Structural And Biochemical Studies Of HIV-1 Integration And Inhibition

Grant Eilers
University of Pennsylvania

Follow this and additional works at: https://repository.upenn.edu/edissertations

Part of the Biochemistry Commons

Recommended Citation
Eilers, Grant, "Structural And Biochemical Studies Of HIV-1 Integration And Inhibition" (2020). Publicly Accessible Penn Dissertations. 3917.
https://repository.upenn.edu/edissertations/3917

This paper is posted at ScholarlyCommons. https://repository.upenn.edu/edissertations/3917
For more information, please contact repository@pobox.upenn.edu.
Structural And Biochemical Studies Of Hiv-1 Integration And Inhibition

Abstract
The human immunodeficiency virus (HIV) affects millions of people worldwide who rely on antiretroviral therapy to prevent the acquired immunodeficiency syndrome and halt further HIV transmission. HIV integrase (IN), one of the three retroviral enzymes, catalyzes the covalent insertion of a DNA copy of the HIV genome into host cell chromatin, an essential step in the viral replication cycle. Integration enables expression of a new generation of viral RNA genomes and subgenomic RNAs encoding viral proteins, and establishes the potential for latency, a major barrier to cure. Despite being considered initially to be "undruggable", IN has been successfully targeted with small-molecule therapeutics, two classes of which are studied here. The strand transfer inhibitors (STIs), which block an essential step in IN catalysis, are FDA-approved and in widespread clinical use. Clinical resistance to STIs has been documented, motivating the design of new generations of STIs with high genetic barriers to resistance and novel approaches to targeting IN. The allosteric inhibitors of integrase (ALLINIs) bind to a site distinct from STIs and do not directly block catalysis, instead act by aberrantly polymerizing IN, disrupting virion maturation. The binding interface of ALLINIs and a structural explanation of their mechanism of action was recently reported by our group. Accurate structural and biochemical data are essential for ongoing drug development and efforts to understand mechanisms of resistance. Here, we report structural and biochemical advances that further our understanding of antiretrovirals targeting HIV IN. We have improved the resolution of structural models of IN-ALLINI polymers, extended these data to multiple members of this class of compounds, and revealed mechanisms of resistance. We have identified a promising clinical opportunity combining ALLINIs and STIs that exploits the hypersensitivity of STI-resistant IN to ALLINI inhibition. Through the creation of an improved in vitro model, we report structural and biochemical data that most closely recapitulates the form of IN found in vivo. Finally, we report progress toward structural and functional characterization of the catalytically active complex of IN and viral DNA, the intasome. Together, this body of work advances our understanding of a key step in the pathogenesis of HIV and provides a foundation for improvements to therapeutics.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Cell & Molecular Biology

First Advisor
Frederic D. Bushman

Second Advisor
Gregory D. Van Duyne

Keywords
Antiretroviral, Crystallography, HIV, Integrase

Subject Categories
Biochemistry

This dissertation is available at ScholarlyCommons: https://repository.upenn.edu/edissertations/3917
STRUCTURAL AND BIOCHEMICAL STUDIES OF HIV-1 INTEGRATION AND INHIBITION

Grant Eilers

A DISSERTATION

in

Cell and Molecular Biology

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2020

Supervisor of Dissertation

Frederic D. Bushman, Ph.D., William Maul Measey Professor in Microbiology

Co-Supervisor of Dissertation

Gregory D. Van Duyne, Ph.D., Jacob Gershon-Cohen Professor of Medical Science

Graduate Group Chairperson

Daniel S. Kessler, Ph.D., Associate Professor of Cell and Developmental Biology

Dissertation Committee

Rahul Kohli, M.D., Ph.D., Associate Professor of Medicine

Mitchell Lewis, D. Phil., John Morgan Professor of Biomedical Research and Education

Paul Bates, Ph.D., Professor of Microbiology

Robert Doms, M.D., Ph.D., Professor of Pathology and Laboratory Medicine
ACKNOWLEDGMENTS

I am indebted to many people for making this thesis possible. To my mentors, Dr. Frederic Bushman and Dr. Gregory Van Duyne, I am grateful for the vibrant collaborative scientific environment you have fostered. Dr. Bushman’s enduring enthusiasm and strategic vision for science are inspiring. I will miss his scientific anecdotes that span the depths of the oceans to outer space. Dr. Van Duyne perfectly combines deep scientific knowledge with the ability to teach and tell a story, making the complicated world of structural biology and biophysics come to life. It has truly been a privilege to work under the guidance of these two mentors.

I am also grateful to my thesis committee. Dr. Rahul Kohli sets the standard for a successful physician scientist, and I could not ask for a better career role model. He was actively engaged in every committee meeting, asking astute questions and offering excellent advice. The many positive encounters I have had with members of his lab are a further testament to the quality of his mentoring. Dr. Mitchell Lewis taught my favorite class in medical school, and I am grateful that the path of my research brought me back into his orbit. His ongoing encouragement to write this thesis was indispensable, and frequent discussions about aerobic training and bicycles were essential for my wellness during graduate school. I want to thank Dr. Paul Bates for his depth of scientific understanding that seemingly encompasses the entirety of virology and his positivity and encouragement throughout the last four years. I also want to thank Dr. Robert Doms for his teaching in medical school, his mentorship as I was choosing a lab and thesis project, and his brightness and enthusiasm in committee meetings.
I particularly want to acknowledge the Medical Scientist Training Program, and Dr. Skip Brass and Maggie Krall in particular, for support and guidance on both an academic and personal level that has been essential throughout graduate school. Skip and Maggie are also responsible for selecting the wonderful group of my fellow students in the training program who share a wealth of scientific knowledge as well as support and friendship.

I want to thank the members of the Van Duyne and Bushman labs for training, advice, assistance, and friendship. I especially want to recognize the following individuals that have worked most closely with me on this thesis: Kushol Gupta, for fielding countless questions, sharing his incomparably broad and deep knowledge of integrase, and guiding me through the complicated world of crystallography software, and Young Hwang, for his advice and training which were instrumental in making me a more rigorous and efficient scientist.

On a personal note, I want to thank my family and friends for encouragement and support throughout the years of work on this thesis. It is important that I recognize the contributions of a fleet of bicycles that reliably carried me to lab and provided silent therapy in difficult times. And finally, a special thanks to Melanie Weingart, who has endured scientific ramblings and provided emotional support, and who, I hope, will continue to for the rest of my life.
ABSTRACT

STRUCTURAL AND BIOCHEMICAL STUDIES OF HIV-1 INTEGRATION AND INHIBITION

Grant Eilers

Frederic D. Bushman and Gregory D. Van Duyne

The human immunodeficiency virus (HIV) affects millions of people worldwide who rely on antiretroviral therapy to prevent the acquired immunodeficiency syndrome and halt further HIV transmission. HIV integrase (IN), one of the three retroviral enzymes, catalyzes the covalent insertion of a DNA copy of the HIV genome into host cell chromatin, an essential step in the viral replication cycle. Integration enables expression of a new generation of viral RNA genomes and subgenomic RNAs encoding viral proteins, and establishes the potential for latency, a major barrier to cure. Despite being considered initially to be “undruggable”, IN has been successfully targeted with small-molecule therapeutics, two classes of which are studied here. The strand transfer inhibitors (STIs), which block an essential step in IN catalysis, are FDA-approved and in widespread clinical use. Clinical resistance to STIs has been documented, motivating the design of new generations of STIs with high genetic barriers to resistance and novel approaches to targeting IN. The allosteric inhibitors of integrase (ALLINIs) bind to a site distinct from STIs and do not directly block catalysis, instead act by aberrantly polymerizing IN, disrupting virion maturation. The binding interface of ALLINIs and a structural explanation of their mechanism of action was recently reported by our group. Accurate structural and biochemical data are essential for ongoing drug development and efforts to understand mechanisms of resistance. Here, we report structural and biochemical advances that further our understanding of antiretrovirals targeting HIV IN. We have improved the resolution of structural models of IN-ALLINI polymers, extended
these data to multiple members of this class of compounds, and revealed mechanisms of resistance. We have identified a promising clinical opportunity combining ALLINIs and STIs that exploits the hypersensitivity of STI-resistant IN to ALLINI inhibition. Through the creation of an improved in vitro model, we report structural and biochemical data that most closely recapitulates the form of IN found in vivo. Finally, we report progress toward structural and functional characterization of the catalytically active complex of IN and viral DNA, the intasome. Together, this body of work advances our understanding of a key step in the pathogenesis of HIV and provides a foundation for improvements to therapeutics.
# TABLE OF CONTENTS

ACKNOWLEDGMENTS ii

ABSTRACT iv

TABLE OF CONTENTS vi

LIST OF ILLUSTRATIONS ix

LIST OF TABLES x

Chapter 1: Introduction 1
  1.1 Human immunodeficiency virus and antiretroviral therapy 1
  1.2 HIV integration and integrase 2
  1.3 Intasome structure 4
  1.4 Integrase strand transfer inhibitors 7
  1.5 LEDGF and allosteric integrase inhibitors 9
  1.6 Thesis objectives 13

Chapter 2: Structural details of allosteric HIV integrase inhibitor binding and resistance 15
  2.1 Abstract 15
  2.2 Introduction 16
  2.3 Results 21
    2.3.1 Structures of replication competent INs containing resistance substitutions with the ALLINI BI-D 21
    2.3.2 Purification of individual IN domains 24
    2.3.3 Interaction of INCD and INCTD in vitro 25
    2.3.4 Co-crystallization of INCD, the ALLINI BI-224436, and INCTD 26
    2.3.5 Structure of the INCD·BI-224436·INCTD complex 28
  2.4 Discussion 29
  2.5 Acknowledgments 33

Chapter 3: Complementary activity of a high-potency ALLINI against strand transfer inhibitor-resistant integrase 34
  3.1 Abstract 34
  3.2 Introduction 34
  3.3 Results 37
    3.3.1 Aggregation of integrase by the ALLINI ViiV-919 37
    3.3.2 Inhibition of wild type integrase activity by ViiV-919 37
    3.3.3 ViiV-919 activity against dolutegravir-resistant integrase 39
    3.3.4 Combined activity of raltegravir and ViiV-919 41
  3.4 Discussion 42
  3.5 Acknowledgments 45

Chapter 4: Influence of the amino-terminal sequence on the structure and function of HIV integrase 46
  4.1 Abstract 46
LIST OF ILLUSTRATIONS

Figure 1.1. Intasome Structure and Diversity 5
Figure 1.2. Structure of IN\_\text{CCD} with LEDGF\_IBD and the ALLINI BI-224436. 10
Figure 1.3. Structure of IN-ALLINI Polymer 12
Figure 2.1. Structure of IN-ALLINI Polymers 24
Figure 2.2. ALLINI-induced aggregation of isolated IN\_\text{CCD} and IN\_\text{CTD} domains 26
Figure 2.3. Structure of IN\_\text{CCD} BI-224436-IN\_\text{CTD} 27
Figure 2.4. Interaction network of IN-ALLINI binding interface 28
Figure 3.1. Aggregation of IN by ALLINIs BI-224436 and ViV-919 38
Figure 3.2. Effect of Integrase Inhibitors on 3\textsuperscript{-}Processing 38
Figure 3.3. Effect of Integrase Inhibitors on Strand Transfer 39
Figure 3.4. Effect of ViV-919 on 3\textsuperscript{-}processing Activity of Dolutegravir-Resistant IN 40
Figure 3.5. Effect of ViV-919 on Strand Transfer Activity of Dolutegravir-Resistant IN 40
Figure 3.6. Effect of Combining Raltegravir and ViV-919 41
Figure 4.1. Structure of HIV-1 integrase with a native N-terminus 54
Figure 4.2. 3\textsuperscript{-}processing activity in vitro 57
Figure 4.3. Strand transfer activity in vitro 59
Figure 4.4. ALLINI-induce aggregation of IN 1F, IN GSH, and IN MF 61
Figure 4.5. Biophysical analysis of IN 1F and IN MF 63
Figure 5.1. Co-purification of IN 1F with LEDGF\_IBD 73
Figure 5.2. LEDGF\_IBD in complex with IN containing a native N-terminus 75
Figure 5.3. IN\_LEDGF\_IBD activity 76
Figure 5.4. Solubility limits of IN and IN\_LEDGF complexes 77
Figure 5.5. Covalent crosslinking of 3WJ intasome preparations 79
Figure 5.6. Agarose electrophoretic mobility shift assays of intasome preparations 80
Figure 5.7. Polyacrylamide electrophoretic mobility shift assays of intasome preparations 82
Figure 6.1. Role of the IN\_\text{CTD} in ALLINI-induced aggregate formation in vivo 90
Figure 6.2. ALLINI chemotypes and structural diversity 96
Supplemental Figure 2.1. Structure of ALLINIs used in this study 102
Supplemental Figure 2.2. Representative images of crystals 102
Supplemental Figure 2.3. Domains contributing to ALLINI binding 103
Supplemental Figure 2.4. Models of IN-ALLINI polymerization in crystallo 103
Supplemental Figure 2.5. Detail of CTD face involved in ALLINI binding 104
Supplemental Figure 3.1. Strand transfer activity of wild type and dolutegravir-resistant IN constructs 107
Supplemental Figure 3.2. Location of STI resistance substitutions 108
Supplemental Figure 4.1. Purification scheme and SDS-PAGE analysis of protein purifications 116
Supplemental Figure 4.2. Models demonstrating two possibilities of NTD orientation 116
Supplemental Figure 4.3. Secondary structure comparison of IN structures 117
Supplemental Figure 4.4. NaCl-dependence of strand transfer activity 118
Supplemental Figure 4.5. Effect of raltegravir on strand transfer activity of IN 1F, IN GSH, and IN 1F 118
Supplemental Figure 4.6. Biophysical and biochemical characterization of IN 1F\textsuperscript{F185H} 119
Supplemental Figure 5.1. Intasome DNA templates 124
LIST OF TABLES

Table 2.1. Crystallographic Statistics 23
Table 4.1. Crystallographic Statistics 55
Table 4.2. Properties determined by sedimentation equilibrium analysis 63
Chapter 1: Introduction

1.1 Human immunodeficiency virus and antiretroviral therapy

The origin of the human immunodeficiency virus (HIV) can be clearly traced to zoonotic transmission from non-human primates to humans in the early 1900s (Keele et al. 2006; Faria et al. 2014). HIV spread undetected until the 1980s, when clusters of cases of unexplained immunodeficiency (Gottlieb et al. 1981; Masur et al. 1981) launched a search for a causative agent. The virus was identified shortly thereafter (Barré-Sinoussi et al. 1983; Gallo et al. 1983), and determined to be the cause of the acquired immune deficiency syndrome (AIDS) through infection and depletion of CD4+ T cells and systemic immune dysfunction (Moir, Chun, and Fauci 2011; McCune 2001). HIV has infected approximately 75 million people worldwide (“UNAIDS Data 2019” 2019), and approximately 32 million people have died from AIDS-related illnesses. It remains a major cause of morbidity and mortality today (Roth et al. 2018).

The first antiretroviral drug, zidovudine (3’-azido-3’-deoxythymidine or AZT) inhibited reverse transcription and was approved in 1987 (Mitsuya et al. 1985; Fischl et al. 1987). It was quickly recognized that a single drug would not effectively control HIV (Larder, Darby, and Richman 1989), spurring the development of combination regimens (Hammer et al. 1996) that are the current paradigm of antiretroviral therapy (Arts and Hazuda 2012). An arsenal of antiretroviral therapeutics targeting multiple steps of the viral replication cycle are available today, and play a major role in mitigating the effects of HIV. The last few years have seen a reduction in the number of new HIV infections and AIDS-related deaths (“UNAIDS Data 2019” 2019), underscoring the success of
antiretroviral therapy. However, many HIV infections remain undiagnosed and millions of people still lack access to antiretroviral therapy, emphasizing the importance of continued research and development in this area.

Although antiretroviral drugs have been developed to target nearly every step in the viral replication cycle, the most commonly used drugs target the three viral enzymes: protease, reverse transcriptase, and integrase (Arts and Hazuda 2012). Although integrase was the last enzyme to be successfully targeted (D. J. Hazuda, Anthony, et al. 2004; Espeseth et al. 2000; D. J. Hazuda, Young, et al. 2004), integrase inhibitors are now in widespread clinical use (A. N. Engelman 2019; Mesplède and Wainberg 2013), and are included in the recommended initial ART regimens (Günthard et al. 2016; “AIDSinfo” 2019; World Health Organization 2018). Research directed toward antiretroviral therapy targeting HIV integrase (the topic of this thesis) will directly benefit the ongoing battle against the morbidity and mortality of the HIV/AIDS epidemic.

1.2 HIV integration and integrase

After entering a host cell and completing the process of reverse transcription, the HIV genome enters the nucleus to be integrated into host chromatin (Robert Craigie and Bushman 2012). This process is an essential step in the retroviral replication cycle, allowing for the production of viral proteins and the transcription of viral genome copies to form another generation of virions (Varmus 1988). The integrated provirus is replicated along with the host cell genome, establishing the potential for latency, a major barrier to the cure of HIV.

Integration is catalyzed by the retroviral integrase enzyme in two steps, 3’ processing and strand transfer (Fujiwara and Mizuuchi 1988; Brown et al. 1989; A.
Engelman, Mizuuchi, and Craigie 1991). In the first step, 3’-processing, integrase cleaves two nucleotides from the 3’ end of the viral DNA, leaving the conserved 5’-CA-3’ sequence. In the second step, the newly exposed 3’ hydroxyl group attacks the phosphate backbone of host cell DNA, leading to the covalent insertion of the viral DNA. Both reactions are transesterification reactions breaking phosphodiester bonds through nucleophilic attack (Delelis et al. 2008). In 3’-processing, the phosphodiester bond is in the viral DNA and the nucleophile is a water molecule. In strand transfer, the phosphodiester bond is in the host cell DNA, and the nucleophile is the newly-exposed 3’ hydroxyl group of the viral DNA. Strand transfer must occur at both viral DNA ends for productive integration, a process termed “concerted” integration. The resulting symmetrical gaps and overhangs are repaired by host cell enzymes (Brin et al. 2000). Both steps of the integration reaction can be reproduced in vitro (Leh et al. 2000; M. Li et al. 2006; M. Li and Craigie 2005; Sinha, Pursley, and Grandgenett 2002; Sinha and Grandgenett 2005; Bushman and Craigie 1991; Bushman, Fujiwara, and Craigie 1990; Hindmarsh et al. 1999). Integrase has additional roles in the viral replication cycle, and likely many more that remain undiscovered, but they are not covered in detail here. One particular auxiliary role, relevant for virion maturation, will be mentioned in a subsequent section.

The integrase enzyme is produced by proteolysis of the Gag-Pol polyprotein by the retroviral protease. Its 288 amino acids fold into three domains: the N-terminal (NTD), catalytic core (CCD), and C-terminal (CTD) domains (X. Li et al. 2011; A. Engelman and Craigie 1992). The NTD adopts a zinc-finger fold through a conserved and functionally essential HHCC motif (Mengli Cai et al. 1997; Zheng, Jenkins, and Craigie 1996; Z. Zhao et al. 2008). The CCD contains the active site, a D,D-35-E motif
within an RNase H-like fold that coordinates divalent metal ions (A. Engelman and Craigie 1992; Kulkosky et al. 1992; Dyda et al. 1994; Wielens et al. 2010). Integrase activity depends on the presence of the divalent metal ionic cofactor (Mg$^{2+}$ or Mn$^{2+}$), coordinated by the D,D-35-E motif (Goldgur et al. 1998; Maignan et al. 1998). The CCD also contributes to DNA binding (T. M. Jenkins et al. 1997; Esposito and Craigie 1998; Heuer and Brown 1997; Drake et al. 1998; A. A. Johnson et al. 2006). The CTD binds DNA through its SH3-like fold (J. C.-H. Chen et al. 2000; Eijkelenboom et al. 1999). Dimers of the NTD and CTD have been observed (Mengli Cai et al. 1997; Eijkelenboom et al. 1999, 1995; Lodi et al. 1995), but the extensive CCD dimerization interface is the best characterized and is expected to dictate the state of IN in vivo (Dyda et al. 1994). However, integrase has been reported to exist in a wide range of multimeric forms (Hare, Di Nunzio, et al. 2009; A. Engelman, Bushman, and Craigie 1993; Peter Cherepanov et al. 2003; Faure et al. 2005; McKee et al. 2008; Pandey, Bera, and Grandgenett 2011). Importantly, although a CCD dimer carries two active sites, the spacing between these sites is incompatible with the spacing between sites of concerted integration. Long before molecular models were available, the active form of integrase was suspected to be a tetramer at minimum (Faure et al. 2005; M. Li et al. 2006; Bera et al. 2009; Hare, Di Nunzio, et al. 2009), with some evidence of larger forms (Bera et al. 2009; Heuer and Brown 1998; Peter Cherepanov et al. 2003).

1.3 Intasome structure

The active form of integrase was initially characterized through purification of viral replication intermediates from infected cells (Farnet and Haseltine 1990; Ellison et al. 1990; Bowerman et al. 1989). These intermediates, termed preintegration complexes.
(PICs) contain multiple subunits of integrase, the viral DNA, and several other viral and host cell proteins (M. D. Miller, Farnet, and Bushman 1997; Bukrinsky et al. 1993; Nermut and Fassati 2003). However, examination of PICs, integration reaction products, and the structure of isolated domains did not elucidate the molecular structure of the active integrase complex, although many models were proposed (Chiu and Davies 2004; Jaskolski et al. 2009). Determination of the structure of the integrase-DNA complex (the intasome) from the prototype foamy virus (Hare, Gupta, et al. 2010; G. N. Maertens, Hare, and Cherepanov 2010) was a major breakthrough in the understanding of integrase structure and function (Figure 1.1). The prototype foamy virus intasome is a tetramer of integrase subunits assembled on the two viral DNA ends. The CCD dimerization interface is essentially identical to that observed with the isolated domains, so the intasome forms what is essentially a dimer-of-dimers. The complex consists of extensive protein-DNA and protein-protein contacts, offering a structural context to explain the regions of integrase implicated in DNA binding (T. M. Jenkins et al. 1997;
Esposito and Craigie 1998; Heuer and Brown 1997; Drake et al. 1998; A. A. Johnson et al. 2006). Although the core of the intasome (the region immediately surrounding the catalytic sites) is well-resolved, domains located distally are not fully resolved, leaving unanswered questions about their function.

Now that the structures of other retroviruses have been determined, the prototype foamy virus intasome can be viewed as the simplest assembly of integrase subunits on the viral DNA ends (Ballandras-Colas et al. 2017). Subsequent structures revealed higher order forms present in viruses closely related to HIV (Figure 1.1) (Ballandras-Colas et al. 2016; Yin et al. 2016; Ballandras-Colas et al. 2017; Cook et al. 2020), culminating in the determination of the HIV intasome itself (Passos et al. 2017, 2020). What remained relatively constant across all retroviral intasomes was the core intasome structure that contains the active sites, important for understanding the function of the strand transfer inhibitors (which will be discussed in more detail later) (A. N. Engelman and Cherepanov 2017). Due to differences in domain architecture and intra-domain linker length, however, the domains comprising the core intasome structure are contributed by different integrase subunits across retroviruses.

The HIV-1 intasome has proven most recalcitrant to structural determination, in spite of significant effort due to its clinical importance. In contrast to other retroviruses, study of the HIV-1 intasome has been met with significant challenges. HIV-1 integrase suffers from poor solubility and low activity in vitro (Peter Cherepanov et al. 2003; Hare, Shun, et al. 2009; M. Li et al. 2014). Intasome preparations have resulted in difficult-to-interpret heterogeneous results (Passos et al. 2017; M. Li et al. 2014). These difficulties have been bypassed though the fusion of a small DNA-binding domain to the NTD of integrase (M. Li et al. 2014), allowing for determination of the first HIV-1
intasome structures (Passos et al. 2017, 2020). The best-resolved intasome species is a tetramer, which is not surprising since the fusion of a DNA-binding domain to the N terminus of integrase mimics the N-terminal extension domain present in the prototype foamy virus integrase, which forms tetrameric intasomes. Additional higher-order intasomes were also observed in the HIV-1 data, with stoichiometries similar to that of the maedi-visna virus intasome, which consists of 16 subunits of integrase (Ballandras-Colas et al. 2017). The core intasome structure of HIV-1 is highly similar to the prototype foamy virus, facilitating the optimization of the strand transfer inhibitors, but there is still an unmet need for structures of the native HIV-1 intasome.

1.4 Integrase strand transfer inhibitors

The initial development of integrase inhibitors was carried out in the absence of molecular models of the intasome (Eric Deprez et al. 2004; D. J. Hazuda et al. 2000; Egbertson 2007; Semenova, Marchand, and Pommier 2008). Due to their primary effect on strand transfer, and comparatively minor effect on 3’ processing, these compounds were termed strand transfer inhibitors (STIs), and presumed to bind to integrase in the strand transfer complex, a form of the intasome that includes both the viral DNA and the host cell DNA. Not until the prototype foamy virus intasome structure was revealed was the mechanism of action of STIs made clear (Hare, Gupta, et al. 2010; G. N. Maertens, Hare, and Cherepanov 2010; Hare, Vos, et al. 2010). These compounds bind in the active site, coordinate the catalytic divalent metal cations, and displace the 3’ end of the viral DNA, preventing nucleophilic attack of the host cell DNA. Due to differences in domain architecture, intasome stoichiometry, and sensitivity to STIs between prototype foamy virus and HIV (Grawenhoff and Engelman 2017; Valkov et al. 2009), significant
effort was invested in obtaining equivalent HIV intasome and intasome-inhibitor structures. Ultimately, these structures were obtained through protein engineering and Cryo-EM (Passos et al. 2017, 2020; M. Li et al. 2014). The HIV structures differ slightly from the prototype foamy virus structures, although, structurally, the active site is highly conserved. The HIV intasome structures illuminate pathways for increasing drug potency and combating mechanisms of resistance.

