NORTHWESTERN UNIVERSITY

Characterizing Critical Residues for the Interaction of Human Immunodeficiency Virus Type 1 Integrase with DNA Substrates

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Microbiology – Immunology
Integrated Graduate Program in the Life Sciences (IGP)

By

James Robert Dolan III

EVANSTON, ILLINOIS

DECEMBER 2008

ABSTRACT

Characterizing Critical Residues for the Interaction of Human Immunodeficiency Virus Type 1 Integrase with DNA Substrates

James Robert Dolan III

A tetramer model for HIV-1 IN with DNA representing the LTR termini was previously assembled which predicted amino acid residues on the enzyme surface that interact with the LTR termini (Chen et al, 2006). A separate structural alignment of HIV-1, SIV, and ASV INs predicted which of these residues were unique. By substituting the unique amino acids found in ASV IN into the structurally related positions of HIV-1, nine amino acid residues were shown to partially change the specificity for 3' processing from HIV-1 to ASV duplex oligo substrates. Using a similar strategy, but with a structural alignment that substituted MPMV for the SIV sequence, six additional residues (Q44, L68, E69, D229, S230 and D253) were identified that interact with the LTR DNA and change the 3' processing specificity of the enzyme. All fifteen residues align along a sixteen base pair length of the LTR termini asymmetrically positioned relative to each strand of the DNA.

The tetramer model for HIV-1 IN with LTR termini was modified to include two IN binding domains for LEDGF/p75. The target DNA was predicted to bind in a surface trench perpendicular to the plane of the LTR DNA binding sites of HIV-1 IN and extending alongside LEDGF. Consistent with this hypothesis is the finding that a HIV-1 IN mutant with a K219S substitution displays an activity phenotype where there is a loss in strand transfer with little change in 3' processing activity towards HIV-1 substrates. Mutations at seven other residues reported in the literature have the same activity phenotype and align along the opposite face of the putative target DNA binding trench.

Recently small-molecule inhibitors have been approved by the FDA for human therapy. When infected cells are put under selection of these inhibitors, mutations arise in IN that result in resistance to the drugs. At least 10 of these sites have been identified. Examination of *in vitro* activities of INs with changes at these residues provides insight into both the mechanism of action and the function of secondary and tertiary mutations.

ACKNOWLEDGEMENTS

Many people have influenced my personal and professional development in the past few years, and I would like to take this space to acknowledge and thank them.

First, I would like to thank my wife Lauren for all of her love and support during graduate school. It has been quite a journey, from our days in Champaign to our future home of New York City. I wouldn't want another partner to share the experience.

Mom and Dad, you have been tremendous role models and provided support and advice when you thought it was appropriate. Thank you for your love, generosity, inspiration and understanding. I would not be who I am today without your guidance.

To my siblings: Emily, Andrew, Nicholas, and Sean. Family is incredibly important and I am glad that I was fortunate enough to have found a graduate program in our hometown so as to enjoy the last few years with you all.

I would like to thank Abby, Shan and Eric for their friendship and the memories that we share from our time in Chicago together. Things won't be the same, but that doesn't mean they won't still be great. We now will have places to stay in some of the greatest cities in America. I look forward to visiting you regularly.

To my friends Kelly, Yvonne, Tom, and Derek: I look forward to seeing you (and drinking some wine) whenever we have the opportunity to get together, maybe in San Francisco or Miami.

I would like to thank my advisor, Dr. Jonathan Leis for allowing me the opportunity to work in his lab on a number of interesting projects and providing me guidance and support during my graduate school career. I would also like to thank current and past members of the lab for their contributions to my graduate experience.

I would like to thank my thesis committee members Dr. Laimins, Dr. Sznajder, and Dr. Smith for their questions, insights, advice, and critiques.

I would also like to thank other faculty within the IGP who have influenced my development, including Dr. Chandel, Dr. Rundell, Dr. Longnecker and Dr. Seifert. Your interest and devotion to scientific research is inspiring. I thank you for all the advice that I have received from you during my time in the IGP.

LIST OF ABBREVIATIONS

ASV – avian sarcoma virus
ATP – adenosine triphosphate
BSA – bovine serum albumin
Ci – curie
Cpm – counts per minute
DMSO – dimethyl sulfoxide
DNA – deoxyribonucleic Acid
DTT – dithiothreitol
EDTA – ethylene diamine tetracetic acid
HIV-1 – human immunodeficiency virus, type 1
IN – integrase
IPTG - isopropyl β-D-1-thiogalactopyranoside
Kan – kanamycin
KCl – potassium chloride
LB – luria broth
LEDGF – Lens epithelium derived growth factor
LTR – long terminal repeat
mA – milliamps
$MgCl_2$ – magnesium chloride
MLV – murine leukemia virus
mmol - millimol
MOPS – 3-(N-morpholino) propane-sulfonic acid
M-PMV – mason-pfizer monkey virus
MW – molecular weight
Oligo – oligodeoxyribonucleotide
PIC – pre-integration complex
V – volts

TABLE OF CONTENTS

LIST OF TABLES, FIGURES, AND GRAPHS

CHAPTER 1

INTRODUCTION

1.1 Impact of HIV-1 Infection on Global Public Health

The 2006 Report on the Global AIDS Epidemic from the UN noted progress to control the expansion of the epidemic. Practices such as HIV testing, condom use, fewer sexual partners and counseling can have a profound impact on HIV infection rates. Of note, treatment strategies have been developed to significantly reduce mother-to-child transmission of the virus. Identifying infected persons through rigorous testing is critical. The United States was discovered to be underestimating the numbers of HIV-infected persons by as much as 40%. Changing sex practices such as using prophylactics and fewer sexual partners have shown promising results in parts of Africa and Asia. Practice of these principles on a global scale will mean great strides in the fight against the current epidemic.

HIV incidence has peaked in most countries in Africa. In addition, India, which is home to >66% of current HIV^+ individuals in Asia, has shown a decline in HIV prevalence for the period of 2001-2005, but therapeutic coverage is still well below 10%. However, some areas in southern Africa are still experiencing severe and expanding epidemics with shocking prevalence rates, some even exceeding 30%. Countries in eastern Asia, most notably China, have been experiencing outbreaks and expanding HIV prevalence. Eastern Europe and central Asia have expanding epidemics, with the numbers of HIV-1 infected individuals showing a twenty-fold increase in less than a decade. Developing nations and those with the largest populations, including Brazil, Russia and much of Southeast Asia, are also showing a trend of increasing HIV-1 prevalence. Lastly, Brazil is home to approximately one-third of the HIV^+ persons in Latin America, which threatens progresses in economic and social areas as well. The objectives of the UNAIDS and the global commitment on HIV/AIDS demonstrate that much has been done, but there is more yet to do.

In developed nations, the prevalence and mortality rates have decreased in recent history due to the development and distribution of a number of pharmacologic inhibitors. The Food and Drug Administration has approved inhibitors for all three enzymes of HIV (Protease, Integrase, and Reverse Transcriptase), in addition to viral fusion inhibitors. These four drug targets are used in cocktail formulations to decrease the opportunity for development of resistance against any of the inhibitors. In the United States, the treatment regimen costs roughly $6000 annually. This lofty financial requirement puts therapeutics out of reach for many countries across the globe. As such, the incidence and morbidity has grown and will continue to grow as long as the system continues to operate within the status quo.

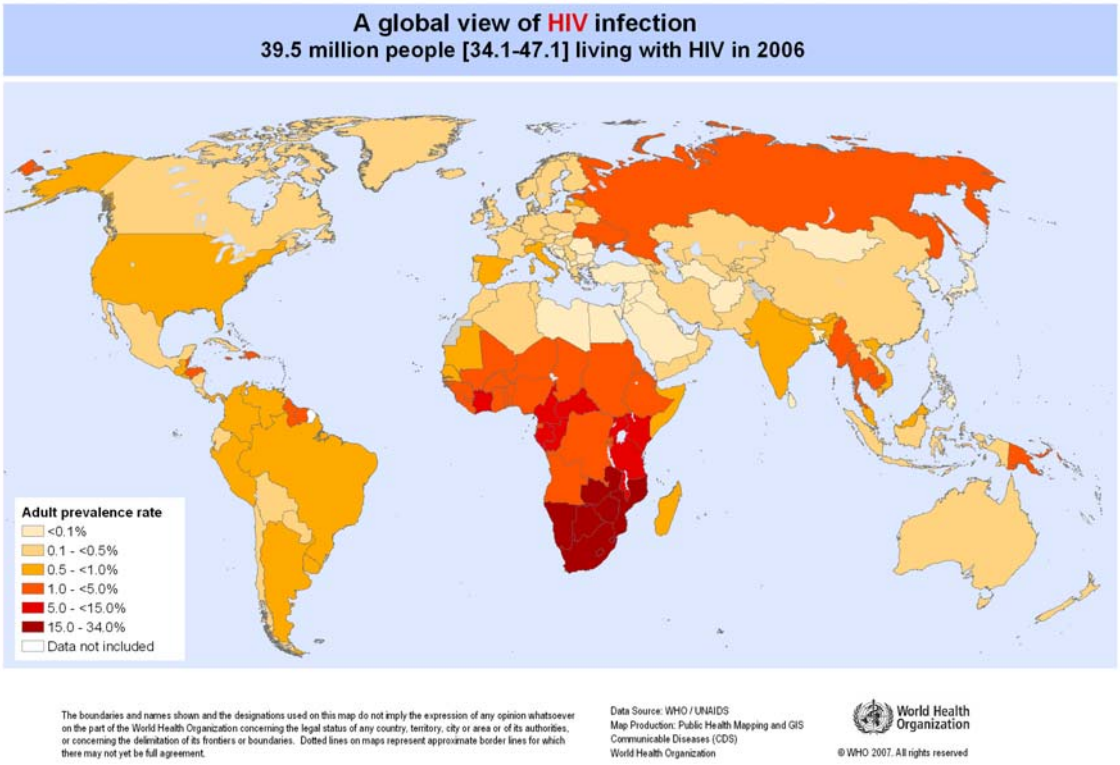


Figure 1.1 Global prevalence of HIV-1 infection in adults

This map from the World Health Organization is a representation of the prevalence of HIV-1 infection around the globe. The color coding does not imply even distribution but rather is the prevalence rate for a given geopolitical area. As seen in the map, several Southern African countries have staggeringly high HIV-1 prevalence rates, exceeding 15%.

1.2 Human Immunodeficiency Virus

The clinical outcome of infection with human immunodeficiency virus (HIV) is called acquired immune deficiency syndrome (AIDS). The start of the AIDS epidemic is credited as June 5, 1981, when a cluster of five homosexual men were reported to have developed *Pneumocystis carinii* pneumonia (PCP) and shared a common exposure through sexual contact (CDC 1981). Several years later, the etiologic agent for the immune disorder was discovered (Barre-Sinoussi, Chermann et al. 1983; Gallo, Salahuddin et al. 1984; Popovic, Sarngadharan et al. 1984). What was once perceived to affect homosexuals and intravenous-drug users has become a global pandemic.

The terminal result of HIV infection is an attenuated immune system providing opportunity for infection and leukemias that are not observed in otherwise healthy individuals. All organ systems are affected by HIV infection, and as a result, opportunistic infections in these systems occur after the immune system is compromised. Tuberculosis and pneumonias are common pulmonary infections that occur after HIV infection. Additionally, a variety of gastrointestinal infections of common and opportunistic bacteria, parasites, and viruses are routinely observed.

Neurologic disorders are also seen in AIDS patients. Infection of the nervous system by *Toxoplasma gondii* results in encephalitis, while cryptococcal meningitis is caused by the fungus *Cryptococcus neoformans* (Luft and Chua 2000). Additionally, a demyelinating disease called progressive multifocal leukoencephalopathy is attributed to infection with the JC virus and fatality usually results in a few months (Sadler and Nelson 1997). A cognitive and motor dysfunction syndrome called HIV-associated dementia (HAD) is the most common cause of dementia among individuals under 40 years old (Burt, Agan et al. 2008).

Furthermore, several cancers develop in AIDS patients due to virus infection (Boshoff and Weiss 2002). Kaposi's sarcoma is caused by the gamma-herpesvirus Kaposi's sarcoma-associated herpes virus (KSHV) and has been a hallmark tumor since the beginning of the HIV epidemic. A number of B-cell lymphomas are attributed to infection by Epstein-Barr virus (EBV) and KSHV. In women, cervical cancer after infection by human papillomavirus (HPV) is another outcome of AIDS. All together, there are many diseases that appear in immunocompromised individuals with late stage HIV infection, often with a fatal outcome.

1.3 Retrovirus replication

A graphic representation of retrovirus replication is shown in Figure 1.2 (Balvay, Lopez Lastra et al. 2007). Replication can be broken down into early and late stages of replication, the steps of which have been well characterized. Briefly, HIV-1 infection is dependent on the interaction of the viral envelope protein gp120 (SU) with a cellular receptor and co-receptor. Upon entry into the cell, the linear single stranded positive sense RNA genome is reverse transcribed into a DNA duplex catalyzed by a viral RNA-dependent DNA-polymerase called reverse transcriptase (RT). The DNA is in association with both viral and host cell proteins in a pre-integration complex (PIC). The PICs are transported to the nucleus, where the viral DNA is covalently inserted into the host chromosome catalyzed by the viral integrase (IN). Once integrated, the DNA is referred to as the provirus and is defined as the end step in early replication. Late steps in replication involve host cell proteins that synthesize viral RNAs which are both spliced and unspliced. After transport into the cytoplasm, the full length genomic RNA serves as the mRNA for the *gag* and *pol* gene products while a spliced RNA serves as the mRNA for the envelope glycoproteins gp120 (SU) and gp41 (TM). A portion of the genome length RNA is packaged into newly formed virions rather than translated. Assembly and budding of virus particles is a complex process involving viral and cellular proteins. Once particles break from the cell surface, the viral encoded protease (PR) cleaves the Gag and GagPol polyproteins. As this process occurs, there is a maturation of the viral particle indicated by condensation of the viral core, resulting in infectious virions.

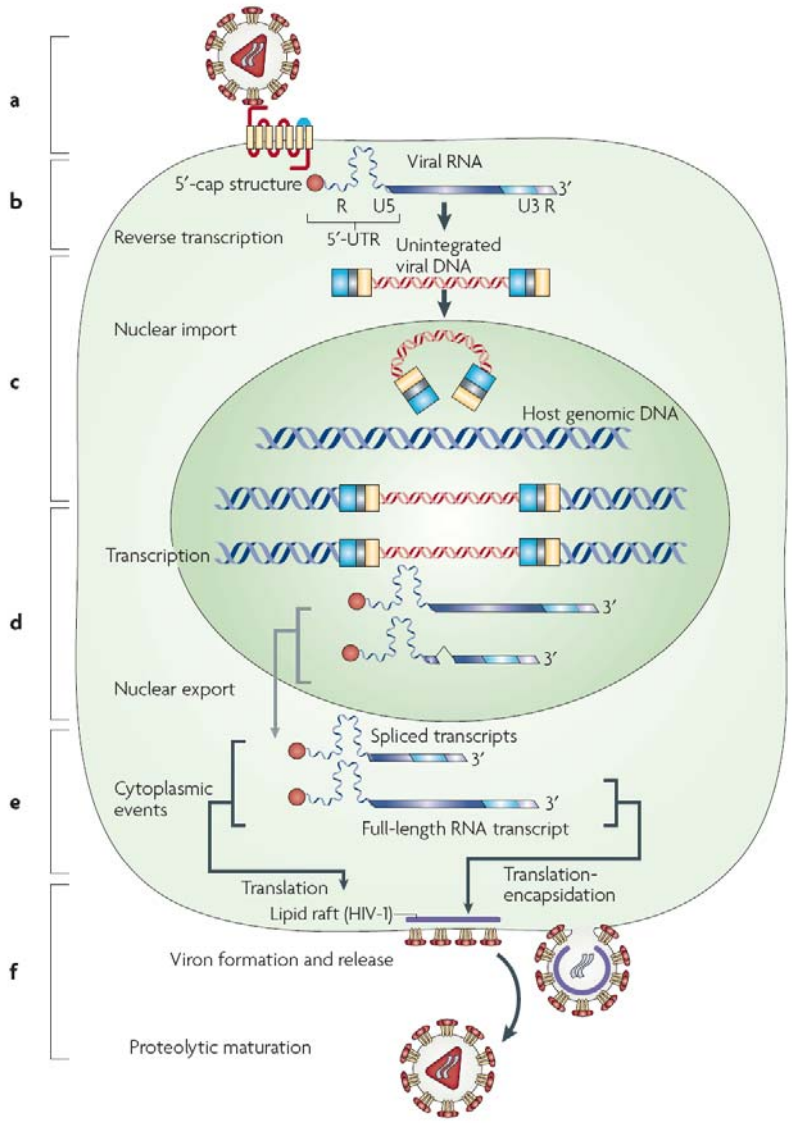


Figure 1.2 Retrovirus replication cycle

This figure highlights the steps of retrovirus replication and was adapted from Balvay et al. (Balvay, Lopez Lastra et al. 2007).

.

Human immunodeficiency virus is a member of the Retrovirus family, genus Lentivirus, subgenus Primate Lentivirus. The subgenus has three members: HIV type 1, HIV type 2, and simian immunodeficiency virus (SIV). The HIV-1 genome is organized as shown in Figure 1.3. The following subsections will explore the genes, gene products, and proteins of HIV.

1.3.1 The *gag, pol, and env* genes

The major genes encoded in all retroviral genomes are *gag* (**g**roup-specific **a**nti**g**en), *pol* (**pol**ymerase) and *env* (**env**elope glycoprotein). The *gag* gene encodes a polyprotein precursor ($Pr55^{Gag}$) that is cleaved by the retroviral protease (PR) to produce the structural proteins capsid (CA or p24), nucleocapsid (NC or p7), and matrix (MA or p17). Additional products from the protease cleavage of $Pr55^{Gag}$ are two spacer peptides, called p1 and p2, and the protein p6. The *pol* gene encodes reverse transcriptase (RT), protease (PR) and integrase (IN). A rare frameshift during Gag translation results in the $Pr160^{GagPol}$ precursor. Protease cleavage by the viral PR produces the mature viral enzymes. The membrane glycoproteins are encoded in the *env* gene, and are also synthesized as a polyprotein precursor called gp160. Unlike $Pr55^{Gag}$ and $Pr160^{GagPol}$, gp160 is cleaved by a cellular protease that produces the surface (SU) glycoprotein gp120 and the transmembrane (TM) glycoprotein gp41.

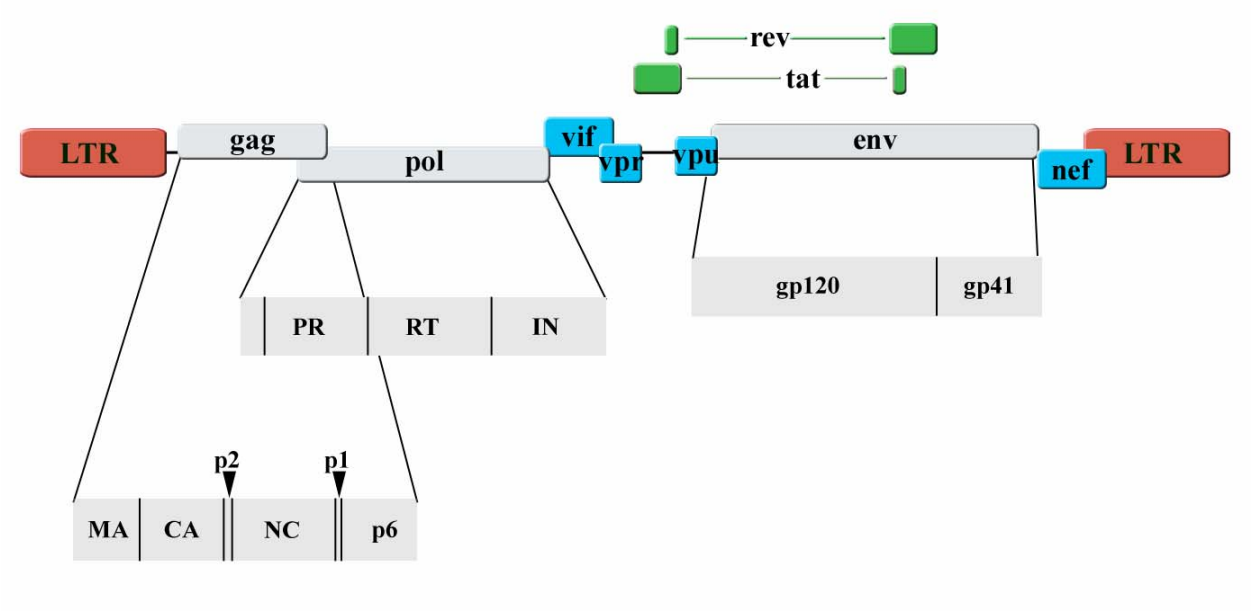


Figure 1.3 HIV-1 genome

The HIV-1 genome is comprised of three genes (colored gray), two transactivating genes required for replication (colored green), and four accessory genes (colored blue). The viral long terminal repeat ends are shown in red. For the gag, pol and env genes, the polyproteins are shown below with vertical lines indicating protease cleavage sites and resulting viral proteins using their two letter designations (Leis, Baltimore et al. 1988).

1.3.2 The *rev* and *tat* genes

The basal activity from the HIV LTR is low but the transcriptional transactivator Tat greatly influences transcription (Dayton, Sodroski et al. 1986; Fisher, Feinberg et al. 1986). Tat acts on an RNA element called the transactivation response region (TAR) which is present at the 5' end of all viral RNAs (Berkhout, Silverman et al. 1989). Additionally, Tat interacts with the cellular protein cyclin T1 and directs the positive-transcriptional-elongation factor b (P-TEF-b) to the TAR (Garber, Wei et al. 1998; Wei, Garber et al. 1998). This recruitment results in phosphorylation of RNA Polymerase II and an increase in transcriptional processivity (Parada and Roeder 1996; Cujec, Cho et al. 1997; Cujec, Okamoto et al. 1997).

Export of viral RNAs from the nucleus to the cytoplasm is handled by the Rev protein (**r**egulator of **e**xpression of **v**irion proteins) (Chang and Sharp 1989; Emerman, Vazeux et al. 1989; Felber, Hadzopoulou-Cladaras et al. 1989). This protein binds to an RNA element in unspliced or partially spliced HIV-1 RNAs (these include the Gag, GagPol, Env, Vif, Vpu and Vpr precursors) called the Rev responsive element (RRE). The Rev-RRE multimer is capable of interacting with the cellular machinery for nuclear export. Rev shuttles back to the nucleus using a nuclear localization signal for further export of viral RNA molecules.

1.3.3 The "accessory" proteins of HIV

The so-called "accessory" proteins encoded within the HIV genome include vpu, vpr, vif, and nef are not essential for virus replication in culture. Vpu (**v**iral **p**rotein **u**) is unique to HIV-1 and functions to both enhance the release of virus particles from the plasma membrane as well as promote the degradation of CD4 through a ubiquitin/proteasome pathway (Strebel, Klimkait et al. 1988; Margottin, Bour et al. 1998). Vpr (**v**iral **p**rotein **r**) also is incorporated into virions and is proposed to be involved in the transport of the viral pre-integration complex to the nucleus

(Karni, Friedler et al. 1998; Popov, Rexach et al. 1998). Additionally, Vpr efficiently induces cell-cycle arrest in G_2 by inhibition of the $p34^{cdc2}$-cyclin B kinase complex (He, Choe et al. 1995; Re, Braaten et al. 1995). Vif (**v**iral **i**nfectivity **f**actor) is believed to be necessary for the production of infectious virus. A mutation in *vif* results in low virus titers (Fisher, Ensoli et al. 1987; Strebel, Daugherty et al. 1987). This infectivity defect was observed in some cell types, but not others, leading to the hypothesis that some cells might have a cellular protein with a compensating activity (Fan and Peden 1992; Gabuzda, Lawrence et al. 1992; Michaels, Hattori et al. 1993; von Schwedler, Song et al. 1993). Nef was believed to be non-essential based on *in vitro* experiments, but analysis *in vivo* demonstrated that this gene product is essential for disease progression. Nef downregulates CD4 and MHC-I from the plasma membrane (Guy, Kieny et al. 1987; Garcia and Miller 1991; Aiken, Konner et al. 1994; Collins, Chen et al. 1998). This serves two functions: it prevents super-infection of cells during replication by removing the CD4 receptor and reduces the ability of CD8 T-cells to identify infected cells by reduction of MHC-I expression at the surface.

While it is clear that a number of factors are important in HIV replication, my work has focused on characterizing the interaction of IN with DNA substrates.

1.4 Integration

The viral encoded integrase (IN) is both necessary and sufficient to catalyze integration (Katz, Merkel et al. 1990; Li and Craigie 2005; Sinha and Grandgenett 2005). Retroviral DNA integration is a multi-step process that occurs in well defined stages. After assembly of a stable complex of IN with specific DNA sequences at the ends of the viral long terminal repeats (LTRs), terminal dinucleotides are removed from each 3' end by endonucleolytic processing. The viral DNA 3' ends are then covalently linked to the host target DNA in a concerted cleavage/ligation reaction. The processing and joining steps from several retroviruses have been analyzed, including avian sarcoma virus (ASV), murine leukemia virus (MLV) and HIV-1 (Brown, Bowerman et al. 1987; Fujiwara and Mizuuchi 1988; Brown, Bowerman et al. 1989; Katzman, Katz et al. 1989; Roth, Schwartzberg et al. 1989; Farnet and Haseltine 1990; Lee and Coffin 1990; Pauza 1990; Farnet and Haseltine 1991; Whitcomb and Hughes 1991). Both steps require a stable complex composed of IN and at least 16 base pairs of both ends of a linear viral DNA. While integrases specifically cleave their respective viral DNA ends, there is little specificity for the target DNA. This results in integration at many sites. By definition then, the binding of viral DNA must be different from the binding of target DNA by IN and must be accounted for in any structural model of the enzyme. In the following sections, I will define in more detail the structural and biochemical properties of the holoenzyme as well as interactions between the enzyme and viral and host DNA substrates. Next I will discuss the biochemistry of the 3' processing and joining followed by host cell proteins that influence the integration reaction. Lastly, I will highlight IN as a chemotherapeutic target and current strategies for development of novel IN inhibitors.

1.5 Integrase Structure and Oligomerization

All IN proteins share a conserved three-domain structure including a N-terminal, catalytic core, and C-terminal domain (Figure 1.4) (Engelman, Englund et al. 1995). The N-terminal domain (residues 1-49) contains a HHCC zinc-binding site (Cai, Zheng et al. 1997; van den Ent, Vos et al. 1999). The IN of HIV-1, HIV-2, and simian immunodeficiency virus (SIV) all have a phenylalanine at position one, and HIV-1 IN has been shown to be subject to degradation by a ubiquitin-proteasome pathway (Mulder and Muesing 2000). This mechanism has been proposed to maintain host chromosome stability and integrity through IN instability. The C-terminal domain (residues 213-288) is less conserved and forms a SH3 fold through three α-helices. This domain contains multimerization determinants as well as non-specific DNA binding activity (Andrake and Skalka 1996). The catalytic core domain (residues 50-212) contains the residues involved in catalysis. The conserved DDE motif in the catalytic core domain chelates the required metal cofactor and forms part of the active site (Polard and Chandler 1995). When mutations are introduced at these residues, all enzymatic activities are significantly impaired (Kulkosky and Skalka 1994).

Retroviral IN can be found in monomer, dimer, and tetramer forms in solution. Dimeric IN can catalyze 3' processing and joining of DNA duplex oligodeoxyribonucleotides, while the homotetramer is required for a concerted DNA integration reaction (Bao, Wang et al. 2003; Faure, Calmels et al. 2005; Li, Mizuuchi et al. 2006). Crystal structures of the individual domains and two-domain fragments have been solved providing information about orientations and interactions of the various domains with each other and with host proteins (Jenkins, Esposito et al. 1997; Wang, Ling et al. 2001; Chiu and Davies 2004; Maroun, Zargarian et al. 2005).

Unfortunately, a crystal structure for the HIV-1 IN holoenzyme in complex with DNA substrates is not available because the enzyme complex is not stable at concentrations required for crystallization. Therefore, little is known about the positioning of the domains in the active oligomer (Petit, Schwartz et al. 1999; Craigie 2001) or intersubunit interaction, though two residues have been implicated in oligomerization (Jenkins, Engelman et al. 1996; Kalpana, Reicin et al. 1999). In the absence of a crystal structure for either HIV-1 IN monomers or higher order oligomers, a number of proposed structures exist in the literature to model HIV-1 IN (Yang, Mueser et al. 2000; Gao, Butler et al. 2001; Podtelezhnikov, Gao et al. 2003; Karki, Tang et al. 2004; Wielens, Crosby et al. 2005). However, these models do not explain all of the cross-linking, mutagenesis, and kinetics data and are not available in the Protein Data Bank (PDB).

To get around this problem, the two domain structures of HIV-1 IN (N-term and CCD, PDB code 1K6Y; C-term and CCD, PDB code 1EX4) were used to assemble a computer model of HIV-1 IN. These structures are shown in Figure 1.4. To validate the computer model, residues predicted to interact with viral DNA were analyzed *in vitro* (Chen, Weber et al. 2006). Molecular dynamics simulations suggest that the C-terminal domains rotate toward the catalytic domain to create a narrower groove for enclosing the viral DNA (Chen, Weber et al. 2006). Additionally, this model accounts for all previously published data of HIV-1 IN, including residues that cross-link viral DNA and mutational analyses that have identified residues important in DNA binding and intersubunit interactions. The computer model of an IN tetramer with 16 base pairs of viral LTR DNA shown in Figure 1.5 is the basis for the *in vitro* analysis of HIV-1 IN described in this work, and is publicly available in the RCSB data bank (PDB code 2G3L).

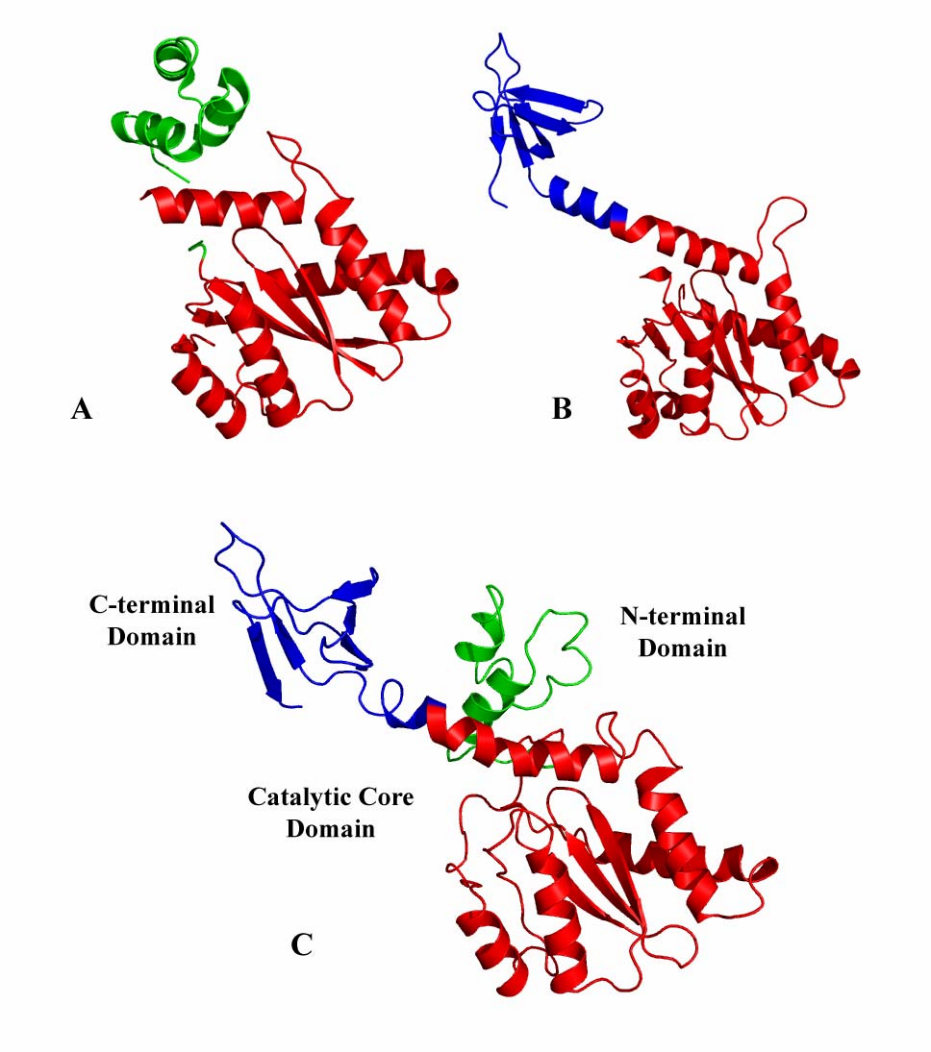


Figure 1.4 HIV-1 IN monomer in cartoon structural representation

HIV-1 IN is modeled using the PyMol software from DeLano Scientific. Crystal structures of the N-terminal and Catalytic Core Domains (PDB code 1K6Y) or the Catalytic Core and C-terminal domains (PDB code 1EX4) are shown (A, B respectively). Using the two solved crystals, a full-length representation of IN was assembled using molecular dynamics simulation, shown in C (PDB code 2G3L). In cartoon representation, the N-terminal domain is colored green (residues 1-49), the Core domain is colored red (residues 50-212), and the C-terminal domain is colored blue (residues 213-270). Residues 271-288 are not represented.

