IMPACT OF NONNUCLEOSIDE REVERSE TRANSCRIPTASE (RT) INHIBITORS ON THE ENZYME-SUBSTRATE INTERACTIONS AND INTRAMOLECULAR DYNAMICS OF WILD-TYPE AND DRUG RESISTANT HIV-1 RT

by

Grant David Schauer

B.A, Biochemistry, University of New Mexico, 2003

M.S., Biomedical Sciences, University of New Mexico, 2006

Submitted to the Graduate Faculty of

the School of Medicine in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2013

UNIVERSITY OF PITTSBURGH

School of Medicine

This dissertation was presented

by

Grant D. Schauer

It was defended on

July 18, 2013

and approved by

Nicolas Sluis-Cremer, Associate Professor, Department of Medicine

Chakra Chennubhotla, Assistant Professor, Department of Computational and Systems Biology

Saleem Khan, Professor, Department of Department of Microbiology and Molecular Genetics

Maumita Mandal, Assistant Professor, Department of Chemistry, Carnegie Mellon University

Dissertation Advisor: Sanford H. Leuba, Associate Professor, Department of Cell Biology

IMPACT OF NONNUCLEOSIDE REVERSE TRANSCRIPTASE (RT) INHIBITORS ON THE ENZYME-SUBSTRATE INTERACTIONS AND INTRAMOLECULAR DYNAMICS OF WILD-TYPE AND DRUG RESISTANT HIV-1 RT

Grant D. Schauer, M.S.

University of Pittsburgh, 2013

HIV-1 Reverse Transcriptase (RT) is entirely responsible for conversion of viral RNA into dsDNA in the host cell cytoplasm, making it an attractive drug target. Nonnucleoside RT Inhibitors (NNRTIs) are highly effective in the treatment and prevention of HIV, yet their mechanism of action remains unknown. Furthermore, NNRTI resistance mutations arise from therapy, complicating treatment, yet the mechanism(s) of how these mutations inhibit polymerization by RT is unclear. In Chapter 2, we examine the role of NNRTI on the dynamics between RT and its Template/Primer (T/P) substrate using a combination of single-molecule and bulk fluorescence techniques which provide an unprecedented glimpse into the dynamics of RT-T/P interaction as well as the intramolecular conformation of RT itself while bound to its substrate. We show that efavirenz, an NNRTI, causes RT to relinquish its grip on the T/P substrate via “molecular arthritis," accompanied by increased shuttling on the substrate, reducing time spent in a polymerase-competent configuration. The K103N mutation relieves the arthritis in the fingers and thumb sub-domains of RT, enabling the efavirenz-bound enzyme to form a stable polymerase-competent complex. We demonstrate that relief of molecular arthritis is likely caused by disruption of a salt bridge between K101 and E138, residues at a key hinge site in the RT heterodimer. The data suggests a unique mechanism of resistance that is mediated by interplay between intramolecular conformational changes in RT and intermolecular dynamics of the RT-template/primer-dNTP complex. In chapter 3, we discuss ongoing experiments to probe

the intramolecular dynamics of RT on the T/P substrate in response to NNRTIs. Combining time-resolved measurements of interdomain distances within RT as it binds and responds to NNRTIs in real time with atomistic molecular dynamics simulations of RT, we are attempting to characterize on-pathway structural intermediates of NNRTI-induced conformational change.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PROLOGUE

"Strengths

- None

Weaknesses

- Single molecule FRET studies of reverse transcriptase in combination with computational analysis will not have the resolution to improve or develop new inhibitors.

- The approach will not provide a clear understanding of the effects of drug resistant mutations on inhibitor binding affinity.

- While studies of this type have value for our understanding of large scale protein motions, they are not, at least at this point, sufficiently detailed to provide practical information for the development of better drugs"

Reviewer 1, NIH R21 Study section

ACKNOWLEDGEMENTS

Although much of this work was performed in solitude in a dark room, I couldn't have accomplished any of this without the constant support of my friends, family, and colleagues, who were with me throughout my isolation.

Many thanks to all of the members of my committee for their time and effort. Thanks especially to Sanford, who was a wonderful maniac and a delight to work with. I will fondly remember our time together; thank you for all the ripe fruit, espresso, and good conversation, and thank you for providing an environment where such exciting work could flourish. Thank you Nic for ultimately trusting me and leaving your comfort zone to get involved in this high risk, high reward project. Chakra, thank you for providing the occasional respite from experiment with some much needed contagious enthusiasm, and for giving me long-term hope in the midst of multiple failures.

Thank you to all involved in the MBSB program, which was exactly the enriching experience I hoped to obtain. In particular, I'd like to thank Angela Gronenborn and Jen Walker for running such a tight ship. I'm thankful for the support of my colleagues. Specifically, I'd like to thank Matt Fagerburg, who helped me tackle the wizard. Matt, you lovely Renaissance man, thank you for your patience and for imparting at least one part of your vast knowledge onto me. Thanks to Rakesh Mishra, who is like a brother to me; thank you for all the memorable times (and for the exciting way only you could retell them), and thanks also for genuinely caring.

Harshad Ghodke, thanks for a bit of sanity on the cusp of completely losing it, and for being at least as critical as I am about my own work. Thanks to Raymond for helping me; I couldn't have made such significant progress without your insight. Thanks to all the other great friends I met in Pittsburgh who were there for me.

Thanks to all my friends and family from home. I really missed you guys. To my mom, dad, sister, aunt, niece, nephew, and grandparents; I love you all so much and have been so lucky to have you. Your endless support and encouragement is what helped me persevere in the face of major defeats. In return, I can at least offer you work that I am very proud of. I have also always been very proud of each of you. To my wife Emily, who I look up to; thank you for supporting me with your strength and individuality, and for understanding me. I love you and am grateful for you.

1.0 INTRODUCTION

Approximately 34 million people worldwide are infected with human immunodeficiency virus (HIV), the etiological agent of acquired immunodeficiency syndrome (AIDS), causing 1.7 million deaths in 2011 alone (UNAIDS, 2012). There are two subtypes of HIV: HIV-1 and HIV-2, with HIV-1 being the most common and virulent strain. Standing for "major," the M group of HIV-1 is responsible for over 90% of clinical cases of HIV worldwide and can be further subdivided into subtypes A-K, which are classified according to their predominant clinical presentation in various regions around the world (Sharp and Hahn, 2011). Currently the most popular drug target against HIV/AIDS, HIV-1 Reverse Transcriptase (RT) transcribes retroviral (+)RNA into human cDNA, a crucial step in the retroviral life cycle. There are currently thirteen FDA approved RT Inhibitors (RTIs) which are typically administered as highly active antiretroviral therapy (HAART), i.e., in combinations with 2 or more other RTIs (Menéndez-Arias, 2013). Although these regimens are highly effective, new drug-resistant mutations are constantly arising which can greatly reduce the efficacy of RTIs, posing a challenge to researchers attempting to design more efficacious RTIs (Esposito et al., 2012; Le Grice, 2012). It is therefore imperative to understand the mechanism(s) of inhibition of this remarkably versatile enzyme in detail, including an understanding of the molecular mechanism(s) of RTI resistance.

1.1 HIV-1 LIFE CYCLE

HIV-1 hijacks the host cell in order to replicate (Sluis-Cremer and Tachedjian, 2008). Upon fusion with a CD4+ T-lymphocyte, mediated by gp41 and gp120 on the viral envelope and CD4 and either the CXCR5 or CCR5 coreceptors on the cell surface, the viral envelope fuses with the host cell membrane, releasing the viral capsid into the cytoplasm. This capsid packages two copies of the same remarkably simple (+) ssRNA genome (see Figure 1.1).

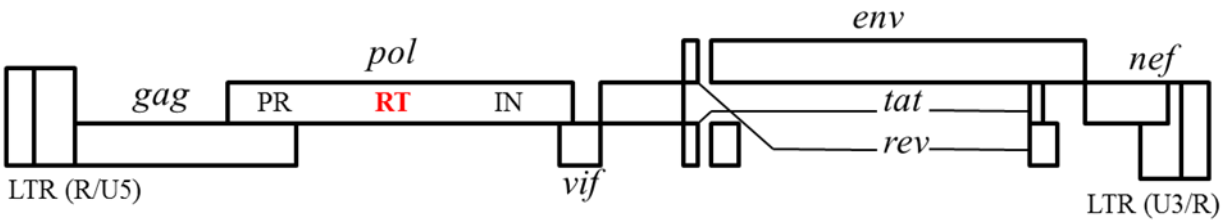

Figure 1 Organization of the HIV-1 genome.

The schematic indicates the relative distribution of genes from 5'-3', including the three major genes, *gag*, *pol*, and *env*, as well as regulatory (*tat* and *rev*) and accessory (*vif*, *vpr*, *vpu*, and *nef*) genes, flanked by the long terminal repeat (LTR) regions designated R, U5, and U3. RT, encoded by the *pol* gene, is highlighted in red, indicated next to the coding sequences for protease (PR) and integrase (IN). Figure adapted from Menendez-Arias, 2013.

The HIV-1 genome consists of several open reading frames (ORFs) flanked by long terminal repeat sequence (LTR), designated U5, R, and U3, a 5' cap (Gppp) and a 3' poly(A)tail (see Figure 1.1). These ORFs encode the three structural genes, *gag*, *pol*, and *env*, as well as regulatory (*tat* and *rev*) and accessory genes (*vif*, *vpr*, *vpu*, and *nef*; see figure 1.1). For the purposes of this review, we will focus on *pol*, the gene which encodes the Pol polyprotein consisting of the enzymes RT, integrase (IN), and protease (PR) (see Lee et al., 2012 for more

information). PR is required to cleave components of the Gag polyprotein, but is also required to cleave RT. IN is required to integrate the dsDNA product of RT into the host cell genome.

The 640 bp LTR flanking the HIV-1 genome contains enhancers and promoters necessary for transcription (Esposito et al., 2012). Particularly important for the life-cycle of HIV-1 RT, the 18 nt PBS, or primer binding site, is a sequence in the U5 region which is complementary to 18 nt of the tRNALys, required to prime reverse transcription initiation (see Figure 1.2).

HIV-1 RT is an RNA- or DNA-dependent DNA polymerase and also contains ribonuclease H (RNase H) activity, containing all the necessary enzymatic activity for conversion of HIV-1 ssRNA into dsDNA, which is then inserted into the human genome with IN (Figure 1.2) (Sarafianos et al., 2009; Sluis-Cremer and Tachedjian, 2008). Reverse transcription (Hu and Hughes, 2012) is initiated at the 3'-end of cellular lysyl-$tRNA^{Lys,3}$, which hybridizes to the primer binding site (PBS) on the HIV-1 genome. RNA-primed RT elongates until the 5'-end of the HIV-1 RNA is reached (Initiation), forming minus-strand strong-stop DNA. Employing the RNase H activity of RT, the remaining HIV-1 genomic RNA is cleaved to allow the nascently synthesized DNA to circularize and hybridize with the repeat sequence (R) at the 3'end of the HIV-1 ssRNA. After this strand transfer, the nascent DNA strand is further elongated by RT. RT hydrolyzes the remaining RNA but leaves behind a purine-rich sequence named the polypurine tract (PPT), which subsequently serves as a primer for the initiation of second strand DNA synthesis. RT then elongates the PPT primer. The RNase H activity of RT removes all remaining RNA, including the tRNA primer and the PPT. Strand transfer takes place by PBS sequence homology. DNA polymerization and strand-displacement followed by further DNA elongation results in an integrase-competent dsDNA, which is flanked by LTR at both ends.

This dsDNA is then a substrate for integrase, which catalyzes the insertion of dsDNA into the human genome (Sluis-Cremer and Tachedjian, 2008).

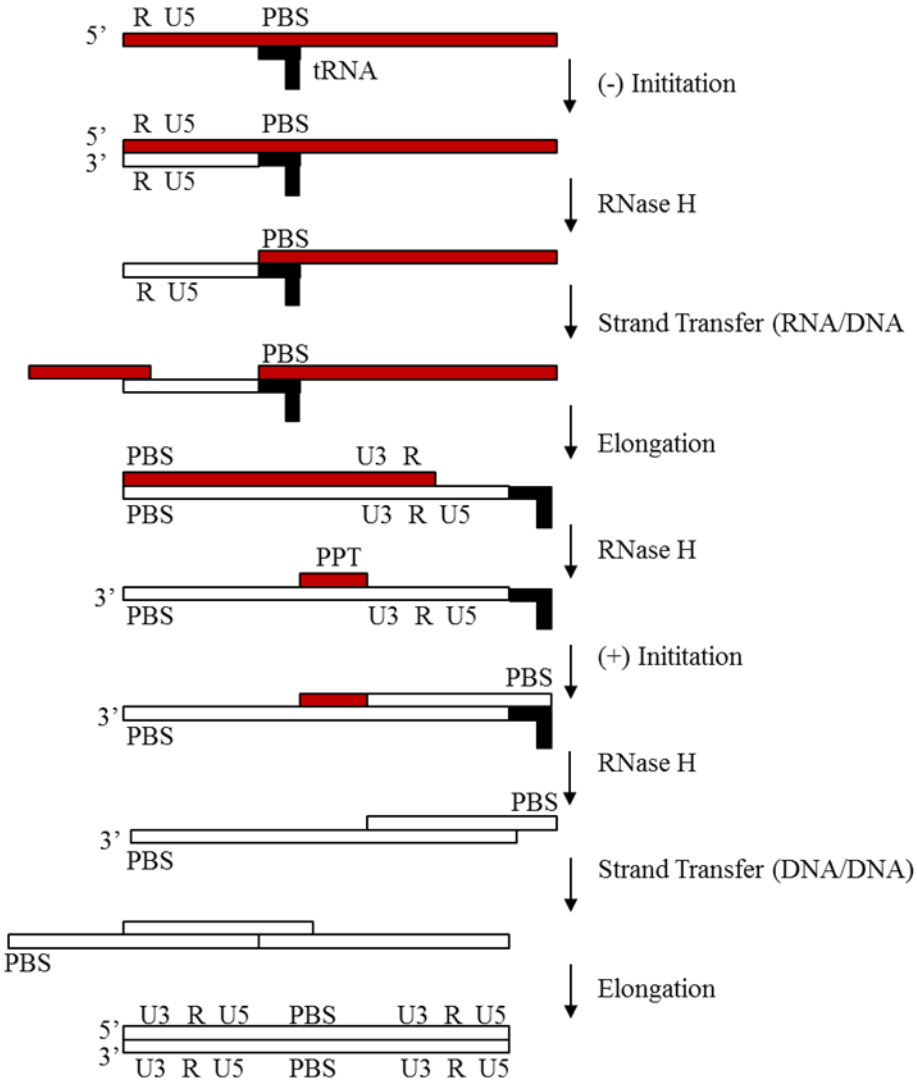

Figure 2. Reverse Transcription.

Schematic of the multistep process of conversion of viral RNA (red) into integrase-competent dsDNA (bottom) for insertion into the human genome. Figure adapted from Sluis-Cremer and Tachedjian, 2008.

1.2 HIV-1 REVERSE TRANSCRIPTASE: STRUCTURE AND FUNCTION

Able to efficiently catalyze DNA polymerization on a variety of substrates (i.e., RNA/DNA, DNA/DNA) and also possessing RNase H activity, RT is an astonishingly versatile enzyme (Sarafianos et al., 2009), perhaps due in large part to its modular structure.

RT is a 110 kD heterodimer composed of two subunits: p66 (560 aa long) and p51 (440 aa long) (Divita et al., 1993a). Both subunits are a product of proteolytic processing of a gag-pol polyprotein by HIV-1 protease (PR). In p51, the RNase-H domain (residues 550-560) has been cleaved, resulting in a structure that shares secondary structural elements with p66. However, since the overall tertiary structure is spatially configured differently than p66, p51 only plays a scaffolding role for p66, the only subunit in the RT heterodimer to possesses polymerase and RNase H activity (Jacobo-Molina et al., 1993; Kohlstaedt et al., 1992).

Like other DNA polymerases, RT is shaped like a right hand (Kohlstaedt et al., 1992). Its subdomains are accordingly named: fingers (residues 1-85 and 118-155), thumb (residues 237-318), palm (residues 86-117), and connection (319-426) subdomains (see Figure 3). In the *apo* structure of RT (e.g., Rodgers et al., 1995) the thumb and fingers domains are nearly in contact, and the thumb resides in the nucleic acid binding cleft (see Figure 7, left for a view of unliganded RT) (Esnouf et al., 1995; Rodgers et al., 1995). To accommodate T/P substrates in the nucleic acid binding cleft, the thumb must extend outward. Pre-steady-state kinetic analysis showed that binding to the T/P and dNTP substrates is a two step process involving initial entry of the T/P into the nucleic acid binding cleft and a slow step dependent on conformational change of the enzyme to accommodate cognate dNTP, which can increase the overall binding affinity to the T/P substrate (KruhØfter et al., 1993). The resultant binary complex contains a T/P substrate simultaneously spanned by a polymerase and an RNase H approximately 18 bp

apart. Remarkably, RT is able to flip on template/primer substrates and can reside in different relative orientations based on the composition of the template/primer alone (Abbondanzieri et al., 2008).

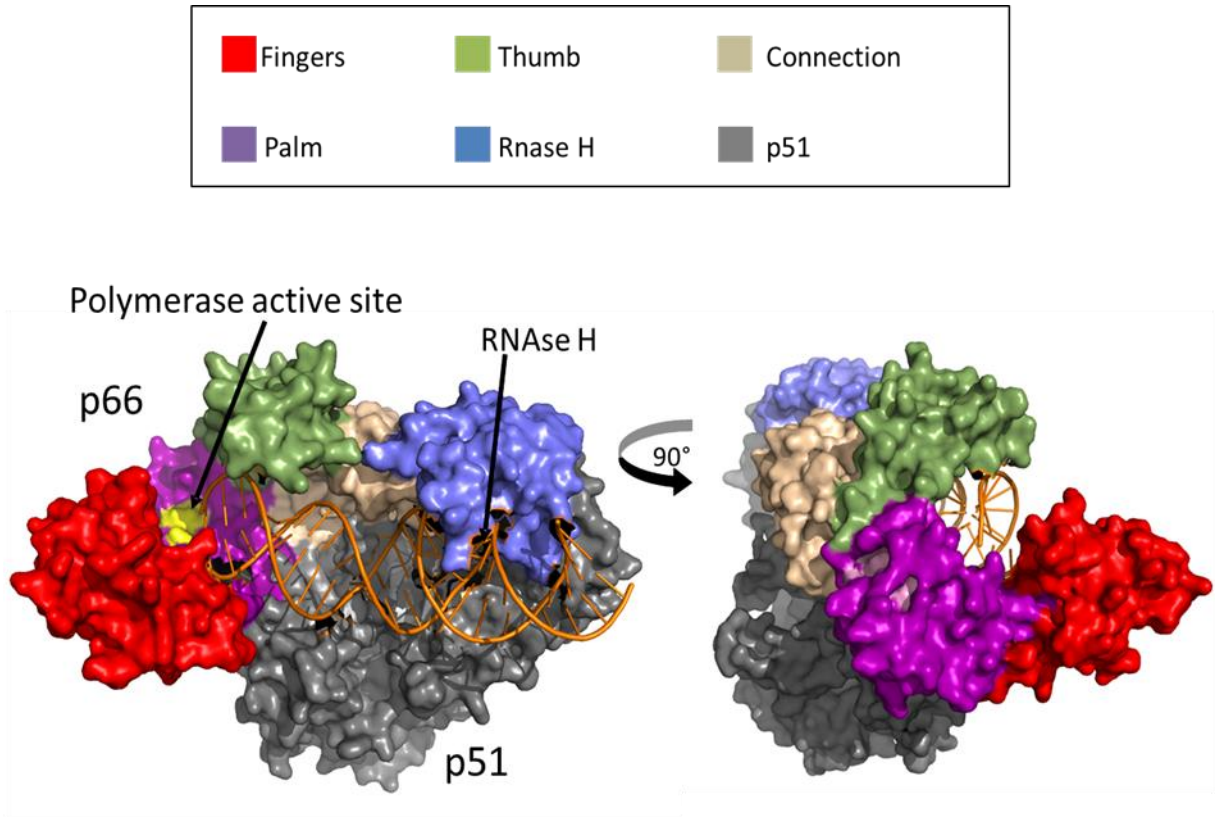

Figure 3. Structure of RT.

RT is pictured as a solvent-accessible surface area model (taken from PDB ID: 1RTD). The DNA/DNA template/primer is shown as cartoon. The p51 subdomain is colored in grey, and the p66 subdomain is subdivided into thumb (green), fingers (red), palm (purple), connection (wheat), and RNase H (blue) domains. The polymerase active site is colored in yellow.

RT clamps down on the incoming nucleotide with its fingers (Huang et al., 1998), specifically, via the Lys65, Arg72, Asp113, and Ala114 residues of the e β3- β4 loop (Le Grice, 2012). Structures of RT complexed with a DNA/DNA template/primer that has been dideoxy

terminated at the 3'-end (to create a dead-end complex in the presence of cognate dNTP) identified the binding site of dNTP: Asp113, Tyr115, Phe116, and Gln151 (Huang et al., 1998). In agreement from pre-steady-state kinetics data and based on data from other polymerases (Kati et al., 1992; Kirmizialtin et al., 2012), it was proposed that fingers clamping of dNTP is a rate-limiting prerequisite to catalysis (Sarafianos et al., 1999). Pre-steady-state kinetic analysis showed that dNTP incorporation rates depend upon the composition of the T/P substrate, and suggested that RT can bind T/P substrates in productive or nonproductive complexes in terms of dNTP incorporation, and that the T/P must isomerize in order for the RT-T/P complex to be converted into a productive one (Wöhrl et al., 1999). It was further shown that the rate-limiting step of fingers bending is on the order of milliseconds and that the conformational change itself governs dNTP specificity (Kirmizialtin et al., 2012).

The bending of the fingers clamp precisely positions dNTP for incorporation, which is coordinated with the positioning of the growing primer by the "primer grip," composed of the β12-β13 hairpin (Jacobo-Molina et al., 1993). Finally, formation of a nascent phosphodiester bond on a growing primer (i.e., DNA polymerization) is coordinated by the catalytic triad in the palm subdomain (D110, D185, D186; part of the conserved YXDD motif), which is a conserved process amongst polymerase and is dependent upon two divalent metals (Steitz, 1998). The process of nucleotide addition ends in release of pyrophosphate, which is accompanied by fingers domain opening. In order for processive replication to continue, the T/P substrate needs to translocate 1bp with respect to RT, a process which is possibly linked potential energy stored in the YMDD loop, akin to a loaded springboard (Götte et al., 2010; Sarafianos et al., 2002).

1.3 REVERSE TRANSCRIPTASE INHBITORS

The US Food and Drug Administration (FDA) has approved 13 RTIs for the treatment of HIV-1 infection, although only 11 of these are currently used (see below). These inhibitors, all of which bind at or near to the DNA polymerase active site of the enzyme, can be classified into two distinct groups: (1) the nucleoside and nucleotide RT inhibitors (NRTIs) and (2) the nonnucleoside RT inhibitors (NNRTIs) (Arts and Hazuda, 2012).

First-line antiretroviral therapy (ART) for the treatment of HIV-1 infection typically includes two NRTIs in combination with an NNRTI or a protease inhibitor. NRTIs and NNRTIs are also routinely used in second-line and salvage ART therapies. HIV-1 resistance to all of the FDA-approved NRTIs and NNRTIs has been documented. An understanding of the mutations associated with RT inhibitor (RTI) resistance, the antagonistic or complementary interactions between RTI resistance mutations, and the mechanisms of HIV-1 resistance to RTIs is of critical importance for the development and formulation of effective ART therapies. Of concern, there has been a significant increase in circulating and transmitted NNRTI drug resistance in resource-limited settings due to the extensive use of NNRTIs in prevention and treatment strategies for HIV-1 infection. Despite this increase in NNRTI drug resistance, the next generation diarylpyrimidine NNRTIs, dapivirine, etravirine and rilpivirine, will be increasingly used in resource-limited settings. As such, there is a continued need to monitor and understand NNRTI resistance (Esposito et al., 2012; Menéndez-Arias, 2013; Sluis-Cremer and Tachedjian, 2008).

1.4 NRTIS

The NRTIs are analogs of naturally occurring dNTPs that lack a 3'-hydroxyl group on the ribose sugar/pseudosugar. They were the first drugs used to treat HIV-1 infection and they remain integral components of nearly all antiretroviral (ART) regimens. To exhibit antiviral activity, NRTIs must be metabolically converted by host-cell kinases to their corresponding triphosphate forms (NRTI-TP).

The NRTI-TP inhibit HIV-1 RT DNA synthesis by acting as chain-terminators of DNA synthesis (Goody et al., 1991). Eight NRTIs have been approved for clinical use, namely 3'-azido-3'-deoxythymidine (zidovudine, AZT), 2',3-dideoxyinosine (didanosine, ddI), 2',3'-dideoxycytidine (zalcitabine, ddC), (-)-β-2',3'-dideoxy-3'-thiacytidine (lamivudine, 3TC), 2'-deoxy-2',3'-didehydrothymidine (stavudine, d4T), (1*S*,4*R*)-4-[2-amino-6-(cyclopropyl-amino)-9H-purin-9-yl]-2-cyclopentene-1-methanol succinate (abacavir, ABC), (*R*)-9-(2-hosphonylmethoxypropyl) adenine (TFV, tenofovir), and 5-fluoro-1-[(*2R*,5*S*)-2-(hydroxymethyl)-1,3-oxathiolan-5-yl]cytosine (emtricitabine, FTC) (Figure 1.3). ddC is less potent than the other NRTIs, has inconvenient dosing schedules, and is associated with serious adverse events. For these reasons it is now rarely used to treat HIV-1 infection. Similarly, the World Health Organization advocated that d4T should be phased out of use because of its long-term, irreversible side-effects. However, d4T is still used in first-line therapy in developing countries due to its low cost and widespread availability.

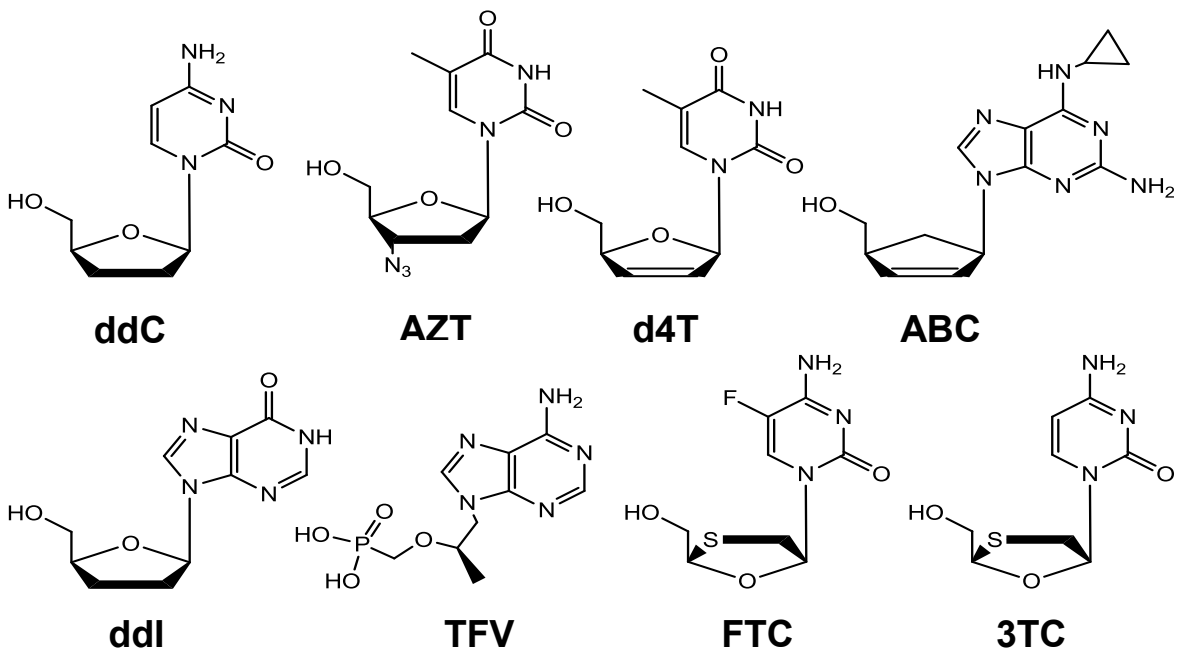

Figure 4. Chemical structures of FDA-approved NRTIs

1.4.1 NRTI Resistance

NRTI-associated resistance mutations can be broadly categorized into two groups depending on their phenotypic mechanism of resistance (Selmi et al., 2003; Sluis-Cremer and Tachedjian, 2008). The mutations K65R, K70E, L74V, Q151M (in complex with A62V, V75I, F77L and F116Y) and M184V increase the selectivity of RT for incorporation of natural dNTP substrate *versus* the NRTI-TP (Deval et al., 2002, 2004a, 2004b, 2004c; Feng and Anderson, 1999; Selmi et al., 2001, 2003; Sluis-Cremer et al., 2007). This resistance mechanism has been termed NRTI-TP discrimination. In comparison, the mutations M41L, D67N, K70R, L210W, T215F/Y and K219Q/E are typically referred to as thymidine analog mutations (TAMs). These mutations augment the ability of HIV-1 RT to excise a chain-terminating NRTI-monophosphate (NRTI-

MP) from a prematurely terminated DNA chain (Arion et al., 1998; Meyer et al., 1999). This resistance mechanism has been termed NRTI-MP excision.

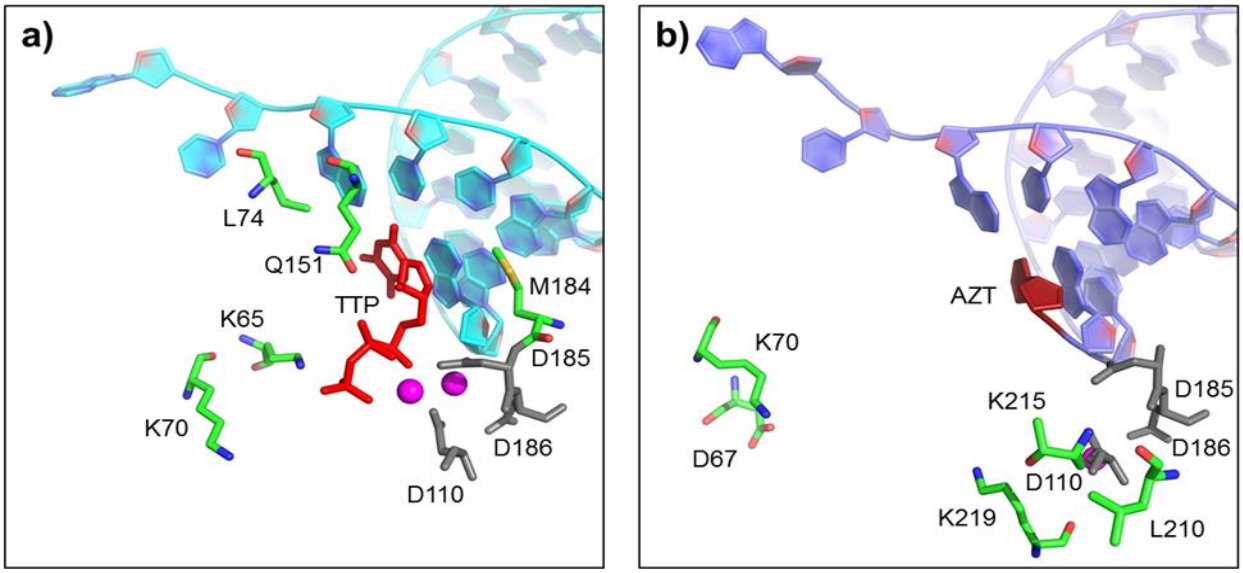

Figure 5. Residues involved in NRTI resistance

Structures of RT taken from a) the TTP-bound ternary complex (PDB ID: 1RTD) or b) an AZT-terminated primer in the P-site (PDB ID: 1N6Q) are shown, highlighting the residues (sticks) associated with NRTI resistance.

1.5 NNRTIS

The NNRTIs are a group of amphiphilic compounds that bind to a hydrophobic pocket in HIV-1 RT that is proximal to but distinct from the polymerase active site called the NNRTI Binding Pocket (NNRTIBP), which is not formed unless an NNRTI is bound (Kohlstaedt et al., 1992; Sluis-Cremer et al., 2004). Because the binding pocket is a full 10 Å away from the polymerase active site and are noncompetitive with dNTP or template/primer binding (Divita et al., 1993b), NNRTI are considered classical allosteric inhibitors of HIV-1 RT DNA polymerization reactions (Seckler et al., 2011; Temiz and Bahar, 2002). The precise mechanism of inhibition by NNRTI

(discussed further in section 1.5.1) remains to be entirely understood (Esposito et al., 2012; Ivetac and McCammon, 2009; Sluis-Cremer and Tachedjian, 2008), but proposed mechanisms have included restriction of thumb flexibility (Kohlstaedt et al., 1992), distortion of the catalytic triad (Ren et al., 1995), primer grip repositioning (Das et al., 1996), repositioning of primer (Das et al., 2012), and loosening of the thumb/fingers clamp on template/primer substrates (Liu et al., 2008).

FDA-approved NNRTIs include 11-cyclopropyl-4-methyl-5,11-dihydro-6*H*-dipyrido[3,2-*b*:2′,3′-*e*][1,4]diazepin-6-one (nevirapine; NVP), *N*-[2-({4-[3-(propan-2-ylamino)pyridin-2-yl]piperazin-1-yl}carbonyl)-1*H*-indol-5-yl]methanesulfonamide (delavirdine; DEL), (4*S*)-6-chloro-4-(2-cyclopropylethynyl)-4-(trifluoromethyl)-2,4-dihydro-1*H*-3,1-benzoxazin-2-one (efavirenz; EFV), 4-[6-Amino-5-bromo-2- [(4-cyanophenyl)amino] pyrimidin-4-yl]oxy-3,5-dimethylbenzonitrile (etravirine; ETV) and 4-{[4-({4-[(*E*)-2-cyanovinyl]-2,6-dimethylphenyl}amino)pyrimidin-2-yl]amino}benzonitrile (rilpivirine; RIL). The efficacy of delavirdine is lower than that of the other NNRTIs, especially EFV, and it also has an inconvenient dosing schedule. These factors have led the U.S. Department of Health and Human Services (DHHS) to recommend against its use as part of initial therapy.

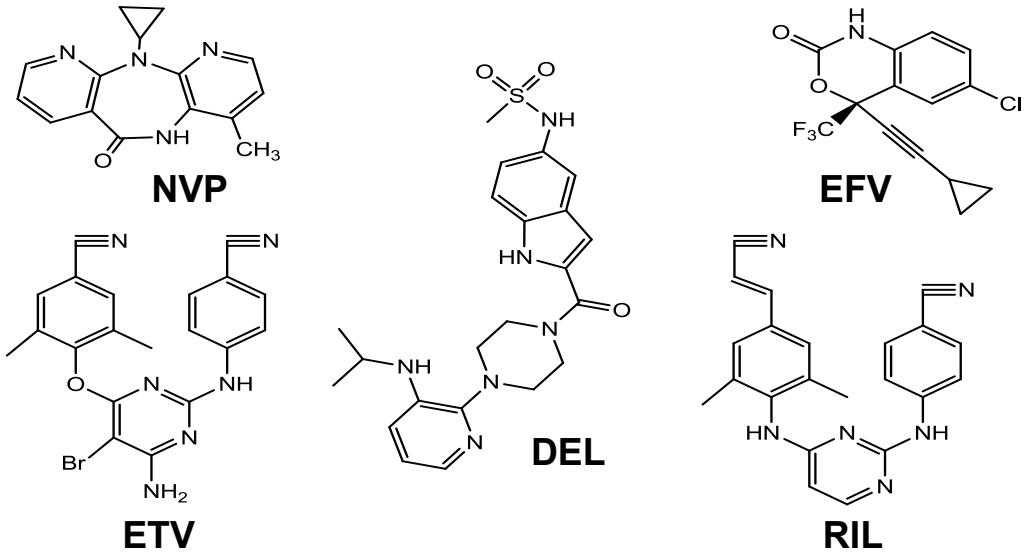

Figure 6. Chemical structures of FDA-approved NNRTIs

Despite their widespread use and remarkable efficacy, the mechanism(s) behind inhibition of reverse transcription by NNRTIs remains unclear.

1.5.1 Mechanism of action of NNRTIS

The mechanism of action of NNRTIs is controversial (Ren and Stammers, 2008; Sarafianos et al., 2009). One possible mechanism is distortion of the catalytic triad (Ren et al., 1995), rendering an inactive polymerase. It has also been postulated that NNRTIs are responsible for repositioning the primer grip, eliminating the coordination of primer to the catalytic triad (Das et al., 1996). Recently, there has been evidence that NNRTIs cause loosening of the thumb/fingers clamp on template/primer substrates (Liu et al., 2008), as well as repositioning of primer (Das et al., 2012).

One persistent observation from the solved crystal structures of RT in the presence of NNRTIs is the hyperextension of the thumb domain and a further opening of the fingers domain accompanied by a marked decrease in flexibility, also known as "molecular arthritis." (see Figure 6), which appears to be closely related to the existence of a hinge region between the p51 and p66 subdomains (Bahar et al., 1999; Ivetac and McCammon, 2009). Restriction of thumb flexibility was considered a key element in the inhibition mechanism (Kohlstaedt et al., 1992). Although it is unclear whether molecular arthritis is functionally tied to inhibition of polymerization, there is evidence that it affects the grip on the template/primer substrate. Liu and colleagues found that, while cognate dNTP stabilizes the polymerase competent configuration, NNRTI destabilizes this configuration, regulating the sliding dynamics of RT on the T/P, indicating that modulation of the grip on the T/P substrate itself by NNRTIs may alter RT-T/P dynamics, disfavoring the polymerase mode (Liu et al., 2008). It is furthermore possible that the function of molecular arthritis is to reposition the catalytic triad in a polymerase incompetent configuration, as was observed in large, multicopy simulations of *apo* and NNRTI-bound RT (Ivetac and McCammon, 2009).

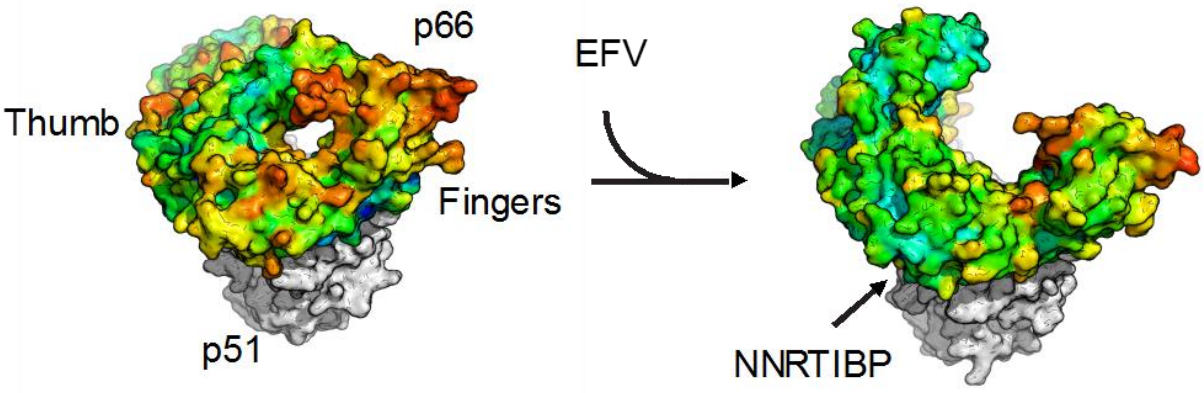

Figure 7. Molecular arthritis.

The structures of *apo* WT RT (left; PDB ID: 1DLO) and WT RT bound to efavirenz (right; PDB ID 1FK9). Structures are represented by solvent-accessible surface area. The p51 subunit is colored in grey and the p66 subunit is colored according to crystallographic B-factors.

Until recently, there only existed structures of RT complexed with dsDNA template/primer (T/P) mimic substrates (mostly DNA/DNA, DNA/PPT) or structures of RT complexed with NNRTIs, however there was no structure of a ternary RT-T/P-NNRTI complex. It has historically been a challenge to find order in the electron density maps of RT-T/P in the presence of NNRTIs. Recently, however, a structure of RT crosslinked to a DNA/RNA T/P substrate and bound to neviripine (NVP) was elucidated (Das et al., 2012), wherein the authors observed a distortion of the primer grip accompanied by thumb hyperextension and a 5.5 Å shift of the 3'-end primer away from the active site, dNTP binding site distortion, and a reduction in RT-T/P contacts. These newly solved structures primarily resemble bunches of grapes and long tubes (Deejay Smumilo, personal communication). Soon after, structures of RT bound to several DNA/RNA T/P substrates in the presence of NVP and EFV were solved without the use of crosslinking agents (Lapkouski et al., 2013). The structure showed that NNRTIs pushed RT toward a degradative/RNase H-competent mode (as opposed to a polymerase competent mode) in the context of a hybrid substrate, which is consistent with the findings of Radzio and Sluis-Cremer, 2008, which indicated that EFV accelerates RNA degradation.