STIs have been successful clinically (A. N. Engelman 2019; Mesplède and Wainberg 2013) and now make up the first-line treatment for HIV (Günthard et al. 2016; “AIDSinfo” 2019; World Health Organization 2018). Currently, four compounds, raltegravir (Summa et al. 2008), elvitegravir (Sato et al. 2006), dolutegravir (Kobayashi et al. 2011), and bictegravir (Tsiang et al. 2016) are FDA-approved. As with all classes of antiretrovirals, resistance is a major barrier to durable effectiveness of the STIs (Mesplède, Quashie, and Wainberg 2012; Wainberg, Mesplède, and Quashie 2012; Geretti, Armenia, and Ceccherini-Silberstein 2012). Resistance to one STI often conveys resistance to other members of the class (D. J. Hazuda, Anthony, et al. 2004; Garrido et al. 2012). Optimized compounds show increased potency and higher barriers to resistance (Tsiang et al. 2016; Naidu et al. 2018; Wiscount et al. 2008; Métifiot et al. 2011; X. Z. Zhao et al. 2016, 2017; Smith et al. 2018), often achieved by rational optimization of inhibitor binding within the intasome active site. Therefore, detailed structural models of the HIV intasome are essential determine differences between the prototype foamy virus and HIV and to dissect mechanisms of resistance, enabling further development of the STIs (Passos et al. 2020). Faced with viral resistance, alternative approaches to targeting integrase are also needed.
1.5 LEDGF and allosteric integrase inhibitors

One alternative approach to targeting integrase was discovered from a study of interactions with host cell proteins. The best-characterized integrase-interacting host cell proteins is the lens epithelium-derived growth factor (LEDGF), a protein normally involved in transcription (Ge, Si, and Roeder 1998). LEDGF was identified in several studies to interact with integrase (Peter Cherepanov et al. 2003; Turlure et al. 2004; Emiliani et al. 2005), and through DNA- and chromatin-binding domains is able to recruit integrase to genomic DNA (G. Maertens et al. 2003). Knockdown of LEDGF decreased the infectivity of HIV in vitro, specifically affecting integration (Llano, Vanegas, et al. 2006; Shun et al. 2007; Emiliani et al. 2005; Busschots et al. 2007; Hombrouck et al. 2007). After knockdown, integration events were no longer focused in actively transcribed regions of the genome, a phenotype that could be rescued by reintroduction of LEDGF (Ciuffi et al. 2005; Marshall et al. 2007; Shun et al. 2007). Conversely, overexpression of the integrase binding domain (IBD) of LEDGF separate from the DNA-binding domains also inhibited integration (De Rijck et al. 2006; Llano, Saenz, et al. 2006), by competing with endogenous full-length LEDGF and modulating the oligomeric state of integrase (Peter Cherepanov et al. 2003; Gupta et al. 2010; McKee et al. 2008; Hare, Di Nunzio, et al. 2009). These observations led to investigation of the LEDGF\textsubscript{IBD}-integrase interaction as a target for drug development (Llano, Saenz, et al. 2006; Peter Cherepanov et al. 2005).

The LEDGF\textsubscript{IBD} binds to a well-characterized site at the CCD dimerization interface (Figure 1.2) (Peter Cherepanov et al. 2005; Hare, Shun, et al. 2009). Through examination of available co-crystal structures and mimicry of the interacting residues on
LEDGF, small molecules were developed to interfere with the integrase-Ledgf_{IBD} interaction (Figure 1.2) (Christ et al. 2010). These compounds contained molecular mimics of the key LEDGF residues that interact with integrase: a carboxyl moiety in place of Asp-366, a phenyl group in place of Ile-365, and a chlorinated quinoline in place of Leu-368. A hydrophobic tert-butoxy moiety was found to be an improvement over the propyl moiety used in the first compounds (Christ et al. 2012), and is buried deeply in a hydrophobic pocket along the CCD α1 helix. Notably, this site had been identified as a potential inhibitor binding site before it was known to be the site of LEDGF binding (Molteni et al. 2001). Multiple groups have synthesized and characterized ALLINIs (Fader et al. 2014; H. Wang et al. 2012; G. Li et al. 2020; Kessl et al. 2012; Sharma et al. 2014; Tsiang et al. 2012; Amadori et al. 2017; Gupta et al. 2014; Wilson et al. 2019; Peese et al. 2019), although all members of this class retain the same key features.

The effect of ALLINIs was expected to be similar to LEDGF knockdown or LEDGF_{IBD} overexpression, that is, inhibiting the integration step and redirecting
integration site specificity (Christ et al. 2010). Surprisingly, the major effect of ALLINIs was observed to be late in the viral replication cycle, after integration into the host cell genome had already occurred (Desimmie et al. 2013; Feng et al. 2013; Le Rouzic et al. 2013; Gupta et al. 2014; Slaughter et al. 2014). Although some activity is evident early in viral replication, it is likely a secondary effect and not responsible for the potency of the ALLINIs (Fontana et al. 2015; Balakrishnan et al. 2013; Sharma et al. 2014). The phenotype observed was a virion maturation defect, leading to the production of non-infectious virions containing abnormal, eccentric electron-dense aggregates (Balakrishnan et al. 2013; Bonnard et al. 2018). Normal virion maturation is a complex process that occurs as the newly budded virion containing all of the necessary components for infection leaves the producer cell. Proteolytic processing of the Gag and Gag-Pro-Pol polyproteins results in condensation of the viral RNA genome and nucleocapsid into a central ribonucleoprotein complex surrounded by the capsid (Sundquist and Kräusslich 2012). Disruption of this step by ALLINI treatment implicated integrase in virion maturation (Kessl et al. 2016; Fontana et al. 2015; Tsiang et al. 2012). Virions produced in the presence of ALLINIs contained mislocalized ribonucleoprotein complexes and eccentrically-located aggregates of integrase (Jurado et al. 2013; Fontana et al. 2015; Kessl et al. 2016). The cause of this phenotype was determined to be abnormal multimerization of integrase by the ALLINIs (Jurado et al. 2013; Gupta et al. 2014; Desimmie et al. 2013; Feng et al. 2013; Gupta et al. 2016; Deng et al. 2016; Koneru et al. 2019), preventing integrase from interacting with the viral RNA (Kessl et al. 2016).

The molecular structure of ALLINI-induced integrase aggregates was not immediately obvious. However, domain truncations (Gupta et al. 2014),
hydrogen/deuterium exchange (Shkriabai et al. 2014), and molecular modeling (Deng et al. 2016) implicated the CTD in ALLINI-induced aggregation. Eventually, a molecular model of the complete integrase-ALLINI complex was determined by our group (Figure 1.3) (Gupta et al. 2016). In this structure, the ALLINI forms extensive contacts with both the CCD and CTD, linking adjacent dimers of integrase together to form an open polymer. This polymer differs from the form of integrase observed in the intasome, and is not expected to be catalytically active. Work is ongoing to determine whether the primary disruptive effect of this polymer is due to sequestration of integrase from its role in condensing the viral RNA or physical disruption of the virion (Gupta et al. 2020, submitted). Although this structure is a dramatic step forward in understanding the mechanism of action of ALLINIs, it lacks high resolution detail, was determined with a heavily-modified integrase construct that does not retain catalytic activity, and could only be achieved with a single ALLINI (Gupta et al. 2016). Addressing these deficiencies is a major goal of this thesis.

![Figure 1.3. Structure of IN-ALLINI Polymer.](image)

(Figure 1.3. Structure of IN-ALLINI Polymer. Co-crystallization of INV15A, F16SH and GSK1264 revealed that the ALLINI GSK1264 (shown in red) binds to an interface between CCD and CTD of adjacent dimers (shown in silver and gold). The interaction is bivalent, with two CCD-CTD binding interfaces forming between each pair of dimers. The polymer continues indefinitely in the crystal lattice, and presumably, in vivo, forming insoluble and catalytically inactive aggregates. (Adapted from Gupta et al. 2016. PDB: 5HOT))
As with all antiretroviral classes, resistance to ALLINIs is a problem that must be addressed. Serial viral passage experiments in the presence of ALLINIs has led to the isolation of multiple resistance mutations (Christ et al. 2010; Tsiang et al. 2012; Christ et al. 2012; Le Rouzic et al. 2013; Balakrishnan et al. 2013; Sharma et al. 2014; Fader et al. 2014). Integrase constructs containing resistance substitutions show decreased aggregation in vitro, supporting a link between a decreased propensity to aggregate and in vivo resistance. Many resistance substitutions are located at or near the inhibitor binding site (Gupta et al. 2016), however, some mutations are located distant from the binding site and convey resistance in yet undiscovered ways. Although some resistance substitutions are shared between ALLINIs, they do not overlap entirely across different compounds, emphasizing the importance of characterizing mechanisms of resistance to enable the development of inhibitors with higher barriers to resistance. Importantly, identified ALLINI resistance substitutions are non-overlapping with STI resistance substitutions (Gupta et al. 2016; Tsiang et al. 2012; Blanco et al. 2011), meaning that ALLINIs offer a complementary approach to existing antiretrovirals targeting integrase.

1.6 Thesis objectives

Structural biology has a proven track record in developing effective antiretrovirals to combat the HIV epidemic. However, significant challenges and opportunities remain in the field of HIV research and antiretroviral drug development. In this body of work, I apply the techniques of structural biology and biochemistry to the task of better understanding HIV integrase and the antiretrovirals targeting integrase. Reported here is significant progress toward a better understanding of the function of HIV integrase and its inhibition by STIs and ALLINIs. In Chapter 2, I extend published data from our lab
detailing the structural mechanism of action of the ALLINIs. This work improves the resolution of structural detail, includes additional ALLINIs, elucidates mechanisms of resistance, and outlines potential routes for compound optimization; progress that will benefit ongoing drug development. In Chapter 3, I report our collaborative work on a highly potent ALLINI that emphasizes the promise of this class of compounds. I show that STI-resistant IN is hypersensitive to this ALLINI, providing a rationale for combination therapy of STIs and ALLINIs. In Chapter 4, I report the structural, biochemical, and biophysical characterization of wild type, native HIV-1 integrase. These data have not been previously reported due to the difficulty of working with HIV-1 integrase in vitro. Our purification approach enabled us to study integrase in the form found in vivo, and set the foundation for more accurate study of integrase in the future. Chapter 5 extends the data from Chapter 4, reporting the formation of integrase complexes with LEDGF and viral DNA. These studies are essential steps toward the structural determination of the catalytically active integrase-DNA complex, the intasome. STIs bind to this complex, and determination of its structure with wild type, native HIV-1 integrase is a major future goal. Together, this work advances the understanding of the complex and multifaceted functions of HIV integrase, and the structures presented here establish a template for ongoing antiretroviral drug development.
Chapter 2: Structural details of allosteric HIV integrase inhibitor binding and resistance

The contents of this chapter are currently being prepared as a manuscript for publication.

2.1 Abstract

HIV integrase (IN), the enzyme that catalyzes viral integration into genomic DNA, has been successfully targeted by inhibitors that bind the active site. A second class of inhibitors, the allosteric integrase inhibitors (ALLINIs), bind a different site and act by a different mechanism. ALLINIs have not yet been implemented clinically, pending improvements in specificity, potency, and barriers to resistance—goals that can be achieved through a better understanding of the structural mechanism of action of ALLINIs. We have previously co-crystallized full length HIV IN bound with an ALLINI and reported the structure of this complex. Here, we extend the structural understanding of ALLINIs through novel IN·ALLINI crystal structures. Two structures of an ALLINI bound to full-length IN harboring resistance substitutions illustrate possible mechanisms of ALLINI resistance. These are the first structures of replication-competent full-length IN bound to an ALLINI, confirming that IN·ALLINI polymers can form in vivo and that polymer formation is the structural mechanism of action of ALLINIs. A ternary complex structure of an ALLINI bound to the catalytic core domain and the carboxy-terminal domain of IN orthogonally confirms the structure of the ALLINI binding site and improves the resolution of structural detail. Examination of these structures identifies routes for optimization of existing ALLINI scaffolds.
2.2 Introduction

Resistance to antiretrovirals presents a barrier to effective control of HIV and prevention of transmission, essential goals to end the worldwide HIV epidemic (Clavel and Hance 2004). Use of potent antiretrovirals with the highest genetic barriers to drug resistance is essential for minimizing the emergence and spread of resistant virus (B. Brenner and Wainberg 2016). Antiretrovirals targeting HIV integration, an essential step in the retroviral replication cycle (Robert Craigie and Bushman 2012), has several advantages including improved tolerability, fewer side effects or drug-drug interactions, and, importantly, a high genetic barrier to resistance (Osterholzer and Goldman 2014; M. M. Miller et al. 2015; B. G. Brenner and Wainberg 2017; Smith et al. 2018; Tsiang et al. 2016; Malet et al. 2018).

The integrase (IN) enzyme is part of the pol gene and cleaved from a polyprotein precursor by the viral protease. The general structure and three-domain organization of IN, consisting of the N-terminal domain (NTD), catalytic core domain (CCD), and C-terminal domain (CTD), is conserved across retroviral families (A. Engelman and Craigie 1992; A. N. Engelman and Cherepanov 2017; Chiu and Davies 2004; Alan Engelman and Cherepanov 2014). IN associates with the free ends of viral DNA to form a macromolecular complex, the intasome, which progresses through three distinct forms during integration (A. Engelman, Bushman, and Craigie 1993; A. N. Engelman and Cherepanov 2017; M. D. Miller, Farnet, and Bushman 1997; Lewinski and Bushman 2005; Cook et al. 2020; Passos et al. 2020, 2017; Ballandras-Colas et al. 2017). During integration, IN catalyzes two sequential reactions: 3'-processing (removal of a GT dinucleotide from the viral DNA ends) and strand transfer (the covalent insertion of viral...
DNA ends into host DNA) (Robert Craigie and Bushman 2012). The reaction is likely completed by host DNA repair enzymes (Yoder and Bushman 2000).

Studies of the enzymatic activity of IN led to the development of the first class of integrase inhibitors, the strand transfer inhibitors (STIs) (D. J. Hazuda et al. 2000; D. Hazuda 2012), four of which, raltegravir, elvitegravir, dolutegravir, and bictegravir, have received FDA approval (Mesplède and Wainberg 2013; A. N. Engelman 2019). Resistance to STIs is well-documented (You et al. 2016; Malet et al. 2008; I. E. A. Wijting et al. 2018; I. Wijting et al. 2017; Cooper et al. 2008) and, importantly, different STIs show overlapping resistance profiles, motivating the search for alternative approaches to targeting IN (Mesplède and Wainberg 2015; Shimura et al. 2008; Molina et al. 2012; Garrido et al. 2012).

The allosteric inhibitors of integrase (ALLINIs) were developed as an alternative approach to targeting IN. ALLINIs bind to a pocket normally occupied by the cellular cofactor LEDGF/p75 (Christ et al. 2010; Fader et al. 2014; H. Wang et al. 2012; G. Li et al. 2020). IN associates with the chromatin-associated host factor, LEDGF, which directs IN toward highly expressed genes (A. R. W. Schröder et al. 2002; Emiliani et al. 2005; Busschots et al. 2007; Peter Cherepanov et al. 2003; De Rijck et al. 2006; Hombrouck et al. 2007; G. Maertens et al. 2004; Shun et al. 2007; Vanegas et al. 2005). Integration of HIV occurs in active transcription units, and knockdown of LEDGF results in a loss of integration site specificity (Ciuffi et al. 2005; G. P. Wang et al. 2007; Lewinski et al. 2006; Vandekerckhove et al. 2006) and inhibits HIV replication (Llano, Saenz, et al. 2006). The LEDGF binding pocket, which is formed by the dimerization of two monomers of IN, was recognized to be a potential target for drug design (Peter Cherepanov et al. 2005; Christ and Debyser 2013; Poeschla 2008; Kessl et al. 2009), leading to the development of
ALLINIs. The major binding interactions are contributed by the pharmacophore carboxylic acid and tert-butoxy groups with the backbone amides of residues Glu-170 and His-171 and the Thr-174 side chain (Gupta et al. 2014; Tsiang et al. 2012; Le Rouzic et al. 2013; Christ et al. 2010; Feng et al. 2013; Jurado et al. 2013; Gupta et al. 2016). The complete molecular structure of the ALLINI binding interface was revealed in a crystal structure of an IN·ALLINI polymer (Gupta et al. 2016, 2014). The CTD of an adjacent dimer of IN contributes to ALLINI binding with Tyr-226, Trp-235, Ile-268, and Lys-266, although the resolution of the currently available structure limits definitive assignment (Gupta et al. 2016). A major goal of the current study is to improve the resolution of the ALLINI binding interface.

The major effect of ALLINIs is observed late in the viral replication cycle, after viral DNA integration into the host cell genome has already occurred (Desimmie et al. 2013; Jurado et al. 2013; Feng et al. 2013; Le Rouzic et al. 2013; van Bel et al. 2014). ALLINIs interfere with maturation, the process by which a newly budded virion containing all of the necessary components for infection undergoes proteolytic processing of the Gag and Gag-Pro-Pol polyproteins, condensing the viral RNA genome and nucleocapsid into a central ribonucleoprotein (RNP) complex surrounded by the capsid. ALLINI-treated producer cells produce non-infectious virions containing abnormal, eccentric electron dense aggregates (Balakrishnan et al. 2013; Gupta et al. 2016, 2014). These virions contain mislocalized RNP and the eccentric electron dense aggregates are composed of IN (Jurado et al. 2013; Fontana et al. 2015; Kessl et al. 2016). Studies using ALLINIs has revealed a role of IN in sequence-specific binding of the viral RNA genome, suggesting that IN plays an active role in virion maturation (Kessl et al. 2016). ALLINIs disrupt this role by promoting polymerization of IN (Gupta et al. 2014, 2016;
Feng et al. 2016; Shkriabai et al. 2014; Deng et al. 2016), by a mechanism similar to that of the anti-neoplastic drug paclitaxel, which promotes the polymerization of tubulin into microtubules (Horwitz 1994). It remains unclear whether the effect of ALLINIs on virion maturation is due to the sequestration of IN into inactive polymers, preventing interaction with viral RNA and thereby preventing formation of the central RNP complex, or due to physical disruption of virions by large aggregates of IN. Ongoing work in our group suggests that IN may be forming a protein gel, a three-dimensional protein aggregate with specific properties, and thereby disrupting proper virion function (Gupta 2020, submitted). A secondary effect of ALLINIs is observed early in the viral replication cycle, and is a manifestation of ALLINIs directly interfering with IN binding to LEDGF (Christ and Debyser 2013; Christ et al. 2012, 2010; Kessl et al. 2012; Le Rouzic et al. 2013; Sharma et al. 2014; Tsiang et al. 2012).

The IN construct used to determine the structure of the IN·ALLINI polymer contained the two substitutions Y15A and F185H, which convey beneficial solution properties (Gupta et al. 2016; Bujacz et al. 1996; T. M. Jenkins et al. 1995). The F185H substitution in the CCD mimics the residue present in ASV IN (Bujacz et al. 1996; T. M. Jenkins et al. 1995), and is replication-competent (A. Engelman et al. 1997). Although the NTD is not seen in the IN·ALLINI structure, the IN construct used in this crystal structure contains the substitution Y15A (known to impair replication in vivo) in the NTD (Takahata et al. 2016). It is therefore important to determine a crystal structure of an IN·ALLINI polymer with replication-competent IN.

Importantly, ALLINIs target a site on IN distinct from the site targeted by STIs, and resistance mutations to ALLINIs do not overlap with STI resistance mutations (Gupta et al. 2016; Blanco et al. 2011; Tsiang et al. 2012), meaning that ALLINIs offer a
non-redundant approach to targeting IN. ALLINI resistance substitutions can be readily identified in serial viral passage experiments (Christ et al. 2010; Tsiang et al. 2012; Christ et al. 2012; Le Rouzic et al. 2013; Balakrishnan et al. 2013; Sharma et al. 2014; Fader et al. 2014). The mechanisms of several resistance substitutions have been studied (Feng et al. 2013; Slaughter et al. 2014; Hoyte et al. 2017), albeit with a focus on the CCD side of the ALLINI binding interface. The vast majority of resistance substitutions occur in the CCD and would be expected to directly affect ALLINI binding to the LEDGF binding site, but effects on CCD-CTD interactions in the context of the complete ALLINI binding interface must be considered (Gupta et al. 2016). We selected substitutions at Trp-131 and Asn-222 for further structural study as they are distant from the ALLINI binding interface and do not have an obvious mechanism of resistance. An understanding of the structural basis for resistance will inform future drug development, and indirectly, inform our understanding of the solution properties of IN (Gupta et al. 2016; Koneru et al. 2019).

Here, we extend published data from our group showing the structure of an IN-ALLINI complex, revealing the specific mechanism of action of ALLINIs and making sense of their primary effect at late stages of the HIV life cycle (Gupta et al. 2016). The published IN-ALLINI complex structure was determined with an IN construct containing the Y15A substitution that impairs replication in vivo (Takahata et al. 2016); here, we report two new structures that are fully replication-competent. These structures include the substitutions W131C or N222K, which convey resistance to ALLINIs. These structures show that the IN-ALLINI polymer accommodates replication-competent IN, and that resistance substitutions do not directly disrupt polymer formation. Instead, resistance substitutions act by decreasing interaction affinities on the way to polymer
formation. These structures were determined with the ALLINI BI-D, structurally similar to the ALLINI GSK1264, used to determine the published IN-ALLINI structure (Gupta et al. 2016), generalizing the structural mechanism of action of these compounds. Furthermore, we have improved the resolution of structural models of the IN-ALLINI interface, achieved by separating IN into individual domains and forming a ternary complex with a ALLINI in vitro. This structure was determined with a third member of the ALLINI class, BI-224436, and recapitulates the general structure of the IN-ALLINI interface. From these improved models, routes for improvement to existing ALLINI scaffolds can be identified, which will be useful for drug optimization in the future.

2.3 Results

2.3.1 Structures of replication competent INs containing resistance substitutions with the ALLINI BI-D

We sought to improve on our previously published structure of IN$^{Y15A}$ with the ALLINI GSK1264 by using replication-competent IN constructs containing ALLINI resistance substitutions and a different ALLINI, BI-D (Supplemental Figure 2.1) (Tsantrizos et al. 2009). Substitutions in IN that convey resistance to ALLINIs are consistently located near the ALLINI binding site, with the vast majority found in the CCD (Christ et al. 2010; Tsiang et al. 2012; Christ et al. 2012; Le Rouzic et al. 2013; Balakrishnan et al. 2013; Sharma et al. 2014; Fader et al. 2014). Substitutions that lie within the binding interface would be expected to prevent ALLINI binding by steric interactions or disrupting electrostatic or hydrophobic interactions. The mechanism of resistance of substitutions that lie outside the binding interface is less clear. In serial viral passage experiments in the presence of ALLINIs, we observed selection for multiple
Resistance substitutions (Gupta et al. 2016). We selected two of these substitutions, W131C (in the CCD) and N222K (in the CTD) for further structural study (Gupta et al. 2016; Tsiang et al. 2012). Full length IN constructs were purified and crystallized as reported previously (Gupta et al. 2016). Crystals of the two resistance substitution-containing IN constructs were grown in the presence of the ALLINI BI-D, and were in the same space group and diffracted to a similar resolution as the crystals of IN^Y15A and the ALLINI GSK1264 (Supplemental Figure 2.2). The previously published IN^Y15A structure (PDB: 5HOT) was used as a search model for molecular replacement. Crystallographic statistics are found in Table 2.1.

Overall, the structure of the IN constructs with resistance substitutions was the same as the IN Y15A structure (Gupta et al. 2016) (Figure 2.1a). The ALLINI BI-D binds at the CCD dimer interface as expected (Jurado et al. 2013; Slaughter et al. 2014). Residues contributed by the CTD are similar to those observed in the IN Y15A structure, although the resolution limits definitive assignment of side chain position. The W131C substitution removes the potential for hydrophobic interactions between Trp-131 and Ile-268 and cation-π interactions between Trp-131 and Arg-224 (Figure 2.1b). However, the presence of cysteine at this site allows for the formation of a hydrogen bond with nearby Tyr-226, which may compensate for the loss of the contributions of Trp-131 to binding affinity. The N222K substitution also appears to disrupt interactions with Trp-131 on the CCD (Figure 2.1b). Selection for the longer, basic Lys-222 may induce a clash with Trp-131, decreasing the affinity of the CTD for the CCD. It is also possible that Lys-222 alters the conformation of the CTD in the absence of CCD binding, through either a disruption of the CCD-CTD linker or through the formation of a salt bridge.
between Lys-222 and Asp-270. Asp-270 is located near Asn-222 in the solution structure of the isolated CTD (Lodi et al. 1995; Eijkelenboom et al. 1995, 1999).

