RETROVIRAL cDNA INTEGRATION: MECHANISM, APPLICATIONS AND INHIBITION

Mark S. T. Hansen, Sandrine Carteau, Christopher Hoffmann*, Ling Li**
and Frederic Bushman

Infectious Disease Laboratory
The Salk Institute for Biological Studies
10010 North Torrey Pines Road
La Jolla, CA 92037

INTRODUCTION

Retroviruses such as human immunodeficiency virus (HIV) are distinguished from other viruses by the combination of two steps in the viral life cycle. The first is reverse transcription, which results in the production of a cDNA copy of the viral RNA genome, and the second is integration, which results in the connection of that cDNA copy to a chromosome of the host. This article centers on the cDNA integration step. Understanding the integration system is important for designing methods for genetic manipulation in animal cells. Integration also represents a key area in developing new anti-retroviral therapy. Furthermore, studies of the integration system shed light on mechanisms in diverse related "mobile DNA" systems. In this article, we focus on recent work in three areas: i) the mechanism of retroviral integration, ii) integration and the development of retroviral vectors and iii) inhibition of integration. For further reviews on retroviral integration see (1-9).

---

* Present address: Oregon Health Sciences University, School of Medicine, 3181 SW Sam Jackson Park Rd., Portland, OR 97201

** Present Address: Gene Therapy Laboratories, Norris Cancer Center, USC School of Medicine, Los Angeles, CA 90033

Genetic Engineering, Vol. 20, edited by J.K. Setlow
Plenum Press, New York, 1998

MECHANISM OF cDNA INTEGRATION

Background and Early Studies

In the 1960s, Howard Temin proposed the provirus hypothesis, which stated that despite the fact that retroviruses contain an RNA genome, they replicate through a DNA intermediate (a provirus) that becomes permanently inserted in the cell genome (10). This proposal was dramatically confirmed by the discovery of the reverse transcriptase enzyme, which copies retroviral RNA into DNA.

Early progress in understanding the proteins involved in retroviral integration came from Grandgenett and co-workers (11), who identified a new protein with endonuclease activity present in avian myeloblastosis virus (AMV). Later biochemical and genetic studies indicated that this protein was encoded by the viral *pol* gene and was necessary for the integration reaction (12-14). On this basis the protein was named integrase.

DNA Cutting and Joining Reactions Involved in Integration

The integration process is now understood to involve a coordinated series of DNA breaking and joining reactions (Figure 1). The substrate for integration is the linear form of the viral cDNA produced by reverse transcription (15, 16). Circular forms of the viral cDNA also accumulate in infected cells and were once erroneously thought to be the substrate for integration. The reverse transcribed viral cDNA is first cleaved to remove dinucleotides from each 3' end, thereby exposing recessed 3' hydroxyls (Figure 1A, part 1 and 2). Next, the 3' ends are attached to opposite strands of the host genome (Figure 1A, part 3). The points of attachment on each strand are offset by four to six base pairs, with the length of the spacing invariant and characteristic for each retrovirus ("concerted integration" of the two ends). Melting of target DNA between the points of joining results in the formation of gaps at each host-viral DNA junction (Figure 1A, part 4). The last step consists of repair of the DNA gap, probably mediated by host DNA repair machinery. As a result of the repair step, four to six bp of target DNA are duplicated (Figure 1A, part 5).

In vitro Assays for Integration: Preintegration Complexes (PICs)

Key progress in understanding integration came from the establishment of the first *in vitro* assay for integration (17). Cellular extracts from Moloney murine leukemia virus (Mo-MLV) infected cells provided a source of integration machinery and DNA from phage lambda served as target. After incubation *in vitro*, integration products were packaged into lambda particles and recovered genetically. Products had structures resembling those generated by infection *in vivo* (Figure 1). The terminal base pairs had been removed from the cDNA ends, and each integrant was flanked by the four base pair duplication of the target DNA characteristic of Mo-MLV. The virus-derived nucleoprotein complexes that carried out integration in this system were operationally defined as "preintegration complexes" (PICs).

PICs have since been isolated from cells infected with HIV-1, HIV-2 and avian sarcoma leukosis virus (ASLV) (18-20). Analysis of reaction products revealed that in each case the gapped integration intermediate shown in Figure 1A part 4 was produced. Evidently the final repair step does not take place efficiently in the extracts studied, though this step can be carried out artificially to permit sequencing of reaction products. For all the host-virus DNA junctions sequenced, a total of 35 reported to date, each had the duplication characteristic of the virus studied, strongly emphasizing the authenticity of *in vitro* reactions with PICs (17, 21-23) (K. Myrick and C. Farnet, unpublished data).

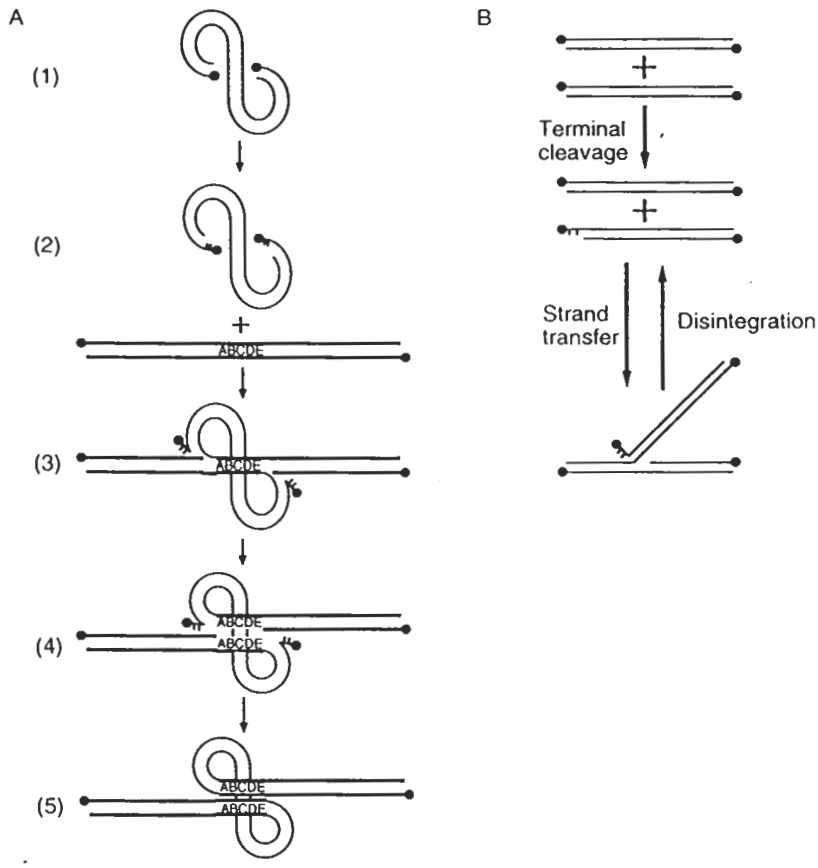

Figure 1. DNA cutting and joining reactions involved in retroviral cDNA integration. (A) Steps of integration of retroviral cDNA *in vivo*. The linear product of reverse transcription (step 1) is first processed by removing two nucleotides from each 3' end (step 2). The recessed 3' hydroxyl groups generated thereby are then attached to protruding 5' ends of breaks made in the target DNA (step 3). The paired bases of the target DNA that are between the points of joining then melt (step 4), producing single-stranded gaps at each end. The gaps are then filled in and sealed to yield an integrated provirus (step 5). (B) Reactions catalyzed by purified integrase.

The next several sections focus on the structure and function of integrase protein. We subsequently return to a more detailed discussion of the organization and function of PICs.

In vitro Assays for Integration: Purified Integrase Protein

Further progress came with the development of *in vitro* assays based on purified integrase. The purified integrases of ASLV and HIV were first found to carry out the terminal cleavage reaction, in which two nucleotides were removed from the 3' ends of model substrates (24-27). Key advances came with the establishment of assays in which purified integrase catalyzed formation of a covalent bond between a model viral DNA and a target DNA (26, 28, 29), recapitulating the covalent step of cDNA integration. These studies established that integrase was the enzyme that connected viral cDNA to target DNA

and not a required cofactor for a hypothetical cellular recombinase. This work also allowed simple assays to be developed for screening candidate small molecules for inhibitory activity (30, 31).

Studies of purified integrase also clarified whether the protein was actually an enzyme. Did integrase act once and then stop, or could integrase release the DNA product and then carry out a second reaction on another substrate molecule? *In vivo*, such an activity might not have been necessary, since the viral DNA, once inserted, was not expected to be released. In fact, integrase has been shown to turn over under some circumstances, albeit sluggishly (32, 33), and to mediate *in vitro* a reversal of the reaction called disintegration (34).

Chemistry of Integrase Action

The chemical mechanisms of the terminal cleavage and strand transfer reactions have been analyzed *in vitro* with chiral reaction substrates. Phosphorothioates were introduced into substrates for terminal cleavage and strand transfer, and the chirality of reaction products determined. It was found that the reactions proceeded with a single inversion of the chiral center, consistent with single-step transesterification mechanisms (35) and at odds with earlier proposals for covalent intermediates (36). According to this view, the cleavage and strand transfer mechanisms are chemically similar, differing only by the choice of nucleophile, a water molecule in the case of the terminal cleavage and a viral DNA 3' end in the case of strand transfer (1).

