CHARACTERIZING INTERACTIONS OF HIV-1 INTEGRASE WITH VIRAL DNA
AND THE CELLULAR COFACTOR LEDGF

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Christopher J. McKee, B.A.

Graduate Program in Molecular, Cellular & Developmental Biology

2010

Dissertation Committee:

Mamuka Kvaratskhelia, Advisor
Mary Jo Burkhard
Karin Musier-Forsyth
Thomas Schmittgen
ABSTRACT

A hallmark of retroviral replication is the permanent integration of the viral genome into the host cell genome. This integration is mediated by the viral enzyme integrase (IN) which is bound to the ends of the viral DNA in the context of a large nucleoprotein complex known as the pre-integration complex (PIC). Using the two catalytic activities of IN, the viral DNA ends are first processed then used for a strand transfer reaction that simultaneously breaks the host DNA backbone and ligates the viral genome into it. For HIV-1, the site of integration into the host genome is non-random and occurs most often in active transcription units. This bias is likely explained by the recruitment of pre-integration complexes to chromatin via interactions between IN and the cellular protein Lens Epithelium-Derived Growth Factor (LEDGF). Failure of either enzymatic reaction or failure of the PIC to engage chromatin is a replicative dead end for the virus.

Integrase has long been considered an attractive therapeutic target due to its essential role in replication and its lack of a cellular counterpart. The first-in-class HIV IN inhibitors block integration by binding specifically to the IN-viral DNA complex, displacing the viral DNA ends from the active site and effectively preventing strand transfer. The specificity of this inhibition led to the rapid emergence of HIV-1 phenotypes resistant to all available strand transfer
inhibitors and highlighted the need for new classes of IN inhibitor. The
development of these new inhibitors will be driven by an improved understanding
of IN structure and of the sequence of pre-integration events. In spite of
tremendous efforts, no high-resolution structural data is available for the full-
length HIV-1 IN. The following text describes my efforts to characterize three
critical molecular interactions that determine the fate of HIV-1 integration: IN with
viral DNA, IN with the cellular cofactor LEDGF, and LEDGF with chromatin.
Using innovative mass spectrometric footprinting techniques I have mapped and
validated biologically essential contacts at IN-viral DNA, IN-IN, and IN-LEDGF
interfaces. These experiments provided structural details that revealed
previously undescribed changes in integrase conformation and subunit dynamics
upon binding viral DNA and LEDGF, respectively. Finally, I have explored
epigenetic modifications in chromatin that are recognized by LEDGF. Specific
binding of LEDGF to regions featuring these modifications could explain, at least
in part, the targeting of active transcription units for HIV-1 integration in infected
cells.
ACKNOWLEDGMENTS

I would like to thank my advisor Mamuka Kvaratskhelia for his mentorship and the many opportunities I have had while working in his lab. Furthermore, my co-workers in the Kvaratskhelia lab have contributed enormously to the work presented herein. I would in particular like to thank Zhuojun Zhao, Jacques Kessl, Jocelyn Eidahl, and Nick Shkriabai for their technical assistance and friendship. I am indebted to our many collaborators at OSU and beyond, especially Alan Engelman of the Dana-Farber Cancer Institute and the members of the Ohio State Center for Retrovirus Research for their experimental contributions and constructive criticism of my work. My research has been funded by grants from the National Institutes of Health and the Jeffrey J. Seilhamer Cancer Foundation, to whom I am extremely grateful for their confidence and financial support. Previously copyrighted materials are reprinted here with permission of the American Society for Biochemistry and Molecular Biology.

Special thanks go to the OSU department of Molecular, Cellular, & Developmental Biology for admitting me to their outstanding program and guiding me to success as a graduate student. I also extend many thanks to those who have assisted me on my circuitous path to graduate school including John
Haddock, Yan Zhang, and Sreekumari Rajeev. Most of all I thank Jessica Dyszel for her many years of support, both technical and moral.
VITA

July, 1978..........................Born- Springfield, Illinois, USA

2001..............................................B.A. Microbiology
Southern Illinois University
Carbondale, IL, USA

2003-2005..................................Microbiologist
Ohio Department of Agriculture
Animal Disease Diagnostic Laboratory
Reynoldsburg, OH, USA

2005-present.............................Graduate Research and Teaching associate
The Ohio State University
Columbus, OH, USA
PUBLICATIONS

Research Publication


*Shared first-authorship

FIELDS OF STUDY

Major Field: Molecular, Cellular, & Developmental Biology
# TABLE OF CONTENTS

Abstract ........................................................................................................................................................................... ii  
Dedication ........................................................................................................................................................................ iv  
Acknowledgments ............................................................................................................................................................... v  
Vita .................................................................................................................................................................................... vi  
List of Tables ........................................................................................................................................................................ x  
List of Figures .......................................................................................................................................................................... xi  

Chapters:  
1. Introduction ........................................................................................................................................................................ 1  
   1.1 The HIV/AIDS challenge and integrase inhibitors to treat HIV/AIDS ..................... 1  
   1.2 The HIV-1 replication cycle ............................................................................... 2  
   1.3 Integration of HIV-1 viral DNA into host chromatin ....................................... 3  
   1.4 Structural aspects of HIV-1 integration ........................................................... 5  
   1.5 LEDGF as a cellular cofactor of HIV-1 integration ....................................... 10  
   1.6 Figures for chapter 1 .................................................................................. 13  

2. Identification of Subunit Specific Viral-DNA Contacts and a DNA-Induced Conformation Change within Functional HIV-1 Integrase Nucleoprotein Complexes ................................................................................................................................. 17  
   2.1 Introduction ........................................................................................................... 17  
   2.2 Materials and methods ....................................................................................... 19  
      2.2.1 Preparation of HIV-1 IN, DNA oligonucleotides, and cross-linked products ............................................................. 19  
      2.2.2 Mass spectrometric footprinting ................................................................. 20  
      2.2.3 CD Spectroscopy ......................................................................................... 20  
      2.2.4 Virus infectivity ......................................................................................... 21  
   2.3 Results .................................................................................................................. 21  
   2.4 Discussion ............................................................................................................. 30  
   2.5 Tables for Chapter 2 ....................................................................................... 37  
   2.6 Figures for Chapter 2 ...................................................................................... 39
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Susceptibility of IN basic residues to modification</td>
<td>37</td>
</tr>
<tr>
<td>2.2 Infectivities of IN mutant viruses</td>
<td>38</td>
</tr>
<tr>
<td>3.1 Summary of mass spectrometric footprinting results</td>
<td>71</td>
</tr>
<tr>
<td>3.2 Summary of size exclusion chromatography results for</td>
<td>72</td>
</tr>
<tr>
<td>wild type IN interactions with the LEDGF IBD</td>
<td></td>
</tr>
<tr>
<td>3.3 Summary of size exclusion chromatography results for</td>
<td>72</td>
</tr>
<tr>
<td>wild type and mutant IN proteins</td>
<td></td>
</tr>
<tr>
<td>4.1 Histone peptides selected for further study</td>
<td>96</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>HIV replication cycle</td>
<td>13</td>
</tr>
<tr>
<td>1.2</td>
<td>Enzymatic activities of HIV integrase</td>
<td>14</td>
</tr>
<tr>
<td>1.3</td>
<td>Assays of integrase enzymatic activity</td>
<td>15</td>
</tr>
<tr>
<td>1.4</td>
<td>Two-domain structures of HIV integrase</td>
<td>16</td>
</tr>
<tr>
<td>2.1</td>
<td>Experimental strategy</td>
<td>39</td>
</tr>
<tr>
<td>2.2</td>
<td>IN mutant enzymatic activity</td>
<td>40</td>
</tr>
<tr>
<td>2.3</td>
<td>IN mutant cross-linking efficiency with modified DNA</td>
<td>41</td>
</tr>
<tr>
<td>2.4</td>
<td>The effects of the modifying reagent on IN function</td>
<td>42</td>
</tr>
<tr>
<td>2.5</td>
<td>Representative segments of MALDI-ToF data</td>
<td>43</td>
</tr>
<tr>
<td>2.6</td>
<td>Representative segments of R199 IN MALDI-ToF data</td>
<td>44</td>
</tr>
<tr>
<td>2.7</td>
<td>Representative MALDI-TOF spectra showing specific protection of K14</td>
<td>45</td>
</tr>
<tr>
<td>2.8</td>
<td>Effects of amino acid substitutions on recombinant IN</td>
<td>46</td>
</tr>
<tr>
<td>2.9</td>
<td>Model of viral DNA bound to HIV IN</td>
<td>47</td>
</tr>
<tr>
<td>2.10</td>
<td>The crystal structure of the CCD-CTD fragment</td>
<td>48</td>
</tr>
<tr>
<td>3.1</td>
<td>The LEDGF proteins used in our studies</td>
<td>73</td>
</tr>
<tr>
<td>3.2</td>
<td>Effects of LEDGF and the isolated IBD on IN 3’ processing and DNA strand transfer activities</td>
<td>74</td>
</tr>
</tbody>
</table>
3.3 Effects of LEDGF and the isolated IBD on IN concerted integration activity ......................................................... 75
3.4 Subunit exchange assay using His-tag IN and SDS-PAGE ................................................................. 77
3.5 Size exclusion chromatography of free IN and IN-IBD complexes ........................................................................................................ 78
3.6 Schematic presentation of the protein footprinting strategy ........................................... 79
3.7 Representative segments of the MALDI-ToF mass spectra from the protein footprinting ........................................................................................................... 80
3.8 Size exclusion chromatography of wild type and indicated Mutants ........................................................................................................................................ 81
3.9 LEDGF binding to wild type and mutant IN proteins .......................................................... 82
3.10 Biochemical characterization of IN mutant proteins ................................................................................................................................. 83
3.11 A model for the IBD bound NTD-CCD tetramer ........................................................................ 84
3.12 IN 3’ processing activities in the presence of the diketo acid inhibitor (118-D-24) 4-[3-(azidomethyl)phenyl]-2-hydroxy-4-oxo-2-butenoic acid ................................................................................................................................. 85
3.13 The NTD-CCD dimer-dimer interface ................................................................................. 85
3.14 A potential interaction hotspot at the IN subunit-subunit interface .......................................... 87
4.1 Representative peptide screening results ....................................................................... 98
4.2 MS data from in vitro pull-down assays ........................................................................ 99
1.1. The HIV/AIDS challenge and integrase inhibitors to treat HIV/AIDS

Human immunodeficiency virus (HIV), the causative agent of acquired immunodeficiency syndrome (AIDS), continues to be a serious public health concern worldwide. HIV infection leads to rapid deterioration of cell-mediated immunity via replication in and consequent destruction of CD4+ immune cells. This deterioration marks the progression towards AIDS and diminished capacity to defend against pathogens. Since its discovery in 1981, nearly 60 million people have been infected with HIV and 25 million have died from AIDS related causes (1). Fortunately the challenge of AIDS has been met with unprecedented drug discovery efforts, yielding 25 FDA approved drugs in 25 years. Access to antiretroviral therapy dramatically prolongs life expectancy and improves quality of life. Antiretroviral drug development is an ongoing process as the rapid evolutionary rate of HIV leads to the emergence of viruses resistant to available drugs. In 2007 the first in a new class of HIV drugs, the integrase inhibitors, was approved by the FDA and has been shown to be remarkably effective (39, 121). Predictably, the drug-resistant viral phenotypes observed in vitro eventually emerged in the clinic and highlighted the need for a new class of integrase
inhibitors with an alternate inhibitory mechanism. The development of these drugs will likely be driven by an improved understanding of integrase structural biology and critical pre-integration events.

1.2 The HIV-1 Replication Cycle

The HIV-1 replication cycle (Figure 1.1) begins with target cell recognition through interactions between viral envelope proteins and host cell surface receptors. Fusion of the viral envelope to the cellular membrane introduces the viral RNA and core proteins into the cytoplasm in the form of a large nucleoprotein complex (37). It is in the context of this nucleoprotein complex that reverse transcription, a process unique to retroviral replication, converts the viral RNA genome into cDNA using the virally-encoded polymerase reverse transcriptase. Following reverse transcription this complex collects several host proteins to form a massive new complex termed the pre-integration complex (PIC). By unknown mechanisms, the HIV-1 PIC is actively transported across the nuclear membrane where it has access to host chromatin. Active transport of the PIC across the nuclear membrane is a peculiarity of HIV-1 and other lentiviruses that enables the infection of target cells without waiting for nuclear disassembly during mitosis (153). This allows infection at all stages of the cell cycle and expands the host range to non-dividing or terminally differentiated cells. Upon entry to the nucleus the PIC, or some remnant of it, is targeted to host chromatin where the viral genome is permanently integrated by the viral integrase (IN) protein. Target site preferences vary among the retroviridae. For example, lentivirus integration occurs most often within active transcription units.
while gammaretroviruses prefer promoter regions and the associated CpG islands (125). After integration the provirus is treated essentially as a host cell gene, transcribed and translated by the host cell machinery (37). Manipulation of host RNA splicing allows for the nuclear export of genome-length RNAs for packaging into new virions as well as spliced RNAs for protein translation. Viral proteins and genomes are targeted to the cell membrane for assembly and budding. Finally, during or shortly after budding, the viral core is processed by the viral protease to yield an infectious particle capable of starting the cycle anew.

1.3. Integration of HIV-1 viral DNA into host chromatin

Permanent integration into the host genome has numerous benefits from the viral perspective. Integration into the host chromosome endows the provirus with long-term stability and provides access to the host transcription machinery. Furthermore, the integrated provirus takes advantage of co-mitotic replication with the host cell. Stable integration has the added benefit of establishing a latent reservoir capable of avoiding immune surveillance and archiving drug-resistant phenotypes (131).

Successful integration requires two spatially and temporally separate IN catalytic activities: 3′-processing and strand transfer (Figure 1.2). The HIV-1 IN protein contains a single active site that carries out both activities (38, 56). Upon completion of reverse transcription, while the pre-integration complex is still in the cytoplasm, integrase binds specifically to the long terminal repeats of the viral genome and cleaves two bases from each 3′ end following a conserved –CA
dinucleotide (38). The strand transfer reaction simultaneously breaks the host DNA phosphodiester backbone and ligates the processed viral DNA ends into opposite strands, across the major groove of the host DNA double helix with an invariable spacing of 5 base pairs between attachment sites. 3‘ processing and strand transfer are essentially the same reaction from a catalytic perspective, with H2O serving as a nucleophile in the first reaction and the exposed 3’-OH in the second. Host DNA repair mechanisms complete the process by removing the unpaired overhangs, filling the gaps and repairing the nicks to yield a permanently integrated genome flanked by a duplication of the five base pair sequence between attachment sites (139, 151). Along with the biologically relevant 3’-processing and strand transfer activities, purified IN exhibits additional activities in vitro. The enzyme can reverse the strand transfer reaction by site selectively cleaving the integrated DNA in a reaction termed disintegration (34). A recent report has indicated that the recombinant protein can also catalyze internal cleavage at a palindromic sequence mimicking LTR-LTR junction (43, 44). However, there is no evidence as of yet that these additional catalytic activities observed in vitro occur in infected cells.

