4454–4461 Nucleic Acids Research, 1994, Vol. 22, No. 21

Efficient concerted integration of retrovirus-like DNA *in vitro* by avian myeloblastosis virus integrase

Ajaykumar C.Vora, Mark McCord, Michael L.Fitzgerald+, Ross B.Inman[1] and Duane P.Grandgenett*

Institute for Molecular Virology, St Louis University Health Sciences Center, 3681 Park Avenue, St Louis, MO 63110 and [1]Institute for Molecular Virology, University of Madison, Madison, WI 53706, USA

Received July 1, 1994; Revised and Accepted September 6, 1994

ABSTRACT

We report the efficient concerted integration of a linear virus-like DNA donor into a 2.8 kbp circular DNA target by integrase (IN) purified from avian myeloblastosis virus. The donor was 528 bp, contained recessed 3′ OH ends, was 5′ end labeled, and had a unique restriction site not found in the target. Analysis of concerted (full-site) and half-site integration events was accomplished by restriction enzyme analysis and agarose gel electrophoresis. The donor also contained the *SupF* gene that was used for genetic selection of individual full-site recombinants to determine the host duplication size. Two different pathways, involving either one donor or two donor molecules, were used to produce full-site recombinants. About 90% of the full-site recombinants were the result of using two donor molecules per target. These results imply that juxtapositioning an end from each of two donors by IN was more efficient than the juxtapositioning of two ends of a single donor for the full-site reaction. The formation of preintegration complexes containing integrase and donor on ice prior to the addition of target enhanced the full-site reaction. After a 30 min reaction at 37°C, ~20–25% of all donor/target recombinants were the result of concerted integration events. The efficient production of full-site recombinants required Mg^{2+}; Mn^{2+} was only efficient for the production of half-site recombinants. We suggest that these preintegration complexes can be used to investigate the relationships between the 3′ OH trimming and strand transfer reactions.

INTRODUCTION

Upon infection of cells by retroviruses, a large-size viral nucleoprotein complex is observed in the cytoplasm. These complexes vary in size from 160S for murine leukemia virus (1,2) to 80S for human immunodeficiency virus (HIV) type-1 (3). The nucleoprotein complexes contain newly synthesized blunt-ended linear viral DNA that is subsequently trimmed by two nucleotides at its 3′ OH terminus by the viral integrase (IN) (4). In these complexes, IN can catalyze the concerted integration of the viral DNA termini into exogenous DNA targets (full-site reaction) mimicking the *in vivo* reaction (1).

Similar trimming and integration reactions can also be catalyzed by purified IN *in vitro*. IN derived from bacterial expression systems or purified from avian myeloblastosis virus (AMV) can trim two nucleotides from the termini of oligonucleotides or DNA restriction fragments containing viral long terminal repeat (LTR) sequences (5–8). The recessed substrates can then be integrated into other DNA targets by IN (6,8,9); the majority of the observed integration events involve the insertion of a single LTR terminus into one strand of the target (half-site reaction). IN from recombinant sources requires Mn^{2+} for efficient catalysis of either the trimming or the strand transfer reactions (5,6,8,9). AMV IN purified from virions can effectively use Mg^{2+} or Mn^{2+} for these reactions (7,9,10). To date, recombinant IN is not capable of efficiently performing the concerted insertion of virus-like substrates into DNA using either divalent cation (6,11,12).

We wanted to better understand the conditions necessary to reconstitute a preintegration complex, containing purified AMV IN and donor, that efficiently performed the concerted integration reaction. AMV IN is capable of the full-site integration reaction using a linear 3.4 kbp virus-like substrate (R35) as donor and λ gtWes as target (10). The reaction required Mg^{2+}, and ~0.25% of the donor substrate was inserted into the target in a concerted fashion. The data demonstrated that AMV IN is capable of catalyzing the full-site reaction, but conditions for formation of preintegration complexes capable of efficient concerted integration were not optimal.

To further optimize conditions for forming IN–donor preintegration complexes capable of concerted integration, the R35 donor and the target were changed. A transient circle is an apparent necessity for concerted integration as well as for

*To whom correspondence should be addressed

+Present address: Colorado State University, Fort Collins, CO 80523, USA

trimming of the viral genome by IN *in vivo* (13). Thus, the size of the donor was reduced to 528 bp (termed M-2) (Fig. 1A) to increase its cyclization probability (*j* factor) (14). AMV IN was also shown to form DNA looped structures in the 3.4 kbp donor, mapping predominately to A/T-rich regions (15). The ability of IN to form non-specific DNA looped structures may hinder the formation of preintegration complexes capable of the full-site reaction. M-2 lacks A/T-rich regions to overcome this problem. The donor contained a unique restriction site (Fig. 1A) not found in the target molecule, thereby permitting isolation and analysis of half-site and full-site integration recombinants by agarose gel electrophoresis. M-2 also contained the *SupF* gene that was used for genetic isolation of individual full-site recombinants to characterize the target site. The target was changed to a small circular DNA (Fig. 1A) to provide an easier analysis of IN–donor complex interactions with target DNA.

The full-site integration reaction can occur via two pathways using the above substrates (Fig.1). First, the termini of a single M-2 molecule, held together by IN to form a transient circle, can be inserted by IN into circular pGEM (Fig. 1C). Second, one terminus each of two individual M-2 molecules, juxtapositioned together by IN, can be inserted in a concerted fashion into pGEM (Fig. 1D). Concerted insertion of M-2 into pGEM accounts for 20–25% of the total quantity of M-2–pGEM recombinants formed, and the formation of IN–M-2 preintegration complexes prior to strand transfer enhances this reaction. The half-site integration reaction involving the insertion of a single M-2 terminus by IN into pGEM can also occur (Fig. 1B).

MATERIALS AND METHODS

Bacteria and plasmids

R35 (10) was the source of the 65 bp LTR insert and the *SupF* gene (469 bp). *Escherichia coli* (strain CA244)(16), that had amber mutations in the *lacZ* gene and for tryptophan biosynthesis, was obtained from the *E.coli* Genetic Stock Center at Yale University. CA244 cells transformed by M-2–pGEM recombinants or R-35 were grown at 37°C in M9 medium that contained autoclaved casamino acids (2 g/l), lactose (0.2%), 0.5 mM isopropyl-β-thio-galactopyranoside, 40 μg/ml 5-bromo-4-chloro-3-indolyl-β-D-galactoside, and ampicillin (30 μg/ml).

