Nucleic Acids Research, 1994, Vol. 22, No. 2 137–144

Premature strand transfer by the HIV-1 reverse transcriptase during strong-stop DNA synthesis

Bep Klaver and Ben Berkhout*
University of Amsterdam, Department of Virology, Academic Medical Center, Meibergdreef 15, 1105 AZ Amsterdam, The Netherlands

Received November 17, 1993; Accepted December 10, 1993

ABSTRACT

Reverse transcription of retroviral genomes starts near the 5′ end of the viral RNA by use of an associated tRNA primer. According to the current model of reverse transcription, the initial cDNA product, termed minus-strand strong-stop DNA, 'jumps' to a repeated sequence (R region) at the 3′ end of the RNA template. The human retroviruses have relatively long R regions (97–247 nucleotides) when compared to murine and avian viruses (16–68 nucleotides). This suggests that the full complement of the R region is not required for strand transfer and that partial cDNA copies of the 5′ R can prematurely jump to the 3′ R. To test this hypothesis, we generated mutants of the human immunodeficiency virus with R region changes and analyzed whether 5′ or 3′ R sequences were inherited by the progeny. We found that in most cases, 5′ R-encoded sequences are dominant, which is consistent with the model of reverse transcription. Using a selection protocol, however, we were also able to identify progeny viruses with R sequences derived from the original 3′ R element. These results suggest that partial strong stop cDNAs can be transferred with R region homologies much shorter than 97 nucleotides.

INTRODUCTION

The long terminal repeats (LTRs) of an integrated retroviral DNA genome contain the start and stop signals for transcription by the cellular RNA polymerase. The viral RNA that is synthesized inherits a subregion of the terminal repeats (R region) that will become important during reverse transcription. Reverse transcription is performed by the virally encoded reverse transcriptase (RT) and occurs within the viral core particle in the cytoplasm of virus-infected cells. In all retroviral systems that have been examined, first-strand (minus-strand) DNA synthesis is primed by a cellular tRNA molecule. The 3′ end of this tRNA is perfectly complementary to an 18-nucleotide region located close to the 5′ end of the viral RNA. Thus, in the initial steps of reverse transcription, a DNA copy of the 5′ repeat (5′ R) is synthesized, which is termed the strong-stop DNA. In the current model of reverse transcription (1, reviewed in 2), a full-length dsDNA molecule is generated in a multi-step process that is initiated by a 'jump' of the strong-stop DNA to the 3′ end of the RNA template, thereby guided by the complementary 3′ R region. Many details of the molecular mechanism of reverse transcription, including this first strand transfer step, are still incompletely understood (see e.g. 3–5).

The actual template for reverse transcription inside viral particles is a dimeric RNA genome consisting of two identical RNA positive strands which are held together in mature virion particles at or near their 5′ ends (reviewed in 6). Although it was originally reported that strong-stop DNA transfer is directed exclusively to the second genomic RNA molecule (7), recent data suggest that intra- and inter-molecular jumps occur with approximately the same frequency (8). It is generally believed that a mechanism exists whereby the 5′ ends and 3′ ends of both genomes are held together in a conformation that allows efficient strand transfer to occur. The higher order structure of the two viral RNA strands or the specific attachment of the RNA templates to the protein core of the virion may facilitate this strand transfer. Consistent with this notion is the finding that *in vitro* strand transfer with purified RT protein and nucleic acid factors is a rather inefficient process at low template concentrations (9).

The efficiency of strong-stop strand transfer may also depend on the number of basepairs that can be formed between the cDNA and the 3′ end of the RNA genome, and this number is determined by the size of the repeat (R) region. Interestingly, the LTRs of various retroviruses show quite different sizes for the R region, ranging from 16 to 247 nucleotides (Table 1). Obviously, strand transfer with a limited overlap of only 16 nucleotides is sufficient in the replication cycle of some viruses (mouse mammary tumor virus, 10; jaagsiekte sheep retrovirus, 11) and even shorter R regions have been reported for endogenous retroviral elements (12). It is therefore possible that retroviruses with a relatively long R region do not necessarily use the full complement of the R region for transfer of the cDNA. Such a premature jump of a partial cDNA ('weak-stop' cDNA) was previously described for the Moloney murine leukemia virus (M-MuLV, 13). More recently, similar observations were reported for the spleen necrosis virus (SNV, 14). We tested this phenomenon using a set of mutant HIV-1 genomes with single base substitutions in the 5′ R or 3′ R regions. As shown by

* To whom correspondence should be addressed

Table 1. Size of repeat (R) regions at termini retroviruses

Size of repeat (R) regions at termini of retroviruses		
16 nucleotides	MMTV	Mouse Mammary Tumor Virus
21	ALV,ASV	Avian Leukosis-Sarcoma Virus
68	Mo-MLV	Moloney Murine Leukemia Virus
71	FIV	Feline Immunodeficiency Virus
78	SNV	Spleen Necrosis Virus
79	EIAV	Equine Infectious Anemia Virus
84	CAEV	Caprine Arthritis-Encephalitis Virus
97	VLV	Visna Virus
97	HIV-1	Human Immunodeficiency Virus type 1
113	BIV	Bovine Immunodeficiency Virus
117	SIV-agm	Simian Immunodeficiency Virus (African green monkey)
173	HIV-2	Human Immunodeficiency Virus type 2
175	SIV-mnd	Simian Immunodeficiency Virus (mandrill)
228	HTLV-I	Human T-cell Leukemia Virus type I
247	HTLV-II	Human T-cell Leukemia Virus type II

sequence analysis, the majority of progeny viral genomes contained R region sequences derived from the 5′ LTR, which is consistent with a transfer of the full-length strong-stop DNA molecule. At low frequency, however, 3′ R sequences were inherited by the viral progeny, which is indicative of a premature jump during strong-stop cDNA synthesis.

MATERIALS AND METHODS

Cells and viruses

The SupT1 T cell line was grown in RPMI 1640 medium containing 10% fetal calf serum (FCS). Introduction of HIV-1 plasmids was carried out by means of electroporation. Briefly, cells were washed once in RPMI with 20% FCS, 2×10^6 cells were mixed with 5 μg DNA in 0.4 cm cuvets and electroporated at 250 Volts (960 μF). Cells were resuspended in complete medium and cultured at 37°C for the time indicated. In long term culturing experiments, cells were split 1 to 10 every 4 days. COS cells were grown in Dulbecco's modified Eagles medium containing 5% serum and transfected by the DEAE-dextran method as previously described (15).