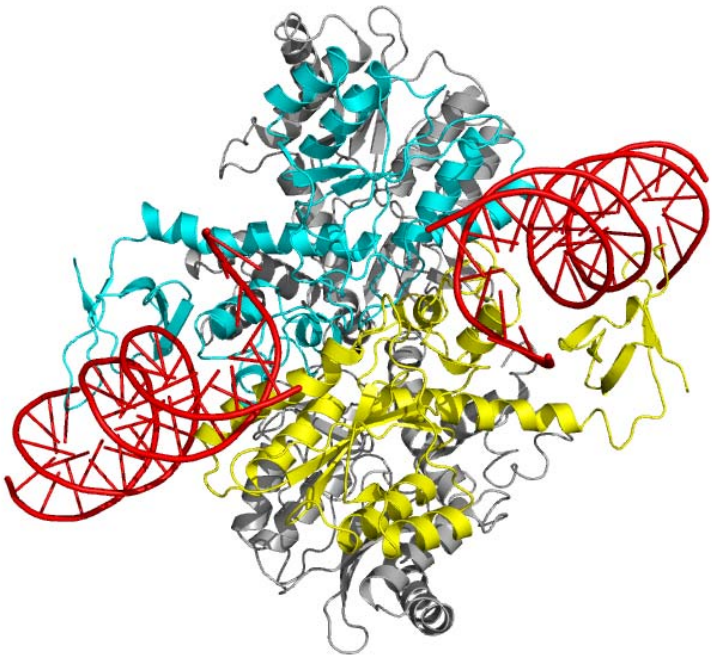


Figure 1.5 HIV-1 IN tetramer model with loaded viral LTR DNA

A tetramer model of HIV-1 IN complexed with viral LTR DNA was assembled by Irene Weber and Robert Harrison. This computer model represents interactions between the HIV-1 IN tetramer with two viral DNA substrates, as well as a groove adjacent to the viral DNA that may accommodate the target DNA substrate. Two subunits that interact with viral DNA are shown in blue and yellow. Two additional subunits are colored gray.

1.6 Enzymatic Activities

The enzymatic activities of IN have been well characterized and can be separated into individual steps *in vitro*. Several of these reactions and their intermediates are used in this work to define and characterize recombinant IN based on predictions from our computer model. An understanding of these catalytic abilities is central to understanding this work; therefore they will be highlighted in the following section.

During viral replication, IN catalyzes two distinct reactions that are important for replication. The first reaction is the removal of two nucleotides from LTR ends immediately adjacent to the conserved 'CA' dinucleotide. This produces a DNA duplex with a 5' overhang and this activity is referred to as 3' end processing. A representation of this activity is shown in Figure 1.6. Additionally, IN catalyzes a concerted integration reaction whereby the processed viral DNA ends are inserted into the host chromosomal DNA. Nucleophilic attack by the 3'-OH of the processed strand of the viral LTR DNA is critical for introducing a break in the chromosomal DNA. This is referred to as the strand transfer or joining reaction and is also depicted in Figure 1.6. When the viral DNA is inserted into the target DNA, it occurs in a staggered manner; the size of the stagger is virus specific. The resulting gapped intermediates are repaired by host proteins introducing the 4- to 6-base pair duplications in the target that flank the integrated viral DNA (Brin, Yi et al. 2000). The determinants for the duplication size are not known. Note that after insertion there are gaps corresponding to the 5' overhanging ends of the viral DNA. These gaps are presumably repaired by host cell enzymes and not IN. The sites of integration in the target DNA are for the most part random (Kulkosky and Skalka 1994).

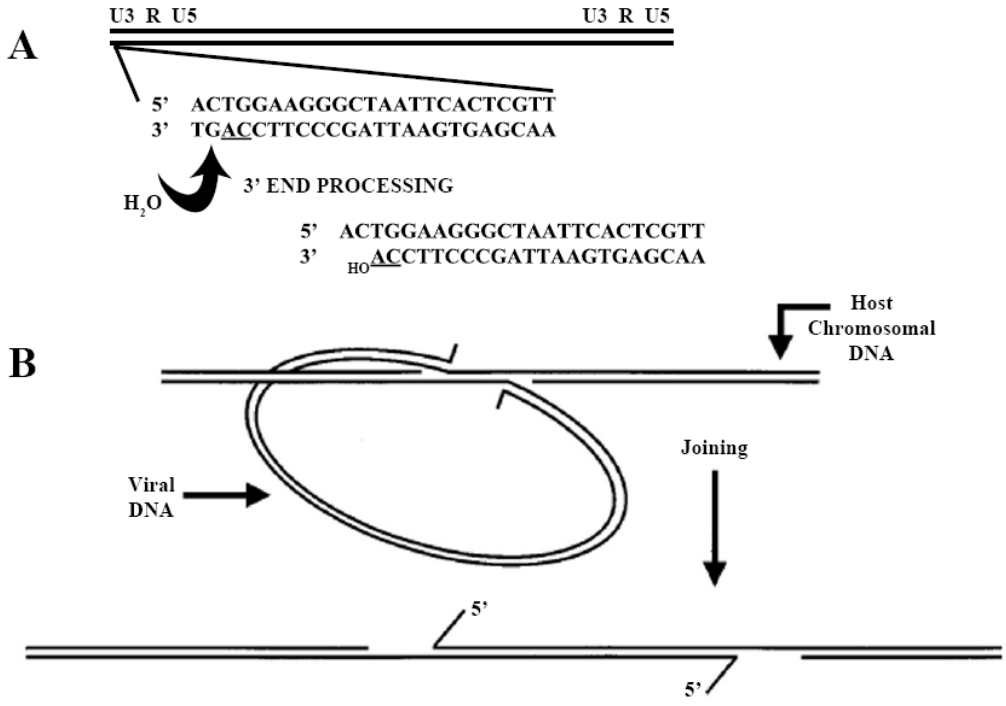


Figure 1.6 Enzymatic activities of IN

The two relevant enzymatic activities for *in vivo* integration are shown (Hindmarsh and Leis 1999). First, removal of two nucleotides immediately adjacent to the conserved CA-dinucleotide (underlined) results in a DNA substrate with a 5'-overhang (A). The exposed hydroxyl group is involved in the nucleophilic attack of the acceptor DNA during the insertion and joining reaction, which is illustrated in B. The insertion produces a duplication in the acceptor DNA and the gapped intermediate is repaired by host enzymes.

Much of the information we have on the molecular mechanism of integration comes from the use of *in vitro* reconstituted systems employing duplex oligodeoxyribonucleotides (oligos) where the products of the reaction are separated by gel electrophoresis. ASV IN, for example, catalyzes the specific removal of the two bases from the 3' end of the strands adjacent to a conserved CA dinucleotide using 15 base pair substrates corresponding to either viral DNA end (called U3 or U5) (Katzman, Katz et al. 1989; Kukolj and Skalka 1995). This *in vitro* reaction approximates the 3' end processing activity of IN. For HIV-1 duplex oligo substrates of comparable size, nucleotide substitutions in the U5 and U3 ends were shown to effect one or both of the catalytic functions of HIV-1 IN. Substitutions in the HIV-1 U5 region, for example, inhibit 3' processing (e.g. positions 1-6 and 9-11) or not (e.g. positions 12-14) (Esposito and Craigie 1998). Other changes at specific nucleic acid positions in the HIV-1 U3 and U5 ends affect each of the catalytic reactions, though changes in one end have a more pronounced effect than in another.

Similar oligo substrates were used to demonstrate the strand transfer reaction, where one oligo integrated into another thereby increasing the size of the radiolabeled substrate after 3' end processing (Craigie, Fujiwara et al. 1990; Katz, Merkel et al. 1990). This reaction is only a model of the joining reaction observed during replication because the DNA serves as both donor and acceptor. The strand transfer reaction can be measured independently from 3' processing using a "preprocessed" duplex DNA oligo substrate in which the two terminal bases adjacent to the CA dinucleotides are removed. Nevertheless, most *in vitro* reactions do not display the concerted nature of the DNA integration reaction as it was first described where deletions were placed 5' to the conserved CA dinucleotides in the MLV U3 LTR region, and when transfected into cells the processing of both LTR ends were affected adversely (Murphy and Goff 1992).

This finding implies that the two ends of the viral DNA were brought together so that mutations in one affected the processing of the other, though the actual insertion of each DNA end into the target DNA may be sequential (Li, Mizuuchi et al. 2006).

Several assay systems that display concerted DNA integration properties have been described and used to demonstrate changes in retroviral DNA integration are context dependent. For example, base pair substitutions placed into the HIV-1 U5 or ASV U3 ends (referred to as the "dominant" LTR ends) caused decreases in the rate of catalysis. In contrast, comparable substitutions at the same positions in the "non-dominant" ends are associated with changes in mechanism from concerted (two-ended viral DNA insertion) to non-concerted (one-ended viral DNA insertion) integration (Fitzgerald, Vora et al. 1992; Lutzke, Vink et al. 1994; Vora, McCord et al. 1994; Aiyar, Hindmarsh et al. 1996; Vora, Chiu et al. 1997; Masuda, Kuroda et al. 1998; McCord, Chiu et al. 1999; Brin and Leis 2002; Brin and Leis 2002; Li and Craigie 2005; Li, Mizuuchi et al. 2006).

IN is capable of catalyzing the reverse reaction referred to as disintegration. It involves the resolution of a hybrid DNA substrate representing the joining intermediate into separate components that represent the viral and target DNAs. This DNA substrate has been effectively used in cross-linking experiments to identify amino acid residues or peptides that specifically interact with portions of the DNA substrate (Heuer and Brown 1997). This polynucleotidyl transferase activity of IN has not been observed *in vivo*, and is not believed to be relevant for replication.

1.7 Host Cell Interacting Proteins

The IN complex is comprised of the viral IN, viral and host DNA substrates, and possibly host proteins. A number of cellular proteins have been implicated in the integration and replication of HIV-1. Some of these proteins are lentiviral specific (e.g. LEDGF/p75) whereas other proteins affect retroviruses as a whole family (e.g. BAF). The components of the PIC still remain to be elucidated, and it is possible that other proteins will be identified.

1.7.1 High mobility group protein A1

High mobility group chromosomal protein A1 (HMGA1, formerly HMG-I(Y)) stimulates integration reactions in vitro (Aiyar, Hindmarsh et al. 1996; Farnet and Bushman 1997). HMG proteins have interaction domains for binding DNA and protein and are involved in nuclear processes including transcription. It is believed that HMGA1 functions in pre-integration complexes and aids in the unwinding of the LTR termini (Hindmarsh, Ridky et al. 1999; Li, Yoder et al. 2000). The role of HMGA1 is still unclear as there is no direct interaction detected between HMGA1 and HIV-1 IN in a co-immunoprecipitation assay (Hindmarsh, Ridky et al. 1999). Furthermore, a genetic knockout of the HMG gene family in a cell line does not appear to affect viral replication (Beitzel and Bushman 2003; Maroun, Delelis et al. 2006).

1.7.2 Barrier-to-autointegration factor

A protein was originally isolated from the pre-integration complexes of Moloney murine leukemia virus and shown to limit autointegration of viral DNA (Lee and Craigie 1994; Lee and Craigie 1998). This barrier-to-autointegration factor (BAF), found as a dimer in solution, was later shown to be involved in HIV integration (Cai, Huang et al. 1998; Chen and Engelman 1998). Reconstitution of salt-stripped PICs with recombinant BAF restores integration activity *in vitro,* but this activity is not due to BAF stimulating the enzymatic activity of IN (Chen and

Engelman 1998; Carteau, Gorelick et al. 1999). This protein has been shown to bridge DNA as well as interact with the nuclear lamina protein LAP-2α, suggesting a role for BAF in nuclear structure under normal conditions (Furukawa 1999; Zheng, Ghirlando et al. 2000).

1.7.3 Integrase interactor 1

The Integrase interactor 1 (INI1) protein was first isolated using a yeast-two-hybrid screen of IN and a cDNA library derived from a macrophage-monocyte cell line (Kalpana, Marmon et al. 1994). The human INI1 protein is a homolog of the yeast transcriptional activator SNF5 and has been shown to be a part of the human SWI/SNF complex (Wang, Cote et al. 1996). INI1 has three highly conserved regions: including two direct imperfect repeats (Rpt1 and Rpt2), a C-terminal coiled-coil domain and a homology region 3 (Morozov, Yung et al. 1998). Overexpression of the integrase-interaction domain (residues 183-294) functions as a dominant-negative inhibitor in cells (Yung, Sorin et al. 2001). Expression of full-length INI1 in cells that are genetically null for this protein results in an increase in viral particle production (Yung, Sorin et al. 2001). Additionally, HIV-1 PICs have been visualized after infection in cells and co-localize with INI1 prior to nuclear translocation (Turelli, Doucas et al. 2001). However, the role of INI1 in HIV-1 replication is still unclear.

1.7.4 Lens epithelium derived growth factor

The lens epithelium derived growth factor (LEDGF/p75) has been shown to interact directly with HIV-1 IN by co-immunoprecipitation and yeast-two-hybrid analysis (Cherepanov, Maertens et al. 2003; Emiliani, Mousnier et al. 2005). LEDGF is a member of the hepatoma-derived growth factor family and has chromatin-tethering activity, suggesting that LEDGF is biologically relevant within the host cell nucleus (Vanegas, Llano et al. 2005). A crystal structure of the LEDGF/p75 integrase binding domain (IBD) bound to the catalytic core domain of HIV-1

IN was solved and the critical residues for this interaction were within the 165-173 region of IN and residues Ile^{365}, Asp^{366} and Phe^{406} of the LEDGF IBD (Cherepanov, Sun et al. 2005). LEDGF has been shown to stimulate IN activity *in vitro* (Raghavendra and Engelman 2007). Another group demonstrated that when LEDGF is added after 3' processing (and presumably after stable IN complex formation) there is no stimulation (Yu, Jones et al. 2007). In cells that have been treated with a small interfering RNA (siRNA) targeted to LEDGF/p75, the nuclear localization of IN and co-localization to chromosomes is abolished (Maertens, Cherepanov et al. 2003). Additionally, LEDGF/p75 contains a nuclear localization sequence (Maertens, Cherepanov et al. 2004). It is proposed that LEDGF interacts with IN to promote nuclear import, aid in chromosomal tethering, influence the site of integration, and stimulate integration of the viral genome into host chromosome.

1.7.5 Histone acetylase p300

The acetylation of IN by p300 has been shown to occur at target lysine residues in the C-terminus of HIV-1 IN (Cereseto, Manganaro et al. 2005). These modifications are thought to influence affinity of IN for target DNA. It has been postulated that this modification, in concert with acetylation of the nucleosomes resulting in uncoiling of chromatin, promotes the integration of lentivirus DNA into actively transcribed regions of the host genome (Cereseto, Manganaro et al. 2005). Studies have also implicated that regions of active transcription are preferential sites for integration, and proteins that interact with these regions can influence integration (Schroder, Shinn et al. 2002).

1.8 Integrase Holoenzyme Properties

A homo-tetramer model for HIV-1 IN with DNA representing 20 base pairs of the U3 and U5 termini was assembled using structural and biochemical data and molecular dynamics simulations (Chen, Weber et al. 2006). Construction of this model used separate two-domain crystal structures of the N-terminal domain and catalytic core (PDB code 1K6Y) and the catalytic core with the C-terminal domain (PDB code 1EX4). Superimposition of the two crystals at the catalytic core led to the assembly of the three domain structure (Figure1.4). The tetramer was assembled using crystal lattice contacts. The viral DNA was placed in the tetramer using critical positions derived from transposase structural data and use of a molecular dynamics simulation. The viral DNAs face one another with their 3' ends positioned adjacent to the active site Asp residues. To validate the model, Chen et al. identified amino acids that were spatially within 10Å from the viral DNA ends. Using a structural alignment of HIV-1, SIV, and ASV INs, unique residues were identified. Construction and purification of 16 HIV-1 enzymes in which the amino acids from ASV were placed into the structurally related positions of HIV-1 IN was carried out, and the chimeric enzymes were tested for changes in specificity for 3' processing of the viral DNA ends using duplex oligo substrates. This analysis demonstrated that there are multiple HIV-1 IN amino acid contacts with the viral DNA and that substitution of ASV IN amino acids at many positions conferred the partial ability to cleave ASV substrates with a concomitant loss in cleavage of the homologous HIV-1 substrate. Substitutions at HIV-1 IN residues that changed specificity include Val^{72}, Ser^{153}, Lys^{160}, Iso^{161}, Gly^{163}, Gln^{164}, Val^{165}, His^{171}, Leu^{172} (Chen, Weber et al. 2006). Figure 1.7 highlights the viral DNA bound in one trench of the IN tetramer, and the labeled residues represent those that were shown to change the specificity of the enzyme for substrate DNA.

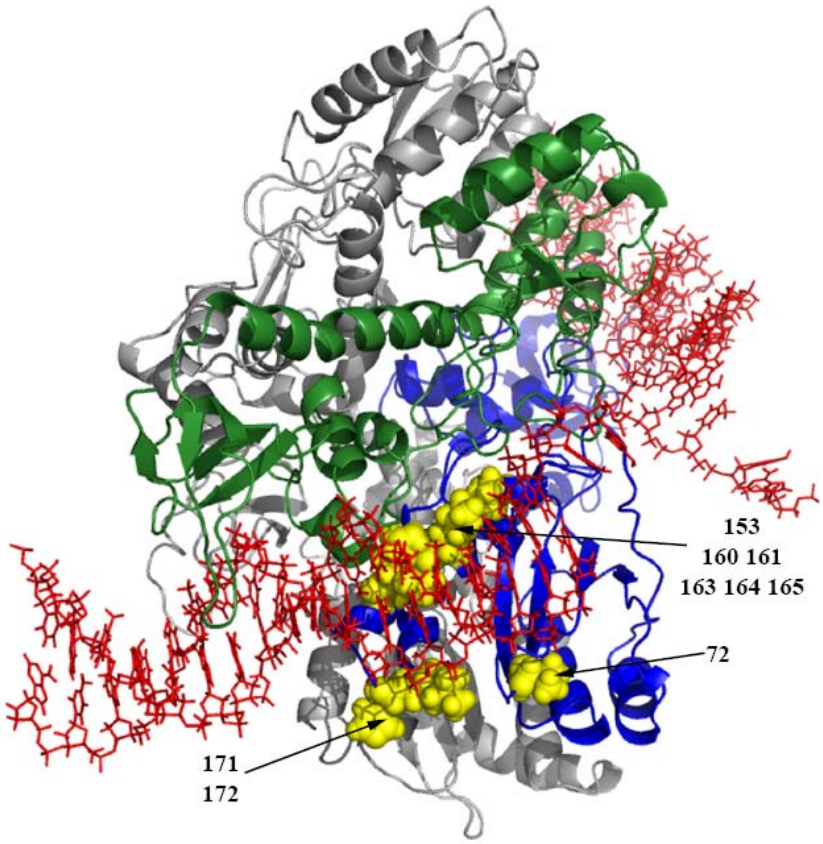


Figure 1.7 Residues that affect HIV-1 IN specificity

The HIV-1 IN tetramer is shown in cartoon representation with the two viral LTR DNA ends represented in red. Residues that are reported to affect the specificity of the LTR:DNA interaction are shown in space-fill representation and colored yellow (Chen, Weber et al. 2006). The residues are Val^{72}, Ser^{153}, Lys^{160}, Iso^{161}, Gly^{163}, Gln^{164}, Val^{165}, His^{171}, and Leu^{172} (using HIV-1 IN numbering).

1.8.1 Viral DNA interaction

Recognition of viral DNA by integrase is specific; however, there is cross-recognition within respective virus families. For example, HIV-1 IN will 3'-process LTR end sequences derived from simian immunodeficiency virus (SIV), another member of the primate lentivirus subgenus. Several techniques including NMR spectroscopy, use of cross-linking agents, and nucleotide analog-modified substrates have been used to probe DNA-IN interactions. The most informative have been cross-linking studies that identified contacts on the enzyme surface with both viral and target DNA, including residues Tyr^{143}, Gln^{148}, Lys^{156}, Lys^{159}, Lys^{160}, Ser^{230}, Glu^{246}, Arg^{262}, Arg^{263}, and Lys^{264} (Jenkins, Esposito et al. 1997; Drake, Neamati et al. 1998; Esposito and Craigie 1998; Heuer and Brown 1998; Gao, Butler et al. 2001). A representation of these residues in our HIV-1 IN tetramer model is shown in Figure 1.8. These studies used a variety of DNA substrates and cross-linking agents, and many of these residues have been confirmed in secondary studies to interact with DNA (either viral or target) through additional mutational and/or chemical investigation. Analysis of nucleotide analog modified substrates to identify contacts between the viral DNA and the enzyme surface has also been informative (Agapkina, Smolov et al. 2006). In this case, a map was assembled of putative contacts between the enzyme surface and either the phosphate-backbone or heterocyclic bases. The processing reaction was also determined to require local destabilization of the third A-T base pair (in the C<u>A</u> dinucleotide) for efficient activity (Agapkina, Smolov et al. 2006). Specific IN residues that were identified in this study include K159 and E152, which interact with the N-7 position of the 3rd adenosine on the processed strand. Residues nearer the active site, such as Q148 and Y143, interact with the terminal nucleotides of the processed strand.

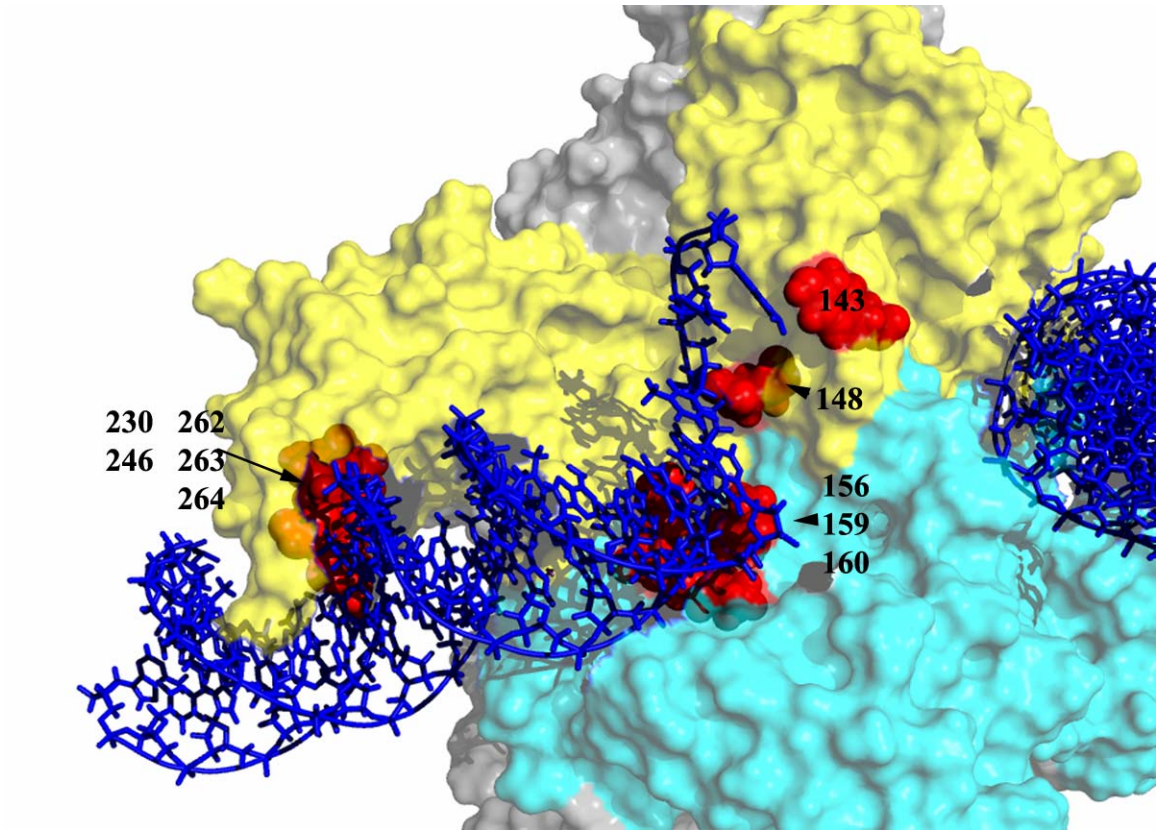


Figure 1.8 Residues that cross-link viral LTR DNA

Residues reported in the literature to cross-link to viral LTR DNA. Interactions between viral DNA and HIV-1 IN have been demonstrated for residues Tyr^{143}, Gln^{148}, Lys^{156}, Lys^{159}, Lys^{160}, Ser^{230}, Glu^{246}, Arg^{262}, Arg^{263}, and Lys^{264} (Jenkins, Esposito et al. 1997; Drake, Neamati et al. 1998; Esposito and Craigie 1998; Heuer and Brown 1998; Gao, Butler et al. 2001). These residues are colored red in space-fill representation and shown only for one LTR end.

1.8.2 Target DNA Interaction

The IN tetramer binds the target DNA in a different trench than the viral DNA and both substrates are bound at the same time during the joining reaction. Because integration occurs at many sites in the target DNA, it is likely that conserved residues such as Lys, Arg, Asn, and Gln would be found in the target DNA site.

The use of cross-linking reagents during the incubation of IN with a non-viral DNA substrate oligo has implicated several peptides representing residues 1-11, 49-69, 139-152, 213-246, and 271-288 to be close enough to the target DNA to cross-link (Heuer and Brown 1998). In the Heuer and Brown study, a hybrid DNA substrate representing the integration intermediate, or Y-substrate, was used to cross-link HIV-1 IN (Heuer and Brown 1997). As shown in Figure 1.9, our model represents a number of these peptides very well for their reported interaction with the hybrid DNA substrate. In particular, Peptide 2 (orange) sits in the active site of the HIV-1 IN holoenzyme and could interact with both the viral and target DNAs. Additionally, Peptide 4 (red) presents as one face of the putative target DNA trench, and may include amino acid residues on the surface that interact with the target DNA substrate.

A few other residues have been identified to interact with the target DNA and influence the joining reaction including residues 124 in ASV (which is structurally related to residue 119 in HIV-1), 130, and 132 (Harper, Skinner et al. 2001; Harper, Sudol et al. 2003; Li and Craigie 2005; Al-Mawsawi, Fikkert et al. 2006). Mutations introduced at these residues result in a loss of the joining reaction with little effect on 3’ processing.

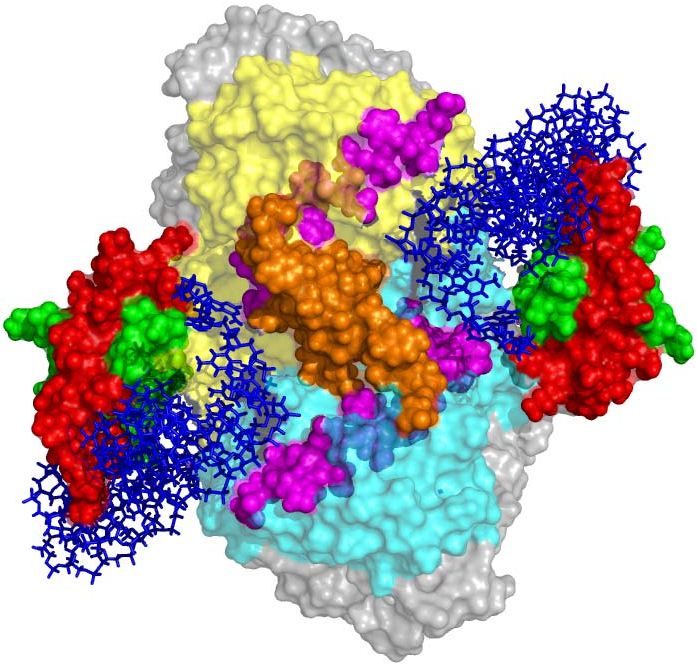


Figure 1.9 Peptides identified by photo-crosslinking and endopeptidase Glu-C digestion

A collection of peptides was previously reported to cross-link to either the HIV LTR or target DNA ends of a hybrid substrate (Heuer and Brown 1997). Peptide 1, residues 1-11, was reported to cross-link to target DNA but is obscured from view in our model. Peptide 2, residues 49-69, is colored magenta and cross-linked to viral and target DNA as did peptide 3, which is colored orange and includes residues 139-152. Peptide 4 only cross-linked to target DNA and is colored red (residues 213-246). Peptides 5, residues 247-270 in green, cross-linked only the viral DNA, but distal to the CA overhang. The final peptide, residues 271-288, cross-linked to viral and target DNA but is not represented in our model.

1.9 Chemotherapeutics: Design and Strategy

Inhibitors of integrase are currently a topic of investigation for antiretroviral therapy. Current anti-retroviral therapies use a combination of inhibitors targeted against the viral enzymes, protease (PR) and reverse transcriptase (RT). Molecules that inhibit viral fusion are also being used in drug regimens. Resistance mutations to the enzyme inhibitors for PR and RT occur frequently enough that escape mutants after selection are common. Therefore, molecules that inhibit integrase are being sought. Several classes of compounds show promise of efficacy against IN at physiologically relevant levels. Diketo acids and naphthyridine carboxamides are two examples of molecules currently under investigation and development as IN inhibitors. An additional bonus of these compounds is that they can be used to help define the different DNA binding sites of IN as many of these drugs have been shown to preferentially disrupt one enzymatic activity but not the other.

The diketo acids are selective inhibitors of IN and act at the level of strand transfer (Hazuda, Felock et al. 2000). The diketo acid moiety showed positive inhibition against IN at nanomolar concentrations, suggesting that these molecules could be developed to function at physiologically relevant levels. Two molecules of this class were used for selection of resistant mutants. Three positions changed in a majority of the viruses that were sequenced. Mutations at residues 66, 153, and 154 were commonly selected. These residues are located near the active site, suggesting that the molecule binds near the catalytic center of the enzyme and presumably does not interfere with 3' processing but affects the joining reaction in a currently undefined manner. Since the initial discovery of this class of compounds, numerous derivatives have been synthesized and examined at their inhibitory potential (Fikkert, Van Maele et al. 2003; Lee and Robinson 2004; Svarovskaia, Barr et al. 2004; Brigo, Lee et al. 2005).

Naphthyridine carboxamides are also small molecules that were selected to inhibit IN at the mechanism of strand transfer (Hazuda, Anthony et al. 2004). These molecules are nanomolar inhibitors, and many derivatives have yielded good pharmacokinetics (Embrey, Wai et al. 2005; Guare, Wai et al. 2006). Similar to the diketo acid selection, one molecule of this set demonstrated good pharmacokinetics and was used for further study. Selection of resistant mutants in the presence of the inhibitor highlighted the putative binding site to be near residues 72, 121, and 125. Although naphthyridine carboxamides are derivatives of diketo acids, there is no cross resistance for the selected mutations. Naphthyridine carboxamides and diketo acids therefore have distinct binding sites for their inhibitory mechanism.

Several of the mutations that show up in drug resistant enzymes affect 3' processing and not just the joining reaction. This would result in partially defective INs that would place a pressure for the selection of second site mutations that compensate for lost 3' processing activity by the initial drug resistance mutation. For example, position 153 is important in the recognition of the viral DNA, and mutation at this position only mildly affects the activity of the enzyme. As such, this model would predict that this mutation would be found by itself when subjected to drug selection with little replication defect (Lee and Robinson 2004; Lu, Limon et al. 2005). In contrast, mutation at position 72 affects the catalytic activity of the 3' processing as well as strand transfer reactions in vitro (Chen, Weber et al. 2006). Not surprisingly, a mutation at this residue is found in combination with mutations at positions 121 and 125 (Hazuda, Anthony et al. 2004). It has been demonstrated that the T125S substitution increases the 3' processing of HIV-1 duplex (Chen, Weber et al. 2006) as well as the joining of a HIV-1 preprocessed substrate (unpublished) so that substitutions at this position could compensate for the low activity caused by the mutation at Val^{72} (Chen, Weber et al. 2006). In identifying critical residues involved in

the particular enzymatic reactions, investigation with small molecules has also proven useful in understanding the biology of the enzyme. Studies involving the binding of high affinity drugs that target the enzymatic activities of IN have demonstrated that binding of viral DNA promotes a distinct active conformation (Espeseth, Felock et al. 2000; Grobler, Stillmock et al. 2002; Pommier, Johnson et al. 2005).

The US FDA approved raltegravir (MK-0518, isentress) in late 2007. This compound serves as a proof-of-principle for the development of therapeutic molecules that target IN. The compound is a new class of compound (pyrimidione carbosamide) and selectively inhibits the strand transfer activity of IN. Two recent studies highlighted the application of the IN-inhibitor raltegravir as being effective in patients with triple-class drug resistance for whom antiretroviral therapy had failed (Cooper, Steigbigel et al. 2008; Steigbigel, Cooper et al. 2008). This is significant in that there are now more options for prophylaxis of HIV^+ individuals, as well as pharmacologics for patients infected with strains resistant to current protease and reverse transcriptase inhibitors. A second compound, GS-9137, was originally derived from antibiotic quinolones (Sato, Motomura et al. 2006). Research into the therapeutic potential of this compound is ongoing, with favorable data being reported (DeJesus, Berger et al. 2006). These data support the idea that there is still room in the therapeutic treatment of HIV-infected patients for new classes of drugs that will be effective, particularly with respect to targeting the retroviral IN. Some analysts predict the HIV therapeutic market at $12B by 2016. The benefits of advancement in this area will be profound in the treatment of HIV^+ patients for prolonging survival post-infection and improving quality of life.

1.10 Conclusion

A great wealth of knowledge and information has been gathered about the biochemistry and structure of retroviral IN recently. Identification of residues that are responsible for interacting with a specific DNA sequence is one recent development, as well as the assembly of a predictive computer model of the IN holoenzyme. These tools advance the understanding of the enzyme system in addition to open new areas of research. We hypothesize that we can use the computer model of the HIV-1 IN tetramer to identify additional residues that specifically interact with the viral DNA substrate. Furthermore, we can use this computer model to predict residues that are involved in the binding of the target DNA substrate. This work has validated the model as an accurate representation of the IN holoenzyme and has withstood the rigors of *in vitro* experimentation. This model should be used to investigate and design specific small molecule inhibitors of IN activity, rather than the current approach of screening molecules in a high-throughput format.

CHAPTER 2

MATERIALS AND METHODS

2.1 Reagents

[γ-^{33}P]-ATP (2500 Ci/mmol) was purchased from Perkin Elmer Life Sciences. HiTrap™ Chelating HP resin and HiTrap™ Heparin HP resin were purchased from GE Healthcare Life Sciences (Piscataway, NJ). T4 polynucleotide kinase was from USB (Cleveland, Ohio). IPTG was from Roche (Indianapolis, IN). The Slide-A-Lyzer Dialysis cassette (10kD MWCO) was obtained from Pierce (Rockford, IL). CentriPrep centrifugal filter devices with YM-10 MW membranes were from Millipore (Bedford, MA). Acrylamide and bisacrylamide solutions were from Bio-Rad (Hercules, CA). SimplyBlue Safe Stain was from Invitrogen (Carlsbad, CA). DE81 filters were purchased from Whatman International Ltd (Kent, UK). Unless specified, all restriction enzymes were purchased from New England Biolabs (Beverly, MA). ASV IN was provided by Dr. Ann Skalka (Fox Chase Cancer Center, Phil, PA). An expression construct for HIV-1 IN 1-288 residues (p28bIN-3CS-F185H) was also obtained from the laboratory of Dr. Ann Skalka and contains the wild type NY5 HIV-1 sequence (Parke Davis clone) from the *Nde*I to the *Hind*III site in the pET28b plasmid vector. The IN sequence encodes 4 substitutions (C56S, C65S, C280S, and F185H) to increase solubility and a six amino acid His-tag separated from the N-terminus of IN by a thrombin cleavage site. Two translation stop codons were added after residue D288.

