations of the next nucleotide as indicated in Fig. 5*A*. The complexes were treated with ammonium iron sulfate (1 mM) as described previously (19). This reaction relies on autoxidation of Fe^{2+} (20) to create a local concentration of the hydroxyl radical, which cleaves the DNA at the nucleotide closest to the Fe^{2+} specifically bound to the RNase H active site.

PP_i- and ATP-dependent Excision and Rescue of $T/P_{EFdA\text{-}MP}$ and T/P_{ddAMP}

PP_i-dependent Excision of $T/P_{EFdA\text{-}MP}$ and T/P_{ddAMP}—Purified $T/P_{EFdA\text{-}MP}$ and T/P_{ddAMP} (20 nM) were incubated at 37 °C with HIV-1 RT (60 nM) in the presence of 150 μM PP_i in RT buffer and 6 mM $MgCl_2$. Aliquots of the reaction were stopped at different times (0–30 min) and analyzed as described above.

PP_i-dependent Rescue of $T/P_{EFdA\text{-}MP}$ and T/P_{ddAMP}—Purified $T/P_{EFdA\text{-}MP}$ and T/P_{ddAMP} (20 nM) were incubated with HIV-1 RT (60 nM) at various concentrations of PP_i (0–150 μM) in RT buffer and 6 mM $MgCl_2$. The assay was performed in the presence of a large excess of competing dATP (100 μM), which prevented reincorporation of EFdA-MP, 0.5 μM dTTP, and 10 μM ddGTP. After an incubation of 10 min at 37 °C, the reactions were stopped and analyzed as described above.

ATP-dependent Rescue of $T/P_{EFdA\text{-}MP}$ and T/P_{ddAMP}—Purified $T/P_{EFdA\text{-}MP}$ and T/P_{ddAMP} (20 nM) were incubated with HIV-1 RT (60 nM) in the presence of 3.5 mM ATP, 100 μM dATP, 0.5 μM dTTP, and 10 μM ddGTP in RT buffer and 10 mM $MgCl_2$. Aliquots of the reaction were stopped at different time points (0–90 min) and analyzed as described above.

Molecular Modeling

Molecular models of two reaction intermediates that involve EFdA were built as follows. 1) A model of the ternary complex of HIV-1 RT·DNA·EFdA-TP (Fig. 7*A*) was built starting with the coordinates of the crystal structure of the HIV-1 RT·DNA·TFV-DP complex. The triphosphate of EFdA-TP was built using as a guide the corresponding atoms of TFV-DP in structure with PDB code 1T05 and of dTTP in PDB code 1RTD. The coordinates of the 4′-ethynyl sugar ring were from our NMR structure of EFdA[5] showing that EFdA is in a North conformation similar to the sugar puckering observed in the crystal structure of 4′-ethynyl-2′-deoxycytidine (21). The structure of the EFdA-TP was assembled from its components using the sketch module of Sybyl 7.0 (Tripos Associates, St. Louis, MO), and minimized by the semiempirical quantum chemical method PM3 (22). The PM3 charges and the docking module of Sybyl 7.0 were used to dock the EFdA-TP at the RT dNTP-binding site (after removing the TFV-DP from 1T05) to generate the ternary complex HIV-1 RT·T/P·EFdA-TP. The final complex structure was minimized for 100 cycles using the AMBER force field with Coleman united charges on the protein and DNA molecules. 2) The model of the RT·$T/P_{EFdA\text{-}MP}$ binary complex with primer 3′-terminal EFdA-MP at the pre-translocation nucleotide-binding site (N-site) (or dNTP-binding site) (Fig. 7*B*) was built using as a starting model our crystal structure of the pre-translocation complex RT·$T/P_{AZT\text{-}MP}$ (PDB code 1N6Q). The structures of AZTMP and the base-pairing dA were replaced by EFdA-MP (built as in EFdA-TP above) and a dT, respectively, using the sketch module of Sybyl 7.0, and energy-minimized using the AMBER force field with Coleman united charges on the protein and DNA molecules.

[5] K. A. Kirby, K. Singh, E. Michailidis, B. Marchand, E. N. Kodama, E. Nagy, N. Ashida, H. Mitsuya, M. A. Parniak, and S. G. Sarafianos, unpublished data.