In infected cells IN functions in the context of a large nucleoprotein complex termed the pre-integration complex (PIC), where a number of viral and cellular proteins contribute to retroviral integration (15, 16, 23, 24, 26, 32, 33, 52, 61, 92, 93, 96, 105, 106, 147). PICs can be extracted from infected cells and used for biochemical assays in vitro (12, 14, 27, 50, 53, 63, 95, 169). However, the abundance of these nucleoprotein complexes is insufficient to perform atomic
structural or even lower resolution biophysical analyses. Therefore, recombinant
IN and model DNA substrates have been employed instead to study protein-
nucleic acid interactions. In their simplest form, these assays utilize purified
recombinant protein and short DNA substrates (~21-mer dsDNA mimicking the
U5 end of viral DNA) to monitor 3'-processing and strand transfer activities in
vitro (Figure 1.3A). These reactions, however, are only capable of monitoring
single-end integrations and not the biologically-relevant concerted integration of
two DNA ends into the target. More intricate assays use longer donor DNA
substrates of several hundred base pairs and a circular target DNA to monitor
both 3’ processing and concerted integration (Figure 1.3B) (98-100, 127, 148,
149). Furthermore, the nucleoprotein complexes formed in this improved
experimental design have properties reminiscent of those in pre-integration
complexes and can be assembled in quantities useful for structural analysis (98,
100).

1.4. Structural aspects of HIV-1 integration

HIV-1 IN is notoriously resistant to crystallization efforts, likely owing to its
poor solubility and inherent flexibility. HIV-1 IN is comprised of three stable
domains: an N-terminal domain (NTD), catalytic core domain (CCD) and C-
terminal domain (CTD)(38). Mutant HIV-1 IN proteins with reduced surface
hydrophobicity have been used to crystallize the individual domains as well as
two-domain structures consisting of the NTD+CTD and CCD+CTD (Figure 1.4)
(21, 48, 109). Of note, the soluble mutants used for the structures exhibit
aberrant inter-subunit dynamics in vitro and yield replication-defective viruses in
vivo (82, 123). No structure to date has included viral or target DNA. The N-terminal domain (NTD, residues 1-50) is a compact α-helical bundle stabilized by the coordination of a zinc atom with a distinctive CCHH zinc finger motifs. It plays an important architectural role, stabilizing both protein-protein and protein-DNA interactions (38, 70). The core domain (residues 51-212) contains a single highly conserved active site in which the three acidic residues D64, D116, and E152 coordinate a divalent metal ion (56). Mutations of these residues severely compromise IN activities both in vitro and in infected cells (56, 74, 91).

Biochemical assays with purified IN revealed that it requires either Mg$^{2+}$ or Mn$^{2+}$ to carry out the reactions with model DNA substrates. Of these, Mg$^{2+}$ is considered to be the physiological cofactor due to its relative abundance in the cells. Several structural studies have shown a single divalent metal bound to the active site of the HIV-1 CCD (67). However, based on the two-metal mechanism for structurally and functionally similar polynucleotidyl transferases (8, 170), it has been proposed that DNA binding stabilizes the second metal in the active site (102). IN uses the same catalytic site for 3’-processing and strand transfer reactions. Therefore, the CCD is likely to harbor both viral and target DNA binding sites. Furthermore, the CCD is also an essential building block for formation of the functional multimeric IN. The CCD-CCD interface is fairly large (~1650 Å) and mutations destabilizing these interactions adversely affect IN catalytic activities (47, 166). The C-terminal domain (residues 213-288) is made up of highly basic β-sheets that bind DNA strongly but non-specifically. The C-terminal domain (CTD) is rich in basic amino acids and adopts an SH3-like fold
Other proteins with the same fold bind the minor groove of DNA in a nonspecific manner (66, 88, 138). Similarly, the CTD is thought to provide a stabilizing platform for DNA substrates. In addition, the CTD has been implicated in functional oligomerization of IN. L241A and L242A mutations along the C-terminal dimer disrupted IN dimerization and compromised catalytic activities (114).

IN selectively binds U5 and U3 termini of viral DNA. Footprinting of PICs isolated from infected cells revealed the terminal 200-250 base pairs of each viral DNA end as primary protein binding sites (27). While random DNA sequences can effectively compete with IN-viral DNA interactions and impair the 3’-processing reactions in vitro (90, 150, 160), in the context of infected cells this is unlikely to significantly deter association with viral DNA ends as PIC assembly takes place in the cytoplasm where competition from non-specific DNA would be minimal. Functional assays have shown that IN can distinguish between the viral DNA ends and nonspecific substrates. Mutational studies in vitro have indicated the importance of CA/TG dinucleotide pair for effective 3’-processing of the viral DNA ends (90). Additional proximal regions of viral DNA have also been implicated in specific recognition of the viral DNA (59). One important feature contributing to selective recognition of the LTR termini by IN could be the DNA end distortion. NMR analysis of a 17 base pair oligonucleotide containing the U5 terminal sequence revealed that base stacking and minor groove were significantly disordered at the cleavage site and a correlation between DNA end
distortion and cleavage activities was observed (137). Introducing mismatch bases at the terminal three positions enhanced base unstacking and unpairing, and substantially stimulated the site specific processing activities.

The alternative experimental strategies to identify the LTR regions important for selective recognition involved application of DNA analogs. Probing effects of various DNA backbone, base, and groove modifications on IN catalytic activities suggested that IN requires flexibility of the phosphodiester backbone at the scissile bond (85). The other study examined 2’-modified nucleosides and 1,3-propanediol insertions in various positions of the U5 sequence (2). Akin to the mutagenesis experiments (59) divalent metal dependent affects were observed upon altering certain regions of the DNA (2). Nucleoside modifications at positions 3, 5 and 6 significantly diminished Mg\(^{2+}\) dependent activities, while Mn\(^{2+}\) dependent reactions were less affected. In contrast, Mg\(^{2+}\) and Mn\(^{2+}\) dependent activities were equally impaired when the modifications were introduced at positions 7-9 (2). Taken together, the biochemical approaches enabled the delineation of several important features of viral DNA essential for formation of the functional nucleoprotein complexes. Nevertheless, the detailed mechanism for selective recognition remains elusive. Ideally, atomic structures of IN complexes with specific and non-specific DNAs would be necessary to fully address this question.

Recombinant HIV-1 IN exists predominantly as a dimer in solution, in equilibrium with monomers, tetramers and higher order oligomers. The IN dimer interface is the most stable, with a predicted \(K_d\) of 67.8 pm (155). The less stable
integrase tetramer requires a separate interface that probably involves all three IN subdomains. One particular two-domain crystal structure shows two symmetry-related CCD+NTD particles in what appears to be a tetramer formation formed by crystal lattice contacts (167). In this structure the dimer interface is unchanged and the tetramer interface is stabilized by extensive electrostatic interactions between the NTD and CCD of monomers in separate dimers. The multimeric state of IN in the context of preintegration complexes is unknown, though strong evidence suggests that monomers assemble on the viral LTRs to form a dimer of dimers that carries out both enzymatic activities (65, 69, 123, 136). In such an assembly, only two of active sites would be catalytically active while the remaining monomers would play solely architectural roles.

The available two-domain IN crystal structures can easily be superimposed to generate a full-length model. These models fail to recreate a plausible active site arrangement as the spacing between active sites is >30 angstroms while the five base pair distance between attachment sites in target DNA is only ~15 angstroms (for B-DNA). The inconsistency is most likely explained by the absence of DNA in any of the structures used to build the models.

The structure of prototype foamy virus IN in complex with viral DNA inspired a unique model in which a tetramer of IN in which the active sites are spaced appropriately to carry out the concerted integration of vDNA ends seen in vivo. This tetrameric structure features two IN dimers face to face, with the inner subunits making extensive contacts with viral DNA and the CTD and NTD of the
same subunits participating in protein-protein interactions (70). Only the CCDs of the outer subunits were resolved in the crystal structure, and they seem to be placed far from the viral DNA substrate. The role of these outer subunits could be to interact with target DNA or cofactors, or they could be playing exclusively an architectural role.

1.5. LEDGF as a cellular cofactor of HIV-1 integration

While purified recombinant IN is capable of carrying out the enzymatic activities in vitro using small DNA substrates, these assays do not reflect the complexities of integration in vivo. The PIC contains numerous viral and cellular proteins and integrates a >9kb viral genome into chromatin, which is itself complex in structure and composition. A number of cellular cofactors are suspected to interact with HIV-1 IN (52, 157). Among these, lens epithelium-derived growth factor (LEDGF) has been identified as critical for effective integration (32, 105). LEDGF is a ubiquitous cellular transcription factor of uncertain purpose found associated with chromatin throughout the cell cycle. It is a 530 amino acid 60kD protein (the “p75” was coined based on its apparent molecular weight on an SDS-PAGE gel). The LEDGF N-terminal domain ensemble (NDE) consists of a PWWP domain, a nuclear localization signal (NLS), and dual copies of the AT-hook DNA binding motif (158). The AT-hook domain and PWWP domain are required for association with chromatin while the NLS is possibly contributes to nuclear import of the PIC (107). The LEDGF C-terminus contains a highly conserved integrase binding domain (IBD) that strongly binds HIV-1 IN as well as various cellular proteins (31, 32, 171). The
solution structure of the LEDGF IBD and a co-crystal structure of the IBD in complex with the IN catalytic core domain have been reported (30, 33). The IBD consists of an 82 residue α-helical bundle that docks into a relatively small cavity at the IN catalytic core domain dimer interface, contacting both IN subunits (30).

Cherepanov et al observed that ectopically expressed FLAG-tagged HIV-1 IN consistently pulled down LEDGF and co-localized with LEDGF in the nucleus (32). Suppression or knockout of LEDGF abolished chromatin association and nuclear localization and reduced integration levels ~10-fold in infected cells (105). In addition, integration into active transcription units was reduced in these cells (36). These findings immediately suggested a role for LEDGF in nuclear import of the PIC and integration target site selection. These findings were especially compelling in light of the fact that the IN-LEDGF interaction is lentivirus specific and suppression of LEDGF abolishes the two hallmarks of lentiviral infection: nuclear import of PICs and an integration preference for active transcription units (29). The role of LEDGF as a nuclear import factor remains controversial. NLS-mutant LEDGF proteins have been shown to sequester IN in the cytoplasm when both proteins are expressed at high levels (163). However, the relevance of these experiments to naturally occurring viral infection is questionable, as the majority of LEDGF in the cell is chromatin associated and LEDGF has not been reported to be packaged in viral particles. In viral replication assays using LEDGF knockout cells, the accumulation of unintegrated HIV genomes in the nucleus indicated the replication block was at the level of integration and not at nuclear translocation (147). The residual infection of
LEDGF knockdown cells suggests a LEDGF-independent mechanism for nuclear translocation (147, 163). Furthermore, PICs isolated from the cytoplasmic fractions of LEDGF knockout cells are fully competent for integration in *in vitro* assays (147). Taken together these findings suggest that LEDGF is not a critical component for assembly or function of the PIC, and that the critical role of LEDGF lies in its ability to direct PICs to active transcription units and tether them to chromatin.

While the ability of LEDGF to tether proteins to chromatin has been demonstrated, the details of LEDGF interactions with chromatin are mostly unexplained. The PWWP module found in the LEDGF N-terminus is present in ~60 eukaryotic proteins and is related in sequence and 3-D structure to a number of other chromatin binding domains collectively known as the Tudor clan domains (119). The PWWP domain of the closely related hepatoma-derived growth factor (HDGF) has been crystallized and the 3-D structure is presumably similar to LEDGF PWWP (113). The defining feature of the HDGF PWWP is a 5 stranded antiparallel β barrel structure that features a surface exposed hydrophobic pocket (113). The function of LEDGF in the cell is unclear, though it seems to act as a general co-activator of transcription and chromatin tether. LEDGF knockout mice survive, although with increased mortality (152). It was originally described as a transcription factor that controlled heat shock and stress response elements, although later experiments were not able to replicate a DNA sequence-specific binding of LEDGF (140, 142, 158).
1.6 Figures for chapter 1

Figure 1.1 HIV replication cycle: HIV viral envelope proteins interact specifically with surface receptors on target cells, leading to membrane fusion and entry of the viral core into the cytoplasm. The viral RNA genome is reverse transcribed into cDNA, which is transported into the nucleus and permanently integrated into the host DNA by the viral enzyme integrase. The integrated provirus is used as a template for the synthesis of RNAs encoding viral proteins as well as full-length RNAs for packaging into new viral particles. Translation of viral proteins and viral particle assembly lead to the budding of new infectious particles capable of infecting other cells. Figure adapted from Ciuffi et al, 2006 (35).
Figure 1.2 Enzymatic activities of HIV integrase: (A) In the cytoplasm HIV IN recognizes the 3'- viral DNA ends and (B) cleaves a GT dinucleotide, leaving recessed ends. (C) In the nucleus the viral pre-integration complex engages chromatin where (D) the exposed 3'-OH groups are used in a nucleophilic attack that breaks the target DNA phosphodiester backbone and attaches the viral DNA ends five base pairs apart on opposite strands. (E) Host DNA repair enzymes fill in the recessed ends and seal the nicks in the DNA backbone. Figure adapted from Pommier et al, 2005 (132).
Figure 1.3 Assays of integrase enzymatic activity: (A) $^{32}$P labeled 21-mer dsDNA mimicking viral DNA ends is used as a substrate for both the 3'-processing and strand transfer reactions. (B) Long $^{32}$P labeled DNAs (>200 base pairs) with ends mimicking viral DNA are pre-incubated with IN to assemble a stable synaptic complex in which the 3'-processing reaction occurs. The target is super-coiled plasmid DNA, which is uncoiled by a single-end integration event or linearized by the successful concerted integration of two viral DNA ends.
Figure 1.4 Two-domain structures of HIV integrase The crystal structures of HIV IN (A) catalytic core domain + C-terminal domain (28) and (B) N-terminal domain + catalytic core domain (167). Individual IN subunits comprising the dimeric structure are colored green and yellow.
CHAPTER 2

SUBUNIT SPECIFIC VIRAL-DNA CONTACTS AND A DNA-INDUCED CONFORMATION CHANGE WITHIN FUNCTIONAL HIV-1 INTEGRASE NUCLEOPROTEIN COMPLEXES

2.1 Introduction

HIV-1 integrase (IN) is commonly viewed as an important therapeutic target for the following reasons: its catalytic activities are required for viral replication, there is no closely related cellular equivalent of IN, and specific IN inhibitors are likely to be effective against viral strains resistant to currently available therapies targeting reverse transcriptase (RT), protease, and virus-cell fusion. Detailed structural information on functional IN-DNA complexes could aid drug design efforts. For example, the promising diketo acid class of inhibitors preferentially bind to the assembled IN-viral DNA complex rather than the free protein (58, 68, 72, 73).

The two chemical reactions catalyzed by HIV-1 IN, 3' processing and DNA strand transfer, have been characterized in detail (reviewed in (13)). First, IN removes two nucleotides from each 3' end of the viral DNA synthesized by RT. In the following step, concerted transesterification reactions covalently join the viral DNA ends into the host genome (57). In vivo, the enzyme acts in the context of a
large nucleoprotein complex with a number of viral and host proteins contributing to the integration process (15, 16, 23, 24, 26, 32, 33, 52, 61, 92, 93, 96, 105, 106, 147).