Functions of Integrase Protein Domains

Integrase (IN) protein is composed of three domains, defined by partial proteolysis (37), functional studies (38, 39) and structural studies (6, 40-45) (Figure 2): i) the N-terminal domain containing the conserved HHCC amino-acid motif (46), ii) the central catalytic domain containing the conserved D,D(35)E active site motif (47-49), and iii) the C-terminal domain. Many studies have investigated the importance of individual amino acid residues in each (33, 37, 50-67). The three-dimensional structure has now been determined for each domain in isolation, and efforts are under way to understand these

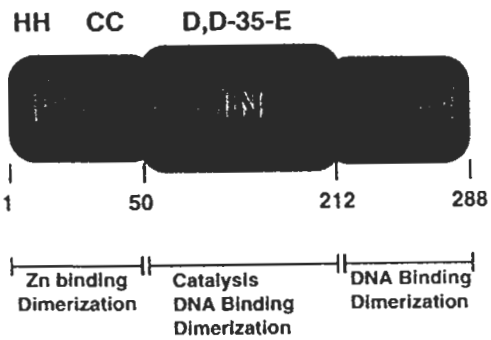

Figure 2. Domains of the integrase protein. Domain boundaries are marked by the amino acid numbers below the pictured monomer. Conserved sequence motifs are shown above the monomer. Proposed functions for each domain are shown beneath.

pieces in full monomers and the higher order multimers thought to mediate integration *in vivo*.

The N-terminal domain is required for full activity, though its precise role is not clear. Derivatives of Rous sarcoma virus (RSV) IN lacking the N-terminal domain retain both cutting and joining activities *in vitro* under specialized conditions (68, 69), suggesting that this domain plays a role that is in some sense secondary in reactions *in vitro*. However, the motif H-X(3-7)-H-X(23-32)-C-X(2)-C is highly conserved in all the retroviral IN proteins (46), indicating an important role. This motif binds zinc (44, 45, 70, 71) which enhances the activity *in vitro* (72).

The structures of the aminoterminal domains of HIV-1 and HIV-2 integrases were recently solved by NMR spectroscopy (44, 45). This work revealed that the N-terminal domains (residues 1-55 in each case) are composed of alpha helices with the zinc coordinated to the HHCC amino acids. The first three helices are seen in both structures, the fourth in HIV-1 only, though this may be due to experimental differences and not differences in the *in vivo* structure. The HIV-1 domain is dimeric, and this is likely for the HIV-2 domain as well, though the HIV-2 NMR study only reported the structure of a monomer. The HIV-1 dimerization interface is composed of helices 1, 3 and 4. The arrangement of the three first helices in both structures is strikingly similar to the helix-turn-helix class of DNA-binding domains, except that the HIV domains bind a zinc atom. This represents a new type of zinc-binding fold. The alpha helix that in the helix-turn-helix proteins is inserted into the major groove for sequence-specific recognition serves as a dimerization interface of the HIV-1 domain. It has been suggested nevertheless that the aminoterminal domain may bind DNA (54, 73), though direct evidence for this is still lacking.

The structure of the central catalytic domain has been solved for both HIV-1 and avian sarcoma virus (ASV) integrases by X-ray crystallography (40-42). Each had a similar fold, composed of mixed alpha-helices and beta-sheets. Each catalytic domain formed a dimer, with similar interfaces burying relatively large protein surfaces, consistent with the idea that each constituted a biological dimer interface. The acidic residues of the D,D-35-E motif were clustered in each, as expected if these residues contribute to forming the active site.

The structure of the integrase catalytic domain closely resembles structures of several other enzymes involved in polynucleotide phosphotransfer reactions, including MuA transposase, *E. coli* DNA polymerase I, RNase H of *E. coli* and HIV, and *E. coli* RuvC. Where tested, each of these enzymes, like HIV-1 integrase, acts chemically by a single-step transesterification (74). Evidently integrase is a member of an ancient group of related phosphotransfer enzymes.

All the members of the superfamily require the presence of Mg or Mn to be active and may employ similar mechanisms to catalyze the transesterification reaction. In some cases metal atoms have been detected by X-ray crystallography bound to acidic residues at sequences related to the D,D-35-E motif integrases. For the exonuclease of *E. coli* DNA polymerase I, a chemical mechanism based on two bound metal atoms has been proposed (75). In this case, the metal atoms have been suggested to promote hydrolysis by facilitating the formation of an attacking hydroxide nucleophile and stabilizing the phosphate leaving group. Similar chemistry can be modified to create an integration mechanism by substitution of the hydroxide nucleophile with a DNA 3' hydroxyl (5). Recent crystallographic studies of the ASV catalytic domain in the presence of zinc have succeeded in visualizing two atoms bound to the conserved acidic residues with the geometry predicted for a two-metal mechanism (42). Though the zinc-bound form does not carry out normal integration, it seems likely that this structure is reflective of the biological complex with Mg, the probable ion used *in vivo*.

The C-terminal domain is involved in nonspecific DNA binding, though its contribution to integration has not been fully clarified (38, 48, 76-80). Its presence is required for cutting and joining activities of integrase but not for the disintegration reaction. Like both the aminoterminal and central catalytic domains, the carboxyterminal domain dimerizes in solution. Its structure has been solved by NMR, and was found to consist of a five-stranded beta-barrel, topologically similar to SH3 domains (6, 43). The related MuA transposase also contains a catalytic domain linked to a carboxyterminal SH3-like domain, consistent with a conserved function for this pair of domains in transposases and integrases (81).

Oligomerization of Integrase

Early analyses of integrase protein purified from AMV suggested that the protein may be dimeric (11). Further studies demonstrated that integrase functions as a dimer or higher order multimer (38, 39, 82, 83). Studies of complementation among mutant proteins *in vitro* provided functional evidence for multimerization. Mutant derivatives of HIV-1 integrase bearing lesions in the D,D-35-E motif are inactive *in vitro*, as are derivatives deleted for the aminoterminal domain. However, mixing of the two mutants *in vitro* results in restoration of considerable activity, supporting the idea that integrase acts as a multimer (38, 39). As yet the nature of the active multimer is unclear, due in part to aggregation of integrase protein under the conditions in which it is active *in vitro*.

Cai et al. proposed a model for an HIV-1 integrase tetramer that accommodates the structural and functional studies (44). Since the structure of each domain was solved separately, the challenge is to model biologically reasonable monomers and higher-order multimers. The amino- and carboxytermini of the catalytic domain are quite close together. Consequently, models cannot be formed in which the aminoterminal domain, the central catalytic domain, and the carboxyterminal domain all dimerize comfortably on one pair of molecules. However, a neat tetramer can be formed if the aminoterminal domain and the carboxyterminal domain each dimerize with a different partner molecule. The present challenge is to obtain data to strengthen this idea and distinguish among the possible isomers.

The relationship of the active sites in the catalytic domain dimer presents a puzzle. The points of joining of the two viral DNA ends in the target DNA are separated by five base pairs for HIV, and integration complexes likely approach the target from the major groove side and join the cDNA ends to points across a single major groove (84, 85). However, the distance between active sites in the catalytic domain dimer does not match this distance in either the HIV or ASV case. Furthermore, none of the monomers in the proposed tetramer of Cai et al. (44) has this geometry either. Possibly an integrase octamer could satisfy the target site geometry (44). Alternatively, the finding that DNA distortion may promote integration (86-89) could imply that target DNA is distorted during the integration reaction, potentially altering the geometry required for integrase to attach each cDNA end to target DNA.

Specialized assembly in PICs is implied by the observation that HIV-1 PICs carry out concerted integration of the two viral DNA ends in the target DNA. Reactions with purified HIV-1 integrase, in contrast, primarily yield unconcerted products modeling events at one viral DNA end only (26, 29). Coupled joining of the two cDNA ends likely reflects correct association of protein-DNA complexes at each end of the unintegrated cDNA. What features of PICs permit efficient coupled joining? One possibility is that HIV-1 integrase protein as purified from bacteria is partially defective. In potential support of this view is the finding that purified integrases of ASV and Mo-MLV seem somewhat more proficient at carrying out coupled reactions (26, 28, 90-93), consistent with the possibility that these

proteins fold more efficiently *in vitro* than HIV-1 integrase. However, other models can also explain these results. It may be that factors other than integrase present in PICs promote coupled joining (see below). Consistent with this idea, lysates of HIV virions can support clear coupled joining *in vitro* (94), though the basis for this difference with recombinant integrase is not yet clear. Yet a third possibility is that the DNA substrates tried so far lack sequences necessary for coupled joining. It could also be that all the required factors are present in active form but the proper assembly conditions have not been identified.

It seems likely that studies of PICs will shed light on some of these issues. We now return to the topic of PICs and review recent findings on the function, organization, and components.

Organization of PICs

Biophysical studies of PICs have begun to elucidate aspects of macromolecular organization. Gel filtration studies of HIV-1 PICs yield a Stokes radius of about 28 nm (95), about half that of a virion. The PICs appear to be very large, with sedimentation coefficients ranging from 160-640s, depending on the solution conditions and the virus under study (95-97). The length of the viral cDNA is 3.3 μm, so it must be greatly compacted to fit in the measured volume of the PIC. Such biophysical measurements, however, can be confounded by artifacts. For example, PICs may aggregate during handling or bind to gel filtration resins during separation, but these data at least provide a starting point for further refinement.

