AN ABSTRACT OF THE DISSERTATION OF

Kogan K. Bao for the degree of Doctor of Philosophy in Biochemistry and Biophysics presented on September 19, 2002.

Title: Kinetic Analysis of Avian Sarcoma Virus Integrase in the Presteady-State.

Abstract approved: Redacted for Privacy

Isaac Wong

Integrase catalyzes insertion of a retroviral genome into the host chromosome. Following reverse transcription, integrase binds specifically to the ends of the duplex retroviral DNA, endonucleolytically cleaves two nucleotides from each 3'-end (the processing activity), and inserts these ends into the host DNA (the joining activity) in a concerted manner. Additionally, it has been observed that integrase can catalyze the removal of inserted viral ends (the disintegration activity) *in vitro*. Presteady-state experiments were performed using synapsed substrates to probe the processing reaction and a disintegration substrate to determine the number of protomers in a functional multimeric complex. In single-turnover studies, a novel "splicing" reaction was observed that revealed complications with accurate quantification of enzymatic activity using the synapsed substrates. The splicing reaction was further used to gain insight into the selection of nucleophiles and electrophiles at the binding site. To reduce the complexity introduced by the integrase-catalyzed splicing reaction, 5'-5' reverse-polarity synapsed substrates were designed that were not susceptible to the splicing reaction and that allowed direct comparison of LTR ends simultaneously bound at the active site. Analysis of the presteady-state assays using these reverse-polarity substrates revealed that the concurrent binding of the biologically relevant U3/U5 combination of viral ends facilitates maximal activity of the processing reaction. A disintegration substrate was used in presteady-state active site titrations to determine a reaction stoichiometry of four integrase protomers per one substrate molecule for the disintegration reaction. A tetrameric active complex was then confirmed using atomic force microscopy to image integrase-DNA complexes during the first catalytic turnover. The observed increase of the tetramer population in the presence of substrate DNA demonstrates that the binding of the disintegration substrate induces assembly of the active tetramer and suggests that tetramer assembly may be an integral and dynamic component of the catalytic pathway.

Kinetic Analysis of Avian Sarcoma Virus Integrase in the Presteady-State

By
Kogan K. Bao

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented September 19, 2002
Commencement June 2003

Doctor of Philosophy dissertation of Kogan K. Bao presented on September 19, 2002.

APPROVED:

Redacted for Privacy

Major Professor, representing Biochemistry and Biophysics

Redacted for Privacy

Head of the Department of Biochemistry and Biophysics

Redacted for Privacy

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Redacted for Privacy

Kogan K. Bao, Author

ACKNOWLEDGMENTS

Prior to my initiation into graduate school, I took note that scientists often referred to their research as being done by "we," or the conclusions drawn as "ours." What was this obsession with the first person plural? In my cynicism, I thought that this reference to oneself as being part of a group was either a mere affectation that was in vogue or a way to diffuse responsibility for ones work. Now, at the end of my graduate education, I have come to realize that all research is a collaborative effort. It is rare to have the liberty to give all the "collaborators" on a project the credit due to them and I am glad to have this forum to give the following people my graditude for their help.

Isaac Wong – whose passion and optimism was the catalyst of this project. His enthusiasm never waned even when I thought I had run dry.

Ann Skalka, Hong Wang, and Dorothy Erie – my co-authors on the manuscripts, with whom it was my honor to work.

Mr. McDonald and Boyd Cook – two gentlemen who dedicated their careers to educating young minds, and who influenced my thinking more than they will ever know.

My family and friends – grouped together because most of you fall into both categories. Your support has contributed as much to the completion of this project as any other element.

CONTRIBUTION OF AUTHORS

Anna Marie Skalka assisted in the editing of all the chapters. Hong Wang and Dorothy Erie assisted in the collection and analysis of the atomic force microscopy data in Chapter 4.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

LIST OF FIGURES

LIST OF FIGURES (continued)

LIST OF TABLES

"When you come to a fork in the road, take it."

—Yogi Berra

Kinetic Analysis of Avian Sarcoma Virus Integrase in the Presteady-State

Chapter 1

A BRIEF OVERVIEW OF ASV INTEGRASE

Numerous diseases including AIDS, leukemia, and a variety of other cancers have retroviral etiology. While the consequences of retroviral infection may be diverse, the insertion of the viral genetic material into host chromosomal DNA is an obligatory replication step for all retroviruses (for a comprehensive review of retroviral biology see refs. [1] and [2]). In addition to "accessory" genes (many retroviruses, including avian sarcoma virus, do not contain any accessory genes), the central body of the retroviral DNA contains the four genes common to all retroviruses, namely *gag*, *env*, *pro*, and *pol*. The *gag* and *env* genes encode internal structural and envelope proteins, respectively. *pro* encodes the viral protease which proteolytically processes some of the viral proteins encoded by the other three genes. Two essential retroviral replication enzymes, reverse transcriptase and integrase, are encoded by the *pol* gene.

After the virus invades the cell and deposits the virion core into the cytoplasm of the host, reverse transcriptase catalyzes the synthesis of a double-stranded linear DNA copy of the viral RNA genome. The strand transfer mechanism of reverse transcriptase [3] generates identical long terminal repeat (LTR) sequences at the two ends of the retroviral DNA. Thus, the product of reverse transcription is organized with the region containing the genes proper flanked by the LTRs (Fig. 1.1). Each LTR consists of three distinct elements: U3, R, and U5. Because the two LTRs are oriented in the same direction within the reverse transcriptase product, the DNA copy of the viral genome is terminated by the U3 sequence at one end and the U5 sequence at the other. Although the sequences of the U3 and U5 elements are not identical, they both terminate at the 3' end with the sequence CAXX, where the CA dinucleotide is highly conserved and the final two nucleotides vary among retroviruses; for avian sarcoma virus (ASV), the dinucleotide is TT.

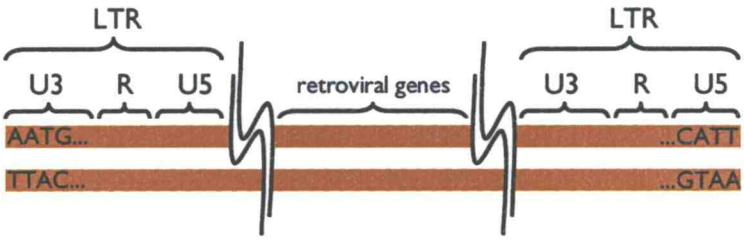

Figure 1.1: Simplified genetic organization of ASV provirus. The retrovial genes *gag*, *pro*, *pol*, and *env* are contained within the body of the viral DNA with the long terminal repeats (LTRs) at each end. Other noncoding sequences important for replication and gene expression (primer-binding site, polypurine tract, etc.) are also located in the main body of the retroviral genome between the two LTRs but are not pictured.

ENZYMATIC ACTIVITIES

Integrase uses the DNA product from reverse transcriptase as the viral substrate for integration in two separate steps that do not require any exogenous energy source, but do require a metal cofactor (either Mn^{2+} or Mg^{2+}). In the first step, integrase specifically recognizes the two blunt-ended U3 and U5 ends and catalyzes the hydrolytic removal of the terminal 3' dinucleotides to expose recessed 3'-hydroxyls [4] (Fig. 1.2, *top*). This site-specific endonucleolytic activity is referred to as the "processing" reaction and occurs while the viral DNA is in the cytoplasm of the infected cell [5]. Integrase is nonspecific in its choice of nucleophiles for the processing reaction, and it has been observed that processing can occur with non-water hydroxyls ranging from glycerol to free and protein-derived alcoholic amino acids (serine and threonine) to the 3'-hydroxyl at the end of the same viral DNA strand (resulting in a 3'-5' cyclic form of the terminal dinucleotide product) [6, 7]. It has been hypothesized that the purpose of the processing reaction is to ensure consistent viral end sequences as reverse transcriptase can incorrectly add extraneous nucleotides to the 3' end of viral DNA [8]. Additionally, the CA that is left exposed is conserved not only among retroviruses but also among retrotransposons and DNA transposable elements in both eukaryotes and prokaryotes [1, 9, 10].

The second step of integration is completed in the nuclear compartment, and involves the covalent joining of the processed viral 3'-ends into staggered positions in the host chromosomal DNA via a concerted cleavage-ligation termed the "joining" reaction (Fig. 1.2, *bottom*). Integrase uses the 3'-hydroxyls exposed during processing of the viral ends as nucleophiles to attack the phosphate backbone of the host target DNA in a transesterification that results in covalent attachment of the viral DNA 3'-ends to the host DNA with unjoined viral 5'-ends, the terminal dinucleotide of which are unpaired, and a gap in the adjacent host DNA. The trimming of the dinucleotide overhang, gap fill-in, and ligation to complete the integration are likely mediated by host DNA repair mechanisms [11, 12]. Integrated viral DNA is termed the provirus and the completion of viral DNA integration produces a short duplication of the host sequence flanking the processed viral DNA that results from the stagger between the sites of joining. The length of this stagger is

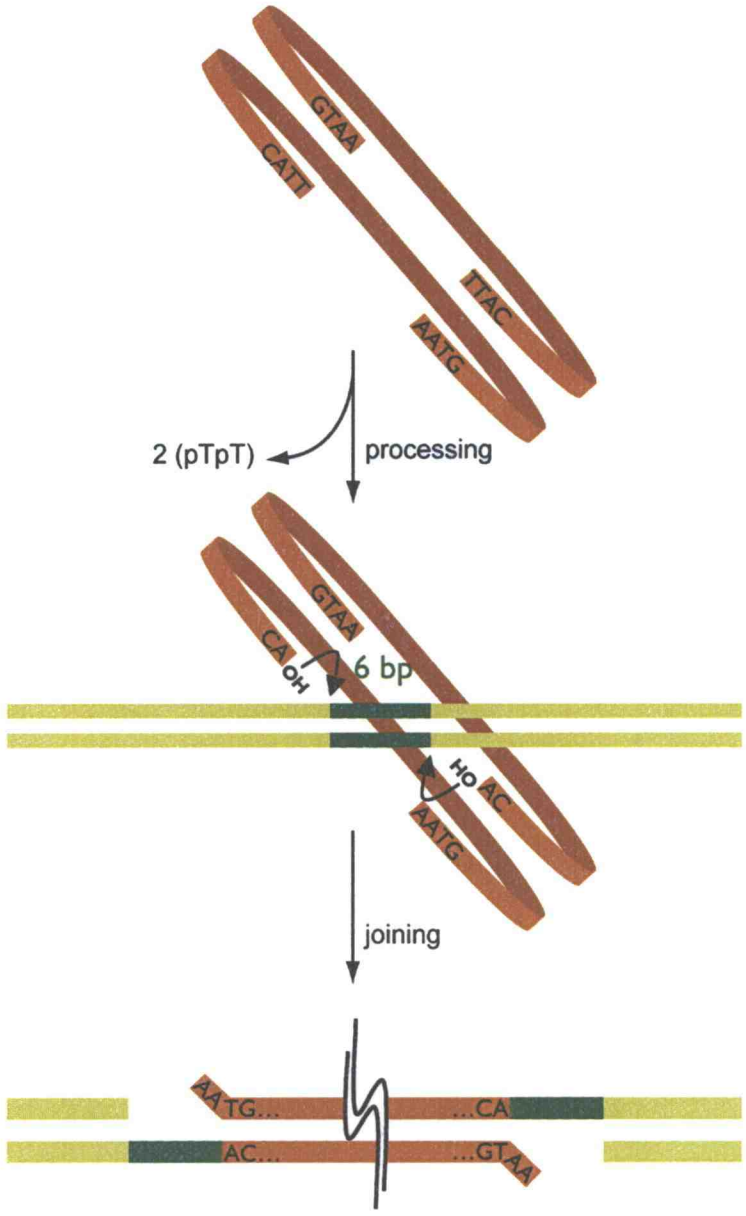

Figure 1.2: *In vivo* activities of ASV integrase. *Top*, the processing activity is the hydrolytic removal of the 3' terminal dinucleotides from the retroviral DNA ends. *Bottom*, the joining activity is the concerted cleavage-ligation whereby the recessed 3'-OHs of the viral DNA (*orange*) are covalently attached to the host DNA (*green*). The result of the subsequent gap fill-in, removal of unpaired nucleotides, and ligation by undefined mechanisms is an integrated provirus flanked by a repeat of host sequences. In the case of ASV integrase, a 6 base pair stagger (*dark green*) along the host DNA separates the sites of joining.

virus-specific and is 6-base pairs for avian sarcoma viral integrase. Although a preference for bent or distorted DNA has been observed *in vitro* [13], the target site on the host DNA for the joining reaction appears to have no significant sequence specificity.

In vitro, integrase also catalyzes a "disintegration" reaction, the apparent reversal of the joining reaction, when presented with a "Y-shaped" substrate which has a structure similar to the joining product from a single processed viral DNA end integrated into target DNA [14]. Integrase-catalyzed disintegration of the Y-shaped substrate reverses the joining to release the separate viral and target DNA components by using the 3'-hydroxyl end of the target fragment, that resides 5' of the nick created in the joining reaction, to regenerate the unmodified target DNA fragment concomitantly releasing the viral DNA. It is difficult to accurately quantify enzymatic activity in assays that probe the forward/joining activity because the products are widely distributed in sizes due to the lack of sequence-specificity in target site selection; the Y-shaped substrate, therefore, has the advantage of having a fixed site of joining thereby circumventing this problem. The sequence and structural requirements for disintegration are less stringent than those for the processing and joining reactions, thus allowing the disintegration reaction to be used as a probe for enzymatic activity of truncated and mutated variants of the enzyme.

STRUCTURE

The full-length ASV integrase contains 286 amino acids, and HIV−1 integrase 288 amino acids. The integrase monomer has three discrete domains, as identified by functional studies [15] and the susceptibility of the linker regions to proteolysis [16]: the N-terminal domain (residues 1 to ~50), the catalytic core domain (residues ~50−210), and the C-terminal domain (residues ~210−270). The structures of each isolated domain as well as 2-domain fragments have been solved using either crystallographic or NMR methods, but the structure of the full-length enzyme remains unknown due to the low solubility of the enzyme under reaction conditions [17, 18]. Furthermore, the spatial relationship between the three domains in the active complex with DNA substrate is unknown due to flexibility

in the linkage between the domains and the absence of a solved structure of a substrate-integrase cocrystal. The latter precludes a crystallographic determination of the number of integrase protomers that are organized in a complete, catalytically active unit.

The N-terminal domain consists of a bundle of α-helices stabilized by a zinc-binding site that is coordinated by conserved pairs of histidine and cysteine residues, termed the HHCC region. The HHCC region is similar to the zinc-coordinating residues of the canonical zinc-binding motif found in some DNA-binding proteins. However, the histidine and cysteine pairs are reversed in orientation relative to the metal-coordinating residues in these zinc-binding motifs [11]. Though the isolated N-terminal domain has been shown to bind zinc [19], the function of the N-terminal domain remains unclear though cross-linking studies have suggested that it may be involved in the binding of DNA [20]. Mutation and deletion studies of the N-terminal domain have concluded that even if it does bind DNA, the HHCC region is not the sole DNA-binding domain nor is it required for catalytic activity [21–23]. Rather, it has been proposed that the HHCC region might be involved in the recognition of viral terminal sequences that is necessary to promote efficient processing and that there are separate viral and target DNA binding sites on the enzyme. The structure of a 2-domain fragment of HIV-1 integrase consisting of the N-terminal domain and the catalytic core has been solved by x-ray crystallography [24], and the N-terminal of this fragment is structurally very similar to that of the isolated domain as determined in solution by NMR [25, 26], although the dimer interface is very different between the two solved structures.

The catalytic core domain encodes the active site residues and contains an invariant metal-binding D,D-35-E motif that, for ASV integrase, is comprised of the acidic residues Asp^{64}, Asp^{121}, and Glu^{157} [20, 27–29]. Similar motifs are also found in retrotransposons and the transposase of some bacterial insertion sequence elements [10, 27, 30]. Mutation of any of these residues severely reduces catalytic activity, and by analogy with models of catalysis by HIV-1 RNase H and the exonuclease domain of DNA polymerase I, it is hypothesized that coordination of divalent metal ions to these residues plays a key role in catalysis [27]. Because mutation of these residues had parallel effects on the processing and joining

activities of the enzyme, it has been further proposed that there is likely a single catalytic site for both reactions. The crystal structures of the isolated catalytic core domains of ASV [29] and HIV-1 [31] have been solved. The residues comprising the active site region exhibit considerable flexibility, suggesting that the binding of substrate is required to impose the precise configuration of residues that is required for catalysis. In addition to the N-terminal plus catalytic core 2-domain fragment discussed above, the structure of an additional 2-domain fragment containing the C-terminal and catalytic core domains has also been solved by x-ray crystallography [32]. It exhibits considerable flexibility in the linkage between the C-terminal and catalytic core domains.

The C-terminal domain, as solved by NMR for HIV−1 integrase[26, 33], consists of a five-stranded β-barrel resembling an Src homology 3 (SH3) domain, although there is no known functional relationship with the SH3 domains of other proteins. The residues of the C-terminal domain are less conserved relative to the other two domains, and its function is less clear. Though it is not required for the disintegration activity, the C-terminal domain is essential for the processing and joining activities [34−36]. In addition, cross-linking studies suggest that the C-terminal domain interacts with a subterminal region of the viral DNA [37−40], although the C-terminal domain has been observed to bind DNA nonspecifically in deletion analysis and independent expression experiments [21, 36, 41, 42]. Furthermore, it has been observed that the nuclear localization signal of ASV integrase resides at the proximal end of the C-terminal domain [43] (though it is located in both the C-terminus and core domains in HIV-1 [44]).

ASSEMBLY STATE

Steady-state studies have provided evidence indicating that integrase functions as a multimer and must minimally be a dimer for processing and joining [45]. Complementation studies [15, 46], sedimentation analyses [45], crosslinking experiments [39, 47], and time-resolved fluorescence anisotropy studies [48] further support that integrase functions as a multimer. Although the structure of the intact 32-kDa integrase

protein is not known, all truncated forms of the enzyme whose structures have been solved have been observed as dimers. However, the active sites of the protomers in these dimeric structures are located on opposite faces of the crystallographic dimer and separated by >50 Å, a distance that seems too large to be spanned by the six base pairs that defines the spacing between the two sites of concerted integration within the host DNA. As a result, the dimeric C-terminal plus catalytic core 2-domain crystal structure has been used by the crystallographers to model a tetrameric functional integration complex, consisting of a dimer of dimers, to accommodate crystallographic contacts and the geometric constraints imposed by the concerted mechanism of retroviral integration [32]. Tn5 transposase shares a similar phosphatidyl transfer mechanism to that of integrase, and the crystal structure of Tn5 transposase in complex with its DNA substrate [49] has also been used along with the N-terminal plus catalytic core 2-domain crystal structure of integrase to model the interaction of a dimer-of-dimers integrase complex with substrate DNA [24].

The two tetrameric integrase models are similar in that two of the four active sites (one from each dimer) do not participate in catalysis, but likely play a supporting structural role. Neither model, however, addresses two related structural details concerning the catalytic mechanism of coordinated cleavage at the active sites: 1) the manner in which the LTR ends, which are bound as electrophiles in the hydrolytic processing reaction, come to reside in the active site as nucleophiles in the joining reaction (in which the host/target DNA serves as the electrophile), and 2) how integrase accommodates the interchange of the electrophile in the reactions being bound in a sequence-specific manner in the processing reaction (i.e. the viral DNA ends) and in a sequence-nonspecific manner in the joining reaction (i.e. the host DNA).

RESEARCH SUMMARY

In vitro integrase processing assays use radiolabeled substrates with sequences derived from the viral ends in reaction(s) typically lasting 30-90 minutes as integrase shows low reactivity in such assays [50−52]. Single-end substrate assays also fail to produce

measurable amounts of concerted integration products, separated by the 6 basepair stagger, as is observed with preintegration complexes purified from infected cells [53, 54]. In an attempt to mimic the geometric organization of the viral LTR ends of the *in vivo* preintegration complex at a more molecular level, Kukolj and Skalka [55] designed a series of substrates that covalently linked two single-end substrates together in a head-to-head configuration using 1-3 nucleotides of single-stranded DNA. It was hypothesized that these single-stranded tethers would provide sufficient flexibility to alleviate torsional or rotational strains arising from the structural alignment of the two viral ends bound within the integrase active site(s). These authors observed enhanced processing efficiency with these synapsed-end substrates *in vitro* and concluded that the tether effectively brought together integrase subunits bound separately to the two cognate sites, thereby coordinating the formation of a requisite higher-order oligomeric structure with enhanced activity. These observations were consistent with the suggestion by Murphy and Goff [56] that integrase must recognize both DNA ends for efficient processing at either end to occur *in vivo*.

While the results with these synapsed substrates clearly illustrated the important relationship between assembly of an integrase multimer and the coordinated binding of both viral DNA ends, the assays were performed in the time regime where the enzyme-catalyzed reaction had undergone multiple turnovers. Although much effort has been expended to demonstrate that integrase functions as a true enzyme in being able to catalyze multiple turnovers under steady-state conditions (24), the physiological relevance of multiple turnover events is questionable considering that only a single round of catalysis is sufficient to achieve integration *in vivo*.

Described here is a single-turnover investigation of integrase using synapsed substrates to probe the processing reaction and a Y-shaped substrate to determine the number of protomers in a functional multimeric complex. Using substrates similar to those of Kukolj and Skalka, a novel "splicing" reaction was observed that revealed complications with accurate quantification of enzymatic activity using these substrates. The splicing reaction was further used to gain insight into the selection of nucleophiles and electrophiles at the binding site and provided the basis for a model of the active site that resolves the

nucleophile/electrophile/specific/nonspecific quandary previously posed. To reduce the complexity introduced by the integrase-catalyzed splicing reaction, 5'–5' reverse-polarity synapsed substrates were designed that were not susceptible to the splicing reaction and that allowed direct comparison of LTR ends simultaneously bound at the active site. Analysis of the presteady-state assays using these reverse-polarity substrates revealed that the concurrent binding of the biologically relevant U3/U5 combination of viral ends facilitates maximal activity of the processing reaction.

It was not possible to determine the reaction stoichiometry by active site titrations using the synapsed substrate because of solubility and aggregation difficulties with integrase in the presence of these substrates. It was, however, possible to use the Y-shaped substrate in presteady-state active site titrations to determine a reaction stoichiometry of four integrase protomers per one substrate molecule for the disintegration reaction. A tetrameric active complex was then confirmed using atomic force microscopy to image integrase-DNA complexes during the first catalytic turnover. The observed increase of the tetramer population in the presence of substrate DNA, in the AFM studies, demonstrates that binding of the Y-shaped substrate induces assembly of the active tetramer and suggests that tetramer assembly may be an integral and dynamic component of the catalytic pathway.

REFERENCES

1. Coffin, J., Hughes, S.H., Varmus, H.E., ed. *Retroviruses.* 1997, Cold Springs Harbor Laboratory Press: New York. 843.

2. Flint, S.J., Enquist, L.W., Krug, R.M., Racaniello, V.R., Skalka, A.M., *Principles of Virology.* 2000, Washington, D.C.: ASM Press. 804.

3. Peliska, J.A. and S.J. Benkovic, *Mechanism of DNA strand transfer reactions catalyzed by HIV-1 reverse transcriptase.* Science, 1992. **258**(5085): p. 1112-8.

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

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

6. Vink, C., et al., *Site-specific hydrolysis and alcoholysis of human immunodeficiency virus DNA termini mediated by the viral integrase protein.* Nucleic Acids Res, 1991. **19**(24): p. 6691-8.

7. Engelman, A., K. Mizuuchi, and R. Craigie, *HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer.* Cell, 1991. **67**(6): p. 1211-21.

8. Miller, M.D., C.M. Farnet, and F.D. Bushman, *Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition.* J Virol, 1997. **71**(7): p. 5382-90.

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

10. Fayet, O., et al., *Functional similarities between retroviruses and the IS3 family of bacterial insertion sequences?* Mol Microbiol, 1990. **4**(10): p. 1771-7.

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

12. Daniel, R., R.A. Katz, and A.M. Skalka, *A role for DNA–PK in retroviral DNA integration.* Science, 1999. **284**(5414): p. 644–7.

13. Muller, H.P. and H.E. Varmus, *DNA bending creates favored sites for retroviral integration: an explanation for preferred insertion sites in nucleosomes.* Embo J, 1994. **13**(19): p. 4704–14.

14. Chow, S.A., et al., *Reversal of integration and DNA splicing mediated by integrase of human immunodeficiency virus.* Science, 1992. **255**(5045): p. 723–6.

15. van Gent, D.C., et al., *Complementation between HIV integrase proteins mutated in different domains.* Embo J, 1993. **12**(8): p. 3261–7.

16. Engelman, A. and R. Craigie, *Identification of conserved amino acid residues critical for human immunodeficiency virus type 1 integrase function in vitro.* J Virol, 1992. **66**(11): p. 6361–9.

17. Asante–Appiah, E. and A.M. Skalka, *HIV–1 integrase: structural organization, conformational changes, and catalysis.* Adv Virus Res, 1999. **52**: p. 351–69.

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

19. Burke, C.J., et al., *Structural implications of spectroscopic characterization of a putative zinc finger peptide from HIV–1 integrase.* J Biol Chem, 1992. **267**(14): p. 9639–44.

20. Bujacz, G., et al., *Binding of different divalent cations to the active site of avian sarcoma virus integrase and their effects on enzymatic activity.* J Biol Chem, 1997. **272**(29): p. 18161–8.

21. Khan, E., et al., *Retroviral integrase domains: DNA binding and the recognition of LTR sequences [published erratum appears in Nucleic Acids Res 1991 Mar 25; 19(6):1358].* Nucleic Acids Res, 1991. **19**(4): p. 851–60.

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

23. Katz, R.A., G. Merkel, and A.M. Skalka, *Targeting of retroviral integrase by fusion to a heterologous DNA binding domain: in vitro activities and incorporation of a fusion protein into viral particles.* Virology, 1996. **217**(1): p. 178–90.

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

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

26. Eijkelenboom, A.P., et al., *The DNA–binding domain of HIV–1 integrase has an SH3–like fold.* Nat Struct Biol, 1995. **2**(9): p. 807–10.

27. Kulkosky, J., et al., *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, 1992. **12**(5): p. 2331–8.

28. Bujacz, G., et al., *The catalytic domain of avian sarcoma virus integrase: conformation of the active–site residues in the presence of divalent cations.* Structure, 1996. **4**(1): p. 89–96.

29. Bujacz, G., et al., *High–resolution structure of the catalytic domain of avian sarcoma virus integrase.* J Mol Biol, 1995. **253**(2): p. 333–46.

30. Khan, E., et al., *Retroviral integrase domains: DNA binding and the recognition of LTR sequences.* Nucleic Acids Res, 1991. **19**(4): p. 851–60.

31. Dyda, F., et al., *Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases.* Science, 1994. **266**(5193): p. 1981-6.

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

33. Lodi, P.J., et al., *Solution structure of the DNA binding domain of HIV-1 integrase.* Biochemistry, 1995. **34**(31): p. 9826-33.

34. Kulkosky, J., et al., *Activities and substrate specificity of the evolutionarily conserved central domain of retroviral integrase.* Virology, 1995. **206**(1): p. 448-56.

35. Bushman, F.D. and B. Wang, *Rous sarcoma virus integrase protein: mapping functions for catalysis and substrate binding.* J Virol, 1994. **68**(4): p. 2215-23.

36. Vink, C., A.M. Oude Groeneger, and R.H. Plasterk, *Identification of the catalytic and DNA-binding region of the human immunodeficiency virus type I integrase protein.* Nucleic Acids Res, 1993. **21**(6): p. 1419-25.

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

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

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

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

41. Mumm, S.R. and D.P. Grandgenett, *Defining nucleic acid-binding properties of avian retrovirus integrase by deletion analysis.* J Virol, 1991. **65**(3): p. 1160-7.

42. Woerner, A.M. and C.J. Marcus-Sekura, *Characterization of a DNA binding domain in the C-terminus of HIV-1 integrase by deletion mutagenesis.* Nucleic Acids Res, 1993. **21**(15): p. 3507-11.

43. Kukolj, G., K.S. Jones, and A.M. Skalka, *Subcellular localization of avian sarcoma virus and human immunodeficiency virus type 1 integrases.* J Virol, 1997. **71**(1): p. 843-7.

44. Gallay, P., et al., *HIV nuclear import is governed by the phosphotyrosine-mediated binding of matrix to the core domain of integrase.* Cell, 1995. **83**(4): p. 569-76.

45. Jones, K.S., et al., *Retroviral integrase functions as a multimer and can turn over catalytically.* J Biol Chem, 1992. **267**(23): p. 16037-40.

46. Engelman, A., F.D. Bushman, and R. Craigie, *Identification of discrete functional domains of HIV-1 integrase and their organization within an active multimeric complex.* Embo J, 1993. **12**(8): p. 3269-75.