Mutant HIV plasmids

The wild-type HIV plasmid pLAI was previously described (16). In order to individually mutate the two LTR regions, we generated 5′ LTR- and 3′ LTR-subclones in Bluescript KS(+). Blue-5′ LTR contains the complete 5′ LTR as an *Xba*I-*Cla*I fragment and Blue-3′ LTR the 3′ LTR as *Xho*I-*Bgl*I fragment. The B123, BΔ, L4 and L5 mutations in the Tat-responsive TAR element were previously reported to abolish LTR-function in pLTR-CAT constructions (17). These mutations were introduced into the Blue-LTR subclones as *Bsp*EI-*Hin*dIII fragment, and reintroduced into pLAI. The 3′ LTR deletions (3′ R-82 and 3′ R-37) were initially introduced into the Blue-3′ LTR plasmid. The R region *Hin*dIII restriction site at position +82 was ligated to cat sequences as previously described (15). Further deletion of R region sequences was accomplished by digestion with *Sac*I (position +37) and *Hin*dIII, subsequent treatment with the Klenow enzyme and ligation. Both truncated 3′ LTRs were recloned into the pLAI plasmid with either a wild-type or L5-mutated 5′ LTR element. All mutations were verified by sequence analysis of the full-length pLAI constructs. The Tat-expression vector pSVTat was previously described (15).

DNA and RNA analysis

Total cellular DNA was isolated by published procedures (18). The 3′ LTR was specifically amplified in a standard PCR reaction (35 cycles of 1 min 95°C, 1 min 55°C and 2 min 72°C) using the sense Nef-region primer 2 (CATGCGGCCGCAATAGA-GTTAGGCAGGGATA) and the antisense U5-region primer CN1 (GGTCTGAGGGATCTCTAGTTACCAGAGTC). For sequence analysis, the dsDNA PCR fragment was digested with *Bam*HI and *Hin*dIII, cloned into Bluescript KS$^+$ and sequenced on a Biosystems automated sequencer using the Dye Deoxy Terminator Cycle kit and LTR-U3 region primer 5′ CE (CTA-CAAGGGACTTTCCGCTGG). Southern blot analysis of the *Bgl*II digests was performed according to standard procedures (19). The probe used was generated by the random priming method on a plasmid containing the complete 5′ LTR. Total cellular RNA was isolated by the hot phenol method and analyzed by primer extension assays as previously described (20). The oligonucleotides used as primer were the U5 region primer (CN1) GGTCTGAGGGATCTCTAGTTACCAGAGTC and the R region primer (TAR80/50) GCTTTATTGAGGCTTAAGC-AGTGGGTTCCC.

RESULTS

5′ R sequences are inherited by the majority of the HIV-1 progeny

According to the mechanism of reverse transcription the 3′ R RNA sequence is not copied into proviral DNA but merely serves as a bridge to facilitate transfer and elongation of nascent cDNA chains that were copied from the 5′ R region. Thus, if a mutant HIV-1 DNA genome with different R sequences is transcribed, the genomic HIV-1 RNA will be hybrid in nature (Figure 1A; indicated by the 5′ R black box and the 3′ R white box), but after one replication cycle both ends of the newly synthesized proviral DNA molecules will contain sequences corresponding to the 5′ R element. Alternatively, if strand transfer takes place with a partial cDNA of the 5′ R region, the progeny will inherit chimeric R regions consisting of the 3′ part of 5′ R and the 5′ part of 3′ R. In order to test whether a full-length (strong-stop) or partial cDNA (weak-stop; 13) is translocated, we changed the sequence of the 5′ R region and analyzed which repeat element was inherited by the progeny.

The HIV-1 R region, 97 nucleotides in length, encodes the TAR RNA stem-loop structure (shown in Figure 1B) that forms the binding site for the Tat *trans*-activator protein. Since an intact TAR element is critical for viral replication (21,22), we initially introduced a phenotypically silent mutation into TAR. The 3-nucleotide bulge at position 23−25 is either UUU or UCU in most HIV-1 isolates (23,24), and no difference in Tat-binding was observed for these two TAR variants (25). We changed the UUU-bulge in the parental HIV-1 provirus-containing plasmid pLAI into UCU, either in the 5′ R alone or in both 5′ R and 3′ R. This base substitution introduces a *Bgl*II restriction site (A-GAUCU) that was used in the subsequent analysis. We will refer to the parental construct as Bgl$^-$/Bgl$^-$, the hybrid mutant as Bgl$^+$/Bgl$^-$ and the double mutant as Bgl$^+$/Bgl$^+$.

The three plasmids were individually transfected into the SupT1 T cell line and there were several days of viral replication. In order to preclude analysis of the input plasmid DNA sequences, the cell-free culture supernatant was used to infect T cells. Chromosomal DNA was isolated at day 3 and the 3′ LTR of