### Table 2.1: Crystallographic Statistics

<table>
<thead>
<tr>
<th>Data Collection*</th>
<th>IN\text{W131C, F185H}·BI-D</th>
<th>IN\text{K222K, F185H}·BI-D</th>
<th>IN\text{CCTD·BI-224436}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>P 6, 2 2</td>
<td>P 6, 2 2</td>
<td>C 1 2 1</td>
</tr>
<tr>
<td>Unit cell dimensions (Å)</td>
<td>104.924 104.924 245.062</td>
<td>106.033 106.033 245.875</td>
<td>90.85 57.756 112.057</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>48.23 - 4.56</td>
<td>48.68 - 4.53</td>
<td>90, 104.026, 90</td>
</tr>
<tr>
<td>R\text{merge}</td>
<td>0.1337 (1.664)</td>
<td>0.1402 (2.116)</td>
<td>0.062 (0.709)</td>
</tr>
<tr>
<td>CC\text{1/2}</td>
<td>99.9 (26.3)</td>
<td>99.7 (30.0)</td>
<td>99.5 (96.1)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>5.4 (4.8)</td>
<td>7.4 (7.7)</td>
<td>4.0 (4.0)</td>
</tr>
<tr>
<td>I/σ</td>
<td>6.75 (0.76)</td>
<td>7.06 (1.03)</td>
<td>12.62 (1.73)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>96.3 (87.53)</td>
<td>99.1 (99.6)</td>
<td>98.23 (95.37)</td>
</tr>
<tr>
<td>Refinement</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reflections (work)</td>
<td>4334</td>
<td>4662</td>
<td>10,574</td>
</tr>
<tr>
<td>Reflections (free)</td>
<td>481</td>
<td>517</td>
<td>1159</td>
</tr>
<tr>
<td>R\text{work}</td>
<td>0.332\text{b}</td>
<td>0.3017\text{b}</td>
<td>0.3484\text{b}</td>
</tr>
<tr>
<td>R\text{free}</td>
<td>0.445\text{b}</td>
<td>0.4238\text{b}</td>
<td>0.4015\text{b}</td>
</tr>
<tr>
<td>Non-Hydrogen atoms</td>
<td>3276</td>
<td>3289</td>
<td>2784</td>
</tr>
<tr>
<td>R.m.s. bonds (Å)</td>
<td>0.042</td>
<td>0.043</td>
<td>0.02</td>
</tr>
<tr>
<td>R.m.s. angles (°)</td>
<td>2.78</td>
<td>2.86</td>
<td>1.81</td>
</tr>
<tr>
<td>Ramachandran (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Favored</td>
<td>95.24</td>
<td>94.99</td>
<td>97.29</td>
</tr>
<tr>
<td>Allowed</td>
<td>4.26</td>
<td>5.01</td>
<td>2.41</td>
</tr>
<tr>
<td>Outliers</td>
<td>0.5</td>
<td>0.0</td>
<td>0.3</td>
</tr>
</tbody>
</table>

*Values in parentheses are for the highest-resolution shell

*bRefinements utilizing the DEN method implemented in CNS and the Rosetta method implemented in Phenix are ongoing.
2.3.2 Purification of individual IN domains

As seen in the structure of full-length IN bound with an ALLINI (Figure 2.1, Supplemental Figure 3a), the ALLINI binding interface is composed entirely of residues from the CCD and CTD. We therefore truncated IN to remove the NTD, which is not seen in the full-length IN·ALLINI complex, in an effort to decrease disorder in the crystal lattice and improve resolution. $\text{IN}_{\text{CCD}}$ and $\text{IN}_{\text{CTD}}$ have previously been purified as
individual domains and their structures have been determined (Dyda et al. 1994; Maignan et al. 1998; J. C.-H. Chen et al. 2000; Goldgur et al. 1998; Eijkelenboom et al. 1995, 1999; Lodi et al. 1995). We purified IN\textsubscript{CCD} (residues 50-212) and IN\textsubscript{CTD} (residues 220-271) (Supplemental Figure 3b). IN\textsubscript{CCD} contained the solubility-enhancing substitution F185K (Dyda et al. 1994). IN\textsubscript{CTD} was truncated at Tyr-271, which is the last residue of the CTD observed in the full-length IN·ALLINI polymer (PDB: 5HOT) to make contact with the CDD (Gupta et al. 2016). The remaining residues (272-288) were removed to decrease disorder in the crystal.

2.3.3 Interaction of IN\textsubscript{CCD} and IN\textsubscript{CTD} in vitro

Addition of ALLINI to full-length IN in vitro results in the formation of a cloudy precipitate composed of IN oligomers that can be measured by light scattering (Gupta et al. 2016, 2014, 2010). Surprisingly, addition of ALLINI to an equimolar solution of IN\textsubscript{CCD} and IN\textsubscript{CTD} also resulted in the formation of a cloudy precipitate (Figure 2.2). This is unexpected, as full-length IN·ALLINI polymers form by bivalent CCD-ALLINI-CTD interactions between pairs of IN dimers (Gupta et al. 2016), and the isolated domains lack the interdomain linker that connects adjacent IN dimers in the full-length IN·ALLINI polymer (Supplemental Figure 4a). We propose a new model for IN\textsubscript{CCD} and IN\textsubscript{CTD} oligomer formation whereby IN\textsubscript{CCD} and IN\textsubscript{CTD} form monovalent CCD-ALLINI-CTD interactions and extend into a polymer by CTD-CTD dimerization (Supplemental Figure 4b). IN\textsubscript{CTD} has been reported to be dimeric in solution (Eijkelenboom et al. 1995; Lodi et al. 1995), and, as seen in both the NMR (Eijkelenboom et al. 1995) and unpublished crystal structure of the CTD (PDB: 5TC2), the dimer interface is opposite to the surface involved in ALLINI binding and is not observed in the structure of full-length IN·ALLINI.
structure (Supplemental Figure 4a). The CTD dimerization interface involves the β2, β3, and β4 sheets and is predominantly hydrophobic, with significant contributions from residues Val-240, Leu-242, Trp-243, and Ile-257 (Eijkelenboom et al. 1995, 1999; Lodi et al. 1995). To test our model of CCD-ALLINI-CTD polymerization, we disrupted the hydrophobic CTD dimerization interface with the substitution L242A. IN$_{CTD}^{L242A}$ was shown to be dimeric by SEC-MALS and AUC (Gupta 2020, submitted). Mixing of IN$_{CCD}$ and IN$_{CTD}^{L242A}$ completely abrogated precipitate formation in the presence of the ALLINIs BI-224436 and BI-D (Figure 2.2).

2.3.4 Co-crystallization of IN$_{CCD}$, the ALLINI BI-224436, and IN$_{CTD}$
Mixtures of IN$_{CCD}$, IN$_{CTD}$, and the ALLINI BI-224436 readily crystallized (Supplemental Figure 2). Crystals could not be replicated in the absence of either domain or ALLINI. The structure was solved by molecular replacement using the IN$_{CCD}$ dimer (PDB: 3L3U) (Wielens et al. 2010) as a search model. Electron density corresponding to IN$_{CTD}$ and ALLINI was visible, completing the ALLINI binding site as seen in the full-length IN·ALLINI polymer (Gupta et al. 2016). A dimer of IN$_{CCD}$ and 2 monomers of the IN$_{CTD}$ could be placed in the asymmetric unit (Figure 2.3a). In agreement with our model (Supplemental Figure 4b), the crystal lattice revealed open

![Figure 2.3. Structure of IN$_{CCD}$-BI-224436-IN$_{CTD}$](image)

Co-crystallization of isolated IN$_{CCD}$ and IN$_{CTD}$ resulted in the formation of an IN·ALLINI binding interface similar to that observed in full-length IN·ALLINI crystal structures. a) Each of the two ALLINI binding sites on the CCD dimer (green) is occupied by a molecule of BI-224436, and each binding interface is capped by a CTD (blue). b) and c) Detailed views of ALLINI binding interface, with domains colored as in (a), with labels identifying secondary structure elements involved in ALLINI binding.

27
polymers consisting of monovalent CCD-ALLINI-CTD interactions and CTD-CTD dimers. The CTD dimerization interface is rotated approximately 90 degrees from the orientation observed in solution (Supplemental Figure 4).

2.3.5 Structure of the IN\textsubscript{ccd}·BI-224436·IN\textsubscript{ctd} complex

The overall structure of the IN-ALLINI binding interface is similar to that observed in the structure of the full-length IN-ALLINI polymer (Gupta et al. 2016). The binding interface is composed of residues from the α1, α3, α4, and α5 helices of the CCD and the β1, β1-2 turn, and β5 sheets of the CTD (Figure 2.3b and Figure 2.4). The key interactions between the BI-224436 pharmacophore and residues Glu-170, His-171, and Thr-174 are observed. The quinoline scaffold makes contacts with residues Thr-124 and Thr-125 and approaches Ala-128 (Fader et al. 2014; Koneru et al. 2019). The large tricyclic arene forms extensive hydrophobic contacts with the CCD (Koneru et al. 2019; Fader et al. 2014). The interactions with the CTD are primarily hydrophobic and π-π interactions involving the quinoline scaffold and the tricyclic arene. The quinoline scaffold projects toward Trp-235 and Tyr-226, and the tricyclic arene contacts Ile-268. Trp-235
and Tyr-226 adopt the conformation seen in one of two subunits of the IN·ALLINI polymer determined with GSK1264, with the hydroxyl group of Tyr-226 pointed away from Trp-235, as seen in the solution NMR structure of the isolated CTD (Lodi et al. 1995) (Supplemental Figure 5). A coordinated flip of these side chains would avoid a steric clash, and may readily occur in solution. Lys-266 is in a position to participate in hydrogen bonding with the ALLINI, but adopts a rotamer that projects more toward the α4-α5 turn of the CCD rather than the ALLINI (Supplemental Figure 5).

Examination of the structures in this study and a high resolution IN_{ccd}·BI-224436 structure (Konuru et al. 2019) allowed for the construction of an interaction network and revealed routes for inhibitor optimization. Two putative hydrogen-bonding networks bridged by water molecules were identified in the structure (Figure 2.4a). His-171 forms hydrogen bonds with the tert-butoxy and carboxylic acid moieties and is positioned to form an interaction with Gln-95 via a water molecule. Similarly, Glu-170 is positioned to form a hydrogen bond with the aromatic nitrogen in the quinoline ring and Trp-235 via a water molecule. Expansion of ALLINI scaffolds to displace water molecules and form additional hydrogen bonds with polar side chains would be expected to improve affinity and potency of these compounds. Although much of the ALLINI is deeply buried between the CCD and CTD and solvent-inaccessible, a solvent channel flanked by Thr-124, Thr-125, and Trp-235 connects the quinoline ALLINI core to bulk solvent. ALLINI scaffolds could be extended in this direction to form additional contacts with the CCD and CTD without disrupting existing CCD and CTD contacts.

2.4 Discussion
The ALLINIs were designed to bind to the LEDGF binding site on the CCD dimer and exclude LEDGF binding (Peter Cherepanov et al. 2005; Christ and Debyser 2013; Poeschla 2008; Kessl et al. 2009), inhibiting integration in a similar manner to that of LEDGF knockdown (Ciuffi et al. 2005; G. P. Wang et al. 2007; Lewinski et al. 2006; Vandekerckhove et al. 2006; Llano, Saenz, et al. 2006). Surprisingly, the major effect of ALLINIs occurs late in the viral replication cycle (Desimmie et al. 2013; Jurado et al. 2013; Feng et al. 2013; Le Rouzic et al. 2013) corresponding to a disruption in virion maturation. This effect is due to polymerization and aggregation of IN. A crystal structure published by our group revealed the structural mechanism of action of ALLINIs (Gupta et al. 2016), but was determined with a replication-incompetent IN construct and lacked high resolution detail. Here, we extend our understanding of ALLINIs with structures of replication-competent IN constructs containing resistance substitutions with an additional ALLINI, BI-D. These structures are essential evidence that supports the formation of IN-ALLINI polymers in vivo and reveal mechanisms of resistance and routes for optimization of ALLINI compounds. We also report a structure consisting of the two IN domains involved in ALLINI binding, the CCD and CTD, with an additional ALLINI, BI-224436. This structure improves the resolution of structural models of the IN-ALLINI binding interface, further identifying routes for ALLINI compound optimization.

The full-length IN-ALLINI structures reported here harbor the resistance substitutions W131C or N222K. These structures reveal hydrophobic and electrostatic disruptions between the CCD and CTD distant from the ALLINI binding site. Instead of directly interfering with ALLINI binding, these substitutions appear to function by decreasing the affinity of CCD-CTD interactions. CCD-CTD interactions are not observed in currently available intasome structures (Passos et al. 2017, 2020), so
substitutions that decrease the affinity of this interaction would not be expected to affect IN activity directly. However, ALLINI resistance substitutions have been shown to decrease fitness in other steps of the viral replication cycle, and are compensated for by additional substitutions in IN (Hoyte et al. 2017). CCD-CTD interactions are dramatically strengthened by ALLINIs (Tse et al. 2020), and optimizing ALLINI scaffolds to increase CCD-CTD affinity would be expected to improve potency. Further study of the spectrum of resistance substitutions (Christ et al. 2010; Tsiang et al. 2012; Christ et al. 2012; Le Rouzic et al. 2013; Balakrishnan et al. 2013; Sharma et al. 2014; Fader et al. 2014) is needed to fully elucidate mechanisms of resistance of ALLINIs and to guide further drug development.

Before the structure of an IN·ALLINI polymer was determined, indirect evidence suggested that the CTD contributed to ALLINI binding (Feng et al. 2016; Shkriabai et al. 2014). The structure of the IN·ALLINI polymer clearly confirmed the contribution of the CTD (Gupta et al. 2016). This CCD-CTD interaction has not been observed in multimeric IN structures in the absence of ALLINI, although this may be due to the use of 3-((3-cholamidopropyl) dimethylammonio)-1-propanesulfonate (CHAPS) in vitro (Z. Chen et al. 2000; E. Deprez et al. 2000). Here, we report a IN\textsubscript{CCD}·BI-224436·IN\textsubscript{CTD} structure that confirms that the entire ALLINI binding site consists of contributions from both the CCD and CTD domains. This orthogonal determination of the ALLINI binding interface shows that it is not a crystallization artifact. The IN\textsubscript{CCD}·BI-224436·IN\textsubscript{CTD} structure improves the resolution of the structural model of the ALLINI binding site, achieved by truncating portions of IN thought to be contributing disorder to the IN·ALLINI polymer crystal structure (the NTD and C-terminal residues in the CTD). The improvement in resolution will benefit ongoing drug development efforts. Here, we identify two
hydrogen-bonding networks that involve the ALLINI scaffold and molecules of water. Modification of existing ALLINI scaffolds to displace these water molecules would be expected to increase affinity and further stabilize the ALLINI in the CCD-CTD interface. The major solvent channel connecting to the ALLINI binding pocket represents an additional avenue of improvement to existing ALLINI scaffolds. Extensions in this direction from the ALLINI core could form additional interactions with residues on both the CCD and CTD. Enhancement of CCD-CTD affinity is important to combat drug resistance from substitutions such as W131C and N222K. As we have shown, IN constructs containing these substitutions are still capable of forming IN·ALLINI polymers, but their formation is disfavored by the disruption of hydrophobic and electrostatic interactions. Enhancement and stabilization of CCD-CTD interactions by expansion of IN-ALLINI contacts is required.

The previously described CTD-CTD dimerization interface (Eijkelenboom et al. 1995; Lodi et al. 1995) is again observed in the IN_{CCD·BI-224436·IN}_{CTD} structure. A molecule of the ALLINI links CCD dimers and CTD dimers together to form an open polymer in a similar manner to that observed in the full-length IN·ALLINI polymer (Gupta et al. 2016). In the full-length IN·ALLINI polymer, however, CTD dimers are not observed, although the dimerization interface is opposite the ALLINI binding interface and fully accessible. We recently reported that CTD dimerization contributes to branching of the IN·ALLINI polymer in vitro, leading to the formation of a branched, gel-like network instead of linear filaments of IN (Gupta 2020, submitted). Residues involved in CTD dimerization have been shown to be essential for IN oligomerization and activity (Lutzke and Plasterk 1998), suggesting that the CTD dimerization interface may be important for uncharacterized oligomeric forms of IN. In the hexadecameric
maedi-visna virus intasome (Ballandras-Colas et al. 2017), the CTD dimerization interface mediates interactions between flanking IN subunits, and therefore may be important in the context of higher-order HIV-1 intasomes (Passos et al. 2017, 2020).

Together, the structures presented here significantly extend published data on the mechanism of action of ALLINIs and pathways of resistance. Improvements in resolution identify opportunities to increase potency and raise barriers to resistance, and use of replication-competent IN connects in vitro data more closely to the retroviral replication cycle in vivo.

2.5 Acknowledgments

We thank Paul Bates, Robert Doms, Rahul Kohli, and Mitchell Lewis for suggestions and comments, and Laurie Zimmerman for help with the manuscript. This work was supported by the Penn Center for AIDS Research, the PennCHOP Microbiome Program, and the Johnson Research Foundation. The crystallographic data were obtained at the 24-ID-E beamline at the Advanced Photon Source at the Argonne National Laboratory and the 17-ID-1 beamline at the National Synchrotron Light Source II at Brookhaven National Laboratory.
Chapter 3: Complementary activity of a high-potency ALLINI against strand transfer inhibitor-resistant integrase

3.1 Abstract

The HIV integrase enzyme is an important target for antiretroviral therapy. Inhibitors that block strand transfer activity are in widespread clinical use, but resistance is a significant barrier to durable control. A second class of inhibitors, the allosteric inhibitors of integrase, offer an alternative approach to targeting integrase. Since resistance to both classes of inhibitors are non-overlapping, we wondered whether these two classes could be used for a complementary therapeutic approach. The widely used strand transfer inhibitor dolutegravir has a high barrier to resistance, but eventually selects for resistance substitutions in integrase. We tested the susceptibility of these resistance substitutions to a current-generation, highly potent ALLINI. Integrase constructs containing dolutegravir-resistant substitutions were more sensitive to the ALLINI as compared to wild type integrase. However, combination of dolutegravir and ALLINI did not result in additive or synergistic effects. These data support the complementary use of strand transfer inhibitors and ALLINIs to combat the development of resistance.

3.2 Introduction

The identification of the human immunodeficiency virus (HIV) as the causative agent of the acquired immune deficiency syndrome (AIDS) initiated a tremendous effort of drug development. Antiretroviral therapy significantly decreased the morbidity and mortality associated with HIV infection (Palella et al. 1998; Mocroft et al. 1998), and
brought the hope of a cure for HIV (Perelson et al. 1997). Unfortunately, HIV remains an incurable disease despite improvements in lifespan and time to the development of AIDS (Detels et al. 1998; Schneider et al. 2005; Smit et al. 2006). Major barriers to a cure remain (Stein, Storcksdieck Genannt Bonsmann, and Streeck 2016), including antiretroviral drug resistance (Clavel and Hance 2004; Tang and Shafer 2012) and latency (Pitman et al. 2018). The problem of resistance has been addressed by the development of multiple classes of antiretroviral drugs targeting nearly every step in viral replication (Gubernick et al. 2016). The multitude of drug targets results in non-overlapping resistance to different antiretroviral drug classes. However, resistance remains a significant cause of treatment failure (Wainberg, Zaharatos, and Brenner 2011; Deeks et al. 2009) and solutions to this problem that do not involve more complicated (Astuti and Maggiolo 2014; Thompson et al. 2012; Maggiolo et al. 2002) or toxic (Carr 2003; Willig et al. 2008) regimens are needed.

Antiretrovirals targeting HIV integrase directly address one major barrier to cure: the establishment of latency. Covalent insertion of the reverse-transcribed viral DNA by integrase is an essential step in the viral replication cycle, allowing for the transcription of a new generation of viral RNA and the establishment of latency. Strand transfer inhibitors (STIs) were developed to directly interfere with integration (Eric Deprez et al. 2004; D. J. Hazuda et al. 2000) by binding to the catalytically active complex of integrase and viral DNA (Hare, Gupta, et al. 2010; G. N. Maertens, Hare, and Cherepanov 2010; Hare, Vos, et al. 2010). The STIs have been clinically successful (Demeulemeester, De Maeyer, and Debyser 2015), and are recommended for first-line regimens (Günthard et al. 2016; “AIDSinfo” 2019; World Health Organization 2018). However, STIs are susceptible to another major barrier to cure: the development of resistance (Mesplède,

Integrase relies on the host cell cofactor LEDGF for recruitment to chromatin and integration targeting (Emiliani et al. 2005; Busschots et al. 2007). Knockdown of LEDGF or overexpression of a dominant negative LEDGF variant inhibited HIV replication (Llano, Saenz, et al. 2006; De Rijck et al. 2006). The LEDGF-IN interaction was identified as a potential site for inhibitor development, leading to the development of the allosteric inhibitors of integrase (ALLINIs) (Christ and Debyser 2013; A. N. Engelman 2019; Poeschla 2008). Although designed to block the LEDGF-IN interaction, important during the integration step early in the viral replication cycle, the major effect of ALLINIs is observed late in the replication cycle, after integration has already occurred (Desimmie et al. 2013; Feng et al. 2013; Le Rouzic et al. 2013; Gupta et al. 2014; Slaughter et al. 2014). This effect is due to the abnormal multimerization of IN within the maturing virion, resulting in defective, noninfectious virions (Jurado et al. 2013; Feng et al. 2013; Desimmie et al. 2013; Le Rouzic et al. 2013; Kessl et al. 2016). The ALLINIs were found to bridge two domains of IN to form an open polymer, a form distinct from catalytically active forms of IN (Gupta et al. 2014, 2016).

Together, the STIs and ALLINIs represent complementary approaches to targeting IN and have non-overlapping resistance profiles (Gupta et al. 2016; Tsiang et al. 2012; Blanco et al. 2011), making them attractive candidates for combination antiretroviral regimens. Here, we examine aspects of the combination of the STIs
raltegravir (Summa et al. 2008) and dolutegravir (Kobayashi et al. 2011) with the ALLINI ViiV-919 (Peese et al. 2019; G. Li et al. 2020; Sivaprakasam et al. 2020). We show that ViiV-919 potently inhibits IN strand transfer activity in addition to promoting IN multimerization. This effect is magnified in the presence of IN substitutions that convey resistance to STIs, suggesting that resistance to STIs comes at the expense of increased susceptibility to ALLINIs. These results motivate the investigation of complementary therapy with STIs and ALLINIs in vivo.

3.3 Results

3.3.1 Aggregation of integrase by the ALLINI ViiV-919

The primary mechanism of action of ALLINIs is the polymerization of IN into dysfunctional aggregates (Gupta et al. 2016). This activity can be recapitulated in vitro by the formation of turbid aggregates when IN and ALLINIs are mixed together (Gupta et al. 2014). The extent of aggregate formation can be measured by the scattering of light at 405 nm, allowing for the quantification of the ALLINI effect. We measured the effect of the ALLINIs BI-224436 and ViiV-919 by this method (Figure 3.1). Both ALLINIs induced the formation of aggregates, but ViiV-919 induced aggregation to a significantly greater extent compared to BI-224436.

3.3.2 Inhibition of wild type integrase activity by ViiV-919

The ALLINIs do not directly block catalytic activity by binding to the active site. In vivo, the major effects of ALLINIs is observed late in viral replication, with the integration step relatively spared from inhibition (Fontana et al. 2015; Balakrishnan et al. 2013; Sharma et al. 2014). However, in vitro, they do aggregate IN into a form different from
the canonical intasome (Gupta et al. 2016). Therefore, we tested the effect of the ALLINI Viiv-919 on the catalytic activity of IN. We first examined inhibition of 3'-processing with a fluorescence polarization assay (Figure 3.2). The negative control lacking IN enzyme showed no 3'-processing over time, while the positive control exhibited 3'-processing. Raltegravir, a strand transfer inhibitor, did not inhibit 3'-processing, as expected. The

Figure 3.1. Aggregation of IN by ALLINIs BI-224436 and Viiv-919. IN and ALLINIs were mixed together and incubated for 15 minutes. The extent of aggregation was measured by the formation of turbidity, detected by light scattering at 405 nm. No aggregation was observed in the no ALLINI condition. BI-224436 and Viiv-919 both induced the formation of IN aggregates, with Viiv-919 inducing significantly more aggregation than BI-224436. **** denotes P≤0.0001

Figure 3.2. Effect of Integrase Inhibitors on 3'-Processing. 3'-processing is measured by the release of a fluorescently-labeled 3' dinucleotide from an oligonucleotide mimic of the viral LTR. Unprocessed oligonucleotide exhibits high fluorescence polarization. The processed and released 3' dinucleotide exhibits low fluorescence polarization. Control reactions are (-) IN, which lacks the IN enzyme, and Vehicle, which contains IN and the solvent DMSO. Raltegravir, a strand transfer inhibitor, has no effect on 3'-processing, BI-224436 has a mild effect, and Viiv-919 completely inhibits 3'-processing.
ALLINI BI-224436 had a minor effect on 3’-processing, while the highly potent ALLINI ViiV-919 completely inhibited 3’-processing activity.

We next examined the effect of ViiV-919 on strand transfer activity (Figure 3.3). Strand transfer activity is measured from the integration of a fluorescently-labeled oligonucleotide into supercoiled plasmid DNA. Single-strand integration events result in a tagged circle product and concerted (double strand) integration events result in a 2 LTR coupled product. The strand transfer raltegravir and the ALLINI ViiV-919 inhibited strand transfer activity with similar potency. The IC$_{50}$ for inhibiting single-strand integration was 390 nM (95% CI: 334 to 455 nM) and 534 nM (95% CI: 485 to 588 nM) for raltegravir and ViiV-919, respectively. The IC$_{50}$ for inhibiting concerted integration was 100 nM (95% CI: 82 to 120 nM) and 106 nM (95% CI: 95 to 117 nM) for raltegravir and ViiV-919, respectively.

![Figure 3.3](image.png)

**Figure 3.3. Effect of Integrase Inhibitors on Strand Transfer.** Activity is measured by quantification of the integration of a fluorescent oligonucleotide into a supercoiled plasmid. Activity was normalized to the vehicle condition. The strand transfer inhibitor raltegravir and ViiV-919 show nearly identical potency.

### 3.3.3 ViiV-919 activity against dolutegravir-resistant integrase

Since we are interested in the utility of combining strand transfer inhibitors and ALLINIs, we tested whether substitutions that convey resistance to dolutegravir also conveyed an increased susceptibility to the ALLINI ViiV-919. The substitutions N155H,
K211R, and E212T were introduced into IN, with and without the polymorphism K156N (Malet et al. 2018; Koneru et al. 2019). We then examined the effect of ViiV-919 on 3'-processing and strand transfer activity. ViiV-919 completely abrogated 3'-processing activity in wild type and dolutegravir-resistant constructs (Figure 3.4).

![Figure 3.4. Effect of ViiV-919 on 3'-processing Activity of Dolutegravir-Resistant IN. Wild type and mutant IN constructs perform 3'-processing, and ViiV-919 completely blocks this activity by all three constructs. ** denotes P<0.01, *** denotes P<0.001.](image)

Differences between wild type and dolutegravir-resistant constructs were observed when examining inhibition of strand transfer (Figure 3.5). The effect on IN lacking the polymorphism K156N was minimal, with the greatest effect on concerted integration observed at low concentrations of ViiV-919 (IC\textsubscript{25} = 63 nM vs. 30 nM for IN\textsuperscript{wt} and IN\textsuperscript{N155H, K156N, K211R, E212T}, respectively), and no significant effect on single-strand integration. However, in the presence of K156N, ViiV-919 more potently inhibited both single-strand (IC\textsubscript{50} = 1.04 µM vs. 103 nM for IN\textsuperscript{wt} and IN\textsuperscript{N155H, K156N, K211R, E212T}, respectively)

![Figure 3.5. Effect of ViiV-919 on Strand Transfer Activity of Dolutegravir-Resistant IN. ViiV-919 more potently inhibits strand transfer activity of dolutegravir resistant constructs as compared to wild type integrase.](image)
and concerted integration ($IC_{50} = 138 \text{ nM}$ vs. $34 \text{ nM}$ for $\text{IN}^{wt}$ and $\text{IN}^{N155H, \ K156N, \ K211R, \ E212T}$, respectively). Notably, the polymorphism K156N has been implicated as important for the development of resistance to dolutegravir (Malet et al. 2018). In addition, the baseline strand transfer activity of the dolutegravir-resistant IN constructs was lower than wild type IN (Supplemental Figure 3.1), suggesting that dolutegravir resistance arises at a fitness cost for the virus, a phenomenon also observed to occur in the context of resistance to ALLINIs (Hoyte et al. 2017) and other antiretroviral drugs (Nijhuis, Deeks, and Boucher 2001).