2.2 Bacterial Strains and Growth Conditions

The protein expression host strain BL21(DE3) were purchased from Novagen (Madison, WI). The BL21(DE3) bacterial strain is sufficient for easy induction and high-level production of recombinant HIV-1 IN under the control of the T7 RNA Polymerase promoter. Selection of mutagenesis for chimera construction was performed in Supercompetent XL1-Blue

cells from Stratagene (LaJolla, CA). Confirmed clone DNA is stored in DH5α at -80^{o}C using chemically competent cells from Invitrogen (Carlsbad, CA). The storage media is a mix of LB and Glycerol at a ratio of 1:1. Unless otherwise noted, bacteria were selected for using LB+Kan^{50} media at 37^{o}C.

2.3 Preparation of Duplex Oligo Substrates

The following oligos were used in the integrase 3' processing activity assay:

HIV-1 U5
5' TGTGGAAAATCTCTAGCAGT 3' (+)
3' ACACCTTTTAGAGATCGTCA 5' (-)

ASV U3
5' GTATTGCATAAGACTACATT 3' (+)
3' CATAACGTATTCTGATGTAA 5' (-)

MLV U5
5' TAGTCAGCGGGGTCTTTCATT 3' (+)
3' AATGAAAGACCCCGCTGACTA 5' (-)

SIV U5
5' GCAGGAAAATCCCTAGCAGT 3'(+)
3' CGTCCTTTTAGGGATCGTCA 5'(-)

The following oligos were used in the integrase strand transfer activity assay to simulate pre-processed LTR ends:

HIV-1 pre-processed U5

5’ CAGTGTGGAAAATCTCGAGCA 3’ (+)
3’ GTCACACCTTTTAGAGCTCGTCA 5’ (-)

ASV pre-processed U3

5’ GAGTATTGCATAAGACTACA 3’ (+)
3’ CTCATAACGTATTCTGATGTAA 5’ (-)

The plus-strand substrates (100pmoles, containing the conserved ‘CA’ dinucleotides) were 5’-end labeled using T4 polynucleotide kinase (30U) and [γ-^{33}P]-ATP as previously described (Chen, Weber et al. 2006). The specific activity of the radiolabeled substrates was diluted to 10^5 cpm/pmol using unlabeled plus-strand oligo and the mixture was purified and recovered from a 20% denaturing polyacrylamide gel. Duplex oligos were formed by annealing to a three-times excess of unlabeled complementary strand as described (Chen, Weber et al. 2006).

2.4 Construction of HIV-1 IN Mutants

Mutagenesis oligos were obtained from Integrated DNA Technologies Inc (Coralville, Iowa) and are listed in Figure 2.1. The mutations were constructed using QuikChange® Site-Directed Mutagenesis Kit from Stratagene (La Jolla, CA) according to manufacturer's directions. This kit was selected because it combines a high-fidelity polymerase with a small amount of starting template DNA to yield high mutation efficiency. Briefly, mutagenesis of the template plasmid was performed by mixing 50ng of template DNA with 125ng of each mutagenic primer. Codon preferences for *E. coli* were used in the oligo design. The reaction mixture was cycled using the program recommended by the protocol. Denaturing of the DNA duplex at 95 ^{o}C was followed by an annealing step at 55 ^{o}C. Extension for one minute per kilobase of DNA template allowed for synthesis of the entire plasmid construct. For amino acid changes these steps were cycled eighteen times. Digestion of the template DNA was achieved by addition of 30U of DpnI with incubation at 37^{o}C for one hour. Selection of successful clones was performed by chemical transformation into XL-1 Blue Ultracompetent E. coli. The presence of all mutations was confirmed by sequencing the complete individual DNA clones. At least 100ng/ul samples of clone DNA were submitted for sequencing. The sequencing primers for HIV-1 IN mutation are shown in Figure 2.2.

The Wizard® Plus SV Miniprep DNA Purification System (Promega, Madison, WI) was used to prepare DNA for cloning. This kit is efficient for quick purification of desired plasmid DNA from E. coli cultures.

Amino Acid Substitutions [a]	Mutagenic Oligodeoxyribonucleotide [b]
S39T K42H	5'-GGAAATCGTGGCT**ACC**TGCGAC**CAT**TGCCAATTGAAAGG-3'
Q44N	5'-CTTGCGACAAATGC**AAC**TTGAAAGGTGAAG-3'
L68E E69P	5'-GGACTCTACTCAT**GAACCG**GGTAAAGTTATCC-3'
L74A	5'-GAGGGTAAAGTTATC**GCG**GTTGCTGTTCACGTG-3'
K156R	5'-ATCGAATCTATGAAC**CGT**GAGCTCAAAAAAATC-3'
E170M	5'-GTACGTGATCAGGCT**ATG**CACCTGAAAACCGCG-3'
Y227I	5'-CCGTGTATAC**ATT**CGTGACTCTAG-3'
D229I S230E	5'-CGTGTATACTACCGT**ATCGAA**AGAGACCCGGTTTGG-3'
D253N	5'-GCTGTTGTTATCCAG**AAC**AACTCTGACATCAAAG-3'
N254D	5'-GTTGTTATCCAGGAC**GAT**TCTGACATCAAAGTG-3'
K258I R262S	5'-CTCTGACATC**ATT**GTGGTACCG**TCT**CGTAAAGCTAAAATC-3'
K211S	5'-GATATCCAGACT**AGC**GAACTGCAGAAACAG-3'
K219S	5'-CAGAAACAGATCACT**AGC**ATCCAGAACTTCCG-3'
Q221S	5'-CACTAAAATC**AGC**AACTTCCGTG-3'

Figure 2.1 Mutagenic Primers for Generation of Chimeric HIV-1 IN

[a] Amino acid substitutions are based on alignment shown in Supplemental Figure 1. For chimeras that were generated for altered specificity, amino acid substitutions are based on residues in structurally related positions in ASV. For target DNA residues, (211, 219, and 221) substitutions to Ser were made.

[b] The bolded and underlined nucleotides indicate the bases involved in the mutagenic changes to introduce changed amino acid residues. Only the plus strand sense oligos are shown.

Primer Name	Primer Sequence
pET-28b(up)	GAT ATA CCA TGG GCA GCA GCC
pET-28b(down)	ACT CAG CTT CCT TTC GGG CTT
HIV IN seq1	CC GGT CAG GAA ACT GCT TAC TTC
HIV IN seq2	GGT GGT TAC AGC GCT GGT GAA CG

Figure 2.2 HIV-1 IN Sequencing Primers for pET-28b vector

Purified clone DNA was submitted for sequencing using these oligos for sequencing the HIV-1 IN coding sequence. Clones with the desired mutation, in the absence of additional mutations identified by sequencing, were used for purification and *in vitro* analysis.

2.5 Purification of HIV-1 IN chimeras

His-tagged HIV-1 IN chimeras were purified as previously described (Yi, Asante-Appiah et al. 1999) with some modification. Briefly, proteins were induced in BL21 (DE3) cells at 20^{o}C by adding IPTG to 0.5mM after the bacteria had grown to optical density at 600nm of 0.8. The bacteria were lysed in 25mM Bis Tris, pH 6.1, 1M KCl, 1M urea, 1% thiodiglycol, 5mM imidazole and then filtered through 0.22μm membrane from Millipore (Billerica, MA). The lysate fraction was applied to a HiTrap^{TM} Chelating HP Ni-affinity column (5ml) and IN was eluted with a 5mM to 1.0M linear imidazole gradient. Fractions containing IN, as detected by absorbance at 280nm and confirmed by SDS-PAGE with staining using the SimplyBlue Safe Stain, were applied to a HiTrap^{TM} Heparin HP column (5ml) and eluted with a 0.25 - 1.0M linear KCl gradient. Selected fractions were concentrated using a Centriprep filter with YM-10 MW membrane and then dialyzed against 25mM Bis Tris, pH 6.1, 0.5M KCl, 1% Thiodiglycol, 1mM DTT, 0.1mM EDTA, and 40% glycerol. The purified protein was aliquoted and stored at –80^{o}C. The protein concentration was determined using a Bio-Rad protein assay as described by the manufacturer. The purification procedure is summarized in a schematic representation in Figure 2.3.

Figure 2.3 Schematic representation of recombinant HIV-1 IN purification.

[a] Mutagenesis construction of recombinant HIV-1 IN was performed as described in Materials and Methods. The mutagenic oligos are cataloged in Figure 2.1. After PCR mutagenesis, clones were screened after transformation into XL-1 Blue Supercompetent cells from Promega.

[b] Colonies were selected for culture and confirmation of desired mutation by sequencing before inoculation into LB culture for protein induction. Culture at 20°C overnight before subculture and growth to log phase (OD_{600} = ~0.8) for addition of IPTG to final concentration of 0.5mM. Cultures were then grown for 16h at 20°C for collection.

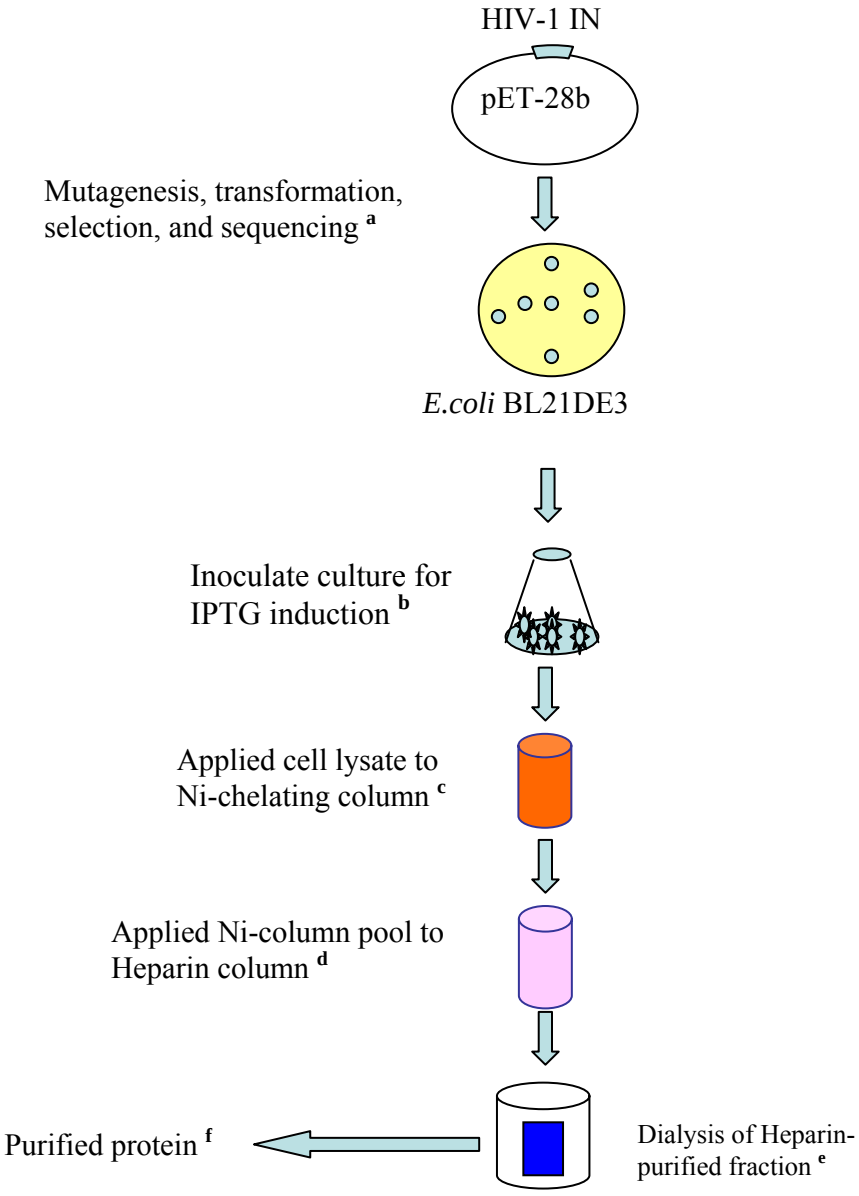
[c] Cell lysates were first applied to a chelating Nickel column and eluted by linear imidazole gradient.

[d] Fractions containing IN were then applied to a Heparin column and eluted by linear salt gradient.

[e] IN containing fractions were then concentrated through a centrifugal filter device before loading into dialysis cassette.

[f] Precipitate was removed from purified fractions by high-speed centrifugation before calculating protein concentration and verifying enzymatic activity.

Purification Scheme of Recombinant HIV-1 IN



2.6 Integrase 3' end processing or strand transfer assay

The processing reactions for the HIV-1 U5 or ASV U3 LTR substrates were carried out as previously described (Jenkins, Engelman et al. 1996). Figure 2.4 shows the outcome of the 3' processing reaction by the cleavage of the terminal nucleotides from the 3' end of the substrate to create a 5' overhang. Reactions were in a volume of 12µl with 25mM MOPs, pH 7.2, 10mM DTT, 15mM potassium glutamate, 5% PEG8000, 5% DMSO, 500ng of HIV-1 or HIV-1 chimeras and 1pmol of labeled duplex substrate as indicated. Reaction mixtures were assembled from individual components and preincubated overnight at 4^{o}C. To start the processing reaction, $MgCl_2$ was added to final concentration at 10mM and reaction mixtures incubated at 37^{o}C for 90 minutes. The reactions were stopped by the addition of 3µl of Stop buffer (95% formamide, 20 mM EDTA, 0.1% xylene cyanol, 0.1% bromphenol blue), heated to 95^{o}C for 5 minutes, and then placed on ice. Products of the reaction were separated through a 20% polyacrylamide denaturing sequencing gel. Labeled reaction products were visualized using KODAK MR film by exposure overnight. For reactions containing ASV IN the final reaction mixture contained 20mM MOPs, pH 7.2, 3mM DTT, 100µg/ml BSA, 500ng ASV IN, and 1pmol labeled duplex substrates as indicated.

For the strand transfer assay, the reaction conditions are identical to those used for 3' end processing, however the substrates used mimic the preprocessed LTR end (designed with 5'-CA dinucleotide overhang) in order to promote strand transfer. This reaction is represented visually in Figure 2.5. The reaction products were analyzed by denaturing gel electrophoresis (Katzman, Katz et al. 1989; Chen, Weber et al. 2006).

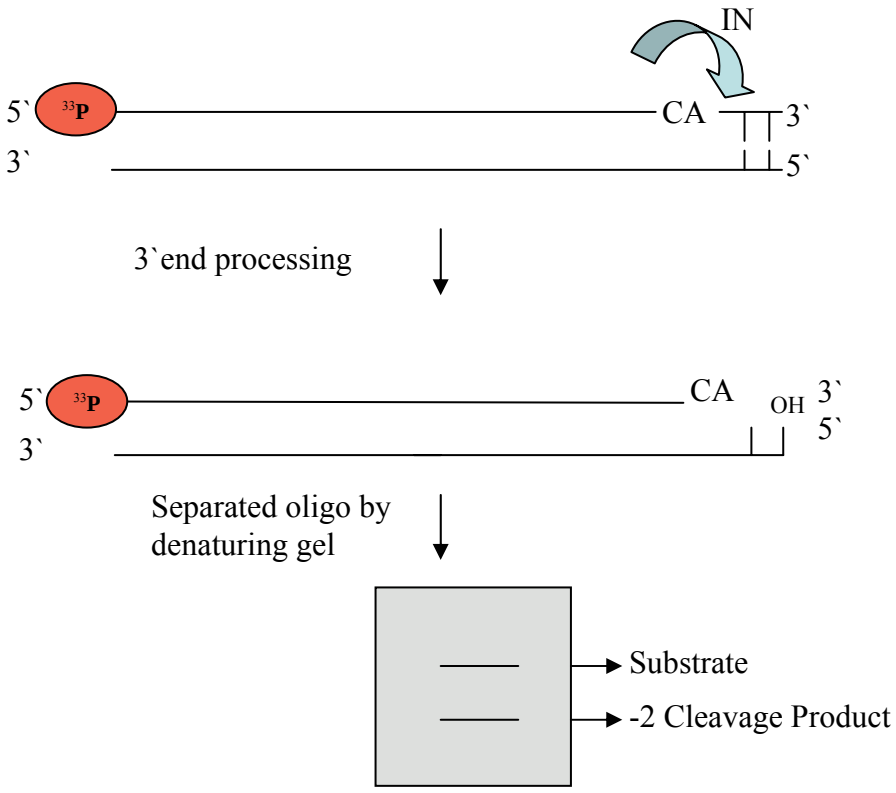


Figure 2.4 HIV-1 Integrase 3' End Processing Assay

Radiolabeled oligo substrates were prepared as described in the Materials and Methods. Once purified, the labeled oligos were mixed with 500ng recombinant HIV-1 for assay. Cleavage of the two terminal nucleotides to generate a 5' overhang generates the hydroxyl group that is critical for the integration of the viral into target DNA. In vitro, the cleavage reaction can be measured by incubating the radiolabeled substrate with purified enzyme and reaction products are separated on a DNA sequencing gel. Substrates that have undergone specific processing are shorter by two nucleotides. Visualization of DNA bands is done by exposure to film.

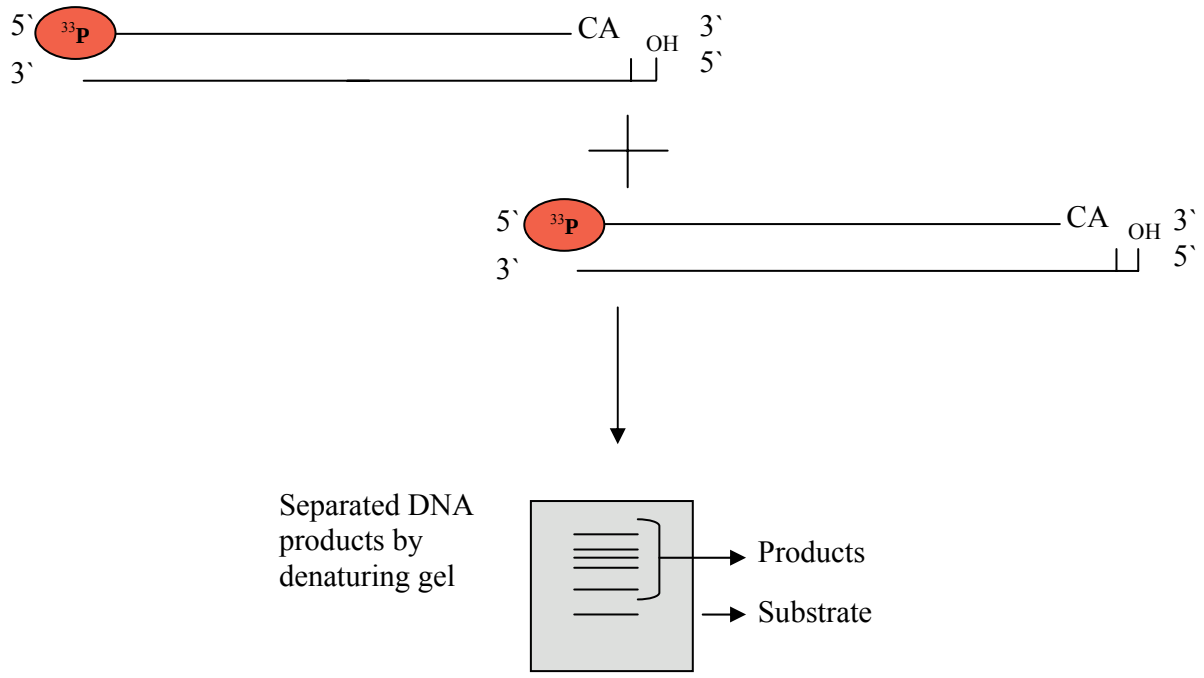


Figure 2.5 HIV-1 Integrase Strand Transfer Assay

Radiolabeled oligo substrates were prepared as described in the Materials and Methods. Once purified, the labeled oligos were mixed with 500ng recombinant HIV-1 for assay. Strand transfer was measured by integration of one oligo into another. Using a pre-processed substrate (simulated by a 5' overhang) allows for the strand transfer activity to be measured in vitro. Strand transfer can only be achieved by loading of substrates into adjacent DNA binding trenches for insertion and joining. After incubation of radiolabeled substrate with purified enzyme, reaction products are separated on a DNA sequencing gel. Strand transfer products migrate more slowly due to increased molecular weight on the gel. Visualization of DNA bands is done by exposure to film.

2.7 Electrophoresis and Imaging

Purified HIV-1 IN protein was visualized by SDS-PAGE electrophoresis. Briefly, 10% polyacrylamide gel solutions were mixed and poured into casting apparatus. This resolving gel was then stacked with a 5% polyacrylamide gel for loading. The wells were flushed with isopropanol and water before loading with protein sample in Tris-Glycine-SDS Running Buffer. The contents of the buffer are 25mM Tris, 192mM glycine, 0.1% SDS. Gels were then run at room temperature at 100V/30mA maximum. Gels were run until the loading dye was run out of the bottom of the gel, and then the gels were transferred to a staining tray for visualization of the protein.

Staining of protein gels was performed using the SimplyBlue Safe Stain from Invitrogen and following the manufacturer's directions for use of a microwave. Briefly, the gel was washed in distilled water before addition of the stain and at each step both the gel and solution were heated to near-boiling using a microwave. Incubation of the gel in the heated distilled water was to wash residual salt and detergent from the gel so as to not interfere with the staining. Heating of the stain solution provided quicker dye incorporation for the protein samples. The stain was washed using distilled water and a 20% NaCl solution before transfer to Whatman paper and drying on BioRad gel dryer.

2.8 Statistical Analysis

Data were analyzed with the GraphPad Prism software suite using a two-tailed Student's t-test for paired data. Statistical significance was established when $p<0.05$. Levels of significance are indicated in the text or figure legends.

2.9 Modeling of LEDGF Complex with IN Tetramer and LTR DNAs

The crystal structure of LEDGF (residues 345-426) bound to the IN catalytic domain dimer (2B4J) was superimposed on the IN catalytic domains in the tetramer model. Then, the program AMMP(Harrison 1993) was used to produce and minimize hydrogen atom positions for the two LEDGF monomers and the monomers were minimized using conjugate gradients with all non-bonded and geometric terms. The all-atom sp4 potential set (Weber and Harrison 1996; Weber and Harrison 1997) was used with the charge generation parameters from Bagossi et al (Bagossi, Zahucky et al. 1999) and a dielectric of 1.0. The LEDGF monomers were combined with the model consisting of the tetramer of full-length IN with Zn and Mg atoms, two 20-mer LTRs, as described in Chen et al (Chen, Weber et al. 2006). The new model with LEDGF was optimized by 500 cycles of conjugate gradients minimization in AMMP to ensure good non-bonded interactions. Figures of the model were made using PyMol (De Lano 2002).

CHAPTER 3

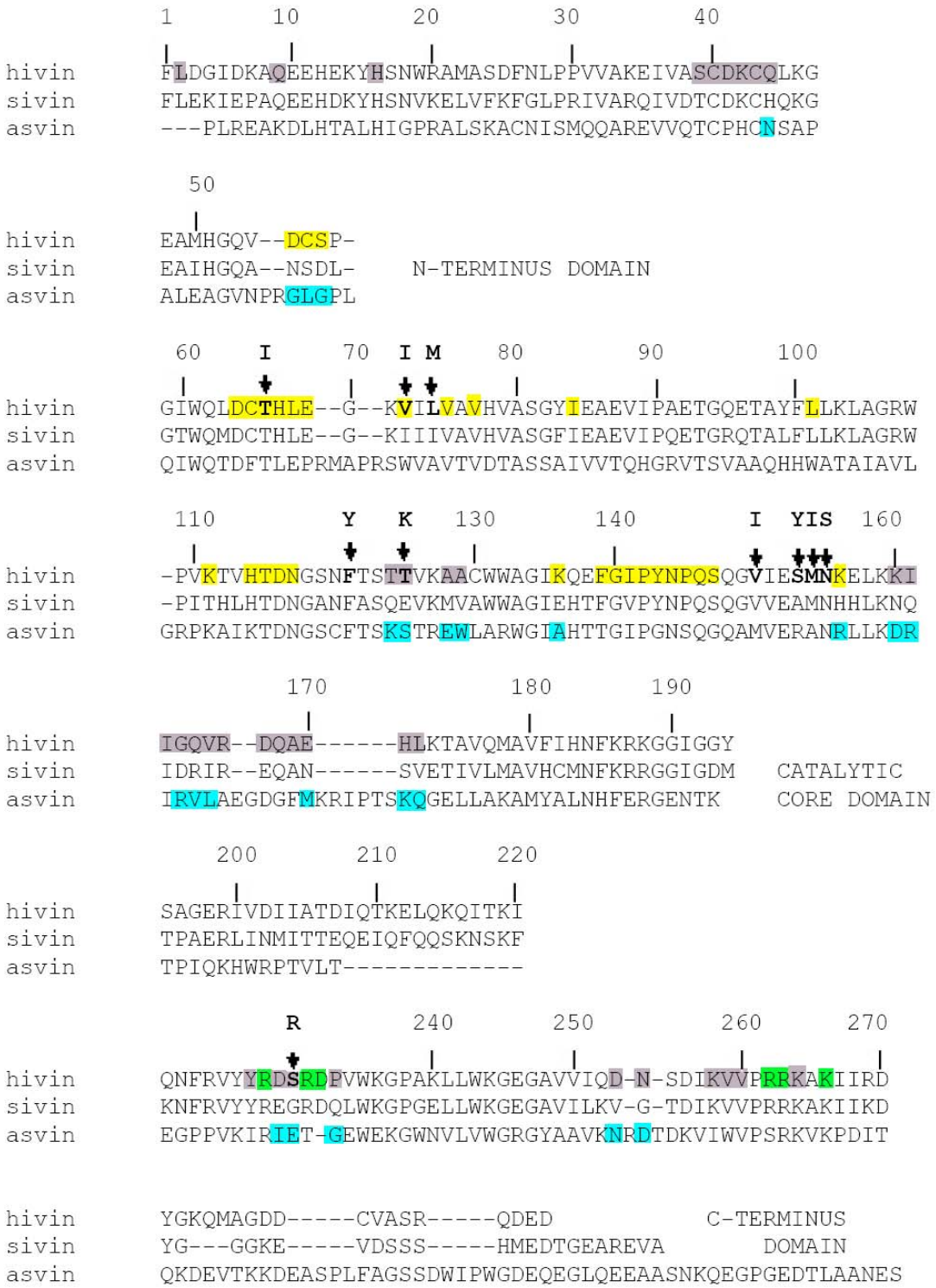
RECOGNITION OF HOMOLOGOUS VIRAL DNA SUBSTRATES

3.1 Overview

The assembled tetramer model of HIV-1 IN provided an opportunity to investigate enzyme/substrate interactions. Residues predicted to bind the viral LTR substrate were examined first because each viral integrase specifically recognizes its homologous substrate. Therefore, structurally unique residues on the enzyme surface are candidates for specifically interacting with the LTR DNA. The structural alignment is shown in Figure 3.1. This provided a foundation using an *in silico* approach to predict a number of residues for analysis *in vitro*. Mutation of the HIV-1 IN amino acid sequence to the corresponding ASV residue in structurally conserved positions provided a template for analyzing the accuracy of the assembled model. The analysis resulted in the identification of nine amino acid residues that are involved in the specific recognition of the LTR ends (Chen, Weber et al. 2006). The residues identified in this analysis, shown in Figure 3.2, are predicted to interact within 12 nucleotides of the end of the LTR sequence. This is in agreement with the report that interactions between HIV-1 IN and the LTR bases occurs within 6 bases of the LTR ends (Agapkina, Smolov et al. 2006). Interactions distal to the CA dinucleotide are reported to be with the phosphate backbone, suggesting a conformational or structural restriction between the enzyme surface and substrate DNA for specific interaction (Agapkina, Smolov et al. 2006).

Figure 3.1 Structural alignment of the primary amino acid sequence of HIV-1, SIV, and ASV INs

Residues shaded in yellow, gray, and green in HIV-1 IN are proposed from the structural model to be in close proximity to base pairs 1-4, 5-15, and 16-20, respectively, of the LTR DNA ends. Blue shaded residues in ASV IN mark those unique to all three INs. A dash indicates no amino acid in the corresponding structures. Bold letters above the HIV-1 sequence indicate residues that change in drug-resistant IN mutants (Hazuda, Felock et al. 2000; Fikkert, Van Maele et al. 2003; Hazuda, Anthony et al. 2004; Hazuda, Young et al. 2004; Lee and Robinson 2004). The alignment was compiled by another group and previously published (Snasel, Krejcik et al. 2005).


1 10 20 30 40
hivin FLDGIDKAQEEHEKYHSNWRAMASDFNLPPVVAKEIVASCDKCQLKG
sivin FLEKIEPAQEEHDKYHSNVKELVFKFGLPRIVARQIVDTCDKCHQKG
asvin ---PLREAKDLHTALHIGPRALSKACNISMQQAREVVQTCPHCNSAP
50
hivin EAMHGQV--DCSP-
sivin EAIHGQA--NSDL- N-TERMINUS DOMAIN
asvin ALEAGVNPRGLGPL
60 I 70 I M 80 90 100
hivin GIWQLDCTHLE--G--KVILVAVHVASGYIEAEVIPAETGQETAYFLLKLAGRW
sivin GTWQMDCTHLE--G--KIIIVAVHVASGFIEAEVIPQETGRQTALFLLKLAGRW
asvin QIWQTDFTLEPRMAPRSWVAVTVDTASSAIVVTQHGRVTSVAAQHHWATAIAVL
110 Y K 130 140 I YIS 160
hivin -PVKTVHTDNGSNFTSTTVKAACWWAGIKQEFGIPYNPQSQGVIESMNKELKKI
sivin -PITHLHTDNGANFASQEVKMVAWWAGIEHTFGVPYNPQSQGVVEAMNHHLKNQ
asvin GRPKAIKTDNGSCFTSKSTREWLARWGIAHTTGIPGNSQGQAMVERANRLLKDR
170 180 190
hivin IGQVR--DQAE------HLKTAVQMAVFIHNFKRKGGIGGY
sivin IDRIR--EQAN------SVETIVLMAVHCMNFKRRGGIGDM CATALYTIC
asvin IRVLAEGDGFMKRIPTSKQGELLAKAMYALNHFERGENTK CORE DOMAIN
200 210 220
hivin SAGERIVDIIATDIQTKELQKQITKI
sivin TPAERLINMITTEQEIQFQQSKNSKF
asvin TPIQKHWRPTVLT----------
R 240 250 260 270
hivin QNFRVYYRDSRDPVWKGPAKLLWKGEGAVVIQD-N-SDIKVVPRRKAKIIRD
sivin KNFRVYYREGRDQLWKGPGELLWKGEGAVILKV-G-TDIKVVPRRKAKIIKD
asvin EGPPVKIRIET-GEWEKGWNVLVWGRGYAAVKNRDTDKVIWVPSRKVKPDIT
hivin YGKQMAGDD-----CVASR-----QDED C-TERMINUS
sivin YG---GGKE-----VDSSS-----HMEDTGEAREVA DOMAIN
asvin QKDEVTKKDEASPLFAGSSDWIPWGDEQEGLQEEAASNKQEGPGEDTLAANES

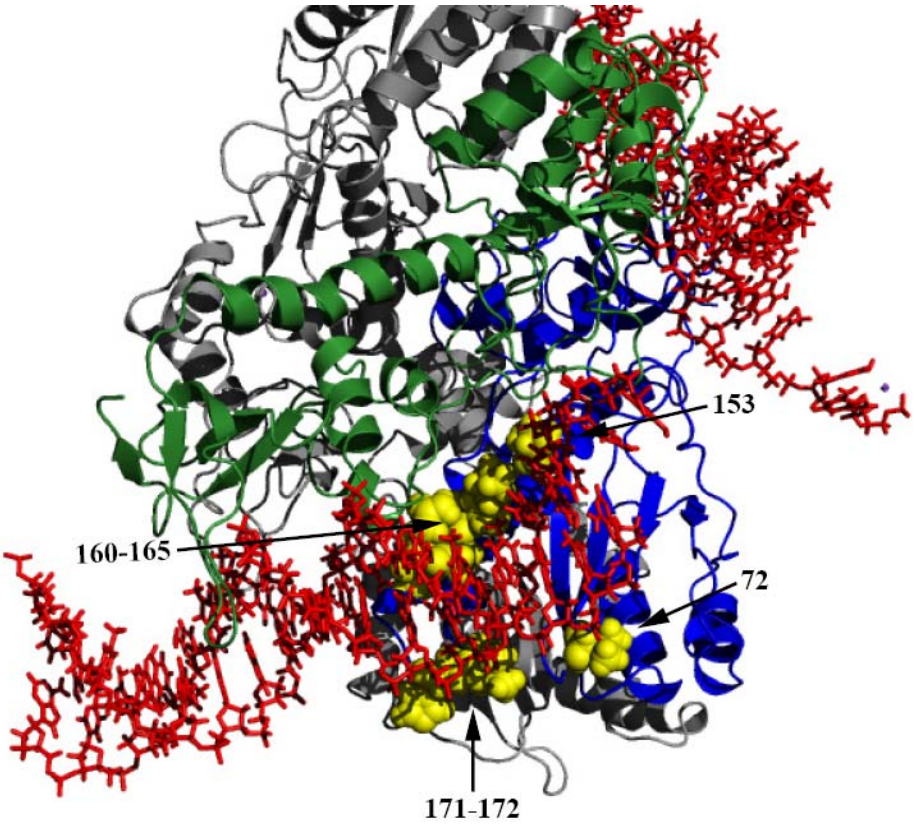


Figure 3.2 Residues that affect LTR specificity represented in HIV-1 IN model

Mutation at the indicated residues to their corresponding amino acid from ASV resulted in at least partial specificity change, indicating they play a role in the specificity reaction between IN and the LTR DNA. Interestingly, residues 72 and 153 are reported in the literature to have mutations during drug-selection using strand transfer inhibitors, suggesting the mechanism of action for these small-molecules may not only function at the level of strand transfer.

