© 1992 Oxford University Press
Nucleic Acids Research, 1992, Vol. 20, No. 23 6261–6266

# HIV-1 promotor insertion revealed by selective detection of chimeric provirus-host gene transcripts

Ines Raineri and Hans-Peter Senn*
Institute for Medical Microbiology, University of Basel, Petersplatz 10, 4003 Basel, Switzerland

Received September 9, 1992; Revised and Accepted November 6, 1992 EMBL accession no. X68242

## ABSTRACT

**To study host gene activation by retroviral promotor insertion, a polymerase chain reaction (PCR) assay was developed. This method allows a sensitive and selective detection of chimeric provirus-host gene transcripts, hallmarks of insertional activation events, which does not rely on an induction of tumor cell growth. We analysed HIV-1 infected cells of a CD4+ T-cell line (H9), infected peripheral blood mononuclear cells and cells in broncho-alveolar washes of AIDS patients. In each case, a variety of chimeric mRNA molecules were detected using a PCR amplification reaction and 5′ primers specific to the HIV-1 LTR and 3′ primers specific to poly A of mRNA. In infected H9 lymphocytes, a mRNA was identified encoding a putative protein of 145 amino-acids that was not expressed in uninfected H9 cells. This shows for the first time that HIV-1 can activate transcription of host cellular genes by promotor insertion in a fashion similar to slow-transforming avian and murine retroviruses.**

## INTRODUCTION

The integration of proviral DNA into the cellular genome represents an insertional mutagenesis mechanism that results in the disruption of contiguous host sequences (1–3). Possible consequences for host genes include their inactivation, their activation by the retroviral promotor and/or enhancer sequence, and gene transduction (4,5). Early findings suggested that provirus integration into the host genome is a random process (6). More recent studies found a predilection for provirus integration in or near active genes and in close proximity to DNase I hypersensitive sites (7–9). In a study on avian leukosis virus, approximately 20% of proviral insertions occurred within an average of 800 sites in the cellular genome (10).

In contrast, common integration sites in tumor cells are explained by clonal outgrowth of cells harboring proviral DNA in particular loci which activate tumor cell growth. These proviral insertions may result in dominant cellular phenotypes, so it is not surprising that the most extensively studied effects of proviral insertions are due to the activation of nearby cellular proto-oncogenes (11). In contrast, it is less clear, whether the activation of genes not directly involved in cellular growth control and differentiation also contribute to pathogenetic processes.

For this purpose, we have established a PCR method which allows a selective detection of chimeric provirus-host gene transcripts. We focused our studies on the capacity of HIV-1 to serve as an insertional mutagen. HIV-1 is a member of the lentivirus family and the insertion of its proviral DNA into the genome of CD4+ cells does not directly provoke tumor cell growth. On the other hand, its propensity to generate defective HIV-1 proviruses (12) might favor the activation of host cell genes by the 3′ LTR promotor/enhancer. In this report we demonstrate that such chimeric HIV-1–host cell gene transcripts occur *in vitro* and *in vivo*. Therefore, our findings show that HIV-1 proviral DNA insertion can result in activation and/or alteration of transcription of nearby host cell genes.

## MATERIAL AND METHODS

### Cells and infection

Cells of the H9 T-cell line (13) were propagated either uninfected or chronically infected with HIV-1 (HTLV IIIB). Peripheral blood mononuclear cells (PBMC) were obtained from heparinized venous blood of normal blood donors by Ficoll-Paque separation. These cells were stimulated with phytohemagglutinin (PHA) (10$\mu$g/ml) for 48 hrs. Subsequently, the culture of these cells was continued in RPMI 1640 containing 10% fetal calf serum, 2mM glutamine, 50 mM $\beta$-2-mercaptoethanol, 1$\mu$g/ml polybrene and 20ng/ml rIL-2. Infection of these cells was performed, at high viral multiplicity, using plasma of an HIV-1 infected newborn child. Unabsorbed virus was removed by washing of cells after 24 hrs. The medium was changed every 3–4 days and the cultures were allogeneically stimulated by the addition of PHA-activated donor cells (PBMC). At different time points (see Results), aliquots of cells were removed and tested for the presence of multiply spliced and chimeric HIV-1 transcripts.