RESULTS

EFdA-TP Is a Highly Potent Inhibitor of HIV-1 RT—EFdA inhibits HIV-1 replication in phytohemagglutinin-activated PBMCs with an EC_{50} of 50 pM (Table 2), consistent with previously published data obtained using T-cell lines (12, 13), and data published after completion of this work (23). The antiviral potency of EFdA is at least 4 orders of magnitude greater than the clinically used adenine nucleotide analog tenofovir and over 400-fold greater than that of AZT when assessed under the same conditions (Table 2). EFdA thus appears to be the most potent nucleoside inhibitor described to date of HIV-1 replication in primary cells. It is also interesting to note that EFdA is substantially more potent than analogs lacking a 3′-OH function (EFddA and EFd4A; Table 2). No cytotoxicity was noted at 10 μM EFdA (data not shown), the highest concentration tested, suggesting an *in vitro* selectivity index of over 200,000. To better understand the molecular basis for the exceptional antiviral potency of EFdA, we carried out a series of detailed *in vitro* evaluations of the impact of the active antiviral form of EFdA, namely EFdA-TP, on DNA synthesis catalyzed by purified HIV-1 RT.

TABLE 2
Inhibition of HIV-1 replication in phytohemagglutinin-activated PBMCs by EFdA, EFdA analogs, and other NRTIs

Compound	EC_{50}[a]
	nM
EFdA	0.05 ± 0.02
EdA	11 ± 7
EFddA	570 ± 92
EFd4A	14 ± 11
Zidovudine (or AZT)	22 ± 7
Tenofovir	3300 ± 1240

[a] Values are means ± S.D. of triplicate determinations and were determined by assessment of reduction in HIV-1 p24 antigen production in infected cells as described under "Experimental Procedures."

We first compared the effect of EFdA-TP with other NRTI-TPs (ddATP, TFV-DP, AZTTP, and ddCTP) on RT-catalyzed DNA synthesis in *in vitro* primer extension assays using a nucleic acid T/P comprising a 100-nucleotide DNA template annealed to a Cy3–5′-labeled 18-nucleotide DNA primer (Table 1). As shown in Fig. 1*B*, EFdA-TP suppressed full-length DNA synthesis by RT in a dose-dependent manner. EFdA-TP was between ~1 and 2 orders of magnitude more effective at inhibiting RT-catalyzed DNA synthesis than any of the NRTIs evaluated. The IC_{50} for suppression of full primer extension by EFdA-TP was 14 nM using the longer T_{d100}/P_{d18} (Fig. 1, *C* and *D*). To confirm the high efficiency at which RT uses EFdA, we performed single nucleotide incorporation assays under steady state conditions (using T_{d31}/P_{d18} or T_{r31}/P_{d18} as a T/P, Table 1). Our results show that under these conditions the incorporation efficiency (k_{cat}/K_m) of EFdA-TP by RT is twice that for the natural dATP substrate and four times that for ddATP, primarily because of changes in the K_m of RT to these substrates (Table 3). Moreover, we found that the increase in incorporation efficiency of EFdA-TP could be even higher at different nucleic acid sub-