Construction of M-2

The 65 bp LTR circle junction was recovered from R35 by *Eco*RI digestion. The isolated fragment was ligated, digested by *Nde*I, and cloned into the *Nde*I site of pBR322 lacking its *Eco*RI site. A plasmid was selected that contained a single U3/U5 insert having *Nde*I termini and an internal *Eco*RI site. This construct was digested with *Eco*RI and the *SupF* gene was cloned into this site by blunt-end ligation. Digestion of the construct by *Nde*I yielded a 528 bp restriction fragment that contained U3 and U5 LTR termini and the internal *SupF* gene (Fig. 1). M-2 contained unique *Bgl*II and *Xba*I sites. M-2 was subsequently cloned into the *Nde*I site of the high copy-number plasmid pUC19 (New England Biolabs) for large scale isolation. The *Xba*I site located near the U3 terminus of M-2 was used to create a donor molecule that lacked its U3 terminus, but was similar in size to M-2.

Labeling of M-2

Supercoiled pUC19 containing M-2 and pGEM (Promega) were isolated by velocity sedimentation on sucrose gradients. M-2 was released from pUC19 by *Nde*I digestion and isolated from low-melt agarose gels. M-2 was dephosphorlyated and 5′ end labeled using [γ-^{32}P]ATP and T4 polynucleotide kinase. The specific activity of M-2 was generally 2000 to 5000 c.p.m./ng. The pUC19 plasmid containing M-2 was also digested by *Xba*I and *Nde*I, releasing M-2 lacking its U3 terminus (Fig. 1A); the 490 bp fragment was labeled at both of its 5′ ends as described above. *Xba*I digestion removed the first two nucleotides of the adjacent *Bgl*II site.

Assay conditions

The standard reaction mixture contained 20 mM Tris–hydrochloride (pH 7.5), 10 mM dithiothreitol, 5 mM $MgCl_2$, 10% dimethyl sulfoxide, 0.05% Nonidet P-40, 5% polyethylene glycol, and 200 mM NaCl (10). The preincubation step included mixing of M-2 with IN at 0°C for 10 min in 20 μl aliquots; preincubation at 10°C or 20°C gave similar results. The molar ratio of dimeric IN to M-2 was usually set at 12:1 (33 ng or 25 nM to 15 ng, respectively). The reaction was initiated by addition of supercoiled pGEM followed by immediate incubation at 37°C. The concentrations of labeled M-2 and pGEM were 15 ng and 100 ng, respectively. The molar ratio of M-2 to pGEM was 1. Scale-up reactions maintained the same concentrations of enyzme and DNA substrates for isolation of various M-2–pGEM recombinants. AMV IN was purified to near homogeneity (7).

Analysis of M-2–pGEM recombinants

The reactions were stopped with sodium dodecyl sulfate and proteinase K at final concentrations of 1% and 1 mg/ml, respectively. The samples were incubated at 37°C for 2 h, subjected to phenol–chloroform (1:1) and ether extractions, and precipitated by ethanol. Aliquots were subjected to electrophoresis on 1 or 1.5% agarose gels that were dried and exposed to X-ray films. The products were quantitated with a Molecular Dynamics PhosphorImager. For *Bgl*II restriction analysis, each 20 μl reaction was digested by 12 units for 2 h. The samples were subjected to electrophoresis on 1.5% agarose gels in a Tris–borate–EDTA buffer with 0.5 μg/ml of ethidium bromide for 13 h at 100 V. Linear DNA fragments (Boehringer Mannheim), pGEM, and R35 were used as molecular weight markers. Linear R35 was also 5′ end labeled with γ-^{32}P and was used to identify linear 3.4 kbp DNA obtained by *Bgl*II digestion of M-2–pGEM recombinants.

Characterization of M-2–pGEM recombinants resulting from concerted integration events

The 3.4 kbp linear DNA obtained by *Bgl*II digestion of all M-2–pGEM recombinants was isolated from scale-up reactions. The digested products were subjected to electrophoresis on 1.5% agarose and the wet gel was exposed to X-ray film. The desired DNA was excised, electroeluted, and was purified by Wizard PCR Prep columns (Promega). The DNA was ligated and transformed into either *E.coli* HB101, Epicuran Coli Sure (Stratagene) or CA244 cells. Colonies were screened for plasmids that were analyzed by size, restriction enzyme digestions, and dideoxy DNA sequencing. Primers for sequencing were homologous within M-2 near both the U3 and U5 termini and were used to sequence the donor–target junctions. The above 3.4 kbp DNA and other M-2–pGEM recombinants were examined by electron microscopy to determine their sizes and structures (15).

RESULTS

Parameters for production and for physical quantitation of concerted integration events

A simple approach to investigate concerted integration of a donor into a target by IN is shown in Figure 1. Linear M-2 has a recessed 3′ CA moiety at its U3 and U5 termini and has a unique *Bgl*II restriction site (Fig. 1A). Concerted integration of one M-2 molecule into circular pGEM-3 (2867 bp) by AMV IN will result in the formation of a nicked 3.4 kbp circle that, when digested by *Bgl*II, produces a 3.4 kbp linear molecule (Fig. 1C). Half-site integration of one M-2 donor into pGEM will result in a stick and ball structure also 3.4 kbp in size (Fig. 1B); *Bgl*II digestion of half-site recombinants will result in the formation of two different stick and ball structures ('A' or 'B'), depending on whether the U3 or U5 terminus was inserted. Digestion of M-2 by *Xba*I (Fig. 1A) will produce a donor capable of making U5 half-site recombinants with pGEM (similar to stick and ball structure 'A') (Fig. 1B) whose length is 3.36 kbp. Lastly, concerted integration can also result from a single insertion event involving two individual M-2 molecules into circular pGEM (Fig. 1D), producing a linear 3.92 kbp recombinant. *Bgl*II digestion of this linear recombinant produces four DNA products whose ratios also depend upon the use of U3 or U5 termini (Fig. 1D). The integration of 5′ end labeled M-2 into circular pGEM can be followed by restriction enzyme digestion and agarose gel electrophoresis (Fig. 2).

