Published in final edited form as:

Subcell Biochem. 2018 ; 88: 189–210. doi:10.1007/978-981-10-8456-0_9.

Nucleoprotein intermediates in HIV-1 DNA integration: structure and function of HIV-1 intasomes

Robert Craigie

Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892

Abstract

Integration of a DNA copy of the viral genome into host DNA is an essential step in the replication cycle HIV-1 and other retroviruses and is an important therapeutic target for drugs. DNA integration is catalyzed by the viral integrase protein and proceeds through a series of stable nucleoprotein complexes of integrase, viral DNA ends and target DNA. These nucleoprotein complexes are collectively called intasomes. Retroviral intasomes undergo a series of transitions between initial formation and catalysis of the DNA cutting and joining steps of DNA integration. Intasomes, rather than free integrase protein, are the target of currently approved drugs that target HIV-1 DNA integration. High-resolution structures of HIV-1 intasomes are needed to understand their detailed mechanism of action and how HIV-1 may escape by developing resistance. Here, we focus on our current knowledge of the structure and function of HIV-1 intasomes, with reference to related systems as required to put this knowledge in context.

1. INTRODUCTION

Integration of a DNA copy of the viral genome into host DNA is an essential step in the replication cycle of HIV-1 and other retroviruses (Coffin et al. 1977). Once integrated, the viral DNA is replicated along with cellular DNA during each cycle of cell division, and populations of long-lived cells with integrated proviruses have obstructed efforts to cure AIDS; although viral replication can now be effectively suppressed by antiviral drugs, these infected cells are a reservoir from which virus reemerges if treatment is interrupted. Biochemical studies have shown that retroviral DNA integration occurs by a mechanism that is shared by a class of DNA mobile genetic elements that are ubiquitous in both prokaryotes and eukaryotes, and by retrotansposons. Although the mechanism of DNA integration is closely related among these classes elements, the source of the DNA to be integrated differs. In the case of DNA transposons, the transpose encoded by the transposon excises the transposon from its original location in the genome and inserts it into a new location. Retrotransposons must first transcribe an RNA copy of their genome which then undergoes an intermediate step of reverse transcription within the same cell to make the DNA copy that is then integrated at a new site. Retroviruses have evolved the additional step of packaging the transcribed viral RNA in the form of a virion that is budded from the infected cell. The virion subsequently infects another cell where reverse transcription and DNA integration

occur. Transposons, retrotansposons, and retroviruses share the common feature that, prior to integration, the two ends of the mobile DNA are tightly associated with the enzyme that catalyzes the DNA integration reaction. This protein is called transposase in the case of transposons and integrase in the case of retrotransposons and retroviruses. In retroviruses and retrotransposons, terminal CA dinucleotides are joined to target DNA, whereas terminal sequences are more divergent among transposons. Complexes between integrase and viral DNA are collectively termed intasomes. Retroviral intasomes undergo a series of transitions between initial formation and catalysis of the DNA cutting and joining steps of DNA integration. Here, we focus on our current knowledge of the structure and function of HIV-1 intasomes, with reference to related systems as required to put this knowledge in context. First, we review key discoveries that led to the recent breakthroughs with high-resolution structural studies of HIV-1 intasomes.

2. MECHANISM OF DNA INTEGRATION

2.1 The preintegration complex (PIC)

The establishment of an in vitro system for retroviral DNA integration by Brown and colleagues in 1987 (Brown et al. 1987) was a pivotal step in biochemical studies of retroviral DNA integration. They discovered that cytoplasmic extracts of cells infected with Moloney murine leukemia virus (MoMLV) supported integration of the viral DNA made by reverse transcription into an exogenously added plasmid DNA in vitro. Importantly, the integrated viral DNA was flanked by a 4 bp repeat of target DNA at the site of integration, a hallmark of correct MoMLV DNA integration. Integration activity sedimented as a very large nucleoprotein complex with an S value of approximately 160S (Bowerman et al. 1989); for comparison, the S value of eukaryotic ribosomes is 80S. These complexes have been termed preintegration complexes (PICs). HIV-1 DNA was subsequently shown to form part of similarly large preintegration complexes (Ellison et al. 1990; Farnet and Haseltine 1990). In addition to viral DNA, PICs contain many proteins derived from the infecting virion and cellular proteins acquired from the cytoplasm of the infected cell (Farnet and Haseltine 1991; Bukrinsky et al. 1993; Li et al. 2001). The organization and composition of PICs is still poorly defined because their low abundance in extracts of infected cells limits the types of studies that can be attempted. It is likely that components of the PIC are jettisoned between initial formation after reverse transcription in the cytoplasm and transport to the nucleus for integration into cellular DNA. However, functional studies of integration activity clearly show that integrase must be tightly associated with the viral DNA ends because integrase is the enzyme that catalyzes integration (see below) and integration activity is retained upon treating PICs with greater than 0.5 M NaCl and separating them from loosely-bound protein factors by gel filtration or sedimentation. These complexes of integrase associated with the viral DNA ends (intasomes) form a sub-structure of the PIC and are the topic of this chapter. Since the low abundance of PICs in infected cells precludes structural studies, recent progress in this area has built upon biochemical studies of intasomes assembled from purified components in vitro.

2.2 DNA cutting and joining steps

In vitro DNA integration assays using PICs isolated from infected cells provided the first definitive evidence that the immediate precursor DNA for integration is linear. At the time, retroviral DNA was known to exist in both linear and circular forms and the circular form was a strong candidate based on analogy with the extensively studied bacteriophage lambda integration system. Analysis of the in vitro product of DNA integration, revealed that only the 3' ends of the viral DNA are joined to target DNA at the site of integration. The unjoined 5' ends of the viral DNA extended two nucleotides beyond the CA terminal nucleotides on the joined viral DNA strands, consistent with a linear viral precursor, and not four nucleotides as would be expected if the viral DNA were circularized prior to integration (Fujiwara and Mizuuchi 1988; Brown et al. 1989). The DNA cutting and joining steps deduced from these studies are depicted in Figure 1. In the first step, 3' end processing, two nucleotides are generally removed from the 3' ends of the initially blunt-ended linear viral DNA (Figure 1B). A target DNA is recruited (Figure 1C) and then in the second step, DNA strand transfer, the hydroxyl groups at each 3' end of the viral DNA attack a pair of phosphodiester bonds in the target DNA. The sites of attack on the two target DNA strands are separated by five nucleotides in the case of HIV-1.

In the integration intermediate (Figure 1D), only the 3' ends of the viral DNA are joined to target DNA and the 5' ends of the viral DNA remain unjoined with a 5 bp overhang. Completion of integration requires removal of each overhanging dinucleotide, filling in of the single strand gaps between viral and target DNA by a polymerase and ligation. The enzymes that carry out these steps have not been determined, but the reactions can be readily carried in vitro out by the known activities of cellular nucleases, polymerases and ligases (Yoder and Bushman 2000).

