HOST PROTEINS IN RETROVIRAL cDNA INTEGRATION

Frederic D. Bushman

Infectious Disease Laboratory
Salk Institute for Biological Studies
La Jolla, California 92037

I. Introduction

Most viruses encode relatively few of the proteins required for their own replication, relying instead on the host cell to provide the necessary machinery. All viruses pirate the host translational apparatus and cellular precursors for macromolecular synthesis. Many other host systems are also subverted to support viral replication. Host cell surface proteins act as receptors, and host molecular chaperones direct protein folding (29). Relatively small viruses such as retroviruses encode relatively few proteins and rely on the host to provide many additional functions, such as apparatus for transcription and RNA processing (23).

For the case of human immunodeficiency virus (HIV), several exciting insights have come with the identification of new host factors important for replication. For example, the discovery that chemokine receptors act as coreceptors for HIV has explained key aspects of viral tropism and resistance (3,54). The finding that the host-encoded cyclophilin protein is important for replication of some HIV-1 strains has led to new models for uncoating of the viral core (38,55,73). The observation

0065-3527/99 $30.00

that HIV replicates in nondividing cells implicates host proteins in nuclear import in nondividing cells (10,32,36,37,42,75). This review focuses on host proteins potentially important for another step in the viral life cycle: the integration of retroviral cDNA into host DNA. For recent reviews of retroviral cDNA integration see Refs. 20, 23, 25, 45, and this volume.

II. Assaying cDNA Integration *in Vitro*

A. Purified Integrase Protein

The pathway of integration *in vivo* is presented in Fig. 7.1A and discussed extensively elsewhere in this volume. The early steps of this

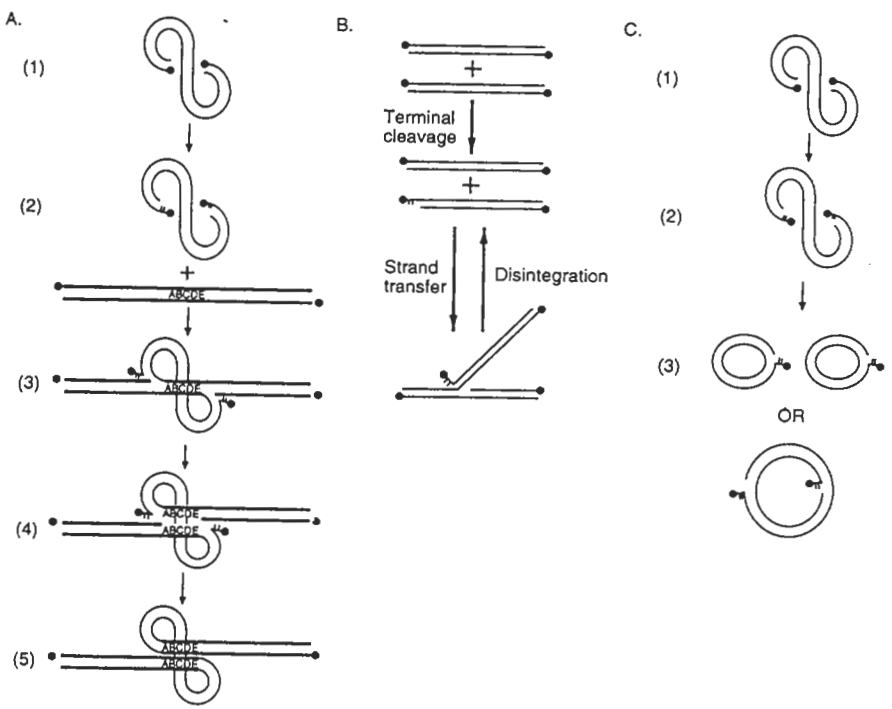

Fig 7.1. Integration and autointegration reactions. (A) Diagram of the inferred pathway of cDNA integration *in vivo*. Dots indicate DNA 5′ ends. Thin lines, retroviral cDNA; thick lines, target DNA. (B) Pathway of single-ended integration. (C) Pathway of autointegration. Integration into either the same strand or the opposite strand of the cDNA yields the pair of subgenomic circles or the inverted circle products, respectively.

pathway can be reproduced *in vitro*, and such assay systems have been used to identify and characterize host factors potentially important for the integration step.

The simplest are assays using integrase purified from cells engineered to overexpress the protein. Purified DNAs can be used as mimics of the viral cDNA end (long terminal repeat, LTR) and target DNA. Incubation of integrase with model substrates results in the cleavage of the LTR termini (24,40,48,69) and covalent joining of LTR DNA to target DNA (16,24,46) (Fig. 7.1B). These assays have the advantage of speed and convenience and have been used extensively in characterizing the function of integrase. Assays with purified integrase have also been used to identify many inhibitors of potential use for treating HIV infection (see the article by Pommier in this volume). However, it is also clear that assays with purified integrase do not fully recapitulate some aspects of integration *in vivo*. For this reason, it has also been useful to study integration *in vitro* using preintegration complexes (PICs), authentic replication intermediates isolated from freshly infected cells.

B. Preintegration Complexes

Another method for carrying out integration *in vitro* begins with retroviral stocks instead of overexpressed integrase protein. Brown and co-workers showed that it is possible to prepare extracts of cells freshly infected with a retrovirus that contained unintegrated cDNA bound to proteins. When presented with a naked target DNA *in vitro*, these PICs were able to integrate the viral cDNA into target DNA, yielding the integration intermediate shown in Fig. 7.1A (step 4) (8). Evidently, the final DNA repair steps (steps 4 and 5 in Fig. 7.1A) were not carried out efficiently in the extracts studied.