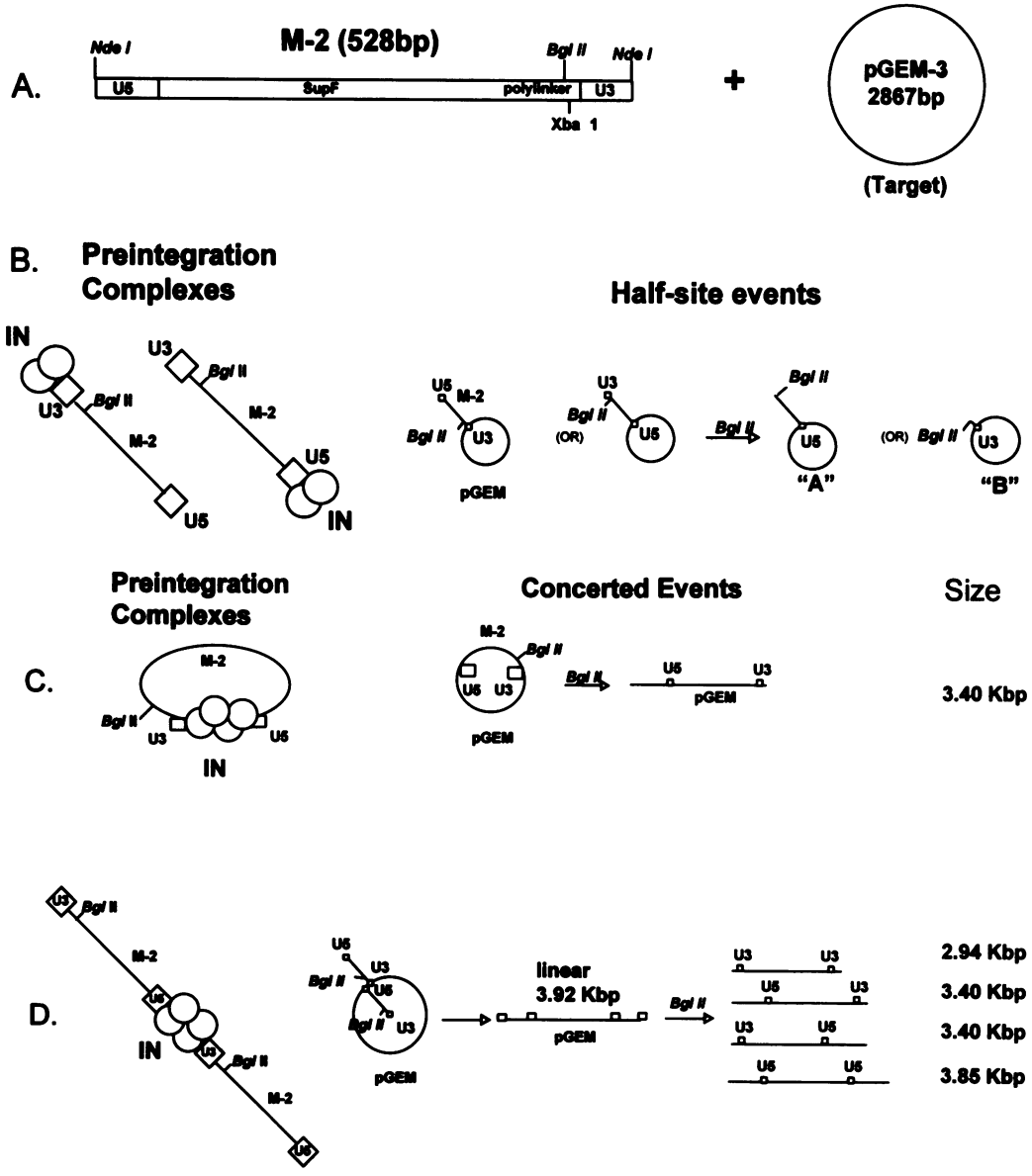

Figure 1. DNA structures and pathways involved in formation of M-2–pGEM recombinants. (**A**) The 528 bp M-2 donor possesses the terminal 24 bp of the U3 LTR and 36 bp of the terminal U5 LTR, flanked by a *Nde*I site. The *SupF* gene is located between the LTR termini. The target is circular pGEM. (**B**) Dimeric AMV IN is shown associated on one LTR terminus of M-2 in a preintegration complex. Half-site insertion of M-2 into pGEM occurs in two orientations; depending upon this orientation, *Bgl*II digestion produces two different ball and stick molecules ('A' and 'B'). (**C**) An IN tetramer juxtapositioning the LTR termini of a single M-2 donor. The concerted integration (full-site) of a single M-2 molecule into circular pGEM is depicted. *Bgl*II digestion of this recombinant produces a linear 3.4 kbp DNA. (**D**) A tetramer of IN holds together two individual M-2 donors in a preintegration complex. The concerted integration of two separate M-2 molecules into pGEM produces a linear 3.92 kbp DNA. Four possible recombinant structures are possible. *Bgl*II digestion results in the formation of four possible linear structures, with the two 3.4 kbp structures comigrating.

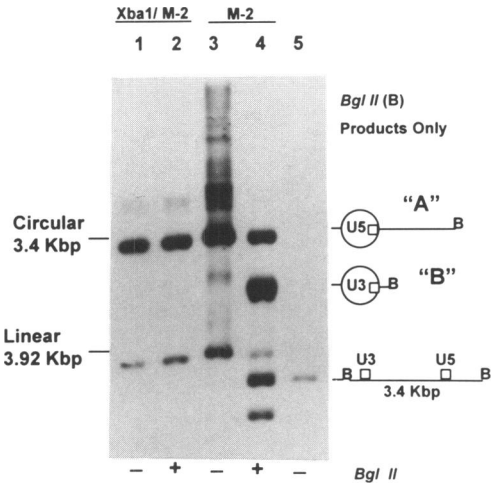

Figure 2. Formation of IN–M-2 preintegration complexes produces concerted M-2–pGEM recombinants. IN and M-2 or *Xba*l-truncated M-2 that lacks its U3 terminus were preincubated together on ice under standard reaction conditions. The target was added just prior to incubation at 37°C for 20 min. The reactions were processed and some were digested with *Bgl*II (see bottom). The samples were subjected to electrophoresis on 1.5% agarose gels. The dried gel was exposed for 22 h with a screen. Lane 5 contains a labeled 3.4 kbp DNA marker. The right side shows *Bgl*II digestion products and the left side shows the size of M-2–pGEM recombinants.

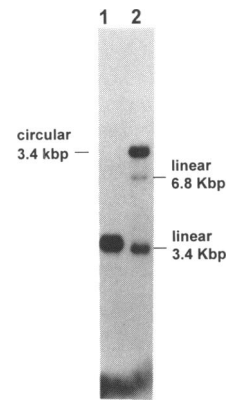

Figure 3. Characterization of linear 3.4 kbp DNA obtained by *Bgl*II digestion of M-2–pGEM recombinants. *Bgl*II digestion of all M-2–pGEM recombinants was performed as shown in Figure 2. Scale-up reactions were used to isolated the linear 3.4 kbp as described in Materials and Methods. Two thirds of the isolated DNA was ligated. The samples were subjected to 1.5% agarose gel electrophoresis. Lane 1, one third of the unligated DNA; lane 2, one half of the ligated DNA. The other half of the ligated DNA was used to transform CA244 cells. Linear and circular DNA size markers were used in separate lanes to verify the size of the labeled products.