3.3.4 Combined activity of raltegravir and ViiV-919

Since both raltegravir and ViiV-919 inhibit strand transfer activity with similar potency, we next examined whether raltegravir and ViiV-919 showed additive or synergistic effects on strand transfer activity. We selected a dose of each compound approximately corresponding to the $IC_{50}$ of concerted integration (100 nM). As expected, both raltegravir and ViiV-919 alone inhibited strand transfer to a similar extent (Figure 3.6). The effect on concerted integration was greater than the effect on single strand

![Figure 3.6. Effect of Combining Raltegravir and ViiV-919.](image)

Concentrations of raltegravir and ViiV-919 corresponding to approximately $IC_{50}$ of concerted integration were selected for combination treatment. As expected, raltegravir and ViiV-919 alone inhibited concerted integration. Combination treatment with the same doses did not result in a greater effect on either single strand or concerted integration.
integration, as observed previously (Figure 3.3). Combination treatment with raltegravir and ViiV-919 at the same doses did not result in additional inhibition, with no evidence of synergistic or additive activity.

3.4 Discussion

Although antiretroviral therapy has significantly reduced the morbidity and mortality caused by HIV, the challenges of blocking the establishment of latency and preventing the development of viral resistance remain. Drugs targeting HIV integrase directly address these challenges by blocking the enzyme responsible for integration and exhibiting high barriers to resistance (Tsiang et al. 2016; Naidu et al. 2018; Wiscout et al. 2008; Métifiot et al. 2011; X. Z. Zhao et al. 2016, 2017; Smith et al. 2018). The strand transfer inhibitors (STIs) have met with significant clinical success, however, resistance even to the latest drugs may emerge (I. E. A. Wijting et al. 2018; I. Wijting et al. 2017; You et al. 2016), motivating the search for complementary approaches to target integrase. The allosteric inhibitors of integrase (ALLINIs) are a response to this challenge (A. N. Engelman 2019); their pleiotropic mechanism of action and non-overlapping resistance profile make them attractive targets for combination with STIs (Kessl et al. 2012). Here, we report details of this complementary approach.

We show that ViiV-919, a pyrazolopyrimidine ALLINI with an attractive PK profile (Peese et al. 2019; G. Li et al. 2020; Sivaprakasam et al. 2020), potently aggregates IN. IN aggregation is the primary mechanism of action of ALLINIs (Kessl et al. 2016; Gupta et al. 2014, 2016), manifesting late in the viral replication cycle. The structure of the IN·ALLINI aggregate is distinct from the catalytically active complex (the intasome) (Gupta et al. 2016), leading us to examine whether the potent ALLINI ViiV-919 also
inhibited IN activity. We found that ViiV-919 completely abrogated 3'-processing activity, to a greater extent than the less-potent ALLINI BI-224436, and inhibited strand transfer activity with a similar potency to the STI raltegravir. These results suggest that potent inhibition of IN by ALLINIs not only disrupts virion maturation but also inhibits catalytic activity.

Resistance to STIs is a cause of antiretroviral therapy failure (Wainberg, Zaharatos, and Brenner 2011; Deeks et al. 2009). Dolutegravir, an STI recommended for first-line antiretroviral therapy (Günthard et al. 2016; "AIDSinfo" 2019; World Health Organization 2018), fails to inhibit IN containing the substitutions N155H, K156N, K211R, and E212T (Malet et al. 2018). We show that IN containing these substitutions (with or without the K156N polymorphism) can not only be inhibited by ViiV-919, but is hypersensitive to the ALLINI, highlighting the complementary role of STIs and ALLINIs. Examination of recent intasome structures suggests a possible basis for hypersensitivity to ALLINIs (Supplemental Figure 3.2). The N155H and directly adjacent K156N change likely directly affect Mg$^{2+}$ and H$_2$O coordination in the active site (Supplemental Figure 3.2a) (Cook et al. 2020), but, importantly, Lys-156 is closely associated with the viral DNA backbone (Supplemental Figure 3.2b). Substitution of the basic lysine with asparagine would be expected to negatively affect IN-DNA affinity, potentially destabilizing the intasome. The K211R and E212T substitutions are located on the helical CCD-CTD linker, a crucial structure for the formation of the IN-ALLINI polymer (Supplemental Figure 3.2c). Disruption of this linker through introduction of a proline “kink” or deletion of Ala-205 conveyed resistance to ALLINIs, but also impaired virus infectivity and enzymatic activity (Koneru et al. 2019). Further study is needed to determine if substitutions in this linker convey hypersensitivity to ALLINIs, as seen in this study.
These investigations should be extended to other STI resistance substitutions to determine if this is a common theme. Unfortunately, we did not observe additive or synergistic effects on strand transfer inhibition with the combination of raltegravir and ViiV-919. A more robust analysis of combination treatment should be carried out in vivo, where potent inhibition of integration by STIs and the pleiotropic effects of ALLINIs may result in additive or synergistic effects. The unique mechanism of action of ALLINIs and their action at both early and late stages of the viral replication cycle make them attractive candidates for combination with other classes of antiretroviral drugs. In addition, since antiretroviral therapy regimens involve a combination of several drugs, all antiretroviral drug candidates must be evaluated for antagonistic effects on other antiretroviral drug classes. Promisingly, no moderate or strong antagonism has been observed between the ALLINI BI-224436 and other antiretroviral classes (Fenwick et al. 2014). The ALLINI BI-224436 showed marked synergy with the nucleoside reverse transcriptase inhibitor (NRTI) tenofovir, but minor antagonism with the NRTI abacavir. IN and RT are known to interact (Dobard, Briones, and Chow 2007; Tekeste et al. 2015), and this may be the basis for the interactions observed, but further study is needed to understand the different effects of two chemotypes from the same class. Moderate synergy was observed with a selection of protease inhibitors (Fenwick et al. 2014). ALLINIs have been shown to aggregate the Gag-Pol protein (Desimmie et al. 2013), which must be accessible for the viral protease to properly initiate virion maturation. Dual inhibition of Gag-Pol processing by ALLINIs and protease inhibitors may explain the observed synergy. However, when IN is supplied to virions in trans (by fusion with Vpr), virions remain fully susceptible to ALLINI inhibition (Liu et al. 1997; Jurado et al. 2013; Fontana et al. 2015), showing that Gag-Pol may not be the target of ALLINIs. Further
study will be needed to clarify the synergy between ALLINIs and protease inhibitors. Combination experiments should be repeated with ALLINIs such as ViiV-919 to identify opportunities for potent combination.

Viral resistance often arises at a fitness cost for the virus (Hoyte et al. 2017; Nijhuis, Deeks, and Boucher 2001), and combination antiretroviral therapy should exploit this fitness defect. These data suggest that STIs and ALLINIs should be combined in future antiretroviral regimens. Rational combination of complementary antiretroviral classes will effectively prevent viral integration and prevent the development of resistance.

3.5 Acknowledgments

We thank members of the Van Duyne and Bushman laboratories for advice and technical assistance. This work was supported by ViiV Healthcare, the Penn Center for AIDS Research, the PennCHOP Microbiome Program, the Johnson Research Foundation, and the National Institutes of Health (R61-HL137063, R01-HL113252, U19-AI117950, UM1-AI126620, and T32-AI007632 to F.D.B., R01-AI129661 to F.D.B. and G.V.D., P30-AI045008 to F.D.B. and K.G., and R01-NS100081 to K.G.). G.E. was supported by the NIH Training in HIV Pathogenesis grant (T32-AI007632).
Chapter 4: Influence of the amino-terminal sequence on the structure and function of HIV integrase

The contents of this chapter have been published:


4.1 Abstract

Antiretroviral therapy (ART) can mitigate the morbidity and mortality caused by the human immunodeficiency virus (HIV). Successful development of ART can be accelerated by accurate structural and biochemical data on targets and their responses to inhibitors. One important ART target, HIV integrase (IN), has historically been studied in vitro in a modified form adapted to bacterial overexpression, with a methionine or a longer fusion protein sequence at the N-terminus. In contrast, IN present in viral particles is produced by proteolytic cleavage of the Pol polyprotein, which leaves a phenylalanine at the N-terminus (IN 1F). Inspection of available structures suggested that added residues on the N-terminus might disrupt proper protein folding and formation of multimeric complexes. We purified HIV-1 IN 1F^{1–212} and solved its structure at 2.4 Å resolution, which showed extension of an N-terminal helix compared to the published structure of IN^{1–212}. Full-length IN 1F showed increased in vitro catalytic activity in assays of coupled joining of the two viral DNA ends compared to two IN variants containing additional N-terminal residues. IN 1F was also altered in its sensitivity to inhibitors, showing decreased sensitivity to the strand-transfer inhibitor raltegravir and increased sensitivity to allosteric integrase inhibitors. In solution, IN 1F exists as monomers and
dimers, in contrast to other IN preparations which exist as higher-order oligomers. The structural, biochemical, and biophysical characterization of IN 1F reveals the conformation of the native HIV-1 IN N-terminus and accompanying unique biochemical and biophysical properties. IN 1F thus represents an improved reagent for use in integration reactions in vitro and the development of antiretroviral agents.

4.2 Introduction

Integration of a reverse-transcribed DNA copy of the HIV RNA genome into a host cell chromosome is an essential step in retroviral replication (Robert Craigie and Bushman 2012). The integrated provirus serves as a template for retroviral gene expression and the production of a new generation of virions. Integration also establishes the potential for latency, a major barrier to the treatment and cure of HIV-1 infection. Integrase (IN), the retroviral enzyme that catalyzes integration, is produced by proteolysis of the viral Gag-Pol polyprotein precursor by the virus-encoded protease. HIV-1 IN is comprised of three domains: the N-terminal (NTD), catalytic core (CCD), and C-terminal (CTD) domains. Some non-lentiviral INs contain an additional N-terminal extension domain (X. Li et al. 2011). The NTD adopts a zinc finger fold containing a conserved HHCC motif essential for activity (Mengli Cai et al. 1997; Zheng, Jenkins, and Craigie 1996; Z. Zhao et al. 2008). The CCD contains a D,D-35-E motif which binds divalent metal ions within an RNase H-like fold that comprises the active site (A. Engelman and Craigie 1992; Kulkosky et al. 1992; Dyda et al. 1994). The CTD adopts an SH3-like fold and is implicated in DNA binding (J. C.-H. Chen et al. 2000; Eijkelenboom et al. 1999). Dimers of each isolated domain have been observed (Mengli Cai et al. 1997; Dyda et al. 1994; Eijkelenboom et al. 1999, 1995; Lodi et al. 1995) and
recombinant full-length IN has been reported to exist in forms ranging from monomer to octamer (Hare, Di Nunzio, et al. 2009; A. Engelman, Bushman, and Craigie 1993; Peter Cherepanov et al. 2003; Faure et al. 2005; McKee et al. 2008; Pandey, Bera, and Grandgenett 2011).

IN carries out two catalytic reactions: 3'-processing and strand transfer (Robert Craigie and Bushman 2012), in a macromolecular complex consisting of multiple IN protomers, viral DNA, cofactors, and host cell proteins termed the intasome (A. Engelman, Bushman, and Craigie 1993; A. N. Engelman and Cherepanov 2017; M. D. Miller, Farnet, and Bushman 1997; Lewinski and Bushman 2005; Cook et al. 2020; Passos et al. 2020, 2017; Ballandras-Colas et al. 2017). Early intasome structures were determined with IN from the prototype foamy virus (PFV) (Hare, Gupta, et al. 2010; Hare, Maertens, and Cherepanov 2012; G. N. Maertens, Hare, and Cherepanov 2010; Yin et al. 2012; Maskell et al. 2015), revealing the structural details of enzymatic activity and the mechanism of action of the strand-transfer inhibitors (STIs), which displace the 3' viral DNA end from the active site, rendering the intasome nonfunctional (Cook et al. 2020; Passos et al. 2020; D. J. Hazuda et al. 2000; D. Hazuda 2012). STIs are in widespread clinical use (Mesplède and Wainberg 2013), however, as with all antiretrovirals, development of resistance is a major barrier to durable inhibition of viral replication (You et al. 2016; Mesplède and Wainberg 2015).

PFV intasome structures and homology modeling (B. C. Johnson et al. 2013; Krishnan et al. 2010) have provided important insight into HIV-1 intasome function; however, PFV IN diverges significantly from HIV-1 IN in sequence identity, interdomain linker length, and the presence of an N-terminal extension domain (Peletskaya et al. 2011). Intasome structures from Rous sarcoma virus (RSV) (Yin et al. 2016), mouse
mammary tumor virus (MMTV) (Ballandras-Colas et al. 2016), and Maedi-visna virus (MVV) (Ballandras-Colas et al. 2017) have furthered our understanding of HIV-1 IN. In contrast to PFV IN (Hare, Gupta, et al. 2010; Gupta et al. 2012), study of HIV-1 IN is challenging due to poor solubility and a propensity to aggregate (Hickman et al. 1994; Timothy M. Jenkins et al. 1996), along with inefficient catalysis in vitro (Sinha and Grandgenett 2005; Sinha, Pursley, and Grandgenett 2002; R. Craigie, Fujiwara, and Bushman 1990; Carteau, Gorelick, and Bushman 1999; Bushman and Craigie 1991; Kalpana et al. 1994; Goodarzi et al. 1995). Fusion of a small DNA-binding protein, Sso7d, that mimics the N-terminal extension domain of PFV to the N-terminus of HIV-1 IN improved catalysis and the solubility of HIV-1 intasomes (M. Li et al. 2014), and enabled structural determination of the HIV-1 core intasome complex by cryo-electron microscopy (cryo-EM) (Passos et al. 2020, 2017). However, a native HIV-1 intasome structure remains elusive.

The first N-terminal residue of HIV-1 IN is a highly conserved phenylalanine (B. Foley et al. 2018; Lightfoote et al. 1986; Steimer et al. 1986) liberated by retroviral protease cleavage from the C-terminus of reverse transcriptase. Viruses containing engineered substitutions at IN F1 are replication-incompetent (Lloyd et al. 2007), showing defects in reverse transcription and integration, characteristic of class II IN mutations such as those that disrupt the HHCC motif (Leavitt et al. 1996; Masuda et al. 1995; A. Engelman et al. 1995; A. Engelman 1999). Another closely studied NTD substitution Y15A also affects reverse transcription and integration (Takahata et al. 2016), and IN^{Y15A} is hypo-oligomeric in solution (Hare, Di Nunzio, et al. 2009; Gupta et al. 2016). Isolated IN NTD^{Y15A} is structurally constrained, adopting only one of two NTD conformational states (the E form) (Nomura, Masuda, and Kawai 2006) while the wild
type NTD adopts both the E and D forms (Mengli Cai et al. 1997). Conformational transition between E and D forms involves significant structural rearrangements in the NTD, including a change in the length of the α1 helix by 6 residues (Mengli Cai et al. 1997). The aberrant phenotypes caused by substitutions at F1 and Y15 led us to investigate the structure and function of the HIV-1 NTD in more detail.

IN is often produced for laboratory studies by bacterial overexpression in vitro with an N-terminal methionine (IN MF) (Gupta et al. 2016, 2014, 2010) or as an N-terminal fusion protein, such as the Sso7d-IN fusion (M. Li et al. 2014; Passos et al. 2017, 2020). Solution structures of the isolated NTD were determined from constructs purified with a cleavable N-terminal affinity tag (Mengli Cai et al. 1997; M. Cai et al. 1998), so that thrombin cleavage of the fusion protein left three residues (G-S-H-) preceding F1 (IN GSH). In the solution structure of IN GSH\textsubscript{NTD} (Mengli Cai et al. 1997), the backbone carbonyl of F1 contributes the first hydrogen bond of the α1 helix. The solution structure of another variant, IN GSH\textsubscript{NTD}^{H12C}, which contains a substitution in the HHCC Zn-binding motif, shows a different N-terminal structure: the carbonyl of F1 is not involved in a hydrogen bond, L2 is displaced, and the α1 helix begins with G4 (M. Cai et al. 1998). The only crystal structure containing the HIV-1 IN\textsubscript{NTD} (PDB: 1K6Y) (J.-Y. Wang et al. 2001) consists of a two-domain truncated form (NTD-CCD) also purified using an N-terminal affinity tag and subsequent thrombin cleavage, leaving 3 residues (G-S-H-) preceding F1 (Timothy M. Jenkins et al. 1996; J.-Y. Wang et al. 2001). In this case as well, the α1 helix is shortened, suggesting that the extra N-terminal residues might be disrupting native folding of the α1 helix.

Four NTDs in two structurally distinct positions exist in the HIV-1 core intasome complex cryo-EM structures determined with Sso7d-IN (Passos et al. 2017, 2020). One
NTD, positioned close to the viral DNA and the CCD responsible for catalysis, forms NTD-NTD interactions in the dodecameric HIV-1 intasome and the hexadecameric MVV intasome (Ballandras-Colas et al. 2017). The α1 helix of this NTD is shortened in the first HIV-1 tetrameric intasome structure where it begins with Asp 3 (Passos et al. 2017). The α1 helix is extended in four of five recent intasome structures, with only one structure showing partial disruption (Passos et al. 2020). The second NTD does not interact with the viral DNA and is distant from the active site. This NTD does not form NTD-NTD interactions in dodecameric or hexadecameric intasomes and shows a range of α1 helical structures: disordered, partially unstructured, and extended (Passos et al. 2020). Intasomes of a closely-related simian immunodeficiency virus were prepared with IN purified with an N-terminal affinity tag and subsequent human rhinovirus 3C protease cleavage, leaving 3 residues (G-P-G-) preceding F1 (Cook et al. 2020). The NTDs in these structures show extended α1 helices.

In this paper, we report a purification scheme of wild type IN with phenylalanine as the N-terminal residue (IN 1F), and associated alterations in the N-terminal structure and IN function. IN 1F was purified with an N-terminal affinity tag, which, when removed, leaves phenylalanine at position 1. We report a two-domain NTD-CCD crystal structure of IN 1F that shows a continuous helical fold beginning with the backbone carbonyl of F1, in contrast to the existing IN GSH$_{\text{NTD-CCD}}$ structure (J.-Y. Wang et al. 2001). IN 1F also shows greater concerted integration activity in vitro compared to IN GSH and IN MF. IN 1F is altered in its sensitivity to inhibitors, showing decreased sensitivity to the strand-transfer inhibitor raltegravir and increased sensitivity to allosteric integrase inhibitors (ALLINIs). Biophysical characterization reveals that IN 1F has oligomeric properties distinct from previously studied recombinant IN constructs. We propose that
HIV-1 IN 1F more closely recapitulates the structure and functions of IN found in authentic HIV infection.

4.3 Results

4.3.1 Cloning and purification of HIV-1 integrase with a native N-terminus

To determine the biochemical and structural properties of HIV-1 IN with a phenylalanine at the N-terminus, we cloned NL4-3 IN into an expression vector containing an N-terminal His7-FLAG-SUMO tag immediately preceding F1. The SUMO protease Ulp1 cleaves at a G-G-/X motif (with the cleavage site indicated by /, with X being any residue except proline) (Mossessova and Lima 2000). This allows for purification of wild-type IN with a native N-terminus (“IN 1F”) by Ulp1 cleavage at the sequence G-G-/-F (Supplemental Figure 4.1). To compare to IN with a non-native N-terminus, we inserted additional N-terminal residues preceding F1. IN GSH contains the three residues (G-S-H) that remain after thrombin cleavage, as used to determine the structure of IN GSH$_{\text{NTD-CCD}}$ (PDB: 1K6Y) (J.-Y. Wang et al. 2001), and IN MF contains an N-terminal methionine found in constructs commonly used for bacterial overexpression (Gupta et al. 2016, 2014, 2010). A nickel-affinity step captures Ulp1 and the cleaved affinity tag and subsequent size-exclusion chromatography yields a highly pure final product (Supplemental Figure 4.1).

4.3.2 Crystallization of an IN 1F NTD-CCD derivative

To investigate structural differences between IN 1F and IN GSH, we created an IN 1F$_{\text{NTD-CCD}}$ construct containing the same solubility-enhancing substitutions (W131D, F139D, and F185K) used to determine the structure of IN GSH$_{\text{NTD-CCD}}$ (J.-Y. Wang et al. 2001).
2001). Affinity purification, Ulp1 cleavage, and size-exclusion chromatography yielded a highly pure final product (Supplemental Figure 4.1) that readily crystallized as described previously (J.-Y. Wang et al. 2001). The structure was solved by molecular replacement, using the existing NTD-CCD structure (PDB: 1K6Y) as a search model. Four copies of both the NTD and the CCD were present in the asymmetric unit (Figure 4.1a), with the inter-domain linker (residues 47-55) unresolved in the electron density. In the structure of IN GSH$\text{NTD-CCD}$, each NTD is assigned to a “distal” position relative to the CCD (Supplemental Figure 4.2). However, in the crystal structure of the HIV-2 IN$\text{NTD-CCD}$ complexed with the lens epithelium derived growth factor (LEDGF) integrase binding domain (IBD) (PDB: 3F9K) (Hare, Shun, et al. 2009), the interdomain linker is well-defined in the electron density, placing the NTDs in a “proximal” position relative to the CCD (Supplemental Figure 4.2). This is also the favored position for the NTDs in small angle X-ray scattering (SAXS) analysis of IN NTD-CCD coexpressed with the LEDGF IBD (Gupta et al. 2010). In the IN 1F$\text{NTD-CCD}$ structure, the unresolved 10-residue linker would be long enough to span the unobstructed distance of 28.7-31.8 Å to position the NTDs in a “proximal” position. We have therefore defined the NTDs in the “proximal” orientation relative to the CCDs, as observed in the HIV-2 IN$\text{NTD-CCD}$ structure (Figure 4.1a). Crystallographic statistics are summarized in Table 4.1.
Figure 4.1. Structure of HIV-1 integrase with a native N-terminus (PDB 6VRG). a) Structure of IN 1F\textsubscript{NTR-CCD}, NTDs are colored in red and CCDs are colored in blue, Zn\textsuperscript{2+} (grey), K\textsuperscript{+} (purple), and Phosphate (orange and red) atoms are shown as spheres. b) Comparison of the α1 helix between IN 1F and IN GSH (PDB 1KSY). The IN 1F NTD adopts a helical structure starting from the carbonyl of F1. The IN GSH NTD shows a disruption of the α1 helix. c) View highlighting differences between IN 1F and IN GSH at the N-terminus; with a deviation of 4.6 Å in the peptide backbone at L2 and a 10.4 Å deviation in side chain position. This change is accompanied by a flip of ~180° in the orientation of the N-terminus. 2Fo-Fc electron density (d) and simulated annealing omit (e) maps contoured at 1.5 σ unambiguously demonstrate the N-terminal structure of IN 1F.
4.3.3 Structure of the IN 1F NTD-CCD construct

The overall structure of IN 1F<sub>NTD-CCD</sub> is highly similar to IN GSH<sub>NTD-CCD</sub> (global RMSD: 0.90 Å). A phosphate ion is found near the active site of each CCD. Each copy of the NTD folds into a 3-helix motif coordinating a Zn<sup>2+</sup> ion with residues H12, H16, C40, and C43. A potassium ion is coordinated by the carbonyl oxygens of V37, A38, C40,
Close inspection of the N-terminus reveals differences between IN 1F and IN GSH (Figure 4.1b). In the asymmetric unit of IN 1F, two secondary structures are observed at the N-terminus. The α1 helix in chains A and C begins as a hydrogen-bonded turn at the backbone carbonyl of F1, while in chains B and D, a canonical alpha helix begins at the backbone carbonyl of F1 (Figure 4.1b, Supplemental Figure 4.3). In IN GSH, the α1 helix does not begin until D3 due to a shift in the L2 side chain by ~10 Å, accompanied by a ~4.6 Å displacement of the peptide backbone at L2 (Figure 4.1b,c, Supplemental Figure 4.3). F1 is in a similar position in IN 1F and IN GSH, where it caps a hydrophobic core in the NTD made up of I5, L28, P29, and V32. The N-terminal amino group also differs between these two structures due to the peptide backbone displacement at L2. In IN GSH, the N-terminal amino group is oriented toward the C-terminal end of the α2 helix, whereas in IN 1F, it is flipped ~180° and oriented toward the α3 helix of a neighboring NTD. The same NTD-NTD interface is observed in dodecameric HIV-1, hexadecameric MVV, and SIV intasome structures (Cook et al. 2020; Passos et al. 2017, 2020; Ballandras-Colas et al. 2017), and the NTDs modeled at this position adopt extended α1 helical structures in four of six structures (Supplemental Figure 4.3). The NTDs that do not form an NTD-NTD interface show a variety of structures: disordered, partially unstructured, and extended (Supplemental Figure 4.3). Difference maps and simulated annealing omit maps calculated around the N-terminus of each protomer of the IN 1F_{NTD-CCD} structure confirmed the observed differences between the N-termini of IN 1F and IN GSH (Figure 4.1d,e).

4.3.4 Activity of IN 1F in vitro
IN carries out two catalytic functions, 3'-processing and strand transfer, which can be replicated in vitro using fluorescently-labeled oligonucleotides that mimic the viral long terminal repeat (LTR). To assay 3'-processing, we used a 3'-fluorescently-labeled double-stranded oligonucleotide mimicking the viral LTR to monitor release of the terminal dinucleotide (5'-GT-3') using fluorescence polarization (Merkel et al. 2009). The unprocessed oligonucleotide emits highly polarized fluorescence. Upon cleavage by IN, the released dinucleotide emits fluorescence with low polarization. In the presence of Mg\(^{2+}\) and Mn\(^{2+}\), IN MF and IN 1F showed similar 3'-processing activities (Figure 4.2).