HIV-1 IN is composed of three distinct structural and functional domains: the N-terminal domain (NTD) (residues 1-50) that contains an HHCC zinc binding motif, the catalytic core domain (CCD) (residues 51-212) containing the DDE motif essential for coordinating catalytic divalent metals, and the C-terminal domain (CTD) (residues 213-288) that is thought to provide a platform for DNA binding. Crystallographic or NMR structural data are available for each of the individual domains (20, 47, 49, 67, 108). In addition, two-domain CCD/CTD (28) and NTD/CCD (166) crystal structures have been determined. However, efforts to obtain detailed structures for the full-length protein or IN-DNA complexes have been impeded by relatively poor protein solubility and/or full-length enzyme flexibility.

To better understand IN-DNA interactions, a number of biochemical approaches have been employed. Site directed mutagenesis experiments indicated that different monomers within the IN multimer provide complementary rather than symmetrical contacts to DNA (54, 159, 161). Photo and chemical cross-linking studies revealed several potential DNA binding residues (59, 75, 76, 83, 84, 120). These biochemical experiments together with crystallographic determination of the two-domain structures prompted molecular modeling research. However, the IN-DNA models obtained by different groups vary
significantly, indicating that the available experimental data comprises an insufficient number of constraints for formulating a common outcome (25, 41, 64, 76, 86, 130, 166, 168).

Here we employed a new experimental strategy combining two established methodologies of site specific protein-DNA cross-linking through disulfide bond formation and mass spectrometric (MS) protein footprinting. This approach enabled us to identify monomer selective contacts with substrate DNA. Importantly, our studies found that upon DNA binding IN undergoes a conformational change involving the alpha helical connection between the core and C-terminal domains. We also uncovered a potential role for the NTD in binding to the viral DNA substrate. The functional importance of the identified residues was confirmed by site directed mutagenesis. Our data provide new and important details on how HIV-1 IN interacts with its viral DNA substrate.

2.2 Materials and Methods

2.2.1 Preparation of recombinant HIV-1 IN, DNA oligonucleotides, and cross-linked products: Full-length IN and mutant proteins were expressed in *E. coli* and purified as described previously (84). The catalytic activities of recombinant proteins were monitored according to the reported procedure (84). Phosphoramidites (O6-phenyl-dI, O4-triazolyl-dU and 2-F-dI) were purchased from Glen Research (Sterling, VA) and synthetically incorporated in the viral DNA sequences using an Applied Biosystems 392 DNA synthesizer. Post-synthetic attachment of the carbon tether (cystamine) to the modified bases, and subsequent activation and purification of oligonucleotides were performed as
published previously (84). For cross-linking reactions, 10 μM IN was incubated with equimolar DNA duplex in 50 mM HEPES, pH 7.0, 100 mM NaCl, 5 mM MgCl₂, and 10% glycerol at 37 °C for 20 min. The reactions were quenched by 20 mM methyl methanethiosulfonate and subjected to surface topology analysis described below.

2.2.2 Mass spectrometric footprinting: Free IN and cross-linked IN-DNA complexes were treated at 37 °C with NHS-biotin or p-Hydroxyphenylglyoxal (HPG) for 30 and 60 min, respectively. The reagent concentrations are indicated in the text and figure legends. The reactions were quenched by excess Lys and Arg using free amino acid forms. Cross-linked products were separated by denaturing SDS-PAGE using nonreducing gel loading and separation buffers. Bands were visualized by Microwave Blue stain (Protiga, Gaithersburg, MD), excised, and destained according to the described procedure (40, 89, 103, 117, 141, 143). In-gel proteolysis was performed using 0.5 μg trypsin to generate small peptide peaks amenable to MS and MS/MS analysis. Tryptic peptides were monitored by a Kratos MALDI-ToF instrument equipped with a curved field reflectron feature (Kratos Analytical Instruments, Manchester, U.K.), enabling generation of post source decay (PSD) amino acid sequence data. MALDI-ToF experiments were performed using α-cyano-4-hydroxy-cinnamic acid as a matrix. For accurate quantitation of the modified peptide peaks, the intensities of unmodified IN tryptic peptides were used as internal controls.

2.2.3 CD Spectroscopy: CD spectra were recorded on an AVIV 202 circular dichroism spectrometer (USA). A 0.01 cm pathlength cell and 1 nm bandwidth
were used to make measurements in the near (215-320 nm) ultraviolet region. The protein concentration was 10 µM in 20 mM PIPES, pH 6.8, 750 mM NaCl, 2 mM β-mercaptoethanol, 0.1 mM EDTA and 1 mM CHAPS.

2.2.4 Virus infectivity: The SpeI-EcoRI fragment from pNL4-3 containing the IN coding sequence sub-cloned into pBlueScript (Invitrogen) was subjected to QuikChange mutagenesis (Stratagene, La Jolla, California), and the mutated fragments were returned to pNL4-3. AgeI-PflMI fragments were subsequently exchanged for the IN-coding fragment in the NLX.Luc.R- strain that carries the luciferase reporter gene (110). Normalized levels of viral infectivities using SupT1 target cells were calculated as previously described (110).

2.3 Results

We chose to employ MS protein footprinting (40, 89, 103, 117, 141, 143) to study IN-viral DNA interactions for the following reasons. The methodology proved to be effective for identifying contact amino acids to cognate nucleic acids for a number of nucleoprotein complexes not amenable to conventional structural biology analyses such as X-ray crystallography or NMR. Of note, the MS footprinting consistently revealed functionally essential interactions (40, 89, 103, 117, 141, 143). Briefly, the approach is based on comparing surface topologies of free protein versus pre-assembled nucleoprotein complexes using amino acid specific reagents such as NHS-biotin and HPG that modify exposed lysine and arginine residues, respectively. The basic residues are targeted due to their likely role in electrostatic contacts with nucleic acids. Reagent concentrations are carefully optimized to ensure mild modification conditions where the integrity of
the functional complex is preserved. Subsequent MS analyses reveal the surface residues readily labeled in free protein yet shielded from modification in the complex due to bound nucleic acid.

Unlike previously examined nucleoprotein complexes, IN-DNA interactions presented us with a new challenge as homologous protein monomers are predicted to provide complementary rather than symmetrical contacts to cognate DNA. Indeed, earlier complementation studies (54, 161) suggested that the viral DNA is coordinated by the CCD of one monomer and the CTD of another monomer. Because of this, it was essential to separate individual monomers based on their differential interactions with DNA to accurately assign contacts within full length multimeric IN. For this, we employed a new experimental strategy combining two technologies established in our laboratories: site-specific protein-DNA cross-linking through disulfide bridging (reviewed in (164)) and MS footprinting (40, 89, 103, 117, 141, 143) (Figure 2.1). Our goal was to obtain two different cross-linked complexes: one with the CCD tethered to the DNA end, and the other with the CTD linked to a subterminal position in the DNA substrate. The two complexes were then exposed to selective amino acid modifying reagents. Subsequent SDS-PAGE enabled us to separate the cross-linked complexes from free protein. As depicted in figure 2.1, monomer 1 cross-linked to DNA through CCD residue E152C was separated from monomer 2 (Figure 2.1, middle). In a parallel experiment, the monomer 2-DNA complex cross-linked through a selective CTD residue was separated from monomer 1 (Figure 2.1, bottom). Cross-linked products were then processed and analyzed separately using MS.
We predicted that protection patterns within monomer 1 and monomer 2 would differ, and the identified amino acids in the respective monomers would reveal a DNA binding channel in the context of full length multimeric IN.

To form disulfide bridged IN-DNA complexes, the alkanethiol tether was placed at selected sites in the viral DNA sequence. DNA oligonucleotides were prepared through synthetic incorporation of commercially available analogs of A, C and G. Crystallographic and NMR studies have shown that placements of cross-linkable base analogs in double stranded DNA do not alter Watson-Crick base pairing or global nucleic acid structures (reviewed in (164)). Furthermore, viral sequences containing the modified bases were effective IN substrates (Figure 2.2).

IN was prepared for cross-linking using a previously reported two-step procedure (64, 84). First, 3 surface Cys residues (C56, C65 and C280) were replaced with Ser to rid non-specific cross-linking between the wild type protein and modified DNA. Of the remaining three cysteines, C40 and C43 coordinate the structural Zn ion, and C130 is hindered from the protein surface (28, 166). Indeed, the triple mutant (C56S, C65S and C280S) exhibited minimal reactivity with cross-linkable DNA (84). In the second step, the reactive Cys residue was introduced at a surface position predictive of specific nucleoprotein complex formation. Based on available mechanistic and structural studies, we chose E152C to tether IN to DNA modified at base G2 (numbering as in Figure 2.2, upper panel). E152 is the catalytic Glu of the DDE motif, and G2 is located at the scissile phosphodiester bond. Furthermore, the reactive SH group in the G
analog points to the minor groove (164), stabilizing the IN catalytic site immediately adjacent to the scissile bond. To confirm specificity, binding of E152C protein to DNA modified at other positions was examined. The results reveal a striking preference for formation of the IN(E152C)xDNA(G2) complex (Figure 2.3A). The reduced reactivity for position 2 in the lower strand could be explained by the fact that the SH group of the C-analog points to the major groove, away from the cleavage site. The observation that the other sites yield only residual cross-linking (Figure 2.3A) indicates the specific nature of the IN(E152C)xDNA(G2) interaction.

Previous studies by Bushman and colleagues (64) identified IN(E246C) as a plausible CTD contact to DNA. Using TNB-thio tethered substrates the authors observed robust but non-specific cross-linking of IN(E246C) to various nucleotide positions. Further experiments with less reactive nucleotide analogs uncovered the relative preference of IN(E246C) for position A7 albeit significantly reduced formation of the nucleoprotein complex. To generate sufficient amounts of cross-linked products for our MS footprinting we used TNB-activated DNAs. In addition to IN(E246C) crosslinking to A7 (Figure 2.3B, lane 5) we probed interactions with additional nucleotides including G16 and G19 located at the distal region of the DNA substrate (Figure 2.3B, lanes 6 and 7). The data indicated comparable reactivities of IN(246C) with various cross-linkable nucleotides (Figure 2.3B). While these non-specific interactions were expected, the available crystal structure of the CCD-CTD fragment (28) allowed us to delineate functional from non-productive IN-DNA interactions. For example, based on the spatial
separation between the CCD and CTD, we could predict that crosslinking IN(E246C) to nucleotides adjacent to the scissile bond (Figure 2.3B, lanes 2 and 3) would position the CCD toward the opposite terminus of the DNA substrate. In contrast, in the functional complex the CCD would have to directly interact with the 3'-processing site, apposing the CTD to the distal regions of the DNA substrate. Therefore, for the footprinting studies we chose to examine IN(E246C) crosslinks to positions A7, G16 and G19.

Of note, the triple (C56S, C65S, and C280S) background mutant and E246C proteins were catalytically active, indicating these mutations did not significantly alter IN folding or its interactions with viral DNA (Figure 2.2B). As expected, replacement of active site residue E152 compromised IN activity, due to the inability of the substituent amino acid to coordinate the essential magnesium ion for catalysis. Therefore, the integrity of E152C IN was assessed as follows. First, E152C efficiently crosslinked with DNA(G2) (Figure 2.3A, lane 3), indicating that its global structure is intact. Chemical modification and subsequent MS analyses importantly revealed very similar surface topologies for the wild type and E152C INs. Furthermore, the two proteins yielded overlapping CD spectra. This collection of data supports the notion that substituting Cys for Glu-152 does not overtly alter IN structure.

The modification reactions were carried out under mild conditions such that the integrity of the nucleoprotein complex was preserved. For this, effects of increasing concentration of NHS-biotin and HPG modifiers on wild type IN activities were examined. Representative data in Figure 2.4 show that treatment
of free IN with NHS-biotin using the 1/50 ratio of protein/reagent inactivated IN function, probably due to modification of essential lysine residues in the vicinity of the enzyme active site (83). To ensure that this was the case as compared to gross disruption of IN structure, we subjected the pre-assembled IN-DNA complex to the same modification. The result was a remarkable retention of original IN activity, indicating that under these conditions, the integrity of the functional complex is preserved (Figure 2.4, lane 3). The optimal IN/HPG ratio was analogously defined as 1/1000. Importantly, centrifugation assays (15,800 g for 1h) indicated that the crosslinked complexes were fully soluble (data not shown).

The two nucleoprotein complexes IN(E152C)xDNA(G2) and IN(E246C)xDNA(A7) were subjected to MS footprinting analysis according to the scheme depicted in Figure 2.1. Representative segments of the mass spectra are depicted in Figure 2.5. C1, C2 and C3 peaks are unmodified IN peptides that provide internal controls for accurate quantitation of modified peptides. The peptide peak 265-269 containing biotinylated K266 is another control, as this lysine is equally exposed in the free protein and IN-DNA complexes (Figure 2.5A). Figure 2.5B reveals monomer selective protection of K160. This residue is readily modified in free IN. However, upon formation of the IN(E152C)xDNA(G2) complex, K160 is shielded from modification. In contrast, the residue remains surface-exposed in the IN(E246C)xDNA(A7) complex. A converse picture is observed for CTD residue K264 (Figure 2.5C). The intensity of the modified peak persists in the IN(E152C)xDNA(G2) complex, but was significantly diminished in
the IN(E246C)xDNA(A7) complex. Taken together, these results alongside the summary in Table 2.1 demonstrate that separate monomers differentially contact substrate DNA. DNA cross-linked via E152C specifically protected K14, K159, K160, and R199, whereas residues K219, K244, R263, K264, and R269 were footprinted only via the E246C link. The protection patterns of E246C with positions A7, G16 and G19 were very similar.

Interestingly, a unique modification pattern was observed for R199 (Figure 2.6). This residue was only marginally modified in the context of free IN. The reactivity was further reduced in the IN(E152C)xDNA(G2) complex. Of note, we observed a remarkable increase in the extent of R199 modification within the IN(E246C)xDNA(A7) complex (Figure 2.6C), indicating that subunit-specific DNA binding significantly exposes this amino acid to the protein surface. R199 was also more reactive in the IN(E246C)xDNA(G16) and IN(E246C)xDNA(G19) complexes as compared with the free protein. However, a pattern could be noticed wherein the extent of R199 modification decreased as the crosslinking points were further distanced from the cleavable DNA end (compare 4C, D and E).

To address the specificity of the IN-DNA interactions, E152C was cross-linked with non-specific double stranded and single-stranded DNA. The non-specific double strand did not yield sufficient cross-linked species for quantitative MS analysis. These results together with the data depicted in Figure 2.3A indicate that in non-specific IN-DNA complexes, E152C is not positioned close enough to establish disulfide cross-links with the DNA. The single stranded
oligonucleotide resulted in higher yields of cross-linked product than its double stranded counterpart, but constituted only ~25% of the specific IN(E152C)xDNA(G2) complex (data not shown). MS comparison of IN(E152C)-ssDNA and IN(E152C)xDNA(G2) products indicated sharp differences in protein-DNA contacts within the two complexes. The residues protected in the latter complex remained surface exposed in the former species. Representative data showing differential accessibilities of K14 in the two complexes are depicted in Figure 2.7. These results suggest that detectable cross-linking between E152C and ssDNA was due probably to the highly flexible nature of the oligonucleotide, rather than its specific binding to IN.