Preintegration complexes differ from the complexes prepared to date with purified integrase in several respects. Preintegration complexes always carry out coupled reactions in which pairs of cDNA ends are joined to target DNA with the correct spacing. Reactions with some purified integrase proteins carry this reaction out inefficiently, yielding a substantial fraction of single-ended joining products (Fig. 7.1B). Purified HIV integrase is particularly poor at coupled integration, although avian sarcoma leukosis virus (ASLV) integrase and Moloney murine leukosis virus (MoMLV) integrase are more efficient (2,13,14,16,30,31,46,76). Assays with PICs also differ from assays with purified integrase in that they are less sensitive to small

molecule inhibitors, an observation potentially useful for studying inhibitors of integration (28). Furthermore, at least in the HIV case, PICs differ from complexes formed with purified integrase in their preferences for different target DNA sites (6,18).

In the HIV system, several proteins have been found to cofractionate with PICs in addition to integrase (10,26,27,36), including the viral matrix and reverse transcriptase proteins (10,27,35,36,42,57). The viral capsid protein is present at low levels, if at all, in PICs of HIV. An earlier proposal that capsid protein might be present in PICs of MoMLV has been called into question (7); further data on this point would be helpful. Other viral proteins have been found in crude fractions containing PICs, although their degree of association is unclear. Active HIV PICs treated with detergent have been reported to contain integrase as the only viral protein present and retaining some integration activity (27). At least one cellular protein, HMG/ IY, is reported to be present in PICs, and is implicated in PIC function (see below).

The organization of PICs is beginning to be clarified. Preintegration complexes are compact particles (7). The diameter of PICs measured by gel filtration is 54 n*M*, but the 10-kb HIV cDNA has a contour length of 6.6 μm, indicating that the cDNA in PICs is present in a condensed state (57). The cDNA ends in PICs are protected by bound proteins from attack by added nucleases or recombination complexes (57,77). The cDNA ends are held together by a protein bridge, as indicated by the finding that the cDNA can be cleaved internally with a restriction enzyme, but such complexes nevertheless remain capable of coordinated integration (57). The complexes at the cDNA ends are large and likely contain integrase (57,77). The more internal cDNA is at least partially exposed, indicative of looser packing than in complexes at the cDNA ends (57,77). Evidently, the PIC represents a remodeled derivative of the viral core, transformed by reverse transcription, release of certain viral proteins, and acquisition of cellular proteins.

III. Host Proteins Involved in Early Steps of cDNA Integration

A. Host Proteins Stimulating Integration by PICs

Much interest has centered on the question of whether host proteins are important for integration *in vivo*. This article will first review proposals for stimulatory proteins arising from *in vitro* studies of PICs and then review studies employing purified integrase. Proteins important for completing integration and proteins influencing integration by binding target DNA will be considered in the following sections.

In one study, PICs of HIV-1 were studied using a depletion/add back procedure (26). It was found that treatment of PICs with high concentrations of salt followed by gel filtration eliminated integration activity, although some of the integrase still remain associated with the cDNA. Addition of extracts from uninfected cells resulted in restoration of activity. This provided an assay method that permitted the purification of the stimulatory host factor, which turned out to be the high-mobility group protein HMG I(Y). Importantly, this small DNA binding protein was found to be associated with PICs prior to treatment with the high concentrations of salt. Complementing activity could be obtained from preparations of PICs and the activity immunodepleted with antibodies against HMG I(Y) but not with antibodies against several other proteins. HMG I(Y) purified from an *Escherichia coli* expression system also stimulated activity. A number of other DNA binding proteins failed to stimulate integration. Of the proteins tested, only the HIV NC protein showed partial activity, although this protein could not be detected in the PICs studied and an anti-NC antibody could not deplete the activity stripped from PICs, arguing against the importance of NC in this context.

B. Host Proteins Blocking Autointegration by PICs

Following the completion of reverse transcription, the PIC is presented with a dangerous challenge. Various assays reveal that the viral cDNA itself can be used as an integration target (Fig. 7.1C). Because this results in a rearrangement of the viral genome and blocks subsequent intermolecular integration, this "autointegration" reaction is suicidal for the retrovirus.

Studies of PICs of MoMLV have led to the proposal that a host protein may be important for blocking autointegration (51). Lee and Craigie treated PICs of MoMLV with high concentrations of salt and subjected them to gel filtration under these conditions (the first use of a salt-stripping approach). Such salt-stripped complexes were found to be incapable of normal intermolecular integration, but autointegration activity was actually promoted. Addition of extract from virions failed to block autointegration. However, addition of cellular extract did block autointegration and restored intermolecular integration, implicating a possible host factor. The factor that blocks autointegration has been isolated and named "barrier to autointegration factor" (BAF) (52). The protein is similar to proteins predicted from DNA sequences of many organisms, but its function has not yet been studied outside the retrovi-

ral work. An unaddressed issue is whether BAF is present in PICs and is blocking autointegration prior to salt stripping. According to one extreme theory, the factor removed from MoMLV PICs by salt stripping may not be BAF but a different protein with a similar function. The methods and reagents developed in the above study should allow this issue to be addressed experimentally.