DNA-Protein Interactions in PICs

The cDNA in PICs is likely compacted at least in part by association with DNA-binding proteins. Integrase protein is known to be present (96, 98-100), and the host cell protein HMG I(Y) also appears to be present (100). Other viral and cellular DNA-binding proteins may also be associated. The ends of the viral cDNA in PICs are insensitve to digestion with exonucleases under conditions where deproteinized cDNA is digested efficiently, indicating that the viral cDNA ends are associated with bound proteins (95). In contrast, internal DNA sequences are exposed to endonucleases, consistent with a looser association with proteins (95, 97).

Association of the cDNA ends can be studied functionally as well. Internal cleavage of the cDNA in HIV PICs with a restriction enzyme does not disrupt coupled joining of the severed DNA ends, indicating that the ends are bridged by proteins (95). The bound protein has not been identified, but integrase is known to have had access to the DNA ends, since the cDNA in HIV-1 PICs have the two 3' nucleotides removed, supporting the idea that integrase itself is bound to the cDNA termini (95).

Structure of the cDNA in PICs

Examination of the viral cDNA in integration-competent HIV-1 PICs revealed that the plus strand DNA was not fully continuous, but in fact contained breaks. HIV contains a polypurine tract sequence that serves as a primer for reverse transcription not only at the usual place, the edge of the right long terminal repeat (LTR), but also in the middle of the genome in the integrase coding region (101). DNA synthesis from upstream stops in this region (102), resulting in a plus strand cDNA in two halves. Additional sites of plus strand initiation can also be detected *in vivo* (103) and *in vitro* (104). Older studies of oncoretroviruses also reported breaks in the plus strand cDNA (4).

Studies of integration by HIV-1 PICs reveal that the breaks in the plus strands are present in active PICs and are preserved in products after integration *in vitro*. Together with time-course studies of synchronous infections *in vivo*, these data suggest that the breaks in the HIV-1 cDNA may not be repaired until after the completion of integration (103). Already host DNA repair enzymes are thought to be involved in sealing gaps at the junctions between viral and host sequences, and it seems likely that the same system repairs internal gaps at the same time.

A Possible Role for the Terminal Cleavage Reaction

Reverse transcription yields a blunt cDNA end, and two nucleotides are removed from each 3' end afterward. Why do retroviruses synthesize extra DNA, then cut it off? It has been suggested that IN binds tightly to the resulting processed ends, supporting the view that forming a stable intermediate is one purpose of the cleavage reaction (83, 86). Recent findings on the structure of the cDNA in HIV PICs suggest another purpose as well. Early after synthesis many of the 3' ends of the unintegrated cDNA are longer than expected from the sequence, apparently due to the attachment of extra bases to otherwise blunt 3' ends. In fact, HIV reverse transcriptase is known to be able to attach extra nucleotides to DNA 3' ends of model substrates *in vitro*, especially the DNA U5 end (104). For the case of cDNA ends generated during infection, the U5 end is also found to be longer than predicted, while the U3 end is as expected (95), supporting a role for reverse transcriptase (RT) in addition *in vivo*. Another reason for the existence of the terminal cleavage reaction may be the removal of such extra bases, which if present might interfere with integration.

Composition of PICs

Identifying the components of PICs and their stoichiometries has been difficult, in part because obtaining large quantities of active PICs has been challenging. The observation that a factor cofractionates does not necessarily mean that it is associated with PICs, since PICs have not yet been purified to homogeneity, and even proteins that are found to be bound to the cDNA might be functionally insignificant or artifactually associated. Proteins that are not detected may be present but at levels below the limit of detection.

However, it is clear that integrase protein is associated with PICs of HIV-1, and important for integration *in vivo* and *in vitro* (96, 98-100). In a study in which PICs were handled in the presence of high concentrations of detergents, integrase was the only viral protein detected (96). In PICs not treated with detergent, matrix protein (MA) has also been found in most studies, while the viral capsid (CA) is absent (95, 105). Of the other viral proteins, RT and Vpr proteins have been proposed to be components, and nucleocapsid protein (NC) has been detected in some partially-purified fractions (95, 96, 98, 99, 106). In the case of HIV-2 and simian immunity virus (SIV), Vpx protein can also be detected associated with the cDNA (19, 107), though no functional difference was found in PICs from Vpx^- viruses (19).

HMG I(Y): a Host Factor Important for Function of PICs *in vitro*

For HIV-1 PICs, a cellular protein, HMG I(Y), has been proposed to be required for integration activity *in vitro* (100). It was found that when HIV-1 PICs were partially purified and subjected to gel filtration in high salt buffer, complexes remained partially intact. However, PICs returned to standard reaction conditions and incubated with a target DNA did not display integration. Integration activity could be restored by incubation with cell extracts, implicating a host protein as the complementing activity. Purification of the

activity yielded HMG I(Y) protein. HMG I(Y) was found to be associated with PICs, and complementing activity stripped from authentic PICs could be depleted with an anti-HMG I(Y) antibody. Recombinant HMG I protein also complemented.

HMG I(Y) is a high mobility group (HMG) protein, defined as extractable from chromatin in 5% perchloric acid and migrating rapidly on acid/urea gels (hence "high mobility"). HMG families include HMG 1/2 , HMG 14/17 and HMG I(Y) (108). HMG Y is produced from the same gene as HMG I, but contains an internal deletion of 11 amino acids due to alternate splicing. HMG I and HMG Y often copurify, and so far are functionally interchangeable, so the two are referred to together as HMG I(Y). Another family member, HMG IC, is expressed only in fetal tissue (109)(K. Yoder, M. Hansen, and F. Bushman, unpublished data).

The mechanism by which HMG I(Y) complements salt-stripped PICs is unclear, although several models can be proposed. Interaction between HMG I(Y) and integrase has not been detected in glutathione S-transferase (GST) pull-down experiments or coimmunoprecipitation experiments, arguing against a direct interaction between the two proteins. HMG I(Y) binds A/T rich DNA sequences (110-114), and such sequences are present in the LTRs of many retroviruses. Each HMG I(Y) monomer has three DNA binding motifs (110), and one HMG I(Y) molecule can bridge two DNAs by binding to sites on each (115), leading to the possibility that HMG I(Y) may act by bringing distant DNA domains into proximity.

One model holds that one HMG I(Y) molecule binds both viral cDNA and target DNA, thus bridging the PIC and target. However, a study of sixty-one integration sites generated by HIV-1 infection did not show an enrichment for HMG I(Y) binding sites near integration sites (116), arguing against this model. Furthermore, order of addition experiments also indicated that HMG I(Y) acted on PICs and not on target DNA (100, 117).

An attractive model holds that HMG I(Y) may serve as an architectural element for the formation of active PICs, helping shape the DNA for optimal function. Such a model for HIV integration parallels previous findings in several site-specific recombination and transposition systems. Normal integrative recombination by phage lambda integrase, for example, requires the participation of the *E. coli* protein integration host factor (IHF), which binds to the viral DNA and bends it into the active conformation [for review see (118)]. Phage Mu transposition requires the participation of *E. coli* Hu protein, which binds to the left end of Mu DNA and helps bring monomers of MuA transposase into the active alignment [see (119) and references therein]. *E. coli* Fis protein, another DNA-bending protein, is important for formation of DNA inversion complexes by Gin and Hin invertases [for review see (120)]. In each case the active recombination complex is composed of several monomers of the recombination enzyme precisely organized with the host proteins on substrate DNA. HMG I(Y) resembles IHF, Hu and Fis in that it is a small host DNA-binding protein that when bound alters the conformation of DNA (114, 121-126), possibly mediating its role in HIV integration.

Host Factors Involved in Integration of Mo-MLV PICs

Several host activities have been identified that influence integration in the Mo-MLV system. Lee and Craigie used a salt-stripping procedure to identify a host activity important not for stimulating integration but for preventing PICs from using their own cDNA as integration targets ["autointegration," see (127) for discussion](128). The autointegration reaction produces circular forms of the viral cDNA that are not competent for normal integration. Thus autointegration is suicidal for the retrovirus. Lee and Craigie treated Mo-MLV PICs with high salt, then the salt-stripped PICs were separated from dissociated proteins by velocity sedimentation. Such salt-stripped complexes had no normal integration

activity but increased autointegration activity. The addition of a cytoplasmic extract of NIH 3T3 cells could restore normal integration and block autointegration. The addition of virion extract could not reconstitute the salt-stripped complexes, leading to the proposal that a host factor provided the autointegration barrier. The identity of the autointegration barrier factor has not yet been reported.

More recently, a similar salt-stripping procedure was used to identify factors stimulating normal Mo-MLV integration (117). Either nuclear or cytoplasmic extracts from NIH 3T3 cells supported both reconstitution of normal integration and suppression of autointegration. Purification of the complementing activity yielded murine HMG I(Y). Interestingly, HMG I(Y) was unable to suppress autointegration, indicating that a separate factor is responsible for this activity.