Cells of AIDS patients obtained from broncho-alveolar exudates were washed with phosphate-buffered saline before mRNA isolation (14) and tested for normal and chimeric HIV-1 transcripts.

### ‘Chimeric’ polymerase chain reaction

Lysates were made from aliquots of $10^5$ infected cells or uninfected control cells. The mRNA was enriched using Dynabeads[R] Oligo$(dT)_{25}$ (Dynal, Milan Biotech Inc.). Beads-coupled oligo-$(dT)_{25}$ primers were used in first strand cDNA

* To whom correspondence should be addressed

synthesis with Moloney MLV reverse transcriptase for 1 h at 42°C, essentially as described (14). However, this reaction mix was prewarmed to 42°C before adding reverse transcriptase. This improved the specificity of oligo(dT)−priming. For priming of the 'minus libraries', oligonucleotides complementary to the HIV-1 U3 region (M185, M186,& M187) were used together with the beads-coupled oligo-$(dT)_{25}$ (Fig. 1). An annealing step of 30 min at 42°C was performed before reverse transcription. After the reaction, mRNA/cDNA duplexes were melted twice at 95°C for 1 min. The supernatants were discarded and second strand cDNA synthesis was initiated with M101 (0.5 pmoles/100μl) and 2 U Taq DNA polymerase in 100 μl 1×TB buffer (50 mM KCl, 20 mM Tris HCl (pH 8.4), 1.5 mM $MgCl_2$ and 0.1 mg/ml BSA (Boehringer) and 200 μM dNTP each). The reaction mix was heated to 95°C, beads were added and the reaction was performed for 20 sec at 94°C, 2 min at 58°C and 3 min at 72°C. To stop the reaction, tubes were immediately put on ice-water, the reaction mix was removed and 50μl 1×TB buffer was added. Second strands were recovered from the solid phase cDNA pool by melting at 95°C for 2 min and separation of supernatant and beads with the help of a magnet holder. New complementary first strands were synthesized using the above mentioned reaction mix preheated to 95°C and oligo-$(dT)_{21}$ anchor primer M108 (0.5 pmoles/100 μl) for the 3′ end. Reaction conditions were 20 sec at 94°C, 2 min at 45°C and 3 min at 72°C. After completion, the new primer pair M102 and M45 (20 pmoles each) was added together with 2U of fresh AmpliTaq$^R$ and PCR was done for 50 cycles using the following conditions: 94°C for 30 sec, 55°C for 2 min and 72°C for 3 min with 5 sec segment autoextension (Thermal Cycler, Perkin-Elmer-Cetus). 30 μl samples were analysed by 1.2% agarose gel electrophoresis and Southern blotting.

### Oligonucleotides

Oligonucleotides were synthesized by phosphoramidite chemistry, using a 391 PCR-Mate DNA Synthesizer (Applied Biosystems Inc.). Sequences of oligonucleotides used in the experiments: M101 5′-dGTAGTGTGTGCCCGTCTGTTG-3′, (99−119); M102 5′-dTGTGACTCTGGTAACTAGAGA-3′, (120−140); M108 5′-dGTTGACAGGTGACAGAGACAGT21−3′; M45 5′-dGACAGGTGACAGAGACAGTT-3′; M103 5′-dCCTCA-GACCCTTTTAGTCAGT-3′, (143−163); M177 5′-dCGACG-CAGGACTCGGCTTGCT-3′, (233−253); M185 5′-dATCA-AGGAT(CG)TCTTGTCTTCTTT-3′, (8707−8685*); M186 5′-dTCTTTCCCTTA(TC)AGCAGGCCATC-3′, (8431−8409*); M187 5′-dCTCCATGTTTTTCCAGGTCT-3′, (8498−8479*); M239 5′-dGTGTATGCCACGCGCTCCT-3′; M240 5′-dCAGGACTCGGTGGACTTCTC-3′; M291 5′-dACTCTT-TGTTACATCAGGTATTGC-3′; M292 5′-dCCAACGTGTC-TAGTTCCATCA-3′; M238 5′-dGTTGTTGGATACTTGCT-GGA-3′. The coordinates listed in parentheses are from HTLV III/LAV (15). Coordinates with asterisks indicate consensus sequences of 6 published HIV-1 sequences.