What is the relationship between the factors identified by salt stripping-reconstitution methods in the HIV and MoMLV cases? A study examined host activities capable of stimulating intermolecular integration by salt-stripped MoMLV PICs (53). In this case, as in the previous work, salt stripping was found to block intermolecular integration and a host factor to restore it. A stimulatory activity was purified and found, as in the HIV case, to be HMG I(Y). Reconstitution of salt-stripped MoMLV PICs with HMG I(Y) had no effect on autointegration, which remained prominent even after addition of HMG I(Y). Evidently BAF can reprogram autointegration into intermolecular integration, while HMG I(Y) stimulates only intermolecular integration.

Oddly, in the HIV case, salt stripping was not found to stimulate autointegration. At least under the conditions studied, salt-stripped HIV PICs did not carry out any integration reactions prior to reconstitution. The basis for this difference between the HIV and MoMLV systems remains unclear.

C. Proteins Influencing Reactions with Purified Integrase

A variety of proteins have been found to stimulate the activity of purified integrase proteins *in vitro,* providing further candidates for host proteins important for integration. An early observation in this area came from analysis of catalysis by integrase from MoMLV virions. In this study it was found that an extract from uninfected host cells stimulated recovery of integration products roughly 50-fold (34). However, analysis of the stimulating activity did not turn out to be straightforward, as it was subsequently found that several DNA binding proteins could also stimulate in this assay. Later studies of purified MoMLV integrase revealed that integration could be stimulated 3- to 13-fold by *E. coli* SSB, RNase A, and phage T4 gene 32 protein (52; R. Craigie and F. Bushman, unpublished). The fact that so many proteins stimulated integration in this assay made it difficult to attribute *in vivo* importance to any particular one.

Several more recent studies have identified stimulatory host proteins with activities suggestive of possible *in vivo* roles. A screen for proteins

that bind tightly to integrase yielded Ini-1, a component of the SWI–SNF complex which is thought to be important for modifying chromatin structure to facilitate transcription. Addition of Ini-1 to reactions *in vitro* containing purified HIV-1 integrase under certain conditions yielded a stimulation of 10- to 20-fold. This might fit with previous proposals that integration of retroviral cDNA is favored near actively transcribed genes, which may be chromosomal regions enriched for Ini-1. However, some studies of integration site selection have not revealed a strong connection between transcription and integration (18,78), calling this component of the model into question. Furthermore, as with the MoMLV case, diverse proteins have been found to stimulate the activity of HIV-1 integrase *in vitro* [such as RNAse T1 (56), NC protein (17), or a mixture of DNA binding proteins (16)]. Thus the results with Ini-1, although provocative, need to be bolstered with further experiments to strengthen the case for importance *in vivo.*

The HMG-1 protein (not to be confused with HMG I(Y), which differs in sequence and function) was found by Aiyar *et al.* to stimulate purified ASLV integrase 2- to 4-fold (2). Order of addition experiments suggested that HMG-1 acts on the integrase–LTR complex and not on the target DNA. Another DNA binding protein tested, *E. coli* HU, did not stimulate as strongly. Thus, HMG-1 stands as another candidate for a stimulatory host cofactor.

What are the relationships among the proteins acting on PICs and purified integrase? HMG I(Y) seems to differ in function from the proteins that stimulate integrase since, for example, HMG-1 cannot substitute for HMG I(Y) in reconstituting salt-stripped PICs (26). However, it is possible that proteins that stimulate purified integrase reflect another activity important for function *in vivo.* Of interest on this point is the finding that PICs of MoMLV that are partially purified but not salt stripped can in some cases be boosted in activity by HMG-1, NC, RNase A, or *E. coli* SSB, possibly mimicking the stimulation seen in reactions with purified integrase (53). The effects of BAF on purified integrase have not yet been reported. It will be useful to determine which of the candidate proteins are present in PICs in addition to HMG I(Y) and, if possible, design experiments to test for stimulatory function *in vivo.* Ultimately, a more detailed understanding of the mechanism of stimulation in each of the experimental paradigms should clarify their relationship and significance.

In a related series of experiments, proteins bound to the MoMLV LTRs were examined after salt stripping and reconstitution using a recombination-based footprinting method (77). Extract from uninfected host cells was found to be important for reconstituting the footprint

after salt stripping, but purified HMG I(Y) or HMG-1 was not sufficient. Evidently, further host factors are important for reassembling PICs in this experimental paradigm.

IV. Host Proteins Involved in Late Stages of Integration Reaction

Integrase joins the recessed 3′ LTR ends to protruding 5′ ends in target DNA (Fig. 7.1A, step 3). Unpairing of the host DNA sequences between the points of joining results in the formation of DNA gaps at each host–virus DNA junction (Fig. 7.1A, step 4). The two nucleotides of the viral cDNA 5′ end will also be unpaired, forming a free "frayed" end. Completion of the integration reaction requires polymerization across the gap, removal of the frayed viral 5′ end, and sealing of the new DNA strand by ligation (Fig. 7.1A, step 5). The enzymes that carry out this reaction are unknown, and the detailed pathway of viral 5′ end joining is unclear.

Perhaps the simplest view is that host DNA repair enzymes are responsible for the late steps of integration. DNA repair systems are ubiquitous and highly active. For example, it is estimated that 10,000 sites of spontaneous depurination need to be repaired per genome per day. Known repair enzymes seem to have the activities needed for repair of gaps at the host–virus DNA junctions (50). However, little evidence is available on the specific enzymes involved. One study reported that inhibition of the poly(ADP-ribose) polymerase, an enzyme apparently involved in regulating DNA repair, caused a reduction in viral titers possibly accounted for by effects on integration. However, clouding interpretation of this result is the lack of direct evidence for an effect on gap repair and unresolved questions on the role of the poly(ADP-ribose) polymerase itself in repair (33).

