Target DNA capture by HIV-1 integration complexes

Michael D. Miller*, Yeou-Cherng Bor† and Frederic Bushman*

*Infectious Disease Laboratory, and †Chemical Evolution Laboratory, Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92024, USA.

Background: The early steps of human immunodeficiency virus 1 (HIV-1) replication involve reverse transcription of the viral RNA and integration of the resulting cDNA into a host chromosome. The DNA integration step requires the integration machinery ('preintegration complex') to bind to the host DNA before connecting the viral and host DNAs. Here, we present experiments that distinguish among three possible pathways of target-DNA capture: repeated binding and release of target DNA prior to the chemical strand-transfer step; binding followed by facilitated diffusion along target DNA (sliding); and integration at the initial target-capture site. The mechanism of target-DNA capture has implications for the design of gene therapy methods, and influences the interpretation of results on the selection of integration target sites *in vivo*.
Results: We present new *in vitro* conditions that allow us to assemble HIV-1 integrase — the virus-encoded recombination enzyme — with a viral DNA and then to trap assembled complexes bound to target DNA. We find that complexes of integrase and viral DNA do not slide along target DNA substantially after binding. We confirm and extend these results by analyzing target capture by a hybrid protein composed of HIV-1 integrase linked to a sequence-specific DNA-binding domain. We find that the integrase domain binds quickly and tightly under the above conditions, thereby obstructing function of the fused sequence-specific DNA-binding domain. We also monitor target-DNA capture by HIV-1 preintegration complexes purified from freshly infected cells. Partially purified complexes commit quickly and stably to the first target DNA added, whereas preintegration complexes in crude cytoplasmic extracts do not. The addition of extracts from uninfected cells to partially purified complexes blocks quick commitment.
Conclusions: Under new conditions favorable for the analysis of target-DNA capture *in vitro*, HIV-1 integrase complexes bind quickly and stably to target DNA without subsequent sliding. Parallel studies of preintegration complexes support a model in which target-site capture *in vivo* is reversible as a result of the action of cellular factors.

Current Biology 1995, **5**:1047–1056

Introduction

In order to replicate, a retrovirus must integrate a DNA copy of its RNA genome into a chromosome of the host [1]. How the retroviral integration system selects a particular integration-acceptor site from among the myriad possible sites *in vivo* is unclear. Integration acceptor sites have been reported to lie most frequently near DNAse I-hypersensitive sites [2,3], consistent with the idea that open chromatin regions are favored for integration. However, a recent study has presented evidence that many chromosomal regions serve as integration-acceptor sites with about equal frequency [4]. Another possible influence on target site selection is the potential binding of the integration apparatus to host proteins bound to cellular DNA. Such a tethering mechanism may explain selective integration by the retrovirus-like yeast retrotransposon Ty3, which integrates near the start site of genes transcribed by RNA Polymerase III [5,6]. Yeast Ty1 also displays highly selective integration [7], mediated by an unknown but possibly related mechanism.

We have demonstrated previously, *in vitro*, that a hybrid protein composed of human immunodeficiency virus 1 (HIV-1) integrase fused to the DNA-binding domain of λ repressor could direct integration selectively to DNA sequences near binding sites for repressor [8]. The finding that this artificial means of tethering an integrase to target DNA suffices for selective integration is consistent with the idea that tethering might mediate selective integration in naturally occurring systems. Furthermore, a cellular protein that binds HIV-1 integrase has been identified and proposed to promote integration *in vivo* by tethering integrase to target DNA [9].

Which, if any, of the above mechanisms is likely to influence the selection of the integration site *in vivo* will be strictly determined by the nature of the association between the target DNA and the integration complex. Three possible association mechanisms are presented in Figure 1. Firstly, the integrase–HIV-1 DNA complex might bind and release target DNA many times before actually carrying out the chemical steps of integration (Fig. 1a). Secondly, integrase might bind to target DNA and slide along it for some distance prior to integration (Fig. 1b). Thirdly, integrase might bind irreversibly to the first target DNA encountered and carry out integration at that site (Fig. 1c). For tethering to affect the selection of an integration target site, some form of target-DNA exploration must take place, by either of the first two mechanisms.

Here, we investigate the mechanism of target-DNA capture using integration reactions *in vitro*. We have

Correspondence to: Frederic Bushman.

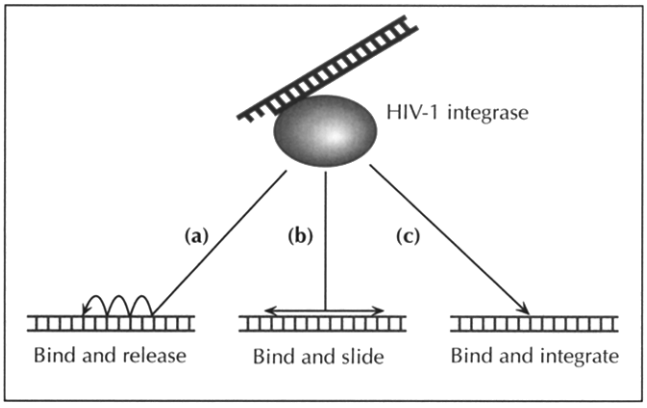

Fig. 1. Candidate mechanisms for target DNA capture by HIV-1 integrase. An HIV-1 DNA end is represented by a thick ladder, and target DNA by a thin ladder. **(a)** The integration complex binds and releases target DNA many times before carrying out the chemical step of integration. **(b)** The integration complex binds to target DNA and slides along it prior to carrying out integration. **(c)** The integration complex binds to target DNA and integrates at that site.

found new conditions that allow the target capture and chemical steps of integration to be distinguished, and used such reactions to investigate the mechanism of target-capture by complexes of integrase and model viral cDNAs. We have also compared the mechanism of target capture using two other sources of integration activity: a fusion protein consisting of integrase and λ repressor, and viral nucleoprotein complexes (pre-integration complexes) purified from infected cells. We find that both of the pathways shown in Figure 1a,c can operate under different conditions, and that components of host cells may influence the pathway of target-site capture.

Results

Strand transfer directed by purified HIV-1 integrase in the presence of Mg^{2+} and DMSO

Experiments to distinguish target-DNA binding from the chemical step of strand transfer were made possible by the development of conditions *in vitro* in which target binding is relatively fast compared with strand transfer. For these reactions, an oligonucleotide matching one end of the unintegrated viral cDNA (the long terminal repeat or LTR) was used as an integration donor (labeled with ^{32}P at the 5′ end), and a circular plasmid DNA was used as an integration target (Fig. 2a) [10–13]. Integration results in the covalent attachment of the LTR oligonucleotide to one strand of the target DNA, yielding a 'tagged circle' product. Integration products migrate in gels with the mobility of the relaxed circular form of the target plasmid, as integration introduces a nick in the plasmid and the mass contributed by the integrated oligonucleotide is small.