3.2 Specific DNA Activity of Purified HIV-1/ASV IN Chimeras

HIV-1 IN mutants were constructed and purified as previously described in Materials and Methods. The chimeras were assembled in a 3CSF185 background, an enzyme with four amino acid substitutions (C56S, C65S, F185H, C280S) to improve solubility. This enabled purification of the chimeras from the soluble fraction. Individually these amino acid substitutions have little or no effect on viral replication (Engelman, Liu et al. 1997; Bischerour, Tauc et al. 2003; Zhu, Dobard et al. 2004). Enzymes purified by this protocol were free of detectable non-specific nuclease (Chen, Weber et al. 2006). As further evidence for the purity of INs prepared by this protocol, 5' ^{33}P-end labeled DNA substrates representing the HIV-1 U5, ASV U3, and MLV U5 LTR termini, were individually incubated with wild type or selected IN chimeras. The 3CSF185H HIV-1 and ASV INs cleave their respective homologous substrates but not heterologous substrates including the MLV substrate (Figure 3.3). As previously reported (Chen, Weber et al. 2006), the G163R Q164V V165L chimera in the 3CSF185H background cleaved both HIV-1 U5 and ASV U3 LTR end substrates (Figure 3.3). However, the chimera did not cleave the MLV U5 substrate indicating the ability to cleave the ASV substrate was not due to a non-specific nuclease activity. A second HIV-1 IN chimera that contains a K211S substitution also in the 3CSF185H background cleaved the HIV-1 U5 but not the ASV U3 or MLV U5 substrates. This amino acid is positioned in the structural model near the putative target DNA binding trench but at a distance from the LTR binding sites so that it was expected to maintain HIV-1 substrate specificity for 3' processing.

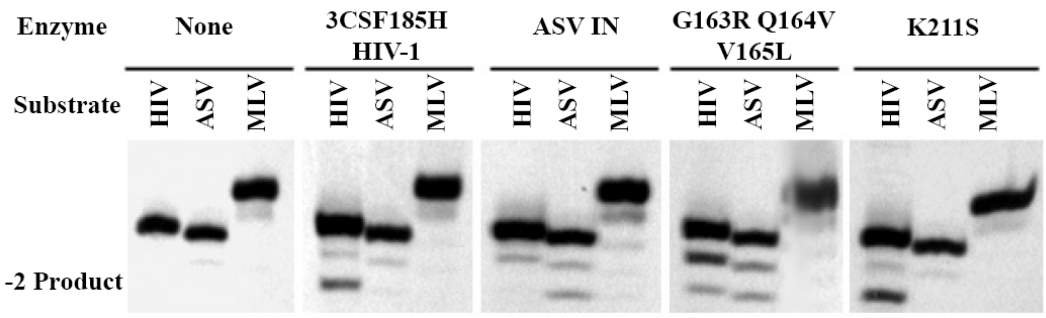


Figure 3.3 Purified HIV-1 chimeras do not cleave MLV substrates.

^{33}P-end labeled duplex oligos representing the HIV-1 U5, ASV U3, and MLV U5 LTR ends were incubated with HIV-1 (3CSF185H), ASV, or mutant HIV-1 INs (the triple substitution mutant G163R Q164V V165L and K211S) and the products separated by denaturing gel electrophoresis as described in Materials and Methods. The production of the -2 cleavage product demonstrates the specific enzymatic processing of the viral DNA substrate.

3.3 Prediction of additional HIV-1 IN residues interacting with the LTR DNA ends

In the original selection of amino acids that could affect LTR end recognition, we used the structural alignment of SIV, HIV-1, and ASV INs to identify those residues that were unique. We subsequently observed that HIV-1 IN was capable of 3' processing a U5 SIV (Aiping Chen, personal communication) but not a MPMV LTR DNA substrate (Figure 3.4). In light of this data, we reexamined our definition of unique residues using the structural alignment of INs by replacing the SIV with the MPMV IN sequence (Snasel, Krejcik et al. 2005). This analysis identified additional unique residues in the same structural position in HIV-1, ASV, and MPMV INs near the LTR DNA ends as (using HIV-1 numbering): Ser^{39}, Lys^{42}, Gln^{44}, Leu^{68}, Glu^{69}, Leu^{74}, Lys^{156}, Glu^{170}, Tyr^{227}, Asp^{229}, Ser^{230}, Asp^{253}, Asn^{254}, Lys^{258}, and Arg^{262} (Figure 3.5). These residues are colored red and labeled in the structural model shown in Figure 3.6. The position of the newly identified residues is striking in that they sit in the viral DNA groove on both sides of the DNA substrate. We hypothesized that at least some of these residues, when mutated to the structural equivalent in ASV, will gain the ability to specifically process the ASV LTR DNA substrate. To test whether any of these residues were involved in recognition of the LTR ends, eleven HIV-1/ASV IN chimeras were constructed that substituted the amino acid from ASV into the structurally related position of HIV-1 IN as described in Materials and Methods.

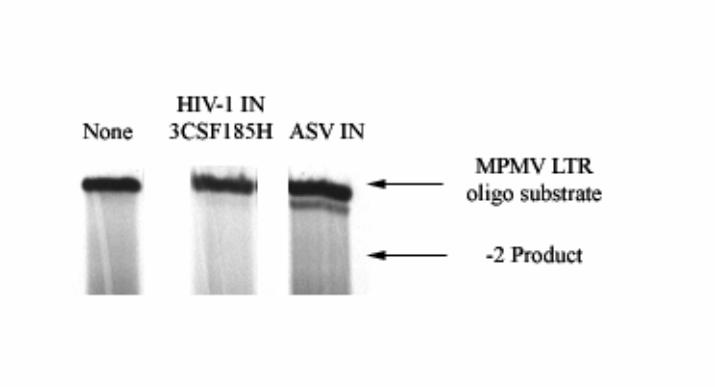
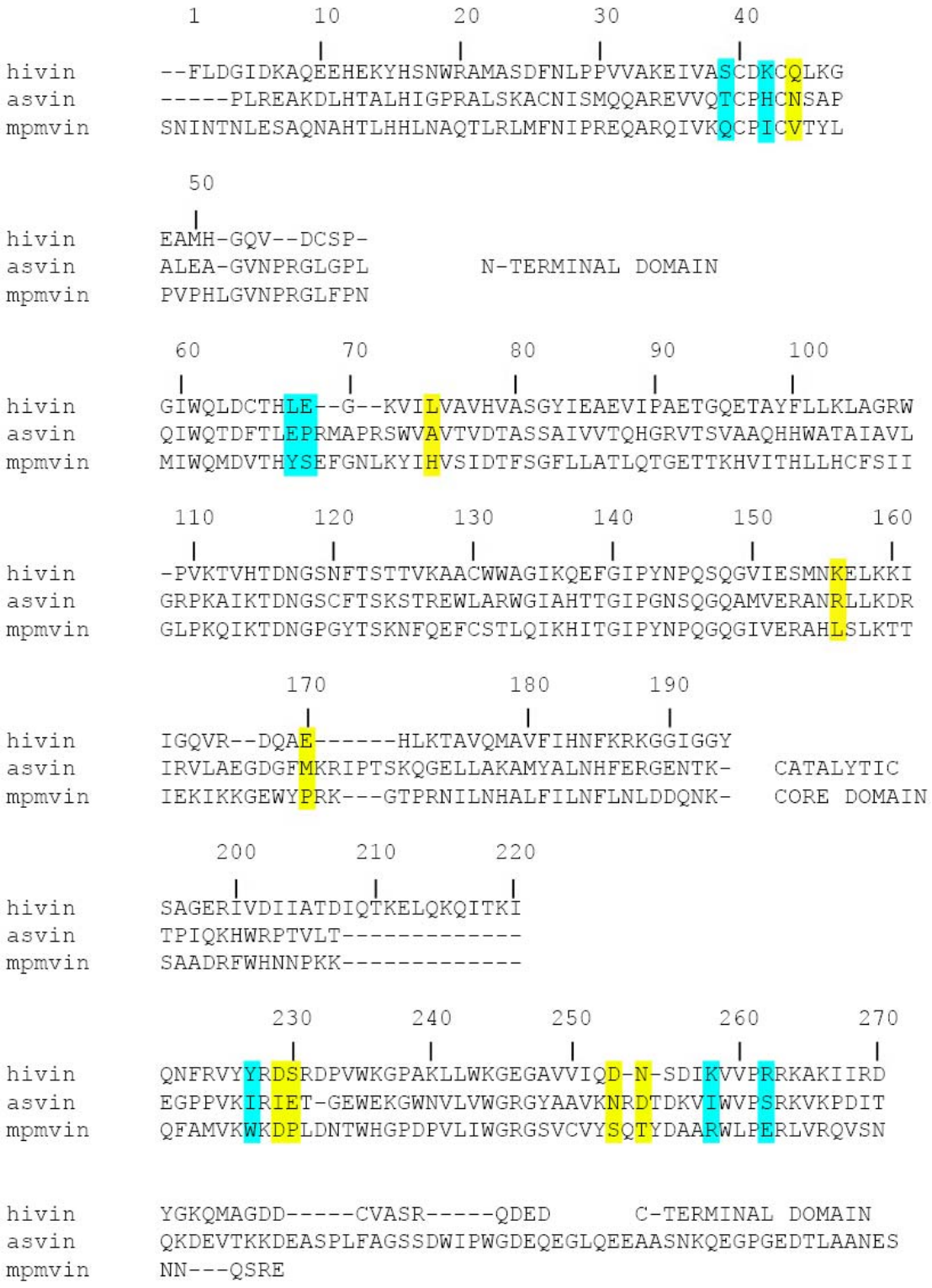


Figure 3.4 3' End Processing of wild-type enzymes on MPMV LTR DNA substrate

Incubation of wild-type HIV-1 (in the soluble 3CSF185H backbone) or ASV IN with a ^{33}P-labeled duplex oligo substrate derived from the U5 LTR DNA end of MPMV show that there is no specific cleavage of the DNA substrate as evidenced by the lack of a -2 product.

Figure 3.5 Structural Alignment of HIV-1, ASV, and MPMV IN

This is a structural alignment of HIV-1, ASV and MPMV that was previously published (Snasel, Krejcik et al. 2005). We identified fifteen additional amino acid residues that were different at structurally conserved positions between the three enzymes. The residues from this analysis are in addition to those that were previously studied (Chen, Weber et al. 2006) and are predicted to interact with the LTR DNA. Residues highlighted in yellow were analyzed by Aiping Chen while those highlighted in blue were analyzed by James Dolan. HIV-1 IN chimeras were assembled by substitution at predicted positions to the corresponding ASV residue and assayed for 3' end processing against the heterologous ASV LTR DNA substrate, as described in the Materials and Methods.


1 10 20 30 40
hivin --FLDGIDKAQEEHEKYHSNWRAMASDFNLPPVVAKEIVASCDKCQLKG
asvin -----PLREAKDLHTALHIGPRALSKACNISMQQAREVVQTCPHCNSAP
mpmvin SNINTNLESAQNAHTLHHLNAQTLRLMFNIPREQARQIVKQCPICVTYL
50
hivin EAMH-GQV--DCSP-
asvin ALEA-GVNPRGLGPL N-TERMINAL DOMAIN
mpmvin PVPHLGVNPRGLFPN
60 70 80 90 100
hivin GIWQLDCTHLE--G--KVILVAVHVASGYIEAEVIPAETGQETAYFLLKLAGRW
asvin QIWQTDFTLEPRMAPRSWVAVTVDTASSAIVVTQHGRVTSVAAQHHWATAIAVL
mpmvin MIWQMDVTHYSEFGNLKYIHVSIDTFSGFLLATLQTGETTKHVITHLLHCFSII
110 120 130 140 150 160
hivin -PVKTVHTDNGSNFTSTTVKAACWWAGIKQEFGIPYNPQSQGVIESMNKELKKI
asvin GRPKAIKTDNGSCFTSKSTREWLARWGIAHTTGIPGNSQGQAMVERANRLLKDR
mpmvin GLPKQIKTDNGPGYTSKNFQEFCSTLQIKHITGIPYNPQGQGIVERAHLSLKTT
170 180 190
hivin IGQVR--DQAE------HLKTAVQMAVFIHNFKRKGGIGGY
asvin IRVLAEGDGFMKRIPTSKQGELLAKAMYALNHFERGENTK- CATALYTIC
mpmvin IEKIKKGEWYPRK---GTPRNILNHALFILNFLNLDDQNK- CORE DOMAIN
200 210 220
hivin SAGERIVDIIATDIQTKELQKQITKI
asvin TPIQKHWRPTVLT----------
mpmvin SAADRFWHNNPKK----------
230 240 250 260 270
hivin QNFRVYYRDSRDPVWKGPAKLLWKGEGAVVIQD-N-SDIKVVPRRKAKIIRD
asvin EGPPVKIRIET-GEWEKGWNVLVWGRGYAAVKNRDTDKVIWVPSRKVKPDIT
mpmvin QFAMVKWKDPLDNTWHGPDPVLIWGRGSVCVYSQTYDAARWLPERLVRQVSN
hivin YGKQMAGDD-----CVASR-----QDED C-TERMINAL DOMAIN
asvin QKDEVTKKDEASPLFAGSSDWIPWGDEQEGLQEEAASNKQEGPGEDTLAANES
mpmvin NN---QSRE

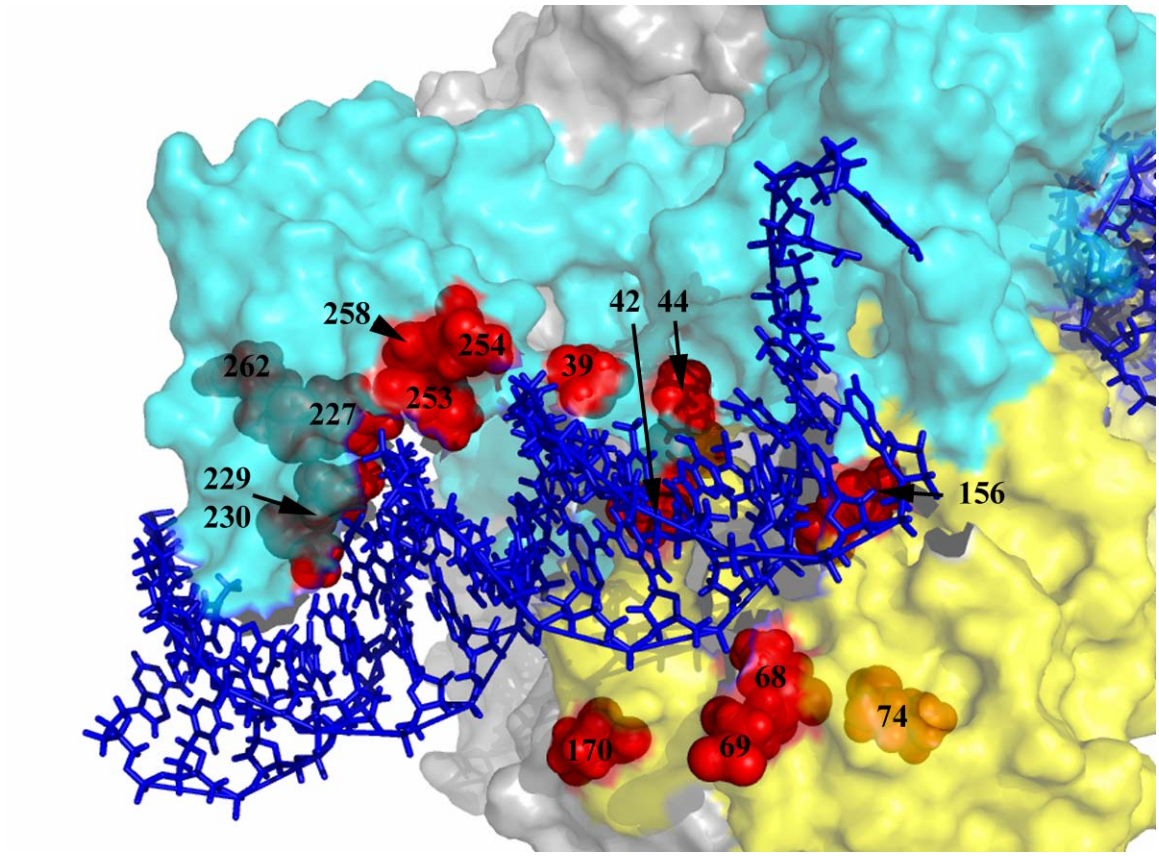


Figure 3.6 Model of new residues from alignment for investigation

Amino acid residues that were identified in the new alignment (Figure 3.5) are displayed in our model. The HIV-1 IN tetramer is represented in cartoon form with viral LTR DNA in proposed trenches. The residues shown in space fill representation and colored red are those that were identified as unique between HIV-1, ASV, and MPMV and are proximal to the viral LTR DNA.

3.4 Effect of Specific Amino Acid Changes on 3' Processing

We screened each HIV-1/ASV chimera for the ability to 3' process HIV-1 and ASV LTR duplex DNA substrates. This analysis was split between Aiping Chen, a former member of the lab, and James Dolan. The initial enzyme set assembled by Aiping included residues Gln^{44}, Leu^{74}, Lys^{156}, Glu^{170}, Asp^{229}, Ser^{230}, Asp^{253}, and Asn^{254}. These residues were mutated to the corresponding structural amino acid in ASV and purified as described in the Materials and Methods. The N254D chimera showed a decrease in activity against the HIV-1 substrate and did not cleave the ASV substrate (Figure 3.7). In contrast, the Q44N, D229I S230E and D253N chimeras all maintained the ability to cleave the HIV-1 substrate, but they gained the ability to cleave the ASV substrates to different extents (Figure 3.7). The D229I S230E mutation was unique among these chimeras in that it possessed more 3' processing activity towards the HIV-1 substrate than wild type 3CSF185H (Figure 3.7). A summary of these mutations is shown in Table 3.12.

We constructed the remaining chimeras to complete the data set. This included mutations at Ser^{39}, Lys^{42}, Leu^{68}, Glu^{69}, Tyr^{227}, Lys^{258}, and Arg^{262}. As shown in Figure 3.8, the L68E E69P chimera gained the ability to cleave the ASV U3 LTR DNA substrate. In contrast, there is no specificity change for the S39T K42H enzyme (it only cleaves the HIV U5 LTR DNA). The K258I R262S enzyme has no 3' end processing activity for either substrate. These enzymes, in combination with the other chimeras, add six residues to the known amino acids that influence the recognition of the LTR DNA.

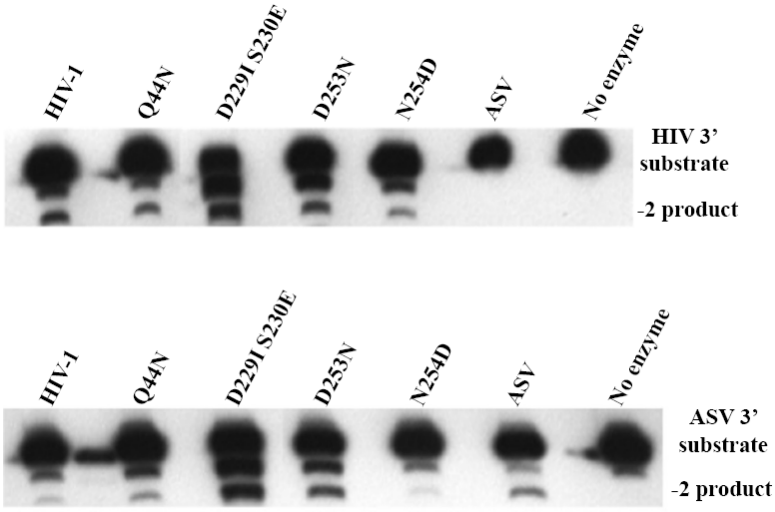


Figure 3.7 3' End processing of new residues performed by Aiping Chen

Wild-type HIV-1 and ASV INs in addition to Q44N, D229I S230E, D253N, N254D HIV-1 IN chimeras were incubated with HIV-1 U5 or ASV U3 LTR substrates and analyzed for 3' end processing activity as described in the Materials and Methods. The presence of a -2 cleavage product is indicative of specific cleavage of the DNA substrate. Chimeras Q44N, D229I S230E, and D253N gained the ability to cleave the ASV DNA substrate. Additionally, the D229I S230E mutant displayed increased enzymatic activity relative to wild-type.

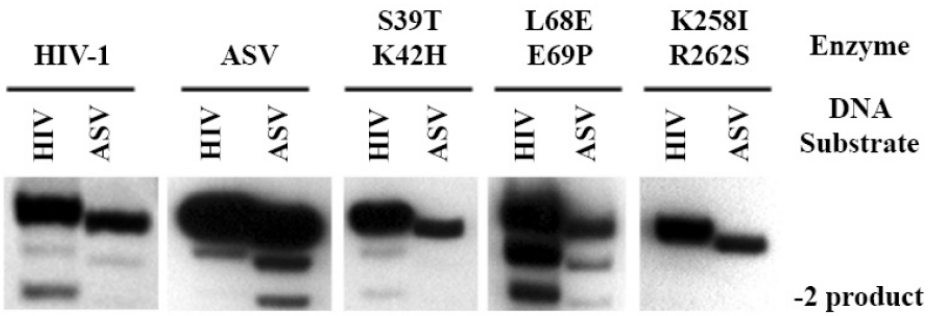


Figure 3.8 3' End processing phenotypes of other residues identified by realignment

Wild-type HIV-1 and ASV INs in addition to S29T K42H, L68E E69P, and K285I R262S HIV-1 IN chimeras were incubated with HIV-1 U5 or ASV U3 LTR substrates and analyzed for 3' end processing activity as described in the Materials and Methods. The presence of a -2 cleavage product is indicative of specific cleavage of the DNA substrate. The L68E E69P chimera shows activity against the ASV U3 LTR DNA, whereas the other chimeras show no change (S39T K42H) or no activity (K258I R262S).

Mutation at positions Leu^{74}, Lys^{156}, Glu^{170} and Tyr^{227} were prepared and the enzymes had varying phenotypes. The Leu^{74} and Tyr^{227} enzymes were inactive when tested *in vitro* (much like the Lys^{258} Arg^{262} enzyme). The purified Lys^{156} enzyme had a non-specific cleavage phenotype meaning it did not discriminate the HIV-1 LTR DNA for specific cleavage of two nucleotides. Additionally, there was no change in the specificity reaction for the Glu^{170} enzyme. The results for the Leu^{74} and Tyr^{227} do not preclude an involvement of these residues in the specificity reaction, but Lys^{156} and Glu^{170} are likely not involved.

Residues Gln^{44}, Leu^{68}, Glu^{69}, Asp^{229}, Ser^{230}, and Asp^{253} have been highlighted on the structural model in Figure 3.9 (magenta residues) along with the previously identified residues that affect 3' processing specificity (red residues) (Chen, Weber et al. 2006). Taken together, these residues strikingly define two linear trenches on the HIV-1 IN molecular surface that accommodate the two LTR ends. The binding trenches are asymmetrically positioned along the two strands of a 16 base pair length of the LTR ends.

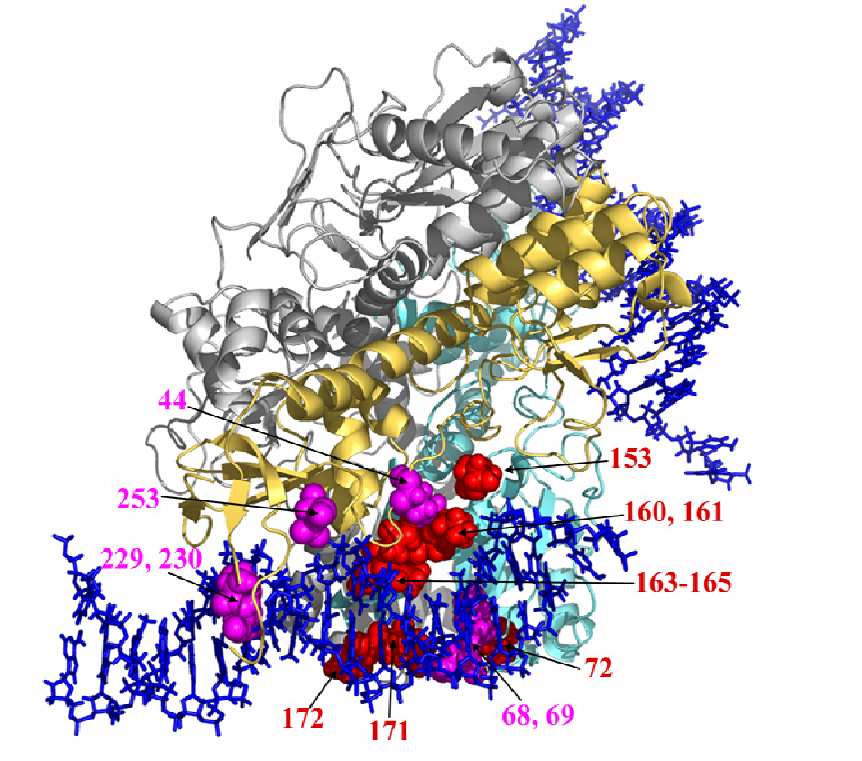


Figure 3.9 Tetramer model of all residues that alter specificity

IN tetramer model with subunits that interact with the two viral DNA ends colored cyan and yellow with the two other subunits colored in gray. The LTR DNA ends are represented by blue helices. The amino acids that specifically recognize the viral DNA are shown in space fill model for only one viral DNA end. Those in red are from Chen et al (Chen, Weber et al. 2006); those newly identified are in magenta.

3.5 Conversion of specificity by cumulative mutation

We combined twelve of the fifteen amino acid exchanges that affect LTR recognition in the processing reaction into a single construct and purified from the soluble fraction. This combined the V72W S153R K160D I161R G163R Q164V V165L H171K L172Q D229I S230E D253N substitutions and is referred to as S7C. As shown in Figure 3.10, this enzyme is active and has substantially more specific 3' processing activity towards the ASV than the HIV-1 DNA substrate. This result suggests that the sum of IN interactions along a 16 base pair length of the viral DNA ends determines its specificity. Additionally, there is still some -1 product in the presence of the HIV-1 U5 LTR DNA substrate, suggesting that there is some non-specific processing of the DNA. This cleavage product is observed in the wild-type enzyme, and other cleavage products are not detected. There may be additional amino acids that are involved in the recognition of the DNA substrate. There are currently at least two known residues that have not been changed in the combination mutant, and these are Leu^{68} and Glu^{69}. These residues are predicted to interact within five base pairs of the LTR end and interact with the nucleotide. We predict that if these residues are added and a soluble enzyme is purified by our protocol, the non-specific product in the HIV-1 U5 LTR DNA lane may be reduced.

A combination enzyme assembled by Aiping Chen contained the exact same amino acid changes that affect the LTR specificity as our S7C mutant, but this enzyme was in a backbone where two solubility enhancing mutations, F185H and C280S, were removed. Her enzyme was reconstituted from an insoluble fraction and is referred to as 2CS-S7C. This enzyme had qualitatively similar activity to the soluble S7C enzyme when incubated with HIV-1 U5 and ASV U3 LTR DNA substrates (Figure 3.11).

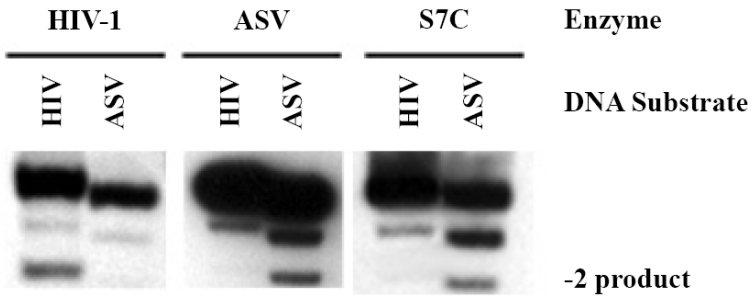


Figure 3.10 Specificity change of S7C combination mutant

S7C is in the soluble 3CSF185H HIV-1 IN background and has 12 amino acid substitutions. This cumulative mutant, with amino acid changes at positions Val^{72}, Ser^{153}, Lys^{160}, Iso^{161}, Gly^{163}, Gln^{164}, Val^{165}, His^{171}, Leu^{172}, Asp^{229}, Ser^{230}, and Asp^{253}. This enzyme loses the ability to specifically process the HIV-1 U5 LTR DNA duplex substrate and processes a significant proportion of the ASV U3 LTR DNA duplex substrate.

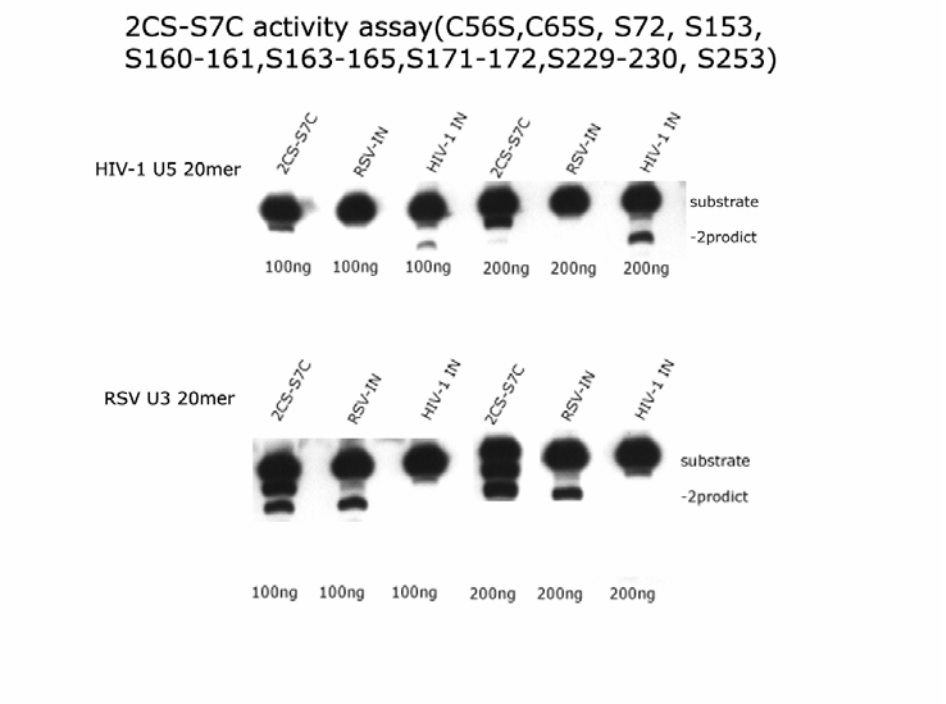


Figure 3.11 Insoluble combination mutant activity

The 2CS-S7C chimera also has 12 amino acid substitutions. This enzyme was purified from the insoluble fraction and renatured from urea before testing *in vitro* 3' end processing activity against the homologous and heterologous DNA substrates. This mutant, created by Aiping Chen, has a similar activity to the 3CS S7C mutant and shows a decreased ability to specifically process the HIV-1 LTR DNA with an enhanced activity against the ASV LTR substrate.

Mutant	LTR end base pair interaction	Substitutions	3' Processing Outcome
S39/42	6-8	S39T K42H	Decrease for HIV-1 substrate; no ASV
S44	3-6	Q44N	Partial Change in Specificity
S68/69	5-8 (1-4?)	L68E E69P	Partial Change in Specificity; Increased activity
S74	1-4	L74A	Inactive enzyme
S156	4-5	K156R	Non-specific cleavage of HIV-1; no ASV
S170	7-8	E170M	No change detected; no ASV
S227	14-15	Y227I	Inactive enzyme
S229/230	15-17	D229I S230E	Partially altered specificity
S253	8-12	D253N	Partial change in specificity
S254	8-12	N254D	Decrease for HIV-1 substrate; no ASV
S258/262	19-20, 13-14	K258T R262S	Inactive enzyme

Figure 3.12 Summary of HIV-1/ASV chimeras

A summary of the enzyme chimeras is shown above with the corresponding amino acid substitutions. The predicted LTR end base pair interactions are also included. The activities of all chimeras are summarized, with specificity changes in the S44, S68/69, S229/230, and S253 chimeras.

3.6 LTR trench residue strand transfer analysis

In addition to the 3' end processing activity, viral IN catalyzes the insertion of duplex DNA substrates in a reaction called strand transfer. This reaction uses a single DNA substrate with a 5' overhang to simulate a processed viral LTR DNA end. Integrase binds to the DNA duplex in both the viral and target DNA trenches, initiates integration and produces a strand transfer product of increased size compared to the original substrate. After identifying a number of residues that influence the specificity of the 3' end processing reaction, we decided to assay these enzyme chimeras in the strand transfer assay. We predicted that because these chimeras had the ability to cleave the ASV LTR substrate in the 3' end processing reaction, they would also have the ability to catalyze strand transfer using a pre-processed heterologous DNA.

Strand transfer assays were set up with HIV-1 IN chimeras using a pre-processed ASV U3 LTR DNA duplex. Reactions were incubated at 37°C for one hour before loading on a DNA sequencing gel to separate any integration products. After repeated attempts, none of the HIV-1 chimeras catalyzed strand transfer using the ASV DNA.

The 3' end processing activities of all the chimeras was previously established, and all enzymes maintained an ability to specifically cleave the HIV-1 U5 LTR DNA in the 3' end processing assay. In order to validate that the HIV-1 IN chimeras could catalyze strand transfer, we then decided to assay these enzymes using a HIV-1 substrate as described in the Materials and Methods. HIV-1 IN chimeras were incubated with pre-processed HIV U5 LTR DNA for one hour before separation on DNA sequencing gel for visualization. As shown in Figure 3.13, a number of the chimeras had the equivalent to wild-type activity in the strand transfer assay using a homologous DNA substrate (S153R, K160D I161R, G163R Q164V V165L, and D229I

S230E). This is interesting because the D229I S230E chimera had more 3' end processing activity but not more strand transfer activity.

There is a contrast in the strand transfer phenotypes of other chimeras, however. The T125S and D253N chimeras have more strand transfer activity relative to wild-type (Figure 3.13). These data are consistent with the 3' end processing phenotypes of these enzymes as they both display an increased enzymatic activity in that assay as well. Unexpectedly, the L68E E69P, V72W, and H171K L172Q chimeras that had 3' end processing activity against the HIV U5 LTR DNA do not have strand transfer activity (Figure 3.13). Altogether, these mutants demonstrate a disconnection between the 3' end processing and strand transfer activities as they show specific 3' end processing of the heterologous DNA substrate, but none have the ability to catalyze strand transfer against the heterologous DNA. These data demonstrate the lack of understanding in the field currently with regard to the strand transfer or integration reaction.