We next used site-directed mutagenesis to confirm the functional importance of the identified amino acids. In fact, several interacting residues revealed from our footprinting results have previously been examined by mutagenesis due to their high degree of conservation among various HIV strains, and were shown to be essential for recombinant IN function and HIV-1 replication (83, 110, 111). Therefore, our efforts focused on analyzing two novel contacts: K14, the first NTD residue to be implicated in viral DNA binding, and R199, involved in the protein conformational change in one subunit while interacting with DNA in the other monomer (Table 2.1). The substitution of Ala or Glu was engineered at each position within the wild type IN background. MS surface mapping and CD spectroscopy revealed proper folding of the recombinant mutant proteins. The results in Figure 2.8 show that the K14A/E and R199A/E mutants exhibited residual levels of specific activity (3’-processing and DNA
strand transfer) and generated non-specific (20-mer) cleavage products. Interestingly, the same phenomenon was observed with mutations of certain IN residues (for example Q148) implicated in specific recognition of the viral DNA substrate (84). An alternative scenario that our preparations could be contaminated with *E. coli* nucleases is less tenable as the D116N active site mutant purified according to the identical procedure did not generate the 20-mer product (Figure 2.8, lane 7). Also noteworthy is that the Ala mutants displayed significantly reduced specific activity, while the Glu mutations yielded more pronounced affects (Figure 2.8, compare lane 4 to lane 3 and lane 6 to lane 5). Neutralizing the positive charge that is needed to stabilize DNA at the specific site by Ala substitutions could reduce binding affinity, while the placement of Glu may result in more efficient displacement of the nucleic acid from its specific site due to electrostatic repulsions.

Each of the four mutations also significantly impacted the infectivity of the virus. The R199A mutant retained ~0.24% of wild-type HIV-1<sub>NL4-3</sub> function in a single-round infection, whereas the other viruses failed to yield detectable infectivities (Table 2.2).

Our experimental results were used to dissect the viral DNA binding channel in the functional IN nucleoprotein complex. For this, a full length IN model was created by superimposing the two available crystal structures: NTD/CCD and CCD/CTD (28, 166). Then the residues identified by our footprinting analysis were located on the IN monomers (Figure 2.9). The assignment of the amino acids to the respective IN subunits was critical to
visualize a plausible DNA binding channel. The DNA substrate was readily positioned along the path of the basic residues (Figure 2.9) with all three protein domains providing direct contacts to nucleic acid.

2.4 Discussion

We have investigated interactions between full-length HIV-1 IN and its viral DNA substrate using a novel experimental strategy combining two established methodologies: site-specific protein-DNA cross-linking through disulfide bridging, and MS protein footprinting. The cross-linking technology has been successfully used in combination with x-ray crystallography to obtain atomic structures of a number of nucleoprotein complexes including the covalently trapped complex of RT with a DNA primer-template and a deoxynucleoside triphosphate (78, 164). Analogous efforts for the HIV-1 IN-DNA complex have so far been hampered by poor protein solubility and/or full length enzyme flexibility. Combining site-specific cross-linking and MS footprinting offered a powerful alternative and led us to the following important findings. Asymmetric contacts between IN and viral DNA were identified, which provided major constraints to dissect the viral DNA binding channel in the functional nucleoprotein complex. In addition, our experiments uncovered a subunit-specific conformational change induced upon DNA binding.

Our work was based on and expanded previous optimization of IN-DNA cross-linking through disulfide bridging (64). We complemented the cross-linking approach with MS surface topology analysis of basic amino acid residues. Our results reveal the importance of these technologies for identification of
asymmetric contacts between IN and viral DNA. While the observed protections near the cross-linking points were expected, the method also revealed subunit selective contacts distant from the disulfide linkages. For example, K14 within the NTD was protected when CCD residue E152C was cross-linked to DNA(G2). In addition, increased reactivity rather than protection of R199 was observed in the IN(E246C) complexes with DNA substrates. Site-specific cross-linking was also essential to delineate specific from non-specific IN-DNA interactions. IN-DNA binding could yield a mixture of active and non-productive complexes. However, functional site-specific modification of IN and DNA enabled us to covalently link and separate specific complexes from non-productive IN-DNA interactions, the latter of which dissociated under denaturing SDS-PAGE (Figure 2.3A). Although our method does not distinguish whether the observed amino acid protections are due to direct protein-nucleic acid or protein-protein contacts, the protections appear to strictly depend on the presence of the specific HIV-1 DNA substrate in the reaction mixture. The surface basic residues identified in this study are moreover likely to engage in electrostatic interactions with the DNA substrate. Indeed, all the interacting amino acids align in a plausible DNA binding channel (Figure 2.9).

Our results for the first time implicate the NTD in direct contact with viral DNA. Previous studies revealed the importance of the NTD for IN catalytic activities via testing mixtures of different protein domains in functional trans-complementation assays (54, 159, 161). This approach, however, fell short of demonstrating direct NTD-viral substrate interactions. We in contrast observed
specific protection of NTD residue K14 in the IN(E152C)xDNA(G2) complex, but
importantly not with the non-specific IN(E152C)-ssDNA complex (Figure 2.7) nor
IN(E246C) complexes (Table 2.1). Other studies reported that the N-terminal 11
residues of HIV-1 IN cross-linked to the target DNA component rather than viral
sequences in a dumbbell disintegration substrate (75, 76). Given that the authors
examined a limited number of cross-linkable sites in the viral DNA, the
contribution by K14 could easily have been missed. In contrast, the Lys and Arg
modifying reagents employed here provide a significantly more detailed picture of
individual amino acid residues. The alternative scenario that K14 became
protected at a point in the reaction pathway that lay significantly downstream of
IN-viral DNA complex formation for example during assembly of the DNA strand
transfer complex could be ruled out for the following reasons. While our in vitro
assays could yield strand transfer reaction products after 3’-processing of the
blunt-ended 21-mer substrate, our crosslinking and modification reactions were
performed for relatively short periods of time under which no 3’-processing or
strand transfer products were detected (data not shown). In addition, the
IN(E152C)xDNA(G2) complex, in which K14 was protected from modification,
was catalytically inactive. Therefore, our findings suggest that K14 plays a role in
interacting with viral DNA. This conclusion is further supported by the site-
directed mutagenesis studies, which showed that substituting K14 with Ala or Glu
compromised both 3’-processing and strand transfer activities in vitro and
abrogated HIV-1 infectivity in cell culture. Interestingly, the recombinant mutant
proteins generated -1 cleavage products (Figure 2.8), a phenotype reminiscent of
the pattern observed with mutations of certain IN residues involved in selective interactions with viral DNA (84). Figure 2.9 indicates that K14 is positioned in the viral DNA binding channel, where it could indeed contribute to cognate DNA recognition.

Another key finding of this study is a DNA-induced protein conformational change involving the extended alpha helix that connects the CCD to the CTD. In a previous study, Asante-Appiah and Skalka identified a metal induced reorganization of HIV-1 IN structure, which affected the recognition of the CCD and CTD, but not the NTD, by domain selective antibodies (7). Bushman and co-workers observed differential cross-linking of CTD residues with blunt ended and processed DNA substrates, suggesting protein structural changes upon cleavage of the viral DNA terminus (64). Roth and co-workers found that a 19 amino acid sequence insertion at E212, located on the helix connecting the CCD and CTD, was tolerated by IN as this large structural change did not significantly impact enzyme function (133). Our observation of the increased reactivity of R199 in the IN-viral DNA complex (Figure 2.6) provides important clues regarding the nature of protein structural changes. R199 is located on the ridged helix (aa 196-221) bridging the CCD to the CTD (see Figure 2.9). The helix is flanked by a flexible loop (aa 187-195) and the highly basic CTD on its N- and C-termini, respectively. Such structural organization enables the flexibility of the loop to be translated to and exploited by the CTD upon formation of the functional nucleoprotein complex.
Our interpretation of the step-wise decrease in R199 reactivity within the IN(E246C)xDNA(G16) and IN(E246C)xDNA(G19) complexes as compared with IN(E246C)xDNA(A7) is the following. Linking the CTD on A7 with the CCD concomitantly committed to the scissile bond at the 3’ end of the cleaved strand would require a significant conformational change in the protein. Accordingly, the CTD accommodated distal G16 and G19 sites with less overall structural rearrangements (Figure 2.6). This observation reflects an extraordinary degree of flexibility within this region of HIV-1 IN. The residual reactivity of R199 in the IN(E246C)xDNA(G16) and IN(E246C)xDNA(G19) complexes could reflect relatively minor adjustments of the CTD to bound DNA. We would argue that the CTD interacts with the energetically more favorable distal regions of the DNA substrate. In Figure 2.9, G19 is located nearby protected K244 and crosslinkable E246 residues. However, the binding site of the CTD can apparently vary. Unlike the CCD that needs to precisely position over the cleavable scissile bond, the CTD is highly flexible and thus engages in non-specific contacts with DNA \textit{in vitro} (Figure 2.3B). In this respect it is noteworthy that insertion of 19 amino acids at the CCD-CTD connection, which could have very well altered the positioning of the CTD on viral DNA, did not significantly affect IN activity (133). We propose that this reflects an inherent flexibility of the CCD-CTD connector loop, which could be essential for the productive assembly and stability of the nucleoprotein complex.
Why does viral DNA bind to only one surface of the IN multimer while the opposite side of the protein remains unaligned? There are at least the following two factors to consider. The crystal structure of the CCD-CTD fragment indicates that the extended α-helix (aa 196-221) is a canonical cylinder in one monomer and is bent by ~45 degrees at Gln209 in the second monomer, thus rendering the IN dimer asymmetric ((28), also see Figure 2.10). Our findings indicate that upon DNA binding only one of the two monomers undergoes a conformational change. Thus, the DNA bound and surface exposed sites within the IN multimer differ significantly. We suggest that the conformational change together with the kink observed in the crystal structure have biological relevance and are important for the assembly of the functional nucleoprotein complex.

IN dimers have been shown to bind each viral DNA end and catalyze the 3’-processing reactions (69, 71). Following GT dinucleotide removal IN remains stably associated with cognate DNA (69) and the two viral DNA-bound dimers presumably form a tetramer to carry out the integration of both viral DNA ends, as occurs during infection (100). Figure 2.9 depicts the IN dimer bound to one viral DNA end. Further experimental data elucidating protein-protein interfaces in the synaptic complex are needed to validate higher order interactions at play during the concerted integration of both viral DNA ends. Conditions for effective full-site integration by recombinant IN and 2-D-gel electrophoretic isolation of synaptic complexes have been reported (100, 149). These techniques can be combined with MS surface topology analyses to better understand the protein-protein interactions essential for the two-ended integration reaction. It will also be
intriguing to apply our footprinting approach to study effects of physiologically relevant cell factors like LEDGF/p75 on IN-DNA interactions. Our identification of a restricted, single viral end-specific DNA binding channel by the combined use of subunit-specific protein-DNA cross-linking and MS footprinting is an important step toward these long-term goals. In addition, the new experimental strategy reported here can be employed by other research groups to analyze a large variety of nucleoprotein complexes.

The new findings reported here also have implications for exploiting the IN-viral DNA structure as a therapeutic target. For example, the most promising diketo class of inhibitors are known to specifically interact with the pre-assembled IN-viral DNA complex and impair the DNA strand transfer activity (58, 68, 132). More recently, IN-DNA interfacial quinolonyl diketo acid derivatives, which potently inhibit both 3’-processing and strand transfer, were reported (46). The novel IN-viral DNA contacts uncovered herein could help to elucidate the mechanisms of these known compounds and facilitate rational evolution of new potent inhibitors.
2.5 Tables for Chapter 2

<table>
<thead>
<tr>
<th>Residue</th>
<th>Free IN</th>
<th>IN(E152C)xDNA(G2)</th>
<th>IN(E246C)xDNA(A7)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>K14</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>K111</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>K159/K160</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>K160</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>R166</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>K188</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>R199</td>
<td>+</td>
<td>-</td>
<td>+++</td>
</tr>
<tr>
<td>K211</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>K219</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>R224</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>R228</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>K244</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>R262</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>R262/R263</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>K264/K266</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>K266</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>R269</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>K273</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

*: IN(E246C) complexes with DNA(A7), DNA(G16) and DNA(G19) yielded very similar protection patterns;
+ , surface exposed residues readily modified;
- , amino acids inaccessible to modification;
++++, increased reactivity (see Figure 2.6);
The amino acids that become shielded from modifiers in the nucleoprotein complex are highlighted.

Table 2.1 Susceptibility of IN basic residues to modification
<table>
<thead>
<tr>
<th>Mutant</th>
<th>Infectivity*</th>
</tr>
</thead>
<tbody>
<tr>
<td>V165A</td>
<td>0.01 ± 0.01</td>
</tr>
<tr>
<td>K14A</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>K14E</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>R199A</td>
<td>0.24 ± 0.13</td>
</tr>
<tr>
<td>R199E</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>D64N/D116N</td>
<td>0.05 ± 0.02</td>
</tr>
</tbody>
</table>

* Percent wild-type infectivity (2.7x10^5 relative light units/µg total cell protein using 4 x 10^7 RT-cpm of HIV-1_{NLX.Luc.R-} and 5 x 10^6 SupT1 cells) of three independent infections with standard error. V165A and D64N/D116N were included as prototypical class II and class I IN mutant viruses, respectively (110).