3. Integrase

3.1 Discovery of integrase

Early evidence for a key role of a virally encoded enzyme in retroviral DNA integration came from genetic experiments (Panganiban and Temin 1984; Donehower and Varmus 1984; Schwartzberg et al. 1984). Mutations that abolished integration, but allowed infection to proceed normally up completion of the viral DNA synthesis by reverse transcription, mapped to 3' end of the viral *pol* gene. This part of *pol* includes a protein that we now know catalyzes the key steps of DNA integration and is called integrase (IN). Mutations that disrupt the termini of the viral DNA resulted in an essentially identical phenotype, suggesting that this protein acts on the viral DNA ends to promote integration, even before any biochemical evidence was obtained (Panganiban and Temin 1983; Colicelli and Goff 1985; Roth et al. 1989). However, biochemical studies of integrase actually began much earlier when Grandgenett (Grandgenett et al. 1978) and colleagues identified a protein from avian sarcoma and leucosis virus (ASLV) (p33) that had a DNA nicking activity on substrate corresponding to the circularized form of ASLV DNA. This protein was later identified as the ASLV integrase protein.

3.2 Early in vitro assay systems

The finding that detergent-disrupted MoMLV virions could support in vitro integration of linear DNA with termini that mimic the viral DNA ends (Fujiwara and Craigie 1989) suggested that integrase might be sufficient to carry out the DNA cutting and joining steps of integration. This was confirmed by assays using purified integrase protein of MoLV (Craigie et al. 1990), ASLV (Katzman et al. 1989) (Katz et al. 1990) and HIV-1 (Sherman and Fyfe 1990; Bushman et al. 1990; Engelman et al. 1991) as the only protein factor. The first assay systems used a linear “mini-viral DNA” substrate with phage lambda DNA as the target for integration (Fujiwara and Craigie 1989; Bushman et al. 1990). Integration events were scored by in vitro packaging of the lambda DNA, infection of *E. coli* and selection for a marker present on the mini-viral DNA substrate. Subsequent assays used oligonucleotides mimicking the ends of the viral DNA as the viral DNA substrate; oligonucleotides also served as the target DNA for integration and the products were detected by gel electrophoresis (Figure 2). In the first step of the reaction (3' end processing) two nucleotides are cleaved from the 3' end of the viral DNA substrate to expose the CA-3' ends that are to be joined to target DNA (Figure 2A). In the next step (DNA strand transfer), the hydroxyl groups at the termini exposed by the 3' end processing reaction attack a pair of phosphodiester bonds in the target DNA (Figure 2B). Integration can occur at essentially any position along the target DNA, so the products are heterogeneous in length. In the case of HIV-1, the products of DNA strand transfer produced in such assays resulted largely from integration of a single viral DNA end into a single strand of target DNA, rather than two viral DNA ends integrated into each target DNA strand with the sites of integration separated by 5 bp as occurs in the normal DNA strand transfer reaction. A separate in vitro activity, termed disintegration, represents the chemical reversal of DNA strand transfer (Chow et al. 1992). Although not thought to occur during virus infection, disintegration activity provided useful information on the domain organization of integrase protein (see below).

3.3 Chemistry of DNA integration

The development of simple assay systems with purified components and a physical readout ignited detailed biochemical studies of integration and laid the foundation to the development of high-throughput screens for inhibitors of integrase. Many DNA recombinases catalyze recombination by cleaving DNA and conserving bond energy by forming a covalent intermediate with one of the cleaved DNA ends. However, in the case of retroviral DNA integration, the 3' hydroxyl of the DNA strand to be joined is exposed prior to the joining reaction. This raised the possibility that DNA strand transfer might occur by a one-step transesterification reaction in which phosphodiester bond cleavage and formation of the new phosphodiester bond occur in one step. This was tested by incorporating chiral phosphorothioate in the target DNA and monitoring the change in chirality after the reaction. If a covalent protein-DNA intermediate is involved, two reaction steps would be required and the chirality would be retained, whereas a single step transesterification would result in inversion of chirality. The chirality was found to invert, showing that integration occurred by a one-step transesterification mechanism (Engelman et al. 1991). In vitro, the product of 3' end processing exists as both a linear dinucleotide and a circular dinucleotide in which the 3' end of the viral DNA itself bends back and attacks the scissile phosphodiester bond. This

enabled the stereochemical course of the 3' end processing reaction to also be followed by putting a chiral phosphorothioate at the site of cleavage. The chirality inverted during the course of this reaction as well, demonstrating that 3' end processing also occurs by a one-step mechanism (Engelman et al. 1991).

3.4 Domain organization and early mutagenesis studies

Partial proteolysis studies revealed that HIV-1 integrase is comprised of three domains (Engelman and Craigie 1992)(Figure 3). Protein sequence alignment revealed three amino residues, D64, D116, and E152 that are absolutely conserved among retroviral integrases, retrotransposon integrases, and some classes of transposases. Mutation of any of these residues abolished 3' processing and DNA strand transfer activity (Kulkosky et al. 1992; Engelman and Craigie 1992; van Gent et al. 1992; Leavitt et al. 1993). These studies, together with inversion of chirality for both DNA strand transfer and 3' end processing, suggested that both reactions share a common chemical mechanism and are catalyzed by the same active site. The residues of the D,D-35-E motif were proposed to coordinate divalent metal ions for catalysis (Kulkosky et al. 1992) and this role was subsequently confirmed by structural studies. The N-terminal domain is well conserved among retroviral integrases and contains a pair of His and Cys residues, the HHCC motif, that binds zinc (Zheng et al. 1996). The C-terminal domain is the least conserved of the three domains.

4. HIV-1 Integrase domain structures

4.1 Catalytic domain structure

Structural studies of HIV-1 integrase and other retroviral integrases have been impeded by the propensity of the protein to aggregate and early attempts at crystallization were unsuccessful. This led investigators to focus their efforts on individual domains. The catalytic core domain of HIV-1 integrase, like the full-length proteins, aggregates in solution. However, the mutation F185K in the Catalytic domain was found to result in a protein that is much more soluble than the wild type Catalytic domain and retains disintegration activity in vitro (Jenkins et al. 1995). The domain with this mutation was successfully crystallized and the structure was solved (Dyda et al. 1994). The structure contained a five-stranded beta sheet flanked by helices (Figure 4A), with two of the previously identified catalytic residues in close proximity, while the third catalytic residue was on a disordered loop. The structure established that integrase belongs to a superfamily of polynucleotidyl transferases that at the time included ribonuclease H (Yang et al. 1990) and the Holliday junction resolvase RuvC (Ariyoshi et al. 1994), but has subsequently expanded to include many additional enzymes. Unfortunately, the structure provided little information on how integrase interacts with DNA. The two active sites in the structure, which were located on opposite sides of the nearly spherical dimer, were separated by approximately 30Å. This spacing is incompatible with the 5 bp spacing of the sites of joining of the two viral DNA ends to two strands of target DNA and suggested that the active multimer must be larger than a dimer.