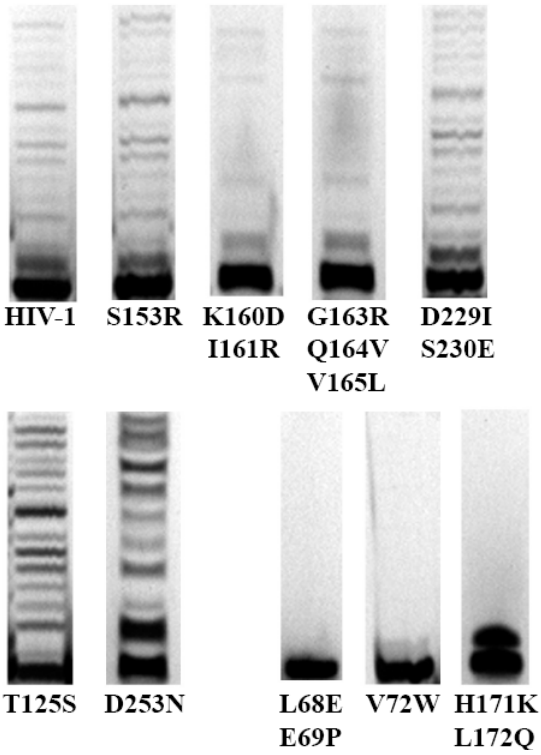


Figure 3.13 Strand transfer activity of HIV-1 IN chimeras that have 3' end processing phenotypes

HIV-1 IN and chimeras were incubated with a pre-processed HIV-1 U5 LTR duplex substrate with two bases removed adjacent to the CA dinucleotide. The chimeras are ordered by those with wild-type activity (top) or altered (higher on bottom left, reduced/none on the bottom right).

3.7 Conclusion

We present a data set that has identified a series of amino acids that were predicted to interact with the viral LTR DNA. These amino acids were changed to their equivalent ASV residue based on a structural alignment and the purified enzymes were tested for their *in vitro* 3' end processing activity. We have identified six additional residues involved in the specificity reaction for a total of fifteen residues on the surface of HIV-1 IN that affect the recognition of the viral LTR DNA.

We also combined a number of these mutations into a single construct and assayed this enzyme's activity against the heterologous substrate. Substitution at twelve residues showed a significant loss in 3' end processing of the HIV LTR DNA substrate while gaining the ability to produce the -2 cleavage product using the heterologous ASV LTR DNA. This is significant because it is the first time that the specificity of a retroviral integrase has been altered in the 3' end processing reaction as well as identified fifteen residues that contribute to the recognition of the viral LTR DNA end. A combination of sequence and structure is necessary for this recognition, and our residues align along at least a 16 base pair length of the viral LTR.

Lastly, while we predicted that the mutations that alter specificity in the 3' end processing reaction, we show that none of the enzyme chimeras can catalyze strand transfer of the heterologous DNA substrate. The understanding of the strand transfer reaction, and the determinants such as the interaction between enzyme and substrates, remain largely undefined. Regardless, several of the residues completely lost the ability to strand transfer either homologous or heterologous DNA substrates, indicating their role in both the 3' end processing and strand transfer reactions.

CHAPTER 4

RECOGNITION OF TARGET DNA SUBSTRATES

4.1 Overview

A functional integration complex requires the binding of host DNA at the site of integration. This substrate interaction is essentially random, with sequence specificity playing little to no role in site selection. As a result, integration events occur throughout the genome. One current strategy for therapeutic development involves disrupting the integration process, presumably through inhibition of strand transfer. Blocking the interaction between the enzyme and host DNA prevents the concerted integration of the viral genome into the host.

The HIV-1 IN interaction with target DNA is largely uncharacterized, and our objective was to identify key residues involved in binding the target using the computer model for the HIV-1 IN tetramer. By placing the LTR termini on the HIV-1 IN surface, a groove is observed on the model that is located between the two LTR ends and approximately perpendicular to the long axis of the LTR DNA that could accommodate the target DNA (Figure 4.1). We hypothesize that residues on the surface in the opposing grooves are involved in the interaction with the target DNA substrate, and that mutating these residues will affect the ability of HIV-1 IN to bind target DNA.

4.2 Structural analysis for conserved residues in the putative target DNA trench

A structural alignment assembled previously (Snasel, Krejcik et al. 2005; Chen, Weber et al. 2006) and used to predict amino acid residues that interact with the viral LTR DNA was again used to identify amino acids that interact with the target DNA. The residues that we focused on were conserved Lys, Gln, Arg, Asn, and His as shown in Figure 4.2. Conserved residues were classified into two categories: conserved in 3+ INs or conserved in at least two INs. A summary of the identified residues is shown in Figure 4.3. To complete the analysis, we placed the

conserved residues in our model. Figure 4.4 is a representation of the surface of the putative trench with residues highlighted as in the alignment. There are many residues that were conserved in at least three enzymes that align along this trench, consistent with the possibility that this trench is involved in binding target DNA. We were particularly interested in those residues that were in the target DNA trench but distal to the active site because mutation at these residues may affect strand transfer which involves the target DNA but not 3' end processing of the LTR DNA. Residues 186, 187, 188, 211, 214, and 219 are of particular interest for us as they are not near the active site and are positioned on opposing faces of the trench.

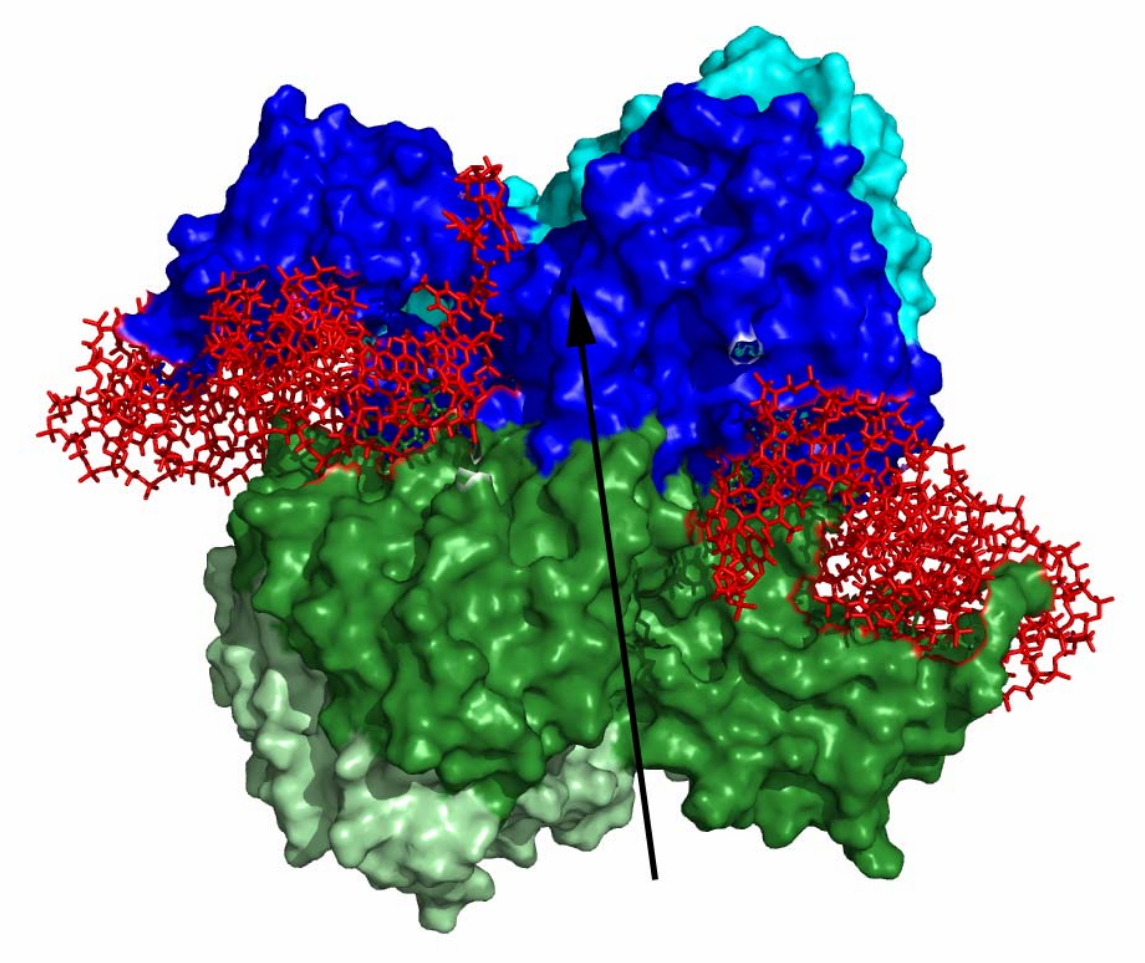
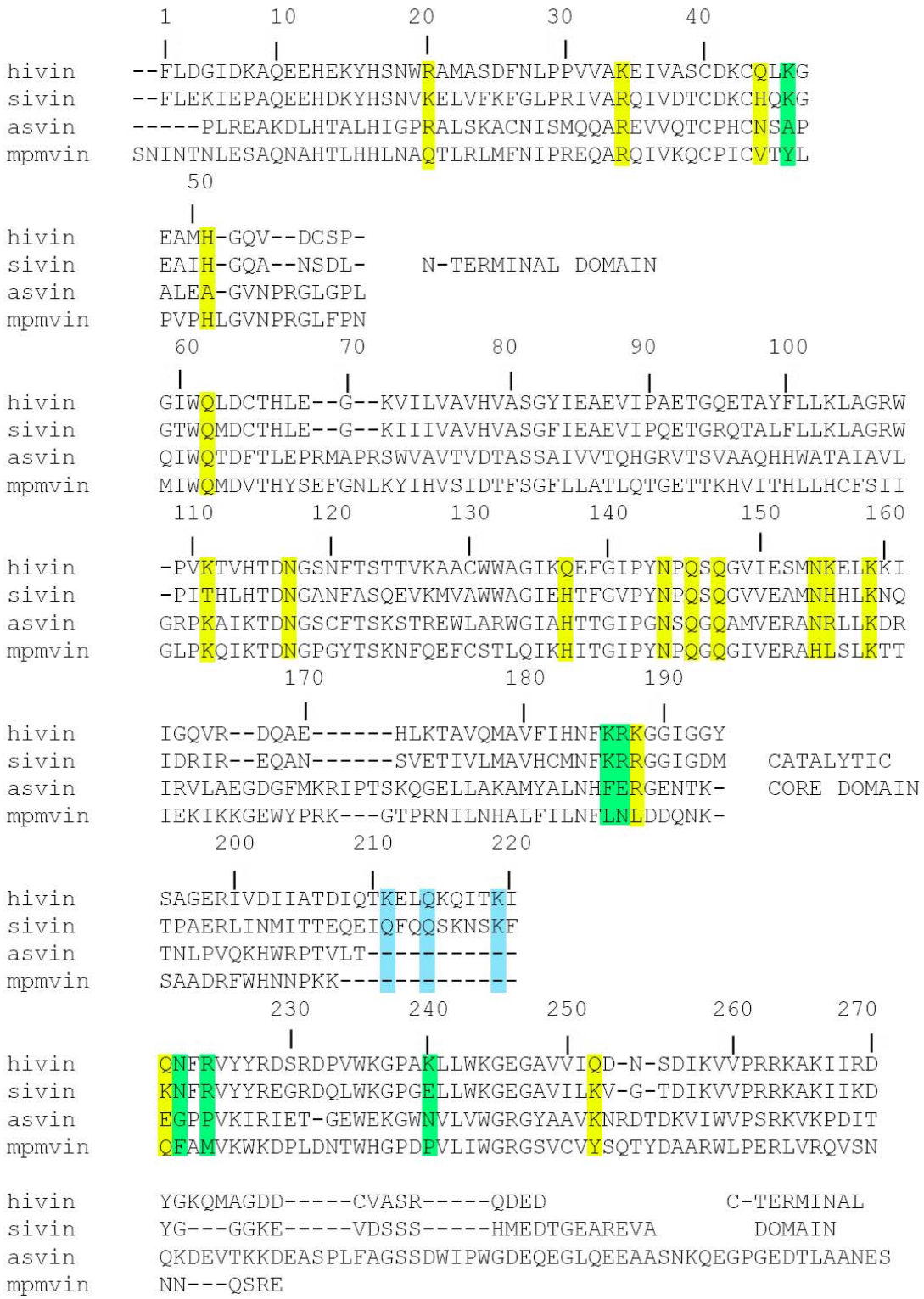


Figure 4.1 Model figure depicting target trench with an arrow

HIV-1 IN tetramer model with bound viral LTR DNA ends. The arrow shows the groove along which the putative residues may be involved in target DNA binding. This representation highlights the juxtaposition of the viral and target DNA binding trenches.

Figure 4.2 New alignment with conserved residue highlights for target residues

Conserved amino acids (Lys, Arg, His, Asn, and Gln) are highlighted on the alignment of HIV-1, SIV, ASV IN and MPMV IN. Residues in yellow are conserved in at least three sequences, and those in green are conserved in at least two INs. Those in blue are conserved in HIV and SIV and do not have structural equivalents in ASV or MPMV. The alignment was compiled by another group and previously published (Snasel, Krejcik et al. 2005).


1 10 20 30 40
hivin --FLDGIDKAQEEHEKYHSNWRAMASDFNLPPVVAKEIVASCDKCQLKG
sivin --FLEKIEPAQEEHDKYHSNVKELVFKFGLPRIVARQIVDTCDKCHQKG
asvin -----PLREAKDLHTALHIGPRALSKACNISMQQAREVVQTCPHCNSAP
mpmvin SNINTNLESAQNAHTLHHLNAQTLRLMFNIPREQARQIVKQCPICVTYL
50
hivin EAMH-GQV--DCSP-
sivin EAIH-GQA--NSDL- N-TERMINAL DOMAIN
asvin ALEA-GVNPRGLGPL
mpmvin PVPHLGVNPRGLFPN
60 70 80 90 100
hivin GIWQLDCTHLE--G--KVILVAVHVASGYIEAEVIPAETGQETAYFLLKLAGRW
sivin GTWQMDCTHLE--G--KIIIVAVHVASGFIEAEVIPQETGRQTALFLLKLAGRW
asvin QIWQTDFTLEPRMAPRSWVAVTVDTASSAIVVTQHGRVTSVAAQHHWATAIAVL
mpmvin MIWQMDVTHYSEFGNLKYIHVSIDTFSGFLLATLQTGETTKHVITHLLHCFSII
110 120 130 140 150 160
hivin -PVKTVHTDNGSNFTSTTVKAACWWAGIKQEFGIPYNPQSQGVIESMNKELKKI
sivin -PITHLHTDNGANFASQEVKMVAWWAGIEHTFGVPYNPQSQGVVEAMNHHLKNQ
asvin GRPKAIKTDNGSCFTSKSTREWLARWGIAHTTGIPGNSQGQAMVERANRLLKDR
mpmvin GLPKQIKTDNGPGYTSKNFQEFCSTLQIKHITGIPYNPQGQGIVERAHLSLKTT
170 180 190
hivin IGQVR--DQAE------HLKTAVQMAVFIHNFKRKGGIGGY
sivin IDRIR--EQAN------SVETIVLMAVHCMNFKRRGGIGDM CATALYTIC
asvin IRVLAEGDGFMKRIPTSKQGELLAKAMYALNHFERGENTK- CORE DOMAIN
mpmvin IEKIKKGEWYPRK---GTPRNILNHALFILNFLNLDDQNK-
200 210 220
hivin SAGERIVDIIATDIQTKELQKQITKI
sivin TPAERLINMITTEQEIQFQQSKNSKF
asvin TNLPVQKHWRPTVLT----------
mpmvin SAADRFWHNNPKK----------
230 240 250 260 270
hivin QNFRVYYRDSRDPVWKGPAKLLWKGEGAVVIQD-N-SDIKVVPRRKAKIIRD
sivin KNFRVYYREGRDQLWKGPGELLWKGEGAVILKV-G-TDIKVVPRRKAKIIKD
asvin EGPPVKIRIET-GEWEKGWNVLVWGRGYAAVKNRDTDKVIWVPSRKVKPDIT
mpmvin QFAMVKWKDPLDNTWHGPDPVLIWGRGSVCVYSQTYDAARWLPERLVRQVSN
hivin YGKQMAGDD-----CVASR-----QDED C-TERMINAL
sivin YG---GGKE-----VDSSS-----HMEDTGEAREVA DOMAIN
asvin QKDEVTKKDEASPLFAGSSDWIPWGDEQEGLQEEAASNKQEGPGEDTLAANES
mpmvin NN---QSRE

Conserved in 3+ INs	Conserved in 2 INs
R20	K46
K34*,#	K186*,#
Q44	R187*,#
H51	K211
Q62	Q214
K111*,#	K219
N117	N222
Q137	R224
N144	K240
Q146	
Q148	
N155	
K156*,#	
K159*,#	
K188*,#	
Q221	
Q252	

Figure 4.3 Table of conserved residues that are exposed on the surface

Residues that are predicted by the tetramer model to be exposed on the enzyme surface and are structurally conserved between HIV-1, SIV and ASV. Indicated residues are either within peptides identified by photo-crosslinking (*, (Heuer and Brown 1997)) or are protected from protease digestion in a footprinting assay (#,(Dirac and Kjems 2001)).

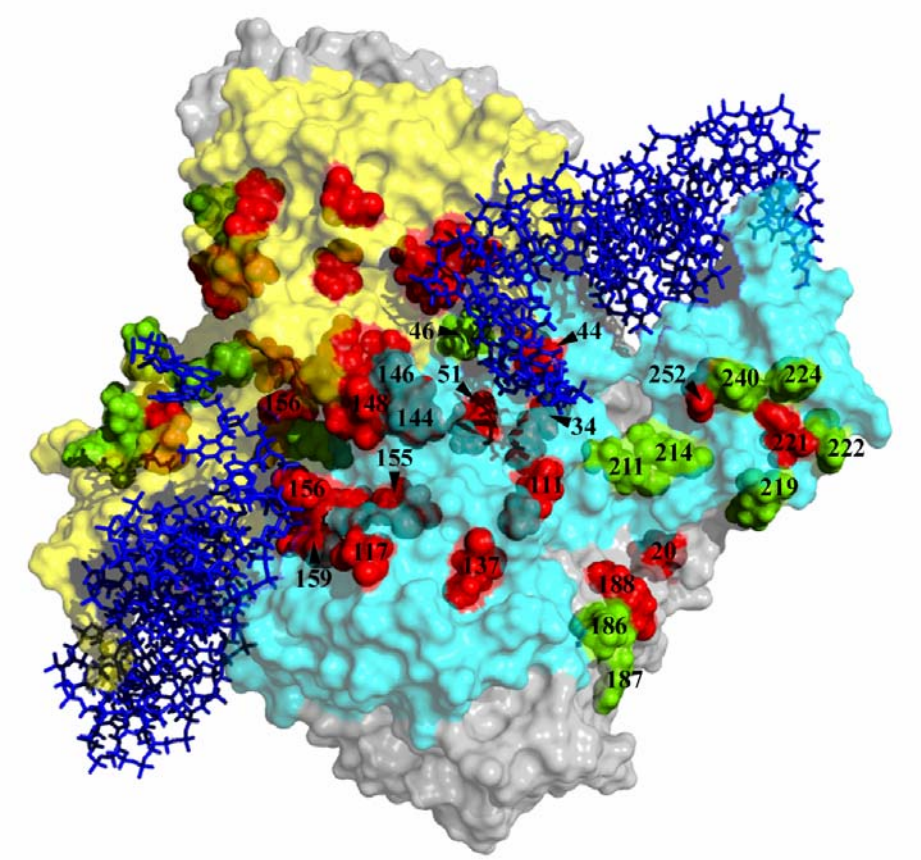


Figure 4.4 Modeling conserved residues for predicting interactions with the target DNA

Residues that are conserved in the structural alignment are represented in the tetramer model of HIV-1 IN. The coloring of labeled residues indicates the conservation and is consistent with the alignment used in Figure 4.3. Briefly, residues conserved in at least three IN sequences are colored in red, while residues conserved in two INs are colored green.

4.3 Published data in support of the tetramer model for target DNA

A report by Heuer and Brown identified a series of peptides that cross-link a hybrid DNA substrate (Heuer and Brown 1997). This DNA is a chimeric construct meant to represent an integration intermediate using components of both the viral and target DNA. They used a photo-cross-linking reagent with an azidophenacyl group to link the DNA substrate to the enzyme and then subjected the complex to Glu-C protease digestion. Peptides were identified after mass-spectrometry and sequencing. In their analysis they isolated six peptides to which cross-links could be formed with viral DNA, target DNA, or both. The six peptides were amino acids 1-11, 49-69, 139-152, 213-246, 247-270, and 271-288. These peptides are represented in the tetramer model of HIV-1 IN with the viral LTR DNA in Figure 4.5. All of the peptides cross-linked to non-viral DNA sequences with the exception of 247-270 (green, Figure 4.5). This peptide only cross-linked to the viral DNA. Peptides 49-69 (magenta) and 139-152 (orange) formed cross-links with both viral and target aspects of the DNA substrate. Interactions with target but not viral DNA sequences were detected for peptides 1-11 (pink, Figure 4.5) and 213-246 (red, Figure 4.5). Peptide 271-288 is not shown because all published crystal structures for this region are highly disordered. The interaction of these peptides to viral and target DNA is in agreement to where they are displayed in the model. For example, the 139-152 peptide sits at the base of the active site between the two LTR ends and is also along the proposed target DNA binding trench. Hence, observing cross-links to both DNA substrates is not surprising. Peptide 213-246 is along one wall of the proposed target DNA binding trench. Again, this prediction is in agreement with the observation that this peptide only cross-linked to target DNA.

The active site, with residues that have been previously reported to cross-link with the LTR DNA ends (Esposito and Craigie 1998; Gerton, Ohgi et al. 1998), correlates well with the 139-152 peptide (orange, Figure 4.5) and includes an essential element of the $DD_{35}E$ motif required for enzymatic activity. Taken together, our model is in agreement with a number of elements from published cross-linking data further supporting the predictive ability to identify residues involved in the interaction with target DNA.

Many of the residues from the structural alignment analysis are in the peptides identified by Heuer and Brown. Figure 4.6 shows the residues from the structural alignment that are predicted to interact with the target DNA and their presence in peptides isolated by Heuer and Brown. Additionally, a number of residues are protected from protease digestion via interaction with DNA, including residues Lys^{34}, Lys^{111}, Lys^{136}, Glu^{138}, Lys^{156}-Lys^{160}, Lys^{185}-Lys^{188}, Asp^{207}, Lys^{215}, Glu^{246}, Lys^{258} and Lys^{273} (Dirac and Kjems 2001). Again, there is good agreement between this data and our tetramer model. Seven residues predicted by the model are protected in the footprinting experiment (Figure 4.3, indicated by #).

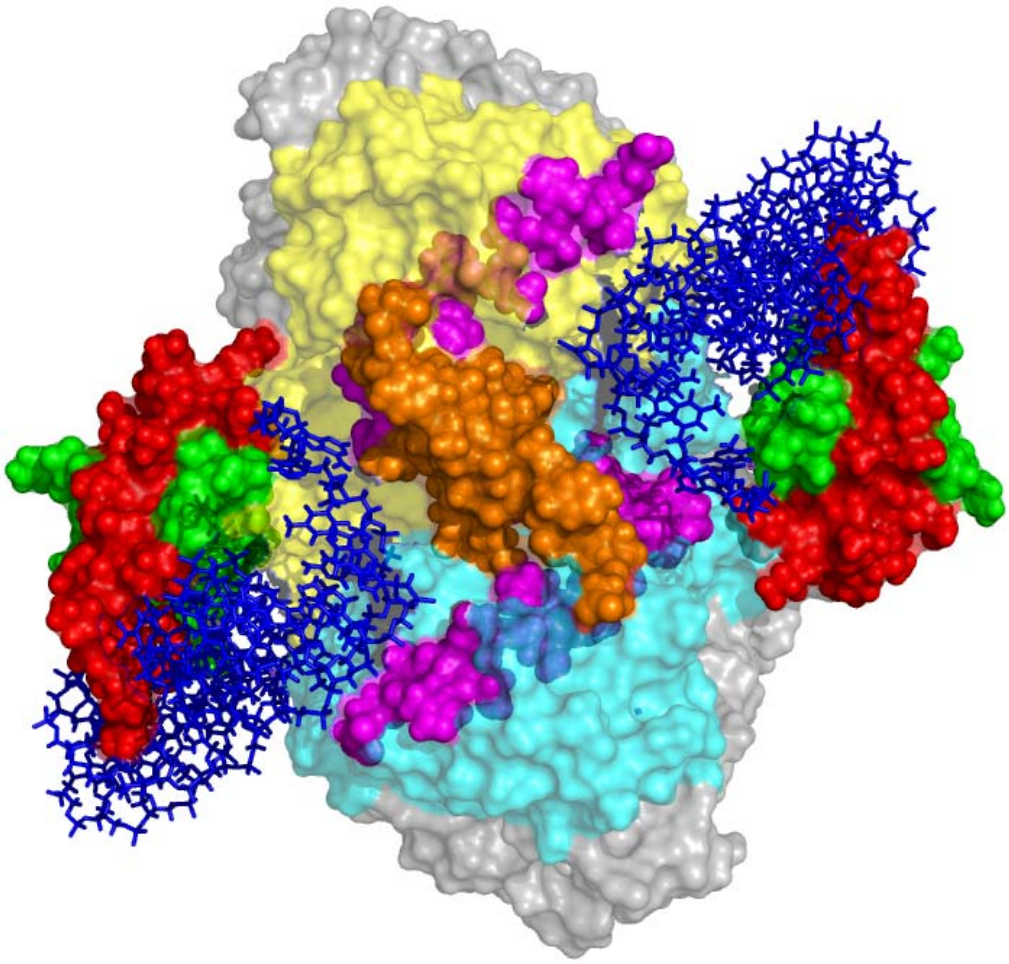


Figure 4.5 Peptides that interact with DNA substrates

Peptide 1-11, which is buried in our tetramer model of HIV-1 IN, was identified to cross-link to elements of the target DNA, as was peptide 213-246 [red]. Peptide 247-270 [green] only cross-linked to elements of the viral LTR DNA. The other peptides (49-69 [purple], 139-152 [orange], and 271-288 [not shown]) cross-linked to both viral LTR and target DNA elements (Heuer and Brown 1997).

Heuer and Brown Peptides	1-11	49-69	139-152	213-246	247-270	271-288
Residues from Alignment that are contained within peptides		H51	N144	Q214	Q252	
		Q62	Q146	K219		
			Q148	Q221		
				N222		
				R224		
				K240		

Figure 4.6 Residues predicted from alignment cross-referenced to Heuer and Brown peptides

The six peptides that were isolated in the Heuer and Brown photo-crosslinking data (Heuer and Brown 1997) are displayed in the top row. Residues that were predicted from our structural alignment to be involved with binding to target DNA are listed below the corresponding peptide. Peptide 4, representing amino acids 213-246, was reported to only interact with target DNA, and our alignment identified a number of residues that are within that peptide.

4.4 Target trench residue analysis

We predict from the structural model that mutations introduced into the target DNA binding site of IN, but distant from the catalytic and LTR binding sites, would have a phenotype where strand transfer would be inactivated without impairment of the 3' processing reaction. Within the putative DNA binding trench, a series of residues including S119 (Konsavage, Burkholder et al. 2005), N120 (Gerton, Ohgi et al. 1998), C130 (Al-Mawsawi, Fikkert et al. 2006; Al-Mawsawi, Sechi et al. 2007), W132 (Al-Mawsawi, Fikkert et al. 2006), F181, and F185 (Al-Mawsawi, Christ et al. 2008; Al-Mawsawi, Hombrouck et al. 2008) display this activity phenotype. These residues align along the length of one wall of the trench (Figure 4.7, magenta residues). If this trench is involved in binding to the target DNA, we would predict that point mutations introduced at residues on the opposite wall would have the same phenotype. We therefore assembled HIV-1 IN mutants with K211S, K219S, and Q221S, substitutions, respectively. Each was tested for 3' end processing and strand transfer against the homologous HIV-1 DNA substrates. The K211S mutant showed near wild-type 3' processing while the K219S mutant showed a reduction in 3' processing relative to the wild-type activity (Figure 4.8). When tested in the strand transfer assay using pre-processed HIV-1 LTR DNA, we observed that the K211S mutant was as active as wild-type while the K219S mutant was inactive. The W131S and Q221S mutants also showed 3' end processing and strand transfer activity similar to that of the wild-type 3CSF185H. These reactions were tested separately and the specific activity of the strand transfer substrate was reduced relative to the assay with K211S and K219S chimeras. Nevertheless, the interpretation of the results is not changed. Thus, the K219S enzyme lost the ability to strand transfer with a decrease in 3' processing while the W131S, K211S and the Q221S mutants had no detectable affect on either activity.

Mutants at Lys^{34}, Lys^{111}, Gln^{137} and Gln^{214} were also prepared to test additional residues in the target DNA trench. These residues are highlighted in Figure 4.9. Purified proteins were assayed in the 3' end processing and strand transfer assays but were inactive for both activities. Interpretation of these residues being involved in the target DNA binding is therefore inconclusive.

Amino acid residues 186-188 were identified in our alignment and represented in our model in the target DNA trench. While these residues may interact with the target DNA, it was reported in the literature that Lys^{186} lies at a subunit interface and is involved in subunit multimerization (Berthoux, Sebastian et al. 2007). Based on our experience with mutation of other residues at subunit interfaces being inactive in 3' end processing and strand transfer assays *in vitro*, Lys^{186}, Arg^{187} and Lys^{188} were not analyzed by mutation to Ser.

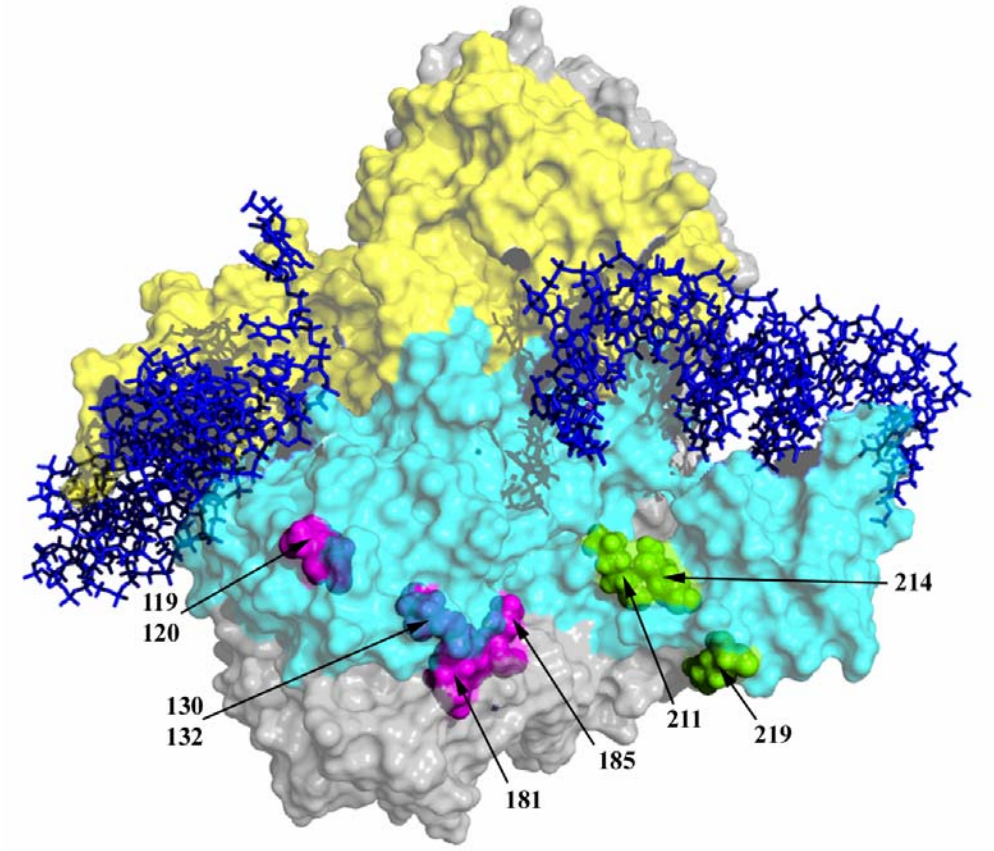


Figure 4.7 Model of residues for target DNA trench

HIV-1 IN tetramer model with residues that are identified in the literature to disrupt the strand transfer reaction are shown in yellow. Residues Ser^{119}, Asn^{120}, Cys^{130}, Cys^{132}, Phe^{181} and Phe^{185} are represented in our model on one side of the putative target DNA binding trench (Gerton, Ohgi et al. 1998; Konsavage, Burkholder et al. 2005; Al-Mawsawi, Fikkert et al. 2006; Al-Mawsawi, Sechi et al. 2007). Residues Lys^{211}, Gln^{214} and Lys^{219} are conserved in our structural alignment and are on the opposite wall of the target DNA trench. As such, these residues were selected for further analysis.

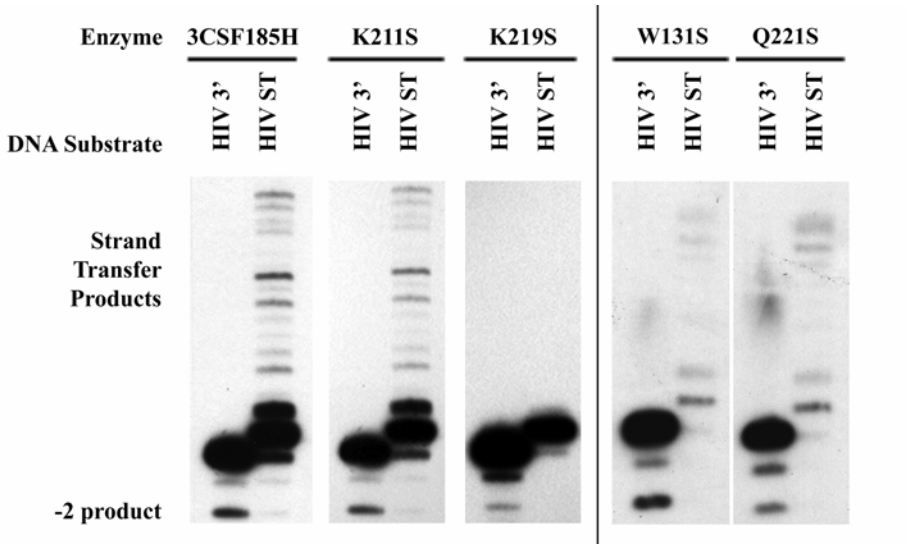


Figure 4.8 ST phenotype of target residues

HIV-1 IN, K211S, K219S and Q221S IN mutants were incubated with HIV-1 U5 LTR duplex substrate to measure 3' end processing (HIV 3') or a pre-processed HIV-1 U5 LTR duplex substrate with two bases removed adjacent to the CA dinucleotide (HIV ST). In the HIV ST lane, products migrating slower than the starting substrate represent integration of one oligo into another.