IN–M-2 preintegration complexes and structural characterization of M-2–pGEM recombinants

We tested conditions shown to be conducive for catalyzing the concerted integration of a linear 3.4 kbp donor into λ gtWES (10). IN and M-2 were preincubated together on ice for 10 min with Mg^{2+} prior to the addition of pGEM and incubation at 37°C (Fig. 2). After 20 min, the samples were subjected to gel electrophoresis on 1.5% agarose for 13 h (Fig. 2, lane 3). Four major labeled M-2–pGEM products were produced. One species comigrated with a nicked circle 3.4 kbp marker and another more heterogeneous group migrated slightly slower. Electron microscopy of the isolated DNA established that the majority of the 3.4 kbp M-2–pGEM recombinants were like the ball and single stick model ('A', Fig. 1B), and the slower migrating group of recombinants were circular pGEM with two or more independent M-2 inserts (data not shown). These structural data were confirmed using *Xba*I-digested M-2 that lacked its U3 terminus (Fig. 1A). Single insertion events of *Xba*I/M-2, containing only a U5 terminus, into pGEM migrating with a mass of 3.36 kbp were observed by gel electrophoresis (Fig. 2, lanes 1 and 2) or by electron microscopy (data not shown).

Two other M-2–pGEM recombinants migrating faster than nicked circular 3.4 kbp were observed (Fig. 2, lane 3); one migrating slightly faster that the 3.4 kbp recombinant whose structure has not yet been established and the fastest migrating species, a linear 3.92 kbp recombinant involving two M-2 molecules (Fig. 1D). A similar size linear product (Fig. 2, lane 1) was produced using the *Xba*I-cleaved M-2 donor that lacks its U3 terminus. We will address the linear 3.92 kbp recombinant later. Donor–donor recombinants, or M-2–M-2 molecules, were not visualized because these products migrated off the gel (see Fig. 5).

It was not possible to differentiate which products resulted from half-site or full-site reactions (Fig. 2, lane 3) without *Bgl*II digestion. Digestion of the M-2–pGEM recombinants by *Bgl*II results in the cleavage of all *Bgl*II sites regardless of whether they are half-site or full-site reactions. With all reactions involving only a single M-2 molecule, one half of the radioactivity will be lost with half-site recombinants upon *Bgl*II digestion while no radioactivity will be lost by *Bgl*II digestion of single M-2 full-site recombinants, because no 5′ end labeled termini are released (Fig. 1B and C). One-half of the radioactivity will be released from linear 3.92 kbp recombinants, that involve two M-2 molecules, upon *Bgl*II digestion (Fig. 1D).

Five major products were produced by *Bgl*II digestion of all M-2–pGEM recombinants (Fig. 2, lane 4). The slowest moving DNA comigrated with nicked circular 3.4 kbp DNA ('A' product) and the next slowest moving DNA ('B' product) nearly comigrated with nicked circular pGEM (2.86 kbp). The 'A' and 'B' products derived by *Bgl*II digestion of M-2–pGEM recombinants were the result of half-site U5 and U3 insertions, respectively. As expected, *Bgl*II digestion of *Xba*I/M-2 recombinants (Fig. 2, lane 2) did not alter the migration of these recombinants nor identify any species containing U3 insertions.

*Bgl*II digestion of M-2–pGEM recombinants produced by AMV IN confirmed earlier data on the preference of U3 over U5 termini for both the trimming and the half-site and full-site strand transfer reactions by a 2 to 1 margin (10,15,17). This same U3 over U5 preference was observed by phosphorimager analysis of *Bgl*II digests of M-2–pGEM recombinants (see products 'A' and 'B') (Fig. 2, lane 4). The total quantities of *Bgl*II-digested DNA 'A' and 'B' were 23 and 45%, respectively.

The other cluster of M-2–pGEM recombinants produced by *Bgl*II digestion were linear molecules of 2.94, 3.4, and 3.85 kbp in length (Fig. 2, lane 4, Fig. 1D). Quantitation showed that these recombinants represented 11, 17, and 4%, respectively, of the total amount of labeled digested DNA.

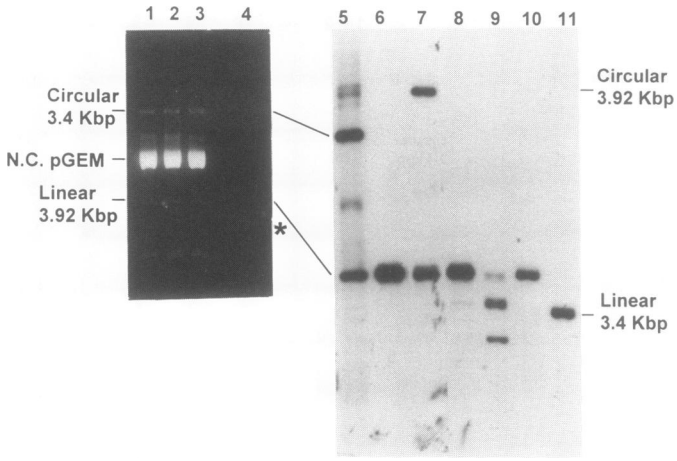

Figure 4. Characterization of linear 3.92 kbp DNA. Ethidium bromide staining of a scale-up M-2−pGEM integration reaction. Standard integration reaction conditions were used. Lanes 1−3 represent a total of thirty 20 μl reactions. The circular 3.4 kbp and the linear 3.92 kbp DNAs were the major labeled products (see lines). Lane 4 contained a 3.4 kbp DNA marker and its location is marked by an asterisk. The lightly stained DNA in lanes 1−3 (below the 3.4 kbp marker) was present in our sucrose gradient purified pGEM preparations. Restriction analysis of ligated and unligated 3.92 kbp DNA. Lane 5, standard reaction; lane 6, purified 3.92 kbp DNA; lane 7, ligated 3.92 kbp DNA; lane 8, *NdeI* digestion of ligated 3.92 kbp DNA; lane 9, *BglII* digestion of unligated 3.92 kbp DNA; lane 10, *BglI* digestion of ligated 3.92 kbp DNA. Lane 11 contained a 3.4 kbp linear DNA marker.