An interesting alternative mechanism for the late steps is that the virus-encoded reverse transcriptase and integrase proteins carry out these reactions. Possibly reverse transcriptase fills in the gapped region, and integrase seals the strand by a DNA-splicing reaction reminiscent of the reverse reaction of integration (disintegration)(22). Consistent with this possibility, both reverse transcriptase and integrase are reportedly present in PICs. This interesting model has not yet been tested decisively.

A report made the remarkable suggestion that integrase itself has a DNA polymerase activity, and that the parts of integrase responsible for this activity may differ from those at the known active site (1). This surprising result awaits conformation in other laboratories.

V. Proteins Influencing Target Site Selection

Integration *in vivo* does not take place into naked target DNA but rather into DNA complexed with proteins in chromatin. The selection of the location for cDNA integration is only partially understood, despite the importance of this process for both the virus and the host (23). From the perspective of the host, it is clear the retroviral integration can activate oncogenes or inactive required host genes, leading to oncogenesis or other pathology. From the perspective of the virus, selection of the integration site is important as genomes integrated into different locations can be transcribed at different rates. Diverse experiments *in vitro* reveal the influence of bound proteins in model targets, and further experiments reveal the importance of the chromatin environment *in vivo*. Studies in these areas are reviewed in turn below.

A. Effects of Primary Sequence in Target DNA in Vitro

Studies *in vivo* and *in vitro* show that integration can take place in most phosphodiesters in a target DNA, although its efficiency can vary, depending on the primary sequence (23). In two studies the sequences found most commonly flanking *in vivo* integration sites for HIV-1 were compiled, revealing a consensus sequence of GT(A/T)AC (18,72). Analogous studies of other retroviruses have also found weakly favored primary sequences (30,65). When the HIV-1 consensus sequence was synthesized and tested as an integration target for HIV-1 PICs, these sites were found to be somewhat favored for integration, revealing a modest effect of optimal primary sequence (18). However, *in vivo* the influence of the target DNA sequence is likely to be secondary to that of bound proteins.

B. Effects of Proteins Bound to Target DNAs in Vitro

Several studies *in vitro* indicate that association of a DNA binding protein with its recognition site can reduce integration at that site. Evidently, a protein bound to target DNA can block integration by simple steric hindrance (11,12,39,47,64).

Interestingly, binding of some proteins that distort the target DNA structure can actually promote integration. The first evidence for this view came from studies of integration into nucleosomal targets *in vitro*. It was found, somewhat surprisingly, that wrapping target DNA in nucleosomes actually promoted integration (63–65). Later studies us-

ing a different type of nucleosome preparation revealed that integration was most favored at certain major groove sites that were the most opened or kinked (61,62).

Analysis of integration into other types of protein–DNA complexes extended these observations. Complexes containing bent DNA formed by binding cAMP activation protein (CAP) protein (59) or integration host factor (IHF) protein (5) also showed enhanced integration at the known sites of maximal widening of the DNA major groove. However, another bent DNA complex, that containing lymphoid enhancer factor (LEF) protein, failed to show favored integration (5). The LEF protein lies over the most opened part of the DNA helix, in this case the minor groove. Evidently, opening of accessible major groove sites can promote the action of integrase.

A speculative model to accommodate these data holds that integrase distorts its DNA substrates during the catalytic cycle. Thus predistorting the substrate DNA may promote the reaction by lowering the activation energy for proceeding along the reaction pathway. Consistent with this theory, DNA distortion has been found or inferred to promote action of integrase in several further experimental contexts (15,21,68).

C. Effects of Chromosomal Proteins on Integration in Vivo

What host chromosomal features dictate selection of integration sites *in vivo*? The studies *in vitro* provide a series of candidates, but experiments *in vivo* are needed to strengthen the existing proposals. A number of reports have investigated this issue, and although some interesting hypotheses have been adduced, site selection *in vivo* remains incompletely characterized.

Early studies, primarily focusing on MoMLV, led to a model in which open chromatin regions at transcription units were favored. Integration was proposed to be favored near DNase I hypersensitive sites (66,74) or CpG islands (67), features proposed to be enriched near actively transcribed regions. Another study proposed that unusual host DNA structures were common near integration sites (58). A study of avian leukosis virus integration frequencies at several chromosomal sites failed to show any major differences among the regions studied (78), in contrast to an earlier report (70). Studies of HIV-1 have suggested that still other chromosomal features are favored. It has been proposed that integration may be favored near repetitive elements [including LINE (long interspersed nuclear elements)-1 elements (71) or Alu islands (72)] or topoisomerase cleavage sites (43).

A study investigated integration of HIV cDNA *in vivo* and identified centromeric heterochromatin as a disfavored target (18). In this study, 61 HIV-1 integration sites generated by experimental infection were sequenced and analyzed. For comparison, a library of control sequences from uninfected target cells was also prepared by similar methods and 104 sequences were determined. It was found that the frequency of recovery of transcription units, LINE elements, or Alu repeats was not clearly different between integration sites and controls. However, alphoid repeats, found only in centromeric heterochromatin, were absent from integration site libraries but present in the controls.