In the presence of Mn^{2+}, purified HIV-1 integrase is capable of removing two nucleotides from the 3′ end of the viral cDNA [10,14–16]. In the presence of Mg^{2+}, this activity was not prominent under our conditions (data not shown). However, strand transfer took place efficiently in the presence of Mg^{2+} using LTR oligonucleotides that lacked the two nucleotides normally removed by integrase. Reaction conditions employing

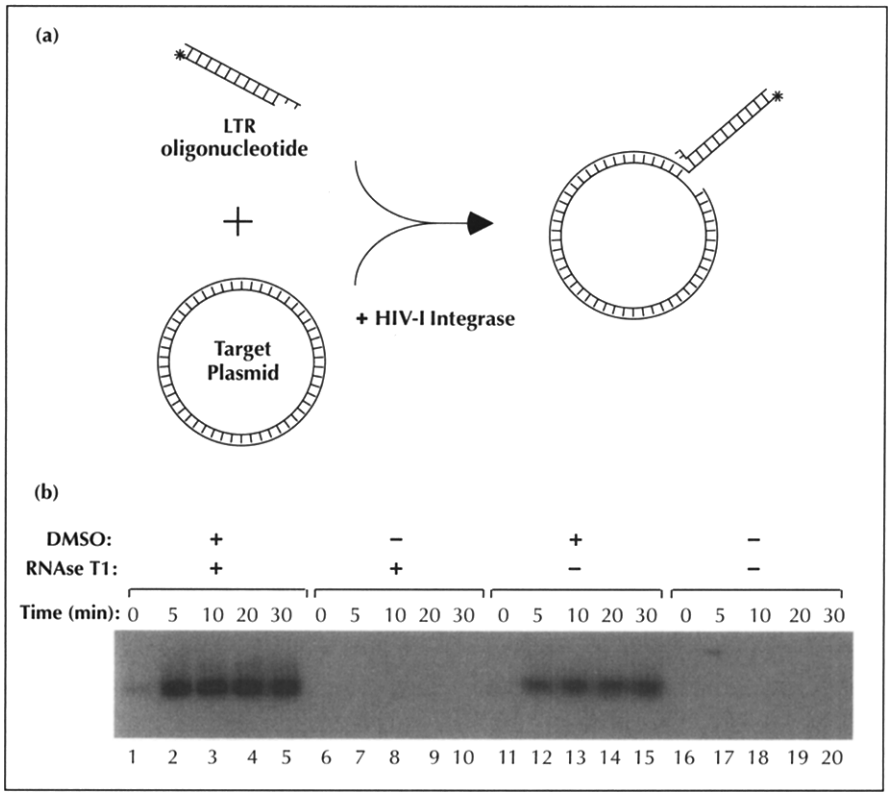

Fig. 2. Integration reactions in the presence of Mg^{2+}. **(a)** The integration reaction. DNAs are drawn as ladders; the asterisk indicates a ^{32}P label at the 5′ end of the LTR oligonucleotide. **(b)** Time courses of integration reactions in the presence or absence of DMSO or RNase T1.

Mg^{2+} instead of Mn^{2+} are of particular interest, as Mg^{2+} is more abundant *in vivo* and hence is likely to be the metal used by integrase during retroviral infection.

Integrase and the LTR oligonucleotide were preincubated at 37 °C for 5 minutes, forming a stable complex as described previously [17]. Target DNA was then added and the incubation continued at 37 °C for 30 minutes. dimethyl sulfoxide (DMSO) stimulated strand transfer substantially under these conditions, and the addition of a small basic protein (RNAse T1) stimulated transfer slightly further. The effect of a small basic protein was tested because such proteins have been reported previously to stimulate integration in an *in vitro* genetic assay [11]. The reaction was largely complete after 10 minutes (Fig. 2b).

Rapid and stable capture of target DNA by purified integrase

To begin to investigate target DNA capture by integrase–LTR complexes, integration reactions were carried out in which two target DNAs were added sequentially, to ask whether the first target added was preferred (Fig. 3). After both targets were added, reaction mixtures were incubated for a further 30 minutes to allow integration products to accumulate. To distinguish integration into each plasmid in mixtures, two plasmids of different sizes (plasmid A of 5.4 kb and plasmid B of 1.2 kb, see Materials and methods) were used.

The simultaneous addition of plasmids A and B yielded radioactive signals with the mobility of the relaxed circular form of each plasmid, as expected for integration into both target DNAs (Fig. 3a,b, lane 1). Two bands arise from integration into plasmid B (labeled B and B′) because the small plasmid B requires a second larger plasmid (B′) to support replication, and both plasmids are present in the preparation. Addition of plasmid B only 20 seconds after plasmid A resulted in integration into plasmid A almost exclusively (Fig. 3a, compare lanes 1,2). A similar preference was seen when plasmid B was added one, two or six minutes after plasmid A. Adding plasmid B before plasmid A (Fig. 3b) resulted in preferential integration into plasmid B, indicating that the order of addition and not the identity of the target determined the target DNA selected.

As a control, it was necessary to show that the observed preferential use of the first target added was due to fast commitment rather than fast integration. Reactions were therefore carried out exactly as above, except that the reactions were stopped at the time of addition of the second target plasmid. After 20 seconds of incubation with the first target DNA, no significant integration was seen (Fig. 3a,b, lane 6). After 1 minute, integration product was detectable and it was prominent by 6 minutes. Thus, the strong bias for integration into the first target DNA added cannot be explained by fast integration into that target within the first 20 seconds. These data support a model in which target capture under these conditions proceeds by either fast irreversible binding (Fig. 1c) or binding with subsequent sliding (Fig. 1b).