4.2 C-terminal domain structure

The structure of the C-terminal domain was solved by NMR (Lodi et al. 1995). The protein was a dimer in solution with each subunit comprised of a five-stranded β barrel with a topology similar to the Src-homology 3 (SH3) domain. (Figure 4B)

4.3 N-terminal domain structure

The N-terminal domain structure was also determined by NMR (Cai et al. 1997). The structure was dimeric with each subunit comprising four-helix bundle with zinc tetrahedrally coordinated to the conserved residues of the HHCC motif, H12, H16, C40 and C43 (Figure 4C). The two subunits in the dimer are arranged parallel to each other and the interface is mainly hydrophobic. The N-terminal domain of HIV-2 integrase has an essentially similar structure (Eijkelenboom et al. 1997).

4.4 Multi-domain structures

The first multi-domain integrase structure to be solved was the Catalytic domain together with the C-terminal domain (Chen et al. 2000a). The structure was a Y-shaped dimer (Figure 5A). The catalytic domain formed the only dimer interface and this interface was essentially the same as in the isolated Catalytic domain dimer. The C-terminal domains were monomeric, 55Å apart and linked to the catalytic domain by alpha helices. The second multi-domain structure to be determined was the N-terminal together with Catalytic domain (Wang et al. 2001). This structure is dimeric with the same Catalytic domain dimer interface as observed in the earlier structures (Figure 5B). Although dimerized, the N-terminal dimer interface was different from that of the solution structure of the N-terminal domain alone. The lack of density for the linker connecting the two domains did not allow unambiguous assignment of the connectivity between the N-terminal and Catalytic domains. The missing linker can easily span the distance in the arrangement shown at the top of Figure 5B. The connectivity shown below cannot be excluded, although it would require extreme stretching of the linker. However, the N-terminal domain is located in a similar flanking position in the structure of the N-terminal together with the Catalytic domain of HIV-2 integrase (Hare et al. 2009; Jaskolski et al. 2009).

The structures of the isolated domains and two-domain structures provided high-resolution information on the individual domains but provided few clues as how viral DNA assembles with integrase in the intasome.

5. Intasomes

5.1 Intasomes

Intasome is a collective term for stable nucleoprotein complexes on the retroviral DNA integration reaction pathway (Figure 6) and is analogous to transpososomes that mediate transposition of a class of DNA transposons. The first intasome on the reaction pathway is the Stable Synaptic Complex (SSC) which comprises a pair of viral DNA ends bridged by integrase. Within the SSC, two nucleotides are removed from the 3' ends of the viral DNA to form the Cleaved Stable Synaptic Complex (cSSC). The cSSC then non-covalently captures

a target DNA to form the Target Capture Complex (TCC). Finally, within the TCC, a pair of transesterification reactions occurs to form the Strand Transfer Complex (STC).

5.2 Biochemical studies of HIV-1 intasomes

The first indication that the ends of the viral DNA are stably bridged by integrase in the preintegration complex (PIC) came with functional studies of Moloney murine leukemia virus PICs. PICs incubated at high ionic strength and separated from released protein factors by sedimentation or gel filtration retain integration activity, suggesting that integrase remains associated with the viral DNA ends (Lee and Craigie 1994).

Early biochemical studies HIV-1 DNA integration used reaction conditions in which most of the integration products resulted from “half-site” integration of a single viral DNA end into one strand of target DNA. Under these conditions, highly stable complexes of integrase with viral DNA ends were not detected. However, under improved reaction conditions that support concerted integration of pairs of viral DNA ends (Hindmarsh and Leis 1999; Sinha et al. 2002; Li and Craigie 2005; Sinha and Grandgenett 2005), highly stable complexes of integrase and viral DNA ends (SSC intasomes) are formed (Li et al. 2006; Pandey et al. 2007). Although the overall efficiency of concerted two-end DNA integration in vitro is low, the limiting step is formation of intasomes; once assembled essentially all the intasomes successfully complete DNA strand transfer to form STCs (Li et al. 2006).

5.3 Overcoming obstacles to high-resolution structural studies of HIV-1 intasomes

Although the structures of all the individual domains of HIV-1 integrase were solved by the mid 1990s, it would be almost two decades before the first HIV-1 intasome structures were determined. The early biochemical assays that resulted in mainly single-site integration appear to be a “loosened-up” reaction system in which the chemical steps of integration occur normally, but stable intasomes are by-passed. Improvements in in vitro assay systems were required for intasomes to be detected biochemically. The biggest obstacle to structural studies was the tendency of integrase and intasomes to aggregate in solution. This problem was partly overcome by the finding that fusing the DNA binding domain of *Sulfolobus solfataricus* chromosomal protein Sso7d to the N-terminus of HIV-1 integrase resulted in a protein that was not only hyperactive in vitro, but also assembled intasomes with greatly improved solubility properties (Li et al. 2014). The fusion protein is also active with oligonucleotide viral DNA substrate, whereas the wild type protein requires several hundred additional base pairs of non-specific DNA for integration activity and assembly of intasomes (Li and Craigie 2005; Li et al. 2006). The reasons for these differences are not understood. Importantly, the Sso7d fusion IN is active in virus, albeit with a somewhat lower efficiency than wild type HIV-1 IN (Li et al. 2014).

6. Structure of prototype foamy virus (PFV) intasomes

A major breakthrough in structural studies of intasomes came when the Cherepanov group solved the crystal structures of the prototype foamy virus (PFV) SSC and STC intasomes (Hare et al. 2010a; Maertens et al. 2010). PFV is a relative of HIV-1 with relatively low sequence similarity, but shares predicted secondary structures and key active site residues