Perhaps counterintuitively, it has been demonstrated that some NNRTIs act as potent chemical enhancers of HIV-1 RT heterodimerization (Tachedjian et al., 2001; Venezia et al.,

2006). To date, efavirenz was found to be the most potent enhancer of RT heterodimerization, whereas nevirapine has a weak effect and delavirdine has no effect at all (Tachedjian et al., 2001). While there doesn't appear to be a correlation between the impact of NNRTI-mediated enhancement of RT heterodimerization and the defects in RT polymerase function (Xia et al., 2007), recent studies have demonstrated effects of some potent NNRTIs, (e.g., efavirenz, dapivirine and etravirine) on the late stages of HIV replication (Tachedjian et al., 2005; Figueiredo et al., 2006).

1.5.2 NNRTI Resistance

Typically, HIV-1 resistance to NNRTIs correlates directly with mutations of one or more RT residues in the NNRTI-binding pocket. Mutations associated with resistance to NVP and EFV include L100I, K101E/P, K103N/S, V106A/M, Y181C/I/V, Y188C/L/H, G190A/E/S and M230L (Stanford University HIV Drug Resistance Database: http://hivdb.stanford.edu/). Although ETV and RIL have been reported to have higher *in vitro* genetic barriers to resistance than EFV or NVP (Andries et al., 2004; Azijn et al., 2010), 17 mutations in HIV-1 RT have been associated with decreased virologic response to ETV (V90I, A98G, L100I, K101E/H/P, V106I, E138A, V179D/F/T, Y181C/I/V, G190A/S and M230L) (Vingerhoets et al., 2010), and 15 mutations with decreased virologic response to RIL (K101E/P, E138A/G/K/Q/R, V179L, Y181C/I/V, H221Y, F227C, and M230I/L) (Anta et al., 2013). In general, these NNRTI resistance mutations can affect inhibitor binding in a number of ways. (1) They can remove one or more favorable interactions between the inhibitor and NNRTI-binding pocket. For example, the Y181C mutation eliminates π-stacking interactions between this residue and the aromatic ring of the NNRTI pharmacophore (Ren et al., 2001). (2) They can introduce steric barriers to

NNRTI binding. For example, the G190E mutation introduces a bulky side-chain which may prevent NNRTI binding by sterically interfering with functional groups, such as the cyclopropyl ring of NVP (Huang et al., 2003; Yap et al., 2007). (3) The mutations may introduce or eliminate inter-residue contacts in the NNRTI-binding pocket, which interfere with the ability of other residues in the pocket to fold down over the NNRTI (Sluis-Cremer et al., 2004).

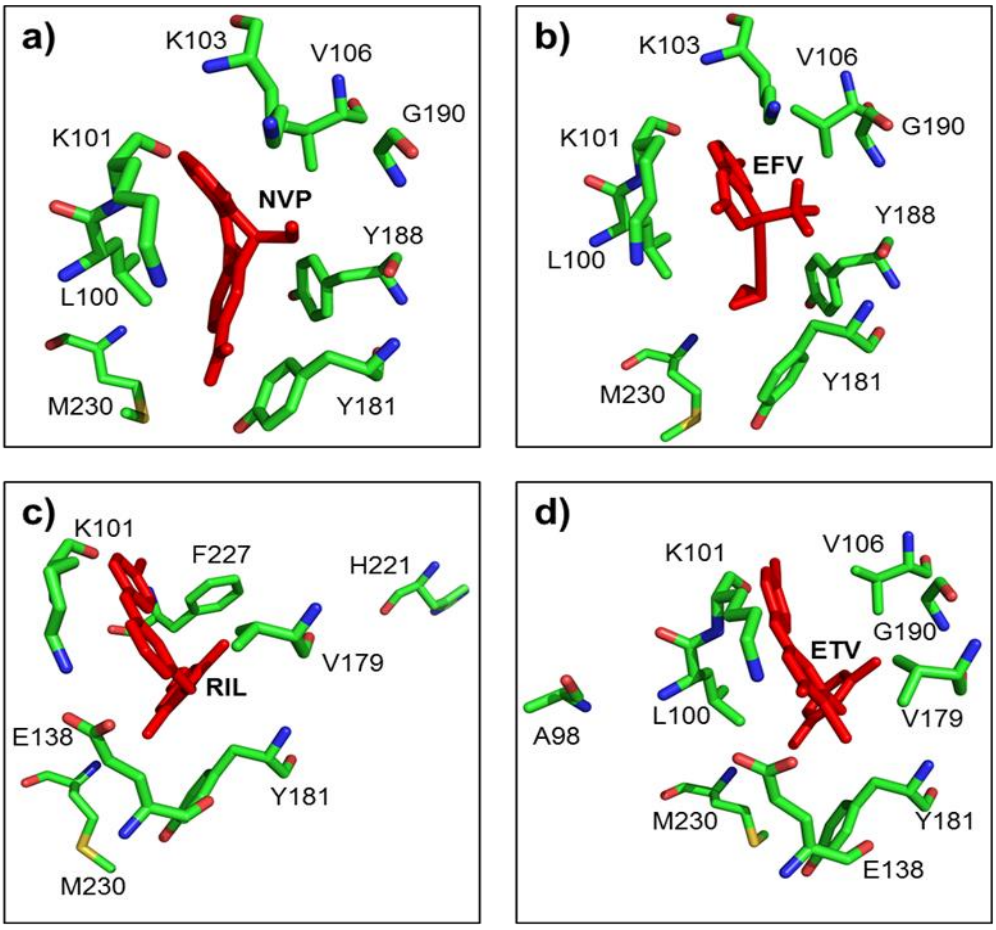

Figure 8. Residues involved in NNRTI resistance

1.5.3 K103N

A major focus of this dissertation, K103N is the most commonly reported NNRTI-resistance mutation (Ren and Stammers, 2008). It confers high level resistance to a wide range of NNRTIs,

notably efavirenz (EFV) and neviripine (NVP) (approximately 20-fold reduction in susceptibility) (Clotet, 1999). Despite its devastating effects on susceptibility to first-gen NNRTIs, the mechanism of resistance by K103N remains unclear. Observations from the first structures (Hsiou et al., 2001; Ren et al., 2001) of K103N RT indicated that neither K103 nor the mutant N103 residue make any contacts with EFV or NVP (although K103 can H-bond with DEL) (Ren and Stammers, 2008).

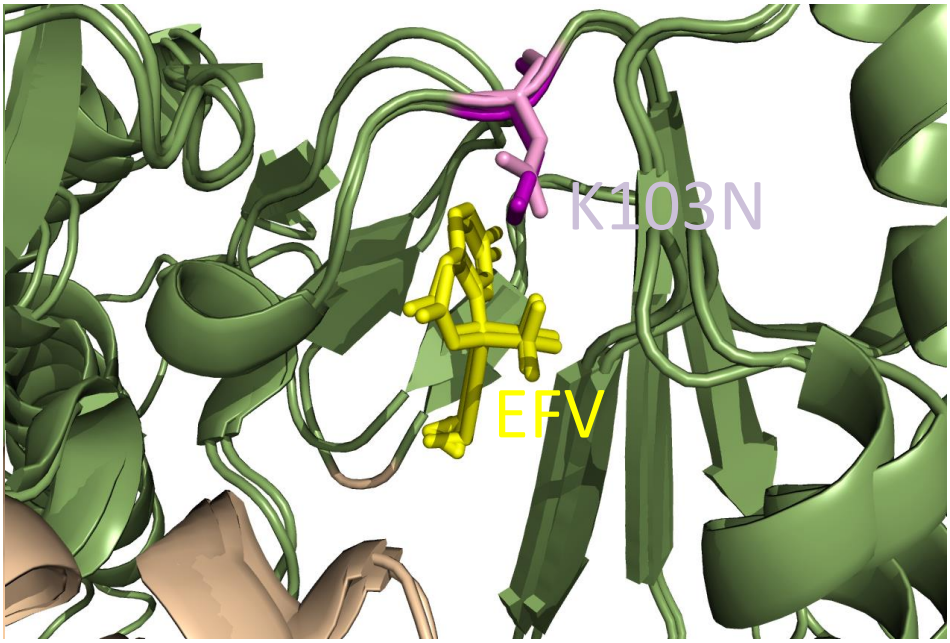

Figure 9. K103N RT Structure

The structures of RT harboring Lys103 (WT RT, purple sticks) to Asp103(K103N RT, pink sticks) are shown superimposed. The secondary structure surrounding the NNRTIBP is also shown for the p66 (green) and p51 (wheat) subunits.

It was initially proposed that the K103N mutation was responsible for adding an extra H-bond with Y188 in the NNRTIBP, stabilizing the *apo* form of RT(Hsiou et al., 2001), however, the H-bond is relatively weak compared to NNRTI binding. It was later inferred from a K103N/Y181C

double mutant that this H-bond was possibly mediated by a Na^+ ion and water molecules in the NNRTIBP (Das et al., 2007).

Unlike first generation NNRTIs, the next generation diarylpyrimidine NNRTIs, etravirine (TMC125) and rilpivirine (TMC278) are effective against K103N RT (Andries et al., 2004; Azijn et al., 2010; Lansdon et al., 2010; Rawal et al., 2012). At the moment, it is unclear what advantage these inhibitors have over single-moiety NNRTIs, and whether the strengthened H-bond network conferred by K103N is broken by this new class of NNRTIs.

1.6 METHODS BACKGROUND

1.6.1 Single-molecule Total Internal Reflection Fluorescence Microscopy (TIRFM)

Traditional spectroscopy employs bulk solutions of fluorophore-labeled biomolecules, however, bulk averaging of signals, a result of acquiring signal from many unsynchronized kinetic processes occurring simultaneously, makes spectra from these techniques difficult to interpret. In order to eliminate bulk averaging, we observe individual, surface tethered, fluorophore-labeled molecules with single-molecule total internal reflection fluorescence microscopy (TIRFM; Fagerburg and Leuba, 2011).

TIRF is an illumination scheme which relies on the difference in refractive indices of water and glass to reflect photonic light off of the interface of those surfaces while sending an evanescent wave of matching frequency into the sample (see Figure 10; Axelrod, 1989). The wave decays exponentially as $I_z = I_o e^{-z/d}$, where I_z is the intensity at distance z from surface, I_o is the intensity at the interface, and d is the penetration depth, only penetrating ~100 nm deep. A

low background, high contrast technique, TIRFM effectively eliminates excitation of other molecules in solution besides the small illuminated patch under observation. We use prism-based TIRFM in these studies, which is ideal for single-molecule experiments (Fagerburg and Leuba, 2011). See Figure 10 for a schematic of TIRFM.

Biomolecules of interest are tethered to homemade quartz flow cells through biotin-streptravidin-biotinPEG linkages. To reduce nonspecific binding of proteins to the surface, flow cells are surface passivated by PEGylation (Roy et al., 2008). Molecules are deposited on the surface at optimal surface density (typically ~100 molecules per 512x512 pixel field of view) and imaged with a highly sensitive electron multiplying CCD (EMCCD) camera. The vacuum-sealed camera CCD is thermoelectrically and liquid cooled to -90C to minimize dark current, or signal arising from non-photonic sources such as heat. Photostability of the fluorophores is maximized with oxygen scavengers (to extend the pre-bleaching observation lifetime of the molecule to the minutes scale) and triplet state quenchers (to eliminate photoblinking).

Small regions on the slide surface (~100 μm x 50 μm) are then excited with the appropriate excitation wavelength. Molecules can optionally be excited by donor excitation (532 nm) or acceptor excitation (647 nm) diode lasers. To solely monitor the fluorescence of Cy3 (as in PIFE), we simply illuminate our sample with the green (532 nm) laser and collect its emission. Alternately, to collect FRET data, we excite the donor and separately collect I_D and I_A (see Figure 10). Using Alternating Laser Excitation (ALEX; Lee et al., 2005), we are additionally able to probe the integrity of the acceptor dye by directly exciting it with acceptor excitation.

Fluorescence emission is collected through a 60x 1.2NA water-immersion objective and first passes through a 550LP filter to remove scattered donor-excitation laser light. The fluorescence emission is next split into two separate paths via a DualView^{TM} unit (Optical

Insights, Inc.) which uses a 610 nm dichroic mirror to split the emission into two fields: emission shorter than 610 nm is sent through the donor signal path and additionally filtered by a 580/40 nm bandpass filter, while wavelengths longer than 610nm are sent through the acceptor signal path, it is also filtered by a 660LP filter. The resultant signals are finally imaged onto separate halves of a thermoelectrically cooled EMCCD (Andor iXon). The resultant images represent diffraction limited spots, i.e., are a point spread function (PSFs) limited by the optics of the system with a maximum resolution of ~250 nm. The crux of single-molecule spectroscopy, however, is to accurately localize the signals of the surface-tethered molecules by fitting the PSFs with 2D Gaussian functions and localizing their origin. Integrating this function extracts the time-resolved intensity from a single fluorophore. In single-molecule FRET experiments, the two wavelength-separated projected regions are then superimposed onto each other using linear transformation as well as functions which account for small amounts of warping and/or aberration in the respective field (Hinterdorfer and Van Oijen, 2009).

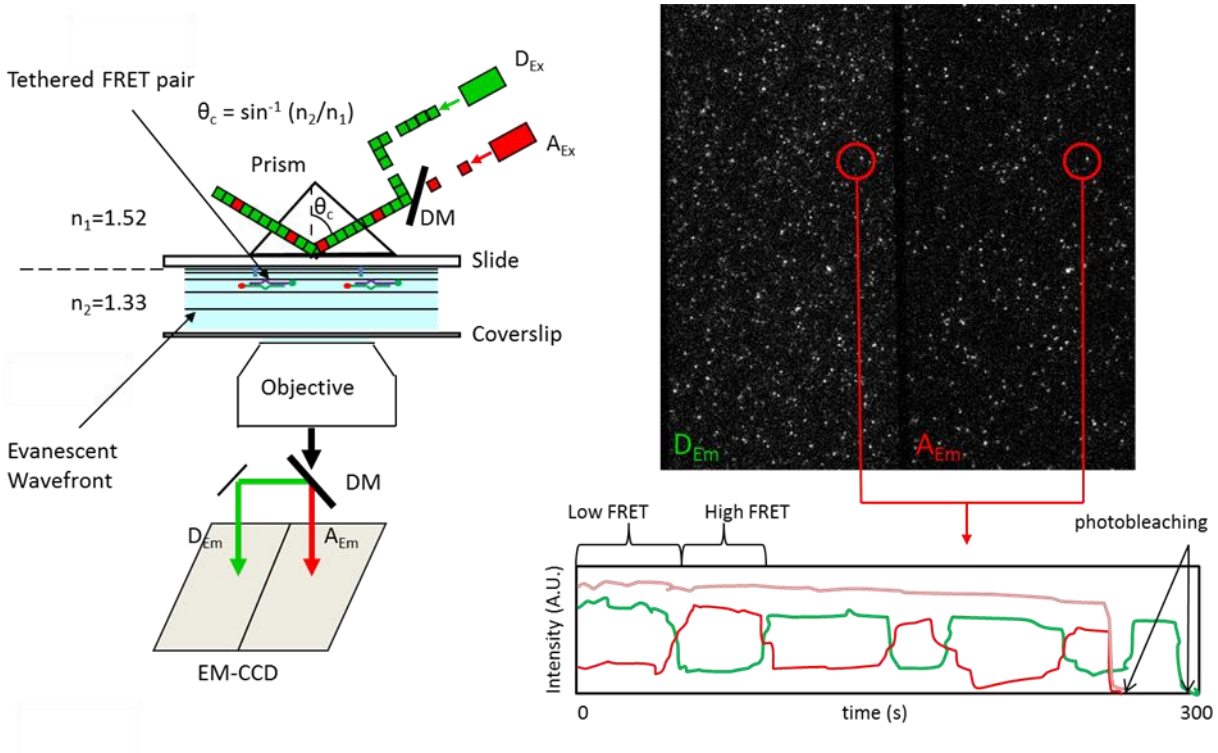

Figure 10. Schematic of single-molecule TIRFM.

The setup described in the text is shown in schematic form. Left: TIRF optics. Negotiated by a dichroic mirror, coherent donor excitation (Dex) and Acceptor excitation (Aex) laser beams are directed through a prism (coupled to the slide with index-matching oil) and into the sample. The schematic shows theoretical biomolecules harboring a Cy3/Cy5 (green/red) FRET pair. Donor emission (Dem) and Acceptor Emission (Aem) are then projected onto the EMCCD. Right: example image, with donor signals on the left, and acceptor signals on the right. Below is a theoretical donor and acceptor trace resulting from donor excitation (red and green), as well as an acceptor trace resulting from acceptor excitation (pink). Trace shows anticorrelated behavior indicative of FRET, as well as photobleaching of both dyes.

Acquiring image stacks at speeds of up to 30 ms/frame for a 512x512 pixel region and 15 ms/frame for a 512x256 pixel region, we are then able to simultaneously observe hundreds of individual fluorescence intensity trajectories per region, making TIRFM a robust technique for monitoring many kinetic processes in parallel and at high temporal resolution.

1.6.2 Fluorescence Anisotropy

Fluorescence anisotropy (Lakowicz, 2006) takes place in a standard bulk spectrofluorimeter which includes linear polarizers mounted to motorized stages. Circularly polarized fluorescence excitation and emission can be linearly polarized in the vertical or horizontal axes by these filters according to their relative orientation. Polarized excitation of a fluorescent sample results in polarized fluorescence. Gradually, polarized fluorescence emission returns to unpolarized fluorescence, depending on the timescale of rotational diffusion; rapidly tumbling fluorophores and/or dye-labeled DNA with small hydrodynamic radii quickly depolarize, whereas slowly tumbling molecules with large hydrodynamic radii slowly depolarize. Anisotropy, the ratio of the polarized-light component to the total light intensity, is related to the speed of depolarization by the equation

$$r=r_o/(1+\tau/\theta) \qquad (1.1)$$

where τ is the fluorescence lifetime of the fluorophore (typically on the range of nanoseconds) and θ is the rotational correlation time of the molecule the fluorophore is bound to, defined as the average time the complex takes to rotate 1 radian, which can vary from picoseconds to hundreds of nanoseconds depending on the hydrodynamic radius of the molecule under study.

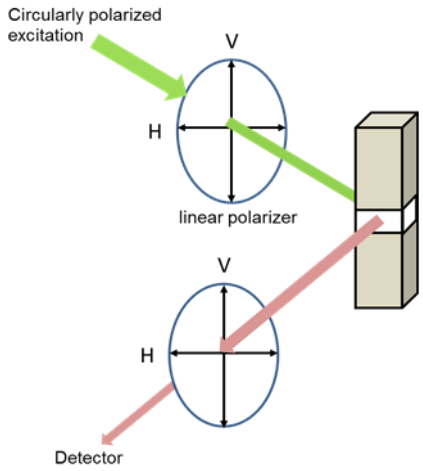

Figure 11. Schematic of fluorescence anisotropy experiment

A basic "L format" setup for anisotropy is shown. Two optical polarizes (ovals) are placed in the path of the excitation beam and the orthogal emission beam.

In the lab reference frame, we define light intensity, e.g., I_{VV} is with both excitation and emission polarizers oriented vertically; I_{HH} is for excitation and emission polarizers oriented horizontally, while I_{HV} uses a horizontal excitation polarizer and the emission uses a vertical polarizer, etc. Average anisotropy <r> is calculated as

$$<r>= (I_{VV}-G\cdot I_{VH})/(I_{VV}+2\cdot G\cdot I_{VH}) \quad (1.2)$$

where the "G factor," G, is defined as $G=I_{HV}/I_{HH}$ and is measured on a per-experiment basis. The rotational correlation time θ is related to both the volume of a rotating sphere V and the molecular weight of the complex M by the equation (Lakowicz, 2006):

$$\theta=\eta V/RT = \eta M\cdot(v+h)/RT \quad (1.3)$$

where η is the solution viscosity, R is the gas constant, T is the temperature, v is the specific volume, and h is the hydration in ml/g.

1.6.3 Molecular mechanics and molecular dynamics simulations

Grounded in theoretical physics, molecular models are a way of describing molecular systems to facilitate their analysis (Leach, 2001; Schlick, 2010). Quantum mechanical models are often used for very small systems since they can offer an extremely accurate prediction of multiple variables for a molecular system. These methods, however, are extremely computationally demanding and currently not amenable to the description of biomolecular systems. To study large systems, we must neglect quantum effects and approximate molecular motions and interactions with Newtonian physics. The Born-Oppenheimer approximation, which is invariably assumed for any type of molecular model, states that since the mass ratio of nuclei to electrons is so large, the electrons can instantaneously adjust to the positions of the nuclei. Separating the motions of nuclei and electrons, the energy of a molecule can then be considered as a function of its nucleus only. This means that we can view a large molecule as a network of nuclei-centered masses (atoms) connected essentially by springs (bonds) which can interact with each other in a closed system, responding by stretching, bending, and rotating about these bonds. This approach is known as molecular mechanics since it is based in classical mechanics, but is also known as the force-field or potential energy method. Most atomic coordinates use a Cartesian coordinate system, and atomic coordinates are usually determined by X-ray crystallography or Nuclear Magnetic Resonance (NMR). Many biological molecules can be found in the RCSB Protein Data Bank (http://www.rcsb.org/pdb/), and at the time of this writing there were 44,320 structures deposited in the Data Bank. A force field is a way to describe the potential energy of a molecule as a function of its coordinates. To develop an empirical description of the atomic potentials of a molecule, known as a parameter set, experimentalists can assign values to molecule-specific parameters by measuring their equilibrium values with

various spectroscopic methods, and these parameters can be further optimized with high-level quantum mechanical calculations. The functional form of a simple empirical force field is usually described as the sum of all intramolecular and intermolecular potentials:

$$V(r^N) = \sum \frac{k}{2}(l_i - l_{i0})^2 + \quad \text{(bonds)}$$

$$\sum \frac{k}{2}(\theta - \theta_{i,0})^2 + \quad \text{(angles)}$$

$$\sum \frac{V_n}{2}(1 + \cos(n\omega - \gamma)) + \quad \text{(torsions)}$$

$$\sum_{i=1}^{N} \sum_{j=i+1}^{N} \left(4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right) \quad \text{(nonbonded)} \qquad (1.4)$$

Where $V(r^N)$ is the total potential energy as a function of the coordinates (r) of N atoms.

Based on Hooke's law, bonds, angles, and torsions are described as masses connected by springs with a harmonic potential. The energy of an atom bonded to another atom varies with the square of the displacement from the reference bond l_0. The deviation of angles also contributes to an atomic potential as the square of angle displacement to the reference angle θ_0. Torsion here is defined as the energy barriers associated with the twisting of a molecule, and is usually defined for 4 interconnected atoms. For the torsional potential, V_n is considered the barrier height of the torsional energy, where n is the multiplicity of the function or the number of minimum points in the potential as the bond is rotated through 360°, and γ is the phase factor that describes the angle at which the torsional energy goes to zero. The nonbonded term includes a Lennard-Jones or 6-

12 potential, and a Coulombic potential. The Lennard-Jones potential models van der Waals interactions, and describes atoms interacting at distance r, with a repulsive term that changes as r^{-12} and a term for attraction that changes as r^{-6}, with a well depth ε and a collision diameter σ. The second term in the nonbonded term is a potential for two charges that is modeled after Coloumb's law, which describes two charges q, separated by distance r, where ε_0 is the permittivity of free space. Charges are distributed as partial charges on each atom so that the total charge of the molecule is 0. Using the nonbonded potential, every atom is able to "feel" every other atom in a simulation.

In Chapters 2 and 3, we implement molecular dynamics simulations to study biological molecules. Molecular dynamics simulations, as is implied by the name, simulate the time-dependant dynamics of molecules. Since we now have described the potential energy for a set of atomic coordinates, it is straightforward to see how we can simulate interactions of these potentials in time. Molecular dynamics is a deterministic method, meaning we can predict the final configuration of the system with the knowledge of the initial variables. In a typical simulation, initial velocities are randomly assigned to each coordinate for a specified temperature such that there is no overall momentum P in the system:

$$P = \sum_{i=1}^{N} m_i v_i = 0 \tag{1.5}$$

Velocities are randomly selected from a Gaussian or Boltzmann distribution at a desired temperature, which gives the probability (p) that an atom i has the velocity v_x in the x direction at temperature T:

$$p\left(v_{ix}\right) = \left(\frac{m_i}{2\,\pi k_B T}\right)^{\frac{1}{2}} \exp\left[-\frac{1}{2}\frac{m_i v_{ix}^2}{k_B T}\right] \tag{1.6}$$

Where the temperature T for the system can be calculated as:

$$T = \frac{1}{(3\,N)}\sum_{i=1}^{N}\frac{|p_i|}{2\,m_i} \tag{1.7}$$

Knowing the masses, velocities, and potential energy of each atom in the system, we can numerically integrate Newton's familiar second law to describe changes for the system in time :

$$\frac{d^2 x_i}{dt^2} = \frac{F_{\mathrm{xi}}}{m_i} \tag{1.8}$$

At each step, the vector sum of the force on each atom (the negative first derivative of the potential function with respect to the coordinates) can be computed for each atom along with the current positions and velocities, and the new positions and velocities are described for a short time step (δt) ahead. The force on a particle changes with position (as dictated by the force field), resulting in a many-body problem that cannot be solved analytically. Using finite difference integration and discretizing the time steps, we can calculate the accelerations of particles from the forces, combining them with the positions and velocities at time t to calculate the positions and velocities at time t + δt, assuming that the force has stayed constant over the length of the time step. There are many finite difference integration methods, and all of them assume that the positions, velocities, and accelerations can be approximated as a Taylor series. Commonly, a leap-frog Verlet integrator is used, using the following definitions:

$$r\left(t + \delta t\right) = r(t) + \delta t v\left(t + \frac{1}{2}\delta t\right) \tag{1.9}$$

$$v\left(t + \frac{1}{2}\delta t\right) = v\left(t - \frac{1}{2}\delta t\right) + \delta t a(t) \tag{1.10}$$

First, the velocities v(t+1/2δt) are calculated from the velocities at time t-1/2δt and the accelerations at time t. Positions r(t+δt) are calculated from the velocities together with the positions at time r(t) using equation (5). The velocities at time t can then be calculated using

$$v(t) = \frac{1}{2}\left[v\left(t + \frac{1}{2}\delta t\right) + v\left(t - \frac{1}{2}\delta t\right)\right] \tag{1.11}$$

So the velocities 'leap' over the positions to give their values at t+1/2δt, and the positions leap over the velocities giving their new values at t+δt. The more accurate velocity Verlet integrator used in Chapter 2 and 3 calculates new positions, velocities, and accelerations of atoms simultaneously and uses the following definitions:

$$r(t + \delta t) = r(t) = \delta t v(t) + \frac{1}{2}\delta t^2 a(t) \tag{1.12}$$

$$v(t + \delta t) = v(t) + \frac{1}{2}\delta t\left[a(t) + a(t + \delta t)\right] \tag{1.13}$$

which requires three stages since calculating the new velocities requires accelerations at time t and t+δt. Velocities at time t + 1/2δt are calculated:

$$v\left(t + \frac{1}{2}\delta t\right) = v(t) + \frac{1}{2}\delta t a(t) \tag{1.14}$$

Next, forces are calculated from the current positions, which gives a(t+δt). Finally, the velocities at time t+δt are calculated by:

$$v(t + \delta t) = v\left(t + \frac{1}{2}\delta t\right) + \frac{1}{2}\delta t a(t + \delta t) \tag{1.15}$$

After many iterations, the result is a trajectory for a system, or the propagation of atomic coordinates through time. Most molecular dynamics simulations tend to use a 1 femtosecond (fs) time step, since this is approximately one tenth of the period of the highest frequency

vibration in the system, the covalent bond between oxygen and hydrogen. Any larger, and the time step used would overlook these vibrations. The SHAKE algorithm is widely used to impose constraints on covalent hydrogen bonds during every time step, essentially freezing this motion and allowing an acceptable time step of 2fs. Both simulations in Chapters 2 and 3 define the system as a canonical NVT ensemble: a constant number of atoms, volume, and temperature. Volume can be controlled by imposing boundaries on the system. In all simulations in Chapter 2 and 3, periodic boundary conditions are used, which describe the simulation as one 'unit cell' with an infinite number of adjacent images of the same cell surrounding it, allowing atomic interactions between cells in order to simulate bulk phase. Desired temperature for the simulations in Chapters 2 and 3 is achieved with a Langevin heat bath, which adds random noise to the system until a desired temperature is reached and can be adjusted by a thermostat which checks the temperature at defined intervals during the simulation. Also, the calculation of nonbonded atomic interactions can be very expensive. Usually, the program keeps a running list of all atoms which interact through space, and calculates nonbonded forces only between those atoms. Cutoff schemes to truncate this list, and to bring the Lennard-Jones potential and the Coulombic potential to zero at some finite distance are commonly used in the simulations in Chapters 2 and 3 since the 'real' potential only gets infinitesimally close to zero as interatomic distance increases. Lastly, Accelerated Molecular Dynamics (AMD; Hamelberg et al., 2004) is implemented in all simulations herein. AMD is a method of adding a biasing potential to the overall dihedral potential term in order to make it shallower, effectively facilitating sampling while maintaining the native free energy landscape of the system.

2.0 QUANTIFYING THE DYNAMIC INTERACTIONS BETWEEN HIV-1 RT AND ITS TEMPLATE/PRIMER (T/P) SUBSTRATE WITH SINGLE-MOLECULE AND BULK FLUORESCENCE SPECTROSCOPY

Non-Nucleoside Reverse Transcriptase Inhibitors (NNRTIs) are an attractive drug class since they are highly effective against RT in a specific and, thus, relatively nontoxic manner. However, NNRTI-resistant strains are continually emerging, highlighting an urgent need to understand the mechanism(s) underlying drug resistance. In RT, K103N is one of the most common mutations arising from efavirenz-based treatment regimens, yet the mechanism by which K103N confers efavirenz resistance is unknown. We show that efavirenz causes RT to relinquish its grip on the template/primer substrate via “molecular arthritis," accompanied by increased shuttling on the substrate, reducing time spent in a polymerase-competent configuration. The K103N mutation relieves the arthritis in the fingers and thumb sub-domains of RT, enabling the efavirenz-bound enzyme to form a stable polymerase-competent complex. We demonstrate that relief of molecular arthritis is likely caused by disruption of a salt bridge between K101 and E138, residues at a key hinge site in the RT heterodimer. Our data suggests a unique mechanism of resistance that is mediated by interplay between intramolecular conformational changes in RT and intermolecular dynamics of the RT-template/primer-dNTP complex.

2.1 INTRODUCTION

In this study, we wished to characterize the dynamics of RT on its T/P. In order to accomplish this, we used a single molecule TIRFM technique known as Protein Induced Fluorescence Enhancement (PIFE; Hwang 2011) to monitor dynamics of individual complexes. This technique has the advantages over the use of Förster Resonance Energy Transfer (FRET) to measure RT dynamics as previously reported (Liu et al., 2008) since it has a shorter distance sensitivity (~0-4 nm for PIFE vs. ~5-8 nM for FRET), does not perturb the structure of RT with fluorophore labeling and corequisite site-directed mutagenesis, and, by restricting the fluorophore to the nucleic acid substrate, allows for the use of physiologically relevant concentrations of RT without saturating the CCD detector.

We furthermore devised a technique to measure the mobility of RT on its template/primer (T/P) substrate which should be extensible to any protein-DNA system. Using this technique, we are able measure two simultaneous quantities: 1) the relative effect of NNRTIs on RT mobility and 2) the binding affinity of NNRTIs. Of particular importance to researchers in the RT field will be the second application, since there are few biophysical techniques able to measure the binding affinity directly, which is made more difficult due to the insolubility of NNRTIs (e.g., isothermal calorimetry cannot be used due to the high concentrations of NNRTI required).

To understand the mechanism of NNRTI resistance by K103N, we combined these novel techniques with single-molecule Forster Resonance Energy Transfer (FRET) to measure, in unprecedented detail, the changes in RT-T/P dynamics, dNTP affinity, and intramolecular FRET. Using this combination of approaches, we were able to demonstrate that the NNRTI resistance mutation K103N works by alleviating molecular arthritis, mediated by two crucial residues at the intersubunit hinge site in RT.

2.2 MATERIALS AND METHODS

2.2.1 Template/Primer (T/P) constructs

All oligonucleotides were ordered from IDT. Oligonucleotides were resuspended to 200 μM in buffer (50 mM Tris-HCl pH 7.5, 25 mM NaCl) and annealed by mixing at a 1.2:1 biotin-Template:Primer ratio (ensuring all Primer-Cy3 signals were Template-bound), heated at 94°C for 5 minutes on a heating block, and allowed to cool on the block for 1 hour before freezing at -20°C.

To facilitate analysis, the T/P substrate used in this study is identical to that in PDB 1RTD, with the addition of an extended template for tethering purposes. The sequences used were:

Template: 5'-biotin-GGGTTTGCTAAGCACCGGCGCCCGAACAGGGACTG

Primer (PIFE): 5'-Cy3-CAGTCCCTGTTCGGGCGC-3'-ddC

Primer (FRET): 5'-CAGTCCCTGTTCGGGCGC-3'-ddC

For experiments involving T/P lengthening, we used the following sequences:

Template_2: 5'-biotin-GGGTTTGCTAGAGAGCGGCGCCCGAACAGGGACTG

Primer (+2b): 5'-Cy3-CAGTCCCTGTTCGGGCGCCGCT-3'-ddC

Primer (+4b): 5'-Cy3-CAGTCCCTGTTCGGGCGCCGCTCT-3'-ddC

The following were used in the anisotropy experiments:

Template: 5'-GGGTTTGCTAAGCACCGGCGCCCGAACAGGGACTG

Primer (Anisotropy): 5’-Fluorescein-CAGTCCCTGTTCGGGCGC-3’-ddC

2.2.2 RT expression, purification, and labeling

WT and mutant RT were expressed and purified as previously described (Brehm et al., 2008; Radzio and Sluis-Cremer, 2008). FRET constructs were prepared by mutating endogenous cysteines C38 and C280 to serines and introducing cysteines into codons 250 (thumb) and 139 (fingers) with site-directed mutagenesis. These positions were chosen due to their optimal separation distance (~5 nM), ideal for reporting FRET in the quasi-linear response regime, and also for their high degree of solvent accessibility (both residues reside in solvent-exposed loops in WT RT; see Figure 5a). We then labeled the resultant C38S/C280S/T139C/D250C RT simultaneously with equimolar amounts of Cy3- and Cy5-maleimides (GE Healthcare) at pH 8.0 according to the manufacturer instructions.

2.2.3 Single-molecule Total Internal Reflection Fluorescence Microscopy (TIRFM)

Single-molecule TIRFM was performed on an Olympus IX-71 configured with prism-based TIRF illumination geometry and imaged with an Andor iXon 897 back-illuminated electron multiplying CCD (EMCCD). Donor and acceptor emission signals were separated by a 610 nm dichroic longpass mirror, a 580/40 nm bandpass filter and a 660 nm longpass filter and subsequently imaged.

2.2.4 Protein Induced Fluorescence Enhancement (PIFE) Experiments

PIFE data was acquired on a 512x256 pixel region of an Electron Multiplying CCD (EMCCD, Andor Technologies) at 30 ms/frame. T/P molecules were surface tethered to the PEGylated flow cell via a biotin:streptavidin:biotin-PEG linkage. Concentrations of T/P used were typically 20 pM for an optimal surface density of ~100 molecules per field of view. 250 nM RT was introduced to the flow cell and incubated for 5 minutes, followed by a wash step with 5X volume of imaging buffer. The power of the 532 nm excitation beam was typically ~8 mW. Fluorescence intensity traces which visibly fluctuated and also exhibited single-step photobleaching were selected for analysis (typically 25%-50% of all molecules, as many had not bleached yet and also possibly since many molecules remained in high PIFE states throughout observation time). PIFE traces were then rescaled to "fold intensity" (as in Figure 13a-c) by fitting each trajectory to two Gaussian functions and dividing the data by the fitted average of the lowest peak in order to rescale it to 1-fold. PIFE histograms (e.g., Figure 13d) were constructed by binning many PIFE traces into 50 bins and fitting with two Gaussians to extract the average fold intensity increase. Data from Figure 20 (red circles) was constructed of the average of 3 of such values. Data was also rescaled for use with the vbFRET program, and fluctuating fluorescence intensity traces were fit to idealized states. Idealized dwell times aggregated into cumulative residence time distributions (employed to eliminate any bias from choice of bin number), and these distributions were fit with single exponential functions to calculate shuttling rates.

2.2.5 Single-molecule Förster Resonance Energy Transfer Experiments

For FRET data, unlabeled T/P was surface tethered by introducing ~2μM T/P into the flow cell, ensuring a dense "carpet" of unlabeled substrates for RT to bind. FRET-pair labeled RT was subsequently introduced at concentrations of ~40 pM. Using this method, we were able to avoid the problem excess background noise typical of saturating the EMCDD with relatively higher (nM range) fluorophore-labeled protein concentrations. Based on the relative distribution of donor and acceptor intensities in the observed field of view with TIRFM, dual-labeling efficiency was approximately 5-10% for all final constructs (data not shown). FRET data was acquired at 100 ms/frame, and peak choice was thresholded on Cy5 intensity such that only molecules exhibiting FRET were chosen. E_{app} was calculated as $I_A/(I_D+I_A)$, where I_D and I_A respectively denote background-corrected donor and acceptor intensities. Noting that E_{app} should remain the same value regardless of the relative arrangement of Cy3 or Cy5 on D250C/T139C, E_{app} histograms were subsequently constructed from all traces resulting in donor and acceptor photobleaching (see Supplemental Figure 3a), fit with 25 bins each, and fit with single Gaussian functions (see Supplemental Figure 3b). Data in the Results section is reported as the mean ± std of these fits. All analysis of single-molecule data was performed using custom-written MATLAB scripts (see Appendix).

2.2.6 Accelerated Molecular Dynamics (AMD) Simulations

All Molecular Dynamics (MD) simulations were performed in NAMD with atomic coordinates of the ternary complex (PDB: 1RTD) modified at the 5-end of the primer with coordinates of

first excited state of Cy3 optimized at the CIS/6-31G(d,p) level (Courtesy David Norman). Systems were solvated in a 125 Å x 96 Å x 96 Å periodic box of TIP3 waters, ionized with 25 mM NaCl, and loaded with the CHARMM27 force field including Cy3 parameters courtesy of Arjan van der Vaart. In their article, Spiriti et al. (Spiriti et al., 2011) stiffened the dihedral parameters of the polymethine chain connecting the heterocyclic rings to explicitly prevent excessive Cy3 isomerization. Upon request, Spiriti and van der Vaart generously provided the original dihedral parameters for the linker, which are herein provided in the Supplemental Information. For simulations which lengthened the distance between the 5'-Cy3 and RT, the DNA duplex was rotated 35.3° and translated 3.4 Å upstream along its principle axis for each increase in 1 bp of separation (initial configurations shown in Supplemental Figure 2a), and periodic boxes were grown as necessary (total size ranging from ~100,000-125,000 atoms). Systems were subjected to 100,000 steps of conjugate gradients minimization, heated to 300 K in increments of 30 K over the course of 1 ns, and allowed to equilibrate for an additional 1 ns in an NPT ensemble with traditional MD. Simulations were subsequently continued using AMD, a method of enhanced sampling which enabling us to capture Cy3 isomerization events, for an additional 50 ns each. For every system simulated, the NAMD parameters for AMD implementation, accelMDE and accelMDalpha, were defined as accelMDE=V_d +4·N_{res} and accelMDalpha=4·(N_{res}/5), where V_d is the average of the dihedral potential over the course of the 1 ns of MD equilibration after heating, and N_{res} is the number of residues in the system. For the MD portion, 0.05 kcal/mol·Å^2 harmonic constraints were added to every Cα and DNA P atom. All constraints within 25 Å of Cy3 were subsequently removed for the AMD portion. All simulations were run with 2 fs time steps (employing the SHAKE algorithm), Particle-mesh Ewald electrostatics, a Langevin Thermostat set to 300 K, and a Langevin Barostat set to

1.01325 atm. Trajectories were analyzed with VMD. θ_{Cy3} was defined as the dihedral angle between the Cy3 atoms C5A,C5B,N1A, and N1B. All AMD simulations were performed in triplicate for each condition.