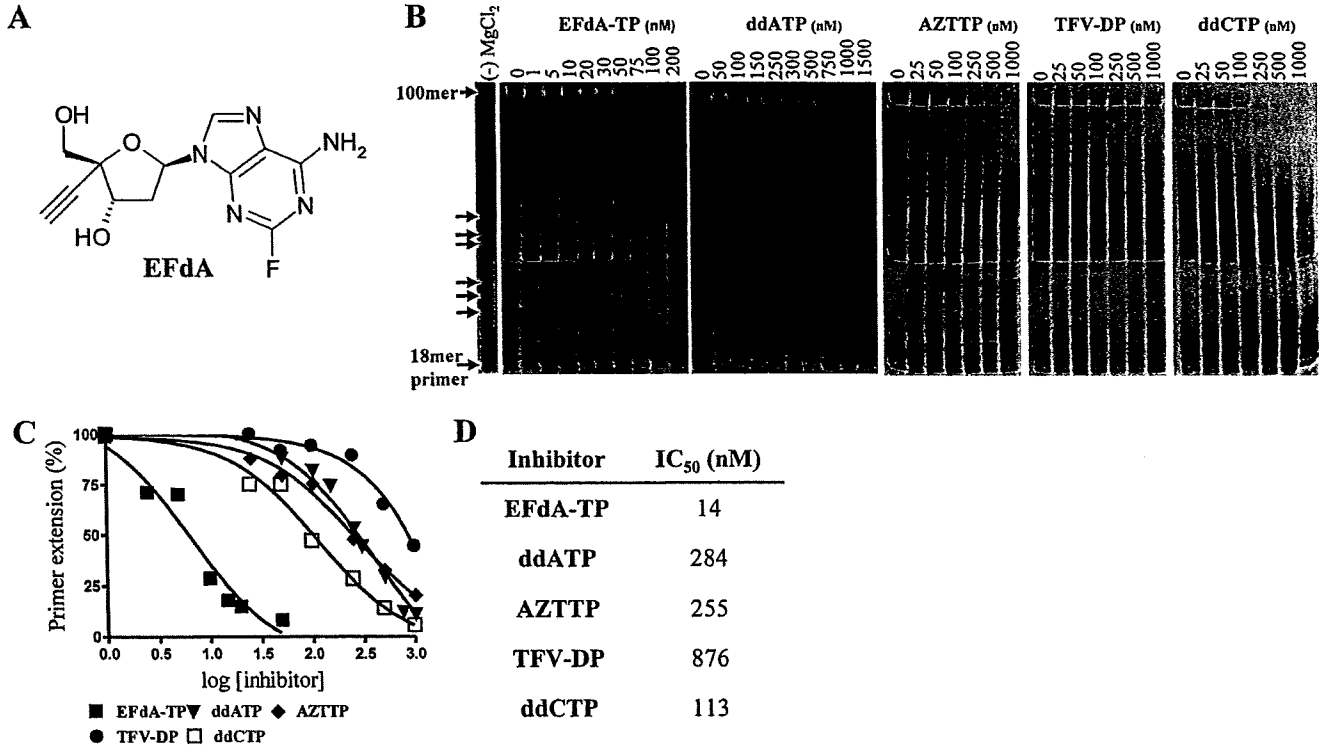


FIGURE 1. **HIV RT inhibition by EFdA-TP and other NRTIs.** *A*, structure of EFdA. *B*, primer extension by HIV-1 RT was observed in the presence of fixed concentrations of 4 dNTPs, T_{d100}/P_{d18}, and $MgCl_2$ and increasing concentrations of EFdA-TP, ddATP, AZTTP, TFV-DP, or ddCTP. The reactions were carried out for 15 min. The *arrows* denote stops of the elongating DNA chain where adenosine analogs (EFdA-TP, ddATP, or TFV-DP) were expected to be incorporated. The *first lane* is a negative control, where no $MgCl_2$ was added; it shows the length of the 18-mer primer. *C*, the 100-mer products synthesized by HIV-1 RT were quantified and plotted against increasing concentrations of various inhibitors. The data points were fitted by GraphPad Prism 4. *D*, IC_{50} values of the nucleotide analogs were determined by quantifying the percent of full extension and fitting the data points to GraphPad Prism 4 using one-site competition nonlinear regression.

TABLE 3

Steady state kinetic parameters (K_m and k_{cat}) for EFdA-MP and dAMP incorporation by HIV-1 RT

Values are means ± S.D. of triplicate determinations and were determined from Michaelis-Menten equation using GraphPad Prism 4. ND, not determined.

dNTP	T/P (DNA/DNA)				T/P (RNA/DNA)			
	K_m	k_{cat}	k_{cat}/K_m	Selectivity[a]	K_m	k_{cat}	k_{cat}/K_m	Selectivity[a]
	nM	min^{-1}	$min^{-1}\cdot nM^{-1}$		nM	min^{-1}	$min^{-1}\cdot nM^{-1}$	
dATP	73.11 ± 11	19.9 ± 0.7	0.272	1	21.3 ± 8	3.1 ± 0.2	0.145	1
EFdA-TP	39.2 ± 3	21.1 ± 0.4	0.538	2	24.1 ± 5	2.3 ± 0.1	0.095	0.7
ddATP	97.0 ± 9	15.4 ± 0.3	0.159	0.6	ND	ND	ND	ND

[a] Selectivity is the ratio of the incorporation efficiency (k_{cat}/K_m) of EFdA-MP or ddAMP over that of dAMP ($[k_{cat}/K_m]_{EFdA\text{-}MP}/[k_{cat}/K_m]_{dAMP}$ or $[k_{cat}/K_m]_{ddAMP}/[k_{cat}/K_m]_{dAMP}$).

