Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Richard A. Stanton Date

Discovery, characterization, and lead optimization of 7-azaindole non-nucleoside HIV-1 reverse transcriptase inhibitors

By

Richard A. Stanton
Doctor of Philosophy

Graduate Division of Biological and Biomedical Science
Molecular and Systems Pharmacology

Raymond F. Schinazi, Ph.D., D.Sc.
Advisor

Baek Kim, Ph.D.
Committee Member

Edward T. Morgan, Ph.D.
Committee Member

Eric A. Ortlund, Ph.D.
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

Discovery, characterization, and lead optimization of 7-azaindole non-nucleoside HIV-1 reverse transcriptase inhibitors

By

Richard A. Stanton
M.Sc., Georgia State University, 2010
B.S., Georgia Institute of Technology, 2002

Advisor: Raymond F. Schinazi, Ph.D., D.Sc.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
In partial fulfillment of the requirement for the degree of
Doctor of Philosophy
In the Graduate Division of Biological and Biomedical Science,
Molecular and Systems Pharmacology
2016

Abstract

Discovery, characterization, and lead optimization of 7-azaindole non-nucleoside HIV-1 reverse transcriptase inhibitors
By Richard A. Stanton

Current antiretroviral therapy can effectively manage HIV infections, but often carries unwanted side effects and can select for resistant viral strains. Most pressingly, it cannot cure infected patients. In an effort to identify new drug candidates, a library of 585 compounds built off a novel 7-azaindole core was screened for anti-HIV activity. Ten hits emerged with submicromolar potency and little toxicity. Two of these selected mutations on the viral polymerase reverse transcriptase (RT), in the binding pocket of non-nucleoside reverse transcriptase inhibitors (NNRTI). NNRTI are a class of allosteric inhibitors of RT, five of which have been FDA approved for clinical use. Cell-free assays verified that three of the hit compounds directly inhibited the polymerase activity of RT in a manner consistent with that of other NNRTI. The most promising compound **i**nhibited RT with submicromolar potency (IC_{50} = 0.73 μM). However, that is still several log fold less than existing NNRTI, necessitating further optimization of this compound.

Unfortunately, optimization of NNRTI using rational design approaches remains difficult in spite of the availability of > 150 solved NNRTI-bound RT crystal structures. Because of the diversity of NNRTI, docking results vary largely between receptor structures. To address this problem, more than 40 chemical descriptors were evaluated for their ability to pre-select a best receptor for NNRTI cross-docking. The receptor selection was based on similarity scores between the bound- and target-ligands generated by each descriptor. The top descriptors were able to double the probability of cross-docking accuracy over random receptor selection. Additionally, recall of known NNRTI from a large library of similar decoys was increased using the same approach.

Applying that method, the lead 7-azaindole and related analogs were docked and the resultant structures were used as a basis for free energy perturbation (FEP) calculations. FEP was used to explore potential modifications of the lead compound that could increase potency. Of the dozens analyzed, three compounds were chosen for synthesis and testing, one of which displayed a two-fold increase in potency against RT (IC_{50} = 0.36 μM). These results suggest that further optimization of the 7-azaindole NNRTI may produce more effective anti-HIV agents.

Discovery, characterization, and lead optimization of 7-azaindole non-nucleoside HIV-1 reverse transcriptase inhibitors

By

Richard A. Stanton
M.Sc., Georgia State University, 2010
B.S., Georgia Institute of Technology, 2002

Advisor: Raymond F. Schinazi, Ph.D., D.Sc.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
In partial fulfillment of the requirement for the degree of
Doctor of Philosophy
In the Graduate Division of Biological and Biomedical Science,
Molecular and Systems Pharmacology
2016

I would like to thank my wife, my mom, my dad, the rest of my family, and my friends for the support they have provided throughout my studies, I am lucky to have them in my life. I would also like to thank my advisor Raymond F. Schinazi, Ph.D., D.Sc., my dissertation committee, the MSP Program, and my many mentors and colleagues.

This work is dedicated to my daughters Summer and her sister and my nephews Jacob, Rex, Keller, Hudson, and Carter.

Glossary

AIDS: acquired immune deficiency syndrome

Autodock Vina: a docking algorithm form the Scripps Research Institute

CC_{50}: 50% cytoxic concentration

CD4: cluster of differentiation 4, a glycoprotein found on the surface of certain subsets of immune cells such as T-helper cells and macrophages that can act as a co-receptor with antigen-presenting cells, and also acts as a co-receptor for HIV virion entry

CEM: cell line derived from human T-lymphoblasts

cluster: in the context of cheminformatics, a compound set based on similarity to a given molecule (the cluster center) using a specific descriptor

$\Delta\Delta G$: difference in free energies of two different compounds

descriptor: measurable attribute of a chemical compound's molecular structure that can be compared to another compound to determine a Tanimoto similarity

DLV: delaviridine, a first generation non-nucleoside reverse transcriptase inhibitor

docking: the computational technique of aligning a small molecule into a binding pocket of a protein (receptor) in order to predict their physiologically relevant orientation to one another

EC_{50}: 50% effective concentration for reduction in viral replication

EFV: efavirenz, a first generation non-nucleoside reverse transcriptase inhibitor

ETV: etravirine, a second generation non-nucleoside reverse transcriptase inhibitor

FEP: free energy perturbation, a technique to compare the relative free energy between two closely related compounds

Gag: group specific antigen, the polyprotein that makes up the proteins of the HIV virion interior (including matrix protein, capsid, nucleocapsid, and others)

HAART: highly active antiretroviral therapy, or the use of a combination of anit-HIV drugs to treat the virus

HIV-1: human immunodeficiency virus-1

IC_{50}: 50% inhibitory concentration for reduction of enzyme activity

MACCS: Molecular Access System structural keys, a publicly available 2D descriptor

MD: molecular dynamics, a computational simulation of the movement of atoms

MOI: multiplicity of infection, ratio of virus added to cells in a culture

MTT: 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide, a dye used in an assay to assess cell viability

NAMD: Nanpscale Atomic Molecular Dynamics, a molecular dynamics program from the University of Illinois at Urbana-Champaign

NNRTI: non-nucleoside reverse transcriptase inhibitors, a class of allosteric inhibitors of HIV-1 reverse transcriptase

NNRTI-BP: non-nucleoside revere transcriptase binding pocket, the allosteric pocket on the HIV-1 reverse transcriptase enzyme that binds inhibitors

NRTI: nucleoside/tide reverse transcriptase inhibitors, a class of inhibitors of HIV reverse transcriptase that act as obligate chain terminators of viral DNA

NVP: nevirapine, a first generation non-nucleoside reverse transcriptase inhibitor

PBM: peripheral blood mononuclear cells, extracted from whole blood and including lympohcytes and monocytes

PDB: Protein Data Bank, a public repository for experimentally solved protein strucutres

pharmacophore: an atom or group of atoms on a molecule that posess a specific chemical property that can affect binding or activity

PI: protease inhibitors, a class of active site inhibitors of HIV protease

Pol: the polyprotein of the HIV viral enzymes (protease, reverse transcriptase, and integrase)

receptor: in the context of docking, the protein structure into which the new ligand is placed

RMSD: root mean squared distance, determined by the average displacement of atoms in one compound compared to another

ROC curve: Receiver Operator Characteristic curve, a plot of the percent of a sample inspected to the amount of the desired sample identified to compare multiple means of differntiation

RPV: rilpivirine, a second generation non-nucleoside reverse transcriptase inhibitor

RT: reverse transcriptase, the HIV viral enzyme that acts as a polymerase to convert single stranded viral RNA into double stranded, proviral DNA through reverse transcription

RTC: reverse transcriptase complex, the multi-component complex where rever transcription occurs, derived from the post-fusion virion found in the cytoplasm

7-azaindole: bicyclic fused 5-member pyrrole and 6 member pyridine ring compound similar to an indole with an N substitution at C-7

Tanimoto similarity: measure of similarity between two chemical compounds on a normalized scale from 0 (no similarity) to 1 (complete similarity, such as from the same molecule)

VERO: cell line derived from African green monkey kidney epithelial cells

Table of Contents

List of Figures, Tables, and Schemes

CHAPTER 1:

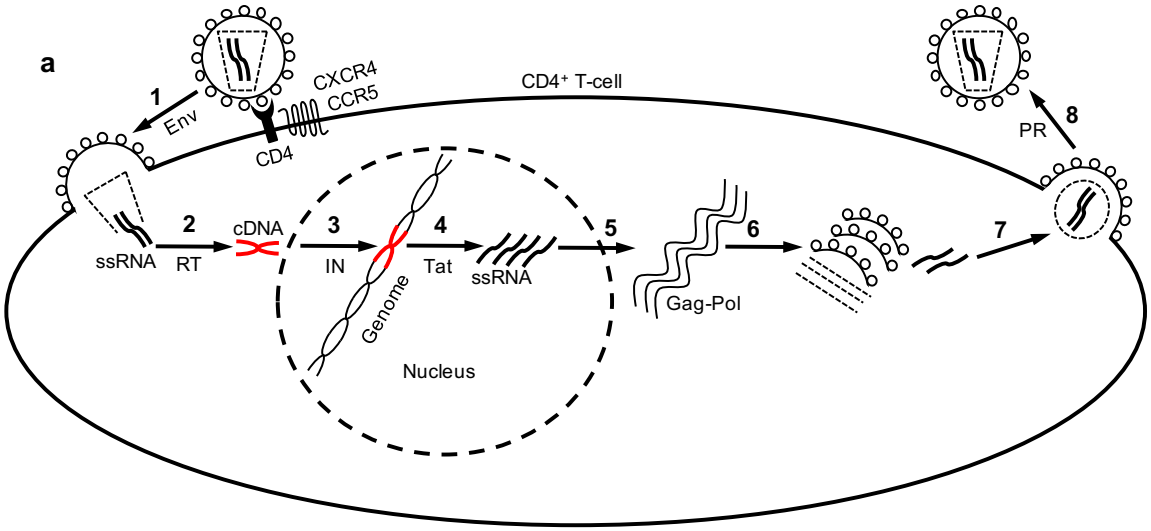
Introduction

1.1 Treatment of HIV Infections

Since the first recognized cases emerged in 1981 (1), acquired immune deficiency syndrome (AIDS) has caused more than 34 million deaths, and there are presently more than 37 million infected individuals with the human immunodeficiency virus-1 (HIV-1) worldwide (2). HIV-1 is a lentivirus that targets $CD4^+$ T-cells, leading to their rapid decline, which results in a loss of immune function as the infection progresses to AIDS (3). While there is no known curative treatment, there are 29 FDA approved antiretroviral drugs. Combinations of these drugs are used simultaneously as highly active antiretroviral therapy, or HAART, which can suppress viral load to levels below detectable limits (4). However, prolonged use of existing drugs is often associated with unpleasant side effects (5) and can select for resistant mutations (4), underscoring the ongoing need for new therapeutic agents.

Shortly after HIV-1 was first isolated from the lymph nodes of an infected patient (6), work began to discover compounds that could safely and effectively inhibit viral replication. Several classes of drugs targeting different steps of the viral replication cycle have since been discovered and put into clinical use (**Fig. 1.1**).

a



b

Step	Process	Enzymes	FDA Approved Drugs
1	Entry	Viral: gp120 and gp41; Host: CD4, CXCR4, and CCR5	*Entry Inhibitor*: maraviroc
			Fusion Inhibitor: enfuvirtide
2	Reverse Transcription	Viral: reverse transcriptase	*Nucleoside Reverse Transcriptase Inhibitors*: zidovudine, didanosine, stavudine, lamivudine, saquinavir, abacavir, tenofovir, emtricitabine
			Nonnucleoside Reverse Transcriptase Inhibitors: nevirapine, delavirdine, efavrienz, etravirine, rilpivirine
3	Integration	Viral: integrase	*Integrase Strand Transfer Inhibitors*: raltegravir, dolutegravir
4	Transcription	Viral: tat; Host: RNA Pol II	
5	Translation	Viral: rev; Host: ribosome	
6	Assembly	Viral: Gag polyprotein	
7	Budding	Viral: $p6^{Gag}$; Host: ESCRT complex	
8	Maturation	Viral: protease	*Protease Inhibitors*: ritonavir, indinavir, nelfinavir, lopinavir, atazanavir, fosamprenavir, tipranavir, darunavir

Figure 1.1. HIV replication cycle (a) and FDA approved inhibitors (b).

The first stage of the viral replication process can be inhibited by CCR5 (7) and fusion inhibitors (8), which prevent virion entry into the host cell. The next step of viral replication is mediated by the enzyme reverse transcriptase (RT), and is inhibited by two classes of FDA approved compounds, nucleoside/tide reverse transcriptase inhibitors (NRTI) (9) and nonnucleoside reverse transcriptase inhibitors (NNRTI) (10). Integrase strand transfer inhibitors act at the next step, preventing integration of the viral cDNA into the host genome (11). Protease inhibitors (PI) act at the final stages, and can halt maturation into infective virions both within and outside the cell (12).

NRTIs were the first widely used monotherapy agents to treats AIDS, and they conferred improvements in CD4 counts and other measures of disease progression (13). However, the rapid emergence of resistant strains led to virologic rebound, making prolonged use of the same drug ineffective (14). The ensuing development of PIs and NNRTIs allowed compounds with different viral targets to be used simultaneously to treat HIV infections, which increased the barrier of resistance as several mutations across the genome would be required to decrease multiple the potencies of multiple drugs (15). The advent of HAART has led to a drastic extension in the life expectancy of HIV-infected individuals, which was recently estimated to be 75 years (16). First line treatment for new HIV infections now always includes multiple drug classes to decrease resistance selection, as WHO recommends two NRTI (TDF + 3TC or FTC) and one NNRTI (EFV) (17), while the NIH suggests the same NRTI backbone but with either an integrase inhibitor (daltegravir) or a PI (darunavir) (18).

While HAART increases the genetic barrier to resistance, it does not provide an impenetrable defense against mutations. In fact, drug resistance mutants to every class of antiretrovirals have been documented in patients (4), and in clinical trials up to 9.1% of patients receiving first line HAART regimens see resistant strains emerge after 96 weeks of treatment (19). The problem of resistance is further complicated by the fact that such strains may also directly transmitted (20).

Another drawback of HAART is the side effects associated with taking multiple drugs simultaneously. Prolonged use of HAART is often associated with nephrotoxicity, bone mineral density loss, neurological problems, cardiovascular disease, and hepatotoxicity (21). Additionally, HAART-related polypharmacy can also increase drug-drug interactions with commonly prescribed medications such as statins, antibiotics, and antipsychotics (18). These side effects can lead to treatment disruptions, which can contribute to virologic rebound (22).

Most pressingly, current antiretroviral regimens cannot completely eliminate the virus from the body. Recent research suggests that the persistence of HIV is due in large part to viral reservoirs, specific tissues where viral replication continues despite drug treatment (23). These reservoirs can persist due to poor drug penetration to the lymph nodes (24) or across the blood brain barrier (25), direct cell-to-cell transfer of virions (26), or changes in viral replication dynamics occurring in certain cells, such as macrophages (27).

Because of these shortcomings, the need for new compounds to treat HIV-1 infections remains steadfast. This work focuses on the discovery, characterization,

and optimization of a new chemical class of NNRTI, which may eventually help address some of the ongoing needs related to current HAART.

1.2 Reverse Transcriptase

The target of NNRTI, reverse transcriptase (RT), is a multifunctional viral enzyme responsible for converting the single stranded RNA genome of HIV into a double stranded proviral DNA (Step 2 from **Fig. 1.1**). After fusion, entry, and uncoating of the virion, reverse transcription occurs in the cytoplasm and within the reverse transcriptase complex (RTC), which includes several viral and host proteins in addition to RT (28-30). Reverse transcription occurs discontinuously over multiple steps (**Fig. 1.2**), and involves the RNA-primed DNA polymerase, ribonuclease (RNase) H1, and DNA-primed DNA polymerase functions of RT, in addition to several strand-transfers (31). RT lacks spell checking and has a very high error rate (>1/1700 incorporated bases) (32), which, along with the high replicative output of the virus (~1-10 billion viroins produced/day) (33), contributes to the rapid emergence of drug resistant mutations in HIV.

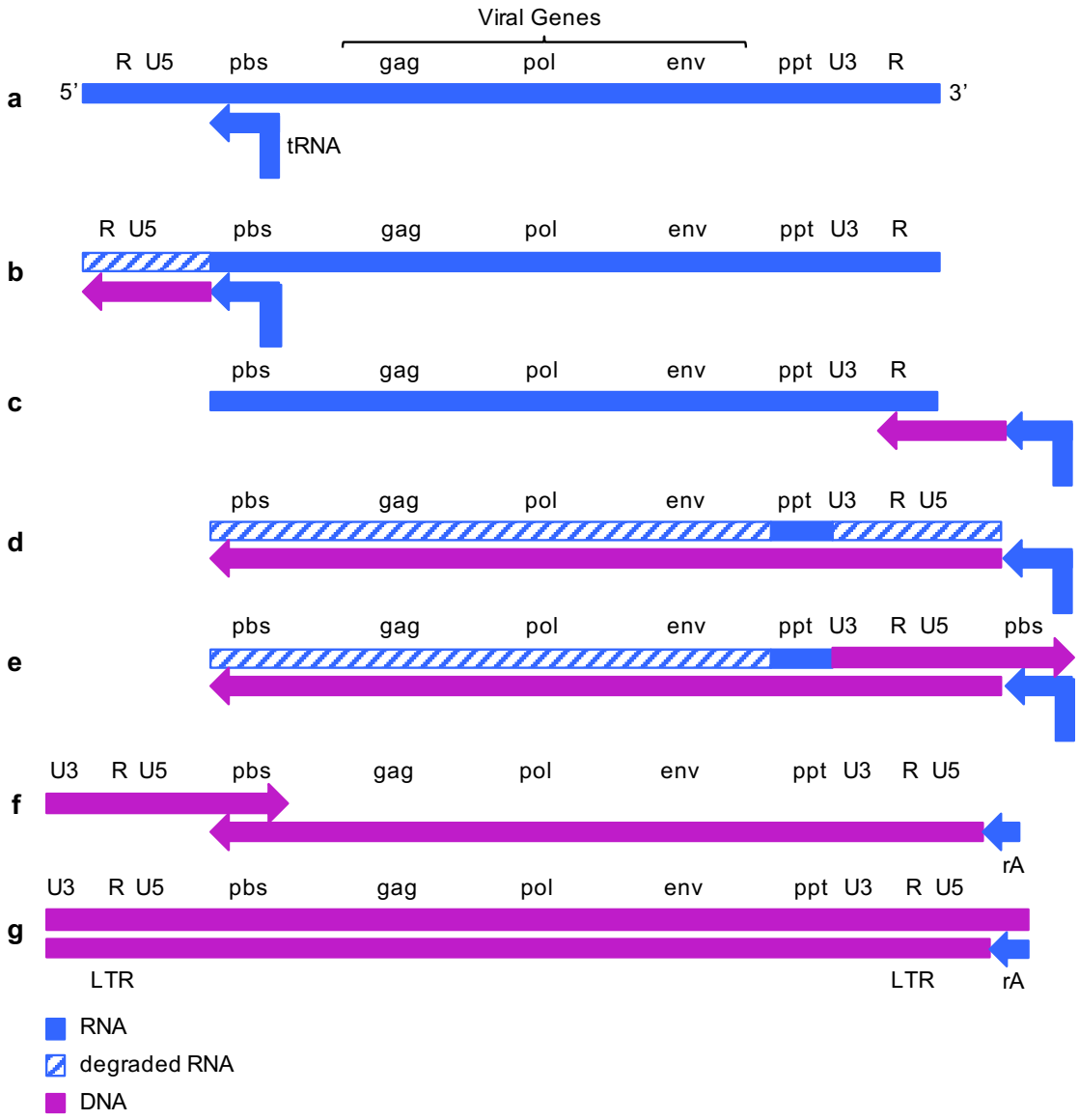


Figure 1.2. HIV reverse transcription occurs as a series of discreet steps. The initial single stranded RNA viral genome is primed with Lys3 tRNA, bound at the primer binding site (pbs) (a). Negative sense DNA strand polymerization begins at pbs, with concurrent Rnase H degradation of RNA (b). Minus strand transfer aligns nascent DNA with R sequence of 3' end of RNA (c). Minus strand DNA polymerization continues along with Rnase H degradation of RNA (d). Plus strand synthesis begins and Rnase H degrades tRNA, save rA from the 30 end (e). Plus strand transfer aligns nascent DNA strands at pbs (f). Extension of the plus and minus strand results in full viral cDNA (g).

RT is 117 kDa heterodimer consisting of an enzymatic p66 subunit and an inactive p55 subunit, which serves as a structural scaffold. It is initially translated as part of the Gag-Pol precursor polyprotein, and then processed by the viral protease during maturation into its functional subunits (34). The enzyme has separate polymerase- and RNase H-active sites at opposite ends of its structure, and like other viral polymerases is shaped like a right hand holding the nucleopolymer with a thumb, fingers, and palm subdomain of p66 (**Fig. 1.3**) (35).

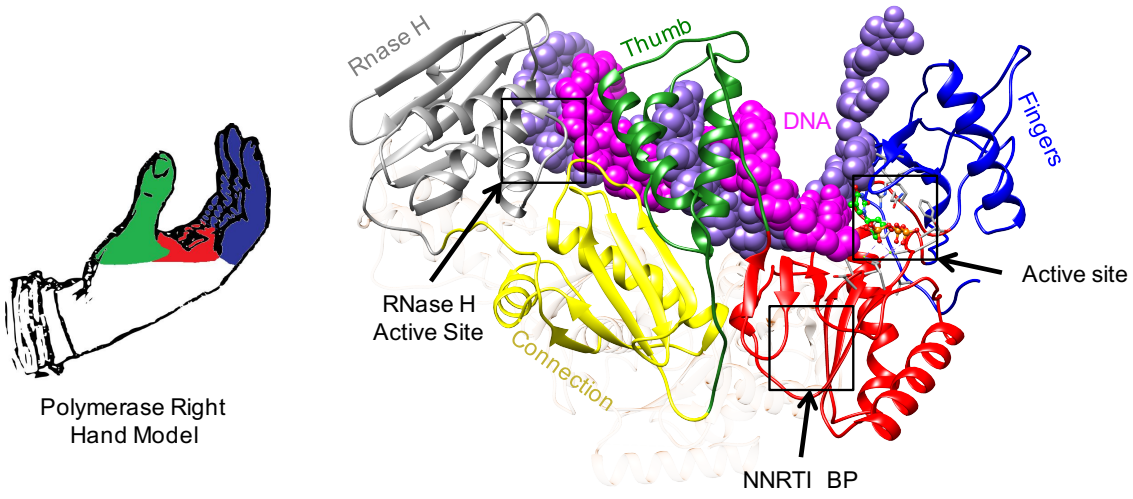


Figure 1.3. Right hand model of HIV-1 reverse transcriptase.

1.3 RT-Targeting Compounds

After HIV-1 was identified as a retrovirus (36), compounds known to inhibit reverse transcription were logically assessed as potential anti-HIV drugs. Accordingly, the first FDA-approved antiretroviral was the NRTI AZT (12). Since that time twelve other compounds compounds targeting RT have been approved for clinical use, with many more from in various stages of development (37). There are two classes of anti-RT compounds used clinically: Nucleoside/tide reverse transcriptase inhibitors (NRTI) and non-nucleoside reverse transcriptase inhibitors (NNRTI). Two other experimental classes of inhibitors have been identified (nucleotide competing reverse transcriptase inhibitors (38) and RNase H inhibitors (39)) but have not been used clinically.

NRTI are the most widely used antiretroviral agents in HIV infected persons and make up the backbone of HAART. Administered as prodrugs, NRTIs undergo intracellular phosphorylation and then compete with natural nucleotides for incorporation into the nascent DNA strand, where they act as obligate chain terminators since they lack a 3'-OH group (9). There are currently eight FDA approved NRTI, and they generally have favorable pharmacokinetic profiles resulting in high plasma concentrations (40). Like all available anti-HIV compounds, their long term use can lead to harmful side effects and resistant mutations (41).

The other class of RT-targeting compounds used clinically, non-nucleoside reverse transcriptase inhibitors (NNRTI) are the main topic of this work. They are a chemically diverse class of allosteric inhibitors of RT, five which are currently FDA approved (42). The first-generation NNRTI (nevirapine (NVP), efavirenz (EFV),

delaviridine (DLV), **Fig. 1.4**) select RT mutants such as K103N in patients, which confers resistance to all three compounds (43). Subsequent efforts to identify compounds active against mutants selected by those compounds led to the development of the two second-generation compounds (etravirine (ETV), and rilpivirine (RPV), **Fig. 1.4**) (44), which have different resistance profiles (45). The following sections address the specifics of binding, mechanism of action, mutants, and side effects of this class of drugs.

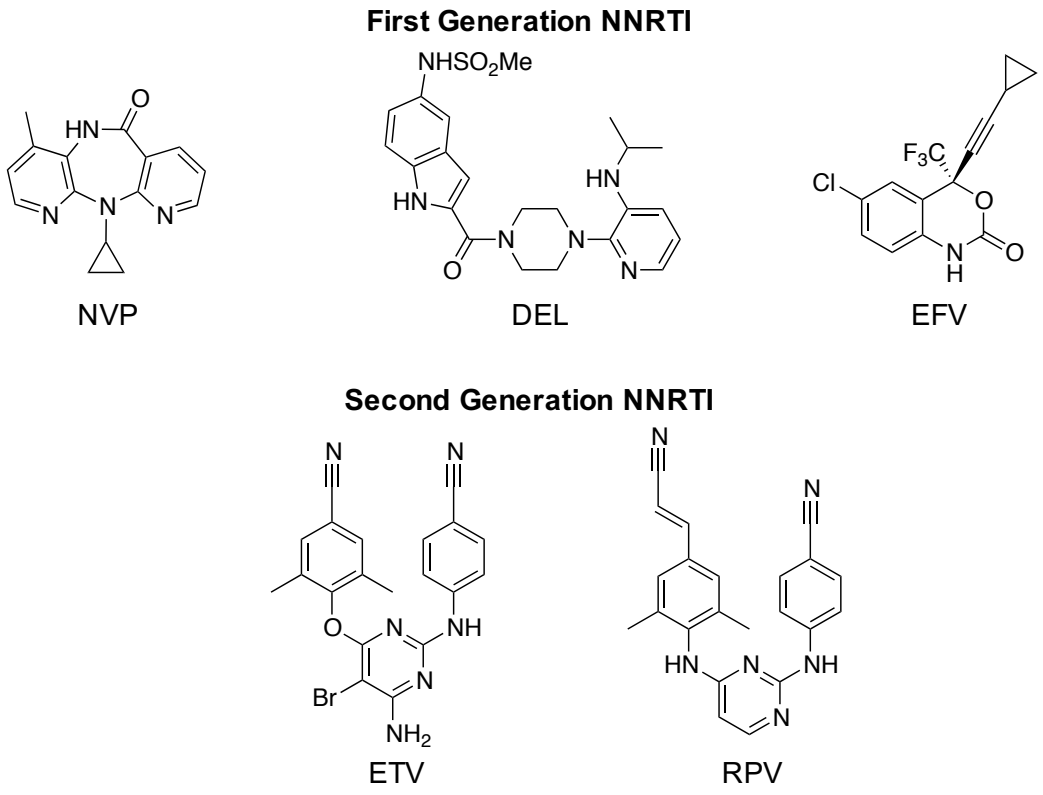


Figure 1.4. FDA approved non-nucleoside reverse transcriptase inhibitors.

1.4 NNRTI Binding

The NNRTI-binding pocket (NNRTI-BP) is found in the palm region of the p66 subunit of RT, ~ 10 Å away from the polymerase active site and 60 Å away from the RNase H1 active site (**Fig. 1.3**). It is formed between two, three-strand antiparallel β sheets, one made of the β6-β10-β11 strands and the other by the β12-β13-β14 strands. The NNRTI-BP does not exist in crystal structures solved without inhibitors (46), suggesting that binding is fully induced fit. The binding pocket is formed by the flipping of the side chains of residues Y181 and Y188 (46), and both of those as well as V106 and W229 are involved in the largely hydrophobic binding of NNRTI (47), though specific residue interaction patterns vary between compounds (**Fig. 1.5**). Hydrogen bonding can with the backbones of residues K101 (EFV (48), ETV, and RPV (49)) or K103 (DLV (50)) can also occur, though again these interactions are not generalizable. So while they all occupy the same pocket, the specific binding patterns between the protein and the small molecules vary from compound to compound, making rational design and optimization of new NNRTI challenging.