2.2.7 Anisotropy Experiments

Anisotropy experiments were performed as previously described on a Cary Eclipse spectrofluorimeter using 485 nm excitation and 520 nm emission, with T/P-fluorescein concentrations (400uL volumes) of ≤5 nM. Anisotropy values were collected with an integration time of 0.25 s for three consecutive readings, and each value represents the average of 3 independent experiments.

2.2.8 Buffer components and small-molecule ligand concentrations

All experiments were performed in 50mM Tris-HCl pH 7.5, 25 mM NaCl, and 5 mM $MgCl_2$. TIRFM experiments additionally contained 0.1 mg/ml glucose oxidase, 0.02 mg/ml catalase, 0.4% wt/v β-d-glucose, and 2 mM Trolox for photoprotective purposes. Unless otherwise indicated (e.g., with titrations), TTP was held at 50 μM in all TIRFM experiments and 100 μM in all anisotropy experiments, and EFV concentration was held at 500 nM in all experiments, ensuring saturated binding conditions for both ligands.

2.3 RESULTS

2.3.1 PIFE-based assay of RT-T/P shuttling dynamics in the context of polymerization

In order to measure RT dynamics on its T/P substrate, we employed a single-molecule technique based on Protein Induced Fluorescence Enhancement (PIFE), a phenomenon recently characterized as a distance-dependent enhancement of the intensity of a fluorophore due solely to its relative proximity to protein (Fischer et al., 2004; Hwang et al., 2011). To probe RT shuttling on T/P as previously reported (Liu et al., 2008), we designed a T/P reporter construct consisting of a 5'-biotinylated Template annealed to a Primer with a 5'-Cy3 with a 3'-ddC (Figure 12).

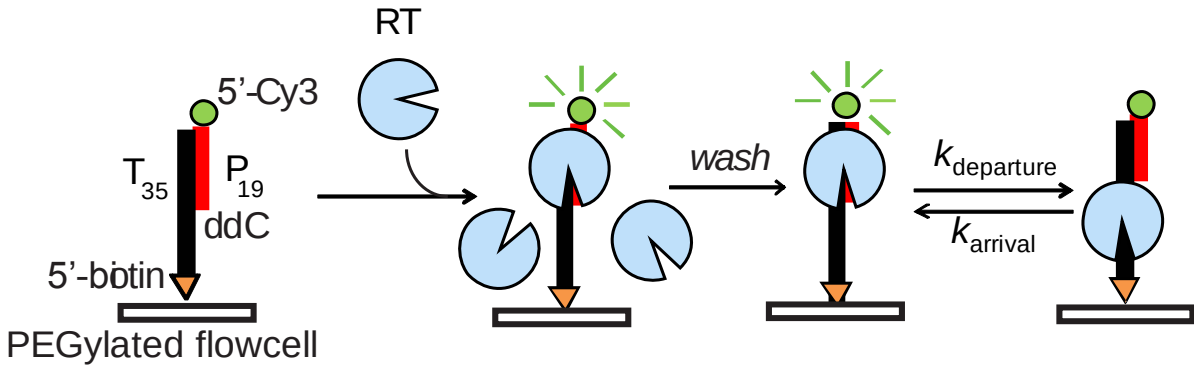

Figure 12. Schematic of PIFE assay of RT shuttling

In our shuttling assay, the T/P reporter construct is composed of a 5'-Cy3 labeled, 5'-dideoxy terminated Primer (red) a 5'-biotinylated Template (black) is surface tethered to a PEGylated flow cell (see Methods), incubated with RT, and washed to eliminate unbound RT. PIFE occurs when RT is proximal to the terminal Cy3.

The 3'-dideoxy nucleotide allowed us to add the next correct dNTP without enzymatic primer extension by RT, resulting in a nonproductive ternary complex. The T/P substrate was surface tethered to a PEGylated quartz flow cell and visualized at the single-molecule level with Total

Internal Reflection Fluorescence Microscopy (TIRFM), and the resulting fluorescence intensity traces were normalized for PIFE analysis (see Methods section). In the absence of RT, the majority (>95%) of T/P intensity trajectories exhibited no observable fluctuations (Figure 13a). After incubating the T/P with RT and subsequently washing the flow cell with 5X volume imaging buffer to eliminate signal from RT association and/or dissociation events, we observed fluctuating PIFE signals with gradual transitions between 1- and 2-fold intensity from the majority (>75%) of the fluorescence traces (Figure 13b), indicative of RT shuttling on the T/P. Upon addition of TTP, the next correct dNTP, we observed the stabilization of the PIFE signal into two discrete low and high states (Figure 13c), resulting in an average intensity increase of ~1.6-fold (Figure 13d). In agreement with our single-molecule experiments, we also observed concentration-dependent PIFE in bulk (Figure 14).

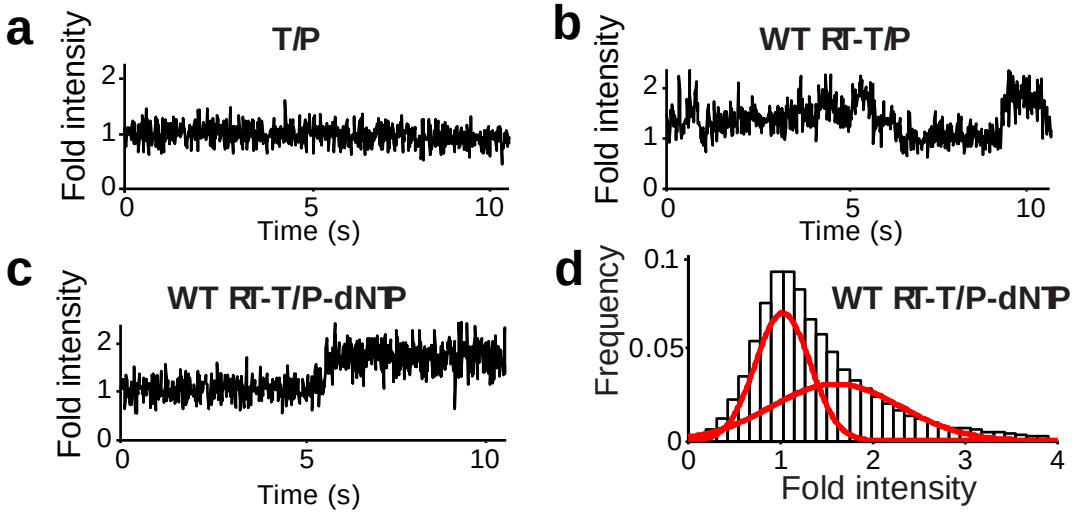

Figure 13. Probing the polymerase-competent mode with PIFE

Representative PIFE traces for the indicated experimental conditions. Fluorescence intensity is normalized to a fold scale, with the lowest intensity set to 1-fold a) T/P-Cy3 in the absence of RT. b) WT-T/P. and c) WT RT-T/P-dNTP. d) Representative PIFE histogram. PIFE signals (n=106) were binned into 50 bins and fit with two Gaussians. The fitted means (n=3) of similar histograms were then used for the final average (e.g., for the 1.58 ±

0.07 point in Figure 19 for the ternary complex). Observation time was set to 10 s for all traces for a direct comparison between traces.

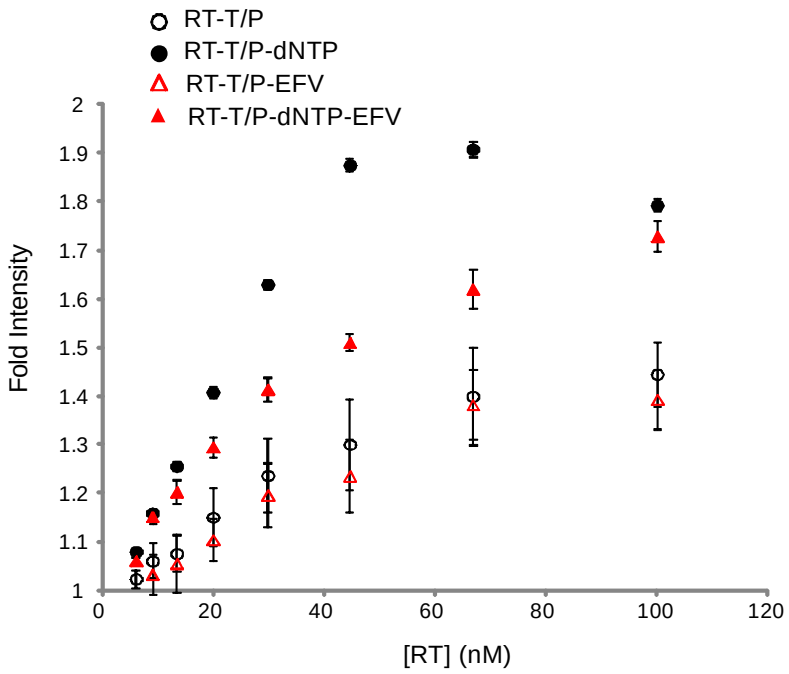

Figure 14. Bulk PIFE shows RT-T/P association

Bulk PIFE experiments were carried out on a standard spectrofluorimeter (see Methods). PIFE is represented as the Cy3 emission intensity at 570 nm at the given RT concentrations, normalized by the intensity at [RT] = 0. Points represent n=3 independent experiments; error bars indicate s.e.m.

To better understand the origin of the high PIFE signal we observed, we performed Accelerated Molecular Dynamics (AMD) simulations on the crystal structure of the RT-T/P-dNTP complex (Huang et al., 1998; PDB ID 1RTD) modified with a terminal Cy3 and placed in explicit solvent (see Methods section) while monitoring θ_{Cy3}, the dihedral angle between the planes of the heterocyclic rings of Cy3 (Figure 16, left). Contrary to the case of T/P in the absence of RT, where the terminal Cy3 had a high degree of conformational freedom, Cy3 was

observed to closely interact with RT in a nonspecific manner, confining the conformational space explored by Cy3 (Figures 15 and 17).

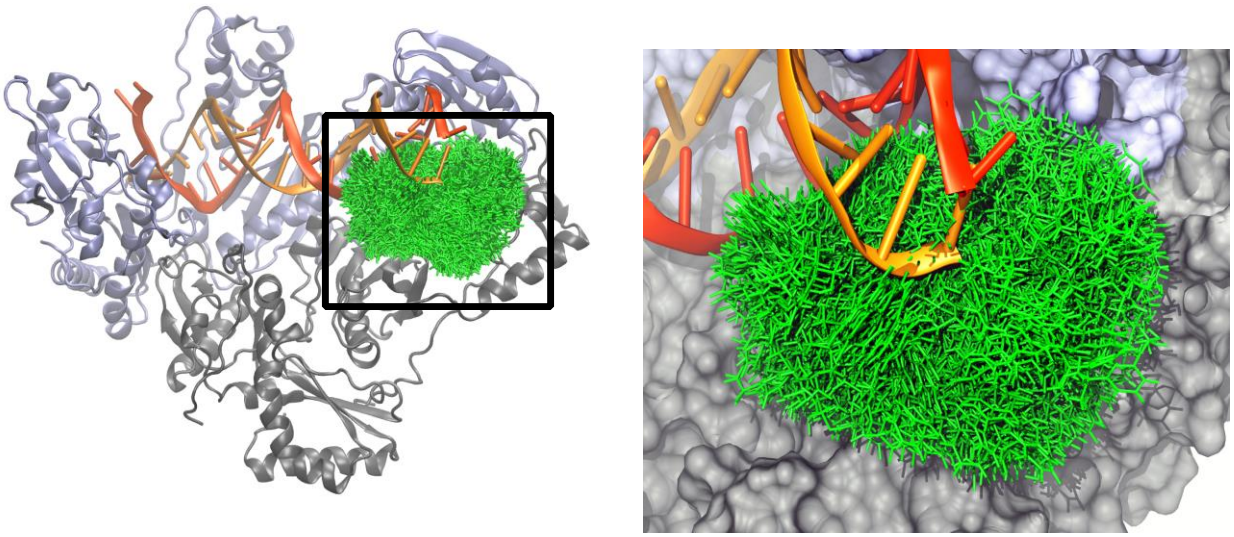

Figure 15. Cy3 Interacts with RT

The RT-T/P complex is shown, highlighting the nonspecific interaction of Cy3 with RT. The conformation of Cy3 is drawn every 0.1 ns for a 50 ns AMD simulation.

This interaction with RT markedly changed the distribution of θ_{Cy3}, enhancing the photoactive *cis* isomer of Cy3 (Figure 16, right). In the case of the ternary configuration, RT increased the overall time Cy3 spent in the *cis* conformation from 39.2% ± 3.3% to 60.4% ± 6.0% for an average relative increase of 1.54 ± 0.15 fold (± s.e.m., n=3 50 ns simulations).

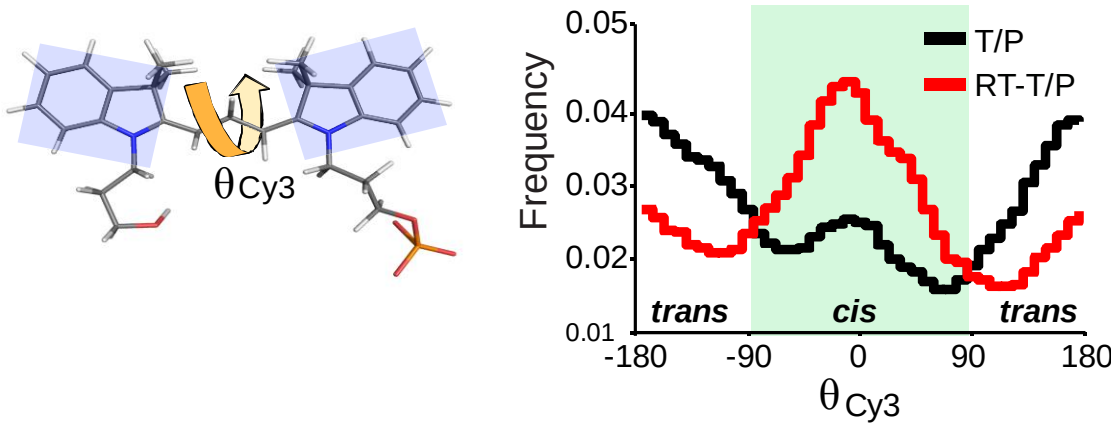

Figure 16. Interaction with RT stabilizes the *cis* conformation of Cy3

Left: Definition of $\theta_{Cy3,}$ the dihedral angle between the planes of the heterocyclic rings of Cy3. Right: distributions of θ_{Cy3} in the presence (red) or absence (black) of RT. The distributions represent data from 3 AMD simulations each. The *cis* isomeric state is highlighted in green.

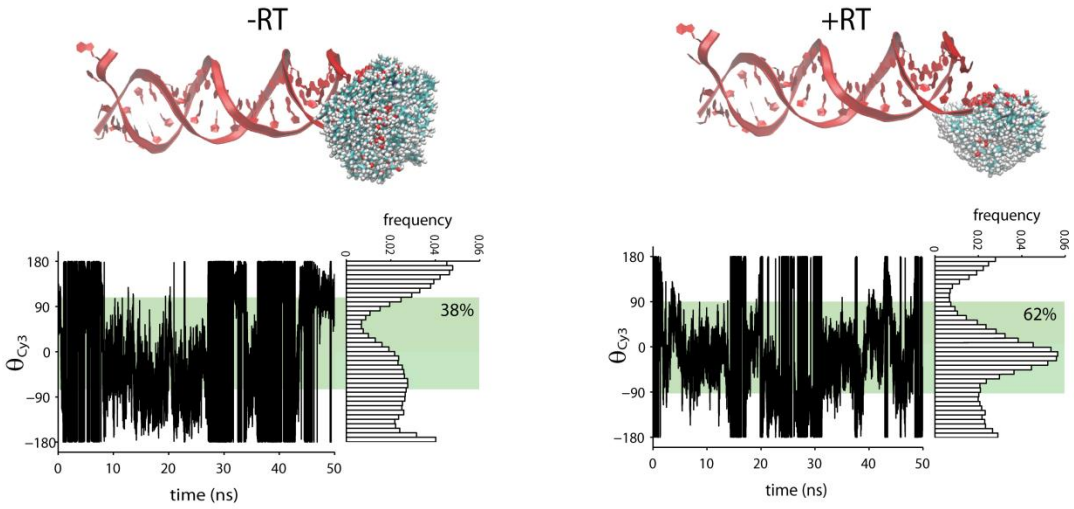

Figure 17. Representative simulation data

AMD simulations with (left) and without (right) RT are shown. Top: Cy3 configurations from every 0.1 ns drawn. Bottom: Representative data over 50 ns simulations tracking θ_{Cy3} is shown. Histograms of θ_{Cy3} were binned, and the percentage of time spent in the *cis* conformation is shown. For the data in Figure 17, n=3 of such simulations were combined into one histogram for each condition (i.e., T/P +/- RT).

In order to accurately quantify the relationship between the PIFE signal we observed and the position of RT on the T/P relative to the terminal Cy3, we next extended the distance of the terminal Cy3 1 to 5 bp away from the starting configuration and performed similar AMD simulations as above (Figure 18). The simulations revealed an inverse relationship between the distance of RT to Cy3 and the enhancement of the *cis* isomer (Figure 1g, red circles). We also performed similar PIFE experiments, wherein we lengthened the T/P either 2 bp or 4 bp upstream (relative to the 3' end of the primer) of the original ternary configuration (Figure 1g, top).

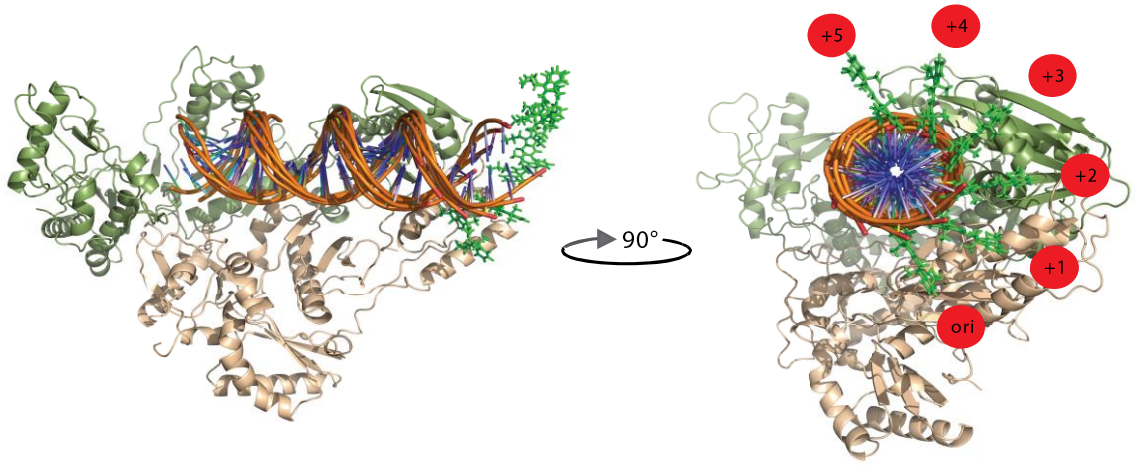

Figure 18. Starting configuration of RT-T/P in AMD simulations

The starting configuration of the RT-T/P complexes in the AMD simulations are shown, where ori stands for origin at the ternary complex (PDB ID 1RTD), and +1-5 represents primers translated 1-5 bp upstream as discussed in the Methods section.

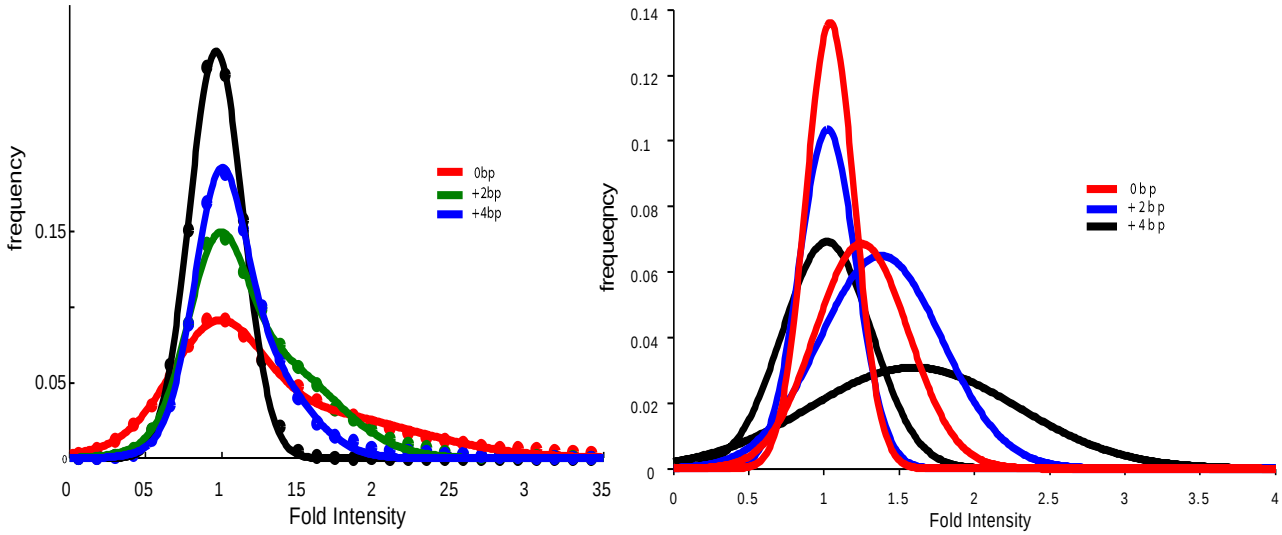

Figure 19. Extracting fold-intensity values

Left: Representative fitted histograms from RT-T/P (red), RT-T/P + 2bp (green), and RT-T/P + 4 bp (blue). Points represent bin centers, and lines represent composite fits from two-component Gaussian fits. Right: Individual Gaussian peaks from curve fits of RT-T/P (red), RT-T/P + 2bp (blue), and RT-T/P + 4 bp (black). The averages of n=3 such peaks were used for the data presented in Figure 19.

Similar to the result for the simulations, we observed a distance dependence on the fold increase in Cy3 intensity in analogous PIFE experiments (Figure 20, black squares; see also Figure 19). Comparing the distance dependence on the experimental fold enhancement of Cy3 intensity with that of the enhancement of the *cis* isomer in the simulations revealed markedly similar trends (R=0.92). Particularly striking was the similarity between the experimental PIFE signal for the ternary complex (1.58 ± 0.07) and the fold enhancement of the *cis* isomer for the ternary configuration in the simulations (1.54 ± 0.15), leading us to conclude that the high PIFE signal observed upon incorporation of the next correct dNTP was a result of RT stably residing in polymerase-competent configuration.

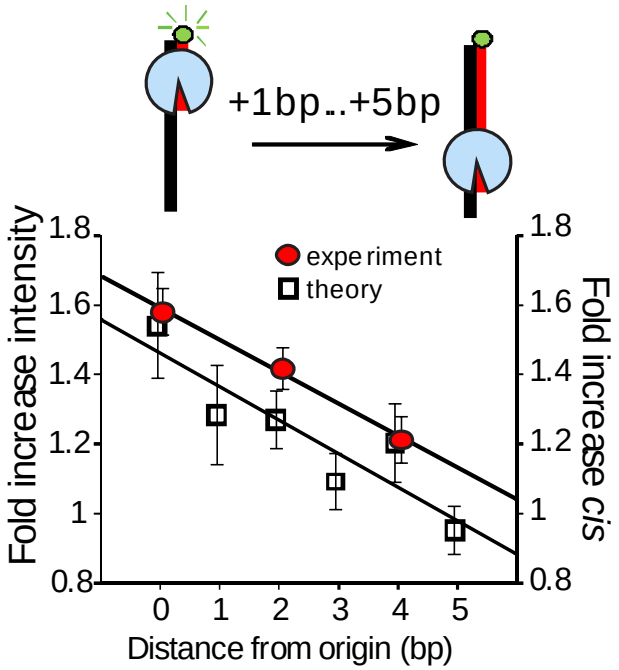

Figure 20. Accurate identification of the polymerase competent configuration

Top: schematic of extension experiments. Cy3 was extended from 1 to 5bp away from RT in the polymerase competent state. Bottom: Fold intensity changes from PIFE experiment (squares, left y-axis) are compared to the fold increase in the *cis* isomer of Cy3 (red circles, right y-axis)

2.3.2 Efavirenz alters shuttling dynamics of WT, not K103N, RT

With the ability to accurately identify the polymerase-competent RT-T/P configuration, we probed the effect of efavirenz (EFV) on the shuttling dynamics of both WT RT and K103N RT with respect to residence time in the polymerase mode via dwell-time analysis of PIFE traces (see Methods).

In the absence of EFV, the ternary complex displayed a stable, high PIFE state (Figure 21a, left) as discussed above, with relatively few transitions observed per trace, (Figure 21c, left, blue). Upon infusion with EFV, RT spent relatively shorter times in the polymerase mode (Figure 21a, right), displaying many more transitions away from this state (Figure 21c, left, red).

In contrast, PIFE traces for the K103N RT-T/P-dNTP complex (Figure 21b) were indistinguishable to those of WT RT in either the absence (Figure 21b, left) or presence (Figure 21b, right) of EFV and displayed a similar number of transitions in both cases (Figure 2c, right). Using dwell-time analysis (see Methods), we calculated the rates to and from the polymerase mode, $k_{arrival}$ and $k_{departure}$, respectively, and the results are summarized in Figure 21d. While addition of EFV to the WT RT-dNTP complex (Figure 21d, left) resulted in significant ~5- and ~6-fold increases in $k_{arrival}$ and $k_{departure}$, respectively, addition of EFV to the K103N RT-dNTP complex (Figure 21d, right) resulted in an insignificant ~2.5-fold change in $k_{arrival}$ and a comparatively smaller ~2-fold change in $k_{departure}$. Taken together, these results suggest that EFV destabilizes the polymerase-competent mode in WT, but not K103N RT.

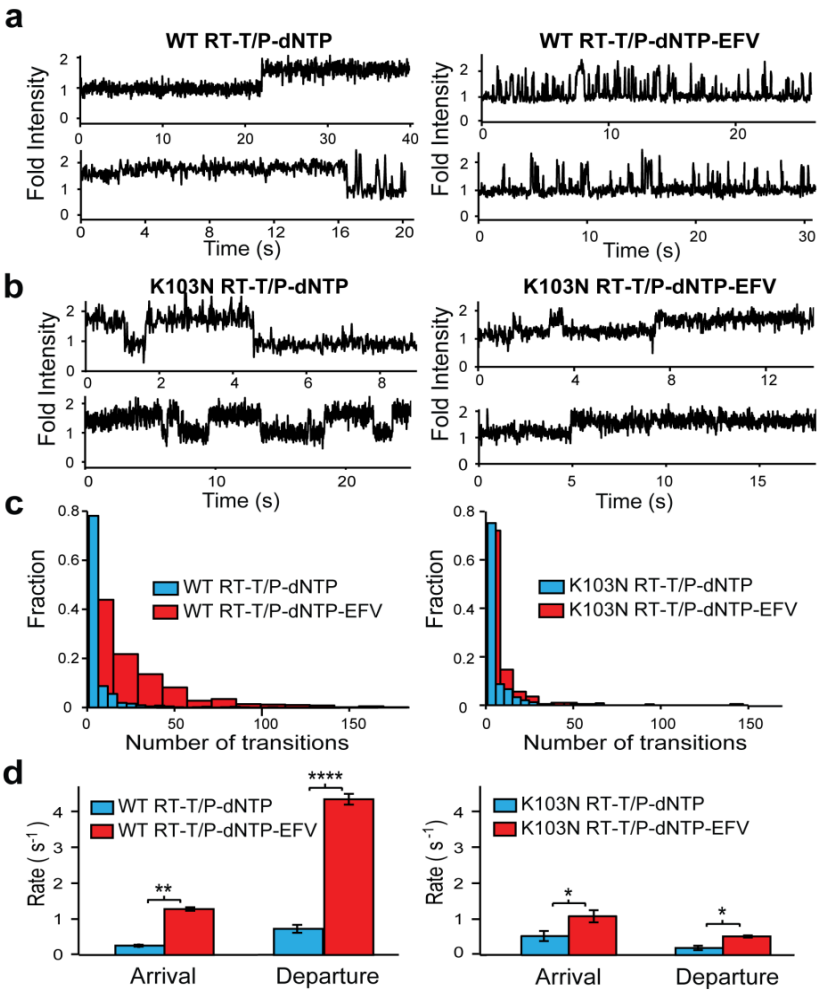

Figure 21. EFV destabilizes the polymerase mode in WT, but not K103N, RT

Representative PIFE traces are shown for **a)** WT RT-T/P-dNTP or **b)** K103N-RT-T/P-dNTP in the absence (left) or presence (right) of EFV. **c)** Transition histograms showing the average number of transitions per trace WT RT-T/P-dNTP (left) or K103N RT-T/P-dNTP (right) in the presence (red) or absence (blue) of EFV. **c)** Rates of arrival from or departure to the polymerase competent state for WT RT-T/P-dNTP (left) or K103N RT-T/P-dNTP (right) in the presence (red) or absence (blue) of EFV. Error bars represent the mean +/- s.e.m. of n=3 fits from analysis of 3 groups derived from n=630 and n=573 total molecules for WT RT-T/P-dNTP in the respective absence or presence of EFV, and n=926 and n=725 total molecules for K103N RT-T/P-dNTP in the respective absence or presence of EFV.

2.3.3 Development of a novel anisotropy-based assay to simultaneously report relative mobility and NNRTI binding affinity

We next asked whether presence of dNTP or EFV could alter the affinity of RT for its T/P substrate. Using fluorescence anisotropy with a T/P substrate identical to that used in the PIFE experiments but with a terminal 5'-flouorescein on the primer in place of the 5'-Cy3, we performed titrations of WT RT and recorded anisotropy binding curves in the presence or absence of saturating concentrations of EFV and/or dNTP (see Methods section). Interestingly, neither ligand significantly affected the binding affinity of RT for T/P, with the dissociation constant K_d in the ~10 nM range for all cases. Furthermore, we unexpectedly observed large changes in the absolute anisotropy accompanying the presence of various ligands (Figure 22).

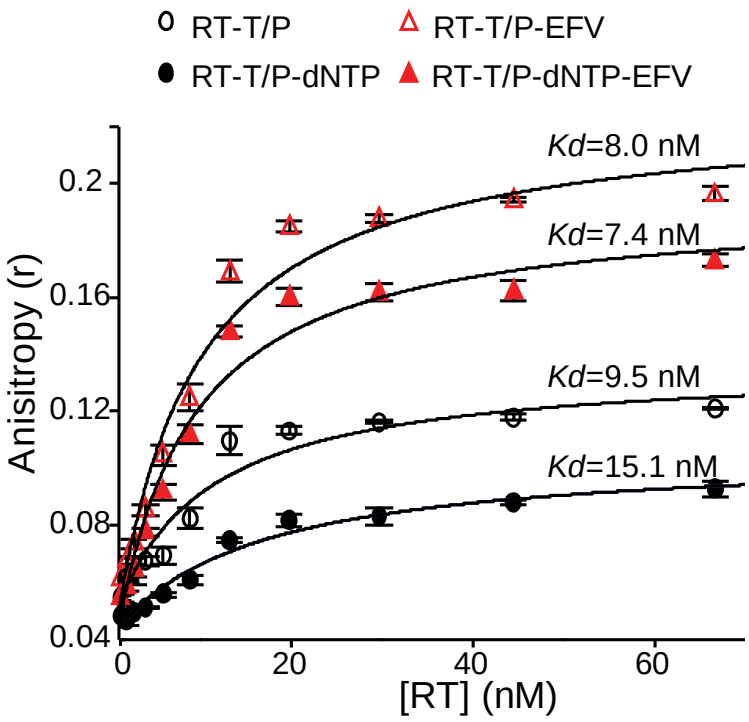

Figure 22. Small molecule ligands do not affect RT-T/P affinity

Bulk anisotropy binding curves. Anisotropy from T/P-fluorescein is monitored during [RT] titration in the presence of the ligands indicated. All K_d values of RT for T/P remained in the ~10 nM range regardless of ligand.

To understand the source of these observed differences, we calculated the anisotropy values expected for various sphere sizes and also for the molecular weight (MW) of the complex (Figure 23) and made two observations: first, due to the large size of the complex, the response of anisotropy to the spherical size of the complex falls outside of the quasi-linear regime for the value expected for the MW, which largely ruled out the possibility that, e.g., increases in the effective hydrodynamic radius due to a broader occupancy of RT on the substrate were responsible for increases in anisotropy. Second, the anisotropy we calculated is much lower than that expected for this complex based on the MW (Figure 23, top right), which is likely due to the contribution from the fast tumbling of fluorescein (Figure 23, bottom right).

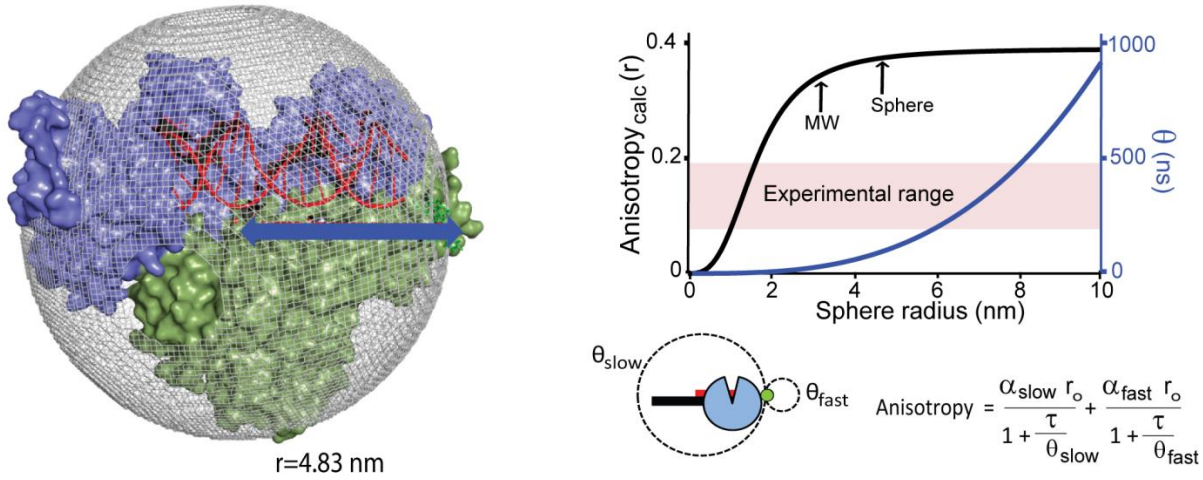

Figure 23. Changes in hydrodynamic radius are insufficient to explain changes in anisotropy

Left: spherical approximation of RT-T/P. Top right: Anisotropy calculated according to first principles calculations (see section 1.6.2. for more info). Arrows point to the anisotropy expected for a sphere of r=4.83 nm and the anisotropy expected for the molecular weight (MW) of the complex, assuming a hydrated sphere. The dependence

of the rotational correlation time θ is also shown in blue (right y-axis). The experimental range for RT-T/P experiments (i.e., Figure 22) is highlighted in pink. Lower right: A schematic of fast (due to fluorescent dye) and slow (due to the protein-DNA complex) tumbling with corresponding θ_{fast} and θ_{slow} are shown. The equation for anisotropy (r) including the linear combination from fast and slow tumbling is shown on the right.

We next considered a model where RT was able to interact with the fluorescein as it shuttled on the T/P substrate (Figure 24, left). In this model, changes in a theoretical standard deviation σ^2 of the distribution of RT on the T/P, starting from the polymerase-competent position, are related to changes in the observed anisotropy due to interactions of RT with the fluorescein, which is expected to confine the dye and increase anisotropy (Figure 24, right). Using this model, we were able to derive the expected distributions of RT on the T/P from the absolute anisotropy value of each case in the saturated condition. The calculated σ^2 for the cases of the indicated complexes (Figure 24, right, dotted lines) were then used to render the normal distributions shown in the left panel of Figure 24. Compared to the RT-T/P-dNTP complex, which has a tight distribution around the polymerase mode as expected, addition of EFV to RT resulted in a broad distribution on the T/P substrate, consistent with our PIFE results. Therefore, we considered absolute anisotropy as a proxy measurement for RT mobility on the T/P.

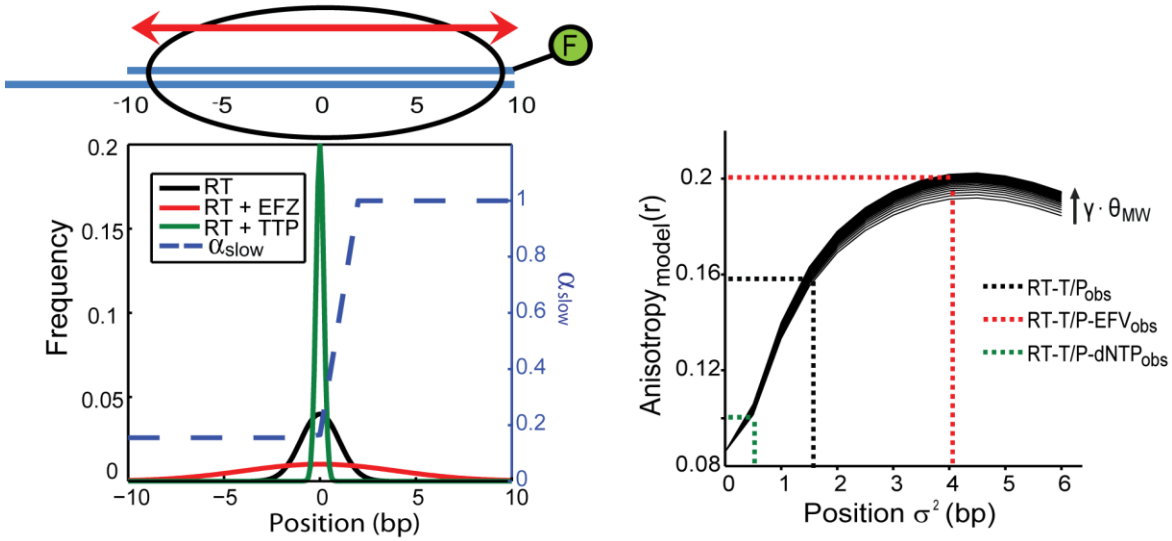

Figure 24. Simplified model of RT shuttling reveals origin of anisotropy changes

Left: A model of RT (oval) shuttling on its T/P substrate (blue) is presented, including putative interaction with the terminal fluorescein (labeled "F" in schematic), where the center of RT is able to move up to 10bp left or right from its origin at the polymerase-competent configuration (red arrow indicates shuttling motion). The chart below the model shows the relationship between RT position and α_{slow} (blue dotted line, right scale bar; value explained in text) and the position of RT, also corresponding to the model above it. The histograms represent the positional occupancy (i.e., mobility) of RT and were calculated from **d)**, which shows the relationship between **σ,** the standard deviation of the position of RT on its T/P substrate, and the resulting anisotropy that would be observed according to the model in Figure 23. As indicated by dotted lines, σ values were interpolated from the experimental anisotropy values taken from the value at 66 nM RT (i.e., saturated binding) and the indicated conditions. α_{slow} is a value which roughly indicates the odds of RT interacting with fluorescein, causing anisotropy resulting from slow tumbling to dominate, thereby raising absolute anisotropy (see Methods).

2.3.4 K103N RT is structurally resistant to efavirenz irrespective of binding affinity

We next probed the effect of the concentration of EFV and Neviripine (NVP) on the mobility of WT and K103N RT and found a dose-dependent increase in absolute anisotropy due to presence of both NNRTIs (Figures 25 and 26, left). WT RT plateaued at a higher anisotropy value in response to EFV than did K103N RT (Figure 25, left).

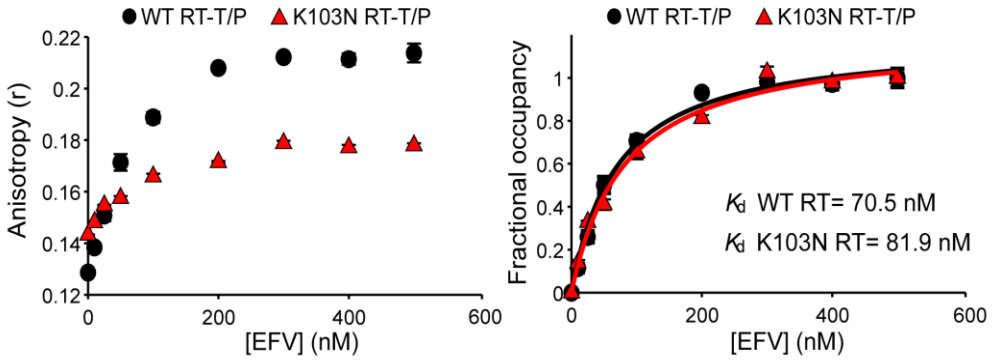

Figure 25. Binding of efavirenz to WT RT and K103N RT

Left: Absolute anisotropy of the WT RT-T/P (black circles) or K103N RT-T/P complex as a function of efavirenz concentration. Right: The curves on the left are normalized and fit to a single-site binding hyperbola, and K_d values from the fits to WT RT-T/P (black) or K103N RT-T/P (red) are shown. Points represent mean +/- s.e.m. values from n=3 separate experiments.