47. Andrake, M.D. and A.M. Skalka, *Multimerization determinants reside in both the catalytic core and C terminus of avian sarcoma virus integrase.* J Biol Chem, 1995. **270**(49): p. 29299-306.

48. Deprez, E., et al., *Oligomeric states of the HIV-1 integrase as measured by time-resolved fluorescence anisotropy.* Biochemistry, 2000. **39**(31): p. 9275-84.

49. Davies, D.R., et al., *Three-dimensional structure of the Tn5 synaptic complex transposition intermediate.* Science, 2000. **289**(5476): p. 77-85.

50. Chow, S.A., *In vitro assays for activities of retroviral integrase.* Methods, 1997. **12**(4): p. 306-17.

51. Scottoline, B.P., et al., *Disruption of the terminal base pairs of retroviral DNA during integration.* Genes Dev, 1997. **11**(3): p. 371–82.

52. Pemberton, I.K., M. Buckle, and H. Buc, *The metal ion–induced cooperative binding of HIV–1 integrase to DNA exhibits a marked preference for Mn(II) rather than Mg(II).* J Biol Chem, 1996. **271**(3): p. 1498–506.

53. Craigie, R., T. Fujiwara, and F. Bushman, *The IN protein of Moloney murine leukemia virus processes the viral DNA ends and accomplishes their integration in vitro.* Cell, 1990. **62**(4): p. 829–37.

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

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

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

Chapter 2

A SPLICING ACTIVITY AND STRUCTURE FUNCTION IMPLICATIONS FOR COGNATE SITE RECOGNITION

Kogan K. Bao, Anna Marie Skalka, and Isaac Wong

Published in *Journal of Biological Chemistry*

9650 Rockville Pike

Bethesda, Maryland 20814-3996

Volume 277, Number 14, Issue of April 5, 2002, pp. 12089–12098

SUMMARY

Integrase catalyzes insertion of a retroviral genome into the host chromosome. After reverse transcription, integrase binds specifically to the ends of the duplex retroviral DNA, endonucleolytically cleaves two nucleotides from each 3'-end (the processing activity), and inserts these ends into the host DNA (the joining activity) in a concerted manner. In first-turnover experiments with synapsed DNA substrates, we observed a novel splicing activity that resembles an integrase joining reaction but uses unprocessed ends. This splicing reaction showed an initial exponential phase ($k_{splicing}$ = 0.02 s^{-1}) of product formation and generated products macroscopically indistinguishable from those created by the processing and joining activities, thus bringing into question methods previously used to quantitate these reactions in a time regime where multiple turnovers of the enzyme have occurred. With a presteady-state assay, however, we were able to distinguish between different pathways that led to formation of identical products. Furthermore, the splicing reaction allowed characterization of substrate binding and specificity. Although integrase requires only a 3'-OH with respect to nucleophiles derived from DNA, it specifically favors the cognate sequence CATT as the electrophile. These experimental results support a two-site "switching" model for binding and catalysis of all three integrase activities.

INTRODUCTION

After infection, retroviruses create a linear DNA copy of their RNA genome that, through the strand-transfer mechanism of reverse transcriptase, places the U3 region of the long terminal repeat (LTR) sequence at one terminus and the U5 region of the LTR sequence at the other terminus [1]. Retroviral replication is dependent on the viral protein integrase catalyzing the recombination of the viral DNA genome into the host genomic DNA. Integrase binds to the two blunt-ended viral LTRs, hydrolyzes the terminal two nucleotides to expose a recessed 3'-OH of the conserved CA dinucleotide at each of the ends (the processing activity), and inserts these "processed" ends into the host DNA (the joining activity) at sites separated by a virus-specific stagger of six base pairs for avian sarcoma virus (ASV). The location of the insertion is nearly random as there is little

sequence specificity for the site of recombination within the host genome [2–4]. The processing and joining activities are biochemically similar in that both use a hydroxyl group as the nucleophile in an endonucleolytic cleavage. In the case of the processing activity (3'-dinucleotide removal), the enzyme is specific in its choice of electrophile (the cognate CATT), whereas the nucleophile (the processed CA-OH) is specified in the joining reaction (strand transfer). Both the ends-processing and joining activities have been reproduced *in vitro* with purified recombinant integrase and oligodeoxyribonucleotides whose sequences are derived from the retroviral U3 and U5 LTR sequences [5–7]. Detailed examination of the processing activity *in vitro* has revealed that integrase requires the physiologically relevant configuration of both U3 and U5 ends to be bound concurrently for maximal efficiency of processing catalysis, with the U3 sequence undergoing both the cleavage and recombination reactions earlier than the U5 sequence [8].

Structural investigations of integrase suggest that the enzyme possesses three structural domains, 1) an N-terminal domain characterized by a zinc-stabilized helix-turn-helix, 2) a central core domain with a D,D-35-E motif, and 3) a C-terminal domain with structure resembling a Src homology 3 (SH-3) domain (for review, see Refs. [9–11]). Although there is much evidence that active integrase functions as a multimer [12–15], it is still undetermined how the three domains interact with each other within a single integrase monomer or a multimer integrase-DNA complex. Wang *et al.* [16] have recently reported the structural solution of a two-domain fragment of human immunodeficiency virus-1 (HIV-1)integrase that suggests a dimer of dimers resembling Tn5 transposase. However, the actual oligomeric state and geometric arrangement of integrase protomers within the active DNA-protein complex remain unconfirmed, and the detailed catalytic mechanism of the coordinated cleavage of four different DNA segments, resulting in the concerted insertion of the two LTR ends, remains unsolved. Two general models of the organization of the multimeric complex necessary to catalyze the insertion of the two ends of DNA have been proposed [17–19]. In one model, separate integrase molecules bind each of the two processed viral DNA ends with two separate integrase molecules binding host DNA to prepare the two host phosphates for attack. In the other model, each

integrase protomer contains separate binding sites for host and viral DNA, and a single active site catalyzes both processing and joining reactions, thus requiring the active integrase complex to only be a dimer. An unanswered question with both of the models, however, is the manner in which the LTR ends, which serve as electrophiles in the processing reaction, come to reside in the active site as nucleophiles in the joining reaction (where the host DNA is the electrophile). This question becomes increasingly complex when one considers the possible combinations of nucleophiles recognized by integrase [20] and the fact that there must be specific and nonspecific DNA binding sites.

In the course of undertaking a presteady-state investigation of the processing reaction using synapsed-end substrates, we developed an assay that features a preincubation step to form DNA-protein complexes before the initiation of catalytic activity. Along with analysis of product formation during the first enzymatic turnover, the presteady-state assay enabled direct comparison of the reactivity of substrates while minimizing the complications involved in the quantitation of enzymatic activity. Using this assay for the processing reaction with synapsed-end substrates modeled after those of Kukolj and Skalka [5], we report here the discovery of a novel splicing reaction with products nearly indistinguishable from those of the processing and joining reactions. The identification of the splicing reaction resolved complications with the accurate quantitation of enzymatic activity. Additionally, the splicing reaction was used as a tool to gain significant insight into the structure-function relationship in the mechanism of sequence recognition by integrase. Specifically, the splicing reaction allowed us to probe the selection of nucleophiles and electrophiles by the structurally-defined, sequence-specific binding sites of integrase. The results from these experiments allow us to propose a model for the configuration of the nucleophile and electrophile within the enzyme active site that satisfies the varying specificity requirements of the three integrase-catalyzed reactions.

For clarity, the remainder of this report will refer to the 3'-end dinucleotide TT-trimming activity as the "processing" reaction, the subsequent sequence-dependent insertion of processed ends into a double-stranded DNA target as the "joining" reaction, and this novel activity with synapsed substrates as the "splicing" reaction.

EXPERIMENTAL PROCEDURES

Reagents and Buffers

Except where noted, all buffers were made with reagent grade chemicals and Milli-Q Plus (Millipore, Bedford, MA) purified distilled-deionized water. Urea, SDS, and dithiothreitol were Ultrapure grade obtained from U. S. Biochemical Corp. Kanamycin sulfate was obtained from Amresco (Solon, OH). EDTA, HEPES, and ammonium sulfate were SigmaUltra grade obtained from Sigma (St. Louis, MO). Spectroscopic grade glycerol was from Aldrich (Milwaukee, WI). Isopropyl β-D-thiogalactopyranoside was from Fermentas (Hanover, MD). Buffer A is 50 mM Tris-HCl, pH 8.0, 10% sucrose (w/v). Buffer B is 50 mM Tris-HCl, pH 7.5, 10% glycerol (v/v). When NaCl is added to either buffers A or B, the resulting buffer is denoted with an A or B followed by the NaCl concentration in mM. Storage buffer is 50 mM Na-HEPES, pH 7.5, 500 mM NaCl, 40% spectroscopic grade glycerol. 5% polyethyleneimine (PEI) was made from 50% (w/v) PEI (Sigma), and the pH was adjusted to 7.5 with concentrated HCl. All buffer stock solutions were filtered through a 0.2-μm polyethersulfone filter (Nalgene, Rochester, NY).

Enzymes and Proteins

Restriction enzymes and T4 DNA ligase were obtained either from Fermentas (Hanover, MD) or Invitrogen. T4 polynucleotide kinase and acetylated bovine serum albumin were obtained from U. S. Biochemical Corp (Cleveland, OH). Lysozyme was obtained from Sigma. ASV integrase was purified as described under "Purification of Integrase."

Synthetic Oligodeoxyribonucleotides

Oligodeoxyribonucleotides were synthesized by the Center for Gene Research and Biotechnology Central Services Laboratory (Oregon State University) and purified by denaturing PAGE (20 or 13% acrylamide, 8 M urea in TBE) as previously described [21]. Reversed-polarity oligodeoxyribonucleotides were synthesized using 5'-β-cyanoethyl phosphoramidites (Glen Research, Sterling, VA). Concentrations were determined spectrophotometrically in Tris-EDTA using the calculated extinction coefficients at 260 nm [22] listed in Table 2.1. Radiolabeled oligodeoxyribonucleotides were 5'-end-labeled

except where otherwise specified and are designated by an asterisk (*). 5'-Radiolabeled oligodeoxyribonucleotides were prepared using 3 units of T4 polynucleotide kinase, 70 μCi of [γ-^{32}P]ATP (Amersham Biosciences, Arlington Heights, IL), and 50 pmol of DNA in 10 μl using the manufacturer's reaction buffer. After incubation at 37 °C for 15 min, reactions were quenched by the addition of 10 μl of 500 mM EDTA. Complementary strands of DNA were then added at equimolar amounts and annealed by heating to 100 °C followed by slow cooling to room temperature over a period of 15−20 min. Kinase was removed by extraction with phenol:chloroform:isoamyl alcohol (25:24:1) followed by a chloroform-only extraction. Unincorporated nucleotides and residual organic material were removed by purification through a Bio-Spin-6 micro-column (Bio-Rad, Hercules, CA). Final yield and purity of radiolabeled substrates were determined by thin-layer liquid chromatography on PEI cellulose plates developed in a mobile phase of 0.75 M LiCl, 1.25 M formic acid, 40% ethanol. Under these conditions, radiolabeled DNA remains at the origin, whereas unincorporated nucleotides are eluted by the mobile phase. Accurate quantitation of DNA yield is achieved by comparing the amount of radioactivity bound at the origin for samples obtained before and after purification steps. Typically, 87−90% of the DNA is recovered with complete removal of unincorporated nucleotides. In addition, comparison of the radioactivity in unincorporated label and DNA in the pre-purification sample shows that the efficiency of the labeling reaction exceeds 90% under these conditions.

The naming convention used for annealed DNA substrates is as follows. 1) Strands with sequences derived from the U5 and U3 ends of the ASV genome are designated with a "5" and "3," respectively, 2) strands of duplex DNA containing ASV integrase cognate sequence, CATT, are designated with a "t," 3) strands containing the complementary GTAA sequence are designated with a "b," 4) synapsed strands are designated with the length of the tether within parentheses, 6) duplex names consist of a list of the names of all oligodeoxyribonucleotide strands annealed separated by slashes (/), and 7) sequences labeled with ^{32}P will be represented in the text with an asterisk (*) at the beginning or end to denote 5'- or 3'-end radiolabeling, respectively.

Table 2.1: Nomenclature and μM Extinction Coefficients (μM^{-1} cm^{-1}) at 260 nm of DNA Substrates

Name[a]	Sequence	ε_{260}
5t	5′-GCTGAAGCAGAAGGCTTCATT-3′	0.20
5t-2	5′-GCTGAAGCAGAAGGCTTCA-3′	0.19
5b	5′-AATGAAGCCTTCTGCTTCAGC-3′	0.19
3t	5′-GCTATTGCATAAGACTACATT-3′	0.21
3b	5′-AATGTAGTCTTATGCAATAGC-3′	0.21
m5t	5′-GCTGAAGCAGAAGGCTTGCAA-3′	0.21
m3b	5′-TTGCTAGTCTTATGCAATAGC-3′	0.20
5b(2)3t	5′-GCTATTGCATAAGACTACATTTAAATGAAGCCTTCTGCTTCAGC-3′	0.43
5b(2)3b	3′-CGATAACGTATTCTGATGTAAT-5′-5′-AAATGAAGCCTTCTGCTTCAGC-3′	0.43
m5b(2)3t	5′-GCTATTGCATAAGACTACATTTATTGCAAGCCTTCTGCTTCAGC-3′	0.41
m5b(2)3t	5′-GCTATTGCATAAGACTAGCAATAAATGAAGCCTTCTGCTTCAGC-3′	0.43

[a]The following is the naming convention used for annealed DNA substrates. Strands with sequences derived from the U5 and U3 ends of the ASV genome are designated with a "5" and "3," respectively. Strands containing the cognate sequence, CATT, are designated with a "t." Strands containing the complementary GTAA sequence are designated with a "b." Synapsed strands are designated with the length of the tether within parentheses. Duplex names consist of a concatenation of the names of all oligodeoxyribonucleotide strands annealed separated by slashes (/). Sequences labeled with ^{32}P will be represented in the text with an asterisk (*) to denote 5'-end radiolabeling.

Bacterial Strains and Plasmid DNA

Plasmid pRC23(IN) in *Escherichia coli* MC1061 has been described [23]. The translated sequence of this construct differs from the sequence recorded for the Schmidt-Ruppin B strain in the GenBank (accession number AF052428). These differences have been attributed to strain variance (R. Katz, personal communication). Because of the possible importance of one of these differences, a glutamic acid at position 256 was altered to match the lysine of the reported sequence by PCR site-directed mutagenesis. The altered gene was subsequently subcloned into pET24a (Novagen, Madison, WI) to produce the recombinant plasmid, pET24a(IN). Overproduction of integrase from this construct was obtained in *E. coli* BL21(DE3) (Novagen) by using isopropyl β-D-thiogalactopyranoside induction at 25 °C. Although recombinant integrase from either variant is biologically competent (R. Katz, personal communication), in pre-steady state *in vitro* assays, integrase with Lys-256 produced 2-fold greater processing and splicing burst amplitudes (data not shown).

Protein Expression

E. coli BL21(DE3)/pET24a(IN) was grown in Luria-Bertani broth with 30 μg/ml kanamycin, 20 mM HEPES, pH 7.5, and 6 drops/liter Sigma antifoam B at 25 °C to A_{600} ≅ 0.8 (~4 × 108 cells/ml) and induced with 1 mM isopropyl β-D-thiogalactopyranoside, 0.5 mM $ZnCl_2$, and additional Sigma antifoam B. The cells were then grown for 3 h at 25 °C and harvested by centrifugation, rinsed in buffer A containing 100 mM NaCl (A100), recentrifuged, and the pellet was stored at -80 °C. The average yield of wet weight cell paste ranged from 4 to 5 g/l of culture. The crude post-induction cell lysate was analyzed by SDS-PAGE stained with Coomassie Blue.

Purification of Integrase

All purification steps were performed at 4 °C or on ice except where noted. Frozen cells (35 g) were thawed and completely resuspended in 80 ml of buffer A20 plus 1 ml of protease inhibitor cocktail (Sigma, P8849), 5 mg of carboxypeptidase inhibitor from potato tuber (Sigma), and 300 μg/ml lysozyme. The suspension was incubated for 25 min at room temperature with constant stirring, placed on ice, and adjusted to 0.52 M NaCl,

1 mM $ZnCl_2$, 5 mM dithiothreitol. An additional 1 ml of protease inhibitor mixture and 5 mg of carboxypeptidase inhibitor were added, and the suspension was stirred for an additional 20 min. The lysate was sonicated for 20 min (20-W output, Fisher Scientific 60 Sonic Dismembrator) to reduce viscosity followed by centrifugation at 43,700 × *g* for 60 min to remove cellular debris, yielding 90 ml of cleared lysate. The cleared lysate was adjusted to 100 mM NaCl by the addition of buffer B with no salt (B0). Integrase was precipitated by the addition of 5% PEI while stirring to a final concentration of 0.05% (v/v). Stirring was continued for an additional 20 min, and the precipitate was collected by centrifugation at 15,300 × *g* for 20 min. The pellet was washed by homogenization using a Dounce homogenizer with 100 ml of buffer B with 350 mM NaCl (B350), and the pellet was recovered by centrifugation at 15,300 × *g* for 15 min. The pellet was then extracted twice using a Dounce Homogenizer with 150 ml of buffer B500 containing 3 mM dithiothreitol and 50 nM $ZnCl_2$ followed by centrifugation at 15,300 × *g* for 20 min each. The combined 300 ml of "PEI eluate" was diluted to 350 mM NaCl with buffer B0 and loaded at a flow rate of 5 ml/min onto a 38-ml Macro-Prep High S (Bio-Rad) column (6.4 × 2.6 cm) pre-equilibrated in B350. The column was washed to base line with 3-4 column volumes of buffer B500, and integrase was eluted with a 225-ml linear NaCl gradient from 0.5 to 1.5 M NaCl in buffer B at the same flow rate. Integrase was eluted as a broad peak centered around 1.1 M NaCl. Fractions from 900 mM to 1.4 M NaCl were pooled. The front portion of the elution peak, corresponding to a NaCl concentration lower than 900 mM, was excluded from the pooled eluate to avoid a nuclease contaminant. Integrase was precipitated with ammonium sulfate at 56.4% saturation and redissolved in 20 ml of storage buffer. The concentrated protein was dialyzed against 3 changes of 500 ml of storage buffer. The final integrase concentration in storage buffer after a 45-min clearing spin at 43,700 × *g* typically ranged from 130 to 200 μM as determined spectrophotometrically using ε_{280} = 59,940 M^{-1} cm^{-1}. This value of ε_{280} was calculated based on the presence of 2 Tyr, 10 Trp, and 4 Cys [24]. UV spectra of integrase obtained in 0−8 M urea showed no changes in A_{280} upon unfolding of the protein. Protein was stored at -80 °C in 0.3−1-ml aliquots. The protocol yielded 2−2.5 mg of purified integrase (>99% homogeneity as judged by SDS-PAGE stained with Coomassie Blue)/g of wet cells. The

purified pooled fractions, when assayed for both single and double-stranded nuclease activity, showed less than 5% nonspecific degradation of labeled DNA at incubation times in excess of 60 min, which exceeds the longest time points used in our experiments.

Presteady-state Burst Assay

Standard reaction mixtures (100 μl) contained 10 mM Na-HEPES, pH 7.5, 4% glycerol, 20 mM Tris, pH 8.0, 10 mM 2-mercaptoethanol, 0.050 mg/ml acetylated bovine serum albumin, and 130 mM NaCl. Integrase was first preincubated with radiolabeled oligonucleotide substrates on ice for 30 min then warmed briefly for 2 min at 37 °C immediately before the start of the assay. A complete range of preincubation (0−2 h) and warm-up times (0−15 min) were tested to ensure the mixtures had achieved equilibrium while minimizing enzyme degradation. Reactions were initiated by the addition of 37 °C $MnCl_2$ to 5 mM. Although there is uncertainty in the literature with regard to the physiological divalent metal cofactor for integrase, Mn^{2+} was used because it has been observed that retroviral integrases exhibit maximal activity in standard assays with this divalent cation [20]. Additionally, no activity was detected with ASV integrase when Mg^{2+} was used at physiological concentrations. A range of $MnCl_2$ concentrations (0−20 mM) were examined to optimize enzymatic activity. At time points ranging from 4 s to 30 min, aliquots of the reaction mixture were withdrawn and added to a quenching solution (8 M urea, 0.25 M EDTA, 20% sucrose) in a 2:1 quench:reaction mix ratio. Generally, seven time points were taken in the first 30 s of the reaction with seven more time points in the ensuing 29 min. The reaction products were then analyzed by denaturing sequencing PAGE (20% acrylamide, 8 M urea, TBE) using a Sequi-Gen GT 38 × 39 cm apparatus (Bio-Rad). Bands in the gel were visualized by using a Molecular Dynamics PhosphorImager (Amersham Biosciences), and the intensity of DNA bands was quantitated using ImageQuant software (Amersham Biosciences).

Quantitation

The intensity of each product band, $I_i(t)$, at each time, t, was first normalized with respect to the sum of intensities in the starting substrate band, $I_0(t)$, plus all product bands according to Equation 1.

$$F_{i,norm}(t) = \frac{I_i(t)}{I_0(t) + \sum_{i=1}^{n} I_i(t)} \quad \text{(Eq. 1)}$$

This normalized product fraction, $F_{i,norm}$, was then corrected for background intensity present at $t = 0$ for the ith band and renormalized for background intensities of all product bands to obtain the final corrected product fraction, $F_{i,corr}$ according to Equation 2.

$$F_{i,corr}(t) = \frac{F_{i,norm}(t) - F_{i,norm}(0)}{1 - \sum_{i=1}^{n} F_{i,norm}(0)} \quad \text{(Eq. 2)}$$

The number of exponential phases observable for any particular experiment increased with the length of the time regime under examination [8]. Accordingly, the resulting time courses were fitted to Equation 3 consisting of n exponential terms, with amplitudes A_i and apparent rate constants $\lambda_{burst,i}$, plus a linear term with an apparent rate constant λ_{lin} to fit to the linear portion of the ensuing exponential phase.

$$y = \sum_{i=1}^{n} A_i(1 - e^{-\lambda_{burst}t}) + \lambda_{lin}t \quad \text{(Eq. 3)}$$

Non-linear least squares fittings were performed using Kaleidagraph software (Synergy, Redding, PA).

Preparation and Purification of Splicing Product for Maxam-Gilbert Sequencing

A 300-µl reaction was performed under standard conditions with 3 µM integrase and 3 µM *5t/5b(2)3t/3b or *5t-2/5b(2)3t/3b (for oligodeoxyribonucleotide definitions see Table I) and allowed to proceed for 6.5 min. The reaction was stopped with quench solution, and the products were separated by denaturing sequencing PAGE in a 20% acrylamide, 8 M urea gel. The product bands were recorded on Kodak X-Omat film, and the largest product band was excised from the gel and isolated with an Elutrap electroelution device (Schleicher & Schuell, Keene, NH). Maxam-Gilbert sequencing was performed on the purified product DNA according to the published protocol [25].

Double-filter Binding

Nitrocellulose filter binding experiments were performed with only minor modifications to the double-filter method previously described [26]. Standard assay buffer without Mn^{2+} was used for binding studies. Integrase was first incubated on ice for 30 min with radiolabeled oligonucleotide substrates. After incubation, the solution was filtered through a combination of nitrocellulose (BA-S83) and DEAE (1 NA-45 or 3 DE-81) membranes (Schleicher & Schuell or Whatman, Maidstone, England) by using a 96-well Bio-Dot microfiltration apparatus (Bio-Rad). Immediately after filtration of the reaction solution, the wells of the apparatus were rinsed twice with an equal volume of ice-cold buffer. The filters were imaged by using a Molecular Dynamics PhosphorImager (Amersham Biosciences) and quantitated as described [26]. For experiments using DE-81 membranes, the radioactivity of all three membranes was quantified and summed.

RESULTS

Presteady-state Assay Reveals Splicing of Synapsed Substrate by ASV Integrase

Fig. 2.1*A* shows results from a typical first-turnover experiment performed with a synapsed DNA substrate, *5t/5b(2)3t/3b, as detected by electrophoresis on an 8 M urea, 20% acrylamide sequencing gel. This substrate contains DNA sequences from the U3 and U5 ends of the viral LTR organized with the cognate CATT sites held together by a two-nucleotide single-stranded tether in a head-to-head configuration [5]. The substrate was radiolabeled at 5t, the 21-mer DNA strand containing the U5 sequence. This 21-mer terminates at the 3'-end with the processing cognate sequence, CATT, which is endonucleolytically cleaved by integrase at the -2 position to exponentially yield an expected 19-mer as the major product at 0.2 s^{-1} (Fig. 2.1*B*) plus a dinucleotide TT product, which is silent in this assay [27–29]. The 5.9 nM burst amplitude reflects only 1.2% of total integrase monomers. Although the concentrations used (500 nM integrase, 500 nM DNA) were known to be not saturating, the propensity of integrase to aggregate at higher concentrations precluded experiments under saturating conditions [8]. In addition, a minor 18-mer side product was also observed. Because the ratio between the two products was

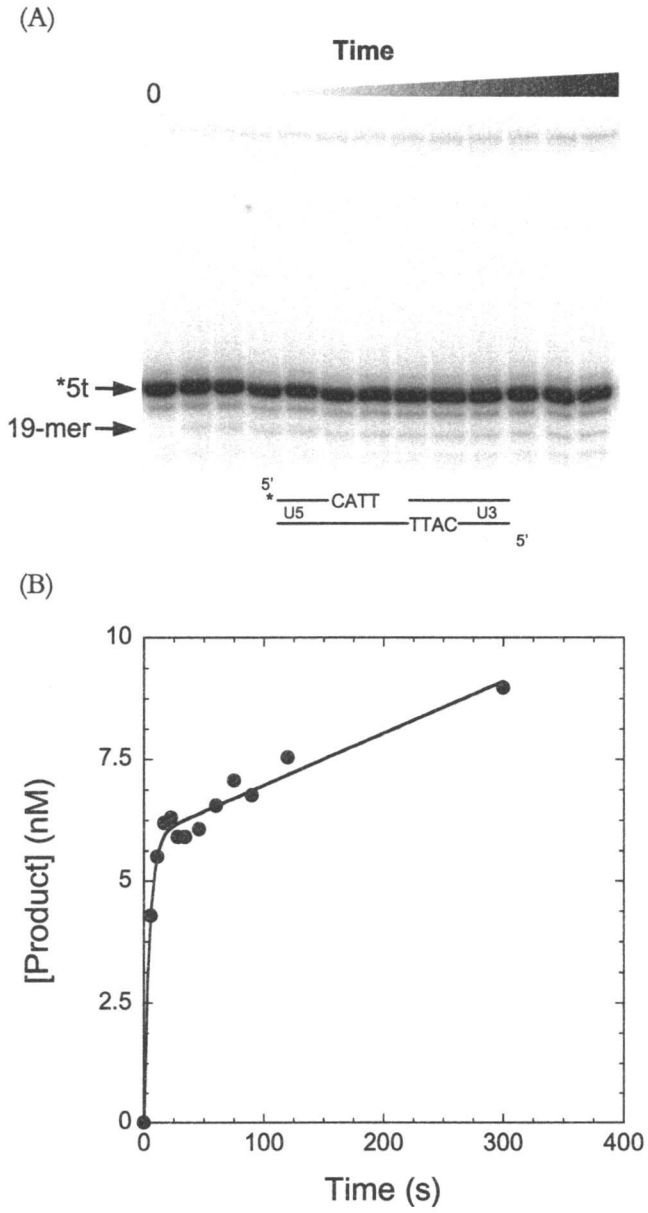

Figure 2.1: Processing reactivity at the terminal CATT cognate site on the 21-mer *5t of the dual-site synapsed substrate. First-turnover processing activity assay at 0.5 μM integrase, 0.5 μM *5t/5b(2)3t/3b, 100 mM NaCl with samples subjected to electrophoresis on a 20% polyacrylamide, 8 M urea, TBE sequencing gel (*A*). Lanes represent reaction times of 0, 5.9, 11.1, 17.1, 22.6, 28.0, 34.1, 46.0, 60.0, 75.0, 90, 120, and 300 s. The amount of processing products was quantified using a PhosphorImager and a plot of the results shown in *B*. The *solid line* represents best fit of the data as described under "Experimental Procedures" with A_1 = 5.9 ± 0.1 nM, λ_1 = 0.20 ± 0.03 s^{-1}, λ_{lin}= 0.010 ± 0.001 nM/s.

approximately constant, the 18-mer was likely produced in parallel to the 19-mer major product rather than from the degradation of the 19-mer. This 18-mer is consistent with "near miss" processing at the minus 3 position observed with Mn^{2+} as the metal cofactor as reported [28, 30].