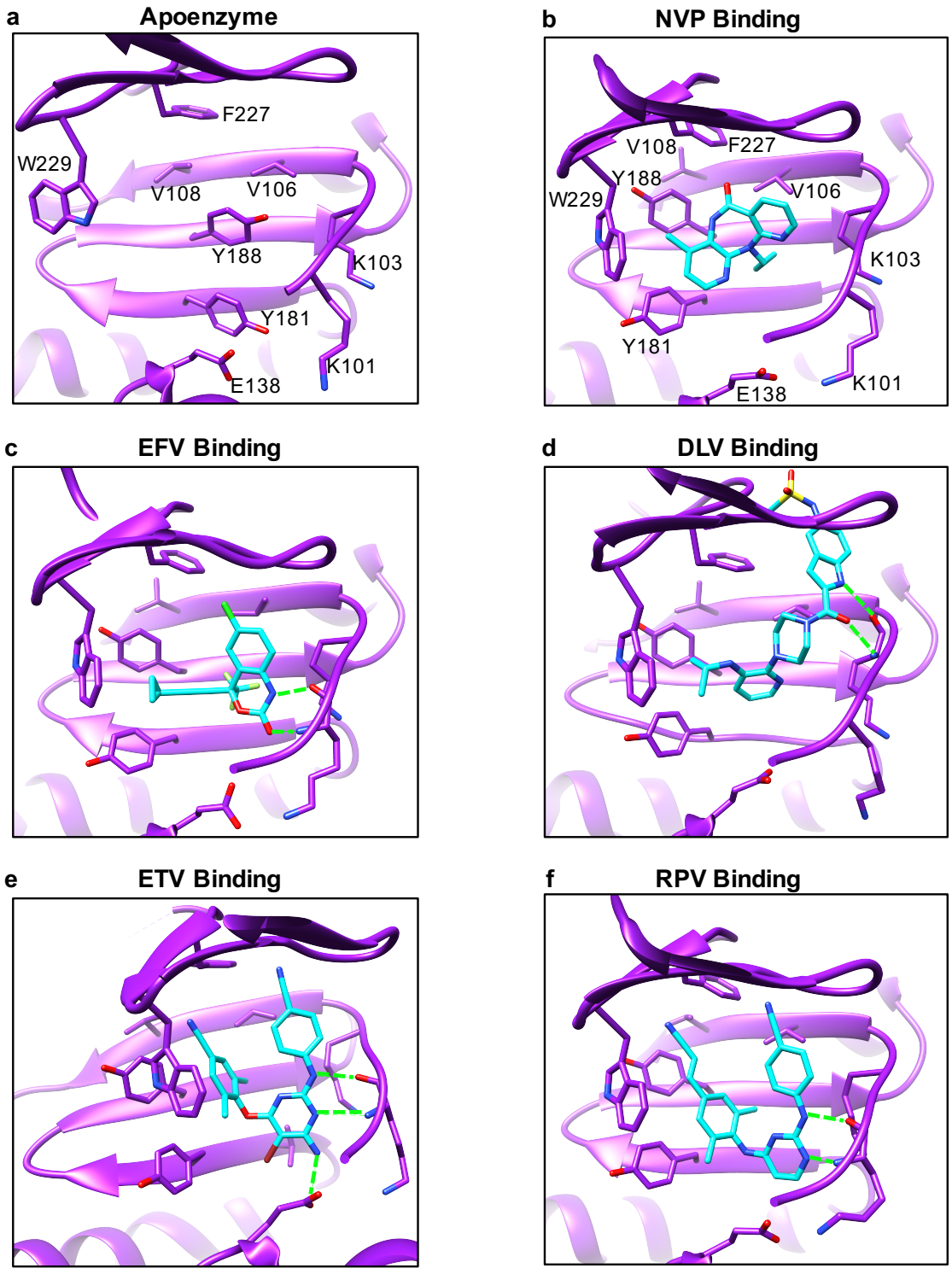
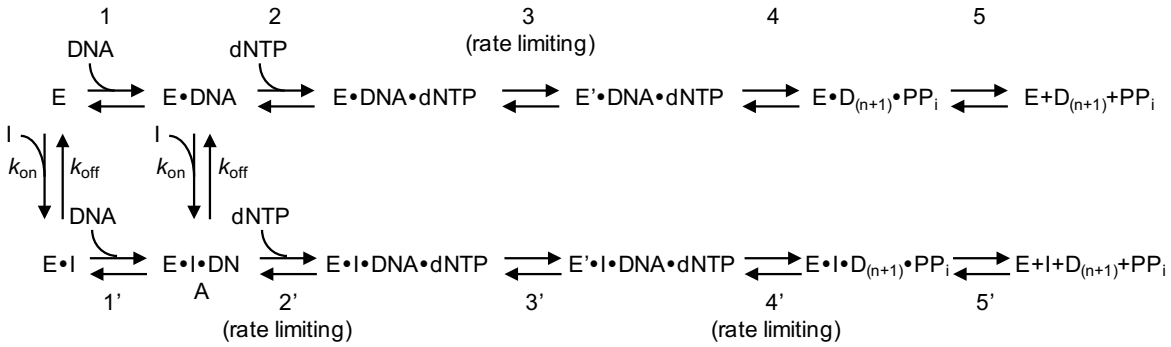


Figure 1.5. Non-nucleoside reverse transcriptase binding pocket without drug (a)or bound to FDA-approved nevirapine (b), efavirenz (c), delavirdine (d), etravirine (e), and rilpivirine (f). Hydrogen bonds are shown as green dashed lines.

1.5 NNRTI Mechanisms of Action

While it is recognized that all NNRTI bind to the same pocket, there may be multiple mechanisms by which they inhibit the enzyme, depending on both the compound and the stage of the viral replication cycle binding occurs (42).

Broadly speaking, the chief mechanism of NNRTI is inhibition of DNA polymerization by RT. This inhibition can be non-competitive and/or uncompetitive depending on the drug (43) and can affect various points in catalytic cycle of the enzyme (**Scheme 1.1**).



Scheme 1.1. HIV RT catalytic cycle with and without NNRTI (I). Studies by Spence et al. (45) and Rittinger et al. (54) suggest that the chemistry step (4') is rate limiting when RT is drug bound, while more recent work from Bec et al. (56) and Schauer et al. (57) suggest that the nucleotide binding step (2') is rate limiting.

Three hypotheses for the allosteric mechanism of inhibition have been suggested by structural studies: (1) The NNRTI-BP may act as a hinge region between the thumb and palm domains of the enzyme, and drug binding could slow down or halt the mobility of the thumb, preventing translocation and elongation of DNA. This emerged from examination of the first structure of RT in complex with NVP determined by X-ray crystallography (**Fig. 1.6**) (35); (2) A rearrangement of the conserved catalytic triad of the polymerase active site (residues 110, 185, and 186, **Fig. 1.7**) occurs upon drug binding, resulting in a rigid, catalytically inactive conformation. This was proposed after comparing several structures of NNRTI-RT complexes to that of the recently solved apoenzyme (51); (3) The conformational changes induced by NNRTI binding lead to a repositioning of the DNA away from the active site of RT, preventing successful catalysis. This conclusion was reached after the structure of RT bound to DNA was solved and compared back to the NNRTI-RT complexes (**Fig. 1.8**) (52).

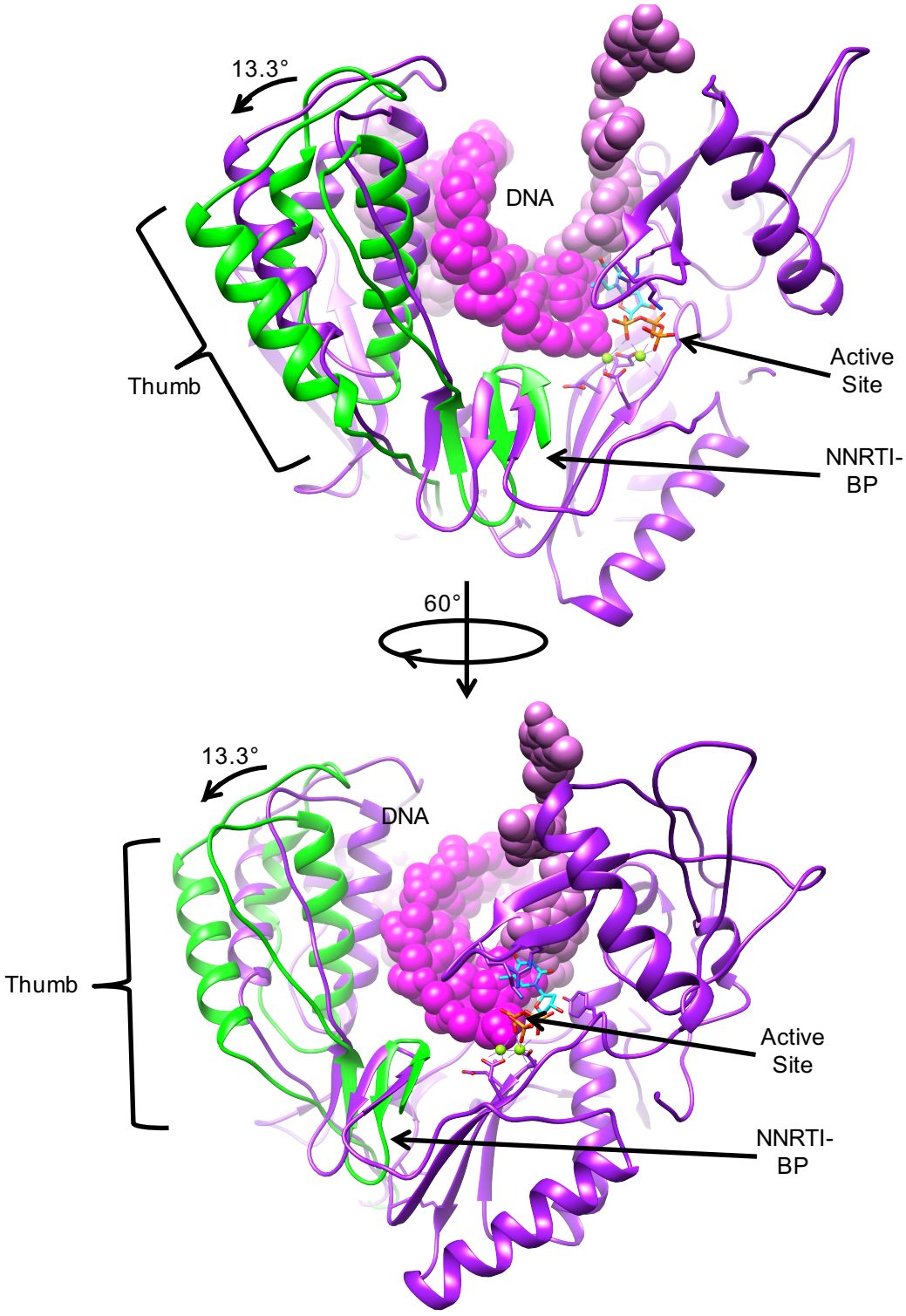


Figure 1.6. Restriction of thumb region mobility of NNRTI bound RT**.** The thumb region (green) of NVP-bound RT is superimposed on the crystal structure of the pre-catlytic tertiary complex (RT + DNA + dNTP).

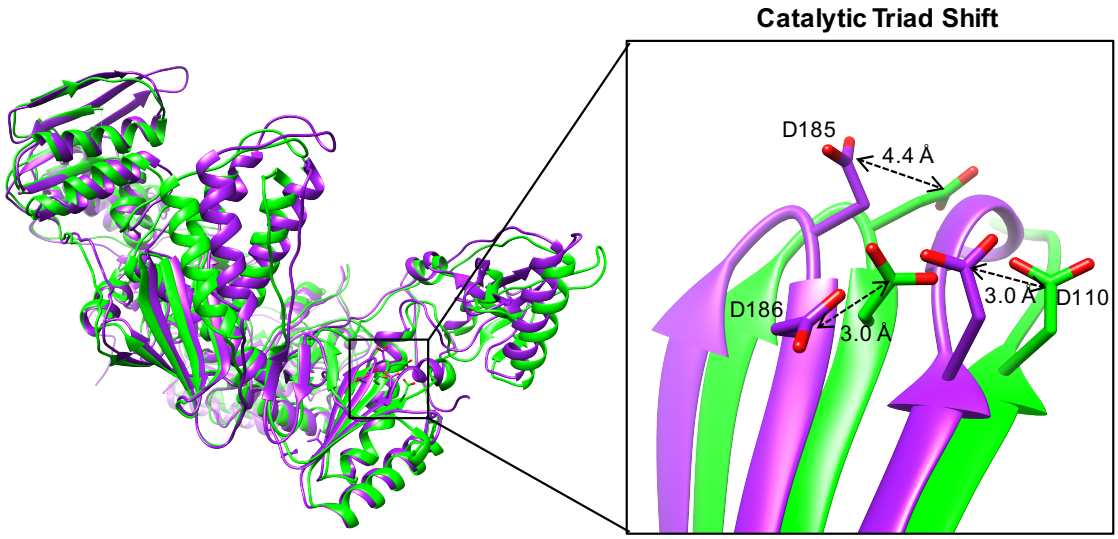


Figure 1.7. Rearrangement of catalytic triad upon NNRTI binding. NVP-bound RT (green) aligned to crystal structure of apoenzyme shows rearrangement of side chains of the residues of D110, D185, and D186, the catalytic triad (inset).

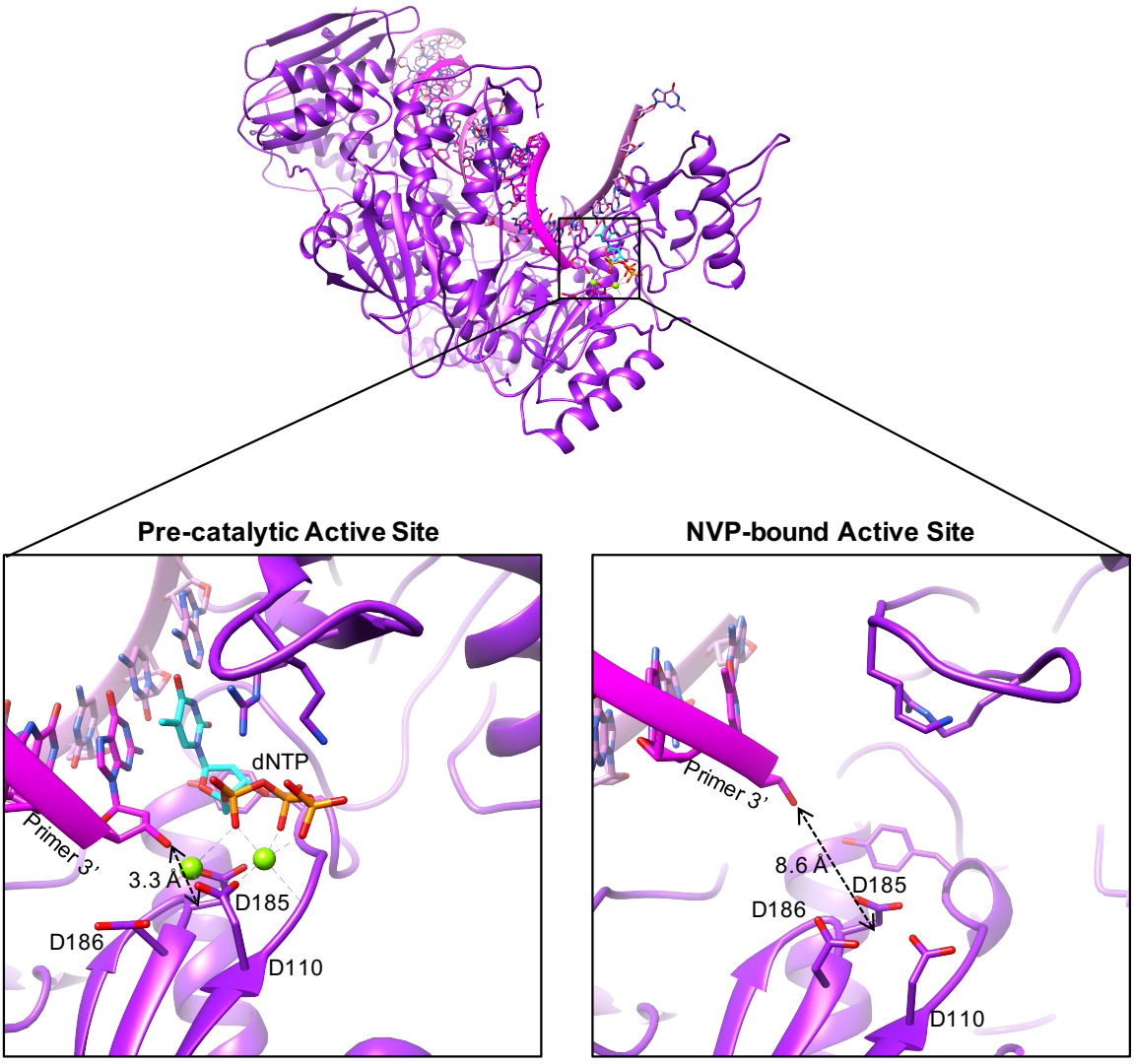


Figure 1.8. Movement of primer away from active site upon NNRTI binding. NVP-bound RT-DNA complex has the 3' end of the primer moved > 5 Å away from the enzyme active site (inset).

Additional structural research has suggested that these hypotheses may not be mutually exclusive. The first two hypotheses, which both suggest that the drugs decrease enzyme dynamicity, were supported by hydrogen-deuterium exchange experiments with RT that showed that EFV-bound enzymes exhibit less mobility in solution than the apoenzyme (53; 54). Later, the crystal structure of NVP bound to an RT-DNA complex suggested that both the second and third hypothesis could co-exist, as the active site featured a rearranged YMDD motif and the 3'-end of the primer strand of DNA was pulled away from the active site (**Figure 1.8**), preventing incoming nucleotides from binding to the crystallized structure (55).

Studies of pre-steady state RT kinetics have also suggested several possible mechanisms of NNRTI action. Early experiments suggested that NVP inhibits only the chemical step of RT-catalyzed DNA polymerization (**Scheme 1.1**), with little effect on the binding affinities for DNA or incoming nucleotides (46; 56). Later work not only showed inhibition of the chemical step, but also suggested that NNRTI-RT-DNA complexes (with NVP, EFV, or DLV as the NNRTI) actually have increased affinity for incoming nucleotides (57).

In contrast to those results, isothermal titration calorimetry (58) and fluourescent (59) studies have instead suggested that NVP and EFV inhibit the binding of incoming nucleotides.

Even simple measurements of inhibition of RT polymerase activity by NNRTI can vary with experimental conditions. Changing the DNA sequence of the primer/template in cell-free RT inhibition assays have been shown to alter the measured potencies of NVP, DLV, and EFV ~10 – 400 fold, depending on the drug

(42). And when using an RNA/DNA hybrid instead, the potency of NVP decreased by 11-fold while increasing 2-fold for EFV (60).

Other studies have shown that NNRTI can affect aspects of RT activity aside from just polymerization as well. NVP, EFV, and ETV can all inhibit RNase H activity with nanomolar potencies (60) and can also affect strand transfer dynamics (61). Some NNRTI have even been shown to act at different stages in the viral replication cycle beyond just reverse transcription, as EFV and ETV were both able to enhance processing of Gag-Pol, which led to a decrease in the release of virions (62). And an experimental compound, IQP-0410, was found to act as an NNRTI while also inhibiting viral entry by targeting the interactions between HIV envelope protein and the host cell CD4/chemokine receptors (63).

Structural, kinetics, and biochemical inhibition studies have suggested multiple, sometimes contradictory mechanisms of NNRTI action. While this makes NNRTI difficult to characterize and complicates the rationale design of new compounds, using multiple mechanisms of inhibition enhances their value as drugs.

1.6 NNRTI Resistance

Given the various binding patterns and mechanisms by which individual NNRTI work, it is unsurprising that there are multiple, compound-specific, resistant RT mutants that can emerge during treatment. At least 19 mutants are considered clinically significant and can confer resistance to NNRTI *in vivo*, and more than 35 others have been identified that can act as accessory mutations to enhance resistance in patients or can be selected *in vitro* (**Table 1.1**) (64).

Table 1.1. HIV-1 RT NNRTI-resistant mutants

Wild Type Residue	Drugs			
	NVP	EFV	ETR	RPV
V90	I	I	I	I
A98	G	G		
L100	V	IV	IV	IV
K101	EPHQ	EPHQ	EPH	EP
K103	NSHTR	NSHR		
V106	AMI	AMI	I	I
V108	I	I		
I132	ML	ML		
E138			KAQGR	KAQGR
V179	DEL	DEL	DEFT	DFTL
Y181	CIV	CIV	CIV	CIV
Y188	LCH	LCH		L
G190	ASEQCTV	ASEQCTV	EQ	EQ
H221				Y
P225	H	H		
F227	L	L	C	C
M230	LC	LC	LC	LC
L234	I	I		
P236**				
K238	TN	TN	N	
Y318	F			
N348	I	I		

*Clinically significant mutants listed in red.
** P236L confers resistance to DLV, which is no longer in clinical use.

While all of the prominent mutants occur in and around the NNRTI-BP, they can act through different mechanisms. These can be broadly categorized into three distinct, but not mutually exclusive groups: (1) Direct binding mutants, such as Y181C, change residue side chains that interact with bound NNRTI (65); (2) Indirect binding mutants, like V108I, which occur on residues that do not have direct interactions with bound NNRTI but instead perturb the orientation of side chains that do (66); (3) Functional mutations, such as K103N, which involve side chains that do not affect NNRTI binding but instead mechanistically disrupt inhibition (59).

The first class of NNRTI-resistant mutants, those that directly affect binding, include some of the first recognized in patients, such as Y181C (67), which rapidly emerged after monotherapy treatment with NVP (68). The side chain of Y181 is involved with π-π stacking with the methyl-pyridine ring of NVP (**Fig. 1.9**), an interaction that is lost when the residue is mutated to cysteine. This results in an of ~500 fold decrease in binding affinity (69) and a concomitant ~100-fold loss of potency by NVP (68).

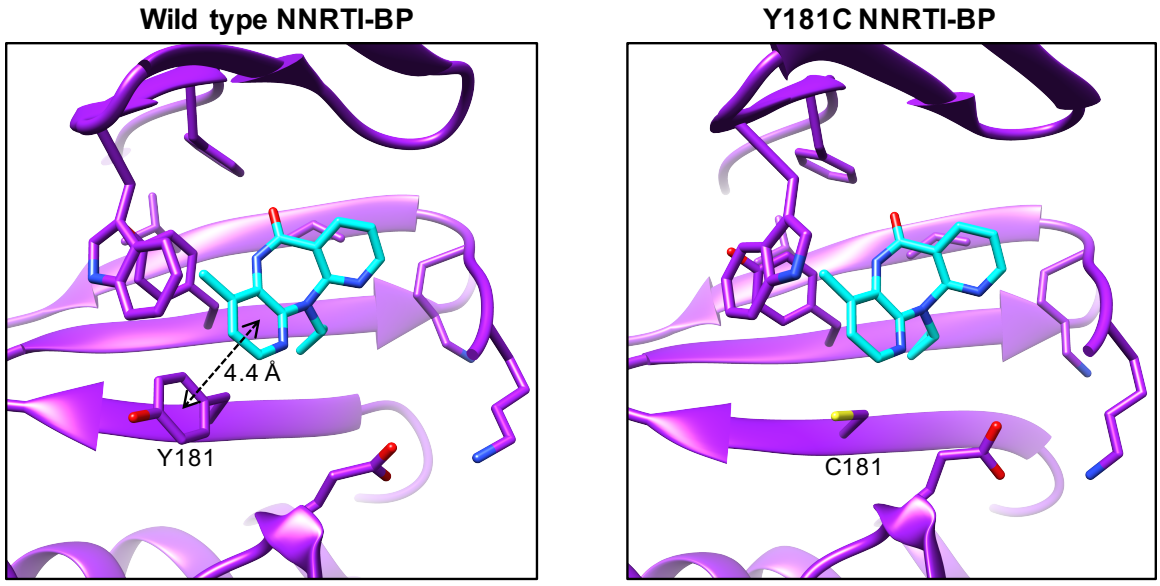


Figure 1.9. Crystal structure of RT mutant Y181C shows loss of important π-π interaction of tyrosine side chain with bound NVP.

The second class of NNRTI-associated mutants indirectly affect binding, and is exemplified by V108I. The side chain of V108 is located > 7 Å away from bound NVP (**Fig. 1.10**), well beyond van der Waals interaction distance. The V108I mutant, however, still causes a 10-fold reduction in NVP potency (70). This is believed to be due to the resultant changes in the conformation of the Y181 and Y188 side chains (66), which alters NVP binding. In addition to NVP, this mutant also occurs in patients treated with EFV (41).

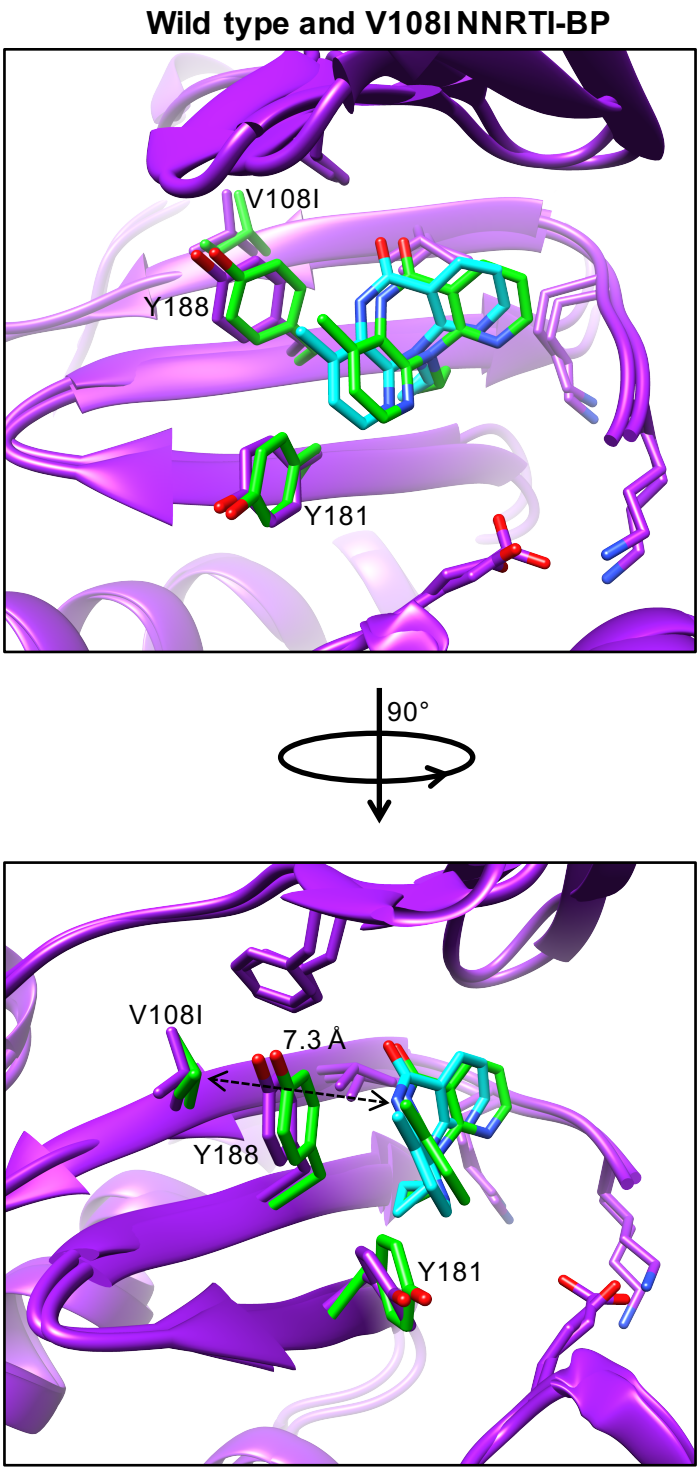


Figure 1.10. Crystal structure of RT mutant V108I shows repositioning of Y181 and Y188 side chains that are important for NVP binding. 108 residue is > 7 Å away from NVP and does not directly interact with bound drug.

The third general class of resistant mutants are resistant to NNRTI but do not drastically affect binding affinity. This class includes the most common NNRTI-associated mutant, K103N (45). The potency of EFV to K103N decreases by > 1000-fold versus wild type, yet the binding affinity is only 1.2-fold less (59). The mechanism(s) by which K103N and other mutants of this class cause resistance is still a matter of intense research (70). The K103N mutant was initially believed to increase stabilization of the apoenzyme via a hydrogen bond formed between N103 and Y188, which would prevent the flip of the latter residue necessary to open the NNRTI-BP (**Fig. 1.11**) (71). This initial hypothesis was also supported by studies of enzyme kinetics that showed decreased k_{on} rates for both NVP and EFV in the mutant (59). More recent studies have pointed to an alternate mechanism, suggesting that upon NVP or EFV binding K101 forms a salt-bridge with E138 that stabilizes an open form of the enzyme, causing it to slide along the primer-template and preventing successful catalysis (59). This theoretical mechanism would be disrupted in the K103N mutant, which pulls the K101 side chain away from E138 (48). The same K101-E138 salt bridge would also be disrupted by an E138K mutation, which also causes only small changes in drug binding affinity and emerges in patients treated with the second-generation drugs ETR and RPV, (45; 72).

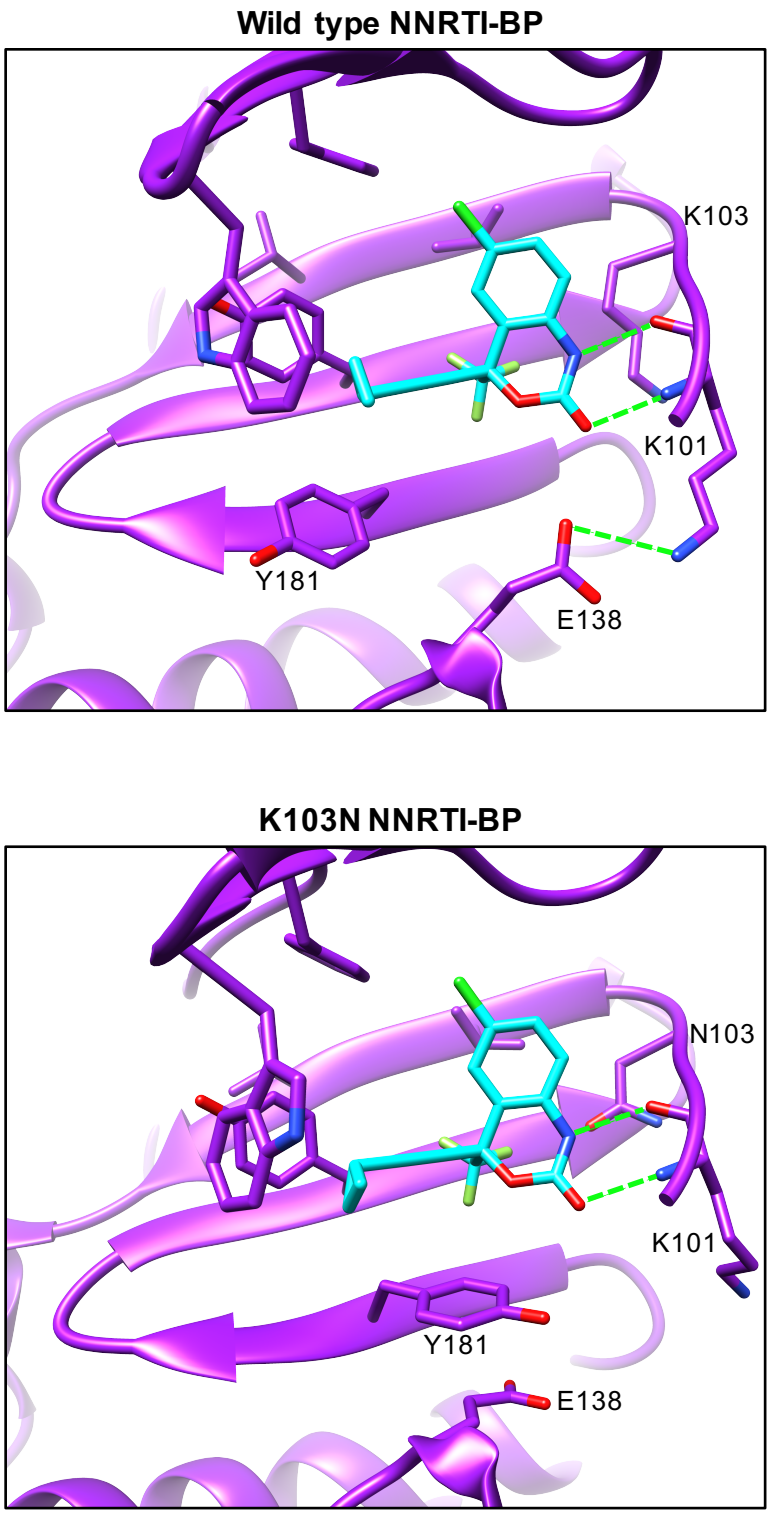


Figure 1.11. K103N RT mutant does not directly affect binding of EFV, but does disrupt hydrogen bond (shown as green dashes) between E138 and K101, which is believed to be important for inhibition.

In general, NNRTI-associated mutants are especially problematic because they have only a small impact on viral fitness (73) and single mutations can confer resistance to multiple drugs from the class (74). While these mutants do show slightly reduced replicative capacity versus the wild type virus in the absence of drugs, they can be >10 times more fit when tested with clinically relevant concentrations of NNRTI (73). Given that the commonly mutated Y181 residue within the NNRTI-BP exists as a cysteine or isoleucine in the wild type forms of other primate lentiviruses, it follows that mutations to such residues only has a minimum effect on viral fitness (75).