Although K103N RT exhibited similar changes in mobility in the presence of both drugs, the mobility of WT RT was higher in the presence of EFV than with NVP (Figure 26, left). When normalized to fractional occupancy (Figure 25 and 26, right), we observed curves which fit well with single-site binding hyperbola (R^2 > 95% for EFV and NVP), allowing us to extract K_d values of NNRTI binding.

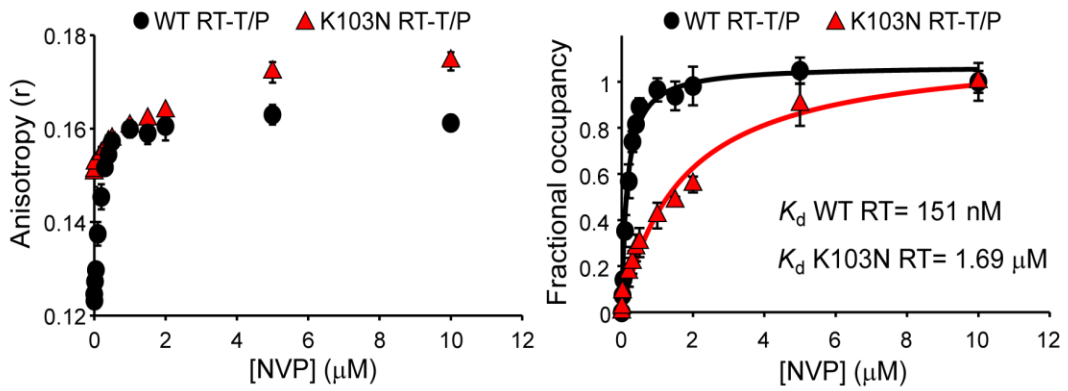

Figure 26. Binding of neviripine to WT RT and K103N RT

Left: Absolute anisotropy of the WT RT-T/P (black circles) or K103N RT-T/P complex as a function of neviripine concentration. Right: The curves on the left are normalized and fit to a single-site binding hyperbola, and K_d values from the fits to WT RT-T/P (black) or K103N RT-T/P (red) are shown. Points represent mean +/- s.e.m. values from n=3 separate experiments.

In agreement with previous findings(Ren et al., 2000), K103N resulted in an ~11-fold reduction in binding affinity of NVP as compared to WT RT (Figure 26, right), supporting our conclusion that anisotropy can be used to accurately measure NNRTI binding. Remarkably, the K103N mutation did not affect EFV binding at all (Figure 25, right).

As expected, when we titrated dNTP into the WT RT-T/P complex, we saw an opposite effect from EFV on RT mobility (Figure 27, left, open circles), which also fit well to a single-site binding hyperbola. This trend persisted in the presence of saturating concentrations of EFV (Figure 27, left, closed circles), leading us to conclude that RT-T/P can incorporate EFV and dNTP simultaneously, and that these independently alter RT mobility. In agreement with our PIFE data, K103N RT similarly responded to dNTP in the presence or absence of EFV (Figure 27, left, open and closed triangles, respectively), but the magnitude of change in anisotropy arising from saturating EFV in the presence of dNTP was significantly lower than that of WT RT

(compare red fits to black fits). Furthermore, the presence of EFV lowered the affinity of dNTP for WT RT-T/P by ~14-fold, whereas EFV affected the dNTP binding affinity to the K103N RT-T/P complex by a modest ~3-fold. Taken together, our results indicate that the K103N mutation allows RT to subvert the changes in RT-T/P dynamics seen with WT RT in response to EFV irrespective of EFV binding affinity.

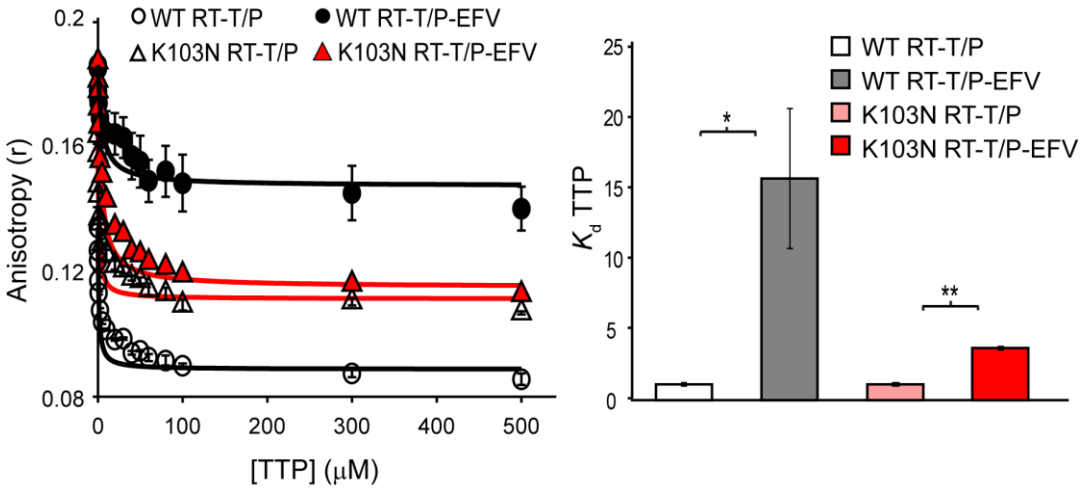

Figure 27. Binding of dNTP to WT or K103N in the presence or absence of EFV

Left: dNTP titrations are shown for WT (circles) or K103N (triangles) in the absence (no fill) or presence (fill) of EFV. Hyperbolic fits consistent with a single-site binding model are shown (black: WT RT, red: K103N RT). Right: K_d of TTP for the indicated RT-T/P complex in the presence or absence of EFV. Data represents mean +/- s.e.m. from n=3 experiments (left) and hyperbolic fits (right) each.

2.3.5 K103N RT does not succumb to arthritis

To understand the source of resistance to changes in shuttling dynamics conferred by K103N, we probed the conformational change of WT and K103N RT in response to EFV by monitoring the distance spanning the thumb and fingers subdomains of RT with single-molecule Förster

Resonance Energy Transfer (FRET) experiments on RT labeled with the Cy3-Cy5 FRET pair on the thumb (D250C-Cy3/5) and fingers (T139C-Cy3/5) domains (Figure 28a; see Methods) after allowing it to bind unlabeled, surface tethered T/P substrates.

WT RT exhibited a FRET E_{app} value of 0.74 ± 0.14 (mean ± std) and did not change appreciably in the presence of dNTP, with E_{app} = 0.73 ± 0.13 (Figure 28b). However, when EFV was added to RT, the thumb and fingers domains became noticeably further apart (E_{app} = 0.57 ± 0.15), even in the presence of dNTP, albeit slightly closer (E_{app} = 0.64 ± 0.15). The marked EFV-dependent changes in E_{app} distributions are entirely consistent with previous observations of "molecular arthritis" (see Table 1 for a summary of results and comparison expected values from crystal structures).

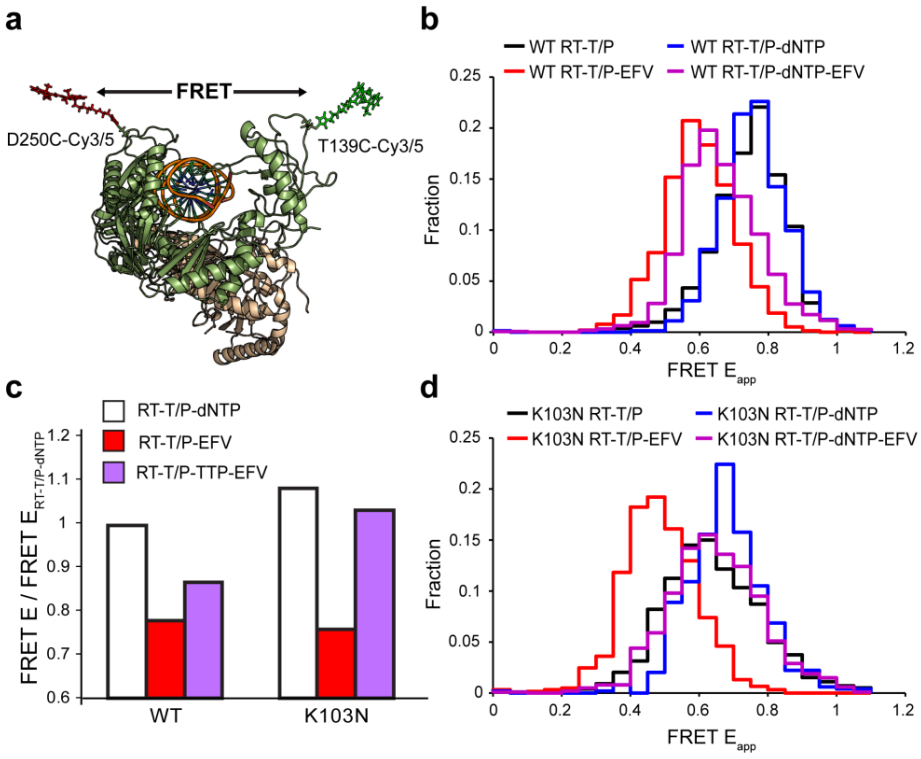

Figure 28. K103N does not succumb to arthritis.

a) Schematic of the D250C/T139 FRET construct. Shown are atomistic models of Cy3 (green) and Cy5 (red) built into the structure of the ternary complex (PDB ID: 1RTD). Note that labeling can be either way, i.e., Cy3 could be

labeled on D250C and Cy5 on T139C, and the FRET signal in our experiments would stay the same. Single-molecule FRET histograms (see Methods) for b) WT RT and d) K103N are presented for the indicated complexes. d) Ratio of the average of the indicated FRET E_{app} signal over the average FRET E_{app} from the RT-T/P-dNTP (ternary) complex.

		FRET (E_{app} ± STD)	**D250C-T139C FRET-derived distance (Å ± STD)**	**D250-T139 Cα-Cα crystal structure distance (Å ± STD) (PDBID)**
WT RT	WT RT-T/P	0.74 ± 0.14	45.6 ± 3.1	51.6 (2HMI)
	WT RT-T/P-dNTP	0.73 ± 0.13	45.7 ± 2.7	50.6 (1RTD)
	WT RT-T/P-EFV	0.57 ± 0.15	51.5 ± 1.7	56.2 (3V81)
	WT RT-T/P-dNTP-EFV	0.64 ± 0.15	49.2 ± 2.3	N/A
K103N RT	K103N RT-T/P	0.60 ± 0.19	50.4 ± 2.4	N/A
	K103N RT-T/P-dNTP	0.65 ± 0.14	48.8 ± 2.1	N/A
	K103N RT-T/P-EFV	0.46 ± 0.15	55.6 ± 1.2	N/A
	K103N RT-T/P-dNTP-EFV	0.62 ± 0.18	49.8 ± 2.5	N/A
E138D RT	E138D RT-T/P	0.69 ± 0.14	47.2 ± 2.6	N/A
	E138D RT-T/P-dNTP	0.66 ± 0.19	48.2 ± 3.1	N/A
	E138D RT-T/P-EFV	0.46 ± 0.12	55.6 ± 1.0	N/A
	E138D RT-T/P-dNTP-EFV	0.52 ± 0.13	53.3 ± 1.3	N/A

Table 1. FRET-derived D250C-T139C distances and comparison to crystal structures.

For all complexes indicated, the calculated E_{app} values are shown, derived from the mean of fitted Gaussian functions to the observed E_{app} data. Projected D250C-T139 distances are shown. When available, D250-T139 alpha carbon distances from crystallographic data are shown, with their respective PDBID indicated.

Although FRET was slightly lower in K103N RT ($E_{app} = 0.60 \pm 0.19$) than in WT RT, the thumb-fingers distance in K103N RT also remained largely unchanged by presence of dNTP ($E_{app} = 0.65 \pm 0.14$) and showed a similar response to EFV as WT RT by shifting to a lower FRET ($E_{app} = 0.46 \pm 0.15$), an E_{app} change of 0.16 change for K103N RT vs. 0.15 for WT RT (Figure 28d). In contrast to WT RT, however, K103N RT remained indistinguishable from both unliganded and dNTP-bound K103N RT when simultaneously incubated with dNTP and EFV ($E_{app} = 0.62 \pm 0.18$). Thus, although the K103N RT-T/P-EFV complex experiences molecular arthritis in the presence of EFV, it maintains its grip on the T/P substrate in the presence of both dNTP and EFV and is likely polymerase competent, escaping the arthritis that WT RT-T/P-dNTP-EFV experiences (Figure 28c; compare purple bars). Since [TTP] in these experiments is well above the K_d calculated for both WT- and K103N-RT-T/P complexes in the presence or absence of EFV (Figure 27, right; see Methods), we conclude that cognate nucleotide can bind WT-RT-T/P complexes in the presence of EFV, yet, due to molecular arthritis, cannot be clamped by the fingers domain into a conformation compatible with primer extension. On the other hand, K103N RT is able to overcome molecular arthritis and remain firmly clamped onto the T/P substrate in an extension-competent configuration.

2.3.6 HIV-1 RT is a tunable molecular clutch

With results indicating that the K103N mutation can alter RT-T/P dynamics and intramolecular conformation irrespective of EFV affinity, we compared the structures of WT and K103N RT bound to EFV and noted one striking difference in K103N RT: the disruption of a salt bridge between E138 on p51 and K101 on p66 likely removes the positive charge in K103 responsible for repelling K101 (Figure 29).

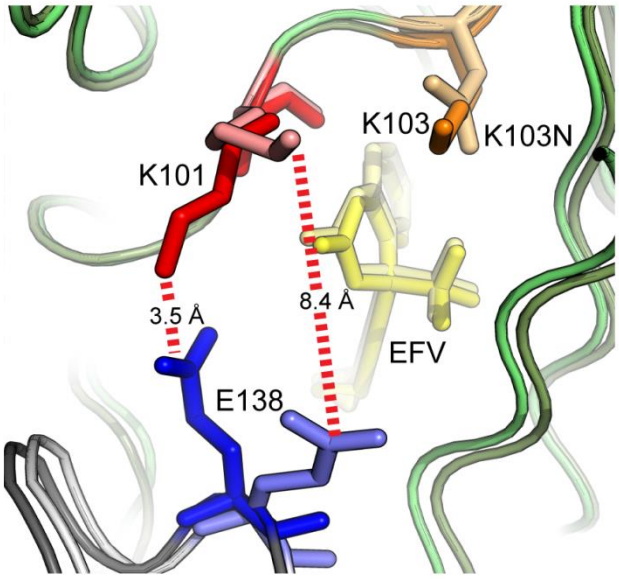

Figure 29. K103N breaks the E138-K101 salt bridge

The NNRTIBP of WT (darker) or K103N (lighter) is shown in ribbon representation. p66 is colored in green, and p51 in grey. The distances between the Oε atom of E138 and the Nζ atom of K101are represented amidst red dotted lines.

In published structures of WT RT (Kohlstaedt et al., 1992; Lansdon et al., 2010; Ren et al., 2000), the E138-K101 salt bridge (defined as a separation of < 4 Å between the Oε atom of E138 and the Nζ atom of K101; Donald et al., 2011) only exists in the presence of NNRTIs such as efavirenz, nevirapine, or rilpivirine. Since E138-K101 resides at the hinge site between p51 and p66, often implicated in the mechanism of RT inhibition by NNRTIs, we next engineered E138D-K101R into RT to form a stronger salt bridge by increasing its polarity, and examined its effects on EFV activity. We first noted that the binding affinity of EFV for E138D-K101R was increased by ~31-fold (Figure 30, left). Furthermore, the anisotropy-derived mobility of E138D-K101R RT on the T/P substrate was constitutively higher than WT-RT and increased to a much higher degree in response to EFV than did WT RT (Figure 30, right).

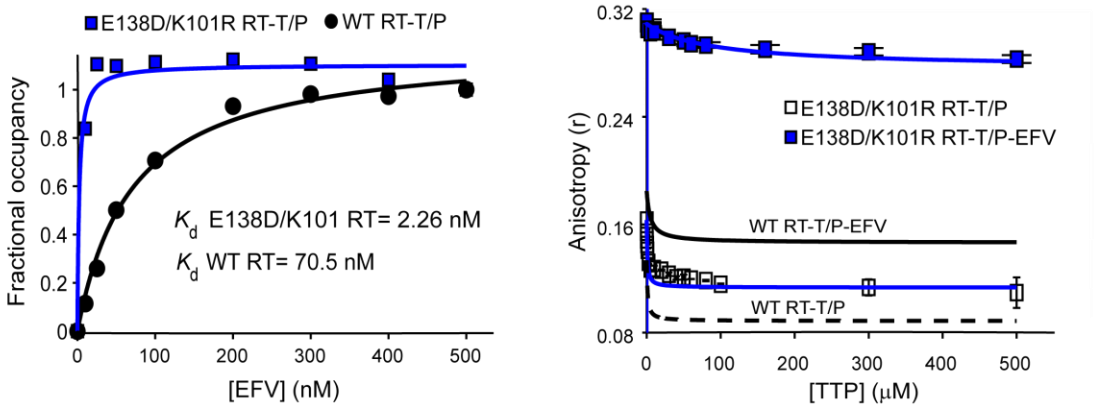

Figure 30. E138D/K101R binds and reacts strongly to EFV

Left: Normalized binding curves of EFV to E138D/K101R are shown, with WT in black circles, and E138D/K101R in blue squares. The data was normalized and fit similar to the binding curves presented in Figures 25 and 26. The indicated K_d values are derived from fits to the hyperbolic, single site binding curves. Right: dNTP titration with E138D/K101R in the presence (hollow squares) or absence (solid blue squares) of EFV are shown. Fits to single site binding models are shown as blue lines. Also shown are the fits from the analogous data from WT RT, i.e., from Figure 27, where the black dotted line is from WT RT-T/P and the solid black line is from WT RT-T/P-EFV.

We next tested the E138D, K101R, and E138D-K101R mutations in two separate *in vitro* assays (see Methods) and found a respective 2.7 ± 0.2-, 1.3 ± 0.1-, and 2.7 ± 0.2-fold increase in susceptibility to EFV as measured by IC50, indicating that the changes observed in the E138D-K101R salt bridge mutant largely arose from the presence of E138D. We furthermore observed a 7.8 ± 4.6-fold increase in susceptibility of E138D-K101R RT (the other mutant viruses had no detectable activity) as measured by EC50 (see Table 2).

Genotype of RT or virus	IC_{50} (nM) for inhibition of HIV-1 RT activity (Fold-resistance)				EC_{50} (nM) for inhibition of HIV-1 replication (Fold-resistance)			
	EFV	NVP	RIL	ETV	EFV	NVP	RIL	ETV
WT	47.6±1.1	2323.7±179.3	22.5±2.2	24.5±2.2	3.1±1.8	245±59.7	1.6±0.5	4.2±1.4
E138D	17.6±2.1 (0.37)	676.3±68.5 (0.29)	8.3±1.2 (0.37)	12.1±1.6 (0.49)				
K101R	37.4±1.8 (0.78)	2936.0±599.3 (1.26)	15.5±0.8 (0.69)	27.3±7.5 (1.11)				
K101R/E138D	17.5±2.4 (0.37)	1102.7±290.7 (0.47)	8.4±1.0 (0.37)	11.2±0.8 (0.46)	0.4±0.2 (0.13)	216±16.8 (0.88)	0.3±0.1 (0.19)	0.7±0.2 0.17)

Table 2. IC50 and EC50 data for E138D, K101R, and K101R/E138D

To quantify the effects of a stronger salt bridge on molecular arthritis, we next introduced the E138D mutation into our D250C/T139C thumb/fingers FRET construct. Figure 6d illustrates that the distance between the thumb and fingers of the E138D-T/P complex (E_{app} = 0.69 ± 0.14) was only slightly larger than that of WT, but responded in much greater magnitude to the presence of EFV than either WT or K103N RT did (E_{app} = 0.46 ± 0.12, compare the FRET E_{app} change of 0.24 for E138D vs. 0.15 and 0.16 for WT and K103N RT).

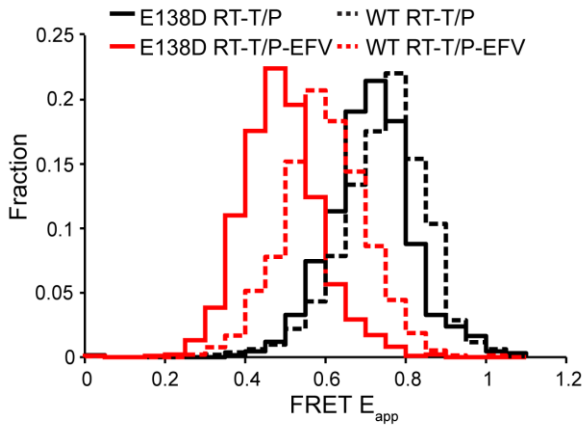

Figure 31. E138D RT is more arthritic than RT in the presence or absence of efavirenz

FRET histograms for E138D RT-T/P (red) or WT-RT (black) both before (dotted) and after (solid) EFV.

2.4 CONCLUSIONS

We investigated mechanism(s) of inhibition of RT by Nonnucleoside RT Inhibitors (NNRTIs) and found that efavirenz (EFV), an NNRTI, did not significantly alter binding affinity of RT for its Template/Primer (T/P) substrate but substantially increased RT mobility on the T/P, reducing the time spent in the polymerase-competent position. In contrast, introduction of cognate dNTP resulted in an increased dwell time in the polymerase competent state in a manner independent of EFV, indicating distinct mechanisms of RT mobility modulation. Using a novel and extensible anisotropy-based assay of RT mobility, we also found that K103N, a drug resistance mutation in RT, reduced neviripine (NVP) affinity by ~10 fold but did not significantly affect EFV binding affinity. Despite unaltered affinity of EFV, addition of EFV to K103N RT did not result in the increased shuttling we typically observed for WT RT, providing evidence that K103N confers structural resistance to EFV distinct from potential changes in the NNRTI binding pocket. Furthermore, we showed that EFV addition results in a larger distance between the thumb and fingers domain of WT RT but not of K103N RT. After structural analysis revealed that NNRTIs generally result in the formation of a salt bridge between K101 and E138 of their respective p66 and p51 RT subunits and that this salt bridge is significantly disrupted in K103N RT, we engineered an electrostatically stronger salt bridge, E138D-K101R into RT and showed that this mutation resulted in an increase in NNRTI susceptibility accompanied by a large increase in T/P mobility and a persistently larger distance between the thumb and fingers domain than that of WT RT.

Taken together, the results indicate that RT is akin to a tunable “molecular clutch” under partial control of the E138-K101 salt bridge, which modifies RT conformation through a crucial hinge site between p51 and p66. EFV-induced molecular arthritis, greatly strengthened by the

salt bridge, is responsible for diminishing RT-T/P contacts, reducing the grip of RT on its T/P substrate and its ability to clamp down on the dNTP in the polymerase competent mode, permitting highly dynamic T/P translocation that is disruptive to DNA polymerization. The K103N mutation alleviates molecular arthritis by breaking this crucial bridge and restoring the ability of RT to grip the T/P substrate in the presence of dNTP and EFV, conferring a structural workaround to the activity-reducing effects of EFV.

3.0 CHARACTERIZING THE INTRA-MOLECULAR DYNAMICS OF HIV-1 RT ON THE T/P SUBSTRATE IN THE ABSENCE AND PRESENCE OF NNRTI USING SPFRET AND MOLECULAR MODELLING

3.1 INTRODUCTION

HIV-1 Reverse Transcriptase (RT) is an attractive drug target as it is critical to the life cycle of HIV-1. Nevertheless, there are major limits to our understanding about this important enzyme which impede the development of new and effective RT Inhibitors (RTIs). Specifically, detailed information about drug- and/or substrate-induced changes in its structural dynamics and its substrate binding kinetics is lacking. In this chapter, we planned to address these deficits by developing methods that combine single-molecule fluorescence spectroscopy with molecular modeling. Utilizing several inter-domain FRET pairs in RT, we propose to model equilibrium structures of RT in the presence of various substrate/drug combinations using spFRET. In addition, we will model the structural intermediates of RT during drug-induced conformational change, using a method under development called Structural Information from FRET (SIFF) that employs Bayesian methods to infer the most probable instantaneous structure in a trajectory from spFRET traces and structural sampling. To obtain absolute, accurate FRET distances, all spFRET measurements will be complemented with data from atomistic MD simulations of our dye-labeled system of interest.

In the last 25 years there has been significant progress in understanding HIV-1 RT structure-function relationships. To date, more than 100 different crystal structures of wild-type (WT) and mutant HIV-1 RTs have been solved in their unliganded forms or in complex with substrates [e.g., template/primer (T/P) and dNTP] or RTIs. These structures have highlighted the large intra-subunit, inter-domain conformational changes that can occur in HIV-1 RT upon substrate or drug binding (Ivetac and McCammon, 2009; Seckler et al., 2011; Sluis-Cremer et al., 2004; Temiz and Bahar, 2002). Biochemical studies have shed insight into the mechanisms associated with DNA polymerization, RNase H cleavage, RTI inhibition and drug resistance. Furthermore, recent single pair Förster Resonance Energy Transfer (spFRET) studies have revealed how nucleic acid sequence and RTI binding can influence the RT-T/P interaction (Abbondanzieri et al., 2008; Liu et al., 2008). For example, HIV-1 RT can flip its orientation 180° on a polypurine tract (PPT) T/P without dissociation upon NNRTI binding (Liu et al., 2008).

Despite this wealth of available structural, biochemical and biophysical data, there remain significant knowledge gaps that have not been addressed. These include:

i. The structure and dynamics of RT in complex with substrates and RTIs; and

ii. The structural trajectories and transitions that occur in RT upon substrate or RTI binding; and

iii. The change in RT-T/P binding kinetics that occurs upon substrate or RTI binding

Such information would have an immediate and profound impact on all levels of the biology of HIV-1 reverse transcription including structure-activity relationships, inhibition, drug resistance and drug discovery.

To date, our ability to follow binding kinetics and to measure the intra-subunit conformational changes that occur in HIV-1 RT upon substrate and/or RTI binding has been

limited by technology. In this regard, we aim to develop and apply single molecule fluorescence methods coupled with computational modeling to study the structural transitions in HIV-1 RT *in vitro* under physiologically relevant conditions. To this end, we proposed the following approaches:

Approach 1: To develop an accurate method to generate structures of on-pathway transitional intermediates. We will observe several RT constructs bearing inter-domain FRET pairs with spFRET using a Scanning Confocal Fluorescence Microscope (Zheng et al., 2007) and/or a Total Internal Reflection Fluorescence Microscope (Graham et al., 2011; Roy et al., 2008) equipped with a computer automated stopped-flow setup. Equilibrium (i.e., "static") atomistic RT structures will be modeled with multiple experimental distance constraints. To model accurate, absolute distances (Allen and Paci, 2009; Corry and Jayatilaka, 2008; Iqbal et al., 2008a; Schroder, 2005; VanBeek et al., 2007), measurements will be complemented with data from atomistic simulations of dye-bound proteins, including orientation factor and linker flexibility. To model on-pathway structural intermediates between free drug-bound RT, drug induced conformational changes will be monitored by spFRET in real time during drug infusion. These inter-domain distance trajectories will in turn serve as inputs to a computational method called Structural Information from FRET (SIFF), which uses Bayesian particle filtering (Blake and Isard, 1998) methods to accurately recover and rank the likeliest atomistic structural trajectory responsible for the reported spFRET trace. Comparison of existing structural data with generated trajectories from subsequent SIFF calculations will provide a clear proof-of-concept for our method's ability to predict the proper conformation, cross-validating our experimental

(spFRET) and computational (SIFF) work and confirming its suitability to characterize less well-studied systems.

Approach 2: Characterize substrate- and RTI-induced conformational changes in RT. The primary goals of this Aim are to discover novel structures of RT complexed with various combinations of nucleic acid substrates (e.g., template/primer and/or NTPs) and/or drugs (NNRTIs, ncRTIs and RNHIs). Steady state traces of RT will be recorded, and equilibrium structures for each state will be modeled with FRET-derived distance constraints. To obtain structures of on-pathway intermediates during drug-induced conformational change, RT will be infused with ligand during real-time observation, and the resultant spFRET trajectories will be analyzed with techniques developed in Approach 1.

3.1.1 Statement of Problem

Although HIV-1 RT is one of the best studied viral polymerases, several knowledge gaps limit our ability to understand the structure and dynamics of enzyme function as well as the mechanisms of RT inhibitor (RTI) inhibition and drug resistance, specifically:

i) While there are several crystal structures of HIV-1 RT in complex with substrates [e.g., T/P or T/P and dNTP] or RTIs [e.g., NNRTIs or RNase H inhibitors (RNHI)], there is no structural information for RT in complex with both substrates and RTIs. In this regard, the structural mechanisms associated with NNRTI inhibition of HIV-1 RT function have largely been deduced from RT-NNRTI binary complexes. Given that several kinetic

analyses have demonstrated that many NNRTIs preferentially bind to RT-T/P or RT-T/P-dNTP complexes, the available structural information may not be entirely relevant. It is crucial, then, that we find an alternate means to obtain structures of these target complexes.

ii) Crystal structures of HIV-1 RT in complex with nonnucleoside RTIs (NNRTIs) reveal large conformational changes in the 66 kDa (p66) subunit of HIV-1 RT, but the nature and dynamics of the structural transitions and trajectories involved in these conformational changes are unknown. While X-ray crystallography and NMR methods provide a wealth of structural information on low energy substates, they are insufficient for predicting *transitional structures* between two or more substates: interpolation between these structures is unsatisfactory since there is an astronomically large number of structural intermediates possible between different substates. Rational design of high-affinity drugs may greatly benefit from reliable structural information from transitional intermediates; therefore, development of a means to identify these transitional intermediates is of critical importance.

iii) It is unclear to what extent RTIs affect the on/off binding kinetics of the RT-T/P complex. Although bulk methods exist to measure kinetics of DNA-protein complexes, they can lead to artifacts from ensemble averaging and involve extensive instrumentation (e.g., stopped flow). Thus, our goal is to develop a means to directly measure the kinetics of DNA-protein reactions at the single-molecule level.

We aimed to develop robust experimental and computational methods that will allow us to (i) model equilibrium structures of various RT complexes of interest and (ii) predict, with

unprecedented accuracy, structures of transitional intermediates during drug- and/or substrate-induced conformational change in RT

In this chapter, we propose coupling single-molecule fluorescence methods with computational modeling to study the structural dynamics and binding kinetics of RT. spFRET is a powerful method, yielding high information content about the kinetics of individual molecules and relative changes between two or more states in the millisecond to minute timescale (Hohlbein et al., 2010; Joo et al., 2008; Lee et al., 2005; Lu, 2005; Michalet et al., 2006; Roy et al., 2008; Wang and Geva, 2005). However, individual spFRET measurements significantly lack detailed structural information. Conversely, *in silico* atomistic simulations are atomically accurate but suffer from temporal sampling deficits. We therefore propose to integrate these two approaches to yield detailed, time-resolved atomistic information. In trying to push the limits of spFRET, we are attempting to optimize the accuracy of our experiments by optimizing s/n with rapidly evolving advances in the field (Chung et al., 2010; Iqbal et al., 2008b; Joo et al., 2008; Liu et al., 2010; Lu, 2005; Santoso et al., 2008; Wang and Geva, 2005; Wang and Lu, 2008; Watkins and Yang, 2004; Wozniak et al., 2008) and by complementing our results with molecular simulation of the system under study. Furthermore, we are using the recently described phenomenon of fluorescence intensity enhancement (Hwang et al., 2011) to develop a powerful technique to record binding events and to accurately measure the on/off rates of molecular complexes, particularly important quantities as they provide information about the energy barrier of the reaction. Notably, the methods developed in the proposed work will directly benefit the RT field and will be broadly applicable to other protein-ligand systems.

3.1.2 Structural Information from FRET (SIFF)

A computational tool under development in the lab of Chakra Chennubhotla (University of Pittsburgh), Structural Information from FRET (SIFF) recovers the instantaneous three-dimensional structure of a natively fluctuating protein from spFRET data. Experimental spFRET methods capture molecular kinetics and relative distance changes between multiple conformational states with millisecond resolution. However, as one-dimensional dye-separation traces, spFRET measurements cannot resolve three-dimensional structural details; countless protein geometries could satisfy any single spFRET-derived distance. Molecular dynamics (MD) simulations make the opposite compromise: structural changes are sampled at atomistic resolution but with necessary temporal deficits; simulation timescales are orders of magnitude smaller than those accessible to spFRET. SIFF eliminates these tradeoffs by iteratively updating a structural ensemble at each spFRET-time point consistent with the reported spFRET distance and based on prior knowledge from the overall structural trajectory. The method is based on Bayesian filtering and predicts the time-evolution of a protein system that is constrained by incomplete (i.e., a single distance constraint) and imperfect (i.e., with quantifiable noise) experimental data; resulting (sequential) on-pathway intermediates agree with the inputted spFRET trace. Reconstructed dynamics are informed by atomistic and coarse-grained simulations, which quantify the three-dimensional conformational responses to reported spFRET distances. SIFF thus simultaneously removes the sampling deficits of molecular simulations and the structural deficits of spFRET experiments. Particularly suited to hinge proteins with large distance changes between domains, the method has already been validated for spFRET traces of *apo* Adenylate Kinase from Haw Yang's lab.

3.2 PROPOSED STUDIES

3.2.1 Development of a method to structurally characterize on-pathway transitional intermediates

We plan to assess our technique using efavirenz and neviripine, FDA-approved NNRTIs, since the binding and kinetics of these drugs are well characterized. Steady state spFRET experiments will be performed on free and NNRTI-bound RT bearing each FRET pair, and the data will be corrected for spectral crosstalk and variations in fluorophore quantum yield on a trace-by-trace basis to ensure accurate and precise FRET E values (McCann et al., 2010). For each dye pair, distance distributions will be calculated from FRET E histograms and R_o as determined above. We will compare these inter-fluorophore distances to estimates from static X-ray structures (ΔE, Table 3), and from simulation (i.e., the distance between the dye-cloud centroids). We will evaluate cross-correlation matrices from both atomistic and coarse grained simulations as inputs into SIFF, and we will assess several ways of incorporating dye structure into the model. Next, we will process individual spFRET traces with the optimized SIFF algorithm, quantitatively comparing the resultant structural trajectories with crystallographic structures and simulation data. This comparison will effectively cross-validate the accuracy of spFRET traces and of SIFF-derived structures and will allow us to make adjustments to either complimentary technique if necessary.

3.2.2 Discovering novel structures of RT complexes

Despite the large number of RT structures deposited in the PDB, there is a conspicuous lack of structures in complex with more than one ligand. To address this deficit, we plan to solve the structures of RT while in complex with various combinations of template/primer (T/P) substrates, NTPs, and drugs in the NNRTI, ncRTI, and RTI classes. We will employ a T/P substrate commonly used in our lab: a 58mer template and a 26mer primer. In some instances, the primer in the T/P substrate will be 3'-end terminated with a dideoxythiamine which lacks the 3'OH necessary for elongation, denoted as T/P_{ddT}. In other cases, the primer in the T/P substrate will be a synthetic polypurine tract (PPT), denoted as T/P_{PPT}. We will perform steady state spFRET experiments with TIRFM to record the interdomain distances of every binary, ternary, and quaternary complex outlined in Table 3. Similar to the use of Nuclear Overhauser Effect (NOE) distance constraints in NMR structure determination, we will use multiple distance constraints derived from designed FRET pairs to model equilibrium RT structures. We will incorporate absolute distance distributions from each FRET pair into our model which has been optimized to best describe the dye centroids. Harmonic constraints will applied to the centroids of the dyes modeled on RT, accounting for the positional error of FRET due to protein dynamics, and structures will be globally optimized by simulated annealing and/or replica exchange molecular dynamics procedures. The resultant structures will be exclusive, atomistic models of all of the various complexes outlined in Table 3, providing deep structural insight into the interplay between substrate and drug binding in RT. Importantly, the success of the D250C/T139C FRET scheme in Chapter 2 provides proof-of-principle that these experiments will work.

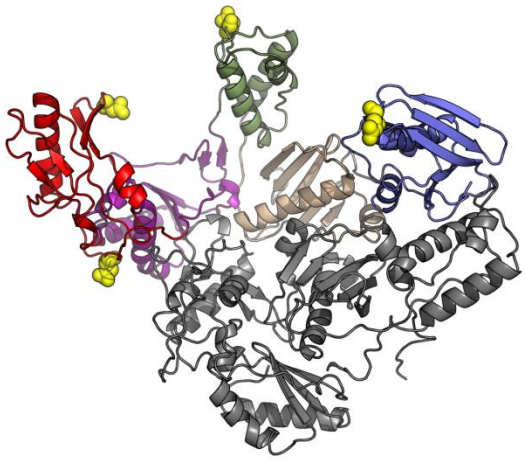

Figure 32. Proposed intra-RT FRET pairs.

The proposed cysteines to be dye labeled are indicated in yellow spheres, where any pair within RT is a potential FRET pair.

p66 Mutant	State	r_{obs} (Å)	predicted FRET E	predicted ΔE
D67C/D250C	free bound	19.5 34.9	1.00 0.94	0.05
E6C/D67C	free bound	34.1 43.2	0.95 0.83	0.13
E6C/E449C	free bound	69.8 72.3	0.21 0.18	0.03
E6C/D250C	free bound	47.0 63.5	0.74 0.32	0.42
D67C/ E449C	free bound	61.2 67.8	0.37 0.24	0.13
D250C/E449C	free bound	48.6 51.8	0.70 0.61	0.09
E6C/L560C	free bound	N/A	N/A	N/A
D67C/L560C	free bound	N/A	N/A	N/A
D250C/L560C	free bound	N/A	N/A	N/A
E449C/L560C	free bound	N/A	N/A	N/A

Table 3. Proposed distances between FRET pairs

Distances between alpha carbons of the indicated FRET pairs in the NVP-free or –bound state. Predicted FRET E is shown, and the predicted change in E upon complexing with NVP, ΔE, is shown.

3.2.3 Identification of transitional structures in RT during drug-induced conformational change

We will monitor spFRET from various RT FRET pairs as several different drugs are infused into the system by stopped-flow. We will start by testing Efavirenz and will move to novel compounds (ncRTIs, RHIs) discovered in our lab and in others in order to elucidate the mechanism of RT inhibition behind each drug. We will first perform experiments on FRET-labeled T139C/D250C to see if we can monitor changes in molecular arthritis in real time. We will next attempt to see conformational change within the other FRET pairs detailed in Table 3. Preliminary calculation (Table 3) suggests that the FRET pair between the thumb and the palm (C6/C250) will yield the highest FRET change upon Efavirenz binding. Though the C67/C250 (Fingers-Thumb) and C6/C449 (Fingers-RNase H) dye pairs are predicted to yield little to no FRET E change upon NNRTI binding, measuring E from both FRET pairs is worthwhile for providing more distance constraints and as an internal control against false-positive observations of transitions: the more FRET pairs that are sampled for each drug, the more information we get regarding conformational change and the more confidence we can have in our SIFF output. We will first observe drug-induced transitions via TIRFM, since it can observe spFRET traces from hundreds of individual molecules at once, as opposed to our SCFM, which focuses on one molecule at a time. At this stage we may decide to increase the temporal resolution of our TIRFM by speeding up the frame rate, sacrificing S/N by using smaller integration times. We may also try pixel binning and/or reducing the effective area projected onto the CCD by acquiring sub-regions. In this way, we could drive our TIRF microscope to acquire up to 1 ms/frame (1 kHz), although we would only be watching 1-10 molecules (as opposed to

hundreds) at a time. In order to sample the transition at higher speeds, we will next observe drug-induced conformational change with our SCFM setup.

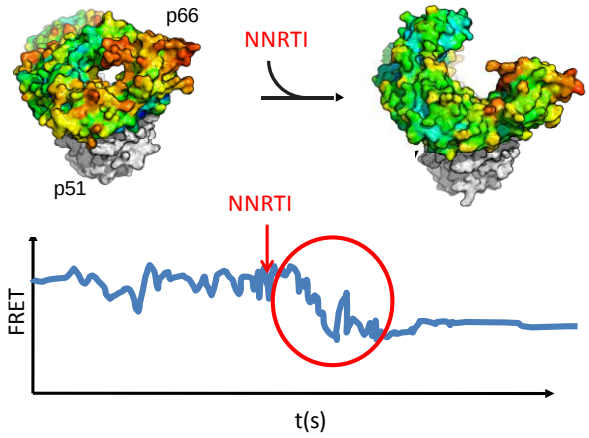

Figure 33. Theoretical transitional FRET trace.