(Valkov et al. 2009). It also has an additional small domain at the N-terminus, the N-terminal extension domain (NED), that is not present in HIV-1 integrase. PFV integrase is much more active than HIV-1 integrase in vitro and, unlike HIV-1 integrase, both PFV integrase and PFV intasomes are highly soluble. The PFV SSC structure is a tetramer comprised of a dimer of dimers (Figure 7). Each dimer has essentially the same Catalytic dimer as in the isolated HIV-1 integrase Catalytic domain structure. All the interactions with viral DNA occur with the domains of the inner subunits within the tetramer. The N-terminal domains of the two inner subunits are exchanged and extend to the active site region of each other. The NED, N-terminal, and C-terminal domains of the outer subunits are disordered in the structure. DNA-protein interactions play a dominant role and act as the “glue” that stabilizes the intasome. Such a structure could not have been predicted from the structures of the isolated integrase domains. Structures of PFV intasomes that have been determined importantly include all complexes on the reaction pathway (SSC, cSSC, TCC, and STC) (Hare et al. 2010a; Maertens et al. 2010; Hare et al. 2012) as well as the cSSC in complex with DNA strand transfer inhibitors (Hare et al. 2010a; Hare et al. 2010b). The structure of the cSSC in complex with compounds such as raltegravir elucidated the mechanism of this class of inhibitor. Binding of the inhibitor displaces the attacking 3' hydroxyl from the active site. Although the PFV structures are excellent for modeling the interaction of HIV-1 integrase inhibitors, sequence differences necessitate structures of HIV-1 intasomes in complex with inhibitors to understand the detailed molecular interactions of HIV-1 intasomes with inhibitors and the structural basis for clinical drug resistance. This is especially true for understanding resistance mutations away from the immediate vicinity of the active site, where differences in the two proteins make modeling less reliable.

7. Structure of HIV-1 intasomes and implications

7.1 Structure of HIV-1 intasomes

Fusion of Sso7d to the N-terminus of HIV-1 IN (Sso7d-IN) was a key step that facilitated the first HIV-1 intasome structures. The other strategy that contributed was the assembly of HIV-1 STC intasomes on branched DNA substrate that mimics the product of DNA strand transfer. It had previously been shown that PFV STCs assembled on such a DNA substrate were identical to STCs made through the normal forward reaction pathway (Yin et al. 2012). HIV-1 STCs aggregate less than HIV-1 SSCs and were therefore chosen for the first structural studies of HIV-1 intasomes. The assembly efficiency is relatively low and multiple purification steps were required to obtain material that was suitable for structural studies. After a final step of gel filtration, the intasomes appeared relatively homogeneous as judged by the gel filtration profile, although the peak exhibited a slight hump suggesting the presence of more than one species (Li et al. 2014). Attempts to crystallize these intasomes failed. With hindsight, crystallization was unlikely to work as we know multiple species are present, even after purification.

The STC intasomes required the presence of high ionic strength (0.5 M NaCl) and glycerol in the buffer to prevent aggregation. Such conditions have generally been regarded as unfavorable for cryo-electron microscopy (cryoEM). However, recent advances in cryoEM techniques partially negate these factors and cryoEM was highly successful in determining

the structure of HIV-1 intasomes even with less than optimal buffer conditions (Passos et al. 2017).

The initial studies focused on the smaller species with the erroneous assumption that larger species were simply aggregates of the same basic unit. The density map resolved to 3.5 Å to 4.5 Å, with the highest resolution being in the core of the structure around the active site. The structure was tetrameric and essentially the same as the previously determined PFV intasome structure (Figure 8). The intasome comprised a dimer of dimers. Like the PFV intasome, most the interactions with DNA are mediated by the inner protomer and the N-terminal domains of each inner protomer extend out to the active site region of the other inner protomer. Unlike the PFV intasome, the C-terminal domains of the outer protomers were resolved and contribute to interaction with viral DNA. The N-terminal domains of the outer protomers were disordered as in PFV intasomes. The target DNA is distorted from B form, as is the target DNA in PFV STC intasomes, with five bp between the active sites as expected from the target site duplication. The Sso7d domain that was added to the N-terminus of integrase to improve the solubility of the intasomes was disordered.

In addition to tetrameric STC intasomes, cryoEM revealed higher order species, the best-resolved of which contained 12 integrase protomers. This higher order STC intasome has the same arrangement of domains interacting with DNA as in the tetrameric STC. However, some of these “positionally conserved” domains are contributed by additional subunits in the higher order STC (Figure 9).

7.2 Reinterpreting earlier domain structures with hindsight of the HIV-1 STC intasome structure

The earlier structures of individual domains and two-domain structures of HIV-1 integrase provided few clues as to intasome structure. Although many modeling studies were attempted based on these structures, none of these came close to predicting the functional DNA-integrase structure. This is hardly surprising given the role of protein-DNA interactions in organizing the intasome and none of the earlier structures contained DNA. The Catalytic domain dimer interface was the only interface consistently observed interface in the structures of retroviral integrases in the absence of DNA. The isolated HIV-1 integrase N-terminal domain was a dimer in solution (Cai et al. 1997). In the context of the N-terminal together with the Catalytic domain, although the N-terminal domain dimerized, the interface was entirely different from that observed with the isolated N-terminal domain; additionally, the N-terminal domain dimer was packed against the Catalytic domain (Wang et al. 2001). A similar ambiguity was seen with the C-terminal domain. Whereas the isolated C-terminal domain was a dimer in solution (Lodi et al. 1995), in the structure of the Catalytic together with the C-terminal domain, the C-terminal domain was monomeric and separated from the Catalytic domain by an alpha helical linker (Chen et al. 2000a). To complicate the picture further, in the structure of the Catalytic domain together with the C-terminal domain of simian immunodeficiency virus (SIV) integrase, only one of the C-terminal domains was discernable and it was packed against the Catalytic domain (Chen et al. 2000b). The variety of domain arrangements seen in these structures was generally thought to reflect the flexibility of the linkers between domains and not necessarily be directly relevant to

intasome architecture. It is striking that many of these interfaces seen in the absence of DNA are present in the HIV-1 STC intasome structures. For example, in the higher order STC, positionally conserved domains share the interface observed in the structure of the N-terminal together with Catalytic domain. Strikingly, in the higher-order STC, the linker between the Catalytic and C-terminal domain is helical and extends away from the Catalytic domain as in the structure of the Catalytic together with C-terminal domain in the absence of DNA. Two of the C-terminal domains also dimerize exactly as in the NMR structure of the isolated C-terminal domain. All these structures together highlight the plasticity of integrase in using a common set of interfaces to assemble different structures.