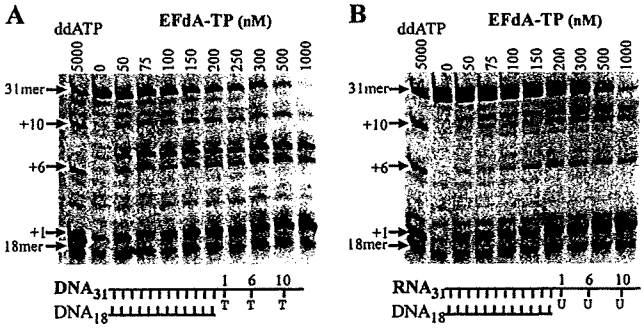


FIGURE 2. **Inhibition of DNA- and RNA-dependent DNA synthesis by EFdA-TP.** *A*, T_{d31}/P_{d18} was incubated with HIV-1 RT for 15 min in the presence of 1 μM dNTPs and $MgCl_2$ and increasing concentrations of EFdA-TP (0–1000 nM). The *first lane* (dd*ATP*) shows the inhibition of primer extension by ddATP to identify points of adenosine analog (ddATP or EFdA-TP) incorporation (*arrows*: +1, +6, and +10). *B*, the primer extension under the same conditions with an RNA/DNA substrate containing an RNA template annealed to a DNA primer (T_{r31}/P_{d18}).

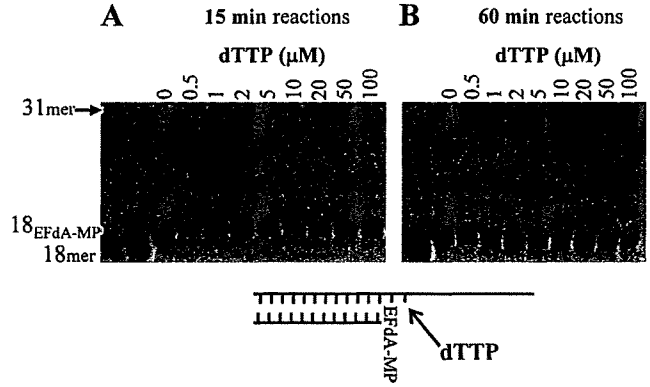


FIGURE 3. **Incorporation of dNTP on EFdA-terminated template/primer ($T/P_{EFdA\text{-}MP}$).** EFdA-TP was first incorporated at T_{d31}/P_{d18} by HIV-1 RT and purified as described under "Experimental Procedures." The incorporation of the next incoming nucleotide on $T/P_{EFdA\text{-}MP}$ was examined in the presence of HIV-1 RT and $MgCl_2$ and increasing concentrations of dTTP. All other dNTPs were at a concentration of 1 μM. The reactions were stopped after 15 min (*A*) and 60 min (*B*).

strate sequences, more than 10 times higher than dATP (data not shown).

The EFdA-TP-mediated reduction in full-length DNA synthesis was accompanied by the concomitant appearance of products corresponding to the primer extension only at the length expected for the incorporation of adenosine nucleotides (indicated by *arrows* in Fig. 1*B*). Neither ddATP nor TFV-DP provided significant accumulation of this small DNA product.

EFdA-TP Inhibits DNA Synthesis Mainly at the Point of Incorporation—The stopping patterns of DNA synthesis were different in the presence of EFdA-TP compared with other dATP analogs such as ddATP and TFV-DP (marked by *arrows* in Fig. 1*B*). Hence, we used a shorter template (T_{d31}/P_{d18} (Table 1)), which allowed unambiguous identification of the stopping sites. As expected, the inhibitory potential of EFdA (and other NRTIs) appears lower in these shorter T/P (IC_{50} for suppression of full primer extension was 104 nM) in which there are fewer opportunities for incorporation (24). This substrate allows incorporation of dA, ddA, or EFdA at positions 1, 6, and 10 (Fig. 2). The results show that EFdA-TP causes major pauses at all possible points of incorporation ((Fig. 2*A*, positions 1, 6, and 10), suggesting that EFdA-TP inhibits RT mainly as an obligate chain terminator. Notably, there was a distinct difference at position +6 of T_{d31}/P_{d18}, where we observed a strong stop not only at the point of incorporation but also at the position following (Fig. 2*A*, positions 6 and 7, respectively). These results suggest that in some cases EFdA-MP may also allow incorporation of an additional nucleotide depending upon the template sequence.