Another activity of host DNA-binding proteins was detected with the use of a different depletion procedure. Mo-MLV PICs that were partially purified but not salt-stripped displayed consistently lower integration activity. The low integration activity could be improved 2- to 3-fold by addition of a variety of DNA-binding proteins, including HMG-1, HMG I(Y), HIV-1 NC, T4 phage gene 32, RNase A or *E. coli* single-stranded binding protein (SSB). This activity may parallel previous reports of many DNA-binding proteins stimulating the activity of purified integrase protein *in vitro* (29, 90, 129-133). It is important to distinguish this stimulatory activity from the reconstitution of salt-stripped PICs, since only HMG I(Y) supported reconstitution of salt-stripped Mo-MLV PICs, emphasizing the specificity. The fact that so many DNA-binding proteins stimulated activity in assays with purified integrase or partially-purified but not salt-stripped PICs raises the possibility that the mechanism involves nonselective coating of DNA. For example, DNA coating may promote the interaction between the viral cDNA and target DNA by shielding negative charges of the DNA phosphates and reducing electrostatic repulsion [though for a different view see (133)]. Thus three different activities have been detected influencing Mo-MLV PICs *in vitro* (117, 128), and efforts are under way to assess the importance of each *in vivo*.

Target Site Selection

An issue of long-standing interest in the retroviral field is the question of host determinants of sites for integration *in vivo*. *In vitro*, integrase and PICs show some preferences for certain DNA sequences, but the effect is modest (23, 92, 116, 134, 135). *In vivo*, however, the host integration acceptor sites are not expected to be present as naked DNA, but rather associated with histones and other DNA-binding proteins in chromatin. *In vitro*, a simple DNA-binding protein bound to target DNA can block integration at that site, probably by steric hindrance (84, 136). DNA-bending proteins such as nucleosomes, in contrast, can promote integration (84, 85, 87, 88, 137). In these cases, integration is most prominent at sites of maximal DNA distortion, consistent with the idea that distortion is involved in the integration mechanism (86, 89).

Previous surveys of *in vivo* integration sites have led to several proposals for factors influencing site selection. Studies of Mo-MLV have supported a model in which open chromatin regions at transcription units were favored, since associated features such as DNase I hypersensitive sites (138, 139) or CpG islands (140) were apparently enriched near integration sites. Another study proposed that unusual host DNA structures were common near integration sites (141). A study of avian leukosis virus integration frequencies at several chromosomal sites failed to show any major differences among the regions studied (142), contrary to a previous report (143). For HIV-1, it has been proposed that integration may be favored near repetitive elements [including LINE elements (144) or Alu islands (134)] or topoisomerase-cleavage sites (145).

An interesting contrast to the retroviral case is provided by the related Ty retrotransposons of yeast. The life cycle of these elements differs from retroviruses in that they lack an extracellular phase, but replicate by transcription, reverse transcription and integration as with retroviruses. The Ty1 and Ty3 retrotransposons show remarkable specificity, integrating primarily upstream of tRNA genes (146, 147). This may seem like an odd choice of target site, but insertions at these locations apparently do not disrupt function of the nearby tRNA genes. Similarly, the Ty5 element integrates near telomeres or the silent mating type locus, again without major disruption of host functions (148). Seventy percent of the *Saccharomyces cerevisiae* genome is composed of coding regions, and yeast spend part of their life cycle in the haploid state, so targeted integration reduces suicidal disruption of the host genome. The degree if any to which retroviruses might exploit similar mechanisms is still being investigated.

The most extensive study of retroviral integration sites to date reported sequences of 61 integration sites for HIV-1 and also further studies of the implications by other methods (116). Integration sites were generated by infection of cultured human T cells, and DNA was harvested 12 to 24 hours after initiating the infection to minimize selection of sites during growth of cells. The experimentally-generated integration sites were compared to a library of 112 control fragments of human DNA that were cloned and sequenced by similar means. At odds with previous reports from smaller scale studies (134, 144), no strong biases were found either for or against integration into transcription units or repetitive sequences such as Alu or LINE elements. However, centromeric alphoid repeats were found to be selectively absent at integration sites, indicating that centromeres are disfavored targets. These issues were then probed with a different method (84, 149), analysis of integration sites generated by infection *in vivo* and PCR to amplify junctions between the HIV LTR and specific repeated sequences. Again, integration could be seen in Alu and LINE elements, but no integration in the alphoid repeats. For comparison, integration was carried out *in vitro* with PICs as a source of integration activity and naked chromosomal DNA as target. In this case integration was seen in each of these repetitive elements, indicating that centromeric alphoid repeats are selectively blocked for integration *in vivo* but not in naked DNA *in vitro*.

Heterochromatic centromeres are seen to be more compact than euchromatin in fixed chromosome spreads (150), and alphoid sequences are more resistant to digestion with DNase I in isolated nuclei than are most DNAs (151, 152). Alphoid repeats are associated with the centromere-specific proteins CENP-A, CENP-B and CENP-C (151, 152). Evidently HIV-1 cDNA integration is obstructed by packaging DNA in centromeric heterochromatin, revealing an unexpected effect of chromatin structure (116).

INTEGRATION AND THE DEVELOPMENT OF RETROVIRAL VECTORS

Requirements for integration dictate in part the functions necessary for construction of retroviral vectors. In this section we briefly review the requirements for integration in retroviral vectors and modifications of the integration system of possible use in gene therapy and other settings. For reviews of retroviral vectors see (7, 153).

Retroviral Components Necessary for Integration

Retroviral vector systems rely on the production of viral proteins and a separate packagable RNA (vector) in the same cell, so that each component can be incorporated into viral particles. Upon infection of target cells, the vector RNA is reverse transcribed and integrated into cellular DNA by the usual retroviral machinery. Besides integrase,

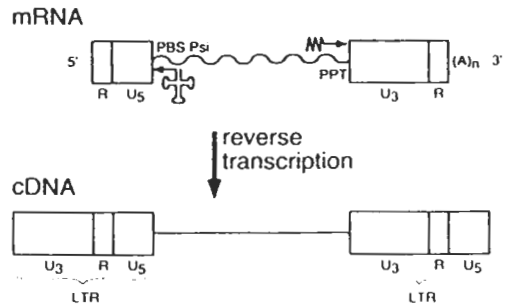

Figure 3. Diagram of the retroviral RNA and proviral DNA highlighting sequences important for integration. The structure of the viral mRNA is shown at the top, the proviral cDNA below. The LTR is composed of three regions, U3, R and U5. Reverse transcription initiates at the primer binding site (PBS) with a host tRNA as primer (cloverleaf structure). Synthesis of the other cDNA strand initiates at least in part at the polypurine tract (PPT) near the U3 region, with viral RNA as primer (zig-zag). Importantly, the sequences that form the LTR ends in the cDNA are encoded at the U5 edge adjacent to the PBS and at the U3 edge adjacent to the PPT.

sequences at the LTR termini of the unintegrated cDNA are important for integration of vectors. Available information indicates that the terminal eight to fifteen nucleotides are important for interaction with the integration system *in vivo* and *in vitro* (24, 27, 86, 90, 154-167), and these sequences must be maintained in vectors. However, as a consequence of the mechanism of reverse transcription, information for the LTR termini in the cDNA does not map to the ends of the mRNA [for reviews of reverse transcription see (4, 168)]. Without going into the details of reverse transcription, the sequences important for cDNA integration are encoded internally in the RNA adjacent to the PBS and PPT sites for initiation of reverse transcription (Figure 3). It is essential that these sequences be maintained in retroviral vectors.

Controlling Target Site Selection

A drawback to the use of retroviral vectors in gene therapy is the potential disruption of the host genome due to integration. Many studies in vertebrate systems establish that integration of retroviral DNA can result in inactivation or ectopic activation of cellular genes, thereby causing cancer or other diseases [for a review see(7)]. For the case of humans, Shiramizu et al. have described examples of non-Hodgkin lymphomas from late-stage AIDS patients in which integration of HIV apparently activated the c-fps/c-fes oncogene (169).

In an effort to develop retroviral vectors capable of integration at predetermined sites, several groups have studied fusions of integrases to sequence-specific DNA-binding domains. Purified fusion proteins can direct quite selective integration into target DNAs containing recognition sites for the cognate DNA-binding domain (136, 170, 171). One integrase fusion could also influence the specificity of integration by PICs when introduced into viruses (172). The current challenge is to improve the specificity enough to create targeting vectors useful *in vivo* (173).

Possibly the utility of integrase-DNA-binding domain fusions may be widened by exploiting exciting recent studies on engineering of specificity in DNA-binding proteins such as Zif268. The Zif268 protein is a member of the zinc finger family that binds DNA as a monomer with high affinity and specificity (174, 175). Several groups have reported striking studies on controlling the specificity of DNA binding by Zif268. A recent study employed a fusion of HIV-1 integrase to Zif268 (172).