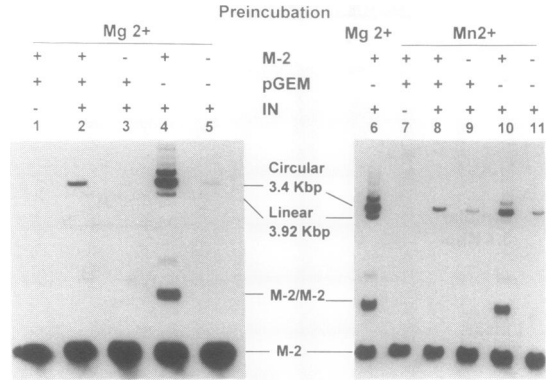

Figure 5. Preincubation conditions and divalent metal ion affect production of M-2−pGEM recombinants. Standard reaction conditions were employed throughout except for the presence or absence of IN, M-2, and pGEM in the preincubation mixture. Following preincubation on ice, the reactions were initiated by the addition of the missing reagent or reagents. Incubation was for only 10 min at 37°C. Lanes 1−6 and lanes 7−11 contained Mg^{2+} (5 mM) and Mn^{2+} (1 mM), respectively. Experiments in lanes 1−5 and lanes 6−11 were performed independently. The middle line indicates which reagents were present in the preincubation mixture, noted with + or − signs, and the various recombinants produced.

Physical and genetic analysis of concerted M-2−pGEM recombinants

We wanted to establish that the linear 3.4 kbp M-2−pGEM recombinants obtained by *BglII* digestion were the result of full-site integration reactions. Several 30 min scale-up reactions were performed and the samples were subjected to *BglII* digestion. The linear 3.4 kbp DNA was identified on wet 1.5% agarose gels by autoradiography and was purified (Fig. 3, lane 1). The 3.4 kbp DNA was ligated and then analyzed again by 1.5% agarose gel electrophoresis (lane 2). The ligated DNA comigrated with nicked circular 3.4 kbp DNA. In three separate experiments, the maximum amount of the circular 3.4 kbp product produced by ligation was 75%, suggesting that some of the DNA ends were damaged upon elution and purification of the linear 3.4 kbp DNA. The 6.8 kbp band probably represents head-to-head ligation of two linear 3.4 kbp molecules.

To further establish the physical structure of the recombinant DNA that was present in the 3.4 kbp band produced by *BglII* digestion, the DNA was analyzed by electron microscopy. Counting 200 molecules, 70% of the molecules were linear structures of 3.2 kbp in length, 1% were branched structures of the same size, 15% were linear molecules of various smaller sizes and 15% were circles of 2.86 kbp in length. The linear molecules of various sizes probably represent degraded 3.4 kbp DNA. The 2.86 kbp circular DNA is pGEM that was not used as a target. Greater than 95% of the target DNA is not used in 30 min reactions as observed by ethidium bromide staining of DNA (see Fig. 4). The contaminating pGEM that comigrated with the labeled 3.4 kbp linear DNA represented a very minor population of topoisomer forms of pGEM induced by the presence of ethidium bromide in the 1.5% agarose gels.

The linear 3.4 kbp recombinant DNA represented concerted integration events. We established, by the genetic selection of individual recombinants, that the recombinants contained the correct host duplication size that is observed upon concerted integration *in vivo*. The ligated 3.4 kbp DNA obtained by *BglII* digestion (Fig. 3) was transformed into *E.coli* (HB101, Epicuran Coli Sure, or CA244 cells). Screening of individual colonies demonstrated that 5−10% of the plasmids were 3.4 kbp in length when either HB101 or Sure cells were used as hosts. Fill-in and repair of the ligated 3.4 kbp DNA by *E.coli* DNA polymerase I before transformation did not significantly increase the percentage of rescued 3.4 kbp plasmids. The rest of the plasmids were 2.86 kbp in length. Restriction enzyme analysis of the plasmids demonstrated that the 3.4 kbp DNA products were concerted integration events and the 2.86 kbp DNA was pGEM. DNA sequence analysis of 12 recombinants of 3.4 kbp in length demonstrated that the majority of them had the correct avian 6 bp duplication size (ten 6, one 5, and one 7 bp) (10,18).

To eliminate the pGEM background in the genetic assay, we transformed the ligated 3.4 kbp DNA (Fig. 3) into CA244 cells which have amber mutations in the *lacZ* gene and for tryptophan biosynthesis (16). M-2 contains the *SupF* gene that, when inserted into pGEM in a concerted manner, will permit the replication of this recombinant plasmid. Without the *SupF* gene, pGEM will not replicate in CA244 cells. Transformation of CA244 resulted in the production of only blue colonies (data not shown), and all of the rescued plasmids were 3.4 kbp in length. DNA sequence analysis of ten 3.4 kbp plasmids verified that the recombinants were the result of concerted integration events.

Defining the origins of the linear 3.4 kbp M-2−pGEM DNA produced by *BglII* digestion

As indicated in Figure 1, concerted integration of M-2 into pGEM could occur by two pathways, one involving one donor molecule producing a 3.4 kbp nicked circle (Fig. 1C), the second involving two donor molecules producing a 3.92 kbp linear molecule (Fig.

1D). We show that AMV IN preferentially used the latter pathway to produce full-site recombinants. Scale-up reactions were performed to isolate sufficient quantities of labeled 3.92 kbp DNA for analysis (Fig. 4, lanes 1–3). Ethidium bromide staining of this scale-up reaction showed that the majority of pGEM (supercoiled pGEM migrated off the gel) was not used as a target in a 30 min reaction. The isolated 3.92 kbp DNA (Fig. 4, lane 6) was subjected to restriction enzyme analysis, and also to ligation followed again by restriction enzyme analysis. *Bgl*II digestion of the linear 3.92 kbp DNA resulted in the predicted three linear structures (Fig. 4, lane 9) as well as minor labeled products just below and above the 3.85 kbp and the 3.4 kbp products, respectively (see below for electron microscopy studies). Ligation of the 3.92 kbp DNA resulted in the formation of a nicked circle migrating with the size of ~3.92 kbp (lane 7). Although only 50% of the DNA in this experiment was ligated, as determined by the phosphorimager, digestion of the ligated DNA by *Nde*I resulted in the formation of only linear 3.92 kbp DNA (lane 8) consistent with its predicted structure (Fig. 1D).

To show that the linear 3.92 kbp DNA resulted from concerted integration events, individual recombinants were isolated. Transformation of the ligated 3.92 kbp DNA (Fig. 4, lane 7) into CA244 cells and subsequent DNA sequence analysis of six plasmids 3.92 kbp in size demonstrated that these M-2–pGEM recombinants were the result of concerted integration events. In summary, DNA sequence analysis of 28 M-2–pGEM recombinants derived from various genetic assays showed that they contained either 5, 6, or 7 bp host duplication sizes that are commonly observed (9,10,18). One exception involving a 2 bp duplication was observed.