### Cloning and sequencing

Blunt-ended, phosphorylated amplified products (amplicons) were cloned into the dephosphorylated Sma I site of a modified pSP64 vector. Inserts were sequenced using the dideoxy-chain termination method (Sequenase$^R$ Version 2.0 (USB)). Sequence analysis of b2/*Hin-1* was done for both strands.

## RESULTS

### Selective detection of chimeric transcripts

In order to detect chimeric provirus-host gene transcripts, a reusable pool of cDNAs ('library') was generated from mRNA of the HIV-1 infected $CD4^+$ T-cell line H9 using a technique described recently (see ref. 14 and Fig. 1). PCR amplification was performed using 5′ primers specific to the U5 region in the viral long terminal repeat (LTR) and 3′ primers hybridizing to poly(A). Analysis of such amplicons revealed a band of 1.7 kb. This DNA represents the amplicon of the short, multiply spliced, viral mRNA (15,16) (Fig. 2b, lanes 2&3). Even when using highly stringent annealing conditions and avoiding 'cold DNA polymerization' (17), smears of variably sized amplicons were readily detectable and suggested the presence of chimeric proviral-host gene sequences. To enrich for such sequences,

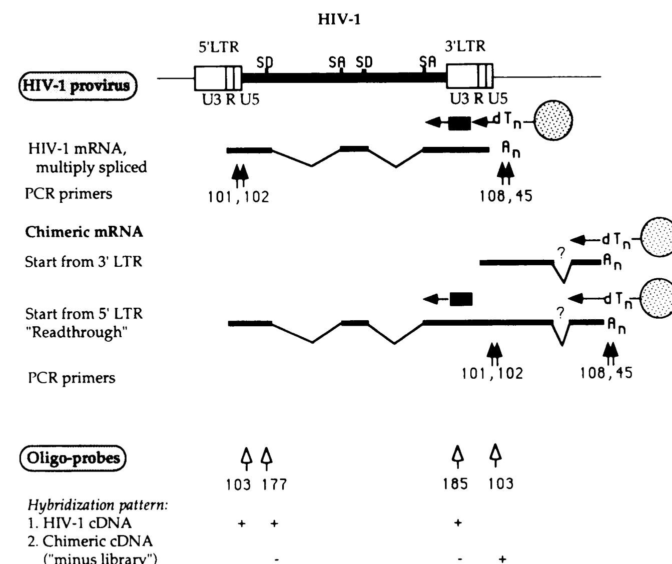


**Figure 1.** Detection of chimeric transcripts. Scheme representing HIV-1 provirus with two long terminal repeats (LTR) inserted into host DNA (thin line) and three types of polyadenylated ($A_n$) transcripts with the corresponding cDNA (horizontal arrows) coupled to magnetic beads (stippled circles). The first mRNA shown belongs to the class of short transcripts which are multiply spliced according to splice donors (SD) and acceptors (SA) in the proviral DNA (not at scale). The detection of singly spliced (4 kb) or full length viral mRNA (9 kb) is beyond the limit of this PCR. The second and third type of mRNA shown are chimeric viral-host transcripts starting either from the 3′ LTR promotor region or as 'readthrough' transcripts from the 5′ LTR. Question marks indicates the unknown splicing pattern of chimeric transcripts. Primers used for single-sided PCR are indicated by vertical arrows and numbers (without the preceeding M symbol used in legends and text). The same upstream and downstream primers can amplify the cDNA of multiply spliced HIV-1 transcripts as well as cDNA of putative chimeric viral-host transcripts. The cDNA of the multiply spliced viral and readthrough transcripts are shown with two types of primers for the reverse transcription: the oligo-(dT) coupled to the beads and the U3 LTR/nef primers M185, M186 and M187 (black solid box) which are unlinked and together with the cDNA primed will become detached during the melting step leading to the 'minus library' (see Methods and Fig. 2a). The amplicons are analysed by oligonucleotide probes indicated by open vertical arrows. M103 corresponds to a sequence downstream from M102 in the viral U5, M177 is located upstream from the first splice donor (SD) and M185 corresponds to a sequence in the viral U3/*nef* gene. The expected hybridization pattern for colonies containing cloned HIV-1 transcript sequences would be positive with all three probes whereas cloned chimeric transcripts will be negative for the M177 and M185 probes but positive for M103.