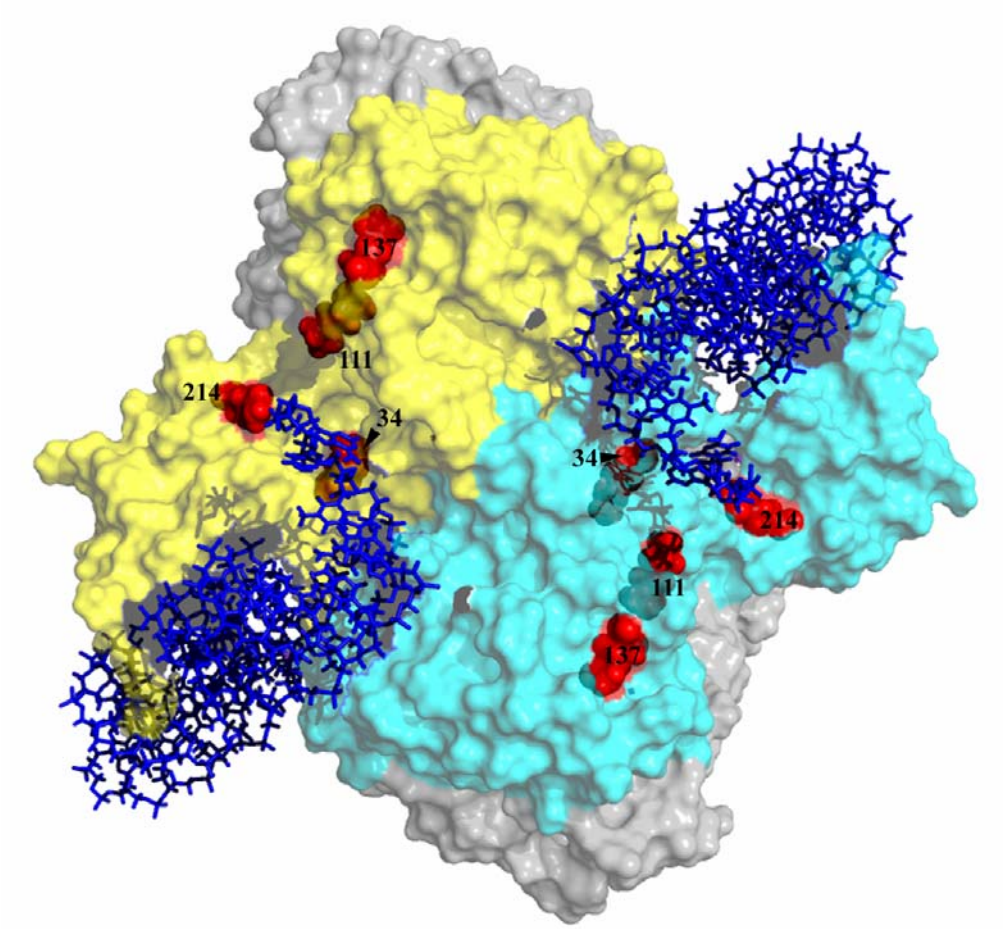


Figure 4.9 Modeling residues that produce enzymes with no activity

Amino acid residues Lys^{34}, Lys^{111}, Gln^{137} and Gln^{214} are represented in the HIV-1 IN tetramer model. These residues are highlighted in red along the target DNA trench. These residues are exposed on the surface of the target DNA trench in two of the four subunits. In the other two subunits, these residues are at interfaces and might be involved in oligomerization.

4.5 Conclusion

Using the tetramer model of HIV-1 IN we have identified a series of amino acid residues that are critical for interaction with the target DNA and strand transfer activity. A structural alignment of HIV-1, SIV and ASV yielded a number of positively-charged amino acid residues that were exposed on the surface of our tetramer model in the putative target DNA trench. Mutation of select residues to Ser resulted in the identification of Lys^{219} as a key amino acid residue that interacts only with the target DNA in a structurally unrelated position to previously published residues. The phenotype of this mutant is in agreement with the data of Heuer and Brown and is contained within a peptide that was determined to only interact with the target DNA (Heuer and Brown 1997). This residue in concert with published residues in the literature defines the target DNA trench in our model.

A number of enzymes were inactive after purification but are not precluded from interacting with the target DNA. As shown in Figure 4.10, these residues are exposed on the surface and may interact with the target DNA. Further analysis of these amino acids in other subunits positions them at interfaces. We conclude that mutation of these residues disrupts the formation of an active tetramer. A summary of all mutations and phenotypes after purification for this part are shown in Figure 4.10.

Mutant	DNA substrate	Substitutions	3' Processing Outcome	Strand Transfer Outcome
S34	Target	K34S	Inactive enzyme	n/a
S111	Target	K111S	Inactive enzyme	n/a
S137	Target	Q137S	Inactive enzyme	n/a
S211	Target	K211S	No change detected	No change detected
S214	Target	Q214S	Inactive enzyme	n/a
S219	Target	K219S	Some defect relative to wild-type	**No ST activity**
S221	Target	Q221S	No change detected	No change detected
S131	Target	W131S	No change detected	No change detected
S223/224	Target	F223S R224S	Inactive enzyme	n/a

Figure 4.10 Enzyme chimeras purified for *in vitro* assay

Summary of HIV-1 IN chimeras purified and tested in both the 3' end processing assay and strand transfer assay as described in Materials and Methods. Phenotypes are summarized for all purified proteins.

CHAPTER 5

INSIGHTS INTO SMALL MOLECULE INHIBITOR SENSITIVITY

5.1 Overview

Small molecule therapeutics for treatment of HIV-1 infection or AIDS is a market worth an estimated $100B by 2010. Current therapy regimens use small molecules that have been developed to inhibit several processes involved in replication, including two out of the three viral enzymes. There is now a FDA approved IN drug in clinical use. Development of inhibitors against IN is critical for the continued battle to slow the HIV-1 epidemic because of the emergence of drug resistance.

As the understanding of IN activity and structure has grown, a number of strategies have been developed for inhibition of the protein. Preventing the conformation flexibility that is required for activity, disrupting IN:IN interactions and oligomerization, interfering with IN:DNA interactions, and inhibition of the catalytic activities involved in integration are the current targets for inhibitor development. Current discovery methods involve high-throughput methods for identifying IN inhibitor candidates. A number of candidate molecules have been tested *in vitro* and in cells for inhibitory activity. Some molecules that have passed this screening process as lead compounds or those for clinical trial are shown in Figure 5.1.

There are several factors that affect the analysis and interpretation of IN inhibitors. Differences between wild-type and mutant INs that are more soluble have presented challenges at the screening level. Some compounds that have been identified with mutant INs *in vitro* have not yielded similar results when tested in cells. Additionally, different inhibitory activities in the presence of Mn^{2+} or Mg^{2+} have been demonstrated for a number of compounds (Marchand, Johnson et al. 2003). The order of addition within the *in vitro* assay is also important as the formation of an IN:DNA complex is critical for eliminating false positive leads (Cherepanov, Este et al. 1997). Taken together, it is thought that the best conditions for identifying small

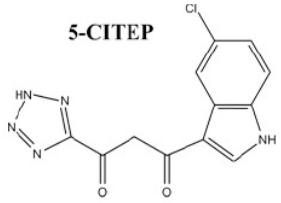
molecules *in vitro* for IN inhibition involves using pre-assembled wild-type HIV-1 IN:viral DNA complexes using Mg^{2+} as the metal co-factor.

A number of compounds exist that are under investigation for inhibition of IN activity. Strand transfer inhibitors of HIV-1 IN include the naphthyridine carboxamides (L-870,810 and L-870, 812), the dihydroquinolone carboxylic acid GS-9137, the styrylquinolone FZ41, and the hydroxypyrimidinone carboxamide MK-0518. (Fesen, Pommier et al. 1994; Levy-Mintz, Duan et al. 1996; Neamati, Hong et al. 1997; Zhao, Neamati et al. 1997; Lutzke and Plasterk 1998; Neamati, Hong et al. 1998; Zouhiri, Mouscadet et al. 2000; Zhuang, Wai et al. 2003; Bonnenfant, Thomas et al. 2004; Deprez, Barbe et al. 2004; Mousnier, Leh et al. 2004; John, Fletcher et al. 2005; Meanwell, Belema et al. 2005; Sato, Motomura et al. 2006). MK-0518 is a FDA approved compound that is the first IN-inhibitor to make it to the market and is known as raltegravir/isentress. Recent clinical studies on raltegravir demonstrate the potential of IN inhibitors for therapy as patients that had never received therapy as well as those with drug-resistant HIV-1 saw dramatic results during a 46-week treatment and observation course (Cooper, Steigbigel et al. 2008; Steigbigel, Cooper et al. 2008).

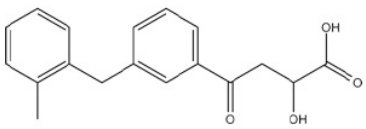
Figure 5.1 Small molecule inhibitors of HIV-1 IN

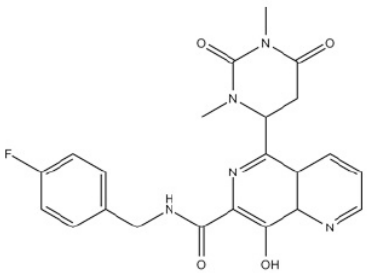
Compounds that have been assayed for specific IN inhibition both *in vitro* and *in vivo*. 5-CITEP is the lead compound that led to the development of diketo acids and the derivatives L-731,988 and S-1360 (diketo analogue). Other compounds that have been investigated for their properties as strand transfer inhibitors of HIV-1 IN include the naphthyridine carboxamides (L-870,810 and L-870, 812), the dihydroquinolone carboxylic acid GS-9137, the styrylquinolone FZ41, and the hydroxypyrimidinone carboxamide MK-0518.

5-CITEP

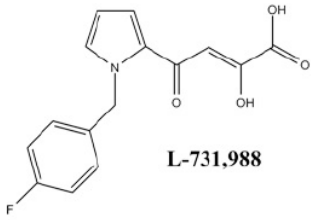


diketo acid

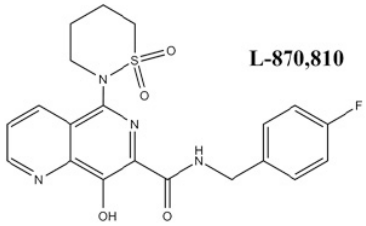




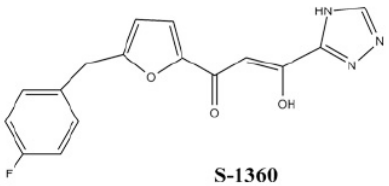
naphthyridine carboxamide



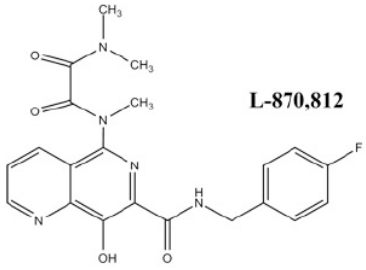
L-731,988



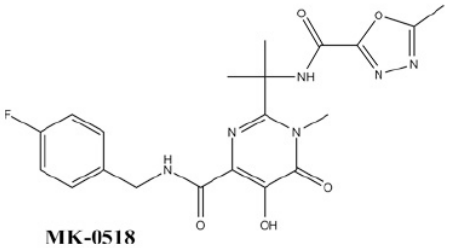
L-870,810



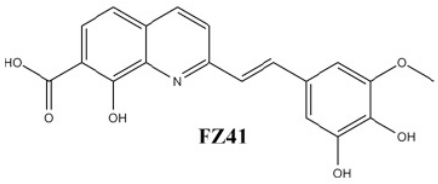
S-1360



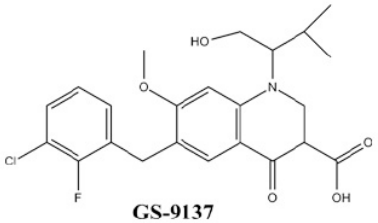
L-870,812



MK-0518



FZ41



GS-9137

5.2 Diketo acid inhibitors

Diketo acids are compounds that have two oxygen atoms that are positioned to chelate the divalent cation in the active site of the IN holoenzyme (Grobler, Stillmock et al. 2002). The DKA compounds are broken into two classes: those that are inhibitory in the presence of both Mn^{2+} and Mg^{2+} or those that inhibit with Mg^{2+} but not Mn^{2+} (Marchand, Johnson et al. 2003).

One of the first compounds of this class was 5-CITEP and this molecule was crystallized with the active site of HIV-1 IN (Lubkowski, Yang et al. 1998; Goldgur, Craigie et al. 1999). One derivative of 5-CITEP, S-1360, was evaluated in clinical trials. The failure of this compound was due to resistance mutation, conferred by residues Thr^{66}, Leu^{74}, Gln^{148}, Ile^{151} and Asn^{155} (Fikkert, 2004). The location of these residues is depicted in Figure 5.2 and are distributed around the active site architecture suggesting a change in the active site can lead to drug resistance. Other diketo acids have been surveyed under cell selection, which reavealed additional residues that confer resistance: Ser^{153}, Met^{154} and Ser^{230} (Espeseth, Felock et al. 2000; Hazuda, Felock et al. 2000; Lee and Robinson 2004).

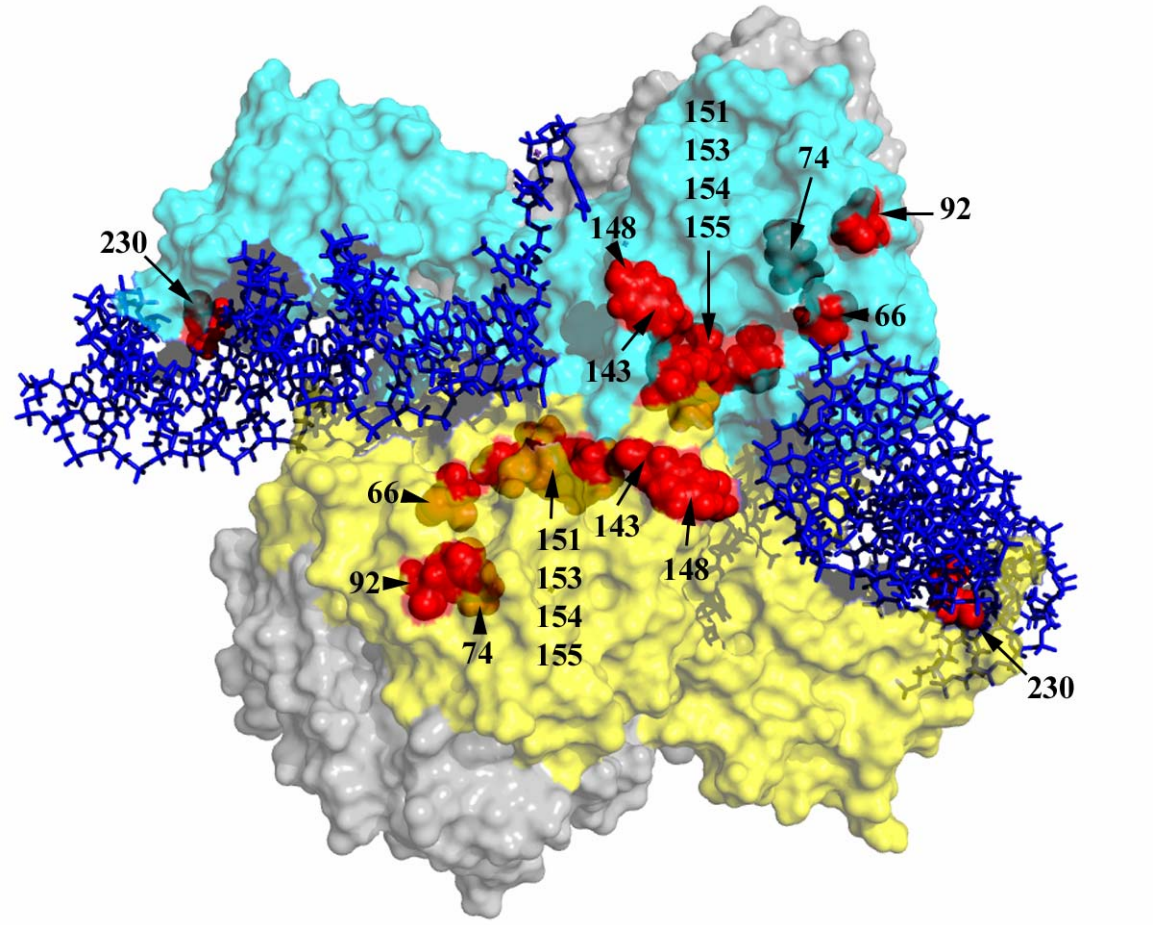


Figure 5.2 Residues that arise during drug selection using diketo acids

Residues indicated in red are reported to mutate during drug selection with diketo acids and derivatives (Espeseth, Felock et al. 2000; Hazuda, Felock et al. 2000; Fikkert, Hombrouck et al. 2004; Lee and Robinson 2004). When these residues are represented in the HIV-1 IN tetramer model, they map to the active site between the LTR ends which is consistent with the proposed mechanism of strand transfer inhibition.

5.3 Naphthyridine carboxamide inhibitors

Napthyridine carboxamides are another class of small molecules that are related to the diketo acids. Substitution of the 1,3-diketone for 8-hydroxy-1,6-naphthyridine or the ketone for 8-hydroxy-1,6-naphthyridin-7-carboxamide yields several derivatives that have shown promise in vitro. The derivatives L-870,810 and L870,812 have been evaluated for inhibitory activity (Hazuda, Anthony et al. 2004; Dayam, Sanchez et al. 2005). Like the diketo acids from which they were derived, the accumulation of resistance mutation in cell-based screening assays confers resistance against these compounds. For L-870,810, mutation at Val^{72}, Phe^{121}, Thr^{125}, and Val^{150} results in resistance to the naphthyridine compound but not other diketo acid molecules (Hazuda, 2004). These data suggested that the two classes of compounds can be developed for therapy because resistant mutations from one class do not confer resistance to the other. Residues selected during naphtyridine carboxamide selection are shown in Figure 5.3, and are more distal to the active site with the exception of Val^{150}.

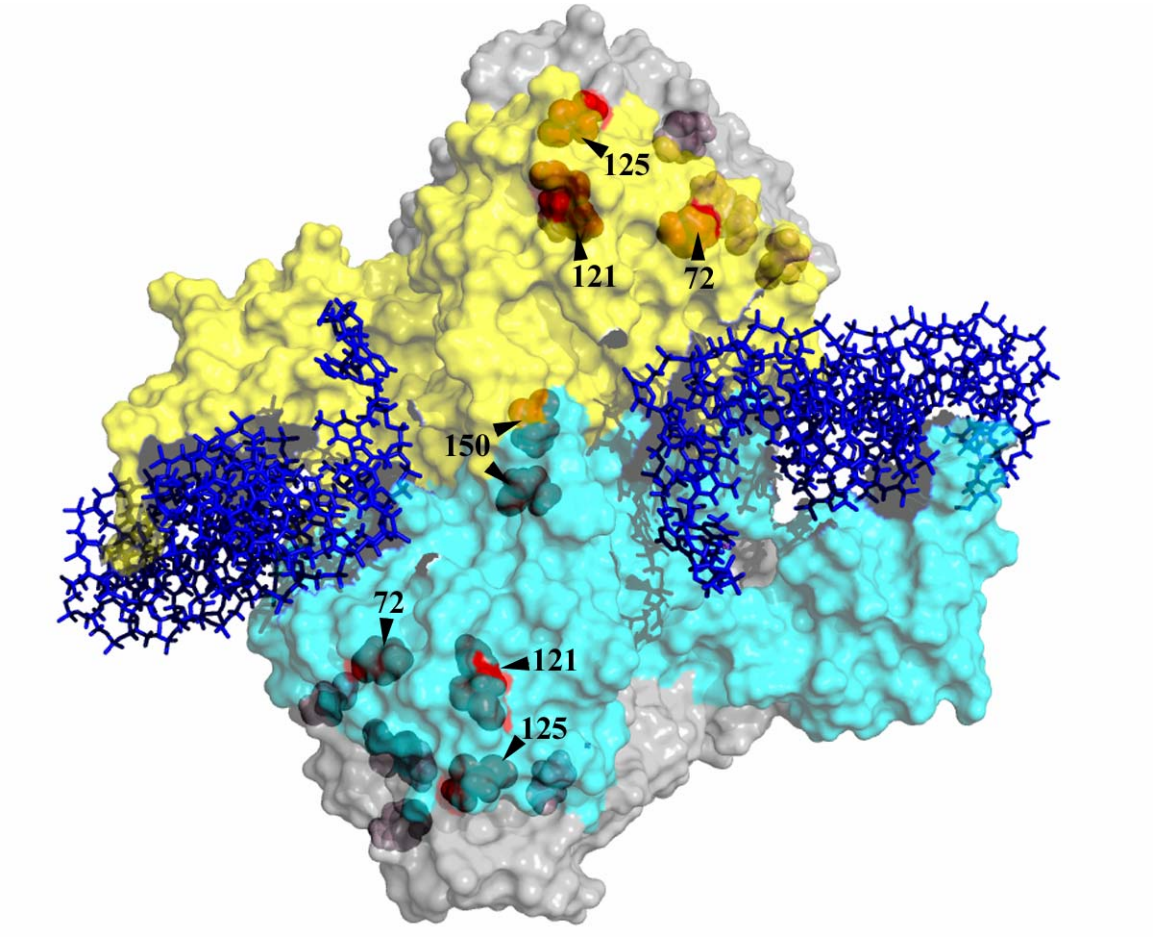


Figure 5.3 Residues that arise during drug selection using napthyridine carboxamides

Drug selection using naphthyridine carboxamide inhibitors results in mutation at the residues indicated in red after nine months (Hazuda, Anthony et al. 2004). These residues arise in several combinations, such as F121Y T125K, V72I F121Y T125K, and V72I F121Y T125K V150I.

5.4 Drug residues that affect LTR specificity

When HIV-1 is replicated in the presence of diketo-acid based compounds, a number of escape mutants are selected that are believed to act against the strand transfer reaction. We previously reported that Val^{72} and Ser^{153} are among the residues that influence LTR selection and these residues are mutated in drug resistant INs (Chen, Weber et al. 2006). Changes at Ser^{230} in HIV-1 IN are also found in drug resistant INs (Fikkert, Van Maele et al. 2003). These chimeras show that these residues are involved in the recognition of the viral DNA, and that while the small compounds may be directed at inhibiting the strand transfer reaction, there may also be effects at the level of 3' end processing.

5.5 Activity complementation during drug selection

When a substitution in integrase occurs that disrupts its catalytic activity (3' processing or strand transfer) and the virus is maintained in the presence of the drug, we hypothesize second site mutations in IN are selected to compensate for the lost activity caused by the original mutation. In the case of Val^{72}, which has reduced 3' end processing of HIV-1 substrates (Chen, Weber et al. 2006), second site substitutions at Phe^{121}, Thr^{125}, and Val^{150} have been reported (Hazuda, Anthony et al. 2004). A chimeric IN with a T125S substitution resulted in an enzyme that increased its 3' processing reaction relative to wild type but did not alter 3' processing activity towards the ASV substrate (Chen, Weber et al. 2006). When the V72W and T125S substitutions were combined into a single enzyme, the resultant chimera had a 3' end processing activity equivalent to wild-type (Figure 5.4). This result demonstrates that at least one mutation at a second site associated with substitutions at residue Val^{72} can compensate for its decreased 3' processing activity.

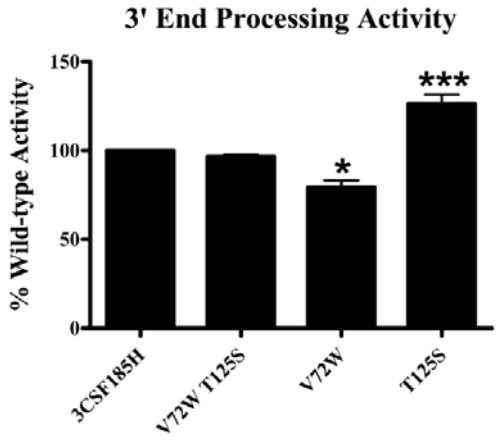


Figure 5.4 Assay of HIV-1 IN V72W T125S chimera 3' end processing activity

The 3' end processing activity of HIV-1 U5 LTR substrate by HIV-1 IN (3CSF185H), V72W, T125S, and V72W T125S enzymes was determined as previously described. The relative processing activity was calculated by densitometric analysis from three different experiments of the -2 products relative to total substrate and compared to 3CSF185H. For statistical significance, *p* values of ≤0.05 (*) or ≤0.001 (***) are indicated.

5.6 Conclusion

We have shown that we can explain, at least for one combination of resistance mutants, that the *in vitro* activity of these mutants in single can be correlated to the need for multiple mutations. Val^{72} mutants have reduced 3' end processing ability, and thus lower replication fitness. To restore this activity, mutation at Thr^{125} is necessary to return to wild-type levels. There is still a gap in the understanding however, as there are two additional residues that occur in this selection. Since the V72W T125S chimera does not exhibit strand transfer using either a pre-processed HIV-1 or ASV LTR DNA, it is likely that mutation at Ile^{121} and Val^{150} are involved with that activity. We attempted to investigate this hypothesis by creating a chimeric enzyme with mutation at Val^{72}, Thr^{125}, and Ile^{121}. Unfortunately, this mutant was insoluble and did not display any activity in the *in vitro* 3' processing or strand transfer assays. A chimera with all four mutations (Val^{72}, Thr^{125}, Ile^{121} and Val^{150}) may be more soluble than the triple mutant. Alternatively, preparing this enzyme chimera in a backbone that does not have solubility-enhancing mutations or purification through a denaturing-renaturing protocol might allow for an active enzyme that can be investigated *in vitro*.

CHAPTER 6

DISCUSSION

6.1 Summary

The goal of our study was to identify contacts between HIV-1 Integrase and both the viral and target DNA substrates. While dimers of IN are capable of catalyzing 3' processing and strand transfer reactions, they do not support a concerted DNA integration reaction (Faure, Calmels et al. 2005). For this reason, we assembled a homotetramer model of HIV-1 IN to predict residues that interact with DNA substrates.

Analysis of our tetramer model for HIV-1 IN with 20 base pairs of the U3 and U5 LTR termini predicted residues that interact with the viral LTR DNA. We hypothesized that mutating the amino acids from HIV-1 to their structurally conserved equivalent from ASV that we could change the specificity of the 3' end processing reaction. The initial study from our lab identified nine residues that were involved in the specific recognition of viral LTR ends (Chen, Weber et al. 2006). It was apparent that this analysis was incomplete because an enzyme chimera which combined these substitutions into a single construct failed to fully convert the 3' end processing reaction from HIV-1 to ASV DNA. In the current study, we extended this work to now include at least fifteen residues involved in the specificity reaction for 3' end processing. Additionally, when at least twelve of these residues were combined into a single chimera there was a noticeable change in specificity toward the ASV substrate with a loss in specific processing of the HIV-1 LTR DNA.

Further examination of the model suggested a surface trench suitable for binding of target DNA, which runs perpendicular to the plane of the LTR binding sites and is appropriately positioned for nucleophilic attack and strand transfer reactions. The proposed binding site for target DNA is consistent with numerous lines of evidence from photo cross-linking and mutational analysis that result in altered or unaltered IN activity. In our study we identified an

amino acid that defines the opposing side of the target DNA trench, as predicted in our model. In combination with all published residues we now have a representation of the target DNA trench using our model of the HIV-1 IN tetramer.

Additionally, we have performed the first *in vitro* investigation to explain activities observed when virus is placed under drug selection. By analyzing the *in vitro* activities of substitutions at amino acid residues that are changed we can correlate the selection with secondary and tertiary mutations. Finally, we have assembled a representation of the HIV-1 IN tetramer docked to two LEDGF molecules in collaboration with Irene Weber and Rob Harrison (Georgia State University). This interaction is thought to facilitate binding to relaxed chromatin influencing the integration of viral into host DNA. This discovery reveals a new application of this model for future exploration.

6.2 Interactions with LTR DNA

On the surface of IN there are a minimum of fifteen residue positions associated with LTR specificity in each LTR DNA binding site. They are predicted to interact asymmetrically along a 15-16 base length of DNA duplex (see Figure 3.8). Figure 6.1 allows for easy identification of the viral DNA trench.

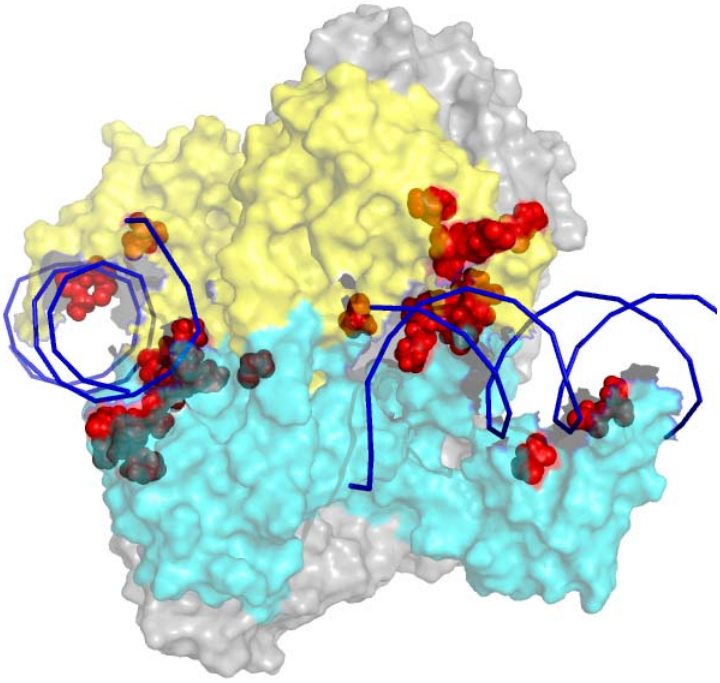


Figure 6.1 LTR residues positioned in the model

Amino acid residues implicated in interacting with the viral LTR DNA end and affecting the substrate specificity for 3' processing in our study are depicted in the HIV-1 IN tetramer model. The model representing the HIV-1 IN tetramer has residues Leu^{68}, Glu^{69}, Val^{72}, Ser^{153}, Lys^{160}, Ile^{161}, Glu^{163}, Gln^{164}, Val^{165}, His^{171}, Leu^{172}, Asp^{229}, Ser^{230} and Asp^{253} highlighted in red. Viral LTR DNA is just displayed in blue ribbon format so as to not obscure the view of the viral LTR DNA trench.

When LTR DNA is bound to IN and the complex treated with DNase, a 16 base pair duplex length of the LTR is protected from digestion (Vora and Grandgenett 2001; Markowitz, Morales-Ramirez et al. 2006). Thus there is agreement between the size of the protected DNA and residues in close proximity to the LTR DNA that change specificity for substrate. In addition, interactions between viral DNA and HIV-1 IN have been demonstrated in cross-linking studies for residues 143, 148, 156, 159, 160, 230, 246, 262, 263, and 264 (Lutzke, Vink et al. 1994; Jenkins, Esposito et al. 1997; Drake, Neamati et al. 1998; Esposito and Craigie 1998; Gao, Butler et al. 2001). These residues are highlighted in Figure 6.2. Many cluster in and around the catalytic site in close proximity to the first 6 bases/base-pairs of the processed LTR ends. Residues 246, 262, 263, and 264 are in close proximity to base-pairs 15-16 of the LTR DNA that interact with residues 229, 230, and 253 identified in this study. Finally, Agapkina et al. used substrate analogs to probe contacts between HIV-1 IN and LTR DNA substrates (Agapkina, Smolov et al. 2006). In this study they identified eleven contacts with the sugar phosphate backbone from residues 5-9 and interactions with three bases asymmetrically distributed between the two strands of the LTR ends. Figure 6.3 shows the DNA interactions with the heterocyclic base (orange) or sugar phosphate backbone (magenta). The fifteen residues we identified that influence specificity for 3' processing reactions are spatially in close proximity to all of these sugar phosphate backbone contacts and most of the base contacts (shown in red space-fill, Figure 6.3). Therefore, we propose that our model is an accurate representation of the HIV-1 tetramer complexed with the viral DNA. In the absence of a crystal structure of the HIV-1 IN holoenzyme, we believe that our model is very useful for investigating the interactions of the enzyme with the viral DNA substrate.

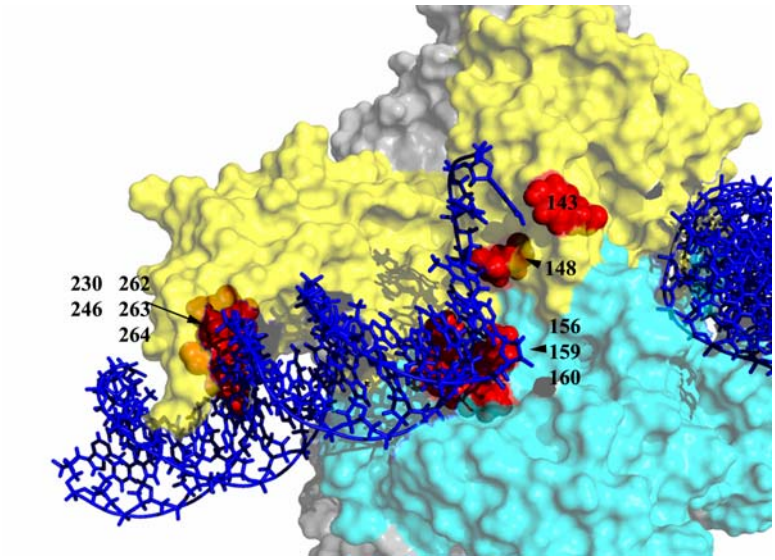


Figure 6.2 Residues reported to cross-link to viral LTR DNA in the literature

Residues reported in the literature to cross-link to viral LTR DNA. Interactions between viral DNA and HIV-1 IN have been demonstrated for residues Tyr^{143}, Gln^{148}, Lys^{156}, Lys^{159}, Lys^{160}, Ser^{230}, Glu^{246}, Arg^{262}, Arg^{263}, and Lys^{264} (Jenkins, Esposito et al. 1997; Drake, Neamati et al. 1998; Esposito and Craigie 1998; Heuer and Brown 1998; Gao, Butler et al. 2001). These residues are colored red in space-fill representation for interaction with one LTR end.