**Figure 4.2. 3'-processing activity in vitro.** a) Diagram of 3'-processing fluorescence polarization assay. The double stranded oligo containing a 3’ fluorescent label exhibits high fluorescence polarization. Cleavage and release of the terminal dinucleotide causes a decrease in fluorescence polarization. 5'-ends are designated by filled circles and the fluorophore is designated by the green star. b) 3'-processing activity of IN 1F compared to IN GSH (left) and IN 1F compared to IN MF (right) in the presence of Mg\(^{2+}\). c) 3’ processing activity of IN 1F compared to IN GSH (left) and IN 1F compared to IN MF (right) in the presence of Mn\(^{2+}\). Data are plotted as mean ± SD. ** denotes \(P < 0.01\) and **** denotes \(P < 0.0001\).

IN carries out two catalytic functions, 3'-processing and strand transfer, which can be replicated in vitro using fluorescently-labeled oligonucleotides that mimic the viral long terminal repeat (LTR). To assay 3'-processing, we used a 3'-fluorescently-labeled double-stranded oligonucleotide mimicking the viral LTR to monitor release of the
terminal dinucleotide (5’-GT-3’) using fluorescence polarization (Merkel et al. 2009). The unprocessed oligonucleotide emits highly polarized fluorescence. Upon cleavage by IN, the released dinucleotide emits fluorescence with low polarization. In the presence of Mg²⁺ and Mn²⁺, IN 1F, IN GSH, and IN MF showed similar 3’-processing activities (Figure 4.2).

To assay strand transfer activity, we used 5’-fluorescently-labeled oligonucleotides mimicking the viral LTR and a supercoiled plasmid mimicking nucleosomal DNA (Figure 4.3a). Concerted integration of two viral LTRs by IN results in linearization of the supercoiled plasmid and incorporation of the fluorescent label. Strand-transfer activity in the presence of Mg²⁺ and Mn²⁺ was influenced by NaCl concentration, with the highest level of concerted integration occurring at 150mM NaCl in the presence of Mg²⁺ and 200-250mM NaCl in the presence of Mn²⁺ (Supplemental Figure 4.4). In identical assay conditions, IN 1F showed superior concerted integration activity, resulting in the formation of 2 LTR coupled products, as compared to IN GSH and IN MF at all time points measured (Figure 4.3b,c). This difference was observed in the presence of either Mg²⁺ or Mn²⁺.
Figure 4.3. Strand transfer activity in vitro. a) Diagram of gel-based strand transfer assay. A pre-processed double-stranded oligo containing a 5' fluorescent label is integrated into a supercoiled target plasmid (pUC19) in the presence of integrase and cofactor (Mg²⁺ or Mn²⁺). Single strand integration results in the formation of a tagged circle product accompanied by the relaxation of supercoiling. Concerted integration results in the formation of a 2 LTR coupled product, accompanied by the linearization of the plasmid. Reaction products are separated by agarose gel electrophoresis. b) Example gels of IN 1F and IN MF strand transfer activity. The slowest-migrating band represents single strand integration events (tagged circle) with the linearized concerted integration product (2 LTR coupled) migrating further. The supercoiled target plasmid is visible in the ethidium bromide stain as the furthest-migrating band. Unintegrated fluorescent oligo is observed at the bottom of the gel. c) Quantification of IN 1F and IN GSH strand transfer activity. Single strand integration events are shown in the top panels and concerted integration events are shown in the bottom panels. IN 1F shows greater concerted integration activity than IN GSH in the presence of either Mg²⁺ or Mn²⁺. d) Quantification of IN 1F and IN MF strand transfer activity. Single strand integration events are shown in the top panels and concerted integration events are shown in the bottom panels. IN 1F shows greater concerted integration activity than IN MF in the presence of either Mg²⁺ or Mn²⁺. e) Inhibition of strand transfer activity by raltegravir. Raltegravir more potently inhibits both the single strand (left) and concerted integration (right) activity of IN GSH and IN MF as compared to IN 1F. Data before normalization are plotted in Supplemental Figure 5. Data are plotted as mean ± SD of 3 replicates. * denotes P < 0.05, ** denotes P < 0.01, and **** denotes P < 0.0001.
A partial reaction, the integration of a single LTR oligo, results in relaxation of the supercoiled plasmid and incorporation of the fluorescent label. Quantification of the fluorescently tagged, relaxed-circular plasmid indicates single-ended integration activity. Single-end activity, resulting in the formation of tagged circle products, was not improved by IN 1F as compared to IN GSH or IN MF (Figure 4.3c).

Treatment with the strand transfer inhibitor raltegravir more potently inhibited both the single-strand and concerted integration activity of IN GSH and IN MF as compared to IN 1F (Figure 4.3d, Supplemental Figure 4.5). The IC₅₀ for inhibiting concerted integration was 125 nM (95% CI: 83-186 nM), 109 nM (95% CI: 83-142 nM), and 370 nM (95% CI: 270-508 nM) for IN GSH, IN MF, IN 1F, respectively. The IC₅₀ for inhibiting single-strand integration was 253 nM (95% CI: 202-318 nM), 223 nM (95% CI: 141-356 nM), and 1.13 µM (95% CI: 0.73-1.8 µM) for IN GSH, IN MF, and IN 1F, respectively.

4.3.5 Response of IN 1F and IN MF to ALLINIs

The allosteric inhibitors of integrase (ALLINIs) (Christ et al. 2010; Fader et al. 2014; H. Wang et al. 2012) are a class of small molecule inhibitors that block the interaction of IN with LEDGF and cause aberrant aggregation of IN (Gupta et al. 2016, 2014; Jurado et al. 2013). ALLINIs aggregate recombinant IN in vitro, causing turbidity that can be measured by light scattering (Gupta et al. 2016, 2014, 2010). Using this approach, we measured the sensitivity of IN 1F, IN GSH, and IN MF to ALLINIs (Figure 4.4). ALLINI-induced aggregation is NaCl-dependent, so we tested aggregation at NaCl concentrations from 250 mM to 1 M. At 1 M NaCl, no aggregation was observed by the
ALLINIs BI-224436 (Fader et al. 2014), BI-D (Tsantrizos et al. 2009), or CX04328 (Compound 6 from Christ et al. (Christ et al. 2010)). At NaCl concentrations where ALLINI-induced aggregation was observed, ALLINIs induced equal or greater
aggregation of IN 1F as compared to IN GSH or IN MF. BI-224436, BI-D, and CX04328 aggregated IN 1F more than IN GSH at 300-500 mM NaCl, 300-400 mM NaCl, and 300-500 mM NaCl, respectively. Significant ALLINI-induced aggregation of IN GSH was only observed at 250 mM NaCl, where IN 1F was observed to aggregate in the absence of ALLINI. BI-224436, BI-D, and CX04328 aggregated IN 1F more than IN MF from 350-500 mM NaCl, 350-400 mM, and 500 mM NaCl, respectively. At lower NaCl concentrations, ALLINIs induced aggregation of IN 1F and IN MF to an equal extent.

4.3.6 Solution properties of IN with a native N-terminus

One possible explanation for the improved activity of IN 1F is that folding associated with the native N-terminal sequence changes the oligomerization state. To investigate this possibility, we analyzed IN 1F and IN MF by size-exclusion chromatography in line with multi-angle light scattering (SEC-MALS) to determine the oligomeric state in solution. Both IN 1F and IN MF showed mass profiles consistent with a monomer-dimer transition (32-64 kD, expected MW of monomer: 32 kD) at eluted concentrations of ~8-10 µM, as well as retention times consistent with a mixture of monomers and dimers (Figure 4.5a). Sedimentation velocity analytical ultracentrifugation (SV-AUC) experiments performed at similar concentrations and temperatures confirmed the presence of monomers and dimers with the presence of two discrete species at ~2.8 S and ~4 S, respectively (Figure 4.5b). Sedimentation equilibrium analytical ultracentrifugation (SE-AUC) analysis at 4°C and similar concentrations also confirmed the presence of monomers and dimers, and global fitting of a monomer-dimer equilibrium yielded a K_d of 60 ± 6 µM and 127 ± 23 µM for IN 1F and IN MF, respectively (Figure 4.5c, Table 4.2). An attempt to fit a dimer-tetramer equilibrium could only be
accomplished with a $K_d > 1$ mM. Therefore, no evidence of tetramers was observed by three biophysical methods, in contrast to prior studies performed under similar conditions with IN expression constructs with an N-terminal methionine (Gupta et al. 2016, 2014, 2010), N-terminal thrombin (Hickman et al. 1994), or human rhinovirus 3C protease (Hare, Di Nunzio, et al. 2009; Hare, Shun, et al. 2009) cleavage sequences.

![Graphs of IN 1F and IN MF](image)

**Figure 4.5. Biophysical analysis of IN 1F and IN MF.** a) SEC-MALS analysis of IN 1F and IN MF. Both the $M_w$ (weight-average molecular mass) from multangle light scattering and retention times are consistent with mixtures of monomers and dimers for both IN 1F and IN MF (expected MW of monomer: 32kDa). b) Sedimentation velocity analysis of IN 1F and IN MF shows distinct populations of monomer and dimer in solution. c(S) distributions derived from the fitting of the Lamm equation are shown. c) Sedimentation equilibrium analysis of IN 1F and IN MF indicates the presence of monomers and dimers in solution at 4°C. Globally fit radial distributions for 8.9 µM (IN 1F) and 6 µM (IN MF) in a monomer-dimer model are shown. Table 2 provides the association properties derived from this analysis.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Concentrations (µM)</th>
<th>Speeds (krpm)</th>
<th>Model Fit</th>
<th>Mass (Da)</th>
<th>$K_d$ (µM)</th>
<th>Global Reduced $\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>IN 1F</td>
<td>8.9, 14.5</td>
<td>16,20,24</td>
<td>M-D</td>
<td>32330</td>
<td>59.8 ± 6</td>
<td>2.7</td>
</tr>
<tr>
<td>IN MF</td>
<td>6.0, 7.9</td>
<td>16,20,24</td>
<td>M-D</td>
<td>32199</td>
<td>126.7 ± 23</td>
<td>1.3</td>
</tr>
</tbody>
</table>

*Calculation of mass of monomer from sequence*
The monomer-dimer behavior of IN 1F in solution also differs from IN with an N-terminal methionine, C-terminal intein cleavage site, and the solubility-enhancing substitution F185H, which exists as a mixture of dimers and tetramers in solution (Gupta et al. 2016, 2010) and is replication-competent in virus (A. Engelman et al. 1997). Introduction of the F185H substitution into IN 1F (IN 1F\textsuperscript{F185H}) resulted in the formation of dimers and a spectrum of higher-order aggregates in solution, as determined by SEC-MALS (Supplemental Figure 4.6). IN 1F\textsuperscript{F185H} retained similar 3'-processing activity as compared to IN 1F, but showed a significant decrease in single strand and concerted strand transfer activity (Supplemental Figure 4.6), indicating the effect of the oligomeric state of IN on strand transfer activity.

4.4 Discussion

In this paper, we report the construction and purification of IN with a native N-terminus (IN 1F). The crystal structure of IN 1F\textsubscript{NTD-CCD} reveals an extended α1 helix starting with F1, as compared to IN GSH\textsubscript{NTD-CCD} with a shortened helix. Despite the remainder of the structure showing little to no difference, this change in the N-terminus is sufficient to improve concerted integration activity. In contrast, the 3’-processing and single strand integration activities were not affected. We also observed a change in sensitivity to IN-targeting antiretroviral drugs. IN 1F was less sensitive to the STI raltegravir and more sensitive to ALLINI-induced aggregation. We suggest that IN 1F will be useful in studies of IN function and response to inhibitors in the future.

The zinc finger fold of the NTD is shared with other DNA-binding proteins (Otwinowski et al. 1988; van Pouderoyen et al. 1997; Xu et al. 1995), with residues homologous to positions 1-3 in IN located adjacent to the phosphate backbone of DNA.
In retroviral intasomes, the NTD binds to the distal viral DNA ends (A. N. Engelman and Cherepanov 2017; Passos et al. 2020, 2017; Ballandras-Colas et al. 2017; Hare, Gupta, et al. 2010; Yin et al. 2016; Ballandras-Colas et al. 2016). However, unlike other helix-turn-helix binding proteins, the NTD does not insert a helix into the major groove of DNA, and F1 is distant from the phosphate backbone. The effect of the N-terminal disruption in IN GSH and IN MF is unclear, because the change is not expected to disrupt a tetrameric intasome. In the hexadecameric maedi-visna virus intasome, however, two pairs of NTDs are closely oriented head-to-head (Ballandras-Colas et al. 2017), forming a nearly identical NTD-NTD interface as that observed in the structure of IN 1F and IN GSH_{NTD-CCD} (J.-Y. Wang et al. 2001). This hydrophobic dimerization interface would involve significant contributions from F1, in contrast to the dimerization interface of the isolated NTD which mainly involves the α3 helix (Mengli Cai et al. 1997). Additional N-terminal residues, such as the N-terminal Sso7d-IN fusion, could induce a steric clash (Passos et al. 2017). It is possible that such a disruption explains the presence of heterogeneous, poorly resolved higher-order intasomes reported in the cryo-EM studies of HIV-1 Sso7d-IN intasomes (Passos et al. 2017, 2020). Additionally, disruption of the α1 helix could affect binding to LEDGF, as the NTD cooperates with the CCD in binding LEDGF (Hare, Di Nunzio, et al. 2009; Hare, Shun, et al. 2009). Destabilization of the intasome and disruption of the IN-LEDGF interaction are possible explanations for the differences in concerted integration activity and STI sensitivity between IN 1F, IN GSH, and IN MF.

Surprisingly, we found IN 1F to be more potently aggregated by ALLINIs compared to IN GSH and IN MF. ALLINIs cause the formation of open polymers of IN mediated by CCD-CTD interactions (Gupta et al. 2016), and it is not immediately clear
how addition of N-terminal residues affects this process. Previously, we have shown that the NTD is dispensable for ALLINI-induced aggregation (Gupta et al. 2014), although others have reported that constructs lacking the NTD are resistant to ALLINI-induced aggregation (Koneru et al. 2019), suggesting that the NTD plays a role in modulating ALLINI-induced aggregation. In multiple structures (Hare, Di Nunzio, et al. 2009; J.-Y. Wang et al. 2001; Hare, Shun, et al. 2009), the NTD interacts with the CCD in a manner expected to clash with the CCD-CTD interactions observed in the ALLINI-induced IN polymer. An effect on competition between the NTD and CTD for CCD binding may explain the difference in ALLINI potency between IN 1F, IN GSH, and IN MF. Recently, IN tetramers have been implicated as the preferred target of ALLINIs (Koneru et al. 2019), but we show that IN 1F, which is a mixture of monomers and dimers in solution, is aggregated by ALLINIs. However, aggregation is NaCl-dependent, and we have not determined the oligomeric state of IN 1F at lower NaCl concentrations. IN GSH remains soluble at NaCl concentrations that lead to aggregation of IN 1F in the absence of ALLINI, demonstrating that additional N-terminal residues can improve solubility. This is consistent with the observation of improved solubility of Sso7d-IN (M. Li et al. 2014) and PFV IN, which harbors an N-terminal extension domain (Hare, Gupta, et al. 2010; Gupta et al. 2012). Additional experiments are needed to determine the details of ALLINI-induced polymer initiation and propagation.

Wild type IN 1F is a mixture of monomers and dimers in solution, which differs from previously reported IN preparations containing substitutions at F185 or additional N-terminal residues which are a mixture of dimers and tetramers (Hare, Di Nunzio, et al. 2009; Timothy M. Jenkins et al. 1996; Gupta et al. 2016, 2010). We found that the substitution F185H in the IN 1F background resulted in the formation of higher-order
species in solution. NTD-CCD interactions between residues such as E11 and K186 have been shown to be important for tetramerization (Hare, Di Nunzio, et al. 2009; Koneru et al. 2019), and we have now shown that modification of the adjacent residue F185 affects oligomerization in the context of a native N-terminus. Notably, the construct used to solve the only HIV IN crystal structure with a naturally-occurring F185, HIV-2 IN\textsubscript{NTD-CCD} co-expressed with LEDGF IBD, was dimeric in solution (Hare, Shun, et al. 2009). In this structure, the interdomain linker is clearly resolved in the electron density, showing that the NTD contacts the CCD in a “proximal” orientation. This is in contrast to the IN GSH\textsubscript{NTD-CCD} structure (PDB: 1K6Y) where the interdomain linker is not resolved, short interdomain linkers are assigned, and each NTD is in a “distal” orientation (J.-Y. Wang et al. 2001). The interdomain linker is not resolved in our IN 1F\textsubscript{NTD-CCD} structure, but we favor longer interdomain linkers, positioning each NTD in a “proximal” orientation, as this is the orientation observed in the HIV-2 IN\textsubscript{NTD-CCD}-LEDGF co-crystal structure (Hare, Shun, et al. 2009). Additional work is needed to understand the effect of substitutions at F185 and K186 on NTD-CCD interactions in dimeric forms of IN.

4.5 Conclusions

HIV IN containing a native N-terminus adopts a distinct structural configuration, shows improved activity in vitro, and manifests altered sensitivity to inhibitors. Because it mimics the form of IN produced by proteolytic cleavage in the maturing virion, IN 1F provides an improved reagent for the study of IN activity in vitro and for use in antiviral drug development.

4.6 Contributions
GE: Conceptualization, Formal Analysis, Investigation, Methodology, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; KG: Conceptualization, Formal Analysis, Investigation, Methodology, Visualization, Software, Writing – Original Draft Preparation, Supervision, Writing – Review & Editing, Funding Acquisition; AA: Formal Analysis, Investigation, Methodology, Visualization; JZ: Investigation, Methodology; YH: Investigation, Methodology; MC: Formal Analysis, Investigation, Software; FDB: Conceptualization, Methodology, Supervision, Writing – Review & Editing, Funding Acquisition; GVD: Conceptualization, Investigation, Methodology, Supervision, Writing – Review & Editing, Funding Acquisition

4.7 Acknowledgments

We thank Paul Bates, Robert Doms, Rahul Kohli, and Mitchell Lewis for suggestions and comments, and Laurie Zimmerman for help with the manuscript. This work was supported by the Penn Center for AIDS Research, the PennCHOP Microbiome Program, the Johnson Research Foundation, and the National Institutes of Health (R61-HL137063, R01-HL113252, U19-AI117950, UM1-AI126620, and T32-AI007632 to F.D.B., R01-AI129661 to F.D.B. and G.V.D., P30-AI045008 to F.D.B. and K.G., and R01-NS100081 to K.G.). G.E. was supported by the NIH Training in HIV Pathogenesis grant (T32-AI007632). SEC-MALS and AUC measurements were performed at the Johnson Research Structural Biology and Biophysics Core Facility. The crystallographic data were obtained at the 17-ID-1 (AMX) beamline at the National Synchrotron Light Source II, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Brookhaven National Laboratory.
Chapter 5: Characterization of the catalytically active form of integrase, the intasome

5.1 Abstract

Integrase, the enzyme responsible for the insertion of the HIV genome into a host cell, is an important target of antiretroviral therapy. The main class of antiretrovirals that inhibit integrase, the strand transfer inhibitors, target the catalytically active complex consisting of multiple copies of integrase, the viral DNA, and the host cell DNA, termed the intasome. The first structures of HIV-1 intasomes have recently been determined, accomplished by N-terminal fusion of a DNA-binding domain not found in vivo that increases activity and solubility. We recently reported the purification and characterization of HIV-1 integrase with a native N-terminus. Here, we report progress toward the goal of structural, biochemical, and biophysical characterization of the native HIV-1 intasome. We were able to purify integrase with the host cell co-factor LEDGF, which improves solubility and activity, and is an important component of intasome preparations. Using several approaches for intasome preparation, we observe the formation of intasomes with multiple readouts. These data set the stage for the structural, biochemical, and biophysical determination of the native HIV-1 intasome, a significant unmet need in the field.

5.2 Introduction

Integration of a reverse-transcribed DNA copy of the viral RNA genome is an essential step in the retroviral replication cycle. Integration enables transcription of a new generation of RNA genomes and subgenomic RNAs encoding viral proteins, as well as
establishing the potential for latency. Integration is performed by the viral integrase (IN) through two sequential reactions, 3’-processing and strand transfer (Robert Craigie and Bushman 2012). In 3’-processing, IN catalyzes the hydrolysis of the viral DNA, releasing a pair of nucleotides from the invariant CA dinucleotide sequence. The newly exposed 3’-hydroxyl group serves as the nucleophile in attacking a host cell DNA 5’-phosphate, covalently linking the viral genome to host chromatin (Fujiwara and Mizuuchi 1988; Brown et al. 1989; A. Engelman, Mizuuchi, and Craigie 1991). Both reactions are transesterification reactions (Delelis et al. 2008) and are carried out in a nucleoprotein complex consisting of IN, the viral DNA ends, and associated host cell proteins termed the intasome (Hare, Maertens, and Cherepanov 2012). One important host cell protein, LEDGF, has been extensively characterized (Peter Cherepanov et al. 2003; Turlure et al. 2004; Emiliani et al. 2005; G. Maertens et al. 2003). Host cell DNA repair enzymes subsequently repair the strand breaks and gaps (Brin et al. 2000).

The first structure of a retroviral intasome was determined from the prototype foamy virus (PFV) using X-ray crystallography (Hare, Gupta, et al. 2010), revealing the architecture of a tetrameric intasome and the structure of the core intasome complex. Intasome structures from other retroviral genera have revealed higher-order multimeric assemblies with a conserved intasome core (Ballandras-Colas et al. 2016; Yin et al. 2016; Ballandras-Colas et al. 2017; Cook et al. 2020), and a series of structures of reaction intermediates elucidated the sequence of catalysis (Hare, Maertens, and Cherepanov 2012; G. N. Maertens, Hare, and Cherepanov 2010; Yin et al. 2012). However, a structure of the clinically important HIV-1 intasome proved difficult to determine. Advances in cryo-electron microscopy and the creation of an engineered hyperactive HIV-1 IN construct ultimately enabled the determination of HIV-1 intasome
structures (M. Li et al. 2014; Passos et al. 2017, 2020). HIV-1 intasome preparations resulted in a heterogeneous mixture of multiple oligomeric species. The best-resolved species was tetrameric, like the PFV intasome, which is not surprising, as the hyperactive HIV-1 IN construct contains an N-terminal fusion domain mimicking the N-terminal extension domain of PFV IN (M. Li et al. 2014). Higher-order species including dodecamers and hexadecamers were also observed, raising questions about the structure of the native HIV-1 intasome. Does native HIV-1 IN adopt multiple oligomeric intasome forms, and if so, what is the functional relevance of this conformational heterogeneity? What are the consequences of the non-native N-terminal fusion? These questions motivate our work to determine the structure of the native HIV-1 intasome.

Additional motivation comes from the fact that millions of people worldwide depend on antiretroviral therapy to combat the morbidity and mortality of HIV and to prevent its spread (“UNAIDS Data 2019” 2019). Antiretrovirals targeting the retroviral IN protein are in widespread clinical use and are part of recommended first-line regimens (Günthard et al. 2016; “AIDSinfo” 2019; World Health Organization 2018). This clinical success belies the fact that the structure of the native HIV-1 intasome -- the target of the strand transfer inhibitors -- remains to be determined. Studies of the conserved intasome core (Hare, Gupta, et al. 2010; Pandey et al. 2014) have revealed the mechanism of action of the strand transfer inhibitors, which bind within the two core catalytic sites, displace the 3’ viral DNA ends, and prevent strand transfer. However, differences between PFV structures reveals avenues for STI optimization, underscoring the importance of accurate structure models for drug development (Cook et al. 2020; Passos et al. 2020).
Recently, our group has reported the purification and structural and biochemical characterization of native HIV-1 IN (IN 1F) for the first time (Eilers et al. 2020). Importantly, this construct does not contain engineered additional residues at the N-terminus, and therefore accurately mimics the form of IN present in virions. IN 1F retains a high level of activity in vitro and has favorable solution properties, making it an attractive candidate for structure determination, complex formation with host cell proteins, and intasome preparation. Here, we report progress in complex formation of IN 1F with the host cell protein LEDGF and intasome preparation. IN 1F forms highly pure, monodisperse, and catalytically active complexes with the host cell protein LEDGF. Utilizing published approaches of intasome preparation, we show that IN 1F can be successfully assembled into IN-DNA complexes. These results set the foundation for future work toward the structural, biochemical, and biophysical characterization of the native HIV-1 intasome.

5.3 Results

5.3.1 Integrase-LEDGF complex formation

We have previously reported the purification of wild type HIV-1 integrase with a native N-terminus (IN 1F) (Eilers et al. 2020). Extensive biophysical characterization of this construct shows that it exists as a monomer-dimer mixture in solution, distinct from previously characterized IN constructs that exist as higher-order oligomers (Gupta et al. 2014, 2016, 2010; Hickman et al. 1994; Hare, Di Nunzio, et al. 2009; Hare, Shun, et al. 2009). We have extensively characterized IN constructs in complex with LEDGF (Gupta et al. 2010) which form tetramers. To extend this data, we co-expressed IN 1F with the LEDGF integrase binding domain (LEDGF_{IBD}), which is known to bind to and stabilize IN
(McKee et al. 2008; Hare, Di Nunzio, et al. 2009) and has been shown to be important for intasome formation (Balandras-Colas et al. 2017; Passos et al. 2017). Initial experiments were performed without removing the N-terminal affinity tag to evaluate consistency with previous results. Affinity purification of LEDGF$_{IBD}$ enabled co-purification of IN 1F and LEDGF$_{IBD}$ (Figure 5.1a). Electrophoretic analysis of this complex revealed

![Image of electrophoretic analysis](image.png)

**Figure 5.1. Co-Purification of IN 1F with LEDGF$_{IBD}$.** a) SDS-PAGE analysis of a complex of LEDGF$_{IBD}$ and HFS-IN 1F obtained by purification of affinity-tagged LEDGF$_{IBD}$ and subsequent cleavage to liberate LEDGF$_{IBD}$, LEDGF$_{IBD}$, and HFS-IN 1F eluted together from size-exclusion chromatography as a highly pure complex with apparent stoichiometry of 2:4 IN:LEDGF$_{IBD}$. Shown is a pool of fractions taken across the major peak at half-height. The IN construct retains an N-terminal affinity tag that can be utilized for subsequent purification steps (e.g. after intasome preparation), or that can be cleaved to liberate the native N-terminus. b) A complex of LEDGF$_{IBD}$ and HFS-IN$_{NTD-CCD}$ could be isolated in the same manner as in panel (a). This complex has an apparent equimolar stoichiometry. c) Size exclusion chromatography in line with multi-angle light scattering (SEC-MALS) analysis of the HFS-IN 1F-LEDGF$_{IBD}$ complex. Retention time relative to MW standards and absolute molecular mass determination indicate a 4:2 IN:LEDGF$_{IBD}$ complex.
an apparent stoichiometry of 2:1 IN:LEDGF_{IBD}. Characterization of this complex by size exclusion chromatography in line with multi-angle light scattering (SEC-MALS) revealed an absolute molecular mass of approximately 253 kDa, consistent with a tetramer of IN bound to two molecules of LEDGF_{IBD} (Figure 5.1c,d). A two-domain construct of IN containing the NTD and CCD (IN_{NTD-CCD}) is competent for binding to LEDGF_{IBD}, and has been previously characterized as a 2:2 IN_{NTD-CCD}:LEDGF_{IBD} complex (Gupta et al. 2010). We successfully isolated a complex of IN_{NTD-CCD} (with a retained N-terminal affinity tag) and LEDGF_{IBD} (Figure 5.1b). Electrophoretic analysis of this complex revealed an apparent stoichiometry of 1:1, consistent with a 2:2 IN_{NTD-CCD}:LEDGF_{IBD} complex.