One derivative of Zif268 contains three tandem zinc fingers, each of which binds sequential three-base-pair sequences in a nine-base-pair recognition site. Rearranging the order of the zinc fingers yielded a protein that recognized correspondingly rearranged triplet sequences in the recognition site. This suggests that each finger is an independent modular unit that recognizes a three-base-pair sequence. Additional studies have involved phage display methods to generate fingers with new binding specificities. At present, the feeling in the field is that Zif268 derivatives may be constructed that recognize many different nine-base-pair sequences, perhaps all possible nine-base-pair sequences (176-183). Further studies focus on attempting to lengthen the Zif268 recognition site by adding further fingers or linking Zif268 to still other DNA-binding domains (184, 185). Together these studies raise the hope that the specificity of an integrase-Zif268 fusion could be designed to direct integration into any desired sequence.

INHIBITORS OF HIV INTEGRATION

Integrase as a Target for Antiviral Agents

The integration system is a particularly attractive target for antiretroviral drugs. The integration step is required for replication, as demonstrated by the finding that HIV derivatives containing lesions in the integrase protein are unable to replicate (57-61). There are no known human enzyme systems that resemble the integration system, raising the hope that integration inhibitors might be relatively non-toxic.

Inhibitors of Purified Integrase

Many small molecule inhibitors have been described that are active against purified integrase protein. Unfortunately, none of these appear to be active *in vivo* against integrase. Here we catalog some of the known inhibitors and the status of present studies.

Early studies identified topoisomerase inhibitors and polyanionic compounds as integrase inhibitors (186-192). However, most of these probably acted nonspecifically by blocking binding of integrase to the DNA substrates. Integrase is not thought to utilize nucleotides during integration, but integrase can bind free nucleotides and nucleotide analogs *in vitro* (193), and these represent another class of inhibitors (73, 194, 195). A large group of inhibitors contain the bis-catechol moiety, a benzene ring with a pair of bis-hydroxyls (196-200). Some of these compounds have been observed to act on the catalytic core domain of integrase (196-198, 200), raising the possibility that they act against the active site (196, 199-201). Some reservations have been expressed about inhibitors of this class. i) They may inhibit assembly of the LTR DNA with integrase instead of chemistry at the active site (202). ii) At least some members of this class may be capable of crosslinking cellular proteins (203), raising fears of nonspecificity and toxicity *in vivo*. Inhibitory molecules have been derived from myriad further compounds, including anthraquinones (200), coumarins (204), flavones (196), dicaffeoylquinic acids (205), phenanthroline-metal complexes (198), cosalane analogs (206), tyrphostins (200, 207), depsides (208), depsidones (208), arylamides (209) and others (210, 211).

The highly-promising combinatorial chemistry approach is just beginning to be applied to the development of integrase inhibitors. In the first example of such a study, a six-mer peptide was isolated from a six-mer library and shown to inhibit integrase (212). This inhibitory peptide is not itself likely to be useful as an antiretroviral agent, but this study does demonstrate that it is possible to isolate integrase inhibitors from a large pool of

synthetically generated molecules. Similarly, randomly-generated RNA molecules have been selected for high-affinity binding to integrase (213).

PICs in Drug Screening

It was recently found that integration assays with PICs are less sensitive to many inhibitors than conventional reactions with purified integrase, emphasizing the utility of PIC assays as a secondary screen. As was discussed above, PICs differ from complexes assembled with purified integrase in their composition and activities, so it is not surprising that the two respond differently to inhibitors. Furthermore, the response to AZT provides evidence that the PIC response is more authentic. AZT inhibits reactions with purified integrase (194), but not reactions with PICs (200). However, AZT is clearly active *in vivo* against RT, not integrase, since viral escape mutants are always found in RT. At least in this case it seems that the assays with PICs are more authentic as well as more selective.

Recently it was reported that *in vitro* reactions in which integrase is prebound to an LTR substrate (214, 215) may more closely model reactions with PICs. This modification may represent a useful improvement of screening methods employing purified integrase.

CENTRAL QUESTIONS FOR FUTURE STUDIES

Of the questions confronting the integrase field, none is more important than that of whether clinically useful inhibitors can be developed. Although the slow pace has been discouraging, it may not be surprising that the development of integrase inhibitors has lagged behind the development of reverse transcriptase and protease inhibitors. For each of the latter two, previously known inhibitors of related enzymes greatly aided the design of the present AIDS therapeutics. For reverse transcriptase, nucleotide analogs provided obvious molecules to analyze, and AZT was found in such a collection. For the HIV protease, naturally-occurring inhibitors were known for the related microbial aspartic-acid proteases, and these provided a point of departure for design of HIV inhibitors. No inhibitors are known for integrase or any related protein, and as yet nothing suggests a clear path for inhibitor design.

Additional questions surround the organization and assembly of preintegration complexes. What proteins are associated, and in what amounts? How are they assembled together? How does the PIC change during the sequential reactions that mediate integration? Studies of composition and function of PICs have begun to yield a detailed picture, and studies of the structures of the proteins involved are advancing, but understanding the full PIC in terms of atomic structures remains a challenging goal.

A particularly exciting question is that of whether further host factors are involved in integration. A role for the DNA-binding protein HMG I(Y) has been proposed (100, 117), but other host factors are also likely involved. For example, unidentified host DNA-repair enzymes probably finish the job of integration, but as yet no candidate proteins have even been proposed. It seems likely that still further host factors may be involved, for example, in localizing the PIC to the host cell nucleus during infection.

Further questions surround the issue of target site selection. What chromosomal features determine the choice of integration sites? Several studies have characterized target site selection *in vitro*, but as yet these have not been tightly connected to studies of site selection *in vivo*, leaving the picture incomplete.

Lastly, can the integration system be modified to generate more useful gene therapy tools? It is clear that fusion of integrase to a sequence-specific DNA-binding domain can influence site selection, but as yet no derivatives have been described that tightly control

specificity *in vivo*. This area seems promising, since naturally-occurring retrotransposons in yeast display just the kind of tight integration site-specificity that would be useful in retroviral vectors . It remains to be seen, however, whether such an efficient targeting system can be constructed in a retroviral vector.

REFERENCES

1 Mizuuchi, K. (1992) J. Biol. Chem. 287, 21273-21276.
2 Brown, P.O. (1990) Curr. Top. Microbiol. Immunol. 157, 19-48.
3 Goff, S.P. (1992) Annu. Rev. Genet. 26, 527-544.
4 Weiss, R., Teich, N., Varmus, H. and Coffin, J. (1984) Cold Spring Harbor Press, Cold Spring Harbor, NY.
5 Yang, W. and Steitz, T.A. (1995) Curr. Biol. 3, 131-134.
6 Plasterk, R.H.A. (1995) Nature Struct. Biol. 2, 87-90.
7 Coffin, J.M. (1996) in Virology (B.N. Fields, D.M. Knipe and R.M. Howley, eds.) pp. 1767-1848, Lippincott-Raven Publishers, Philadelphia, PA.
8 Farnet, C.M. and Bushman, F.D. (1996) AIDS, 10 (Suppl. A), 3-11.
9 Chow, S.A. (1997) Methods 12, 306-317.
10 Cooper, G.M., Temin, R.G. and Sugden, B. (1995) The DNA Provirus. Howard Temin's Scientific Legacy. ASM Press, Washington, DC.
11 Grandgenett, D.P., Vora, A.C. and Schiff, R.D. (1978) Virology 89, 119-132.
12 Schwartzberg, P., Colecilli, J. and Goff, S.P. (1984) Cell 37, 1043-1052.
13 Panganiban, A.T. and Temin, H.M. (1984) Proc. Nat. Acad. Sci. U.S.A. 81, 7885-7889.
14 Donehower, L.A. and Varmus, H.E. (1984) Proc. Nat. Acad. Sci. U.S.A. 81, 6461-6465.
15 Brown, P.O., Bowerman, B., Varmus, H.E. and Bishop, J.M. (1989) Proc. Nat. Acad. Sci. U.S.A. 86, 2525-2529.
16 Fujiwara, T. and Mizuuchi, K. (1988) Cell 54, 497-504.
17 Brown, P.O., Bowerman, B., Varmus, H.E. and Bishop, J.M. (1987) Cell 49, 347-356.
18 Farnet, C.M. and Haseltine, W.A. (1990) Proc. Nat. Acad. Sci. U.S.A. 87, 4164-4168.
19 Hansen. M. and Bushman, F.D. (1997) J. Virol. (in press).
20 Ha Lee, Y.M. and Coffin, J.M. (1991) Mol. Cell. Biol. 11, 1419-1430.
21 Ellison, V.H., Abrams, H., Roe, T., Lifson, J. and Brown, P.O. (1990) J. Virol. 64, 2711-2715.
22 Ha Lee, Y.M. and Coffin, J.M. (1990) J.Virol. 64, 5958-5965.
23 Pryciak, P.M., Sil, A. and Varmus, H.E. (1992) EMBO J. 11, 291-303.
24 Katzman, M., Katz. R.A., Skalka, A.M. and Leis, J. (1989) J. Virol. 63, 5319-5327.
25 Sherman, P.A. and Fyfe, J.A. (1990) Proc. Nat. Acad. Sci. U.S.A. 87, 5119-5123.
26 Craigie, R., Fujiwara, T. and Bushman, F. (1990) Cell 62, 829-837.
27 Bushman, F.D. and Craigie, R. (1991) Proc. Nat. Acad. Sci. U.S.A. 88, 1339-1343.
28 Katz, R.A., Merkel, G., Kulkosky, J., Leis, J. and Skalka, A.M. (1990) Cell 63, 87-95.
29 Bushman, F.D., Fujiwara, T. and Craigie, R. (1990) Science 249, 1555-1558.
30 Craigie, R., Mizuuchi, K., Bushman, F.D. and Engelman, A. (1991) Nucl. Acids Res. 19, 2729-2734.
31 Hazuda, D.J., Hastings, J.C., Wolfe, A.L. and Emini, E.A. (1994) Nucl. Acids Res. 22, 1121-1122.
32 Chow, S.A. and Brown, P.O. (1994) J. Virol. 68, 3896-3907.
33 Kulkosky, J., Jones, K.S., Katz, R.A., Mack, J.P.G. and Skalka, A.M. (1992) Mol. Cell. Biol. 12, 2331-2338.
34 Chow, S.A., Vincent, K.A., Ellison, V. and Brown, P.O. (1992) Science 255, 723-726.