While the low genetic barrier of resistance to the drugs is certainly problematic, the NNRTI-specific patterns of resistance do offer the potential to treat individual mutants with tailored drug regimens. The impetus behind the development of the diarylpyrimidine compounds (including ETR and RPV) was to combat the K103N and Y181C mutants, which conferred resistance to both NVP and EFV (10). Both ETR and RPV compounds are active against the individual and dual K103N/Y181C mutants, yet themselves select for E138K in patients (45). However, E138K and Y181C are antagonistic, as a virus with both mutations becomes unfit (76). Furthermore, NVP and EFV maintain potency against E138K (77), suggesting that an NNRTI-combination therapy could potentially subdue drug resistance.

1.7 Side Effects of NNRTI

While the emergence of resistant mutants is a drawback to use of NNRTI, side effects can also limit their effectiveness and lead to poor adherence, which can lead to viral rebound (21).

Central nervous system (CNS) toxicity is the most common side effect of NNRTI. But that is largely because EFV is the most commonly used NNRTI, and more than 50% of patients on the drug report CNS-related symptoms such as headaches, amnesia, abnormal dreams, hallucinations, and even suicide ideation; leading to 6-8% to discontinue use (78). EFV can also cause neural tube defects and other problems in fetal development (79) and is therefore not recommended for use by women of child bearing age (17; 18). Additionally, the FDA trial of the experimental NNRTI IDX-899 were halted at Phase IIb after four patients experienced seizures (80). NVP has been associated with nightmares and vivid dreams, but at much lower rates than EFV (81), and it is in fact recommended for inhibition vertical-transmission (17; 18).

Other side effects that can occur with NNRTI are generally milder. Skin rashes and short-term hepatoxicity have been reported in 5% of patients receiving NVP (82). The second generation NNRTI, ETV and RPV, have also been associated with milder side effects such as skin rashes and gastrointestinal problems (83), but these compounds have lower rates of toxicity-related discontinuation versus EFV in two clinical trials (84).

Drug-drug interactions are also a concern for patients receiving NNRTI. All NNRTI used clinically are substrates of cytochrome P450 enzymes, primarily 2B6 and 3A4, and can act as both inhibitors and inducers of the enzymes (18). These can lead to changes in the bioavailable concentrations of both the NNRTI themselves as well as common drugs such as antacids, antidepressants, statins, birth control and other antivirals (18). Given the high rates of complications and co-infections such as

HCV that can occur with HIV, these drug-drug interactions can limit the use of NNRTI (84).

1.8 NNRTI and Viral Reservoirs

Recently published work suggests that the lymph nodes are an important viral reservoir that contributes to the persistence of HIV (23). While data on other NNRTI is limited, it has been shown that the concentration of EFV in lymphoid tissue of HIV infected patients is > 99% lower than in the blood, allowing ongoing viral replication to occur (23; 24). Like EFV, the second generation NNRTI ETR and RPV also have > 99% protein binding (83), which suggests that these compounds may also have trouble penetrating lymph nodes at effective concentrations. NVP does have decreased protein binding (~60%) (85), yet since it is not known to eliminate HIV from the body it is most likely not found in effective concentrations in lymph nodes.

1.9 Opportunities for NNRTI Development

The value of NNRTI was clearly demonstrated over 20 years ago when patients treated with a combination of NVP and two NRTI experienced a sustained suppression of HIV in one of the earliest clinical trials of HAART (86). Since that time four additional NNRTI have been FDA approved, but concerns over resistant mutants and side effects when compared to protease or integrase inhibitors led the NIH to move HAART regimens containing NNRTI from the Recommended to the Alternative category of first-line HIV treatment (18).

Unlike other classes of anti-HIV compounds, NNRTI are not analogs of natural enzyme substrates or peptides, which suggests that they could potentially

have superior specificity with limited side-effects versus the other drugs. Thus there is a need for new NNRTI, but in order to reach the clinic any new compound would have to offer improvements with regards to activity against resistant mutants, toxicity, and penetration into viral reservoirs. The balance of this paper details the first steps taken towards achieving those ends, focusing on the discovery, characterization, and initial optimization of a novel NNRTI chemotype built off of a 7-azaindole scaffold.

CHAPTER 2:

Discovery and Characterization of 7-azaindole Non-Nucleoside HIV-1 Reverse Transcriptase Inhibitors

Portions of the research presented in this chapter are included in a manuscript submitted for publication in *Bioorganic & Medicinal Chemistry Letters*.

Richard A. Stanton, Xiao Lu, Mervi Detorio, Catherine Montero, Emily T. Hammond, Maryam Ehteshami, Robert A. Domaoal, James H. Nettles, Michel Feraud and Raymond F. Schinazi.

"Discovery, characterization, and lead optimization of 7-azaindole non-nucleoside HIV-1 reverse transcriptase inhibitors."

Contributions to this chapter:

Richard A. Stanton cloned and purified the enzymes, performed all cell-free RT inhibition assays and *in silico* calculations, and wrote and edited the chapter.

Mervi Detorio and Catherine Montero performed the cell-based antiviral, cytoxicity, and resistance selection assays.

Emily T. Hammond, Maryam Ehteshami, and Robert A. Domaoal contributed training and assistance towards the cloning and purification of the enzymes and cell-free RT inhibition assays.

James H. Nettles assisted with data analysis and editing of the chapter.

Michel Feraud supplied the 7-azaindole compound library.

Raymond F. Schinazi conceived of the project, provided all experimental resources, and assisted editing of the chapter.

2.1 Abstract

A library of 585 compounds built off a 7-azaindole core was screened for anti-HIV-1 activity, and ten hits emerged with submicromolar potency and therapeutic index >100. Of these, three were identified as nonnucleoside reverese transcriptase (RT) inhibitors and were assayed against relevant resistant mutants. Lead compound **8** inhibited RT with submicromolar potency (IC_{50} = 0.73 μM) and also maintained activity against the clinically important RT mutants K103N and Y181C (IC_{50} = 9.2, 3.5 μM) in cell-free assays. These results highlight the discovery of a unique scaffold with the potential to move forward as next-generation anti-HIV-1 agents.

2.2 Introduction

Since the first recognized cases emerged in 1981, AIDS has caused more than 34 million deaths and there are currently more than 37 million individuals infected with HIV-1 worldwide (2). There are 29 FDA approved antiretroviral drugs, and when used in multi-drug cocktails (called highly active antiretroviral therapy, or HAART) viral load can be suppressed to below detectable limits (4). However, there is no known curative treatment and prolonged use of existing compounds is often associated with unpleasant side effects(5) and can select for resistant mutations (41), underscoring the need for new drugs.

As part of the ongoing effort to identify novel antiretroviral compounds, a library of small molecules built off a 7-azaindole (pyrrolo[2,3-b]pyridine) core (**Fig. 2.1**) was screened for anti-HIV-1 activity. This versatile scaffold is a biostere of natural purines (87) and the motif is found in drugs used to treat the influenza virus

(88), autoimmune disorders (89), and cancer (90). The library tested here included 585 compounds with a variety of substitutions off of six positions on the core (**Fig. 2.1**) and molecular masses ranging from 133 – 530 Da. This work describes the screening, characterization, and mutant profiles of these compounds.

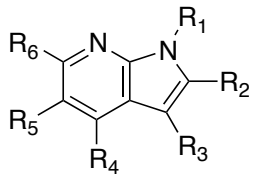


Figure 2.1. 7-azaindole core has six available positions for substitutions.

2.3 Results and Discussion

The compounds were screened for antiviral activity in physiologically relevant primary human peripheral blood mononuclear (PBM) cells infected with HIV-1_{LAI} as previously described (91). These experiments revealed a broad range of activity, including twenty compounds exhibiting submicromolar potency, per their 50% effective concentration for viral reduction (EC_{50}, **Fig. 2.2a**).

The compounds were simultaneously screened for toxicity against the PBM cells as well as against CEM (human T-lymphoblastoid derived cell line) and VERO cells (derived from African green monkey kidney epithelium) using an MTT proliferation assay (92). The 7-azaindoles were largely nontoxic, as 73% of the compounds tested were not cytoxic (as defined by the 50% cytoxic concentration (CC_{50})) at the highest concentration tested (100 μM) against PBM cells (**Fig. 2.2b**).

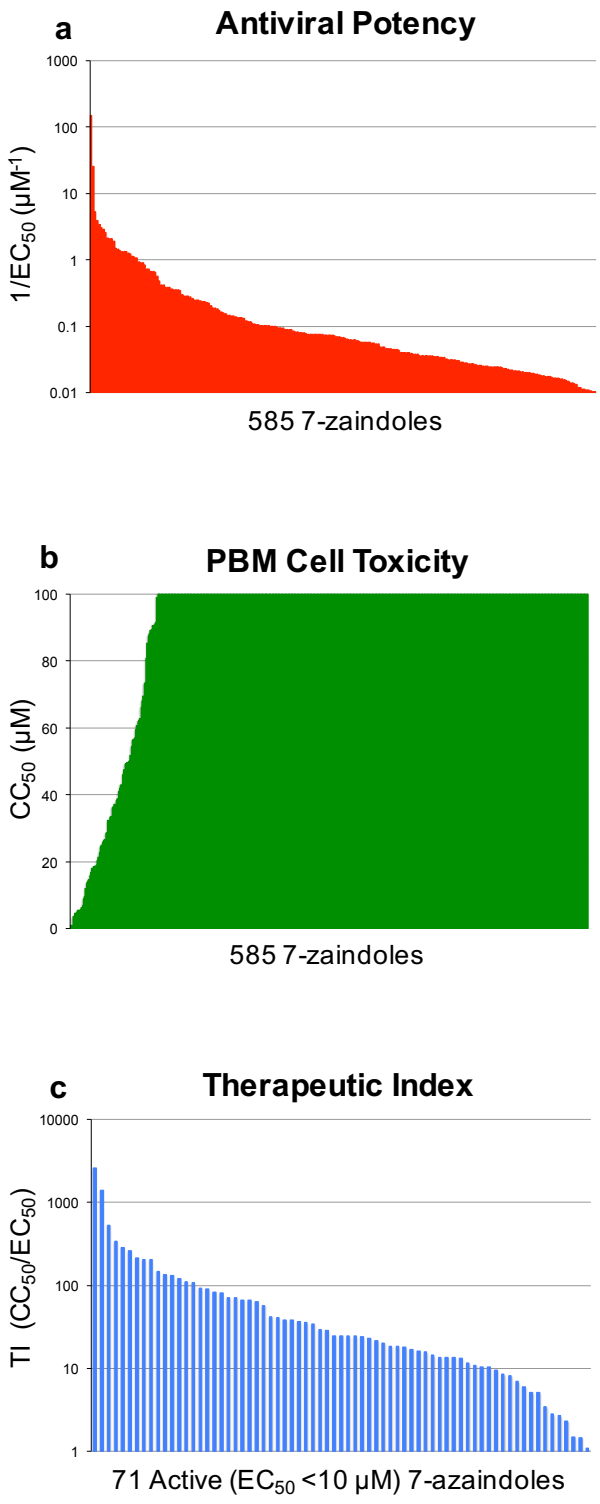


Figure 2.2. Antiviral potency (a) and toxicity (b) of all 7-azaindoles tested and therapeutic index of compounds with potency <10 μM (c).

Ten of the compounds (**Figure 2.3**) with submicromolar antiviral potency showed no toxicity against PBM cells at the maximum concentration tested, giving a therapeutic index (ratio of toxicity to potency) > 100 (**Fig. 2.2c**). Five of those also had no detectable toxicity to all three of the cell lines tested (**Table 2.1**).

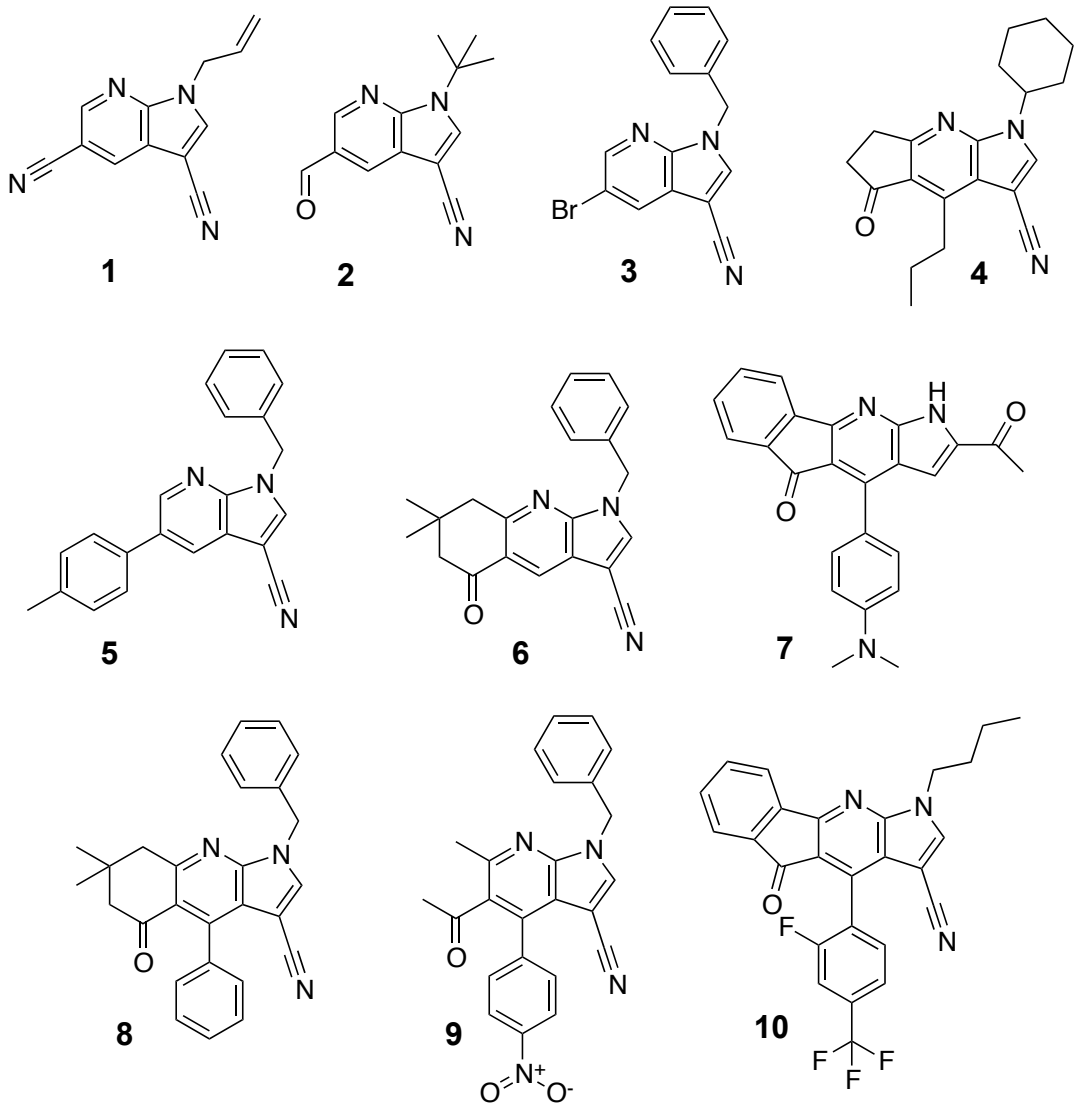


Figure 2.3. Ten most potent 7-azaindoles arranged by molecular mass.

Table 2.1. Activity and toxicity of potent 7-azaindoles (all values in μM, ± SD)

Compound	Antiviral Potency (EC_{50})*	RT Inhibition (IC_{50})**	Cell Cytoxicity (CC_{50})		
			PBM	CEM	VERO
1	0.76 ± 0.48	>100	>100	>100	>100
2	0.49 ± 0.33	14 ± 3.7	>100	>100	8.1
3	0.47 ± 0.30	>100	>100	>100	>100
4	0.69 ± 0.11	>100	>100	>100	>100
5	0.49 ± 0.33	>100	>100	>100	8.1
6	0.91 ± 0.52	12 ± 5.2	>100	3.7	>100
7	0.93 ± 0.10	15 ± 1.2	>100	24	11
8	0.35 ± 0.26	0.73 ± 0.32	>100	>100	>100
9	0.19 ± 0.05	6.3 ± 0.79	>100	>100	>100
10	0.83 ± 0.39	0.58 ± 0.17	>100	48	6.8
NVP	0.04 ± 0.03	0.16 ± 0.01	>100	>100	>100
EFV	0.001 ± 0.0004	0.01 ± 0.005	>100	>100	>100

*Average of at least three independent experiments in HIV-infected PBM cells.
**Average of three independent experiments measuring inhibition of RT polymerase activity in FRET-assay.

In order to determine their antiviral target, a mutation selection assay was conducted using the two most potent compounds, **8** and **9**. Since most anti-HIV-1 hits from chemical screens target the viral polymerase, reverse transcriptase (RT) (10), the first 300 residues of that enzyme were sequenced from the supernatant of PBM cells infected with HIV-1_{LAI} and treated with each compound, as previously described (92). Both drugs selected for mutations on RT, as the V108I mutant emerged after 69 weeks of treatment with **8** and E138K arose after 39 weeks of treatment with **9** (**Fig. 2.4**). These mutations are known to confer resistance to non-nucleoside reverse transcriptase inhibitors (NNRTI), a class of allosteric inhibitors of RT that are an important component of HAART (41). The mutations occurred in the NNRTI-binding pocket, which is found ~10 Å away from the enzyme active site (35). The V108I mutant is associated with resistance to first generation NNRTI (nevirapine (NVP) and efavirenz (EFV)), while the E138K mutant emerges with second generation of drugs (etravirine and rilpivirine) (45), suggesting that the structural differences between **8** and **9** may result in different binding characteristics.

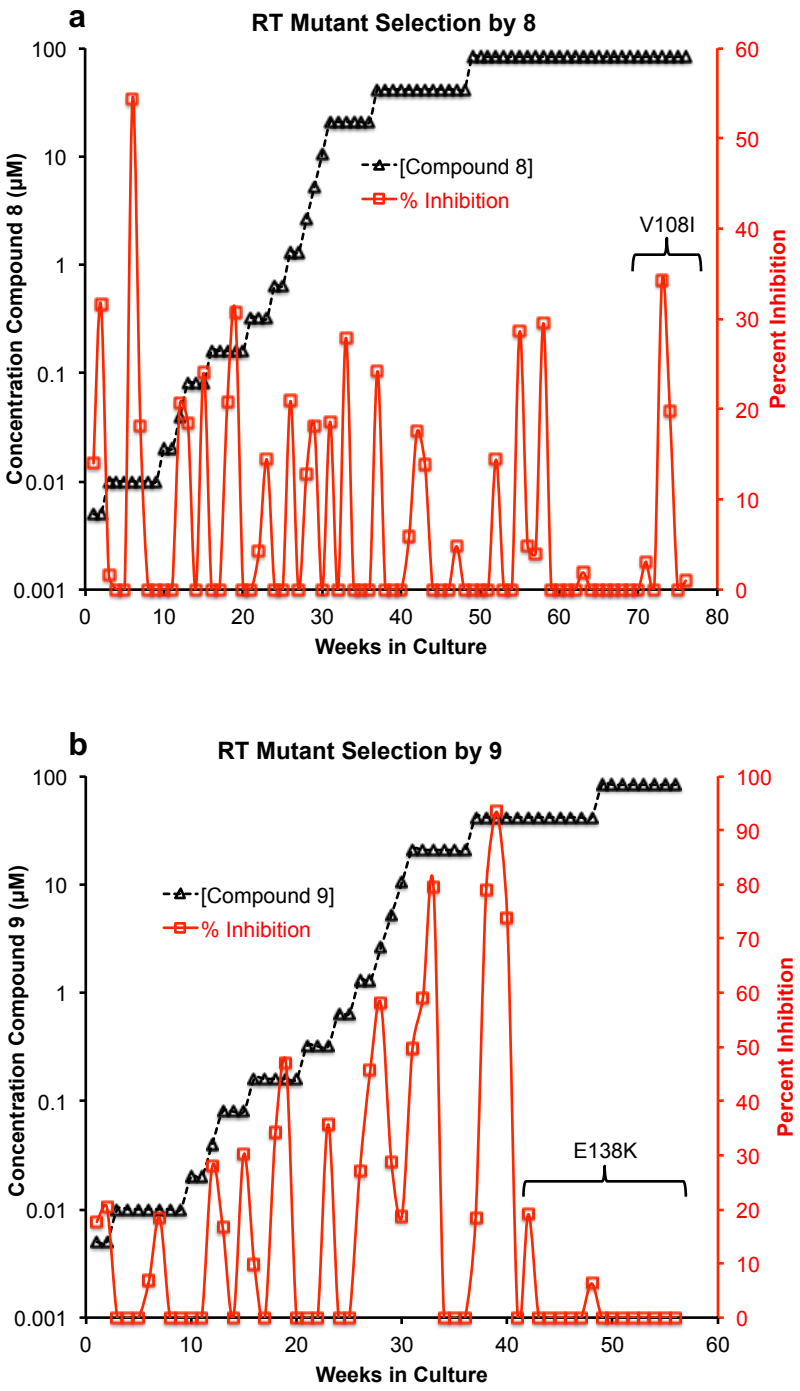


Figure 2.4. Selection of NNRTI resistant mutations by potent 7-azaindoles **8** (a) and **9** (b).

Since the resistance selection assay suggested that the 7-azaindoles are acting as NNRTI, the ten compounds with a therapeutic index >100 were tested for their ability to directly inhibit RT, using a cell-free DNA-dependent DNA polymerization FRET assay (93-95) (details in Supplementary Materials). Only compounds **8** and **10** were found to inhibit the polymerase activity of RT with submicromolar potency, with 50% inhibition concentration (IC_{50}) values of 0.73 and 0.58 μM, while **9** had single-digit micromolar inhibition (**Fig. 2.5** and **Table 2.1**). Those were similar to their EC_{50} values from the cell-based assay, suggesting direct RT inhibition as their primary mechanism of action. Compounds **2**, **6**, and **7** modestly inhibited RT with IC_{50}'s > 10 μM. Interestingly, compounds **1**, **3**, **4**, and **5** showed no inhibition of the enzyme, suggesting the possibility of a different mechanisms of action for their antiviral activity. Given that compounds featuring the 7-azaindole motif have been used against a wide variety of viral and endogenous targets (88; 89; 96; 97), it is not unreasonable that these could have different mechanisms of HIV-1 inhibition aside from acting as NNRTI.

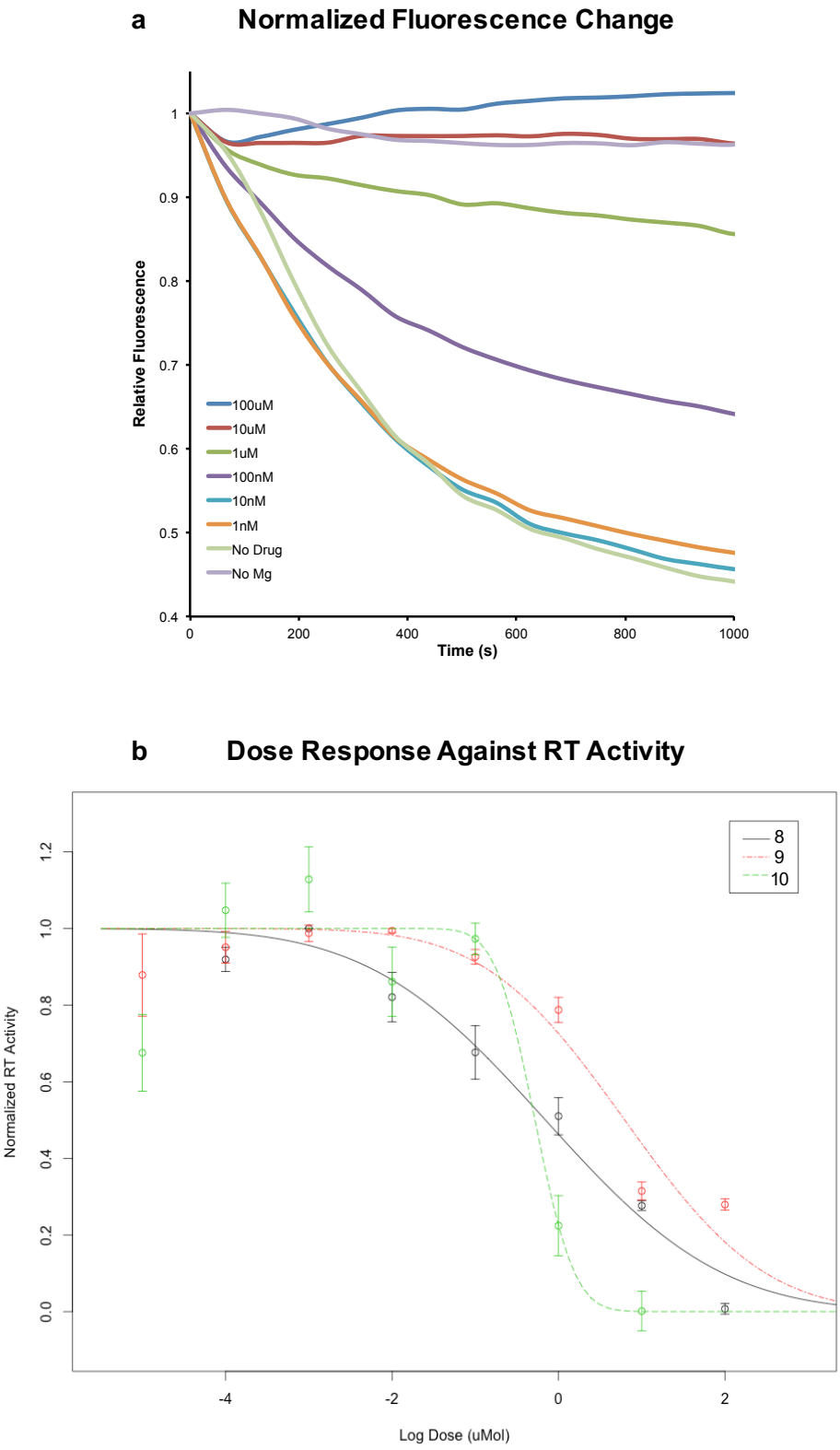


Figure 2.5. For the FRET-based assay normalized changes in fluorescence (a) were converted to RT activity in order to determine dose-response values (IC_{50}) for 7-azaindoles (b).

Because NNRTI inhibit RT via binding to the enzyme at an allosteric site, they do not compete with the enzyme's natural substrates, free nucleotides (46). To determine if that is also the case for the active 7-azaindoles, the potency of **8** was measured with varying concentrations of nucleotides using a gel-based RT primer-extension assay (98) (details in Materials and Methods), as the FRET assay is sensitive to high nucleotide concentrations and therefore unsuitable in this case. Increasing nucleotide concentration did not decrease the IC_{50} of the compound (**Fig. 2.6** and **Table 2.2**), indicating that **8** is not competing with the substrates of RT, consistent with the established NNRTI mechanism. The potency of **8** was actually enhanced with increasing concentration of natural nucleotides, a phenomenon previously observed with other NNRTI (99; 100).

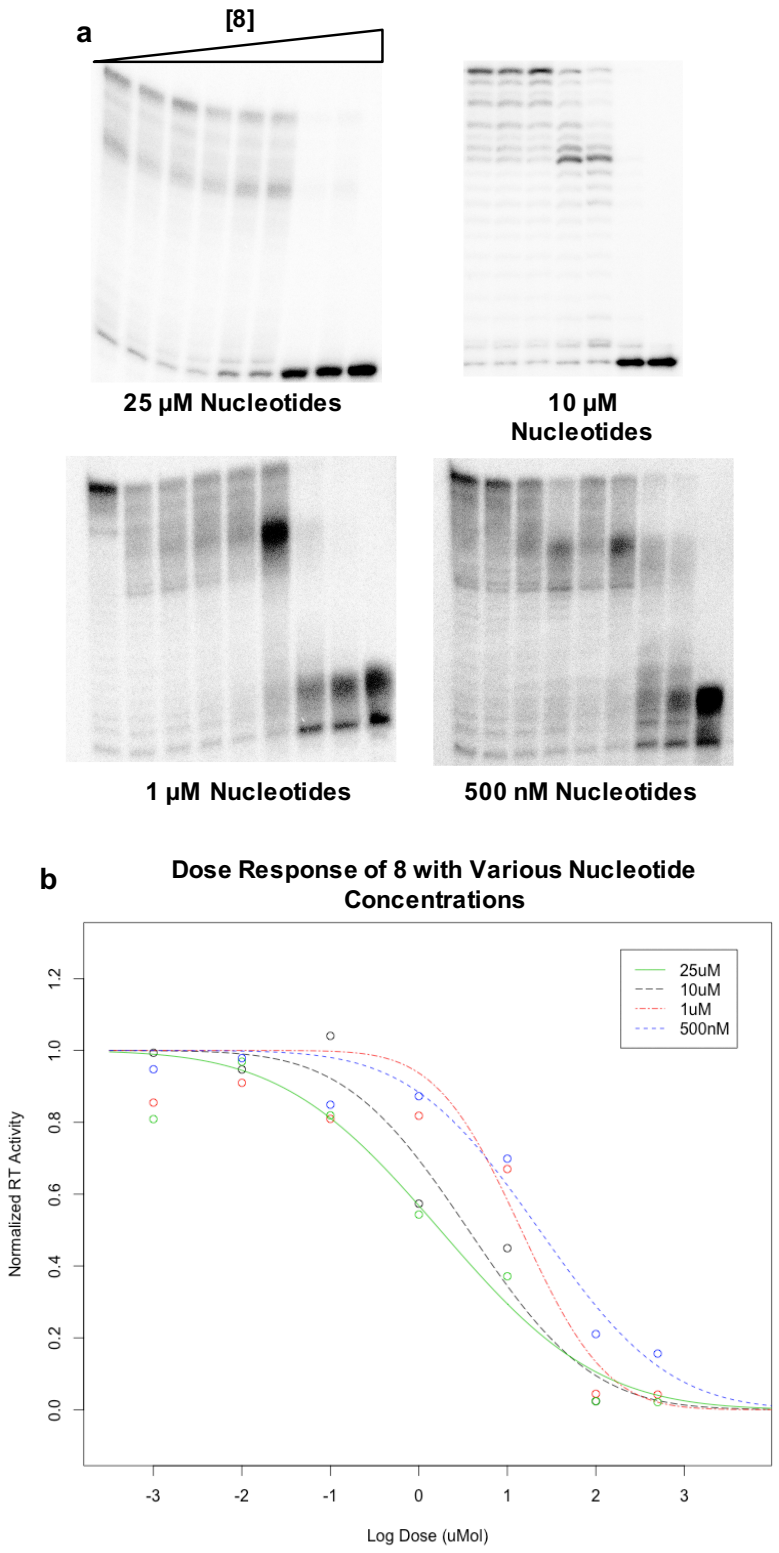


Figure 2.6. RT inhibition by 8 does not compete with nucleotides. RT activity was determined by gel-based primer-extension assay under varying nucleotide concentrations (a), which was then used to determine dose-response values (IC_{50}) **8** (b).