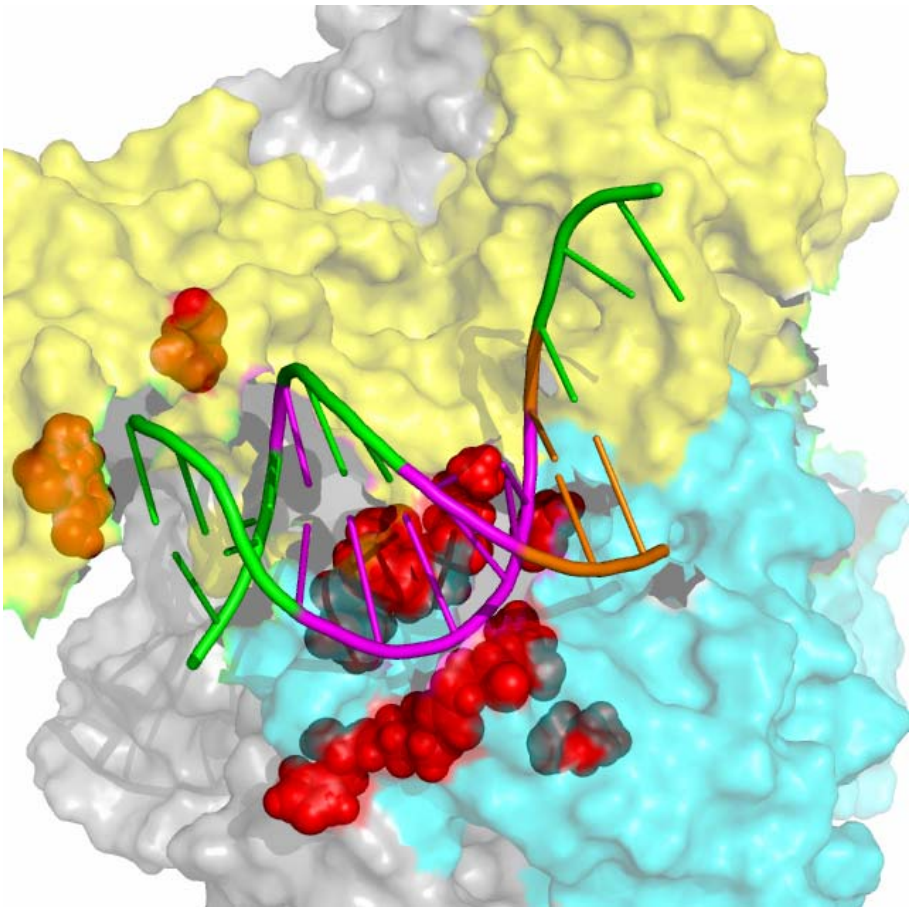


Figure 6.3 Interactions based on nucleotide analog investigation

Using a novel DNA analog for studying HIV-1 IN:DNA interactions, Agapkina et al. identify metal-dependent and metal-independent interactions between enzyme and substrate (Agapkina, Smolov et al. 2006). Twelve base-pairs of the DNA double-helix are displayed in green, with interactions that involve contact with the heterocyclic bases are colored in orange. Interactions with the sugar phosphate backbone are colored in purple. Residues identified in our analysis to be involved in the specific recognition of viral LTR DNA are highlighted in red.

To assess amino acid residues that were involved in the specific recognition of the viral DNA, we chose to mutate structurally unique residues from HIV-1 to the corresponding ASV amino acid. The HIV-1 chimeras were assembled in a 3CSF185H background, an enzyme which has four amino acid substitutions to improve solubility. Individually these amino acid substitutions have little or no effect on viral replication (Engelman, Liu et al. 1997; Bischerour, Tauc et al. 2003; Zhu, Dobard et al. 2004). Nevertheless, we assembled a S7C IN chimera in which the substitutions at positions 185 and 280 were restored to wild-type residues. Unfortunately this resulted in an enzyme that was insoluble; 3' processing activity could be restored by renaturation from urea and the activity of this enzyme was qualitatively similar to that shown for S7C in Figure 3.10. This indicates that these two substitutions did not affect the specificity of the 3' processing reaction. While we have been able to alter the specificity for 3' processing of LTR substrates, we have not been able to demonstrate changes in strand transfer activity. This is because several of the individual amino acid changes that we introduced into the HIV-1 IN that alter 3' processing disrupt the strand transfer activity towards the HIV-1 substrates for unknown reasons. When we combined multiple substitutions into the soluble form of S7C, it too was unable to support a strand transfer reaction with either the HIV-1 U5 LTR or ASV U3 LTR preprocessed end substrates. One possible explanation for this behavior might be that because IN is a tetramer, single substitutions targeted to one residue will necessarily change in all four subunits and some of these other substitutions might disrupt DNA binding, oligomerization, introduce conformational distortions, or have other effects that result in a loss of strand transfer activity.

6.3 Implications for IN inhibitor understanding and design

There are a series of naphthyridine carboxamide and diketo acid related drugs that act in the nanomolar range to inhibit HIV-1 IN (Hazuda, Felock et al. 2000; Grobler, Stillmock et al. 2002; Marchand, Zhang et al. 2002; Pluymers, Pais et al. 2002; Fikkert, Van Maele et al. 2003; Fikkert, Hombrouck et al. 2004; Hazuda, Anthony et al. 2004; Lee and Robinson 2004; Embrey, Wai et al. 2005; Guare, Wai et al. 2006). Drug resistant enzymes with changes in at least ten different sites were identified in the literature. Five of these residues were unique in the structural alignment of different INs and located near the LTR ends in the structural model. While these drugs were thought to act at strand transfer and not 3' processing, we found that HIV-1 IN residues S153 and V72 (Chen, Weber et al. 2006) and S230 were among those positions involved in LTR end recognition and 3' processing. Because drug resistant sites affect specific recognition of the viral DNA ends and change the rate of processing of HIV-1 substrates, we predict that amino acid changes at some of these sites will lead to partially defective integrases in cells. This should subsequently result in selection of second site substitutions that compensate for the loss in 3' processing activity caused by the initial drug resistant amino acid substitutions. If correct, we predict that depending upon the extent of change in 3' processing observed towards HIV-1 duplex substrates (Chen, Weber et al. 2006), our data will be correlated to the appearance of individual or multiple residue substitutions detected in drug resistant enzymes. For example, in the case of a position 153 chimera, it gains the ability to 3' process the ASV LTR end duplex with only a small decrease in its ability to 3' process the HIV-1 duplex substrate (Chen, Weber et al. 2006). As such, we would predict that this mutation would be found by itself and should have only a small affect on replication of HIV-1 in cells, as observed (Lee and Robinson 2004; Lu, Limon et al. 2005). In contrast, the substitution at position 72 (V72W) caused a larger decrease

in the ability to process the HIV-1 U5 duplex substrate (Chen, Weber et al. 2006). On this basis, we would predict that the V72I drug resistant mutation would appear in the presence of other substitutions that compensate for the loss in its 3' processing towards HIV-1 substrates. Second site mutations of F121Y and T125K subsequently appear in HIV-1 IN containing the V72I mutation (Hazuda, Anthony et al. 2004). A T125S substitution increases the 3' processing of U5 HIV-1 duplex (Chen, Weber et al. 2006) as well as joining of a HIV-1 preprocessed substrate. In our study, we show that when the T125S substitution is combined with the V72W mutation, this produces an enzyme with near wild type levels of 3' processing, suggesting that a second site mutation compensates for the decrease in 3' processing caused by the initial drug resistant mutation. A second illustrative example involves position 230 where substitutions at this residue also affect recognition of the viral DNA ends. With the caveat that the exchange of S230E was analyzed as a double mutant in combination with D229I, it gained the ability to cleave the ASV substrate, but in contrast to chimeras with changes at positions 72 or 153, it displayed an increase in activity towards HIV-1 substrates. Changes at position 230 are reported to appear in conjunction with T66I and M74L substitutions in cells (Fikkert, Van Maele et al. 2003). We have not analyzed substitutions at position 66 in vitro, but Lee and Robinson reported that the T66I substitution caused a small decrease in 3' processing (Lee and Robinson 2004). We tested a M74A substitution that resulted in a significant loss of 3' processing of HIV-1 substrates. Taken together, this suggests that substitution at position 230 might compensate for the loss in 3' processing caused by mutations at position 74 and 66. Our study is the first such analysis that tested the enzymatic activities of residues that arise during drug selection *in vitro*. This analysis can be extended to future drug resistance residues for interpreting the mechanism of action for the small molecule.

6.4 Interaction between the HIV-1 IN tetramer and LEDGF/p75

Finally, we docked the LEDGF IN binding domains using an energy minimalization simulation and report that there is potential for two sites of interaction between the HIV-1 IN tetramer and LEDGF/p75 (Figure 6.4). Structural data implicate residues 102, 128, 129 and 132 in one subunit and residues 174 and 178 in a second subunit to form a pocket for interaction with LEDGF (Cherepanov, Ambrosio et al. 2005). These observations are consistent with interactions observed in our model. Additional residues that may be involved in this interaction based on mutagenesis and structural data include 131, 161, 165, 166, 168, and 170-173 (Busschots, Vercammen et al. 2005; Cherepanov, Ambrosio et al. 2005; Rahman, Lu et al. 2007). These are also located at or near the interaction interface. Residues reported to interact directly with LEDGF are colored red in the model, while those residues that disrupt IN:LEDGF interaction after mutation are colored in yellow. The positioning of LEDGF as an extension of the target DNA binding trench (indicated by arrows, Figure 6.4) would be consistent with its role in interacting with chromatin to influence target site selection (Vanegas, Llano et al. 2005; Ciuffi and Bushman 2006; Llano, Saenz et al. 2006; Llano, Vanegas et al. 2006; Raghavendra and Engelman 2007).

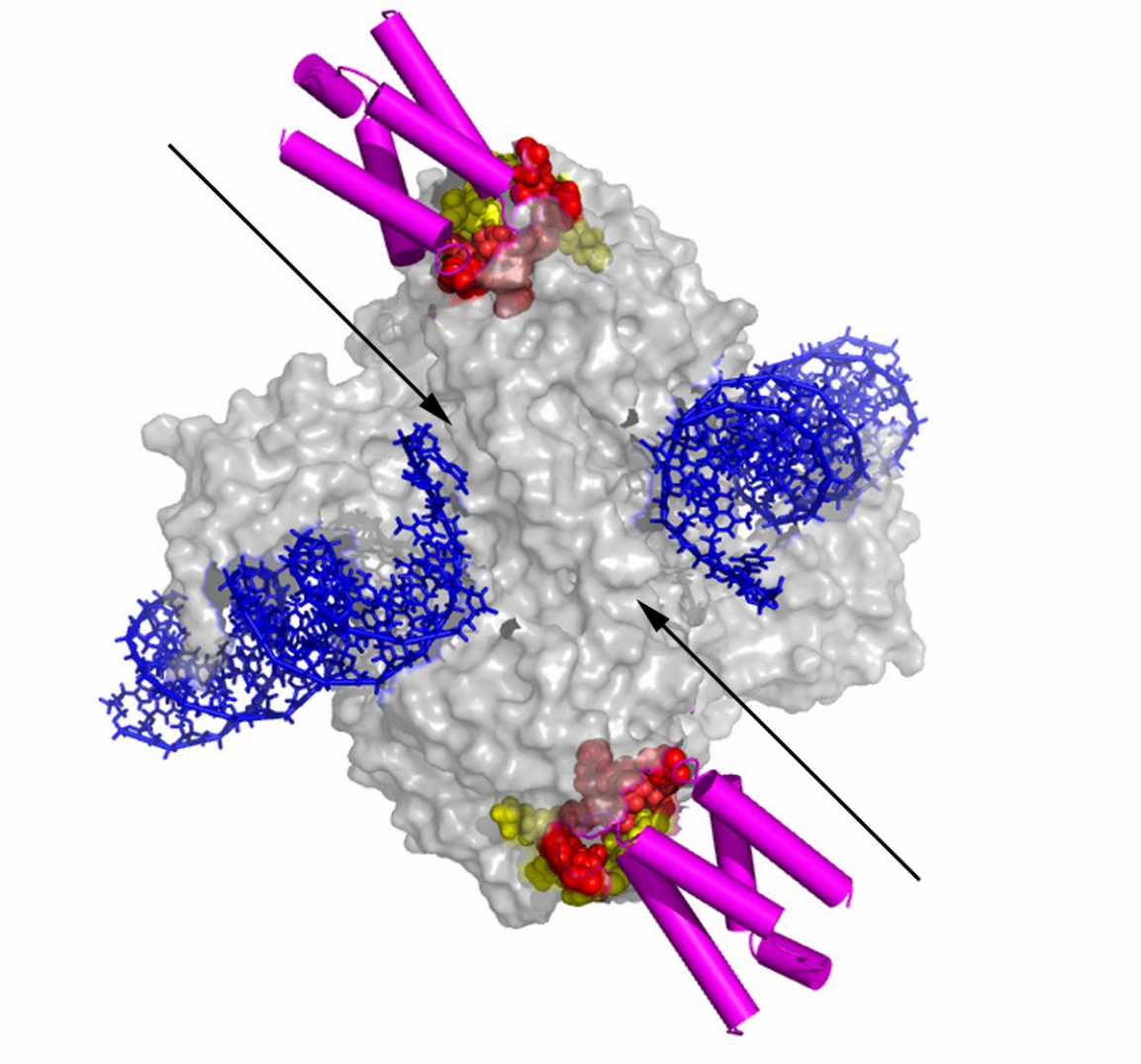


Figure 6.4 Modeling of HIV-1 IN with LEDGF/p75.

Using an energy minimalization simulation, LEDGF/p75 was docked with the HIV-1 IN tetramer model. The LEDGF/p75 IN interacting domains are shown in green, with the proposed tetramer of HIV-1 IN shown in gray. Residues reported to interact directly with LEDGF are visualized in red, while residues that affect the IN:LEDGF interaction when mutated are shown in yellow (Busschots, Vercammen et al. 2005; Cherepanov, Ambrosio et al. 2005; Rahman, Lu et al. 2007).

6.5 Defining the target DNA trench

In examining the structural model, we identified a trench on the HIV-1 IN surface with its long axis almost perpendicular to those accommodating the viral DNA ends (Chen, Weber et al. 2006). We speculate that the target DNA fits into this groove. Moreover, interactions with LEDGF will further stabilize this IN:target DNA complex and would be consistent with the role for LEDGF in promoting the interaction of the integration complex with host chromosomal DNA (Cherepanov, Maertens et al. 2003; Maertens, Cherepanov et al. 2003; Busschots, Vercammen et al. 2005; Ciuffi, Llano et al. 2005; Emiliani, Mousnier et al. 2005; Llano, Saenz et al. 2006). The target DNA is positioned between the viral DNA and this orientation will facilitate the nucleophilic attack of the 3' hydroxyl ends of the respective CA strands into each strand of the target DNA. There are several lines of evidence that support this hypothesis.

First, amino acid residues S119, N120, C130, W132, and K159 are reported to interact with target DNA based upon activity and drug sensitivity data (Gerton, Ohgi et al. 1998; Harper, Skinner et al. 2001; Konsavage, Burkholder et al. 2005; Lu, Limon et al. 2005; Al-Mawsawi, Fikkert et al. 2006). These residues strikingly align along one surface wall of the proposed target DNA binding trench. A N120S mutant is reported to increase 3' processing and strand transfer activities while N120Q and N120K mutants show little effect on processing but some decrease in strand transfer (Gerton, Ohgi et al. 1998; Lu, Limon et al. 2005). The C130S and W132A/G/R substitutions are reported to have normal 3' processing but little or no joining activity (Al-Mawsawi, Fikkert et al. 2006). A C130S substitution in combination with three other mutations shows loss in strand transfer but with some decrease in 3' processing (Zhu, Dobard et al. 2004). An ASV mutant, structurally equivalent to HIV-1 S119, has normal 3' processing but barely detectable strand transfer activity (Konsavage, Burkholder et al. 2005). More recently, Nouri

Neamati (USC) made four HIV-1 IN mutants with W132Y, M178C, F181G, and F185G substitutions, respectively. The enzymes with mutations at positions Trp^{132}, Phe^{181}, and Phe^{185} did not support a strand transfer reaction but had wild type or near wild type levels of 3' processing. In contrast, the mutation at Met^{178} showed decreases in both activities (Al-Mawsawi, Christ et al. 2008; Al-Mawsawi, Hombrouck et al. 2008). As shown in Figure 6.5, Trp^{132}, Phe^{181}, and Phe^{185} lie on the IN surface in the putative target DNA binding trench (green residues) and are aligned with other residues that cause similar activity defects. In contrast, residue Met^{178} lies below the surface of the putative target DNA binding trench. As reported here, the K219S mutation also loses strand transfer with little effect on 3' processing activity. However, in contrast to the above residues, Lys^{219} is found on the opposite wall of the trench (Figure 6.5, red residues). A K219A mutation has been analyzed for its effect on HIV-1 replication and was reported to have a limited affect (Lu, Limon et al. 2004). It is not known why it did not show a stronger phenotype. This may be related to alanine rather than serine being substituted in this study or reflects differences in sensitivity between *in vitro* and cell culture replication assays.

Second, the target DNA binding site contains peptides previously shown to be cross-linked to the target DNA portion of a disintegration substrate modified with an azidophenacyl group. After UV photo activation, cross-links between the DNA substrate and the six endoproteinases Glu-C digested peptides. In terms of our model, the peptide 139-152 represents the active site between the two LTR ends and contains Gln^{148} as well as Gln^{137}, Gln^{146}, and Asn^{144}. A second peptide, 213-247 is found at a distance from the catalytic site in the putative target DNA binding site and contains Lys^{219} but not Lys^{211}. In contrast to K219S, we report the K211S mutation has no affect on activity of IN *in vitro*. The other peptides identified in the

report were implicated in binding viral, target, or both DNA substrates in agreement with the model's predictions (Figure 6.2).

Third, we find a series of residues (Arg, Lys, Gln, and Asn) that appear along the length of the putative DNA binding pocket, which are found in known DNA binding sites of other enzymes (Winkler, Banner et al. 1993; Jeltsch, Wenz et al. 1996; Yang, Horton et al. 2005) and could therefore be involved in binding to the target DNA (Figure 4.4). In contrast to the residues interacting with the LTR ends, these residues are conserved among INs to different extents. Because IN inserts the viral DNA into many sites in the target DNA, conservation would be consistent with non-specific target DNA site selection. Five of these residues (Gln^{62}, Asn^{117}, Gln^{148}, Asn^{155}, and Lys^{159}) have been mutated and cause defects to 3' processing, strand transfer, and disintegration (Lutzke, Vink et al. 1994; Engelman, Liu et al. 1997; Jenkins, Esposito et al. 1997; Esposito and Craigie 1998; Gerton, Ohgi et al. 1998). These residues are predicted to lie in the catalytic site between the ends of the two LTRs so they could interact with both the viral and the target DNAs. This conclusion is supported by a recent study that showed that Gln^{148} cross-linked to the ends of the LTRs (Johnson, Santos et al. 2006).

Fourth, when we examined the positions in the structural model of 50 amino acid residues described in the literature where mutations result in either little or no effect on or decreases in both 3' processing and joining activities, none lay in the proposed target DNA binding trench (Vincent, Ellison et al. 1993; Lutzke, Vink et al. 1994; Engelman, Liu et al. 1997; Gerton, Ohgi et al. 1998; Sayasith, Sauve et al. 2000; Bischerour, Leh et al. 2003; Priet, Navarro et al. 2003). The only exception is when the targeted amino acids were positioned between the two LTR ends where they could interact with both viral and target DNAs. Additionally, Puglia et al. reported an analysis of HIV-1 IN where in-frame insertions of small peptides were placed

at 56 sites (Puglia, Wang et al. 2006). The mutants were analyzed for changes in the joining reaction (but not 3' processing because a preprocessed substrate was used). We examined the positions of these mutations in our structural model and can interpret their reported activity changes. For example, when the bulky insertions are near the enzyme surface but not near the viral DNA ends or the proposed target DNA binding site, the model predicts and Puglia et al. report that there is no effect on activity (Puglia, Wang et al. 2006). In contrast, when the peptide insertions are at the surface near either the viral DNA or target DNA binding sites, we predict that there should be a disruption to the joining reaction and this is observed. When insertions are buried within the structure we predict distortions that disrupt all activities by affecting the oligomerization of HIV-1 IN and this too is seen. Taken together, these results are consistent with the model and the hypothesis for the binding of target DNA.

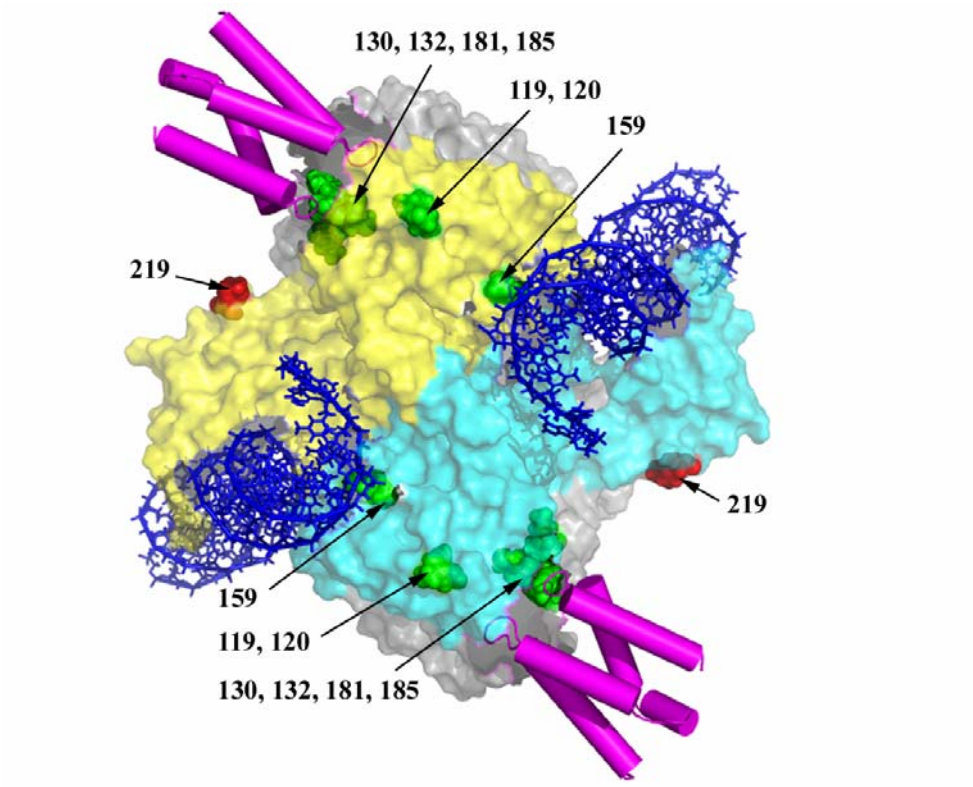


Figure 6.5 Positioning of residues reported in the literature that affect target DNA

The four subunits of IN are shown in surface representation to reveal the proposed trench for binding target DNA. The IN subunits are colored with cyan and yellow indicating subunits interacting with the LTRs. The LTR DNA ends are shown in blue. Residues 119, 120, 130, 132, 159(Gerton, Ohgi et al. 1998; Harper, Skinner et al. 2001; Konsavage, Burkholder et al. 2005; Lu, Limon et al. 2005; Al-Mawsawi, Fikkert et al. 2006), 181, and 185 (Al-Mawsawi, Christ et al. 2008; Al-Mawsawi, Hombrouck et al. 2008) that affect target DNA binding are shown in green space fill models. Residue 219 (red) is in peptide 213-247 known to cross-link to target DNA in a disintegration model substrate (Heuer and Brown 1997). The LEDGF/p75 IN binding domains are shown in magenta.

6.6 Closing Remarks

This project has extended work that was previously performed by identifying novel amino acids that are important for the *in vitro* 3' end processing enzymatic activity of HIV-1 IN. Additionally, we are the first to report an architecture for the binding of target DNA based on the *in vitro* analysis of a single amino acid that disrupted the strand transfer activity. Finally, we demonstrated that the *in vitro* phenotypes of residues can be correlated to their activity and be predictive of compensating mutations when put under drug selection.

Aside from the reported data, several things can be taken away from this project. First, an appreciation for the sensitivity within the enzyme oligomer was only developed after repeated mutagenesis and purification of amino acids in certain regions. More specifically, purification of mutation at Lys^{111} and Gln^{137} was unsuccessful, even after adaptation of the purification protocol with varying the ionic strength and detergent conditions of the purification buffer. Only upon further investigation of these residues within the tetramer model did we determine that, while exposed in two subunits, they also lie at an interface that must be critical for multimerization. Our analysis identified many residues that are conserved in at least four retroviral integrases. Keeping in mind that the IN holoenzyme is composed of four subunits, this approach of identifying highly conserved residues that are both exposed and buried may be very useful if adapted in a drug discovery program. By applying this information to design compounds to target those residues that are both exposed and critical for oligomerization, the development of resistance mutations could be slowed. While one limitation for this is that our model is indeed just a model, we have shown it to be highly predictive for interactions between the enzyme and substrate.

An extension of this understanding can be applied in the drug discovery of small molecule inhibitors of HIV-1 IN. As noted previously, small-molecule inhibitors are of great interest now for therapeutic development. Instead of large-scale *in vitro* analysis of compounds, we propose using our computer model for *in silico* analysis to more efficiently identify molecules that would be effective against the IN multimers. The strand transfer inhibitor (STI) class of compounds might benefit most from this approach because these compounds currently have targets that are at or near the active site. The mechanism of action for many of these compounds is the chelation of the divalent cation and inhibition of enzymatic activity. Using an *in silico* approach, targeting residues that disrupt the interaction with the target DNA by steric hindrance or perturbing the interaction of IN with host proteins like LEDGF/p75 will have profound effects on advancing the therapeutic options for controlling HIV infection.

Finally, gene therapy is a potential application for the retroviral integrase. Current retroviral gene therapies are not feasible because of the random nature of the integration event. Hypothetically, if you can target the vector to a specific locus, the negative effects of gene deregulation and chromatin disruption/distortion are avoided. Stated more simply, if you can target the defective gene specifically with a functional replacement, with little to no risk of integration anywhere else in the chromosome, you have an effective therapy. What this requires, and what this project worked towards, was an understanding of the interaction between integrase and the DNA substrates. Theoretically, engineering the target DNA binding site to specifically recognize a DNA sequence will allow for directed integration of the therapeutic vector. We have shown that the specificity of the viral DNA trench can be altered by mutation; translation of that work to the target DNA trench, and actually incorporating a specificity requirement to the target substrate interaction is a natural extension of this work.

Altogether, our report advances the understanding of HIV-1 activity for both the 3' end processing and strand transfer activities. We have proposed, tested, and validated a computer model for the interaction of HIV-1 IN with DNA substrates. Lastly, we have tested residues that arise during drug selection *in vitro* and can explain their phenotypic development based on these activities.

References

Agapkina, J., M. Smolov, et al. (2006). "Probing of HIV-1 integrase/DNA interactions using novel analogs of viral DNA." J Biol Chem **281**(17): 11530-40.

Aiken, C., J. Konner, et al. (1994). "Nef induces CD4 endocytosis: requirement for a critical dileucine motif in the membrane-proximal CD4 cytoplasmic domain." Cell **76**(5): 853-64.

Aiyar, A., P. Hindmarsh, et al. (1996). "Concerted integration of linear retroviral DNA by the avian sarcoma virus integrase in vitro: dependence on both long terminal repeat termini." J Virol **70**(6): 3571-80.

Al-Mawsawi, L. Q., F. Christ, et al. (2008). "Inhibitory profile of a LEDGF/p75 peptide against HIV-1 integrase: insight into integrase-DNA complex formation and catalysis." FEBS Lett **582**(10): 1425-30.

Al-Mawsawi, L. Q., V. Fikkert, et al. (2006). "Discovery of a small-molecule HIV-1 integrase inhibitor-binding site." Proc Natl Acad Sci U S A **103**(26): 10080-5.

Al-Mawsawi, L. Q., A. Hombrouck, et al. (2008). "Four-tiered pi interaction at the dimeric interface of HIV-1 integrase critical for DNA integration and viral infectivity." Virology **377**(2): 355-63.

Al-Mawsawi, L. Q., M. Sechi, et al. (2007). "Single amino acid substitution in HIV-1 integrase catalytic core causes a dramatic shift in inhibitor selectivity." FEBS Lett **581**(6): 1151-6.

Andrake, M. D. and A. M. Skalka (1996). "Retroviral integrase, putting the pieces together." J Biol Chem **271**(33): 19633-6.

Bagossi, P., G. Zahucky, et al. (1999). "Improved Parameters for Generating Partial Charges: Correlation with Observed Dipole Moments." J Mol Model **5**: 143-152.

Balvay, L., M. Lopez Lastra, et al. (2007). "Translational control of retroviruses." Nat Rev Microbiol **5**(2): 128-40.

Bao, K. K., H. Wang, et al. (2003). "Functional oligomeric state of avian sarcoma virus integrase." J Biol Chem **278**(2): 1323-7.

Barre-Sinoussi, F., J. C. Chermann, et al. (1983). "Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS)." Science **220**(4599): 868-71.

Beitzel, B. and F. Bushman (2003). "Construction and analysis of cells lacking the HMGA gene family." Nucleic Acids Res **31**(17): 5025-32.

Berkhout, B., R. H. Silverman, et al. (1989). "Tat trans-activates the human immunodeficiency virus through a nascent RNA target." Cell **59**(2): 273-82.

Berthoux, L., S. Sebastian, et al. (2007). "The role of lysine 186 in HIV-1 integrase multimerization." Virology **364**(1): 227-36.

Bischerour, J., H. Leh, et al. (2003). "Disulfide-linked integrase oligomers involving C280 residues are formed in vitro and in vivo but are not essential for human immunodeficiency virus replication." J Virol **77**(1): 135-41.

Bischerour, J., P. Tauc, et al. (2003). "The (52-96) C-terminal domain of Vpr stimulates HIV-1 IN-mediated homologous strand transfer of mini-viral DNA." Nucleic Acids Res **31**(10): 2694-702.

Bonnenfant, S., C. M. Thomas, et al. (2004). "Styrylquinolines, integrase inhibitors acting prior to integration: a new mechanism of action for anti-integrase agents." J Virol **78**(11): 5728-36.

Boshoff, C. and R. Weiss (2002). "AIDS-related malignancies." Nat Rev Cancer **2**(5): 373-82.

Brigo, A., K. W. Lee, et al. (2005). "Comparative molecular dynamics simulations of HIV-1 integrase and the T66I/M154I mutant: binding modes and drug resistance to a diketo acid inhibitor." Proteins **59**(4): 723-41.

Brin, E. and J. Leis (2002). "Changes in the mechanism of DNA integration in vitro induced by base substitutions in the HIV-1 U5 and U3 terminal sequences." J Biol Chem **277**(13): 10938-48.

Brin, E. and J. Leis (2002). "HIV-1 integrase interaction with U3 and U5 terminal sequences in vitro defined using substrates with random sequences." J Biol Chem **277**(21): 18357-64.

Brin, E., J. Yi, et al. (2000). "Modeling the late steps in HIV-1 retroviral integrase-catalyzed DNA integration." J Biol Chem **275**(50): 39287-95.

Brown, P. O., B. Bowerman, et al. (1987). "Correct integration of retroviral DNA in vitro." Cell **49**(3): 347-56.

Brown, P. O., B. Bowerman, et al. (1989). "Retroviral integration: structure of the initial covalent product and its precursor, and a role for the viral IN protein." Proc Natl Acad Sci U S A **86**(8): 2525-9.

Burt, T. D., B. K. Agan, et al. (2008). "Apolipoprotein (apo) E4 enhances HIV-1 cell entry in vitro, and the APOE epsilon4/epsilon4 genotype accelerates HIV disease progression." Proc Natl Acad Sci U S A **105**(25): 8718-23.

Busschots, K., J. Vercammen, et al. (2005). "The interaction of LEDGF/p75 with integrase is lentivirus-specific and promotes DNA binding." J Biol Chem **280**(18): 17841-7.

Cai, M., Y. Huang, et al. (1998). "Solution structure of the cellular factor BAF responsible for protecting retroviral DNA from autointegration." Nat Struct Biol **5**(10): 903-9.

Cai, M., R. Zheng, et al. (1997). "Solution structure of the N-terminal zinc binding domain of HIV-1 integrase." Nat Struct Biol **4**(7): 567-77.

Carteau, S., R. J. Gorelick, et al. (1999). "Coupled integration of human immunodeficiency virus type 1 cDNA ends by purified integrase in vitro: stimulation by the viral nucleocapsid protein." J Virol **73**(8): 6670-9.

CDC (1981). "Pneumocystis pneumonia --- Los Angeles. ." MMWR(30): 250-2.

Cereseto, A., L. Manganaro, et al. (2005). "Acetylation of HIV-1 integrase by p300 regulates viral integration." Embo J **24**(17): 3070-81.

Chang, D. D. and P. A. Sharp (1989). "Regulation by HIV Rev depends upon recognition of splice sites." Cell **59**(5): 789-95.

Chen, A., I. T. Weber, et al. (2006). "Identification of amino acids in HIV-1 and avian sarcoma virus integrase subsites required for specific recognition of the long terminal repeat Ends." J Biol Chem **281**(7): 4173-82.

Chen, H. and A. Engelman (1998). "The barrier-to-autointegration protein is a host factor for HIV type 1 integration." Proc Natl Acad Sci U S A **95**(26): 15270-4.

Cherepanov, P., A. L. Ambrosio, et al. (2005). "Structural basis for the recognition between HIV-1 integrase and transcriptional coactivator p75." Proc Natl Acad Sci U S A **102**(48): 17308-13.

Cherepanov, P., J. A. Este, et al. (1997). "Mode of interaction of G-quartets with the integrase of human immunodeficiency virus type 1." Mol Pharmacol **52**(5): 771-80.