A theoretical FRET trajectory before and after injection of NNRTI during observation. In this scenario, molecular arthritis causes the thumb and fingers distance to become greater, decreasing the FRET value, while reducing the variability in the FRET signal due to stiffening of the domains relative to each other.

Figure 30 illustrates an example of predicted experimental results. In the depiction of an ideal FRET trace from an experiment using the C6/C250 dye pair as an example, NNRTI is infused some seconds after acquisition has started, inducing an observable state-to-state transition (depicted as a circle), indicating large conformational rearrangements in RT. It has been reported that both Thumb and Finger domain mobilities are attenuated by NNRTI binding. The free and drug-bound states should hence be resolvable by quantifiable changes in FRET distribution width. We hope to capture dozens of time-resolved drug binding events for every FRET-pair construct and drug combination. We posit that sampling the unique structural trajectories of conformational change will significantly enhance rational drug design, so the ultimate goal will be capture as many structures of transitional intermediates as possible.

3.2.4 Investigation of the interplay between RT conformation and ligand affinity

We hypothesize that, as the overall structure of RT begins to resemble the drug-bound structure during drug-induced conformational change, the affinity of drug for its binding site will increase. Moreover, we anticipate that, during a drug binding event, the spFRET traces should abide by the conformational selection hypothesis when undergoing a drug induced transition. Specifically, drug-bound conformations should be "reachable" from the unbound state by virtue of RT's structurally encoded intrinsic dynamics. If our efforts to sample structural trajectories of conformational change are successful, we will have a unique and exciting opportunity to test these hypotheses.

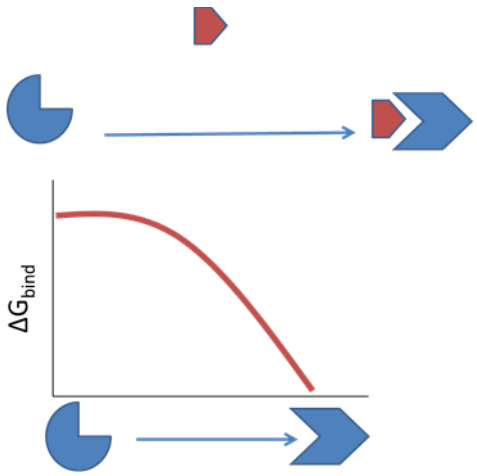

Figure 34. Theoretical effect of conformational change on ligand binding affinity

The theoretical response of the calculated ligand binding free energy ΔG_{bind} to docking to the structures along an on-pathway transition from *apo* to *holo* forms, where blue represents a protein whose binding site shape is becoming more "ligand-like" as it makes the transition to a binding-competent conformation.

Using the structural trajectories we obtain, we will computationally dock the appropriate ligand onto each on-pathway intermediate using Autodock, which outputs an estimate of ligand binding free energy, ΔG_{bind} for a particular ligand/receptor pair (see Figure 34). Also indicated in Figure 34, we will quantify the probability of a ligand-free conformation "reaching" a drug-bound conformation from our simulations. Independent of using transitional intermediates, we will perform similar experiments on the novel equilibrium structures from our modeling efforts. For instance, we can test the hypothesis that Efavirenz has a preference for binary and tertiary RT complexes while Nevirapine has a preference for free enzyme.

3.2.5 Determining the role of conformational dynamics on polymerization activity

The mechanism of polymerization by RT remains unclear. Steady-state and pre-steady-state kinetics have pointed to the possibility that the rate-limiting step of polymerization is either nucleotide incorporation or T/P substrate dissociation. Nucleotide incorporation is believed to take place in the polymerase active site as the p66 fingers domain clamps down on the incoming dNTP, precisely orienting the primer for elongation, followed by T/P translocation. In this paradigm, the rate limiting step of polymerization is the conformational dynamics of RT itself. Following RT:T/P complexes with spFRET, we will infuse dNTPs into the flow cell and observe conformational changes in the fingers domain relative to the thumb domain, attempting to relate the observed motions to the number of dNTP additions required for complete polymerization. If this preliminary experiment is successful, we can easily move to combining dye-labeled T/P substrates with singly-labeled RT p66 (D250C, for instance) in order to watch stepwise changes in spFRET trajectories as the substrate is translocated, with the purpose of decoupling the

kinetics of translocation with those of dNTP addition. If successful, the experiments will provide much needed insight into the structural dynamics of RT activity.

3.2.6 Modeling accurate, absolute FRET distances with atomistic RT simulations

Although it is reasonable to compare *relative* changes between two E measurements, inherent uncertainty in FRET E renders the method insufficient as an *absolute* quantitative ruler. Using the equation for FRET Efficiency (E):

$$E = 1/ (1+ (R_0/r)^6) \quad (3.1)$$

we can directly calculate distance changes between single D-A pairs, r. In the above equation, R_o is the Förster distance, or the distance at which E =0.5, and is defined as

$$R_o = 8.8\text{x}10^{-28} \kappa^2 \eta^{-4} Q_o J \quad (3.2)$$

where n is the refractive index, Q_0 is the fluorescence quantum yield of the donor without acceptor, and J is the integral of the spectral overlap between the donor emission and acceptor excitation spectra. In equation 2, κ^2 is the orientation factor of the first excited state dipole and is calculated as

$$\kappa^2 = (sin\theta_D \, sin\theta_A \, cos\Phi - 2 \, cos\theta_D \, cos\theta_A)^2 \quad (3.3)$$

where θ_D and θ_A are the angles between the transition dipole moments of donor and acceptor and the vector connecting the two dyes, and Φ is the angle between the planes of the fluorophores (see Figure 2).

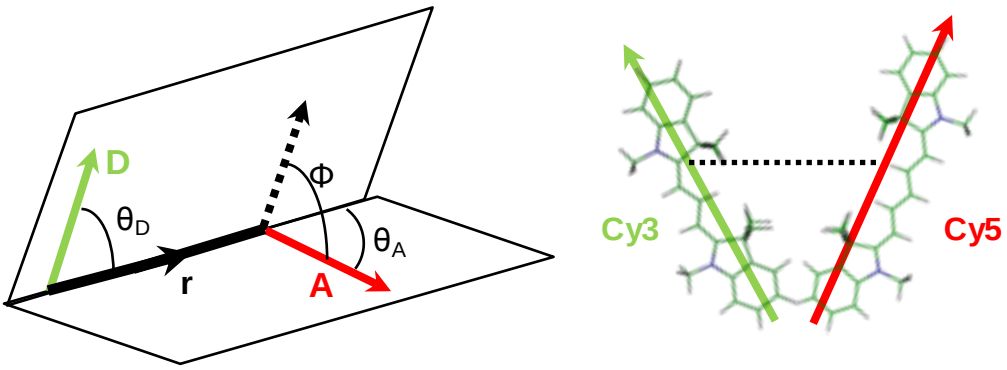

Figure 35. Orientation factor and transition dipoles of FRET

Left: vectors from the text are explained in schematic format. The dipole of Cy3/donor are indicated as green vector and the dipole of Cy5/acceptor as red. Right: The transition dipole moments of Cy3 and Cy5, projected onto the molecule.

The value of κ^2 can range from 0 to 4, depending on the relative dye orientations. Most researchers assume a κ^2 value of 2/3 for isotropically tumbling dyes despite the fact that a high correlation between κ^2 and R_o has been observed. In our system and in many others, the fluorophores are fixed to the system of interest with limited conformational freedom and are expected to display anisotropic behavior. Clearly, uncertainty in FRET-derived distances (beyond measurement error from shot noise and dye-linker flexibility) largely stems from using an arbitrary value for R_o. Indeed, we typically use a commonly referenced R_o of 5.8 nm to approximate distances, calculated for the Cy3-Cy5 D-A pair assuming $\kappa^2 = 2/3$ (Bastiaens and Jovin, 1996). Thus, in order to calculate accurate *absolute* distances, i.e., distances which are not *relative* to another measurement, it is necessary to have an accurate estimate of κ^2. It is straightforward to calculate the instantaneous κ^2 value from a snapshot of an MD simulation, so one can easily determine the mean κ^2 value over the course of the simulation. Currently,

however, there are no publically available force field parameters for Cy3 and Cy5, so it was necessary to parameterize the dyes ourselves (see topology and parameter files in Appendix).

In order to model the geometry of a fluorescing dye, one must model the excited state geometry of the molecule, which differs than that of the ground state. Starting with previously optimized Cy3 and Cy5 coordinates courtesy of David Norman's group (Iqbal et al., 2008b) (used to reduce computation time) and parameters courtesy of J. Spiriti (Spiriti et al., 2011), we parameterized Cy3 and Cy5 with the CHARMM General Force Field (CGENFF) (Vanommeslaeghe et al., 2010). Employing the newly parameterized Cy3 and Cy5 residues, we will perform explicit solvent Molecular Dynamics (MD) simulations of RT with either Cy3 or Cy5 residues on all of the four locations with known structure (residues 6, 67, 250, and 449). Using simple vectors between two atoms that best describe the excited state dipole, we can calculate mean κ^2 as in eq. 3 and, for each time point in a given FRET trajectory, derive an absolute distance r from eqs. 3.2 and 3.1.

3.3 RESULTS

Preliminary simulations focused on obtaining a quantitative measurement of the accuracy of FRET simulations. The first simulation we performed was on two tripeptides separated by 5 Å and fixed at the alpha carbon, in order to measure the distribution of FRET due to dye flexibility alone in the absence of protein domain motions (see Figure 36). As a proof-of-principle, this experiment shows that we are able to measure FRET *in silico* with reasonable accuracy. We were also able to measure κ^2, which averaged 0.87 over the course of the 100 ns simulation.

This value is close to the expected value of 0.66 for the case of an isotropically tumbling dye, indicating that the dye linkers give sufficient flexibility *per se*, and that any gain in anisotropy is likely due to interaction with the protein. Interaction of the dye with the protein is likely given (1) the results in Chapter 2 indicating that Cy3 nonspecifically associates with RT and (2) in our FRET experiments, the laser power must be cut very low (<2 mW) to avoid photobleaching compared to the power used for dye-labeled oligonucleotides.

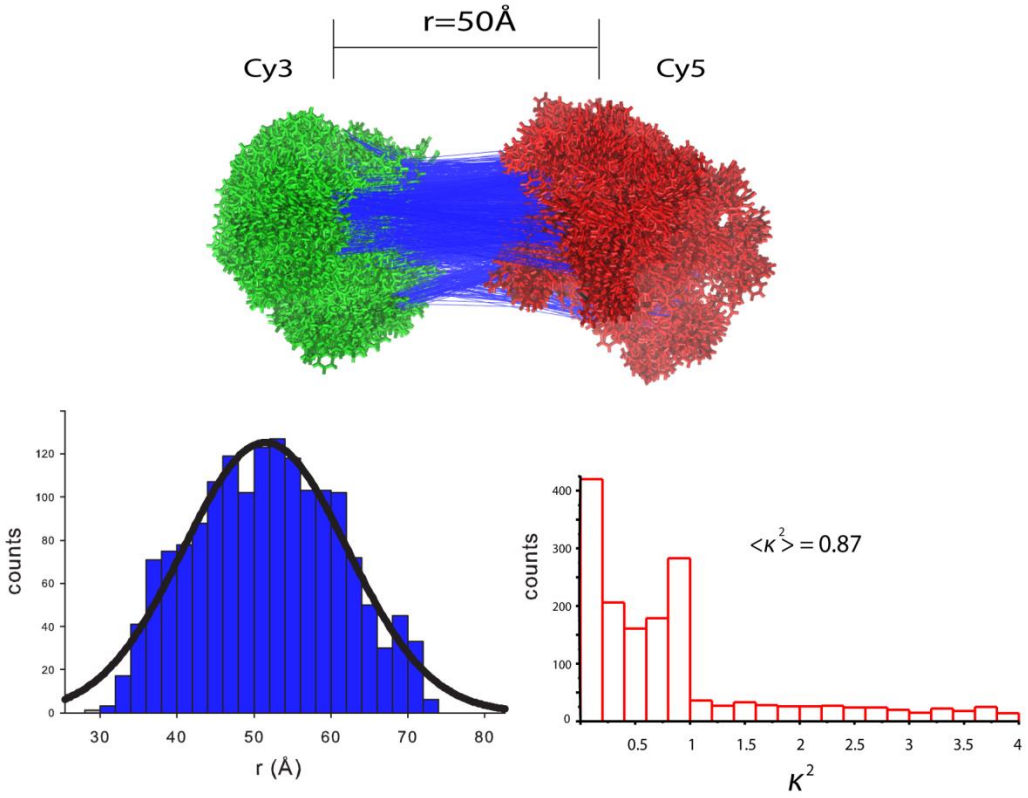

Figure 36. Simulation of FRET

FRET simulations for the two tripeptides held at 5 Å apart are presented. The top schematic indicates the positions of the dyes over the course of the trajectory (taken every 0.1 ns). The bottom left is a histogram of all instantaneous FRET-derived distance values calculated over the trajectory. The bottom right figure shows a histogram of κ^2 also calculated instantaneously for every frame of the simulation.

Currently, we are running AMD simulations of the recent *apo* and *holo* structures of RT-T/P with or without NVP (Das et al., 2012) (PDB IDs 3VD6 and 3V81), bound to Cy3 and Cy5 at D250C and T139C on the respective thumb and fingers domains (as shown in Figure 28a) which required removal of the DNA-fingers crosslink and the terminal AZT on the primer, and also required parameterization of NVP (using CGENFF) and Cy3- and Cy5-maleimides. As of the time of this writing, the simulations have reached about 200 ns each and are gaining ~4.5 ns/day each. We then intend to carry out the experiments outlined in 3.2.3, using our stopped-flow apparatus to inject NVP during observation time.

4.0 CONCLUSIONS

HIV-1 RT must form a stable ternary RT-T/P-dNTP complex to facilitate phosphodiester bond formation and elongation of the DNA primer. Our data shows that EFV substantially increases the sliding of RT on the T/P, which effectively reduces the enzyme's dwell time in the polymerase-competent ternary complex. EFV was also found to decrease the affinity of dNTP for the RT-T/P complex. The observed increase in RT sliding and decrease in dNTP binding provide a plausible mechanism by which EFV (and other NNRTIs) inhibit HIV-1 reverse transcription. In pre-steady-state kinetic studies, the EFV-bound RT-T/P complex exhibits an increase in dNTP affinity. However, it should be noted that the anisotropy and pre-steady-state kinetic assays measure different events. In the anisotropy assays we assess dNTP binding to an RT-T/P complex that cannot undergo catalysis because the DNA primer is chain-terminated. As such, we likely capture both productive and non-productive dNTP binding. In contrast, the K_d determined from pre-steady-state kinetic studies only measures productive dNTP binding to an RT-T/P complex actively undergoing catalysis. Unexpectedly, we found that the K103N mutation in RT does not prevent EFV from binding to the RT-T/P complex but allows the enzyme to form a stable polymerase-competent RT-T/P-dNTP ternary complex. However, the K103N mutation did decrease the affinity of NVP for the RT-T/P complex. In comparison to NVP, EFV is a smaller drug and can reposition itself within the pocket, whereas the bulky fused ring system of NVP remains relatively rigidly in place and no such rearrangements are possible.

Overall, this is consistent with earlier predictions from crystallographic data (Ren et al., 2000). The single-molecule FRET data unambiguously shows that the K103N mutation relieves the molecular arthritis in the fingers and thumb sub-domains of RT such that it can form a stable RT-T/P-dNTP complex. The molecular arthritis phenotype is likely driven by the formation of salt bridge between K101 and E138 at the hinge region of the RT dimer interface. Of note, RIL is active against HIV-1 containing K103N, and our anisotropy studies reveal that the drug efficiently binds to and mobilizes K103N RT. Interestingly, in the structure of K103N RT in complex with rilpivirine the salt bridge between K101 and E138 is intact (Lansdon et al., 2010). In this structure, the central, electrostatically positive nitrogenous ring of the three ring diarylpryimidine moiety appears to be responsible for repositioning K101 into a salt bridge-competent configuration in the structure of K103N bound to rilpivirine (Lansdon et al., 2010). Taken together with our data, this observation suggests that a positive interloping substituent can take place of K103, whose purpose in the context of NNRTI activity is to precisely position K101 into register to constitute the E138-K101 salt bridge.

In addition to sampling on-pathway transitional intermediates in NNRTI-induced conformational change, we plan to use the simulation results from Chapter 3 to directly monitor the connection between E138-K101 salt bridge integrity and thumb/finger dynamics, with the working hypothesis that the two are directly related. Since it is possible that the E138-K101 salt bridge is only partially responsible for the overall effect of NNRTI, this analysis will directly demonstrate the magnitude of interdependence of the global dynamics of RT on the NNRTI-induced salt bridge.

Our data highlights an unexpected and novel mechanism of HIV-1 resistance to NNRTIs and will be critical in transforming NNRTI drug design from a shotgun/library-based approach to

a rational approach which incorporates molecular mechanism(s) of action as well as resistance mechanism(s). Since using our unique combination of techniques to probe the mechanism of a single NNRTI resistance mutation yielded much needed insight into the mechanism of NNRTIs as well as the mechanism of NNRTI resistance, essential to this future endeavor into rational NNRTI design will be the understanding of other resistance mutations in the context of other NNRTIs. For instance, we have begun to investigate the effects of RIL on RT T/P dynamics and intradomain conformation, and are interested in the mechanism of various RIL resistance mutations such as E138A/G/K/Q/R and/or K101E/P, which appear to target the salt bridge and its effects on RT dynamics. Results from these studies will be critical in the design of novel, effective inhibitors against an ever-emerging battery of NNRTI resistance mutations in HIV-1 RT.

APPENDIX A

MATLAB SCRIPTS USED FOR DATA ANALYSIS

********** ********** PIFE.m **********

```
function PIFE (fname,m,n)
%% usage: redlaser ('fname',n)

%where fname is root file name: i.e. for example 1, example 2, etc.
%usage is redlaser ('example',1)

workdir= 'F:\PIFE\130328.traces\'  %where the .traces file exists
timeunit=0.015; %define CCD frame increment in seconds
fold_intensity=0; %set this to 1 to convert intensity histograms to "fold" histograms via the
PIFE fit routine.


%set "high" to be divisible by bin_number
bin_number=25; %number of bins to be used in all histograms
high=2000; low = 0; bin_interval = (high-low)/bin_number; %define upper/lower limit of intensity
values
PIFE_space=low:bin_interval:high;

%FILTERING AND CORRECTION
filter=0;  %set to 1 to use savitsky-golay filter on FRET signal
filter_window=5; %must be odd number >1
number_polys=2;

vb_cat='y';   %Use this to go through multiple regions, pressing 'v' and taking regions
close all;
cd(workdir);
%for n=startn:endn
%% Definitions
           global base_mean;
           donor_count=0;
           vb_count=0;
           donor_regions_all_points=0;
           vb_regions_all_points=0;
           vb_struct={};

                   if nargin==1
                       n=0; manyregions='n';
                   elseif nargin==2 | nargin==3
                       manyregions='y';  %multi-region (static histogram) run
                   end

%% For every slide region
for region=m:n
```

```
        if nargin==2 | nargin==3
            fnames=[fname '_' num2str(region)];
        elseif nargin==1
            fnames= fname;
        end


        fid=fopen([fnames '.traces'],'r');
        len=fread(fid,1,'int32');
        time=timeunit:timeunit:len*timeunit;
        Ntraces=fread(fid,1,'int16');
        disp(['Slide region ' num2str(region) ': Number of time traces: ' num2str(Ntraces/2) ',
Number of frames is ' num2str(len),', framerate = ' num2str(timeunit),'s'])
        raw=fread(fid,Ntraces*len,'int16');
        fclose(fid);
        tracker=['./traces/' fnames '_track.dat'];
        fname_fret_reg_tally=['./traces/' fnames '_reg_FRET.dat'];
        fname_vbFRET=['./traces/' fnames '_vbFRET.mat'];
        fname_PIFE=['./traces/' fnames '_PIFE.mat'];
      %% convert into donor traces

      index=(1:Ntraces*len);
      Data=zeros(Ntraces,len);
      donor=zeros(Ntraces/2,len);
%       vb donor=zeros(Ntraces/2,len);
%       vb_acceptor=ones(Ntraces/2,len);
      vbFRET_scale_factor=zeros(Ntraces/2);
      vbFRET_acceptor=zeros(Ntraces/2,len);
      vbFRET_donor=zeros(Ntraces/2,len);
      track=zeros(len+1,2);
      Data(index)=raw(index);

  for i=1:(Ntraces/2)
      for j=1:len
          donor(i,j)=Data(i*2-1,j); % + beta*donor(i,:)
%           vb_donor(i,j)=donor(i,j)+1;%trying to trick vbFRET; failed.
%           vb_FRET(i,j)=(1/(vb_donor(i,j)-1))^-1;
      end

%       vbFRET_acceptor(i,:)=vb_FRET(i,:);
%       vbFRET_donor(i,:)=1-vbFRET_acceptor(i,:);  %trying to trick vbFRET; failed.

        vbFRET_scale_factor(i)=max(donor(i,:));
        vbFRET_acceptor (i,:) = donor(i,:)/vbFRET_scale_factor(i); %scale to 1 for
vbFRET.vbFRET acceptor=vb FRET;

  end

        vbFRET_donor = 1-vbFRET_acceptor;  %just a requirement of vbFRET

                if (region==0) | (region==m)
                    PIFE_all_regions=donor;
                else
                    PIFE_all_regions=vertcat(PIFE_all_regions,donor);
                end


  %% calculate, plot and save AVERAGE traces

   dAvg=sum(donor,1)/Ntraces*2;

   figure(1); %hdl1=gcf; figure(hdl1);
   plot(time,dAvg,'g');
   title('Average donor');
   zoom on;
   avgOutput=[time' dAvg'];
   avgFileName=[fnames '_avg.dat'];
   save(avgFileName,'avgOutput','-ascii');

  if manyregions=='n'
      figure(2)
```

```
      title('Avg intensity of all traces for first 100 frames');
      subplot(1,2,1);
      hist(dAvg(1:100),20,'FaceColor','g','EdgeColor','k');
      legend(['donor average= ' num2str(mean(dAvg(1:100)))])
      subplot(1,2,2);
      hist(dAvg(1:100),20,'FaceColor','r','EdgeColor','k');
      legend(['acceptor average= ' num2str(mean(dAvg(1:100)))])
  end

  MSGID =  'MATLAB:MKDIR:DirectoryExists'; warning ('off',MSGID);


%% Define trajectory regions for multiple slide regions
if manyregions=='n' || vb_cat=='y'
mkdir traces;
rk = 0; lkb = 0; lknb = 0; lk2= 0; lk = 0; i=0; saved=0; k=0;

% Enter trace window.  needs while loop (not for loop).
while i <((Ntraces/2)),

i=i+1;

figure(3);  %%TRACE window

subplot(2,1,1); %D trace
   plot(time,donor(i,:),'g');
   line([0 max(time)],[0.1 0.1], [0,0], 'Color','r', 'linewidth',2);

   temp=axis;

  temp(3)=min( min(donor(i,:))) ; temp(4)=max( max(donor(i,:))) ;
  grid on; zoom on; axis(temp);
  title(['Molecule ' num2str(i)]);

subplot(2,1,2);% donor histogram
  hist(donor(i,:),25);
  temp=axis; title(['Intensity histogram for trace ' num2str(i)]);

%% MAKE A SELECTION
ans=input('save([),pass(CR),back(p), or (e)nd?','s');
switch lower(ans)
    case 'e'
            break

    case 'p'
        i=i-2;


    case 'g'  %GOTO molecule
     nml = input('which molecule?');
     i=nml - 1;

    case 'f'

      donor(i,:)=sgolayfilt(donor(i,:),number_polys,filter_window);
      i=i-1;

    case 'r' %define region.  should do this after 'b' and 'c'.  save to array

         clear X Y click_first click_last mollength ginput started;
         [X, Y, mousebutton]= ginput;

        mollength= length(X);

        if mollength==2 %if 2 clicks, define regionp

            for clk=1:mollength
              if clk==1
                    totalclickfirst_reg= X(clk);
                    totalclickfirst_reg= fix(totalclickfirst_reg/timeunit);
                elseif clk==2
```

```
                totalclicklast_reg= X(clk);
                totalclicklast_reg= fix(totalclicklast_reg/timeunit);
            end
            end

            totalclicktime_reg= abs(totalclicklast_reg - totalclickfirst_reg)';

          clear time reg donor reg  %reload region variables each time since dimensions will vary
            time_reg= time(totalclickfirst_reg:totalclicklast_reg);  %+1
            donor_reg=donor(i,totalclickfirst_reg:totalclicklast_reg);

      i = i -1;
      end

  case 'x'
        %region=region+1
        clear X Y click_first click_last mollength ginput started;
        [X, Y, mousebutton]= ginput;
        click_first= 0;
        click_last= 0;
        started=0;

        mollength= length(X);

        if (mod(mollength,2)==0) & (mollength > 0)

            for clk=1:mollength

                if started==0
                    click_first= click_first+1;
                    totalclickfirst(click_first)= X(clk);
                    totalclickfirst= fix(totalclickfirst/timeunit);
                    totalclickfirst

                    started=1;

                else
                    click_last= click_last+1;
                    totalclicklast(click_last)= X(clk);
                    totalclicklast= totalclicklast;
                    totalclicklast= fix(totalclicklast/timeunit);
                    totalclicklast
                    started=0;
                end
            end
            totalclicktime= abs(totalclicklast- totalclickfirst)';
            totalclicktime


           %%%%%%%%%%%%%%%save the picked traces%%%%%%%%%%%%%%%%%

            fname2=['./traces/' fnames '_tr_' num2str(i) '_region.dat'];
            output=[time(1:totalclicktime+1)' donor(i,totalclickfirst:totalclicklast)'];
            save(fname2,'output','-ascii') ;
           %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
        end

case 's'
        save (fname_vbFRET,'-struct', 'vb_struct');  %temporary save (good if you crash the
program a lot by e.g. clicking outside of the region);

case {'v', '['}
         clear X Y click_first click_last mollength ginput started;
        [X, Y, mousebutton]= ginput;

        mollength= length(X);

        if mollength==1 %if 1 clicks, define region and save as vbFRET input
            for clk=1:mollength
              if clk==1
                    totalclickfirst_reg= timeunit;
```

```
                totalclickfirst_reg= fix(totalclickfirst_reg/timeunit);
                totalclicklast_reg= X(clk);
                totalclicklast_reg= fix(totalclicklast_reg/timeunit);
            end
            end

        totalclicktime_reg= abs(totalclicklast_reg - totalclickfirst_reg)';

        if totalclicktime_reg > len;
            disp ('You clicked outside of the region')
            %clear clear totalclickfirst_reg; clear totalclicklast_reg;
            i=i-1;


        end

        clear time_reg donor_reg donor_intensity  %reload region variables each time since
dimensions will vary

        if totalclicktime_reg <len
           time_reg= time(totalclickfirst_reg:totalclicklast_reg);  %+1

        %for fetching later


        %these regions are for vbFRET.. weirdness to feed Id as E
        donor_reg=vbFRET_donor(i,totalclickfirst_reg:totalclicklast_reg);
        acceptor_reg=vbFRET_acceptor(i,totalclickfirst_reg:totalclicklast_reg);

        %this region is for PIFE_fit
        donor_region = donor(i,totalclickfirst_reg:totalclicklast_reg);


        %clear output_vbFRET;
        output_vbFRET=[donor_reg' acceptor_reg'];
        vb_count=vb_count+1;

      end
    elseif mollength==2   % if you just press enter, save the whole trace

        output_vbFRET=[vbFRET_donor(i,:)' vbFRET_acceptor(i,:)'];
        vb_count=vb_count+1;
     elseif mollength==0
     end

f fold_intensity==1;
    %now make a histogram and send it to be curve fit
    figure(10);

    s=size(donor_region,2);

     xlim= donor_region<=high;
     donor_region=donor_region(xlim);
     xlim2= donor_region >= low;
     donor_region=donor_region(xlim2);


       [freq,x_data]=hist(donor_region,PIFE_space);
       x_data=x_data';
       y_data=(freq./s(1,1))';
       PIFE_fit(x_data,y_data);



ans2= input('Choose PIFE 1-fold value: Use (;,l)owest peak mean, (d)efine, or (p)oint to
it?','s');
      switch lower(ans2)
          case {'l',';'}
              donor_region=donor_region./base_mean;
              figure(11);
              hist(donor_region,bin_number);
```

```
          case 'd'
              ans3=input('Enter base value: ');
              donor_region=donor_region./ans3;
              figure(11);
              hist(donor_region,bin_number);

          case 'p'  %you align the x-axis with the noise to define bckgnd
             figure(3);


             [Xb,Yb] = ginput;
             if numel(Yb)==0;break;end
             donor_base = Yb(1);
             donor_region=donor_region./donor_base;
             figure(11);
             hist(donor_region,bin_number);

      end
end
            %Creat a structure called PIFE with labels, donor intensity,
            %normalized (fold) intensity, and the fold-factor required to divide the intensity to
get the
            %fold value.  Saved as fname_PIFE.m
            PIFE.labels{vb_count}= [fnames '_mol_'  num2str(i)];
            PIFE.Cy3_intensity{vb_count}=donor(i,:);
            PIFE.fold_intensity{vb_count}=donor_region;

PIFE.fold_factor{vb_count}=PIFE.Cy3_intensity{vb_count}(:,1)./PIFE.fold_intensity{vb_count}(:,1);

            %Create a structure called vb_struct with labels and corrected
            %donor/acceptor intensity required to spoof vbFRET into
            %thinking the trace is FRET.  Feed directly to vbFRET.
            vb_struct.data{vb_count}= output_vbFRET;
            vb_struct.labels{vb_count} = [fnames '_mol_'  num2str(i)];
            vb_struct.scale_factor_intensity{vb_count}=vbFRET_scale_factor(i); %corrects vbFRET E
to Intensity

vb_struct.scale_factor_fold{vb_count}=(vb_struct.data{vb_count}(1,2).*vb_struct.scale_factor_inte
nsity{vb_count})./PIFE.fold_factor{vb_count}; %corrects vbFRET E to Fold I

            if vb_count==1
                vb_regions_all_points=output_vbFRET;
                donor_regions_all_points=donor_region';
            else
                vb_regions_all_points=[vb_regions_all_points; output_vbFRET];
                donor_regions_all_points=[donor_regions_all_points; donor_region'];
            end

        %Ntraces=0; continue;

   end %end switch
end

elseif manyregions=='y'
close all
continue

end

 end

%% makes total histogram of all slide regions
 if size(PIFE_all_regions)>=1;
    mkdir all_points_hist;

      PIFE_all_regions=reshape(PIFE_all_regions,1,[]);
      s=size(PIFE_all_regions,2);

           xlim= PIFE_all_regions<=high;
      PIFE_all_regions=PIFE_all_regions(xlim);
             xlim2= PIFE_all_regions >= low;
```

```
      PIFE_all_regions=PIFE_all_regions(xlim2);

      [freq,xout]=hist(PIFE_all_regions,PIFE_space);
      histdata=[xout' (freq./s(1,1))'];
      figure (4);
      bar(xout,freq./s(1,1),1,'w'); %plot normalized histogram
      xlabel ('I');
      labn=['p ' '(n=' num2str(s(1,1)) ')'] ;
      ylabel (labn);
      titulo=['All points histogram' workdir '\' fname];
      title (titulo,'interpreter','none'); %suppress latex interpretation in title
      savefile=[workdir 'all_points_hist\' fname '_all_points_hist.dat']; %save hist data as
fname_hist.dat
      save(savefile, 'histdata', '-ascii', '-tabs'); %tab delimited
end

%% vbREGions are tallied using "["), makes a histogram of these regions.

if vb_count >= 1
       save (fname_vbFRET,'-struct', 'vb_struct');
       save (fname_PIFE,'-struct','PIFE');

       %
mkdir vb_regions_hist;
%% vbHist
     s=size(vb_regions_all_points,2);
         xlim= vb_regions_all_points<=high;
     vb_regions_all_points=vb_regions_all_points(xlim);
     xlim2= vb_regions_all_points >= low;
     vb_regions_all_points=vb_regions_all_points(xlim2);


        [freq,xout]=hist(vb_regions_all_points,PIFE_space);
        histdata=[xout' (freq./s(1,1))'];
        figure (5);
        bar(xout,freq./s(1,1),1,'w'); %plot normalized histogram

        xlabel ('I'); labn=['p ' '(n=' num2str(s(1,1)) ')'] ;
        ylabel (labn);
        titulo=['Regions histogram: ' workdir '\' fname]; title (titulo,'interpreter','none');
%suppress latex interpretation in title
        savefile=[workdir 'vb_regions_hist\' fname '_vb_regions_hist.dat']; %save hist data as
fname_hist.dat
        save(savefile, 'histdata', '-ascii', '-tabs'); %tab delimited

 %% donor region hist (normalized to 1)
figure(8);
hist(donor_regions_all_points,50);

fold_bin_number=50; %number of bins to be used in all histograms
     fold_low=0; fold_high = 6; fold_bin_interval = (fold_high - fold_low)/fold_bin_number;
%define upper/lower limit of intensity values

s=size(donor_regions_all_points,1);
         xlim= donor_regions_all_points<=fold_high;
     donor_regions_all_points=donor_regions_all_points(xlim);
     xlim2= donor_regions_all_points >= fold_low;
     donor_regions_all_points=donor_regions_all_points(xlim2);


     donor_space=fold_low:fold_bin_interval:fold_high;

        [freq,xout]=hist(donor_regions_all_points,donor_space);
        histdata=[xout' (freq./s(1,1))'];
        figure (6);
        bar(xout,freq./s(1,1),1,'w'); %plot normalized histogram

        xlabel ('fold increase in intensity'); labn=['p ' '(n=' num2str(s(1,1)) ')'] ;
        ylabel (labn);
        titulo=['Fold histogram: ' workdir '\' fname]; title (titulo,'interpreter','none');
%suppress latex interpretation in title
```