To analyze the general locations where M-2 integrated into pGEM, it was necessary to digest the target instead of the donor in the 3.92 kbp recombinants. *Bgl*I digestion (M-2 contains no *Bgl*I site but pGEM contains one site) of the ligated 3.92 kbp DNA (Fig. 4, lane 7) produced linear 3.92 kbp DNA and a series of products equal to or shorter than linear 3.4 kbp DNA (Fig. 4, lane 10). Apparently, *Bgl*I digestion of the ligated 3.92 kbp DNA resulted in the quantitative transfer of this DNA to the linear 3.92 kbp product; the unligated 3.92 kbp DNA remaining in this sample (lane 7) was digested by *Bgl*I to products of sizes of ~3.4 kbp and less (lane 10). Production of these smaller size products by *Bgl*I showed that integration into pGEM was nonspecific. A longer exposure of this gel showed darker bands suggesting nonrandom integration (10,18,19).

We wanted to physically establish what DNA structures were present in the linear 3.92 kbp product. Electron microscopy of linear 3.92 kbp DNA produced after a 30 min reaction, as shown in Figure 4 (lanes 1–3), confirmed that the size of the linear recombinant DNA was correct. Electron microscopy (count of 129 molecules) of the purified DNA (Fig. 4, lane 6) showed that 66% were linear 3.8 kbp molecules, 10% were branched structures (linear 2.8 kbp with a 528 bp arm), 11% were a mixture of shorter linear molecules, and 12% were circles of 2.8 kbp in length. A short exposure of the gel (Fig. 4) also showed that not all of the labeled products associated with the 3.92 kbp products were linear DNA. As shown in Figure 4 (lane 7), ~10% of the labeled products migrated slightly slower than the linear 3.92 kbp band, suggesting the presence of slower migrating branched structures (15). The branched molecules were the result of half-site reactions involving a single M-2 donor and pGEM. These branched half-site reaction products were apparently due to breakage of the DNA during the DNA processing steps. Fewer branched molecules were observed after only a 10 min reaction (Fig. 6A) in contrast to the above 30 min reaction (Fig. 4). In summary, these studies confirmed that the majority of the labeled DNA migrating at the 3.92 kbp position was linear DNA of the correct size.

Circular 3.4 kbp recombinants that resulted from the concerted integration of one M-2 donor into pGEM can also give rise to linear 3.4 kbp DNA after *Bgl*II digestion (Fig. 1C). Digestion of purified circular 3.4 kbp resulted in the production of 'A' and 'B' products (Fig. 1B), and linear 3.4 kbp DNA. However, the linear DNA represented only 4% of the total radioactivity of the original undigested circular 3.4 kbp product (data not shown). Calculations showed that ~10% of the linear 3.4 kbp DNA produced by *Bgl*II digestion of a total M-2–pGEM reaction was derived from the concerted insertion of a single M-2 molecule into pGEM. The rest of the linear 3.4 kbp DNA resulted from the *Bgl*II digestion of the linear 3.92 kbp DNA (see Fig. 7 also).

Recognition of donor and target molecules by IN

We wanted to investigate how the donor and target molecules affected the production of concerted integration products. The approach was based on the presence or absence of M-2, pGEM, or IN in the preincubation mixture. After preincubation, the strand transfer reaction was limited to 10 min at 37°C to define only the early catalytic events. The influence of Mg^{2+} and Mn^{2+} in the preincubation mixtures on the formation of various recombinants was also examined.

Assay conditions with Mg^{2+} present included the mixing of IN/M-2 on ice prior to strand transfer (Fig. 5, lanes 4 and 6). The samples were subjected to gel electrophoresis on 1% instead of 1.5% agarose in order to visualize the M-2 donor, M-2–M-2 recombinants as well as all the M-2–pGEM recombinants. The preincubation of M-2 with IN produced a significant quantity of recombinants containing only M-2 molecules. The DNA migrating immediately above the M-2–M-2 recombinants was presumably M-2–M-2 structures with an additional integrated M-2 donor (lanes 4 and 6). The inclusion of pGEM with M-2 in the preincubation mixture before the addition of IN prevented the formation of these M-2–M-2 recombinants (Fig. 5, lane 2). Under these preincubation conditions, the same M-2–pGEM recombinants produced in the absence of pGEM in the preincubation mixture were again observed, but at only 20% of the previous amount (compare lane 2 to lane 4). The preincubation of IN with pGEM (lane 3) or IN alone (lane 5) prior to strand transfer similarly prevented the formation of M-2–M-2 or M-2–pGEM recombinants.

The formation of IN–M-2 preintegration complexes on ice does not absolutely require the presence of Mg^{2+} in the preincubation mixture. The efficiency for forming all M-2–pGEM or M-2–M-2 recombinants was reduced ~20% in the absence of Mg^{2+} in comparison to having Mg^{2+} present in the preincubation mixture (data not shown).

When IN–M-2 complexes were preincubated on ice in the presence of Mn^{2+}, the efficiency for producing M-2–M-2 and 3.4 kbp circular recombinants was not significantly reduced (14%) in comparison to the Mg^{2+} reaction (Fig. 5, compare lane 10 to lane 6). The other preincubation conditions with Mn^{2+} (lanes 8, 9, and 11) also produced most of the same products as those obtained with Mg^{2+} (lanes 2, 3, and 5). But, with Mn^{2+}, the production of the linear 3.92 kbp DNA was almost eliminated under all of the preincubation conditions (lanes