35 Engelman, A., Mizuuchi, K., and Craigie, R. (1991) Cell 67, 1211-1221.
36 Katzman, M., Mack, J.P.G., Skalka, A.M. and Leis, J. (1991) Proc. Nat. Acad. Sci. U.S.A. 88, 4695-4699.
37 Engelman, A. and Craigie, R. (1992) J. Virol. 66, 6361-6369.
38 Engelman, A., Bushman, F.D. and Craigie, R. (1993) EMBO J. 12, 3269-3275.
39 van Gent, D.C., Vink, C., Oude Groeneger, A.A.M. and Plasterk, R.H.A. (1993) EMBO J. 12, 3261-3267.
40 Dyda, F., Hickman, A.B., Jenkins, T.M., Engelman, A., Craigie, R. and Davies, D.R. (1994) Science 266, 1981-1986.
41 Bujacz, G., Jaskolski, M., Alexandratos, J., Wlodawer, A., Merkel, G., Katz, R.A. and Skalka, A.M. (1995) J. Mol. Biol. 253. 333-346.
42 Bujacz, G., Alexandratos, J., Wlodawer, A., Merkel, G., Andrake, M., Katz, R.A. and Skalka, A.M. (1997) J. Biol. Chem. 272, 18161-18168.
43 Lodi, P.J., Ernst, J.A., Kuszewski. J., Hickman, A.B., Engelman, A., Craigie, R., Clore, G.M. and Gronenborn, A.M. (1995) Biochemistry 34, 9826-9833.
44 Cai, M., Zheng, R., Caffrey, M., Craigie, R., Clore, G.M. and Gronenborn, A.M. (1997) Nature Struct. Biol. 4, 567-577.
45 Eijkelenboom, A.P.A.M., van den Ent, F.M.I., Vos, A., Doreleijers, J.F., Hard, K., Tullius, T., Plasterk, R.H.A., Kaptein, R. and Boelens, R. (1997) Curr. Biol. (in press).
46 Johnson, M.S., McClure, M.A., Feng, D.-F., Gray, J. and Doolittle, R.F. (1986) Proc. Nat. Acad. Sci. U.S.A. 83, 7648-7652.
47 Rowland, S.-J. and Dyke, K.G.H. (1990) Mol. Microbiol. 4, 961-975.
48 Khan, E., Mack, J.P.G., Katz, R.A., Kulkosky, J. and Skalka, A.M. (1990) Nucl. Acids Res. 19, 851-860.
49 Fayet, O., Ramond, P., Polard, P., Prere, M.F. and Chandler, M. (1990) Mol. Microbiol. 4, 1771-1777.
50 van Gent, D.C., Oude Groeneger, A.A.M. and Plasterk, R.H.A. (1992) Proc. Nat. Acad. Sci. U.S.A. 89, 9598-9602.
51 Drelich, M.R., Wilhelm, R. and Mous, J. (1992) Virology 188, 459-468.
52 Jonsson, C.B. and Roth, M.J. (1993) J. Virol. 67, 5562-5571.
53 Leavitt, A.D., Shiue, L. and Varmus, H.E. (1993) J. Biol. Chem. 268, 2113-2119.
54 Vincent, K.A., Ellison, V., Chow, S.A. and Brown, P.O. (1993) J. Virol. 67, 425-437.
55 LaFemina, R.L., Schneider, C.L., Robbins, H.L., Callahan, P.L., LeGrow, K., Roth, E., Schleif, W.A. and Emini, E.A. (1992) J. Virol. 66, 7414-7419.
56 Schauer, M. and Billich, A. (1992) Biochem. Biophys. Res. Commun. 185, 874-880.
57 Clavel, F., Hoggan, M.D., Willey, R.L., Strebel, K., Martin, M. and Repaske, R. (1989) J. Virol. 63, 1455-1459.
58 Adachi, A., Ono, N., Sakai, H., Ogawa, K., Shibata, R., Kiyomasu, T., Masuike, H. and Ueda, S. (1991) Arch. Virol. 117, 45-58.
59 Sakai, H., Kawamura, M., Sakuragi, J.-I., Sakuragi, S., Shibata, R., Ishimoto, A., Ono, N., Ueda, S. and Adachi, A. (1993) J. Virol. 67, 1169-1174.
60 Stevenson, M., Haggerty, S., Lamonica, C.A., Meier, C.M., Welch, S.-K. and Wasiak, A.J. (1990) J. Virol. 64, 2421-2425.
61 Stevenson, M., Stanwick, T.L., Dempsey, M.P. and Lamonica, C.A. (1990) EMBO J. 9, 1551-1560.
62 Englund, G., Theodore, T.S., Freed, E.O., Engelman, A. and Martin, M.A. (1995) J. Virol. 69, 3216-3219.
63 Wiskerchen, M. and Muesing, M.A. (1995) J. Virol. 69, 376-386.
64 Cannon, P.M., Wilson, W., Byles, E., Kingsman, S.M. and Kingsman, A.J. (1994) J. Virol. 68, 4768-4775.

65 Engelman, A., Englund, G., Orenstein, J.M., Martin, M.A. and Craigie, R. (1995) J Virol. 69, 2729-2736.
66 Taddeo, B., Haseltine, W.A. and Farnet, C.M. (1994) J. Virol. 68, 8401-8405.
67 Shin, C.-G., Taddeo, B., Haseltine, W.A. and Farnet, C.M. (1994) J. Virol. 68, 1633-1642.
68 Bushman, F.D. and Wang, B. (1994) J. Virol. 68, 2215-2223
69 Kulkosky, J., Katz, R.A., Merkel, G. and Skalka, A.M. (1995) Virology 206, 448-456
70 Bushman, F.D., Engelman, A., Palmer, I., Wingfield, P. and Craigie, R. (1993) Proc Nat. Acad. Sci. U.S.A. 90, 3428-3432
71 Burke, C.J., Sanyal, G., Bruner, M.W., Ryan, J.A., LaFemina, R.L., Robbins, H.L., Zeft, A.S., Middaugh, C.R. and Cordingley, M.G. (1992) J. Biol. Chem. 267, 9639-9644.
72 Zheng, R., Jenkins, T.M. and Craigie, R. (1996) Proc. Nat. Acad. Sci. U.S.A. 93, 13659-13664
73 Mazumder, A.N., Neamati, J.O., Ojwang, S., Sunder, S., Rando, R.F. and Pommier, Y (1996) Biochemistry 35, 13762-13771.
74 Mizuuchi, K. and Adzuma, K. (1991) Cell 66, 129-140.
75 Beese, L.S. and Steitz, T.A. (1991) EMBO J. 10, 25-33.
76 Vink, C., Oude Groeneger, A.M. and Plasterk, R.H.A. (1993) Nucl. Acids Res. 21, 1419-1425.
77 Woerner, A.M., Klutch, M., Levin, J.G. and Marcus-Sekura, C.J. (1992) AIDS Res. Hum. Retrovir. 8, 297-304
78 Woerner, A.M. and Marcus-Sekura, C.J. (1993) Nucl. Acids Res. 21, 3507-3511.
79 Mumm, S.R. and Grandgenett, D.P. (1991) J. Virol. 65, 1160-1167
80 Puras Lutzke, R.A., Vink, C. and Plasterk, R.H.A. (1994) Nucl. Acids Res. 22, 4125-4131.
81 Rice, P. and Mizuuchi, K. (1995) Cell 82, 209-220.
82 Jones, K.S., Coleman, J., Merkel, G.W., Laue, T.M. and Skalka, A.M. (1992) J. Biol Chem. 267, 16037-16040.
83 Ellison, V., Gerton, J., Vincent, K.A. and Brown, P.O. (1995) J. Biol. Chem. 270, 3320-3326.
84 Pryciak, P.M. and Varmus, H.E. (1992) Cell 69, 769-780.
85 Pruss, D., Bushman, F.D. and Wolffe, A.P. (1994) Proc. Nat. Acad. Sci. U.S.A. 91, 5913-5917.
86 Bushman, F.D. and Craigie, R. (1992) Proc. Nat. Acad. Sci. U.S.A. 89, 3458-3462.
87 Bor, Y.-C., Bushman, F. and Orgel, L. (1995) Proc. Nat. Acad. Sci. U.S.A. 92, 10334-10338.
88 Muller, H.-P. and Varmus, H.E. (1994) EMBO J. 13, 4704-4714.
89 Scottoline, B.P., Chow, S., Ellison, V. and Brown, P.O. (1997) Genes Dev. 11, 371-382.
90 Bushman, F.D. and Craigie, R. (1990) J. Virol. 64, 5645-5648.
91 Fitzgerald, M.L., Vora, A.C., Zeh, W.G. and Grandgenett, D.P. (1992) J. Virol. 66, 6257-6263.
92 Fitzgerald, M.L. and Grandgenett, D.P. (1994) J. Virol. 68, 4314-4321.
93 Kukolj, G. and Skalka, A.M. (1995) Genes Dev. 9, 2556-2567
94 Goodarzi, G., Im, G.-J., Brackmann, K. and Grandgenett, D. (1995) J. Virol. 69, 6090-6097.
95 Miller, M.D., Farnet, C.M. and Bushman, F.D. (1997) J. Virol. 71, 5382-5390
96 Farnet, C.M. and Haseltine, W.A. (1991) J. Virol. 65, 1910-1915
97 Bowerman, B., Brown, P.O., Bishop, J.M. and Varmus, H.E. (1989) Genes Dev. 3, 469-478.