EFdA-TP inhibits both RNA- and DNA-dependent RT-catalyzed full-length DNA polymerization to comparable extents (Fig. 2). The selectivity for incorporation of EFdA-MP over dAMP ($(k_{cat}/K_m)_{EFdA\text{-}TP}/(k_{cat}/K_m)_{dATP}$) was slightly increased when we used DNA/DNA *versus* RNA/DNA template/primers (Table 3, 2 *versus* 0.7). Notably, the overall pausing pattern due to inhibition by EFdA-TP differs depending on whether the template is RNA or DNA. Inhibition differences based on type and sequence of template are currently under investigation.

Extension of EFdA-terminated T/P ($T/P_{EFdA\text{-}MP}$)—The data in Figs. 1 and 2 suggest that EFdA-TP can act as a terminator of RT-catalyzed DNA synthesis in a manner similar to that of other NRTI-TPs. We therefore examined the efficiency of nucleotide additions to primers that were synthesized to already possess a 3′-terminal EFdA-MP. Fig. 3 shows that the primer extension from EFdA-MP is very limited and is evident only at very high and nonphysiological concentrations (>50 μM) of the next nucleotide and with extended reaction times (Fig. 3*B*, 60-min incubation). It therefore appears that EFdA acts as a *de facto* chain terminator, despite the presence of a 3′-OH function.

RT Binds T/P_{ddAMP} and $T/P_{EFdA\text{-}MP}$ at Similar Efficiencies—A possible reason for the inability of RT to efficiently extend the EFdA-MP-terminated primer is that it binds $T/P_{EFdA\text{-}MP}$ with less affinity than it does a T/P lacking a 3′-terminal EFdA-MP nucleotide. We therefore used gel mobility shift assays to compare the stabilities of the binary complexes of RT with T/P possessing either EFdA-MP ($T/P_{EFdA\text{-}MP}$) or ddAMP

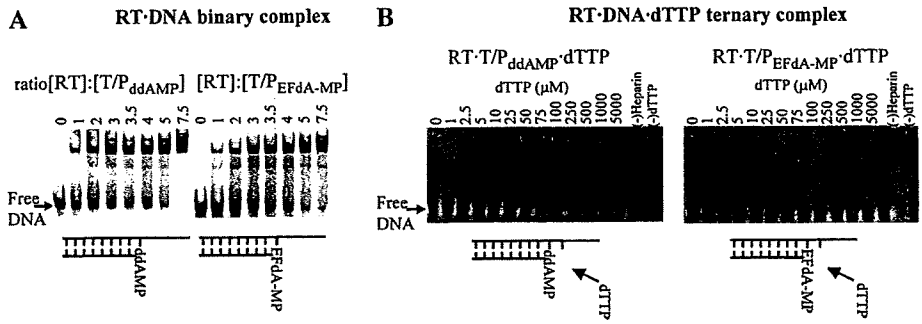


FIGURE 4. **Effect of ddA or EFdA on formation of binary and ternary complexes.** *A*, formation of a binary complex between RT and T/P$_{ddAMP}$ or T/P$_{EFdA-MP}$. Purified T/P$_{ddAMP}$ or T/P$_{EFdA-MP}$ (20 nM) was incubated with HIV-1 RT at the indicated molar ratios and resolved by nondenaturing gel electrophoresis. *B*, formation of a ternary complex between RT and T/P$_{ddAMP}$ or T/P$_{EFdA-MP}$ and incoming dTTP. The stability of the ternary complexes was analyzed by incubating 100 nM RT and 9 nM T/P$_{ddAMP}$ or T/P$_{EFdA-MP}$ in the presence of increasing dTTP concentrations and heparin, which acted as an enzyme trap. In the absence of dTTP, the T/P·RT binary complex is unstable (*lane 0*), as RT dissociates from the T/P and is trapped by heparin.