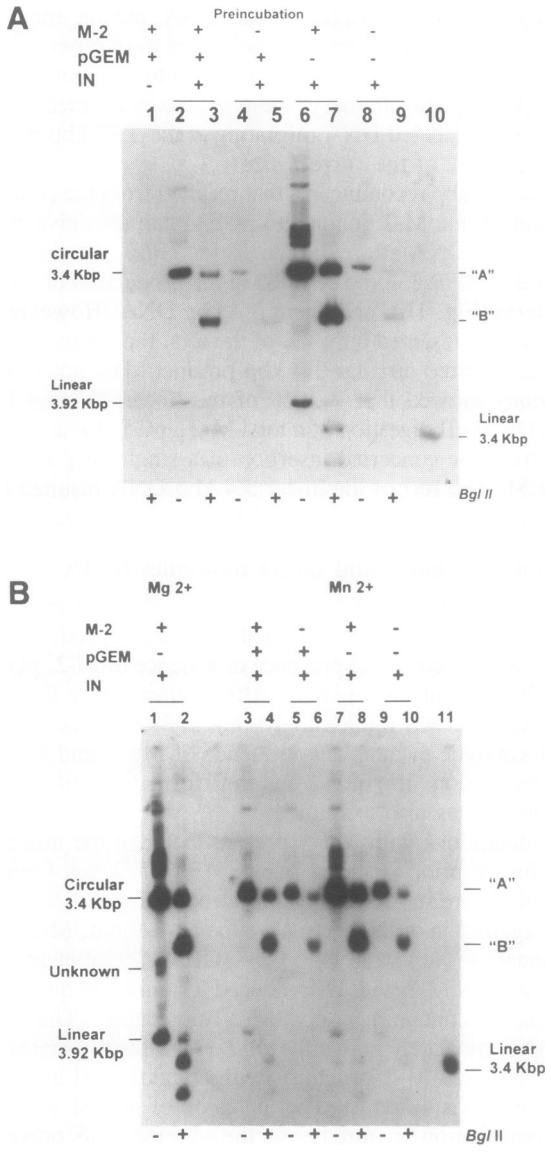

Figure 6. Preincubation of donor or target with IN and its influence on the concerted integration reaction. (**A**) Standard assay conditions in the presence of Mg^{2+} were employed. The samples shown in Figure 5 (lanes 1–5) were grouped into pairs with one half of the sample digested with *Bgl*II, followed by electrophoresis on 1.5% agarose gels. The left side of the Figure designates undigested products while the right side identifies digested products. Lane 1 contained no IN with the other lanes identifying what reagents were present in the preincubation mixtures. Lane 10 contained a linear 3.4 kbp marker. The bottom line indicates which samples were digested with *Bgl*II. (**B**) Standard reaction conditions were used except that 1 mM Mn^{2+} was present. The samples shown in Figure 5 (lanes 6–11) were grouped into pairs with one half of the samples digested with *Bgl*II, followed by electrophoresis on 1.5% agarose gels. The left side of the Figure identifies undigested products while the right side designates digested products. Lanes 1 and 2 contained Mg^{2+} instead of Mn^{2+} to serve as an identical control to lanes 7 and 8. The top lines indicate which reagents were present in the preincubation mixtures. Lane 11 contained a linear 3.4 kbp marker. The bottom line identifies which samples were digested with *Bgl*II. The unknown DNA was also marked.

8–11). These data suggest that Mn^{2+} affects the ability of IN to efficiently catalyze the concerted integration reaction.

To examine what affect preincubation conditions had on formation of specific M-2–pGEM recombinants, the reaction samples shown in Fig. 5 were digested with *Bgl*II (Fig. 6). Preincubation of IN with M-2 in the presence of Mg^{2+} prior to strand transfer was the most efficient method to produce concerted integration products, as well as half-site reactions (Fig. 6A, lanes 6 and 7). The presence of pGEM with M-2 in the preincubation mixture reduced the efficiency for producing all M-2–pGEM recombinants (lanes 2 and 3) by ~80% with no apparent affect on what recombinants were produced. This efficiency was further reduced upon preincubation of IN with only pGEM (lanes 4 and 5) or with IN itself (lanes 8 and 9).

Preincubation of IN–M-2 complexes in the presence of Mn^{2+} (Fig. 6B, lanes 7 and 8) produced different results than those obtained with Mg^{2+} (Fig. 6B, lanes 1 and 2). With Mn^{2+}, the quantity of 3.4 kbp and >3.4 kbp M-2–pGEM recombinants produced were similar to that produced in the presence of Mg^{2+}. However, few concerted integration products were produced in the presence of Mn^{2+} as observed by the lack of linear of 3.92 kbp (lane 7) or the lack of linear 3.4 kbp DNA obtained by *Bgl*II digestion (lane 8). These observed results were independent of which preincubation conditions were used with Mn^{2+} present (Fig. 6B, lanes 3–6 and lanes 9 and 10).

Kinetics of donor use and M-2–pGEM recombinant production

The rate at which specific M-2–pGEM recombinants were formed (Fig. 7) was investigated under standard conditions of IN–M-2 preintegration complex formation with Mg^{2+} (see Fig. 6B, lane 1). Approximately 70% of the M-2 donor was integrated into other M-2 molecules or pGEM within 30 min (Fig. 7A). Maximum production of M-2–M-2 recombinants was evident after 10 min followed by their gradual disappearance. The half-site M-2–M-2 recombinants are presumably a Y-type 'disintegration' substrate (20) for AMV IN. Alternatively, the M-2–M-2 recombinants could serve as a donor substrate for insertion into pGEM, although these complex structures were rarely found by electron microscopy (data not shown). Maximum production of the circular 3.4 kbp and the linear 3.92 kbp DNAs occurred by ~40 min with no apparent decrease thereafter. The nicked 3.4 kbp circles containing one half-site integrated M-2 did not appear to be used for the 'disintegration' reaction. The formation of the linear 3.4 kbp product produced by *Bgl*II digestion of the total M-2–pGEM population paralleled the formation of the linear 3.92 kbp recombinants (Fig. 7B).

DISCUSSION

Physical assay for measuring concerted integration events

AMV IN promotes the efficient concerted integration of M-2 into circular pGEM by two pathways. The least efficient pathway for forming full-site recombinants involves the insertion of one M-2 molecule into the target (Fig. 1C). The most efficient mechanism involves the use of two individual M-2 molecules by IN (Fig. 1D); ~90% of the concerted recombinants produced use this latter pathway. Mu transposase A (21) can also catalyze the concerted integration of linear donor molecules into circular target substrates similar to that shown by AMV IN.

How does a presumed tetramer of IN use two individual M-2 molecules for a concerted integration event?

The simplest model would involve the correct and stable juxtapositioning of the LTR terminus of two individual M-2 molecules by IN for concerted integration into pGEM (Fig. 1D). The reason for the preferred formation of a preintegration complex by IN involving two M-2 donors instead of one for

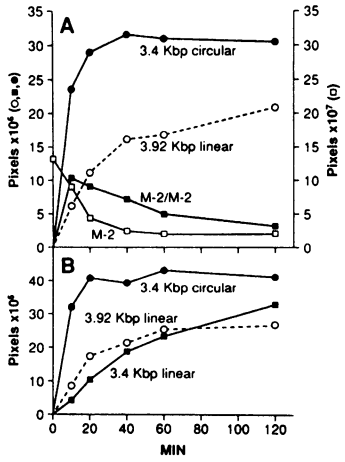

Figure 7. Kinetics of M-2 use and synthesis of donor−donor and donor−target recombinants. (**A**) Standard reaction conditions were used with preincubation of M-2 with IN prior to incubation at 37°C in the presence of pGEM. Aliquots were taken at the indicated times and the samples were subjected to electrophoresis on 1% agarose gels. Upon drying, the gel was subjected to phosphorimager analysis. The utilization of M-2 as donor and synthesis of the various recombinants were quantitated as pixels. (**B**) Aliquots of the above reactions were also analyzed on 1.5% agarose gels. Half of each aliquot was digested with *BglII* prior to electrophoresis to quantitate linear 3.4 kbp DNA.