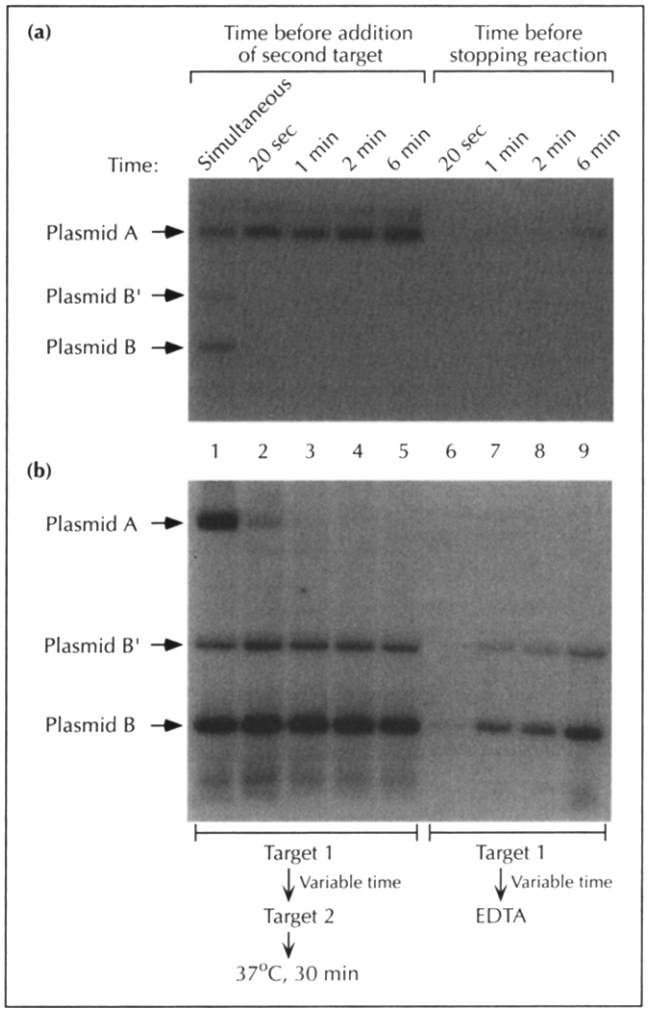

Fig. 3. Stable commitment of integrase–LTR complexes to target DNA. Reactions were carried out in the presence of 5 mM $MgCl_2$ and 20 % DMSO. **(a)** Addition of plasmid A followed by addition of plasmid B at various times thereafter. **(b)** Addition of plasmid B followed by addition of plasmid A at various times. Times before addition of the second plasmid are indicated at the top of the autoradiograms (lanes 1–5); times before stopping the reaction are indicated above lanes 6–9. Reactions containing Mn^{2+} were too fast to be analyzed by this method (data not shown). See text for details.

Lack of sliding on target DNA following target capture

To determine whether integrase–LTR complexes slide along target DNA in the above conditions, integrase–LTR complexes were prebound to a linear target DNA, a second target was ligated onto the first, and integration into the second target DNA was then monitored. Sliding of the complex from one target to the other would result in accumulation of integration product in the second target. To allow integration in each target to be monitored, the two targets were separated by digestion with a restriction enzyme after the reaction was stopped. Such an approach was used previously to demonstrate that the phage Mu B protein (a sequence-nonspecific DNA-binding protein) did not slide along DNA after binding [18].

Integrase–LTR complexes were formed by incubating integrase and an LTR oligonucleotide at 37 °C, and two

target DNAs — ϕX174 DNA (5.4 kb) and pMM105 DNA (3.1 kb), each linearized by digestion with *Pst*I — were added under various conditions. Firstly, the two target DNAs were added simultaneously, and then incubated for 30 minutes at 37 °C. Integration into both targets was observed, as expected (Fig. 4a, lane 1). Reactions were then carried out at 0 °C — conditions in which ligase remains active whereas integrase is inactive. Integrase–LTR complexes were prebound to the two target DNAs for 20 seconds, immediately frozen in liquid nitrogen, and then thawed and incubated on ice for one hour. After this treatment, no integration product accumulated (Fig. 4, lane 2). When such a reaction mixture was subsequently warmed at 37 °C for 30 minutes, high levels of integration products were obtained (Fig. 4, lane 3).

We also carried out mock ligation reactions in which integrase–LTR complexes were prebound to one target DNA and the second target was added 20 seconds later. The mixtures were then quickly chilled by plunging in liquid nitrogen, thawed on ice for one hour, and then warmed to 37 °C for 30 minutes. As expected, integration product was detected predominantly in the first target added, whether it was the larger ϕX174 DNA target (Fig. 4, lane 4) or the smaller pMM105 target (Fig. 4, lane 7).

To ask whether integrase–LTR complexes could slide along the target DNA, these latter integration reactions were repeated with the addition of ligase and ATP during the one hour incubation on ice. After this step, the bulk target DNA and the integration product were both present as high molecular weight forms, indicating that substantial ligation had taken place (Fig. 4a,b, lanes 5,8). Cleavage of these ligated DNAs with *Pst*I regenerated the two linear DNA forms (Fig. 4a,b, lanes 6,9). Ligation of the DNAs did not induce the accumulation of integration product in the second target DNA added, indicating that substantial sliding along target DNA did not take place during the one hour incubation on ice or the time between warming of the reaction mixtures at 37 °C and the completion of strand transfer. Note that previous work determined that the lac repressor of *Escherichia coli* can 'scan' about 10^3 base pairs per second by sliding [19]; thus, sufficient time should have elapsed in our integration reactions for sliding from one target DNA to the other prior to strand transfer. Our data therefore indicate that, under these conditions, target DNA capture involves irreversible binding to target DNA without subsequent sliding over long distances (consistent with the pathway illustrated in Fig. 1c).

A fusion protein of integrase and a sequence-specific DNA-binding domain as a probe of target site capture

The conclusions reached in the preceding section were challenged by a second experimental approach, using a fusion of HIV-1 integrase (IN) to the DNA-binding domain of λ repressor (λR). This fusion protein (λR–IN) was shown previously to direct selective integration into