Table 2.2. RT inhibition of 8 with varying nucleotide concentrations (all values in μM, ± SD)

Nucleotide Concetration	RT Inhibition (IC_{50})*
25	0.73 ± 0.32
10	0.96 ± 1.3
1	8.9 ± 1.4
0.5	10 ± 1.2

*Average of three independent experiments measuring inhibition of RT polymerase activity in gel-based assay.

An important consideration for developing new NNRTI is potency against clinically relevant RT mutations (101). The three most active compounds against wild type RT were tested against two clinically important NNRTI-resistant mutants, K103N and Y181C, as well as the two mutants selected by compounds **8** and **9,** V108I and E38K. Enzymes were expressed and purified as previously described (102) and drug activity was measured with the FRET assay. All of the compounds experienced some potency loss against the mutants, although they all maintained measurable activity against the K103N mutant (**Table 2.3**). Compounds **8** and **9** also demonstrated some activity against the Y181C and E138K mutants. However, none of the compounds inhibited the V108I mutant selected by **8**, indicating that there may be conserved binding features among all three molecules that are disrupted by this mutation.

Table 2.3. Activity of 7-azaindole NNRTI against RT mutants (all values in μM, ± SD)

Compound	K103N		V108I		E138K		Y181C	
	RT Inhibition*	Fold Change**	RT Inhibition	Fold Change	RT Inhibition	Fold Change	RT Inhibition	Fold Change
8	9.2 ± 5.2	12.6	>100	N.D.	69 ± 8.2	94	3.6 ± 2.7	4.9
9	15 ± 1.5	2.4	>100	N.D.	59 ± 1.1	9.4	78 ± 1.6	12
10	28 ± 1.2	49	>100	N.D.	>100	N.D.	>100	N.D.

*Average of three independent experiments measuring inhibition of RT polymerase activity in FRET-assay. **Fold change in potency vs. wild type enzyme.

To help establish a structure activity relationship consistent with the experimental results, the three compounds active against RT were docked using an approach described in more detail in **Chapter 3** (103; 104). The compounds all docked with a conserved orientation of the 7-azaindole core, which was validated by the positions of the substitutions on **8** and **9** relative to their selected mutations (**Fig. 2.7**). The mutations occur at opposite ends of the NNRTI-binding pocket, and the dimethylcycloketone motif of **8** is positioned proximal to residue V108 while the nitro group on **9** is adjacent to E138. The V108I mutant is known to shift the position of residue Y188 (66), which may directly affect binding of the 7-azaindole core, as that RT mutant was resistant to inhibition at the highest tested concentration of all three compounds.

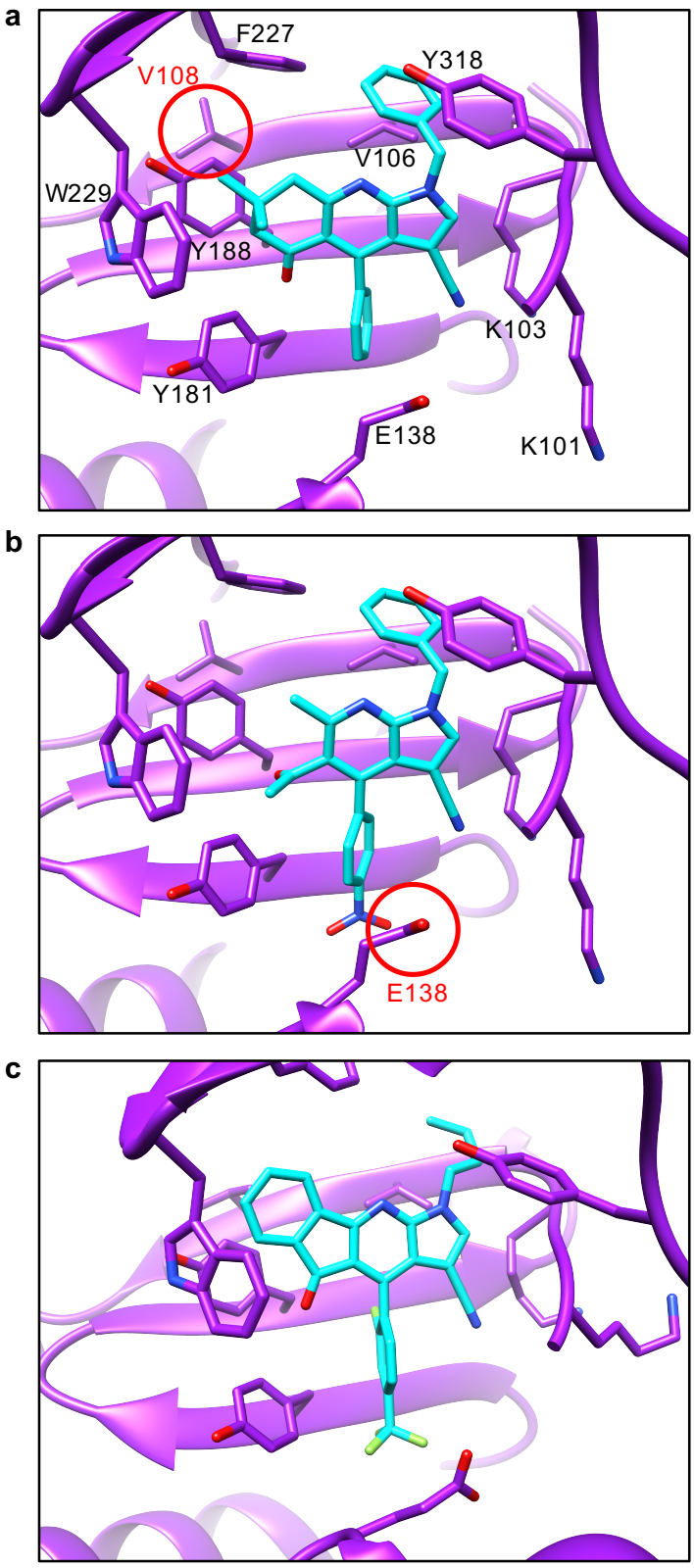


Figure 2.7. Docking of 7-azaindoles **8** (a) and **9** (b) into PDB ID: 2B6A and **10** (c) into PDB ID: 3IS9. Residues mutated after treatment with **8** (V108I) and **9** (E138K) are highlighted in their respective images.

The binding poses also suggest that larger substitutions off of the pyridine ring increase interactions deeper within the pocket towards residue W229. This could help explain the decrease in RT inhibition by **9** versus **8** and **10**. Other compounds lacking these substitutions also exhibited decreased potency against the enzyme. Additionally, it appears that the lower phenyl ring (off of position R_4 in Fig. 1) on **8**, **9**, and **10** may be involved with π-stacking with residue Y181. This motif is notably absent from the compounds with no observed RT polymerase inhibition.

2.4 Conclusions

A unique class of NNRTI has been discovered, as 7-azaindoles demonstrated potent and specific anti-HIV activity with low toxicity, selected for NNRTI-associated mutations, and directly inhibited the polymerase activity of RT while not competing with nucleotides for the enzyme active site. The lead compounds maintained activity against clinically relevant mutants. Additionally, the observation that some of the 7-azaindoles inhibited HIV-1 with submicromolar potency and low toxicity while not directly affecting RT polymerase activity suggest that they may have alternative mechanisms of action. Based on this observation, compounds could potentially be built off of the 7-azaindole scaffold with multiple anti-HIV mechanisms of action, which could help address the problem of NNRTI-associated mutations. Efforts towards identifying the alternative targets of antiviral effect and further optimization of the NNRTI activity of the 7-azaindoles are ongoing.

2.5 Materials and Methods

Viral Replication Inhibition in PBM Cells

Primary human peripheral blood mononuclear (PBM) cells were cultured as previously described (105). Briefly, PBM cells were isolated from buffy coats from healthy seronegative donors and and stimulated with phytohemagglutinin in RPMI medium. The cells were infected with HIV-1 at an MOI of 0.1 and serial dilutions of the drugs were added to triplicate cultures. The cultures were maintained for 6 days after infection, at which point the supernatants were sampled for reverse transcriptase (RT) activity using the RT assay previously described (106). EC_{50} values were derived from the resultant dose-response curves.

Cytoxicity Assays

Cytoxicity was measured in PBM, CEM (human T-lymphoblastoid derived), and VERO (African green monkey kidney epithelium) cells using an MTT-dye reduction assay as previously described (92). Dose-response curves for each drug were determined in triplicate using a serial dilution, from which the CC_{50} values were derived.

RT Expression and Purification

RT was cloned from the HXB2 HIV-1 strain in a Novagen pET28a expression plasmid (MilliporeSigma, Darmstadt, DE) with an N-terminal hexahistidine-tagged p66 subunit. Mutations within the NNRTI-binding pocket were introduced using a QuikChange Site-Directed Mutagensis Kit (Agilent Technologies, Santa Clara, CA). The enzymes were expressed in *E. coli* BL21 cells and purified with Ni^{2+}-NTA chromatography followed by SP anion exchange, as previously described(102; 107).

FRET-Based HIV Reverse Transcriptase DNA Polymerase Assay

The RT inhibition assay was adapted from a protocol described by Cauchon et al. (93). A 17-mer DNA primer was annealed to 2.5-fold excess 42-mer, Alexa Fluor 488 5'-end labeled DNA template (both from Integrated DNA Technologies, Coraville, IA). Reaction mixes included 80 nM of primer/template, 500 nM of nucleotides (including the Alexa Fluor 555-aha-dUTP quencher from Life Technologies, Grand Island, NY), 10 nM RT, with varying concentrations of drugs dissolved in DMSO, all in a buffer of 50 mM Tris, 80 mM KCl, 1 mM dithiothreitol, 12 mM $MgCl_2$. 10 µL of the enzyme and primer/template mix was incubated for 10 minutes with the drugs, and the reaction was initiated by the addition of 10 µL of the nucleotide mix, for a final volume of 20 µL. The reaction was run on 96 well plates for 30 minutes at 37° C in a Roche Lightcycler 480, which monitored fluorescence once/minute with excitation/detection wavelengths of 483/533 nM. Steady state rates of reaction were estimated from fluorescent resonance energy transfer (FRET)-associated decreases in the fluorescent signal over time under concentrations of drug from 1 nM up to 100 µM, with a DMSO only and a no-$MgCl_2$ reaction included as controls. Each experiment was done in triplicate with four technical replicates for each drug concentration. The raw fluorescence data was normalized and IC_{50} values were determined using the drift (94) package in R (95).

Gel-Based Primer Extension Assay

The primer extension assay was modified slightly from a previous description (98). A 5'-end ^{32}P-labeled 24-mer primer was annealed to 2.5-fold excess of a 48-mer template (both from Integrated DNA Technologies). The reaction

mixture included varying concentrations of nucleotides along with 20 nM primer/template, 1.5 nM RT in a buffer of 50 mM Tris and 50 mM NaCl with 10 mM $MgCl_2$. The enzyme was incubated with primer/template and various concentrations of drugs dissolved in DMSO for 5 minutes and the reaction was initiated with the addition of nucleotides for a final volume for 40 μL. The reactions were run for various times (15 minutes for 500 nM nucleotides, 10 minutes for 1 μM, 5 minutes for 10 μM, and 2 minutes for 25 μM) so as to reach 50% extension. Reactions were quenched with 0.3 M (final) EDTA. Samples were loaded onto a 14% polyacrymalide/8 M urea gel and visualized using PharosFX (Bio-Rad). The products were normalized and IC_{50} values were determined as before.

CHAPTER 3:

Ligand Similarity Guided Receptor Selection Enhances Docking Accuracy and Recall for Non-Nucleoside HIV-1 Reverse Transcriptase Inhibitors

This research was published in *Journal of Molecular Modeling*.

Richard A. Stanton, James H. Nettles, and Raymond F. Schinazi.

"Ligand similarity guided receptor selection enhances docking accuracy and recall for non-nucleoside HIV reverse transcriptase inhibitors."

J Mol Model. 2015 Nov; 21(11):282.

Contributions to this chapter:

Richard A. Stanton performed all *in silico* calculations and programming, wrote, and edited the chapter.

James H. Nettles assisted with data analysis and editing of the chapter.

Raymond F. Schinazi conceived of the project, provided all experimental resources, and assisted editing of the chapter.

3.1 Abstract

Non-nucleoside reverse transcriptase inhibitors **(**NNRTI) are allosteric inhibitors of human immunodeficiency virus type 1 (HIV-1) reverse transcriptase (RT), the viral polymerase. Despite the availability of > 150 NNRTI-bound RT crystal structures, rational design of new NNRTI remains challenging because of the variability of their induced fit, hydrophobic binding patterns. Docking NNRTI yields inconsistent results that vary markedly depending on the receptor structure used, as only 27% of the >20k cross-docking calculations performed using known NNRTI were accurate. In order to determine if a hospitable receptor for docking could be selected *a priori*, more than 40 chemical descriptors were evaluated for their ability to pre-select a best receptor for NNRTI cross-docking. The receptor selection was based on similarity scores between the bound- and target-ligands generated by each descriptor. The top descriptors were able to double the probability of cross-docking accuracy over random receptor selection. Additionally, recall of known NNRTI from a large library of similar decoys was increased using the same approach. The results demonstrate the utility of pre-selecting receptors when docking into difficult targets.

3.2 Introduction

Docking novel small molecules into protein targets is an important step towards identifying leads and providing a rationale for optimization of new compounds (108). Unfortunately, current docking approaches have limitations predicting the correct binding poses and affinities for certain targets (109; 110). This becomes especially apparent when dealing with induced fit effects, whereby

ligand binding leads to large changes in the conformation of the receptor pocket (111; 112).

Induced fit binding effects often occur with allosteric inhibitors, which bind to pockets distal from the active site (113). Allosteric binding sites often undergo more marked structural rearrangements upon binding than those of active site inhibitors (114), making accurate docking much more difficult.

There are many known allosteric targets (115), and among those HIV-1 reverse transcriptase (RT) is of particular clinical interest. RT plays an essential role in HIV replication, converting the viral genome from RNA into DNA prior to insertion into the host nucleus. Non-nucleoside reverse transcriptase inhibitors (NNRTI) are a class of widely-used allosteric inhibitors of the enzyme, five of which (nevirapine (NVP), delavirdine (DLV), efavirenz (EFZ), rilpivirine (RPV), and etravirine (ETV)) are FDA approved. NNRTI bind via an induced fit mechanism to a pocket that does not exist in the absence of inhibitors, about 10 Å away from the active site. This binding causes a conformational shift to a non-productive form of the enzyme, thereby inhibiting activity (42).

In spite of the vast amount of available data (>25k compounds have been tested as NNRTI in the NIAID ChemDB (116) and more than >150 crystal structures of RT-NNRTI complexes are in the RCSB Protein Data Bank (PDB)) (117), virtual screening and lead optimization of new NNRTI remains challenging. Known NNRTI are structurally diverse, and as with other allosteric inhibitors, their induced fit binding mechanism yields large variations in protein-ligand interaction patterns (70; 118), making ligand-based drug design approaches problematic. And since the

binding is largely hydrophobic and relatively non-specific (47), scoring of docked poses is difficult (119) and cross-docking NNRTI gives results that vary widely in accuracy from receptor to receptor (120). Studies of other induced fit systems have shown that docking into the crystal structures of receptors bound to similar ligands increases accuracy, but these focused on only active-site inhibitors (121-123).

Building on these observations and taking advantage of the structural information available for NNRTI, a combination of ligand- and structure-based approaches were used to predict receptors for accurate cross-docking and improve the selection of active compounds through virtual screening. Forty-four chemical descriptors were assessed to test predictions for cross-docking 87 known NNRTI based solely on the similarity between the native- and target-ligand. Then, the best performing descriptor was used as a guide to select receptors for docking a mixed set of known NNRTI and inactive decoys to determine if the approach could enhance recall of active compounds in a virtual screen.

3.3 Results and Discussion

Analysis of NNRTI-Bound RT Structures

Eighty-seven of the >150 NNRTI-bound RT crystal structures (**Figure 3.1**) available in the RCSB Protein Data Bank (124) were selected for the initial study, as these included the wild type enzyme bound to a single NNRTI.

There is not very much variability in the 20 NNRTI binding site residues of the different structures as measured by the root mean square deviation (RMSD) between all atoms (both side-chain and backbone) of these residues, as the average value is only 1.11 Å (**Table 3.1**). However, changes in the positions of only a few of

those 20 residues can lead to vastly different interaction patterns for bound NNRTI (125; 126). For example, the crystal structures of NVP- and DLV-bound RT shown in **Figure 3.1b** and **3.1c** have a low binding site RMSD (only 1.81 Å), yet their binding pocket volumes differ by > 20%, and their protein-ligand interaction patterns are quite distinct. These differences arise largely due to the movement of the side chains of only two key residues, K102 and Y318, which allow the larger DLV molecule to interact with an additional region of the pocket. Thus the low average variability of the binding pockets of the NNRTI structures (as measured by the RMSD of their respective atoms) does not necessarily reflect similar pocket conformations and interaction patterns.

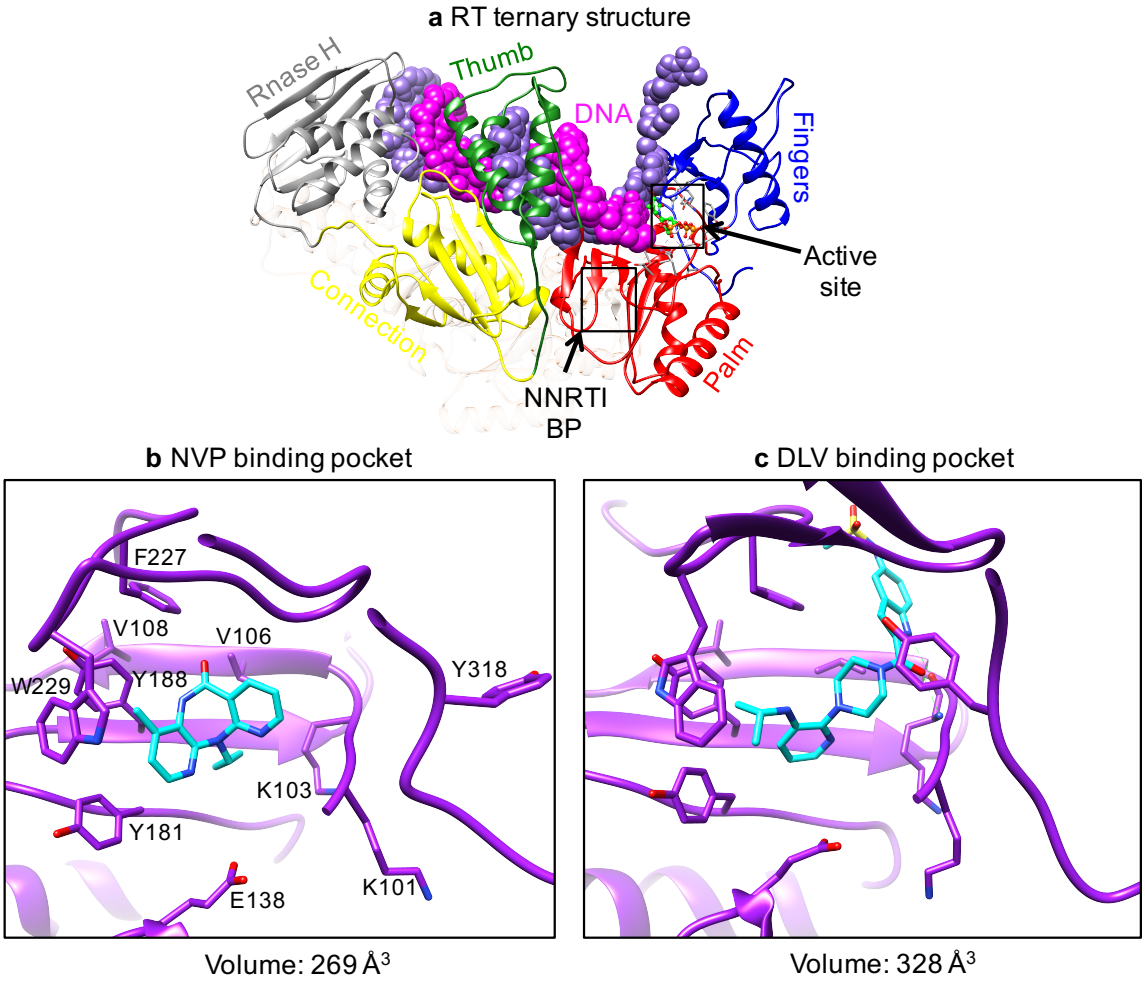


Figure 3.1. NNRTI binding pocket variability. Reverse transcriptase from solved ternary structure of PDB ID: 1RTD (a). The binding pockets of clinical NNRTI NVP (PDB ID: 3HVT) and DLV (PDB ID: 1KLM) are inset (b, c).

Table 3.1. Comparison of NNRTI-bound RT structures

	Backbone RMSD (Å)	Binding Pocket RMSD (Å)
Average	2.66	1.11
Maximum	4.85	2.79
Minimum	0.23	0.19
Std. Dev.	0.83	0.33

Overall, the NNRTI from the set are quite chemically distinct, with an average Tanimoto similarity of only 0.41 as measured by MACCS structural keys (**Table 3.2**). However, 40 of the 87 structures have at least one closely related analog within the set, while nine of the structures actually feature overlapping NNRTI (three structures are bound to RPV, and two each are bound to NVP, EFV, and ETV).

Table 3.2. Comparison of NNRTI from Solved Structures

Comparison Set	MACCS Similarity
Cluster Centers	0.328
Different Cluster	0.389
Same Cluster	0.521
Overall Avg.	0.41
Avg. Max (Non-Self)	0.81
Avg. Max (Non-Cluster)	0.65

In order to explore methods of selecting a best receptor for cross-docking, receptors that shared the same ligand or closely related analogs were excluded from the analysis, so as to better replicate the scenario of working with novel compounds. To accomplish this the NNRTI were clustered into ten bins (with cluster centers shown in **Figure 3.2**, compounds **11**-**20**), which varied in size from only two members (cluster **18**) up to twenty (cluster **15**), and only cross-docking results from receptors selected from outside each cluster were considered. For example, the best receptor for cross-docking a compound from cluster **11** was selected from clusters **12-20**, and so on. This ensured that the receptor selection was not skewed by cross-docking into the receptor of the same ligand or one from the same chemical series.

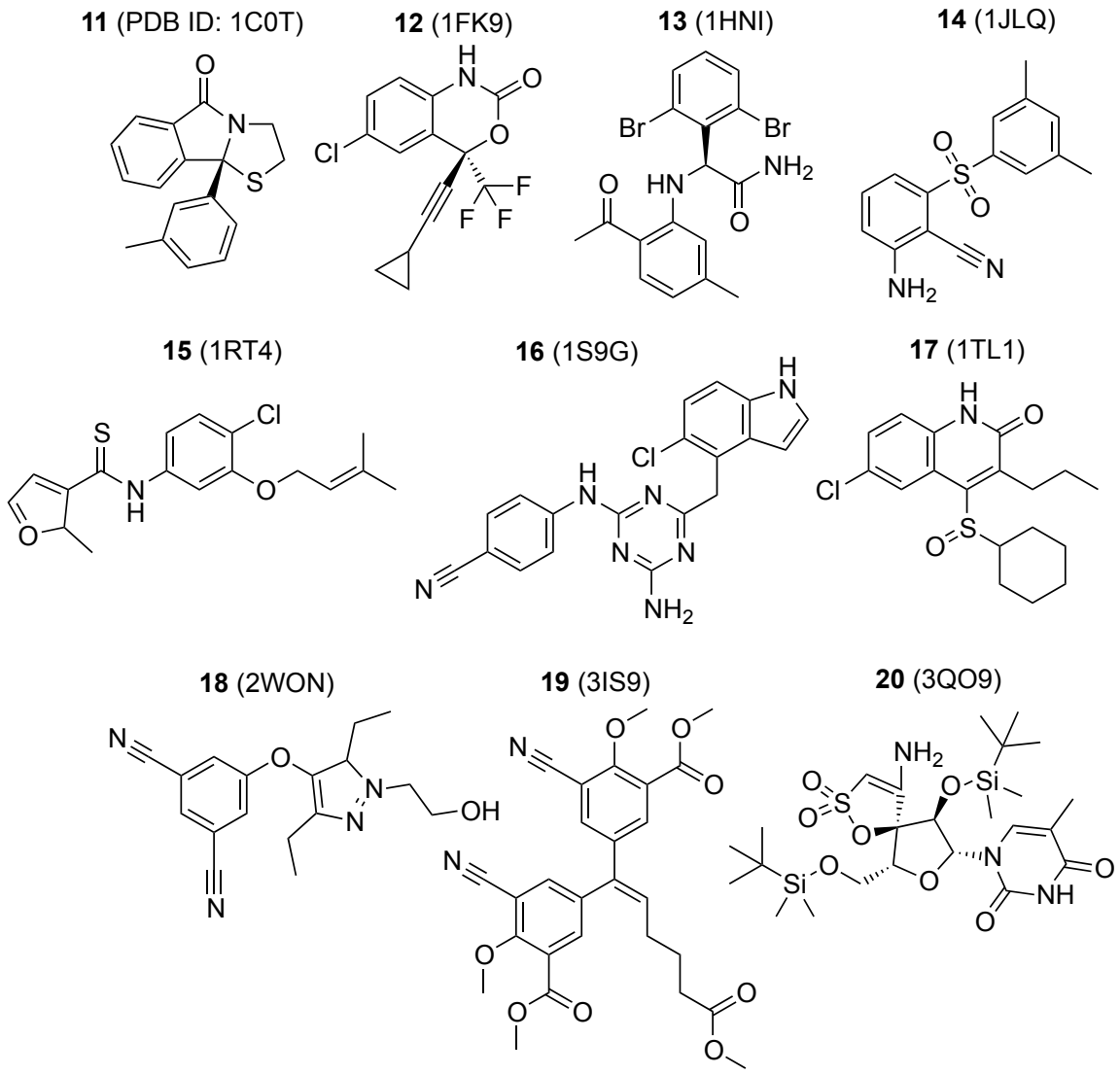
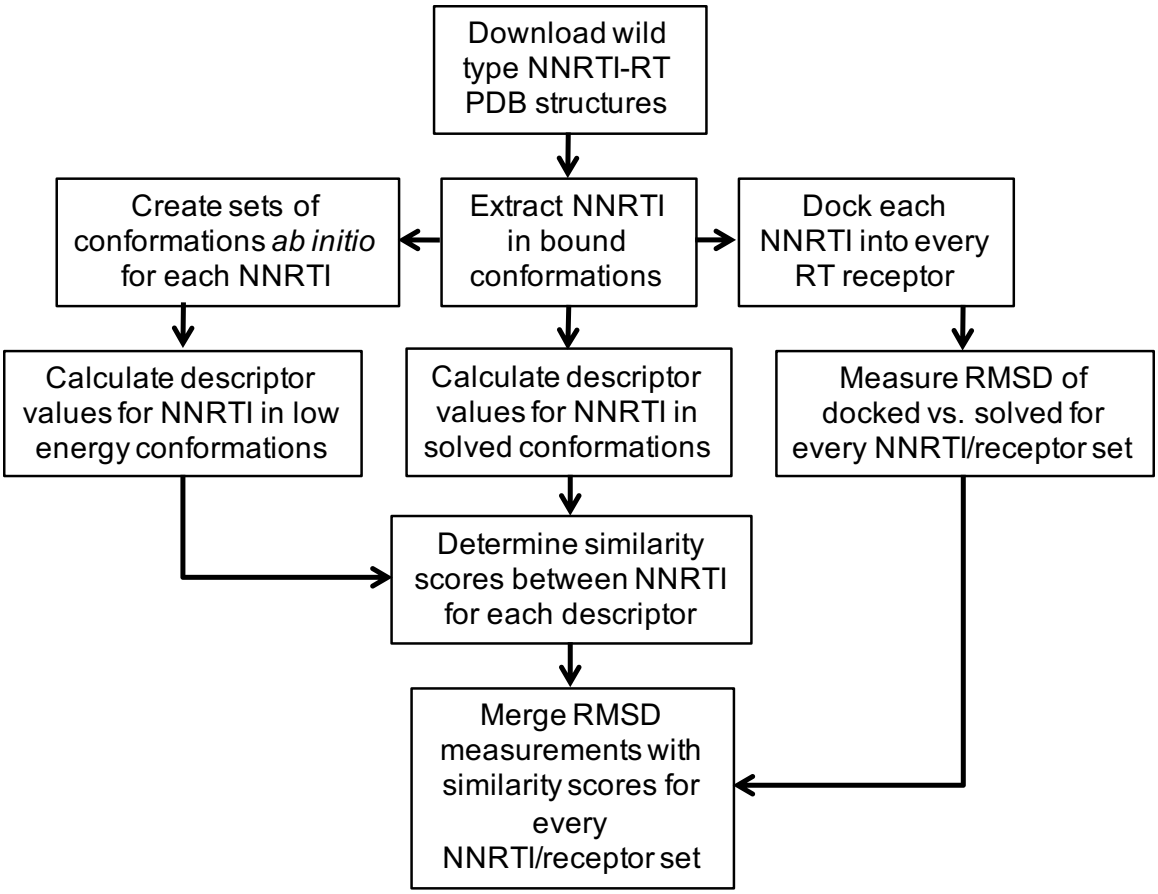


Figure 3.2. Cluster center compounds for the ten clusters made from the MACCS similarity scores of 87 NNRTI used in the study, with their PDB ID of their respective structures in parentheses.

Cross Docking Results

The workflow shown in **Scheme 3.1** was used to create a database of ligand structure similarity and cross-docking results. The 87 NNRTI were docked into the 87 RT structures using Autodock Vina (103) and then the accuracy of the resultant docked poses was determined using the RMSD of the heavy atoms of the docked and solved binding conformations. Since Autodock Vina uses a random ligand conformation to initiate docking, we ran each calculation in triplicate, resulting in a database of 22,707 RMSD measurements (87 ligands x 87 receptors x 3 trials). Only the top-scoring poses from each ligand-receptor set for this study were considered.



Scheme 3.1. Workflow to create merged database of cross-docking accuracy and native/target ligand similarity scores for 44 chemical descriptors applied to 87 NNRTI.

A docking pose was designated as accurate if its RMSD was < 2 Å from that of the solved binding pose of the ligand. The self-docking accuracy of the 87 NNRTI into their own binding pockets was 85% (74 of the 87, **Figure 3.3**), indicating that Autodock Vina can accurately reproduce and rank binding poses of NNRTI given the correct receptor conformation. However, cross-docking was much less accurate. Docking into the receptors of compounds from the same cluster was only 36% accurate, while cross-docking into non-cluster receptors was even less effective at 27%, and varied widely from receptor to receptor (**Table 3.3**).