7.3 Comparison with intasome structures of other retroviruses

The PFV SSC was the first intasome structure to be solved for any retrovirus (Hare et al. 2010a). Given the similarity of PFV integrase to other retroviral integrases, it seemed reasonable to assume that other retroviral intasomes would have a similar organization. However, in PFV intasomes the C-terminal domains of the protomers involved in interactions with the viral and target DNA extend away from the Catalytic domain of one inner protomer towards the Catalytic domain of the other inner protomer (Figure 7). This is possible because PFV integrase has a long linker between the Catalytic and C-terminal domain. This represented a conundrum because the Catalytic to C-terminal domain linkers of some retroviral integrases are too short to accommodate the same arrangement of domains as in the PFV tetrameric intasomes (Figure 10). The solution became apparent when the structures of Rous sarcoma virus (RSV) (Yin et al. 2016) and mouse mammary tumor virus (MMTV) intasomes (Ballandras-Colas et al. 2016) were determined. The intasomes had the same arrangement of domains around the viral DNA as PFV intasomes, but the positionally conserved C-terminal domains were contributed by an additional pair of flanking dimers in an octameric intasome. Maedi visna virus (MVV) intasomes were found to be hexadecameric (Ballandras-Colas et al. 2017), but again had the same set of positionally conserved domains interacting with DNA. Structures of both tetrameric and hexadecameric intasomes have been determined for HIV-1. The HIV-1 hexadecamer has weak density in positions of the extra protomers in the MVV hexadecamer; it is therefore possible that this HIV-1 higher-order intasome is a hexadecamer like MVV, with some protomers disordered. Biophysical studies also suggest the presence of additional higher-order intasome species, the structures of which are yet to be determined. At the time of writing, the functional significance of multiple species of HIV-1 intasomes is unclear. However, they all use essentially the same protein-protein and protein-DNA interfaces at the core of the structure, but differ in the protomers that contribute the positionally conserved domains.

7.4 Beyond intasomes

HIV-1 intasomes are sufficient for catalysis of the chemical steps of HIV-1 DNA integration in the absence of other protein factors. Although outside the scope of this chapter, it would be remiss to not mention the important role of other viral and cellular protein factors in the overall DNA integration process. Two of the steps that clearly involve recruitment of such factors are nuclear entry (reviewed in (Matreyek and Engelman 2013)) and targeting of intasomes to specific regions of chromatin (reviewed in (Debyser et al. 2015)). One of the

best understood cellular factors in HIV-1 DNA integration is Lens epithelium–derived growth factor p75 splice variant (LEDGF). LEDGF contains a domain that binds HIV-1 integrase and a domain that binds chromatin. Intasomes are therefore preferentially targeted to regions of chromatin that are enriched for LEDGF, and this is a major mechanism behind the preferential integration of HIV-1 in transcriptionally active chromatin. Roles for many other cellular factors in retroviral DNA integration can be expected to be uncovered.

8. Perspectives and drug discovery

The first clinically approved HIV-1 inhibitors were developed using lead compounds identified by high-throughput screens independent of any structural information on HIV-1 integrase. They were found to target cSSC intasomes rather than free integrase protein. The structure of the PFV cSSC in complex with inhibitors showed that they clearly work by displacing the 3'-OH, that is poised to attack the target DNA, away from the active site. Sequence similarity allows modeling of the HIV-1 the intasome active site based on the PFV intasome structure (Krishnan et al. 2010; Johnson et al. 2013). However, sequence divergence makes modeling away from the immediate vicinity of the active site less reliable. Structures of HIV-1 cSSC intasomes will be required to understand the detailed mechanism of action of HIV-1 integrase inhibitors and how mutations can confer drug resistance.

Acknowledgements

We thank Alan Engelman and Wei Yang for valuable input to the manuscript. This work was supported by the Intramural Program of the National Institute of Diabetes and Digestive Diseases of the National Institutes of Health, and by the Intramural AIDS Targeted Antiviral Program of the Office of the Director of the NIH.

References

- Ariyoshi M, Vassilyev DG, Iwasaki H, Nakamura H, Shinagawa H, Morikawa K (1994) Atomic structure of the RuvC resolvase: a holliday junction-specific endonuclease from *E. coli*. *Cell* 78:1063–1072 [PubMed: 7923356]
- Ballandras-Colas A, Browne M, Cook NJ, Dewdney TG, Domeler B, Cherepanov P, Lyumkis D, Engelman AN (2016) Cryo-EM reveals a novel octameric integrase structure for betaretroviral intasome function. *Nature* 530 (7590):358. doi:10.1038/nature16955 [PubMed: 26887496]
- Ballandras-Colas A, Maskell DP, Serrao E, Locke J, Swuec P, Jonsson SR, Kotecha A, Cook NJ, Pye VE, Taylor IA, Andresdottir V, Engelman AN, Costa A, Cherepanov P (2017) A supramolecular assembly mediates lentiviral DNA integration. *Science* 355 (6320):93–95. doi:10.1126/science.aah7002 [PubMed: 28059770]
- Bowerman B, Brown PO, Bishop JM, Varmus HE (1989) A nucleoprotein complex mediates the integration of retroviral DNA. *Genes Dev* 3:469–478 [PubMed: 2721960]
- Brown PO, Bowerman B, Varmus HE, Bishop JM (1987) Correct integration of retroviral DNA in vitro. *Cell* 49:347–356 [PubMed: 3032450]
- Brown PO, Bowerman B, Varmus HE, Bishop JM (1989) Retroviral integration: structure of the initial covalent product and its precursor, and a role for the viral IN protein. *Proc Natl Acad Sci USA* 86:2525–2529 [PubMed: 2539592]
- Bukrinsky MI, Sharova N, McDonald TL, Pushkarskaya T, Tarpley WG, Stevenson M (1993) Association of integrase, matrix, and reverse transcriptase antigens of human immunodeficiency virus type 1 with viral nucleic acids following acute infection. *Proc Natl Acad Sci USA* 90:6125–6129 [PubMed: 7687060]
- Bushman FD, Fujiwara T, Craigie R (1990) Retroviral DNA integration directed by HIV integration protein in vitro. *Science* 249:1555–1558 [PubMed: 2171144]