specific Fe^{2+} footprinting assay (19) to determine whether the primer 3′-terminal EFdA-MP of the RT·T/P$_{EFdA-MP}$ complex resides primarily in the pre- (N-site) or post-translocation (P-site) state in the absence and the presence of varying levels of the incoming complementary dNTP, which serves to "force" the 3′-primer terminus from the pre-translocation (N-site) to the post-translocation site (P-site). As shown on Fig. 5*A* (*left panel*), in the absence of dTTP (*first lane*), primers with a 3′-terminal ddAMP are located in both the pre- and post-translocation sites in approximately equal amounts, and the addition of the next complementary nucleotide (dTTP) forces the primer 3′-end almost entirely into the post-translocation site. In contrast, in the absence of dTTP, primers with a 3′-terminal EFdA-MP are located exclusively in the pre-translocation site (Fig. 5*A*, *right panel, first lane*). The next complementary nucleotide (dTTP) is unable to shift the position of the 3′-EFdA-MP except at very high (nonphysiological) concentrations.

(T/P$_{ddAMP}$) at the 3′-primer terminus (T/P chain terminated by EFdA-MP or ddAMP). As shown in Fig. 4*A*, RT binds T/P$_{EFdA-MP}$ with an apparent affinity comparable with that of the normal T/P (K_d for RT·T/P$_{EFdA-MP}$ = 51 nM; K_d for RT·T/P$_{ddA-MP}$ = 42 nM). This observation suggests that RT is inhibited at a downstream step in the polymerization reaction.

RT Is Unable to Form a Stable Ternary Complex with T/P$_{EFdA-MP}$ and dNTP—The next step in the DNA polymerization mechanism is the binding of the next complementary dNTP to RT·DNA, thus forming the ternary complex that precedes catalysis. To determine whether EFdA exerts its inhibitory effect by interfering with the formation of a stable ternary complex with the incoming dNTP, we used a gel-based nondenaturing electrophoresis assay (25) that detects the ternary complex formed by RT·T/P and the next complementary dNTP (in this case, dTTP). In this assay, the stability of the ternary complex is assessed by the persistence of the RT·DNA·dNTP complex upon addition of a competing heparin trap (25). As seen in Fig. 4*B* (*left panel*), the ternary complex formed by RT with T/P$_{ddAMP}$ and dTTP is quite stable. In contrast, no significant amount of ternary complex was noted in assays using T/P$_{EFdA-MP}$ even at very high dTTP concentrations (Fig. 4*B*, *right panel*).

Incorporation of EFdA-TP into DNA (T/P$_{EFdA-MP}$) Decreases Translocation of RT—The inability of RT to form a stable ternary complex with T/P$_{EFdA-MP}$ and the next complementary dNTP could arise from several factors including (i) the inability of the 3′-terminal EFdA-MP primer to efficiently translocate from the N-site (which is also the pre-translocation site) to the post-translocation primer site (P-site), thereby preventing the next incoming nucleotide from binding; and (ii) the fact that T/P$_{EFdA-MP}$ *can* translocate, but the presence of unnatural substituents in the primer 3′-terminal EFdA-MP (4′-ethynyl and 2-fluoro) may alter geometric and electronic parameters at the primer end, thus preventing efficient incorporation of the next incoming dNTP such that catalysis cannot occur. These two different possibilities place the nucleic acid at different positions/registers with respect to a site-specific landmark in RT, the metal-binding ribonuclease H (RNase H) active site, as shown in the schematic of Fig. 5*B*. We therefore used a site-

Because it is physically impossible for the incoming dNTP to bind at the N-site (dNTP-binding site) when it is occupied by the 3′-primer terminus of the non-translocated T/P$_{EFdA-MP}$, the present data demonstrate that the apparent termination of RT-catalyzed DNA synthesis upon incorporation of EFdA-MP arises from the inability of the 3′-EFdA-MP-terminated primer/template (T/P$_{EFdA-MP}$) to efficiently translocate to the P-site and allow incorporation of the next dNTP. Moreover, the latter is unable to force translocation of the 3′-EFdA-MP-terminated primer strand. The inability of the 3′-EFdA-MP-terminated primer to translocate to allow binding of the next complementary dNTP effectively prevents continued elongation of the nascent viral DNA, despite the presence of the 3′-OH on EFdA. Therefore, we propose that EFdA acts as a *translocation-defective reverse transcriptase inhibitor.*