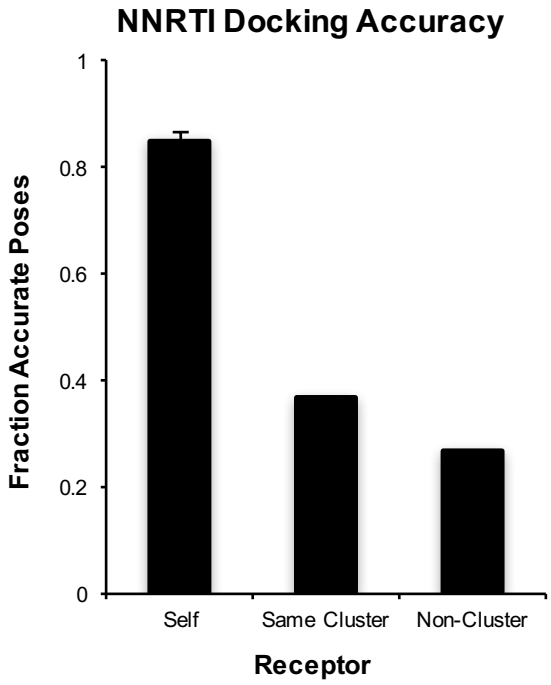


Figure 3.3. Overall cross docking results of 87 NNRTI. Each bar represents the fraction of low-scoring docked poses with an RMSD < 2 Å away from the solved binding pose. The average results from three docking runs per ligand-receptor pair is shown.

Table 3.3. Cross-docking accuracy per PDB

	Non-Cluster Cross Docking Accuracy	
PDB ID	***Receptor***	***Ligand***
1BQM	0.101	0.610
1C0T	0.417	0.550
1C0U	0.358	0.442
1C1B	0.613	0.342
1C1C	0.396	0.409
1DTQ	0.299	0.074
1DTT	0.367	0.063
1EET	0.278	0.551
1EP4	0.401	0.190
1FK9	0.184	0.877
1HNI	0.177	0.576
1HNV	0.140	0.285
1IKW	0.241	0.890
1JLQ	0.328	0.132
1KLM	0.271	0.046
1REV	0.259	0.364
1RT1	0.458	0.480
1RT2	0.471	0.258
1RT4	0.414	0.288
1RT5	0.348	0.313
1RT6	0.303	0.096
1RT7	0.217	0.030
1RTH	0.210	0.272
1RTI	0.289	0.213
1S6P	0.108	0.059
1S6Q	0.113	0.279
1S9E	0.191	0.172
1S9G	0.387	0.039
1SUQ	0.069	0.201
1TKT	0.354	0.652
1TKX	0.293	0.359
1TKZ	0.258	0.571
1TL1	0.343	0.525
1TL3	0.364	0.727
1TV6	0.000	0.000
1TVR	0.254	0.101

	Non-Cluster Cross Docking Accuracy	
PDB ID	***Receptor***	***Ligand***
1VRT	0.284	0.481
1VRU	0.172	0.515
2B5J	0.222	0.040
2B6A	0.325	0.143
2BAN	0.175	0.163
2BE2	0.404	0.377
2I5J	0.000	0.000
2JLE	0.176	0.250
2OPP	0.254	0.785
2RF2	0.238	0.446
2RKI	0.198	0.203
2VG5	0.238	0.075
2VG6	0.308	0.046
2VG7	0.196	0.125
2WON	0.321	0.197
2YKN	0.251	0.322
2YNG	0.364	0.268
2YNH	0.369	0.202
2YNI	0.354	0.237
2ZD1	0.230	0.446
3C6T	0.317	0.133
3C6U	0.376	0.165
3DI6	0.348	0.434
3DLE	0.414	0.318
3DLG	0.318	0.313
3DRP	0.353	0.230
3DYA	0.329	0.236
3E01	0.392	0.147
3FFI	0.364	0.232
3HVT	0.136	0.457
3I0R	0.329	0.192
3I0S	0.263	0.167
3IRX	0.257	0.000
3IS9	0.201	0.000
3LAK	0.309	0.221
3LAL	0.267	0.133

Table 3.3. Cross-docking accuracy per PDB (cont.)

	Non-Cluster Cross Docking Accuracy	
PDB ID	***Receptor***	***Ligand***
3LAM	0.413	0.213
3LAN	0.409	0.236
3M8P	0.029	0.588
3M8Q	0.154	0.192
3MEC	0.049	0.377
3MEE	0.240	0.324
3NBP	0.044	0.162
3QO9	0.044	0.036
3T19	0.397	0.139
4G1Q	0.240	0.328
4H4M	0.311	0.076
4H4O	0.262	0.062
4I2P	0.172	0.417
4I7F	0.292	0.078
4IG0	0.232	0.092

Cross Docking with a Flexible Receptor

Common techniques to improve cross-docking accuracy include incorporating side-chain (such as with rotamer exploration) (127) or full residue flexibility (*via* molecular dynamics (MD)) (128). We assessed the utility of those methods using a single RT-NNRTI structure (1KLM, bound to DLV) (50) as a representative receptor, as it had the same non-cluster cross-docking accuracy (27%, **Table 3.3**) as the overall average for every receptor used. Incorporating flexible side chains while docking actually lead to a decrease in the number of accurate poses (**Figure 3.4**). Performing a post-docking minimization (using a conjugate gradient method in NAMD for 10 ps) or minimization plus a short MD run (1 ns) on the docked structures both led to slightly more accurate poses, but this improvement was not statistically significant.

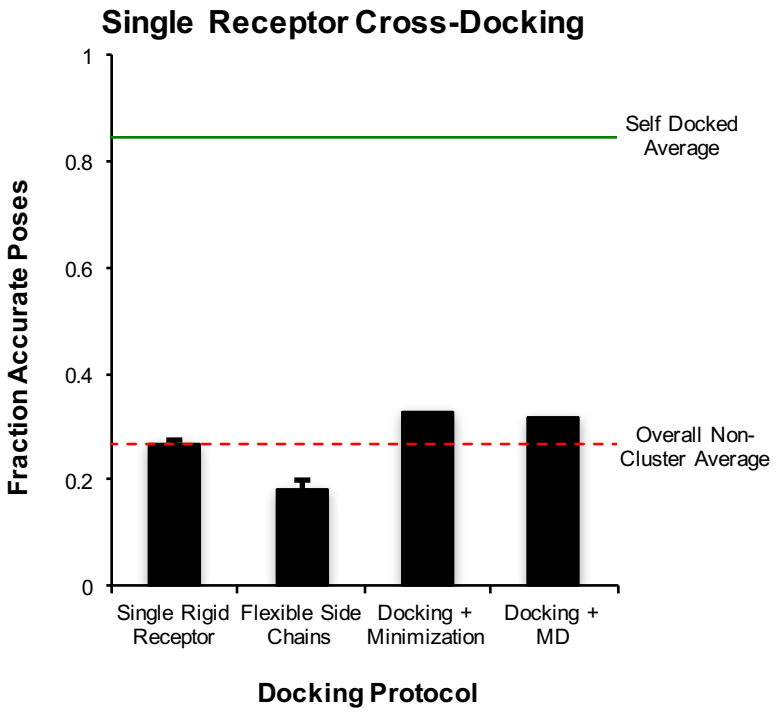


Figure 3.4. Cross-docking into the pocket of DLV-bound RT (PDB ID: 1KLM) with several additional steps to compensate for the induced fit effect. The overall fractions of accurate self, and non-cluster results are shown as lines for comparison.

Assessment of Chemical Descriptors for Receptor Selection

While the overall cross-docking accuracy for the entire set of NNRTI using non-cluster receptors is low, for 95% of the ligands (83 of the 87) there is at least one non-cluster receptor for which the docked pose is accurate. This suggests that there may be overlapping binding pocket conformations induced by chemically dissimilar NNRTI. Assuming that these pocket conformations are driven by similar protein-ligand interaction patterns, it stands to reason that there may be functional groups with similar properties and orientations on otherwise dissimilar NNRTI (since they are from different clusters).

To determine if the chemical similarities inducing similar binding patterns could be identified *a priori*, 44 descriptors were assessed for their ability to predict receptors for accurate cross-docking of NNRTI based on the chemical similarity scores between the compound to be docked and the NNRTI originally bound to the receptor structure.

The similarity amongst the NNRTI was determined for each descriptor (**Scheme 3.1**). Then for each NNRTI the most similar non-cluster compound was selected and the accuracy of docking into the receptor of that compound was assessed. For example, using the number of heavy atoms as the descriptor, the most similar non-cluster compound to **11** is **14**, as they both have 20 heavy atoms and belong to different clusters. **11** docks into the receptor of **14** with an RMSD of 0.41 Å as compared to the solved pose, so that is counted as an accurate pose, since RMSD < 2 Å. This process was repeated for all 87 NNRTI for each descriptor, and the fraction of accurate poses for each descriptor was determined and compared back to the

overall average of 27%. Statistical significance was determined using a one-tailed Student's *t*-test with a threshold of only 0.005 to decrease the likelihood of false positives since more than 40 descriptors were being tested.

Three 1D descriptors (e.g., number of heavy atoms), twenty 2D fingerprint-based descriptors (e.g., ECFC_6 from Pipeline Pilot (129)) eight 3D shape-based descriptors (e.g., surface area), and twelve 3D pharmacophore-based descriptors (e.g., piDAPH4 from MOE (130)) were assessed. In order to facilitate 3D comparisons a set of low energy conformations for each of the compounds was created using Low Mode MD. Then those were compared to the solved conformations of the NNRTI using each descriptor, and only the highest similarity score from each conformation set was kept.

Overall, nine of the descriptors tested were able to select receptors for accurate cross-docking at a statistically improved rate over the non-cluster average. Of these, seven were 2D fingerprint- and two were 3D pharmacophore-based descriptors (**Table 3.4**). Three of the successful 2D fingerprint-based (ECFC_2, ECFC_4, and ECFC_6) were able to select accurately docked receptors for ≥50% of the NNRTI. The publicly available FP2 descriptor from Open Babel (131) also performed well, as the receptor selected using that descriptor was accurate for 46% of the NNRTI tested. The 3D pharmacophore-based descriptors that could accurately predict receptors for cross-docking at a rate significantly above the average were a 4-point pharmacophore (piDAPH4) and 2-point atom type descriptor (TAD).

Table 3.4. Cross-docking accuracy* into receptors selected by similarity descriptors

Descriptor	*Type*	*Source*	*Fraction Accurate*
ECFC_6	2D	Pipeline Pilot	0.53*
ECFC_2	2D	Pipeline Pilot	0.53*
ECFC_4	2D	Pipeline Pilot	0.50*
ECFP_4	2D	Pipeline Pilot	0.48*
FP2	2D	Obabel	0.46*
ECFP_6	2D	Pipeline Pilot	0.46*
ECFP_2	2D	Pipeline Pilot	0.46*
piDAPH4	3D	MOE	0.46*
TAD	3D	MOE	0.46*
FCFP_6	2D	Pipeline Pilot	0.44
FCFC_4	2D	Pipeline Pilot	0.44
FCFP_2	2D	Pipeline Pilot	0.43
FCFP_4	2D	Pipeline Pilot	0.43
TGT	2D	MOE	0.43
gpiDAPH3	2D	MOE	0.42
FCFC_6	2D	Pipeline Pilot	0.42
piDAPH3	3D	MOE	0.41
FCFC_2	2D	Pipeline Pilot	0.41
TGD	2D	MOE	0.39
TAT	3D	MOE	0.39
Surface Area	Shape	MOE	0.38
Volume	Shape	N/A	0.36
MACCS	2D	Obabel	0.36
Spectrophore - Shape Deviations	Shape	Obabel	0.35
ESshape3D	3D	MOE	0.34
Spectrophore - Overall	3D	Obabel	0.33
VDW Area	Shape	N/A	0.32
Esshape3D_HYD	3D	MOE	0.31
VDW Volume	Shape	MOE	0.31
BIT_MACCS	2D	MOE	0.31
FP3	2D	Obabel	0.31
Diameter	Shape	MOE	0.29

* $p < 0.005$

Table 3.4. Cross-docking accuracy into receptors selected by similarity descriptors (cont.)

Descriptor	*Type*	*Source*	*Fraction Accurate*
Spectrophore - Partial Charges	Shape	Obabel	0.29
logP	1D	N/A	0.28
# Heavy Atoms	1D	N/A	0.28
Spectrophore - Lipophilicity	Shape	Obabel	0.28
Spectrophore - Electrophilicity	Shape	Obabel	0.27
Density	Shape	N/A	0.26
FP4	2D	Obabel	0.26
Globularity	Shape	N/A	0.25
MolecularWeight	1D	N/A	0.23
Hydrophobic Surface Area	Shape	MOE	0.23
ACPC	3D	Zhang Initi.	0.16

The top descriptors were those that used abstractions of atoms and functional groups to create fingerprints or pharmacophore patterns in 2D and 3D, and were superior to the simpler 1D and shape-only 3D descriptors at selecting appropriate receptors for docking (**Table 3.4** and **Figure 3.5**). Descriptors that more heavily weighted electronic features (such as ACPC (132) or Spectrophore-Partial Charges (131)) did not perform well in this context, since those characteristics are less important for the hydrophobic binding of NNRTI. Ultimately, performance of any descriptors at receptor selection will largely depend on the specific molecular interactions that drive binding and inhibition, and how well the descriptors can recognize those attributes.

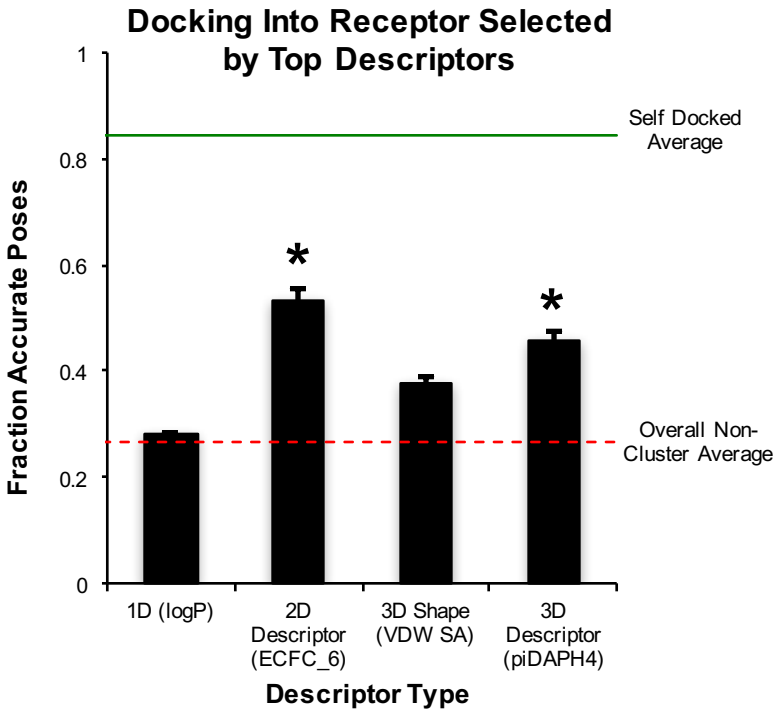


Figure 3.5. The performance of the top descriptors varies for predicting the best receptor for cross-docking NNRTI. Both the top 2D and 3D descriptors were able to select receptors for accurate cross-docking at significantly improved rates over the non-cluster average (*, $p < 0.005$).

And while more 2D than 3D descriptors were successful at receptor selection, this appears to be a product of the descriptor characteristics themselves (i.e., how they classify and weight individual atoms or groups) rather than the limitations of generating *ab initio* 3D conformations relevant to binding. In fact, the 3D conformations generated for the NNRTI were largely consistent with their binding poses, as 92% had an RMSD < 2 Å from the bound conformation. Furthermore, a comparison of three MOE descriptors applied as both 2D and 3D showed no consistent enhancement for one approach over the other (**Figure 3.6**). Together, these observations suggest that for NNRTI the method of abstraction by descriptors is more important than the 3D overlap of specific atoms or groups.

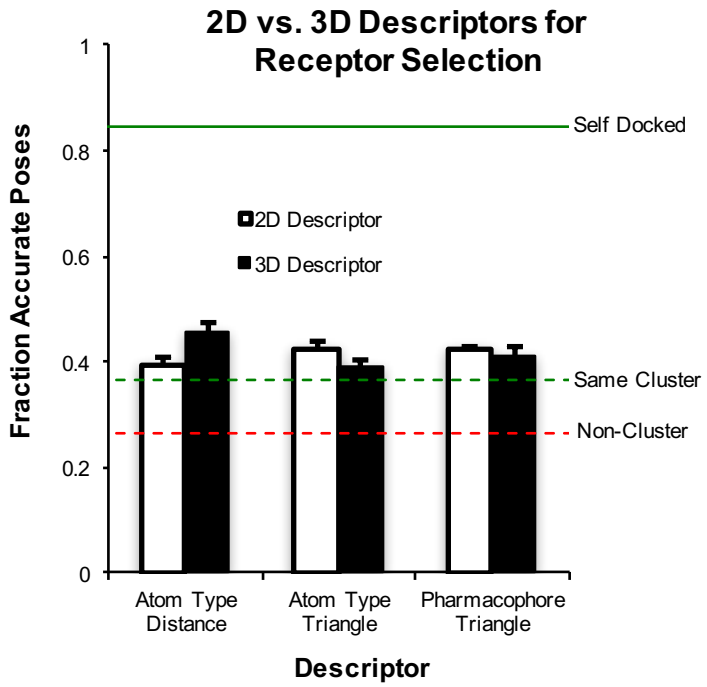
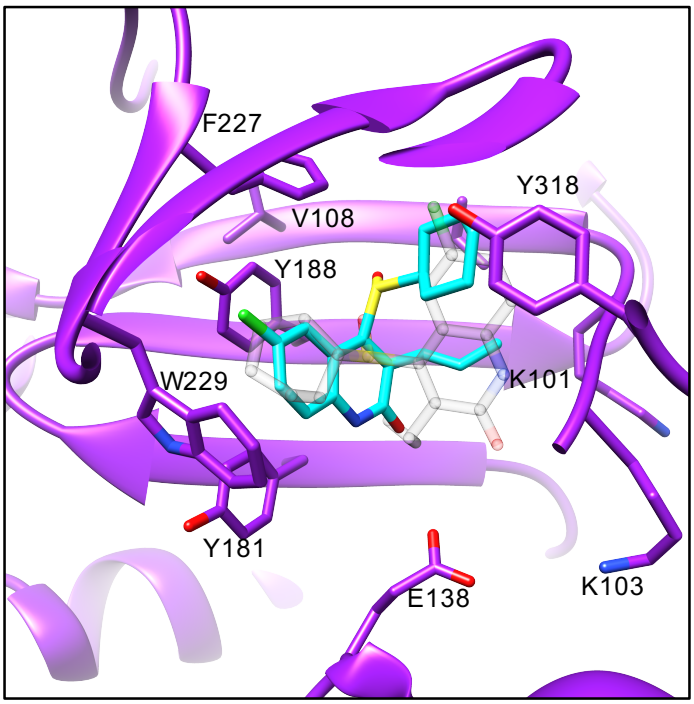


Figure 3.6. A comparison of receptor selection by descriptors available in both 2D and 3D formats from MOE. The Atom Type Distance (2D: TGD/3D: TAD), Atom Type Triangle (TGT/TAT), and Pharmacophore Triangle (gpiDAPH3/piDAPH3) descriptors were used.

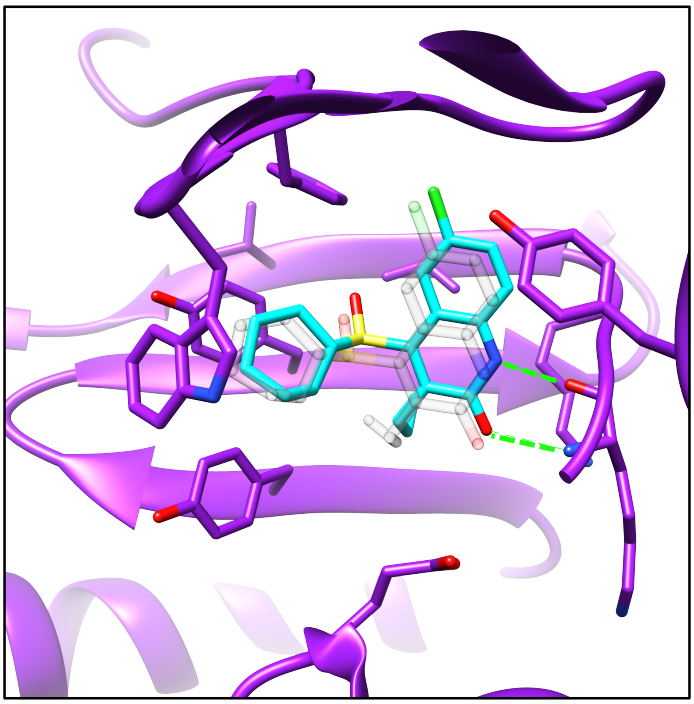
The impact receptor selection can play on docking accuracy is illustrated in **Figure 3.7**. Compound **17** self-docks accurately, with an RMSD < 1 Å (data not shown). Yet when docking into the representative receptor described previously (1KLM), the orientation of **17** relative to its solved binding pose was flipped, and this position cannot be corrected with minimization and molecular dynamics (**Figure 3.7a**). However, docking into the receptor selected by the top performing descriptor (ECFC_6, **Figure 3.7b**) yielded a ligand orientation very similar to that of the solved pose (RMSD = 0.68 Å). This docked pose included two hydrogen bonds with the backbone carbonyl and amine groups of K101 with the same orientation (within 0.2 Å and 11° between donor-acceptor pairs) as those found in the actual solved, bound state of the NNRTI. In this case ligand-similarity mediated receptor selection revealed relevant protein-ligand interaction motifs that were missed when docking into other receptors, and knowing these interaction patterns can provide a basis for structure activity relationship studies and rational compound optimization.

a 17 in 1KLM after docking and 1ns MD



RMSD from solved conformation: 4.98 Å

b 17 docked in ECFC_6 selected pocket (1C1C)



RMSD: 0.68 Å

Figure 3.7. Docking of 17 into a representative RT-NNRTI structure (a) and into a receptor selected based off of ligand similarity (b). Solved pose is shown in transparent grey for comparison, with hydrogen bonds are shown as green dashes.

Assessment of Receptor Selection for Active Recall

NNRTI with solved binding conformations are useful for assessing cross-docking accuracy, as the new docked pose can be readily compared back to that of the solved one. However, when considering novel molecules without known binding poses (such as when running a virtual screen to identify leads), the docking score is often used as a predictor of potency. The scoring function in Autodock Vina performs reasonably well when cross docking known NNRTI, as it enhanced prediction of accurate poses from the cross docking of the 87 NNRTI into both a single receptor (which it selected at a statistically significant rate of 48%), and for all of the non-self receptors from the set (**Figure 3.8**). However, when dealing with large chemical libraries it is very cumbersome to dock all potential ligands into 87 different receptors in order to find the very best scoring ligand-receptor pairs. Thus we sought to determine if the best descriptor at predicting receptors for NNRTI cross-docking accuracy could also enhance recall of known NNRTI from a set of inactive decoys.

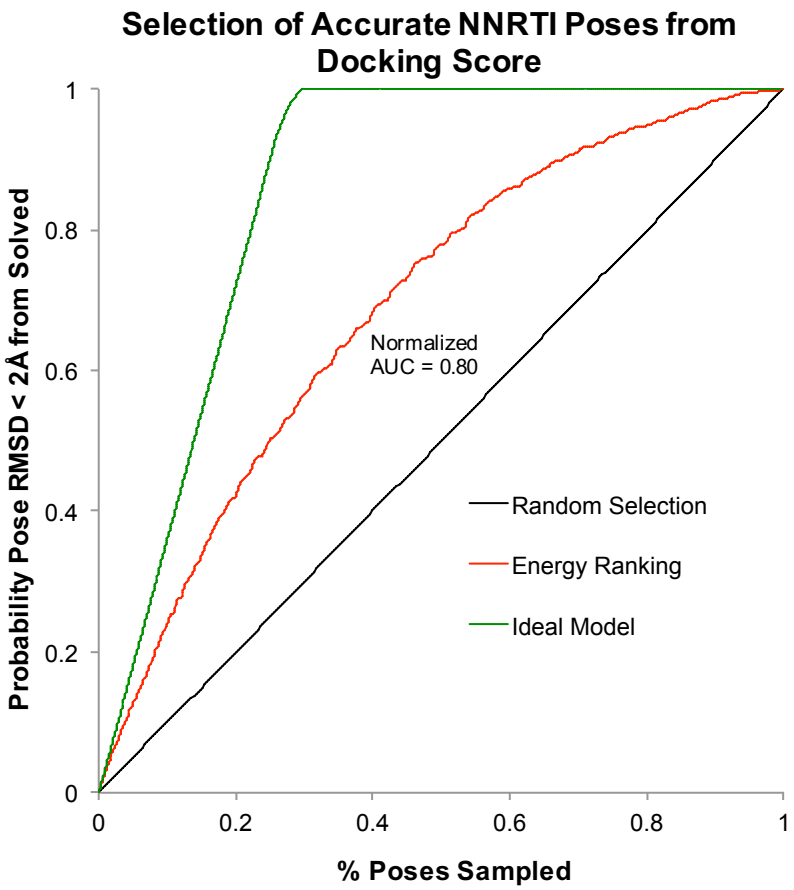


Figure 3.8. ROC curves showing recall of accurate poses of 87 solved NNRTI from cross-docking into all 87 available receptors using the docking score from AutoDock Vina.

A set of 1,653 NNRTI decoys were selected from the DUDE-E database (133) and combined with the set of known NNRTI to form a virtual library of 1740 test compounds. The library was constructed such that the 87 known actives would constitute 5% of the total number of compounds. Receiver operating characteristic (ROC) curves were then used to assess various methods of recalling the known NNRTI from the virtual library.

A set of similarity scores between a single target NNRTI (DLV) and all 1,740 of the molecules from the virtual library was created, using the top performing descriptor (ECFC_6) from the previous exercise. Then these were rank-ordered to determine if the similarity scores to DLV correlated with the 87 active NNRTI. These results were also compared to the rank-ordered docking scores created from docking the compounds into a single, representative receptor (1KLM, the receptor of DLV).

Multiple targets for similarity scoring and docking were also used. Applying a similar approach as outlined in **Scheme 3.1**, the ECFC_6 descriptor was implemented to create a database of similarity scores between all 1,740 compounds of the virtual library and each of the 87 known NNRTI. For each inactive compound in the virtual library only the top similarity score from the 87 created for each NNRTI were used. But for the active NNRTI in the virtual library only the top non-cluster similarity score was considered, as was done previously (i.e., for a compound in cluster **11** only the top similarity score between it and compounds from clusters **12-20** was used). These values were then rank-ordered to determine if higher similarity scores correlated with active NNRTI. Each of the library compounds was

also docked into the receptor of the NNRTI with the highest similarity, and this data was used to create an additional rank-ordered list of the subsequent docking scores.

The results (**Figure 3.9** and **Table 3.5**) show the advantage of using multiple NNRTI for similarity-based receptor selection for cross-docking. This approach captured a statistically significant 71% of the active NNRTI at a 1% sample rate, which was more than twice the enrichment of the next best method.

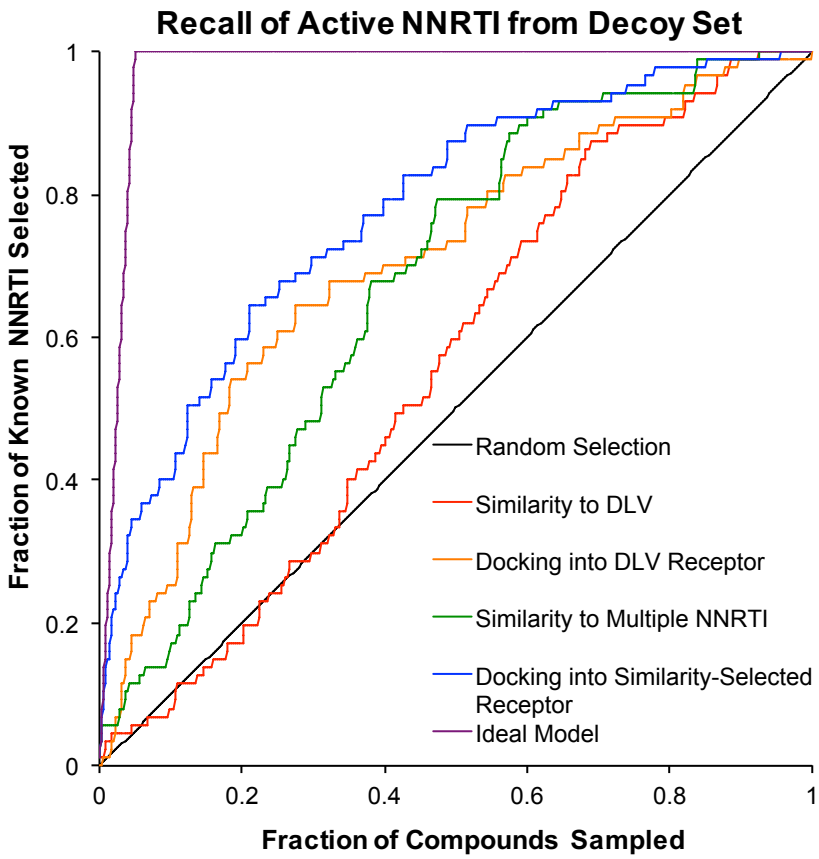


Figure 3.9. ROC curves showing recall of 87 known NNRTI from virtual library of 1740 NNRTI-like compounds.

Table 3.5. ROC curve data from recall of active NNRTI from decoy set

Selection Method	*Fraction of Ideal at 1% Sample*	*Normalized AUC*
Similarity to DLV	0.18	0.58
Docking into DLV Receptor	0.06	0.72
Similarity to Multiple NNRTI	0.29	0.69
Docking into Similarity-Selected Receptor	0.71*	0.80

* $p < 0.005$

3.4 Conclusions

The results suggest that a sequential application of ligand- and structure-based approaches by selecting a best receptor for docking *via* ligand-similarity scores can increase both the probability of selecting active NNRTI from a virtual screen and predicting the correct binding pose thereof. Additionally, using multiple targets and receptors of NNRTI for similarity scoring and docking improved virtual screening results, consistent with the variability in composition and binding of NNRTI. These results can be applied directly to guide hit discovery and lead optimization efforts of new NNRTI via receptor selection for docking, and the workflow outlined can also be used to assess, benchmark, and customize new descriptors for receptor selection in other drug systems targeting viruses other than HIV.

3.5 Methods

Analysis of NNRTI-Bound RT structures

One hundred and eighty-nine RT structures were downloaded from the RCSB PDB (117), and from these a subset of 87 were selected for analysis, as they included the wild type apoenzyme bound to a single NNRTI and had no missing residues within the NNRTI binding site. The p66 subunit of each of these structures was aligned to the p66 subunit of the wild type ternary (enzyme, DNA duplex, and incoming nucleotide) 1RTD (134) structure using the MatchMaker function in UCSF Chimera (135). This aligned coordinate system was used for visual inspection and comparison. The RMSD values of the protein backbone of the p66 subunit were determined along with that of all the atoms of the binding site residues (these were

residues 95, 100-103, 106, 179-181, 188-191, 227, 229, 234-237, and 318 from the p66 subunit and 138 from the p51 subunit) for all of the structures against one another using the Match function in UCSF Chimera, performing a realignment of all the other structures against each template before measurement. Binding pocket volumes were measured with Accelrys' Discovery Studio Site Finder function (136).