Table 2.2 Infectivities of IN mutant viruses
Figure 2.1 Experimental strategy: Free IN (space-fill model of the two domain CCD/CTD structure is used for illustration) and the two cross-linked IN-DNA complexes are examined in parallel experiments. The two IN mutants each containing a reactive Cys residue in either the CCD (E152C) or CTD (E246C) are cross-linked to DNA substrates that contain modified nucleotides placed at respective positions. Free IN and the two complexes are subjected to treatment by small chemical modifiers (M). Surface residues in free IN and complexes are modified, but amino acids interacting with DNA are shielded from modification. Complexes and free IN monomers are separated by SDS-PAGE. Monomer-DNA complex bands are excised and subjected to in-gel proteolysis. Subsequent comparative MS analyses reveal modification patterns in free protein and individual monomers.
Figure 2.2 IN mutant enzymatic activity: The upper panel shows the 21-mer U5 viral DNA end substrate used in our studies. The modified nucleotides are in bold and the IN cleavage site is indicated by a vertical arrow. (A) Wild type IN 3' processing and strand transfer activities using native and thiol cross-linkable oligonucleotides. Lane 1: no IN control; lane 2: unmodified DNA sequence+IN; lane 3: DNA(G2)+IN; lane 4: DNA(A7)+IN; lane 5: DNA(G16)+IN; lane 6: DNA(G19)+IN. (B) 3'-processing and strand transfer activities of the following proteins: lane 1: no IN control; lane 2: wild type IN; lane 3: the background triple (C56S, C65S and C280S) mutant IN, lane 4: the IN(E152C) mutant; lane 5: the IN(E246C) mutant. Of note, while E152C (lane 4) and E246C (lane 5) mutations where introduced in the background SSS sequence (lane 3), these proteins for simplicity are referred in the figure and text as IN(E152C) and IN(E246C). 21-mer substrate (21-S), 19-mer cleavage (19-P) and strand transfer (STP) products are indicated.
Figure 2.3 IN mutant cross-linking efficiency with modified DNA: The upper panel shows the 21-mer U5 viral DNA end substrate used in our studies. The modified nucleotides are in bold and the IN cleavage site is indicated by a vertical arrow. (A) Cross-linking efficacy of IN(E152C) to the following modified DNAs: 1) no DNA control, 2) DNA(C2), 3) DNA(G2), 4) DNA(G5), 5) DNA(A7), 6) DNA(G16) and 7) DNA(G19). (B) Cross-linking efficacy of IN(E246C) to the following modified DNAs: 1) no DNA control, 2) DNA(C2), 3) DNA(G2), 4) DNA(G5), 5) DNA(A7), 6) DNA(G16) and 7) DNA(G19). Of note, while E152C and E246C mutations where introduced in the background SSS sequence, these proteins for simplicity are referred in the figure and text as IN(E152C) and IN(E246C).
Figure 2.4 The effects of the modifying reagent on IN function: Lane 1: reactions without any modifications, lane 2: IN was pre-treated with NHS-biotin (IN/NHS-biotin ratio of 1/50) prior to DNA substrate addition. Lane 3, NHS-biotin added after IN-DNA complex formation. 21-mer substrate (21-S),19-mer cleavage (19-P) and strand transfer (STP) products are indicated.
Figure 2.5 Representative segments of MALDI-ToF data: Free IN, IN(E152C)xDNA(G2) and IN(E246C)xDNA(A7) complexes were treated with NHS-biotin (A, B and C). IN peptides as well as the affected residues are indicated. Peaks C1, C2 and C3 are unmodified tryptic peptide peaks of IN, which serve as internal controls.
Figure 2.6. Representative segments of R199 IN MALDI-ToF data: Free IN (A), IN(E152C)xDNA(G2) (B), IN(E246C)xDNA(A7) (C), IN(E246C)xDNA(G16) (D) and IN(E246C)xDNA(G19) (E) complexes were treated with HPG in parallel reactions. The aa189-211 peptide containing HPG modified R199 is indicated. Peak C4 is an unmodified tryptic peptide of IN, which served as an internal control.
Figure 2.7 Representative MALDI-TOF spectra showing specific protection of K14: In parallel experiments IN(E152C)xDNA(G2) and IN(E152C)-ssDNA complexes were subjected to NHS-biotin treatment. The presented spectra indicate that K273 remained surface assessable in both complexes, while K14 was specifically protected in IN(E152C)xDNA(G2) but not IN(E152C)-ssDNA. C5 is an unmodified tryptic peptide of IN, which served as an internal control.
Figure 2.8 Effects of amino acid substitutions on recombinant IN activities: Upper image depicts strand transfer activities. Positions of 21-mer substrate and reaction products (STP) are indicated. Lower image displays 3’-processing activities. The positions of 21-mer substrate, and specific (19-mer) and non-specific (20-mer) products are shown.
Figure 2.9 Model of viral DNA bound to HIV IN: Separate IN subunits comprising the dimeric structure are colored yellow and green. Amino acid contacts identified by mass spectrometric footprinting are labeled, with the substrate DNA (magenta) overlaid. The IN active site residues are highlighted in red.
Figure 2.10  The crystal structure of the CCD-CTD fragment (28): Monomers 1 and 2 are colored in yellow and green, respectively. In monomer 1, the 196-221 $\alpha$-helix is shown in cyan and the asymmetric kink at Gln209 is indicated. In monomer 2, the same helix is depicted in red, the 187-195 loop is in magenta, and side chain of R199 is in blue.
CHAPTER 3
DYNAMIC MODULATION OF HIV-1 INTEGRASE STRUCTURE AND FUNCTION BY CELLULAR LEDGF PROTEIN

3.1 Introduction
Integration of the reverse transcribed RNA genome into a host chromosome is an obligatory step for HIV-1 replication (reviewed in (13)). This process is catalyzed by the retroviral enzyme integrase (IN) in two reaction steps. In the first step, which is called 3’ processing and takes place shortly after the cDNA is made, IN hydrolyzes a GT dinucleotide from each end of the viral DNA. In the second step, IN catalyzes concerted integration of the processed viral DNA ends into chromosomal DNA. The sites of attack on the two target DNA strands are separated by 5 bp, which leads to dissociation of the small double-stranded DNA fragment between the attachment sites. The subsequent repair of the intermediate species by cellular enzymes completes the integration reaction.

HIV-1 IN consists of three distinct structural and functional domains. The N-terminal domain (NTD) (residues 1-50) contains conserved pairs of histidine and cysteine residues that bind zinc (18, 20), which contributes to IN multimerization and its catalytic function (94, 176). The catalytic core domain
(CCD) (residues 51-212) contains three acidic residues, D64, D116 and E152, which play a key role in coordinating active site divalent metal ions (47, 67). The C-terminal domain (CTD) (residues 213-288) also contributes to functional IN multimerization (5, 82). Results of structural biology studies revealed each individual domain as a dimer (20, 47, 49, 67, 108) and more recent two-domain crystal structures comprised of the CCD and CTD (28) or NTD and CCD (166) likewise unveiled dimeric organizations. Functional studies suggested that a dimer of full-length IN could suffice to process each 3’ end, whereas a tetramer is required to integrate both viral DNA ends into chromosomal DNA (62, 69, 100). Efforts to determine the complete IN structure have been impeded by limited protein solubility and/or the inherent flexibility of the three-domain enzyme. Full-length IN interestingly exists as a mixture of monomers, dimers, tetramers, and higher-order species in the absence of DNA (45, 82, 160, 165).

While in vitro analysis with model DNA substrates demonstrated that IN alone could catalyze 3’ processing and DNA strand transfer reactions, the in vivo function of the enzyme is likely to be regulated by a number of viral and cellular proteins. Following the completion of reverse transcription, the newly synthesized cDNA remains associated with several viral proteins and recruits host factors to form the preintegration complex (PIC) (16, 32, 60, 79, 97, 101, 104, 115, 124, 157). Of these, transcriptional co-activator p75, also known as lens epithelium-derived growth factor (LEDGF), is the principal cellular interactor of HIV-1 IN (32, 115, 157). A number of recent studies have indicated that LEDGF is critically important for effective HIV-1 integration and viral replication (42, 77, 105, 162).
RNA interference (RNAi)-mediated knock-down of endogenous LEDGF to below detectable levels resulted in reduction of infection to 3.5% of that observed in the presence of normal cells (105). Similarly significantly reduced levels of HIV-1 infection were detected in LEDGF knockout mouse embryo fibroblasts (118, 147). Expression of recombinant HIV-1 IN in human cells revealed that LEDGF protects the viral protein from proteasomal degradation and tethers it to chromosomal DNA (51, 104, 106, 107, 115). Accordingly, LEDGF primarily functions during HIV-1 infection to tether PICs to active genes during integration (147). In vitro assays with purified recombinant proteins furthermore demonstrated that LEDGF binds directly to IN, which significantly stimulates its enzymatic activities (29, 31, 32, 127, 134, 158, 172).

The N-terminal part of LEDGF contains a PWWP domain, nuclear localization signal, and dual copy of the AT-hook DNA binding motif (reviewed in (55); Figure 3.1). These conserved elements primarily mediate LEDGF association with chromatin (107, 158). An evolutionarily conserved domain in the C-terminal region (residues 347-429) mediates the interaction with IN and was thus termed the IN-binding domain (IBD) (31, 163). The solution structure of the LEDGF IBD and its co-crystallization with the IN CCD has been recently reported (30, 33). Interestingly, the IBD docks into a relatively small cavity at the CCD dimer interface, contacting both IN subunits (30). The importance of the interacting amino acids revealed from the crystal structure has been validated by site directed mutagenesis in the context of full-length recombinant proteins (19, 33, 51, 135, 157) and by the out-growth of resistant viral strains in the presence
of a dominant-interfering LEDGF fragment (77). However, mutagenesis studies have also indicated that full-length IN-LEDGF interactions extend beyond the contacts observed in the co-crystal structure of the isolated domains (115).

We have undertaken a number of innovative biochemical approaches to characterize the structural and mechanistic foundations between the full-length interacting partners, which has revealed a highly dynamic nature for the interactions between free IN subunits. LEDGF moreover strongly stabilized the IN subunit-subunit contacts. Mass spectrometric surface topology studies furthermore uncovered novel protein-protein contacts, which lie outside of the central IBD-CCD co-crystal structure. Mutational analysis confirmed the importance of the identified residues and indicated a strong correlation between IN tetramer formation and high affinity LEDGF binding. These findings provide new insight into how LEDGF modulates HIV-1 IN structure/function, and highlight the potential to exploit the highly dynamic nature of IN subunit interactions as a novel therapeutic target.

3.2 Materials and Methods

3.2.1 Expression Plasmids and Recombinant Proteins: HIV-1 IN proteins were expressed from pKBIN6Hthr, which was derived from pKB-IN6H (28) by replacing amino acids VDKLAAALE upstream from the C-terminal His$_6$ affinity tag with LVPRGSALE (thrombin cleavage site underlined) by PCR-directed mutagenesis. Mutations were also introduced into pKBIN6Hthr using PCR, and the coding regions of plasmids created via PCR were verified by DNA sequencing. Wild-type and mutant IN proteins were purified according to the
previously described procedure (29). Purified recombinant LEDGF, mutant (mt) LEDGF, and IBD (Figure 3.1) were obtained as described previously (31, 33, 115).

3.2.2 IN 3' Processing and DNA Strand Transfer Activities: The \(^{32}\)P labeled 21-mer synthetic double stranded DNA (50 nM) mimicking the U5 viral end sequence was used as substrate. The concentrations of wild type and mutant IN proteins as well as LEDGF and LEDGF IBD included in the reactions are indicated in the figure legends. The reactions were carried out at 37 °C for 1 hr in buffer containing: 50 mM MOPS (pH 7.2), 2 mM β-Mercaptoethanol, 10 mM MnCl\(_2\), 1 mM CHAPS, 50 mM NaCl, and stopped with 50 mM EDTA. Reaction products were subjected to denaturing polyacrylamide gel electrophoresis and visualized using a Storm 860 phosphorimager (Amersham Biosciences).

3.2.3 Concerted Integration Assay: These assays were performed as described previously (134). Briefly, the 972 bp Scal-DraIII restriction fragment from pU3U5 (129) served as donor DNA and was 5' end labeled with \(^{32}\)P ATP and T4 polynucleotide kinase. HIV-1 IN (400 nM) was assembled with the labeled donor substrate (18 nM) in the presence of 20 mM HEPES (pH 7.0), 5 mM DTT, 10 mM MgCl\(_2\), 25 µM ZnCl\(_2\), 100 mM NaCl, 5% DMSO, 10% PEG 6000. Ligands (IBD or LEDGF) were added before pre-incubation for 20 min at room temperature. Reactions (25 µl final volume) were initiated by adding 500 ng of circular target DNA (pGEM, Promega) and the mixtures were incubated for 1 hr at 37 °C. Reactions were stopped by adding 10 mM EDTA, 0.2% SDS, and 1 mg/ml proteinase K. After ethanol precipitation, samples were subjected to 0.6%
agarose gel electrophoresis for 6 hours at 50 V. The gels were dried and the labeled DNA products were detected using the Storm 860 phosphorimagery.

3.2.4 Subunit Exchange Assay: His-tagged IN (1 µM) was pre-incubated with or without ligand (2 µM LEDGF or 2 µM mtLEDGF) in exchange buffer (25 mM HEPES, pH 7.1, 200 mM NaCl, 4% glycerol, 2 mM β-mercaptoethanol) for 30 min at room temperature. Tag-free IN (1 µM) was then added and incubated for the indicated times. Aliquots were then briefly centrifuged 2 min at 1,000 g and supernatants were pulled down by Ni-nitrilotriacetic acid (NTA) resin (GE Healthcare) for 10 min in the presence of BSA (0.1 mg/ml). The IN-bound resin was then washed three times with buffer containing 50 mM HEPES (pH 7.1), 200 mM NaCl, 2 mM MgCl$_2$, 100 mM imidazole, and 0.1% (v/v) Nonidet P40. The bound proteins were subjected to SDS-PAGE separation and visualized by Coomassie stain.

3.2.5 Mass Spectrometric Footprinting: In parallel reactions, free IN and IN+LEDGF were first incubated at room temperature for 30 min and then subjected to treatments at 37 °C with 1 mM N-hydroxysuccinimide (NHS)-biotin for 30 min or 20 mM p-Hydroxyphenylglyoxal (HPG) for 60 min. These concentrations of modifying reagents were chosen because comparative pulldown experiments with untreated and modified IN-LEDGF complexes indicated that under these conditions the integrity of the preassembled protein-protein complex was fully preserved (data not shown). NHS-biotin treatment was carried out in buffer containing 50 mM HEPES (pH 8.0), 150 mM NaCl, 10 mM MgCl$_2$. The HPG modifications were performed in 50 mM HEPES (pH 8.0), 50
mM boric acid, 150 mM NaCl. The reactions were quenched by excess Lys and Arg using free amino acid forms. IN-LEDGF complexes were selectively pulled down using Ni-NTA resin. The bound proteins were separated by denaturing SDS-PAGE and visualized by Microwave Blue stain (Protiga, Gaithersburg, MD). IN bands were excised, destained and subjected to in-gel proteolysis with 0.5 µg trypsin. The tryptic peptides were analyzed with the Axima-CFR MALDI-ToF instrument (Shimadzu) using α-cyano-4-hydroxy-cinnamic acid as a matrix.

3.2.6 Size Exclusion Chromatography: Experiments were performed with a Superdex 200 10/300 GL column (GE Healthcare) at 0.5 ml/min in buffer containing 50 mM HEPES (pH 7.4), 750 mM NaCl, and 10% glycerol. The column was calibrated with the following proteins: conalbumin (75,000 Da), carbonic anhydrase (29,000 Da), ribonuclease A (13,700 Da), and aprotinin (6,500 Da). Proteins were detected by absorbance at 280 nm.

3.2.7 Molecular Modeling: The model of the NTD-CCD tetramer bound to the LEDGF IBD was generated by overlaying the CCDs within PDB structures 2B4J (30) and 1K6Y (166) using the Insight II software package (Accelrys Inc., San Diego) on a Silicon Graphics O2 work station. The constructed model was then energy-minimized by the same software package using the CFF91 force field and steepest descent method.

3.2.8 LEDGF Binding Affinities to Wild-type and Mutant Inhs: LEDGF (50-650 nM) was incubated with 100 nM His-tagged IN (WT or mutant) in binding buffer (50 mM HEPES (pH 7.1), 200 mM NaCl, 2 mM MgCl₂, 100 mM imidazole, 0.1% (v/v) Nonidet P40) for 60 min at room temperature. Samples were then briefly
centrifuged for 2 min at 1,000 g and supernatants were pulled down by Ni-NTA resin for 30 min in the presence of BSA (0.1 mg/ml). The resin was then washed three times with the same buffer and the bound proteins were separated by SDS-PAGE. LEDGF was detected by western blot analysis using a mouse monoclonal LEDGF antibody (BD Biosciences) and quantified using Image software (NIH). Plotting and curve fitting was performed with Origin 8 software (OriginLab). Nonspecific signal was not detected when LEDGF was incubated with Ni-NTA beads in the absence of IN (data not shown).