- Cai M, Zheng R, Caffrey M, Craigie R, Clore GM, Gronenborn AM (1997) Solution structure of the N-terminal zinc binding domain of HIV-1 integrase. *Nat Struct Biol* 4:567–577 [PubMed: 9228950]
- Chen JCH, Krucinski J, Miercke LJW, Finer-Moore JS, Tang AH, Leavitt AD, Stroud RM (2000a) Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: a model for viral DNA binding. *Proc Natl Acad Sci USA* 97:8233–8238 [PubMed: 10890912]
- Chen ZG, Yan YW, Munshi S, Li Y, Zugay-Murphy J, Xu B, Witmer M, Felock P, Wolfe A, Sardana V, Emini EA, Hazuda D, Kuo LC (2000b) X-ray structure of simian immunodeficiency virus integrase containing the core and C-terminal domain (residues 50–293) - an initial glance of the viral DNA binding platform. *J Mol Biol* 296:521–533 [PubMed: 10669606]
- Chow SA, Vincent KA, Ellison V, Brown PO (1992) Reversal of integration and DNA splicing mediated by integrase of human immunodeficiency virus. *Science* 255:723–726 [PubMed: 1738845]
- Coffin JM, Hughes SH, Varmus HE (1977) *Retroviruses*. Cold Spring Harbor Laboratory Press, New York
- Colicelli J, Goff SP (1985) Mutants and pseudorevertants of Moloney murine leukemia virus with alterations at the integration site. *Cell* 42:573–580 [PubMed: 4028161]
- Craigie R (2012) The molecular biology of HIV integrase. *Future Virol* 7 (7):679–686. doi:10.2217/fvl.12.56 [PubMed: 23024700]
- Craigie R, Fujiwara T, Bushman F (1990) The IN protein of Moloney murine leukemia virus processes the viral DNA ends and accomplishes their integration in vitro. *Cell* 62:829–837 [PubMed: 2167180]
- Debyser Z, Christ F, De Rijck J, Gijsbers R (2015) Host factors for retroviral integration site selection. *Trends in Biochemical Sciences* 40 (2):108–116. doi:10.1016/j.tibs.2014.12.001 [PubMed: 25555456]
- Donehower LA, Varmus HE (1984) A mutant murine leukemia virus with a single missense codon in pol is defective in a function affecting integration. *Proc Natl Acad Sci USA* 81:6461–6465 [PubMed: 6208550]
- Dyda F, Hickman AB, Jenkins TM, Engelman A, Craigie R, Davies DR (1994) Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science* 266:1981–1986 [PubMed: 7801124]
- Eijkelenboom AP, van den Ent FM, Vos A, Doreleijers JF, Hard K, Tullius TD, Plasterk RH, Kaptein R, Boelens R (1997) The solution structure of the amino-terminal HHCC domain of HIV-2 integrase: a three-helix bundle stabilized by zinc. *Current Biology* 7:739–746 [PubMed: 9368756]
- Ellison V, Abrams H, Roe T, Lifson J, Brown P (1990) Human immunodeficiency virus integration in a cell-free system. *J Virol* 64:2711–2715 [PubMed: 2335814]
- Engelman A, Craigie R (1992) Identification of conserved amino acid residues critical for human immunodeficiency virus type 1 integrase function in vitro. *J Virol* 66:6361–6369 [PubMed: 1404595]
- Engelman A, Mizuuchi K, Craigie R (1991) HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. *Cell* 67:1211–1221 [PubMed: 1760846]
- Farnet CM, Haseltine WA (1990) Integration of human immunodeficiency virus type 1 DNA in vitro. *Proc Natl Acad Sci USA* 87:4164–4168 [PubMed: 2349226]
- Farnet CM, Haseltine WA (1991) Determination of viral proteins present in the human immunodeficiency virus type 1 preintegration complex. *J Virol* 65:1910–1915 [PubMed: 2002549]
- Fujiwara T, Craigie R (1989) Integration of mini-retroviral DNA: a cell-free reaction for biochemical analysis of retroviral integration. *Proc Natl Acad Sci USA* 86:3065–3069 [PubMed: 2524066]
- Fujiwara T, Mizuuchi K (1988) Retroviral DNA integration: structure of an integration intermediate. *Cell* 54:497–504 [PubMed: 3401925]
- Grandgenett DP, Vora AC, Schiff RD (1978) A 32,000-dalton nucleic acid-binding protein from avian retrovirus cores possesses DNA endonuclease activity. *Virology* 89:119–132 [PubMed: 210568]
- Hare S, Gupta SS, Valkov E, Engelman A, Cherepanov P (2010a) Retroviral intasome assembly and inhibition of DNA strand transfer. *Nature* 464:232–236 [PubMed: 20118915]

- Hare S, Maertens GN, Cherepanov P (2012) 3'-Processing and strand transfer catalysed by retroviral integrase in crystallo. *EMBO J* 31 (13):3020–3028. doi:10.1038/emboj.2012.118 [PubMed: 22580823]
- Hare S, Shun MC, Gupta SS, Valkov E, Engelman A, Cherepanov P (2009) A novel co-crystal structure affords the design of gain-of-function lentiviral integrase mutants in the presence of modified PSIP1/LEDGF/p75. *Plos Pathogens* 5 (1). doi:e1000259 10.1371/journal.ppat.1000259
- Hare S, Vos AM, Clayton RF, Thuring JW, Cummings MD, Cherepanov P (2010b) Molecular mechanisms of retroviral integrase inhibition and the evolution of viral resistance. *Proc Natl Acad Sci USA* 107 (46):20057–20062. doi:10.1073/pnas.1010246107 [PubMed: 21030679]
- Hindmarsh P, Leis J (1999) Reconstitution of concerted DNA integration with purified components. In: *Advances In Virus Research*, Vol 52, vol 52. *Advances In Virus Research*. pp 397–410 [PubMed: 10384244]
- Jaskolski M, Alexandratos JN, Bujacz G, Wlodawer A (2009) Piecing together the structure of retroviral integrase, an important target in AIDS therapy. *Febs Journal* 276 (11):2926–2946. doi: 10.1111/j.1742-4658.2009.07009.x [PubMed: 19490099]
- Jenkins TM, Hickman AB, Dyda F, Ghirlando R, Davies DR, Craigie R (1995) Catalytic domain of human immunodeficiency virus type 1 integrase: identification of a soluble mutant by systematic replacement of hydrophobic residues. *Proc Natl Acad Sci USA* 92:6057–6061 [PubMed: 7597080]
- Johnson BC, Metifiot M, Ferris A, Pommier Y, Hughes SH (2013) A homology model of HIV-1 integrase and analysis of mutations designed to test the model. *J Mol Biol* 425 (12):2133–2146. doi:10.1016/j.jmb.2013.03.027 [PubMed: 23542006]
- Katz RA, Merkel G, Kulkosky J, Leis J, Skalka AM (1990) The avian retroviral IN protein is both necessary and sufficient for integrative recombination in vitro. *Cell* 63:87–95 [PubMed: 2170022]
- Katzman M, Katz RA, Skalka AM, Leis J (1989) The avian retroviral integration protein cleaves the terminal sequences of linear viral DNA at the in vivo sites of integration. *J Virol* 63:5319–5327 [PubMed: 2555556]
- Krishnan L, Li XA, Naraharisetty HL, Hare S, Cherepanov P, Engelman A (2010) Structure-based modeling of the functional HIV-1 intasome and its inhibition. *Proc Natl Acad Sci USA* 107 (36): 15910–15915. doi:10.1073/pnas.1002346107 [PubMed: 20733078]
- Kulkosky J, Jones KS, Katz RA, Mack JP, Skalka AM (1992) Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. *Mol Cell Biol* 12:2331–2338 [PubMed: 1314954]
- Leavitt AD, Shiue L, Varmus HE (1993) Site-directed mutagenesis of HIV-1 integrase demonstrates differential effects on integrase functions in vitro. *J Biol Chem* 268:2113–2119 [PubMed: 8420982]
- Lee MS, Craigie R (1994) Protection of retroviral DNA from autointegration: involvement of a cellular factor. *Proc Natl Acad Sci USA* 91:9823–9827 [PubMed: 7937898]
- Li L, Olvera JM, Yoder KE, Mitchell RS, Butler SL, Lieber M, Martin SL, Bushman FD (2001) Role of the non-homologous DNA end joining pathway in the early steps of retroviral infection. *EMBO J* 20 (12):3272–3281 [PubMed: 11406603]
- Li M, Craigie R (2005) Processing the viral DNA ends channels the HIV-1 integration reaction to concerted integration. *J Biol Chem* 280:29334–29339 [PubMed: 15958388]
- Li M, Jurado KA, Lin S, Engelman A, Craigie R (2014) Engineered Hyperactive Integrase for Concerted HIV-1 DNA Integration. *Plos One* 9 (8). doi:10.1371/journal.pone.0105078
- Li M, Mizuuchi M, Burke TR, Craigie R (2006) Retroviral DNA integration: reaction pathway and critical intermediates. *EMBO J* 25 (6):1295–1304 [PubMed: 16482214]
- Lodi PJ, Ernst JA, Kuszewski J, Hickman AB, Engelman A, Craigie R, Clore GM, Gronenborn AM (1995) Solution structure of the DNA binding domain of HIV-1 integrase. *Biochemistry* 34:9826–9833 [PubMed: 7632683]
- Maertens GN, Hare S, Cherepanov P (2010) The mechanism of retroviral integration from X-ray structures of its key intermediates. *Nature* 468 (7321):326–329. doi:10.1038/nature09517 [PubMed: 21068843]