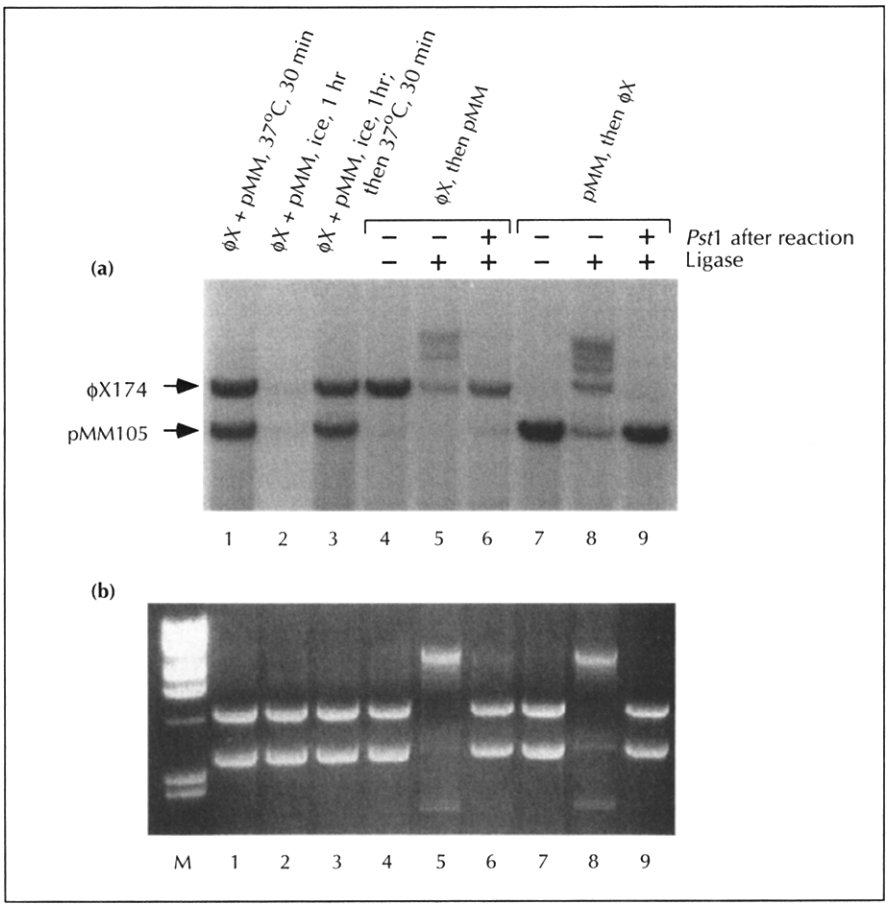

Fig. 4. Assay of sliding on target DNA by bound integrase–LTR complexes. **(a)** Autoradiogram revealing integration product. **(b)** Bulk DNA in the same electrophoresis gel, visualized by staining with ethidium bromide. For all reactions, integrase was preincubated with an end-labeled oligonucleotide for 5 min at 37 °C prior to addition of target DNA. Lane 1: integration products generated by adding two linear DNA targets (ϕX+∓÷174 and pMM105) simultaneously. Lane 2: integration products generated in similar reactions incubated for 1 h on ice. Lane 3: integration products generated in reactions incubated on ice for one hour, then warmed to 37 °C for 30 min. Lane 4: addition of ϕX174 DNA, followed by addition of pMM105 DNA 20 sec later; reactions were chilled for 1 h, then warmed to 37 °C for 30 min. Lane 5: addition of ϕX174 DNA, followed by addition of pMM105 DNA 20 sec later; reactions were incubated on ice for 1 h in the presence of ligase, then warmed for 30 min at 37 °C. Lane 6: similar to lane 5 except the ligated integration products were separated by cleavage with *Pst*1 prior to electrophoresis. Lanes 7–9 are identical to lanes 4–6, except that pMM105 DNA was added first.

DNA targets containing recognition sites for λ repressor (operators). λR–IN was prebound to target DNA, and the LTR oligonucleotide was added last to start the reaction [8]. These conditions allowed the λR domain to bind and release many sequences before locating the operators and binding tightly at that site. However, under conditions *in vitro* in which the integrase part of the fusion protein binds quickly and tightly to target DNA, the λR domain may not be able to direct selective integration efficiently because tight binding by the integrase domain will prevent the λR domain from locating the operators. The outcome of the competition between the λR and integrase parts of the fusion, monitored by selective integration, allows the mode of binding of the integrase part to be assessed.

Initially, target DNA order-of-addition studies were performed using λR–IN in the presence of Mg^{2+} and DMSO. λR–IN bound tightly to the first target DNA added and, consistent with the results using wild-type integrase, did not exchange onto a second target DNA added just 20 seconds later (data not shown). To examine in detail the integration sites selected under different conditions, reactions were carried out *in vitro* and analyzed using the polymerase chain reaction (PCR) (Fig. 5a,b) [8,20,21]. Integration reactions were conducted using an LTR oligonucleotide that was not end-labeled. Products of integration reactions were deproteinized and used as templates for PCR. PCR primers were selected such that one primer was complementary to a target DNA sequence, and the other was complementary to the LTR oligonucleotide. The target primer was labeled with ^{32}P on the 5′ end. PCR amplification of integration products generated a population of molecules that were then denatured and analyzed on a DNA-sequencing-type gel. Each band on the gel corresponded to integration at a specific phosphodiester. The frequency of integration at a particular site was reflected in the intensity of the corresponding band on the final autoradiogram, and the location of each integration site was determined by co-electrophoresis with DNA sequencing reactions. Integration was examined at the right operator (O_R) region of phage λ DNA, which contains three binding sites for λ repressor. PCR primers to either side of O_R were used to assay integration into each target DNA strand (Fig. 5b, lanes 1–6).

Figure 5b presents the results of such an analysis of integration sites used by HIV integrase and λR-IN. Reactions were carried out in the presence of $MgCl_2$ and DMSO. Two different orders of addition were compared: integrase or λR–IN was prebound to target DNA, and then LTR oligonucleotide was added; or integrase or λR–IN was preassembled with the LTR DNA, and the complex subsequently incubated with target DNA.

For wild-type integrase assembled with the LTR oligonucleotide in the presence of $MgCl_2$ and DMSO, a band can be seen at most positions in the integration product ladder, although the frequency of integration at each phosphodiester varied over a broad range (Fig. 5b, lanes 1,4). For unknown reasons, preincubating wild-type integrase with the target DNA under these conditions prior to the addition of the LTR oligonucleotide inhibited the reaction (data not shown).

In the case of reactions containing λR–IN, different results were obtained depending on how the integration complexes were assembled. When λR–IN was prebound to target DNA and the LTR oligonucleotide added last, selective integration was seen (Fig. 5b, lanes 3,6). Little integration was seen into the regions containing the repressor binding sites O_R1 and O_R2, as expected if λR–IN was bound at these sites. A similar 'footprint' over this region can be seen in reactions containing wild-type integrase and λ repressor [8]. Integration took place predominantly at a set of highly preferred sites adjacent to the operators. Farther away from O_R, integration events were less frequent than in the products of reactions with wild-type integrase. The hotspots for integration with λR–IN lie on the same face of the DNA helix as the repressor binding sites, consistent with a model in which the λR–IN complex bound at O_R captures the target DNA by looping out the intervening sequences [8].

When λR–IN was preassembled with an LTR oligonucleotide in the presence of Mg^{2+} and DMSO, however, the operator DNA was only partially protected from integration, and the strong enhancements seen when λR-IN was prebound to target were greatly reduced in intensity (Fig. 5b, lanes 2,5). Phosphodiesters far from O_R were used relatively efficiently by λR–IN following assembly with the LTR DNA, as with wild-type integrase. Thus, selective integration was inhibited by assembling λR–IN with LTR DNA. The λR–IN–LTR complex apparently bound tightly to target DNA without subsequent sliding (Fig. 1c), whereas λR–IN alone, when incubated with target DNA, was capable of some form of target DNA exploration (consistent with the pathways shown in Fig. 1a,b).

Capture of target DNA by purified HIV-1 preintegration complexes

The finding that complexes of HIV integrase, assembled with LTR DNA in the presence of $MgCl_2$ and DMSO, bind quickly and stably to target DNA raised the question of whether the large nucleoprotein complexes that carry out integration *in vivo* behave similarly. Such subviral particles ('preintegration complexes') containing the reverse-transcribed viral cDNA and associated proteins can be obtained from freshly infected cells and are capable of integrating the endogenous viral cDNA into an added target DNA *in vitro* [22–24].

To investigate target capture by preintegration complexes, it was first necessary to determine the kinetics of integration (Fig. 6a). Linear phage ϕX174 target DNA was added to complexes either in crude extracts or partially purified, and samples were taken at time points over 45 minutes. The samples were then deproteinized,