To examine this issue by another method, integration near repeated sequences was examined using repeat-specific polymerase chain reaction (PCR) assays. One PCR primer was complementary to a viral cDNA end, the other to a repeated sequence in human DNA. Detection of amplification products revealed integration near each repeat class studied. To isolate the effect of wrapping human DNA in chromatin, integration sites generated by infection *in vivo* were compared to integration sites made *in vitro* by integration of PICs into naked human DNA. It was found that integration was readily detectable near Alu elements or LINE elements in either sample, but integration was found near alphoid repeats only in samples from *in vitro* integration into deproteinized human DNA. No integration near alphoid repeats *in vivo* was detected by this method. Thus, two experimental approaches indicated that HIV cDNA integration is disfavored *in vivo* in centromeric heterochromatin.

Several lines of evidence indicate that centromeric heterochromatin is organized differently than euchromatin, potentially explaining the effects on integration site selection. Heterochromatic centromeres are seen to be more compact than euchromatin centromeres in fixed chromosome spreads (9), and alphoid sequences are more resistant to digestion with DNase I in isolated nuclei than are most DNAs (60,79). Alphoid repeats are associated with the centromere-specific proteins CENP-A, CENP-B, and CENP-C (60,79). These data provide unexpected evidence for the long-standing possibility that certain types of chromatin may obstruct cDNA integration. Similar experimental approaches may reveal further features of chromosomes that influence integration and suggest new hypotheses for tests *in vitro*.

D. Target Site Selection by Retrovirus-like Ty Transposons of Yeast

Studies of Ty retrotransposons in yeast reveal that integration can be highly site specific. The Ty retrotransposons multiply by transcription,

reverse transcription, and integration using reverse transcriptase and integrase enzymes similar to their retroviral counterparts (4). Ty elements differ from retroviruses by the absence of an extracellular phase. For this reason, Ty retrotransposons must be highly selective in their choice of integration sites; integration into a required cellular gene can be suicidal for the transposon.

Although about 70% of the *Saccharomyces cerevisiae* genome encodes proteins, integration by Ty elements into genes is rare. Ty1 integrates selectively several hundred base pairs upstream of host Pol III transcribed genes (44). Ty3 is most selective of all, integrating just at the start site of transcription of Pol III transcribed genes (19,49). Ty5 shows a different specificity, integrating in telomeric DNA and in silent mating cassette DNA (80,81). No such strong biases have yet been identified for retroviruses (18,41,66,67,72,74,78), although the work with yeast draws attention to this possibility. It will be interesting to see whether further mapping of retroviral integration sites combined with sequencing and analysis of the human genome yields evidence of unexpected biases in retroviral integration as well.

VI. Summary and Future Directions

Many host-encoded proteins seem likely to be important for retroviral cDNA integration. Candidate proteins have been proposed to stimulate integration, block autointegration, carry out late DNA repair steps, and modulate target site selection. Future studies will need to focus on strengthening the evidence for *in vivo* importance of each candidate and advancing our understanding of the mechanism.

References

1. Acel, A., Udashkin, B. E., Wainberg, M. A., and Faust, E. A. (1998). Efficient gap repair catalyzed in vitro by an intrinsic DNA polymerase activity of human immunodeficiency virus type 1 integrase. *J. Virol.* **72,** 2062–2071.
2. Aiyar, A., Hindmarsh, P., Skalka, A. M., and Leis. J. (1996). Concerted integration of linear retroviral DNA by the avain sarcoma virus integrase in vitro: Dependence on both long terminal repeat termini. *J. Virol.* **70,** 3571–3580.
3. Berger, E. A., (1997). HIV entry and tropism: The chemokine receptor connection. *AIDS* **11** (Suppl. A), S3–S16.
4. Boeke, J. D. (1989). Transposable elements in *Saccharomyces cerevisiae. In* (D. E. Berg, and M. M. Howe, eds.), "Mobile DNA" pp. 335–374. Am. Soc. Microbiol., Washington, DC.
5. Bor, Y.-C., Bushman, F., and Orgel. L. (1995). In vitro integration of human immunodeficiency virus type 1 cDNA into targets containing protein-induced bends. *Proc. Natl. Acad. Sci. U.S.A.* **92,** 10334–10338.