Unexpectedly, a third product appeared that was larger than the starting 21-mer. Isolation from a semi-preparative scale assay reaction and sequencing (Fig. 2.2*A*) showed that this product was a 46-mer derived from the direct splicing of an unprocessed *5t 21-mer into a 5b(2)3t 44-mer strand at its internal cognate CATT site (Fig. 2.2*C*). Interestingly, the electrophilic site of splicing occurred at the expected processing position for the U3 cognate sequence, CATT, of 5b(2)3t.

To determine whether this splicing reaction maintains the same sequence specificity as the processing reaction or if the position of attack resulted from the opportunistic placement of the 3'-OH of *5t as a convenient nucleophile near the processing site of 5b(2)3t, an alternate substrate, *5t-2/5b(2)3t/3b, with a slightly different geometry was assayed, and the large product was isolated and its nature examined. In place of the radiolabeled *5t 21-mer previously used, this substrate contained a radiolabeled 19-mer, *5t-2, lacking the terminal TT dinucleotide of *5t. This substitution placed the attacking 3'-OH two base pairs farther away from the CATT of 5b(2)3t, where splicing was observed with the 21-mer. If the observed splicing into 5b(2)3t at the cognate CATT site by *5t was simply a result of spatial constraints, then the site of attack by the 19-mer *5t-2 would be expected to change in accordance with the distance change. Conversely, if splicing into the cognate CATT was sequence-specific, then *5t-2 would attack at the same location as *5t despite being two nucleotides farther removed from the site of splicing. Because *5t-2 was identical in sequence to the processing product 19-mer, no bands corresponding to smaller products were expected or observed. However, a product larger than the starting material was again detected (data not shown). Sequencing showed this product to be a 44-mer (Fig. 2.2*B*), resulting from the attack of the terminal 3'-CA-OH of *5t-2 on 5b(2)3t at the identical position, between the A and T, of the internal cognate sequence (Fig. 2.2*D*). Surprisingly, Kukolj and Skalka [5] did not observe products consistent with the splicing

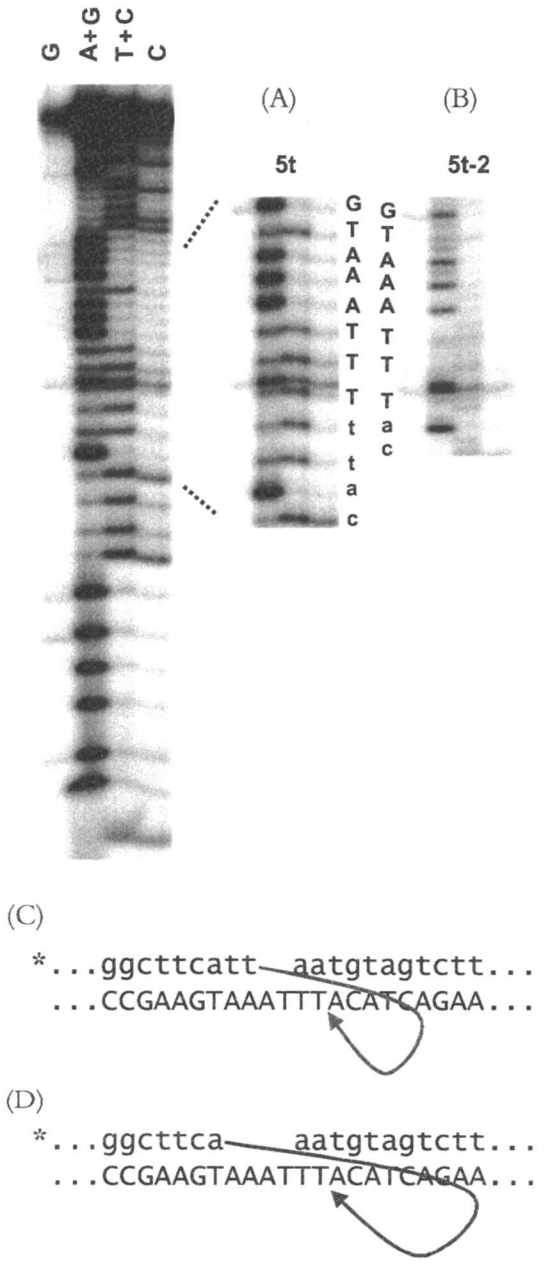

Figure 2.2: Sequence analysis of the splicing junction 46-mer (5t) and 44-mer (5t-2) products. DNA in the large product bands was isolated and purified from a 390-s activity assay at 3 μM integrase, 3 μM *5t/5b(2)3t/3b (*A*) or *5t-2/5b(2)3t/3b (*B*), 100 mM NaCl. Maxam-Gilbert sequencing of the product identified the position of nucleophilic attack by the unmodified 3'-OHs to be between the A and T of the internal cognate CATT for both *5t/5b(2)3t/3b (*C*) and *5t-2/5b(2)3t/3b (*D*).

reaction for a substrate with spacing similar to *5t-2. At present, no obvious rationale can be provided for this apparent discrepancy.

The 44- and 46-mer products result from nucleophilic attack by a 3'-OH from two different terminal sequences, CA and CATT, respectively, at an apparently sequence–specific electrophilic site despite the fact that the two substrates placed the attacking 3'-OH nucleophiles at different distances from the site of splicing. The relative efficiencies of the two reactions were measured in first-turnover experiments to investigate this phenomenon quantitatively (Fig. 2.3). Formation of splicing products in both reactions showed identical initial exponential phases with an exponential product formation rate of 0.02 s^{-1} in the early time regime. Interestingly, the amplitudes and, more importantly, the rates of product formation of the initial exponential phase, representing the first turnover of catalytically competent pre-formed complexes, were comparable for both substrates despite the possible difference in the location of the nucleophiles relative to the site of splicing. The cognate sequence specificity of the electrophile binding site alone appears to be sufficient to determine the specificity of the splicing reaction.

Sequence Requirements of the Splicing Reaction

Two mutant single-site synapsed substrates were synthesized by changing either the U5 or U3 cognate CATT sequence to GCAA to examine nucleophile *versus* electrophile selectivity. The substrate *m5t/m5b(2)3t/3b contained a radiolabel on the 21-nucleotide–long m5t strand, in which the terminal CATT sequence of the 5t strand was substituted with GCAA. A complementary mutation was made on the DNA strand of the substrate containing the tether, m5b(2)3t, to maintain base pairing. The internal U3-derived cognate sequence, where splicing occurs, was unmodified. Alternatively, the substrate *5t/5b(2)m3/m3b was synthesized by making the CATT to GCAA substitution at the internal cognate sequence, derived from the U3 LTR, while leaving the sequence at the U5-derived terminus unaltered.

Fig. 2.4 shows the results from assays, at low (100 mM, *A*) and high (400 mM, *B*) NaCl concentrations, with the unmodified dual-site substrate along with those of the two mutant single site substrates. The single-site substrate *m5t/m5b(2)3t/3b (*circles*),

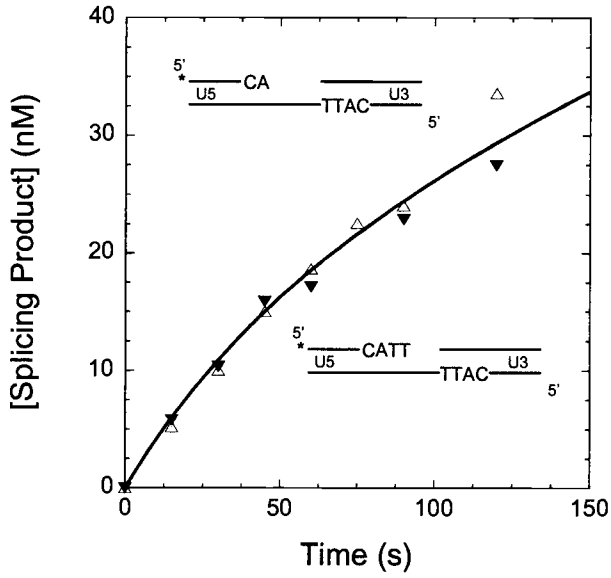

Figure 2.3: Splicing reactivity of processed and unprocessed ends. First-turnover splicing activity assays at 3 μM integrase, 3μM DNA substrate, 100 mM NaCl with samples electrophoresed on a 20% polyacrylamide, 8 M urea, TBE sequencing gel. A plot of the results shows identical splicing for both *5t/5b(2)3t/3b (*closed triangles*) and *5t-2/5b(2)3t/3b (*open triangles*). The *solid line* represents best fit of the data as described under "Experimental Procedures" with A_1 = 20 ± 3 nM, λ_1 = 0.02 ± 0.01 s^{-1}, λ_{lin} = 0.13 ± 0.02 nM/s.

comprising a nonspecific splicing nucleophile and a cognate splicing electrophile, was observed to have significantly higher reactivity than the conventional dual-site substrate *5t/5b(2)3t/3b (*squares*). In contrast, the single-site substrate with a cognate splicing nucleophile and a nonspecific splicing electrophile sequence, *5t/5b(2)m3b/m3t (*diamonds*), had barely measurable splicing activity. Fig. 2.4*B* also shows that the amounts of splicing product observed in the presteady-state assays are greatly increased at 400 mM NaCl. Although the rate constant for the initial appearance of these products was smaller, the actual rate of product formation when expressed as mole of product formed/s was similar. This suggests that the observed reduction in rate constant may merely be a reflection of the increase in amplitude. Consistent with this result, preliminary experiments beyond the scope of this report indicate that the inhibitory effects due to the aggregation of integrase into higher assembly states can be partially reversed by the addition of salt [8].

Because the choice of nucleophile used in the splicing reaction can be CATT-OH, a processed CA-OH (Fig. 2.3), or the non-cognate GCAA-OH, these results showed that the nucleophile binding site on integrase is not sequence-specific. The cognate CATT sequence was, however, required at the electrophile site in order for efficient splicing to occur. Given the lack of discrimination of the enzyme regarding selection of a nucleophile in these reactions, the 5'-OH in spatial proximity to the splicing site was also examined for reactivity. This hydroxyl is at the 5' terminus of the 21-mer strand, 3b, annealed to the 44-mer 5b(2)3t that contains the internal CATT splicing site. In an experiment where 3b* was 3'-end-radiolabeled using [α-^{32}P]TTP and wild type T7 DNA polymerase, no new bands representing any reaction products were observed under the conditions tested (data not shown). This result indicated that although any 3'-OH can act as a nucleophile, the 5'-OH is non-reactive.

Effect of Synapsed Ends on Equilibrium Binding

The equilibrium binding of integrase to these substrates was investigated by the nitrocellulose double-filter binding method [26]. In addition to the usual dual-site synapsed substrate *5t/5b(2)3t/3b, the single-site 20-mer substrate *3t/3b and the single-site synapsed substrate *m5t/m5b(2)3t/3b were examined, and the data were

Figure 2.4: Reactivity of mutant single-site synapsed substrates. *A*, first-turnover activity assay at 3 μM ASV integrase, 3 μM DNA substrate, and 100 mM NaCl. The time-dependent appearance of splicing products for *m5t/m5b(2)3t/3b (*circles*) and *5t/5b(2)m3t/m3 (*diamonds*) are plotted. Additionally, splicing products from the dual-site synapsed substrate *5t/5b(2)3t/3b (*squares*) are shown for comparison. The *solid lines* for the splicing products from *m5t/m5b(2)3t/3b and *5t/5b(2)3t/3b represent best fit of the data to Equation 3 with A_1 = 16.3 ± 0.8 nM, λ_1 = 0.070 ± 0.005 s^{-1}, λ_{lin} = 0.25 ± 0.01 nM/s (*circles*) and A_1 = 16.9 ± 0.7 nM, λ_1 = 0.02 ± 0.01 s^{-1}, λ_{lin} = 0.100 ± 0.004 nM/s (*squares*). The *solid line* for the splicing products from *5t/5b(2)m3t/m3b represent best fit of the data to a line of the form $y = 0.034t + 0.012$ ($r = 0.996$). *B*, first-turnover activity assay at 3 μM ASV integrase, 3 μM DNA substrate, and 400 mM NaCl. The time-dependent appearance of splicing products for *m5t/m5b(2)3t/3b (*circles*) and *5t/5b(2)m3t/m3 (*diamonds*) are plotted. Splicing products from the dual-site synapsed substrate *5t/5b(2)3t/3b (*squares*) are shown for comparison. The *solid lines* for the splicing products from *m5t/m5b(2)3t/3b and *5t/5b(2)3t/3b represent best fit of the data to Equation 3 with A_1 = 2.02 ± 0.06 μM, λ_1 = 0.010 ± 0.001 s^{-1}, λ_{lin} = (1.1 ± 0.5)×10^{-4} μM/s (*circles*) and A_1 = 1.6 ± 0.2 μM, λ_1 = (2.9 ± 0.4)×10^{-3} s^{-1}, λ_{lin} = (2 ± 1)×10^{-4} μM/s (*squares*). The *solid line* for the splicing products from *5t/5b(2)m3t/m3b represent best fit of the data to a line of the form $y = (4.2\times10^{-5})t + 0.001$ ($r = 0.99$).

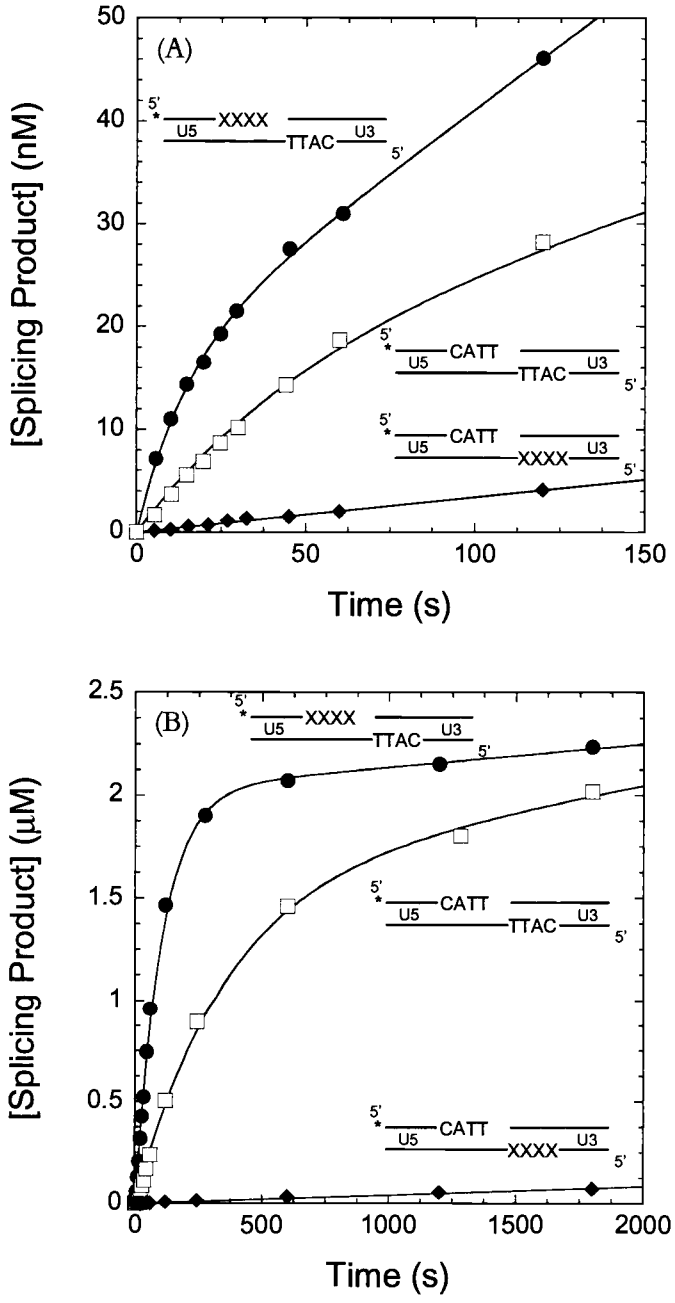
(A)
[Splicing Product] (nM)
Time (s)
5'
XXXX
U5
U3
TTAC
CATT
(B)
[Splicing Product] (μM)
Time (s)

Figure 2.4

phenomenologically fitted to $y = [IN]^n/(K + [IN]^n)$, which describes binding with infinite cooperativity to n sites. The parameter n, akin to the Hill coefficient, is a direct measure of the apparent cooperativity of binding and provides a lower limit for the minimal number of binding sites required for a fit. The two single-site substrates titrated identically despite their difference in length (Fig. 2.5). By comparison, the dual-site synapsed substrate showed both an increase in binding affinity and an increase in apparent cooperativity of binding as indicated by a shift to the left and an increase in the steepness of transition from $n = 1.6 \pm 0.06$ to 2.0 ± 0.08. In the absence of an independent measure of the assembly state of integrase, these titration data alone were insufficient to define a specific model for binding. However, these results were indicative of a change in binding modes by integrase due to the presence of two cognate sites on the DNA substrate. In contrast, when only a single cognate site was present on a synapsed substrate, binding was identical to that of two different substrate molecules despite a significant difference in their sizes.

Intramolecular versus Intermolecular

Theoretically, the splicing reaction could have occurred either intermolecularly, between two different synapsed substrates, or intramolecularly, with the *5t strand splicing into the CATT of the 5b(2)3t annealed to it (Fig. 2.6*A*). To address this issue, an integrase activity assay was performed using an equimolar solution of radiolabeled "preprocessed" *5t-2/5b(2)3t/3b and unlabeled full-length 5t/5b(2)3t/3b. The rationale for the assay stems from the assumption that both the 5t and the 5b(2)3t strands can serve as electrophiles because they both contain a cognate sequence CATT. Therefore, in an intermolecular splicing reaction, the 3'-OH from a given 19-mer *5t-2 strand would have the choice of splicing into either the internal CATT of 5b(2)3t (Fig. 2.6*A*, *Intermolecular option I*) or the terminal CATT-OH of 5t (Fig. 2.6*A*, *Intermolecular option II*) to yield products with expected sizes of 44 and 21 nucleotides, respectively. Although a substantial amount of 44-mers was observed in the assay, no 21-mer products were detected (Fig. 2.6*B*).

To rule out the trivial possibility that the splicing reaction could not occur at a terminal cognate site, a reverse-polarity substrate was synthesized utilizing a 5'−5' phosphate linkage in the tether region. The substrate *5t-2/5b(2)3b/3t replaces the internal CATT of

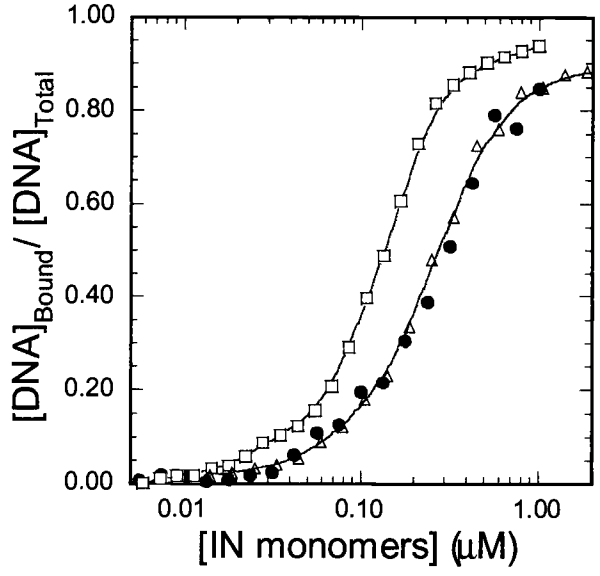

Figure 2.5: Substrate binding titrations. Nitrocellulose/DEAE filter binding studies were performed at 0.2 μM DNA substrate and integrase concentrations ranging up to 2.5 μM. The *solid lines* represent fits to the equation: $y = [IN]^n / (K + [IN]^n)$, which describes binding with infinite cooperativity to n sites. DNA substrates used were the single-site ⋆3t/3b (*closed circles*, $n = 1.6 \pm 0.06$), the single-site synapsed ⋆m5t/m5b(2)3t/3b (*open triangles*, $n = 2.0 \pm 0.08$), and the dual-site synapsed ⋆5t/5b(2)3t/3b (*open squares*).

Figure 2.6: Intermolecular *versus* intramolecular splicing mechanisms. *A*, intermolecular splicing can occur either at an internal CATT (*option I*) or at a terminal CATT sequence (*option II*). In contrast, intramolecular splicing can only occur at an internal CATT sequence. *B*, single-turnover activity assay at 3.0 μM integrase, 3.0 μM DNA substrate, 100 mM NaCl with samples subjected to electrophoresis in a 20% polyacrylamide, 8 M urea, TBE sequencing gel. Lanes represent reaction times of 0, 15, 30, 45, 60, 75, 90, 120, 240, and 360 s. The substrate was made by simultaneously annealing 5b(2)3t 1:1 with 3b, 1:0.5 with 5t, and 1:0.5 with *5t-2 to create an equimolar population of *5t-2/5b(2)3t/3b and 5t/5b(2)3t/3t. The radiolabel on the 19-mer *5t-2 strand allows detection of splicing products at either the terminal CATT-OH of 5t (intermolecular option I; expected product size of 21-mer) or the internal CATT of 5b(2)3t (intermolecular option II or intramolecular; expected product size of 44-mer). No 21-mer products above the background amount present at $t = 0$ were detected over time.

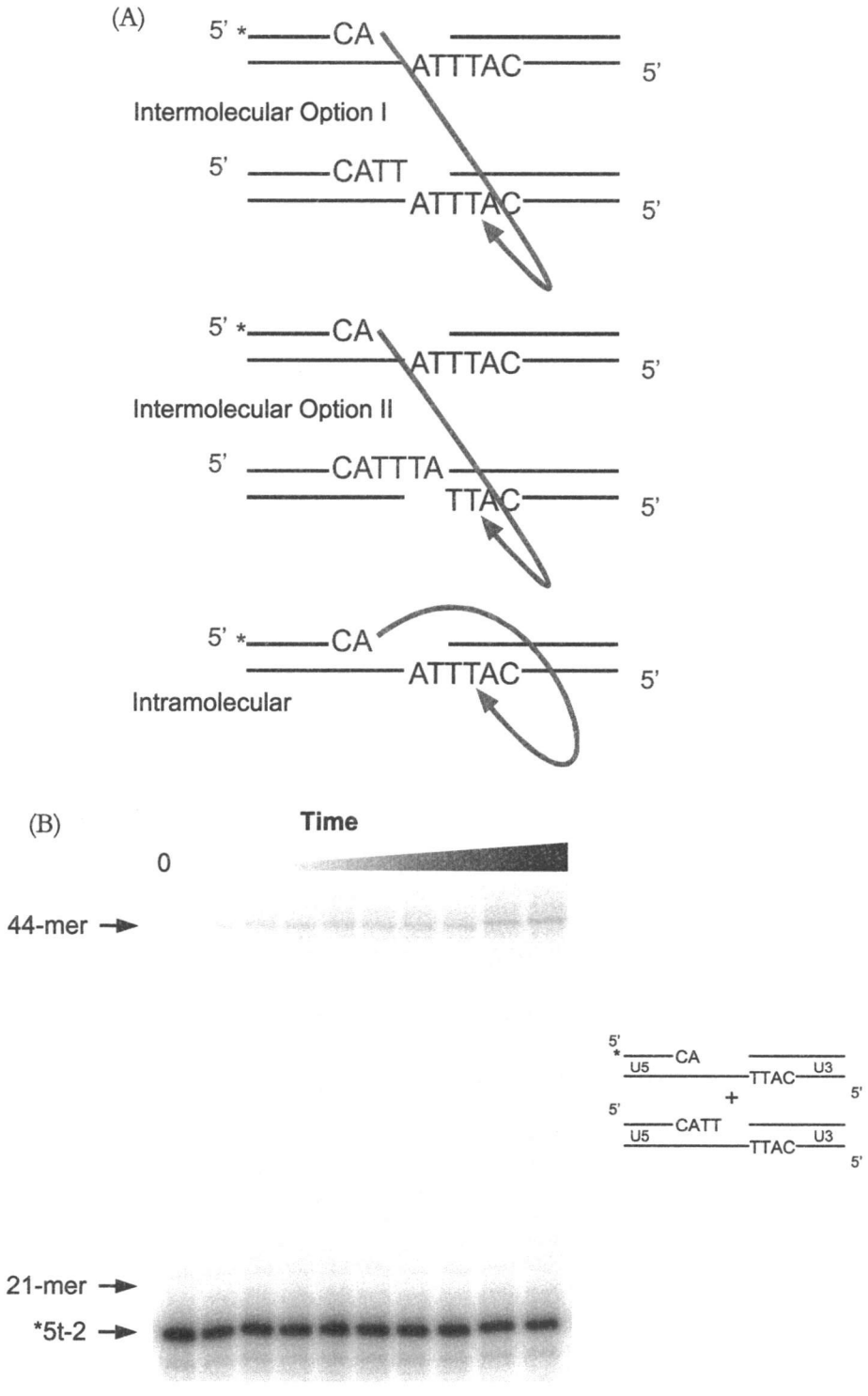
(A)
5' *
CA
ATTTAC
5'
Intermolecular Option I
5'
CATT
ATTTAC
5'
5' *
CA
ATTTAC
5'
Intermolecular Option II
5'
CATTTA
TTAC
5'
5' *
CA
ATTTAC
5'
Intramolecular
(B)
Time
0
44-mer
5'
*
CA
U5
TTAC
U3
5'
+
5'
CATT
U5
TTAC
U3
5'
21-mer
*5t-2

Figure 2.6

the synapsed substrate with a terminal CATT-OH (Fig. 2.7*A*). The preprocessed 19-mer *5t-2 strand was radiolabeled; splicing into the only available cognate site would extend this 19-mer to a 21-mer. Fig. 2.7*B* shows an exponential phase of product formation with product sizes ranging from 21 to 25 appearing within 20 s, a range of products smaller than the 19-mer starting material appearing after a lag of 1 min, and negligible amounts of products larger than a 25-mer even at the longest reaction time. The insignificant amounts of larger products are consistent with a lack of nonspecific joining by the radiolabeled preprocessed end. The delayed appearance of the smaller products was likely the consequence of sequential processing-joining reactions originating from the cognate site of the unlabeled 3t 21-mer. More importantly, the exponential appearance of 21−25-mer initially revealed the presence of splicing activity with this substrate. This demonstrated that the terminal CATT-OH functions as a viable splicing electrophile, although integrase apparently is unable to splice exclusively into the minus 2 position. In contrast, the reaction with equimolar labeled *5t-2/5b(2)3t/3b and unlabeled 5t/5b(2)3t/3b failed to produce any products in the 21- to 25-mer range. The lack of splicing products characteristic of intermolecular option II (Fig. 2.6*A*) suggests an intramolecular mechanism of splicing.

Processing Versus Splicing of the Internal Cognate Site

Assays in which the synapsed 5b(2)3t* 44-mer was radiolabeled yielded radiolabeled 19-mer product from either the splicing or the processing reactions since both are targeted at the identical phosphate bond between the A and T of the internal CATT cognate sequence. To determine the relative contributions of the two reactions to total 19-mer products observed, a first-turnover experiment was performed with 5t/5b(2)3t*/3b. The radiolabel was placed exclusively on the synapsed 5b(2)3t* 44-mer, and time points were collected within the first turnover. In the presteady state, the 19-mer products from the splicing and processing reactions were expected to resolve into two distinct exponential phases with characteristic rate constants of 0.02 and 0.20 s^{-1}, respectively. Fig. 2.8*A* shows the results of such an experiment. Best fit for the amount of 19-mer products formed over time revealed a single exponential phase with an amplitude of 5.8 nM and an apparent rate constant of 0.02 s^{-1}, which is characteristic of the splicing reaction. No exponential phase

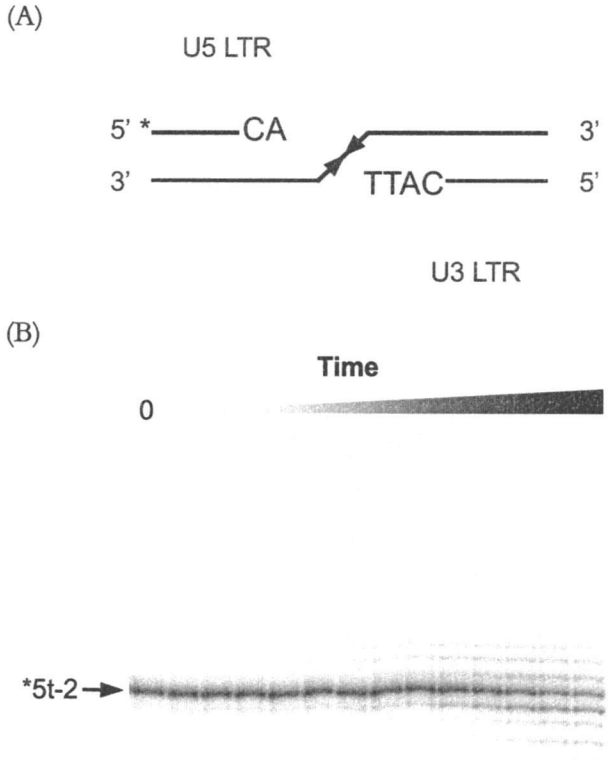

Figure 2.7: Reactivity of a processed reverse-polarity substrate. *A*, a reverse-polarity synapsed substrate containing sequences derived from ASV U5 and U3 LTR sequences attached via a 2-nucleotide single-stranded tether with a 5'−5' phosphate linkage (*facing arrows*). *B*, single-turnover activity assay at 5.0 μM integrase, 0.5 μM *5t-2/5b(2)3b/3t, 400 mM NaCl with samples subjected to electrophoresis on a 20% polyacrylamide, 8 M urea, TBE sequencing gel. Lanes represent reaction times of 0, 5.8, 10.6, 15.3, 20.1, 25.7, 30.9. 45.3. 60. 120. 240. 600. 1200. and 1802 s.