%SKELETOR

```
%saves fname_fold_hist.dat for later use
        savefile=[workdir 'traces\' fname '_fold_hist.dat']; %save hist data as fname_hist.dat
        save(savefile, 'histdata', '-ascii', '-tabs'); %tab delimited

end
%%
if manyregions=='y'
       % PIFE_histogram(workdir,fname,n);
end


***************PIFE_fit.m*****************************************************


function pife_fit(x_data,y_data)
global cf1; global cf2; global base_mean;
%close all;

n=2;

if n==2
cf1 = fit(x_data,y_data,'gauss2');
 zero_means=[abs(cf1.b1),abs(cf1.b2)];
  if min(zero_means)==cf1.b1
            a=cf1.a1;b=cf1.b1;c=cf1.c1;
        elseif min(zero_means)==cf1.b2
            a=cf1.a2;b=cf1.b2;c=cf1.c2;
  end



%disp(['b2=',num2str(cf1.b2)]);
if cf1.b1<cf1.b2
    base_mean=cf1.b1;
elseif cf1.b1>cf1.b2
    base_mean=cf1.b2;
end

disp(['mean=',num2str(base_mean)]);
%
% kill=a*exp(-((x_data-b)/c).^2);
% killed=y data-kill;
%
% cf2=fit(x_data,killed,'gauss1');

%figure (10);
plot(x_data,y_data,'bo')
hold all
plot(cf1,'r')
%plot(x_data,killed,'bx')
%plot(cf2,'g')

legend('data', 'gauss2')
hold off

elseif n==3;
cf1 = fit(x_data,y_data,'gauss3'); % "natural" fit

zero_means=[abs(cf1.b1),abs(cf1.b2),abs(cf1.b3)];

        if min(zero_means)==cf1.b1
```

```
            a=cf1.a1;b=cf1.b1;c=cf1.c1;
        elseif min(zero_means)==cf1.b2
            a=cf1.a2;b=cf1.b2;c=cf1.c2;
        elseif min(zero_means)==cf1.b3
            a=cf1.a3;b=cf1.b3;c=cf1.c3;
        end

kill=a*exp(-((x_data-b)/c).^2);
killed=y_data-kill;

cf2=fit(x_data,killed,'gauss2');

%figure (10);
plot(x_data,y_data,'bo')
hold all
plot(cf1,'r')
plot(x_data,killed,'bx')
plot(cf2,'g')

legend('original data', 'gauss3','minus zero peak','gauss2')
hold off

cftool(x_data,y_data);
end


**********path_fit.m*********************************************

function path_fit (fname);
%Fit vbFRET PATH files to low/high dwell time distributions & on/off rates
workdir= 'F:\PIFE\130126.traces\'   ;  %use "\" at end!
timeunit=0.015;  %CCD integration time
std_threshold=3;  %outside of this STDEV range = outlier
rounds_outlier_removal=2; %number of iterations of outlier removal
n_bins_tao=100;  %specify the number of bins.
exp_trunc=20;  %usually this is too many bins, so only use the first exp_trun to fit w/
exponential
n_bins_k=15;  %this was 15

cd(workdir);
fid=([fname '.dat']); %need the beginning space?
tao_filename=(['tao_' fname '.dat']); %name of savefile
raw=load(fid);
n=raw(end,1); %number of PIFE traces saved
begin=raw(1,1);
%total_frames=size(raw,1);

total_transition_count=0; dumped=0;
for i=begin:n;  %start looping through molecules
  mol_index=find(raw(:,1)==i);
  mol_dat=raw(mol_index,2);   %mol_dat is a 1-column vector of 2 states for every molecule

  low_state=min(mol_dat);  %define low/high states
  high_state=max(mol_dat);

  %transition count=0;
  low_to_high_transition_count=0;high_to_low_transition_count=0;
  low_run_length=1; high_run_length=1; %initialize

  for j=1:size(mol_dat,1)-1;  %find contiguous states

              if mol_dat(j) == mol_dat(j+1) && j < size(mol_dat,1)  %if the next one is the same
                    if mol_dat(j)==low_state
                       low_run_length=low_run_length+1;
                    else
                       high_run_length=high_run_length+1;
                    end
               elseif mol_dat(j) ~= mol_dat (j+1) && j < size(mol_dat,1)
          %transition_count = transition_count+1;
          total_transition_count=total_transition_count+1;
```

```
        if mol_dat(j)==low_state
            low_to_high_transition_count=low_to_high_transition_count+1;
            tao.low(low_to_high_transition_count,i)=timeunit*low_run_length;
        end
                if low_to_high_transition_count == 0
                  tao.low(1,i)=0;    %this was happening by default but not if the first or last
only had 1 transition.
                end



        if mol_dat(j)==high_state
            high_to_low_transition_count= high_to_low_transition_count+1;
            tao.high(high_to_low_transition_count,i)=timeunit*high_run_length;
        end
                  if high_to_low_transition_count == 0
                     tao.high(1,i)=0;    %this was happening by default but not if the first or
last only had 1 transition.
                  end


        %tao.high cat(high to low transiton count,1)=timeunit*high run length;  %1 column vector
of concatenated taos (because the structure was adding zeros and I didn't feel like picking them
out of each column)
        %tao.low cat(low to high transition count,1)=timeunit*low run length;

        low_run_length=1; high_run_length=1;
              end

  tao.transitions(i)=low_to_high_transition_count + high_to_low_transition_count;
  end

end
tao.high_cat(:,1)=nonzeros(tao.high);
tao.low_cat(:,1)=nonzeros(tao.low);

%Remove traces w/out a transition
zero_ind=find(tao.transitions==0);
%if vbFRET only found one state, revise the number
       tao.low(:,zero_ind)=[];
       tao.high(:,zero_ind)=[];
       tao.transitions(:,zero_ind)=[];
n2=n-size(zero_ind,2);


for i=1:n2
tao.frames_kon(1,i)=size(nonzeros(tao.low(:,i)),1);
tao.frames_koff(1,i)=size(nonzeros(tao.high(:,i)),1);
%tao_dwell (A or B) = total time in A or B)/(number of transitions)
%i.e., tao=average of all state dwell times, rate k=1/tao

%kon=1/tao low
%tao.kon(1,i)=1./(sum(nonzeros(tao.low(:,i)))./tao.transitions(1,i));
tao.kon(1,i)=1./(mean(nonzeros(tao.low(:,i))));   %average rate for each molecule, 1 per column
%tao.kon_errors(2,i)=std(1./(nonzeros(tao.low(:,i))));
%tao.kon2(1,i)=mean(1./nonzeros(tao.low(:,i)))

%koff=1/tao high
%tao.koff(1,i)=1./(sum(nonzeros(tao.high(:,i)))./tao.transitions(1,i));
tao.koff(1,i)=1./(mean(nonzeros(tao.high(:,i)))); %average rate for each molecule, 1 per column
%tao.koff_errors(2,i)=std(1./(nonzeros(tao.high(:,i))));
%tao.koff2(1,i)=mean(1./nonzeros(tao.high(:,i)));

%not utilized, just for comparison
tao.kon_total=1./(sum(tao.low_cat)/total_transition_count);
tao.koff_total=1./(sum(tao.high_cat)./total_transition_count);
end

tao.kon_mean=mean(tao.kon);
tao.koff_mean=mean(tao.koff);
tao.kon_mean_std=std(tao.kon);
```

```
tao.koff_mean_std=std(tao.koff);

tao.kon_ave=1./(mean(nonzeros(tao.low(:))));  %non-weighted average of ALL tao, 1/<tao>
tao.koff_ave=1./(mean(nonzeros(tao.high(:))));
tao.koff_ave_std=std(nonzeros(tao.low(:)));
tao.kon_ave_std=std(nonzeros(tao.high(:)));

%now I have a distribution of all calculated rates, but a singular average of these
%rates needs to be weighted by total observation time.  NOTE: STDEV's are
% weighted using an ubiased estimator.
%See en.wikipedia.org/wiki/Weighted_mean#Weighted_sample_variance
%For the distributions I will plot all of them along with a distribution
%after discarding the half of the data with the lowest observation time
  total_frames_kon=sum(tao.frames_kon); total_frames_koff=sum(tao.frames_koff);
  kon_weight= tao.frames_kon./total_frames_kon;
  koff_weight= tao.frames_koff./total_frames_koff;

  %weighted_kon(1,i)= tao.kon(1,i)*kon_weight(1,i);
  %weighted_koff(1,i)=tao.koff(1,i)*koff_weight(1,i);
  weighted_kon=tao.kon.*kon_weight;
  weighted_koff=tao.koff.*koff_weight;

  tao.kon_outliers=tao.kon;
  tao.koff_outliers=tao.koff;
  %weighted kon(1,i)=tao.kon(1,i).* ( size(nonzeros(tao.low(:,i)),1) ./ size(tao.low cat,1) ) ;
  %weighted_koff(1,i)=tao.koff(1,i).* (size(nonzeros(tao.high(:,i)),1)./ size(tao.low_cat,1)  );


tao.kon_weighted_mean=sum(weighted_kon);
tao.koff_weighted_mean=sum(weighted_koff);

%Weighted STD w/ unbiased estimator.  Use these ones.
v2_on=sum(kon_weight.^2); v2_off=sum(kon_weight.^2); %used for unbiased estimate
tao.kon_weighted_std_unbiased= (1./(1-v2_on)) .* (sqrt(sum(kon_weight.*((tao.kon -
tao.kon_weighted_mean).^2))));
tao.koff_weighted_std_unbiased= (1./(1-v2_off)) .* (sqrt(sum(koff_weight.*((tao.koff -
tao.koff_weighted_mean).^2))));

for i=1:rounds_outlier_removal;

%Grubb's Z score
Z_kon=abs(tao.kon_weighted_mean-tao.kon)./tao.kon_weighted_std_unbiased;
Z_koff=abs(tao.koff_weighted_mean-tao.koff)./tao.koff_weighted_std_unbiased;

outlier_index_kon=find(Z_kon>std_threshold);
outlier_index_koff=find(Z_koff>std_threshold);
outlier_index_master=unique([outlier_index_kon outlier_index_koff]);

%Throw out the outliers, if any (~typically 1 or 2).
tossed_koff=size(outlier_index_koff,2); tossed_kon=size(outlier_index_kon,2);
tossed_total=size(outlier_index_master,2);

tao.kon(:,outlier_index_master)=[]; tao.koff(:,outlier_index_master)=[];
tao.low(:,outlier_index_master)=[]; tao.high(:,outlier_index_master)=[];
tao.frames_kon(:,outlier_index_master)=[];
tao.frames_koff(:,outlier_index_master)=[];

disp(['Round ' num2str(i) ':'])
disp(['Detected ' num2str(tossed_kon) ' kon  values outside of ' num2str(std_threshold) ' STDEVs
(' num2str(tao.kon_weighted_std_unbiased,2) ')' char(10) 'Detected ' num2str(tossed_koff) ' koff
values outside of ' num2str(std_threshold) ' STDEVs (' num2str(tao.koff_weighted_std_unbiased,2)
')' char(10) 'Tossed ' num2str(tossed_total) ' traces total']);


%Recalculate weighted mean/std

    clear kon_weight koff_weight
    %tao.kon weighted mean tao.koff weighted mean tao.kon weighted std tao.koff weighted std

    total_frames_no_outliers_kon=sum(tao.frames_kon);
    total_frames_no_outliers_koff=sum(tao.frames_koff);
```

```
    kon_weight= tao.frames_kon./total_frames_no_outliers_kon;
    koff_weight=tao.frames_koff./total_frames_no_outliers_koff;

    weighted_kon= tao.kon.*kon_weight;
    weighted_koff=tao.koff.*koff_weight;

    tao.kon_weighted_mean_post=sum(weighted_kon);
    tao.koff_weighted_mean_post=sum(weighted_koff);

    v2_on=sum(kon_weight.^2); v2_off=sum(kon_weight.^2); %used for unbiased estimate
    tao.kon_weighted_std_unbiased_post= (1/(1-v2_on)) .* (sqrt(sum(kon_weight.*((tao.kon -
tao.kon_weighted_mean_post).^2))));
    tao.koff_weighted_std_unbiased_post= (1/(1-v2_off)) .* (sqrt(sum(koff_weight.*((tao.koff -
tao.koff_weighted_mean_post).^2))));

    tao.kon_mean_post=mean(tao.kon);
    tao.koff_mean_post=mean(tao.koff);
    tao.kon_mean_std_post=std(tao.kon);
    tao.koff_mean_std_post=std(tao.koff);

    tao.kon_ave_post=1./(mean(nonzeros(tao.low(:))));  %non-weighted average of tao, POST STD
FILTERING
    tao.koff_ave_post=1./(mean(nonzeros(tao.high(:))));
    tao.koff_ave_std_post=std(nonzeros(tao.low(:)));
    tao.kon_ave_std_post=std(nonzeros(tao.high(:)));


end
%% frequency histograms
%

MSGID =  'MATLAB:MKDIR:DirectoryExists'; warning ('off',MSGID);
mkdir tao; mkdir cdf;
s_tao_low=size(tao.low_cat,1);  %total number of low states
s_tao_high=size(tao.high_cat,1);%total number high states
s_k=size(tao.kon,2);  %total number of molecules


fname_tao=[workdir 'tao\' fname '_TAO.mat'];
save (fname_tao,'-struct', 'tao');

close all;

        %TAO LOW
        [freq,xout]=hist(tao.low_cat,n_bins_tao);
        yout=freq./s_tao_low(1,1); %change to frequency
        histdata=[xout' yout'];
        xout_fit=xout(1:exp_trunc);
        yout_fit=yout(1:exp_trunc);
        options=fitoptions('exp1');
        set(options,'Robust','Off','MaxIter',5000);
        [cf,gof]=fit(xout_fit',yout_fit','exp1',options);
        x_curve=0:0.01:10;
        fit_low=cf.a.*exp(x_curve'.*cf.b);
        figure (1);
        %plot(xout,yout);
        bar(xout(1:50),yout(1:50),1,'w'); %plot normalized histogram
        hold all; plot(x_curve,fit_low,'r','LineWidth',2); %plot exponential fit
        xlim([0 4]);
        ylim([0 1.1*max(yout)]);
        %axes;
        %set(axes_handle,'XLim',[0 2])
        xlabel ('tao (s)'); labn=['p ' '(n=' num2str(s_tao_low(1,1)) ')'] ;
        ylabel (labn);
        kon_pre_exp=cf.a;
        kon_exp=abs(cf.b);
        text(0.5,0.1, ['kon=' num2str(kon_exp)]);
        titulo=['Low state dwell times: ' workdir '\' fname char(10) char(10) 'kon='
num2str(kon_exp) ', rsquare= ' num2str(gof.rsquare)]; title (titulo,'interpreter','none');
%suppress latex interpretation in title
```

```
        savefile=[workdir 'tao\' fname '_tao_low_hist.dat']; %save hist data as fname_hist.dat
        save(savefile, 'histdata', '-ascii', '-tabs'); %tab delimited
        assignin('base','low_hist',histdata);
        clear freq xout yout cf histdata ;

          %TAO_HIGH

        [freq,xout]=hist(tao.high_cat,n_bins_tao);
        yout=freq./s_tao_high(1,1);
        histdata=[xout' yout'];
        xout_fit=xout(1:exp_trunc);
        yout_fit=yout(1:exp_trunc);
        options=fitoptions('exp1');
        set(options,'Robust','Off','MaxIter',5000);
        [cf,gof]=fit(xout_fit',yout_fit','exp1',options);
        x_curve=0:0.01:10;
        fit_high=cf.a.*exp(x_curve'.*cf.b);
        figure (2);
        bar(xout(1:50),yout(1:50),1,'w'); %plot normalized histogram
        hold all; plot(x_curve,fit_high,'r','Linewidth',2); %plot exponential fit
        xlim([0 4]);
        ylim([0 1.1*max(yout)])
        xlabel ('tao (s)'); labn=['p ' '(n=' num2str(s_tao_high(1,1)) ')'] ;
        ylabel (labn);
        koff_pre_exp=cf.a;
        koff_exp=abs(cf.b);
        text(0.5,0.3, ['koff=' num2str(koff_exp)]);
        titulo=['High state dwell times: ' workdir '\' fname char(10) char(10) 'koff='
num2str(koff_exp) ', rsquare= ' num2str(gof.rsquare)]; title (titulo,'interpreter','none');
%suppress latex interpretation in title
        savefile=[workdir 'tao\' fname '_tao_high_hist.dat']; %save hist data as fname_hist.dat
        save(savefile, 'histdata', '-ascii', '-tabs'); %tab delimited
        assignin('base','high_hist',histdata);
        clear freq xout yout cf histdata ;
        assignin('base','koff_cat',tao.low_cat);

        %KON
        [freq,xout]=hist(tao.kon,n_bins_k);
        histdata=[xout' (freq./s_k(1,1))'];
        figure (3);
        bar(xout,freq./s_k(1,1),1,'w'); %plot normalized histogram
        hold on; bar(tao.kon_weighted_mean,max(freq./s_k),0.2,'k')
        xlabel ('kon (s^-^1)'); labn=['p ' '(n=' num2str(s_k(1,1)) ')'] ;
        ylabel (labn);
        %text(1,(0.95*max(freq./s k)),['Weighted mean= ' num2str(tao.kon weighted mean,3) ' +/- '
num2str(tao.kon_weighted_std_unbiased,3)])
        titulo=['kon: ' workdir '\' fname char(10) char(10) 'Weighted mean= '
num2str(tao.kon_weighted_mean,3) ' +/- ' num2str(tao.kon_weighted_std_unbiased,3)]; title
(titulo,'interpreter','none'); %suppress latex interpretation in title
        figure(12); title ('kon');
        [a,a2]=ecdf(tao.kon);
        cdf_data=[a2,a];
        stairs(a2,a);
        savefile=[workdir 'tao\' fname '_kon_hist.dat']; %save hist data as fname_hist.dat
        save(savefile, 'histdata', '-ascii', '-tabs'); %tab delimited
        clear freq xout histdata
        savefile=[workdir 'cdf\' fname '_kon_CDF.dat'];
        save(savefile, 'cdf_data', '-ascii', '-tabs'); %tab delimited

        %KOFF
        [freq,xout]=hist(tao.koff,n_bins_k);
        histdata=[xout' (freq./s_k(1,1))'];
        figure (4);
        bar(xout,freq./s_k(1,1),1,'w'); %plot normalized histogram
        hold on; bar(tao.koff_weighted_mean,max(freq./s_k),0.2,'k')
        xlabel ('koff (s^-^1)'); labn=['p ' '(n=' num2str(s_k(1,1)) ')'] ;
        ylabel (labn);
        %text(1,(0.95*max(freq./s k)),['Weighted mean= ' num2str(tao.koff weighted mean,3) ' +/-
' num2str(tao.koff_weighted_std_unbiased,3)])
```

```
        titulo=['koff: ' workdir '\' fname char(10) char (10) 'Weighted mean= '
num2str(tao.koff_weighted_mean,3) ' +/- ' num2str(tao.koff_weighted_std_unbiased,3)]; title
(titulo,'interpreter','none'); %suppress latex interpretation in title
        figure(11);title ('koff');
        [a,a2]=ecdf(tao.koff);
        cdf_data=[a2 a];
        stairs(a2,a);
        savefile=[workdir 'tao\' fname '_koff_hist.dat']; %save hist data as fname_hist.dat
        save(savefile, 'histdata', '-ascii', '-tabs'); %tab delimited
        savefile=[workdir 'cdf\' fname '_koff_CDF.dat'];
        save(savefile, 'cdf_data', '-ascii', '-tabs'); %tab delimited

        clear freq xout histdata


        %TRANSITIONS HISTOGRAM
        [freq,xout]=hist(tao.transitions,20);
        histdata=[xout' (freq./s_k(1,1))'];
        figure (3);
        bar(xout,freq./s_k(1,1),1,'w'); %plot normalized histogram
        hold on; bar(tao.kon_weighted_mean,max(freq./s_k),0.2,'k')
        xlabel ('# transitions per molecule'); labn=['p ' '(n=' num2str(s_k(1,1)) ')'] ;
        ylabel (labn);
        %text(1,(0.95*max(freq./s_k)),['Weighted mean= ' num2str(tao.kon_weighted_mean,3) ' +/- '
num2str(tao.kon weighted std unbiased,3)])
        titulo=['Transitions: ' workdir '\' fname char(10) char(10) 'Mean= '
num2str(mean(tao.transitions),3) ' +/- ' num2str(std(tao.transitions),3)]; title
(titulo,'interpreter','none'); %suppress latex interpretation in title

        savefile=[workdir 'tao\' fname '_transitions_hist.dat']; %save hist data as
fname_hist.dat
        save(savefile, 'histdata', '-ascii', '-tabs'); %tab delimited
        clear freq xout histdata
        %savefile= [workdir 'tao\' fname '_koff_hist.dat'];

%% Send a variable to the workspace
%First row contains : 1)weighted k 2) k from exp fit 3) k from all dwell times and 4) non-
weighted average of all k's.

kon=[tao.kon_mean tao.kon_mean_post tao.kon_weighted_mean , tao.kon_weighted_mean_post ,
tao.kon_ave , tao.kon_ave_post , kon_exp];
kon(2,:)=[tao.kon_mean_std tao.kon_mean_std_post tao.kon_weighted_std_unbiased ,
tao.kon_weighted_std_unbiased_post , tao.kon_ave_std , tao.kon_ave_std_post , kon_pre_exp];

koff=[tao.koff_mean tao.koff_mean_post tao.koff_weighted_mean tao.koff_weighted_mean_post
tao.koff_ave tao.koff_ave_post koff_exp];
koff(2,:)=[tao.koff_mean_std tao.koff_mean_std_post tao.koff_weighted_std_unbiased
tao.koff_weighted_std_unbiased_post tao.koff_ave_std tao.koff_ave_std_post koff_pre_exp];

assignin('base','kon',kon);
assignin('base','koff',koff);
```

**********tiffdir.m**********

```
        function tiffdir (workdir)

%% TiffDir
Give it the directory name and it will mop up any and all spooled output from Solis.
% Images are rotated, and if they're split into two (X1,X2), they are sewn together.
% Assumes you spool to disk in Solis using "basename," which is incremented automatically e.g.

% basename   ---> rot basename 0
% basename_1 ---> rot_basename_1
% basename_2 ---> rot_basename_2    etc., and for long files:

% longfile basename
% longfile_basename_X2   ------> rot_longfile_basename_0
% longfile_basename_1
```

```
% longfile_basename_1_X2 ------> rot_longfile_basename_1
% longfile basename 2
% longfile_basename_2_X2 ------> rot_longfile_basename_2  etc.

% IMPORTANT NOTES:
% 1) MAKE SURE to remove all index/ghost files (i.e. <1kb tif from spool).
% 2) NEVER use an "X" when you name your files.
% 3) DO NOT end a spool file name with a number or a symbol (or else you have to add "0" to the
name of the first file after processing, useful for processing with redlaser_dir)
% 4) tiffdir will crash if folder contains tiffs with different pixel dimensions (i.e. 512x512
and 512x256)
% 5) Program assumes it only goes to X2. This is the case for our system: spooling splits files
% after ~4000 frames, and we have a rough 8000 frame limit since we can't address >4gb RAM
% with the current LibTiff.  Future versions of Matlab will support BigTiff.
%
% Copyleft Grant Schauer 05/10
auto_delete='y';

if nargin==0  %if you don't give it an argument, it will let you select it in a GUI

workdir2=uigetdir('J:\Sean\10-8')
workdir=[workdir2 '\'];
end
if workdir == 0  %if you write "redlaser 0" at the command line it will use these in/out
directories:
 workdir='I:\Grant\PIFE\120812\';  %add the '\' at the end
 outdir= 'J:\Grant\PIFE\120812\';
 mkdir (outdir);
else
    outdir=workdir;
end
cd(workdir);
list=ls(workdir);  %list of all files in directory including "." and ".." (always 1 and 2).
no_files=size(list,1); %number of files, including "." and ".."
x_ind=zeros(no_files,1);
count=0;
nframes_total=0;
all_x=0;
%% START WITH THE X1/X2 files
disp(['Total number of files: ' num2str(no_files-2)])
total_time=tic;
for i=1:no_files   %for every file,

   current_file=list(i,:);
   count=count+1;
   hi= find(current_file=='X'); %find all thefiles with an "X"
    if hi>0
      x_ind(count,:) = find(current_file=='X'); %column vector, containing 0 if no X, and n if X
where n= coulumn # of 'X'
    end

end
count=0;
filenum_x=find(x_ind); %rows that contain files w/ "X"

for i=1:size(filenum_x,1);  %for all X files, find partner and concatentate

            j=x_ind(filenum_x(i)); %the character position of the actual "X"

      %Solis will name e.g. filename and filename_X2
      current_x_file=list(filenum_x(i),:);

      %Define the name of X1 based on the name of X2 (discovered by "X" in the filename)
      x2_file=current_x_file;
      x1_file=[x2_file(1:j-2),x2_file(j+2:end)];  %subtract the "_X2_" out of the name

      all_x(i,1:size(x1_file,2))=x1_file(1,:);  %keep track of all X-files you found
      all_x=char(all_x);  %final all_x is a list of all X1 files.

x_time=tic;
  for x = 1:2;  % Start processing X1/X2 files
```

```
        if x==2
          tifname=x2_file;
        else
          tifname=x1_file;
        end

       if isletter(current_x_file(j-2))==0 %if the X1 file is numbered, keep it as is
          outfile=[outdir 'rot_', x1_file];
       else
          outfile =[outdir 'rot_' current_x_file(1:j-2),'_0', current_x_file(j+2:end)]; %if not,
add a 0
       end
    info=imfinfo(tifname);
    nFrames=(numel(info)); % grab number of frames

                if x==1
                    disp([num2str(nFrames) ' frames in X1: ' tifname ])
                else
                    disp([num2str(nFrames) ' frames in X2: ' tifname ])
                end

        for frame=1:nFrames
           %unfortunately A cannot be pre-allocated.
           A(:,:) = flipud(rot90(imread(tifname, frame, 'Info',info),3));
           if frame==1 && x==1;
             imwrite(A,outfile, 'tif', 'Compression','none')
           else
             imwrite(A,outfile, 'tif', 'Compression','none','WriteMode','append');
           end

        end

  nframes_total=nframes_total+nFrames;

  end

  disp([num2str(nframes_total),'-frame rotated/appended file written to ' outfile ])
  nframes_total=0;
   t(1)=toc(x_time);
   disp(['Time elapsed= ' num2str(t) ' seconds.'])
  disp(' ')

clear A;
end

%% Now do the rest of them.
%If you tallied any X's, that row position in variable 'check' will be 0.

check=ones(no_files,1);  %indexes all X's as 0's.

check(find(x_ind),1)=0; %X2=0

if all_x > 0  % has to be here or else strtrim below complains
 for i = 1:no_files;
  for j = 1:size(all_x,1);
      length=size(all_x(j),2);

      if isequal(strtrim(list(i,:)),strtrim(all_x(j,:)))==1;
        check(i)=0;  %X1=0
      end
  end
 end
end

filenum_nonx = find(check==1);  %vector of all row #'s without an X (or an implicit X1).

for i=3:size(filenum_nonx,1);  %start at 3 to avoid (./..)
nox_time=tic;
tifname=list(filenum_nonx(i),:);
dot_location=find(tifname=='.');
```

```
            if isletter(tifname(dot_location-1))==0
               outfile=[outdir 'rot_', tifname];
            end

            if isletter(tifname(dot_location-1))==1
               outfile =[outdir 'rot_' tifname(1:dot_location-1),'_0.tif'];
            end

  if strcmp('rot',tifname(1,1:3))==0  %don't process rot files (if you run this more than once).

    info=imfinfo(tifname);
    nFrames=numel(info); % grab number of frames
    disp([num2str(nFrames) ' frames in ' tifname]);

    %read each frame, rotate clockwise and upside-down, write to rot_file
       for frame=1:nFrames
            A(:,:) = flipud(rot90(imread(tifname, frame, 'Info',info),3));
            if frame==1;
               imwrite(A,outfile,'tif','Compression','none')
            else
               imwrite(A,outfile,'tif','Compression','none','WriteMode','append');
            end

       end
       disp([num2str(nFrames) '-frame rotated file written to ' outfile ])
       t(1)=toc(nox_time);
       disp(['Time elapsed= ' num2str(t) ' seconds.  ' num2str(size(filenum_nonx,1)-i) ' files
and ~' num2str((size(filenum_nonx,1)-i)*t/3600) ' hours left'])
       disp(' ');
   % delete(tifname);

  else
  end

end
t=toc(total_time);
disp(['Processed ' num2str(no_files-2) ' files in ' num2str(t./60) ' minutes']);
disp(' ');
clear A;
%% delete files part
beep2(990,0.8)

if auto_delete~='y';
    disp('Would you like to delete all of the original files?');
    reply=input('(y)es or (n)o? ','s');
    if isempty(reply)
        ans = 'n';
    end
    switch lower(reply)
    case {'y','yes','sure','ok','fine','why not?'}
        for i = 3:no_files
            if strcmp('rot',strtrim(list(i,1:3)))==0  %don't delete rot files if run twice.
              delete(strtrim(list(i,:)))
            else
             continue
            end
        end
    otherwise
    end

elseif auto_delete=='y'
     for i = 3:no_files
            if strcmp('rot',strtrim(list(i,1:3)))==0  %don't delete rot files if run twice.
              delete(strtrim(list(i,:)))
            else
             continue
            end
     end
end

end
```