6. Bor, Y.-C., Miller, M., Bushman, F., and Orgel. L. (1996). Target sequence preferences of HIV-1 integration complexes in vitro. *Virology.* **222,** 238–242.
7. Bowerman, B., Brown, P. O., Bishop, J. M., and Varmus. H. E. (1989). A nucleoprotein complex mediates the integration of retroviral DNA. *Genes Dev.* **3,** 469–478.
8. Brown, P. O., Bowerman, B., Varmus, H. E. and Bishop, J. M. (1987). Correct integration of retroviral DNA in vitro. *Cell (Cambridge, Mass.)* **49:** 347–356.
9. Brown, S. W. (1966). Heterochromatin. *Science* **151,** 417–425.
10. Bukrinsky, M. I., Sharova, N., McDonald, T.L., Pushkarskaya, T., Tarpley, G.W., and Stevenson. M. (1993). Association of integrase, matrix, and reverse transcriptase antigens of human immunodeficiency virus type 1 with viral nucleic acids following acute infection. *Proc. Natl. Acad. Sci. U.S.A.* **90,** 6125–6129.
11. Bushman, F., and Miller, M. D. (1997). Tethering human immunodeficiency virus type 1 preintegration complexes to target DNA promotes integration at nearby sites. *J. Virol.* **71,** 458–464.
12. Bushman, F. D. (1994). Tethering human immunodeficiency virus 1 integrase to a DNA site directs integration to nearby sequences. *Proc. Natl. Acad. Sci. U.S.A.* **91,** 9233–9237.
13. Bushman, F. D., and Craigie, R. (1990). Sequence requirements for integration of Moloney murine leukemia virus DNA in vitro. *J. Virol.* **64,** 5645–5648.
14. Bushman, F. D., and Craigie, R. (1991). Activities of human immunodeficiency virus (HIV) integration protein *in vitro*: Specific cleavage and integration of HIV DNA. *Proc. Natl. Acad. Sci. U.S.A.* **88:** 1339–1343.
15. Bushman, F. D., and Craigie, R. (1992). Integration of human immunodeficiency virus DNA: Adduct interference analysis of required DNA sites. *Proc. Natl. Acad. Sci. U.S.A.* **89,** 3458–3462.
16. Bushman, F. D., Fujiwara, T., and Craigie, R. (1990). Retroviral DNA integration directed by HIV integration protein in vitro. *Science* **249,** 1555–1558.
17. Carteau, S., Batson, S. C., Poljak, L., Mouscadet, J.-F., Rocquigny, H., Darlix, J.-L., Roques, B. P., Kas, E., and Auclair, C. (1997). Human immunodeficiency virus type 1 nucleocapsid protein specifically stimulates Mg^{2+}-dependent DNA integration in vitro. *J. Virol.* **71,** 6225–6229.
18. Carteau, S., Hoffmann, C., and Bushman, F. D. (1998). Chromosome structure and HIV-1 cDNA integration: Centromeric alphoid repeats are a disfavored target. *J. Virol.* **72,** 4005–4014.
19. Chalker, D. L., and Sandmeyer, S. B. (1992). Ty3 integrates within the region of RNA polymerase III transcription initiation. *Genes Dev.* **6,** 117–128.
20. Chow, S. A. (1997). In vitro assays for activities of retroviral integrase. *Methods: Companion to Methods Enzymol.* **12,** 306–317.
21. Chow, S. A., and Brown, P. O. (1994). Substrate features important for recognition and catalysis by human immunodeficiency virus type 1 integrase identified by using novel DNA substrates. *J. Virol.* **68,** 3896–3907.
22. Chow, S. A., Vincent, K. A., Ellison, V., and Brown, P. O. 1992. Reversal of integration and DNA splicing mediated by integrase of human immunodeficiency virus. *Science* **255:** 723–726.
23. Coffin, J. M., Hughes, S. H., and Varmus, H. E. (1997). "Retroviruses." Cold Spring Harbor Lab. Press, Cold Spring Harbor, NY.
24. Craigie, R., Fujiwara, T., and Bushman, F. (1990). The IN protein of Moloney murine leukemia virus processes the viral DNA ends and accomplishes their integration in vitro. *Cell (Cambridge, Mass.)* **62,** 829–837.