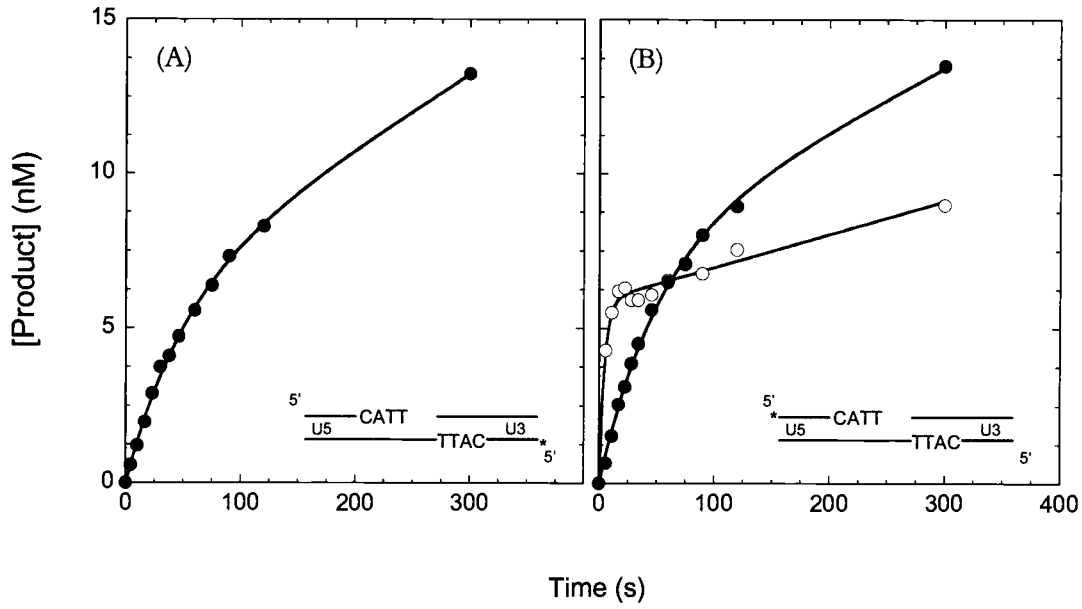

Figure 2.8: Splicing *versus* processing at the internal cognate site. Single-turnover activity assay at 0.3 μM integrase, 0.3 μM DNA substrate, 100 mM NaCl as described under "Experimental Procedures." The amount of processing products (*open circle*) and splicing products (*closed circle*) was quantified using a PhosphorImager. *A*, the results for the substrate 5t/*5b(2)3t/3b. *B*, the results for the substrate *5t/5b(2)3t/3b. The *solid line* represents best fit of the data using Equation 3 with the parameters A_1 = 5.8 ± 0.4 nM, λ_1 = 0.020 ± 0.001 s^{-1}, λ_{lin} = 0.020 ± .001 nM/s (*A*) and $A_{19\text{-mer}}$ = 5.9 ± 0.1 nM, $\lambda_{19\text{-mer, 1}}$ = 0.20 ± 0.03 s^{-1}, $\lambda_{19\text{-mer, lin}}$ = 0.010 ± 0.001 nM/s, $A_{46\text{-mer}}$ = 6.9 ± 0.4 nM, $\lambda_{46\text{-mer, 1}}$ = 0.20 ± 0.02 s^{-1}, $\lambda_{46\text{-mer, lin}}$ = 0.020 ± 0.002 nM/s (*B*).

with the faster rate constant of 0.20 s^{-1}, characteristic of the processing reaction, was observed. Furthermore, the amplitude of this exponential phase was, within error, identical to that observed for the production of the 46-mer splicing product in experiments performed with *5t/5b(2)3t/3b with respect to both rate of product formation and amplitude (Fig. 2.8*B*). This result shows that the internal CATT cognate site is not processed as a terminal cognate site in a hydrolysis reaction but, instead, undergoes the splicing reaction exclusively.

DISCUSSION

Presteady-state Analysis of a Novel Splicing Reaction

Integrase catalyzes two distinct reactions *in vivo*, a 3' dinucleotide trimming or processing reaction and a joining reaction whereby the processed viral 3'-ends become integrated into the host genome. Kukolj and Skalka [5] originally designed synapsed-end substrates to improve integrase processing *in vitro*. The rationale for these substrates, where two cognate sites are tethered in a head-to-head fashion, was to facilitate the assembly and binding of an active complex with both LTR ends bound to integrase. Under the steady-state assay conditions reported, these substrates gave enhanced product formation relative to short oligodeoxyribonucleotide duplexes containing only single ends.

Using similarly designed synapsed-end substrates in our presteady-state studies of the integrase-catalyzed processing reaction, we discovered a novel intramolecular splicing reaction resembling an integration reaction that occurs without prior requisite processing. This splicing reaction is characterized by the following features. 1) The sequence of the DNA strand containing the 3'-OH nucleophile is not critical; however, a 5'-OH cannot replace the 3'-OH or H_2O; 2) a CATT cognate site is preferentially targeted as the electrophile in the reaction; 3) splicing occurs intramolecularly with respect to the synapsed substrates to yield a long hairpin product plus a short fragment, which is structurally indistinguishable from a processing product; and 4) splicing occurs as an exponential phase with a characteristic rate constant of 0.020 s^{-1}, which is 10 times slower than processing.

Although the splicing reaction appears to resemble the joining activity of integrase, the sequence specificity for a cognate CATT as the electrophile represents a significant mechanistic difference between the two. The integrase-catalyzed joining reaction shows very little sequence specificity with regard to target site selection [4, 31], although it does seem to show a structural preference, *e.g.* for bent DNA [32] or cruciform stem-loops [31]. In contrast, the splicing activity of integrase is specific for the cognate sequence CATT such that splicing preferentially occurs between the A and T of this sequence. If the splicing reaction were representative of the true joining reaction, then selective integration at CATT sites in the host genome would be expected. This has not been observed. We have examined the sequence of putative "hot" sites [33] and were not able to correlate sites of integration with the existence of CATT sequences (data not shown).

On the other hand, the preference for a cognate CATT as the electrophile is shared in common with the processing reaction, the difference being that in the splicing reaction a 3'-OH from DNA serves as the nucleophile, a role played by H_2O in the processing reaction. However, even in this aspect, the two reactions are similar. The promiscuity regarding the choice of nucleophiles in the processing reaction is well documented; non-water nucleophiles ranging from glycerol to serine and threonine [29] and even the 3'-OH of the leaving dinucleotide [18] have been observed. In the case of the splicing reaction, we have additionally shown that the DNA sequence attached 5' to the attacking 3'-OH is irrelevant. Thus, the splicing reaction may in fact represent a "rogue reaction" closely akin to processing, at least with regard to the specificity requirements of the electrophile and nucleophile binding sites.

An unexplained feature of the splicing reaction is the observation that no processing, as defined by H_2O cleavage, occurred at the internal CATT site. It would be expected that the internal CATT is a viable site for processing as previous studies [29] have shown that processing can occur at cognate sites with extended 3'-tails. Perhaps exclusive splicing arises from strong competition due to the adventitious proximity of the 3'-OH inherent in the design of the substrate. Additionally, it has been observed that ASV integrase distorts the linear DNA ends before the processing reaction and that fraying of the

DNA ends leads to an increase in processing activity [34]. It is possible that the design of the synapsed-end substrates is such that the internal cognate site is not as susceptible to fraying.

Implications for Previous Steady-state Analysis with Synapsed-end Substrates

The presteady-state results illustrate the problems inherent in the "long" reaction time assays routinely favored. Products found after multiple exponential phases of product formation represent the time-averaged accumulation of products formed from multiple catalytic events. Assays that rely on the quantitation of products at a single point or a series of points in a longer time regime assume a single congruent reaction pathway for the generation of each product. In addition, the comparison between the accumulation of two different products in this time regime represents an inaccurate characterization of the reactivity of the enzyme with respect to the two reactions. This is especially true for reactions such as those catalyzed by integrase, when an exponential formation of product is apparent before reaching steady state. The initial exponential phase represents the actual reactivity of the enzyme with respect to both binding specificity and catalytic rate. In addition, the initial exponential phase arises as a direct consequence of a slow post-chemistry step in the reaction mechanism and generally reflects a slow rate-limiting product release step [35]. In these instances, a faster steady-state turnover typically reflects only a faster rate of product release [21, 36–39]. A more complete discussion of the use of the initial exponential phase to assess substrate reactivity is contained in the second paper of this series [8].

Analysis of the exponential rates and amplitudes characteristic of the initial exponential phase, arising from both true processing and the rogue splicing reaction, provided the basis for the discovery of exclusive splicing activity at the internal CATT site. In contrast, previous experiments relying on measurements made in a time regime where the enzyme had undergone multiple exponential phases of product formation [5] erroneously attributed all the cleavage products at this site to a processing reaction and, as a result, incorrectly estimated the extent of processing seen with these substrates. Furthermore, under these assay conditions, the splicing reaction has almost twice the linear

turnover rate when compared with the processing reaction despite having a similar initial exponential amplitude with a 10-fold slower exponential rate of product formation. Therefore, the net accumulation of splicing products after multiple exponential phases of product formation, mistakenly quantified as processing products, does not represent an accurate measure of the relative extent of the two reactions but may merely reflect the fact that the splicing product is bound less tightly by the enzyme than the processed product.

Although the design of the original synapsed-end substrate introduced complications with respect to quantitation of integrase activity, the binding data reported here do support the hypothesis that the tethering of two cognate sequences facilitates assembly of an integrase-DNA complex. These data show a higher level of cooperativity and tighter binding for the synapsed substrate *versus* either a single-ended untethered substrate (3t/3b) or a synapsed substrate with only one cognate CATT present. In addition, the faster 0.2 s^{-1} burst rate constant for the processing reaction was observed only using substrates that contained two cognate sites. A more ideal design of synapsed substrates, therefore, requires both cognate CATTs to be situated at the 3'-termini of their respective strands. This requirement is readily accommodated by the incorporation of a 5'-5' linkage in the tether segment, as found in the reverse polarity substrate, 5t/5b(2)3b/3t. Mechanistic studies using this new class of synapsed substrates are presented in the accompanying paper of this series [8].

Similarities to V(D)J Recombinases

Interestingly, the splicing reaction of integrase closely resembles a reaction catalyzed by the V(D)J recombinases RAG1 and RAG2 [40–43]. This site-specific recombination requires a 3'-OH at the end of the attacking strand for joining, uses this 3'-OH as the nucleophile in an intramolecular attack at a sequence specific site to form a hairpin, requires divalent metal ions, and is independent of ATP [42]. The parallel observed between these two enzyme-catalyzed transesterification reactions provides the first functional evidence in support of the hypothesis that V(D)J recombinases belong to the same transposase structural superfamily that includes human immunodeficiency virus integrase, ASV integrase, MuA

[42], and would therefore predict a strong structural similarity between these enzymes and retroviral integrases.

A Model for Functional Switching of Binding Sites

The characterization of the splicing reaction also provides significant insights into the structure-function relationship of integrase. In particular, it highlights the important distinction between structurally defined binding sites such as the cognate sequence recognition site, where CATT is bound, *versus* functionally defined sites, such as the electrophile or the nucleophile binding sites. Our results show that in both the splicing and the processing reaction, only the electrophile is bound in the structurally-fixed sequence-specific CATT binding site, whereas the nucleophilic 3'-OH occupies a nonspecific binding site. On the other hand, in the true joining reaction, this relationship between functional and structural sites is reversed. Namely, nonspecific DNA serves as the electrophile, whereas sequence-specific processed-end CA-OHs act as nucleophiles.

In an extreme case, this structure-function switch would require having different sets of active sites for the processing reaction and the joining reaction. In such a model, the processing would take place in a structurally coupled cognate site-specific electrophile binding site. After processing, the viral ends would then be transferred to a nearby cognate site-specific nucleophile binding site. In addition, joining would also require the binding of target host DNA in a third site, one where nonspecific DNA is bound as the electrophile. The host DNA cannot bind in the electrophile site used for processing, since it is cognate-sequence specific.

The current results suggest the possibility of a model that requires only two structurally fixed sites per retroviral DNA end, a sequence-specific site and a sequence-nonspecific site. In contrast to the more complex model, the functionalities of these two sites are not fixed but would switch during the course of catalysis. This is similar to classical acid-base mechanisms for phosphodiester hydrolysis proposed for ribonuclease [44] and alkaline phosphatase [45]. In the initial processing state, the constellation of catalytic groups in the sequence-specific site is poised to activate the bound DNA as an electrophile, whereas the sequence-nonspecific site functions as a nucleophile-activating site.

A switch in the functionality of the two sites accompanies the catalysis of the processing reaction. This switch obviates the need to translocate the processed DNA to a separate nucleophile-activating site. At the same time, the coupled switch would create an electrophile-activating site in the nonspecific binding site in preparation for binding nonspecific target DNA.

The model makes chemical sense, as illustrated in Fig. 2.9 for a minimal constellation consisting of an acid-base pair. In this hypothetical scenario, a base in the nonspecific site activates H_2O as the nucleophile, whereas a corresponding acid in the sequence-specific site activates the phosphate backbone of the bound CATT sequence as the electrophile. In the course of the reaction, the base in the nonspecific site becomes protonated, and the acid in the specific site is deprotonated. This change in the protonation state of the catalytic acid and base in the two sites effectively switches the functionality of the two structural sites. As a result, the processed ends of the viral DNA automatically become bound in the nucleophile site without any need for translocation into a separate site.

This switching model predicts that the processing and joining reactions should be chronologically coupled because the correct association of functional and structural sites for the joining reaction requires prior catalysis of the processing reaction. This may be the basis for the contrast between the observed sequence specificity of the splicing reaction and the relative lack of sequence specificity of the joining reaction. In this context, the hybrid nature of this splicing reaction, with specificity akin to the processing reaction but functionality related to the joining reaction, reflects the fact that integrase exists predominantly in a form where the sequence-specific site is in an electrophile-activating configuration. Thus, integrase is generally poised to bind a cognate sequence as the target of nucleophilic attack and use any convenient functional hydroxyl, the spatially proximal 3'-OH of an unprocessed DNA strand in the case of the synapsed-end DNA molecules, as the nucleophile. The result is catalysis of the observed splicing reaction with these synapsed-end substrates.

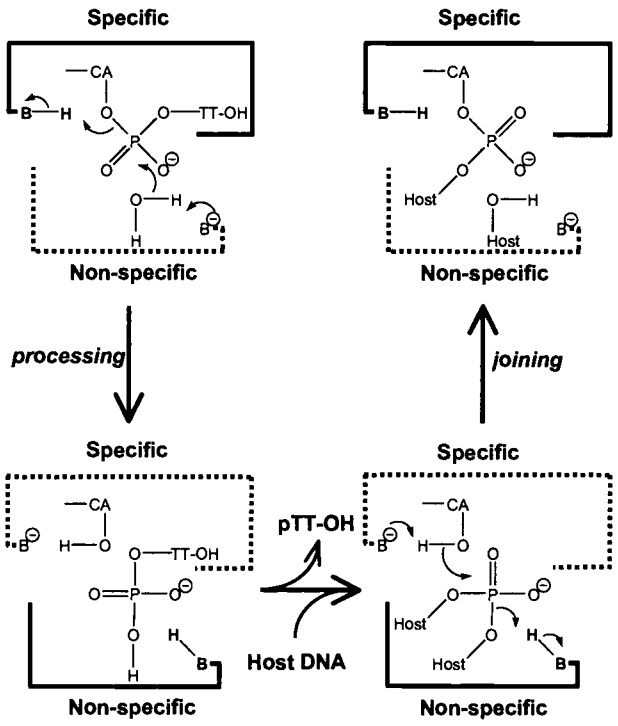

Figure 2.9: A two-site model for the structure-function switch in integrase. The sequence-specific binding site is depicted at the *top* of each reaction and the sequence-nonspecific site is depicted at the *bottom* of each reaction. Whether a particular site is poised to bind an electrophile or nucleophile in the reaction is denoted by a *solid box* or a *dashed box*, respectively, and corresponds to the protonation state of an acid-base pair in this hypothetical "simplest case" mechanism.

REFERENCES

1. Peliska, J.A. and S.J. Benkovic, *Mechanism of DNA strand transfer reactions catalyzed by HIV-1 reverse transcriptase.* Science, 1992. **258**(5085): p. 1112-8.

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

3. Ellison, V. and P.O. Brown, *A stable complex between integrase and viral DNA ends mediates human immunodeficiency virus integration in vitro.* Proc Natl Acad Sci U S A, 1994. **91**(15): p. 7316-20.

4. Bor, Y.C., et al., *Target-sequence preferences of HIV-1 integration complexes in vitro.* Virology, 1996. **222**(1): p. 283-8.

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

6. Chow, S.A., *In vitro assays for activities of retroviral integrase.* Methods, 1997. **12**(4): p. 306-17.

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

8. Bao, K.K., A.M. Skalka, and I. Wong, *Presteady-state Analysis of Avian Sarcoma Virus Integrase. II. Reverse-Polarity Substrates Identify Preferential Processing of the U3-U5 Pair.* J Biol Chem, 2002. **277**(14): p. 12099-108.

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

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

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

12. Deprez, E., et al., *Oligomeric states of the HIV–1 integrase as measured by time–resolved fluorescence anisotropy.* Biochemistry, 2000. **39**(31): p. 9275–84.

13. Jones, K.S., et al., *Retroviral integrase functions as a multimer and can turn over catalytically.* J Biol Chem, 1992. **267**(23): p. 16037–40.

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

15. Pemberton, I.K., M. Buckle, and H. Buc, *The metal ion–induced cooperative binding of HIV–1 integrase to DNA exhibits a marked preference for Mn(II) rather than Mg(II).* J Biol Chem, 1996. **271**(3): p. 1498–506.

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

17. Katz, R.A. and A.M. Skalka, *The retroviral enzymes.* Annu Rev Biochem, 1994. **63**: p. 133–73.

18. Engelman, A., K. Mizuuchi, and R. Craigie, *HIV–1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer.* Cell, 1991. **67**(6): p. 1211–21.

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

20. Skinner, L.M., et al., *Nucleophile selection for the endonuclease activities of human, ovine, and avian retroviral integrases.* J Biol Chem, 2001. **276**(1): p. 114–24.

21. Wong, I., S.S. Patel, and K.A. Johnson, *An induced-fit kinetic mechanism for DNA replication fidelity: direct measurement by single-turnover kinetics.* Biochemistry, 1991. **30**(2): p. 526-37.

22. Cantor, C.R., M.M. Warshaw, and H. Shapiro, *Oligonucleotide interactions. 3. Circular dichroism studies of the conformation of deoxyoligonucleotides.* Biopolymers, 1970. **9**(9): p. 1059-77.

23. Terry, R., et al., *Properties of avian sarcoma-leukosis virus pp32-related pol-endonucleases produced in Escherichia coli.* J Virol, 1988. **62**(7): p. 2358-65.

24. Gill, S.C. and P.H. von Hippel, *Calculation of protein extinction coefficients from amino acid sequence data.* Anal Biochem, 1989. **182**(2): p. 319-26.

25. Ausubel, F.M., *Current protocols in molecular biology.* 1988, New York: Greene Pub. Associates and Wiley-Interscience. <1-2, 4> (loose-leaf).

26. Wong, I. and T.M. Lohman, *A double-filter method for nitrocellulose-filter binding: application to protein-nucleic acid interactions.* Proc Natl Acad Sci U S A, 1993. **90**(12): p. 5428-32.

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

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

29. Vink, C., et al., *Site-specific hydrolysis and alcoholysis of human immunodeficiency virus DNA termini mediated by the viral integrase protein.* Nucleic Acids Res, 1991. **19**(24): p. 6691-8.

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

31. Katz, R.A., K. Gravuer, and A.M. Skalka, *A preferred target DNA structure for retroviral integrase in vitro.* J Biol Chem, 1998. **273**(37): p. 24190–5.

32. Muller, H.P. and H.E. Varmus, *DNA bending creates favored sites for retroviral integration: an explanation for preferred insertion sites in nucleosomes.* Embo J, 1994. **13**(19): p. 4704–14.

33. Shih, C.C., J.P. Stoye, and J.M. Coffin, *Highly preferred targets for retrovirus integration.* Cell, 1988. **53**(4): p. 531–7.

34. Katz, R.A., et al., *Role of DNA end distortion in catalysis by avian sarcoma virus integrase.* J Biol Chem, 2001. **276**(36): p. 34213–20.

35. Muller, B., et al., *Rapid solution assays for retroviral integration reactions and their use in kinetic analyses of wild–type and mutant Rous sarcoma virus integrases.* Proc Natl Acad Sci U S A, 1993. **90**(24): p. 11633–7.

36. Patel, S.S., I. Wong, and K.A. Johnson, *Pre–steady–state kinetic analysis of processive DNA replication including complete characterization of an exonuclease–deficient mutant.* Biochemistry, 1991. **30**(2): p. 511–25.

37. Johnson, K.A., *Rapid quench kinetic analysis of polymerases, adenosinetriphosphatases, and enzyme intermediates.* Methods Enzymol, 1995. **249**: p. 38–61.

38. Johnson, K.A., *Conformational coupling in DNA polymerase information transfer.* Philos Trans R Soc Lond B Biol Sci, 1992. **336**(1276): p. 107–12.

39. Fierke, C.A. and G.G. Hammes, *Transient kinetic approaches to enzyme mechanisms.* Methods Enzymol, 1995. **249**: p. 3–37.

40. Jeggo, P.A., G.E. Taccioli, and S.P. Jackson, *Menage a trois: double strand break repair, V(D)J recombination and DNA– PK.* Bioessays, 1995. **17**(11): p. 949–57.

41. Melek, M., M. Gellert, and D.C. van Gent, *Rejoining of DNA by the RAG1 and RAG2 proteins.* Science, 1998. **280**(5361): p. 301–3.

42. Roth, D.B. and N.L. Craig, *VDJ recombination: a transposase goes to work.* Cell, 1998. **94**(4): p. 411–4.

43. Ramsden, D.A., et al., *Distinct DNA sequence and structure requirements for the two steps of V(D)J recombination signal cleavage.* Embo J, 1996. **15**(12): p. 3197–206.

44. Richards, F.M. and H.W. Wyckoff, in *The Enzymes*, P.D. Boyer, Editor. 1971, Academic Press: New York. p. 647–806.

45. Reid, T. and I. Wilson, in *The Enzymes*, P.D. Boyer, Editor. 1971, Academic Press: New York. p. 373–416.

Chapter 3

REVERSE-POLARITY SUBSTRATES IDENTIFY PREFERENTIAL PROCESSING OF THE U3-U5 PAIR

Kogan K. Bao, Anna Marie Skalka, and Isaac Wong

Published in *Journal of Biological Chemistry*

9650 Rockville Pike

Bethesda, Maryland 20814-3996

Volume 277, Number 14, Issue of April 5, 2002, pp. 12099-12108

SUMMARY

The integrase-catalyzed insertion of the retroviral genome into the host chromosome involves two reactions *in vivo*: 1) the binding and endonucleolytic removal of the terminal dinucleotides of the viral DNA termini and 2) the recombination of the ends with the host DNA. Kukolj and Skalka [1] have previously shown that tethering of the termini enhances the endonucleolytic activities of integrase. We have used 5'−5' phosphoramidites to design reverse-polarity tethers that allowed us to examine the reactivity of two viral long terminal repeat-derived sequences when concurrently bound to integrase and, additionally, developed presteady-state assays to analyze the initial exponential phase of the reaction, which is a measure of the amount of productive nucleoprotein complexes formed during preincubation of integrase and DNA. Furthermore, the reverse-polarity tether circumvents the integrase-catalyzed splicing reaction [2] that obscures accurate analysis of the reactivities of synapsed DNA substrates. Consequently, we were able to establish a lower limit of 0.2 s^{-1} for the rate constant of the processing reaction. The analysis showed the physiologically relevant U3/U5 pair of viral ends to be the preferred substrate for integrase with the U3/U3 combination favored over the U5/U5 pair.

INTRODUCTION

The insertion of a DNA copy of the viral RNA genome into the host chromosome is a critical step in the reproduction cycle of retroviruses [3]. Integrase catalyzes this reaction via a 2-step process, 1) the processing reaction, which is the recognition and endonucleolytic "trimming" of DNA sequences at the two 3'-ends of linear viral DNA and 2) the joining reaction, which is the concerted cleavage-ligation of the processed ends into the host chromosomal DNA. In avian sarcoma virus (ASV), the processing reaction produces site-specific cuts at the CATT sequence of the viral 3'-ends, removing the terminal TT dinucleotide to create new recessed 3'-OH ends. The two 3'-OH groups then serve as nucleophiles in the joining reaction to attack the phosphate bonds of the cellular target DNA in a single-step transesterification to produce a gapped covalent intermediate with 2-nucleotide overhangs. Overhang removal, gap fill-in, and ligation to complete the

integration are likely mediated by host repair mechanisms [4], although participation by viral proteins has been suggested [5–7]. The final product is a 4-base pair shortened viral genome inserted in the host DNA flanked by 6-base pair inverted repeats derived from the six base pairs separating the sites of concerted joining. The CA near the ends of the viral long terminal repeats (LTR) is conserved among retroviruses, whereas the length of the flanking host repeats is virus-specific and is thought to be a structural consequence of a particular integrase [8–11]. For clarity, the remainder of this report will refer to the dinucleotide TT-trimming activity as the "processing" reaction and the subsequent sequence-dependent insertion of processed ends into a double stranded DNA target as the "joining" reaction. The novel recombination activity specific to synapsed DNA substrates [2] will be referred to as the "splicing" reaction.

The U5 and U3 regions in ASV LTRs include two separate nearly perfect inverted repeat sequences in these noncoding sections of the retroviral RNA genome. In the course of retroviral replication, the RNA genome is reverse-transcribed into a duplex DNA copy, and as a result of the strand-transfer mechanism of reverse transcriptase, the U5 and U3 sequences become the termini of the DNA genome [12]. Purified integrase along with Mn^{2+} or Mg^{2+} as a cofactor is sufficient to catalyze both processing and joining reactions in *in vitro* assays using synthetic oligodeoxynucleotide substrates with DNA sequences derived from the U3 and/or U5 viral LTR ends [1, 13, 14].

Typical *in vitro* integrase processing assays use radiolabeled substrates, with the reaction(s) allowed to proceed for 30–90 min before quenching since integrase shows low reactivity in such assays [13, 15, 16]. More recently, Vora and Grandgenett [17] studied the integrase-catalyzed joining reaction using an assay time of 5 min after a period of preincubation, aimed at "more closely reflecting the effects of the initial assembly events." In either case, the reaction products and unreacted substrates are separated by electrophoresis on agarose or denaturing polyacrylamide gels to resolve the dinucleotide-shortened processing products and the extended joining products from the original substrates. Subsequent quantitation of the two products is used to measure the extent of catalytic activity. Single-end substrate assays show poor processing efficiency, and

they also fail to produce measurable amounts of concerted integration products (where, in the case of ASV, two viral ends are inserted six base pairs apart) as is observed with preintegration complexes purified from infected cells [18, 19]. Consequently, products larger than the original substrate in these assays that use a substrate containing only one LTR end-derived sequence are considered to be products of only half-reactions, as only one viral DNA end is joined to a target site [19]. Assays involving substrates designed to resemble the linear retroviral genome, with an LTR end-derived sequence at either terminus, have resulted in >95% of the joining products coming from the recombination of two separate substrate molecules integrating into a single target molecule [14, 20, 21]. Despite the lack of similarity to the nucleoprotein complex assembled *in vivo*, the assays involving such trimolecular reactions have been used to suggest the preference of integrase for a U3/U3 combination of ends over that of a U5/U5 combination, with the specificity for the biologically significant U3/U5 arrangement intermediate between the two [14]. More recently, Brin and Leis [22], using a reconstituted HIV-1 integration system, demonstrated that concerted DNA integration requires the presence of both U3 and U5 ends in the donor DNA. DNase I footprint analysis of the assembled integrase-DNA complex required for integration has revealed that the region of protection at the U3 LTR end (~20 base pairs) was at least twice that at the U5 LTR end (<10 base pairs), suggesting that the nucleoprotein complex is asymmetric when assembled in a fashion capable of full-site integration [17].