APPENDIX B

CYANINE PARAMETERS

Cy3 and Cy5 CHARMM Topology files (can be streamed with CHARMM top file)

```
RESI CY5         +1.000 !
!!Will not use sulfonate groups
!GROUP            ! SULFONATE B
!ATOM SB     SG3O1   1.350 !    0.620
!ATOM OSB1   OG2P1  -0.651 !    0.000
!ATOM OSB2   OG2P1  -0.651 !    0.000
!ATOM OSB3   OG2P1  -0.651 !    0.000
!GROUP            ! SULFONATE A
!ATOM SA     SG3O1   1.350 !    0.620
!ATOM OSA4   OG2P1  -0.651 !    0.000
!ATOM OSA5   OG2P1  -0.651 !    0.000
!ATOM OSA6   OG2P1  -0.651 !    0.000


GROUP !from residue 3HIN, with charge adjustments from BSAT.
!(3H-indole with adjustments for benzenesulfonate, rather than explicitly using sulfonate).
ATOM N1A   NG2R52  -0.11 ! +0.50 added to this charge, atom type changed
ATOM C2A   CG2R52   0.29
ATOM C3A   CG3C50   0.09 !              H5
ATOM C4A   CG2RC0   0.25 !     \       / \\
ATOM C5A   CG2R61  -0.36 ! H32--C3---C4   C6--H6
ATOM H5A   HGR61    0.20 !      |    ||    |
ATOM C6A   CG2R61  -0.22 !      C2   C9   C7--H7
ATOM H6A   HGR61    0.21
ATOM C7A   CG2R61  -0.21 !   H2   N1    C8
ATOM H7A   HGR61    0.21 !              |
ATOM C8A   CG2R61  -0.34 !              H8
ATOM H8A   HGR61    0.26
ATOM C9A   CG2RC0   0.23

GROUP !common methyl group from other compounds
ATOM CM1A   CG331   -0.27 !
ATOM HM1A  HGA3     0.09 !
ATOM HM2A  HGA3     0.09 !
ATOM HM3A  HGA3     0.09 !
GROUP
ATOM CM2A   CG331   -0.27 !
ATOM HM4A  HGA3     0.09 !
ATOM HM5A  HGA3     0.09 !
ATOM HM6A  HGA3     0.09 !

GROUP
ATOM CM3B   CG331   -0.27 !
ATOM HM7B  HGA3     0.09 !
```

```
ATOM HM8B  HGA3     0.09 !
ATOM HM9B  HGA3     0.09 !

GROUP !this is the linker chain between the two rings, from 1,3,5-hexatriene, HEP3
ATOM CX1   CG2DC2 -0.15 !          /    \    /
ATOM HX1  HGA4    0.15 !        H31     C5=C6
GROUP
ATOM CX2   CG2DC2 -0.15 !               /    \
ATOM HX2  HGA4    0.15 !             H51     H62
GROUP
ATOM CX3   CG2DC2 -0.15 !               /    \
ATOM HX3  HGA4    0.15 !             H51     H62
GROUP
ATOM CX1A   CG2DC2 -0.15
ATOM HX1A   HGA4    0.15
GROUP
ATOM CX1B   CG2DC2 -0.15
ATOM HX1B   HGA4    0.15

GROUP !AROMATIC system B from residue 3HIN, with charge adjustments from BSAT
ATOM N1B   NG2R52  -0.11 ! +0.50 added to this charge, atom type changed
ATOM C2B   CG2R52   0.29
ATOM C3B   CG3C50   0.09 !              H5
ATOM C4B   CG2RC0   0.25 !     \       / \\
ATOM C5B   CG2R61  -0.36 ! H32--C3---C4   C6--H6
ATOM H5B   HGR61    0.20 !      |    ||    |
ATOM C6B   CG2R61  -0.22 !      C2   C9   C7--H7
ATOM H6B   HGR61    0.21
ATOM C7B   CG2R61  -0.21 !   H2   N1    C8
ATOM H7B   HGR61    0.21 !              |
ATOM C8B   CG2R61  -0.34 !              H8
ATOM H8B   HGR61    0.26
ATOM C9B   CG2RC0   0.23
GROUP
ATOM CM1B   CG331   -0.27 !
ATOM HM1B  HGA3     0.09 !
ATOM HM2B  HGA3     0.09 !
ATOM HM3B  HGA3     0.09 !
GROUP
ATOM CM2B   CG331   -0.27 !
ATOM HM4B  HGA3     0.09 !scrab
ATOM HM5B  HGA3     0.09 !
ATOM HM6B  HGA3     0.09 !


!GROUP !modified from spiriti. this is the linker chain to DNA, from n-propanol

GROUP !(the only unmodified group from Spiriti)
ATOM CL3B  CG324    -0.18  !this atom type changed b/c next to partially + nitrogen
ATOM H31B  HGA2      0.09
ATOM H32B  HGA2      0.09


!!!!!!!!!!!!!!!!!!!!
!MALEIMIDE LINKER

GROUP !this is the linker chain to DNA, from n-propanol
!ATOM OL1A  OG311    -0.65  !spiriti
!ATOM HO1A  HGP1      0.42  !spiriti
ATOM CL1A CG321 -0.18
!ATOM CL1A  CG321     0.05
ATOM H11A  HGA2      0.09
ATOM H12A  HGA2      0.09

GROUP
ATOM CL2A  CG321    -0.18
ATOM H21A  HGA2      0.09
ATOM H22A  HGA2      0.09

GROUP
ATOM CL3A   CG324    -0.18  !this atom type changed b/c next to partially + nitrogen
```

```
ATOM H31A  HGA2      0.09
ATOM H32A  HGA2      0.09

***new additions below**      PENALTY

GROUP
ATOM CL4    CG321  -0.18 !    0.228
ATOM H1     HGA2    0.090 !    0.000
ATOM H2     HGA2    0.090 !    0.000

GROUP
ATOM CL5    CG321  -0.18 !    0.427
ATOM H3     HGA2    0.090 !    0.000
ATOM H4     HGA2    0.090 !    0.000

GROUP !From Correy C37/041/N40/HAF
ATOM CL6    CG2O1   0.73 !    1.616
ATOM O3     OG2D1  -0.49 !    0.000
ATOM NL1    NG2S1  -0.58 !    3.147
ATOM H5     HGP1    0.34 !    0.563


GROUP
ATOM CL7    CG321  -0.18 !   11.279
ATOM H6     HGA2    0.090 !    0.850
ATOM H7     HGA2    0.090 !    0.850

GROUP
ATOM CL8    CG321  -0.18 !   15.198                              O1
ATOM H8     HGA2    0.090 !    0.835
ATOM H9     HGA2    0.090 !    0.835


GROUP !maleimide group from Correy et al
ATOM NL2    NG2R51 -0.20 !                                 |
ATOM CL9    CG2R53  0.58 !       H1  H3  O3      H6  H8    CL12
ATOM CL12   CG2R53  0.58 !       |   |   |       |   |    /    \
ATOM CL11   CG2R51 -0.18 !       CL4-CL5-CL6-NL1-CL7-CL8-NL2    CL11-H10
ATOM CL10   CG2R51 -0.18 !       |   |       |   |   |    \     /
ATOM O1     OG2D1  -0.48 !      H2  H4      H5  H7  H9    CL9-CL10-H11
ATOM O2     OG2D1  -0.48 !                                  |
ATOM H10    HGR51   0.18 !
ATOM H11    HGR51   0.18 !                                O2


!BOND SB   OSB1
!BOND SB   OSB2
!BOND SB   OSB3
!BOND SB   C6B
!BOND SA   OSA4
!BOND SA   OSA5
!BOND SA   OSA6
!BOND SA   C6A
!^NO sulfonate

!BOND CX2  CX1
BOND HX3  CX3
BOND HX2  CX2
BOND HX1  CX1
BOND HX1A CX1A
BOND HX1B CX1B

BOND CX1 CX1A
BOND CX1A CX1B
BOND CX1B CX2


BOND N1B  C9B
BOND N1B  CL3B
BOND CX3  CX2
BOND C9B  C8B
```

```
BOND C4B  C9B
BOND C4B  C5B
BOND C8B  H8B
BOND C8B  C7B
BOND C7B  H7B
BOND C6B  H6B
BOND C6B  C7B
BOND C5B  C6B
BOND C5B  H5B
BOND C3B  C4B
BOND C3B  CM1B
BOND C3B  CM2B
BOND CM1B HM4B
BOND CM1B HM5B
BOND CM1B HM6B
BOND CM2B HM1B
BOND CM2B HM2B
BOND CM2B HM3B
BOND C2B  N1B
BOND C2B  CX3
BOND C2B  C3B
BOND CL3B H31B
BOND CL3B H32B
BOND CL3B CM3B
BOND CM3B HM7B
BOND CM3B HM8B
BOND CM3B HM9B
BOND N1A  C9A
BOND N1A  CL3A
BOND CX1  C2A
BOND C9A  C4A
BOND C9A  C8A
BOND C4A  C5A
BOND C8A  H8A
BOND C8A  C7A
BOND C7A  H7A
BOND C7A  C6A
BOND C5A  C6A
BOND C6A  H6A
BOND C5A  H5A
BOND C3A  C4A
BOND C3A  CM2A
BOND C3A  CM1A
BOND CM2A HM4A
BOND CM2A HM5A
BOND CM2A HM6A
BOND CM1A HM1A
BOND CM1A HM2A
BOND CM1A HM3A
BOND C2A  N1A
BOND C2A  C3A
BOND CL3A CL2A
BOND CL2A CL1A
BOND CL1A CL4
BOND CL4  CL5
BOND CL5  CL6
BOND CL6  NL1
BOND CL6  O3
BOND NL1  CL7
BOND CL7  CL8
BOND CL8  NL2
BOND NL2  CL9
BOND NL2  CL12
BOND CL9  CL10
BOND CL9  O2
BOND CL12 CL11
BOND CL12 O1
BOND CL11 H10
BOND CL10 CL11
BOND CL10 H11
BOND H31A CL3A
```

```
BOND H32A CL3A
BOND H21A CL2A
BOND H22A CL2A
BOND H11A CL1A
BOND H12A CL1A
BOND H1   CL4
BOND H2   CL4
BOND H3   CL5
BOND H4   CL5
BOND H5   NL1
!BOND H6 NL1
BOND H6   CL7
BOND H7   CL7
BOND H8   CL8
BOND H9   CL8



IMPR CL6    CL5    NL1    O3   !used to maintain planarity
IMPR CL9    CL10   NL2    O2   ! used to maintain planarity
IMPR CL12   CL11   NL2    O1    !used to maintain planarity;

RESI CY3         +1.000 !
!!Will not use sulfonate groups
!GROUP            ! SULFONATE B
!ATOM SB     SG3O1   1.350 !    0.620
!ATOM OSB1   OG2P1  -0.651 !    0.000
!ATOM OSB2   OG2P1  -0.651 !    0.000
!ATOM OSB3   OG2P1  -0.651 !    0.000
!GROUP            ! SULFONATE A
!ATOM SA     SG3O1   1.350 !    0.620
!ATOM OSA4   OG2P1  -0.651 !    0.000
!ATOM OSA5   OG2P1  -0.651 !    0.000
!ATOM OSA6   OG2P1  -0.651 !    0.000


GROUP !from residue 3HIN, with charge adjustments from BSAT.
!(3H-indole with adjustments for benzenesulfonate, rather than explicitly using sulfonate).
ATOM N1A   NG2R52  -0.11 ! +0.50 added to this charge, atom type changed
ATOM C2A   CG2R52   0.29
ATOM C3A   CG3C50   0.09 !              H5
ATOM C4A   CG2RC0   0.25 !     \       / \\
ATOM C5A   CG2R61  -0.36 ! H32--C3---C4   C6--H6
ATOM H5A   HGR61    0.20 !      |    ||    |
ATOM C6A   CG2R61  -0.22 !      C2   C9   C7--H7
ATOM H6A   HGR61    0.21
ATOM C7A   CG2R61  -0.21 !   H2   N1    C8
ATOM H7A   HGR61    0.21 !              |
ATOM C8A   CG2R61  -0.34 !              H8
ATOM H8A   HGR61    0.26
ATOM C9A   CG2RC0   0.23

GROUP !common methyl group from other compounds
ATOM CM1A   CG331   -0.27 !
ATOM HM1A  HGA3     0.09 !
ATOM HM2A  HGA3     0.09 !
ATOM HM3A  HGA3     0.09 !
GROUP
ATOM CM2A   CG331   -0.27 !
ATOM HM4A  HGA3     0.09 !
ATOM HM5A  HGA3     0.09 !
ATOM HM6A  HGA3     0.09 !

GROUP
ATOM CM3B   CG331   -0.27 !
ATOM HM7B  HGA3     0.09 !
ATOM HM8B  HGA3     0.09 !
ATOM HM9B  HGA3     0.09 !

GROUP !this is the linker chain between the two rings, from 1,3,5-hexatriene, HEP3
ATOM CX1   CG2DC2 -0.15 !          /    \    /
```

```
ATOM HX1  HGA4    0.15 !        H31     C5=C6
GROUP
ATOM CX2   CG2DC2 -0.15 !               /    \
ATOM HX2  HGA4    0.15 !             H51     H62
GROUP
ATOM CX3   CG2DC2 -0.15 !               /    \
ATOM HX3  HGA4    0.15 !             H51     H62


GROUP !AROMATIC system B from residue 3HIN, with charge adjustments from BSAT
ATOM N1B   NG2R52  -0.11 ! +0.50 added to this charge, atom type changed
ATOM C2B   CG2R52   0.29
ATOM C3B   CG3C50   0.09 !              H5
ATOM C4B   CG2RC0   0.25 !     \       / \\
ATOM C5B   CG2R61  -0.36 ! H32--C3---C4   C6--H6
ATOM H5B   HGR61    0.20 !      |    ||    |
ATOM C6B   CG2R61  -0.22 !      C2   C9   C7--H7
ATOM H6B   HGR61    0.21
ATOM C7B   CG2R61  -0.21 !   H2   N1    C8
ATOM H7B   HGR61    0.21 !              |
ATOM C8B   CG2R61  -0.34 !              H8
ATOM H8B   HGR61    0.26
ATOM C9B   CG2RC0   0.23
GROUP
ATOM CM1B   CG331   -0.27 !
ATOM HM1B  HGA3     0.09 !
ATOM HM2B  HGA3     0.09 !
ATOM HM3B  HGA3     0.09 !
GROUP
ATOM CM2B   CG331   -0.27 !
ATOM HM4B  HGA3     0.09 !scrab
ATOM HM5B  HGA3     0.09 !
ATOM HM6B  HGA3     0.09 !


!GROUP !modified from spiriti. this is the linker chain to DNA, from n-propanol

GROUP !(the only unmodified group from Spiriti)
ATOM CL3B  CG324    -0.18  !this atom type changed b/c next to partially + nitrogen
ATOM H31B  HGA2      0.09
ATOM H32B  HGA2      0.09


!!!!!!!!!!!!!!!!!!!!
!MALEIMIDE LINKER

GROUP !this is the linker chain to DNA, from n-propanol
!ATOM OL1A  OG311    -0.65  !spiriti
!ATOM HO1A  HGP1      0.42  !spiriti
ATOM CL1A CG321 -0.18
!ATOM CL1A  CG321     0.05
ATOM H11A  HGA2      0.09
ATOM H12A  HGA2      0.09

GROUP
ATOM CL2A  CG321    -0.18
ATOM H21A  HGA2      0.09
ATOM H22A  HGA2      0.09

GROUP
ATOM CL3A   CG324    -0.18  !this atom type changed b/c next to partially + nitrogen
ATOM H31A  HGA2      0.09
ATOM H32A  HGA2      0.09

***new additions below**      PENALTY

GROUP
ATOM CL4    CG321  -0.18 !    0.228
ATOM H1     HGA2    0.090 !    0.000
ATOM H2     HGA2    0.090 !    0.000
```

```
GROUP
ATOM CL5    CG321  -0.18 !    0.427
ATOM H3     HGA2    0.090 !    0.000
ATOM H4     HGA2    0.090 !    0.000

GROUP !From Correy C37/041/N40/HAF
ATOM CL6    CG2O1   0.73 !    1.616
ATOM O3     OG2D1  -0.49 !    0.000
ATOM NL1    NG2S1  -0.58 !    3.147
ATOM H5     HGP1    0.34 !    0.563


GROUP
ATOM CL7    CG321  -0.18 !   11.279
ATOM H6     HGA2    0.090 !    0.850
ATOM H7     HGA2    0.090 !    0.850

GROUP
ATOM CL8    CG321  -0.18 !   15.198                              O1
ATOM H8     HGA2    0.090 !    0.835
ATOM H9     HGA2    0.090 !    0.835


GROUP !maleimide group from Correy et al
ATOM NL2    NG2R51 -0.20 !                                  |
ATOM CL9    CG2R53  0.58 !       H1  H3  O3      H6  H8    CL12
ATOM CL12   CG2R53  0.58 !       |   |   |       |   |    /    \
ATOM CL11   CG2R51 -0.18 !       CL4-CL5-CL6-NL1-CL7-CL8-NL2    CL11-H10
ATOM CL10   CG2R51 -0.18 !       |   |       |   |   |    \     /
ATOM O1     OG2D1  -0.48 !      H2  H4      H5  H7  H9    CL9-CL10-H11
ATOM O2     OG2D1  -0.48 !                                  |
ATOM H10    HGR51   0.18 !
ATOM H11    HGR51   0.18 !                                 O2


!BOND SB   OSB1
!BOND SB   OSB2
!BOND SB   OSB3
!BOND SB   C6B
!BOND SA   OSA4
!BOND SA   OSA5
!BOND SA   OSA6
!BOND SA   C6A
!^NO sulfonate

BOND CX2  CX1
BOND N1B  C9B
BOND N1B  CL3B
BOND CX3  CX2
BOND C9B  C8B
BOND C4B  C9B
BOND C4B  C5B
BOND C8B  H8B
BOND C8B  C7B
BOND C7B  H7B
BOND C6B  H6B
BOND C6B  C7B
BOND C5B  C6B
BOND C5B  H5B
BOND C3B  C4B
BOND C3B  CM1B
BOND C3B  CM2B
BOND CM1B HM4B
BOND CM1B HM5B
BOND CM1B HM6B
BOND CM2B HM1B
BOND CM2B HM2B
BOND CM2B HM3B
BOND C2B  N1B
BOND C2B  CX3
BOND C2B  C3B
```

```
BOND CL3B H31B
BOND CL3B H32B
BOND CL3B CM3B
BOND CM3B HM7B
BOND CM3B HM8B
BOND CM3B HM9B
BOND N1A  C9A
BOND N1A  CL3A
BOND CX1  C2A
BOND C9A  C4A
BOND C9A  C8A
BOND C4A  C5A
BOND C8A  H8A
BOND C8A  C7A
BOND C7A  H7A
BOND C7A  C6A
BOND C5A  C6A
BOND C6A  H6A
BOND C5A  H5A
BOND C3A  C4A
BOND C3A  CM2A
BOND C3A  CM1A
BOND CM2A HM4A
BOND CM2A HM5A
BOND CM2A HM6A
BOND CM1A HM1A
BOND CM1A HM2A
BOND CM1A HM3A
BOND C2A  N1A
BOND C2A  C3A
BOND CL3A CL2A
BOND CL2A CL1A
BOND CL1A CL4
BOND CL4  CL5
BOND CL5  CL6
BOND CL6  NL1
BOND CL6  O3
BOND NL1  CL7
BOND CL7  CL8
BOND CL8  NL2
BOND NL2  CL9
BOND NL2  CL12
BOND CL9  CL10
BOND CL9  O2
BOND CL12 CL11
BOND CL12 O1
BOND CL11 H10
BOND CL10 CL11
BOND CL10 H11
BOND H31A CL3A
BOND H32A CL3A
BOND H21A CL2A
BOND H22A CL2A
BOND H11A CL1A
BOND H12A CL1A
BOND H1   CL4
BOND H2   CL4
BOND H3   CL5
BOND H4   CL5
BOND H5   NL1
!BOND H6 NL1
BOND H6   CL7
BOND H7   CL7
BOND H8   CL8
BOND H9   CL8
BOND HX3  CX3
BOND HX2  CX2
BOND HX1  CX1
IMPR CL6    CL5    NL1    O3   !used to maintain planarity
IMPR CL9    CL10   NL2    O2   ! used to maintain planarity
IMPR CL12   CL11   NL2    O1
```

```
* CY3/CY5 parameters.
* Generated by G.Schauer from CGENFF/paramchem/iqbal/spiriti(!)/correy(!)
*
BONDS  !All were checked 3/22/11
CG2R51 CG2R53  360.00     1.4000 ! CY3_MAL_ , from CG2R51 CG2R52, penalty= 5  !CL9/Cl12
CG321  NG2R51  400.00     1.4580 ! CY3_MAL_ , from CG331 NG2R51, penalty= 6  !NL2-CL8
!^can't find CT2-NH2 in correy (?)
CG324  NG2R52  300.00     1.4330 ! RETINOL SCH2, Schiff's base, protonated #eq# !equilibrium dist
redcuced to agree better with iqbal
CG331  CG3C50  222.50     1.5280 ! TF2M, viv, methylcyclopentane
CG2RC0 CG3C50  305.00     1.5200 ! 3HIN, 3H-indole
CG2R52 CG3C50  350.00     1.5050 ! 2PRZ, 2-pyrazoline
CG2RC0 NG2R52  310.00     1.3650 ! NA G, adm jr. 11/97
CG2R52 CG2DC2  450.00     1.3000 ! mode 13, RETINOL 13DB, Butadiene @@@@@ Kenno: 1.47 --> 1.45
@@@@@

!**NEW bonds*** !bonds from NL1/NL2 (NG2S1/NG2R51) are missing b/c of improper/planar designation
CG321  NG2R52  263.00     1.4740 ! CY3_MAL_ , from CG321 NG311, penalty= 20
!CG321  NG3C51  263.00     1.4740 ! CY3_MAL_ , from CG321 NG311, penalty= 20

CG2R52 NG2R52  420.00     1.3550 ! CY3_MAL_ , from CG2R52 NG311, penalty= 20
!CG2R52 NG3C51  420.00     1.3550 ! CY3_MAL_ , from CG2R52 NG311, penalty= 20

!from PRES SAX
SM   CS    205.000     1.740

!CS CG2R53  360.00     1.4000 ! CY3_MAL_ , from CG2R51 CG2R52, penalty= 5  !CL9/Cl12
CG2R53   CS    305.000     1.52
!CC CS

CG2R51   CS    305.000     1.33
!CA CS
ANGLES
CG2DC2 CG2R52 CG3C50   10.00     123.50 !25.30    123.50 ! RETINOL MECH !vib calc, mode 9. from
spriti.  paramchem says 60, 122

CG2R52 NG2R52 CG2RC0   60.00    107.20 ! copied from 3HIN, 3H-indole !equilibrium angle changed
to agree better with iqbal
!^CG2R52 NG2R52 CG2RC0  145.00    108.00 ! CY3_MAL_ , from CG2R51 NG2R52 CG2R53, penalty= 2
!^^ iqbal's angle is 120.54

CG2R52 CG3C50 CG2RC0  105.00    105.00 ! 3HIN, 3H-indole, mode 13
CG2RC0 CG2RC0 CG3C50  143.00    110.00 ! 3HIN, 3H-indole, mode 13
CG2R61 CG2RC0 CG3C50   60.00    130.00 ! 3HIN, 3H-indole

CG3C50 CG2R52 NG2R52  170.00    112.00 !copied from 2PRZ, 2-pyrazoline; 3HPR, 3H-pyrrole N2-C3-C4
!^CG3C50 CG2R52 NG2R52   45.80    122.00 ! CY3_MAL_ , from CG321 CG2R51 NG2R52, penalty= 10

CG2RC0 NG2R52 CG324    62.30    112.30 ! PROT 107.5->120.0 to make planar Arg (KK) !modified
equilibrium
!^iqbal ~120

CG2R52 NG2R52 CG324    62.30    140.50 ! PROT 107.5->120.0 to make planar Arg (KK) !modified
equilibrium
!^CG2R52 NG2R52 CG324   101.00    111.90 ! CY3_MAL_ , from CG2R52 NG2R52 CG3C54, penalty= 1.2
!^iqbal ~120

CG2R61 CG2RC0 NG2R52  130.00    120.5 ! modified to match iqbal
!^spiriti CG2R61 CG2RC0 NG2R52  130.00    130.00 ! copied from ZIMI, benzimidazole

CG3C50 CG331 HGA3     33.43    110.10   22.53   2.17900 ! RETINOL TMCH/MECH
CG2DC2 CG2DC2 CG2DC2   17.30    123.00 ! RETINOL 13DP, Pentadiene @@@@@ Kenno: 123.5-->123.0
@@@@@ ! vib calc, mode 3
CG2DC2 CG2DC2 CG2R52   40.00    123.00 ! RETINOL 13DP, Pentadiene @@@@@ Kenno: 123.5-->123.0
@@@@@ ! vib calc, mode 3

CG2DC2 CG2R52 NG2R52   10.00    125.60 !12.40    125.60 ! RETINOL SCH3, Schiff's base, protonated
!vib calc, mode 9
```

```
!^iqbal says 129.04

CG2R52 CG3C50 CG331    32.00    112.20 ! RETINOL MECH
CG2RC0 CG3C50 CG331    32.00    112.20 ! RETINOL MECH
!^CG2RC0 CG3C50 CG331    65.00    108.20 ! CY3_MAL_ , from CG2RC0 CG3C52 CG3C52, penalty= 16

!CG331 CG3C52 CG331    58.35    113.50   11.16   2.561 ! RETINOL TMCH/MECH, also neopentane
CG331  CG3C50 CG331    58.35         113.50   11.16   2.561 ! RETINOL TMCH/MECH, also neopentane

CG2RC0 CG2RC0 NG2R52  100.00    105.70 ! CY3_MAL_ , from CG2RC0 CG2RC0 NG2R51, penalty= 25
!^spiriti says CG2RC0 CG2RC0 NG2R52  100.00    110.00 ! NA Ade 5R) bridgeC5... but above matches
iqubal angle

NG2R52 CG321  HGA2     33.43    110.10 ! NA FOR 9-M-G(C), adm jr.

NG2R52 CG324  HGA2     42.00    110.10 ! CY3_MAL_ , from NG2P1 CG324 HGA2, penalty= 2
!spiriti NG2R52 CG324  HGA2     33.43    110.10 ! NA FOR 9-M-G(C), adm jr.

CG2R52 CG2DC2 HGA4     25.50    119.00 ! RETINOL BTE2, 2-butene !vib calc

CG321  CG324  NG2R52   70.00    113.70 ! NA 9-E-G, adm jr. !this angle must be above 110.0 or
causes severe problems

!unused spiriti CG321  CG324  NG2R52   70.00    113.70 ! NA 9-E-G, adm jr. !this angle must be
above 110.0 or causes severe problems
!unused spiriti NG2R52 CG331  HGA3     33.43    110.10 ! NA FOR 9-M-G(C), adm jr.
****new angles*****
CG2DC2 CG2R52 NG2R52   10.00    125.60 !12.40    125.60 ! RETINOL SCH3, Schiff's base, protonated
!vib calc, mode 9
!CG2DC2 CG2R52 NG3C51   60.00    122.00 ! CY3_MAL_ , from CG2DC2 CG2R52 NG311, penalty= 3

!CG3C50 CG2R52 NG3C51   60.00    122.00 ! CY3_MAL_ , from CG2D1 CG2R52 NG311, penalty= 67
CG3C50 CG2R52 NG2R52  170.00    112.00 !copied from 2PRZ, 2-pyrazoline; 3HPR, 3H-pyrrole N2-C3-C4

CG2R51 CG2R51 CG2R53   90.00    106.00 ! CY3_MAL_ , from CG2R51 CG2R51 CG2R52, penalty= 1
CG2R53 CG2R51 HGR51    15.00    127.60 ! CY3_MAL_ , from CG2R52 CG2R51 HGR51, penalty= 1

CG2DC2 CG2R52 CG3C50   10.00     123.50 !25.30    123.50 ! RETINOL MECH !vib calc, mode 9
!CG2DC1 CG2R52 CG3C50  115.00    109.00 ! CY3_MAL_ , from CG2R51 CG2R51 CG3C52, penalty= 28.7

CG2DC2 CG2R52 NG2R52   10.00    125.60 !12.40    125.60 ! RETINOL SCH3, Schiff's base, protonated
!vib calc, mode 9
!CG2DC1 CG2R52 NG2R52  121.00    110.00 ! CY3_MAL_ , from CG2R51 CG2R52 NG2R52, penalty= 22.5


! Starting with some of the maleimide linkers from Correy:
CG2R51 CG2R53 NG2R51 70.00 105.60
!CG2R51 CG2R53 NG2R51  130.00    106.00 ! CY3_MAL_ , from CG2R51 CG2R51 NG2R51, penalty= 15
!NH2  CC   CA     70.000   105.6000  (correy)

CG2R51 CG2R53 OG2D1 80.00 131.50
!CG2R51 CG2R53 OG2D1    55.00    125.50 ! CY3_MAL_ , from CG2DC1 CG2R53 OG2D1, penalty= 22.5
!O    CC   CA     80.000   131.5000 (correy)

CG321  CG321  NG2R51   50.00    113.00
!CG321  CG321  NG2R51   70.00    113.50 ! CY3_MAL_ , from CG321 CG321 NG2S1, penalty= 11
!NH2  CT2  CT2    50.000   113.0000

CG321  CG321  NG2R52   43.70    112.20 ! CY3_MAL_ , from CG331 CG321 NG311, penalty= 3.9
!CG321  CG321  NG3C51   43.70    112.20 ! CY3_MAL_ , from CG331 CG321 NG311, penalty= 3.9

NG2R52 CG321  HGA2     33.43    110.10 ! NA FOR 9-M-G(C), adm jr.
!NG2R52 CG321  HGA2     32.40    109.50   50.00   2.13000 ! CY3_MAL_ , from NG311 CG321 HGA2,
penalty= 3
!NG3C51 CG321  HGA2     32.40    109.50   50.00   2.13000 ! CY3_MAL_ , from NG311 CG321 HGA2,
penalty= 3

CG331  CG324  NG2R52   67.70    110.00 ! CY3_MAL_ , from CG321 CG324 NG2P1, penalty= 2.9

CG2R52 CG3C50 CG2RC0  105.00    105.00 ! 3HIN, 3H-indole, mode 13
!CG2R52 CG3C50 CG2RC0   70.00    105.00 ! CY3_MAL_ , from CG2R52 CG3C52 CG2RC0, penalty= 10
```

```
CG2R52 CG3C50 CG2RC0  105.00    105.00 ! 3HIN, 3H-indole, mode 13
!CG2R52 CG3C50 CG331    80.00     99.00 ! CY3_MAL_ , from CG2R52 CG3C52 CG3C52, penalty= 16

CG2R53 NG2R51 CG2R53 110.0 111.43
!CG2R53 NG2R51 CG2R53  130.00    107.50 ! CY3_MAL_ , from CG2R51 NG2R51 CG2R53, penalty= 3
!CC   NH2  CC    110.000   111.4300 (correy)

CG2R53 NG2R51 CG321  50.00 123.20
!CG2R53 NG2R51 CG321    70.00    127.80 ! CY3_MAL_ , from CG2R53 NG2R51 CG331, penalty= 0.9
!CC   NH2  CT2    50.000   124.2000 (correy)

CG2R52 NG2R52 CG2RC0   60.00    107.20 ! copied from 3HIN, 3H-indole !equilibrium angle changed
to agree better with iqbal
!CG2R52 NG3C51 CG2RC0   40.00    109.00 ! CY3_MAL_ , from CG2R61 NG311 CG2R61, penalty= 42.5

CG2RC0  NG2R52  CG321    62.30    112.30 ! PROT 107.5->120.0 to make planar Arg (KK) !modified
equilibrium
!CG2RC0 NG3C51 CG321    60.00    106.90 ! CY3_MAL_ , from CG2RC0 NG3C51 CG3C52, penalty= 6

CG2R52  NG2R52  CG321    62.30    140.50 ! PROT 107.5->120.0 to make planar Arg (KK) !modified
equilibrium
!CG2R52 NG3C51 CG321    45.00    104.80 ! CY3_MAL_ , from CG2R51 NG3C51 CG3C52, penalty= 28.5

CG2R52 NG2R52 CG324   101.00    111.90 ! CY3_MAL_ , from CG2R52 NG2R52 CG3C54, penalty= 1.2
!CG2R52  NG2R52  CG324    62.30    140.50 ! PROT 107.5->120.0 to make planar Arg (KK) !modified
equilibrium

CG2RC0 NG2R52 CG324   101.00    111.90 ! CY3_MAL_ , from CG2R52 NG2R52 CG3C54, penalty= 4.2
!CG2RC0  NG2R52  CG324    62.30    112.30 ! PROT 107.5->120.0 to make planar Arg (KK) !modified
equilibrium

NG2R51 CG321 HGA2  44.000   107.9000
!NH2  CT2  HA     44.000   107.9000

!from PRES SAX
CS   SM   CT2    55.000   100.00

SM   CS   CG2R53     40.000   118.80
!SM   CS   CC     40.000   118.80

SM   CS   CG2R51     40.000   133.25
!SM   CS   CA     40.000   133.25

CS CG2R53 OG2D1   80.000   126.75
!CS   CC   O      80.000   126.75

CG2R51 CS CG2R53  52.000   108.00
!CA   CS   CC     52.000   108.00

CS CG2R53 NG2R51  70.000   106.12
!CS   CC   NH2    70.000   106.12

CG2R53 CG2R51 CS  52.000   108.56
!CS   CA   CC     52.000   108.56

CS CG2R51 HGR51   30.000   129.92
!CS   CA   HP     30.000   129.92

DIHEDRALS
!
!V(dihedral) = Kchi(1 + cos(n(chi) - delta))
!
!Kchi: kcal/mole
!n: multiplicity
!delta: degrees
!delta=180: minimum at 0, delta at 0, minimum at 180
!delta=180: minimum at 0, delta at 0, minimum at 180
!for mult = 2, delta at 180 = min at 0,180,  delta = 0 means min at 90

!the next two may be duplicative as sch2 has only one of these dihedrals
```

```
!I use !! to describe unneeeded dihedrals from Spiriti (GDS)

CG2DC2 CG2R52 NG2R52  CG324      1.0000  2   180.00 ! RETINOL SCH2, Schiff's base, protonated
!vib calc mode 4
!CG2DC2 CG2R52 NG2R52 CG324      6.0000  2   180.00 ! CY3_MAL_ , from CG2R51 CG2R52 NG2R52
CG3C54, penalty= 23.7
!CG2DC1 CG2R52 NG2R52 CG324      6.0000  2   180.00 ! CY3_MAL_ , from CG2R51 CG2R52 NG2R52
CG3C54, penalty= 23.7

CG2DC2 CG2R52 NG2R52  CG2RC0   1.5000  2   180.00 ! RETINOL SCH2, Schiff's base, protonated !vib
calc mode 4
!CG2DC2 CG2R52 NG3C51 CG2RC0     0.5000  2   180.00 ! CY3_MAL_ , from CG2DC2 CG2R52 NG311 CG2R52,
penalty= 44

HGA4   CG2DC2 CG2R52  NG2R52     0.5000  2   180.00 ! RETINOL SCH3, Schiff's base, protonated
!vib calc mode 4
!HGA4   CG2DC1 CG2R52 NG2R52     0.6000  2   180.00 ! CY3_MAL_ , from HGA4 CG2DC1 CG2R61 CG2R61,
penalty= 77


CG3C50  CG2R52  NG2R52  CG324    1.0000  2   180.00 ! RETINOL SCH2, Schiff's base, protonated
!vib calc mode 4
!CG3C50 CG2R52 NG2R52 CG324      6.0000  2   180.00 ! CY3_MAL_ , from CG2R51 CG2R52 NG2R52
CG3C54, penalty= 72.7

HGA2   CG324  NG2R52  CG2R52     0.1500  3   180.00 ! RETINOL SCH2, Schiff's base, protonated
!HGA2   CG324  NG2R52 CG2R52     0.1500  3   180.00 ! CY3_MAL_ , from HGA3 CG334 NG2P1 CG2D1,
penalty= 48

HGA2   CG324  NG2R52  CG2RC0     0.1500  3   180.00 ! RETINOL SCH2, Schiff's base, protonated
!HGA2   CG324  NG2R52 CG2RC0     0.1500  3   180.00 ! CY3_MAL_ , from HGA3 CG334 NG2P1 CG2D1,
penalty= 48

!from mech, retinol anlaogs
!double all constants for half mode 2

!CG3C50  CG2R52 CG2DC2 CG2DC2     0.9000  1     0.00 ! RETINOL MECH

CG3C50  CG2R52 CG2DC2 CG2DC2     2.1000  2   180.00 ! RETINOL MECH
!CG3C50 CG2R52 CG2DC2 CG2DC1     1.5000  1   180.00 ! CY3_MAL_ , from OG301 CG2R52 CG2DC2 CG2DC1,
penalty= 42

CG3C50  CG2R52 CG2DC2 CG2DC2     0.2200  3     0.00 ! RETINOL MECH
CG3C50  CG2R52 CG2DC2 CG2DC2     0.2500  5   180.00 ! RETINOL MECH
!CG3C50 CG2R52 CG2DC2 CG2DC1     1.5000  1   180.00 ! CY3_MAL_ , from OG301 CG2R52 CG2DC2 CG2DC1,
penalty= 42
!CG3C50 CG2R52 CG2DC2 CG2DC1    15.0000  2   180.00 ! CY3_MAL_ , from OG301 CG2R52 CG2DC2 CG2DC1,
penalty= 42


CG3C50  CG2R52 CG2DC2 CG2DC2     0.1000  6     0.00 ! RETINOL MECH
!CG3C50  CG2R52 CG2DC2 CG2DC2     0.7000  1     0.00 ! RETINOL MECH
!CG3C50  CG2R52 CG2DC2 CG2DC2     1.7000  2   180.00 ! RETINOL MECH
!CG3C50  CG2R52 CG2DC2 CG2DC2     0.1700  3     0.00 ! RETINOL MECH
!CG3C50  CG2R52 CG2DC2 CG2DC2     0.2000  5   180.00 ! RETINOL MECH
!CG3C50  CG2R52 CG2DC2 CG2DC2     0.0800  6     0.00 ! RETINOL MECH

CG3C50  CG2R52 CG2DC2 HGA4       0.5000  2   180.00 ! RETINOL MECH !vib calc mode 4
!CG3C50 CG2R52 CG2DC2 HGA4       3.0000  2   180.00 ! CY3_MAL_ , from OG301 CG2R52 CG2DC2 HGA4,
penalty= 42

NG2R52 CG2R52 CG3C50  CG331      0.5000  2     0.00 ! RETINOL TMCH !vib calc mode 4
NG2R52 CG2R52 CG3C50  CG331      0.4000  3     0.00 ! RETINOL TMCH !vib calc mode 4
!NG2R52 CG2R52 CG3C50 CG331      2.8000  3   180.00 ! CY3_MAL_ , from NG2R50 CG2R52 CG3C52
CG3C52, penalty= 42

CG2DC2 CG2R52 CG3C50  CG331      0.3000  3     0.00 ! RETINOL MECH !vib calc mode 4
!CG2DC2 CG2R52 CG3C50 CG331      0.1000  2   180.00 ! CY3_MAL_ , from CG2DC2 CG2R52 NG301 CG3C51,
penalty= 306.4
```

```
CG2RC0 CG2RC0 CG3C50  CG331      0.5000  2     0.00 ! RETINOL TMCH !this one causes "almost
linear" errors
CG2RC0 CG2RC0 CG3C50  CG331      0.4000  3     0.00 ! RETINOL TMCH !this one causes "almost
linear" errors
!CG2RC0 CG2RC0 CG3C50 CG331      1.0300  3   180.00 ! CY3_MAL_ , from CG2RC0 CG2RC0 CG3C52
CG3C52, penalty= 16

CG2R61 CG2RC0 CG3C50  CG331      0.5000  2     0.00 ! RETINOL TMCH !half mode 2
CG2R61 CG2RC0 CG3C50  CG331      0.4000  3     0.00 ! RETINOL TMCH !half mode 2
!CG2R61 CG2RC0 CG3C50 CG331      3.0000  2   180.00 ! CY3_MAL_ , from CG2R61 CG2RC0 CG3C52
CG3C52, penalty= 16

CG2R52 CG3C50  CG331  HGA3       0.1600  3     0.00 ! RETINOL TMCH
!CG2R52 CG3C50  CG331  HGA3       0.1600  3     0.00 ! RETINOL TMCH
!HGA3   CG331  CG3C50 CG2R52     0.1600  3     0.00 ! CY3_MAL_ , from CG2D1 CG321 CG331 HGA3,
penalty= 68

CG2RC0 CG3C50  CG331  HGA3       0.1600  3     0.00 ! RETINOL TMCH
!CG2RC0 CG3C50  CG331  HGA3       0.1600  3     0.00 ! RETINOL TMCH
!HGA3   CG331  CG3C50 CG2RC0     0.1600  3     0.00 ! CY3_MAL_ , from CG2R51 CG321 CG331 HGA3,
penalty= 66

HGA3   CG331  CG3C50 CG331      0.1600  3     0.00 ! CY3_MAL_ , from HGA3 CG331 CG3C51 CG3C52,
penalty= 12
!CG331  CG3C50  CG331  HGA3       0.1600  3     0.00 ! RETINOL TMCH
!HGA3   CG331  CG3C50 CG331      0.1600  3     0.00 ! CY3_MAL_ , from HGA3 CG331 CG3C51 CG3C52,
penalty= 12

CG2DC2 CG2R52 CG3C50  CG2RC0      0.0000  3     0.00 ! RETINOL MECH !vib calc, mode 4, the 5
member ring is more rigid than the 6 member.
!CG2DC2 CG2R52 CG3C50 CG2RC0     0.5000  2   180.00 ! CY3_MAL_ , from CG2DC2 CG2R52 NG311 CG2R52,
penalty= 319

!from 3HIN
NG2R52 CG2R52 CG3C50 CG2RC0     1.3000  3   180.00 ! 3HIN, 3H-indole !vib calc, mode 4-5
!NG2R52 CG2R52 CG3C50 CG2RC0     3.5000  3   180.00 ! CY3_MAL_ , from NG2R50 CG2R52 CG3C52
CG2RC0, penalty= 36

CG3C50 CG2R52 NG2R52 CG2RC0     6.0000  2   180.00 ! 3HIN, 3H-indole !vib calc, mode 4-5
!CG3C50 CG2R52 NG2R52 CG2RC0     2.5000  2   180.00 ! CY3_MAL_ , from CG321 CG2R51 NG2R52 CG2R53,
penalty= 14

CG2R61 CG2R61 CG2RC0 CG3C50     0.0000  2   180.00 ! 3HIN, 3H-indole
!CG2R61 CG2R61 CG2RC0 CG3C50     0.0000  2   180.00 ! CY3_MAL_ , from CG2R61 CG2R61 CG2RC0
CG3C52, penalty= 1.2

CG2R61 CG2R61 CG2RC0 NG2R52     1.5000  2   180.00 ! ZIMI, benzimidazole
!CG2R61 CG2R61 CG2RC0 NG2R52     3.0000  2   180.00 ! CY3_MAL_ , from CG2R61 CG2R61 CG2RC0
NG2R51, penalty= 25

CG2R61 CG2RC0 CG2RC0 CG3C50     2.5000  2   180.00 ! 3HIN, 3H-indole !vib calc mode 5
!CG2R61 CG2RC0 CG2RC0 CG3C50     6.5000  2   180.00 ! CY3_MAL_ , from CG2R61 CG2RC0 CG2RC0
CG3C52, penalty= 1.2

CG2RC0 CG2RC0 CG3C50 CG2R52     0.5000  3   180.00 ! 3HIN, 3H-indole !vib calc mode 5
!CG2RC0 CG2RC0 CG3C50 CG2R52     1.0000  3   180.00 ! CY3_MAL_ , from CG2RC0 CG2RC0 CG3C52
CG2R51, penalty= 32.5

CG2R61 CG2RC0 NG2R52 CG2R52     4.0000  2   180.00 ! 3HIN, 3H-indole
!CG2R61 CG2RC0 NG2R52 CG2R52    12.0000  2   180.00 ! CY3_MAL_ , from CG2R51 CG2R51 NG2R52
CG2R53, penalty= 40

CG2R61 CG2RC0 CG3C50 CG2R52     3.5000  3     0.00 ! 3HIN, 3H-indole
!CG2R61 CG2RC0 CG3C50 CG2R52     0.9000  3     0.00 ! CY3_MAL_ , from CG2R61 CG2RC0 CG3C52
CG2R51, penalty= 32.5
!CG2R61 CG2RC0 CG3C50 CG2R52     3.5000  3     0.00 ! CY3_MAL_ , from CG2R61 CG2RC0 CG3C52
CG2R52, penalty= 10

CG2RC0 CG2RC0 NG2R52 CG2R52     4.0000  2   180.00 ! 3HIN, 3H-indole !half mode 2
!CG2RC0 CG2RC0 NG2R52 CG2R52    12.0000  2   180.00 ! CY3_MAL_ , from CG2R51 CG2R51 NG2R52
CG2R53, penalty= 32
```

```
CG3C50 CG2RC0 CG2RC0 NG2R52     6.5000  2   180.00 ! 3HIN, 3H-indole !vib calc mode 4-5
!CG3C50 CG2RC0 CG2RC0 NG2R52     6.5000  2   180.00 ! CY3_MAL_ , from CG3C52 CG2RC0 CG2RC0
NG2R50, penalty= 27.2

CG2R61 CG2RC0 CG2RC0 NG2R52     1.5000  2   180.00 ! ZIMI, benzimidazole
!CG2R61 CG2RC0 CG2RC0 NG2R52     1.5000  2   180.00 ! CY3_MAL_ , from CG2R61 CG2RC0 CG2RC0
NG2R51, penalty= 25

HGR61  CG2R61 CG2RC0 NG2R52     0.8000  2   180.00 ! ZIMI, benzimidazole
!HGR61  CG2R61 CG2RC0 NG2R52     3.0000  2   180.00 ! CY3_MAL_ , from HGR61 CG2R61 CG2RC0 NG2R51,
penalty= 25

HGR61  CG2R61 CG2RC0 CG3C50     0.0000  2   180.00 ! 3HIN, 3H-indole
!HGR61  CG2R61 CG2RC0 CG3C50     0.0000  2   180.00 ! CY3_MAL_ , from HGR61 CG2R61 CG2RC0 CG3C52,
penalty= 1.2

!9-ethyl-guanine
CG2RC0 CG2RC0 NG2R52 CG324      9.0000  2   180.00 ! 9MAD, 9-Methyl-Adenine, kevo for gsk/ibm,
half mode 4
!CG2RC0 CG2RC0 NG2R52 CG324      6.0000  2   180.00 ! CY3_MAL_ , from CG2R51 CG2R52 NG2R52
CG3C54, penalty= 37.2

CG2R61 CG2RC0 NG2R52 CG324      9.0000  2   180.00 ! 9MAD, 9-Methyl-Adenine, kevo for gsk/ibm,
half mode 4
!CG2R61 CG2RC0 NG2R52 CG324      6.0000  2   180.00 ! CY3_MAL_ , from CG2R51 CG2R52 NG2R52
CG3C54, penalty= 45.2

!others

HGA3   CG331  CG3C50 CG331      0.1600  3     0.00 ! CY3_MAL_ , from HGA3 CG331 CG3C51 CG3C52,
penalty= 12
!CG331  CG3C50  CG331  HGA3       0.1600  3     0.00 ! RETINOL TMCH

NG2R52 CG324  CG331  HGA3       0.1950  3     0.00 ! CY3_MAL_ , from HGA2 CG321 CG324 NG2P1,
penalty= 8

CG2RC0 CG2RC0 CG3C50 CG2R52     0.5000  3   180.00 ! 3HIN, 3H-indole !vib calc mode 5
!CG2RC0 CG2RC0 CG3C50 CG2R52     2.0000  3   180.00 ! CY3_MAL_ , from CG2RC0 CG2RC0 CG3C52
CG2R52, penalty= 10

!to match vib freqs I had to make these so soft, might get isomerization
CG2R52 CG2DC2 CG2DC2 CG2DC2     0.5600  1   180.00 ! RETINOL HEP3, 1,3,5-heptatriene
CG2R52 CG2DC2 CG2DC2 CG2DC2     4.0000  2   180.00 ! RETINOL HEP3, 1,3,5-heptatriene !mode 1-2,
vib calc
!CG2DC2 CG2DC1 CG2DC1 CG2R52     0.5600  1   180.00 ! CY3_MAL_ , from CG2DC2 CG2DC1 CG2DC1
CG2DC2, penalty= 22.5
!CG2DC2 CG2DC1 CG2DC1 CG2R52     7.0000  2   180.00 ! CY3_MAL_ , from CG2DC2 CG2DC1 CG2DC1
CG2DC2, penalty= 22.5

!use the same one to make about C2A-CX1 for more double bond character
NG2R52 CG2R52 CG2DC2 CG2DC2     0.5600  1   180.00 ! RETINOL HEP3, 1,3,5-heptatriene
NG2R52 CG2R52 CG2DC2 CG2DC2     9.0000  2   180.00 ! RETINOL HEP3, 1,3,5-heptatriene !mode 1-2,
vib calc
!CG2DC1 CG2DC1 CG2R52 NG2R52     3.0000  2   180.00 ! CY3_MAL_ , from CG2D1O CG2DC1 CG2R53
NG2R51, penalty= 38

!All 4 bonds across the linker are as stiff as the central bond in 1,3,5-hexatriene.
!This probably makes the linker too stiff but it's better than being too floppy in this case.
CG2DC2 CG2DC2 CG2DC2 HGA4       5.2000  2   180.00 ! RETINOL 13DB, 1,3-Butadiene !modes 1-2, vib
calc
CG2R52 CG2DC2 CG2DC2 HGA4       5.2000  2   180.00 ! RETINOL 13DB, 1,3-Butadiene !modes 1-2, vib
calc
!CG2R52 CG2DC1 CG2DC1 HGA4       5.2000  2   180.00 ! CY3_MAL_ , from CG2DC2 CG2DC1 CG2DC1 HGA4,
penalty= 22.5

CG2RC0 CG2R61 CG2R61 SG3O1      3.1000  2   180.00 ! based on toluene, adm jr., 3/7/92
!CG2RC0 CG2R61 CG2R61 SG3O1      3.1000  2   180.00 ! CY3_MAL_ , from CG2R61 CG2R61 CG2R61 SG3O1,
penalty= 1.5
!spiriti didn't use S03
```

```
!arginine
!!HGA2   CG321  CG324  NG2R52     0.1950  3     0.00 ! PROT alkane update, adm jr., 3/2/92

!HGA2   CG324  NG2R52  CG2R52      0.0000  6   180.00 ! PROT methylguanidinium, adm jr., 3/26/92
!HGA2   CG324  NG2R52  CG2RC0   0.0000  6   180.00 ! PROT methylguanidinium, adm jr., 3/26/92

!These are especially important as they are along the angles we wish to study.
!!CG321  CG321  CG324  NG2R52     0.1950  3     0.00 ! PROT alkane update, adm jr., 3/2/92
(arginine)
!!CG321  CG324  NG2R52  CG2R52         2.2500  1     0.00 ! PROT methylguanidinium, adm jr.,
3/26/92  !B3LYP/6-31+g(d,p) dihedral scan
!!CG321  CG324  NG2R52  CG2R52         0.9000  2   180.00 ! PROT methylguanidinium, adm jr.,
3/26/92
!!CG321  CG324  NG2R52  CG2RC0         2.2500  1   180.00 ! PROT methylguanidinium, adm jr.,
3/26/92
!!CG321  CG324  NG2R52  CG2RC0         0.9000  2   180.00 ! PROT methylguanidinium, adm jr.,
3/26/92

!these are not from arginine but also along the linker.
!!CG324  CG321  CG321  OG303      0.1950  3     0.00 ! PROT alkane update, adm jr., 3/2/92
!!CG324  CG321  CG321  OG311      0.1950  3     0.00 ! PROT alkane update, adm jr., 3/2/92
!!CG324  CG321  CG321  ON2        0.1950  3     0.00 ! PROT alkane update, adm jr., 3/2/92

!NG3C51=NG2R52

CG2RC0 CG2RC0 NG2R52 CG321      1.9600  3     0.00 ! CY3_MAL_ , from CG2RC0 CG2RC0 NG3C51 CG3C52,
penalty= 6
!CG2RC0 CG2RC0 NG3C51 CG321      1.9600  3     0.00 ! CY3_MAL_ , from CG2RC0 CG2RC0 NG3C51
CG3C52, penalty= 6

CG2R61 CG2RC0 NG2R52 CG321      4.0000  2   180.00 ! CY3_MAL_ , from CG2R61 CG2RC0 NG3C51 CG3C52,
penalty= 6
!CG2R61 CG2RC0 NG2R52 CG324      9.0000  2   180.00 ! 9MAD, 9-Methyl-Adenine, kevo for gsk/ibm,
half mode 4

CG321  CG321  CG321  NG2R52     1.0000  3     0.00 ! CY3_MAL_ , from CG321 CG321 CG324 NG3P2,
penalty= 16
!CG321  CG321  CG324  NG2R52     0.1950  3     0.00 ! PROT alkane update, adm jr., 3/2/92
(arginine)
!CG321  CG321  CG321  NG3C51     1.0000  3     0.00 ! CY3_MAL_ , from CG321 CG321 CG324 NG3P2,
penalty= 16

CG3C50 CG2RC0 CG2RC0 NG2R52     6.0000  2   180.00 ! CY3_MAL_ , from CG3C52 CG2RC0 CG2RC0 NG3C51,
penalty= 1.2
!CG3C50 CG2RC0 CG2RC0 NG3C51     6.0000  2   180.00 ! CY3_MAL_ , from CG3C52 CG2RC0 CG2RC0
NG3C51, penalty= 1.2

CG2DC2 CG2R52 NG2R52  CG2RC0   1.5000  2   180.00 ! RETINOL SCH2, Schiff's base, protonated !vib
calc mode 4
!CG2DC2 CG2R52 NG3C51 CG2RC0     0.5000  2   180.00 ! CY3_MAL_ , from CG2DC2 CG2R52 NG311 CG2R52,
penalty= 44
!CG2DC2 CG2R52 NG2R52  CG2RC0  1.5000  2   180.00 ! RETINOL SCH2, Schiff's base, protonated !vib
calc mode 4

CG2DC2 CG2R52 NG2R52 CG321       1.0000  2   180.00
!CG2DC2 CG2R52 NG3C51 CG321      0.1000  2   180.00 ! CY3_MAL_ , from CG2DC2 CG2R52 NG301 CG3C51,
penalty= 31.4
!CG2DC2 CG2R52 NG2R52  CG324      1.0000  2   180.00 ! RETINOL SCH2, Schiff's base, protonated
!vib calc mode 4

CG3C50 CG2R52 NG2R52 CG2RC0     6.0000  2   180.00 ! 3HIN, 3H-indole !vib calc, mode 4-5
!CG3C50 CG2R52 NG3C51 CG2RC0     0.1000  2   180.00 ! CY3_MAL_ , from CG2D1 CG2R52 NG311 CG2R52,
penalty= 108
!CG3C50 CG2R52 NG2R52 CG2RC0     6.0000  2   180.00 ! 3HIN, 3H-indole !vib calc, mode 4-5

CG3C50  CG2R52  NG2R52  CG324    1.0000  2   180.00 ! RETINOL SCH2, Schiff's base, protonated
!vib calc mode 4
!CG3C50 CG2R52 NG3C51 CG321      0.1000  2   180.00 ! CY3_MAL_ , from CG2D1 CG2R52 NG301 CG3C51,
penalty= 95.4
!CG3C50  CG2R52  NG2R52  CG324    1.0000  2   180.00 ! RETINOL SCH2, Schiff's base, protonated
!vib calc mode 4
```

```
HGA2   CG324  NG2R52  CG2R52     0.1500  3   180.00 ! RETINOL SCH2, Schiff's base, protonated
!HGA2   CG321  NG3C51 CG2R52     0.0000  3   180.00 ! CY3_MAL_ , from HGA3 CG331 NG311 CG2N1,
penalty= 58
!HGA2   CG324  NG2R52  CG2R52     0.1500  3   180.00 ! RETINOL SCH2, Schiff's base, protonated

HGA2   CG324  NG2R52  CG2RC0     0.1500  3   180.00 ! RETINOL SCH2, Schiff's base, protonated
!HGA2   CG321  NG3C51 CG2RC0     0.0000  3   180.00 ! CY3_MAL_ , from HGA3 CG331 NG311 CG2N1,
penalty= 59
!HGA2   CG324  NG2R52  CG2RC0     0.1500  3   180.00 ! RETINOL SCH2, Schiff's base, protonated

CG2R61 CG2RC0 NG2R52 CG2R52     4.0000  2   180.00 ! 3HIN, 3H-indole
!CG2R61 CG2RC0 NG3C51 CG2R52     0.4400  2     0.00 ! CY3_MAL_ , from CG2R61 CG2R61 NG311 CG2R61,
penalty= 61!
!CG2R61 CG2RC0 NG2R52 CG2R52     4.0000  2   180.00 ! 3HIN, 3H-indole

CG2RC0 CG2RC0 NG2R52 CG2R52     4.0000  2   180.00 ! 3HIN, 3H-indole !half mode 2
!CG2RC0 CG2RC0 NG3C51 CG2R52     0.4400  2     0.00 ! CY3_MAL_ , from CG2R61 CG2R61 NG311 CG2R61,
penalty= 62.5
!CG2RC0 CG2RC0 NG2R52 CG2R52     4.0000  2   180.00 ! 3HIN, 3H-indole !half mode 2

NG3C51 CG321  CG321  HGA2       0.1600  3     0.00 ! CY3_MAL_ , from NG311 CG321 CG331 HGA3,
penalty= 9

NG2R52 CG2R52 CG2DC2 CG2DC2     0.5600  1   180.00 ! RETINOL HEP3, 1,3,5-heptatriene
!NG3C51 CG2R52 CG2DC2 CG2DC1     2.5000  2   180.00 ! CY3_MAL_ , from NG311 CG2R52 CG2DC2 CG2O1,
penalty= 32
!NG2R52 CG2R52 CG2DC2 CG2DC2     0.5600  1   180.00 ! RETINOL HEP3, 1,3,5-heptatriene

NG2R52 CG2R52 CG3C50 CG2RC0     1.3000  3   180.00 ! 3HIN, 3H-indole !vib calc, mode 4-5
!NG3C51 CG2R52 CG3C50 CG2RC0     0.1000  2   180.00 ! CY3_MAL_ , from CG2D1 CG2R52 NG311 CG2R52,
penalty= 383
!NG2R52 CG2R52 CG3C50 CG2RC0     1.3000  3   180.00 ! 3HIN, 3H-indole !vib calc, mode 4-5

NG2R52 CG2R52 CG3C50  CG331      0.5000  2     0.00 ! RETINOL TMCH !vib calc mode 4
!NG3C51 CG2R52 CG3C50 CG331      0.1000  2   180.00 ! CY3_MAL_ , from CG2D1 CG2R52 NG301 CG3C51,
penalty= 370.4
!NG2R52 CG2R52 CG3C50  CG331      0.5000  2     0.00 ! RETINOL TMCH !vib calc mode 4

CG321  CG321  NG2R52  CG2R52    2.2500  1     0.00 ! PROT methylguanidinium, adm jr., 3/26/92
!B3LYP/6-31+g(d,p) dihedral scan
!CG321  CG321  NG3C51 CG2R52     0.0500  3   180.00 ! CY3_MAL_ , from CG3C52 CG3C52  CG2R51,
penalty= 108.5
!CG321  CG324  NG2R52  CG2R52   2.2500  1     0.00 ! PROT methylguanidinium, adm jr., 3/26/92
!B3LYP/6-31+g(d,p) dihedral scan

CG321  CG321  NG2R52  CG2RC0     0.9000  2   180.00 ! PROT methylguanidinium, adm jr., 3/26/92
!CG321  CG324  NG2R52  CG2RC0    0.9000  2   180.00 ! PROT methylguanidinium, adm jr., 3/26/92
!CG321  CG321  NG3C51 CG2RC0     1.4500  3   180.00 ! CY3_MAL_ , from CG3C52 CG3C52 NG3C51
CG2RC0, penalty= 86

!!NG3C51 CG2R52 CG2DC2 HGA4       3.0000  2   180.00 ! CY3_MAL_ , from OG301 CG2R52 CG2DC2 HGA4,
penalty= 33
!!duplicate of HGA4 CG2DC2 CG2R52  NG2R52     0.5000  2   180.00 ! RETINOL SCH3, Schiff's base,
protonated !vib calc mode 4

!CG2DC1=CG2DC2

HGA4   CG2DC1 CG2R52 CG3C50      0.5000  2   180.00
!HGA4   CG2DC1 CG2R52 CG3C50     0.6000  2   180.00 ! CY3_MAL_ , from HGA4 CG2DC1 CG2R61 CG2R61,
penalty= 110
!CG3C50  CG2R52 CG2DC2 HGA4       0.5000  2   180.00 ! RETINOL MECH !vib calc mode 4

CG2DC2 CG2R52 CG3C50  CG331      0.3000  3     0.00 ! RETINOL MECH !vib calc mode 4
!CG2DC2 CG2R52 CG3C50  CG331      0.3000  3     0.00 ! RETINOL MECH !vib calc mode 4
!CG2DC1 CG2R52 CG3C50 CG331      0.0500  3   180.00 ! CY3_MAL_ , from CG2R51 CG2R51 CG3C52
CG3C52, penalty= 43.5

CG2DC2 CG2R52 CG3C50  CG2RC0      0.0000  3     0.00 ! RETINOL MECH !vib calc, mode 4, the 5
member ring is more rigid than the 6 member.
```

```
!CG2DC1 CG2R52 CG3C50 CG2RC0     1.5000  3   180.00 ! CY3_MAL_ , from CG2R51 CG2R51 CG3C52
CG2RC0, penalty= 37.5
!CG2DC2 CG2R52 CG3C50  CG2RC0      0.0000  3     0.00 ! RETINOL MECH !vib calc, mode 4, the 5
member ring is more rigid than the 6 member.

CG3C50  CG2R52 CG2DC2 CG2DC2     0.9000  1     0.00 ! RETINOL MECH
CG3C50  CG2R52 CG2DC2 CG2DC2     2.1000  2   180.00 ! RETINOL MECH
CG3C50  CG2R52 CG2DC2 CG2DC2     0.2200  3     0.00 ! RETINOL MECH
CG3C50  CG2R52 CG2DC2 CG2DC2     0.2500  5   180.00 ! RETINOL MECH
CG3C50  CG2R52 CG2DC2 CG2DC2     0.1000  6     0.00 ! RETINOL MECH
!CG2DC1 CG2DC1 CG2R52 CG3C50     3.0000  2   180.00 ! CY3_MAL_ , from CG2D1O CG2DC1 CG2R53
NG2R51, penalty= 83

NG2R52 CG2R52 CG2DC2 CG2DC2     9.0000  2   180.00 ! RETINOL HEP3, 1,3,5-heptatriene !mode 1-2,
vib calc
!CG2DC1 CG2DC1 CG2R52 NG2R52     3.0000  2   180.00 ! CY3_MAL_ , from CG2D1O CG2DC1 CG2R53
NG2R51, penalty= 38
!NG2R52 CG2R52 CG2DC2 CG2DC2     9.0000  2   180.00 ! RETINOL HEP3, 1,3,5-heptatriene !mode 1-2,
vib calc

!duplicate CG2DC2 CG2R52 NG2R52  CG324      1.0000  2   180.00 ! RETINOL SCH2, Schiff's base,
protonated !vib calc mode 4
!CG2DC1 CG2R52 NG2R52 CG324      6.0000  2   180.00 ! CY3_MAL_ , from CG2R51 CG2R52 NG2R52
CG3C54, penalty= 23.7
!CG2DC2 CG2R52 NG2R52  CG324      1.0000  2   180.00 ! RETINOL SCH2, Schiff's base, protonated
!vib calc mode 4


CG2R52 CG2DC2 CG2DC2 CG2DC2     0.5600  1   180.00 ! RETINOL HEP3, 1,3,5-heptatriene
CG2R52 CG2DC2 CG2DC2 CG2DC2     4.0000  2   180.00 ! RETINOL HEP3, 1,3,5-heptatriene !mode 1-2,
vib calc
!CG2DC2 CG2DC1 CG2DC1 CG2R52     0.5600  1   180.00 ! CY3_MAL_ , from CG2DC2 CG2DC1 CG2DC1
CG2DC2, penalty= 22.5
!CG2DC2 CG2DC1 CG2DC1 CG2R52     7.0000  2   180.00 ! CY3_MAL_ , from CG2DC2 CG2DC1 CG2DC1
CG2DC2, penalty= 22.5

CG321  CG321  CG324  NG2R52     0.1950  3     0.00 ! PROT alkane update, adm jr., 3/2/92
(arginine)

HGA2   CG321  CG324  NG2R52     0.1950  3     0.00 ! PROT alkane update, adm jr., 3/2/92

CG321  CG324  NG2R52  CG2R52    2.2500  1     0.00 ! PROT methylguanidinium, adm jr., 3/26/92
!B3LYP/6-31+g(d,p) dihedral scan

CG321  CG324  NG2R52  CG2RC0    2.2500  1   180.00 ! PROT methylguanidinium, adm jr., 3/26/92
CG321  CG324  NG2R52  CG2RC0    0.9000  2   180.00 ! PROT methylguanidinium, adm jr., 3/26/92

!matched to correy
HGA2   CG321  NG2R51 CG2R53     0.0000  3     0.00 ! CY3_MAL_ , from HGA3 CG331 NG2R51 CG2R53,
penalty= 6
!HA   CT2  NH2  CC       0.0000  3      0.00

CG2R53 CG2R51 CG2R51 HGR51       4.200   2    180.00
!CG2R53 CG2R51 CG2R51 HGR51      2.6000  2   180.00 ! CY3_MAL_ , from CG2R52 CG2R51 CG2R51 HGR51,
penalty= 1
!CC   CA   CA   HP       4.200   2    180.00

CG2R53 CG2R51 CG2R51 CG2R53     3.1000   2   180.00
!CG2R53 CG2R51 CG2R51 CG2R53    15.0000  2   180.00 ! CY3_MAL_ , from CG2R51 CG2R51 CG2R51
CG2R51, penalty= 6
!CC   CA   CA   CC       3.1000  2    180.00

CG2R51 CG2R53 NG2R51 CG2R53     3.1      2   180.00
!CC   NH2  CC   CA              3.1000  2    180.00

CG2R51 CG2R53 NG2R51 CG321      0.0000  1     0.00 ! CY3_MAL_ , from CG2R51 CG2R51 NG2R51 CG3C51,
penalty= 21.4
!CT2  NH2  CC   CA              2.5000  2    180.00

CG2R51 CG2R51 CG2R53 NG2R51      3.1000  2    180.00
```

```
!CG2R51 CG2R51 CG2R53 NG2R51     4.0000  2   180.00 ! CY3_MAL_ , from CG2R51 CG2R51 CG2R51
NG2R51, penalty= 15
!NH2  CC   CA   CA               3.1000  2    180.00

NG2R51 CG321  CG321  HGA2       0.1950  3     0.00 ! CY3_MAL_ , from NG2S1 CG321 CG321 HGA2,
penalty= 11
!^can't find this dihedral

OG2D1  CG2R53 NG2R51 CG2R53      3.1000  2   180.00
!OG2D1  CG2R53 NG2R51 CG2R53     2.5000  2   180.00 ! CY3_MAL_ , from OG2D1 CG2R53 NG2R51 CG2RC0,
penalty= 7
!O    CC   NH2  CC               3.1000  2    180.00

CG2R51 CG2R51 CG2R53 OG2D1       3.1000  2   0.00
!CG2R51 CG2R51 CG2R53 OG2D1      8.5000  2   180.00 ! CY3_MAL_ , from CG2R51 CG2R51 CG2R51
OG2R50, penalty= 39
!O    CC   CA   CA               3.1000  2    0.00

CG321  CG321  NG2R51 CG2R53      0.2000  1    180.00
!CG321  CG321  NG2R51 CG2R53     1.8000  1     0.00 ! CY3_MAL_ , from CG321 CG311 NG2R53 CG2R53,
penalty= 64
!CT2  CT2  NH2  CC               0.2000  1    180.00

CG2R52 CG3C50  CG331  HGA3       0.1600  3     0.00 ! RETINOL TMCH
!HGA3   CG331  CG3C50 CG2R52     0.1600  3     0.00 ! CY3_MAL_ , from CG2D1 CG321 CG331 HGA3,
penalty= 68
!CG2R52 CG3C50  CG331  HGA3       0.1600  3     0.00 ! RETINOL TMCH
!HGA3   CG331  CG3C50 CG2R52     0.1600  3     0.00 ! CY3_MAL_ , from CG2R51 CG321 CG331 HGA3,
penalty= 67


CG2RC0 CG3C50  CG331  HGA3       0.1600  3     0.00 ! RETINOL TMCH
!HGA3   CG331  CG3C50 CG2RC0     0.1600  3     0.00 ! CY3_MAL_ , from CG2R51 CG321 CG331 HGA3,
penalty= 66
!CG2RC0 CG3C50  CG331  HGA3       0.1600  3     0.00 ! RETINOL TMCH

OG2D1  CG2R53 NG2R51 CG321       2.5000  2    180.00
!OG2D1  CG2R53 NG2R51 CG321     11.0000  2   180.00 ! CY3_MAL_ , from NG2R50 CG2R53 NG2R51 CG331,
penalty= 46.9
!O    CC   NH2  CT2      2.5000  2    180.00

CG331  CG324  NG2R52  CG2R52    2.2500  1     0.00 ! PROT methylguanidinium, adm jr., 3/26/92
!B3LYP/6-31+g(d,p) dihedral scan
CG331  CG324  NG2R52  CG2R52    0.9000  2   180.00 ! PROT methylguanidinium, adm jr., 3/26/92
!CG331  CG324  NG2R52 CG2R52     0.0000  6   180.00 ! CY3_MAL_ , from CG321 CG324 NG2P1 CG2N1,
penalty= 53.9
!!CG321  CG324  NG2R52  CG2R52         2.2500  1     0.00 ! PROT methylguanidinium, adm jr.,
3/26/92  !B3LYP/6-31+g(d,p) dihedral scan
!!CG321  CG324  NG2R52  CG2R52         0.9000  2   180.00 ! PROT methylguanidinium, adm jr.,
3/26/92

CG331  CG324  NG2R52  CG2RC0    2.2500  1   180.00 ! PROT methylguanidinium, adm jr., 3/26/92
CG331  CG324  NG2R52  CG2RC0    0.9000  2   180.00 ! PROT methylguanidinium, adm jr., 3/26/92
!CG331  CG324  NG2R52 CG2RC0     0.0000  6   180.00 ! CY3_MAL_ , from CG321 CG324 NG2P1 CG2N1,
penalty= 53.9
!!CG321  CG324  NG2R52  CG2RC0         2.2500  1   180.00 ! PROT methylguanidinium, adm jr.,
3/26/92
!!CG321  CG324  NG2R52  CG2RC0         0.9000  2   180.00 ! PROT methylguanidinium, adm jr.,
3/26/92


!!CG321  CG321  CG324  NG2R52     0.1950  3     0.00 ! PROT alkane update, adm jr., 3/2/92
(arginine)

NG2R51 CG2R53 CG2R51 HGR51  4.2000  2  180.00
!NH2  CC   CA   HP       4.2000  2    180.00

!HERE Is the new dihedral for Cy5

CG2DC2 CG2DC2 CG2DC2 CG2DC2     0.5600  1   180.00
CG2DC2 CG2DC2 CG2DC2 CG2DC2     7.0000  2   180.00
```

```
!CG2DC2 CG2DC1 CG2DC1 CG2DC2     0.5600  1   180.00 ! RETINOL HEP3, 1,3,5-heptatriene
!CG2DC2 CG2DC1 CG2DC1 CG2DC2     7.0000  2   180.00 ! RETINOL HEP3, 1,3,5-heptatriene


!HERE are the new dihedrals for the linker:
CG2O1  CG321  CG321  CG321      0.1950  3     0.00 ! CY3_MAL_ , from CG2O1 CG321 CG321 CG311,
penalty= 0.6
NG2R51 CG321  CG321  NG2S1      0.1900  3     0.00 ! CY3_MAL_ , from CG2R61 CG321 CG321 NG2S1,
penalty= 47

!!and HERE are the dihedrals created by PRES SAX
CS   SM   CT2  HA       0.158   3      0.0
!DIHE 1HB1 1CB 1SG 2CL10
!DIHE 1HB2 1CB 1SG 2CL10

CT2  SM   CS   CG2R51       2.4     2    180.0
!DIHE 1CB 1SG 2CL10 2CL11

CT2  SM   CS   CG2R53       2.4     2    180.0
!DIHE 1CB 1SG 2CL10 2CL9

SM   CS   CG2R53   OG2D1        3.1     2    180.0
!DIHE 2O2 2CL9 2CL10 1SG

SM   CS   CG2R53   NG2R51      3.1     2    180.0
!DIHE NL2 2CL9 2CL10 1SG

SM   CS   CG2R51   HGR51       2.4     2    180.0
!DIHE 2H10 2C11 2CL10 1SG

SM   CS   CG2R51   CG2R53       3.1     2    180.0
!DIHE 2CL12 2CL11 2CL10 1SG

CS   SM   CT2  CT1      0.31    3      0.0
!DIHE 1CA 1CB 1SG 2CL10  !(added/GDS)

CS CG2R51 CG2R53 OG2D1   3.1     2    180.0
!CS   CA   CC   O        3.1     2    180.0

CG2R51 CS CG2R53 OG2D1   3.1     2    180.0
!O    CC   CS   CA       3.1     2    180.0

CG2R53 CS CG2R51 CG2R53  3.1     2    180.0
!CC   CA   CS   CC       3.1     2    180.0

CG2R53 CS CG2R51 HGR51   4.2     2    180.0
!HP   CA   CS   CC       4.2     2    180.0

CG321 NG2R51 CG2R53 CS   2.5     2    180.0
!CS   CC   NH2  CT2      2.5     2    180.0

NG2R51  CG2R53  CG2R51  CS  3.1   2  180.0
!CS   CA   CC   NH2      3.1     2    180.0

NG2R51 CG2R53 CS CG2R51  3.1     2    180.0
!CA   CS   CC   NH2      3.1     2    180.0

CG2R53 NG2R51 CG2R53 CS  3.1     2    180.0
!CS   CC   NH2  CC       3.1     2    180.0

OG2D1 CG2R53 CG2R51 HGR51    4.2000   2    180.00
!O    CC   CA   HP       4.2000  2    180.00

IMPROPERS
CG2R53 CG2R51 NG2R51 OG2D1     90.0000  0     0.00 ! CY3_MAL_ , from CG2R53 CG2DC1 NG2R51 OG2D1,
penalty= 57
CG2R53 CS NG2R51 OG2D1     90.0000  0     0.00 ! CY3_MAL_ , from CG2R53 CG2DC1 NG2R51 OG2D1,
penalty= 57

NONBONDED
```