Phosphorolytic Excision of EFdA-MP—Two major mechanisms account for HIV resistance to NRTIs (26). One is based on NRTI discrimination, where the mutant RT preferentially incorporates the natural dNTP rather than the NRTI-TP. The other major resistance mechanism involves ATP-mediated phosphorolytic excision of the incorporated chain-terminating NRTI from the 3′-end of the primer (27, 28). We and others have previously shown that for excision to occur, the 3′-end of the primer must be positioned at the pre-translocation or N-site of RT (19, 29, 30). As we have already shown, the 3′-EFdA-MP-terminated primer strand binds predominantly in a pre-translocation mode. This suggests that EFdA-terminated primers might be especially susceptible to RT-catalyzed phosphorolytic removal of the terminating EFdA-MP. To assess this possibility we carried out primer unblocking experiments using nucleic acid substrates having at the 3′-primer terminus either EFdA-MP or ddAMP (T/P$_{EFdA-MP}$ or T/P$_{ddAMP}$, respectively). The quantitation of results in Fig. 6*A*

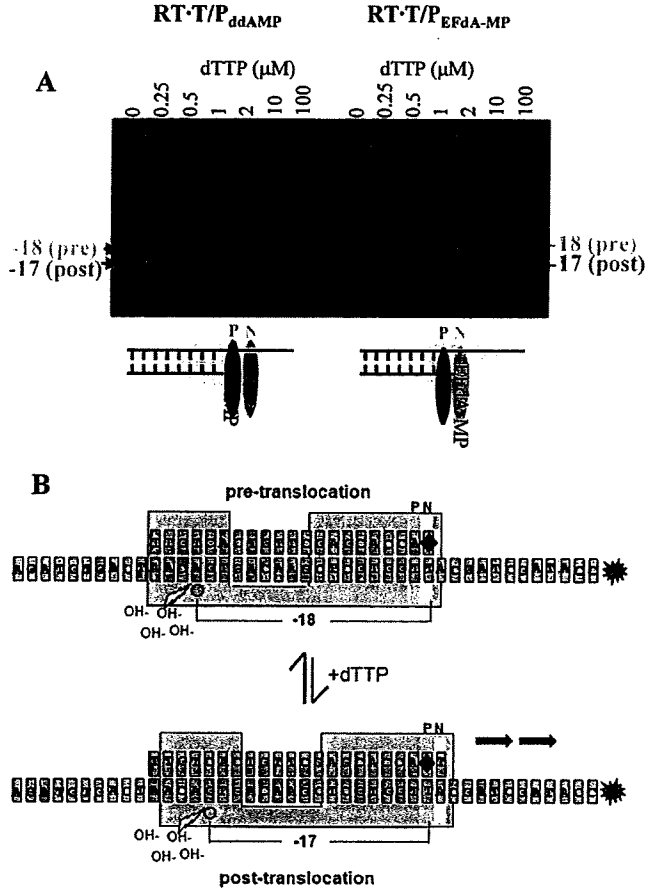


FIGURE 5. **Determination of the translocation state of RT bound to T/P$_{ddAMP}$ and T/P$_{EFdA-MP}$ template/primers.** *A*, the translocation state of RT after EFdA-MP incorporation was determined using site-specific Fe^{2+} footprinting. T/P$_{ddAMP}$ or T/P$_{EFdA-MP}$ (100 nM) with 5′-Cy3 label on the DNA template (see Fig. 5*B*) was incubated with HIV-1 RT (600 nM) and various concentrations of the next incoming nucleotide (dTTP) (as indicated). The complexes were treated for 5 min with ammonium iron sulfate (1 mM) and resolved on a polyacrylamide 7 M urea gel. An excision at position −18 indicates a pre-translocation complex, whereas the excision at position −17 represents a post-translocation complex. The scheme below the gel images indicates that in the absence of incoming dNTP, T/P$_{ddAMP}$ is bound mostly in a post-translocation state, whereas T/P$_{EFdA-MP}$ is bound in a pre-translocation state, with EFdA-MP positioned at the N-site. *B*, schematic of the excision assay. Depending on whether the 3′-primer terminus is positioned at the pre-translocation (N-site) or post-translocation (P-site) site, cleavage is observed on the 5′-labeled template strand at positions −18 or −17, respectively. The addition of varying levels of the incoming complementary dNTP serves to force the 3′-primer terminus from the N-site to the P-site.