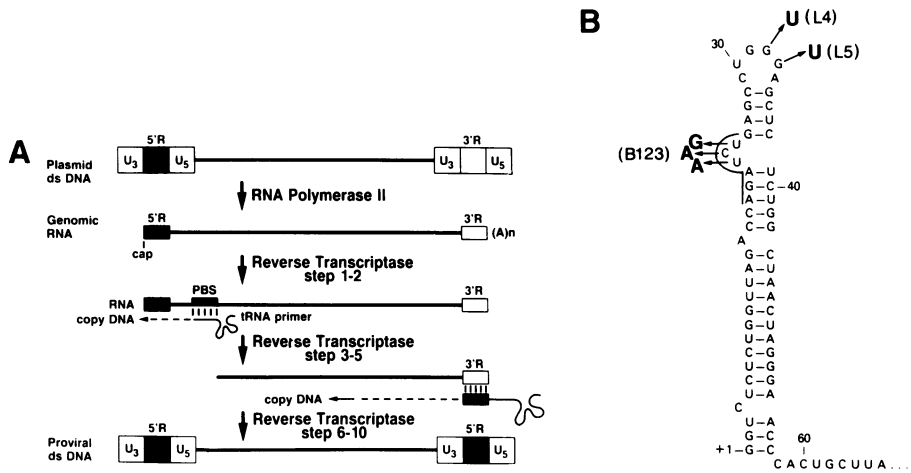

Figure 1. Replication cycle of a transfected retroviral DNA genome and predicted inheritance of the 5′ R region (A). Secondary RNA structure of the TAR RNA element present in the R region of the HIV-1 virus (B). **A.** Structure of LTRs of a full-length HIV-1 genome cloned in a plasmid, the genomic RNA produced by the cellular RNA polymerase II and the proviral DNA generated by the virus-encoded Reverse Transcriptase. Each LTR contains three subregions designated U3, R and U5. RNA transcripts are initiated at the 5′ U3-R border and terminated at the 3′ R-U5 border. Consequently, the 5′U5 and 3′U3 sequences are inherited by the progeny in both LTRs. According to the model of reverse transcription, the 5′ R region (marked as a black box) will be copied into strong-stop DNA. The 3′ R RNA sequence is used only as an acceptor during transfer of this partial cDNA to the 3′ end of the retroviral RNA genome. Step 1 = binding tRNA primer, step 2 = synthesis strong-stop cDNA, step 3 = RNaseH action, step 4 = first jump, step 5 = continued cDNA synthesis, step 6 = initiation +strand strong-stop, step 7 = removal tRNA primer, step 8 = second jump, step 9 = elongation - strand, step 10 = completion ± strands. **B.** RNA structure of the HIV-1 R region. +1 is the transcription initiation site, and only the first 66 of the 97-nucleotide R region are shown. The stem-loop structure is termed the TAR element, which forms the binding site for the *trans*-activator protein Tat. Both the trinucleotide bulge (UCU) and hexanucleotide loop (CUGGGA) contain critical nucleotide sequences. Some of the TAR R region mutations used in this study are indicated: the B123 mutation changes the UCU bulge into AAG, the L4 mutation is a single nucleotide substitution in the loop (G to U change at TAR position 32), and L5 is a G to U change at position 33). The bulge-deletion mutant BΔ is not indicated. The *Bgl*II recognition sequence (AGAUCU) overlapping the bulge element is marked by a line.

integrated proviruses was selectively amplified by PCR (schematized in Figure 2A). Digestion of the 1440-basepair PCR product with *Bgl*II will generate a 638-basepair fragment for the Bgl$^-$ genotype and a 507-basepair fragment for the Bgl$^+$ genotype (Figure 2B, lanes 2 and 7, respectively). The hybrid Bgl$^+$/Bgl$^-$ plasmid generated a progeny with the Bgl$^+$ phenotype (lane 3). This experiment clearly shows that the majority of the progeny had inherited the 5′ R-sequences, indicating that the 5′ R element was copied up to at least position 23 during reverse transcription.

In control experiments, we verified that the Bgl$^-$ marker had no negative effect on viral replication rates. First, we could not detect any difference in virus growth curves between the Bgl$^+$/Bgl$^+$ and Bgl$^-$/Bgl$^-$ viruses. Second, the Bgl$^-$ genotype was found to be stable up to at least six weeks of culture (data not shown).

Efficient reversion of 5′ R-defective mutants in the presence of a wild-type 3′ LTR

The inheritance of the Bgl$^+$ genotype by the Bgl$^+$/Bgl$^-$ virus indicates that the 5′ R is genetically dominant over the 3′ R. These results, however, do not rule out that premature strand transfer occurs at low frequency. In order to select for such an event, we set up a protocol that restricts outgrowth of viruses that inherit the 5′ R element. The strategy was simple to introduce into the 5′ LTR the neutral Bgl$^+$ marker in combination with a second mutation that inactivates the TAR element. The L4,Bgl$^+$/Bgl$^-$ construct contains such an additional mutation in the 5′ TAR loop (G to U at position 32, see Figure 1B). This L4 mutation was previously shown to inhibit Tat-mediated *trans*-activation of the HIV-1 LTR promoter in transient transfection assays (17,26). Upon transfection of T cells, most viruses produced will inherit the defective L4,Bgl$^+$ genotype. The inability of this virus to efficiently spread through the culture is apparent from the reduced levels of Bgl$^+$ DNA detected in the day 3 sample compared to the control construct Bgl$^+$/Bgl$^-$ (Figure 2B; compare lane 4 to 3). The major difference with the control sample is the presence of a significant amount of 3′ R-derived Bgl$^-$ DNA in the proviral DNA (lane 4). Furthermore, the ratio of Bgl$^-$ to Bgl$^+$ DNA steadily increased upon prolonged culturing for 7 and 14 days (lanes 5 and 6, respectively). Sequence analysis of the DNA synthesized by these revertant viruses unequivocally demonstrated that the viruses contained the wild-type 3′ R sequences (not shown). These results are consistent with the hypothesis that partial cDNAs, in this case at least 32 nucleotides shorter than the full-length strong-stop cDNA, can be translocated to the 3′ R.

We reasoned that if reverse transcription could circumvent the 5′L4 mutation by means of an early bypass to the wild-type 3′ R element, other TAR mutations should be reversed with approximate equal efficiency. First of all, we tested another TAR loop mutation (L5; G to U at position 33) that was previously reported to dramatically inhibit LTR-mediated transcription (17,26). Mutant HIV-1 genomes were constructed with either a single L5 mutation in the 5′ or 3′ LTR (L5-wt and wt-L5, respectively) or a double L5-L5 mutation. These plasmids and the wild-type construct were transiently transfected into COS cells for analysis of 5′ LTR-function. The amount of viral RNA produced was quantitated by primer extension analysis (Figure 3). Indeed, viral transcription was severely reduced by the L5 mutation in the 5′ LTR (constructs L5-wt and L5-L5, lanes 2 and 3, respectively), and this inhibition was not relieved in the presence of a co-transfected Tat-encoding plasmid (lanes 6 and 7).

The ability of the different plasmids to produce virus was further demonstrated in long-term T cell transfection assays.

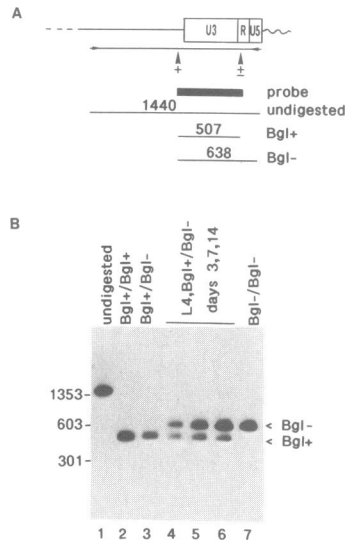

Figure 2. Inheritance of a 5′ LTR-encoded *Bgl*II site in the 3′ LTR of the viral progeny. **A.** Schematic of the 3′ LTR DNA of an integrated HIV-1 provirus, the PCR-product and the different *Bgl*II-digestion products. The position of an invariable *Bgl*II site (+) and the introduced R region site (±) is indicated by arrowheads. The probe used in the Southern blot analysis (panel B) covers the complete U3 region and will detect fragments of either 507- or 638-nucleotides long (Bgl^+ and Bgl^- phenotype, respectively). **B.** SupT1 cells were transfected with the various HIV-1 plasmids indicated on top of the panel. The filtered culture supernatants were used to infect T cells and total cellular DNA was isolated 3 days later (except lanes 5 and 6; day 7 and day 14 samples, respectively). The PCR-amplified LTR-DNA (panel A) was digested with *Bgl*II, size-fractionated on agarose gel, transferred to a zetaprobe membrane, and hybridized to a ^{32}P-labeled LTR-U3 region DNA probe (see panel A). The position of DNA size markers is indicated on the left and the *Bgl*II-phenotype is indicated on the right. Lane 1 shows an undigested control sample.