BIBLIOGRAPHY

Abbondanzieri, E.A., Bokinsky, G., Rausch, J.W., Zhang, J.X., Le Grice, S.F.J., and Zhuang, X. (2008). Dynamic binding orientations direct activity of HIV reverse transcriptase. Nature *453*, 184–189.

Allen, L.R., and Paci, E. (2009). Orientational averaging of dye molecules attached to proteins in Förster resonance energy transfer measurements: Insights from a simulation study. J. Chem. Phys. *131*, 065101.

Andries, K., Azijn, H., Thielemans, T., Ludovici, D., Kukla, M., Heeres, J., Janssen, P., De Corte, B., Vingerhoets, J., Pauwels, R., et al. (2004). TMC125, a novel next-generation nonnucleoside reverse transcriptase inhibitor active against nonnucleoside reverse transcriptase inhibitor-resistant human immunodeficiency virus type 1. Antimicrob. Agents Chemother. *48*, 4680–4686.

Anta, L., Llibre, J.M., Poveda, E., Blanco, J.L., Alvarez, M., Pérez-Elías, M.J., Aguilera, A., Caballero, E., Soriano, V., de Mendoza, C., et al. (2013). Rilpivirine resistance mutations in HIV patients failing non-nucleoside reverse transcriptase inhibitor-based therapies. AIDS Lond. Engl. *27*, 81–85.

Arion, D., Kaushik, N., McCormick, S., Borkow, G., and Parniak, M.A. (1998). Phenotypic mechanism of HIV-1 resistance to 3'-azido-3'-deoxythymidine (AZT): increased polymerization processivity and enhanced sensitivity to pyrophosphate of the mutant viral reverse transcriptase. Biochemistry *37*, 15908–15917.

Arts, E.J., and Hazuda, D.J. (2012). HIV-1 Antiretroviral Drug Therapy. Cold Spring Harb. Perspect. Med. *2*.

Axelrod, D. (1989). Total internal reflection fluorescence microscopy. Methods Cell Biol. *30*, 245–270.

Azijn, H., Tirry, I., Vingerhoets, J., de Béthune, M.-P., Kraus, G., Boven, K., Jochmans, D., Van Craenenbroeck, E., Picchio, G., and Rimsky, L.T. (2010). TMC278, a next-generation nonnucleoside reverse transcriptase inhibitor (NNRTI), active against wild-type and NNRTI-resistant HIV-1. Antimicrob. Agents Chemother. *54*, 718–727.

Bahar, I., Erman, B., Jernigan, R.L., Atilgan, A.R., and Covell, D.G. (1999). Collective Motions in HIV-1 Reverse Transcriptase: Examination of Flexibility and Enzyme Function,. J. Mol. Biol. *285*, 1023–1037.

Bastiaens, P.I., and Jovin, T.M. (1996). Microspectroscopic imaging tracks the intracellular processing of a signal transduction protein: fluorescent-labeled protein kinase C beta I. Proc. Natl. Acad. Sci. USA. *93*, 8407 –8412.

Blake, A., and Isard, M. (1998). Active Contours. Springer Verl.

Brehm, J.H., Mellors, J.W., and Sluis-Cremer, N. (2008). Mechanism by which a Glutamine to Leucine Substitution at Residue 509 in the Ribonuclease H Domain of HIV-1 Reverse Transcriptase Confers Zidovudine Resistance†. Biochemistry *47*, 14020–14027.

Chung, H.S., Louis, J.M., and Eaton, W.A. (2010). Distinguishing between Protein Dynamics and Dye Photophysics in Single-Molecule FRET Experiments. Biophys. J. *98*, 696–706.

Clotet, B. (1999). Efavirenz: resistance and cross-resistance. Int. J. Clin. Pract. Suppl. *103*, 21–25.

Corry, B., and Jayatilaka, D. (2008). Simulation of Structure, Orientation, and Energy Transfer between AlexaFluor Molecules Attached to MscL. Biophys. J. *95*, 2711–2721.

Das, K., Ding, J., Hsiou, Y., Clark, A.D., Jr, Moereels, H., Koymans, L., Andries, K., Pauwels, R., Janssen, P.A., Boyer, P.L., et al. (1996). Crystal structures of 8-Cl and 9-Cl TIBO complexed with wild-type HIV-1 RT and 8-Cl TIBO complexed with the Tyr181Cys HIV-1 RT drug-resistant mutant. J. Mol. Biol. *264*, 1085–1100.

Das, K., Sarafianos, S.G., Clark Jr, A.D., Boyer, P.L., Hughes, S.H., and Arnold, E. (2007). Crystal Structures of Clinically Relevant Lys103Asn/Tyr181Cys Double Mutant HIV-1 Reverse Transcriptase in Complexes with ATP and Non-nucleoside Inhibitor HBY 097. J. Mol. Biol. *365*, 77–89.

Das, K., Martinez, S.E., Bauman, J.D., and Arnold, E. (2012). HIV-1 reverse transcriptase complex with DNA and nevirapine reveals non-nucleoside inhibition mechanism. Nat Struct Mol Biol *19*, 253–259.

Deval, J., Selmi, B., Boretto, J., Egloff, M.P., Guerreiro, C., Sarfati, S., and Canard, B. (2002). The molecular mechanism of multidrug resistance by the Q151M human immunodeficiency virus type 1 reverse transcriptase and its suppression using alpha-boranophosphate nucleotide analogues. J. Biol. Chem. *277*, 42097–42104.

Deval, J., Navarro, J.-M., Selmi, B., Courcambeck, J., Boretto, J., Halfon, P., Garrido-Urbani, S., Sire, J., and Canard, B. (2004a). A loss of viral replicative capacity correlates with altered DNA polymerization kinetics by the human immunodeficiency virus reverse transcriptase bearing the K65R and L74V dideoxynucleoside resistance substitutions. J. Biol. Chem. *279*, 25489–25496.

Deval, J., White, K.L., Miller, M.D., Parkin, N.T., Courcambeck, J., Halfon, P., Selmi, B., Boretto, J., and Canard, B. (2004b). Mechanistic basis for reduced viral and enzymatic fitness of HIV-1 reverse transcriptase containing both K65R and M184V mutations. J. Biol. Chem. *279*, 509–516.

Deval, J., White, K.L., Miller, M.D., Parkin, N.T., Courcambeck, J., Halfon, P., Selmi, B., Boretto, J., and Canard, B. (2004c). Mechanistic basis for reduced viral and enzymatic fitness of HIV-1 reverse transcriptase containing both K65R and M184V mutations. J. Biol. Chem. *279*, 509–516.

Divita, G., Restle, T., and Goody, R.S. (1993a). Characterization of the dimerization process of HIV-1 reverse transcriptase heterodimer using intrinsic protein fluorescence. FEBS Lett. *324*, 153–158.

Divita, G., Mueller, B., Immendoerfer, U., Gautel, M., Rittinger, K., Restle, T., and Goody, R.S. (1993b). Kinetics of interaction of HIV reverse transcriptase with primer/template. Biochemistry *32*, 7966–7971.

Donald, J.E., Kulp, D.W., and DeGrado, W.F. (2011). Salt bridges: Geometrically specific, designable interactions. Proteins Struct. Funct. Bioinforma. *79*, 898–915.

Esnouf, R., Ren, J., Ross, C., Jones, Y., Stammers, D., and Stuart, D. (1995). Mechanism of inhibition of HIV-1 reverse transcriptase by non-nucleoside inhibitors. Nat Struct Mol Biol *2*, 303–308.

Esposito, F., Corona, A., and Tramontano, E. (2012). HIV-1 Reverse Transcriptase Still Remains a New Drug Target: Structure, Function, Classical Inhibitors, and New Inhibitors with Innovative Mechanisms of Actions. Mol. Biol. Int. *2012*, 1–23.

Fagerburg, M.V., and Leuba, S.H. (2011). Optimal practices for surface-tethered single molecule total internal reflection fluorescence resonance energy transfer analysis. Methods Mol. Biol. Clifton NJ *749*, 273–289.

Feng, J.Y., and Anderson, K.S. (1999). Mechanistic studies examining the efficiency and fidelity of DNA synthesis by the 3TC-resistant mutant (184V) of HIV-1 reverse transcriptase. Biochemistry *38*, 9440–9448.

Fischer, C.J., Maluf, N.K., and Lohman, T.M. (2004). Mechanism of ATP-dependent Translocation of E. coli UvrD Monomers Along Single-stranded DNA. J. Mol. Biol. *344*, 1287–1309.

Goody, R.S., Müller, B., and Restle, T. (1991). Factors contributing to the inhibition of HIV reverse transcriptase by chain-terminating nucleotides in vitro and in vivo. FEBS Lett. *291*, 1–5.

Götte, M., Rausch, J.W., Marchand, B., Sarafianos, S., and Le Grice, S.F.J. (2010). Reverse transcriptase in motion: Conformational dynamics of enzyme–substrate interactions. Biochim. Biophys. Acta BBA - Proteins Proteomics *1804*, 1202–1212.

Graham, B.W., Schauer, G.D., Leuba, S.H., and Trakselis, M.A. (2011). Steric exclusion and wrapping of the excluded DNA strand occurs along discrete external binding paths during MCM helicase unwinding. Nucleic Acids Res. *39*, 6585–6595.

Grice, S.F.J.L. (2012). Human Immunodeficiency Virus Reverse Transcriptase: 25 Years of Research, Drug Discovery, and Promise. J. Biol. Chem. *287*, 40850–40857.

Hamelberg, D., Mongan, J., and Mccammon, J.A. (2004). Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. J. Chem. Phys. *120*, 11919–11929.

Hinterdorfer, P., and Van Oijen, A. (2009). Handbook of single-molecule biophysics (Dordrecht; New York: Springer).

Hohlbein, J., Gryte, K., Heilemann, M., and Kapanidis, A.N. (2010). Surfing on a new wave of single-molecule fluorescence methods. Phys. Biol. *7*, 031001.

Hsiou, Y., Ding, J., Das, K., Clark Jr, A.D., Boyer, P.L., Lewi, P., Janssen, P.A.., Kleim, J.-P., Rösner, M., Hughes, S.H., et al. (2001). The Lys103Asn mutation of HIV-1 RT: a novel mechanism of drug resistance. J. Mol. Biol. *309*, 437–445.

Hu, W.-S., and Hughes, S.H. (2012). HIV-1 Reverse Transcription. Cold Spring Harb. Perspect. Med. *2*.

Huang, H., Chopra, R., Verdine, G.L., and Harrison, S.C. (1998). Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. Science *282*, 1669–1675.

Huang, W., Gamarnik, A., Limoli, K., Petropoulos, C.J., and Whitcomb, J.M. (2003). Amino acid substitutions at position 190 of human immunodeficiency virus type 1 reverse transcriptase increase susceptibility to delavirdine and impair virus replication. J. Virol. *77*, 1512–1523.

Hwang, H., Kim, H., and Myong, S. (2011). Protein induced fluorescence enhancement as a single molecule assay with short distance sensitivity. Proc. Natl. Acad. Sci. USA *108*, 7414–7418.

Iqbal, A., Arslan, S., Okumus, B., Wilson, T.J., Giraud, G., Norman, D.G., Ha, T., and Lilley, D.M.J. (2008a). Orientation dependence in fluorescent energy transfer between Cy3 and Cy5 terminally attached to double-stranded nucleic acids. Proc. Natl. Acad. Sci. USA *105*, 11176–11181.

Iqbal, A., Wang, L., Thompson, K.C., Lilley, D.M.J., and Norman, D.G. (2008b). The Structure of Cyanine 5 Terminally Attached to Double-Stranded DNA: Implications for FRET Studies†. Biochemistry *47*, 7857–7862.

Ivetac, A., and McCammon, J.A. (2009). Elucidating the Inhibition Mechanism of HIV-1 Non-Nucleoside Reverse Transcriptase Inhibitors through Multicopy Molecular Dynamics Simulations. J. Mol. Biol. *388*, 644–658.

Jacobo-Molina, A., Ding, J., Nanni, R.G., Clark, A.D., Lu, X., Tantillo, C., Williams, R.L., Kamer, G., Ferris, A.L., and Clark, P. (1993). Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 A resolution shows bent DNA. Proc. Natl. Acad. Sci. USA *90*, 6320–6324.

Joo, C., Balci, H., Ishitsuka, Y., Buranachai, C., and Ha, T. (2008). Advances in Single-Molecule Fluorescence Methods for Molecular Biology. Annu. Rev. Biochem. *77*, 51–76.

Kati, W.M., Johnson, K.A., Jerva, L.F., and Anderson, K.S. (1992). Mechanism and fidelity of HIV reverse transcriptase. J. Biol. Chem. *267*, 25988–25997.

Kirmizialtin, S., Nguyen, V., Johnson, K.A., and Elber, R. (2012). How Conformational Dynamics of DNA Polymerase Select Correct Substrates: Experiments and Simulations. Structure *20*, 618–627.

Kohlstaedt, L.A., Wang, J., Friedman, J.M., Rice, P.A., and Steitz, T.A. (1992). Crystal structure at 3.5 A resolution of HIV-1 reverse transcriptase complexed with an inhibitor. Science *256*, 1783–1790.

KruhØfter, M., Urbanke, C., and Grosse, F. (1993). Two step binding of HIV-1 reverse transcriptase to nucleic acid substrates. Nucleic Acids Res. *21*, 3943 –3949.

Lakowicz, J.R. (2006). Principles of Fluorescence Spectroscopy (Springer).

Lansdon, E.B., Brendza, K.M., Hung, M., Wang, R., Mukund, S., Jin, D., Birkus, G., Kutty, N., and Liu, X. (2010). Crystal Structures of HIV-1 Reverse Transcriptase with Etravirine (TMC125) and Rilpivirine (TMC278): Implications for Drug Design. J. Med. Chem. *53*, 4295–4299.

Lapkouski, M., Tian, L., Miller, J.T., Le Grice, S.F.J., and Yang, W. (2013). Complexes of HIV-1 RT, NNRTI and RNA/DNA hybrid reveal a structure compatible with RNA degradation. Nat. Struct. Mol. Biol. *20*, 230–236.

Leach, A.R. (2001). Molecular modelling: principles and applications (Harlow [etc.]: Prentice Hall).

Lee, N.K., Kapanidis, A.N., Wang, Y., Michalet, X., Mukhopadhyay, J., Ebright, R.H., and Weiss, S. (2005). Accurate FRET Measurements within Single Diffusing Biomolecules Using Alternating-Laser Excitation. Biophys. J. *88*, 2939–2953.

Lee, S.-K., Potempa, M., and Swanstrom, R. (2012). The Choreography of HIV-1 Proteolytic Processing and Virion Assembly. J. Biol. Chem. *287*, 40867–40874.

Liu, S., Abbondanzieri, E.A., Rausch, J.W., Grice, S.F.J.L., and Zhuang, X. (2008). Slide into Action: Dynamic Shuttling of HIV Reverse Transcriptase on Nucleic Acid Substrates. Science *322*, 1092–1097.

Liu, Y., Park, J., Dahmen, K.A., Chemla, Y.R., and Ha, T. (2010). A Comparative Study of Multivariate and Univariate Hidden Markov Modelings in Time-Binned Single-Molecule FRET Data Analysis. J. Phys. Chem. B.

Lu, H.P. (2005). Probing Single-Molecule Protein Conformational Dynamics. Accounts Chem. Res. *38*, 557–565.

McCann, J.J., Choi, U.B., Zheng, L., Weninger, K., and Bowen, M.E. (2010). Optimizing Methods to Recover Absolute FRET Efficiency from Immobilized Single Molecules. Biophys. J. *99*, 961–970.

Menéndez-Arias, L. (2013). Molecular basis of human immunodeficiency virus type 1 drug resistance: Overview and recent developments. Antiviral Res. *98*, 93–120.

Meyer, P.R., Matsuura, S.E., Mian, A.M., So, A.G., and Scott, W.A. (1999). A mechanism of AZT resistance: an increase in nucleotide-dependent primer unblocking by mutant HIV-1 reverse transcriptase. Mol. Cell *4*, 35–43.

Michalet, X., Weiss, S., and Jäger, M. (2006). Single-molecule fluorescence studies of protein folding and conformational dynamics. Chem. Rev. *106*, 1785–1813.

Radzio, J., and Sluis-Cremer, N. (2008). Efavirenz accelerates HIV-1 reverse transcriptase ribonuclease H cleavage, leading to diminished zidovudine excision. Mol. Pharmacol. *73*, 601–606.

Rawal, R.K., Murugesan, V., and Katti, S.B. (2012). Structure-activity relationship studies on clinically relevant HIV-1 NNRTIs. Curr. Med. Chem. *19*, 5364–5380.

Ren, J., and Stammers, D.K. (2008). Structural basis for drug resistance mechanisms for non-nucleoside inhibitors of HIV reverse transcriptase. Virus Res. *134*, 157–170.

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

Ren, J., Milton, J., Weaver, K.L., Short, S.A., Stuart, D.I., and Stammers, D.K. (2000). Structural Basis for the Resilience of Efavirenz (DMP-266) to Drug Resistance Mutations in HIV-1 Reverse Transcriptase. Structure *8*, 1089–1094.

Ren, J., Nichols, C., Bird, L., Chamberlain, P., Weaver, K., Short, S., Stuart, D.I., and Stammers, D.K. (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.

Rodgers, D.W., Gamblin, S.J., Harris, B.A., Ray, S., Culp, J.S., Hellmig, B., Woolf, D.J., Debouck, C., and Harrison, S.C. (1995). The structure of unliganded reverse transcriptase from the human immunodeficiency virus type 1. Proc. Natl. Acad. Sci. USA *92*, 1222–1226.

Roy, R., Hohng, S., and Ha, T. (2008). A practical guide to single-molecule FRET. Nat Meth *5*, 507–516.

Santoso, Y., Hwang, L.C., LeÂ reste Ludovic, and Kapanidis, A.N. (2008). Red light, green light: probing single molecules using alternating-laser excitation. Biochem. Soc. Trans. *036*, 738–744.

Sarafianos, S.G., Das, K., Ding, J., Boyer, P.L., Hughes, S.H., and Arnold, E. (1999). Touching the heart of HIV-1 drug resistance: the fingers close down on the dNTP at the polymerase active site. Chem. Biol. *6*, R137–146.

Sarafianos, S.G., Clark, A.D., Das, K., Tuske, S., Birktoft, J.J., Ilankumaran, P., Ramesha, A.R., Sayer, J.M., Jerina, D.M., Boyer, P.L., et al. (2002). Structures of HIV-1 reverse transcriptase with pre- and post-translocation AZTMP-terminated DNA. EMBO J. *21*, 6614–6624.

Sarafianos, S.G., Marchand, B., Das, K., Himmel, D.M., Parniak, M.A., Hughes, S.H., and Arnold, E. (2009). Structure and Function of HIV-1 Reverse Transcriptase: Molecular Mechanisms of Polymerization and Inhibition. J. Mol. Biol. *385*, 693–713.

Schlick, T. (2010). Molecular modeling and simulation: an interdisciplinary guide (New York: Springer).

Schroder, G. (2005). Simulation of Fluorescence Anisotropy Experiments: Probing Protein Dynamics. Biophys. J. *89*, 3757–3770.

Seckler, J.M., Barkley, M.D., and Wintrode, P.L. (2011). Allosteric Suppression of HIV-1 Reverse Transcriptase Structural Dynamics upon Inhibitor Binding. Biophys. J. *100*, 144–153.

Selmi, B., Boretto, J., Sarfati, S.R., Guerreiro, C., and Canard, B. (2001). Mechanism-based suppression of dideoxynucleotide resistance by K65R human immunodeficiency virus reverse transcriptase using an alpha-boranophosphate nucleoside analogue. J. Biol. Chem. *276*, 48466–48472.

Selmi, B., Deval, J., Boretto, J., and Canard, B. (2003). Nucleotide analogue binding, catalysis and primer unblocking in the mechanisms of HIV-1 reverse transcriptase-mediated resistance to nucleoside analogues. Antivir. Ther. *8*, 143–154.

Sharp, P.M., and Hahn, B.H. (2011). Origins of HIV and the AIDS Pandemic. Cold Spring Harb. Perspect. Med. *1*.

Sluis-Cremer, N., and Tachedjian, G. (2008). Mechanisms of inhibition of HIV replication by non-nucleoside reverse transcriptase inhibitors. Virus Res. *134*, 147–156.

Sluis-Cremer, N., Temiz, N.A., and Bahar, I. (2004). Conformational Changes in HIV-1 Reverse Transcriptase Induced by Nonnucleoside Reverse Transcriptase Inhibitor Binding. Curr. HIV Res. *2*, 323–332.

Sluis-Cremer, N., Sheen, C.-W., Zelina, S., Torres, P.S.A., Parikh, U.M., and Mellors, J.W. (2007). Molecular mechanism by which the K70E mutation in human immunodeficiency virus type 1 reverse transcriptase confers resistance to nucleoside reverse transcriptase inhibitors. Antimicrob. Agents Chemother. *51*, 48–53.

Spiriti, J., Binder, J.K., Levitus, M., and van der Vaart, A. (2011). Cy3-DNA Stacking Interactions Strongly Depend on the Identity of the Terminal Basepair. Biophys. J. *100*, 1049–1057.

Steitz, T.A. (1998). Structural biology: A mechanism for all polymerases. Nature *391*, 231–232.

Temiz, N.A., and Bahar, I. (2002). Inhibitor binding alters the directions of domain motions in HIV-1 reverse transcriptase. Proteins Struct. Funct. Genet. *49*, 61–70.

VanBeek, D.B., Zwier, M.C., Shorb, J.M., and Krueger, B.P. (2007). Fretting about FRET: Correlation between κ and R. Biophys. J. *92*, 4168–4178.

Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., Darian, E., Guvench, O., Lopes, P., Vorobyov, I., et al. (2010). CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. J. Comput. Chem. *31*, 671–690.

Vingerhoets, J., Tambuyzer, L., Azijn, H., Hoogstoel, A., Nijs, S., Peeters, M., de Béthune, M.-P., De Smedt, G., Woodfall, B., and Picchio, G. (2010). Resistance profile of etravirine: combined analysis of baseline genotypic and phenotypic data from the randomized, controlled Phase III clinical studies. AIDS Lond. Engl. *24*, 503–514.

Wang, D., and Geva, E. (2005a). Protein Structure and Dynamics from Single-Molecule Fluorescence Resonance Energy Transfer. J. Phys. Chem. B *109*, 1626–1634.

Wang, D., and Geva, E. (2005b). Protein Structure and Dynamics from Single-Molecule Fluorescence Resonance Energy Transfer Protein Structure and Dynamics from Single-Molecule Fluorescence Resonance Energy. J. Phys. Chem. B.

Wang, X., and Lu, H.P. (2008). 2D Regional Correlation Analysis of Single-Molecule Time Trajectories. J. Phys. Chem. B *112*, 14920–14926.

Watkins, L.P., and Yang, H. (2004). Information Bounds and Optimal Analysis of Dynamic Single Molecule Measurements. Biophys. J. *86*, 4015–4029.

Wöhrl, B.M., Krebs, R., Goody, R.S., and Restle, T. (1999). Refined model for primer/template binding by HIV-1 reverse transcriptase: pre-steady-state kinetic analyses of primer/template binding and nucleotide incorporation events distinguish between different binding modes depending on the nature of the nucleic acid substrate. J. Mol. Biol. *292*, 333–344.

Wozniak, A.K., Schroder, G.F., Grubmuller, H., Seidel, C.A.M., and Oesterhelt, F. (2008). Single-molecule FRET measures bends and kinks in DNA. Proc. Natl. Acad. Sci. USA *105*, 18337–18342.

Yap, S.-H., Sheen, C.-W., Fahey, J., Zanin, M., Tyssen, D., Lima, V.D., Wynhoven, B., Kuiper, M., Sluis-Cremer, N., Harrigan, P.R., et al. (2007). N348I in the connection domain of HIV-1 reverse transcriptase confers zidovudine and nevirapine resistance. PLoS Med. *4*, e335.

Zheng, H., Goldner, L.S., and Leuba, S.H. (2007). Homebuilt single-molecule scanning confocal fluorescence microscope studies of single DNA/protein interactions. Methods San Diego Calif *41*, 342–352.