98 Gallay, P., Swingler, S., Song, J., Bushman, F. and Trono, D. (1995) Cell 83, 569-576.
99 Bukrinsky, M.I., Sharova, N., McDonald, T.L., Pushkarskaya, T., Tarpley, G.W. and Stevenson, M. (1993) Proc. Nat. Acad. Sci. U.S.A. 90, 6125-6129.
100 Farnet, C. and Bushman, F.D. (1997) Cell 88, 1-20.
101 Charneau, P., Alizon, M. and Clavel, F. (1992) J.Virol. 66, 2814-2820.
102 Charneau, P., Mirambeau, G., Roux, P., Paulous, S., Buc, H. and Clavel, F. (1994) J. Mol. Biol. 241, 651-662.
103 Miller, M.D., Wang, B. and Bushman, F.D. (1995) J. Virol. 69, 3938-3944.
104 Patel, P.H. and Preston, B.D. (1994) Proc. Nat. Acad. Sci. U.S.A. 91, 549-553.
105 Gallay, P., Swingler, S., Aiken, C. and Trono, D. (1995) Cell 80, 379-388.
106 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) Proc. Nat. Acad. Sci. U.S.A. 91, 7311-7315.
107 Fletcher, T.M., Brichacek, B., Sharova, N., Newman, M.A., Stivahtis, G., Sharp, P.M., Emerman, M., Hahn, B.H. and Stevenson, M. (1996) EMBO J. 15, 6155-6165.
108 Bustin, M. and Reeves, R. (1996) in Progress in Nucleic Acid Research and Molecular Biology (W.E Cohn and K. Moldave, eds.), pp. 35-100, Academic Press, San Diego, CA.
109 Zhou, X., Benson, K.F., Ashar, H.R. and Chada, K. (1995) Nature 376, 771-774.
110 Reeves, R. and Nissen, M.S. (1990) J. Biol. Chem. 265, 8573-8582.
111 Maher, J.F. and Nathans, D. (1996) Proc. Nat. Acad. Sci. U.S.A. 93, 6716-6720.
112 Strauss, F. and Varshavsky, A. (1984) Cell 37, 889-901.
113 Elton, T.S., Nissen, M.S. and Reeves, R. (1987) Biochem. Biophys. Res. Commun. 143, 260-265.
114 Reeves, R. and Wolffe, A.P. (1996) Biochemistry 35, 5063-5074.
115 Huth, J.R., Bewley, C.A., Nissen, M.S., Evans, J.N.S., Reeves, R., Gronenborn, A.M. and Clore, G.M. (1997) Nature Struct. Biol. 4, 657-665.
116 Carteau, S., Hoffmann, C. and Bushman, F.D. (1997) (unpublished data).
117 Li, L., Farnet, C.M., Anderson, W.F. and Bushman, F.D. (1997) (unpublished data).
118 Nash, H. (1996) in Regulation of Gene Expression in *Escherichia coli* (E.C.C. Lin and A.S. Lynch, eds.), R. G. Landees, Austin, TX.
119 Lavoie, B.D., Shaw, G.S., Millner, A. and Chaconas, G. (1996) Cell 85, 761-771.
120 Merker, P., Muskhelishvili, G., Deufel, A., Rusch, K. and Kahmann, R. (1993) Cold Spring Harbor Symp. Quant. Biol. 58, 505-513.
121 Nissen, M.S. and Reeves, R. (1995) J. Biol. Chem. 270, 4355-4360.
122 Reeves, R. and Nissen, M.S. (1993) J. Biol. Chem. 268, 21137-21146.
123 Lehming, N., Thanos, D., Brickman, J.M., Ma, J., Maniatis, T. and Ptashne, M. (1994) Nature 371, 175-179.
124 Thanos, D. and Maniatis, T. (1992) Cell 71, 777-789.
125 Thanos, D. and Maniatis, T. (1995) Cell 83, 1091-1100.
126 Falvo, J.V., Thanos, D. and Maniatis, T. (1995) Cell 83, 1101-1111.
127 Farnet, C.M. and Haseltine, W.A. (1991) J. Virol. 65, 6942-6952.
128 Lee, M.S. and Craigie, R. (1994) Proc. Nat. Acad. Sci. U.S.A. 91, 9823-9827.
129 Fujiwara, T. and Craigie, R. (1989) Proc. Nat. Acad. Sci. U.S.A. 86, 3065-3069.
130 Miller, M., Bor, Y.-C. and Bushman, F.D. (1995) Curr. Biol. 5, 1047-1056.
131 Carteau, S., Batson, S.C., Poljak, L., Mouscadet, J.-F., Rocquigny, H., Darlix, J.-L., Roques, B.P., Kas, E. and Auclair, C. (1997) J. Virol. 71, 6225-6229.
132 Kalpana, G.V., Marmon, S., Wang, W., Crabtree, G.R. and Goff, S.P. (1994) Science 266, 2002-2006.
133 Aiyar, A., Hindmarsh, P., Skalka, A.M. and Leis, J. (1996) J. Virol. 70, 3571-3580.
134 Stevens, S.W. and Griffith, J.D. (1996) J. Virol. 70: 6459-6462.

135 Bor, Y.-C., Miller, M., Bushman, F. and Orgel, L. (1996) Virology 222, 238-242
136 Bushman, F.D. (1994) Proc. Nat. Acad. Sci. U.S.A. 91, 9233-9237
137 Pruss, D., Reeves, R., Bushman, F.D. and Wolffe, A.P. (1994) J. Biol. Chem. 269, 25031-25041.
138 Rohdewohld, H., Weiher, H., Reik, W., Jaenisch, R. and Breindl, M. (1987) J. Virol. 61, 336-343.
139 Vijaya, S., Steffan, D.L. and Robinson, H.L. (1986) J. Virol. 60, 683-692
140 Scherdin, U., Rhodes, K. and Breindl, M. (1990) J. Virol. 64, 907-913
141 Milot, E., Belmaaza, A., Rassart, E. and Chartrand, P. (1994) Virology 201, 408-412
142 Withers-Ward, E.S., Kitamura, Y., Barnes, J.P. and Coffin, J.M. (1994) Genes Dev. 8, 1473-1487.
143 Shih, C.-C., Stoye, J.P. and Coffin, J.M. (1988) Cell 53, 531-537
144 Stevens, S.W. and Griffith, J.D. (1994) Proc. Nat. Acad. Sci. U.S.A. 91, 5557-5561
145 Howard, M.T. and Griffith, J.D. (1993) J. Mol. Biol. 232, 1060-1068
146 Chalker, D.L. and Sandmeyer, S.B. (1992) Genes Dev. 6, 117-128
147 Ji, H., Moore, D.P., Blomberg, M.A., Braiterman, L.T., Voytas, D.F., Natsoulis, G. and Boeke, J.D. (1993) Cell 73, 1-20
148 Zou, S., Wright, D.A. and Voytas, D.F. (1995) Proc. Nat. Acad. Sci. U.S.A. 92, 920-924
149 Kitamura, Y., Ha Lee, Y.M. and Coffin, J.M. (1992) Proc. Nat. Acad. Sci. U.S.A. 89, 5532-5536.
150 Brown, S.W. (1966) Science 151, 417-425.
151 Pluta, A.R., Mackay, A.M., Ainsztein, A.M., Goldberg, I.G. and Earnshaw, W.C. (1995) Science 270, 1591-1594
152 Wolffe, A.P. (1995) Curr. Biol. 5, 452-454
153 Verma, I.M. (1994) Mol. Med. 1, 2-3.
154 Sherman, P.A., Dickson, M.L. and Fyfe, J.A. (1992) J. Virol. 66, 3593-3601.
155 Vink, C., van Gent, D.C., Elgersma, Y. and Plasterk, R.H.A. (1991) J. Virol. 65, 4636-4644.
156 LaFemina, R.L., Callahan, P.L. and Cordingly, M.G. (1991) J. Virol. 65, 5624-5630.
157 Hazuda, D.J., Wolfe, A.L., Hastings, J.C., Robbins, H.L., Graham, P.L., LaFemina, R.L. and Emini, E.A. (1994) J. Biol. Chem. 269, 3999-4004.
158 Hong, T.E., Murphy, E., Groarke, J. and Drlica, K. (1993) J. Virol. 67, 1127-1131.
159 van den Ent, F.M., Vink, C. and Plasterk, R.H.A. (1994) J. Virol. 68, 7825-7832
160 Leavitt, A.D., Rose, R.B. and Varmus, H.E. (1992) J. Virol. 66, 2359-2368.
161 Panganiban, A.T. and Temin, H.M. (1983) Nature 306, 155-160.
162 Roth, M.J., Schwartzberg, P.L. and Goff, S.P. (1989) Cell 58, 47-54
163 Reicin, A.S., Kalpana, G., Paik, S., Marmon, S. and Goff, S. (1995) J. Virol. 69, 5904-5907
164 Colicelli, J. and Goff, S.P. (1985) Cell 42, 573-580.
165 Colicelli, J. and Goff, S.P. (1988) J. Mol. Biol. 199, 47-59.
166 Murphy, J.E., De Los Santos, T. and Goff, S.P. (1993) Virology 195, 432-440
167 Masuda, T., Plannelles, V., Krogstad, P. and Chen, I.S.Y. (1995) J. Virol. 69, 6687-6696
168 Varmus, H.E. and Brown, P.O. (1989) in Mobile DNA (D.E. Berg and M.M. Howe, eds.), pp. 53-108, A.S.M., Washington, DC.
169 Shiramizu, B., Herndier, B.G. and McGrath, M.S. (1994) Cancer Res. 54, 2069-2072.
170 Goulaouic, H. and Chow, S.A. (1996) J. Virol. 70, 37-46.
171 Katz, R.A., Merkel, G. and Skalka, A.M. (1996) Virology 217, 178-190
172 Bushman, F. and Miller, M.D. (1997) J. Virol. 71, 458-464.
173 Bushman, F.D. (1995) Science 267, 1443-1444.