3.3 Results

We previously reported that LEDGF significantly stimulated the in vitro activities of HIV-1 IN whereas the isolated IBD failed to do so (31, 158). As these assays utilized relatively long blunt-ended viral DNA substrates and DNA strand transfer product formation as read-out, we reanalyzed the effects of these two proteins on IN function using an oligonucleotide-based assay that monitors the formation of 3’ processing and DNA strand transfer reaction products on denaturing sequencing gels (Figure 3.2A) (17). The results in panel B (lanes 1-6) revealed stimulation of IN DNA strand transfer activity by the LEDGF IBD under these assay conditions. It should be noted that in this setting the 19-mer 3’ processing reaction product is the substrate for the second catalytic step (Figure 3.2A). To dissect if the IBD directly enhanced IN 3’ processing activity, a selective IN strand transfer inhibitor (154, 174) was included in the experiment (Figure 3.12). The 19-mer reaction product accumulated under these conditions, revealing significant stimulation of IN 3’ processing activity by the LEDGF IBD.
For control experiments we analyzed the D366N point mtLEDGF, which is defective for IN binding in vitro (33) and in yeast cells (135), but retains LEDGF DNA binding activity (Figure 3.1). In contrast with the LEDGF IBD, formation of 3’ processing and DNA strand transfer reaction products decreased with increasing mtLEDGF concentrations (Figure 3.2B, lanes 7-12), probably due to competition between mtLEDGF and IN for binding to DNA. A bell-shape enhancement of IN activities was observed in the presence of wild type LEDGF (Figure 3.2B, lanes 13-18), likely due to the competition with IN for DNA binding at high LEDGF:IN ratios. The IBD, which is the weaker stimulant, could display this effect at relatively high stoichiometry (compare lanes 6 and 18 in Figure 3.2B) because it lacks the N-terminal LEDGF regions that mediate DNA binding (158) (Figure 3.1).

We next analyzed the activity of the IBD using a longer donor DNA substrate and a second circular target DNA, a design that can distinguish the formation of single-end half-site (HS) integration products from those that form by the pairwise concerted integration of two viral DNA ends (labeled full-site (FS) in Figure 3.3A) (127, 134). As expected (127, 134), wild type LEDGF modestly stimulated FS and HS product formation at LEDGF:IN ratios of <1, while higher LEDGF concentrations selectively inhibited FS product formation (Figure 3.3B, lanes 7-9). The selective reduction of FS product formation was also observed with increasing concentrations of IBD (Figure 3.3B, lanes 4-6).
These experiments (Figures 3.2, 3.3, and 3.12) demonstrated that direct binding of LEDGF or its IBD could differentially influence IN activities. The co-crystal structure did not show any significant changes in the tertiary structure of the CCD upon IBD binding, but did reveal that the host factor engaged both monomers of the IN dimer at the CCD interface (30). Therefore, one possible mechanism could be that LEDGF binding influences the dynamics of IN subunit-subunit interactions. Indeed, prior analyses using IBD-based peptides hinted at this possibility (3, 71). We therefore devised the following experiment to analyze the dynamics of IN subunit exchange (Figure 3.4A). His-tagged IN (the IN2-IN2 multimer) was mixed with tag-free IN (IN1-IN1). Three IN populations could then form upon subunit exchange: IN1-IN1, IN1-IN2, and IN2-IN2. Of these, IN2-IN2 and IN1-IN2 could be pulled down by NTA resin through affinity binding with the His-tag, while the tag-free IN1-IN1 would be washed out. Indeed, when incubated separately IN2 was affectively bound by NTA beads, while the interaction was not detected using IN1 (data not shown). However, the IN1 protein was quantitatively recovered following its pre-incubation with IN2 due to the effective exchange of IN protein subunits. Kinetic analyses determined exchange within 10 min of mixing, reflecting the highly dynamic nature of protein-protein interactions between free IN subunits under these conditions (Figure 3.4B, lanes 1-4).

We next asked how the IBD and full length LEDGF would affect IN subunit exchange. For this, tag-free IN1 pre-incubated with LEDGF was then exposed to His-tagged IN2. LEDGF effectively prevented IN subunit exchange
(Figure 3.4B, lanes 5-8). Very similar results were obtained with the IBD (data not shown). In contrast, due to the inability to effectively bind IN, mtLEDGF failed to affect the dynamics of IN subunit exchange (Figure 3.4B, lanes 9-12).

These results indicated that LEDGF or its IBD markedly affected IN subunit-subunit interactions but did not distinguish whether the cofactor stabilizes (“locks”) IN into a specific multimeric state or prevents the multimer from forming by interfering with subunit-subunit interactions. To delineate how the proteins modulated IN subunit exchange, we employed size exclusion chromatography (Figure 3.5). The lowest detectable concentration of IN (2.5 µM) was used to approach those employed above in activity-based assays. Under these conditions, wild type IN exhibited two main peaks with the predominant and minor species corresponding to tetramer and dimer, respectively (Figures 3.5A). The “shoulder” observed to the right of the dimer peak suggests that some monomeric protein was also present under these conditions. The addition of the IBD resulted in a single shifted peak with a retention time consistent with two LEDGF IBD molecules bound to the IN tetramer (Figures 3.5A). Since the CCD used in the crystallographic studies contained the F185K solubilizing mutation, we also assayed the F185K/C280S double mutant IN (dmlIN), which is a soluble version of the full-length protein (82). By contrast to the wild type, dmlIN contained greater quantities of the dimer than tetramer (Figure 3.5B). Similar with the unmutated IN, the IBD shifted the equilibrium
sharply in favor of the dmIN tetramer-IBD complex compared with the dimer dmIN-IBD complex. These experiments demonstrate that the IBD preferentially binds and stabilizes the IN tetramer.

To obtain more detailed information on the interaction between full length IN and LEDGF, we turned to our mass spectrometric protein footprinting approach (Figure 3.6). The method enables the comparison of surface topologies of free protein versus a protein-ligand complex using small chemical amino acid selective modifiers. The concentrations of the modifying reagents are optimized to ensure mild reaction conditions such that the integrity of the protein-ligand complex is preserved. Subsequent SDS-PAGE separation, proteolysis, and mass spectrometric analyses are carried out to reveal surface amino acids readily modified in free protein but shielded from modification by the interacting partner (Figure 3.6). This approach has proved instrumental for studying a number of nucleic acid-protein interactions including the IN-viral DNA complex, and consistently revealed biologically relevant contact amino acids (40, 89, 103, 117, 141, 143, 144, 175). Here, we used the method for the first time to examine protein-protein interactions. For this, in parallel reactions with free IN and pre-formed IN-LEDGF complexes were subjected to treatments with Lys and Arg modifying reagents. Despite tight binding, the solution mixture of IN and LEDGF was likely to contain unliganded proteins in addition to the protein complex. To ensure for selective analysis, we utilized His-tagged LEDGF and tag-free IN proteins. Following chemical modification, unliganded LEDGF and IN-LEDGF complexes recovered using Ni-NTA resin, and in parallel free IN, were separated
by SDS-PAGE. IN amino acids protected from modification by LEDGF binding were then deciphered by comparing the modification patterns obtained with the different species of gel-isolated IN (Figure 3.6). Control experiments utilized mtLEDGF that lacks the ability to bind IN.

Representative mass spectrometric fragments are shown in Figure 3.7, and two distinct sets of modification patterns were readily observed. Three peaks containing modified R107, K186, and K14 were detected with free IN and IN+mtLEDGF but were markedly reduced from the IN-LEDGF complex (Figure 3.7A-C). In contrast, two peaks containing modified R228 and K273, as well as an unmodified residue 265-273 fragment, persisted in all three samples. A detailed summary of the modification patterns is presented in Table 3.1. Of the 12 Lys and 9 Arg residues readily modified in free IN, K14, R107, R166, K186, R187, and K188 were selectively protected by LEDGF binding. These results are fully consistent with previous deletion analyses that revealed the CCD as the primary viral recognition determinant with the NTD donating secondary affinity-enhancing contacts. The CTD, by contrast, was dispensable for high affinity LEDGF-IN binding (115).

The footprinting results were next analyzed in the context of available protein structures to gain further insight into the details of the LEDGF-IN interaction. Protections of R107 and R166 are consistent with available data from unliganded (28, 47, 67, 166) as well as IBD-bound (30) CCD dimers. For example, R166 is part of the so-called α4/5 connector that forms the primary IBD recognition determinant from one of the IN monomers (30). The guanidino group
of R107 furthermore directly participates in IN dimerization through interactions with E85 of the other CCD subunit (30, 47, 67). Intriguingly, our results indicated a relatively rapid exchange between IN subunits, consistent with R107 accessibility in the free protein. In contrast, LEDGF binding stabilized the interacting IN subunits, rendering R107 inaccessible to chemical modification (Figures 3.4 and 3.5).

Our footprinting studies revealed new protein-protein contact residues K14, K186, R187, and K188 in the full-length IN-LEDGF complex. Of these, the latter three were unliganded and freely surface exposed in the IBD-CCD co-crystal structure (30). The co-crystal was comprised of a CCD dimer, while the results of size exclusion chromatography indicated affinity of the IBD for the full-length IN tetramer (Figure 3.5). Interestingly, the two domain NTD/CCD crystal structure revealed interactions between two dimers, with K14 from the NTD of one dimer and K186/R187/K188 from the CCD of another dimer contributing to a tetrameric interface (166). In particular, Figure 3.13 depicts the hydrogen bonding network between the two dimers involving the side chains of K14, K186, and R187. The primary amine of K188 is not directly involved in dimer-dimer interactions. However, two acidic residues (E198 and D25) located at ~4.5 Å from K188 could effectively restrict the access of NHS-biotin to this site (Figure 3.13). Therefore, shielding of K14, K186, R187, and K188 in the context of the complete IN-LEDGF complex could be due to interactions between two IN dimers. Also consistent with a role for the basic triad (K186, R187, K188) in IN tetramerization is the observation that the mutation of the adjacent F185 to Lys
shifted the oligomeric state of IN in favor of a dimer while wtIN under the same conditions was predominantly tetrameric (compare dmIN and wtIN chromatograms in Figure 3.5).

To test our hypothesis that protections of K14, K186, R187, and K188 resulted from LEDGF mediated stabilization of tetrameric IN, we conducted site directed mutagenesis experiments (Figures 3.8-3.10). Single point mutations of the target residues significantly compromised tetramer formation (Figure 3.8). For example, K14A, K186A, and K187A mutants were predominantly dimeric, even at relatively high protein concentration (10 µM) (Figure 3.8 and Table 3.3). The K188A substitution had relatively modest affect on tetramer formation (Figure 3.8) consistent with the lesser role of this residue in direct dimer-dimer interactions (Figure 3.13).

We next examined the effects of these mutations on the LEDGF-IN interaction. For this, increasing concentrations of wild type tag-free LEDGF were incubated with His-tagged IN proteins and the fractions of LEDGF recovered by Ni-NTA pull-down were quantitated (Figures 3.9A and 3.9B). These experiments yielded an apparent $K_d$ of ~200 nM for the interaction between wild type IN and LEDGF. The K14A, K186A, and R187A mutants exhibited significantly reduced affinity for LEDGF, while the K188A protein was relatively more effective at binding (Figures 3.9A and 3.9B). In fact, the comparison of size exclusion chromatography (Figure 3.8) and binding (Figure 3.9A and 3.9B) results suggested a strong correlation between IN tetramer formation and high affinity LEDGF binding.
The HIV-1 IN dimer can suffice to process viral DNA 3’ ends whereas the
tetramer has been implicated in DNA strand transfer activity (62, 69, 100); results
of a separate study using chimera IN however indicated that the tetramer was
required for efficient 3’ processing activity (10). Because the relevant protein-
protein interfaces within the catalytic complexes are for the most part unknown,
we investigated whether the tetrameric interface important for high affinity
LEDGF binding played a significant role in IN catalysis. Impressively, the K14A,
K186A, and R187A substitutions adversely affected 3’ processing and DNA
strand transfer activities (Figure 3.10). In contrast, the wild type level of activities
was observed with the K188A protein. Taken together, our results suggest that
the tetramer interface involving basic residues K14, K186, and R187 is important
for HIV-1 IN 3’ processing and DNA strand transfer activities.

While the NTD-CCD tetramer was observed in a crystal lattice at high
concentrations of the two-domain protein, it is less clear how LEDGF could
promote tetramer formation in solution at significantly lower protein
concentrations. To address this, we superimposed the structures of the CCD-IBD
complex (two IBD molecules bound to the CCD dimer, see (30)) and the two
domain (NTD+CCD) tetramer (166) using molecular modeling (Figure 3.11). The
results indicated asymmetric interactions of the IBD molecules with the
NTD+CCD tetramer. For example, two (colored magenta) of the four IBD
molecules could effectively bridge between the two IN dimers by coordinating the
CCD of one dimer and establishing additional electrostatic interactions with the
NTD of another dimer (Figure 3.11). These additional contacts could contribute to
the high affinity IBD-IN interactions and stabilize the IN tetramer. The other two IBD molecules (colored grey) interacted with the CCD dimer interfaces but could not establish additional charge-charge contacts with the NTDs due to the spatial separation between the IN domains (Figure 3.11). The lack of such interactions may reduce the binding affinity for these IBD molecules. In other words, we propose that the IN tetramer has two high affinity and two lower affinity binding sites for the LEDGF IBD.

3.4 Discussion

The present studies revealed a highly dynamic nature of interactions between IN subunits, which could be essential for its biological function as well as exploited as a novel therapeutic target (see below). The cellular cofactor LEDGF strongly modulated the dynamic structure of HIV-1 IN by stabilizing subunit-subunit interactions. Unlike the published co-crystal structure of the isolated domains that indicated binding of two IBD molecules to the CCD dimer, our experiments with the full length proteins demonstrated the importance of the IN tetramer for high affinity LEDGF binding. These results are consistent with previous findings that endogenous LEDGF protein associated with tetrameric recombinant HIV-1 IN in human cells (32).

Our results also indicate that the stabilized IN tetramer is very effective at catalyzing IN 3’ processing and DNA strand transfer activities (Figures 3.2, 3.3, and 3.10). While free IN exists in solution as a mixture of monomer, dimer,
tetramer, and high-order species, the IBD strongly promoted formation of the IN
tetramer (Figure 3.5) and stimulated IN catalytic function (Figures 3.2, 3.3, and 3.12).