Although N-terminal modifications do not abrogate IN catalytic activity (M. Li et al. 2014) and allow for the isolation and structural determination of intasome structures (Passos et al. 2017, 2020), we sought to isolate IN·LEDGF_{IBD} complexes using IN with a native N-terminus. Ulp1 cleavage of the N-terminal affinity tag of IN 1F liberates an N-terminal phenylalanine, as found in vivo after proteolytic processing of the pol polyprotein (Eilers et al. 2020). To accomplish this, we co-expressed IN, LEDGF_{IBD}, and Ulp1, resulting in cleavage of the affinity tag from IN in vivo. Purification of the complex was achieved by affinity purification and subsequent cleavage of the LEDGF_{IBD} affinity tag. We successfully isolated a complex of IN and LEDGF_{IBD} with an apparent stoichiometry of 2:1 IN:LEDGF_{IBD} (Figure 5.2a). The addition of the detergent 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS) was essential for the solubility of this complex. The same approach was used to isolate a complex of IN_{NTD-CCD} and LEDGF_{IBD} (Figure 5.2b). This complex has an apparent stoichiometry of 2:1 IN:LEDGF_{IBD}, in contrast to the 1:1 stoichiometry of the uncleaved complex (Figure 5.1b).
5.3.2 Strand transfer activity of Integrase-LEDGF Complexes

IN constructs containing N-terminal modifications have been shown to be hyperactive in vitro (M. Li et al. 2014). Therefore, we tested the strand transfer activity of an HFS-IN·LEDGF\textsubscript{IBD} complex without cleavage of the N-terminal affinity tag (Figure 5.3a). This complex showed both single-strand and concerted integration activity (Figure 5.3b). Single-strand and concerted integration activity were retained after cleavage to liberate the native N-terminus (Figure 5.3c).
5.3.3 Intasome preparation

Intasomes of multiple retroviruses have been prepared in vitro (Hare, Gupta, et al. 2010; Yin et al. 2016; Ballandras-Colas et al. 2016, 2017; Passos et al. 2017, 2020; Pandey et al. 2014; Cook et al. 2020) through a combination of decreasing the ionic strength of the buffer and kinetic trapping by blocking catalysis. Catalysis can be blocked by strand transfer inhibitors (STIs), divalent ions that support assembly but not catalysis (Ca$^{2+}$), substitutions in essential active site residues, and artificially stabilized reaction intermediates. We have utilized each of these techniques in our attempts to prepare native HIV-1 intasomes. The ultimate goal of intasome preparation is biophysical characterization and structure determination, which we have not yet accomplished. Instead, summarized here, is our progress in developing small-scale readouts of intasome preparation, essential for the determination of optimal preparation conditions.

Figure 5.3. IN-LEDGF$_{BID}$ activity. a) Diagram of strand transfer activity assay. Integration of a fluorescently labeled (green star) double-stranded DNA oligonucleotide mimic of the viral LTR results in the formation of tagged circle products (resulting from single-strand integration) and 2 LTR coupled products (resulting from concerted integration). b) Strand transfer activity of HFS-IN-LEDGF$_{BID}$ complex where N-terminal affinity tag remains uncleaved. Integration products were separated on 1.5% agarose gels and imaged for fluorescence (left) and then stained with ethidium bromide (right). Significant single-strand and concerted activities are observed. c) Strand transfer activity of IN-LEDGF$_{BID}$ complex after cleavage, liberating a native N-terminus. Both single-strand and concerted activities are observed.
that can be scaled-up for biophysical analysis, X-ray crystallography, or cryo-electron microscopy.

As intasome preparation schemes require a decrease in the ionic strength of the buffer, we first determined the limits of solubility of IN 1F and IN 1F·LEDGF\textsubscript{IBD} complexes in the absence of DNA. IN 1F alone was highly susceptible to decreasing ionic strength, beginning to precipitate spontaneously at NaCl concentrations below 500 mM (Figure 5.4a). The addition of LEDGF\textsubscript{IBD} significantly improved solubility. IN·LEDGF\textsubscript{IBD} complexes began to precipitate below 225 mM (Figure 5.4b). These concentrations were used to guide subsequent intasome formation experiments. When DNA oligonucleotides mimicking the viral LTR ends (Supplemental Figure 5.1a) were added to IN·LEDGF\textsubscript{IBD} complexes, turbidity decreased, suggesting that the formation of IN-DNA complexes stabilizes the protein in solution (Figure 5.4c).

![Figure 5.4. Solubility limits of IN and IN-LEDGF complexes.](image)

**Figure 5.4. Solubility limits of IN and IN-LEDGF complexes.** The solubility of IN constructs was measured by the detection of turbidity (measured by light scattering at 405 nm) in decreasing ionic strength buffers. a) IN 1F begins to form turbid aggregates below 500 mM NaCl. b) IN-LEDGF\textsubscript{IBD} complexes are more stable in solution, showing aggregate formation below 225 mM NaCl. c) The addition of oligonucleotides mimicking the viral LTRs reduces aggregate formation, suggesting that IN-DNA complexes are more stable in solution. This effect is shown at two NaCl concentrations.

Based on these results, we performed small-scale intasome preparations by dialysis to low ionic strength or direct addition of low ionic strength buffer. Intasomes were trapped and analyzed with a variety of readouts. The goal is the development of a small-scale, high-throughput assay to determine optimal conditions. Biophysical and
structural analysis require large quantities of highly pure complexes, and adjustment of multiple parameters is required to find the ideal conditions for intasome preparation.

5.3.4 Intasome readout: crosslinking assays

Covalent crosslinking has been used to visualize the formation of intasomes (Passos et al. 2017; Yin et al. 2016) and crosslinked intasomes may even retain activity (Faure et al. 2005). We applied the crosslinking technique to intasome preparations by dialysis to low ionic strength. The DNA template was a mimic of the strand transfer reaction product, termed the three-way junction (Supplemental Figure 5.1b). This template mimics the covalently-linked viral DNA and host DNA and is formed by annealing five separate oligonucleotides (Ren et al. 2007). After dialysis, covalent crosslinkers were added and resulting products were separated by gel electrophoresis. Crosslinking with two concentrations of two crosslinkers, disuccinimidyl glutarate (DSG) and disuccinimidyl suberate (DSS), revealed the presence of higher MW species in the absence of 3WJ DNA, corresponding to multimers of IN·LEDGF<sub>IBD</sub> complexes (Figure 5.5). The addition of 3WJ DNA did not result in the formation of distinct higher MW species, the expected result if intasomes were successfully forming. This result highlights a major difficulty in intasome readouts: distinguishing between high molecular weight protein-protein or protein-DNA aggregates and intasomes. We therefore investigated alternative readouts for intasome formation.

5.3.5 Intasome readout: agarose electrophoretic mobility shift assays

Gel electrophoresis is a commonly used technique for the detection of protein-nucleic acid interactions (Garner and Revzin 1981; Fried and Crothers 1981),
and polyacrylamide gels are often preferred over agarose gels for improved resolution (Hellman and Fried 2007). However, agarose gels are more rapid and less toxic than polyacrylamide gels and are therefore useful for initial experiments (Chandrasekhar, Souba, and Abcouwer 1998; Brody et al. 2004; Ream, Lewis, and Lewis 2016). Agarose electrophoretic mobility shift assays (EMSA) have been used to characterize the products of intasome preparations (M. Li et al. 2020). We utilized fluorescently-labeled oligonucleotides mimicking the viral LTRs (Supplemental Figure 5.1c) to assemble intasomes, which were subjected to agarose gel electrophoresis and detected by fluorescence imaging. No target DNA was added to enable trapping of intasomes after assembly. In the absence of IN or the absence of LTR DNA, no high molecular weight species were observed (Figure 5.6a, lanes 1 and 2). Co-incubation of LTR DNA with
increasing concentrations of IN resulted in the formation of high molecular weight species that did not migrate out of the loading well (Figure 5.6a, lanes 3-8). Addition of LEDGF$_{IBD}$ facilitated the migration of IN-DNA complexes out of the wells in a
dose-dependent manner (Figure 5.6b, lanes 3-5). The complexes formed a smear in the gel, possibly due to heterogeneity or poor resolution in the agarose matrix. LEDGF has been shown to be important for intasome formation (Ballandras-Colas et al. 2017; Passos et al. 2017), and enhanced migration of IN-DNA complexes in the presence of LEDGF<sub>IBD</sub> suggests that it modulates multimerization. When Ca<sup>2+</sup> (which allows for intasome assembly but not catalysis) (Figure 5.6b, lanes 6,7) or the STI raltegravir (Figure 5.6b, lanes 8,9) were added to IN-DNA·LEDGF<sub>IBD</sub> complexes, migration was enhanced further. A discrete but poorly resolved band near the 10 kb marker was observed, likely representing well-formed intasomes, alongside a highly diffuse smear of smaller products, likely representing partial intasomes or non-intasome IN·DNA·LEDGF<sub>IBD</sub> complexes (Figure 5.6b, lanes 6-9). Kinetic trapping with Ca<sup>2+</sup> (Ellison and Brown 1994; Ballandras-Colas et al. 2016) and STIs (Pandey et al. 2014; Hare, Gupta, et al. 2010; Passos et al. 2020; Cook et al. 2020) has been shown to stabilize intasomes, and their effect on migration observed here further supports this conclusion. Distinguishing and separating nonspecific IN-DNA complexes from well-formed intasomes is an essential step in structure determination, so we attempted to characterize the discrete high molecular weight species in more detail.

5.3.6 Intasome readout: polyacrylamide electrophoretic mobility shift assay

To better resolve the high molecular weight species observed in agarose EMSA, we utilized polyacrylamide EMSA (native PAGE). IN·LEDGF<sub>IBD</sub> complexes were dialyzed to low ionic strength buffer in the presence of 3WJ DNA, then separated on PAGE without boiling or the addition of Sodium dodecyl sulfate (SDS). A heterogeneous smear of 3WJ DNA was observed in the absence of IN·LEDGF<sub>IBD</sub>, and no bands were visible in
the absence of 3WJ DNA when stained with either a DNA or protein-reactive dye (Figure 5.7a,b). Co-incubation of IN and 3WJ DNA resulted in the appearance of a discrete high molecular weight band, consistent with intasome formation (Figure 5.7a,b). SDS-PAGE analysis of the soluble portion of this intasome preparation revealed the presence of bands corresponding to IN and LEDGF<sub>IBD</sub>, as expected (Figure 5.7c), showing that both proteins were involved in intasome formation.

![Image](image.png)

**Figure 5.7. Polyacrylamide electrophoretic mobility shift assays of intasome preparations.** Intasomes were prepared by addition of IN-LEDGF<sub>IBD</sub> to 3WJ DNA and dialysis to low ionic strength buffer. Soluble portions of dialysis samples were separated by 6% native PAGE and stained to detect nucleic acid (a) or protein (b). Arrow indicates high molecular weight band only observed when both IN-LEDGF<sub>IBD</sub> and 3WJ DNA are present. This band is not present in control samples that contain only 3WJ DNA or IN-LEDGF<sub>IBD</sub>. c) Separation of IN-LEDGF<sub>IBD</sub> co-purification and post-dialysis sample from intasome preparation. Both IN and LEDGF<sub>IBD</sub> are present in the soluble fraction that contains the protein-DNA complex observed in panels a) and b).

### 5.3.7 Intasome readout: biophysical techniques

Retroviral intasomes adopt a wide range of oligomeric forms, from tetramers to hexadecamers (A. N. Engelman and Cherepanov 2017), with different viruses showing characteristic forms. The oligomeric state of HIV-1 intasomes remains unclear (M. Li et
al. 2014; Passos et al. 2017, 2020), despite significant effort to determine the oligomeric state and low-resolution reconstruction of IN subunit arrangements of the native HIV-1 intasome. Solution biophysics approaches such as SEC-MALS or analytical ultracentrifugation (AUC) can be used to determine the oligomeric state of prepared intasomes, and subsequent analysis by small-angle X-ray scattering (SAXS) can be used to generate low-resolution structural models. Unfortunately, our attempts to analyze intasomes by these approaches have been unsuccessful due to multiple issues. Despite our efforts to develop high-throughput, small-scale readouts to identify optimal preparation conditions, native HIV-1 intasomes are still formed with low efficiency. Additionally, preparations are heterogeneous, with free IN, LEDGF$_{IBD}$, and DNA interfering with analysis of intasomes. Distinguishing intasomes from nonspecific IN·DNA aggregates, especially in a heterogeneous background, is difficult. Small scale, low concentration intasome preparations are prone to dissociation during analysis. Stabilization with STIs during dialysis to low ionic strength is infeasible due to the amount of compound required. Additional work is needed to identify optimal conditions for intasome formation and stabilization, which can then be scaled-up for biophysical characterization and subsequent structure determination.

5.4 Discussion

We have recently reported the purification and characterization of wild type, native HIV-1 IN (IN 1F) (Eilers et al. 2020). This construct is catalytically active and has favorable solution properties, making it the ideal candidate for complex formation with host cell proteins, intasome preparation, and structure determination. Here, we report progress in these areas. IN interacts with multiple host cell proteins (Emig-Agius et al.
2014), of which LEDGF is the most extensively characterized (Peter Cherepanov et al. 2003; Turlure et al. 2004; Emiliani et al. 2005). LEDGF binds to and stabilizes IN (McKee et al. 2008; Hare, Di Nunzio, et al. 2009), an interaction that has been shown to be important for intasome formation (Ballandras-Colas et al. 2017; Passos et al. 2017). We purified complexes of IN and LEDGF using both full-length IN and IN_{NTD-CCD} truncations, which include the minimal domains involved in LEDGF binding (Peter Cherepanov et al. 2005; Hare, Shun, et al. 2009). These complexes are highly pure and monodisperse, and full-length IN·LEDGF complexes are fully competent for concerted integration in vitro. They can be prepared with or without the presence of an N-terminal affinity tag, useful for additional purification steps either before or after DNA complex formation. The N-terminal α helix of the IN NTD is opposite the surface that contacts LEDGF, so it is not surprising that the presence of an affinity tag at the N-terminus does not disrupt complex formation. Further characterization will be required to determine whether an N-terminal affinity tag decreases complex stability. In future experiments, a technique for accurately quantifying IN when present in IN·LEDGF_{IBD} complexes will need to be adopted to compare activity levels to isolated IN conditions. However, as we have already shown, IN 1F is hyperactive when compared to IN with N-terminal artifacts (Eilers et al. 2020). Therefore, we expect that IN·LEDGF_{IBD} complexes are similarly hyperactive.

Successful intasome preparations involve a combination of decreasing buffer ionic strength (enabling protein-nucleic acid interactions) and kinetic trapping by blocking catalysis (preventing turnover and disassembly of catalytic complexes) (Hare, Gupta, et al. 2010; Yin et al. 2016; Ballandras-Colas et al. 2016, 2017; Passos et al. 2017, 2020; Pandey et al. 2014). IN·LEDGF_{IBD} complexes were stable at near-physiologic ionic strength, and showed increased solubility in the presence of DNA mimics of the viral
LTRs. These results suggested the formation of IN-DNA complexes, so we characterized these complexes through a variety of approaches. Covalent crosslinking revealed the presence of high molecular weight complexes of IN and LEDGF in the presence or absence of DNA, with no distinct species observed in the presence of DNA. We are therefore unable to demonstrate successful formation of intasome complexes by this technique, but additional investigation with linkers of diverse lengths and chemical reactivities should be attempted. Electrophoretic mobility shift assays are useful for detection of protein-DNA interactions (Garner and Revzin 1981; Fried and Crothers 1981), and we utilized two varieties of this approach to analyze intasome preparations: rapid, low resolution separation by agarose gel electrophoresis and high resolution separation by polyacrylamide gel electrophoresis. Both techniques revealed the presence of high molecular weight species consisting of both protein and DNA, suggestive of intasomes. Unfortunately, detailed characterization of the structure or oligomeric state of these complexes is not possible by this technique, and our attempts at further characterization by biophysical approaches (SEC-MALS and AUC) have been inconclusive. Optimization and scaling-up of preparation methods will enable future biophysical characterization.

Ultimately, the goal of this work is the determination of the structure and solution properties of the native HIV-1 intasome. Biophysical characterization and low-resolution structure determination, such as that obtained by small-angle X-ray scattering, will be a useful starting point, but high-resolution structure determination by X-ray crystallography or cryo-electron microscopy is desired. The preliminary work presented here sets the foundation for these efforts in the future.
5.5 Acknowledgments

We thank members of the Van Duyne and Bushman laboratories for advice and technical assistance. This work was supported by the Penn Center for AIDS Research, the PennCHOP Microbiome Program, the Johnson Research Foundation, and the National Institutes of Health (R61-HL137063, R01-HL113252, U19-AI117950, UM1-AI126620, and T32-AI007632 to F.D.B., R01-AI129661 to F.D.B. and G.V.D., P30-AI045008 to F.D.B. and K.G., and R01-NS100081 to K.G.). G.E. was supported by the NIH Training in HIV Pathogenesis grant (T32-AI007632).
Chapter 6: Future Directions and Concluding Remarks

In the studies presented here, we sought to advance the structural, biochemical, and biophysical understanding of retroviral integrase. HIV remains a global burden on human health despite the development of highly effective antiretroviral therapy, and integrase is one of the most important therapeutic targets. Our efforts were focused on understanding the mechanism of action of, and mechanisms of resistance to, the two classes of integrase inhibitors, the strand transfer inhibitors (STIs) and the allosteric inhibitors of integrase (ALLINIs). Our work in Chapter 2 provides a detailed view of the ALLINI binding interface, essential for the rational design and improvement of ALLINIs. In future experiments, ALLINI resistance substitutions should be introduced into this crystal scaffold to obtain higher-resolution structures. Previously, the affinity of ALLINIs has been determined by displacement of LEDGF (Hou et al. 2008), or indirectly by measuring aggregation. Separating the CCD and CTD faces of the ALLINI binding interface allows for dissection of the binding affinity contributions of both domains. Additionally, the individual domains have much-improved solution characteristics, enabling the study of binding at physiological conditions. Future work should focus on further improvements to the resolution of IN·ALLINI structures and a better understanding of the affinity and kinetics of inhibitor binding.

In Chapter 3, we studied a highly potent ALLINI, ViiV-919, and suggest a therapeutic niche for the ALLINI compounds in combination with STIs. Compounds in this class have not progressed beyond phase I clinical trials (Fenwick et al. 2014; Fader et al. 2014), suggesting insurmountable toxicity. Current generation ALLINIs show a
more favorable toxicity profile (Peese et al. 2019; G. Li et al. 2020; Sivaprakasam et al. 2020). To motivate ongoing interest in the ALLINIs, future work should identify opportunities for combinations with clinically successful antiretrovirals. We have shown that STI-resistant integrase is hypersensitive to ViiV-919, suggesting that the strategic addition of ALLINIs to a therapeutic regimen may prevent the emergence of resistance. These studies should be expanded to additional resistance mutations and antiretroviral classes, and the development of structural models to explain the hypersensitivity of STI-resistant integrase will allow for rational combinations in the future.

Chapter 4 reports the characterization of wild type, native HIV-1 integrase. This construct recapitulates the authentic form of integrase found in vivo, and is therefore an invaluable tool for the study of integrase structure, biochemistry, and biophysics in vitro. Of particular interest is the structural determination of the HIV-1 intasome, and we present progress toward this goal in Chapter 5. Methods for intasome preparation, purification, and structure determination have been detailed in several recent reports, and are readily applicable to IN 1F (M. Li et al. 2014; Passos et al. 2017; Ballandras-Colas et al. 2017; Passos et al. 2020; Cook et al. 2020). Intasome structures determined with IN 1F will have significant immediate impact.

The work in this thesis provides the foundation for future studies aimed at understanding HIV integrase and improving inhibitors that target this enzyme. It also highlights the multifaceted role of a single viral enzyme in orchestrating a complex replication cycle. In the following sections, I will describe areas of ongoing and future study, with the goal of further understanding HIV integrase structure, biochemistry, and biophysics.
6.1 Ternary complex studies - dimerization-deficient, affinity/kinetics, better resolution crystals

Separate purification of the two domains involved in ALLINI binding, the CCD and CTD, improved the resolution of structural models of the IN-ALLINI binding interface (Chapter 2). This approach was successful, in part, due to a reduction in crystal disorder compared to the full-length IN-ALLINI structure (Gupta et al. 2016). Co-crystallization of IN and ALLINIs remains difficult, as it competes with unproductive aggregation and precipitation of IN. We believe that precipitation that occurs in the presence of IN$_{CCD}$, IN$_{CTD}$, and ALLINI is mediated by CTD-CTD dimerization (Supplemental Figure 2.4), and have shown a decrease in precipitation when this dimerization interface is mutated. Our group recently purified a series of dimerization-deficient CTDs (Gupta et al. 2020, submitted) and showed that they form a ternary complex with CCD and ALLINI that remains soluble. These ternary complexes are candidates for future crystallization trials.

Additionally, ternary complex formation is unaffected by the presence of an affinity tag. The affinity tag serves as a “handle” for surface immobilization or fluorescent labeling, requirements for analysis with interaction techniques such as surface plasmon resonance (SPR), biolayer interferometry (BLI), or microscale thermophoresis (MST). The small scale, rapid, and highly sensitive nature of MST is ideally suited to measure binding affinities of IN$_{CCD}$, IN$_{CTD}$, and ALLINIs. As a proof-of-principle, we have utilized a fluorophore compatible with MST that binds non-covalently to a polyhistidine affinity tag (Bartoschik et al. 2018; Sparks and Fratti 2019). Labeling is rapid and does not require subsequent desalting or cleanup. The high affinity and slow dissociation of such labels allow for measurements at the low concentrations necessary to determine the affinity constants of high-affinity ALLINIs.
Examination of the IN·ALLINI polymer crystal structure suggested that CTD-CTD dimers could form in solution, leading to unproductive polymer branching and the formation of insoluble aggregates (Figure 6.1) (Gupta et al. 2020, submitted). We have also incorporated dimerization-deficient CTD substitutions into full-length IN constructs, both in the background of IN constructs that crystallized previously in the presence of ALLINIs (Gupta et al. 2016), and in the background of IN 1F (Eilers et al. 2020). Our aim is to improve the resolution of IN·ALLINI crystal structures by eliminating the formation of unproductive CTD-CTD interactions in solution, thereby favoring the formation of productive crystal contacts. Efforts to purify and crystallize these constructs are ongoing. A crystal structure with all three domains of IN has yet to be determined, and would clarify intra- and inter-molecular domain interactions and provide a context for understanding the complex solution behavior of IN. Future efforts should be directed toward this goal.

Figure 6.1. Role of the IN_{CTD} in ALLINI-induced aggregate formation in vivo (from Gupta et al. 2020, submitted). A. Model of the IN-ALLINI polymer as observed in the crystal lattice. No CTD-CTD dimerization interfaces are observed. In solution, however, CTD-CTD interactions are possible, and would cause branching of the linear IN-ALLINI polymer. B. Overlay of CTD-CTD dimer crystal structure with IN-ALLINI crystal structure, showing that CTD-CTD interactions could occur without disrupting the CCD-CTD ALLINI binding interface. The blue structure (PDB: 1HV) is the NMR solution structure of the IN_{CTD} dimer.
6.2 Complementarity of STIs and ALLINIs - structural studies of ALLINI resistance in Intasome

Current antiretroviral development is an especially daunting task. With the availability of multiple regimens that effectively and durably control HIV replication, new antiretrovirals must not only show pharmacodynamic potency, but minimal toxicity and highly favorable pharmacokinetics with long-term bioavailability (Cihlar and Fordyce 2016). Combination of compounds with additive or synergistic effects is a desirable approach for antiretroviral therapy (Klein et al. 2003; Snyder et al. 2000). ALLINIs and STIs have been shown to be additive or synergistic against viral replication in vitro (Fenwick et al. 2014; Christ et al. 2012) as they target multiple steps in the viral replication cycle, highlighting their potential for combination antiretroviral therapy. Additionally, ALLINIs retain activity against STI-resistant IN (Christ et al. 2010; Le Rouzic et al. 2013), and as we show in Chapter 3, may actually show increased potency in this context. Our data in Chapter 3 should be extended to additional STI-ALLINI combinations, STI resistance substitutions, and full-cycle replication assays.

We do not yet fully understand the structural basis for ALLINI hypersensitivity in STI-resistant IN. Recent structures of the intasome offer structural explanations of complex pathways of STI resistance (Cook et al. 2020; Passos et al. 2020), and these multimeric assemblies highlight the multitude of intra- and inter-subunit interactions adopted by IN (Ballandras-Colas et al. 2017; Passos et al. 2017). Reciprocal mapping of STI resistance substitutions in the context of the IN·ALLINI polymer and ALLINI resistance substitutions in the context of the intasome will identify potential disruptions and opportunities for drug design that exploits hypersensitivity. Recent data has begun to probe the complex interactions in higher-order intasomes (Chivukula et al. 2020),
highlighting the importance of IN substitutions in intasome formation. Many of these substitutions, such as those that disrupt the CCD-CTD linker, would also be expected to affect ALLINI binding. These data highlight the pleiotropic nature of IN substitutions, and begin to explain how resistance to one class of inhibitors could convey hypersensitivity to another class. ALLINI and STI resistance substitutions often arise with a tradeoff in viral fitness, and extensive effort may be necessary to understand the consequences (Hoyte et al. 2017; Madison et al. 2017). Further investigation in this area will reveal strategies for improved inhibitor design and rational inhibitor combinations.