In an attempt to mimic the geometric organization of the viral LTR ends of the *in vivo* preintegration complex at a more molecular level, Kukolj and Skalka [1] designed a series of substrates that covalently linked two single-end substrates together in a head-to-head configuration using 1−3 nucleotides of single-stranded DNA (see Fig. 3.1*A*). It was hypothesized that these single-stranded tethers would provide sufficient flexibility to alleviate torsional or rotational strains arising from the structural alignment of the two viral ends bound within the integrase active site(s). By designing the substrates asymmetrically with respect to the length of the two ends and 5' radiolabeling both ends, it was possible to quantitate processing products for both ends simultaneously in addition to what appeared to

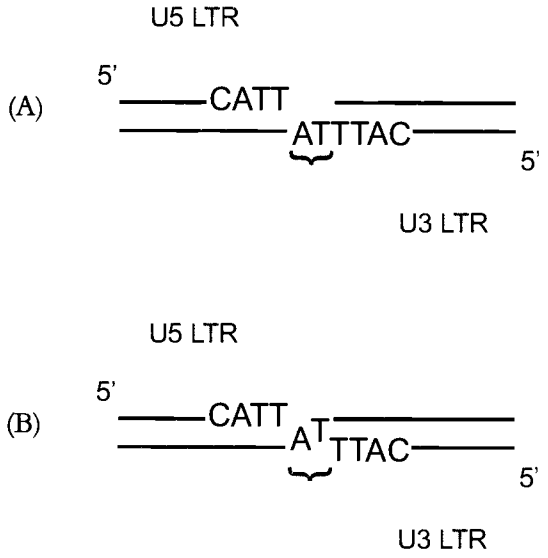

Figure 3.1: Substrate diagram. Dual-cognate site synapsed substrate designed with sequences from the termini of the ASV U5 and U3 LTR sequences attached via a 2-nucleotide (TA) single-stranded tether (indicated by a *bracket*). *A*, a normal-polarity substrate. *B*, a reverse-polarity substrate.

be extended joining products. Using *in vitro* integrase assays similar to those described above, these authors observed enhanced processing efficiencies with these synapsed-end substrates and concluded that the tether effectively brought together integrase subunits bound separately to the two cognate sites, thereby coordinating the formation of a requisite higher order oligomeric structure with enhanced activity. These observations were consistent with the suggestion by Murphy and Goff [23] that integrase must recognize both DNA ends for efficient processing at either end to occur *in vivo*.

Whereas the results with these synapsed substrates clearly illustrated the important relationship between assembly of an integrase multimer and the coordinated binding of both viral DNA ends, the assays were performed in the time regime where the enzyme-catalyzed reaction had undergone multiple turnovers. Although much effort has been expended to demonstrate that integrase functions as a true enzyme in being able to catalyze multiple turnovers under steady-state conditions [24], the physiological relevance of multiple turnover events is questionable considering that only a single round of catalysis is sufficient to achieve integration *in vivo*.

To complete a first-turnover investigation of the processing reaction using synapsed-end substrates, 5'−5' reverse-polarity substrates were designed (see Fig. 3.1*B*) that allow the simultaneous binding of two LTR ends at the active site. Additionally, these substrates were not susceptible to the integrase-catalyzed splicing reaction and consequently simplified the comparisons of the LTR ends. Analysis of presteady-state assays of these reverse-polarity substrates revealed that, although the U3 LTR sequence appears to be preferred by avian integrase when the termini are studied individually [25−28], the U3/U5 combination of retroviral ends is the preferred substrate for integrase microscopically within a single turnover. Results from the first-turnover exponential analyses also have important bearing on previous interpretations of multiple-turnover product distributions. An examination of the exponential phases and their usefulness to further understanding of the mechanism of integrase activity and the binding of substrates is also discussed.

EXPERIMENTAL PROCEDURES

Synthetic Oligodeoxyribonucleotides

Oligodeoxyribonucleotides were synthesized by the Center for Gene Research and Biotechnology Central Services Laboratory (Oregon State University). Reverse-polarity oligodeoxyribonucleotides were synthesized using 5'-β-cyanoethyl phosphoramidites (Glen Research, Sterling, VA). Concentrations were determined spectrophotometrically in Tris-EDTA using the calculated extinction coefficients at 260 nm [29] listed in Table I.

The naming convention used for annealed DNA substrates is as follows. Strands with sequences derived from the U5 and U3 ends of the ASV genome are designated with a "5" and "3," respectively; strands of duplex DNA containing ASV integrase cognate sequence, CATT, are designated with a "t"; strands containing the complementary GTAA sequence, are designated with a "b"; synapsed strands are designated with the length of the tether within parentheses; duplexes are denoted as the concatenation of the names of the component single-stranded oligodeoxyribonucleotides separated by slashes (/); sequences 5'-end-radiolabeled with ^{32}P will be specified in the text with an asterisk (*) at the beginning.

Quantitation

The intensity of each product band, $I_i(t)$, at each time, t, was first normalized with respect to the sum of intensities in the starting substrate band, $I_0(t)$, plus all product bands according to Equation 1 to determine the normalized product fraction $F_{i,\,norm}(t)$.

$$F_{i,norm}(t) = \frac{I_i(t)}{I_0(t) + \sum_{i=1}^{n} I_i(t)} \qquad \text{(Eq. 1)}$$

$F_{i,\,norm}$, was then corrected for background intensity present at t=0 for the ith band and renormalized for background intensities of all product bands to obtain the final corrected product fraction, $F_{i,\,corr}$ according to Equation 2.

Table 3.1: Nomenclature and μM Extinction Coefficients (μM^{-1} cm^{-1}) at 260 nm of DNA Substrates

Name[a]	Sequence	ε_{260}
5t	5'-GCTGAAGCAGAAGGCTTCATT-3'	0.20
5b	5'-AATGAAGCCTTCTGCTTCAGC-3'	0.19
3t	5'-GCTATTGCATAAGACTACATT-3'	0.21
3b	5'-AATGTAGTCTTATGCAATAGC-3'	0.21
5b(2)3t	5'-GCTATTGCATAAGACTACATTTAAATGAAGCCTTCTGCTTCAGC-3'	0.43
5b(2)3b	3'-CGATAACGTATTCTGATGTAAT-5'-5'-AAATGAAGCCTTCTGCTTCAGC-3'	0.43
5b(2)5b	3'-CGACTTCGTCTTCCGAAGTAAA-5'-5'-TAATGAAGCCTTCTGCTTCAGC-3'	0.41
3b(2)3b	3'-CGATAACGTATTCTGATGTAAT-5'-5'-AAATGTAGTCTTATGCAATAGC-3'	0.44

[a]The following is the naming convention used for annealed DNA substrates. Strands with sequences derived from the U5 and U3 ends of the ASV genome are designated with a "5" and "3," respectively. Strands containing the cognate sequence, CATT, are designated with a "t." Strands containing the complementary GTAA sequence are designated with a "b." Synapsed strands are designated with the length of the tether within parentheses. Duplex names consist of a concatenation of the names of all oligodeoxyribonucleotide strands annealed separated by slashes (/). Sequences labeled with ^{32}P will be represented in the text with an asterisk (*) to denote 5'-end radiolabeling.

$$F_{i,corr}(t) = \frac{F_{i,norm}(t) - F_{i,norm}(0)}{1 - \sum_{i=1}^{n} F_{i,norm}(0)} \quad \text{(Eq. 2)}$$

Experimental time courses were fitted to Equation 3 consisting of n exponential terms, with amplitudes A_i and apparent rate constants λ_i plus a linear term with an apparent rate constant λ_{lin} to fit to the linear portion of the ensuing exponential phase.

$$y = \sum_{i=1}^{n} A_i(1 - e^{-\lambda_i t}) + \lambda_{lin} t \quad \text{(Eq. 3)}$$

Non-linear least squares fittings were performed using Kaleidagraph software (Synergy, Redding, PA).

Reagents, Buffers, Purification of Oligodeoxyribonucleotides, 5'^{32}P Labeling, Pre-Steady State Assays, Product Analysis by Denaturing Acrylamide Gel Electrophoresis, and ASV Integrase Overexpression and Purification

These materials and protocols were identical to those described in detail in the first paper of this series [2].

RESULTS

Effect of Reverse-Polarity Tethers on Splicing Activity

The normal-polarity substrate, used originally to develop the presteady-state assay for integrase [2], positioned the cognate CATT of the U3 sequence internally (see Fig. 3.1*A*). As a result, this sequence became the site of a spurious site-specific splicing reaction [2]. Because the site preference of the splicing reaction is identical to that of the processing reaction, *i.e.* the internal CATT, its presence interfered with accurate quantitation of enzymatic processing activity. A reverse-polarity substrate (Fig. 3.1*B*), incorporating a 5'–5' reverse-polarity tether and containing the same sequence as that of the normal-polarity substrate (Table 3.1), was therefore designed specifically to circumvent the splicing reaction while maintaining the advantages of tethering the viral LTR sequences (Fig. 3.1*B*).

At high concentration of NaCl (400 mM), the activity of integrase with normal-polarity substrates was predominated by the splicing activity (Fig. 3.2*A*). To

demonstrate that the reverse-polarity substrate is not susceptible to this splicing reaction, presteady-state assays comparing different DNA substrates with and without a tether were performed at high salt conditions favorable to the splicing reaction. Reactions were performed as described under "Experimental Procedures," with all substrates radiolabeled on *5t, a 21-mer that is endonucleolytically cleaved by integrase at the normal processing, minus 2 position and at a second, minus 3 position to yield radiolabeled 19- and 18-mer products. The splicing reaction would yield a 46-mer product with the normal-polarity substrate *5t/5b(2)3t/3b and an anticipated 23/24-mer with the reverse-polarity substrate *5t/5b(2)3b/3t.

Fig. 3.2*A* shows that the major product, at 400 mM NaCl, with the normal-polarity substrate was the 46-mer splicing product, while the 19-mer processing and shorter products were present in minor but detectable amounts. In contrast, reactions performed under identical conditions with the reverse-polarity substrate, *5t/3b(2)5b/3t, yielded predominantly processing products (19- and 18-mers) (Fig. 3.2*B*). Integrase-mediated splicing activity of this reverse-polarity substrate would have resulted in +2 and +3 products of 23- and 24-nucleotide lengths, respectively; products of such sizes accumulated in minute amounts and only at the two longest reaction times examined (1200 s and 1800 s). By comparison, the processing products appeared within 5−10 s and in much greater quantities. These data demonstrated that even under reaction conditions chosen specifically to favor splicing over processing, the 5'−5' tethering preferentially mediates integrase-catalyzed processing.

To rule out the possibility that the increased NaCl concentration itself was capable of having promoted the initial exponential phase of processing activity with the reverse-polarity substrate, a control experiment was performed with an unsynapsed, single-ended substrate *5t/5b. Fig. 3.2*C* shows clearly the lack of any detectable integrase activity with this substrate under these conditions. Even at the longest time point assayed, 1800 s, *5t/5b was minimally processed. These results showed that without the scaffolding provided by the tethering of the two ends, integrase was unable to utilize the cognate sequence for enzymatic activity at this higher NaCl concentration.

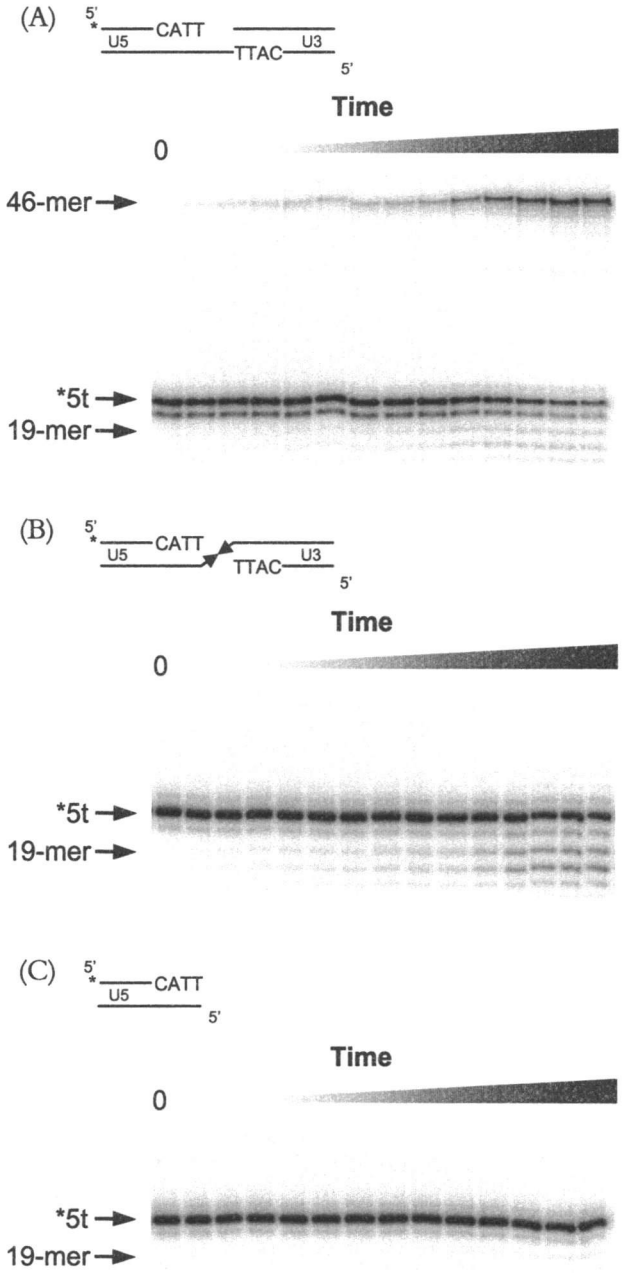

Figure 3.2: Reactivity of substrates at conditions favoring the splicing reaction. Single-turnover activity assay at 5.0 μM integrase, 0.5 μM DNA, 400 mM NaCl with samples subjected to electrophoresis on a 20% polyacrylamide, 8 M urea, TBE sequencing gel. *A*, for *5t/5b(2)3t/3b, lanes represent reaction times of 0, 5.8, 11.1, 16.7, 22.4, 28.4, 34.5, 45.4, 60, 120, 245, 600, 1283, and 1800 s. *B*, for *5t/5b(2)3b/3t, lanes represent reaction times of 0, 5.8, 10.8, 15.9, 21.1, 26.0, 31.2, 46.7, 60, 120, 240, 601, 1200, and 1800 s. *C*, for *5t/5b, lanes represent reaction times of 0, 5.5, 10.1, 14.9, 25.5, 30.8, 45.2, 120, 243, 300, 582, 1199, and 1800 s.

Effect of Tethering on Reactivity

Fig. 3.3*A* shows the results from a typical first-turnover experiment performed as described under "Experimental Procedures." Assay results with synapsed 5t/5b(2)3b/*3t and unsynapsed *3t/3b substrates (see Table I) are shown to illustrate the effect of tethering two viral end sequences in a head-to-head manner via a 5'−5' linkage. The radiolabeled 21-mer, *3t, with a sequence corresponding to the viral U3 LTR, used in both substrates, contained the 3' terminal cognate sequence CATT-OH. In the processing reaction, this sequence was endonucleolytically cleaved by integrase at the minus 2 position to yield a 19-mer and a TT dinucleotide as the major products. Additionally, an 18-mer minor product was also observed that is consistent with the less specific processing activity at the minus 3 position observed when Mn^{2+} is used as the metal cofactor [19, 30, 31]. Both 19- and 18-mer products were summed and plotted to quantitate total integrase-catalyzed processing activity.

Kinetic experiments revealed multi-phasic reactions at all integrase concentrations tested. The time-dependent appearance of the processing product was best described by a series of presteady-state exponential phases followed by a linear phase according to Equation 3. The number of exponential phases observed was dependent on both the substrate used and the duration of time course examined. For all time courses, we have chosen to report the model-independent, actual molar concentration of product observed without further interpretation for accuracy.

Compared to an unsynapsed substrate (*3t/3b), the initial exponential product formation phase of the synapsed substrate (5t/5b(2)3b/*3t) reaction occurred with larger amplitude and a significantly faster rate. The difference in burst amplitudes and rate constants for the synapsed *versus* unsynapsed substrates directly reflected differences in the extent of productive complex formation between enzyme and DNA during the preincubation period (before the initiation of the reaction with Mn^{2+}). Furthermore, the synapsed substrate contained two CATT-OH cognate sites at which processing could have occurred, one at the 3'-end of the 3t strand and one at the 3'-end of the 5t strand. Because only the 3t strand was radiolabeled, however, processing at the cognate site of the U5

Figure 3.3: Processing reactivity of a synapsed substrate *versus* an unsynapsed substrate. *A*, first-turnover processing activity assay at 2 μM integrase, 0.5 μM DNA, 130 mM NaCl with aliquots withdrawn and quenched after addition of $MnCl_2$ at various time points. Samples were then subjected to electrophoresis on a 20% polyacrylamide, 8 M urea, TBE sequencing gel. To assess the effect of tethering, the amount of processing products using synapsed 5t/5b(2)3b/*3t (*closed circles*) and unsynapsed *3t/3b (*closed squares*) as substrates was quantified using a PhosphorImager and the results plotted. The curves represent best fit of the data to two exponentials as described under "Experimental Procedures" with $A_1 = 0.0230 \pm 0.0002$ μM, $\lambda_1 = 0.189 \pm 0.008$ s^{-1}, $A_2 = 0.027 \pm 0.001$ μM, $\lambda_2 = 0.0027 \pm 0.0003$ s^{-1} (*dashed line*) and $A_1 = 0.0076 \pm 0.0003$ μM, $\lambda_1 = 0.069 \pm 0.005$ s^{-1}, $A_2 = 0.036 \pm 0.001$ μM, $\lambda_2 = 0.0020 \pm 0.0001$ s^{-1} (*solid line*). A series of first-turnover processing activity assays, performed as above, at 0.5 μM DNA with integrase concentration ranging from 0–5.0 μM. The data were fit to Equation 3 and the best fit exponential amplitudes (*B*) and exponential rates of product formation (*C*) of the initial phase plotted for 5t/5b(2)3b/*3t (*closed circles*) and unsynapsed *3t/3b (*closed squares*). The *solid lines* represent linear fits of the amplitude and rate data with slopes of 0.0049 ($r = 0.99$) and 0.0287 ($r = 0.99$) respectively. The *dotted lines* connect the data points for ease of comparison.

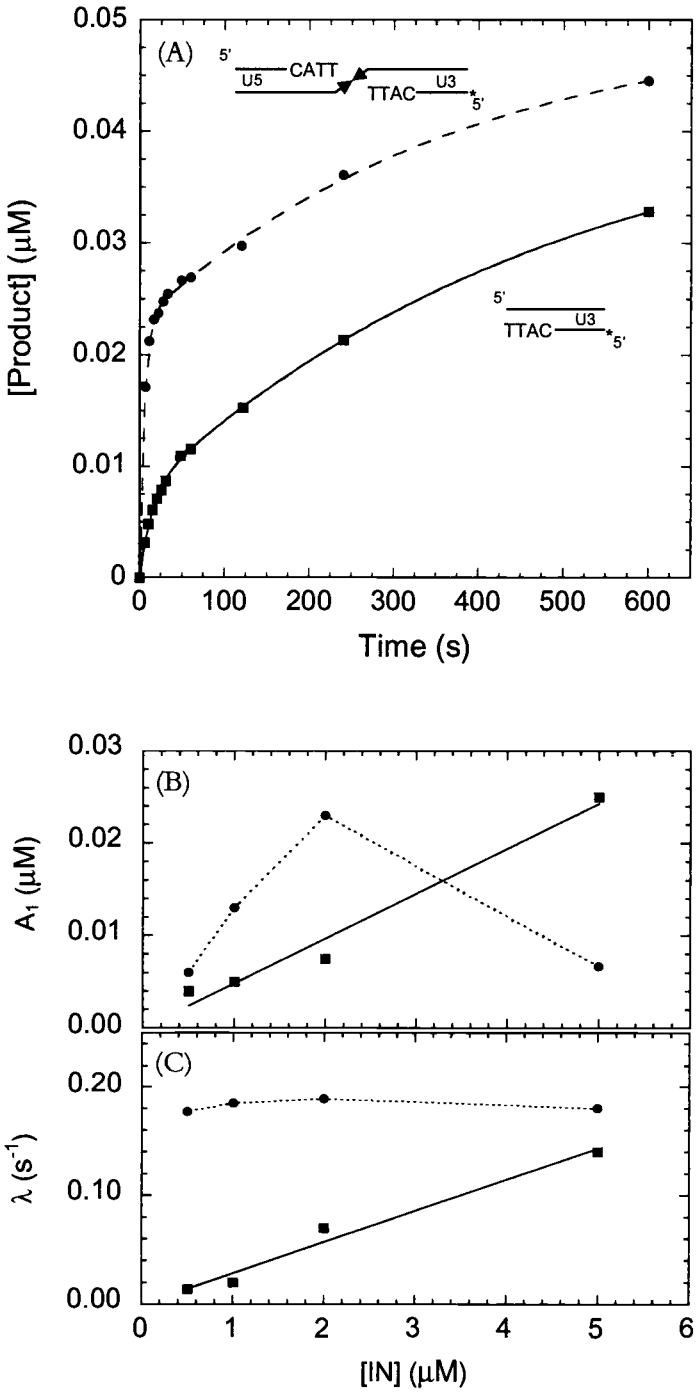
(A)
5'
U5
CATT
TTAC
U3
*5'
5'
TTAC
U3
*5'
[Product] (μM)
Time (s)
(B)
A1 (μM)
(C)
λ (s-1)
[IN] (μM)

Figure 3.3

sequence was silent in these assays. As a consequence, the amplitude measured for the synapsed substrate did not include the amount of productive complexes formed at the U5 cognate site, and the large difference in amplitudes observed in this experiment actually underestimated the true enhancement of productive nucleoprotein complex formation attributable to the tethering of the two viral LTR ends. At time points extended beyond the early exponential phases, both time courses paralleled each other for the synapsed and unsynapsed substrates, indicating similar rates of reaction in subsequent steps of the enzyme. Experiments performed using radiolabeled *5t showed similar but lower reactivity under the same conditions.

To assess the effect of enzyme concentration on the differences in reactivities of the two substrates, titrations of enzyme concentration were performed for both substrates. Typical results are shown in Figs. 3.3, *B* and *C*, for assays conducted at integrase concentrations varying from 0.5 to 5.0 μM and a substrate concentration of 0.5 μM. The data were fit to Equation 3, and the best fit amplitudes and rate constants of the initial phase were plotted as a function of integrase concentrations. In the case of *3t/3b (*closed squares, solid lines*), both the apparent rate constant, λ_1, and amplitude, A_1, increased linearly with increasing integrase concentration with slopes of 0.0287 ($r = 0.99$) and 0.0049 ($r = 0.99$), respectively. In contrast, although the amplitude of the initial exponential phase in the synapsed substrate reaction increased with increasing protein concentration (with the exception of the highest protein concentration where protein aggregation becomes a problem), the rate constant for the initial exponential phase remained 0.2 s^{-1}, independent of integrase concentration for the range of concentrations tested. In contrast, the single-end substrate only approached this rate at the highest integrase concentration range. These data show that the lower limit of the rate constant for the processing reaction is 0.2 s^{-1}.

Effect of U3 and U5 Sequences on Processing and Joining

The reverse-polarity substrates allowed exclusive examination of the processing activity and, therefore, made possible the direct comparison of the reactivities of the U5 and U3 sequences. The reactivities of all four possible combinations of synapsed end sequences, U3/U3, U3/U5, U5/U5, and U5/U3, were separately measured using *3t/3b(2)3b/*3t,

5t/5b(2)3b/*3t, *5t/3b(2)5b/*5t, and *5t/5b(2)3b/3t, respectively. The results from presteady-state assays, as described under "Experimental Procedures," were obtained and compared over a range of both NaCl (130–500 mM) and integrase (0.5–20 μM) concentrations. Fig. 3.4*A* shows a typical direct comparison of the time-dependent appearance of processing products of the U3 end in the context of a synapsed U3/U3 pair (*open circles*, *3t/3b(2)3b/*3t) *versus* that of the same U3 sequence in a synapsed U3/U5 combination (*closed circles*, 5t/5b(2)3b/*3t). At 5 μM integrase, 0.5 μM reverse-polarity substrate DNA, and 130 mM NaCl, the best fits of the time-dependent appearance of processing products were characterized by two initial exponential phases followed by a slower linear phase representing the beginning of a third exponential phase. However, because the substrate used to assay the U3/U3 combination, *3t/3b(2)5b/*3t, contained two radiolabeled 21-mers whereas 5t/5b(2)3b/*3t was radiolabeled on only one DNA strand, it was necessary to divide the best fit of the U3/U3 data by two. Comparison of the adjusted curve (*dotted curve*) with the best fit of the data from the U3/U5 substrate (*solid curve*) revealed that the U3 end is processed to a greater extent when bound concurrently with a U5 end than when bound with another U3 sequence.

Similarly, Fig. 3.4*B* shows the results from parallel presteady-state experiments comparing the reactivity of the U5 end in the context of either a U5/U5 pair (*open squares*, *5t/3b(2)5b/*5t) or a U5/U3 combination of ends (*closed squares*, *5t/3b(2)5b/3t). As described above, the signal measured using *5t/3b(2)5b/*5t, the substrate used to assay the U5/U5 pair, represented twice the reactivity of a single U5 end because the substrate contained two radiolabeled 21-mer strands (*5t). Again, the LTR sequence bound by integrase in an asymmetrical combination (U5/U3, *solid line*) was more reactive than the same sequence concurrently bound with an identical sequence (U5/U5, *dotted line*) after the adjustment of the U5/U5 data by a factor of 0.5.

More dramatic results were observed when assays using the same substrates were performed at higher concentrations of integrase. However, because of the poor solubility of integrase at low NaCl concentrations [32], it was necessary to increase the NaCl concentration to 400 mM NaCl to enable assays to be performed at increased

Figure 3.4: Processing reactivity of integrase with different combinations of LTR ends. Single-turnover splicing activity assays at 5 μM integrase, 0.5 μM DNA substrate, 130 mM NaCl with samples subjected to electrophoresis in a 20% polyacrylamide, 8 M urea, TBE sequencing gel. *A*, a plot of the results from substrates radiolabeled at the *3t strand shows increased processing for the asymmetric U3/U5 (5t/5b(2)3b/*3t, *closed circles*) combination over that of the U3/U3 (*3t/3b(2)3b/*3t, *open circles*) after the best fit of the U3/U3 data has been multiplied by 0.5 (*dotted curve*) to allow for accurate comparison. The lines represents best fit of the data to Equation 3 with A_1= 0.0106 ± 0.0005 μM, λ_1 = 0.19 ± 0.02 s^{-1}, A_2 = 0.037 ± 0.003 μM, λ_2 = 0.0085 ± 0.0009 s^{-1}, λ_{lin} = (7.0 ± 0.5) × 10^{-5} s^{-1} (*dashed line*) and A_1 = 0.0067 ± 0.0005 μM, λ_1 = 0.19 ± 0.02 s^{-1}, A_2 = 0.016 ± 0.002 μM, λ_2 = 0.011 ± 0.002 s^{-1}, λ_{lin} = (8.9 ± 0.4) × 10^{-5} s^{-1} (*solid line*). *B*, a plot of the results from substrates radiolabeled at the *5t strand shows increased processing for the asymmetric U5/U3 (*5t/5b(2)3b/3t, *closed squares*) combination over that of the U5/U5 (*5t/5b(2)5b/*5t), *open squares*) after the best fit of the U5/U5 data has been multiplied by 0.5 (*dotted curve*) to allow for accurate comparison. The curves represents best fit of the data to Equation 3 with A_1 = 0.0020 ± 0.0003 μM, λ_1 = 0.19 ± 0.02 s^{-1}, A_2 = 0.013 ± 0.004 μM, λ_2 = 0.006 ± 0.002 s^{-1}, λ_{lin} = (4.8 ± 0.7) × 10^{-5} s^{-1} (*dashed line*) and A_1= 0.0017 ± 0.0002 μM, λ_1 = 0.19 ± 0.02 s^{-1}, A_2 = 0.0063 ± 0.0006 μM, λ_2 = 0.012 ± 0.002 s^{-1}, λ_{lin} = (7.8 ± 0.1) × 10^{-5} s^{-1} (*solid line*).