Our footprinting and molecular modeling studies uncovered novel intra-
and inter-protein-protein contacts in the full length IN complex with LEDGF
(Figure 3.7). Site directed mutagenesis experiments confirmed the importance of
K14, K186 and K187 residues for tetramer formation and high affinity LEDGF
binding (Figures 3.8-3.10). Furthermore, the K14A, K186A, and R187A mutations
also compromised IN 3' processing and DNA strand transfer activities. These
results suggest the importance of the common tetramer interface for IN
interactions with DNA substrates and LEDGF. Consistently, our previous
footprinting analysis of the IN-DNA complex revealed a role for K14 in DNA
binding (175). The protections in the nucleoprotein complex could arise from
direct protein-DNA or DNA induced protein-protein interactions (175). The
present studies clarify that K14 could be a critical dimer-dimer contact essential
for effective binding of IN with both DNA and LEDGF. Of note, our findings
corroborate results of a number of virus-based assays that indicated essential
roles for K14, K186, and R187 (9, 22, 110, 128, 156, 175) in HIV-1 infection
whereas the K188A mutation resulted in reduced but reproducible levels of virus
spread (110). Therefore, we propose that the IN tetramer is the biologically
relevant form responsible for catalytic activities and high affinity binding to
LEDGF.
The fact that LEDGF selectively impaired concerted integration of two HIV-1 DNA ends (Figure 3.3, see also (127, 134)) has been rather puzzling. How can this observation be reconciled \textit{in vivo}? The following two scenarios can be considered. Free IN could first assemble onto the viral DNA ends before encountering LEDGF (134). The high degree of flexibility within individual IN subunits (54, 161, 175) as well as their dynamic interplay (Figure 3.4) could be critical for effective assembly of the synaptic complex, where the two catalytic sites position themselves for concerted integration. The preassembled IN tetramer-viral DNA complex would engage LEDGF, which in turn would tether the nucleoprotein complex to active genes without significantly affecting structural arrangements of IN with its DNA substrates. Accordingly, \textit{in vitro} experiments with model DNA substrates and purified proteins indicated the importance of the order of IN binding to DNA and LEDGF for effective concerted integration (127, 134). An alternative possibility is that the viral DNA-IN-LEDGF complex engages an as yet unidentified cellular chromatin factor(s). Binding this hypothetical cellular partner could trigger structural changes within the nucleoprotein complex to enable concerted integration. It has been reported that the IN tetramer catalyzes the sequential joining of the two viral DNA ends to target DNA \textit{in vitro} (100), and that LEDGF strongly facilitates single-end HIV-1 integration (29, 127, 134). Protein-protein interactions at chromatin could prompt LEDGF to partly or fully disengage the PIC to allow IN to regain its flexibility and complete the integration of the second viral DNA end.
Our structure-function studies have significant implications for exploiting the highly dynamic nature of HIV-1 IN as a novel therapeutic target. For example, inhibition of functional IN could be accomplished by the following mechanisms: 1) interfering with subunit-subunit assembly and 2) restricting IN flexibility by “locking” the protein into a functionally compromised multimeric state. Of note, such compounds would be complementary to the clinically approved IN active site inhibitor Raltegravir, as the resistant mutations developed to this drug are significantly distanced from the protein-protein interfaces (116). In general, protein-protein interactions are thought to be rather challenging drug intervention targets for the following reasons. The shape of such interfaces is typically flat and comprises large surface areas (~750-1,500 Å) (reviewed in (6)). However, the subset of interfaces that contribute to high-affinity binding could be significantly smaller and affected by small molecule inhibitors. Indeed, significant progress has been made in discovering a number of protein-protein inhibitors in recent years (reviewed in (6)).

Our results indicate that IN monomers could also be a plausible target for interfacial inhibitors. For example, the fact that IN subunits exchange rapidly (Figure 3.4) provides biological targets for effective binding of small molecules. Consistently, our mass spectrometric footprinting revealed that R107, which is located at the dimer interface, is readily modified by HPG (Figure 3.7). While HPG cannot be viewed as a lead compound, detailed analysis of the CCD structure revealed the following intriguing role for this residue. R107 together with G106 form a bulge that effectively docks into a cavity within the interacting
subunit (Figure 3.10) to stabilize the CCD dimer. The structural dimensions depicted in Figure 3.10 suggest that this site could be selectively targeted by small molecules. Of note, discovery of such structural pockets provide a strong rationale for exploiting IN-IN (Figure 3.10) and IN-LEDGF (30) interactions as therapeutic targets.

An alternative intriguing mechanism could be to restrict the catalytically important flexibility of IN by stabilizing a multimeric state. For example, we showed that the LEDGF IBD forms a stable complex with the IN tetramer, which is active for 3’ processing and single-end DNA strand transfer but not concerted integration \textit{in vitro} (Figure 3.3). Previous reports showed that overexpression of LEDGF IBD proteins in target cells effectively impaired HIV-1 replication (42, 105). Furthermore, the IBD was significantly more effective at suppressing HIV-1 replication in LEDGF-deficient cells (555-fold) as compared to cells containing normal LEDGF levels (~30 fold) (105). These observations cannot be fully explained by a competition between the IBD and endogenous LEDGF for IN binding. Instead, our findings provide mechanistic clues that direct binding of the IBD to the IN tetramer restricts protein flexibility required to form the fully functional DNA strand transfer complex.

While the studies with the LEDGF IBD provide proof-of-concept for a new mechanism of IN inhibition, the main interest is to discover small molecule inhibitors. Such compounds do not have to compete with IN subunit-subunit or IN-LEDGF interactions and overcome large energy barriers created by protein-protein interfaces. Instead, they could specifically exploit the structural “pockets”
present in tetrameric or dimeric proteins to stabilize the interacting subunits and compromise the catalytically essential dynamic structure. Recent reports indicate that IBD-derived short peptides impaired IN catalytic activities in vitro (3, 71) and inhibited HIV-1 infection (71). In fact, earlier observations that certain small molecule inhibitors selectively bind at the IN dimer interface gain particular interest in the context of recent elucidation of structural and mechanistic details of LEDGF-IN interactions. For example, 3,4-dihydroxyphenyltriphenylarsionium bromide has been reported to bind the IN dimer at a site that overlaps the IBD binding pocket (126). Our previous studies revealed highly selective binding of another small molecule, methyl N,O-bis(3,4-diacetoxycinnamoyl)serinate, to the IN dimer at the site that is immediately adjacent to the IBD binding pocket (144). A coumarin-based inhibitor was likewise determined to bind IN nearby the IBD interaction site (4). Further research in this direction may well lead to the discovery of novel types of clinically useful HIV-1 IN inhibitors.
### 3.5 Tables for Chapter 3

<table>
<thead>
<tr>
<th>Modified amino acids</th>
<th>Free IN</th>
<th>IN+LEDGF</th>
<th>IN+mtLEDGF</th>
</tr>
</thead>
<tbody>
<tr>
<td>K7</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>K14</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>R107</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>K111</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>K160</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>R166</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>K186</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>R187</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>K188</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>K215</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>K219</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>R224</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>R228</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>R231</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>K236</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>R262</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>R263</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>K264</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>K266</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>R269</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>K273</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

+, surface residues susceptible to modification.  
-, residues protected from modification due to protein-protein contacts.

**Table 3.1 Summary of mass spectrometric footprinting results**
<table>
<thead>
<tr>
<th>Protein</th>
<th>Elution volume, ml</th>
<th>Estimated MW, Da</th>
<th>Suggested oligomeric state</th>
</tr>
</thead>
<tbody>
<tr>
<td>Free IN (peak Tet)</td>
<td>13.573</td>
<td>102,687</td>
<td>Tetramer</td>
</tr>
<tr>
<td>Free IN (peak Dim)</td>
<td>15.100</td>
<td>50,060</td>
<td>Dimer</td>
</tr>
<tr>
<td>Free LEDGF IBD</td>
<td>17.48</td>
<td>15,116</td>
<td>Monomer</td>
</tr>
<tr>
<td>IN-IBD complex</td>
<td>13.076</td>
<td>129,740</td>
<td>IN tetramer+two IBD</td>
</tr>
</tbody>
</table>

Table 3.2 Summary of size exclusion chromatography results for wild type IN interactions with the LEDGF IBD: Theoretical molecular weights of individual IN and IBD subunits are 32,199 Da and 14,459 Da, respectively.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Elution volume, ml</th>
<th>Estimated MW, Da</th>
<th>Suggested oligomeric state</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild type IN (peak Tet)</td>
<td>13.320</td>
<td>104,113</td>
<td>Tetramer</td>
</tr>
<tr>
<td>K14A (peak Tet)</td>
<td>13.123</td>
<td>107,420</td>
<td>Tetramer</td>
</tr>
<tr>
<td>K14A (peak Dim)</td>
<td>14.855</td>
<td>51,841</td>
<td>Dimer</td>
</tr>
<tr>
<td>K186A (peak Tet)</td>
<td>13.129</td>
<td>107,174</td>
<td>Tetramer</td>
</tr>
<tr>
<td>K186A (peak Dim)</td>
<td>14.848</td>
<td>51,988</td>
<td>Dimer</td>
</tr>
<tr>
<td>R187A (peak Tet)</td>
<td>13.109</td>
<td>108,085</td>
<td>Tetramer</td>
</tr>
<tr>
<td>R187A (peak Dim)</td>
<td>14.897</td>
<td>50,932</td>
<td>Dimer</td>
</tr>
<tr>
<td>K188A (peak Tet)</td>
<td>12.857</td>
<td>120,167</td>
<td>Tetramer</td>
</tr>
<tr>
<td>K188A (peak Dim)</td>
<td>14.872</td>
<td>51,453</td>
<td>Dimer</td>
</tr>
</tbody>
</table>

Table 3.3 Summary of size exclusion chromatography results for wild type and mutant IN proteins
3.6 Figures for Chapter 3

**Figure 3.1 The LEDGF proteins used in our studies:** The upper wild type protein exhibits two distinct chromatin and IN binding activities (green and blue, respectively). The middle D366N mutant (mt) protein is severely impaired for IN binding (33, 135). In contrast, the IBD interacts with IN but lacks the ability to bind DNA and chromatin (107, 158).
Figure 3.2 Effects of LEDGF and the isolated IBD on IN 3’ processing and DNA strand transfer activities: (A) Schematic of oligonucleotide-based integration assay. (B) Experimental results. The positions of 21-mer substrate (21-mer S) and the products of 3’-processing (19-mer P) and strand transfer (STP) reactions are indicated. Lanes 1, 7 and 13: 21-mer DNA substrate; Lanes 2, 8 and 14: 50 nM DNA + 0.5 µM IN. The following concentrations of the indicated LEDGF proteins were included in the reactions: lane 3, 9, and 15: 0.25 µM, lane 4, 10, and 16: 0.5 µM, lane 5, 11, and 17: 1 µM, and lane 6, 12, and 18: 2 µM.
Figure 3.3 Effects of LEDGF and the isolated IBD on IN concerted integration activity: (A) Schematic diagram of concerted integration assay. Integration of one donor (B) DNA into the target plasmid yields half-site (HS) or single-end integration products. Integration of a pair of D ends into both strands of target DNA results in full-site (FS) or concerted integration. The D can otherwise integrate into separate D molecules to form D-D products. (D) Agarose gel image of reaction products. Lane 1: $^{32}$P labeled lambda/HindIII DNA mass markers; lane 2: control reaction without IN; lane 3, reaction with 400 nM IN; lane 4, 5 and 6, contained 0.25 µM, 0.5 µM and 1.0 µM IBD, respectively; lanes 7, 8 and 9 contained 0.25 µM, 0.5 µM and 1.0 µM LEDGF, respectively.
Figure 3.4 Subunit exchange assay using His-tag IN and SDS-PAGE: Design (A) and results (B). (A) Subunit exchange between IN multimers was tested by mixing two different IN proteins: IN1, a tag-free form and IN2, containing the 6xHis tag at its C-terminus. The full length proteins are depicted as dimers (IN1-IN1) and (IN2-IN2). Both proteins contained wild type IN sequences and displayed identical catalytic activities. After mixing, unbound proteins were washed from the resin and bound complexes were analyzed by SDS-PAGE. IN1 and IN2 were clearly separated based on their molecular weight differences. (B) The data in lanes 1-4 indicate that IN2 was able to quantitatively pull-down tag-free IN1. The recovered multimers contained a mixture of IN2-IN2 and IN1-IN2, while tag-free IN1-IN1 was washed out from the Ni-NTA resin. To assess the impact of LEDGF on IN subunit exchange, the IN2 multimer was pre-incubated with LEDGF and then exposed to IN1 (lanes 5-8). The results show that LEDGF interacted with IN2 and effectively prevented IN subunit exchange (lanes 5-8). In contrast, mtLEDGF did not bind IN2 or affect subunit exchange (lanes 9-12). Total amounts of input IN1, IN2, LEDGF, and mtLEDGF proteins are shown in lanes 13, 14, 15 and 16, respectively.
Figure 3.4 (Figure legend on previous page)
Figure 3.5 Size exclusion chromatography of free IN and IN-IBD complexes: 2.5 µM IN and 5 µM IBD were used in these experiments. (A) Chromatograms of wild type IN protein in its free form (top) and complexed with IBD (bottom). (B) Elution profiles of the soluble double mutant (F185K/C280S) IN protein in its free form (top) and complexed with IBD (bottom). Peaks corresponding to tetramer (Tet) IN, dimer (Dim) IN, as well as IBD-bound IN tetramer (Tet+IBD) and dimer (Dim+IBD), are indicated.
Figure 3.6 Schematic presentation of the protein footprinting strategy: The structures of the CCD and the IBD-CCD complex are used for illustration while the experiments were performed with full length IN and LEDGF. In parallel experiments free IN and the IN-LEDGF complex were subjected to treatment by small chemical modifiers (M). Surface residues in free IN and the complex were modified, but the interacting amino acids in the complex were shielded from modification. His-tag LEDGF and tag-free IN proteins were used. Following the modification reactions the complex was pulled down by NTA beads, which enabled recovery of only the LEDGF bound form of IN from the reaction mixture. The interacting proteins were then separated by SDS-PAGE. The IN band was excised and subjected to in-gel proteolysis. Subsequent comparative mass spectrometry (MS) analyses revealed modification patterns in free protein and the complex.
Figure 3.7 Representative segments of the MALDI-ToF mass spectra from the protein footprinting: (A) Free IN was treated with HPG or NHS-biotin. (B) IN was preincubated with mtLEDGF and then exposed to treatments with HPG or NHS-biotin. (C) IN-LEDGF complexes were pre-formed and then exposed to HPG or NHS-biotin treatments. Start and end amino acid numbers of the detected peptide peaks are indicated. IN residues affected by modification are depicted in brackets.
Figure 3.8 Size exclusion chromatography of wild type and indicated mutants: All protein concentrations were 10 µM. The elution time points corresponding to tetrameric and dimeric IN are indicated by arrows.
Figure 3.9 LEDGF binding to wild type and mutant IN proteins: (A) Western blot results for LEDGF binding to wild type and mutant IN proteins. Increasing concentrations of LEDGF were incubated with 100 nM of the indicated IN. The concentrations of LEDGF were as follows: lane 1: 50 nM, lane 2: 100 nM, lane 3: 150 nM, lane 4: 200 nM, lane 5: 250 nM, lane 6: 300 nM, lane 7: 350 nM, lane 8: 400 nM, lane 9: 500 nM, lane 10: 650 nM. (B) Quantitative analysis of the western blot results.
Figure 3.10 Biochemical characterization of IN mutant proteins: 3’ Processing and DNA strand transfer activities of the indicated IN proteins; other labeling is the same as in Figure 3.2B.
Figure 3.11 A model for the IBD bound NTD-CCD tetramer: Two available crystal structures of the NTD-CCD tetramer and the IBD-CCD complex were superimposed. Individual subunits of IN are colored cyan, green, yellow and orange. The acidic residues, which coordinate catalytic metals, in the green and yellow subunits are in spheres and colored red. For clarity only K14 in the green subunit and the basic triad (K186, R187, K188) in the yellow subunit are depicted as spheres. These residues contribute to interactions between the two cyan-green and yellow-orange dimers. The two IBD molecules colored magenta establish symmetrical high affinity interactions, bridging the two dimers together. The lower part of the picture shows that the IBD bound to the CCDs of the yellow-orange dimer is positioned to interact with the α-helix within the green NTD that contains K14. The side chains of D6, E10, and E13 in the green NTD are depicted as red sticks. These amino acids potentially establish charge-charge interactions with the side chains of K401, K402, R404, and R405 (blue sticks) of the IBD helix. The additional two lower affinity binding IBD molecules (grey), which coordinate the CCDs of either cyan-green or yellow-orange dimers, are significantly distanced from the tetramer interfaces.
Figure 3.12 IN 3’ Processing activities in the presence of the diketo acid inhibitor (118-D-24) 4-[3-(azidomethyl)phenyl]-2-hydroxy-4-oxo-2-butenoic acid: (154, 174). All reactions contained 5 µM inhibitor, which selectively impaired IN strand transfer activity and allowed accumulation of the 19-mer 3’ processing reaction product (data not shown). Lane 1: free DNA; lane 2: 50 nM DNA + 0.5 µM IN. The following concentrations of IBD were added to the reactions in lanes 3-6: lane 3: 0.25 µM; lane 4: 0.5 µM; lane 5: 1 µM; and lane 6: 2 µM. The positions of 21-mer substrate (21-mer S) and the product of 3’ processing (19-mer P) are indicated.