shows that the rate of hydrolysis of ddAMP- and EFdA-MP-terminated primers was 0.5 and 1.6 min^{-1}, respectively. We also considered whether the EFdA-TP formed upon pyrophosphorolytic removal of the 3′-terminal EFdA-MP was promptly reincorporated. We tested this possibility using so-called phosphorolysis "rescue" assays, where in addition to the PP_i that would react with the EFdA-MP from the 3′-primer terminus to produce EFdA-TP, we also included a high concentration of dATP that would compete with and prevent reincorporation of EFdA-TP. In the case of high excision activity, we expected to see higher bands corresponding to rescued and extended primers. Indeed, substantially more primer extension was noted in PP_i- or ATP-mediated rescue assays using 3′-terminal EFdA-MP primers (Fig. 6, *C* and *D*, *right panels*) than in assays using primers with 3′-terminal ddAMP (Fig. 6, *C* and *D*, *left panels*). The rate for the ATP-dependent rescue of EFdA-MP-terminated primer was 0.063 min^{-1}. The corresponding rate for the ATP-dependent rescue of the ddAMP-terminated primer could not be calculated because the reaction was very slow. Collectively, these data suggest that 3′-terminal EFdA-MP is phosphorolytically excised more efficiently than ddAMP, consistent with its preferential positioning at the phosphorolysis-susceptible pre-translocation N-site.

Molecular Basis of RT Inhibition by TDRTIs—To understand the molecular basis of RT inhibition by EFdA-TP, we built molecular models of complexes that represent the following intermediates of the DNA polymerization reaction: 1) a pre-catalytic RT·DNA·EFdA-TP ternary complex and 2) a complex that corresponds to the product of EFdA-MP incorporation *prior* to translocation. Previously we had solved crystal structures representing both types of these intermediates for other NRTIs (29, 31), and the model building was guided by the structural characteristics of these complexes. Moreover, we built the EFdA sugar ring in both models in a 3′-endo (North) conformation based on our unpublished NMR experimental data, which clearly show that the equilibrium of the geometries of the EFdA sugar ring overwhelmingly favors the 3′-endo conformation (North). The RT·DNA·EFdA-TP ternary complex model was built using as a starting structure the coordinates of our RT·DNA·TFV-DP crystal structure (PDB code 1T05), not only because it is the highest resolution structure of an RT ternary complex but also because it is the only structure of RT in complex with an analog of deoxyadenosine triphosphate (31). The model of the RT·DNA·EFdA-TP complex represents the step of EFdA-TP binding to the preformed RT·DNA complex (Fig. 7*A*). It had no significant differences from the crystal structures of the ternary complexes of RT with DNA and TFV-DP (PDB code 1T05) or dTTP (PDB code 1RTD). It shows that the 4′-ethynyl of EFdA-TP is favorably positioned in a hydrophobic pocket formed by Ala-114, Tyr-115, Phe-160, and Met-184 and the aliphatic portion of Asp-185. These interactions are similar to those that have been proposed for binding of 4′-Ed4T, a related NRTI thymidine analog that also has a 4′-substitution but no 3′-OH group (32). These interactions are also consistent with the observed high efficiency of EFdA-MP incorporation by RT (Fig. 1*D*). Similar interactions stabilize the RT·DNA$_{EFdA-MP}$ pre-translocation binary complex, which has the primer 3′-terminal EFdA-MP positioned at the N-site (Fig. 7*B*, pre-translocation complex). These favorable interactions are also consistent with the enhanced binding of T/P$_{EFdA-MP}$ in a pre-translocated mode (Fig. 5).

DISCUSSION

The single most distinguishing feature of NRTIs used in HIV therapy is the absence of a 3′-OH. This property results in termination of further viral DNA synthesis upon incorporation of the inhibitor into the nascent viral DNA. We have shown (10, 13) that certain nucleoside analogs that retain the 3′-OH group can exert potent antiviral activity. One of these, EFdA, inhibits HIV-1 replication in PBMCs with a potency that is several orders of magnitude greater than that of any of the current