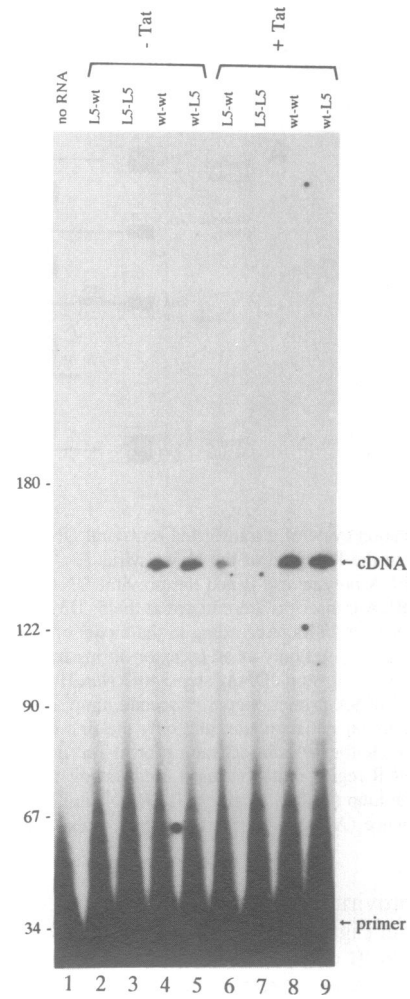

Figure 3. Primer extension analysis of HIV-1 RNA isolated from COS cells transfected with the indicated plasmids. All transfections were performed in the absence and presence of a Tat-encoding plasmid (− and + Tat, respectively). A 30-nucleotide, end-labeled U5 region primer was used (primer alone in lane 1). The position of primer and cDNA product is indicated on the right. End-labeled pBR322-HpaII fragments were used as molecular size markers are indicated on the left.

Consistent with the RNA data, no virus was produced by L5-L5 and L5-wt transfected cells (Figure 4A). However, the defect in virus production seen for the hybrid L5-wt construct was only transient in nature. Viruses appeared around day 8, displaying wild-type growth curves, and sequence analysis showed that this virus population had inherited the wild-type 3′ R sequences. A similar delay in virus production was seen for other 5′ LTR HIV mutants carrying either a three-nucleotide substitution or deletion in the 5′ TAR bulge (Figure 4B; B123-wt and BΔ-wt, respectively). All these hybrid mutants produced revertant viruses with approximately equal efficiencies and all revertants contained the 3′ R-encoded wild-type TAR sequences (not shown). In contrast to the 5′ LTR-mutants, no defect in virus production was measured for the wt-L5 plasmid, which is consistent with the notion that most of the progeny will inherit the 5′ R. These combined results led us to propose a mechanism for the inheritance of 3′ R sequences that occurs at relatively low frequency (schematically depicted in Figure 5).

DNA recombination cannot explain the observed reversion

A trivial explanation for the efficient loss of 5′ R mutations in transfections with hybrid HIV plasmids is that recombination of LTR DNA sequences occurred in transfected cells. Relatively high frequencies have been reported for such intracellular homologous recombination events (27−29). In our system, recombination can occur intramolecularly between two LTRs of one plasmid or intermolecularly between two different HIV plasmids. To directly test whether DNA recombination results in the rescue of 5′ LTR-mutants, we generated two additional hybrid constructs with different mutations in their 5′ and 3′ TAR elements (B123-L5 and L5-B123, Figure 6A). The B123-L5 and L5-B123 mutants are expected to behave identical in a DNA-mediated recombination/reversion event, but reversion during reverse transcription is possible only for the B123-L5 template (illustrated in Figure 6A). The results showed that virus was produced by the B123-L5 mutant, but not by L5-B123 (Figure 6B). The LTR DNA sequence of the revertant virus confirmed that a fully wild-type sequence was regenerated. The differential behaviour of the two mutants indicates that premature strong-stop DNA transfer rather than DNA recombination is involved in the efficient rescue of 5′ R-mutated HIV-1 genomes. Interestingly, reversion of the B123-L5 mutant is inefficient compared to the B123-wt construct (not shown). A possible explanation for this is the limited window in which strand transfer is productive for the B123-L5 mutant. While any translocation before the 5′B123 mutation (position 24) will revert the B123-wt construct, synthesis of a wild-type R on the B123-L5 template

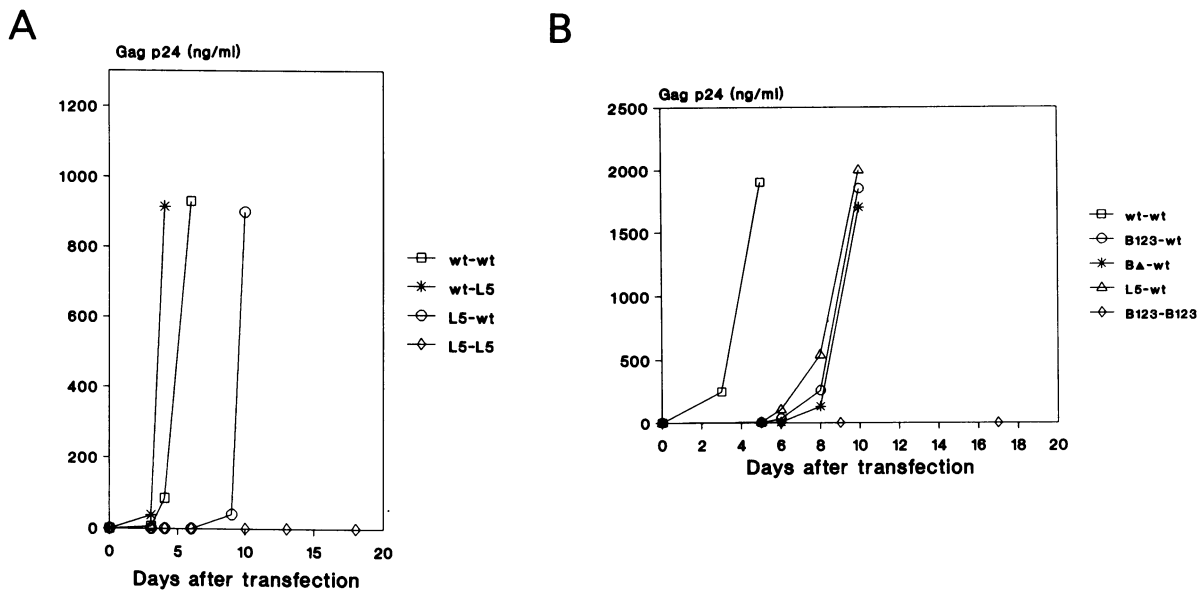

Figure 4. Kinetics of virus production with 5′ LTR- and 3′ LTR-mutated HIV-1 genomes. SupT1 cells were transfected with 5 μg of the indicated plasmids and production of progeny virions was monitored by the appearance of Gag p24 protein in the culture supernatant. Gag p24 levels were quantitated in a standard ELISA assay. All DNA transfections were performed on at least three separate occasions. The results shown are from one representative experiment.

is restricted to translocations that occur in the 8-nucleotide long segment between TAR nucleotides 24 and 33 (the position of the B123 and L5 markers, respectively).