amplicons representing normal HIV-1 transcripts were eliminated. To prepare such HIV 'minus libraries', we used a competitive reverse transcription strategy (see legend to Fig. 2), which resulted in an essentially complete depletion of amplicons representing multiply spliced HIV-1 transcripts (Fig.2b, lanes 5&6). In contrast, the smear representing other amplified cDNAs was still present. These results also show that the smears do not represent breakdown products of the 1.7 kb HIV-1 amplicons. Furthermore, these amplicons are not the result of contaminating cellular or viral DNA since their presence was dependent on reverse transcription of mRNA (data not shown). Finally, Southern blot analysis showed that viral sequences in the smear of HIV 'minus library' amplicons (Fig. 3b&c, lanes 5&6) hybridized only with the U5-specific oligoprobe M103 (Fig. 3b) but not with the U3-specific probe M185 (Fig. 3c). These results also suggest that the smear shown in Fig. 2b results, at least in part, from an amplification of chimeric proviral-host cell RNA templates.

## Chimeric proviral-host gene transcripts in HIV-1 infected H9 T-cells

Amplicons were cloned and oligonucleotide hybridization analysis revealed that 76 of 110 clones exhibited a pattern reminiscent of chimeric transcripts. These amplicons scored positive with the U5 probe M103 but negative with a probe recognizing a region in the leader sequence of gag, 5′ of the first known splice donor (M177) as well as with the U3 probe M185 (Fig. 1a). 13 clones were sequenced (Tab. 1a) and in all cases, the transition from viral to host sequences could be defined to one of four bases at the border of the integrated U5, marked by a CA dinucleotide (18). The absence of *gag* leader sequences further confirmed that the cloned transcripts contained U5 sequences of the 3′ LTR. The sizes of the amplified cDNAs ranged from few hundreds to more than one thousand base pairs (Tab. 1a). Most (10/13) of these cloned cDNAs exhibited a 3′ end with the consensus polyadenylation signal AATAAA (Table 1a). Three of 13 cDNAs showed a G or T in the second position. The sequence of clone f1 was represented twice as evident from sequencing an additional 10 clones, all exhibiting a viral/host breakpoint at the correct end of the viral U5 region (data not shown).

## Clone b2 exhibits an open reading frame

One of these clones, b2, has an open reading frame (ORF) of 435 bp followed by a 3′ untranslated region of 361 bp and a putative polyadenylation site (EMBL data base accession no. X68242). The sequence surrounding the AUG codon at position 128, suggests that it may function as an initiation codon (19). Sequence search by FASTA (20) using the Swissprot database revealed a a significant homology of the peptide of 145 amino acids to the product of the ovarian tumor gene (*otu*) of Drosophila melanogaster (21). The two protein sequences display an identity of 32.5%, with a similarity of 77.2% due to conservative amino