25. Farnet, C. M., and Bushman, F. D. (1996). HIV cDNA integration: Molecular biology and inhibitor development. *AIDS* **10** (Suppl. A), 3–11.
26. Farnet, C., and Bushman, F. D. (1997). HIV-1 cDNA integration: Requirement of HMG I(Y) protein for function of preintegration complexes in vitro. *Cell (Cambridge, Mass.)* **88,** 1–20.
27. Farnet, C. M., and Haseltine, W. A. (1991). Determination of viral proteins present in the human immunodeficiency virus type 1 preintegration complex. *J. Virol.* **65,** 1910–1915.
28. Farnet, C., Lipford, R., Wang, B., and Bushman, F. D. (1996). Differential inhibition of HIV-1 preintegration complexes and purified integrase protein by small molecules. *Proc. Natl. Acad. Sci. U.S.A.* **93,** 9742–9747.
29. Fields, B. N., and Kinpe, D. M. (1990). "Virology." Raven Press. New York.
30. Fitzgerald, M. L., and Grandgenett, D. P. (1994). Retroviral integration: In vitro host site selection by avian integrase. *J. Virol.* **68,** 4314–4321.
31. Fitzgerald, M. L., Vora, A. C., Zeh, W. G., and Grandgenett, D. P. (1992). Concerted integration of viral DNA termini by purified avian myeloblastosis virus integrase. *J. Virol.* **66,** 6257–6263.
32. Fouchier, R. A. M., Meyer, B. E., Simon, J. H. M., Fischer, U., and Malim, M. H. (1997). HIV-1 infection of nondividing cells: Evidence that the amino-terminal basic region of the viral matrix protein is important for Gag processing but not for post-entry nuclear import. *EMBO J.* **16,** 4531–4539.
33. Friedberg, E. C., Walker, G. C., and Siede, W. (1995). "DNA Repair and Mutagenesis." Am. Soc. Microbiol., Washington, DC.
34. Fujiwara, T., and Craigie, R. (1989). Integration of mini-retroviral DNA: A cell-free reaction for biochemical analysis of retroviral integration. *Proc. Natl. Acad. Sci. U.S.A.* **86,** 3065–3069.
35. Gallay, P., Swingler, S., Aiken, C., and Trono, D. (1995). HIV-1 infection of nondividing cells: C-terminal phosphorylation of the viral matrix protein is a key regulator. *Cell (Cambridge, Mass.)* **80,** 379–388.
36. Gallay, P., Swingler, S., Song, J., Bushman, F., and Trono, D. (1995). HIV nuclear import is governed by the phosphtyrosine-mediated binding of matrix to the core domain of integrase. *Cell (Cambridge, Mass.)* **17,** 569–576.
37. Gallay, P., Hope, T., Chin, D., and Trono, D. (1997 . HIV-1 infection of nondividing cells through the recognition of integrase by the importin/karyopherin pathway. *Proc. Natl. Acad. Sci. U.S.A.* **94,** 9825–9830.
38. Gamble, T. R., Vajdos, F. F., Yoo, S., Worthylake, D. K., Houseweart, M., Sundquist, W. I., and Hill, C. P. (1996). Crystal structure of human cyclophillin A bound to the amino-terminal domain of HIV-1 capsid. *Cell (Cambridge, Mass.)* **87,** 1285–1294.
39. Goulaouic, H., and Chow S. A. (1996). Directed integration of viral DNA mediated by fusion proteins consisting of himan immunodeficiency virus type 1 integrase and *Escherichia coli* LexA protein. *J. Virol.* **70,** 37–46.
40. Grandgenett, D. P., Vora, A. C., and Schiff, R. D. (1978). A 32,000-dalton nucleic acid-binding protein from avian retrovirus cores possesses DNA endonuclease activity. *Virology* **89,** 119–132.
41. Hartung, S., Jaenisch, R., and Breindl, M. (1986). Retrovirus insertion inactivates mouse a1(I) collagen gene by blocking initiation of transcription. *Nature (London)* **320,** 365–367.
42. Heinzinger, N. K., Bukrinsky, M. I., Haggerty, S. A., Ragland, A. M., Kewalramani, V., Lee, M.-A., Gendelman, H. E., Ratner, L., Stevenson, M., and Emerman, M. (1994). The Vpr protein of human immunodeficiency virus type 1 influences nuclear localization of viral nucleic acids in nondividing host cells. *Proc. Natl. Acad. Sci. U.S.A.* **91,** 7311–7315.
43. Howard, M. T., and Griffith, J. D. (1993). A cluster of strong topoisomerase II cleavage sites is located near an integrated human immunodeficiency virus. *J. Mol. Biol.* **232,** 1060–1068.
44. Ji, H., Moore, D. P., Blomberg, M. A., Braiterman, L. T., Voytas, D. F., Natsoulis, G., and Boeke, J. D. (1993). Hotspots for unselected Ty1 transposition events on yeast chromosome III are near tRNA genes and LTR sequences. *Cell (Cambridge, Mass.)* **73,** 1–20.
45. Katz, R. A., and Skalka, A. M. (1994). The retroviral enzymes. *Annu. Rev. Biochem.* **63,** 133–173.
46. Katz, R. A., Merkel, G., Kulkosky, J., Leis, J., and Skalka, A. M. (1990). The avian retroviral IN protein is both necessary and sufficient for integrative recombination in vitro. *Cell (Cambridge, Mass.)* **63,** 87–95.
47. Katz, R. A., Merkel, G., and Skalka, A. M. (1996). 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* **217,** 178–190.
48. Katzman, M., Katz, R. A., Skalka, A. M., and Leis, J. (1989). The avian retroviral integration protein cleaves the terminal sequences of linear viral DNA at the in vivo sites of integration. *J. Virol.* **63,** 5319–5327.
49. Kirchner, J., Connolly, C. M., and Sandmeyer, S. B. (1995). In vitro position-specific integration of a retroviruslike element requires Pol III transcription factors. *Science* **267,** 1488–1491.
50. Kornberg, A., and Baker T. (1991). "DNA Replication." Freeman, New York.
51. Lee, M. S., and Craigie, R. (1994). Protection of retroviral DNA from autointegration: Involvement of a cellular factor. *Proc. Natl. Acad. Sci. U.S.A.* **91,** 9823–9827.
52. Lee, M. S., and Craigie, R. (1998). A previously unidentified host protein protects retroviral DNA from autointegration. *Proc. Natl. Acad. Sci. U.S.A.* **95,** 1528–1533.
53. Li, L., Farnet, C. M., Anderson, W. F., and Bushman, F. D. (1998). Modulation of activity of MoMuLV preintegration complexes by several host factors in vitro. *J. Virol.* **72,** 2125–2131.
54. Liu, R., Paxton, W. A., Choe, S., Ceradini, D., Martin, S. R., Horuk, R., MacDonald, M. E., Stuhlmann, H., Koup, R. A., and Landau, N. R. (1996). Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell (Cambridge, Mass.)* **86,** 367–377.
55. Luban, J., Bossolt, K. L., Franke, E. K., Kalpana, G. V., and Goff, S. P. (1993). Human immunodeficiency virus type 1 gag protein binds to cyclophilin A and B. *Cell (Cambridge, Mass.)* **73,** 1067–1078.
56. Miller, M., Bor, Y.-C., and Bushman, F. D. (1995). Target DNA capture by HIV-1 integration complexes. *Curr. Biol.* **5,** 1047–1056.
57. Miller, M. D., Farnet, C. M., and Bushman, F. D. (1997). Human immunodeficiency virus type 1 preintegration complexes: Studies of organization and composition. *J. Virol.* **71,** 5382–5390.
58. Milot, E., Belmaaza, A., Rassart, E., and Chartrand, P. (1994). Association of a host DNA structure with retroviral integration sites in chromosomal DNA. *Virology* **201,** 408–412.
59. Muller, H.-P., and Varmus, H. E. (1994). DNA bending creates favored sites for retroviral integration: An explanation for preferred insertion sites in nucleosomes. *EMBO J.* **13,** 4704–4714.