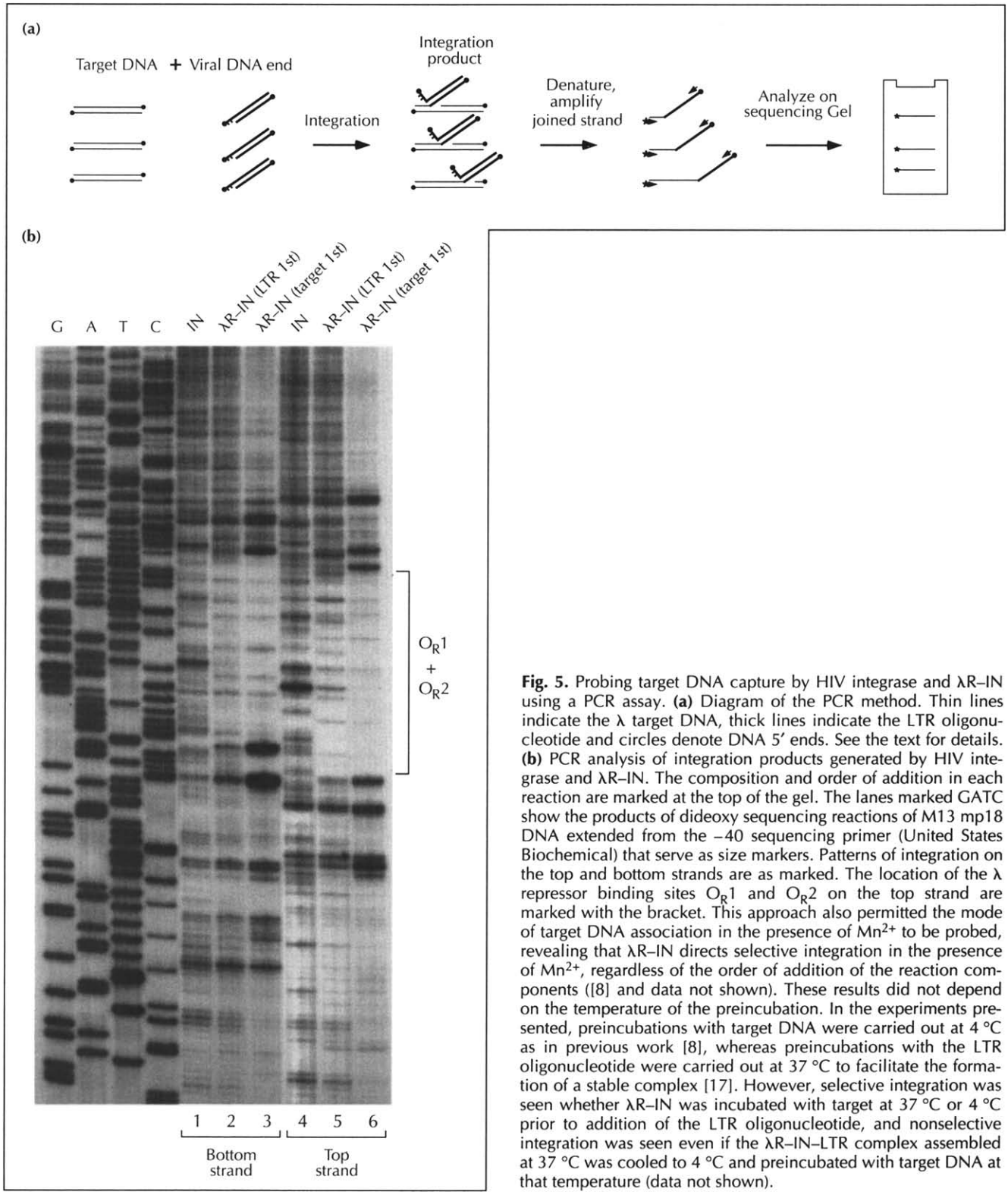

Fig. 5. Probing target DNA capture by HIV integrase and λR–IN using a PCR assay. **(a)** Diagram of the PCR method. Thin lines indicate the λ target DNA, thick lines indicate the LTR oligonucleotide and circles denote DNA 5′ ends. See the text for details. **(b)** PCR analysis of integration products generated by HIV integrase and λR–IN. The composition and order of addition in each reaction are marked at the top of the gel. The lanes marked GATC show the products of dideoxy sequencing reactions of M13 mp18 DNA extended from the −40 sequencing primer (United States Biochemical) that serve as size markers. Patterns of integration on the top and bottom strands are as marked. The location of the λ repressor binding sites O_R1 and O_R2 on the top strand are marked with the bracket. This approach also permitted the mode of target DNA association in the presence of Mn^{2+} to be probed, revealing that λR–IN directs selective integration in the presence of Mn^{2+}, regardless of the order of addition of the reaction components ([8] and data not shown). These results did not depend on the temperature of the preincubation. In the experiments presented, preincubations with target DNA were carried out at 4 °C as in previous work [8], whereas preincubations with the LTR oligonucleotide were carried out at 37 °C to facilitate the formation of a stable complex [17]. However, selective integration was seen whether λR–IN was incubated with target at 37 °C or 4 °C prior to addition of the LTR oligonucleotide, and nonselective integration was seen even if the λR–IN–LTR complex assembled at 37 °C was cooled to 4 °C and preincubated with target DNA at that temperature (data not shown).

separated by electrophoresis in a native agarose gel, blotted to nylon membranes, and probed with labeled sequences complementary to the HIV LTR. Integration products were seen as bands with mobility equal to that of the HIV DNA (9.7 kb) plus the ϕX174 target DNA (5.4 kb). Integration took place quickly, with a half-maximal time of about three minutes for assays of both cytoplasmic extracts and partially purified complexes (Fig. 6a,b; data not shown). A titration of phage ϕX174 target DNA established that 0.5 μg was sufficient to saturate the reaction (Fig. 6c).

To probe target-site capture, integration reactions were carried out in which a target DNA was added to pre-integration complexes as above, and then a second target was added after 20 seconds. Reactions were subsequently

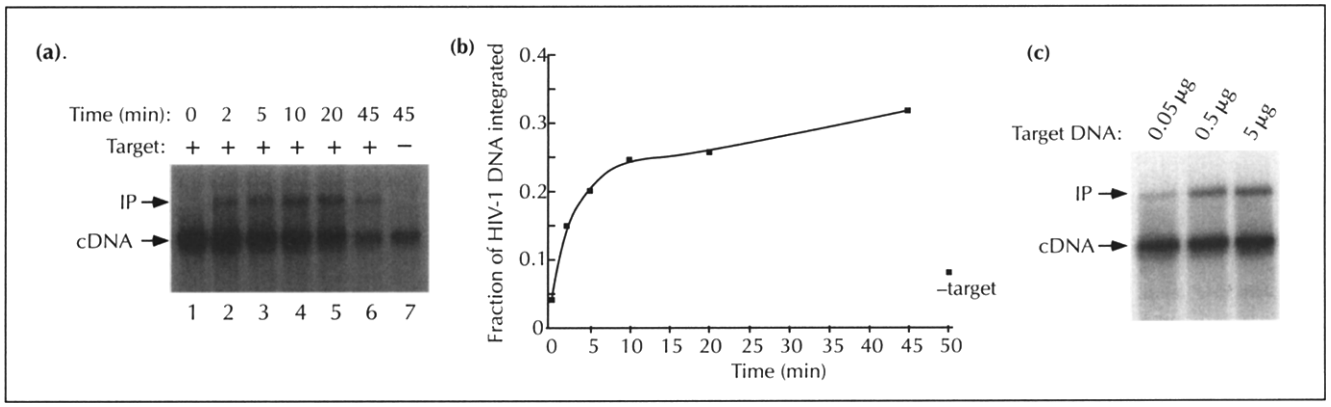

Fig. 6. Characterization of integration reactions containing HIV-1 preintegration complexes. **(a)** Time course of integration. Time after addition of the phage ϕX174 target DNA is shown above the autoradiogram. The band corresponding to integration product (IP) is marked. **(b)** Quantification of the kinetics of integration reactions containing preintegration complexes. Quantification was carried out using a Molecular Dynamics Phosphorimager. Each point represents the average of two experiments. **(c)** Titration of phage ϕX174 target DNA. The amount of ϕX174 target DNA added is marked above the gel.

incubated for 30 minutes to allow integration product to accumulate. Linear phage ϕX174 DNA was used as one target, and sheared salmon sperm DNA as the other. Integration into the salmon sperm DNA yielded only a weak background smear of integration products, as the fragments of salmon sperm DNA are heterogeneous in size, whereas integration into ϕX174 DNA yielded a discrete product band as described above.