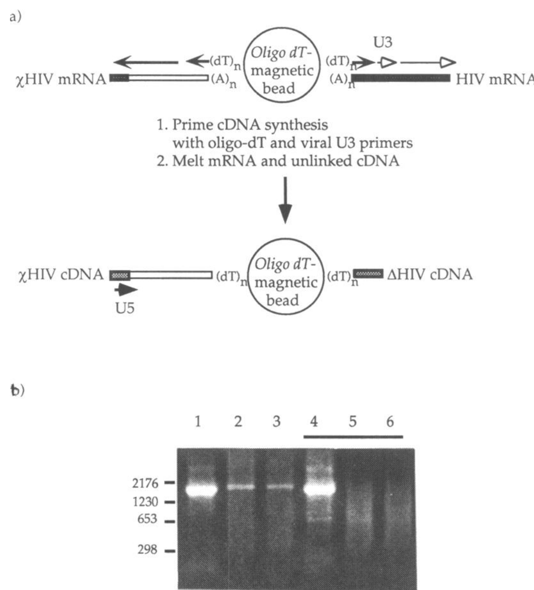


**Figure 2. a)** Schematic diagram of establishing a 'minus library'. Shown in the upper half are two polyadenylated transcripts attached to oligo-dT on a magnetic bead. HIV-1 sequences are stippled. Filled arrows mark the direction of cDNA synthesis, open arrows indicate the co-priming by U3-specific oligonucleotides on regular HIV mRNA. The lower half shows the resulting cDNA linked to the solid phase. The cDNA of regular HIV mRNA is truncated to the point where U3 oligonucleotides have primed in (ΔHIV). Chimeric cDNA (χHIV) can further be amplified by using 5′ primers specific to viral U5 and 3′ oligo-(dT) anchor primers. **b)** Agarose gel showing PCR analysis of cDNA from HIV-1 infected H9 cells. HIV-1 cDNA from unmodified cDNA pools (lanes 2&3) and 'minus libraries' (lanes 5&6) were amplified with the nested upstream primers M101 and M102 and the downstream oligo-(dT) anchor primer M108 and anchor oligonucleotide M45 (lane 2,3, 5,6). Vimentin gene expression (lanes 1&4) was used as control and amplified with nested 5′ primers M239 and M240 and the 3′ anchor primers M108 and M45, resulting in a fragment of 1.5 kb (39). Sizes of DNA markers are given to the left, in base pairs.

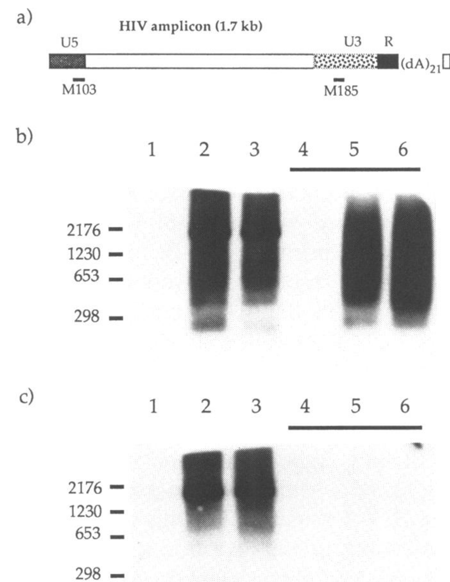


**Figure 3. a)** Scheme of the amplicon generated from the multiply spliced HIV-1 transcript. Bars indicate the regions recognized by oligonucleotide probes M103 and M185. **b)** Southern blot of the gel shown in Fig. 2b and hybridization analysis with $^{32}$P-labeled oligonucleotide probes M103 and c) M185. The filters were hybridized for 2 hr at 50°C and washed at a final stringency of 2.5×SSPE, 0.1% SDS for 10 min at 50°C and exposed to X-ray film for 30 min. Lanes 1−3 show the amplication of a unmodified cDNA pool, lanes 4−6 the amplification of HIV-1 'minus libraries'.