6.3 Intasome pursuits

The time between the publication of the first intasome structure from the prototype foamy virus (Hare, Gupta, et al. 2010; G. N. Maertens, Hare, and Cherepanov 2010) to the publication of the first HIV-1 intasome structure (Passos et al. 2017) underscores the difficult of studying HIV-1 integrase. Protein engineering and cryo-EM overcame the difficulties of working with HIV-1 in vitro, enabling structural determination (M. Li et al. 2014). Additional structures of HIV-1 intasomes and structurally similar Maedi-Visna virus and simian immunodeficiency virus have been published (Ballandras-Colas et al. 2017; Cook et al. 2020; Passos et al. 2020), but as of the writing of this thesis, a structure of the native, unengineered HIV-1 intasome has not been determined. The work in Chapters 4 and 5 describe the purification and characterization of native HIV-1 integrase, which shows desirable properties in vitro and has yielded promising results in intasome preparations. This construct should be used in future work to determine the structure of the native HIV-1 intasome.
Two major hurdles must be overcome to determine the structure of the native HIV-1 intasome. The first hurdle is the efficient, reproducible preparation of stable intasomes. Protocols for intasome preparations for related retroviruses and engineered HIV-1 integrase constructs are widely available in the literature and are briefly reviewed in section 5.3.3. Approaches that combine buffer exchange to low ionic strength with kinetic trapping by Ca$^{2+}$ or strand transfer inhibitors are likely to be successful, but, experience has shown that HIV-1 integrase exhibits significantly different behavior in vitro as compared to the integrases of other retroviruses. Modification of existing protocols with respect to ionic strength, pH, concentration of protein and DNA, and additives (e.g. detergents, Zn$^{2+}$, etc.) will likely be needed. The number of potential combinations is vast, therefore requiring a readout of successful intasome preparation that is rapid and compatible with small-scale reactions. A simple turbidity assay is presented in section 5.3.3, but this is not specific to intasomes as it will also detect non-specific IN-DNA interactions that affect solubility. Initial experiments with covalent crosslinking were not successful, but this approach should be pursued further as it is cost-effective and compatible with small-scale reactions. Agarose and polyacrylamide gel electrophoretic mobility shift assays (EMSA) are most promising as an intasome readout as they utilize a small amount of material, can compare multiple conditions simultaneously, are cost-effective, and can identify discrete, monodisperse species. EMSA should be used as an initial screening readout, and promising conditions for intasome preparation should be scaled-up, analyzed by biophysical approaches as detailed in section 5.3.7, and subjected to structural determination.

A second hurdle exists before structure determination is possible, and this hurdle will depend on the technique selected. X-ray crystallography and cryo-electron
microscopy (cryo-EM) have both been used to determine high-resolution intasome structures. X-ray crystallography requires the growth of crystals, which requires extensive screening of conditions. Protein-DNA complexes additionally require experimentation with DNA length and configuration, rapidly multiplying the number of experiments required. Here we use both short double-stranded oligonucleotides mimicking the viral LTR ends (Katzman et al. 1989; R. Craigie, Fujiwara, and Bushman 1990) and a more complex mimic of the three-way junction intermediate (Ren et al. 2007). The minimum DNA length for enzymatic activity is well-understood, and long DNA constructs support activity (P. Cherepanov et al. 1999), so the optimal length for crystallization will have to be determined empirically. Cryo-EM bypasses the hurdle of crystallization, but may require additional optimization steps such as the addition of detergents or the prevention of polymerization (Cook et al. 2020). As shown in section 5.3.2, integrase that retains an N-terminal tag has concerted integration activity and therefore must be competent to form intasomes. The use of such a tag may be essential to achieve the purity needed for cryo-EM structural analysis. Through X-ray crystallography or cryo-EM, successful determination of the structure of the native HIV intasome will be a high impact contribution to our understanding of the biochemistry of integrase and mechanisms for its inhibition.

6.4 The uncertain future of ALLINIs

Ten years have passed since the report of the first allosteric integrase inhibitor (Christ et al. 2010) with no compounds in this class successfully completing clinical trials. Some compounds have failed clinical development due to existing resistance or low barriers to resistance (Trivedi et al. 2020). Since the withdrawal of BI-224436 from
clinical trials, ALLINI development efforts have been directed toward increasing potency and bypassing viral resistance (Sugiyama et al. 2020; G. Li et al. 2020; Peese et al. 2019). Results from clinical trials of these compounds have not been reported, giving rise to a pessimistic outlook on the future of ALLINIs.

Recent ALLINIs have favorable stability profiles in vitro (Patel et al. 2020; Peese et al. 2019) and in model organisms (Sugiyama et al. 2020; G. Li et al. 2020). However, discordant results between in vitro stability and in vivo pharmacokinetics (G. Li et al. 2020; Peese et al. 2019), as well as significant loss of activity in the presence of human plasma proteins (Patel et al. 2020; G. Li et al. 2020) highlight the challenges faced in ALLINI development. In addition, ALLINIs are thought to activate the cytochrome P450 pathway, potentially causing detrimental drug-drug interactions. In one case (Sivaprakasam et al. 2020), structural modification successfully mitigated the effects on cytochrome P450, but, as of the publication of this thesis, no compounds from that publication are currently under investigation in clinical trials. Without a clear understanding of the pharmacokinetics of ALLINIs in vitro and in model organisms, the future of this class of compounds is uncertain.

With very few exceptions, ALLINIs consist of a tert-butoxy and carboxylic acid group attached to an aromatic core that is decorated by a wide variety of chemical moieties (Supplemental Figure 2.1). Aromatic cores that have been reported are pyridine, thiophene, quinoline, isoquinoline, thienopyridine, pyrazolopyrimidine, hydro-naphthyridine, and benzene (A. N. Engelman 2019; G. Li et al. 2020; Sugiyama et al. 2020) (Figure 6.2a). The moieties that decorate this core range from relatively simple to extremely complex macrocycles, and this chemical space has been thoroughly explored (A. N. Engelman 2019) (Figure 6.2b). Early ALLINIs included substituents other
than tert-butoxy and carboxylic acid groups, but the tert-butoxy and carboxylic groups have become completely invariant in subsequent ALLINI development. It is possible that the combination of a tert-butoxy and carboxylic acid motif linked to an aromatic core causes dose-limiting toxicity in humans, and since this portion of the molecule is shared between otherwise diverse chemotypes of the ALLINI class, diversity must be explored in this region if the pharmacokinetic properties of ALLINIs are to be optimized. Abandonment of the ALLINIs without fully exploring chemical diversity would be tragic.

![Diagram of chemical structures](image)

**Figure 6.2. ALLINI chemotypes and structural diversity.** a) The core aromatic structures of published ALLINIs. Connected to these pharmacophores are the conserved tert-butoxy and carboxylic acid moieties as well as a diverse array of chemical substituents. b) An example of an early ALLINI (CX05168, from Christ et al. 2010) compared to an example of a current-generation ALLINI (Compound 4b from Sivaprakasam et al. 2020). Note the propyl group in CX05168 in place of the tert-butoxy group in Compound 4b.

### 6.5 Concluding Remarks

The work summarized in this thesis represents the focused study of the structural, biochemical, and biophysical properties of a single viral enzyme. This enzyme, the retroviral integrase, has proven to be of central importance as a
pharmaceutical target in the fight against the global HIV pandemic. For this reason alone, research should continue in earnest beyond the scope of this thesis. Furthermore, much of this thesis was prepared during the impact of the global coronavirus pandemic, underscoring the importance of studying the structural, biochemical, and biophysical properties of viral enzymes. The clinical success of antiretroviral therapeutics targeting HIV IN should encourage further focused research on viral enzymes. From another perspective, elucidation of the functions of viral enzymes has provided vast benefits. Through the study of its three enzymes, protease, reverse transcriptase, and integrase, HIV has enabled breakthroughs in biochemistry, genetics, gene therapy, clinical diagnostics, synthetic biology, and countless other aspects of medicine and research. In Chapter 2, we report advances in our understanding of the ALLINI class of compounds. The unique mechanism of action of these compounds, which bind within an interface between molecules of IN, represents a new avenue of drug development (M. Gao and Skolnick 2012; Glasgow et al. 2019). A molecule which acts in a similar manner, the anti-neoplastic drug paclitaxel (Horwitz 1994), has been successful clinically. A class of inhibitors that bind to the HIV capsid (CA) (Singh et al. 2019; Carnes, Sheehan, and Aiken 2018; Link et al. 2020) bind in a similar manner between two molecules of CA. Due to their potency, these compounds have the potential to change the paradigm of HIV treatment, further underscoring the importance of research in this area. Chapter 3 reports our research on a highly optimized member of the ALLINI class and shows the promise of combinatorial therapy. Further investigation of combinatorial approaches to therapy will maximize therapeutic effect while minimizing toxicity. Our work detailing the structural mechanism of action of the ALLINIs and mechanisms of resistance will advance drug development in this paradigm in the future. Chapter 4 is a major step
toward accurate in vitro recapitulation of HIV IN. IN 1F will undoubtedly serve as an important resource for research on HIV in the future. Already, our work in Chapter 5 demonstrates the improvements enabled with the use of IN 1F in the study of the catalytically active complex of IN with DNA.

Our work could be critiqued for an overly narrow focus on a single enzyme. However, a dogged effort was required to develop the first class of antiretrovirals targeting IN, as it was until recently considered to be “undruggable” (Cheng et al. 2007). Today, antiretrovirals targeting IN are included in recommended first-line regimens and are some of the most potent compounds with the highest barrier to resistance. Many years of work were necessary to attain this success, which therefore argues for a continued narrow focus. Additionally, study of the ALLINIs revealed an essential role of IN late in the viral replication cycle, and have revealed the pleiotropic functions of other HIV proteins. Recently, significant progress in structural biology has been achieved through the use of cryo-electron microscopy. Our future studies will have to extend beyond X-ray crystallography and adopt this technique to determine the structure of large IN-protein and IN-DNA complexes. To conclude, this thesis adds to a body of research and drug development efforts undertaken over the past 30+ years, and hopefully contributes to the reduction in morbidity and mortality caused by HIV. Falling victim to its own success, HIV research efforts are shifting elsewhere, and eventually work will be redirected to different infectious diseases with a greater impact on human health. However, as this thesis shows, much remains to be discovered through the study of HIV, and the discoveries so accomplished will be widely applicable to other viruses in the future.
IN constructs were expressed and purified as described previously (Ciuffi et al. 2006; Diamond and Bushman 2006; Gupta et al. 2010, 2016) (Gupta 2020, submitted). Briefly, full-length IN and C-terminal domain constructs used for crystallization were expressed from pET-Duet vectors (Novagen) with C-terminal Mxe intein, chitin binding domain, and polyhistidine tag. Gravity purification was performed through nickel-nitriloacetic acid (Qiagen) and chitin (New England Biolabs) resins, followed by intein cleavage activated by 100 mM β-mercaptoethanol. Liberated proteins were purified through size-exclusion chromatography using a Superdex 75 HiLoad 16/60 column (GE Healthcare) in 20 mM HEPES-NaOH pH 7.5, 1M NaCl, 7 mM CHAPS, 10 µM ZnOAc$_2$, and 10 mM dithiothreitol (DTT).

Individual IN domains (CCD constructs for both crystallization and aggregation assays and CTD constructs for aggregation assays) were inserted into an expression vector containing an N-terminal His7-Flag-Sumo tag (Eilers et al. 2020) (Gupta 2020, submitted). Proteins were purified with nickel-nitriloacetic acid resin, and either subjected to size-exclusion chromatography (for aggregation assays) or a second nickel-nitriloacetic column after fusion proteins were liberated by cleavage with Ulp1. CCD constructs were purified through size-exclusion chromatography using a Superdex 75 HiLoad 16/60 column (GE Healthcare) in 20 mM HEPES-NaOH pH 7.5, 300 mM NaCl, and 2 mM DTT.
Purified proteins were concentrated in an Amicon Ultra-15 (Millipore), flash frozen with 10% glycerol, and stored at -80°C.

Site-directed mutagenesis (QuikChange, Stratagene) was used to generate point mutations (W131C, N222K, L242A) used in this study.

**Crystallization and structure determination**

Full-length IN constructs containing resistance substitutions were crystallized as previously published (Gupta et al. 2016). The ALLINI BI-D (Fader et al. 2014) was purchased from MedChemExpress (Monmouth Junction, NJ, U.S.A.), resuspended in 1,3-dimethyl-2-imidazolidinone (DMI, Fluka), and was present at a final concentration of 1 mM for crystallization. IN was mixed 1:1 with a solution containing 30% 2-methyl-2,4 pentanediol (MPD), and 0.1 M sodium citrate pH 5.6 in standard hanging drop format at 21°C. Crystals appeared after 24-48 hours and grew to maximal size within one week. Crystals were harvested and frozen directly from hanging drops. Diffraction data was collected at the 24-ID-E beamline of the Advanced Photon Source (Argonne National Laboratory).

Data reduction was performed using XDS (Kabsch 2010). Molecular replacement and refinement were carried out in Phenix (Liebschner et al. 2019; Adams et al. 2010) and CNS (Brünger 2007; Brünger et al. 1998). The structures were solved using 5HOT as a search model (Gupta et al. 2016). The asymmetric unit contained a dimer of IN and two ALLINI molecules. Manual building and refinement were carried out in COOT (Emsley et al. 2010). Later rounds of refinement utilized the DEN method (Brünger, Adams, et al. 2012; Brunger, Das, et al. 2012; G. F. Schröder, Levitt, and Brunger 2014) with high-resolution structures of IN\textsubscript{CCD} (Gupta et al. 2014) and IN\textsubscript{CTD} (Z. Chen et al. 2015).
2000) as reference models and the *Rosetta* method as implemented in *Phenix* (DiMaio et al. 2013). Figures were created with *PyMOL* (DeLano 2004).

\[ \text{IN}_{\text{CCD}}, \text{IN}_{\text{CTD}}, \text{and BI-224436 were crystallized by mixing 100 µM IN}_{\text{CCD}}, 100 µM \text{IN}_{\text{CTD}}, \text{and 500 µM BI-224436 in 20 mM Tris pH 7.4, 250-300 mM NaCl, and 1 mM DTT with 25% ethylene glycol. The ALLINI BI-224436 (Fenwick et al. 2014) was purchased from MedChemExpress (Monmouth Junction, NJ, U.S.A.) and resuspended in acetonitrile. Crystals also grew in 0.2 M magnesium formate, but diffracted poorly. Addition of 100 µM ammonium acetate improved crystal growth and yielded large single crystals > 200 µm in length within one week. Crystals were harvested and frozen directly from hanging drops. Diffraction data was collected at the 17-ID-1 (AMX) beamline at the National Synchrotron Light Source II (Brookhaven National Laboratory).}

Data reduction was performed using *DIALS* (Winter et al. 2018). Molecular replacement and refinement was carried out in *Phenix* (Liebschner et al. 2019; Adams et al. 2010) and *COOT* (Emsley et al. 2010). A partial solution was determined using the \( \text{IN}_{\text{CCD}} \) dimer (PDB: 3L3U) (Wielens et al. 2010) as a search model. The \( \text{IN}_{\text{CTD}} \) (PDB: 1IHV) (Lodi et al. 1995) could be placed with a subsequent search. A dimer of \( \text{IN}_{\text{CCD}} \), two molecules of the ALLINI, and two copies of \( \text{IN}_{\text{CTD}} \) occupy the asymmetric unit. Initial steps of refinement included \( \text{IN}_{\text{CCD}} \) and \( \text{IN}_{\text{CTD}} \) reference models and NCS restraints. Later steps utilized the *Rosetta* method as implemented in *Phenix* (DiMaio et al. 2013).

**Aggregation assay for ALLINIs**

Assays were performed as previously described (Gupta et al. 2014, 2016) in a plate reader (Victor 3V, Perkin Elmer), measuring the turbidity of the reaction solution at 405 nm. Final reaction conditions were 20 mM HEPES, pH 7.5, 15 µM IN, 175 mM NaCl,
1.2 mM CHAPS, and 30 μM ALLINI. Significance was evaluated by two-way ANOVA with P values reported from Tukey’s multiple comparisons test.

Supplemental Figures

**Supplemental Figure 2.1. Structure of ALLINIs used in this study.** The tert-butoxy and carboxylic acid moieties are conserved across all members of this class.

**Supplemental Figure 2.2** Representative images of crystals that yielded the structures reported in this study.
Supplemental Figure 2.3. Domains contributing to ALLINI binding. a) Structure of IN-ALLINI polymer demonstrating that the ALLINI binding interface is entirely composed of residues from the CCD (green) and CTD (blue). The ALLINI is shown red. b) IN_{CCD} and IN_{CTD} were purified as individual domains. SDS-PAGE analysis shows highly pure products at the expected molecular weight.

Supplemental Figure 2.4. Models of IN-ALLINI polymerization in crystallo. a) Open polymer observed in full-length IN-ALLINI crystals. Aligned with this structure is the structure of the IN_{CTD} dimer (cyan, highlighted by boxes) (PDB: 1iHV), showing that the CTD dimerization interface is opposite to the CTD face involved in ALLINI binding. b) Open polymer observed in IN_{CCD}, ALLINI, IN_{CTD} crystals. This polymer forms by IN_{CCD} and IN_{CTD} dimerization.
**Supplemental Figure 2.5. Detail of CTD face involved in ALLINI binding.** In the published structure of an IN-ALLINI polymer (PDB: 5HOT), two conformations of W235 and Y226 are observed. a) In one conformation, Y226 is oriented toward W235, which adopts a rotamer that avoids steric clash. b) In the other conformation, Y226 is oriented away from W235, which adopts a rotamer that projects into space occupied by Y226 in the alternate conformation. c) The conformation observed in the solution NMR structure of the isolated IN<sub>CTD</sub> is the same as in (b). The position of K266 in 5HOT is shown in (a), and the different rotamer observed in the IN<sub>CTD</sub>-BI-224436-IN<sub>CTD</sub> is shown in (b).
Supporting Information for Chapter 3

Protein expression and purification

IN constructs were expressed as previously described (Gupta et al. 2016, 2010; Ciuffi et al. 2006; Diamond and Bushman 2005, 2006; Eilers et al. 2020). Briefly, expression constructs were introduced into *E. coli* BL21(DE3), grown in 2xYT at 37°C to an optical density of 1.8-2.2, and induced by isopropyl-ß-D-1-thiogalactopyranoside (IPTG). Expression was allowed to continue for 5 hours at 20°C before pelleting and storage at -80°C.

Full-length IN 1F was purified as previously described (Eilers et al. 2020) by separation on a nickel-nitrilotriacetic acid resin column (Qiagen), cleavage with Ulp1 to liberate the His7-Flag-Sumo tag, separation on a second nickel-nitrilotriacetic acid resin column, and size-exclusion chromatography using a Superdex 75 HiLoad 16/60 column (GE Healthcare). IN was concentrated at 4°C in an Amicon Ultra-15 (Millipore), glycerol was added to a final concentration of 10% (w/v), and aliquots were flash-frozen in liquid nitrogen for storage at -80°C.

LEDGF$_{IBD}$ (Ciuffi et al. 2006) was purified as previously described (Eilers et al. 2020).

Small molecule inhibitors

BI-224436 was purchased from MedChemExpress, ViiV-919 was provided by ViiV Healthcare, and raltegravir was a gift from Merck. Inhibitors were resuspended in DMSO.

Aggregation assay for ALLINIs
Assays were performed as previously described (Gupta et al. 2014, 2016; Eilers et al. 2020). Reaction conditions included 20 mM HEPES-NaOH pH 7.5, 15 μM IN, 300 mM NaCl, 7 mM CHAPS, and 30 μM ALLINI. Turbidity was measured at one-minute intervals in a plate reader (Victor 3V, Perkin Elmer) as the absorbance of the reaction solution at 405 nm. Data plotted are the maximum absorbance values reached during a 15 minute assay. Significance was evaluated by one-way ANOVA with P values reported from Tukey’s multiple comparisons test. Data analysis was carried out in *Prism* (GraphPad).

**Integrase 3’-processing assay**

Assays were performed as previously described (Merkel et al. 2009; Guiot et al. 2006; Eilers et al. 2020). Final assay conditions were 400 nM IN with 20 mM HEPES-NaOH pH 7.5, 100 nM Alexafluor 488-labeled LTR substrate, 50 mM NaCl, 10 mM MgCl$_2$, 10 μM Zn(OAc)$_2$, 10 mM DTT, and 5 μM inhibitor. Reactions were incubated at 37°C and stopped by addition of SDS to a final concentration of 0.25%. After 15 minutes, fluorescence polarization was analyzed with a plate reader (Victor 3V, Perkin Elmer). Significance was evaluated by one-way ANOVA with P values reported from Tukey’s multiple comparisons test. Data analysis was carried out in *Prism* (GraphPad).

**Integrase strand transfer assay**

Assays were performed as previously described (Bushman, Fujiwara, and Craigie 1990; Bushman and Craigie 1991; K. Gao, Butler, and Bushman 2001; Gupta et al. 2016; Eilers et al. 2020). Final assay conditions were 3 μM IN, 20 mM HEPES-NaOH pH 7.5, 0.5 μM Alexafluor 488-labeled LTR substrate, 0.5 μM LEDGF$_{IBD}$, 150 mM NaCl,
10 mM MgCl$_2$, and 10 μM Zn(OAc)$_2$. Inhibitors were added from 10x stocks in DMSO to final concentrations as specified in the text. After 30 minutes at 37°C, pUC19 was added to a final concentration of 15 nM. Reactions were incubated for 1 hr at 37°C, then quenched by addition of 0.5% SDS, 15 mM EDTA, and 1 mg/mL proteinase K. Products were analyzed on 1.5% agarose gels and quantified by ImageJ. Data analysis was carried out in Prism (GraphPad). Dose-response curve fits used a four-parameter logistic regression with a variable Hill slope. Significance was evaluated by one-way ANOVA with P values reported from Tukey’s multiple comparisons test. Data analysis was carried out in Prism (GraphPad).

**Supplemental Figures**

**Supplemental Figure 3.1. Strand Transfer Activity of Wild Type and Dolastatvir-Resistant IN Constructs.** The baseline activity of wild type IN is greater than the activity of IN constructs containing IN resistance substitutions. This suggests that dolastatvir resistance substitutions arise with a fitness cost for the virus.
Supplemental Figure 3.2. Location of STI Resistance Substitutions. a) Location of dolutegravir resistance substitutions in context of dodecameric intasome (PDB: 6U8Q) b) Detail of Lys-156 association with viral DNA backbone. c) Location of dolutegravir resistance substitutions in context of IN-ALLIN polymer (PDB: 5HOT). The substitutions K211R, E212T are located on the CCD-CTD linker.
Supporting Information for Chapter 4

Construction of IN expression vectors

The NL4-3 HIV-1 IN coding sequence was amplified by PCR, fused to an N-His7-Flag-Sumo tag using 4-primer PCR, and cloned into a pCDFDuet expression vector. The fusion junction contains the sequence “G-G-F”, where cleavage by the SUMO protease Ulp1 occurs after the second glycine, liberating IN with a phenylalanine at position 1. IN GSH and IN MF were created by insertion of additional codons preceding the native phenylalanine by inverse PCR (IN GSH) or site-directed mutagenesis (IN MF). IN 1F<sub>NTD-CCD</sub><sup>F185K, W131D, F139D</sup> was constructed by truncation of the full-length construct and insertion of a synthetic cassette containing the amino acid substitutions. The lens epithelium derived growth factor (LEDGF) integrase binding domain (IBD) (residues 347-471) was cloned into a pETDuet expression vector with the Mxe intein, a chitin binding domain, and a His6 tag as previously described (Ciuffi et al. 2006).

Protein expression and purification

IN constructs were expressed as previously described with some modification (Gupta et al. 2016, 2010; Ciuffi et al. 2006; Diamond and Bushman 2005, 2006). Expression plasmids were transformed into E. coli BL21(DE3) and grown in 800mL of 2xYT at 37°C to an optical density of 1.8-2.2. Expression was induced by addition of isopropyl-β-D-1-thiogalactopyranoside (IPTG) and allowed to continue for 5 hours at 20°C. Bacteria were then pelleted and frozen at -80°C.

Full-length IN constructs were purified as described previously (Gupta et al. 2010; Ciuffi et al. 2006; Diamond and Bushman 2006). Briefly, lysates were loaded onto
nickel-nitrilotriacetic acid resin (Qiagen). Fusion proteins were eluted with 20 mM HEPES-NaOH pH 7.5, 1 M NaCl, 7 mM 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS), 10 µM ZnOAc₂, 5 mM β-mercaptoethanol, and 250 mM imidazole. Fusion proteins were liberated from IN by overnight cleavage with the SUMO protease Ulp1 (Life Sensors) at 4°C, with simultaneous dialysis against 20 mM HEPES-NaOH pH 7.5, 1 M NaCl, 7 mM CHAPS, 10 µM ZnOAc₂, and 5 mM β-mercaptoethanol. The affinity tag was separated from IN by a second nickel-nitrilotriacetic acid purification step and further purified using a Superdex 75 HiLoad 16/60 column (GE Healthcare) at room temperature, eluted isocratically in 20 mM HEPES-NaOH pH 7.5, 1 M NaCl, 7 mM CHAPS, 10 µM ZnOAc₂, and 2 mM dithiothreitol (DTT). IN 1F<sub>NTD-CCD</sub><sup>F185K, W131D, F139D</sup> was lysed in 50 mM sodium/potassium phosphate pH 7.0, 500 mM NaCl, 2 mM β-mercaptoethanol, and 10 mM imidazole, loaded onto nickel-nitrilotriacetic acid resin, and eluted with 50 mM sodium/potassium phosphate pH 7.0, 500 mM NaCl, 2 mM β-mercaptoethanol, and 250 mM imidazole. The affinity tag was liberated from IN by overnight cleavage with the SUMO protease Ulp1, with simultaneous dialysis against 20 mM sodium/potassium phosphate pH 7.0, 500 mM NaCl, and β-mercaptoethanol. The affinity tag was separated from IN by a second nickel-nitrilotriacetic acid purification step and further purified using a Superdex 75 HiLoad 16/60 column (GE Healthcare) at room temperature, eluted isocratically in 20 mM HEPES-NaOH pH 7.5, 500 mM NaCl, and 2 mM β-mercaptoethanol. IN was concentrated at 4°C in an Amicon Ultra-15 (Millipore), glycerol was added to a final concentration of 10% (w/v), and aliquots were flash-frozen in liquid nitrogen for storage at -80°C.
LEDGF IBD was purified using nickel-nitrilotriacetic acid (Qiagen) and chitin (New England Biolabs) resins. After fusion proteins were liberated by intein cleavage in 50 mM DTT overnight at 4°C, LEDGF IBD preparations were further purified using a Superdex 75 HiLoad 16/60 column (GE Healthcare) at room temperature, eluted isocratically in 20 mM HEPES-NaOH pH 7.0, 1 M NaCl, 7 mM CHAPS, 10 µM ZnOAc2, and 10 mM β-mercaptoethanol. The LEDGF IBD was concentrated at 4°C in an Amicon Ultra-15 (Millipore), glycerol was added to a final concentration of 10% (w/v), and aliquots were flash-frozen in liquid nitrogen for storage at -80°C.