For the case of preintegration complexes in cytoplasmic extracts, the addition of ϕX174 DNA first, followed by salmon sperm DNA 20 seconds later, yielded integration product as in reactions containing ϕX174 target DNA only (Fig. 7a, lanes 1–3). When salmon sperm DNA was added 20 seconds before ϕX174 DNA, integration products were also seen into the ϕX174 target DNA, indicating that the preintegration complexes in cytoplasmic extracts did not stably commit to the first target DNA added (Fig. 7a, lanes 4–6). A time course of integration carried out in parallel confirmed that little integration took place within the first 20 seconds (data not shown).

However, a different result was obtained with preintegration complexes partially purified by gel filtration. Integration into the ϕX174 target DNA was seen only when it was added first. No integration was detected when the ϕX174 DNA was added 20 seconds after the salmon sperm DNA (Fig. 7b, compare lanes 3,4). These results are not due to fast integration into the first target added, as a parallel time course showed that little integration

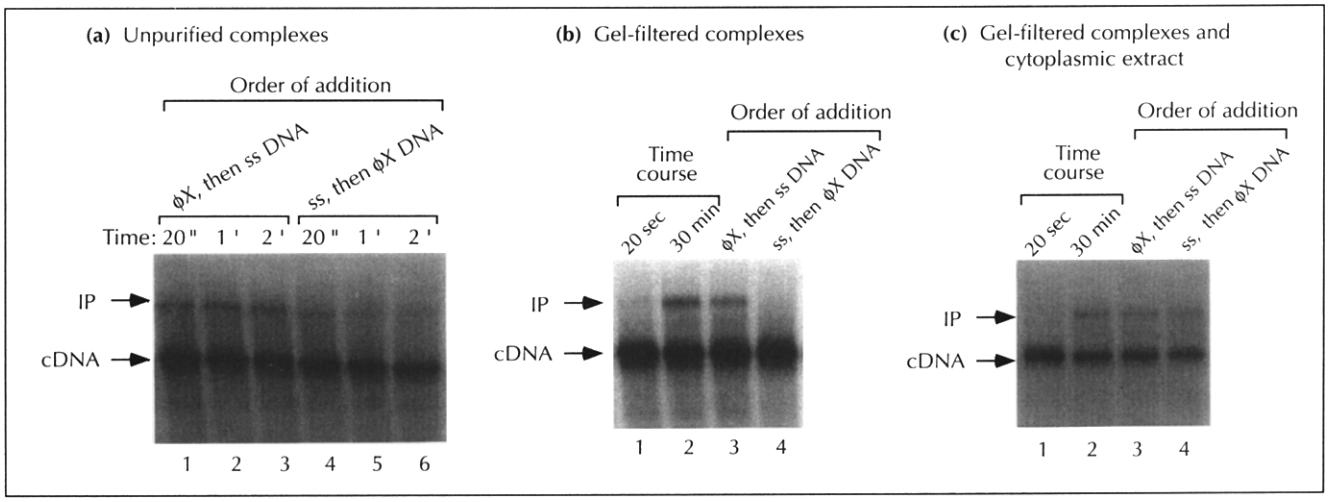

Fig. 7. Effects of order of target-DNA addition and of cytoplasmic extracts on integration reactions containing preintegration complexes. **(a)** Results obtained using cytoplasmic extracts containing preintegration complexes. The order of addition of the two target DNAs — phage ϕX174 DNA (ϕX) or salmon sperm (ss) DNA — is marked above the gel lanes. Times correspond to the interval between addition of the two targets. 'IP' marks integration product, 'cDNA' marks the unintegrated HIV DNA. **(b)** Results obtained using preintegration complexes partially purified by gel filtration. **(c)** Results obtained using partially purified preintegration complexes mixed with cytoplasmic extracts from uninfected cells. Target DNAs were added 20 seconds apart. The methods for the analysis of target-DNA capture described could not be used for reactions containing Mn^{2+} because the reaction proceeded too quickly (data not shown).

product accumulated in the first 20 seconds (Fig. 7b, lanes 1,2). Thus, preintegration complexes, partially purified by gel filtration, bound quickly and tightly to target DNA, but no such commitment was seen with preintegration complexes in crude extracts.

To begin to examine whether host factors removed by gel filtration were responsible for the difference in target capture, the activity of preintegration complexes purified by gel filtration was assessed in the presence of added cytoplasmic extracts from uninfected cells. Integration into the ϕX174 target DNA was seen whether it was added before or after the salmon sperm DNA challenger (Fig. 7c, compare lanes 3,4). Addition of the lysis buffer used to prepare the cytoplasmic extract did not block target commitment (data not shown). Evidently, components of the host cell extract altered target DNA capture so as to permit integration into the second DNA added. Partially purified complexes therefore committed quickly and tightly to the first target DNA added (consistent with Fig. 1b,c), whereas the presence of cytoplasmic extract permitted exchange (consistent with Fig. 1a).

Discussion

The capture of target DNA by an integrase–LTR complex prior to integration could, in principle, take place by any of three mechanisms. The integrase–LTR complex might bind and release candidate target DNA sequences several times (Fig. 1a); the complex might bind to target DNA and slide along it (Fig. 1b); or the complex might integrate at the initial site of binding (Fig. 1c). The results we have described show that complexes of purified integrase assembled with LTR DNA in the presence of $MgCl_2$ and DMSO bind quickly to target DNA, without dissociation prior to integration (consistent with the pathways shown in Fig. 1b,c). Further experiments established that the target-bound complex does not slide far along the DNA (consistent with the pathway shown in Fig. 1c).

We also investigated target capture using a fusion protein composed of the HIV-1 integrase fused to the DNA-binding domain of λ repressor. In this case, the outcome of the competition between the two DNA-binding domains under different conditions provides insights into the function of each. λR–IN preassembled with LTR DNA failed to direct selective integration at λR-binding sites in the presence of $MgCl_2$ and DMSO, although selective integration was seen in other conditions. Evidently, in the presence of $MgCl_2$ and DMSO, the integrase part of the fusion binds tightly to target DNA without subsequent sliding (consistent with Fig. 1c), thereby preventing the repressor domain from locating its recognition site efficiently.