**Table 1.** Sequences of the chimeric transcripts

| | Clone number | HIV-1 U5 / host junction | cDNA 3'end (host) | size(bp) |
|---|---|---|---|---|
| a) | H9-HIV | | | |
| 1. | a1 | *agtgtggaaaatctctagca*atcatattcttaactgataag.. | ..aagagtttatggagaattataataaataatattttgg$(a)_n$ | 330 |
| 2. | a6 | *agtgtggaaaatctctagca*aaaaaagaagcacagtttttc.. | ..gaataaagtttaaaatatgtggactgttcccaaatac$(a)_n$ | 400 |
| 3. | a8 | *agtgtggaaaatctctagca*ttgcagtgggagcttcttgct.. | ..atttaaataaagcattatcccattgttacttctttt$(a)_n$ | 370 |
| 4. | b2 | *agtgtggaaaatctctagcagt*ttacactaagaactctttg.. | ..ccagtaatgaaataaaatggccgtttggatttccttc$(a)_n$ | 973 |
| 5. | b3 | *agtgtggaaaatctctagca*gcaaaataaattaactaatga.. | ..ttaagtgaaagtgaggaaataaagttaagtttgtttt$(a)_n$ | 1130 |
| 6. | b5 | *agtgtggaaaatctctagcagt*tataataaactatattta.. | ..atgacacttataaatcctaataaactgttaaacccac$(a)_n$ | 1080 |
| 7. | b8 | *agtgtggaaaatctctagcagt*tatacacttcaactttcaa.. | ..gcatgtaagctaatatgtcagtaaagcattccaaagc$(a)_n$ | 160 |
| 8. | d10 | *agtgtggaaaatctctagca*cttatatcagcatacttagcg.. | ..ctgagaattccaaaattaaacaacatatgttgtctgt$(a)_n$ | 500 |
| 9. | e1 | *agtgtggaaaatctctagcagt*gggaaccgctgcagggtaa.. | ..aaaaggatattattaataaaacacacctttcttttcc$(a)_n$ | 870 |
| 10. | e5 | *agtgtggaaaatctctagca*ttaagcacatggcttttgaaa.. | ..atcatcaggtaccaagaagtaaacaattgttggcatt$(a)_n$ | 520 |
| 11. | f1 | *agtgtggaaaatctctagcagt*aacttaatcacatcctcct.. | ..tacatgttcatgttcagtaaataaacaattttaagt$(a)_n$ | 320 |
| 12. | f4 | *agtgtggaaaatctctagca*gcatacatatttggagaagaa.. | ..aaagcagttagagtcttaaataaaacaacttttatg$(a)_n$ | 320 |
| 13. | f10 | *agtgtggaaaatctctagca*ctgtcttaaggaggaagaagg.. | ..tttcatgtgattaacataataaaaactttgcgaagtc$(a)_n$ | 1800 |
| b) | PBMC-HIV | | | |
| 1. | 9a | *agtatggaaaatctctagca*gacggagytcgctctgtaccc.. | ..aatgtgaattttacctcaataaacagcatgtaaatg$(a)_n$ | 850 |
| 2. | 11c | *agtgtggaaaatctctagca*gcttacaataactatagtttt.. | ..ttcttacctacaaattaaaatataggtttgtaacat$(a)_n$ | 350 |
| 3. | 8d | *agtgtggaaaatctctagcagt*ttcttgtgggagttttgt.. | ..tagtttataccaataaaggattgtttggataattat$(a)_n$ | 380 |
| c) | BAL/AIDS | | | |
| 1. | PN5 | *agtgtggaaaatctctagca*agctttggcttcccaaagtgc.. | ..taatagtagctaaataaacaagctctacttaaaaag$(a)_n$ | 144 |
| 2. | PN10 | *gtgtggaaaaatctctagca*caaaccagagactagagagca.. | ..gtcctaatgattcgctcaataaacatgtttgaatcc$(a)_n$ | 1100 |

HIV-1 consensus:*

*agtgtggaaaatctctagcagtggcgcccgaacagggacct*

*Proviral sequences are in italics.

*Proviral sequences are in italics.

acid changes in a 123 amino acid overlap. This putative gene locus was tentatively called *Hin-1* (HIV-1 integration 1).

PCR amplification of Hin-1 cDNA with 5′ and a 3′ *Hin-1* specific primers generated a 246 bp fragment. In contrast, the amplification of genomic DNA resulted in a band of about 800 bp suggesting that *Hin-1* expresses a spliced transcript (Fig. 4a). To clarify whether the insertional activation of *Hin-1* is unique for HIV-1 infected H9 cells and whether *Hin-1* is expressed in uninfected H9 cells and/or activated PBMC, cDNAs were prepared and amplified with a set of U5 and *Hin-1* specific primers or with *Hin-1* primers only. Bands of the expected sizes of 297 bp and 246 bp, respectively, were detected only in HIV-1 infected (Fig. 4b, lanes 8&9), but not in uninfected H9 cells (Fig. 4b, lanes 2&3) or allogeneically stimulated PBMC (Fig. 4b, lanes 5&6). Therefore, the expression of *Hin-1* is due to activation by HIV-1 insertion. Expression of *Hin-1* in the infected H9 culture was observed for a period of at least six months indicating that the insertionally tagged cell clone was stable in culture.