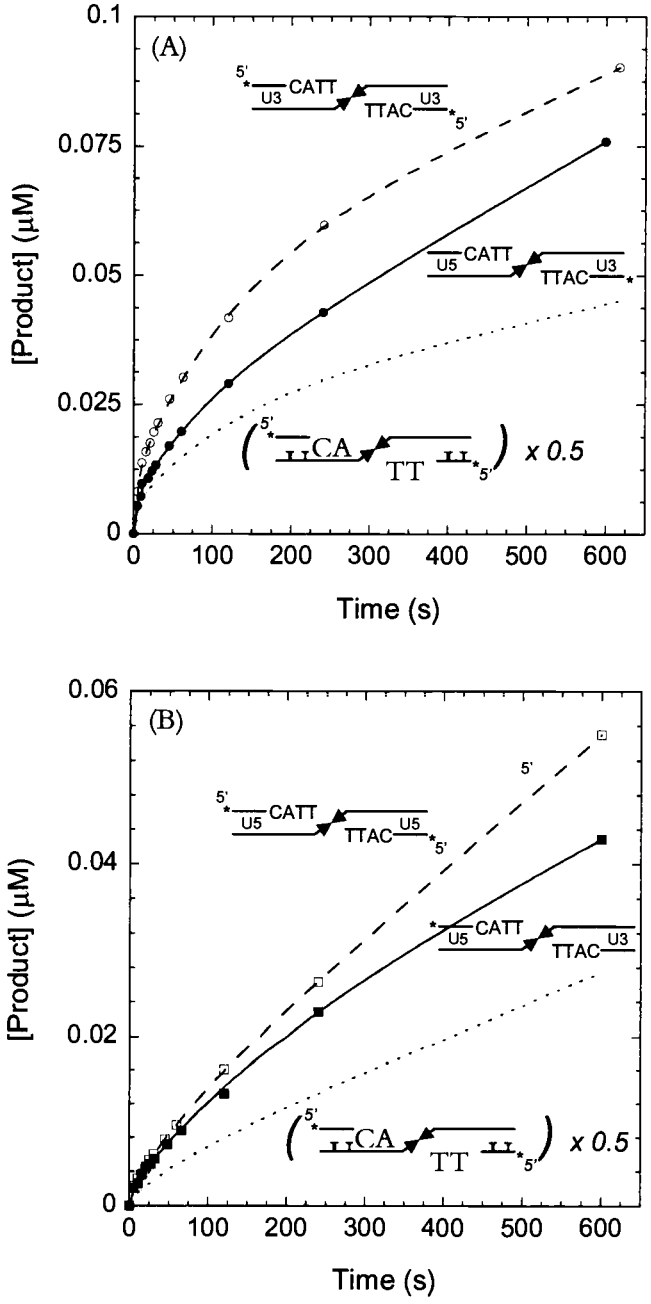
(A)
[Product] (μM)
Time (s)
0.1
0.075
0.05
0.025
0
0
100
200
300
400
500
600
5'
U3
CATT
TTAC
U3
5'
U5
CATT
TTAC
U3
5'
CA
TT
5'
x 0.5
(B)
[Product] (μM)
Time (s)
0.06
0.04
0.02
0
5'
U5
CATT
TTAC
U5
5'
5'
U5
CATT
TTAC
U3
5'
CA
TT
5'
x 0.5

Figure 3.4

Figure 3.5: Processing reactivity of integrase with different combinations of LTR ends at high NaCl concentrations. Single-turnover splicing activity assays at 20 μM integrase, 0.5 μM DNA substrate, 400 mM NaCl with samples subjected to electrophoresis in a 20% polyacrylamide, 8 M urea, TBE sequencing gel. *A*, a plot of the results from substrates with both U3 and U5 ends present shows increased processing at the U3 end, 5t/5b(2)3b/*3t (*closed circles*), over the U5 end *5t/5b(2)3b/3t (*closed squares*). The lines represents best fit of the data to Equation 3 with $A = 0.057 \pm 0.001$ μM, $\lambda = 0.033 \pm 0.002$ s^{-1}, $\lambda_{lin} = (7.7 \pm 0.4) \times 10^{-5}$ s^{-1} (*solid line*) and $A = 0.027 \pm 0.001$ μM, $\lambda = 0.030 \pm 0.001$ s^{-1}, $\lambda_{lin} = (8.1 \pm 0.2) \times 10^{-5}$ s^{-1} (*dashed line*). *B*, a plot of the results with substrates with either U3 or U5 ends shows increased processing for a substrate with only the U3 sequence, *3t/3b(2)3b/*3t (*closed circles*), over a substrate with only the U5 sequence, *5t/5b(2)5b/*5t (closed squares). The curves represent best fit of the data to Equation 3 with $A = 0.046 \pm 0.001$ μM, $\lambda = 0.034 \pm 0.001$ s^{-1}, $\lambda_{lin} = (6.4 \pm 0.3) \times 10^{-5}$ s^{-1} (*solid line*) and $A = 0.028 \pm 0.001$ μM, $\lambda = 0.015 \pm 0.001$ s^{-1}, $\lambda_{lin} = (5.7 \pm 0.2) \times 10^{-5}$ s^{-1} (*dashed line*). Additionally, the fits of the data from Figure 4*A* are summed (*dotted lines*) to allow direct comparison of the processing reactivity of an asymmetric substrate *versus* that of the symmetric substrates.

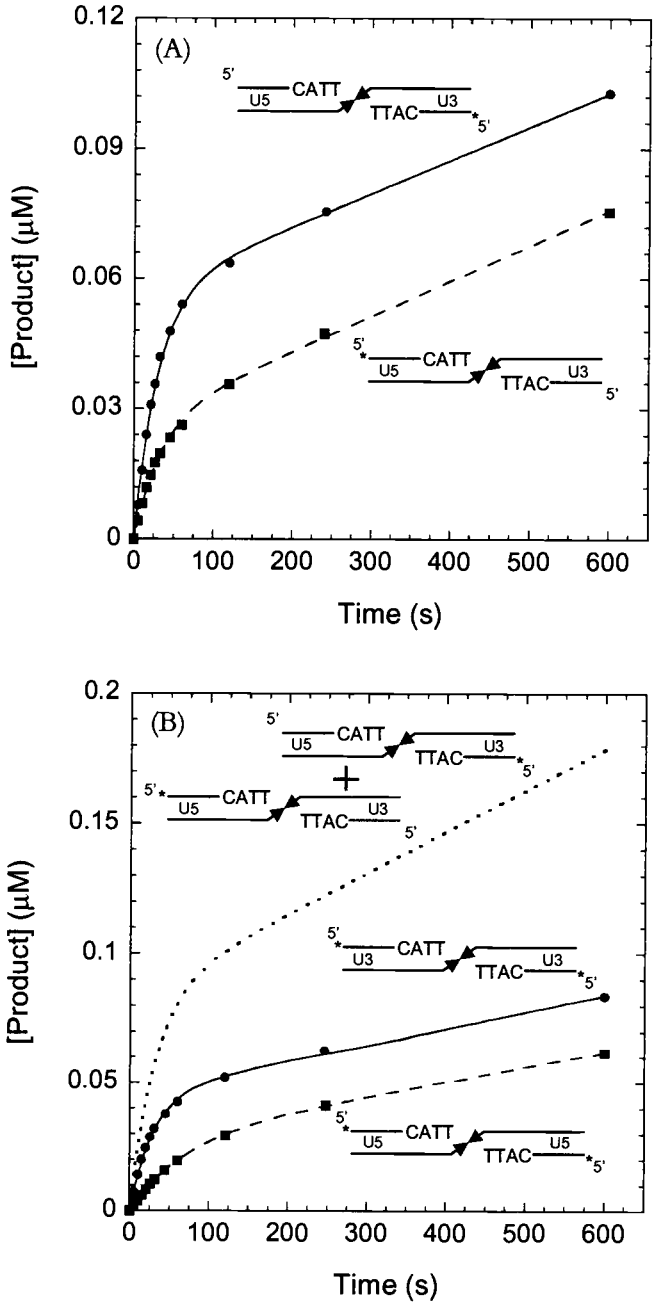
(A)
5'
U5
CATT
TTAC
U3
*5'
5'
*
U5
CATT
TTAC
U3
5'
[Product] (μM)
0.12
0.09
0.06
0.03
0
0
100
200
300
400
500
600
Time (s)
(B)
5'
U5
CATT
TTAC
U3
*5'
+
5'*
U5
CATT
TTAC
U3
5'
5'
*
U3
CATT
TTAC
U3
*5'
5'
*
U5
CATT
TTAC
U5
*5'
[Product] (μM)
0.2
0.15
0.1
0.05
0
0
100
200
300
400
500
600
Time (s)

Figure 3.5

concentrations of integrase (up to 20 μM). Fig. 3.5 shows a comparison of these substrates assayed at 20 μM integrase and 400 mM NaCl. To determine the relative reactivities of the two LTR termini when bound in the physiologically relevant asymmetric combination, the substrate containing both U3 and U5 ends, 5t/5b(2)3b/3t, was assayed with radiolabeling of either the 3t or 5t DNA strand. The results (Fig. 3.5*A*) were similar to those observed at the lower salt/integrase concentrations in that the U3-derived side of the substrate (*closed circles*) had greater reactivity than did the U5-derived side (*closed squares*). Although the initial rate of product formation was similar for the two ends, nearly twice the amount of U3 ends were processed during the initial exponential phase relative to the U5 end of the same substrate.

To examine if this preference in favor of U3 processing is an intrinsic property of the U3 sequence and to verify the results obtained at lower NaCl/integrase concentrations, additional experiments were performed to compare the reverse-polarity substrates containing either two synapsed U3 ends, *3t/3b(2)3b/*3t, or two synapsed U5 ends, *5t/5b(2)5b/*5t . As expected, Fig. 3.5*B* shows that the substrate which contained two synapsed U3 sequences (*closed circles, solid line*) was processed to a greater extent than was the substrate that contained two synapsed U5 sequences (*closed squares, dashed line*). However, to compare these results to those in Fig. 3.5*A*, it was necessary to sum the best fits from the data of the U3/U5 (5t/5b(2)3b/*3t) and U5/U3 (*5t/5b(2)3b/3t) substrates to determine the total reactivity of the U3 and U5 termini when they are concurrently bound. This adjustment was necessary because the radiolabeling of these substrates at either end produced different results (Fig. 3.5*A*) representing the reactivity at only one of the two available cognate sites, whereas the data shown in Fig. 3.5*B* provide a measurement of the reactivity of both available cognate sites within those particular substrates. When the adjusted curve (Fig. 3.5*B*, *dotted curve*) was used for comparison, a surprising and dramatic result was obtained; the combined U3 and U5 amplitudes of a substrate containing both LTR ends greatly exceeded the amplitudes for substrates containing either U3 or U5 sequences exclusively. The amount of reactivity at the cognate sites in decreasing order is: U3 when synapsed with U5, U5 when synapsed with U3, U3 when synapsed with another

U3, and U5 when synapsed with another U5. These results are similar to those obtained from data at lower NaCl/integrase concentrations (Fig. 3.4), however, the reactions at higher NaCl/integrase concentrations displayed increased reactivities for all substrates and enhanced the differences between the LTR ends.

The results at 400 mM NaCl were optimal for demonstrating these preferences; however, similar results, albeit less dramatic, were observed at all NaCl conditions examined. Fig. 3.6 shows polyacrylamide gels of reaction aliquots from assays performed at 130 mM NaCl with the U3/U5 (5t/5b(2)3b/*3t, *A*) and U5/U3 (*5t/5b(2)3b3t, *B*) substrates. As expected, the U3 end was processed to a greater extent relative to the U5 end. Interestingly, a ladder of bands (indicated by a "J" in Fig. 3.6) larger than the starting material appeared after the initial appearance of the processing products. These larger products were present in a broad distribution of sizes, suggesting that they were joining products based on the observation that the integrase-catalyzed joining reaction lacks much sequence specificity [20, 33, 34]. Unfortunately, the distribution of the radioactive label over a broad range of product sizes prevented accurate quantitation of the individual bands. However, a qualitative visual inspection of the gels revealed the rapid appearance of U3 joining products within 15 s and a delayed onset of U5 joining products for 120 s even though the formation of both U3 and U5 processing products occurred within the first 5-10 s of the reaction. These results extend previous reports that integrase favors U3 over U5 [14, 25], and illustrate the importance of having both U3 and U5 sequences present for the correct assembly of an integrase-LTR productive complex. This suggests that after processing, the joining activity of integrase is ordered such that the U3 end of the viral LTR is inserted prior to the U5 LTR.

DISCUSSION

In this work, we report the design, synthesis, and use of reverse-polarity synapsed substrates in presteady-state analysis of the ASV integrase reaction mechanism. Both unsynapsed and reverse-polarity synapsed substrates displayed multiple exponential phases

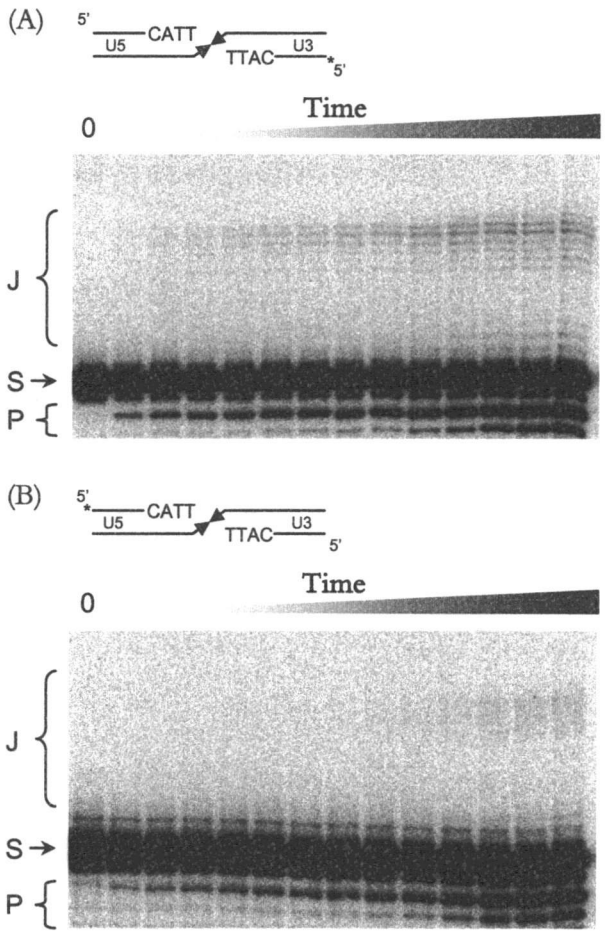

Figure 3.6: Reactivity of synapsed substrates at 130 mM NaCl. Single-turnover activity assay at 5.0 μM integrase, 0.5 μM DNA, 130 mM NaCl samples were subjected to electrophoresis on a 20% polyacrylamide, 8 M urea, TBE sequencing gel. An *S* indicates the starting 21-mer radiolabeled strand of the substrate; a *P* indicates the shortened processing products; extended joining products are indicated with a *J*. *A*, with 5t/5b(2)3b/*3t as the substrate, lanes represent reaction times of 0, 5.2, 9.6, 14.0, 18.5, 23.4, 28.4, 45.1, 61, 120, 241, 600, 1200, and 1800 s. *B*, with *5t/5b(2)3b/3t as the substrate, lanes represent reaction times of 0, 5.4, 10.7, 16.3, 21.4, 27.3, 32.1, 48.3, 66, 120, 240, 600, 1203, and 1800 s.

leading to the initial formation of the dinucleotide-shortened processing product. Analysis of the earliest/fastest exponential phase revealed that the reverse-polarity tethering of viral LTR ends led to a faster rate and greater amplitude, consistent with an increase of productive complex formation. In addition to the benefits of a faster exponential rate and greater amplitude, the reverse-polarity substrate enabled comparative examination of combinations of LTR sequences when concurrently bound to integrase. The results showed integrase to have a preference for an asymmetric pair of ends (U3/U5). The mechanistic implications of these findings will be discussed.

First-Turnover Kinetics of Integrase-Mediated Processing

In presteady-state analyses, the enzyme is treated as a reactant and is used at concentrations comparable to or in excess over that of the substrate. Under first-turnover conditions, experiments are performed with enzyme in excess over substrate to allow the direct observation of the conversion of substrates to intermediates and products through a single catalytic cycle along the reaction pathway to the release of product. Preincubation circumvents the effect of the substrate binding and product release rates from the initial appearance of product and allows measurement of the slowest chemistry or conformational change step. In contrast, steady-state kinetics can define only the rate of conversion of bulk substrate to product as a function of a catalytic quantity of enzyme. Because the products measured are time-averaged over many catalytic turnovers, in steady-state experiments it is generally difficult to resolve intermediates and, thus, the chronology of events at the catalytic site.

Because of solubility and aggregation difficulties with the enzyme in the presence of DNA, it was not possible to determine the reaction stoichiometry in active site titration experiments (for an example of such experiments, see Ref. [35]). Accordingly, we were unable to unambiguously identify the number of exponential phases corresponding to a complete single turnover event, and it was not possible to predict the expected total concentration of bound complexes in a single turnover.

The lack of active site titration data also introduced ambiguities that made direct mechanistic information unobtainable. It was not possible to resolve whether the multiple

exponentials represented different serial steps occurring at different rates, different conformations of the enzyme-substrate complex operating in parallel with different reactivities, or the interconversion between active and non-active complexes. The processing reaction can be fitted to a single exponential phase followed by a linear phase, which represents the initial portion of the subsequent exponential phase only within the first 120 s of the reaction coordinate. For reaction times greater than this, a given product begins to become the time-averaged products resulting from additional exponential phases and/or steps. We have reported a parallel integrase-catalyzed splicing reaction with synapsed substrates that results in a product structurally indistinguishable from an integrase-catalyzed processing product [2]. However, the two different origins of the products were distinguishable using presteady-state assays because the processing and splicing reactions proceed as two separate exponential phases with correspondingly unique rate constants (λ's). Similar results were obtained whether the rates were measured using 5'-end or 3'-end-radiolabeled substrates. These results showed that comparisons of enzyme reactivities based only on the quantitation of integrase-catalyzed products accumulated from "steady-state" measurements made at relatively longer reaction times can be inaccurate.

In experiments where enzyme and DNA substrates were not pre-incubated prior to the addition of $MnCl_2$ to initiate the reaction, the fastest exponential phase was best fit with an exponential rate of only 0.02 s^{-1}. In contrast, experiments in which integrase and DNA were preincubated showed an initial exponential phase of product formation at 0.2 s^{-1}, indicating that some of the slower exponentials seen in product formation are due to slower enzyme/substrate binding and/or conformational change steps leading up to the chemistry step. Preincubation permitted the slower formation of a catalytically productive complex to occur prior to initiation of the reaction (by addition of the metal cofactor), thereby circumventing these slower binding steps and allowing detection of the faster steps in the initial phase of the time courses. The amplitude of the fastest phase therefore represents a population of complexes that is productive and that is depleted rapidly. For this reason, the amplitude of the first occurring and, consequently, fastest rate-containing exponential was

used to compare the relative reactivities of the DNA substrates as it has the smallest number of additional reactions complicating the quantitation of product.

The reaction quench mixture contained both EDTA to chelate the metal cofactor and urea to denature the enzyme, stopping further flux of DNA through the reaction pathway and releasing all DNA bound to the enzyme. As such, the signal measured was not ambiguous and reflected directly the amount of DNA hydrolyzed by the enzyme. The product providing the signal to measure this fastest phase is the physiologically relevant processing product, and thus, its rate of 0.2 s^{-1} represents the lower limit estimate of the rate of the chemistry step for integrase-catalyzed processing. Both the rate of reaction and amplitude of the fastest phase for single-site substrates approach those of the synapsed substrate at higher concentrations of integrase, suggesting that this first exponential step lies on the same mechanistic pathway for both substrates and is not unique to the synapsed substrates.

Amplitude

The need for a large excess of integrase to obtain a significant initial exponential phase is unfortunate but consistent with the following known DNA binding properties of integrase. 1) In nitrocellulose filter binding experiments, a minimum of 10-fold excess of integrase over DNA is required to achieve saturating conditions [2], and 2) both processing and joining activities of integrase require a multimeric form of the enzyme [17, 24, 36] with up to 11 protomers [16] required per LTR end . The typical range of initial processing exponential amplitudes was on the order of 0.5−12% of the total DNA in the range of integrase concentrations examined. Unfortunately, this amplitude is too small to be useful in a detailed mechanistic study. To examine the possibility that the small amplitude size was due to inactive protein, we have compared the amplitude size from integrase purified from different preparations, different clones, and different enzyme purification protocols as well as protein preparations from different laboratories (using different vectors, different clones, and different purification protocols). In all cases, the amplitude size was similar and reproducible (data not shown). The only differences observed in integrase activity, as measured by a

change in the size of the amplitude of the first exponential, was upon the inclusion of detergent during the preparation of the enzyme.

One explanation for the low amplitude is the formation of nonproductive aggregates. In addition to the observed low solubility of the enzyme [1, 32, 37], we and others have observed that although purified integrase exists as monomers, dimers, and tetramers, large aggregates/multimers do form upon the introduction of DNA at integrase concentrations above 20 nM (Ref. [17], data not shown). Furthermore, we have examined DNA substrates of varying lengths and found that the processing exponential amplitude size decreases with increasing substrate length (manuscript in preparation), suggesting that noncognate DNA sequences facilitate the formation of nonproductive integrase/DNA aggregates. In the integrase titration experiments (Fig. 3.3*B*), we suspect the suppressed amplitude at the highest concentration of integrase to have been due to nonproductive protein aggregation in the presence of the longer DNA strands of the synapsed substrate.

Another possible explanation for the small amplitude size is the reported dependency of integrase processing activity on the disruption of the terminal base pairs of the substrate [15, 38]. If the processing activity were dependent on DNA distortion, the amplitude size would be a reflection of the subpopulation of productive integrase/DNA complexes in which the substrate DNA ends are frayed. Although the absolute sizes of the amplitudes were not useful, the relative amplitudes permitted examination of the relative reactivity of substrates.

Effect of Reverse-Polarity Tethering on Integrase-Mediated Processing

While the oligomeric state of functionally active integrase is not known, previous reports have suggested that integrase functions minimally as a dimer [17, 24, 36, 39, 40] or a dimer of dimers [40]. The presteady-state data are consistent with integrase functioning as a multimer in its active form. The initial apparent exponential rate measured for the ends-processing of the reverse-polarity synapsed substrate was 0.2 s^{-1} and remained constant throughout the integrase concentration range tested (0.5−5.0 μM). In contrast, the exponential rate measured for the unsynapsed substrate increased linearly from 0.014 to 0.14 s^{-1} with increasing integrase over the same concentration range. Additionally, although the

exponential amplitudes of both synapsed and unsynapsed substrates increased monotonically with increasing enzyme concentration, the size of the amplitudes was larger for the synapsed substrate for all but the highest integrase concentration. These results indicate that tethering of the LTR ends facilitates increased formation of productive integrase-substrate complexes during the preincubation period relative to the separate individual LTR ends. Furthermore, the correlation between increasing protein concentration and increasing activity suggests the assembly of a higher ordered active integrase multimer.

In addition to the processing and joining activities catalyzed by integrase, substrates that tether together two viral LTRs are also susceptible to a splicing reaction that produces a specifically sized product [2]. The "rogue" splicing and physiologically relevant joining activities differ in that the latter generates products in a range of sizes due to the nonspecific nature of joining site selection. Although the synapsing of the two ends does enhance integrase catalytic activity, the susceptibility of synapsed substrates to the splicing activity complicates accurate quantitation of the extent of the other two reactions [1, 2]. The lack of detectable integrase-catalyzed splicing products from reverse-polarity substrates, even at reaction conditions shown to favor the splicing activity, suggests that these substrates are viable tools for studying the molecular details of the processing activity. Although a small amount of products larger than the starting material was visible in these reactions, these products did not appear until much later, after the initial appearance of processing products. Additionally, the distribution of sizes of these larger products indicates that they are most likely the result of joining rather than splicing activity. We conclude that the reverse-polarity substrates are an improvement over previous normal-polarity substrates because the lack of splicing activity allows for uncomplicated quantitation of presteady-state experiments to make mechanistic rather than phenomenological observations with regard to integrase reactivity.

Comparison of Concurrently Bound Viral LTR Ends

The utilization of reverse-polarity synapsed substrates along with selective radioactive labeling allowed for experiments in which integrase reactivity with a particular DNA sequence could be measured as a function of the other LTR sequence. Burst analysis

allows this comparison to be carried out under single-turnover conditions, thus ensuring that the data reflected a single binding and catalysis event. Experiments were conducted to examine the extent of integrase processing reactivity with substrates containing either two U3 ends, a U3 and a U5 end, or two U5 ends. The data from presteady-state experiments identified the order of integrase preference for LTR end combinations as U3/U5 > U3/U3 > U5/U5 in addition to confirming previous reports of the preference of integrase for the U3 sequence over the U5 sequence [1, 14, 25].

Our results represent the first direct evidence of the preference for a physiologically relevant U3/U5 combination of LTR ends in *in vitro* studies on the processing reaction. However, in light of the fact that the asymmetric U3/U5 pair is the "proper" *in vivo* substrate for the molecule, this preference is not altogether surprising. Additionally, this result is consistent with results from DNase protection analysis, which indicated that integrase likely binds asymmetrically to the U3 and U5 LTR ends when the two ends are coupled for correct full-site integration [17]. Examination of the joining reaction products (Fig. 3.6) further suggests that these preferences extend beyond the processing activity and has implications on the temporal order of the joining reaction. Specifically, the U3 sequence had not only a significantly larger burst amplitude when examined for processing activity, but putative joining products from processed U3 ends also appeared at reaction times much earlier when compared to those from the U5 end. Although the two ends appear to be integrated concurrently on a longer and more macroscopic time scale [17, 22], our data suggest that the U3 cognate sequence is joined to the host DNA before the U5 end, even within a single turnover.

Comparative data of the U3 and U5 ends revealed that productive complex formation is maximal for substrates with both end-sequences present. While this result is consistent with the data reported by Vora *et al.* [14], our conclusions differ from theirs. We believe that the "macroscopic" nature of the analysis used in their study obscured the preference of integrase for the asymmetric U3/U5 combination of ends. In that work, integrase processing and joining activities were assayed by detection of the integration of radiolabeled 480-base pair mini-viral substrates containing the LTR ends into a 2,867-base

pair circular target. Products accumulated after 10 min resulting from both full-site integration, in which two LTR ends are inserted into host DNA, and half-site integration, in which only one LTR end is inserted, were summed to arrive at an order of substrate preference. Full-site integration with these mini-viral substrates could have taken place through either a trimolecular donor reaction, where ends from two different donor molecules are joined to a single target DNA, or a bimolecular reaction, where the two ends of a single molecule are joined to a single target DNA to form the final integration product. It has been observed that the bimolecular reaction is less than 5% as efficient [14, 20, 21, 27]. In contrast, the end-synapsed substrates used in our presteady-state assay are bound predominantly in a unimolecular fashion [2], thus allowing direct measurement of the reactivity of one LTR end sequence with another specific LTR end sequence bound at the active site. In addition, using single-turnover conditions allowed the resolution of the specific origins of the processing and joining products and their intermediates. Although from a macroscopic point of view, the integration of the two LTR ends into host DNA is temporally concerted in a single binding event, our presteady-state data show that within this single binding event the U3 end is processed before the U5 end and undergoes the joining activity earlier as well. A more recent study indicating that concerted DNA integration requires the presence of both U3 and U5 ends in the donor DNA [22] provides additional direct evidence consistent with our results. However, by using the presteady-state assay, we were able to make a distinction between the microscopic order of the events within the integration reaction *versus* the seemingly macroscopic concerted nature of this reaction.

REFERENCES

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

2. Bao, K.K., A.M. Skalka, and I. Wong, *Presteady–state Analysis of Avian Sarcoma Virus Integrase. I. A Splicing Activity and Structure–Function Implications for Cognate Site Recognition.* J Biol Chem, 2002. **277**(14): p. 12089–98.

3. Katz, R.A. and A.M. Skalka, *The retroviral enzymes.* Annu Rev Biochem, 1994. **63**: p. 133–73.

4. Daniel, R., R.A. Katz, and A.M. Skalka, *A role for DNA–PK in retroviral DNA integration.* Science, 1999. **284**(5414): p. 644–7.

5. Chow, S.A., et al., *Reversal of integration and DNA splicing mediated by integrase of human immunodeficiency virus.* Science, 1992. **255**(5045): p. 723–6.

6. Sherman, P.A., M.L. Dickson, and J.A. Fyfe, *Human immunodeficiency virus type 1 integration protein: DNA sequence requirements for cleaving and joining reactions.* J Virol, 1992. **66**(6): p. 3593–601.

7. Kulkosky, J., et al., *Activities and substrate specificity of the evolutionarily conserved central domain of retroviral integrase.* Virology, 1995. **206**(1): p. 448–56.

8. Hishinuma, F., et al., *Nucleotide sequence of acceptor site and termini of integrated avian endogenous provirus ev1: integration creates a 6 bp repeat of host DNA.* Cell, 1981. **23**(1): p. 155–64.

9. Muesing, M.A., et al., *Nucleic acid structure and expression of the human AIDS/lymphadenopathy retrovirus.* Nature, 1985. **313**(6002): p. 450–8.