NNRTI Cross Docking Database

The 87 RT receptors and NNRTI were prepared for docking using AutoDock Tools 1.5.4 (137) and each ligand was then docked into each receptor using AutoDock Vina 1.1.2 (103) with a search space cube of 25 $Å^3$ around the center of mass of the native NNRTI bound to each receptor. Water molecules present in and around the binding pocket were removed, as they were not conserved and found in only 33 of the structures. Autodock Vina uses random conformations as starting positions for docking, so runs were performed in triplicate to reduce effects of any random artifacts. Docking poses were analyzed using ViewDock in UCSF Chimera and measured against the solved structure after realignment based on the binding site residues. All of the NNRTI were also docked using flexible side chains for the binding site residues into the DLV binding pocket (PDB ID: 1KLM). In all docking exercises only the top-scoring pose was kept.

Minimization and MD

The top-scoring pose of all 87 NNRTI docked into 1KLM were prepared for minimization and molecular dynamics simulations using VMD 1.9.1. (138) Residues within 20 Å of the binding pocket center were extracted and a solvent box of TIP3P water model with a 5 Å pad was added to each enzyme-drug complex. NAMD 2.9

(139) was used to minimize the structures for 10 ps followed by 1 ns of molecular dynamics, using timesteps of 1 fs. The simulation used an NPT ensemble at 310 K and a Langevin piston with a target pressure of 1 bar. The van der Waals cut-off distance was set to 11 Å and the particle mesh Ewald method was used for electrostatics. The resultant NNRTI binding poses were measured as before to determine the RMSD from the solved pose.

Similarity Clusters

The 87 compounds were split into 10 clusters based on maximum dissimilarity between clusters using MACCS descriptors in Pipeline Pilot (129). When comparing compounds for receptor selection only NNRTI outside of the cluster of the query molecule were considered, so as to eliminate selection of receptors from closely related analogs.

Chemical Descriptor Similarity Scores

The ligands from the 87 aligned PDB structures were extracted to create a molecular database of their solved conformations. In order to make comparisons of compounds based on their non-bound low energy conformations, the ligands were first converted to 2D structures to remove any bias from their existing conformations, and then LowModeMD (140) from MOE (130) was implemented to create sets of low energy conformations. We used an MMFF94x potential with a dielectric constant of 1 and a non-bonded interaction cutoff of 8 Å, a rejection limit of 100, RMS gradient of 0.1, iteration limit of 1000, and an MM energy minimization iteration limit of 200. Only the 10 lowest energy conformations with an RMSD

similarity lower limit of 1 Å were kept in order to ensure a broad sampling of possible conformations for similarity searching.

MOE was used to calculate values of the 1D (number of heavy atoms, molecular weight, logP) and shape-based descriptors (diameter, volume, density, globularity, van der Waals area, van der Waals surface area, can der Waals volume, and hydrophobic surface area) for all NNRTI. These values were then used to create normalized similarity ratios for each pair of the compounds.

The 2D descriptor comparisons were made using three programs: OpenBabel (131) (for FP2, FP3, and FP4), MOE (for BIT_MACCS, gpiDAPH3, TGD, and TGT), and Pipeline Pilot (for MACCS, ECFC_2, ECFC_4, ECFC_6, ECFP_2-_6, FCFP_2-_6, and FCFC_2-_6). The 3D descriptor comparisons used the conformations previously generated and were also done using three different programs: OpenBabel (for Spectrophpres), ACPC (132), and MOE (for piDAPH3, piDAPH4, ESshape3D, ESshape3D_HYD, TAD, and TAT). Tanimoto similarity coefficients were created for each compound pair using each descriptor.

Selection of Receptors Based on Ligand Similarity

For each NNRTI, the most similar compound (excluding those from their own cluster) was selected based on the maximum similarity score for a given descriptor. The results of docking each NNRTI into the binding pocket of the receptor of the most similar compound were then tabulated for each descriptor, and the fraction of accurate (RMSD < 2 Å from the solved binding pose) poses for each NNRTI-receptor set was subsequently calculated.

Receptor Selection and Docking of Decoys

A random subset of 1,653 NNRTI decoys were selected from the DUDE-E database (133) and combined them with the 87 known NNRTI for a set of 1740 compounds for virtual screening. The top performing descriptor (ECFC6) was used to select the receptors of the most similar NNRTI for docking as described previously. Once docked, receiver operating characteristic (ROC) curves were made in Microsoft Excel to compare recall enhancement of the of the 87 known NNRTI from the entire set of 1740 compounds. ROC curves were made comparing the use of just the descriptor itself (with a single or multiple NNRTI-similarity targets), the scores from docking into the receptors selected by the descriptor, or into a single receptor.

Statistical Analysis

A one-tailed, homoscedastic Student's *t*-test was applied with a significance threshold of 0.005 using Microsoft Excel to the receptor selection results from each descriptor. The cross-docking results for each receptor in the entire set were used to determine the standard deviation of the population. For the virtual screening exercise the same statistical test was used, but the standard deviation of the population was calculated from ROC curves made from docking into three different receptors.

CHAPTER 4:

Free Energy Perturbation Guided Lead Optimization of 7-azaindole Non-Nucleoside HIV-1 Reverse Transcriptase Inhibitors

Portions of the research presented in this chapter are included in a manuscript submitted for publication in *Bioorganic & Medicinal Chemistry Letters*.

Richard A. Stanton, Xiao Lu, Mervi Detorio, Catherine Montero, Emily T. Hammond, Maryam Ehteshami, Robert A. Domaoal, James H. Nettles, Michel Feraud and Raymond F. Schinazi.

"Discovery, characterization, and lead optimization of 7-azaindole non-nucleoside HIV-1 reverse transcriptase inhibitors."

Contributions to this chapter:

Richard A. Stanton cloned and purified the enzymes, performed all cell-free RT inhibition assays and *in silico* calculations, and wrote and edited the chapter.

Xiao Lu synthesized and purified the analogs of the lead compound.

Mervi Detorio and Catherine Montero performed the cell-based antiviral and cytoxicity assays.

Emily T. Hammond, Maryam Ehteshami, and Robert A. Domaoal contributed training and assistance towards the cloning and purification of the enzymes and cell-free RT inhibition assays.

Michel Feraud supplied the 7-azaindole compound library.

Raymond F. Schinazi conceived of the project, provided all experimental resources, and assisted editing of the chapter.

4.1 Abstract

Previous work identified a lead non-nucleoside HIV-1 reverse transcriptase (RT) inhibitor (NNRTI) from a library of 585 molecules built off of a common 7-azaindole core. While the lead compound inhibited RT with submicromolar potency (IC_{50} = 0.73 μM), it is still several log fold less active than current NNRTI. In an effort to improve potency, free energy perturbation calculations were used to computationally assess potential modifications prior to synthesis. This led to the development of a new analog with two-fold increase in potency against RT (IC_{50} = 0.36 μM), validating this approach for future optimization efforts.

4.2 Introduction

Screening of a library of 585 compounds built off of a 7-azaindole core against HIV-1 revealed several promising hits, the most potent of which act as non-nucleoside reverse transcriptase inhibitors (NNRTI), as discussed in **Chapter 2**. These 7-azaindoles demonstrated potent antiviral activity and also inhibited DNA polymerization by HIV reverse transcriptase (RT) while maintaining low toxicity against several cell lines tested, and even maintained some activity against clinically relevant RT mutants.

While most promising compounds in the library inhibited RT with submicromolar IC_{50} values, they are still more than fifty-fold less potent than the FDA-approved NNRTI EFV (**Table 2.1**), which is itself ~ three-fold less active than the second generation compound RPV (141). Newer experimental NNRTI have even demonstrated potency beyond that of RPV, as well (142). So while an NNRTI built off of a novel scaffold is a promising development that may offer advantages over

existing drugs, the potency of the 7-azaindoles must be improved before such compounds can be considered as clinical candidates.

Optimization of the lead 7-azaindole compounds is a complicated endeavor, given the versatility of the core and the wide range of modifications it can accommodate. There are a variety of computational means to estimate the binding potency of ligands, but their accuracy varies depending on the target (143). The following work outlines the validation of one such method, free energy perturbation (FEP), as a tool to optimize the 7-azaindoles as NNRTI.

4.3 Results and Discussion

Lead optimization of the 7-azaindole NNRTI focused on compound **8**, as it demonstrated submicromolar potency against HIV replication in cell culture and RT directly in cell-free assays and was nontoxic against all three cell lines tested (**Table 2.1**). Further, there are already several closely related analogs within the library that could be used to explore the structure activity relationship of the compound and validate computational methods of predicting potency (**Fig. 4.1**).

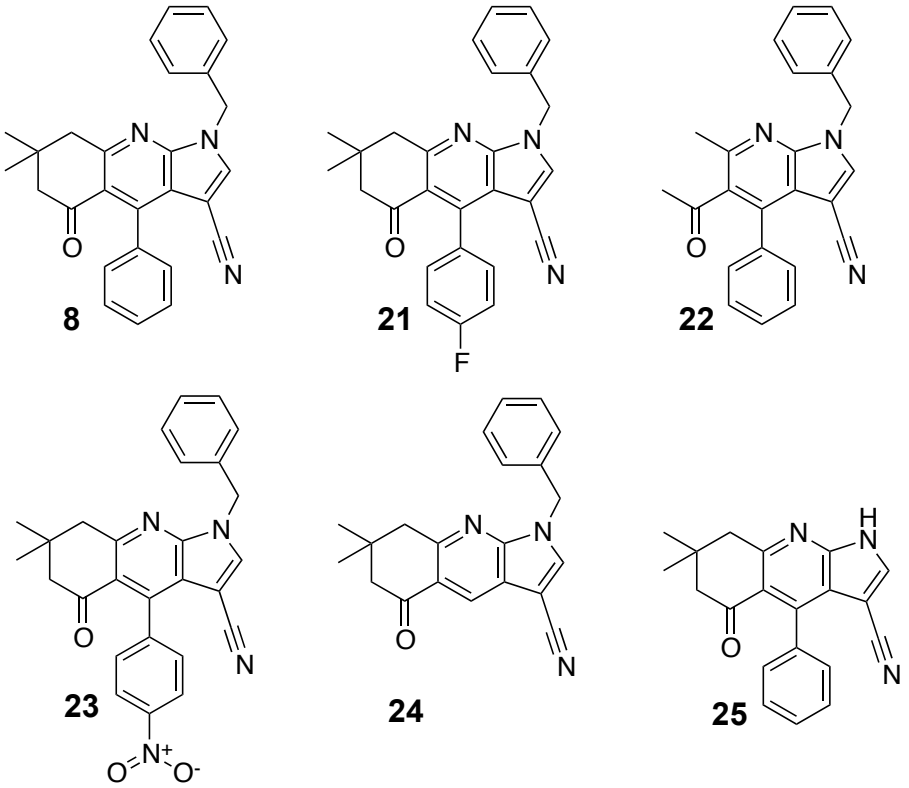


Figure 4.1. Structures of 8 and related analogs from 7-azaindole library.

Docking and Scoring

The analogs of **8** were all docked into the NNRTI binding pocket of receptors selected from the PDB based on their ECFC_6 similarity scores (detailed in **Chapter 3**).

The scoring function from the program used for docking (Autodock Vina) estimates the change in Gibbs free energy of binding (ΔG) for a drug to its target (103). The dissociation constant K_d is related to ΔG by **Equation 1**, where R is the ideal gas constant and T represents the temperature:

$$\Delta G = -RT(\ln K_d) \quad \textbf{Equation 4.1}$$

The ratio of the dissociation constants for two compounds (referred to as 1 and 2) can be determined by the difference between their respective ΔG values ($\Delta\Delta G$) via the Boltzmann distribution law:

$$\Delta\Delta G = -RT(\ln \frac{K_{d2}}{K_{d1}}) \quad \textbf{Equation 4.2}$$

A negative $\Delta\Delta G$ value means that compound 2 has a smaller K_d and therefore has a higher binding affinity for the target than compound 1, which should lead to an increase in potency assuming both compounds equivalently act as inhibitors upon binding.

Since the Autodock Vina scoring function was able to differentiate between known NNRTI and inactive decoy combines on a large scale (**Chapter 3**, **Fig. 3.9**), it was assessed to determine if it could also predict the potency of potential 7-azaindoles prior to their synthesis, using the analogs of **8** as a test set. Unfortunately the $\Delta\Delta G$ values did not agree with the experimental data for all of these compounds.

Compounds **24** and **25** were both predicted to have increased potency based on their $\Delta\Delta G$ scores, contrary to the experimental results (**Table 4.1**).

Table 4.1. Docking predicted and experimental activity of analogs of **8** (± SD)

Compound	Predicted		Experimental Results*	
	ΔΔG (kcal/mol)	Fold Change	RT IC_{50} (μM)	Fold Change
21	1.1 ± 0.06	6.3	2.9	3.9
22	2.1 ± 0.06	30	>100	>100
23	0.10 ± 0.10	1.2	1.4	1.9
24	-1.0 ± 0.10	0.20	12.2	16.6
25	-2.4 ± 0.06	0.02	53.0	72.5

*Average of three independent experiments measuring inhibition of RT polymerase activity in FRET-assay.

It has been recognized that the accuracy of docking algorithms and their scoring functions can vary from target to target, and therefore Autodock Vina may not be best suited for this particular case (110). Since the scoring function used here considers only rigid NNRTI binding pockets, it may overlook subtle side chain rearrangements that occur to accommodate different ligands as well as any contributions from the solvent (144). To address these shortcomings a more sophisticated approach to estimating binding energy changes was assessed.

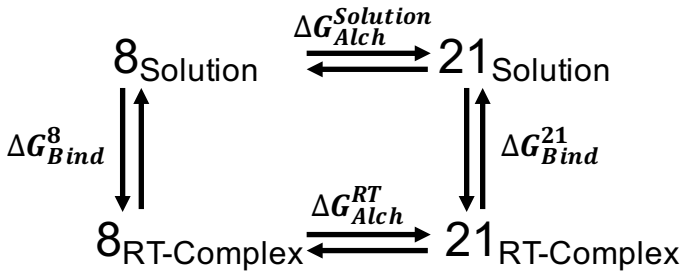
Validation of Free Energy Perturbation

Free energy perturbation (FEP) is a computational method of estimating the difference in binding energy between two structurally similar small molecules. FEP has been used extensively on other NNRTI (145; 146) and has demonstrated utility in predicting potency changes related to modifications of a lead compound (147).

In FEP one molecule is converted to another "alchemically" via a series of discreet non-physical pseudo-atom intermediates, described as functions a morphing parameter λ which increases from 0 (representing the original molecule) to 1 (the final molecule). For instance, converting the para-hydrogen atom in **8** to the fluorine in **21** over a series of ten λ-steps would involved changing both the atom and bond characteristics, such as increasing the atomic mass of the atom from 1 to 19 Da (or 1.8 Da per λ-step), increasing the van der Waals radius from 120 to 150 pM (3 pM per λ-step), increasing the bond length from 1.08 to 1.36 Å (0.028 Å per λ-step), decreasing the partial charge from 0.115 to -0.204 *e* (-.0319 *e* per λ-step), and so on. For each level of λ a molecular dynamics (MD) simulation is run to provide an ensemble of structures from which an estimate of the change in energy

from the previous level is derived. Integrating these changes over all levels of λ gives the total alchemical free energy change (ΔG_{Alch}) for the full conversion of the original to the new molecule.

FEP is a more computationally rigorous process than the Autodock Vina scoring function, but it is better suited for comparing changes in binding affinity for closely related molecules since it can consider conformational changes within the binding pocket as well as contributions of the solvent on the free energy of binding. To that end, FEP calculations were run for compounds in solution and in the NNRTI binding pocket. These calculations yield two ΔG_{Alch} values, as shown in the thermodynamic cycle for converting **8** to **21** (**Scheme 1** and **Fig. 4.2**).



Scheme 4.1. Thermodynamic cycle of alchemical FEP conversion of compound **8** to **21**.

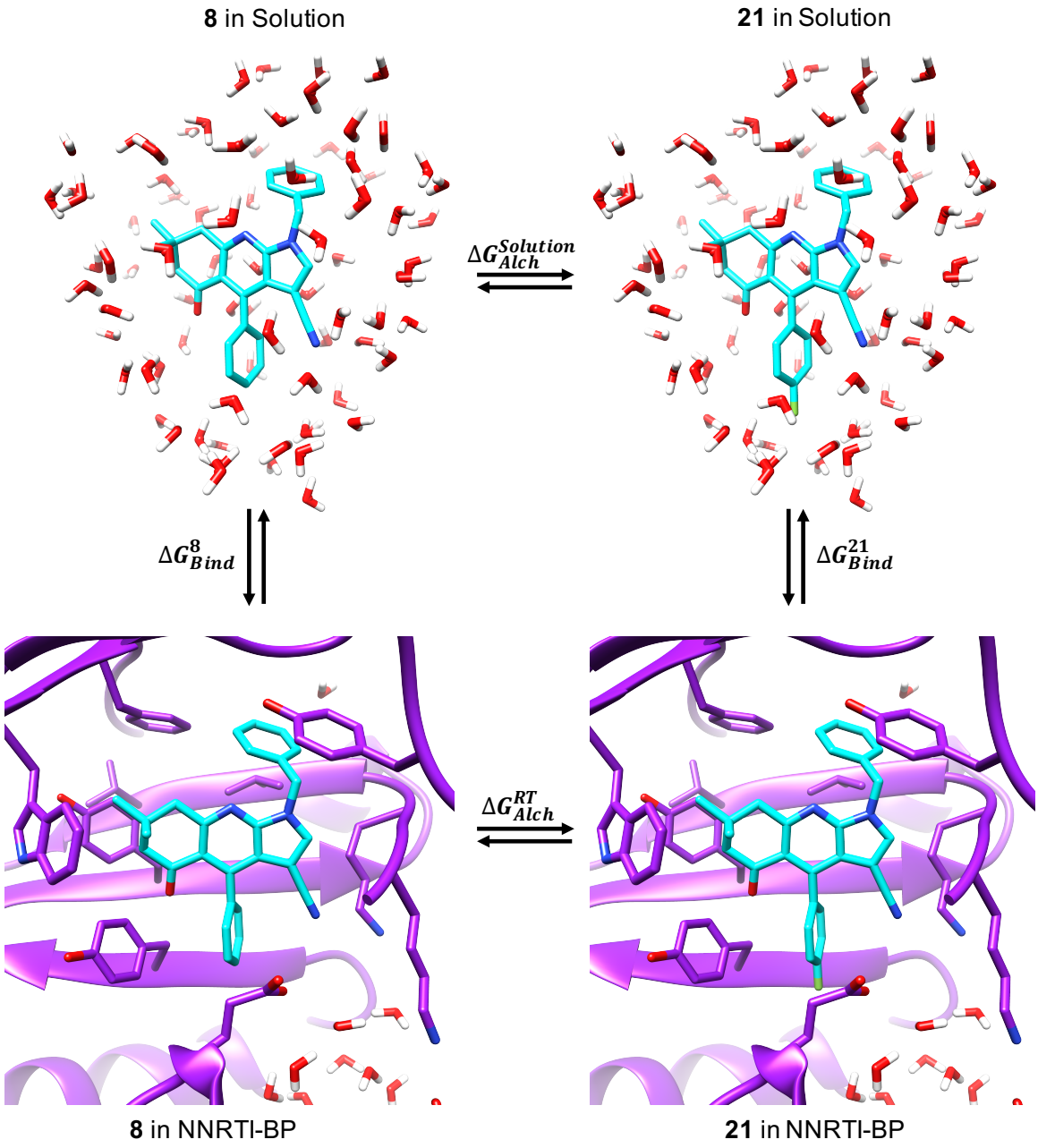


Figure 4.2. Thermodynamic cycle of structures for FEP conversion of compound **8** to **21**.

As discussed above, it is challenging to accurately predict the absolute free energy of binding (i.e., the ΔG_{Bind} values from **Scheme 4.1**) for a ligand to a protein. This is due to a variety of factors, including the difficulty of estimating the pathway of the conformational changes that are induced by the ligand upon binding, which occurs on a large scale with NNRTI. Estimating the ΔG_{Alch} values for two closely related compounds is simplified using the assumption that these conformational trajectories of ligand binding for the two compounds was equivalent, and thus one only has to estimate the energies related the bound state. And since the thermodynamic cycle represented in **Scheme 4.1** is a closed loop, the $\Delta\Delta G$ value for the differences in binding energies between the two NNRTI can be determined using only the values of ΔG_{Alch}:

$$\Delta\Delta G = \Delta G^{21}_{Bind} - \Delta G^{8}_{Bind} = \Delta G^{RT}_{Alch} - \Delta G^{Soln}_{Alch} \qquad \textbf{Equation 4.3}$$

The change in energy for the ten λ steps for the conversion of **8** to **21** is plotted in **Figure 4.3**, and resulted in an overall $\Delta\Delta G$ of 0.49 kcal/mol, which translates to a 2.2 fold decrease in affinity for the NNRTI binding pocket, per **Equation 4.2**. The experimental potency against RT by **21** was 3.9 times greater than that of **8**, in qualitative agreement with the FEP prediction. The FEP predictions for the other analogs of **8** also agreed qualitatively with the experimental results, as the larger fold-change decreases in IC_{50} values for **22**, **24**, and **25** versus the smaller changes for **21** and **23** were reflected by both the direction and scale of the $\Delta\Delta G$ value changes (**Table 4.2**).

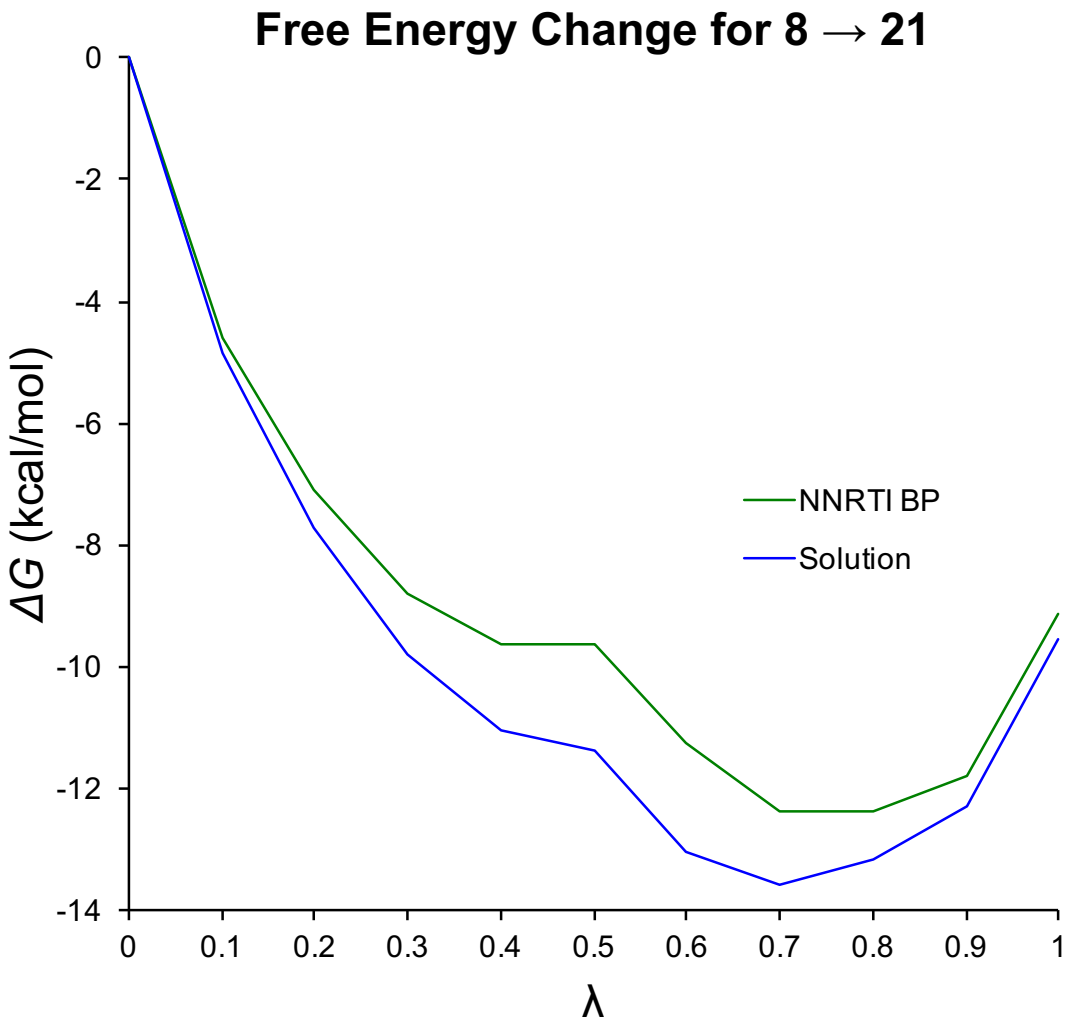


Figure 4.3. Plot of free energy change (ΔG) calculated via FEP for conversion of **8** to **21** using 10-λ steps.

Table 4.2. FEP predicted and experimental activity of analogs of **8** (± SD)

Compound	Predicted		Experimental Results*	
	ΔΔG (kcal/mol)	Fold Change	RT IC_{50} (μM)	Fold Change
21	0.49 ± 0.50	2.2	2.9	3.9
22	28 ± 0.67	>100	>100	>100
23	0.65 ± 0.55	2.9	1.4	1.9
24	6.2 ± 1.3	>100	12.2	16.6
25	9.9 ± 0.71	>100	53.0	72.5

*Average of three independent experiments measuring inhibition of RT polymerase activity in FRET-assay.

Optimization of Lead Compound

The experimental results discussed in **Chapter 2** suggested that the lower phenyl ring of **8** is important for NNRTI activity. This ring is also synthetically accessible, so a series of potential fluoro-, chloro-, bromo-, and methyl-substitutions to that position on **8** were evaluated with FEP to identify analogs with increased potency against RT.

Six of the analogs were predicted to have improved potency (**Table 4.3**). Three of those were selected for synthesis (compounds **28**, **31**, and **43**) as they represented different substitution types (fluoro-, chloro-, and methyl-, respectively). Additionally, the increase in binding affinity predicted for compounds **28** and **31** was due to the change in free energy within the binding site (a more negative ΔG_{Alch}^{RT} term from **Equation 4.3**), while for **43** the change was driven by an increase in the change of free energy in solution (a more positive $\Delta G_{Alch}^{Solution}$ term from **Equation 4.3**), so several mechanisms of predicted potency changes could be assessed with the molecules.

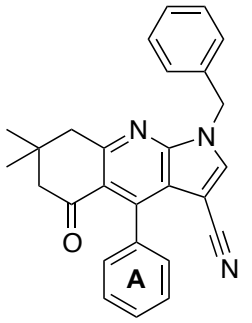
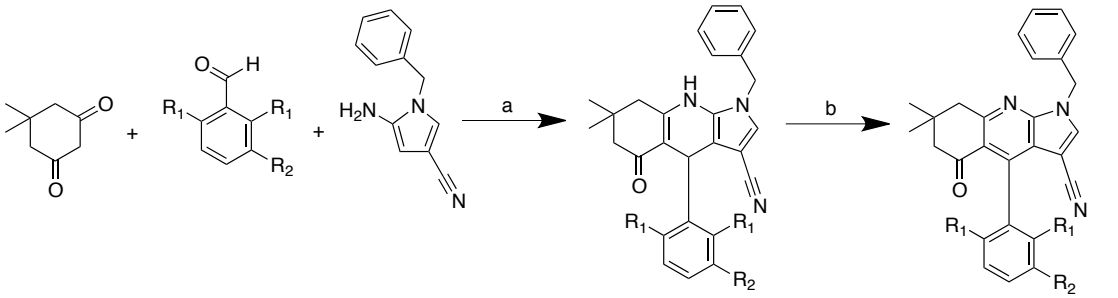


Table 4.3. FEP predicted activity of analogs of **8** (± SD)

Compound	A	ΔG_{Alch}^{RT} (kcal/mol)	$\Delta G_{Alch}^{Solution}$ (kcal/mol)	$\Delta\Delta G$ (kcal/mol)	Fold Change*
26	2-fluorophenyl	-0.06 ± 0.48	2.5 ± 0.43	-2.5 ± 0.45	0.016
27	3-fluorophenyl	-8.1 ± 0.54	-9.0 ± 0.42	0.91 ± 0.48	4.4
28	2,6-difluorophenyl	-3.6 ± 0.67	-1.3 ± 0.66	-2.3 ± 0.67	0.026
29	3,5-difluorophenyl	-54 ± 0.52	-56 ± 0.52	2.6 ± 0.52	72
30	2-chlorophenyl	1.1 ± 0.56	0.14 ± 0.75	1.0 ± 0.66	5.0
31	3-chlorophenyl	-9.0 ± 0.48	-7.0 ± 0.55	-2.0 ± 0.45	0.036
32	4-chlorophenyl	-5.0 ± 0.43	-6.4 ± 0.47	1.4 ± 0.45	10
33	2,6-chlorophenyl	7.7 ± 0.74	3.8 ± 0.73	4.0 ± 0.74	620
34	3,5-dichlorophenyl	-9.2 ± 0.62	-16 ± 0.60	7.2 ± 0.61	11,000
35	2-bromophenyl	4.2 ± 0.59	5.0 ± 0.75	-.80 ± 0.67	0.27
36	3-bromophenyl	0.49 ± 0.59	-4.3 ± 0.56	4.8 ± 0.58	2400
37	4-bromophenyl	-2.1 ± 0.53	-5.3 ± 0.56	3.2 ± 0.54	180
38	2,6-dibromophenyl	7.4 ± 0.77	-1.9 ± 0.78	9.22 ± 0.78	3,100,000
39	3,5-dibromophenyl	-2.5 ± 0.65	-3.8 ± 0.54	1.30 ± 0.60	8.1
40	2-methylphenyl	1.6 ± 0.68	1.8 ± 0.61	-0.17 ± 0.65	0.76
41	3-methylphenyl	-6.5 ± 0.69	-9.2 ± 0.49	2.8 ± 0.59	91
42	4-methylphenyl	-3.4 ± 0.57	-7.2 ± 0.45	3.8 ± 0.51	470
43	2,6-dimethylphenyl	-0.9 ± 0.83	2.2 ± 0.69	-3.1 ± 0.76	0.006
44	3,5-dimethylphenyl	-18 ± 0.68	-20 ± 0.54	2.4 ± 0.61	49

*Predicted fold change in potency from **8**.

The synthesis of these three compounds was carried out according to **Scheme 4.2**. One-pot, three-component condensation of dimedone **45** with different substituted benzaldehydes **46a-c** and 5-amino-1-benzyl-1H-pyrrole-3-carbonitrile **47** in refluxing methanol was followed by 2,3-dichloro-5,6-dicyanobenzoquinone (DDQ) oxidation, providing target compounds in fair to good yields (148).



45 **46a** $R_1 = CH_3$, $R_2 = H$ **47** **48a** $R_1 = CH_3$, $R_2 = H$ **28** $R_1 = F$, $R_2 = H$
46b $R_1 = F$, $R_2 = H$ **48b** $R_1 = F$, $R_2 = H$ **31** $R_1 = H$, $R_2 = Cl$
46c $R_1 = H$, $R_2 = Cl$ **48c** $R_1 = H$, $R_2 = Cl$ **43** $R_1 = CH_3$, $R_2 = H$

Scheme 4.2. Reagents and conditions for synthesis of analogs of **8**: (a) Ar, M.S. (3 Å), MeOH reflux 4-6 h, 42-48%; (b) DDQ, DCM, rt, 1-2 h, 50-85%.