Finally, we tested whether the *Hin-1* mRNA starts from the 5′ LTR of the HIV-1 provirus. For this purpose, PCR analysis was performed using primers specific to a sequence in U3 (3′ of the recognition site of oligonucleotides M185, M186 & M187), U5 and *Hin-1*. Whereas U3→*Hin-1* PCR was negative, U5→*Hin-1* PCR gave amplicons of the expected size (data not shown). Therefore, these results suggest that the chimeric mRNA initiated at the 3′ LTR of HIV-1.

## Chimeric transcripts in freshly infected peripheral blood mononuclear cells

To analyse whether cells, *de novo* infected with a distinct HIV-1 isolate, also produce chimeric mRNAs, plasma from a newborn child who died of AIDS was used to infect normal peripheral

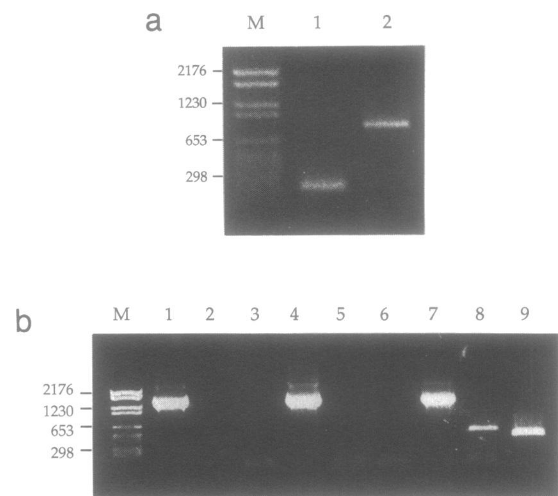


**Figure 4. a)** Comparison between cDNA and genomic DNA of *Hin-1* by PCR analysis. Primer M291 (sense orientation) and M292 (antisense orientation, 25 pmoles/100 µl each) were used to amplify the cloned chimaeric cDNA of *Hin-1* (lane 1) and a sample (1µg) of human genomic DNA (lane 2). Size markers are shown in lane M. **b)** PCR analysis of *Hin-1* transcription. The following primer pairs are used: M291 and M292 for *Hin-1* transcripts (lanes 3,6&9), M102 and M292 for detection of chimaeric *Hin-1* transcripts (lanes 2,5&8) and, as control, M239 and M238 (antisense orientation) for the expression of vimentin (lanes 1,4&7) in uninfected H9 cells (lanes 1−3), in allogeneically stimulated peripheral blood mononuclear cells (lanes 4−6) and in HIV-1 infected H9 cells (lanes 7−9). Agarose gel electrophoresis of the amplification products is shown including standard size markers (lane M). The cDNA of mRNA from cultured cells was isolated and prepared as described in Material and Methods. PCR amplification was done for 50 cycles at the following temperatures: 94°C for 45 sec, 50°C for 2 min and 72°C for 3 min.

blood mononuclear cells. The cultures were subsequently tested for the presence of chimeric mRNA on day 1, 4, 11 and 15. HIV-1 specific transcripts were identified as early day 4 post infection. In contrast, the smear of amplicons indicative of chimeric transcripts was first visible on day 11. Cloning and sequencing further confirmed the chimeric nature of the transcripts (Table 1b). One of these, clone 9a, contained an *Alu* repeat motif (22). Chimeric proviral LTR-*Hin-1* transcripts could not be demonstrated, as expected.