174 Pavletich, N.P. and Pabo, C.O. (1991) Science 252, 809-817.
175 Pavletich, N.P. and Pabo, C.O. (1993) Science 261, 1701-1707.
176 Choo, Y. and Klug, A. (1994) Proc. Nat. Acad. Sci. U.S.A. 91, 11163-11167.
177 Rebar, E.J. and Pabo, C.O. (1994) Science 263, 671-673.
178 Jamieson, A.C., Kim, S.-H. and Wells, J.A. (1994) Biochemistry 33, 5689-5695.
179 Choo, Y. and Klug, A. (1994) Proc. Nat. Acad. Sci. U.S.A. 91, 11168-11172.
180 Desjarlais, J.R. and Berg, J.M. (1993) Proc. Nat. Acad. Sci. U.S.A. 90, 2256-2260.
181 Desjarlais, J.R. and Berg, J.M. (1992) Proteins 12, 101-104.
182 Desjarlais, J.R. and Berg, J.M. (1994) Proc. Nat. Acad. Sci. U.S.A. 91, 11099-11103.
183 Wu, H., Yang, W.-P. and Barbas, C.F. (1995) Proc. Nat. Acad. Sci. U.S.A. 92, 344-348.
184 Pomerantz, J.L., Sharp, P.A. and Pabo, C.O. (1995) Science 267, 93-96.
185 Pomerantz, J.L., Pabo, C.O. and Sharp, P.A. (1995) Proc. Nat. Acad. Sci. U.S.A. 92, 9752-9756.
186 Billich, A., Schauer, M., Frank, S., Rosenwirth, B. and Billich, S. (1992) Antiviral Chem. Chemother. 3, 113-119.
187 Fesen, M.R., Kohn, K.W., Leteurtre, F. and Pommier, Y. (1993) Proc. Nat. Acad. Sci. U.S.A. 90, 2399-2403.
188 Carteau, S., Mouscadet, J.F., Goulaouic, H., Subra, F. and Auclair, C. (1993) Biochem. Biophys. Res. Commun. 192, 1409-1414.
189 Mouscadet, J.F., Carteau, S., Goulaouic, H., Subra, F. and Auclair, C. (1994) J. Biol. Chem. 269, 21635-21638.
190 Mouscadet, J.F., Ketterle, C., Goulaouic, H., Carteau, S., Subra, F., Le Bret, M. and Auclair, C. (1994) Biochemistry 33, 4187-4196.
191 Carteau, S., Mouscadet, J.F., Goulaouic, H., Subra, F. and Auclair, C. (1993) Arch. Biochem. Biophys. 305, 606-610.
192 Cushman, M. and Sherman, P. (1992) Biochem. Biophys. Res. Commun. 185, 85-90.
193 Lipford, J.R., Worland, S.T. and Farnet, C.M. (1994) J. Acquired Immune Defic. Syndr. 7, 1215-1223.
194 Mazumder, A., Cooney, D., Agbaria, R., Gupta, M. and Pommier, Y. (1994) Proc. Nat. Acad. Sci. U.S.A. 91, 5771-5775.
195 Mazumder, A., Uchida, H., Neamati, N., Sunder, S., Jaworska-Maslanka, M., Wickstrom, E., Zeng, F., Jones, R.A., Mandes, R.F., Chenault, H.K. and Pommier, Y. (1997) Mol. Pharmacol. 51, 567-575.
196 Fesen, M.R., Pommier, Y., Leteurtre, F., Hiroguchi, S., Yung, J. and Kohn, K.W. (1994) Biochem. Pharmacol. 48, 595-608.
197 Mazumder, A., Raghavan, K., Weinstein, J., Kohn, K.W. and Pommier, Y. (1995) Biochem. Pharmacol. 49, 1165-1170.
198 Mazumder, A., Gupta, M., Perrin, D.M., Sigman, D.S., Rabinovitz, M. and Pommier, Y. (1995) AIDS Res. Hum. Retroviruses 11, 115-125.
199 LaFemina, R.L., Graham, P.L., LeGrow, K., Hastings, J.C., Wolfe, A., Young, S.D., Emini, E.A. and Hazuda, D.J. (1995) Antimicrob. Agents Chemother. 39, 320-324.
200 Farnet, C.M., Wang, B., Lipford, J.R. and Bushman, F.D. (1996) Proc. Nat. Acad. Sci. U.S.A. 93, 9742-9747.
201 Raghavan, K., Buolamwini, J.K., Fesen, M.R., Pommier, Y., Kohn, K.W. and Weinstein, J.N. (1995) J. Med. Chem. 38, 890-897.
202 Wolfe, A.L., Felock, P.J., Hastings, H.C., Blau, C.U. and Hazuda, D.J. (1996) J. Virol. 70, 1424-1432.
203 Stanwell, C., Ye, B., Yuspa, S.H. and Burke, T.R. (1996) Biochem. Pharmacol. 52, 475-480.
204 Zhao, H., Neamati, N., Hong, H., Mazumder, A., Wang, S., Sunder, S., Milne, G.W.A., Pommier, Y. and Burke, T.R. (1997) J. Med. Chem. 40, 242-249.

205 Robinson, W.E., Reinecke, M.G., Abdel-Malek, S., Jia, Q. and Chow, S.A. (1996) Proc. Nat. Acad. Sci. U.S.A. 93, 6326-6331.
206 Cushman, M., Golegiewski, W.M., Pommier, Y., Mazumder, A., Reymen, D., De Clercq, E., Graham, L. and Rice, W.G. (1995) J. Med. Chem. 38, 443-452.
207 Mazumder, A., Gazit, A., Levitzki, A., Nicklaus, M., Yung, J., Kohlhagen, G. and Pommier, Y. (1995) Biochemistry 34, 15111-15122.
208 Neamati, N., Hong, H., Mazumder, A., Wang, S., Sunder, S., Nicklaus, M.C., Milne, G.W.A., Proksa, B. and Pommier, Y. (1997) J. Med. Chem. 40, 942-951.
209 Zhao, H., Neamati, N., Mazumder, A., Sunder, S., Pommier, Y. and Burke, T.R. (1997) J. Med. Chem. 40, 1186-1194.
210 Nicklaus, M.C., Neamati, N., Hong, H., Mazumder, A., Sunder, S., Chen, J., Milne, G.W.A. and Pommier, Y. (1997) J. Med. Chem. 40, 920-929.
211 Eich, E., Pertz, H., Kaloga, M., Schulz, J., Fesen, M.R., Mazumder, A. and Pommier, Y. (1996) J. Med. Chem. 39, 89-95.
212 Puras Lutzke, R.A., Eppens, N.A., Weber, P.A., Houghten, R.A. and Plasterk, R.H.A. (1995) Proc. Nat. Acad. Sci. U.S.A. 92, 11456-11460.
213 Allen, P., Worland, S. and Gold, L. (1995) Virology 209, 327-336.
214 Ellison, V. and Brown, P.O. (1994) Proc. Nat. Acad. Sci. U.S.A. 91, 7316-7320.
215 Hazuda, D.J., Felock, P.J., Hastings, J.C., Pramanik, B. and Wolfe, A.L. (1997) J. Virol. 71, 7005-7011.