- Matreyek KA, Engelman A (2013) Viral and Cellular Requirements for the Nuclear Entry of Retroviral Preintegration Nucleoprotein Complexes. *Viruses-Basel* 5 (10):2483–2511. doi:10.3390/v5102483
- Pandey KK, Bera S, Zahm J, Vora A, Stillmock K, Hazuda D, Grandgenett DP (2007) Inhibition of human immunodeficiency virus type I concerted integration by strand transfer inhibitors which recognize a transient structural intermediate. *J Virol* 81 (22):12189–12199. doi:10.1128/jvi.02863-06 [PubMed: 17804497]
- Panganiban AT, Temin HM (1983) The terminal nucleotides of retrovirus DNA are required for integration but not virus production. *Nature* 306:155–160 [PubMed: 6316141]
- Panganiban AT, Temin HM (1984) The retrovirus pol gene encodes a product required for DNA integration: identification of a retrovirus int locus. *Proc Natl Acad Sci USA* 81:7885–7889 [PubMed: 6083562]
- Passos DO, Li M, Yang RB, Rebensburg SV, Ghirlando R, Jeon Y, Shkriabai N, Kvaratskhelia M, Craigie R, Lyumkis D (2017) Cryo-EM structures and atomic model of the HIV-1 strand transfer complex intasome. *Science* 355 (6320):89–92. doi:10.1126/science.aah5163 [PubMed: 28059769]
- Roth MJ, Schwartzberg PL, Goff SP (1989) Structure of the termini of DNA intermediates in the integration of retroviral DNA: dependence on IN function and terminal DNA sequence. *Cell* 58:47–54 [PubMed: 2546673]
- Schwartzberg P, Colicelli J, Goff SP (1984) Construction and analysis of deletion mutations in the pol gene of Moloney murine leukemia virus: a new viral function required for productive infection. *Cell* 37:1043–1052 [PubMed: 6204767]
- Sherman PA, Fyfe JA (1990) Human immunodeficiency virus integration protein expressed in *Escherichia coli* possesses selective DNA cleaving activity. *Proc Natl Acad Sci USA* 87:5119–5123 [PubMed: 2164223]
- Sinha S, Grandgenett DP (2005) Recombinant human immunodeficiency virus type 1 integrase exhibits a capacity for full-site integration in vitro that is comparable to that of purified preintegration complexes from virus-infected cells. *J Virol* 79 (13):8208–8216 [PubMed: 15956566]
- Sinha S, Pursley MH, Grandgenett DP (2002) Efficient concerted integration by recombinant human immunodeficiency virus type 1 integrase without cellular or viral cofactors. *J Virol* 76 (7):3105–3113 [PubMed: 11884535]
- Valkov E, Gupta SS, Hare S, Helander A, Roversi P, McClure M, Cherepanov P (2009) Functional and structural characterization of the integrase from the prototype foamy virus. *Nucleic Acids Res* 37 (1):243–255. doi:10.1093/nar/gkn938 [PubMed: 19036793]
- van Gent DC, Groeneger AA, Plasterk RH (1992) Mutational analysis of the integrase protein of human immunodeficiency virus type 2. *Proc Natl Acad Sci USA* 89:9598–9602 [PubMed: 1409671]
- Wang JY, Ling H, Yang W, Craigie R (2001) Structure of a two-domain fragment of HIV-1 integrase: implications for domain organization in the intact protein. *EMBO J* 20 (24):7333–7343 [PubMed: 11743009]
- Yang W, Hendrickson WA, Crouch RJ, Satow Y (1990) Structure of ribonuclease H phased at 2 Å resolution by MAD analysis of the selenomethionyl protein. *Science* 249:1398–1405 [PubMed: 2169648]
- Yin Z, Lapkouski M, Yang W, Craigie R (2012) Assembly of prototype foamy virus strand transfer complexes on product DNA bypassing catalysis of integration. *Protein Science* 21 (12):1849–1857 [PubMed: 23011895]
- Yin Z, Shi K, Banerjee S, Pandey KK, Bera S, Grandgenett DP, Aihara H (2016) Crystal structure of the Rous sarcoma virus intasome. *Nature* 530 (7590):362. doi:10.1038/nature16950 [PubMed: 26887497]
- Yoder KE, Bushman FD (2000) Repair of gaps in retroviral DNA integration intermediates. *J Virol* 74 (23):11191–11200. doi:10.1128/jvi.74.23.11191-11200.2000 [PubMed: 11070016]
- Zheng R, Jenkins TM, Craigie R (1996) Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization, and enhances catalytic activity. *Proc Natl Acad Sci USA* 93:13659–13664 [PubMed: 8942990]

Figure 1.

The DNA cutting and joining steps of retroviral DNA integration. The viral DNA made by reverse transcription is linear and blunt ended (Figure 1A). In the first step, 3' end processing, two nucleotides are removed from each 3' end of the viral DNA (Figure 1B). This exposes the terminal CA dinucleotide (colored in red) that is conserved among all retroviruses. In the second step, DNA strand transfer, the 3' hydroxyl groups at the two viral DNA ends attack a pair phosphodiester bonds on each strand of target DNA. In the case of HIV-1, 5 base pairs separate the sites of attack on the two target DNA strands. In the resulting integration intermediate (Figure 1D), the 3' ends of the viral DNA are covalently joined to target DNA, while the 5' ends of the viral DNA and the 3' ends of the target DNA are unjoined. The viral integrase protein catalyzes both 3' end processing and DNA strand transfer. Cellular proteins complete the integration process by removing the overhanging dinucleotide at the 5' ends of the viral DNA, filling in the 5 bp single strand connection, and ligation of the 5' ends of the viral DNA to the 3' ends of the unjoined target DNA strand. This figure is reproduced from Figure 1 of (Craigie 2012).