After synthesis, the compounds were tested for RT inhibition, antiviral potency, and cytoxicity as described in **Chapter 2** (data in **Table 4.4**). Compound **28** demonstrated improved potency against RT, ~two-fold greater than **8** (IC_{50} = 0.36 μM), though it did have slightly decreased antiviral potency. Compound **31** exhibited decreased RT inhibition but was still active against HIV replication (EC_{50} = 1.0 μM). Compound **43** showed little RT inhibition and had no detectable anitiviral effect. Unfortunately, all three new compounds had detectable toxicity, unlike the nontoxic lead compound, **8**.

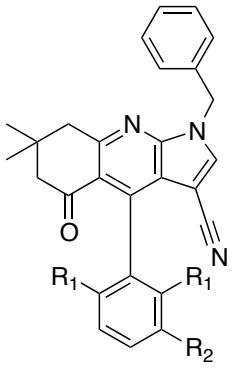


Table 4.4. Activity and toxicity of synthesized analogs of **8** (all values in μM, ± SD)

Compound	R_1	R_2	RT Inhibition (IC_{50})*	Antiviral Potency (EC_{50})**	Cell Cytoxicity (CC_{50})		
					PBM	CEM	VERO
8	H	H	0.73 ± 0.32	0.35 ± 0.26	>100	>100	>100
28	F	H	0.36 ± 0.01	0.46 ± 0.23	>100	>100	57.6
31	H	Cl	5.8 ± 0.95	1.0 ± 0.05	61.4	33.6	>100
43	CH_3	H	26 ± 12	>100	>100	>100	11.3

*Average of three independent experiments measuring inhibition of RT polymerase activity in FRET-assay.

**Average of at least three independent experiments in HIV-infected PBM cells.

The increased potency of **28** does not appear to be due to any direct interactions with the fluoro-substitutions, but may instead be due to stabilization of a conformation that allows increased π-π stacking interactions with side chain of Y181, as both the angle and distance between the two rings decreased after the MD simulations versus **8** (**Fig. 4.4**). Unfortunately, this interaction may be vital for the anti-RT activity of the compound, as **28** showed no activity against the Y181C RT mutant.

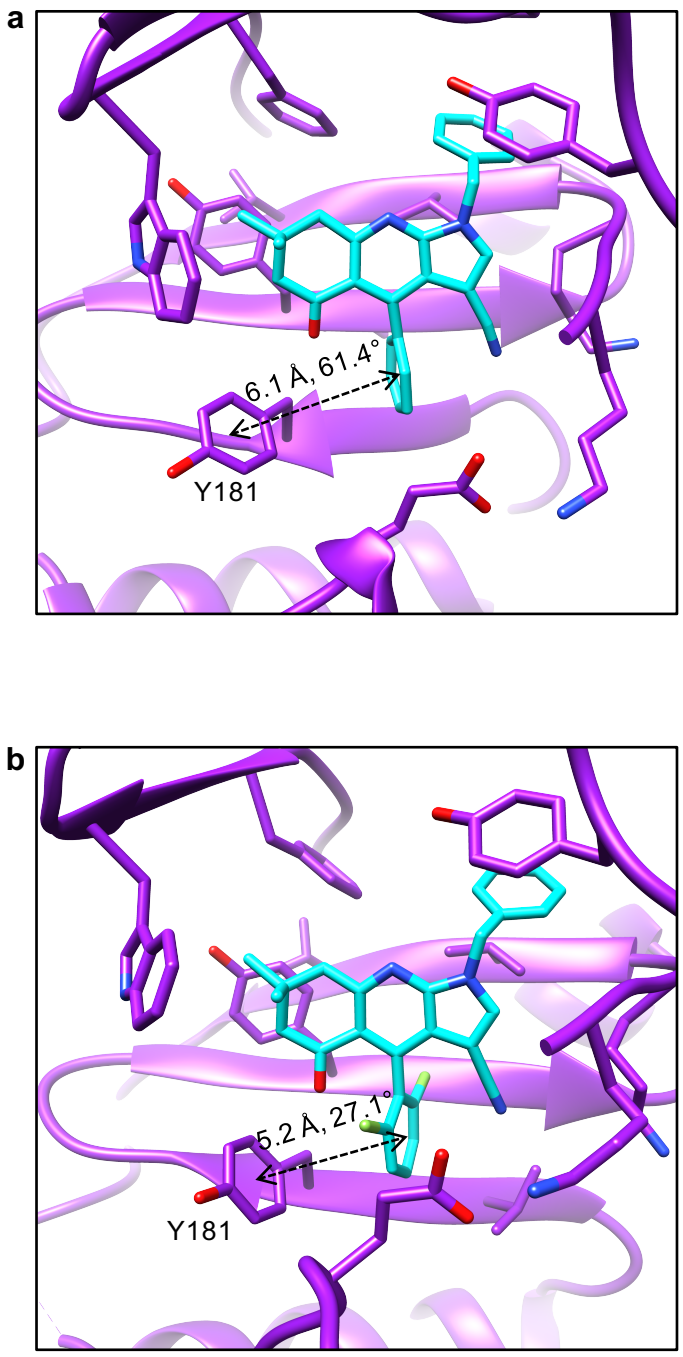


Figure 4.4. Compound **8** after short MD simulation (a), **28** shows increased π-π stacking interactions with side chain of residue Y181 in NNRTI-BP after FEP simulation.

While the FEP-directed lead optimization of **8** failed to produce a more viable drug candidate, the results do provide insights into the utility and limitations of the approach for further lead optimization. The increased anti-RT potency of **28** suggests that compounds with lower $\Delta\Delta G$ driven by lower ΔG_{Alch}^{RT} should be considered before compounds such as **43**, which had lower $\Delta\Delta G$ values as a result of a higher $\Delta G_{Alch}^{Solution}$ value.

The lack of increased antiviral potency of **28** in the cell-based assay, contra its increased anti-RT potency versus **8**, suggests that factors such as protein binding, rates of cell entry or exit, or metabolism may have effected the intracellular concentration of the compound. Such external factors are not considered by FEP, which only focuses on drug binding.

Additionally, the lack of activity of **43** may have been due to the observed poor solubility of the compound. In retrospect this is in line with the increase in the $\Delta G_{Alch}^{Solution}$ value predicted from the FEP conversion of **8** to **43**, which would imply that more energy is required to solubilize the latter. While a more positive $\Delta G_{Alch}^{Solution}$ term does lead to a more negative $\Delta\Delta G$ (per **Equation 3**) value and therefore suggest increased binding affinity, the FEP calculations are built on the assumption that the compound has already entered the binding pocket, which may not occur when it has poor solubility.

Another shortcoming of the FEP approach used is that it relies on MD simulations to estimate ligand-protein interaction energies. The docked pose of **8** suggests that binding is largely driven by hydrophobic interactions such as London Dispersion Forces (LDF), which are very difficult to accurately estimate

computationally (149). LDF occur because of temporary dipole-dipole interactions induced by the motion of electrons, while the MD simulations used in the FEP calculations do not include electronic motion, and instead account only for the motions of whole atoms with fixed diameters and partial charges. The absence of electrons in MD simulations means that the energetic contributions of LDF to binding are not explicitly enumerated and are instead generalized as a function of distance. Including electronic motion to determine the LDF would require much more complex quantum mechanics simulations, which would in turn require much more resources and time for calculations, effectively negating the advantages gained through the rapid evaluation of potential modifications via FEP.

More generally to NNRTI, the relationship between binding affinity and inhibitor potency is not always directly correlated. So even if FEP could predict changes in binding affinity with exact accuracy, this may not lead to more potent NNRTI. As was discussed in **Chapter 1**, the binding affinity of the first generation NNRTI efavirenz (EFV) is the same for wild type RT as it is for the K103N mutant, yet the potency of the drug drops >100-fold (59).

4.4 Conclusions

In spite of the limitations of FEP, it was able to predict that the di-fluoro modification of compound **28** would make it more potent against RT than **8**. And while **28** is still more than several-fold less active than existing NNRTI, the results do provide a guide for interpreting the FEP calculations of other modifications off of the 7-azaindole core, which could lead to more drastic increases in potency.

4.5 Methods

Free Energy Perturbation (FEP)

Compound **8** was docked into the NNRTI binding pocket of PDB ID: 2B6A using AutoDock Vina 1.1.2 (103). This docked structure was then prepared for minimization and molecular dynamics simulations using VMD 1.9.1.(138) Residues within 20 Å of the binding pocket center were extracted and the terminal residues were neutralized. A solvent box of TIP4P water with a 5 Å pad was added to the enzyme-drug complex. The CHARMM force field(150) was used to parameterize the protein while topology files for the drugs were created using ParamChem.(151) NAMD 2.9 (139) was used for the simulations, and after an initial conjugate gradient minimization of 10 ps the FEP protocol was implemented for both forward and reverse perturbations. Ten λ-windows were used for the perturbations, each running for 1 ns of MD simulation. The simulation used an NPT ensemble at 310 K and a Langevin piston with a target pressure of 1 bar. The van der Waals cut-off distance was set to 11 Å and the particle mesh Ewald method was used for electrostatics. The perturbation simulations were run in both the enzyme and in water alone, and the total free energy change of binding ($\Delta\Delta G$) was estimated by subtracting the energy change occurring in water from that occurring in the enzyme, and fold changes in affinity were predicted using a Boltzmann distribution (144). Errors were estimated using the simple overlap sampling method from the ParseFEP toolikit (152).

CHAPTER 5:

Concluding Remarks

Contributions to this chapter:

Richard A. Stanton cloned and purified the enzymes, performed all cell-free RT inhibition assays and docking calculations, and wrote and edited the chapter.

Franck Amblard suggested several new molecules for *in silico* assessment, including compound **49** which is discussed in the chapter.

5.1 Introduction

The discovery and characterization of 7-azaindoles as HIV non-nucleoside reverse transcriptase inhibitors (NNRTI) is an important first step towards addressing some of the shortcomings of current antiretroviral treatments, as the unique scaffold could confer advantages over existing dugs. However, the compounds would need to have increased potency against wild type and mutant viral strains to be seriously considered as clinical candidates, while remaining safe and tolerable. Ideally they would also be able to penetrate viral reservoirs as well. The following section outlines the future work that needed to help achieve those goals.

5.2 Increasing Potency

While the most promising 7-azaindoles had submicromolar antiviral potency and no detectable toxicity against multiple cell lines, their therapeutic indices are still several log fold less than existing NNRTI. After identifying **8** as the strongest lead candidate, optimization efforts to improve potency began, focusing on the lower phenyl ring off of position R4 of the 7-azaindole scaffold (**Fig. 2.1**). Modifications of this ring were chosen because of its synthetic accessibility, making it ideal for a proof of principle study of the utility of free energy perturbation (FEP) approaches to produce more potent compounds. FEP calculations suggested several substitutions to improve binding affinity, which led to the synthesis of three new compounds. One of them (**28**) did have increased potency against RT activity in a cell-free assay, but showed no improvements in the cell-based antiviral screen.

An obvious step to help drive further optimization of **8** would be to determine its bound conformation to RT through x-ray crystallography. While **8** was docked following the protocol detailed in **Chapter 3**, verifying the binding pose experimentally would show true ligand-protein interaction patterns, reveal logical positions for modification, and would even help increase FEP accuracy by providing a relevant starting conformation.

Given the dynamic nature of NNRTI binding, though, solving the crystal structure would not necessarily provide an unambiguous explanation of the structure-activity relationship of the 7-azaindoles. Comparing two solved structures of nevirapine (NVP) bound to RT reveals differences within both the binding pocket and the structure at large, making it difficult to draw firm conclusions as to specific interactions and the allosteric mechanism of inhibition (**Figure 5.1**). The rearrangement of the K103 and Y318 side chains in the two structures underscore the variability of the pocket and suggest that a single solved structure of an NNRTI bound to RT does not always provide the full picture of binding and inhibition. Nonetheless, having direct evidence of a binding orientation of **8** within the pocket would be very useful for further optimization.

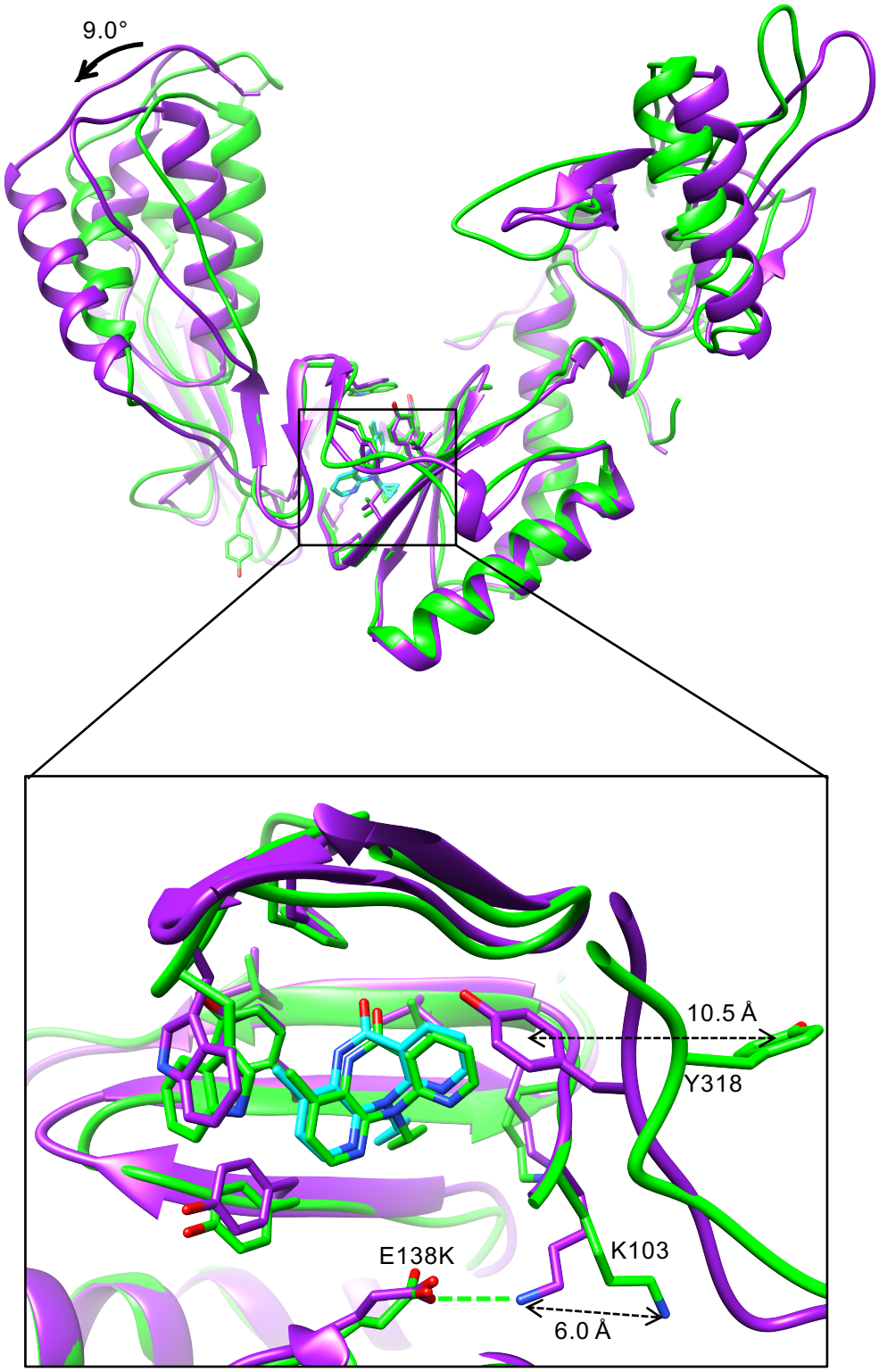


Figure 5.1. Superposition of two wild type RT-NVP bound structures (PDB ID: 1VRT (purple) and 3HVT (green)) shows both large scale (movement of thumb region) and binding pocket (inset, movement of Y318 and K103-E138K hydrogen bond displacement) conformational differences.

Modifications at other positions of **8** also should also be explored. Several dozen have already been assessed through FEP and of those, compound **49**, which includes a sulfonamide linker, was predicted to have a $\Delta\Delta G$ = -32.3 kcal/mol. That would correlate with an increased binding affinity of > 10^6 fold greater than **8**. An assessment of the docked pose of **49** suggests that this increased affinity is driven by more favorable π-π stacking with the Y181 side chain. However, the compound could potentially retain potency against the common Y181C RT mutant via hydrogen bonds formed between that side and the sulfonamide linker (**Figure 5.2**).

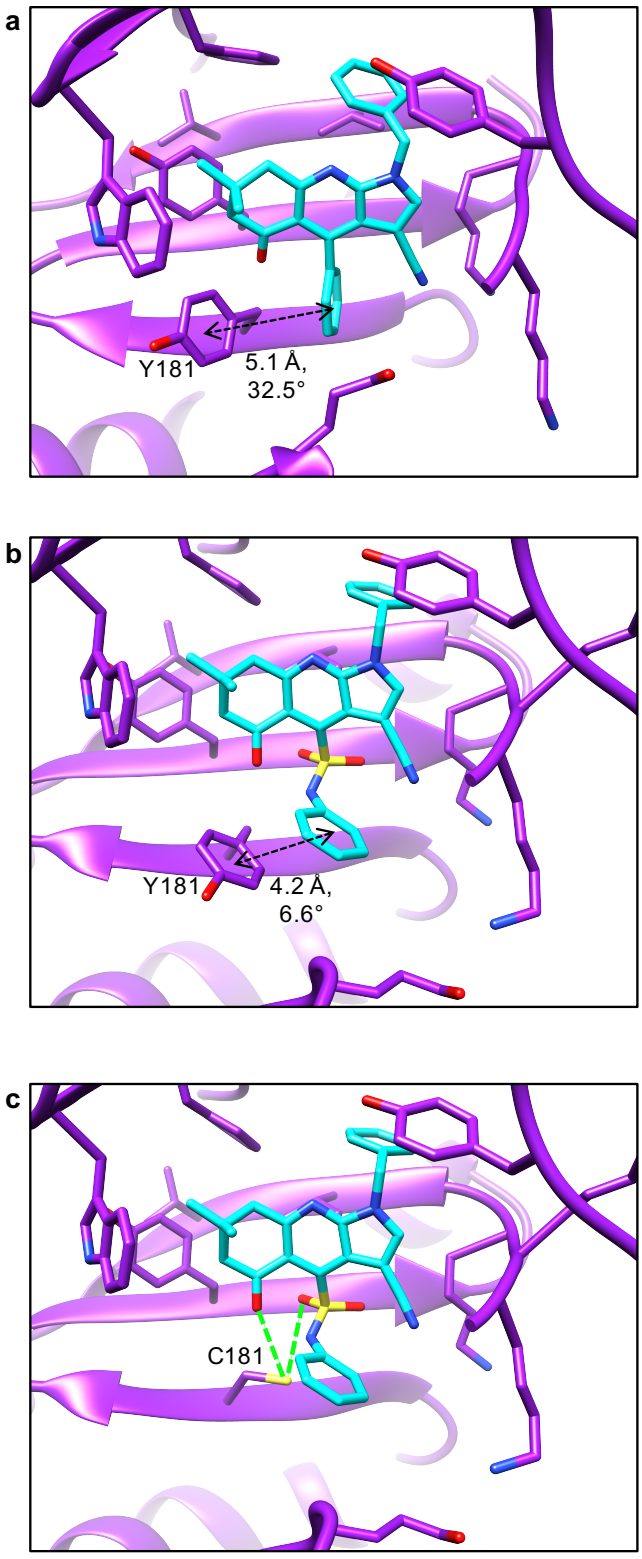


Figure 5.2. Compound **8** (a) and **49** (b) docked into wild type RT-NNRTI binding pocket. **49** shows improvements in the strength of the π-π bond after MD simulation. **49** shows multiple hydrogen bonds to side chain of common Y181C mutant.

5.3 Activity Against Resistant Mutants

A compound more potent against wild type RT and just Y181C, though, would not be of too much use clinically if it was inactive against other resistant viral strains mutants. The second generation of NNRTI used clinically (etravirine (ETV) and rilpivirine (RPV)) were designed to maintain activity against the prominent Y181C and K103N mutants selected by NPV and EFV. Unfortunately, they can select for different mutants, such as E138K (as discussed in section in **Chapter1, 1.6 NNRTI Resistance**).

Any new NNRTI would ideally remain active against all of the clinically relevant mutants while not selecting for any others, either. However, given that more than 50 individual NNRTI-resistance mutations have been identified, and they can occur both alone or in combinations, there are potentially hundreds of possible mutants. It is unlikely that a single drug would remain equipotent against all of these, but there are strategies that could limit the effects of mutants, such as targeting interactions with the protein backbone and immutable side chain residues, like W229.

The docked pose of **8** includes hydrophobic interactions with W229, but lacks any hydrogen bonds with the peptide backbone. The basic structure of the 7-azaindole core could potentially bind so as to form two hydrogen bonds with the backbone of residue 101 as occurs with EFV (**Figure 5.3**). Such a pose cannot occur with compound **8** or related analogs because the benzyl group at position R1 off of the pyrrole nitrogen blocks the potential hydrogen bond donor. There are compounds in the initial library that lacked substitutions at that position, but none

showed potency < 10 μM in the antiviral screen (data not shown). This suggests that such a binding orientation may not be an effective means of inhibition by the 7-azaindole moiety, though the lack of potency from those compounds could be due to other substitutions preventing proper alignment of the compounds for hydrogen bond formation.

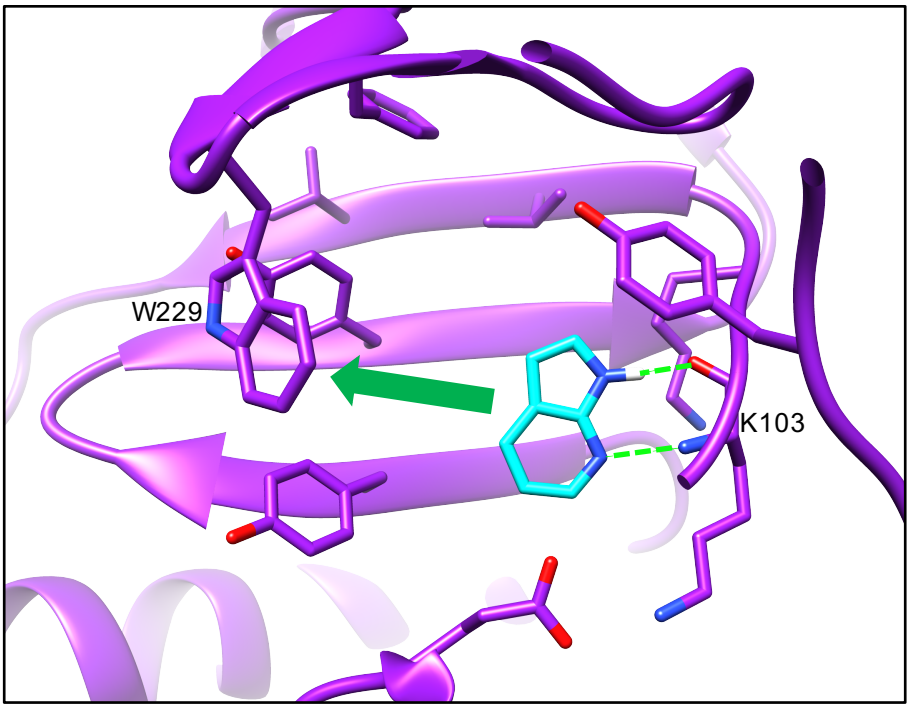


Figure 5.3. 7-azaindole core without substitution at the R_1 position would potentially allow two hydrogen bonds to be formed with protein backbone of residue K103. Green arrow points to potential hydrophobic substitutions could be formed with W229.

Another strategy for maintaining potency against mutants is to build in molecular flexibility to the inhibitor, allowing it to adjust to changes that occur with mutations within the binding pocket. This could be a feasible approach for analogs of **8**, and the prospective compound **49** discussed above does have an extended linker on the lower phenyl ring, which could allow for increased movement to accommodate binding to mutant enzymes.

Unfortunately, though, the FEP approach described in **Chapter 4** may not be very useful for predicting activity against mutants. Calculations were run for compounds **8**, **9**, and **10** to predict changes in potency against the four RT mutants used for experimental testing (K103N, V108I, E138K, and Y181C, as discussed in **Chapter 2**). While the results were able to reproduce the rank ordering of the compound potency changes for K103N and Y181C mutants, they did not for the V108I and E138K mutants (**Table 5.1**). Given that NNRTI-related mutants can confer resistance through several mechanisms beyond just decreasing drug binding affinity, it is expected that a method that estimates only binding affinity would not be sufficient for predicting activity against said mutants.

Table 5.1 FEP scores and fold change in potency of 7-azaindole NNRTI against RT mutants

Compound	K103N		V108I		E138K		Y181C	
	$\Delta\Delta G$ (kcal/mol)	Fold Change*	$\Delta\Delta G$ (kcal/mol)	Fold Change	$\Delta\Delta G$ (kcal/mol)	Fold Change	$\Delta\Delta G$ (kcal/mol)	Fold Change
8	5.6 ± 0.76	13	-1.3 ± 0.27	>140	10 ± 0.92	94	4.0 ± 0.57	4.9
9	3.5 ± 0.73	2.4	2.3 ± 0.29	>16	11 ± 1.0	9.4	7.4 ± 0.64	12
10	8.4 ± 0.67	49	-0.05 ± 0.29	>170	24 ± 0.94	>170	7.9 ± 0.55	>170

*Fold change in potency of inhibition vs. wild type enzyme.

Regardless of the strategy used to come up with new compounds, they should be screened using simultaneous cell-based assays against wild type HIV as well as strains with with K103N-, Y181C-, and E138K-RT mutants, since those represent the major mutations selected by the current clinical NNRTI. Using cell-based assays as the initial screen means that all possible mechanisms of inhibition by NNRTI will be included in the measure of potency (**Chapter 1, 1.5 NNRTI Mechanisms of Action of NNRTI**), which will also be affected by cell entry, exit, and metabolic profiles. This could mean that a strong enzyme inhibitor would be overlooked because of poor metabolic properties, but such a compound would need to be modified anyways, and the cell-based screens are more physiologically relevant than cell-free anti-RT assays.

5.4 Safety and Tolerability

Unlike other classes of antiretrovirals like nucleoside RT inhibitors and protease inhibitors, NNRTI are not direct analogs of any endogenous enzyme ligands and therefore should be expected to have less side effects. There are toxicity issues associated with all clinically used NNRTI, ranging from serious neurological issues with EFV to mild rashes with NVP and RPV (**Chapter 1, 1.8 Side Effects of NNRTI**). The 7-azaindoles were screened for cytoxicity against three cell lines, and more than half of the compounds displayed CC_{50}s > 100 μM. There does not appear to be any intrinsic toxicity associated with the 7-azaindole motif, either, as several compounds containing the moiety are used clinically. This suggests that NNRTI built off of the scaffold should be well tolerated, although any promising candidates that displayed desired potency against wild type and mutant strains would need to

studied further to determine ADME properties in animal models as part of the preclinical evaluation.

5.5 Penetration of Viral Reservoirs

While current antiretroviral therapy can suppress viral load below detectable levels, it cannot eradicate HIV from the body. While the exact causes remain enigmatic, the persistence of HIV is believed to be due in part to viral reservoirs that have lower drug concentrations than the plasma and are therefore permissive to ongoing viral replication. Tissues such as the lymph nodes can contain > 99% lower drug concentrations, and thus have been identified as a possible viral reservoir (**Chapter 1, 1.9 NNRTI and Viral Reservoirs**). An NNRTI that could both penetrate and persist within lymph nodes would have immediate advantages over existing compounds. And while the specific properties needed for such a compound remain unknown (153), the 7-azaindole core is significantly different from other existing NNRTI and therefore may offer advantages related to distribution. If a sufficiently potent compound with activity against relevant mutants and low toxicity was discovered, its ability to reach and remain in lymph nodes could be assessed with animal models.

5.6 Concluding Statements

Drug discovery and development is a challenging endeavor due to the multiple, interconnected variables that must each be optimized before a compound can even be considered as a clinical candidate. A new NNRTI should be an improvement over existing compounds with regards to potency against wild type and mutant strains of HIV, safety and tolerability, and activity in viral reservoirs.

While they do not yet appear to possess such advantages, the compounds built off of the 7-azaindole scaffold can potently inhibit HIV replication without apparent toxicity in cell-based antiviral screens, and FEP-directed optimization coupled with the ease of synthesis suggests that more promising compounds can be efficiently designed and tested.

In order to achieve those ends a variety of experiments and screening systems can be set up. The structure of **8** bound to RT should be determined via x-ray crystallography to help establish a clear map of protein-ligand interactions, which can help in designing and evaluating new compounds via FEP. A cell-based screen including multiple NNRTI-resistant viral strains should be run in parallel with cytoxicity assays to rapidly identify new compounds with improved potency and safety. Further research into small-molecule distribution in lymph nodes needs can be done to identify properties and motifs that can reach and persist within tissues that act as viral reservoirs, and compounds possessing those properties or groups can be selected or added into existing hits.

Ultimately, more analogs of **8** will need to be designed, synthesized, and tested. Even for such a well characterized class of drugs as NNRTI, there are clear limits to what can be accurately predicted *in silico*. Recent lead optimization efforts for novel NNRTI have required the synthesis hundreds (for IDX-899 (154) and doravirine (155)) or even thousands of analogs (for lersivirine) (156) built off of single hit molecules to iteratively select the most desirable properties for a clinical candidate. Even then, two of those compounds ultimately failed in human trials (doravirine is still in Phase II testing). While such a low rate of success is daunting;

smart, focused, and thorough work to optimize compounds built off of the 7-azaindole core could result in a more effective NNRTI.