Crystallization and structure determination

Crystals were grown by vapor diffusion as previously described (J.-Y. Wang et al. 2001). Briefly, 4 µL of protein at 5-10 mg/mL in 0.5 M NaCl, 20 mM HEPES pH 7.5, 100 µM ZnCl₂, 5% (w/v) glycerol, and 5 mM DTT was mixed with 4 µL of reservoir solution containing 0.7 M NaH₂PO₄, 1.0 M K₂HPO₄ and 0.1 M acetate pH 4.6. Two crystal forms were observed, flat hexagons and long tetragonal crystals, with only the latter exhibiting high resolution diffraction. Crystals were cryo-protected in 0.8 M NaH₂PO₄, 1.2 M K₂HPO₄, 0.2 M NaCl, and 20% glycerol and flash-frozen in liquid nitrogen. Diffraction data was collected at 100 K using an Eiger 9M pixel-array detector on beamline 17-ID-1 (AMX) at Brookhaven National Laboratory (M. S. Miller et al. 2019; Fuchs et al. 2016).

Diffraction data were reduced with DIALS (Winter et al. 2018). Molecular replacement, refinement, and the generation of simulated annealing omit maps were carried out in Phenix (Liebschner et al. 2019). The structure was solved by molecular replacement using 1K6Y as a search model. The asymmetric unit contained four monomers (each containing a Zn²⁺, K⁺, and phosphate ion) and 226 waters. The
structure was refined to a \( R \) and \( R_{\text{free}} \) of 22.5\% and 25.3\%, respectively. Molecular models were visualized with \textit{PyMOL} (Schrödinger 2015) and secondary structure was analyzed with \textit{Define Secondary Structure of Proteins (DSSP)} (Touw et al. 2015; Kabsch and Sander 1983).

**Integrase 3’-processing assay**

The 3’-processing assay was adapted from those described previously (Merkel et al. 2009; Guiot et al. 2006). HIV integrase at 60 \( \mu \)M in 20 mM HEPES-NaOH pH 7.5, 1 M NaCl, 7 mM CHAPS, 10 mM DTT, and 10 \( \mu \)M Zn(OAc)$_2$ was diluted to a final assay concentration of 400 nM with 20 mM HEPES-NaOH pH 7.5, 100 nM Alexafluor 488-labeled LTR substrate, 50 mM NaCl, 10 mM MgCl$_2$ or MnCl$_2$, 10 \( \mu \)M Zn(OAc)$_2$, and 10 mM DTT. Final assay conditions were identical for IN 1F, IN GSH, and IN MF. Unprocessed U5 LTR substrates with a 3’ Alexafluor 488 N-hydroxysuccinimide (NHS) ester label were prepared by annealing the following oligonucleotides (Integrated DNA Technologies):

- 5’-ACCCTTTTAGTCAGTGTGGAAAATCTCTAGCAGT-Alexa488-3’
- 5’-ACTGCTAGAGATTTTCCACACTGACTAAAAGGT-3’

Reactions were incubated at 37°C. SDS was added to a final concentration of 0.25\% to stop the reaction and liberate cleaved dinucleotide. After 15 minutes, fluorescence polarization was analyzed with a plate reader (Victor 3V, Perkin Elmer). Significance was evaluated by two-way ANOVA with P values reported from Tukey’s multiple comparisons test. Data analysis was carried out in \textit{Prism} (GraphPad).

**Integrase strand transfer assay**
The strand-transfer assay was adapted from those described previously (Bushman and Craigie 1991; Gupta et al. 2016; K. Gao, Butler, and Bushman 2001; Bushman, Fujiwara, and Craigie 1990). HIV integrase at 60 μM in 20 mM HEPES-NaOH pH 7.5, 1 M NaCl, 7 mM CHAPS, 10 mM DTT, and 10 μM Zn(OAc)₂ was diluted to a final assay concentration of 3 μM with 20 mM HEPES-NaOH pH 7.5, 0.5 μM Alexafluor 488-labeled LTR substrate, 0.5 μM LEDGF IBD, 50-250 mM NaCl, 10 mM MgCl₂ or MnCl₂, and 10 μM Zn(OAc)₂. Final assay conditions were identical for IN 1F, IN GSH, and IN MF. Processed U5 LTR substrates with a 5′ Alexafluor 488 N-hydroxysuccinimide (NHS) ester label were prepared by annealing the following oligonucleotides (Integrated DNA Technologies):

- 5′-Alexa488-ACCCTTTAGTCAGTGTGGAAAATCTCTAGCA-3′
- 5′-ACTGCTAGAGATTTTCCACACTAAAAGGGT-3′

After 30 minutes at 37°C, 15 nM pUC19 plasmid was added. Reactions were carried out for 1-4 hours at 37°C, then quenched using 0.5% SDS, 15 mM EDTA, and 1 mg/mL proteinase K for 30 minutes at 37°C. Reaction products were separated on 1.5% agarose gels in Tris-acetate buffer and imaged using a Typhoon (Amersham) imager. Gels were then stained with ethidium bromide and imaged using a Gel Doc (Bio-Rad) imager. Reaction products were quantified by ImageJ and data analysis was carried out in Prism (GraphPad). Significance was evaluated by two-way ANOVA with P values reported from Tukey’s multiple comparisons test. Dose-response curve fits were performed in Prism (GraphPad) using a three-parameter logistic regression with the Hill slope fixed at -1. The integrase inhibitor raltegravir was a gift from Merck.

Aggregation assay for ALLINIs
Assays were performed as previously described (Gupta et al. 2016, 2014) with some modification. Final reaction conditions were 20 mM HEPES-NaOH pH 7.5, 15 μM IN, 250-1000 mM NaCl, 7 mM CHAPS, and 30 μM ALLIN. The ALLINs BI-224436, BI-D, and CX04328 (HIV-1 integrase inhibitor 2) were purchased from MedChemExpress and resuspended in DMSO. Turbidity was measured after 20 minutes as the absorbance of the reaction solution at 405 nm in a plate reader (Victor 3V, Perkin Elmer). Significance was evaluated by two-way ANOVA with P values reported from Tukey’s multiple comparisons test.

**Size-exclusion chromatography in-line with multi-angle light scattering (SEC-MALS)**

Absolute molecular weights were determined by multi-angle light scattering coupled with refractive interferometric detection (Wyatt Technology Corporation) and a Superdex 200 10/300 column (GE Healthcare) at 25°C equilibrated in 20 mM HEPES-NaOH pH 7.5, 500 mM - 1 M NaCl, 7 mM CHAPS, 10 μM ZnOAc2, and 10 μM β-mercaptoethanol, as previously described (Gupta et al. 2010).

**Sedimentation velocity analytical ultracentrifugation (SV-AUC)**

SV-AUC experiments were performed at 25°C with an XL-A analytical ultracentrifuge (Beckman-Coulter) and a TiAn60 rotor with two-channel charcoal-filled epon centerpieces and quartz windows. Experiments were performed in 20 mM HEPES-NaOH pH 7.5, 1 M NaCl, 7 mM CHAPS, 10 μM ZnOAc2, and 10 μM β-mercaptoethanol. Complete sedimentation velocity profiles were collected every 30 s for 200 boundaries at 40,000 rpm. Data were fit using the c(s) distribution model of the Lamm equation as implemented in the program SEDFIT (Schuck 2000). After optimizing
meniscus position and fitting limits, the sedimentation coefficients and best-fit frictional ratio \((f/f_0)\) were determined by iterative least squares analysis. Sedimentation coefficients were corrected to \(s_{20,w}\) based on the calculated solvent density \((\rho)\) and viscosity \((\eta)\) derived from chemical composition by the program \textit{SEDNTERP} (Hayes et al. 2003).

**Sedimentation equilibrium analytical ultracentrifugation (SE-AUC)**

SE-AUC experiments were performed with an XL-A analytical ultracentrifuge (Beckman-Coulter) and a TiAn60 rotor with two-channel charcoal-filled epon centerpieces and quartz windows. Data were collected at 4°C with detection at 280 nm at multiple concentrations in 20 mM HEPES-NaOH pH 7.5, 1 M NaCl, 7 mM CHAPS, 10 \(\mu\text{M}\) ZnOAc\(_2\), and 10 \(\mu\text{M}\) \(\beta\)-mercaptoethanol. Analyses were carried out using global fits to data acquired at multiple speeds for each concentration with strict mass conservation using the program \textit{SEDPHAT} (Vistica et al. 2004). Error estimates for equilibrium constants were determined from a 1,000-iteration Monte Carlo simulation. The partial specific volume \(\left(\bar{\tau}\right)\), solvent density \((\rho)\), and viscosity \((\eta)\) were derived from chemical composition by \textit{SEDNTERP} (Hayes et al. 2003). SE-AUC data are summarized in Table 4.2.
Supplemental Figures

Supplemental Figure 4.1. Purification scheme and SDS-PAGE analysis of protein purifications. a) Purification scheme of IN with a native N-terminus (IN 1F). The poly-histidine (His7) affinity tag allows for capture of fusion proteins on Ni²⁺-NTA resin. Subsequent cleavage by the SUMO protease Ulp1 frees wild type IN with a phenylalanine at position 1. b) Coomassie-stained SDS-PAGE analysis of IN constructs after Ulp1 cleavage and size-exclusion chromatography. Expected protein size is 32 kDa. c) Coomassie-stained SDS-PAGE analysis of IN 1F,NTD,CCD with the substitutions F185K, W131D, and F139D that enable crystallization. Expected protein size is 23 kDa.

Supplemental Figure 4.2. Models demonstrating two possibilities of NTD orientation. IN is depicted as a dimer with both NTDs in either the "distal" or "proximal" orientation. Models are based on PDB structures 1K6Y, 5HOT, and 6VRG.
Supplemental Figure 4.3. Secondary structure comparison of IN structures. a) Secondary structure annotation of NTDs (residues 1-50) of IN 1F (PDB: 6VRG), IN GSH (PDB: 1K6Y), and Sso7d-IN (PDB: 5U1C, 6PUT, 6PUW, 6PUY, 6PUZ, and 6V3K) by DSSP. H = Alpha Helix, G = 3_{10} Helix, T = Hydrogen bonded turn, and S = Bend. b) Comparison of Sso7d-IN α1 helix structures. Chain A is predicted to form NTD-NTD interactions in dodecameric HIV-1 and hexadecameric MVV intasomes. Chain B is not predicted to form NTD-NTD interactions.
**Supplemental Figure 4.4. NaCl-dependence of strand transfer activity.** a) In the presence of Mg$^{2+}$, IN 1F and IN MF are most active at low NaCl concentrations, with activity disappearing above a NaCl concentration of 200 mM. The highest level of concerted integration activity is observed at 150 mM NaCl. b) In the presence of Mn$^{2+}$, IN 1F and IN MF are most active at NaCl concentrations higher than in the presence of Mg$^{2+}$. The highest level of concerted integration activity is observed at 250 mM NaCl.

**Supplemental Figure 4.5. Effect of raltegravir on strand transfer activity of IN 1F, IN GSH, and IN 1F.** Data are the same as in Figure 3e but plotted as molecules integrated per minute of single strand (left) and concerted integration (right) without normalization. Data plotted as mean ± SD of 3 replicates.
Supplemental Figure 4.6. Biophysical and biochemical characterization of IN 1F^{T188H}. a) SEC-MALS analysis of IN 1F^{T188H} shows a mixture of dimers and higher-order aggregates in solution. b) Quantification of fluorescence polarization assay for 3′ processing. IN 1F^{T188H} performs 3′-processing more rapidly than IN 1F, a difference that reaches statistical significance, but has unclear biological relevance. c) Example gel image of results from strand transfer assay in the presence of Mg^{2+}. d) Quantification of strand transfer activity. The single strand and concerted strand transfer activity of IN 1F^{T188H} is significantly decreased as compared to wild type IN 1F. Data plotted as mean ± SD of 3 replicates. * denotes P<0.05, *** denotes P=0.0005, **** denotes P<0.0001.
**Supporting Information for Chapter 5**

**IN·LEDGF coexpressions**

Expression plasmids containing full-length or truncated IN 1F constructs (Eilers et al. 2020) and LEDGF\textsubscript{IBD} (Ciuffi et al. 2006) were transformed into \textit{E. coli} BL21(DE3) with or without a pET28a plasmid encoding the Ulp1 protease (Mossessova and Lima 2000) with an N-terminal, thrombin-cleavable polyhistidine tag. The IN 1F constructs contain an N-terminal His7-FLAG-SUMO tag, that, when co-expressed with Ulp1 in vivo or cleaved by Ulp1 in vitro, is cleaved off, liberating IN with a native phenylalanine at position 1. The LEDGF\textsubscript{IBD} contains a C-terminal Mxe intein, chitin-binding domain, and polyhistidine tag. Since IN 1F constructs lack a chitin-binding domain, purification by chitin affinity selectively binds LEDGF\textsubscript{IBD} and its binding partners. NL4-3 IN was used for co-expressions of full-length IN with LEDGF\textsubscript{IBD}. For coexpressions of full-length IN or IN\textsubscript{NTD-CCD} with LEDGF\textsubscript{IBD} and Ulp1, NL4-3 IN\textsuperscript{T124N} was used. T124N is a commonly-occurring polymorphism (B. T. Foley et al. 2018) that has been observed to be positively selected in the presence of ALLINIs (Sharma et al. 2014; Gupta et al. 2016; Hoyte et al. 2017). Coexpressions were expressed as described previously, with some modification (Ciuffi et al. 2006; Diamond and Bushman 2006; Gupta et al. 2014, 2010). Cultures of \textit{E. coli} BL21(DE3) containing expression plasmids was grown in 800mL of 2\texttimes YT with 10 µM ZnSO\textsubscript{4} and 2 mM MgSO\textsubscript{4} at 37°C to an optical density of 1.8-2.2. After addition of isopropyl-β-D-1-thiogalactopyranoside (IPTG), expression was allowed to continue for 5 hours at 20°C. Bacteria were then pelleted and frozen at -80°C.

**IN·LEDGF copurifications**

IN and LEDGF were purified as previously described (Eilers et al. 2020)
Lysates of IN-LEDGF<sub>IBD</sub> coexpressions in 20 mM sodium/potassium phosphate pH 7.0 and 300 mM NaCl and protease inhibitor were purified using chitin (New England Biolabs) resin. LEDGF<sub>IBD</sub> was liberated by intein cleavage in the presence of 5mM CHAPS with 100 mM β-mercaptoethanol at 4°C overnight. A second chitin purification step separated the affinity tag from liberated LEDGF<sub>IBD</sub> and bound IN. IN-LEDGF<sub>IBD</sub> complexes were concentrated at 4°C in an Amicon Ultra-15 (Millipore) before injection onto a Superdex 75 HiLoad 16/60 column (GE Healthcare) at room temperature. Complexes were eluted isocratically in 20 mM HEPES-NaOH pH 7.5, 300 mM NaCl, 5 mM CHAPS, 10 μM ZnOAc<sub>2</sub>, and 5 mM β-mercaptoethanol or dithiothreitol and concentrated at 4°C in an Amicon Ultra-15 (Millipore).

**Size-exclusion chromatography in-line with multi-angle light scattering (SEC-MALS)**

Complexes were analyzed by separation on a Superdex 200 10/300 column (GE Healthcare) at 25°C, eluted isocratically in 20 mM HEPES-NaOH pH 7.5, 300 mM NaCl, 5 mM CHAPS, 10 μM ZnOAc<sub>2</sub>, and 10 μM β-mercaptoethanol. Absolute molecular weights were determined by multi-angle light scattering coupled with refractive interferometric detection (Wyatt Technology Corporation), as previously described (Gupta et al. 2010).

**Integrase strand transfer assay**

Assays were performed as previously described (Eilers et al. 2020). Final assay conditions were approximately 3 μM IN, 1.5 μM LEDGF<sub>IBD</sub> (as estimated from A<sub>280</sub> and SDS-PAGE), 20 mM HEPES-NaOH pH 7.5, 0.5 μM Alexafluor 488-labeled LTR substrate, 150 mM NaCl, 10 mM MgCl<sub>2</sub>, and 10 μM Zn(OAc)<sub>2</sub>. Pickle products were
analyzed on 1.5% agarose gels, imaged for fluorescence and then stained with ethidium bromide.

**Turbidity assay for integrase stability**

We have previously reported an integrase aggregation assay based on the measurement of turbidity in solution (Gupta et al. 2014, 2016; Eilers et al. 2020). This assay was adapted to measure the precipitation of integrase in low ionic strength conditions. Final conditions were 30 µM IN, 10 µM LTR DNA, 20 mM HEPES pH 7.5, and 7 mM CHAPS for IN alone and IN·LEDGF\textsubscript{IBD} complexes. CHAPS was dialyzed out of IN·LEDGF\textsubscript{IBD} complexes before mixing with LTR DNA, further emphasizing the solubilizing effect of the LTR DNA. Turbidity was measured for 30 minutes as the absorbance of the reaction solution at 405 nm in a plate reader (Victor 3V, Perkin Elmer) and plotted throughout the course of the experiment (Figure 5.4a,b) or as turbidity at the end of the 30 minute experiment (Figure 5.4c).

**Intasome preparations**

Intasomes were prepared as previously described (Hare, Gupta, et al. 2010; Pandey et al. 2014; Yin et al. 2016; Ballandras-Colas et al. 2016, 2017; Passos et al. 2017, 2020; Cook et al. 2020) with some modification. DNA templates were prepared by mixing oligonucleotides at equimolar ratios in the presence of 10 mM Tris, 0.1 mM EDTA, and 100 mM NaCl. Mixtures were heated to 95°C and then slowly cooled. For low ionic strength dialysis, mixtures of ~50 µM IN, 25 µM LEDGF\textsubscript{IBD} (as estimated from A\textsubscript{280} and SDS-PAGE), and 6.6 µM 3WJ DNA (Supplemental Figure 5.1b) were dialyzed to 5 mM HEPES-NaOH pH 7.5, 100 mM NaCl, 10 µM ZnSO4, and 5 mM DTT in 3K MWCO
dialysis containers (Thermo Scientific) at 4°C overnight. Samples were briefly centrifuged to remove precipitated protein and DNA. Samples were then taken for chemical crosslinking, polyacrylamide EMSA, and SDS-PAGE analysis. For experiments that were spiked to low ionic strength, mixtures of 0.5-5 μM IN, 1-3 μM LEDGF_{IBD} (where indicated), and 0.5 μM fluorescently labeled oligonucleotide (Supplemental Figure 5.1c) in 20 mM HEPES-NaOH pH 7.5, 10 mM MgCl₂ or CaCl₂ (where indicated), and 2 μM or 20 μM raltegravir (where indicated). The integrase inhibitor raltegravir was a gift from Merck. Mixtures were incubated at 37°C for 1 hour. Samples were then taken for agarose EMSA.

**Chemical Crosslinking**

For crosslinking experiments, the crosslinkers disuccinimidyl glutarate (DSG) and disuccinimidyl suberate (DSS) were resuspended in DMSO and added to a final concentration of 0.25 mM (~5x molar excess over IN) and 1 mM (~20x molar excess over IN). Crosslinking was allowed to proceed for 30 minutes at room temperature before quenching with 25 mM final Tris pH 7.5. for 15 minutes at room temperature. Samples were analyzed by SDS-PAGE.

**Agarose electrophoretic mobility shift assay**

Samples were taken from low ionic strength intasome preparations and loading buffer containing glycerol and bromophenol blue was added. Samples were separated on 3% agarose gels in 0.5x Tris-Borate running buffer. Ethidium bromide-free buffers and apparatus were used. Gels were run at 15 v/cm for 30 minutes and then imaged for
fluorescence using a Typhoon (Amersham) imager. Gels were then stained with ethidium bromide and imaged using a Gel Doc (Bio-Rad) imager.

**Polyacrylamide electrophoretic mobility shift assay**

Samples were taken from overnight dialysis intasome preparations and loading buffer containing glycerol and bromophenol blue was added. Samples were separated on 6% polyacrylamide gels in 1x Tris-Borate-EDTA running buffer. Gels were run at 5-10 v/cm for 2 hours at 4°C and then stained with Gel Star (Lonza) for detection of nucleic acid or coomassie blue for detection of protein. Gel Star-stained gels were imaged using a Typhoon (Amersham) imager, and coomassie-stained gels were imaged using a Gel Doc (Bio-Rad) imager.

**Supplemental Figures**

Supplemental Figure 5.1. Intasome DNA Templates. a) Oligonucleotides mimicking the viral LTR ends, used to form a preintegration complex. This oligo complex was used in aggregation experiments. b) Annealed three-way junction complex mimicking the covalent linkage of viral DNA to host cell DNA. This oligo complex was used in crosslinking and polyacrylamide EMSA experiments. c) Fluorescently labeled oligonucleotides mimicking the viral LTR ends. This oligo complex was used in agarose EMSA experiments. Filled circles designate 5’-ends, and the green star indicates the fluorophore.


Balakrishnan, Mini, Stephen R. Yant, Luong Tsai, Christopher O’Sullivan, Rujuta A. Bam, Angela Tsai, Anita Niedziela-Majka, Kirsten M. Stray, Roman Sakowicz, and Tomas Cihlar. 2013. “Non-Catalytic Site HIV-1 Integrase Inhibitors Disrupt Core Maturation and Induce a Reverse Transcription Block in Target Cells.” PloS One 8 (9): e74163.


Bera, Sibes, Krishan K. Pandey, Ajaykumar C. Vora, and Duane P. Grandgenett. 2009. “Molecular Interactions between HIV-1 Integrase and the Two Viral DNA Ends within
the Synaptic Complex That Mediates Concerted Integration.” *Journal of Molecular Biology* 389 (1): 183–98.


Brenner, Bluma, and Mark A. Wainberg. 2016. “We Need to Use the Best Antiretroviral Drugs Worldwide to Prevent HIV Drug Resistance.” *AIDS*.


Bukrinsky, M. I., N. Sharova, T. L. McDonald, T. Pushkarskaya, W. G. Tarpley, and M.


Cherepanov, Peter, Andre L. B. Ambrosio, Shaila Rahman, Tom Ellenberger, and Alan


around Protein-Protein Interfaces Suggests a General Mechanism for Pocket Formation.” Proceedings of the National Academy of Sciences of the United States of America 109 (10): 3784–89.


Jenkins, Timothy M., Alan Engelman, Rodolfo Ghirlando, and Robert Craigie. 1996. “A


9826–33.


Molina, Jean-Michel, Anthony Lamarca, Jaime Andrade-Villanueva, Bonaventura Clotet,


Thompson, Melanie A., Michael J. Mugavero, K. Rivet Amico, Victoria A. Cargill, Larry
Entry into and Retention in Care and Antiretroviral Adherence for Persons with HIV:
Evidence-Based Recommendations from an International Association of Physicians
294.

Touw, Wouter G., Coos Baakman, Jon Black, Tim A. H. te Beek, E. Krieger, Robbie P.
Joosten, and Gert Frient. 2015. “A Series of PDB-Related Databanks for Everyday

Trivedi, Jay, Dinesh Mahajan, Russell J. Jaffe, Arpan Acharya, Debashis Mitra, and
Siddappa N. Byrareddy. 2020. “Recent Advances in the Development of Integrase
Inhibitors for HIV Treatment.” *Current HIV/AIDS Reports* 17 (1): 63–75.

Tsantrizos, Youla, Murray Bailey, François Bilodeau, Rebekah Carson, René Coulombe,

Tse, Celine, Lauren Wickstrom, Mamuka Kvaratskhelia, Emilio Gallicchio, Ronald Levy,
and Nanjie Deng. 2020. “Exploring the Free-Energy Landscape and
Thermodynamics of Protein-Protein Association.” *Biophysical Journal*, August.
https://doi.org/10.1016/j.bpj.2020.08.005.

Tsiang, Manuel, Gregg S. Jones, Joshua Goldsmith, Andrew Mulato, Derek Hansen,
Elaine Kan, Luong Tsai, et al. 2016. “Antiviral Activity of Bictegravir (GS-9883), a
Novel Potent HIV-1 Integrase Strand Transfer Inhibitor with an Improved Resistance
Profile.” *Antimicrobial Agents and Chemotherapy* 60 (12): 7086–97.

Tsiang, Manuel, Gregg S. Jones, Anita Niedziela-Majka, Elaine Kan, Eric B. Lansdon,
Wayne Huang, Magdeleine Hung, et al. 2012. “New Class of HIV-1 Integrase (IN)
Inhibitors with a Dual Mode of Action.” *The Journal of Biological Chemistry* 287 (25):
21189–203.

Proteins and Human Immunodeficiency Virus DNA Integration.” *Frontiers in


Valkov, Eugene, Saumya Shree Gupta, Stephen Hare, Anna Helander, Pietro Roversi,
Characterization of the Integrase from the Prototype Foamy Virus.” *Nucleic Acids

Vandekerckhove, Linos, Frauke Christ, Bénédicte Van Maele, Jan De Rijck, Rik Gijsbers,
Chris Van den Haute, Myriam Witvrouw, and Zeger Debyser. 2006. “Transient and
Stable Knockdown of the Integrase Cofactor LEDGF/p75 Reveals Its Role in the
Replication Cycle of Human Immunodeficiency Virus.” *Journal of Virology* 80 (4):
1886–96.

Vanegas, Maria, Manuel Llano, Sharon Delgado, Daniah Thompson, Mary Peretz, and