10. Starcich, B., et al., *Characterization of long terminal repeat sequences of HTLV–III.* Science, 1985. **227**(4686): p. 538–40.

11. Vink, C., et al., *Analysis of the junctions between human immunodeficiency virus type 1 proviral DNA and human DNA.* J Virol, 1990. **64**(11): p. 5626–7.

12. Peliska, J.A. and S.J. Benkovic, *Mechanism of DNA strand transfer reactions catalyzed by HIV–1 reverse transcriptase.* Science, 1992. **258**(5085): p. 1112–8.

13. Chow, S.A., *In vitro assays for activities of retroviral integrase.* Methods, 1997. **12**(4): p. 306–17.

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

15. Scottoline, B.P., et al., *Disruption of the terminal base pairs of retroviral DNA during integration.* Genes Dev, 1997. **11**(3): p. 371–82.

16. Pemberton, I.K., M. Buckle, and H. Buc, *The metal ion–induced cooperative binding of HIV–1 integrase to DNA exhibits a marked preference for Mn(II) rather than Mg(II).* J Biol Chem, 1996. **271**(3): p. 1498–506.

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

18. Craigie, R., T. Fujiwara, and F. Bushman, *The IN protein of Moloney murine leukemia virus processes the viral DNA ends and accomplishes their integration in vitro.* Cell, 1990. **62**(4): p. 829–37.

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

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

21. Fitzgerald, M.L. and D.P. Grandgenett, *Retroviral integration: in vitro host site selection by avian integrase.* J Virol, 1994. **68**(7): p. 4314–21.

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

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

24. Jones, K.S., et al., *Retroviral integrase functions as a multimer and can turn over catalytically.* J Biol Chem, 1992. **267**(23): p. 16037−40.

25. Fitzgerald, M.L., A.C. Vora, and D.P. Grandgenett, *Development of an acid−soluble assay for measuring retrovirus integrase 3'−OH terminal nuclease activity.* Anal Biochem, 1991. **196**(1): p. 19−23.

26. Vora, A.C. and D.P. Grandgenett, *Assembly and catalytic properties of retrovirus integrase−DNA complexes capable of efficiently performing concerted integration.* J Virol, 1995. **69**(12): p. 7483−8.

27. Fitzgerald, M.L., et al., *Concerted integration of viral DNA termini by purified avian myeloblastosis virus integrase.* J Virol, 1992. **66**(11): p. 6257−63.

28. Grandgenett, D.P., et al., *Comparison of DNA binding and integration half−site selection by avian myeloblastosis virus integrase.* J Virol, 1993. **67**(5): p. 2628−36.

29. Cantor, C.R., M.M. Warshaw, and H. Shapiro, *Oligonucleotide interactions. 3. Circular dichroism studies of the conformation of deoxyoligonucleotides.* Biopolymers, 1970. **9**(9): p. 1059−77.

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

31. Terry, R., et al., *Properties of avian sarcoma−leukosis virus pp32−related pol−endonucleases produced in Escherichia coli.* J Virol, 1988. **62**(7): p. 2358−65.

32. Coleman, J., et al., *Characterization of the self association of Avian sarcoma virus integrase by analytical ultracentrifugation.* J Biol Chem, 1999. **274**(46): p. 32842–6.

33. Ellison, V. and P.O. Brown, *A stable complex between integrase and viral DNA ends mediates human immunodeficiency virus integration in vitro.* Proc Natl Acad Sci U S A, 1994. **91**(15): p. 7316–20.

34. Bor, Y.C., et al., *Target–sequence preferences of HIV–1 integration complexes in vitro.* Virology, 1996. **222**(1): p. 283–8.

35. Patel, S.S., I. Wong, and K.A. Johnson, *Pre–steady–state kinetic analysis of processive DNA replication including complete characterization of an exonuclease–deficient mutant.* Biochemistry, 1991. **30**(2): p. 511–25.

36. Deprez, E., et al., *Oligomeric states of the HIV–1 integrase as measured by time–resolved fluorescence anisotropy.* Biochemistry, 2000. **39**(31): p. 9275–84.

37. Chen, J.C., et al., *Crystal structure of the HIV–1 integrase catalytic core and C–terminal domains: a model for viral DNA binding.* Proc Natl Acad Sci U S A, 2000. **97**(15): p. 8233–8.

38. Katz, R.A., et al., *Role of DNA end distortion in catalysis by avian sarcoma virus integrase.* J Biol Chem, 2001. **276**(36): p. 34213–20.

39. Pemberton, I.K. and O.M. Singh, *In vitro biochemical activities associated with recombinant HIV–1 integrase.* Biochem Soc Trans, 1993. **21**(1): p. 51S.

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

Chapter 4

STRUCTURAL AND FUNCTIONAL DETERMINATION OF THE OLIGOMERIC STATE OF AVIAN SARCOMA VIRUS INTEGRASE

Kogan K. Bao, Hong Wang, Dorothy A. Erie, Anna Marie Skalka, and Isaac Wong

Adapted from the manuscript submitted for publication (August, 2002)

SUMMARY

Retroviral integrase, one of only three enzymes encoded by the virus, catalyses the essential step of inserting a DNA copy of the viral genome into the host during infection. Using the avian sarcoma virus (ASV) integrase, we demonstrate that the enzyme functions as a tetramer. In presteady-state active site titrations, four integrase protomers were required for a single catalytic turnover. Volumetric determination of integrase-DNA complexes imaged by atomic force microscopy (AFM) during the initial turnover additionally revealed substrate-induced assembly of a tetramer. These results suggest that tetramer formation may be a requisite step *during* catalysis with ramifications for antiviral design strategies targeting the structurally homologous human immunodeficiency virus type 1 (HIV-1) integrase.

INTRODUCTION

Integrase catalyses two consecutive transesterification reactions during its *in vivo* function [1, 2]. In the "processing" reaction, the reverse transcriptase-generated DNA copy of the viral genome is trimmed by the endonucleolytic removal of the 3'-terminal dinucleotides from its ends. These processed 3'-ends are then inserted into the host DNA in the "joining" reaction via a concerted cleavage-ligation reaction [3–6]. Purified integrase catalyses both reactions on synthetic oligodeoxyribonucleotide substrates containing viral DNA end sequences in the presence of either Mn^{2+} or Mg^{2+} as a cofactor [7–10].

In vitro, integrase also catalyses the apparent reversal of the DNA joining reaction – the "disintegration" activity – on Y-shaped oligodeoxynucleotide substrates resembling the structure of the cleavage-ligation product [11] (Fig. 4.1). This reaction is routinely used to assay integrase *in vitro* [8, 12–14]. For our presteady-state kinetic assays, we constructed a Y-shaped substrate representing the result of a 21-base pair U3 viral DNA end undergoing integrase-catalysed processing followed by joining into a 44-base pair target DNA, leaving a 19-nucleotide fragment 5' of the nick at the site of insertion on the target (Fig. 4.1). Integrase-catalysed disintegration on this substrate reverses the joining by using the 3'-OH

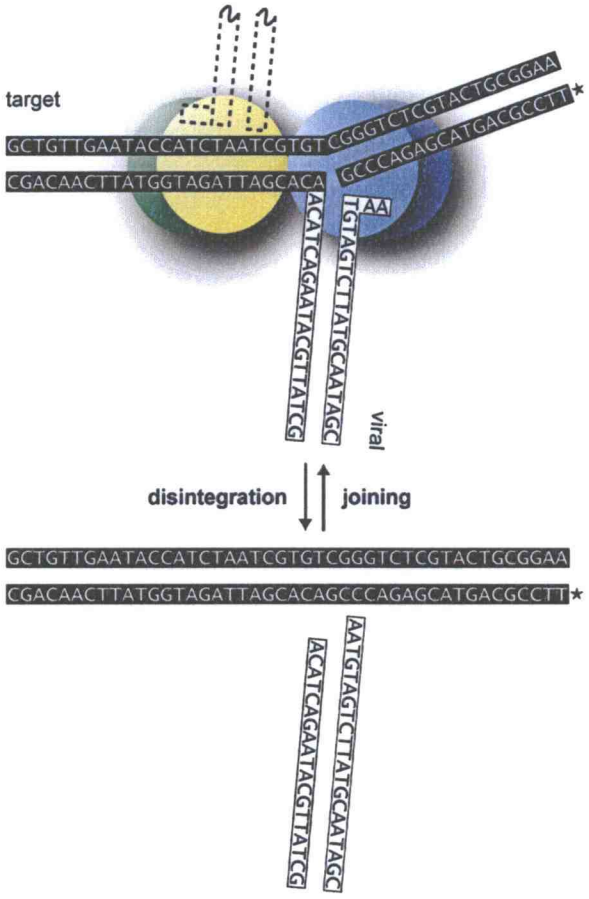

Figure 4.1: Diagram of the integrase-catalyzed disintegration and joining reactions. *Top,* depicted is the explicit sequence of the Y-shaped disintegration substrate along with a generic tetramer model for integrase. *Black letters* are sequences derived from the U3 LTR of ASV; *white letters* are a randomly generated target sequence. The asterisk (*) denotes the position of the 5'-^{32}P-radiolabel. The position of the DNA in the tetramer model is for illustrative purposes only and is not intended to imply positions of direct DNA-protein contact.

of this 19-nucleotide fragment to reseal the nick, concurrently regenerating the 44-base pair target DNA fragment while releasing the processed viral U3 DNA.

EXPERIMENTAL PROCEDURES

Synthetic Oligodeoxyribonucleotides and 5'-^{32}P-Labeling

Oligodeoxyribonucleotides were synthesized by the Center for Gene Research and Biotechnology Central Services Laboratory (Oregon State University). Purification by denaturing PAGE, spectrophotometric determination of concentrations, and 5'-radiolabeling were as previously described [9].

ASV Integrase Overexpression and Purification

Integrase was overexpressed in *Escherichia coli* BL21(DE3), purified as described [9], and stored at -80 °C in 50 mM HEPES, pH 7.5, 500 mM NaCl, and 40% glycerol.

Pre-Steady State Assays and Product Analysis by Denaturing Acrylamide Gel Electrophoresis

Standard reactions were carried out at 37 °C in 20 mM Tris, pH 8.0, 10 mM Na-HEPES, pH 7.5, 4% glycerol, 10 mM 2-mercaptoethanol, 0.050 mg ml^{-1} acetylated BSA, 250 mM NaCl and 5 mM $MnCl_2$. Preincubations were carried out in the absence of $MnCl_2$ for 30 min and the reaction subsequently initiated by the addition of 37 °C $MnCl_2$ to 5 mM. A complete range of preincubation times and temperatures were tested to ensure that equilibrium had been achieved. Gel analysis, image quantitation, and non-linear least squares fittings were performed according to Bao *et al.* [9, 10].

Imaging by Atomic Force Microscopy

Imaging was performed with a Nanoscope IIIa instrument (Digital Instrument, Santa Barbara, CA) using the tapping mode in air. Nanosensor Pointprobe noncontact/tapping mode sensors with a nominal spring constant of 48N/m and resonance frequencies of 190kHz were used for all images. The protein and DNA molecules were deposited onto freshly cleaved mica (Spruce Pine Mica Co., Spruce Pine, NC), immediately washed with deionized distilled water, and dried with a stream of N_2 (gas). Depositions of disintegration reactions during the first catalytic turnover were obtained by preincubating 4

μM integrase with 5 μM substrate DNA in 20 mM Tris, pH 8.0, 10 mM HEPES, pH 7.5, and 250 mM NaCl for a minimum of 30 min on ice, diluting the reaction to 64 nM integrase and 150 mM NaCl immediately before initiating the reaction by the addition of $MnCl_2$ to 5 mM. Aliquots were then deposited within 10 s of initiating the reaction. All images were collected with a scan rate of 3.2 Hz, at 512 × 512 resolution, and a scan size of 1 μm^2. Volume analysis of AFM data was performed with the freeware program ImageSXM (based on NIH Image developed at the National Institutes of Health) using image plane fitting, image analysis, volume calculations, and conversion to molecular weights according to MW_{app} = (Volume + 25) ÷ 1.31 as described [15, 16].

RESULTS

Single-turnover and active site titration disintegration assays

For single-turnover experiments, the substrate was 5'-radiolabeled on the 19-nucleotide target fragment (denoted by the asterisk in Fig. 4.1) which is converted to the 44-nucleotide product during disintegration (*inset*, Fig. 4.2*A*). Substrate (0.25 μM) was preincubated with excess ASV integrase (1.0 μM) for 30 min in the absence of Mn^{2+} to allow equilibrium formation of productive complexes [9, 10]. Reactions were initiated by the addition of Mn^{2+} to 5 mM. Fig. 4.2*A* (*closed circles*) shows biphasic conversion of the radiolabeled 19-nucleotide substrate to the expected 44-nucleotide product. Nonlinear least-squares fit of the data to a double-exponential function yielded observed rate constants of 0.05 s^{-1} and 0.002 s^{-1} with amplitudes of 17% and 60%, respectively. When the reaction was initiated by mixing integrase and substrate without preincubation but in the presence of Mn^{2+}, we observed only the slower 0.002 s^{-1} exponential phase (Fig. 4.2*A*, *open circles*) with an amplitude (75 ± 0.9%) statistically equal to the sum of the amplitudes (77 ± 1.4%) obtained with preincubation. The 0.05 s^{-1} phase therefore reflected productive integrase-substrate complexes formed during the preincubation period in the absence of Mn^{+2}, whereas the slower 0.002 s^{-1} reaction reflected the rate-limiting assembly of productive complexes in the presence of Mn^{2+}. More importantly, the two phases represent sequential reaction steps along the same kinetic pathway. Furthermore, the observation of

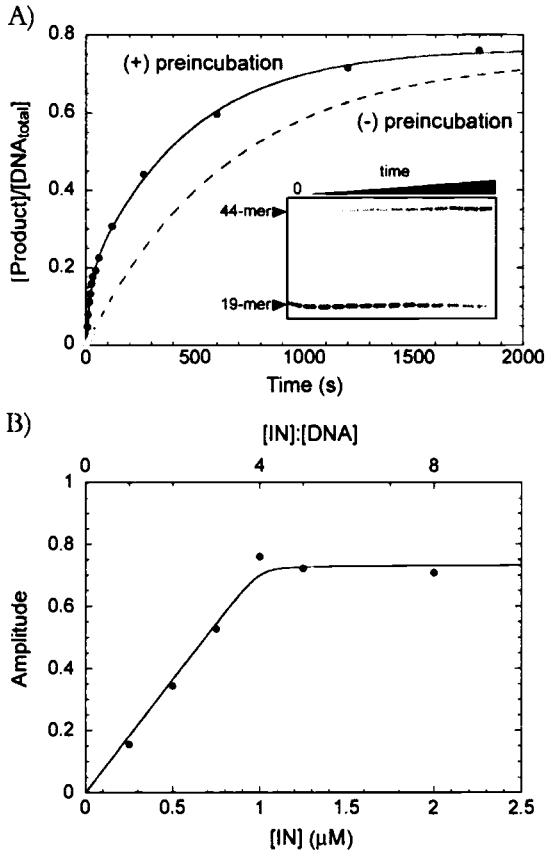

Figure 4.2: Presteady-state disintegration assays of ASV integrase. *A*, single-turnover time course of the disintegration reaction at 1 μM integrase and 0.25 μM substrate DNA were obtained by quantifying the amount of 44-mer formed following chromatographic separation by denaturing PAGE (*inset*). Experiments were conducted with (*closed circles*) and without (*open circles*) preincubating enzyme and substrate prior to initiating the reaction. The ~75% amplitude represented complete conversion of substrate to the expected product in these experiments because a silent, competing hydrolysis reaction consumed and quantitatively accounted for the remaining 20–25% of substrate (data not shown). The *solid* and *dashed lines* represent the best fit of the preincubated and non-preincubated data, respectively, to sums of exponentials with $A_{1,\ preinc} = 0.17 \pm 0.01$, $A_{2,\ preinc} = 0.60 \pm 0.01$, $\lambda_{1,\ preinc} = 0.05\ s^{-1} \pm 0.006$, $\lambda_{2,\ preinc} = 0.002\ s^{-1} \pm 0.0001$, $A_{nonpreinc} = 0.74 \pm 0.009$, and $\lambda_{nonpreinc} = 0.002\ s^{-1} \pm 0.00004$. *B*, a series of presteady-state burst assays were performed at 0.25 μM DNA, 250 mM NaCl, and increasing integrase concentrations. The total exponential amplitudes corresponding to the first turnover are plotted as a function of integrase concentration. The *solid line* is a fit of the data to a quadratic Langmuir isotherm using an arbitrarily fixed $K_d < 1$ nM sufficient to generate a stoichiometric titration curve. At this NaCl concentration, integrase aggregated at concentrations of > 2.5 μM, resulting in an apparent loss of amplitude (data not shown).

only 17% pre-formed productive integrase-DNA complexes in the absence of Mn^{2+} suggests that the formation of these complexes is stimulated by the metal cofactor [17].

To determine the number of integrase protomers required to catalyze disintegration of a single substrate, a series of disintegration reactions was performed at 0.25 μM substrate DNA with varying integrase concentrations from 0.25 μM to >2.5 μM. The total exponential amplitude, which represents the extent of kinetically competent productive complexes formed, increased linearly with added integrase up to 1 μM, reaching a plateau at concentrations >1 μM (Fig. 4.2*B*). Because a product release step slower than 0.002 s^{-1} was rate limiting for *multiple* catalytic turnovers (data not shown), incomplete conversion of substrate in the first catalytic turnover was observed at substoichiometric enzyme concentrations (<1 μM) where the amplitude of the first turnover measured the substrate binding capacity of the integrase present. Above 1 μM integrase, the substrate binding capacity of the added integrase exceeded the concentration of substrate; consequently, the maximum reaction amplitude obtained (~75%) became independent of integrase concentration. The complete conversion of substrate to product within a single turnover required a four-fold molar excess of integrase over DNA, defining a reaction stoichiometry of four protomers per substrate molecule. Identical results were obtained in control experiments performed at different substrate and NaCl (200-400 mM) concentrations, using integrase purified by different protocols (data not shown).

Atomic force microscopy volume determination

To verify a tetrameric active assembly state, we determined the molecular weight of integrase complexes in the presence and absence of DNA by AFM volumetric analysis using the linear relationship between measured AFM volumes and molecular weights previously established for proteins ranging from 41 kDa to 670 kDa [15, 16]. A typical 1 μm^2 AFM image of integrase alone showed primarily particles with volumes consistent with monomers and dimers (Fig. 4.3*A*, *top*). By comparison, images of actively functioning integrase, directly deposited from disintegration reactions during the initial turnover, show significantly increased numbers of tetramer-sized particles (Fig. 4.3*A*, *bottom*). Analysis of the calculated sizes for all particles imaged confirmed that in the absence of DNA, integrase

Figure 4.3: Imaging and volumetric analysis of single molecules of ASV integrase by atomic force microscopy. *A*, AFM images of integrase deposited at a relatively dilute concentration (64 nM) to maintain an optimal particle density for imaging allowed visualization of a distribution of mostly monomers and dimers with an occasional tetramer (*left*) that was shifted in favor of tetramers in the presence of 80 nM substrate DNA (*right*). Images were obtained at 150 mM NaCl and 5 mM $MnCl_2$. The more stringent condition of excess DNA was chosen in order to specifically disfavor the formation tetramers; images of reactions at higher integrase and NaCl concentrations also showed enhanced tetramerization in the presence of substrate, albeit with increased background noise. The monomer peaks in the integrase scans were not observed in control scans of the DNA alone even though the M.W. of the DNA is larger than that of the monomer (data not shown), presumably because of the low contour height of the rod–like DNA. *Scale bar*, 200 nm. Images were created with the freeware program WSxM (www.nanotec.es). *B*, volume histograms of 405 particles analyzed in the absence of substrate (*top*) and 381 particles analyzed in the presence of substrate (*bottom*) show an increase in the tetramer:monomer ratio upon the addition of DNA. Integrase monomer and dimers were not well resolved from each other under these conditions. The *inset axis* represents the molecular volume to molecular weight conversion. *C*, for a more direct comparison of the percentage of integrase molecules found within the distribution of multimeric states, the volume histogram from *B* with substrate present was adjusted to reflect the mass of each particle. Nearly half of the total integrase molecules were found in the tetrameric state. The *inset axis* represents the molecular volume to molecular weight conversion.

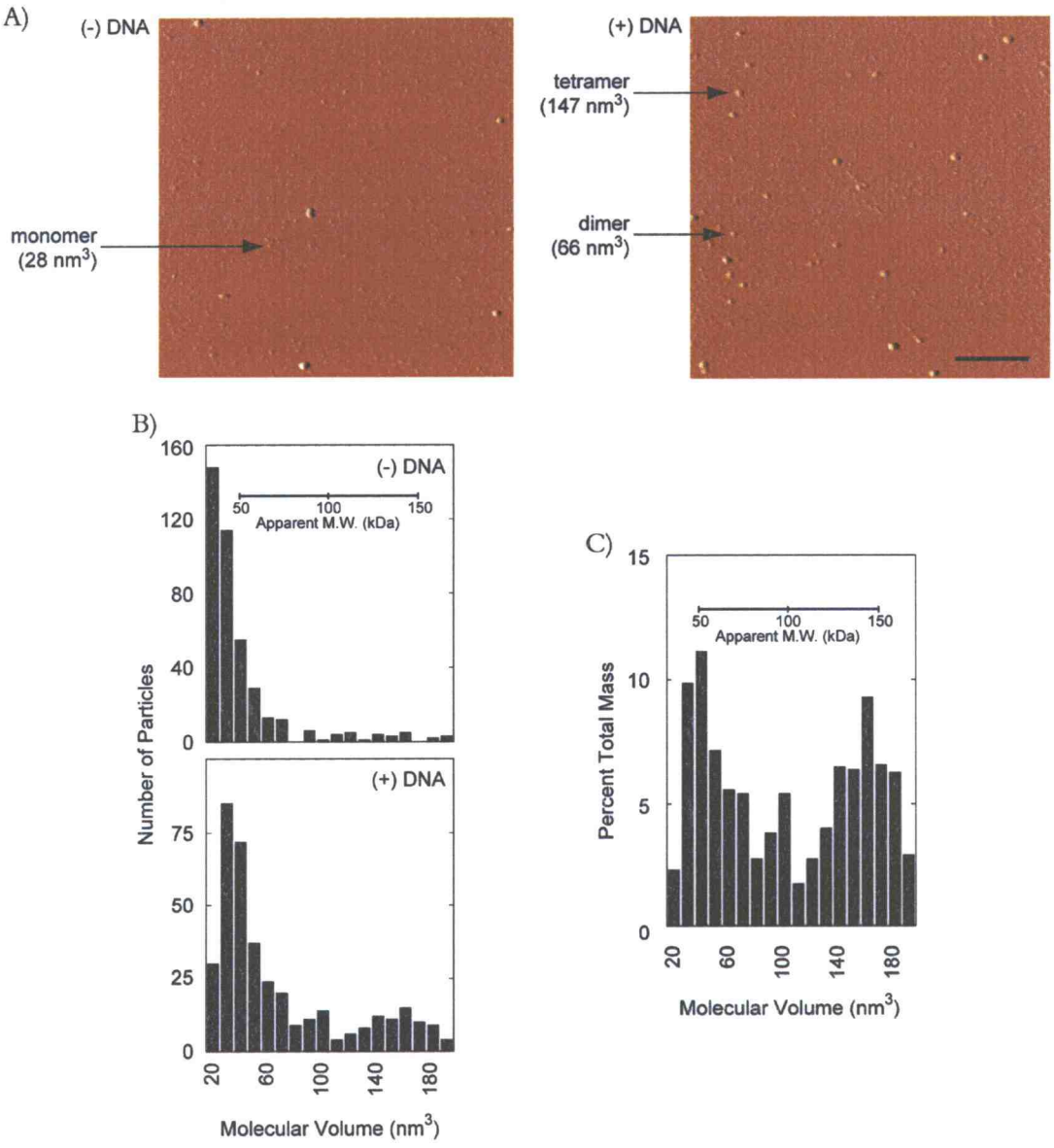
A)
(-) DNA
monomer
(28 nm3)
(+) DNA
tetramer
(147 nm3)
dimer
(66 nm3)
B)
(-) DNA
50
100
150
Apparent M.W. (kDa)
160
120
80
40
0
(+) DNA
75
50
25
0
Number of Particles
20
60
100
140
180
Molecular Volume (nm3)
C)
15
10
5
0
Percent Total Mass
Molecular Volume (nm3)

Figure 4.3

appeared predominantly as monomers and dimers (Fig. 4.3*B*, *top*). In contrast, a new peak with a mean molecular volume of 154 ± 7.5 nm^3, corresponding to a calculated molecular weight of 138 kDa, became apparent in the presence of DNA (Fig. 4.3*B*, *bottom*), indicating that the binding of substrate DNA induced formation of an integrase tetramer. The calculated mass of the tetrameric particles was slightly larger than expected, due likely to the bound DNA, although the actual mass contributed by the DNA was underestimated presumably because part of the volume contributed by the DNA is topologically hidden within the cavity of the DNA binding site. Fig. 4.3*C* illustrates that in the presence of DNA, ~50% of all the integrase was tetrameric when the population distribution is adjusted to more accurately reflect the number of integrase protomers subsumed within each volumetric sub-population by accounting for the mass of the particles. These results, coupled with an active site titration stoichiometry of four, provide compelling evidence for the substrate-induced assembly of a functionally active tetrameric quaternary structure.

DISCUSSION

Tetrameric models for functional integrase

Although the structure of the full-length 32-kDa integrase protein is unknown, the central catalytic core domain [18–20] and all 2-domain fragments consisting of the core plus either the C- or N-terminus [21–24], from a variety of retroviral sources, form dimers in the crystal structures. However, the active sites of the subunits in these dimeric structures are located on opposite faces of the crystallographic dimers, seemingly too far apart (>50 Å) to be spanned by the 5–6-base pair stagger defining the spacing between the two sites of concerted integration within the host DNA [23]. Comparison of the available integrase structures along with that of the homologous bacterial Tn5 transposase has led to several hypothetical "dimer of dimers" models (Fig. 4.4). In these models, the two active sites of the two outward-facing "distal" protomers are not used. Instead, the viral ends are bound at the two inward-facing active sites of the tetramer, on either side of the dimer-dimer interface [23–25]. Concerted integration of both ends would therefore require concurrent binding of the target DNA across the dimer-dimer interface (Fig. 4.1, *top*). Although the

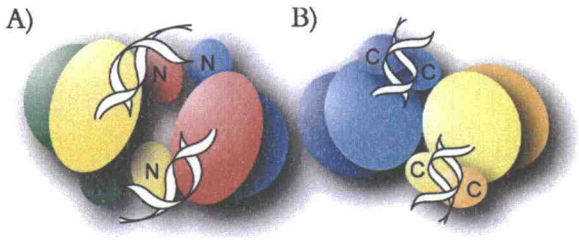

Figure 4.4: Hypothetical dimer of dimers models of an integrase tetramer. Models show concurrent binding of two viral DNA ends on either side of the dimer-dimer interface. *A*, The tetramer model proposed by Wang *et al.* [24] based on a core plus N-terminus 2-domain dimer structure of HIV-1 integrase. The large and small ellipsoids represent core and N-terminus domains, respectively. *B*, The model proposed by Yang *et al.* [23] based on a core plus C-terminus 2-domain dimer structure of Rous Sarcoma Virus integrase. The large and small ellipsoids represent core and C-terminus domains, respectively. The color schemes of the published structures have been retained to facilitate ease of comparison.

molecular details of DNA-binding must await the structure of a co-crystal, our observation that the Y-shaped substrate, with only a single viral-end mimic, induced active tetramer assembly would suggest significant contribution by the target DNA in mediating the dimer-dimer interaction. Whereas these results support a bridging position for the target DNA consistent with the proposed tetramer models, our data cannot rule out an octameric (dimer of tetramers) model [23–25]; however, we feel that such an elaboration may not be necessary.

Implications for structural studies and antiviral design

Earlier attempts to determine the reaction stoichiometry by active site titrations using synapsed substrates containing only viral DNA [9, 10] have been frustrated by substrate-induced aggregation, a problem common to retroviral integrases [26]. Our results using the disintegration substrate suggest that the presence of target DNA spanning the tetramer along with a viral end mitigated these problems. Alternatively, the disintegration substrate conferred sufficient binding stability at the higher NaCl concentrations required to prevent aggregation. Furthermore, the observed increase of the tetramer population in the presence of substrate DNA in the AFM studies demonstrates that binding of the Y-shaped substrate induces assembly of the active tetramer. These results suggest that disintegration-like substrates, containing both target and viral DNA, may facilitate structural solutions of the active integrase unit by promoting tetramer formation as well as minimizing non-productive aggregation.