References

1. Gottlieb MS, Schroff R, Schanker HM, Weisman JD, Fan PT, et al. 1981. Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency. *The New England journal of medicine* 305:1425-31
2. Organization WH. 2014. World Health Statistics 2014. *Global Health Observatory (GHO) data*
3. Coffin J, Swanstrom R. 2013. HIV pathogenesis: dynamics and genetics of viral populations and infected cells. *Cold Spring Harbor perspectives in medicine* 3:a012526
4. Arts EJ, Hazuda DJ. 2012. HIV-1 antiretroviral drug therapy. *Cold Spring Harbor perspectives in medicine* 2:a007161
5. Apostolova N, Blas-Garcia A, Esplugues JV. 2011. Mitochondrial toxicity in HAART: an overview of in vitro evidence. *Curr Pharm Des* 17:2130-44
6. Barre-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, et al. 1983. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* 220:868-71
7. Wood A, Armour D. 2005. The discovery of the CCR5 receptor antagonist, UK-427,857, a new agent for the treatment of HIV infection and AIDS. *Progress in medicinal chemistry* 43:239-71
8. Matthews T, Salgo M, Greenberg M, Chung J, DeMasi R, Bolognesi D. 2004. Enfuvirtide: the first therapy to inhibit the entry of HIV-1 into host CD4 lymphocytes. *Nat Rev Drug Discov* 3:215-25

9. Hurwitz SJ, Schinazi RF. 2012. Practical Considerations For Developing Nucleoside Reverse Transcriptase Inhibitors. *Drug discovery today. Technologies* 9:e183-e93

10. de Bethune MP. 2010. Non-nucleoside reverse transcriptase inhibitors (NNRTIs), their discovery, development, and use in the treatment of HIV-1 infection: a review of the last 20 years (1989-2009). *Antiviral Res* 85:75-90

11. Summa V, Petrocchi A, Bonelli F, Crescenzi B, Donghi M, et al. 2008. Discovery of raltegravir, a potent, selective orally bioavailable HIV-integrase inhibitor for the treatment of HIV-AIDS infection. *J Med Chem* 51:5843-55

12. De Clercq E. 2009. Anti-HIV drugs: 25 compounds approved within 25 years after the discovery of HIV. *Int J Antimicrob Agents* 33:307-20

13. Fischl MA, Richman DD, Grieco MH, Gottlieb MS, Volberding PA, et al. 1987. The efficacy of azidothymidine (AZT) in the treatment of patients with AIDS and AIDS-related complex. A double-blind, placebo-controlled trial. *The New England journal of medicine* 317:185-91

14. Larder BA, Darby G, Richman DD. 1989. HIV with reduced sensitivity to zidovudine (AZT) isolated during prolonged therapy. *Science* 243:1731-4

15. Collier AC, Coombs RW, Schoenfeld DA, Bassett RL, Timpone J, et al. 1996. Treatment of human immunodeficiency virus infection with saquinavir, zidovudine, and zalcitabine. AIDS Clinical Trials Group. *The New England journal of medicine* 334:1011-7

16. Nakagawa F, Lodwick RK, Smith CJ, Smith R, Cambiano V, et al. 2012. Projected life expectancy of people with HIV according to timing of diagnosis. *AIDS* 26:335-43

17. 2015. Fact sheet: What's new in HIV treatment. ed. WH Organization

18. 2016. Guidelines for the Use of Antiretroviral Agents in HIV-1-Infected Adults and Adolescents. ed. NIH

19. Hill AM, Moecklinghoff C, DeMasi R. 2015. When can HIV clinical trials detect treatment effects on drug resistance? *International journal of STD & AIDS* 26:268-78

20. Coffin JM. 1995. HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. *Science* 267:483-9

21. Guaraldi G, Prakash M, Moecklinghoff C, Stellbrink HJ. 2014. Morbidity in older HIV-infected patients: impact of long-term antiretroviral use. *AIDS reviews* 16:75-89

22. Sarmento-Castro R, Vasconcelos C, Aguas MJ, Marques R, Oliveira J. 2011. Virologic suppression in treatment-experienced patients after virologic rebound or failure of therapy. *Current opinion in HIV and AIDS* 6 Suppl 1:S12-20

23. Lorenzo-Redondo R, Fryer HR, Bedford T, Kim EY, Archer J, et al. 2016. Persistent HIV-1 replication maintains the tissue reservoir during therapy. *Nature* 530:51-6

24. Fletcher CV, Staskus K, Wietgrefe SW, Rothenberger M, Reilly C, et al. 2014. Persistent HIV-1 replication is associated with lower antiretroviral drug concentrations in lymphatic tissues. *Proc Natl Acad Sci U S A* 111:2307-12

25. Zhang YL, Ouyang YB, Liu LG, Chen DX. 2015. Blood-brain barrier and neuro-AIDS. *European review for medical and pharmacological sciences* 19:4927-39

26. Luo X, He JJ. 2015. Cell-cell contact viral transfer contributes to HIV infection and persistence in astrocytes. *Journal of neurovirology* 21:66-80

27. Gavegnano C, Schinazi RF. 2009. Antiretroviral therapy in macrophages: implication for HIV eradication. *Antivir Chem Chemother* 20:63-78

28. Iordanskiy S, Berro R, Altieri M, Kashanchi F, Bukrinsky M. 2006. Intracytoplasmic maturation of the human immunodeficiency virus type 1 reverse transcription complexes determines their capacity to integrate into chromatin. *Retrovirology* 3:4

29. Fassati A, Goff SP. 2001. Characterization of intracellular reverse transcription complexes of human immunodeficiency virus type 1. *J Virol* 75:3626-35

30. Nermut MV, Fassati A. 2003. Structural analyses of purified human immunodeficiency virus type 1 intracellular reverse transcription complexes. *J Virol* 77:8196-206

31. Hu WS, Hughes SH. 2012. HIV-1 reverse transcription. *Cold Spring Harbor perspectives in medicine* 2

32. Preston BD, Poiesz BJ, Loeb LA. 1988. Fidelity of HIV-1 reverse transcriptase. *Science* 242:1168-71

33. Perelson AS, Neumann AU, Markowitz M, Leonard JM, Ho DD. 1996. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science* 271:1582-6

34. Swanstrom R, Wills JW. 1997. Synthesis, Assembly, and Processing of Viral Proteins. In *Retroviruses*, ed. JM Coffin, SH Hughes, HE Varmus. Cold Spring Harbor (NY). Number of.

35. Kohlstaedt LA, Wang J, Friedman JM, Rice PA, Steitz TA. 1992. Crystal structure at 3.5 A resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science* 256:1783-90

36. Popovic M, Sarin PS, Robert-Gurroff M, Kalyanaraman VS, Mann D, et al. 1983. Isolation and transmission of human retrovirus (human t-cell leukemia virus). *Science* 219:856-9

37. Cihlar T, Fordyce M. 2016. Current status and prospects of HIV treatment. *Current opinion in virology* 18:50-6

38. Ehteshami M, Nijhuis M, Bernatchez JA, Ablenas CJ, McCormick S, et al. 2013. Formation of a quaternary complex of HIV-1 reverse transcriptase with a nucleotide-competing inhibitor and its ATP enhancer. *J Biol Chem* 288:17336-46

39. Beilhartz GL, Gotte M. 2010. HIV-1 Ribonuclease H: Structure, Catalytic Mechanism and Inhibitors. *Viruses* 2:900-26

40. Hurwitz SJ, Otto MJ, Schinazi RF. 2005. Comparative pharmacokinetics of Racivir, (+/-)-beta-2',3'-dideoxy-5-fluoro-3'-thiacytidine in rats, rabbits, dogs, monkeys and HIV-infected humans. *Antivir Chem Chemother* 16:117-27

41. Sluis-Cremer N, Wainberg MA, Schinazi RF. 2015. Resistance to reverse transcriptase inhibitors used in the treatment and prevention of HIV-1 infection. *Future microbiology* 10:1773-82

42. Sluis-Cremer N, Tachedjian G. 2008. Mechanisms of inhibition of HIV replication by non-nucleoside reverse transcriptase inhibitors. *Virus research* 134:147-56

43. De Clercq E. 1998. The role of non-nucleoside reverse transcriptase inhibitors (NNRTIs) in the therapy of HIV-1 infection. *Antiviral Res* 38:153-79

44. Das K, Clark AD, Jr., Lewi PJ, Heeres J, De Jonge MR, et al. 2004. Roles of conformational and positional adaptability in structure-based design of TMC125-R165335 (etravirine) and related non-nucleoside reverse transcriptase inhibitors that are highly potent and effective against wild-type and drug-resistant HIV-1 variants. *J Med Chem* 47:2550-60

45. Sluis-Cremer N. 2014. The emerging profile of cross-resistance among the nonnucleoside HIV-1 reverse transcriptase inhibitors. *Viruses* 6:2960-73

46. Spence RA, Kati WM, Anderson KS, Johnson KA. 1995. Mechanism of inhibition of HIV-1 reverse transcriptase by nonnucleoside inhibitors. *Science* 267:988-93

47. Krzeminska A, Paneth P, Moliner V, Swiderek K. 2014. Binding Isotope Effects as a Tool for Distinguishing Hydrophobic and Hydrophilic Binding Sites of HIV-1 RT. *J Phys Chem B*

48. Ren J, Milton J, Weaver KL, Short SA, Stuart DI, Stammers DK. 2000. Structural basis for the resilience of efavirenz (DMP-266) to drug resistance mutations in HIV-1 reverse transcriptase. *Structure* 8:1089-94

49. Lansdon EB, Brendza KM, Hung M, Wang R, Mukund S, et al. 2010. Crystal structures of HIV-1 reverse transcriptase with etravirine (TMC125) and rilpivirine (TMC278): implications for drug design. *J Med Chem* 53:4295-9

50. Esnouf RM, Ren J, Hopkins AL, Ross CK, Jones EY, et al. 1997. Unique features in the structure of the complex between HIV-1 reverse transcriptase and the bis(heteroaryl)piperazine (BHAP) U-90152 explain resistance mutations for this nonnucleoside inhibitor. *Proc Natl Acad Sci U S A* 94:3984-9

51. Esnouf R, Ren J, Ross C, Jones Y, Stammers D, Stuart D. 1995. Mechanism of inhibition of HIV-1 reverse transcriptase by non-nucleoside inhibitors. *Nature structural biology* 2:303-8

52. Hsiou Y, Ding J, Das K, Clark AD, Jr., Hughes SH, Arnold E. 1996. Structure of unliganded HIV-1 reverse transcriptase at 2.7 A resolution: implications of conformational changes for polymerization and inhibition mechanisms. *Structure* 4:853-60

53. Seckler JM, Barkley MD, Wintrode PL. 2011. Allosteric suppression of HIV-1 reverse transcriptase structural dynamics upon inhibitor binding. *Biophysical journal* 100:144-53

54. Seckler JM, Howard KJ, Barkley MD, Wintrode PL. 2009. Solution structural dynamics of HIV-1 reverse transcriptase heterodimer. *Biochemistry* 48:7646-55

55. Das K, Martinez SE, Bauman JD, Arnold E. 2012. HIV-1 reverse transcriptase complex with DNA and nevirapine reveals non-nucleoside inhibition mechanism. *Nature structural & molecular biology* 19:253-9

56. Rittinger K, Divita G, Goody RS. 1995. Human immunodeficiency virus reverse transcriptase substrate-induced conformational changes and the mechanism of inhibition by nonnucleoside inhibitors. *Proc Natl Acad Sci U S A* 92:8046-9

57. Xia Q, Radzio J, Anderson KS, Sluis-Cremer N. 2007. Probing nonnucleoside inhibitor-induced active-site distortion in HIV-1 reverse transcriptase by transient kinetic analyses. *Protein science : a publication of the Protein Society* 16:1728-37

58. Bec G, Meyer B, Gerard MA, Steger J, Fauster K, et al. 2013. Thermodynamics of HIV-1 reverse transcriptase in action elucidates the mechanism of action of non-nucleoside inhibitors. *J Am Chem Soc* 135:9743-52

59. Schauer GD, Huber KD, Leuba SH, Sluis-Cremer N. 2014. Mechanism of allosteric inhibition of HIV-1 reverse transcriptase revealed by single-molecule and ensemble fluorescence. *Nucleic Acids Res* 42:11687-96

60. Hang JQ, Li Y, Yang Y, Cammack N, Mirzadegan T, Klumpp K. 2007. Substrate-dependent inhibition or stimulation of HIV RNase H activity by non-nucleoside reverse transcriptase inhibitors (NNRTIs). *Biochem Biophys Res Commun* 352:341-50

61. Hang JQ, Li Y, Yang Y, Zommer H, Mackenzie R, et al. 2006. Compound specific differences in the levels of resistance to HIV-RT catalysed strand transfer,

DNA polymerase and RNase H activities to inhibition by NNRTIs. *Antiviral therapy* 11:S145

62. Figueiredo A, Moore KL, Mak J, Sluis-Cremer N, de Bethune MP, Tachedjian G. 2006. Potent nonnucleoside reverse transcriptase inhibitors target HIV-1 Gag-Pol. *PLoS pathogens* 2:e119
63. Buckheit RW, Jr., Hartman TL, Watson KM, Kwon HS, Lee SH, et al. 2007. The structure-activity relationships of 2,4(1H,3H)-pyrimidinedione derivatives as potent HIV type 1 and type 2 inhibitors. *Antivir Chem Chemother* 18:259-75
64. *HIV Drug Resistance Database*. hivdb.stanford.edu
65. Ren J, Nichols C, Bird L, Chamberlain P, Weaver K, et al. 2001. Structural mechanisms of drug resistance for mutations at codons 181 and 188 in HIV-1 reverse transcriptase and the improved resilience of second generation non-nucleoside inhibitors. *J Mol Biol* 312:795-805
66. Ren J, Nichols CE, Chamberlain PP, Weaver KL, Short SA, Stammers DK. 2004. Crystal structures of HIV-1 reverse transcriptases mutated at codons 100, 106 and 108 and mechanisms of resistance to non-nucleoside inhibitors. *J Mol Biol* 336:569-78
67. Richman D, Shih CK, Lowy I, Rose J, Prodanovich P, et al. 1991. Human immunodeficiency virus type 1 mutants resistant to nonnucleoside inhibitors of reverse transcriptase arise in tissue culture. *Proc Natl Acad Sci U S A* 88:11241-5
68. Debyser Z, De Vreese K, Knops-Gerrits PP, Baekelandt V, Bhikhabhai R, et al. 1993. Kinetics of different human immunodeficiency virus type 1 reverse

transcriptases resistant to human immunodeficiency virus type 1-specific reverse transcriptase inhibitors. *Mol Pharmacol* 43:521-6

69. Spence RA, Anderson KS, Johnson KA. 1996. HIV-1 reverse transcriptase resistance to nonnucleoside inhibitors. *Biochemistry* 35:1054-63

70. Ren J, Stammers DK. 2008. Structural basis for drug resistance mechanisms for non-nucleoside inhibitors of HIV reverse transcriptase. *Virus research* 134:157-70

71. Hsiou Y, Ding J, Das K, Clark AD, Jr., Boyer PL, et al. 2001. The Lys103Asn mutation of HIV-1 RT: a novel mechanism of drug resistance. *J Mol Biol* 309:437-45

72. Singh K, Marchand B, Rai DK, Sharma B, Michailidis E, et al. 2012. Biochemical mechanism of HIV-1 resistance to rilpivirine. *J Biol Chem* 287:38110-23

73. Hu Z, Kuritzkes DR. 2014. Altered viral fitness and drug susceptibility in HIV-1 carrying mutations that confer resistance to nonnucleoside reverse transcriptase and integrase strand transfer inhibitors. *J Virol* 88:9268-76

74. Antinori A, Zaccarelli M, Cingolani A, Forbici F, Rizzo MG, et al. 2002. Cross-resistance among nonnucleoside reverse transcriptase inhibitors limits recycling efavirenz after nevirapine failure. *AIDS research and human retroviruses* 18:835-8

75. Tebit DM, Lobritz M, Lalonde M, Immonen T, Singh K, et al. 2010. Divergent evolution in reverse transcriptase (RT) of HIV-1 group O and M lineages:

impact on structure, fitness, and sensitivity to nonnucleoside RT inhibitors. *J Virol* 84:9817-30

76. Xu HT, Oliveira M, Asahchop EL, McCallum M, Quashie PK, et al. 2012. Molecular mechanism of antagonism between the Y181C and E138K mutations in HIV-1 reverse transcriptase. *J Virol* 86:12983-90

77. Ren J, Nichols CE, Stamp A, Chamberlain PP, Ferris R, et al. 2006. Structural insights into mechanisms of non-nucleoside drug resistance for HIV-1 reverse transcriptases mutated at codons 101 or 138. *The FEBS journal* 273:3850-60

78. Abers MS, Shandera WX, Kass JS. 2014. Neurological and psychiatric adverse effects of antiretroviral drugs. *CNS drugs* 28:131-45

79. Ford N, Calmy A, Mofenson L. 2011. Safety of efavirenz in the first trimester of pregnancy: an updated systematic review and meta-analysis. *AIDS* 25:2301-4

80. 2011. FDA puts a hold on development of ViiV HIV drug. In *AIDSMEDS*

81. Morlese JF, Qazi NA, Gazzard BG, Nelson MR. 2002. Nevirapine-induced neuropsychiatric complications, a class effect of non-nucleoside reverse transcriptase inhibitors? *AIDS* 16:1840-1

82. Medrano J, Barreiro P, Tuma P, Vispo E, Labarga P, et al. 2008. Risk for immune-mediated liver reactions by nevirapine revisited. *AIDS reviews* 10:110-5

83. Usach I, Melis V, Peris JE. 2013. Non-nucleoside reverse transcriptase inhibitors: a review on pharmacokinetics, pharmacodynamics, safety and tolerability. *Journal of the International AIDS Society* 16:1-14

84. Nelson MR, Elion RA, Cohen CJ, Mills A, Hodder SL, et al. 2013. Rilpivirine versus efavirenz in HIV-1-infected subjects receiving emtricitabine/tenofovir DF: pooled 96-week data from ECHO and THRIVE Studies. *HIV clinical trials* 14:81-91

85. Lamson MJ, Sabo JP, MacGregor TR, Pav JW, Rowland L, et al. 1999. Single dose pharmacokinetics and bioavailability of nevirapine in healthy volunteers. *Biopharmaceutics & drug disposition* 20:285-91

86. Montaner JS, Reiss P, Cooper D, Vella S, Harris M, et al. 1998. A randomized, double-blind trial comparing combinations of nevirapine, didanosine, and zidovudine for HIV-infected patients: the INCAS Trial. Italy, The Netherlands, Canada and Australia Study. *Jama* 279:930-7

87. Popowycz F, Routier S, Joseph B, Merour JY. 2007. Synthesis and reactivity of 7-azaindole (1H-pyrrolo[2,3-b]pyridine). *Tetrahedron* 63:1031-64

88. Clark MP, Ledeboer MW, Davies I, Byrn RA, Jones SM, et al. 2014. Discovery of a novel, first-in-class, orally bioavailable azaindole inhibitor (VX-787) of influenza PB2. *J Med Chem* 57:6668-78

89. Farmer L, Ledeboer MW, Hoock T, Arnost MJ, Bethiel RS, et al. 2015. Discovery of VX-509 (Decernotinib): A Potent and Selective Janus kinase (JAK) 3 Inhibitor for the Treatment of Autoimmune Diseases. *J Med Chem*

90. Tsai J, Lee JT, Wang W, Zhang J, Cho H, et al. 2008. Discovery of a selective inhibitor of oncogenic B-Raf kinase with potent antimelanoma activity. *Proc Natl Acad Sci U S A* 105:3041-6

91. Schinazi RF, Sommadossi JP, Saalmann V, Cannon DL, Xie MY, et al. 1990. Activities of 3'-azido-3'-deoxythymidine nucleotide dimers in primary lymphocytes infected with human immunodeficiency virus type 1. *Antimicrob Agents Chemother* 34:1061-7

92. Schinazi RF, Lloyd RM, Jr., Nguyen MH, Cannon DL, McMillan A, et al. 1993. Characterization of human immunodeficiency viruses resistant to oxathiolane-cytosine nucleosides. *Antimicrob Agents Chemother* 37:875-81

93. Cauchon E, Falgueyret JP, Auger A, Melnyk RA. 2011. A high-throughput continuous assay for screening and characterization of inhibitors of HIV reverse-transcriptase DNA polymerase activity. *J Biomol Screen* 16:518-24

94. Ranke J. 2012. drfit: Dose-response data evaluation. R package version 0.05-96. http://cran.r-project.org/package=drfit.

95. 2013. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.r-project.org/.

96. Cai ZW, Wei D, Schroeder GM, Cornelius LA, Kim K, et al. 2008. Discovery of orally active pyrrolopyridine- and aminopyridine-based Met kinase inhibitors. *Bioorg Med Chem Lett* 18:3224-9

97. Kim KS, Zhang L, Schmidt R, Cai ZW, Wei D, et al. 2008. Discovery of pyrrolopyridine-pyridone based inhibitors of Met kinase: synthesis, X-ray crystallographic analysis, and biological activities. *J Med Chem* 51:5330-41

98. Lenzi GM, Domaoal RA, Kim DH, Schinazi RF, Kim B. 2015. Mechanistic and Kinetic Differences between Reverse Transcriptases of Vpx Coding and Non-coding Lentiviruses. *J Biol Chem* 290:30078-86

99. Fletcher RS, Syed K, Mithani S, Dmitrienko GI, Parniak MA. 1995. Carboxanilide derivative non-nucleoside inhibitors of HIV-1 reverse transcriptase interact with different mechanistic forms of the enzyme. *Biochemistry* 34:4346-53

100. Maga G, Ubiali D, Salvetti R, Pregnolato M, Spadari S. 2000. Selective interaction of the human immunodeficiency virus type 1 reverse transcriptase nonnucleoside inhibitor efavirenz and its thio-substituted analog with different enzyme-substrate complexes. *Antimicrob Agents Chemother* 44:1186-94

101. Das K, Arnold E. 2013. HIV-1 reverse transcriptase and antiviral drug resistance. Part 2. *Current opinion in virology* 3:119-28

102. Schinazi RF, Massud I, Rapp KL, Cristiano M, Detorio MA, et al. 2011. Selection and characterization of HIV-1 with a novel S68 deletion in reverse transcriptase. *Antimicrob Agents Chemother* 55:2054-60

103. Trott O, Olson AJ. 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry* 31:455-61

104. Stanton RA, Nettles JH, Schinazi RF. 2015. Ligand similarity guided receptor selection enhances docking accuracy and recall for non-nucleoside HIV reverse transcriptase inhibitors. *Journal of molecular modeling* 21:282

105. Gavegnano C, Detorio M, Montero C, Bosque A, Planelles V, Schinazi RF. 2014. Ruxolitinib and tofacitinib are potent and selective inhibitors of HIV-1 replication and virus reactivation in vitro. *Antimicrob Agents Chemother* 58:1977-86

106. Schinazi RF, Cannon DL, Arnold BH, Martino-Saltzman D. 1988. Combinations of isoprinosine and 3'-azido-3'-deoxythymidine in lymphocytes infected with human immunodeficiency virus type 1. *Antimicrob Agents Chemother* 32:1784-7

107. Nguyen LA, Domaoal RA, Kennedy EM, Kim DH, Schinazi RF, Kim B. 2015. Pre-steady state kinetic analysis of HIV-1 reverse transcriptase for non-canonical ribonucleoside triphosphate incorporation and DNA synthesis from ribonucleoside-containing DNA template. *Antiviral Res* 115:75-82

108. Kitchen DB, Decornez H, Furr JR, Bajorath J. 2004. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov* 3:935-49

109. Schneider G. 2010. Virtual screening: an endless staircase? *Nat Rev Drug Discov* 9:273-6

110. Warren GL, Andrews CW, Capelli AM, Clarke B, LaLonde J, et al. 2006. A critical assessment of docking programs and scoring functions. *J Med Chem* 49:5912-31

111. Huang SY, Zou X. 2010. Advances and challenges in protein-ligand docking. *International journal of molecular sciences* 11:3016-34

112. Cozzini P, Kellogg GE, Spyrakis F, Abraham DJ, Costantino G, et al. 2008. Target flexibility: an emerging consideration in drug discovery and design. *J Med Chem* 51:6237-55

113. Changeux JP. 2011. 50th anniversary of the word "allosteric". *Protein science : a publication of the Protein Society* 20:1119-24

114. Smerdon SJ, Jager J, Wang J, Kohlstaedt LA, Chirino AJ, et al. 1994. Structure of the binding site for nonnucleoside inhibitors of the reverse transcriptase of human immunodeficiency virus type 1. *Proc Natl Acad Sci U S A* 91:3911-5

115. Huang Z, Mou L, Shen Q, Lu S, Li C, et al. 2014. ASD v2.0: updated content and novel features focusing on allosteric regulation. *Nucleic Acids Res* 42:D510-6

116. NIAID. 1989. *ChemDB HIV, Opportunistic Infection and Tuberculosis Therapeutics Database (*http://chemdb.niaid.nih.gov/*)*.

117. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, et al. 2000. The Protein Data Bank. *Nucleic Acids Res* 28:235-42

118. Das K, Bauman JD, Rim AS, Dharia C, Clark AD, Jr., et al. 2011. Crystal structure of tert-butyldimethylsilyl-spiroaminooxathioledioxide-thymine (TSAO-T) in complex with HIV-1 reverse transcriptase (RT) redefines the elastic limits of the non-nucleoside inhibitor-binding pocket. *J Med Chem* 54:2727-37

119. Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, et al. 2006. Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J Med Chem* 49:6177-96

120. Ragno R, Frasca S, Manetti F, Brizzi A, Massa S. 2005. HIV-reverse transcriptase inhibition: inclusion of ligand-induced fit by cross-docking studies. *J Med Chem* 48:200-12

121. Tuccinardi T, Botta M, Giordano A, Martinelli A. 2010. Protein kinases: docking and homology modeling reliability. *J Chem Inf Model* 50:1432-41

122. Rockey WM, Elcock AH. 2006. Structure selection for protein kinase docking and virtual screening: homology models or crystal structures? *Current protein & peptide science* 7:437-57

123. Verdonk ML, Mortenson PN, Hall RJ, Hartshorn MJ, Murray CW. 2008. Protein-ligand docking against non-native protein conformers. *J Chem Inf Model* 48:2214-25

124. Bernstein FC, Koetzle TF, Williams GJ, Meyer EF, Jr., Brice MD, et al. 1977. The Protein Data Bank: a computer-based archival file for macromolecular structures. *J Mol Biol* 112:535-42

125. Nichols SE, Domaoal RA, Thakur VV, Tirado-Rives J, Anderson KS, Jorgensen WL. 2009. Discovery of wild-type and Y181C mutant non-nucleoside HIV-1 reverse transcriptase inhibitors using virtual screening with multiple protein structures. *J Chem Inf Model* 49:1272-9

126. Paris KA, Haq O, Felts AK, Das K, Arnold E, Levy RM. 2009. Conformational landscape of the human immunodeficiency virus type 1 reverse transcriptase

non-nucleoside inhibitor binding pocket: lessons for inhibitor design from a cluster analysis of many crystal structures. *J Med Chem* 52:6413-20

127. Leach AR. 1994. Ligand docking to proteins with discrete side-chain flexibility. *J Mol Biol* 235:345-56

128. Sherman W, Day T, Jacobson MP, Friesner RA, Farid R. 2006. Novel procedure for modeling ligand/receptor induced fit effects. *J Med Chem* 49:534-53

129. 2012. Pipeline Pilot. San Diego, CA: Accelrys Inc.

130. MOE (2011.10). Montreal, QC, Canada: Chemical Computing Group Inc.

131. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. 2011. Open Babel: An open chemical toolbox. *Journal of cheminformatics* 3:33

132. Berenger F, Voet A, Lee XY, Zhang KY. 2014. A rotation-translation invariant molecular descriptor of partial charges and its use in ligand-based virtual screening. *Journal of cheminformatics* 6:23

133. Mysinger MM, Carchia M, Irwin JJ, Shoichet BK. 2012. Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *J Med Chem* 55:6582-94

134. Huang H, Chopra R, Verdine GL, Harrison SC. 1998. Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science* 282:1669-75

135. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, et al. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of computational chemistry* 25:1605-12

136. 2012. Discovery Studio. San Diego, CA: Accelrys Inc.

137. Sanner MF. 1999. Python: a programming language for software integration and development. *J Mol Graph Model* 17:57-61

138. Humphrey W, Dalke A, Schulten K. 1996. VMD: visual molecular dynamics. *Journal of molecular graphics* 14:33-8, 27-8

139. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, et al. 2005. Scalable molecular dynamics with NAMD. *Journal of computational chemistry* 26:1781-802

140. Labute P. 2010. LowModeMD--implicit low-mode velocity filtering applied to conformational search of macrocycles and protein loops. *J Chem Inf Model* 50:792-800

141. Guillemont J, Pasquier E, Palandjian P, Vernier D, Gaurrand S, et al. 2005. Synthesis of novel diarylpyrimidine analogues and their antiviral activity against human immunodeficiency virus type 1. *J Med Chem* 48:2072-9

142. Frey KM, Puleo DE, Spasov KA, Bollini M, Jorgensen WL, Anderson KS. 2015. Structure-based evaluation of non-nucleoside inhibitors with improved potency and solubility that target HIV reverse transcriptase variants. *J Med Chem* 58:2737-45

143. Singh N, Warshel A. 2010. Absolute binding free energy calculations: on the accuracy of computational scoring of protein-ligand interactions. *Proteins* 78:1705-23

144. De Vivo M, Masetti M, Bottegoni G, Cavalli A. 2016. The Role of Molecular Dynamics and Related Methods in Drug Discovery. *J Med Chem*

145. Bollini M, Domaoal RA, Thakur VV, Gallardo-Macias R, Spasov KA, et al. 2011. Computationally-guided optimization of a docking hit to yield catechol diethers as potent anti-HIV agents. *J Med Chem* 54:8582-91

146. Zeevaart JG, Wang L, Thakur VV, Leung CS, Tirado-Rives J, et al. 2008. Optimization of azoles as anti-human immunodeficiency virus agents guided by free-energy calculations. *J Am Chem Soc* 130:9492-9

147. Wang L, Wu Y, Deng Y, Kim B, Pierce L, et al. 2015. Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *J Am Chem Soc* 137:2695-703

148. Vilches-Herrera M, Knepper I, de Souza N, Villinger A, Sosnovskikh VY, Iaroshenko VO. 2012. One-Pot, Three-Component Synthesis of 7-Azaindole Derivatives from N-Substituted 2-Amino-4-cyanopyrroles, Various Aldehydes, and Active Methylene Compounds. *Acs Combinatorial Science* 14:434-41

149. Wagner JP, Schreiner PR. 2015. London dispersion in molecular chemistry--reconsidering steric effects. *Angew Chem Int Ed Engl* 54:12274-96

150. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, et al. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* 102:3586-616

151. *ParamChem*. https://cgenff.paramchem.org/

152. Liu P, Dehez F, Cai W, Chipot C. 2012. A Toolkit for the Analysis of Free-Energy Perturbation Calculations. *Journal of chemical theory and computation* 8:2606-16

153. Edagwa BJ, Zhou T, McMillan JM, Liu XM, Gendelman HE. 2014. Development of HIV reservoir targeted long acting nanoformulated antiretroviral therapies. *Curr Med Chem* 21:4186-98

154. Dousson C, Alexandre FR, Amador A, Bonaric S, Bot S, et al. 2016. Discovery of the Aryl-phospho-indole IDX899, a Highly Potent Anti-HIV Non-nucleoside Reverse Transcriptase Inhibitor. *J Med Chem* 59:1891-8

155. Cote B, Burch JD, Asante-Appiah E, Bayly C, Bedard L, et al. 2014. Discovery of MK-1439, an orally bioavailable non-nucleoside reverse transcriptase inhibitor potent against a wide range of resistant mutant HIV viruses. *Bioorg Med Chem Lett* 24:917-22

156. Mowbray CE, Burt C, Corbau R, Gayton S, Hawes M, et al. 2009. Pyrazole NNRTIs 4: selection of UK-453,061 (lersivirine) as a development candidate. *Bioorg Med Chem Lett* 19:5857-60