Reversion of 5′ R-mutations is sensitive to truncation of the 3′ R region

The previous experiments suggested to us that the full complement of the 97-nucleotides long R region is not essential for the HIV-1 virus to carry out cDNA translocations. Based on these results, it is expected that the non-coding 3′ R can be truncated from the 3′ side without a significant loss in the efficiency of strand transfer. This was tested with two proviral constructs carrying shortened 3′ R regions of 82 and 37 nucleotides instead of the 97 nucleotides for wild-type HIV. Indeed, much of the 3′ part of the 3′ R region could be removed without apparent deleterious effect to virus replication. The two mutant DNAs were infectious and produced virus with wild-type kinetics (Klaver and Berkhout, unpublished data). These results show that a limited homology of 37 nucleotides is sufficient to allow strong-stop DNA transfer.

Reversion of 5′L5-mutants, which involves a premature cDNA translocation before TAR nucleotide 33 is copied, is expected to be more sensitive to 3′ R-truncation (Figure 7A). In the case of the L5−3′ R-37 double mutant for instance, such a weak-stop cDNA will only have a 4 base complementarity to the truncated 3′ R. The three DNA constructs depicted in Figure 7A were transfected in T cells and the kinetics of appearance of revertant virus was monitored (Figure 7B). Mutant L5−3′ R-82, containing a small 3′ deletion in the 3′ R, produced normal levels of revertant virus when compared to the control plasmid L5-wt (L5−3′ R-97 in Figure 7), and analysis of the progeny DNA of this revertant indicated that a full-length, wild-type R sequence was regenerated (not shown). The mutant with the more extensive 3′ R deletion (L5−3′ R-37) produced no virus up to 27 days after transfection. Thus, a 37-nucleotides long 3′ R region, which is sufficient for a normal strong-stop cDNA jump, cannot support the weak-stop-mediated reversion of the 5′L5-mutation.

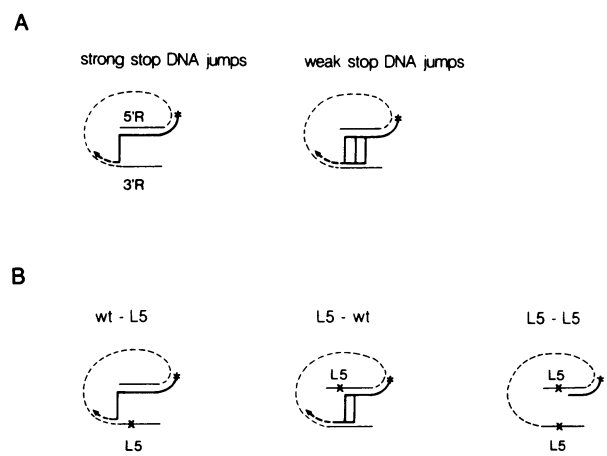

Figure 5. Model of premature strong-stop DNA transfer. The RNA template is schematically depicted with the 5′ and 3′ R regions in close proximity and in parallel orientation. The position of the initiator $tRNA^{lys,3}$ molecule indicated by * and the point of strand-transfer is depicted as a cross-over from 5′ R to 3′ R. Panel **A** shows a full-length (strong-stop) cDNA transfer and three possible jump positions of a partial (weak-stop) cDNA. The situation shown is a simplification in that only one RNA strand of the dimeric genome is used to visualize an intramolecular jump. Alternatively, a more complex, but essentially similar model can be drawn for an inter-molecular jump between two RNA molecules. It is important to note that the interpretation of our reversion data is not influenced by the type of jump because both genomic RNAs are identical in sequence. Panel **B** shows the postulated RT route that will generate a wild-type progeny on templates with a single L5 mutation. The regular strong-stop pathway will efficiently ignore the L5 mutation in the wt-L5 mutant, but a weak-stop jump is necessary for reversion of the L5-wt construct. No RT-mediated reversion of the L5-L5 double mutant is possible.

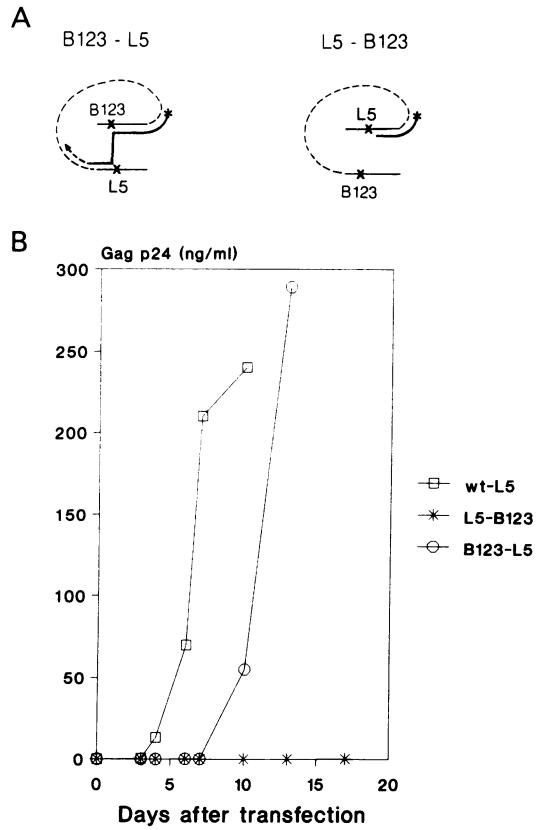

Figure 6. Schematic representation of the B123-L5 and L5-B123 double mutants (panel A) and kinetics of virus production in transfected SupT1 cells (panel B). The results shown are from one representative experiment. All DNA transfections were performed on at least two separate occasions. Culture supernatant samples were assayed for Gag p24 protein using a standard ELISA assay.