Figure 3.13 (next page) The NTD-CCD dimer-dimer interface: (167). The side chains of K14 (lower panel), K186 (upper panel), and R187 (upper panel) are engaged in hydrogen bonding networks (blue dashed lines) to stabilize IN dimer-dimer interactions. The corresponding contact amino acids and distances are indicated. The primary amine of K188 remains unliganded (upper panel) but is located in close vicinity (~4.5 Å) to E198 of the same subunit and D25 of another dimer. The colors and relative orientations of individual subunits are similar to those used in Figure 3.11.
Figure 3.13 (Figure legend on previous page)
Figure 3.10 A potential interaction hotspot at the IN subunit-subunit interface: A and B depict two separate monomers colored green and yellow. R107 and G106 form a bulge structure (A) that docks into the cavity comprised of Y83, I84, E85, A86, V180, and N184 (colored blue) of another subunit (B) to stabilize the IN dimer. The height (H), depth (D), and width (W) for this potential small molecule binding pocket are indicated.
4.1 Introduction

The human protein LEDGF has been identified as a key cellular binding partner of HIV-1 integrase and is essential for effective integration of HIV-1 viral DNA into the host cell genome. The N-terminal end of LEDGF contains a PWWP domain, dual copies of the DNA binding AT hook motif, and a nuclear localization signal (158). This N-terminal domain ensemble has been shown to be responsible for LEDGF interactions with chromatin (145). The C-terminal portion of the protein contains the integrase binding domain (IBD) that binds strongly to HIV-1 IN and numerous cellular proteins (158, 171). The protein uses the polarity of these two functions to act as a bifunctional tether maintaining cargo proteins in close proximity to chromatin. LEDGF is proposed to tether the HIV pre-integration complex to chromatin, thereby steering integration into the active transcription units targeted by LEDGF/p75. While the AT-hook motifs of LEDGF are sufficient to bind strongly to naked DNA, the contribution of the PWWP domain is only evident in the context of chromatinized DNA templates (11, 146).
Although the binding partner of the PWWP domain is unknown, it has been established as critical chromatin interaction motif (107, 146). The PWWP module is present in ~60 eukaryotic proteins and is related in sequence and 3-D structure to a number of other chromatin binding domains collectively known as the Tudor clan domains (119). The PWWP domain of the closely related hepatoma-derived growth factor (HDGF) has been crystallized and the 3-D structure is presumably similar to LEDGF PWWP (113). The defining feature of the HDGF PWWP is a five stranded anti-parallel β barrel structure that features a surface exposed hydrophobic pocket (113). This structure was used by Shun et al to design point mutations in LEDGF PWWP domain that abolished interactions with chromatin (145). These LEDGF PWWP mutants were assumed to deform a corresponding hydrophobic pocket or a nearby positively charged surface. While single or double amino acid substitutions are proven to disrupt interactions with chromatin, it is unclear what specific interaction is being disrupted.

We have hypothesized that the PWWP domain recognizes specific post-translational modifications of histone proteins in chromatin. In chromatin the genomic DNA is organized around an octamer of histones to form a series of nucleosomes (112). Generally, the nucleosome is made up a heterodimer of histones 2A and 2B combined with a histone H3/H4 tetramer. Histone proteins can be modified post translationally in order to rearrange chromatin structure or to serve as landmarks for recruitment of proteins to particular sites (87). Histone amino-terminal tails that protrude from the nucleosome seem to be the most
heavily modified regions of the histone, likely due to their accessibility to modifying enzymes (112). However, numerous modifications have been identified in the structured central regions of the histones and at the protein-DNA interface (173). To test this hypothesis we used a histone peptide array to screen for specific interactions of the LEDGF PWWP domain with modified histone peptides. The library consists of 188 short (21-mer) peptides from histones H3, H2A, H2B, and H4 that are either unmodified or contain common post-translational modifications found in chromatin. This preliminary screen produced several possible interactions which were subjected to further examination to determine the specificity of the interactions.

4.2 Materials and Methods

4.2.1 Expression Plasmids and Recombinant Protein Expression Plasmids and Recombinant Proteins: To express LEDGF PWWP domain the pFT-1 plasmid was altered by site directed mutagenesis to insert sequential stop codons in the coding sequence at the sites corresponding to LEDGF residues 111 and 112. All mutations were introduced using QuikChange XL site-directed mutagenesis kit (Stratagene, La Jolla, CA) according to kit protocol. The resulting plasmid was transformed into BL21(DE3) cells (Stratagene, La Jolla, CA). Cells were grown at 37°C to an OD600 of 0.6 before induction with 500 µM IPTG. Cells were harvested 3 hours post-induction and stored at -80°C. Thawed cell paste was suspended in 50ml HEPES, pH 7.5, 150mM NaCl, 20mM imidazole supplemented with EDTA-free Complete protease inhibitor (Roche) and lysed by
sonication. The resulting lysate was clarified by centrifugation and incubated with Ni-NTA agarose resin (GE Healthcare) at 4 °C for 1 hour with gentle agitation. The resin was loaded onto a gravity column and washed thoroughly with 50ml HEPES, pH 7.5, 150mM NaCl, 20mM imidazole prior to elution with HEPES (pH 7.5), 150mM NaCl, 500mM imidazole. For tag-free preparations of LEDGF PWWP proteins the polyhistidine tag was removed at this stage by overnight incubation with PreScission protease (GE Healthcare) at 4 °C. Peak fractions from the Ni-NTA column were pooled, diluted 1:2 in 50mM HEPES (pH 7.5), loaded on a 5ml heparin HiTrap column (GE Healthcare), with a linear gradient of NaCl from 150-1M. Finally, the proteins were purified by size exclusion chromatography using a 120ml and HiLoad 16/60 Superdex 200 gel filtration column (GE Healthcare) and 50mM HEPES (pH 7.5), 150mM NaCl as the elution buffer. The HP1 chromodomain protein was expressed and purified exactly as above, using an expression vector generously provided by Sepideh Khorasanizadeh.

4.2.2 Histone Peptide Microarrays: The biotinylated histone peptides comprising the library are affixed to streptavidin-coated microtiter plates via an N-terminal biotin (Alta Biosciences, UK). These plates are thoroughly blocked with PBS (pH 7.4), 135mM NaCl, 10ug/ml BSA, 0.1% Tween-20 (blocking buffer) prior to a sixteen hour incubation at 4°C with 100 µL of 5 µM recombinant polyhistidine-tagged LEDGF PWWP (LEDGF residues 1-110) suspended in blocking buffer. After extensive washing with PBS (pH 7.4), 135mM NaCl, 0.1%
Tween-20 (wash buffer) the plates are incubated with 1:3000 dilution mouse α-6xHis 1° antibody (Sigma) in blocking buffer for 1 hour at 20°C. After extensive washing, a 1:1000 dilution of horseradish peroxidase (HRP)-conjugated α-mouse IgG 2° antibody (GE Healthcare) was added to the wells. Wells containing a peptide:PWWP complex are identified spectrophotometrically by the blue color formed during the oxidation of tetramethylbenzidine (TMB) by hydrogen peroxide in the presence of the HRP.

4.2.3 Pull-down Assays: Pull-down of tag-free histone peptides (Gene Med Biosynthesis) with 6xHis-tagged proteins was accomplished by assembly of protein-peptide complexes in 50mM Tris, pH 7.4, 150 mM NaCl, during one hour incubation at room 4°C. The Ni-NTA agarose resin (GE Healthcare) used was washed with 50 mM Tris, pH 7.4, 250 mM NaCl, 2 mM MgCl2. Protein-peptide complexes were spun for 2 minutes at 2000xg to remove precipitate and the supernatant was incubated with the washed resin for 30 minutes at 4°C with gentle agitation. Following the incubation the resin was washed with 50 mM Tris, pH 7.4, 150 mM NaCl, 20 mM imidazole followed by H2O. Protein-peptide complexes were eluted from the resin by the addition of 10 µL of 750 mM imidazole solution followed by vigorous mixing. The eluted proteins and peptides complexes were desalted using a ZipTip C18 tip (Millipore) and visualized by MALDI-ToF mass spectrometry using α-cyano-4-hydroxycinnamic acid as a matrix.

92
4.3 Results

In order to rapidly screen a large number of histone post-translational modifications we employed a library of synthetic peptides affixed to the wells of microtiter plates. In 96-well microtiter plate format, these peptides were exposed to recombinant polyhistidine tagged LEDGF PWWP domain and unbound protein was washed away. The peptide-bound protein was visualized colorimetrically with an enzyme-linked antibody specific to the protein’s polyhistidine tag (Figure 4.1). These preliminary screens revealed a subset of nine potentially interacting peptides (Table 4.1) that were subjected to further examination to eliminate artifactual results. To do this, the peptides were pulled down in vitro with affinity-tagged LEDGF PWWP. The histone peptide H3meK9 was used as a control for these experiments as it is known to interact with the HP1 chromodomain and not with LEDGF PWWP ((80, 81) and unpublished data). This established that our method could identify a specific interaction and discriminate non-specific interactions. In this assay the assembled peptide-protein complexes were immobilized on Ni-NTA agarose resin so that unbound proteins and peptides could be washed away. The complexes were eluted with imidazole solution, desalted, and analyzed by MALDI-ToF mass spectrometry to visualize the bound peptides. This assay was able to reproducibly show that in the control reactions H3meK9 was bound to selectively to HP1-CR and not LEDGF PWWP. Furthermore these assays showed a highly specific interaction of the histone H2B peptide KKRKRSRKESYSVYVY(acetyl-K)VLKQ with LEDGF PWWP (Figure 4.2).
4.4 Discussion and future directions

The peptides identified were found in the central regions of the histone proteins and not at the heavily modified tail regions. While these experiments suggest that LEDGF PWWP indeed binds these modified peptides, the $K_d$ of the interaction remains to be determined. Furthermore, incremental truncation of the peptide is needed to determine the minimal sequence required to maintain the interaction.

The nature of the protein-peptide interface is unknown at this time. We are interested to find out if the point mutations in LEDGF-PWWP that abolished chromatin association (146) would have a similar effect on the binding of this peptide. Of note, the chromatinized templates used by Shun et al for the *in vitro* chromatin binding assays were assembled with a preparation of native HeLa cell histones in which the full complement of post-translational modifications is conceivably represented. In addition to the mutant analysis, the surface topology of the protein-peptide interface can be probed by the mass spectrometric footprinting analysis developed in our lab (122). This method has been used successfully in the past to provide detailed information about protein-protein and protein-nucleic acid interfaces (123, 175). The most convincing structural information would come from a protein-peptide co-crystal structure. While the full-length LEDGF protein has not so far been amenable to crystallization efforts, crystallization of the isolated PWWP domain may be more feasible. Indeed, the highly homologous HDGF PWWP domain has been successfully crystallized (113).
It is of paramount importance to test the significance of the interaction to LEDGF function in vivo. As LEDGF operates as a chromatin tethering protein, chromatin localization using fluorescent-tagged ectopically-expressed protein is likely the best measure of LEDGF function. Cells expressing the fluorescent protein can be permeabilized and treated with test peptide to determine if competition between peptide and chromatin reduces chromatin localization. This strategy has been used successfully in the past to validate the HP1-chromodomain interaction with acetyl-lysine 9 of histone H3 (81). Due to the critical role of LEDGF-chromatin interactions for HIV integration, peptide-treated cells could be treated for sensitivity to HIV infection as another test of biological relevance.
### Tables for chapter 4

<table>
<thead>
<tr>
<th>Well</th>
<th>Sequence</th>
<th>Histone</th>
</tr>
</thead>
<tbody>
<tr>
<td>C5</td>
<td>VKRISGLIYETRGVLKVFLE</td>
<td>H4</td>
</tr>
<tr>
<td>C6</td>
<td>VKRI(S-phosph)GLIYETRGVLKVFLE</td>
<td>H4</td>
</tr>
<tr>
<td>C7</td>
<td>VKRISGLIYETRGVL(K-Ac)VFLE</td>
<td>H4</td>
</tr>
<tr>
<td>E11</td>
<td>LAIRNDEELN(K-Ac)LLGKVTLIAQG</td>
<td>H2A</td>
</tr>
<tr>
<td>F1</td>
<td>LAIRNDEELN(K-Ac)LLG(K-Ac)VTIAQG</td>
<td>H2A</td>
</tr>
<tr>
<td>H6</td>
<td>KKRKRSRKE(S-phosph)YSVYVY(K-Ac)VLKQ</td>
<td>H2B</td>
</tr>
<tr>
<td>H7</td>
<td>KKRKRSRKEYSVYVY(K-Ac)VLKQ</td>
<td>H2B</td>
</tr>
<tr>
<td>G11</td>
<td>VREIAQDF(K-Me2)TDLRFQSSAVMA</td>
<td>H3</td>
</tr>
</tbody>
</table>

Table 4.1 Histone peptides selected for further study
4.5 Figures for chapter 4

Figure 4.1 (next page) Representative peptide screening results: (A) H2/H4 library (B) H3 library. A library of biotinylated modified and unmodified histone peptides were immobilized on a streptavidin plates and assayed for interaction with recombinant, polyhistidine-tagged LEDGF PWWP domain. Peptides binding to LEDGF PWWP in this assay are identified by formation of a blue color (see materials and methods). Peptides selected for further study are listed in table 4.1.
Figure 4.1 (Figure legend on previous page)
Figure 4.2 MS data from *in vitro* pull-down assays: Test peptides were incubated with polyhistidine-tagged LEDGF PWWP domain. Ni-NTA agarose resin was used to immobilize tagged protein and unbound peptide was washed away. Protein-peptide complexes were eluted from the resin with imidazole solution, desalted, and subjected to analysis by MALDI-ToF MS. (A) The histone peptide KKRKRSRKE SYSVYVY(K-Ac)VLKQ. (B) Background binding of the peptide to Ni-NTA resin. (C) Binding of the peptide to LEDGF PWWP. (D) Binding of peptide to the HP1 chromodomain.


24. Carteau, S., R. J. Gorelick, and F. D. Bushman. 1999. Coupled integration of human immunodeficiency virus type 1 cDNA ends by


148. **Sinha, S., and D. P. Grandgenett.** 2005. Recombinant human immunodeficiency virus type 1 integrase exhibits a capacity for full-site...