The observed substrate DNA- and metal cofactor-dependent oligomerization further suggests that the assembly of the active tetramer may be an integral and dynamic component of the catalytic pathway. The dynamic nature of the dimer-dimer interface should make it an ideal target for inhibitor design [13, 14, 26, 27]. The diketo acid family of inhibitors targeting the active site of integrase has recently been shown to inhibit the V(D)J recombinase, RAG1/2 [27]. We have recently characterized a novel RAG1/2-like "splicing" activity of ASV integrase that further suggests that any active-site directed inhibitor of integrase could similarly interfere with T- and B-cell development [9]. If the substrate binding-induced assembly of an active tetramer lies along the kinetic pathway of

catalysis, then inhibitors that disrupt the dimer-dimer interface may provide an enzyme-specific means of disrupting catalytic activity while avoiding possible adverse effects resulting from directly targeting the active site.

REFERENCES

1. Grandgenett, D.P. and S.R. Mumm, *Unraveling retrovirus integration.* Cell, 1990. **60**(1): p. 3–4.

2. Katz, R.A. and A.M. Skalka, *The retroviral enzymes.* Annu Rev Biochem, 1994. **63**: p. 133–73.

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

4. Bushman, F.D., T. Fujiwara, and R. Craigie, *Retroviral DNA integration directed by HIV integration protein in vitro.* Science, 1990. **249**(4976): p. 1555–8.

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

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

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

8. Chow, S.A., *In vitro assays for activities of retroviral integrase.* Methods, 1997. **12**(4): p. 306–17.

9. Bao, K.K., A.M. Skalka, and I. Wong, *Presteady–state Analysis of Avian Sarcoma Virus Integrase. I. A Splicing Activity and Structure–Function Implications for Cognate Site Recognition.* J Biol Chem, 2002. **277**(14): p. 12089–98.

10. Bao, K.K., A.M. Skalka, and I. Wong, *Presteady–state Analysis of Avian Sarcoma Virus Integrase. II. Reverse–Polarity Substrates Identify Preferential Processing of the U3–U5 Pair.* J Biol Chem, 2002. **277**(14): p. 12099–108.

11. Chow, S.A., et al., *Reversal of integration and DNA splicing mediated by integrase of human immunodeficiency virus.* Science, 1992. **255**(5045): p. 723–6.

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

13. Mazumder, A., et al., *Antiretroviral agents as inhibitors of both human immunodeficiency virus type 1 integrase and protease.* J Med Chem, 1996. **39**(13): p. 2472–81.

14. Puras Lutzke, R.A., et al., *Identification of a hexapeptide inhibitor of the human immunodeficiency virus integrase protein by using a combinatorial chemical library.* Proc Natl Acad Sci U S A, 1995. **92**(25): p. 11456–60.

15. Ratcliff, G.C. and D.A. Erie, *A novel single–molecule study to determine protein––protein association constants.* J Am Chem Soc, 2001. **123**(24): p. 5632–5.

16. Xue, Y., et al., *A minimal exonuclease domain of WRN forms a hexamer on DNA and possesses both 3'– 5' exonuclease and 5'–protruding strand endonuclease activities.* Biochemistry, 2002. **41**(9): p. 2901–12.

17. Asante–Appiah, E. and A.M. Skalka, *A metal–induced conformational change and activation of HIV–1 integrase.* J Biol Chem, 1997. **272**(26): p. 16196–205.

18. Dyda, F., et al., *Crystal structure of the catalytic domain of HIV–1 integrase: similarity to other polynucleotidyl transferases.* Science, 1994. **266**(5193): p. 1981–6.

19. Bujacz, G., et al., *High–resolution structure of the catalytic domain of avian sarcoma virus integrase.* J Mol Biol, 1995. **253**(2): p. 333–46.

20. Goldgur, Y., et al., *Three new structures of the core domain of HIV–1 integrase: an active site that binds magnesium.* Proc Natl Acad Sci U S A, 1998. **95**(16): p. 9150–4.

21. Chen, Z., et al., *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, 2000. **296**(2): p. 521-33.

22. Chen, J.C., et al., *Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: a model for viral DNA binding.* Proc Natl Acad Sci U S A, 2000. **97**(15): p. 8233-8.

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

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

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

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

27. Melek, M., et al., *Effect of HIV integrase inhibitors on the RAG1/2 recombinase.* Proc Natl Acad Sci U S A, 2002. **99**(1): p. 134-7.

BIBLIOGRAPHY

Aiyar, A., Hindmarsh, P., Skalka, A. M., and Leis, J., *Concerted integration of linear retroviral DNA by the avian sarcoma virus integrase in vitro: dependence on both long terminal repeat termini.* J Virol, 1996. **70**(6): p. 3571–3580.

Andrake, M. D., and Skalka, A. M., *Multimerization determinants reside in both the catalytic core and C terminus of avian sarcoma virus integrase.* J Biol Chem, 1995. **270**(49): p. 29299–29306.

Andrake, M. D., and Skalka, A. M., *Retroviral integrase, putting the pieces together.* J Biol Chem, 1996. **271**(33): p. 19633–19636.

Asante–Appiah, E., and Skalka, A. M., *A metal–induced conformational change and activation of HIV–1 integrase.* J Biol Chem, 1997. **272**(26): p. 16196–16205.

Asante–Appiah, E., and Skalka, A. M., *HIV–1 integrase: structural organization, conformational changes, and catalysis.* Adv Virus Res, 1999. **52**: p. 351–369.

Ausubel, F. M., *Current protocols in molecular biology.* 1988, New York: Greene Pub. Associates and Wiley–Interscience. <1–2, 4> (loose–leaf).

Bao, K. K., Skalka, A. M., and Wong, I., *Presteady–state Analysis of Avian Sarcoma Virus Integrase. I. A Splicing Activity and Structure–Function Implications for Cognate Site Recognition.* J Biol Chem, 2002a. **277**(14): p. 12089–12098.

Bao, K. K., Skalka, A. M., and Wong, I., *Presteady–state Analysis of Avian Sarcoma Virus Integrase. II. Reverse–Polarity Substrates Identify Preferential Processing of the U3–U5 Pair.* J Biol Chem, 2002b. **277**(14): p. 12099–12108.

Bor, Y. C., Miller, M. D., Bushman, F. D., and Orgel, L. E., *Target–sequence preferences of HIV–1 integration complexes in vitro.* Virology, 1996. **222**(1): p. 283–288.

Brin, E., and Leis, J., *HIV–1 integrase interaction with U3 and U5 terminal sequences in vitro defined using substrates with random sequences.* J Biol Chem, 2002. **277**(21): p. 18357–18364.

Brown, P. O., Bowerman, B., Varmus, H. E., and Bishop, J. M., *Retroviral integration: structure of the initial covalent product and its precursor, and a role for the viral IN protein.* Proc Natl Acad Sci U S A, 1989. **86**(8): p. 2525–2529.

Bujacz, G., Alexandratos, J., Wlodawer, A., Merkel, G., Andrake, M., Katz, R. A., and Skalka, A. M., *Binding of different divalent cations to the active site of avian sarcoma virus integrase and their effects on enzymatic activity.* J Biol Chem, 1997. **272**(29): p. 18161–18168.

Bujacz, G., Jaskolski, M., Alexandratos, J., Wlodawer, A., Merkel, G., Katz, R. A., and Skalka, A. M., *High-resolution structure of the catalytic domain of avian sarcoma virus integrase.* J Mol Biol, 1995. **253**(2): p. 333–346.

Bujacz, G., Jaskolski, M., Alexandratos, J., Wlodawer, A., Merkel, G., Katz, R. A., and Skalka, A. M., *The catalytic domain of avian sarcoma virus integrase: conformation of the active-site residues in the presence of divalent cations.* Structure, 1996. **4**(1): p. 89–96.

Burke, C. J., Sanyal, G., Bruner, M. W., Ryan, J. A., La Femina, R. L., Robbins, H. L., Zeft, A. S., Middaugh, C. R., and Cordingley, M. G., *Structural implications of spectroscopic characterization of a putative zinc finger peptide from HIV-1 integrase.* J Biol Chem, 1992. **267**(14): p. 9639–9644.

Bushman, F. D., Fujiwara, T., and Craigie, R., *Retroviral DNA integration directed by HIV integration protein in vitro.* Science, 1990. **249**(4976): p. 1555–1558.

Bushman, F. D., and Wang, B., *Rous sarcoma virus integrase protein: mapping functions for catalysis and substrate binding.* J Virol, 1994. **68**(4): p. 2215–2223.

Cai, M., Zheng, R., Caffrey, M., Craigie, R., Clore, G. M., and Gronenborn, A. M., *Solution structure of the N-terminal zinc binding domain of HIV-1 integrase.* Nat Struct Biol, 1997. **4**(7): p. 567–577.

Cantor, C. R., Warshaw, M. M., and Shapiro, H., *Oligonucleotide interactions. 3. Circular dichroism studies of the conformation of deoxyoligonucleotides.* Biopolymers, 1970. **9**(9): p. 1059–1077.

Chen, J. C., Krucinski, J., Miercke, L. J., Finer-Moore, J. S., Tang, A. H., Leavitt, A. D., and Stroud, R. M., *Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: a model for viral DNA binding.* Proc Natl Acad Sci U S A, 2000a. **97**(15): p. 8233-8238.

Chen, Z., Yan, Y., Munshi, S., Li, Y., Zugay-Murphy, J., Xu, B., Witmer, M., Felock, P., Wolfe, A., Sardana, V., *et al., 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, 2000b. **296**(2): p. 521-533.

Chow, S. A., *In vitro assays for activities of retroviral integrase.* Methods, 1997. **12**(4): p. 306-317.

Chow, S. A., Vincent, K. A., Ellison, V., and Brown, P. O., *Reversal of integration and DNA splicing mediated by integrase of human immunodeficiency virus.* Science, 1992. **255**(5045): p. 723-726.

Coffin, J., Hughes, S.H., Varmus, H.E., ed. *Retroviruses.* 1997, Cold Springs Harbor Laboratory Press: New York. 843.

Coleman, J., Eaton, S., Merkel, G., Skalka, A. M., and Laue, T., *Characterization of the self association of Avian sarcoma virus integrase by analytical ultracentrifugation.* J Biol Chem, 1999. **274**(46): p. 32842-32846.

Craigie, R., *HIV integrase, a brief overview from chemistry to therapeutics.* J Biol Chem, 2001. **276**(26): p. 23213-23216.

Craigie, R., Fujiwara, T., and Bushman, F., *The IN protein of Moloney murine leukemia virus processes the viral DNA ends and accomplishes their integration in vitro.* Cell, 1990. **62**(4): p. 829-837.

Daniel, R., Katz, R. A., and Skalka, A. M., *A role for DNA-PK in retroviral DNA integration.* Science, 1999. **284**(5414): p. 644-647.

Davies, D. R., Goryshin, I. Y., Reznikoff, W. S., and Rayment, I., *Three-dimensional structure of the Tn5 synaptic complex transposition intermediate.* Science, 2000. **289**(5476): p. 77-85.

Deprez, E., Tauc, P., Leh, H., Mouscadet, J. F., Auclair, C., and Brochon, J. C., *Oligomeric states of the HIV-1 integrase as measured by time-resolved fluorescence anisotropy.* Biochemistry, 2000. **39**(31): p. 9275-9284.

Dyda, F., Hickman, A. B., Jenkins, T. M., Engelman, A., Craigie, R., and Davies, D. R., *Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases.* Science, 1994. **266**(5193): p. 1981-1986.

Eijkelenboom, A. P., Lutzke, R. A., Boelens, R., Plasterk, R. H., Kaptein, R., and Hard, K., *The DNA-binding domain of HIV-1 integrase has an SH3-like fold.* Nat Struct Biol, 1995. **2**(9): p. 807-810.

Ellison, V., and Brown, P. O., *A stable complex between integrase and viral DNA ends mediates human immunodeficiency virus integration in vitro.* Proc Natl Acad Sci U S A, 1994. **91**(15): p. 7316-7320.

Engelman, A., Bushman, F. D., and Craigie, R., *Identification of discrete functional domains of HIV-1 integrase and their organization within an active multimeric complex.* Embo J, 1993. **12**(8): p. 3269-3275.

Engelman, A., and Craigie, R., *Identification of conserved amino acid residues critical for human immunodeficiency virus type 1 integrase function in vitro.* J Virol, 1992. **66**(11): p. 6361-6369.

Engelman, A., Mizuuchi, K., and Craigie, R., *HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer.* Cell, 1991. **67**(6): p. 1211-1221.

Esposito, D., and Craigie, R., *Sequence specificity of viral end DNA binding by HIV-1 integrase reveals critical regions for protein-DNA interaction.* Embo J, 1998. **17**(19): p. 5832-5843.

Fayet, O., Ramond, P., Polard, P., Prere, M. F., and Chandler, M., *Functional similarities between retroviruses and the IS3 family of bacterial insertion sequences?* Mol Microbiol, 1990. **4**(10): p. 1771-1777.

Fierke, C. A., and Hammes, G. G., *Transient kinetic approaches to enzyme mechanisms.* Methods Enzymol, 1995. **249**: p. 3-37.

Fitzgerald, M. L., and Grandgenett, D. P., *Retroviral integration: in vitro host site selection by avian integrase.* J Virol, 1994. **68**(7): p. 4314–4321.

Fitzgerald, M. L., Vora, A. C., and Grandgenett, D. P., *Development of an acid–soluble assay for measuring retrovirus integrase 3'–OH terminal nuclease activity.* Anal Biochem, 1991. **196**(1): p. 19–23.

Fitzgerald, M. L., Vora, A. C., Zeh, W. G., and Grandgenett, D. P., *Concerted integration of viral DNA termini by purified avian myeloblastosis virus integrase.* J Virol, 1992. **66**(11): p. 6257–6263.

Flint, S. J., Enquist, L.W., Krug, R.M., Racaniello, V.R., Skalka, A.M., *Principles of Virology.* 2000, Washington, D.C.: ASM Press. 804.

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

Gallay, P., Swingler, S., Song, J., Bushman, F., and Trono, D., *HIV nuclear import is governed by the phosphotyrosine–mediated binding of matrix to the core domain of integrase.* Cell, 1995. **83**(4): p. 569–576.

Gill, S. C., and von Hippel, P. H., *Calculation of protein extinction coefficients from amino acid sequence data.* Anal Biochem, 1989. **182**(2): p. 319–326.

Goldgur, Y., Dyda, F., Hickman, A. B., Jenkins, T. M., Craigie, R., and Davies, D. R., *Three new structures of the core domain of HIV–1 integrase: an active site that binds magnesium.* Proc Natl Acad Sci U S A, 1998. **95**(16): p. 9150–9154.

Grandgenett, D. P., Inman, R. B., Vora, A. C., and Fitzgerald, M. L., *Comparison of DNA binding and integration half–site selection by avian myeloblastosis virus integrase.* J Virol, 1993. **67**(5): p. 2628–2636.

Grandgenett, D. P., and Mumm, S. R., *Unraveling retrovirus integration.* Cell, 1990. **60**(1): p. 3–4.

Heuer, T. S., and Brown, P. O., *Mapping features of HIV-1 integrase near selected sites on viral and target DNA molecules in an active enzyme-DNA complex by photo-cross-linking.* Biochemistry, 1997. **36**(35): p. 10655-10665.

Heuer, T. S., and Brown, P. O., *Photo-cross-linking studies suggest a model for the architecture of an active human immunodeficiency virus type 1 integrase-DNA complex.* Biochemistry, 1998. **37**(19): p. 6667-6678.

Hindmarsh, P., and Leis, J., *Retroviral DNA integration.* Microbiol Mol Biol Rev, 1999. **63**(4): p. 836-843.

Hishinuma, F., DeBona, P. J., Astrin, S., and Skalka, A. M., *Nucleotide sequence of acceptor site and termini of integrated avian endogenous provirus ev1: integration creates a 6 bp repeat of host DNA.* Cell, 1981. **23**(1): p. 155-164.

Jeggo, P. A., Taccioli, G. E., and Jackson, S. P., *Menage a trois: double strand break repair, V(D)J recombination and DNA- PK.* Bioessays, 1995. **17**(11): p. 949-957.

Jenkins, T. M., Esposito, D., Engelman, A., and Craigie, R., *Critical contacts between HIV-1 integrase and viral DNA identified by structure-based analysis and photo-crosslinking.* Embo J, 1997. **16**(22): p. 6849-6859.

Johnson, K. A., *Conformational coupling in DNA polymerase information transfer.* Philos Trans R Soc Lond B Biol Sci, 1992. **336**(1276): p. 107-112.

Johnson, K. A., *Rapid quench kinetic analysis of polymerases, adenosinetriphosphatases, and enzyme intermediates.* Methods Enzymol, 1995. **249**: p. 38-61.

Jones, K. S., Coleman, J., Merkel, G. W., Laue, T. M., and Skalka, A. M., *Retroviral integrase functions as a multimer and can turn over catalytically.* J Biol Chem, 1992. **267**(23): p. 16037-16040.

Katz, R. A., DiCandeloro, P., Kukolj, G., and Skalka, A. M., *Role of DNA end distortion in catalysis by avian sarcoma virus integrase.* J Biol Chem, 2001. **276**(36): p. 34213-34220.

Katz, R. A., Gravuer, K., and Skalka, A. M., *A preferred target DNA structure for retroviral integrase in vitro.* J Biol Chem, 1998. **273**(37): p. 24190-24195.

Katz, R. A., Merkel, G., Kulkosky, J., Leis, J., and Skalka, A. M., *The avian retroviral IN protein is both necessary and sufficient for integrative recombination in vitro.* Cell, 1990. **63**(1): p. 87–95.

Katz, R. A., Merkel, G., and Skalka, A. M., *Targeting of retroviral integrase by fusion to a heterologous DNA binding domain: in vitro activities and incorporation of a fusion protein into viral particles.* Virology, 1996. **217**(1): p. 178–190.

Katz, R. A., and Skalka, A. M., *The retroviral enzymes.* Annu Rev Biochem, 1994. **63**: p. 133–173.

Katzman, M., Katz, R. A., Skalka, A. M., and Leis, J., *The avian retroviral integration protein cleaves the terminal sequences of linear viral DNA at the in vivo sites of integration.* J Virol, 1989. **63**(12): p. 5319–5327.

Khan, E., Mack, J. P., Katz, R. A., Kulkosky, J., and Skalka, A. M., *Retroviral integrase domains: DNA binding and the recognition of LTR sequences.* Nucleic Acids Res, 1991a. **19**(4): p. 851–860.

Khan, E., Mack, J. P., Katz, R. A., Kulkosky, J., and Skalka, A. M., *Retroviral integrase domains: DNA binding and the recognition of LTR sequences [published erratum appears in Nucleic Acids Res 1991 Mar 25; 19(6):1358].* Nucleic Acids Res, 1991b. **19**(4): p. 851–860.

Kukolj, G., Jones, K. S., and Skalka, A. M., *Subcellular localization of avian sarcoma virus and human immunodeficiency virus type 1 integrases.* J Virol, 1997. **71**(1): p. 843–847.

Kukolj, G., and Skalka, A. M., *Enhanced and coordinated processing of synapsed viral DNA ends by retroviral integrases in vitro.* Genes Dev, 1995. **9**(20): p. 2556–2567.

Kulkosky, J., Jones, K. S., Katz, R. A., Mack, J. P., and Skalka, A. M., *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, 1992. **12**(5): p. 2331–2338.

Kulkosky, J., Katz, R. A., Merkel, G., and Skalka, A. M., *Activities and substrate specificity of the evolutionarily conserved central domain of retroviral integrase.* Virology, 1995. **206**(1): p. 448–456.

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

Lodi, P. J., Ernst, J. A., Kuszewski, J., Hickman, A. B., Engelman, A., Craigie, R., Clore, G. M., and Gronenborn, A. M., *Solution structure of the DNA binding domain of HIV–1 integrase.* Biochemistry, 1995. **34**(31): p. 9826–9833.

Mazumder, A., Wang, S., Neamati, N., Nicklaus, M., Sunder, S., Chen, J., Milne, G. W., Rice, W. G., Burke, T. R., Jr., and Pommier, Y., *Antiretroviral agents as inhibitors of both human immunodeficiency virus type 1 integrase and protease.* J Med Chem, 1996. **39**(13): p. 2472–2481.

Melek, M., Gellert, M., and van Gent, D. C., *Rejoining of DNA by the RAG1 and RAG2 proteins.* Science, 1998. **280**(5361): p. 301–303.

Melek, M., Jones, J. M., O'Dea, M. H., Pais, G., Burke, T. R., Jr., Pommier, Y., Neamati, N., and Gellert, M., *Effect of HIV integrase inhibitors on the RAG1/2 recombinase.* Proc Natl Acad Sci U S A, 2002. **99**(1): p. 134–137.

Miller, M. D., Farnet, C. M., and Bushman, F. D., *Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition.* J Virol, 1997. **71**(7): p. 5382–5390.

Muesing, M. A., Smith, D. H., Cabradilla, C. D., Benton, C. V., Lasky, L. A., and Capon, D. J., *Nucleic acid structure and expression of the human AIDS/lymphadenopathy retrovirus.* Nature, 1985. **313**(6002): p. 450–458.

Muller, B., Jones, K. S., Merkel, G. W., and Skalka, A. M., *Rapid solution assays for retroviral integration reactions and their use in kinetic analyses of wild–type and mutant Rous sarcoma virus integrases.* Proc Natl Acad Sci U S A, 1993. **90**(24): p. 11633–11637.

Muller, H. P., and Varmus, H. E., *DNA bending creates favored sites for retroviral integration: an explanation for preferred insertion sites in nucleosomes.* Embo J, 1994. **13**(19): p. 4704–4714.

Mumm, S. R., and Grandgenett, D. P., *Defining nucleic acid–binding properties of avian retrovirus integrase by deletion analysis.* J Virol, 1991. **65**(3): p. 1160–1167.

Murphy, J. E., and Goff, S. P., *A mutation at one end of Moloney murine leukemia virus DNA blocks cleavage of both ends by the viral integrase in vivo.* J Virol, 1992. **66**(8): p. 5092–5095.

Patel, S. S., Wong, I., and Johnson, K. A., *Pre–steady–state kinetic analysis of processive DNA replication including complete characterization of an exonuclease–deficient mutant.* Biochemistry, 1991. **30**(2): p. 511–525.

Peliska, J. A., and Benkovic, S. J., *Mechanism of DNA strand transfer reactions catalyzed by HIV–1 reverse transcriptase.* Science, 1992. **258**(5085): p. 1112–1118.

Pemberton, I. K., Buckle, M., and Buc, H., *The metal ion–induced cooperative binding of HIV–1 integrase to DNA exhibits a marked preference for Mn(II) rather than Mg(II).* J Biol Chem, 1996. **271**(3): p. 1498–1506.

Pemberton, I. K., and Singh, O. M., *In vitro biochemical activities associated with recombinant HIV–1 integrase.* Biochem Soc Trans, 1993. **21**(1): p. 51S.

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

Puras Lutzke, R. A., Eppens, N. A., Weber, P. A., Houghten, R. A., and Plasterk, R. H., *Identification of a hexapeptide inhibitor of the human immunodeficiency virus integrase protein by using a combinatorial chemical library.* Proc Natl Acad Sci U S A, 1995. **92**(25): p. 11456–11460.

Ramsden, D. A., McBlane, J. F., van Gent, D. C., and Gellert, M., *Distinct DNA sequence and structure requirements for the two steps of V(D)J recombination signal cleavage.* Embo J, 1996. **15**(12): p. 3197–3206.

Ratcliff, G. C., and Erie, D. A., *A novel single–molecule study to determine protein––protein association constants.* J Am Chem Soc, 2001. **123**(24): p. 5632–5635.

Reid, T., and Wilson, I., in *The Enzymes*, P. D. Boyer, Editor. 1971, Academic Press: New York. p. 373–416.

Richards, F. M., and Wyckoff, H. W., in *The Enzymes*, P. D. Boyer, Editor. 1971, Academic Press: New York. p. 647–806.

Roth, D. B., and Craig, N. L., *VDJ recombination: a transposase goes to work.* Cell, 1998. **94**(4): p. 411–414.

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

Scottoline, B. P., Chow, S., Ellison, V., and Brown, P. O., *Disruption of the terminal base pairs of retroviral DNA during integration.* Genes Dev, 1997. **11**(3): p. 371–382.

Sherman, P. A., Dickson, M. L., and Fyfe, J. A., *Human immunodeficiency virus type 1 integration protein: DNA sequence requirements for cleaving and joining reactions.* J Virol, 1992. **66**(6): p. 3593–3601.

Shih, C. C., Stoye, J. P., and Coffin, J. M., *Highly preferred targets for retrovirus integration.* Cell, 1988. **53**(4): p. 531–537.

Skinner, L. M., Sudol, M., Harper, A. L., and Katzman, M., *Nucleophile selection for the endonuclease activities of human, ovine, and avian retroviral integrases.* J Biol Chem, 2001. **276**(1): p. 114–124.

Starcich, B., Ratner, L., Josephs, S. F., Okamoto, T., Gallo, R. C., and Wong–Staal, F., *Characterization of long terminal repeat sequences of HTLV–III.* Science, 1985. **227**(4686): p. 538–540.

Terry, R., Soltis, D. A., Katzman, M., Cobrinik, D., Leis, J., and Skalka, A. M., *Properties of avian sarcoma–leukosis virus pp32–related pol–endonucleases produced in Escherichia coli.* J Virol, 1988. **62**(7): p. 2358–2365.

van Gent, D. C., Vink, C., Groeneger, A. A., and Plasterk, R. H., *Complementation between HIV integrase proteins mutated in different domains.* Embo J, 1993. **12**(8): p. 3261–3267.

Vincent, K. A., Ellison, V., Chow, S. A., and Brown, P. O., *Characterization of human immunodeficiency virus type 1 integrase expressed in Escherichia coli and analysis of variants with amino–terminal mutations.* J Virol, 1993. **67**(1): p. 425–437.

Vink, C., Groenink, M., Elgersma, Y., Fouchier, R. A., Tersmette, M., and Plasterk, R. H., *Analysis of the junctions between human immunodeficiency virus type 1 proviral DNA and human DNA.* J Virol, 1990. **64**(11): p. 5626–5627.

Vink, C., Oude Groeneger, A. M., and Plasterk, R. H., *Identification of the catalytic and DNA–binding region of the human immunodeficiency virus type I integrase protein.* Nucleic Acids Res, 1993. **21**(6): p. 1419–1425.

Vink, C., Yeheskiely, E., van der Marel, G. A., van Boom, J. H., and Plasterk, R. H., *Site–specific hydrolysis and alcoholysis of human immunodeficiency virus DNA termini mediated by the viral integrase protein.* Nucleic Acids Res, 1991. **19**(24): p. 6691–6698.

Vora, A., and Grandgenett, D. P., *DNase protection analysis of retrovirus integrase at the viral DNA ends for full–site integration in vitro.* J Virol, 2001. **75**(8): p. 3556–3567.

Vora, A. C., Chiu, R., McCord, M., Goodarzi, G., Stahl, S. J., Mueser, T. C., Hyde, C. C., and Grandgenett, D. P., *Avian retrovirus U3 and U5 DNA inverted repeats. Role Of nonsymmetrical nucleotides in promoting full–site integration by purified virion and bacterial recombinant integrases.* J Biol Chem, 1997. **272**(38): p. 23938–23945.

Vora, A. C., and Grandgenett, D. P., *Assembly and catalytic properties of retrovirus integrase–DNA complexes capable of efficiently performing concerted integration.* J Virol, 1995. **69**(12): p. 7483–7488.

Wang, J. Y., Ling, H., Yang, W., and Craigie, R., *Structure of a two–domain fragment of HIV–1 integrase: implications for domain organization in the intact protein.* Embo J, 2001. **20**(24): p. 7333–7343.

Woerner, A. M., and Marcus-Sekura, C. J., *Characterization of a DNA binding domain in the C-terminus of HIV-1 integrase by deletion mutagenesis.* Nucleic Acids Res, 1993. **21**(15): p. 3507-3511.

Wong, I., and Lohman, T. M., *A double-filter method for nitrocellulose-filter binding: application to protein-nucleic acid interactions.* Proc Natl Acad Sci U S A, 1993. **90**(12): p. 5428-5432.

Wong, I., Patel, S. S., and Johnson, K. A., *An induced-fit kinetic mechanism for DNA replication fidelity: direct measurement by single-turnover kinetics.* Biochemistry, 1991. **30**(2): p. 526-537.

Xue, Y., Ratcliff, G. C., Wang, H., Davis-Searles, P. R., Gray, M. D., Erie, D. A., and Redinbo, M. R., *A minimal exonuclease domain of WRN forms a hexamer on DNA and possesses both 3'- 5' exonuclease and 5'-protruding strand endonuclease activities.* Biochemistry, 2002. **41**(9): p. 2901-2912.

Yang, Z. N., Mueser, T. C., Bushman, F. D., and Hyde, C. C., *Crystal structure of an active two-domain derivative of Rous sarcoma virus integrase.* J Mol Biol, 2000. **296**(2): p. 535-548.