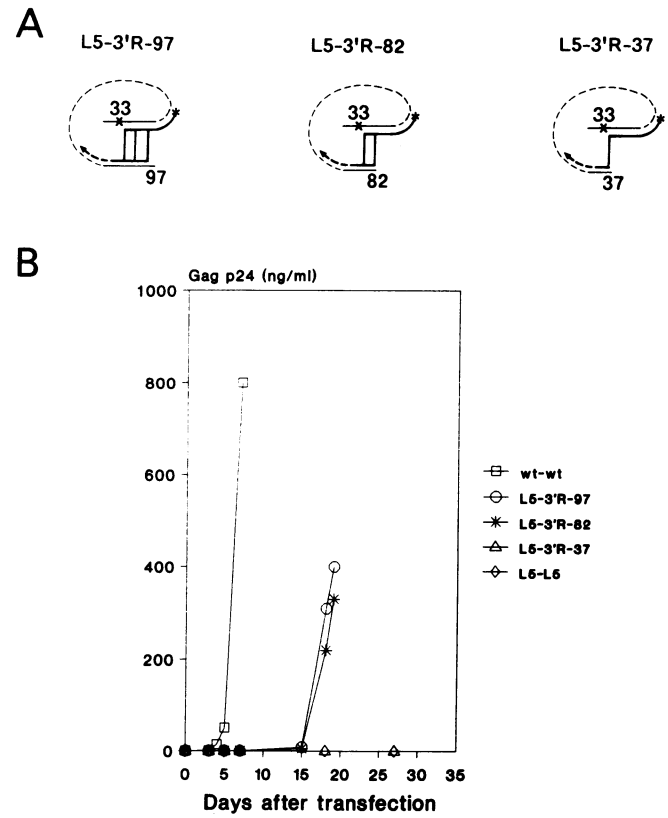

Figure 7. Schematic representation of the L5-wt (L5−3′ R-97), L5−3′ R-82 and L5−3′ R-37 mutants (panel A) and kinetics of virus production in transfected SupT1 cells (panel B). The results shown are from one representative experiment. All DNA transfections were performed on at least two separate occasions. Culture supernatant samples were assayed for Gag p24 protein using a standard ELISA assay.

In vitro analysis of weak-stop cDNAs synthesized on the HIV-1 RNA template

The results described above show that the HIV-1 RT enzyme does not always produce a full-length cDNA on the 5′ R RNA template. It is possible that specific sequences or structures within the viral RNA template do inhibit elongation of the RT enzyme and that these sites form hot spots for premature strand transfer. In order to identify such pause/stop sites we carried out *in vitro* reverse transcription reactions with the natural HIV-1 RNA template. Figure 8 shows the cDNA products obtained when an end-labeled DNA primer was extended on HIV-1 RNA under different experimental conditions. The major reaction product is the strong stop DNA when the reaction is performed at optimal temperature (40°C; lane 9) and with an optimal amount of RT enzyme (200 Units; lane 14). When the incubation temperature was decreased to 10°C (lane 6), there was no detectable strong-stop cDNA product, but instead some partial 'weak-stop' cDNAs were detected. Similar pause sites were detected in assays with limiting amounts of the RT enzyme (0.5 and 2 Units; lanes 11 and 12). The size of this major 'weak-stop' cDNA species corresponds with a termination of reverse transcription at the base of the TAR hairpin at nucleotide position 55 (Figure 1B). In addition, minor pause sites around nucleotide 25 were detected in a similar experiment in which more primer DNA was used (lanes 15−19, stops are most abundant in lane 16). Thus, specific regions of the HIV-1 5′ R RNA do cause reverse transcriptase to pause. Since the major pause site is located at the base of the relatively stable TAR hairpin, it is possible that this RNA structure slows down reverse transcription, thereby triggering a premature jump.

DISCUSSION

The current model of reverse transcription (1) predicts that 5′ and 3′ R regions of viral progeny will be identical to the parental 5′ R, and not to the parental 3′ R. Thus, mutations introduced into the 5′ R will be inherited by the viral progeny, whereas mutations in the 3′ R region will be lost. We have addressed this phenomenon using HIV-1 constructs with phenotypically silent and non-silent mutations in the 5′ R element. Our results show that silent 5′ R markers are efficiently transferred to the progeny, which is in accord with the mechanistic model of reverse transcription. Studies with the non-silent 5′ LTR-inactivating mutations, however, clearly demonstrate that a reversion mechanism is operating which leads to inheritance of the wild-type 3′ R region. This reversion does not occur when the inactivating mutation is present in both LTRs of the HIV-1 plasmid. These results are best explained by occasional translocation of non-full-length cDNAs (weak-stops) that have not yet copied the 5′ R mutation. Such a mechanism was originally proposed by Lobel and Goff (13) to explain the

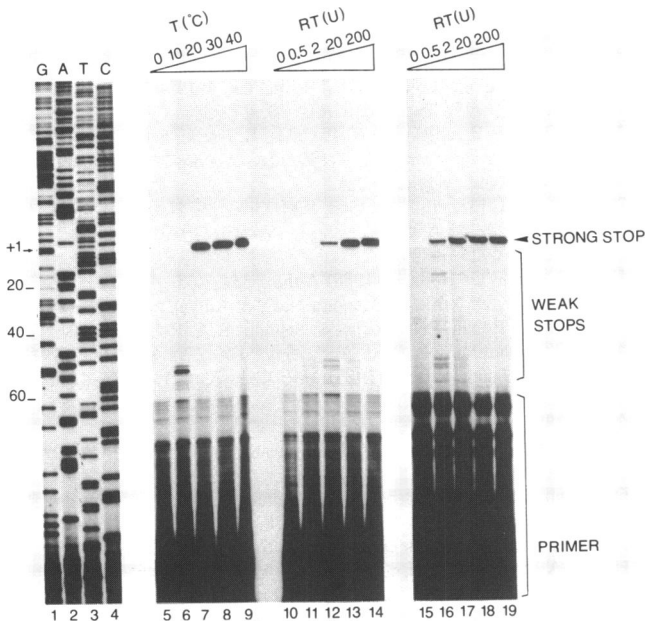

Figure 8. *In vitro* reverse transcription assay with HIV-1 RNA isolated from HIV-infected cells. The HIV-1 RNA was annealed to the end-labeled R region primer TAR80/50 and extended under various conditions. The effect of the incubation temperature (lanes 5–9) and RT concentration (lanes 10–14 and 15–19) was tested. A 10-fold higher amount of primer DNA was used in lanes 15–19 compared to lanes 10–14. Dideoxy sequencing of an LTR-plasmid using the same end-labeled primer was aligned with the primer extension data (lanes 1–4, numbering on the left corresponds to original TAR coordinates, +1 is the transcription initiation site). The position of the primer, weak stop and strong stop cDNA products are indicated on the right.