Figure 2.

HIV-1 integration assay with oligonucleotide DNA substrates. (A) 3' end processing. In the presence of a divalent metal ion, integrase cleaves two nucleotides from the 3' ends of the viral DNA (shown in red) to expose the terminal CA-OH that, in the next step, attacks a phosphodiester bond in the target DNA. (B) DNA strand transfer. In the presence of a divalent metal ion and an oligonucleotide lacking the dinucleotide that is normally removed in the 3' end processing reaction, integrase catalyzes a phosphoryl transfer reaction that covalently joins viral to target DNA (shown in blue). Cleaved oligonucleotides produced by the 3' end processing reaction can also go on to carry out DNA strand transfer. The sites of integration into target DNA are random to a first approximation, although there are sequence preferences. This in vitro assay is a "loosened up" version of the fully authentic reaction in which a pair of viral DNA ends are joined to target DNA on opposite strands. Integrase, which functions as a multimer, is depicted as a circle for simplicity.

Figure 3.

The three domains of HIV-1 integrase. The catalytic core domain is highly conserved in retroviral integrases, retrotransposon integrases and the transposases of many DNA transposons. It includes a triad of conserved acidic residues, the D,D-35-E motif, that are key residues of the active site. The N-terminal domain is conserved among retroviral and retrotransposon integrases and contains a pair of conserved His and Cys residues, the HHCC motif, that coordinate a Zn^{2+} ion. The C-terminal domain is less well conserved than the other two domains.

Figure 4.

Structures of the domains of HIV-1 integrase. (A) The Catalytic domain (pdb: 1ITG) is dimeric with an extensive mainly hydrophobic interface of $1300 \text{ \AA}^2$. The monomers have a fold consisting of a five-stranded β -sheet surrounded by helices, a fold that was first observed in RNAase H, and subsequently in many polynucleotidyl transferases. The active site residues D64 and D116 are shown in red. The third residue of the D,D – 35, E motif, E152 is disordered in this structure. The active sites are located on opposite sides of the nearly spherical dimer. (B) The C-terminal domain (pdb: 1IHV). The C-terminal domain is a dimer in solution with each subunit comprised of a five stranded β -barrel resembling an SH3

domain. (C) The N-terminal domain (pdb: 1WJV). It is a four-helix bundle with Zn^{2+} (red spheres) coordinated to His12, His16, Cys40 and Cys43. The structure is dimeric with a predominantly hydrophobic interface.

Figure 5.

Two-domain structures of HIV-1 integrase. (A) Catalytic together with C-terminal domain (pdb: 1EX4). The Catalytic domain shares the same dimer interface as in the structure of the isolated Catalytic domain. The C-terminal helix of the isolated Catalytic domain structure extends through the linker region to position the C-terminal domains far from the Catalytic domain dimer and the dimer interface seen in the solution structure of the isolated C-terminal domain is not present. The Catalytic domains and linkers are colored in cyan and magenta and the C-terminal domains in green. (B) N-terminal together with the Catalytic domain (pdb: 1K6Y). The absence of density in the linker region makes the assignment of connectivity ambiguous. The connectivity shown at the top is easily accommodated by the length of the missing linker region. The alternative connectivity shown below is possible, but

the linker would be required to adopt an extremely stretched conformation. However, it is worth noting that this connectivity has been observed in the corresponding structure of HIV-2 integrase (Hare et al. 2009). The subunits are colored magenta and cyan. (B)

Figure 6.

Retroviral DNA integration is mediated by a series of intasomes along the reaction pathway. (A) Integrase, together with a pair of viral DNA ends, in the presence of divalent metal ions, forms a Stable Synaptic Complex (SSC), the first intasome on the reaction pathway. (B) Within the SSC, integrase cleaves two nucleotides from each 3' end of the viral DNA to form the Cleaved Stable Synaptic Complex (cSSC). (C) The cSSC then captures a target DNA molecule into which the viral DNA ends are to be integrated to form the Target Capture Complex (TCC). This initial association of the cSSC with target DNA is non-

covalent. (D) The hydroxyl groups at the 3' ends of the viral DNA attack a pair of phosphodiester bonds in the target DNA to accomplish DNA strand transfer and form the Strand Transfer Complex (STC). Completion of integration requires disassembly of the STC, removal of the two overhanging nucleotides from the 5' ends of the viral DNA, filling in of the single strand connections between viral and target DNA by a polymerase, and ligation. These latter steps are carried out by cellular enzymes.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 7. Schematic depiction of the PFV intasome. The PFV intasome is a tetramer in which only the inner two protomers of the tetramer (blue and tan) interact with viral DNA (red). The N-terminal extension domain (NED), N-terminal (NTD), and C-terminal domains of the outer protomers are disordered

Figure 8.

Structure of the tetrameric HIV-1 STC intasome (pdb: 5U1C). Most of the contacts between protein and DNA are made by the inner protomers, colored in magenta and cyan. The C-terminal domains of the outer protomers (colored green and orange) interact with viral DNA while the N-terminal domains of the outer protomers are disordered. The overall architecture is similar to that of PFV intasomes except for the additional contact of the C-terminal domain of the outer protomers with viral DNA. (B) The same view with the DNA removed. (C) As in B, but viewed from the top. The large empty space in the middle that is occupied

by DNA (not shown) highlights the critical role of protein-DNA interactions in holding the intasome together.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 9. Schematic representation of the domain arrangements in the HIV-1 tetrameric and higher order STC intasomes. The domains of the inner protomers are colored light blue and the outer protomers in dark blue. Arrows show the positionally conserved C-terminal domains. In the tetrameric structure they are contributed by the inner protomers, whereas in the higher order structure they are contributed by flanking dimers. Dotted circles represent regions of poorly defined density. This figure is reproduced from Supplementary Figure S16 of (Passos et al. 2017).

Figure 10.

Comparison of the linker lengths in the following retroviral integrases: HIV-1, PFV, Maedi-visna virus (MVV), Rous sarcoma virus (RSV) and mouse mammary tumor virus (MMTV). Whereas the N-terminal to Catalytic domain linkers exhibit little variation in length, the Catalytic to C-terminal domain linkers range from 49 residues in PFV to 6 residues in RSV and MMTV. Relative linker lengths are drawn to scale. The NED of PFV integrase is omitted for clarity.