The results of target-commitment experiments using preintegration complexes from HIV-infected cells were different depending on whether or not a source of host factors was present. Partially purified complexes committed quickly and stably to the first target DNA added (consistent with Fig. 1b,c), whereas preintegration complexes in crude cytoplasmic extracts did not display fast template commitment (consistent with Fig. 1a). The addition of a cytoplasmic extract to partially purified complexes blocked quick commitment, consistent with the idea that the mechanism of target-site selection may be influenced by cellular factors *in vivo*.

DNA sliding

Previous studies have revealed that several sequence-specific DNA-binding proteins probably locate their recognition sites by first binding to nonspecific DNA and then diffusing laterally (sliding) — the *E. coli* lac repressor, the restriction enzyme *Eco*R1 and λ Cro protein have all been reported to use this mechanism [19,25,26]. However, the phage Mu B protein, a recombination enzyme cofactor, does not slide along DNA detectably after binding [18]. Mu B, together with Mu A transposase, catalyzes transposition of Mu DNA. Mu B binds to target DNA sites in a sequence-nonspecific fashion and, once bound, stimulates DNA strand transfer by Mu A. In a ligation assay such as that presented in Figure 4, Mu B did not slide detectably along DNA after binding. The finding that the Mu and HIV recombination systems can bind DNA rapidly without subsequent sliding suggests that such binding may be a general property of these enzymes *in vitro*.

Although integrase–LTR complexes do not slide for long distances in the presence of Mg^{2+}, the detailed pattern of integration sites used indicates that the complex may slide within a limited region. In the PCR analysis of integration sites shown in Figure 5, as in previous studies [27–29], different sites are used as integration targets with varying efficiencies. It seems unlikely that such variation is due to differences in the accessibility of sites in the naked target DNAs. Perhaps the integrase–LTR complex explores a few nearby base pairs in the course of target DNA capture, with the observed hotspots representing sequences that are particularly favorable for binding or catalysis. Alternatively, complexes bound to unfavorable sites may be quickly released and rebound elsewhere. In our experiments, such exchange would have to take place prior to the addition of competitor DNA (within 20 seconds).

An integrase–λ repressor fusion protein as a probe of target DNA capture

The fusion protein composed of the DNA-binding domain of λ repressor and the HIV integrase (λR–IN) provides a novel tool for studying the mechanism of target-DNA capture. For a sequence-specific DNA-binding protein to locate its recognition sequence, it must either bind and release many nonspecific DNA sequences or slide along nonspecific DNA once bound. Such scanning of DNA sites is necessary in order to locate the specific recognition sequence. If the protein is restricted in its exploration of DNA, then specific binding may not take place. Thus, the competition between the integrase and λR parts of λR–IN, as revealed by

occupancy of the repressor site, serves as a reporter for the mode of binding by the integrase domain.

We found that the efficiency of selective integration was influenced by the reaction conditions employed. λR–IN–LTR complexes assembled in the presence of Mg^{2+} and DMSO displayed little selective integration, as expected from the finding that complexes of wild-type integrase bound to LTR DNA under these conditions commit quickly to target DNA without sliding. However, under these conditions, prebinding of λR–IN to target DNA prior to adding the LTR oligonucleotide did permit selective integration. Perhaps prebinding of the LTR oligonucleotide alters the behavior of the target-DNA binding site, converting it from relatively low to relatively high affinity.

Implications for integration *in vivo*

Our findings indicate that target DNA capture is different in reactions containing partially purified preintegration complexes than in reactions containing complexes and cellular factors. The mechanism of action of the cellular factors is unknown, as is the biological significance of this observation. One model for the mechanism invokes the effects of host proteins bound to target DNA. It has been shown previously that the DNA-binding domain of the yeast GAL4 protein binds its recognition site 10-fold less tightly in nucleosomal DNA than in naked DNA [30]. Perhaps the presence of host DNA-binding proteins bound to target DNA destabilizes the initial association of preintegration complexes with the target and thereby permits exchange between targets.

The mechanism of target-site capture is a critical determinant of whether fusion proteins, such as λR–IN, that contain integrase linked to sequence-specific DNA-binding domains will direct selective integration *in vivo*, an issue of interest in developing gene therapy methods [31]. It seems most likely that conditions *in vivo* resemble those seen with preintegration complexes in the presence of host factors, in which stable target-site commitment is not seen (Fig. 1a). If so, integrase fusions may be able to locate specific sites and direct selective integration *in vivo*.

Materials and methods

DNA manipulations

Plasmid A (Fig. 3) was p5XCRE — pGEM with five CRE sites cloned into the polylinker site (a gift from M. Montminy and P. Brindle). Plasmid B was pBW18, a derivative of pUC19 in which the *amp* gene has been replaced with a *SupF* selectable marker. Plasmid pMM105, used as target in Figure 4, was constructed by inserting a synthetic oligonucleotide matching in sequence LAV bases 9605–9706 into pBSIISK$^+$ (Stratagene). This plasmid was chosen as target because it is smaller than phage ϕX174 DNA and, like ϕX174 DNA, it contains a single *PstI* site.

Integration assays

HIV-1 integrase was purified essentially as described [15]; λR–IN was purified as described [8]. Integration reactions in the presence of 5 mM $MnCl_2$ were carried out as described [12]. Reactions carried out in the presence of 5 mM $MgCl_2$ contained, in addition, 25 mM Hepes pH 7.5, 10 mM BME, 10 % glycerol, 0.1 mg ml^{-1} BSA, 13 ng μl^{-1} HIV-1 integrase, 0.1 ng μl^{-1} LTR oligonucleotide (FB79 [8]) 20 % DMSO and 10 U μl^{-1} RNAse T1. Plasmid target DNA was added at a concentration of 50 ng μl^{-1}. Reactions were stopped by addition of one-fourth volume 100 mM EDTA, 1 % SDS, in gel-loading buffer. Reaction products were separated by electrophoresis on 1 % agarose gels in TAE containing ethidium bromide. Integration products were visualized by autoradiography after drying the gels. PCR-based assays of integration site selection in phage λ DNA were carried out as described [8].

Purification and assay of preintegration complexes

Preintegration complexes were purified and analyzed as described in [24] and modified in [32]. In some cases, extracts containing preintegration complexes were further purified by gel filtration. Spin columns were prepared using about 2 ml BSA-coated Sephacryl S400 in Buffer K. Cytoplasmic extract (200 μl) in Buffer K plus 0.5 % Triton was loaded per column and columns were centrifuged for 3 min at 1000 rpm in a Sorval RT6000B centrifuge. This procedure removed about 95 % of the total protein from the extracts containing preintegration complexes as measured by the BCA Protein Assay (Pierce). Integration assays were performed as described by adding 0.5 μg target DNA to preintegration complexes in Buffer K followed by incubation at 37 °C. Integration products were deproteinized, separated on agarose gels, and visualized by Southern blotting as described [24,32]. Cytoplasmic extracts from uninfected cells were prepared exactly as were extracts from infected cells.

Acknowledgements: We thank members of the Salk Institute Infectious Disease Laboratory for suggestions and comments on the manuscript. This work was supported by grant R01 AI37489-01 from the National Institute of Allergy and Infectious Disease to F.D.B. and grant R94-SAL-104 from the University of California Universitywide AIDS Research Program to F.D.B. M.D.M. was supported by Fellowship 1 F32 AI09068-01 from the National Institutes of Allergy and Infectious Diseases. Y.C.B. was supported by grant AI 29850-03 to L. Orgel. F.D.B. is a Special Fellow of the Leukemia Society of America.