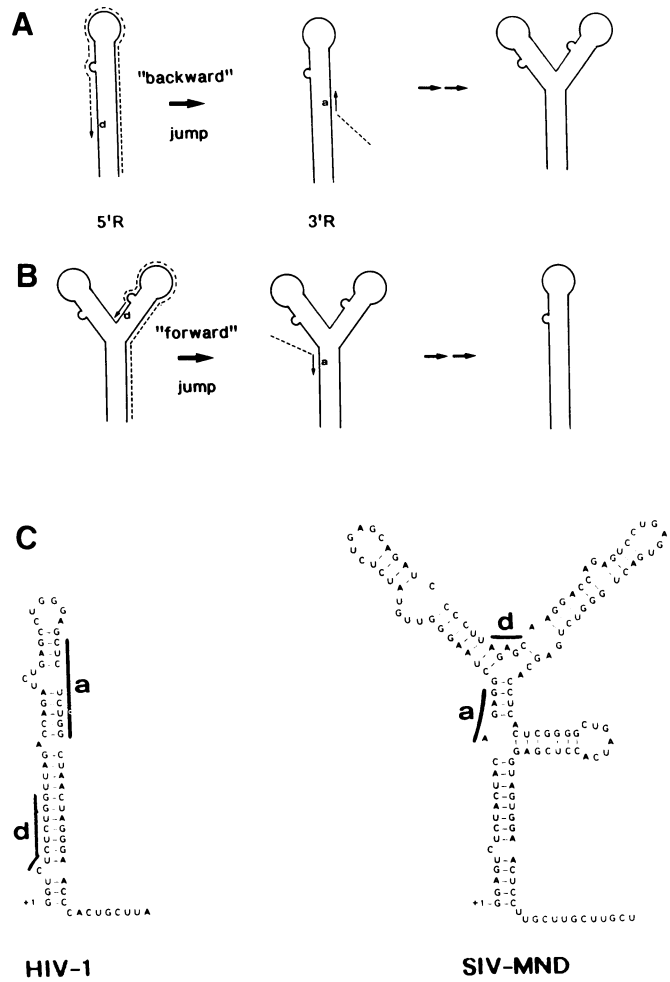

Figure 9. Weak-stop DNA transfer and the evolution of TAR RNA structures. **A.** A model for the transition from simple to duplex TAR stem-loop structures during reverse transcription. A simple, monomeric TAR element (prototype HIV-1, panel C) is shown as 5′ R template for reverse transcription (interrupted line). The growing point of polymerization is indicated by an arrow. A premature transfer from a donor sequence (d) in 5′ R to a partially complementary, downstream acceptor sequence (a) in the 3′ R is indicated. This 'backward' jump will lead to duplication of the upper part of the TAR stem-loop. **B.** A model for the one-step transition from complex to simple TAR structures. The branched TAR element (prototype SIV-MND, panel C) is shown as 5′ R template for reverse transcription. Upon copying of donor sequence d in the 5′ R, the weak-stop cDNA jumps to a partially complementary, upstream acceptor sequence (a) in the 3′ R. This 'forward' jump results in deletion of the first loop of the tandem TAR structure. **C.** Actual nucleotide sequence and structure model of TAR RNA of HIV-1 and SIV-MND (23,24). Putative donor and acceptor sequences are indicated.

behaviour of LTR mutants of the Moloney murine leukemia virus (M-MuLV). In general, it is well known that RT enzymes frequently switch templates during chain elongation. This property, combined with the dimeric nature of retroviral genomes, results in high rates of genetic recombination.

A major implication of our results is that the retroviral DNA translocation can occur with only limited basepair potential between the cDNA-donor and RNA-acceptor. This idea is supported by a comparative analysis of the length of a variety of retroviral R elements (Table 1). This survey indicates that an efficient strong-stop DNA jump is possible with as little as 16 basepairs. Dunn *et al* (30), studying hybrids between Rous sarcoma virus and cellular sequences, have identified a cDNA product that was generated by a first strand transfer with only four bases of perfect homology. More recently, several groups analyzed the strand transfer reaction *in vitro* using purified RT enzyme and nucleic acid molecules (9,31,32). These studies indicated that the efficiency of the jump was dependent on the length of the complementary sequence. No strand switching was detected with an overlap of 10 nucleotides, but significant strand transfer was measured with R regions of 20 nucleotides. Furthermore, the molecular model for retroviral DNA synthesis invokes a second DNA transfer reaction that uses an 18 basepair interaction (1,7) and detailed analysis of RNA recombination events in retroviruses showed that template switching can be mediated by only a few or even no bases of homology (33–37). These combined observations strongly suggest HIV does not require the full complement of its 97-nucleotides-long R region to facilitate the strong-stop DNA translocation.

Multiple factors may influence the processivity of reverse transcription *in vivo* and influence the site and/or frequency of strand switching. For instance, reverse transcription of the HIV-1 genome was shown to be aborted in quiescent primary T cells (38). Mitogenic stimulation of these cells could induce virus production, suggesting that RT activity can be modulated by a cellular factor. Alternatively, the sequence and/or structure of the RNA template may interfere with elongation of reverse transcription (39). Our *in vitro* experiments suggest that the HIV-1 TAR hairpin is such an inhibitory structure and recent data by DeStefano *et al* (31) support the idea that pauses can induce strand transfers. Another possibility is that base-modifications in the viral template influence the processivity of the RT enzyme (1,40–43).

Interestingly, a specific A to I base-modification has been reported at position 26 of the HIV-1 TAR element (44,45). Thus, one expects to find a stable A to G change in the genomes of naturally occurring HIV-1 isolates, since I directs incorporation of C into minus-strand cDNA and C encodes G in the plus-strand viral RNA. However, available sequence information on many HIV-1 isolates does not support this hypothesis (24). The premature jump-mechanism could explain the inheritance of wild-type 3′ TAR sequences from 5′ TAR-modified templates.

Our results are consistent with the notion that transfer of weak-stop cDNAs occur at low frequency in the replicative cycle of wild-type HIV virus. We may cautiously suggest that this RNA recombination mechanism has been involved in the evolution of TAR motifs within the HIV-SIV lentiviral family. Although the HIV-SIV viruses are generally believed to share a very similar Tat-*trans*-activation mechanism (reviewed in 46), considerable heterogeneity in the length and the structure of the TAR elements was reported (23). For instance, the difference in size of the R region between HIV-1 and the simian immunodeficiency virus isolated from mandrills (SIV-MND) can be largely accounted for by the duplication of the upper part of the HIV-1 TAR stem-loop (Figure 9C). The weak-stop DNA transfer reaction, combined with illegitimate annealing to partially complementary 3′ R sequences, could provide a one-step mechanism for the evolution of simple to complex TAR structures and *vice versa* (Figure 9A and 9B, respectively).

ACKNOWLEDGEMENTS

We thank Wim van Est for photography work, Ms J.H. van Horssen-Medema for artwork, and Atze Das and Charles Boucher for critical reviews of the manuscript. This work was supported by the Netherlands Organization for Scientific Research (NWO) and the Dutch Cancer Society (KWF).