60. Pluta, A. R., Mackay, A. M., Ainsztein, A. M., Goldberg, I. G., and Earnshaw, W. C. (1995). The centromere: Hub of chromosomal activites. *Science* **270,** 1591–1594.
61. Pruss, D., Bushman, F. D., and Wolffe, A. P. (1994). Human immunodeficiency virus integrase directs integration to sites of severe DNA distortion within the nucleosome core. *Proc. Natl. Acad. Sci. U.S.A.* **91,** 5913–5917.
62. Pruss, D., Reeves, R., Bushman, F. D., and Wolffe, A. P. (1994). The influence of DNA and nucleosome structure on integration events directed by HIV integrase. *J. Biol. Chem.* **269,** 25031–25041.
63. Pryciak, P., Muller, H.-P., and Varmus, H. E. (1992). Simian virus 40 minichromosomes as targets for retroviral integration in vivo. *Proc. Natl. Acad. Sci. U.S.A.* **89,** 9237–9241.
64. Pryciak, P. M., and Varmus, H. E. (1992). Nucleosomes, DNA-binding proteins, and DNA sequence modulate retroviral integration target site selection. *Cell (Cambridge, Mass.)* **69,** 769–780.
65. Pryciak, P. M., Sil, A., and Varmus, H. E. (1992). Retroviral integration into minichromosomes in vitro. *EMBO J.* **11,** 291–303.
66. Rohdewohld, H., Weiher, H., Reik, W., Jaenisch, R., and Breindl, M. (1987). Retrovirus integration and chromatin structure: Moloney murine leukemia proviral integration sites map near DNase I-hypersensitive sites. *J. Virol.* **61,** 336.
67. Scherdin, U., Rhodes, K., and Breindl, M. (1990). Transcriptionally active genome regions are preferred targets for retrovirus integration. *J. Virol.* **64,** 907.
68. Scottoline, B. P., Chow, S., Ellison, V., and Brown, P. O. (1997). Disruption of the terminal base pairs of retroviral DNA during integration. *Genes Dev.* **11,** 371–382.
69. Sherman, P. A., and Fyfe, J. A. (1990). Human immunodeficiency virus integration protein expressed in *Escherichia coli* possesses selective DNA cleaving activity. *Proc. Natl. Acad. Sci. U.S.A.* **87,** 5119–5123.
70. Shih, C.-C., Stoye, J. P., and Coffin, J. M. (1988). Highly preferred targets for retrovirus integration. *Cell (Cambridge, Mass.)* **53,** 531–537.
71. Stevens, S. W., and Griffith, J. D. (1994). Human immunodeficiency virus type 1 may preferentially integrate into chromatin occupied by L1Hs repetitive elements. *Proc. Natl. Acad. Sci. U.S.A.* **91,** 5557–5561.
72. Stevens, S. W., and Griffith, J. D. (1996). Sequence analysis of the human DNA flanking sites of human immunodeficiency virus type 1 integration. *J. Virol.* **70,** 6459–6462.
73. Thali, M., Bukovsky, A., Kondo, E., Rosenwirth, B., Walsh, C. T., Sodroski, J., and Gottlinger, H. G. (1994). Functional association of cyclophilin A with HIV-1 virions. *Nature (London)* **372,** 363–365.
74. Vijaya, S., Steffan, D. L., and Robinson, H. L. (1986). Acceptor sites for retroviral integrations map near DNaseI-hypersensitive sites in chromatin. *J. Virol.* **60,** 683–692.
75. von Schwedler, U., Kornbluth, R. S., and Trono, D. (1994). The nuclear localization signal of the matrix protein of human immunodeficiency virus type 1 allows the establishment of infection in macrophages and quiescent T lymphocytes. *Proc. Natl. Acad. Sci. U.S.A.* **91,** 6992–6996.
76. Vora, A. C., and Grandgenett, D. P. (1995). Assembly and catalytic properties of retrovirus Integrase-DNA complexes capable of efficiently performing concerted integration. *J. Virol.* **69,** 7483–7488.
77. Wei, S.-Q., Mizuuchi, K., and Craigie, R. (1997). A large nucleoprotein assembly at the ends of the viral DNA mediates retroviral DNA integration. *EMBO J.* **16,** 7511–7520.

78. Withers-Ward, E. S., Kitamura, Y., Barnes, J. P., and Coffin, J. M. (1994). Distribution of targets for avian retrovirus DNA integration in vivo. *Genes Dev.* **8,** 1473–1487.
79. Wolffe, A. P. (1995). Histone deviants. *Curr. Biol.* **5,** 452–454.
80. Zou, S., and Voytas, D. F. (1997). Silent chromatin determines target preferences of the Saccharomyces retrotransposon Ty5. *Proc. Natl. Acad. Sci. U.S.A.* **94,** 7412–7416.
81. Zou, S., Wright, D. A., and Voytas, D. F. (1995). The Saccharomyces Ty5 retrotransposon family is associated with origins of DNA replication at the telomeres and the silent mating locus HMR. *Proc. Natl. Acad. Sci. U.S.A.* **92,** 920–924.