References

1. Vink C, Plasterk RHA: **The human immunodeficiency virus integrase protein.** *Trends Genet* 1993, **9**:433–437.
2. Vijaya S, Steffan DL, Robinson HL: **Acceptor sites for retroviral integrations map near DNaseI-hypersensitive sites in chromatin.** *J Virol* 1986, **60**:683–692.
3. Rohdewohld H, Weiher H, Reik W, Jaenisch R, Breindl M: **Retrovirus integration and chromatin structure: Moloney murine leukemia proviral integration sites map near DNase I-hypersensitive sites.** *J Virol* 1987, **61**:336.
4. Withers-Ward ES, Kitamura Y, Barnes JP, Coffin JM: **Distribution of targets for avian retrovirus DNA integration *in vivo*.** *Genes Dev* 1994, **8**:1473–1487.
5. Chalker DL, Sandmeyer SB: **Ty3 integrates within the region of RNA polymerase III transcription initiation.** *Genes Dev* 1992, **6**:117–128.
6. Kirchner J, Connolly CM, Sandmeyer SB: ***In vitro* position-specific integration of a retrovirus-like element requires Pol III transcription factors.** *Science* 1995, **267**:1488-1491.
7. Ji H, Moore DP, Blomberg MA, Braiterman LT, Voytas DF, Natsoulis G, *et al.*: **Hotspots for unselected Ty1 transposition events on yeast chromosome III are near tRNA genes and LTR sequences.** *Cell* 1993, **73**:1–20.
8. Bushman FD: **Tethering human immunodeficiency virus 1 integrase**

to a DNA site directs integration to nearby sequences. *Proc Natl Acad Sci USA* 1994, **91**:9233–9237.

9. Kalpana GV, Marmon S, Wang W, Crabtree GR, Goff SP: **Binding and stimulation of HIV-1 integrase by a human homolog of yeast transcription factor SNF5.** *Science* 1994, **266**:2002–2006.
10. Craigie R, Fujiwara T, Bushman F: **The IN protein of Moloney murine leukemia virus processes the viral DNA ends and accomplishes their integration *in vitro*.** *Cell* 1990, **62**:829-837.
11. Bushman FD, Fujiwara T, Craigie R: **Retroviral DNA integration directed by HIV integration protein *in vitro*.** *Science* 1990, **249**:1555–1558.
12. Bushman FD, Craigie R: **Activities of human immunodeficiency virus (HIV) *in vitro*: specific cleavage and integration of HIV DNA.** *Proc Natl Acad Sci USA* 1991, **88**:1339–1343.
13. Katz RA, Merkel G, Kulkosky J, Leis J, Skalka AM: **The avian retroviral IN protein is both necessary and sufficient for integrative recombination *in vitro*.** *Cell* 1990, **63**:87–95.
14. Katzman M, Katz RA, Skalka AM, Leis J: **The avian retroviral integration protein cleaves the terminal sequences of linear viral DNA at the *in vivo* sites of integration.** *J Virol* 1989, **63**:5319–5327.
15. Sherman PA, Fyfe JA: **Human immunodeficiency virus integration protein expressed in *Escherichia coli* possesses selective DNA cleaving activity.** *Proc Natl Acad Sci USA* 1990, **87**:5119–5123.
16. Vora AC, Fitzgerald ML, Grandgenett DP: **Removal of 3'-OH-terminal nucleotides from blunt-ended long terminal repeat termini by the avian retrovirus integration protein.** *J Virol* 1990, **64**:5656–5659.
17. Ellison V, Brown PO: **A stable complex between integrase and viral DNA ends mediates human immunodeficiency virus integration *in vitro*.** *Proc Natl Acad Sci USA* 1994, **91**:7316–7320.
18. Adzuma K, Mizuuchi K: **Interaction of proteins located at a distance along DNA: mechanism of target immunity in the Mu DNA strand-transfer reaction.** *Cell* 1989, **57**:41–47.
19. Winter RB, Berg OG, von Hippel PH: **Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. The *Escherichia coli* lac repressor-operator interaction: kinetic measurements and conclusions.** *Biochemistry* 1981, **20**:6961–6977.
20. Pryciak PM, Varmus HE: **Nucleosomes, DNA-binding proteins, and DNA sequence modulate retroviral integration target site selection.** *Cell* 1992, **69**:769–780.
21. Kitamura Y, Lee YM, Coffin JM: **Non-random integration of retroviral DNA *in vitro*: effect of CpG methylation.** *Proc Natl Acad Sci USA* 1992, **89**:5532–5536.
22. Brown PO, Bowerman B, Varmus HE, Bishop JM: **Correct integration of retroviral DNA *in vitro*.** *Cell* 1987, **49**:347–356.
23. Ellison VH, Abrams H, Roe T, Lifson J, Brown PO: **Human immunodeficiency virus integration in a cell-free system.** *J Virol* 1990, **64**:2711–2715.
24. Farnet CM, Haseltine WA: **Integration of human immunodeficiency virus type 1 DNA *in vitro*.** *Proc Natl Acad Sci USA* 1990, **87**:4164–4168.
25. Jack WE, Terry BJ, Modrich P: **Involvement of outside DNA sequences in the major kinetic path by which *Eco*R1 endonuclease locates and leaves its recognition sequence.** *Proc Natl Acad Sci USA* 1982, **79**:4010–4014.
26. Kim JG, Takeda Y, Matthews BW, Anderson WF: **Kinetic studies of Cro repressor-operator DNA interaction.** *J Mol Biol* 1987, **196**: 149–158.
27. Leavitt AD, Rose RB, Varmus HE: **Both substrate and target oligonucleotide sequences affect *in vitro* integration mediated by human immunodeficiency virus type 1 integrase protein produced in *Saccharomyces cerevisiae*.** *J Virol* 1992, **66**:2359–2368.
28. Pruss D, Reeves R, Bushman FD, Wolffe AP: **The influence of DNA and nucleosome structure on integration events directed by HIV integrase.** *J Biol Chem* 1994, **269**:25031–25041.
29. Muller H-P, Varmus HE: **DNA bending creates favored sites for retroviral integration: an explanation for preferred insertion sites in nucleosomes.** *EMBO J* 1994, **13**:4704–4714.
30. Taylor ICA, Workman JL, Schuetz TJ, Kingston RE: **Facilitated binding of GAL4 and heat shock factor to nucleosomal templates: differential function of DNA-binding domains.** *Genes Dev* 1991, **5**:1285.
31. Bushman FD: **Targeting retroviral integration.** *Science* 1995, **267**:1443–1444.
32. Miller MD, Wang B, Bushman FD: **Human immunodeficiency virus type 1 preintegration complexes containing discontinuous plus strands are competent to integrate *in vitro*.** *J Virol* 1995, **69**:3938–3944.

Received: **1 June 1995**; revised: **28 June 1995**.
Accepted: **25 July 1995**.