REFERENCES

1. Gilboa, E., S.W. Mitra, S. Goff and D. Baltimore (1979) *Cell* **18**, 93–100.
2. Varmus, H. and R. Swanstrom (1984) in Molecular Biology of Tumor Viruses. Eds. R. Weiss, N. Teich, H. Varmus and J. Coffin. Cold Spring Harbor Laboratory, USA, pp 369–512.
3. Cook, P.R. (1993) *J. Gen. Virol.* **74**, 691–697.
4. Li, P., A.J. Stephenson, L.J. Kuiper and C.J. Burrell (1993) *Virology* **194**, 82–88.
5. Panganiban, A.T. (1990) *Seminars in Virology* **1**, 187–194.
6. Coffin, J. (1984) in Molecular Biology of Tumor Viruses. Eds. R. Weiss, N. Teich, H. Varmus and J. Coffin. Cold Spring Harbor Laboratory, USA, pp 244–367.
7. Panganiban, A.T. and D. Fiore (1988) *Science* **241**, 1064–1069.
8. Hu, W.-S. and H.M. Temin (1990) *Science* **250**, 1227–1233.
9. Luo, G. and J. Taylor (1990) *J. Virol.* **64**, 4321–4328.
10. Moore, R., M. Dixon, R. Smith, G. Peters and C. Dickson (1987) *J. Virol.* **61**, 480–490.
11. York, D.F., R. Vigne, D.W. Verwoerd and G. Querat (1992) *J. Virol.* **66**, 4930–4939.
12. Lania, L., A. Di Cristofano, M. Strazzullo, G. Pengue, B. Majello and G. La Mantia (1992) *Virology* **191**, 464–468.
13. Lobel, L.I. and S.P. Goff (1985) *J. Virol.* **53**, 447–455.
14. Ramsey, C.A. and A.T. Panganiban (1993) J. Virol. 67, 4114–4121.
15. Berkhout, B., R.H. Silverman and K.-T. Jeang (1989) *Cell* **59**, 273–282.
16. Peden, K., M. Emerman and L. Montagnier (1991) *Virology* **185**, 661–672.
17. Berkhout, B. and K.-T. Jeang (1989) *J. Virol.* **63**, 5501–5504.
18. Boom, R., C.J.A. Sol, M.M.M. Salimans, C.L. Jansen, P.M.E. Wertheim-van Dillen and J. van der Noordaa (1990) *J. Clin. Microbiol.* 28, 495–503.
19. Maniatis, T., E.F. Fritsch and J. Sambrook (Ed.) in Molecular cloning: a laboratory manual (1982) Cold Spring Harbor, New York:
20. Berkhout, B. and K.-T. Jeang (1991) *Nucl. Acids Res.* 19, 6169–6176.
21. Harrich, D., J. Garcia, R. Mitsuyasu and R. Gaynor (1990) *EMBO J.* **9**, 4417–4423.
22. Leonard, J., C. Parrott, A.J. Buckler-White, W. Turner, E.K. Ross, M.A. Martin and A.R. Rabson (1989) *J. Virol.* **63**, 4919–4929.
23. Berkhout, B. (1992) *Nucl. Acids Res.* **20**, 27–31.
24. Myers, G., J.A. Berzofsky, G.N. Pavlakis, B. Korber and R.F. Smith (Ed.) in Human Retroviruses and AIDS (1992) Los Alamos National Laboratory, Los Alamos, New Mexico, USA.
25. Dingwall, C., I. Ernberg, M.J. Gait, S.M. Green, S. Heaphy, J. Karn, A.D. Lowe, M. Singh, M.A. Skinner and R. Valerio (1989) *Proc. Natl. Acad. Sci. USA* **86**, 6925–6929.
26. Feng, S. and E.C. Holland (1988) *Nature* **334**, 165–167.
27. Srinivasan, A., D. York, R. Jannoun-Nasr, S. Kalyanaraman, S.D. Swan, J. Benson, C. Bohan, P.A. Luciw, S. Schnoll, R.A. Robinson, S.M. Desai and S.G. Devare (1989) *Proc. Natl. Acad. Sci. USA* **86**, 6388–6392.
28. Subramani, S. and P. Berg (1983) *Mol. Cell. Biol.* **3**, 1040–1052.
29. Wake, C.T. and J.H. Wilson (1979) *Proc. Natl. Acad. Sci. USA* **76**, 2876–2880.
30. Dunn, M.M., J. C. Olsen and R. Swanstrom (1992) *EMBO J.* **66**, 5735–5743.
31. DeStefano, J.J., L.M. Mallaber, L. Rodriguez-Rodriguez, P.J. Fay and R.A. Bambara (1992) *J. Virol.* **66**, 6370–6378.
32. Pelinska, J.A. and S.J. Benkovic (1992) *Science* **258**, 1112–1118.
33. Hajjar, A.M. and M.L. Linial (1993) J. Virol. 67, 3845–3853.
34. Pathak, V.K. and H.M. Temin (1990) *Proc. Natl. Acad. Sci. USA* **87**, 6024–6028.
35. Pulsinelli, G.A. and H.M. Temin (1991) *J. Virol.* **65**, 4786–4797.
36. Swain, A. and J.M. Coffin (1992) *Science* 225, 841–845.
37. Zhang, J. and H.M. Temin (1993) *J. Virol.* **67**, 1747–1751.
38. Zack, J.A., A.M. Haislip, P. Krogstad and I.S.Y. Chen (1992) *J. Virol.* **66**, 1717–1725.
39. Klarmann, G.J., C.A. Schauber and B.P. Preston (1993) *J. Biol. Chem.* 268, 9793–9802.
40. Beemon, K.L. and J.M. Keith (1976) in Animal Virology (Ed. D. Baltimore *et al*) Academic Press, New York, pp. 97–105.
41. Furiuchi, Y., A.J. Shatkin, E. Stavnezer and J.M. Bishop (1975) *Nature* **257**, 618–620.
42. Stolzfus, C.M. and K. Dimock (1976) *J. Virol.* **18**, 586–595.
43. Youvan, D.C. and J.E. Hearst (1979) *Proc. Natl. Acad. Sci. USA* **76**, 3751–3754.
44. Blanchard, A.D., R. Powell, M. Braddock, A.J. Kingsman and S.M. Kingsman (1992) *J. Virol.* **66**, 6769–6772.
45. Sharmeen, L., B. Bass, N. Sonenberg, H. Weintraub and M. Groudine (1991) *Proc. Natl. Acad. Sci. USA* **88**, 8096–8100.
46. Cullen, B.R. (1992) *Microbiol. Rev.* **56**, 375–394.