Cherepanov, P., G. Maertens, et al. (2003). "HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells." J Biol Chem **278**(1): 372-81.

Cherepanov, P., Z. Y. Sun, et al. (2005). "Solution structure of the HIV-1 integrase-binding domain in LEDGF/p75." Nat Struct Mol Biol **12**(6): 526-32.

Chiu, T. K. and D. R. Davies (2004). "Structure and function of HIV-1 integrase." Curr Top Med Chem **4**(9): 965-77.

Ciuffi, A. and F. D. Bushman (2006). "Retroviral DNA integration: HIV and the role of LEDGF/p75." Trends Genet **22**(7): 388-95.

Ciuffi, A., M. Llano, et al. (2005). "A role for LEDGF/p75 in targeting HIV DNA integration." Nat Med **11**(12): 1287-9.

Collins, K. L., B. K. Chen, et al. (1998). "HIV-1 Nef protein protects infected primary cells against killing by cytotoxic T lymphocytes." Nature **391**(6665): 397-401.

Cooper, D. A., R. T. Steigbigel, et al. (2008). "Subgroup and resistance analyses of raltegravir for resistant HIV-1 infection." N Engl J Med **359**(4): 355-65.

Craigie, R. (2001). "HIV integrase, a brief overview from chemistry to therapeutics." J Biol Chem **276**(26): 23213-6.

Craigie, R., T. Fujiwara, et al. (1990). "The IN protein of Moloney murine leukemia virus processes the viral DNA ends and accomplishes their integration in vitro." Cell **62**(4): 829-37.

Cujec, T. P., H. Cho, et al. (1997). "The human immunodeficiency virus transactivator Tat interacts with the RNA polymerase II holoenzyme." Mol Cell Biol **17**(4): 1817-23.

Cujec, T. P., H. Okamoto, et al. (1997). "The HIV transactivator TAT binds to the CDK-activating kinase and activates the phosphorylation of the carboxy-terminal domain of RNA polymerase II." Genes Dev **11**(20): 2645-57.

Dayam, R., T. Sanchez, et al. (2005). "Diketo acid pharmacophore. 2. Discovery of structurally diverse inhibitors of HIV-1 integrase." J Med Chem **48**(25): 8009-15.

Dayton, A. I., J. G. Sodroski, et al. (1986). "The trans-activator gene of the human T cell lymphotropic virus type III is required for replication." Cell **44**(6): 941-7.

De Lano, W. L. (2002). The PyMol Molecular Graphics System.

DeJesus, E., D. Berger, et al. (2006). "Antiviral activity, pharmacokinetics, and dose response of the HIV-1 integrase inhibitor GS-9137 (JTK-303) in treatment-naive and treatment-experienced patients." J Acquir Immune Defic Syndr **43**(1): 1-5.

Deprez, E., S. Barbe, et al. (2004). "Mechanism of HIV-1 integrase inhibition by styrylquinoline derivatives in vitro." Mol Pharmacol **65**(1): 85-98.

Dirac, A. M. and J. Kjems (2001). "Mapping DNA-binding sites of HIV-1 integrase by protein footprinting." Eur J Biochem **268**(3): 743-51.

Drake, R. R., N. Neamati, et al. (1998). "Identification of a nucleotide binding site in HIV-1 integrase." Proc Natl Acad Sci U S A **95**(8): 4170-5.

Embrey, M. W., J. S. Wai, et al. (2005). "A series of 5-(5,6)-dihydrouracil substituted 8-hydroxy-[1,6]naphthyridine-7-carboxylic acid 4-fluorobenzylamide inhibitors of HIV-1 integrase and viral replication in cells." Bioorg Med Chem Lett **15**(20): 4550-4.

Emerman, M., R. Vazeux, et al. (1989). "The rev gene product of the human immunodeficiency virus affects envelope-specific RNA localization." Cell **57**(7): 1155-65.

Emiliani, S., A. Mousnier, et al. (2005). "Integrase mutants defective for interaction with LEDGF/p75 are impaired in chromosome tethering and HIV-1 replication." J Biol Chem **280**(27): 25517-23.

Engelman, A., G. Englund, et al. (1995). "Multiple effects of mutations in human immunodeficiency virus type 1 integrase on viral replication." J Virol **69**(5): 2729-36.

Engelman, A., Y. Liu, et al. (1997). "Structure-based mutagenesis of the catalytic domain of human immunodeficiency virus type 1 integrase." J Virol **71**(5): 3507-14.

Espeseth, A. S., P. Felock, et al. (2000). "HIV-1 integrase inhibitors that compete with the target DNA substrate define a unique strand transfer conformation for integrase." Proc Natl Acad Sci U S A **97**(21): 11244-9.

Esposito, D. and R. Craigie (1998). "Sequence specificity of viral end DNA binding by HIV-1 integrase reveals critical regions for protein-DNA interaction." Embo J **17**(19): 5832-43.

Fan, L. and K. Peden (1992). "Cell-free transmission of Vif mutants of HIV-1." Virology **190**(1): 19-29.

Farnet, C. M. and F. D. Bushman (1997). "HIV-1 cDNA integration: requirement of HMG I(Y) protein for function of preintegration complexes in vitro." Cell **88**(4): 483-92.

Farnet, C. M. and W. A. Haseltine (1990). "Integration of human immunodeficiency virus type 1 DNA in vitro." Proc Natl Acad Sci U S A **87**(11): 4164-8.

Farnet, C. M. and W. A. Haseltine (1991). "Circularization of human immunodeficiency virus type 1 DNA in vitro." J Virol **65**(12): 6942-52.

Faure, A., C. Calmels, et al. (2005). "HIV-1 integrase crosslinked oligomers are active in vitro." Nucleic Acids Res **33**(3): 977-86.

Felber, B. K., M. Hadzopoulou-Cladaras, et al. (1989). "rev protein of human immunodeficiency virus type 1 affects the stability and transport of the viral mRNA." Proc Natl Acad Sci U S A **86**(5): 1495-9.

Fesen, M. R., Y. Pommier, et al. (1994). "Inhibition of HIV-1 integrase by flavones, caffeic acid phenethyl ester (CAPE) and related compounds." Biochem Pharmacol **48**(3): 595-608.

Fikkert, V., A. Hombrouck, et al. (2004). "Multiple mutations in human immunodeficiency virus-1 integrase confer resistance to the clinical trial drug S-1360." Aids **18**(15): 2019-28.

Fikkert, V., B. Van Maele, et al. (2003). "Development of resistance against diketo derivatives of human immunodeficiency virus type 1 by progressive accumulation of integrase mutations." J Virol **77**(21): 11459-70.

Fisher, A. G., B. Ensoli, et al. (1987). "The sor gene of HIV-1 is required for efficient virus transmission in vitro." Science **237**(4817): 888-93.

Fisher, A. G., M. B. Feinberg, et al. (1986). "The trans-activator gene of HTLV-III is essential for virus replication." Nature **320**(6060): 367-71.

Fitzgerald, M. L., A. C. Vora, et al. (1992). "Concerted integration of viral DNA termini by purified avian myeloblastosis virus integrase." J Virol **66**(11): 6257-63.

Fujiwara, T. and K. Mizuuchi (1988). "Retroviral DNA integration: structure of an integration intermediate." Cell **54**(4): 497-504.

Furukawa, K. (1999). "LAP2 binding protein 1 (L2BP1/BAF) is a candidate mediator of LAP2-chromatin interaction." J Cell Sci **112 (Pt 15)**: 2485-92.

Gabuzda, D. H., K. Lawrence, et al. (1992). "Role of vif in replication of human immunodeficiency virus type 1 in CD4+ T lymphocytes." J Virol **66**(11): 6489-95.

Gallo, R. C., S. Z. Salahuddin, et al. (1984). "Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS." Science **224**(4648): 500-3.

Gao, K., S. L. Butler, et al. (2001). "Human immunodeficiency virus type 1 integrase: arrangement of protein domains in active cDNA complexes." Embo J **20**(13): 3565-76.

Garber, M. E., P. Wei, et al. (1998). "HIV-1 Tat interacts with cyclin T1 to direct the P-TEFb CTD kinase complex to TAR RNA." Cold Spring Harb Symp Quant Biol **63**: 371-80.

Garcia, J. V. and A. D. Miller (1991). "Serine phosphorylation-independent downregulation of cell-surface CD4 by nef." Nature **350**(6318): 508-11.

Gerton, J. L., S. Ohgi, et al. (1998). "Effects of mutations in residues near the active site of human immunodeficiency virus type 1 integrase on specific enzyme-substrate interactions." J Virol **72**(6): 5046-55.

Goldgur, Y., R. Craigie, et al. (1999). "Structure of the HIV-1 integrase catalytic domain complexed with an inhibitor: a platform for antiviral drug design." Proc Natl Acad Sci U S A **96**(23): 13040-3.

Grobler, J. A., K. Stillmock, et al. (2002). "Diketo acid inhibitor mechanism and HIV-1 integrase: implications for metal binding in the active site of phosphotransferase enzymes." Proc Natl Acad Sci U S A **99**(10): 6661-6.

Guare, J. P., J. S. Wai, et al. (2006). "A series of 5-aminosubstituted 4-fluorobenzyl-8-hydroxy-[1,6]naphthyridine-7-carboxamide HIV-1 integrase inhibitors." Bioorg Med Chem Lett **16**(11): 2900-4.

Guy, B., M. P. Kieny, et al. (1987). "HIV F/3' orf encodes a phosphorylated GTP-binding protein resembling an oncogene product." Nature **330**(6145): 266-9.

Harper, A. L., L. M. Skinner, et al. (2001). "Use of patient-derived human immunodeficiency virus type 1 integrases to identify a protein residue that affects target site selection." J Virol **75**(16): 7756-62.

Harper, A. L., M. Sudol, et al. (2003). "An amino acid in the central catalytic domain of three retroviral integrases that affects target site selection in nonviral DNA." J Virol **77**(6): 3838-45.

Harrison, R. W. (1993). Stiffness and energy conservation in molecular dynamics: An improved integrator. **14:** 1112-1122.

Hazuda, D. J., N. J. Anthony, et al. (2004). "A naphthyridine carboxamide provides evidence for discordant resistance between mechanistically identical inhibitors of HIV-1 integrase." Proc Natl Acad Sci U S A **101**(31): 11233-8.

Hazuda, D. J., P. Felock, et al. (2000). "Inhibitors of strand transfer that prevent integration and inhibit HIV-1 replication in cells." Science **287**(5453): 646-50.

Hazuda, D. J., S. D. Young, et al. (2004). "Integrase inhibitors and cellular immunity suppress retroviral replication in rhesus macaques." Science **305**(5683): 528-32.

He, J., S. Choe, et al. (1995). "Human immunodeficiency virus type 1 viral protein R (Vpr) arrests cells in the G2 phase of the cell cycle by inhibiting p34cdc2 activity." J Virol **69**(11): 6705-11.

Heuer, T. S. and P. O. Brown (1997). "Mapping features of HIV-1 integrase near selected sites on viral and target DNA molecules in an active enzyme-DNA complex by photo-cross-linking." Biochemistry **36**(35): 10655-65.

Heuer, T. S. and P. O. Brown (1998). "Photo-cross-linking studies suggest a model for the architecture of an active human immunodeficiency virus type 1 integrase-DNA complex." Biochemistry **37**(19): 6667-78.

Hindmarsh, P. and J. Leis (1999). "Retroviral DNA integration." Microbiol Mol Biol Rev **63**(4): 836-43, table of contents.

Hindmarsh, P., T. Ridky, et al. (1999). "HMG protein family members stimulate human immunodeficiency virus type 1 and avian sarcoma virus concerted DNA integration in vitro." J Virol **73**(4): 2994-3003.

Jeltsch, A., C. Wenz, et al. (1996). "Engineering novel restriction endonucleases: principles and applications." Trends Biotechnol **14**(7): 235-8.

Jenkins, T. M., A. Engelman, et al. (1996). "A soluble active mutant of HIV-1 integrase: involvement of both the core and carboxyl-terminal domains in multimerization." J Biol Chem **271**(13): 7712-8.

Jenkins, T. M., D. Esposito, et al. (1997). "Critical contacts between HIV-1 integrase and viral DNA identified by structure-based analysis and photo-crosslinking." Embo J **16**(22): 6849-59.

John, S., T. M. Fletcher, 3rd, et al. (2005). "Development and application of a high-throughput screening assay for HIV-1 integrase enzyme activities." J Biomol Screen **10**(6): 606-14.

Johnson, A. A., W. Santos, et al. (2006). "Integration requires a specific interaction of the donor DNA terminal 5'-cytosine with glutamine 148 of the HIV-1 integrase flexible loop." J Biol Chem **281**(1): 461-7.

Kalpana, G. V., S. Marmon, et al. (1994). "Binding and stimulation of HIV-1 integrase by a human homolog of yeast transcription factor SNF5." Science **266**(5193): 2002-6.

Kalpana, G. V., A. Reicin, et al. (1999). "Isolation and characterization of an oligomerization-negative mutant of HIV-1 integrase." Virology **259**(2): 274-85.

Karki, R. G., Y. Tang, et al. (2004). "Model of full-length HIV-1 integrase complexed with viral DNA as template for anti-HIV drug design." J Comput Aided Mol Des **18**(12): 739-60.

Karni, O., A. Friedler, et al. (1998). "A peptide derived from the N-terminal region of HIV-1 Vpr promotes nuclear import in permeabilized cells: elucidation of the NLS region of the Vpr." FEBS Lett **429**(3): 421-5.

Katz, R. A., G. Merkel, et al. (1990). "The avian retroviral IN protein is both necessary and sufficient for integrative recombination in vitro." Cell **63**(1): 87-95.

Katzman, M., R. A. Katz, et al. (1989). "The avian retroviral integration protein cleaves the terminal sequences of linear viral DNA at the in vivo sites of integration." J Virol **63**(12): 5319-27.

Konsavage, W. M., Jr., S. Burkholder, et al. (2005). "A substitution in rous sarcoma virus integrase that separates its two biologically relevant enzymatic activities." J Virol **79**(8): 4691-9.

Kukolj, G. and A. M. Skalka (1995). "Enhanced and coordinated processing of synapsed viral DNA ends by retroviral integrases in vitro." Genes Dev **9**(20): 2556-67.

Kulkosky, J. and A. M. Skalka (1994). "Molecular mechanism of retroviral DNA integration." Pharmacol Ther **61**(1-2): 185-203.

Lee, D. J. and W. E. Robinson, Jr. (2004). "Human immunodeficiency virus type 1 (HIV-1) integrase: resistance to diketo acid integrase inhibitors impairs HIV-1 replication and integration and confers cross-resistance to L-chicoric acid." J Virol **78**(11): 5835-47.

Lee, M. S. and R. Craigie (1994). "Protection of retroviral DNA from autointegration: involvement of a cellular factor." Proc Natl Acad Sci U S A **91**(21): 9823-7.

Lee, M. S. and R. Craigie (1998). "A previously unidentified host protein protects retroviral DNA from autointegration." Proc Natl Acad Sci U S A **95**(4): 1528-33.

Lee, Y. M. and J. M. Coffin (1990). "Efficient autointegration of avian retrovirus DNA in vitro." J Virol **64**(12): 5958-65.

Leis, J., D. Baltimore, et al. (1988). "Standardized and simplified nomenclature for proteins common to all retroviruses." J Virol **62**(5): 1808-9.

Levy-Mintz, P., L. Duan, et al. (1996). "Intracellular expression of single-chain variable fragments to inhibit early stages of the viral life cycle by targeting human immunodeficiency virus type 1 integrase." J Virol **70**(12): 8821-32.

Li, L., K. Yoder, et al. (2000). "Retroviral cDNA integration: stimulation by HMG I family proteins." J Virol **74**(23): 10965-74.

Li, M. and R. Craigie (2005). "Processing of viral DNA ends channels the HIV-1 integration reaction to concerted integration." J Biol Chem **280**(32): 29334-9.

Li, M., M. Mizuuchi, et al. (2006). "Retroviral DNA integration: reaction pathway and critical intermediates." Embo J **25**(6): 1295-304.

Llano, M., D. T. Saenz, et al. (2006). "An essential role for LEDGF/p75 in HIV integration." Science **314**(5798): 461-4.

Llano, M., M. Vanegas, et al. (2006). "Identification and characterization of the chromatin-binding domains of the HIV-1 integrase interactor LEDGF/p75." J Mol Biol **360**(4): 760-73.

Lu, R., A. Limon, et al. (2004). "Class II integrase mutants with changes in putative nuclear localization signals are primarily blocked at a postnuclear entry step of human immunodeficiency virus type 1 replication." J Virol **78**(23): 12735-46.

Lu, R., A. Limon, et al. (2005). "Genetic analyses of DNA-binding mutants in the catalytic core domain of human immunodeficiency virus type 1 integrase." J Virol **79**(4): 2493-505.

Lubkowski, J., F. Yang, et al. (1998). "Structure of the catalytic domain of avian sarcoma virus integrase with a bound HIV-1 integrase-targeted inhibitor." Proc Natl Acad Sci U S A **95**(9): 4831-6.

Luft, B. J. and A. Chua (2000). "Central Nervous System Toxoplasmosis in HIV Pathogenesis, Diagnosis, and Therapy." Curr Infect Dis Rep **2**(4): 358-362.

Lutzke, R. A. and R. H. Plasterk (1998). "Structure-based mutational analysis of the C-terminal DNA-binding domain of human immunodeficiency virus type 1 integrase: critical residues for protein oligomerization and DNA binding." J Virol **72**(6): 4841-8.

Lutzke, R. A., C. Vink, et al. (1994). "Characterization of the minimal DNA-binding domain of the HIV integrase protein." Nucleic Acids Res **22**(20): 4125-31.

Maertens, G., P. Cherepanov, et al. (2004). "Identification and characterization of a functional nuclear localization signal in the HIV-1 integrase interactor LEDGF/p75." J Biol Chem **279**(32): 33421-9.

Maertens, G., P. Cherepanov, et al. (2003). "LEDGF/p75 is essential for nuclear and chromosomal targeting of HIV-1 integrase in human cells." J Biol Chem **278**(35): 33528-39.

Marchand, C., A. A. Johnson, et al. (2003). "Metal-dependent inhibition of HIV-1 integrase by beta-diketo acids and resistance of the soluble double-mutant (F185K/C280S)." Mol Pharmacol **64**(3): 600-9.

Marchand, C., X. Zhang, et al. (2002). "Structural determinants for HIV-1 integrase inhibition by beta-diketo acids." J Biol Chem **277**(15): 12596-603.

Margottin, F., S. P. Bour, et al. (1998). "A novel human WD protein, h-beta TrCp, that interacts with HIV-1 Vpu connects CD4 to the ER degradation pathway through an F-box motif." Mol Cell **1**(4): 565-74.

Markowitz, M., J. O. Morales-Ramirez, et al. (2006). "Antiretroviral activity, pharmacokinetics, and tolerability of MK-0518, a novel inhibitor of HIV-1 integrase, dosed as monotherapy for 10 days in treatment-naive HIV-1-infected individuals." J Acquir Immune Defic Syndr **43**(5): 509-15.

Maroun, M., O. Delelis, et al. (2006). "Inhibition of early steps of HIV-1 replication by SNF5/Ini1." J Biol Chem **281**(32): 22736-43.

Maroun, R. G., L. Zargarian, et al. (2005). "A structural study of model peptides derived from HIV-1 integrase central domain." Rapid Commun Mass Spectrom **19**(18): 2539-48.

Masuda, T., M. J. Kuroda, et al. (1998). "Specific and independent recognition of U3 and U5 att sites by human immunodeficiency virus type 1 integrase in vivo." J Virol **72**(10): 8396-402.

McCord, M., R. Chiu, et al. (1999). "Retrovirus DNA termini bound by integrase communicate in trans for full-site integration in vitro." Virology **259**(2): 392-401.

Meanwell, N. A., M. Belema, et al. (2005). "Developments in antiviral drug design, discovery and development in 2004." Curr Drug Targets Infect Disord **5**(4): 307-400.

Michaels, F. H., N. Hattori, et al. (1993). "The human immunodeficiency virus type 1 (HIV-1) vif protein is located in the cytoplasm of infected cells and its effect on viral replication is equivalent in HIV-2." AIDS Res Hum Retroviruses **9**(10): 1025-30.

Morozov, A., E. Yung, et al. (1998). "Structure-function analysis of integrase interactor 1/hSNF5L1 reveals differential properties of two repeat motifs present in the highly conserved region." Proc Natl Acad Sci U S A **95**(3): 1120-5.

Mousnier, A., H. Leh, et al. (2004). "Nuclear import of HIV-1 integrase is inhibited in vitro by styrylquinoline derivatives." Mol Pharmacol **66**(4): 783-8.

Mulder, L. C. and M. A. Muesing (2000). "Degradation of HIV-1 integrase by the N-end rule pathway." J Biol Chem **275**(38): 29749-53.

Murphy, J. E. and S. P. Goff (1992). "A mutation at one end of Moloney murine leukemia virus DNA blocks cleavage of both ends by the viral integrase in vivo." J Virol **66**(8): 5092-5.

Neamati, N., H. Hong, et al. (1998). "Salicylhydrazine-containing inhibitors of HIV-1 integrase: implication for a selective chelation in the integrase active site." J Med Chem **41**(17): 3202-9.

Neamati, N., H. Hong, et al. (1997). "Potent inhibitors of human immunodeficiency virus type 1 integrase: identification of a novel four-point pharmacophore and tetracyclines as novel inhibitors." Mol Pharmacol **52**(6): 1041-55.

Parada, C. A. and R. G. Roeder (1996). "Enhanced processivity of RNA polymerase II triggered by Tat-induced phosphorylation of its carboxy-terminal domain." Nature **384**(6607): 375-8.

Pauza, C. D. (1990). "Two bases are deleted from the termini of HIV-1 linear DNA during integrative recombination." Virology **179**(2): 886-9.

Petit, C., O. Schwartz, et al. (1999). "Oligomerization within virions and subcellular localization of human immunodeficiency virus type 1 integrase." J Virol **73**(6): 5079-88.

Pluymers, W., G. Pais, et al. (2002). "Inhibition of human immunodeficiency virus type 1 integration by diketo derivatives." Antimicrob Agents Chemother **46**(10): 3292-7.

Podtelezhnikov, A. A., K. Gao, et al. (2003). "Modeling HIV-1 integrase complexes based on their hydrodynamic properties." Biopolymers **68**(1): 110-20.

Polard, P. and M. Chandler (1995). "Bacterial transposases and retroviral integrases." Mol Microbiol **15**(1): 13-23.

Pommier, Y., A. A. Johnson, et al. (2005). "Integrase inhibitors to treat HIV/AIDS." Nat Rev Drug Discov **4**(3): 236-48.

Popov, S., M. Rexach, et al. (1998). "Viral protein R regulates docking of the HIV-1 preintegration complex to the nuclear pore complex." J Biol Chem **273**(21): 13347-52.

Popovic, M., M. G. Sarngadharan, et al. (1984). "Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS." Science **224**(4648): 497-500.

Priet, S., J. M. Navarro, et al. (2003). "Reversion of the lethal phenotype of an HIV-1 integrase mutant virus by overexpression of the same integrase mutant protein." J Biol Chem **278**(23): 20724-30.

Puglia, J., T. Wang, et al. (2006). "Revealing domain structure through linker-scanning analysis of the murine leukemia virus (MuLV) RNase H and MuLV and human immunodeficiency virus type 1 integrase proteins." J Virol **80**(19): 9497-510.

Raghavendra, N. K. and A. Engelman (2007). "LEDGF/p75 interferes with the formation of synaptic nucleoprotein complexes that catalyze full-site HIV-1 DNA integration in vitro: Implications for the mechanism of viral cDNA integration." Virology **360**(1): 1-5.

Rahman, S., R. Lu, et al. (2007). "Structure-based mutagenesis of the integrase-LEDGF/p75 interface uncouples a strict correlation between in vitro protein binding and HIV-1 fitness." Virology **357**(1): 79-90.

Re, F., D. Braaten, et al. (1995). "Human immunodeficiency virus type 1 Vpr arrests the cell cycle in G2 by inhibiting the activation of p34cdc2-cyclin B." J Virol **69**(11): 6859-64.

Roth, M. J., P. L. Schwartzberg, et al. (1989). "Structure of the termini of DNA intermediates in the integration of retroviral DNA: dependence on IN function and terminal DNA sequence." Cell **58**(1): 47-54.

Sadler, M. and M. R. Nelson (1997). "Progressive multifocal leukoencephalopathy in HIV." Int J STD AIDS **8**(6): 351-7.

Sato, M., T. Motomura, et al. (2006). "Novel HIV-1 integrase inhibitors derived from quinolone antibiotics." J Med Chem **49**(5): 1506-8.

Sayasith, K., G. Sauve, et al. (2000). "Characterization of mutant HIV-1 integrase carrying amino acid changes in the catalytic domain." Mol Cells **10**(5): 525-32.

Schroder, A. R., P. Shinn, et al. (2002). "HIV-1 integration in the human genome favors active genes and local hotspots." Cell **110**(4): 521-9.

Sinha, S. and D. P. Grandgenett (2005). "Recombinant human immunodeficiency virus type 1 integrase exhibits a capacity for full-site integration in vitro that is comparable to that of purified preintegration complexes from virus-infected cells." J Virol **79**(13): 8208-16.

Snasel, J., Z. Krejcik, et al. (2005). "Integrase of Mason-Pfizer monkey virus." Febs J **272**(1): 203-16.

Steigbigel, R. T., D. A. Cooper, et al. (2008). "Raltegravir with optimized background therapy for resistant HIV-1 infection." N Engl J Med **359**(4): 339-54.

Strebel, K., D. Daugherty, et al. (1987). "The HIV 'A' (sor) gene product is essential for virus infectivity." Nature **328**(6132): 728-30.

Strebel, K., T. Klimkait, et al. (1988). "A novel gene of HIV-1, vpu, and its 16-kilodalton product." Science **241**(4870): 1221-3.

Svarovskaia, E. S., R. Barr, et al. (2004). "Azido-containing diketo acid derivatives inhibit human immunodeficiency virus type 1 integrase in vivo and influence the frequency of deletions at two-long-terminal-repeat-circle junctions." J Virol **78**(7): 3210-22.

Turelli, P., V. Doucas, et al. (2001). "Cytoplasmic recruitment of INI1 and PML on incoming HIV preintegration complexes: interference with early steps of viral replication." Mol Cell **7**(6): 1245-54.

van den Ent, F. M., A. Vos, et al. (1999). "Dissecting the role of the N-terminal domain of human immunodeficiency virus integrase by trans-complementation analysis." J Virol **73**(4): 3176-83.

Vanegas, M., M. Llano, et al. (2005). "Identification of the LEDGF/p75 HIV-1 integrase-interaction domain and NLS reveals NLS-independent chromatin tethering." J Cell Sci **118**(Pt 8): 1733-43.

Vincent, K. A., V. Ellison, et al. (1993). "Characterization of human immunodeficiency virus type 1 integrase expressed in Escherichia coli and analysis of variants with amino-terminal mutations." J Virol **67**(1): 425-37.

von Schwedler, U., J. Song, et al. (1993). "Vif is crucial for human immunodeficiency virus type 1 proviral DNA synthesis in infected cells." J Virol **67**(8): 4945-55.

Vora, A. and D. P. Grandgenett (2001). "DNase protection analysis of retrovirus integrase at the viral DNA ends for full-site integration in vitro." J Virol **75**(8): 3556-67.

Vora, A. C., R. Chiu, et al. (1997). "Avian retrovirus U3 and U5 DNA inverted repeats. Role Of nonsymmetrical nucleotides in promoting full-site integration by purified virion and bacterial recombinant integrases." J Biol Chem **272**(38): 23938-45.

Vora, A. C., M. McCord, et al. (1994). "Efficient concerted integration of retrovirus-like DNA in vitro by avian myeloblastosis virus integrase." Nucleic Acids Res **22**(21): 4454-61.

Wang, J. Y., H. Ling, et al. (2001). "Structure of a two-domain fragment of HIV-1 integrase: implications for domain organization in the intact protein." Embo J **20**(24): 7333-43.

Wang, W., J. Cote, et al. (1996). "Purification and biochemical heterogeneity of the mammalian SWI-SNF complex." Embo J **15**(19): 5370-82.

Weber, I. T. and R. W. Harrison (1996). "Molecular mechanics calculations on HIV-1 protease with peptide substrates correlate with experimental data." Protein Eng **9**(8): 679-90.

Weber, I. T. and R. W. Harrison (1997). "Molecular mechanics calculations on Rous sarcoma virus protease with peptide substrates." Protein Sci **6**(11): 2365-74.

Wei, P., M. E. Garber, et al. (1998). "A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA." Cell **92**(4): 451-62.

Whitcomb, J. M. and S. H. Hughes (1991). "The sequence of human immunodeficiency virus type 2 circle junction suggests that integration protein cleaves the ends of linear DNA asymmetrically." J Virol **65**(7): 3906-10.

Wielens, J., I. T. Crosby, et al. (2005). "A three-dimensional model of the human immunodeficiency virus type 1 integration complex." J Comput Aided Mol Des **19**(5): 301-17.

Winkler, F. K., D. W. Banner, et al. (1993). "The crystal structure of EcoRV endonuclease and of its complexes with cognate and non-cognate DNA fragments." Embo J **12**(5): 1781-95.

Yang, Z., J. R. Horton, et al. (2005). "Structure of HinP1I endonuclease reveals a striking similarity to the monomeric restriction enzyme MspI." Nucleic Acids Res **33**(6): 1892-901.

Yang, Z. N., T. C. Mueser, et al. (2000). "Crystal structure of an active two-domain derivative of Rous sarcoma virus integrase." J Mol Biol **296**(2): 535-48.

Yi, J., E. Asante-Appiah, et al. (1999). "Divalent cations stimulate preferential recognition of a viral DNA end by HIV-1 integrase." Biochemistry **38**(26): 8458-68.

Yu, F., G. S. Jones, et al. (2007). "HIV-1 Integrase Preassembled on Donor DNA Is Refractory to Activity Stimulation by LEDGF/p75." Biochemistry.

Yung, E., M. Sorin, et al. (2001). "Inhibition of HIV-1 virion production by a transdominant mutant of integrase interactor 1." Nat Med **7**(8): 920-6.

Zhao, H., N. Neamati, et al. (1997). "Hydrazide-containing inhibitors of HIV-1 integrase." J Med Chem **40**(6): 937-41.

Zheng, R., R. Ghirlando, et al. (2000). "Barrier-to-autointegration factor (BAF) bridges DNA in a discrete, higher-order nucleoprotein complex." Proc Natl Acad Sci U S A **97**(16): 8997-9002.

Zhu, K., C. Dobard, et al. (2004). "Requirement for integrase during reverse transcription of human immunodeficiency virus type 1 and the effect of cysteine mutations of integrase on its interactions with reverse transcriptase." J Virol **78**(10): 5045-55.

Zhuang, L., J. S. Wai, et al. (2003). "Design and synthesis of 8-hydroxy-[1,6]naphthyridines as novel inhibitors of HIV-1 integrase in vitro and in infected cells." J Med Chem **46**(4): 453-6.

Zouhiri, F., J. F. Mouscadet, et al. (2000). "Structure-activity relationships and binding mode of styrylquinolines as potent inhibitors of HIV-1 integrase and replication of HIV-1 in cell culture." J Med Chem **43**(8): 1533-40.

James R Dolan

332 E 71ST ST APT 3B NEW YORK, NY 10021
JamesDolan2008@u.northwestern.edu

EDUCATION

2002-Present	Northwestern University Feinberg School of Medicine, Chicago, IL Graduate GPA: 3.25 Thesis work to be completed May 2008
2002	GRE Score: 2060 Analytical: 760 Quantitative: 790 Verbal: 510
1998-2002	University of Illinois-Urbana Champaign, Urbana, IL Undergraduate GPA: 3.38

RESEARCH EXPERIENCE

2002-2008 **Northwestern University Feinberg School of Medicine.** Chicago, IL.
Dissertation research, laboratory of Jonathan Leis, PhD.
Examined various aspects of retroviral replication important for future therapeutic development against HIV infection.
Investigated the cellular machinery required for budding of newly formed virions. Currently working on identification of key amino acid residues of the HIV-1 Integrase protein that are involved in forming contacts with viral and target DNA for integration.

2000-2002 **University of Illinois Department of Microbiology**. Urbana, IL.
Undergraduate research, laboratory of James Slauch, PhD.
Investigated a divalent cation transporter in *Salmonella typhimurium* required for pathogenesis.

PUBLICATIONS

Rous Sarcoma Virus: The Late Stages of Replication. James Dolan and Jonathan Leis. **Recent Advances in RNA Virus Replication**, 2006. (Book Chapter)

Retroviral Integrase. James Dolan and Jonathan Leis. **Retrovirus Replication,** 2007. (Book Chapter)

TEACHING EXPERIENCE

09/2004 – 12/2004 **Teaching Assistant,** Graduate Molecular Biology and Genetics. Led research paper discussions and instructed first year graduate students in a number of current molecular biology techniques.

LEADERSHIP

2006	Graduate Leadership Council, Northwestern University
2005-Present	Senior Student Representative, Integrated Graduate Program Executive Committee, Northwestern University
2005-2006	Co-President, Northwestern University Chicago Graduate Student Association
2004-2005	Student Representative, Integrated Graduate Program Executive Committee, Northwestern University
2004-2005	Vice-President of Social Relations, Northwestern University Chicago Graduate Student Association
2003-2004	Social Committee, Northwestern University Chicago Graduate Student Association

HONORS AND AWARDS

2004-2006	**Viral Replication Training Grant.** Department of Microbiology and Immunology, Northwestern University Feinberg School of Medicine
2002	**Distinction in Microbiology.** Department of Microbiology, University of Illinois, Urbana-Champaign
2001	**Undergraduate Fellowship for Summer Research.** Department of Microbiology, University of Illinois, Urbana-Champaign

PRESENTATIONS

Oral

2007 "Characterizing the HIV-1 Integrase using a computer model." Microbiology-Immunology Departmental Seminar, Northwestern University Feinberg School of Medicine. Chicago, IL.

2006 "Replication and Budding of Rous Sarcoma Virus." Virology Club Seminar Series, Northwestern University Feinberg School of Medicine. Chicago, IL.

Poster

2004 "Identification and characterization of cellular proteins involved in Rous sarcoma virus budding." 16th Annual Retreat, Department of Microbiology-Immunology, Northwestern University Feinberg School of Medicine, Lake Geneva, WI.