concerted integration is unknown. Perhaps, like Mu transposase A, the formation of concerted recombinants involving two donors is related to the donor:target ratio (21) or to the concentration of M-2 containing at least one IN dimer per LTR terminus. Other possible insights into how IN selects a pathway for concerted integration involves an understanding of how IN interacts with M-2 in solution. The juxtapositioning of two individual M-2 donors, each containing an IN dimer per LTR terminus (Fig. 1D), is a bimolecular reaction. The juxtapositioning of two LTR termini involving only one M-2 donor by a tetramer of IN is a unimolecular reaction (Fig. 1C). The equilibrium association constants for formation of these two different structures, both capable of performing the concerted integration reaction, are unknown. Another constraint influencing the juxtapositioning of the LTR termini of a single donor by IN is the cyclization probability of the 528 bp M-2. Experimental analysis indicates that this *j* factor reaches a maximum of $\sim 10^{-7}$ M for DNA fragments of ~700 bp (14,22). Other factors which may also affect IN−M-2 interactions include protein concentration and the energetics of both protein−protein interactions and DNA bending.

Trimming, half-site strand transfer, and the 'disintegration' reaction require that HIV-1 IN be dimeric (11,23). It is currently assumed that concerted integration requires a tetramer of IN for forming a stable complex between the donor and the target. We are currently investigating if a tetramer of IN is required not only for the concerted reaction but also for the formation of stable preintegration complexes containing only IN and M-2.

Formation of IN−M-2 preintegration complexes fosters concerted integration events

The initial formation of nucleoprotein complexes in the cytoplasm *in vivo* is presumably virion directed and likely does not require any cellular proteins or nucleic acids. The most efficient mechanism *in vitro* for forming preintegration complexes, capable of concerted integration, required the preincubation of AMV IN homodimer (24) with M-2 prior to strand transfer. The apparent preloading of IN on an LTR terminus promoted protein−protein interactions between two such complexes, regardless of whether one or two M-2 molecules were involved in a concerted reaction. The presence of the target in the preincubation mixture with M-2 seemed only to decrease the total quantity of concerted events (Fig. 6A), suggesting that the target may affect the stability of IN−M-2 preintegration complexes or is a simple competition effect. Currently, our assay procedure does not distinguish between IN−M-2 complexes formed upon preincubation or those formed immediately upon incubation at 37°C. Our data and others suggests that preloading the LTR termini with dimeric IN or other oligomers, similar to that which probably occurs within viral nucleoprotein complexes *in vivo*, enhances the ability of these complexes to perform the 3′ OH trimming of both LTR termini *in vivo* (13) and concerted integration *in vitro*. By using different radioactive end labeling of the 5′ and 3′ OH termini of M-2, it may also be possible to investigate the relationship between the 3′ OH trimming and strand transfer reactions *in vitro*.

ACKNOWLEDGEMENTS

We thank Maria Schnos for her technical assistance with the electron microscopy studies. This work was supported by grants from the National Institutes of Health (AI31334 and CA16312 to D.P.G. and GM14711 to R.B.I.).

REFERENCES

1. Brown, P.O.,Bowerman, B., Varmus, H. E., and Bishop, J.M. (1987) Cell, **49**, 347−356.
2. Bowerman, B., Brown, P.O., Bishop, J.M., and Varmus, H.E. (1989) Genes Dev., **3**, 469−478.
3. Farnet, C.M.,and Haseltine, W.A. (1991) J. Virol., **65**, 1910−1915.
4. Fujiwara, T., and Mizuuchi, K. (1988) Cell, **54**, 497−504.
5. Katzman, M.,Katz, R.A., Skalka, A.M., and Leis, J. (1989) J.Virol., **63**, 5319−5327.
6. Bushman, F.D.,Fujiwara, T.,and Craigie, R. (1990) Science, **249**, 1555−1558.
7. Vora, A.C., Fitzgerald, M.L., and Grandgenett, D.P. (1990) J. Virol., **64**, 5656−5659.
8. Sherman, P.A., and Fyfe, J. A. (1990) Proc. Natl. Acad. Sci. USA, **87**, 5119−5123.
9. Katz, R.A., Merkel, G., Kulkosky, J., Leis, J., and Skalka, A.M. (1990) Cell, **63**, 87−95.
10. Fitzgerald, M.L., Vora, A.C., Zeh, W.G., and Grandgenett, D.P. (1992) J. Virol., **66**, 6257−6263.
11. Engelman,A., Bushman, F.D., and Craigie, R. (1993) EMBO J., **12**, 3269−3275.
12. Vink, C. van der Linden, K.H., Plasterk, R.H.A. (1994) J. Virol., **68**, 1468−1474.
13. Murphy, J.E., and Goff, S.P. (1992) J.Virol., **66**, 5092−5095.
14. Hochschild,A. (1991) Methods Enzymol., **208**, 343−361.
15. Grandgenett, D.P., Inman, R.B., Vora, A.C., and Fitzgerald, M.L. (1993) J.Virol., **67**, 2628−2636.
16. Brenner,S., and Beckwith, J.R. (1965) J. Mol. Biol., **13**, 629−637.
17. Fitzgerald, M.L., Vora, A.C.,and Grandgenett, D.P. (1991) Anal. Biochem., **196**, 19−23.
18. Fitzgerald, M. L., and Grandgenett, D.P. (1994) J. Virol., **68**, 4314−4321.
19. Kitamura, Y., Ha Lee, Y.M., and Coffin, J.M. (1992) Proc. Natl. Acad. Sci. USA, **89**, 5532−5536.
20. Chow, S.A., Vincent, K.A., Ellison, V., and Brown, P.O. (1992) Science, **255**, 723−726.
21. Craigie, R. and Mizuuchi, K. (1987) Cell, **51**, 493−501.
22. Shore, D., Langowski, J. , and Baldwin, R.L. (1981) Proc. Natl. Acad. Sci. USA, **78**, 4833−4837.
23. van Gent, D.C.,Vink, C., Oude Groeneger, A.A.M., and Plasterk, R.H.A. (1993) EMBO J., **12**, 3261−3267.
24. Grandgenett, D.P., Vora, A.C., and Schiff, R. (1978) Virology, **89**, 119−132.