**Chimeric transcripts in cells of AIDS patients**

To investigate the occurrence of chimeric transcripts *in vivo*, cellular RNA from broncho-alveolar exudate cells was analyzed assuming that it would be enriched in HIV-1 mRNAs (23). Samples of three patients were analyzed. HIV-1 DNA sequences specific for the gag region were detected by standard diagnostic PCR in 2 of 3 samples and for the pol region in all of them. Chimeric transcripts were found in only 1/3 samples (Table 1c, clones PN5, PN10). While amplicons representing multiply spliced HIV-1 transcripts (2 kb class) were lacking in this case they were found in another sample (1/3) (data not shown).

## DISCUSSION

We described a PCR method allowing a selective amplification of cDNAs representing chimeric provirus-host gene transcripts. We have used this strategy to identify HIV-1 LTR promotor mediated activation of host gene transcription. The selectivity of the method relies on the exclusion of HIV-1 transcripts from the cDNA pool. The procedure for reverse transcription of mRNAs uses co-priming with oligo(dT) linked to magnetic beads and HIV-1 U3 region specific oligonucleotides lacking beads. Whereas the oligo(dT)-primed cDNAs are coupled to the solid phase during the procedure of cDNA synthesis, co-primed HIV-1 cDNAs and mRNAs are unlinked and subsequently removed by melting. The efficiency of this step is probably dependent on parameters including the sequence variability and the secondary structure of the primed mRNA. Priming in the stem of stem-loop structures is known to be very inefficient. Therefore, the analysis of the mRNA for putative secondary structure formation using a computer program like FOLD (24) might be helpful.

'Chimeric PCR' can be used to study cis-activation either by endogenous or other exogenous retroviruses. Also, it might be used for the detection of rare alternate splicing forms as well as for the identification of an unknown 3′ joining partner in chromosome translocations, provided that the length of the novel chimeric RNA does not exceed several kb. 'Chimeric PCR' might also be used to elucidate other insertional mutagenesis events such as those caused by vectors intended to be used in gene replacement therapy or in endogenous vaccination.

We have shown that the method is useful for detecting chimeric mRNAs in chronically as well as *de novo* HIV-1 infected cells in culture, and the analysis of cells obtained from broncho-alveolar washes of patients with AIDS. In one case we showed that a chimeric transcript exhibited an open reading frame and potentially encodes a protein of 145 amino-acids (clone b2/*Hin-1*). This observation extends the general finding of provirus-mediated host gene activation events to the family of lentiviruses, in particular, HIV-1.

The chimeric nature of the transcripts was confirmed by sequencing the viral-host cDNA junction. The 3′ boundary of the integrated HIV-1 proviral DNA is delineated by the conserved dinucleotide CA and a direct repeat of 5 base pairs of host cell sequences (15). In two third of the cases (12/18), the 3′ boundary was marked by this CA dinucleotide. In 6/18 junctions, however, we found that the integration did not result in the loss of the two bases GT that are also characteristically found in unintegrated proviral HIV-1 DNA (25). Sequencing of the 3′ ends of the cloned amplicons revealed that all transcripts were correctly polyadenylated. Fourteen out of 18 transcripts revealed a consensus polyadenylation signal (AATAAA), whereas the other 4 transcripts exhibited a single base-pair change in the second position of the consensus signal. Such changes are found in 12% of the mRNA species in vertebrates (26).

*In vitro*, the frequency of chimeric HIV-1 transcripts is notable and their overall number seems comparable with the total amount of HIV-1 2kb transcripts. In patient samples, we only sporadicly detected chimeric transcripts. This is expected and in accordance with a relatively low overall number of HIV-1 expressing cells in AIDS patients. Moreover, the findings in one patient lacking multiply spliced HIV-1 transcripts but containing chimeric HIV-1 transcripts suggest that the chimeric proviral-host gene transcripts might be frequently produced *in vivo*.

The synthesis of chimeric transcripts may rely on a transcriptional activation from the 3′ LTR. In the intact provirus, however, there is a clear dominance of the 5′ promotor. This is ascribed either to internal viral sequences controlling LTR activity or to transcription from the 5′ LTR interfering with transcription from the 3′ LTR. Deletions and mutations are capable to render the 5′ LTR transcriptionally silent. This would prevent promotor interference (27) and enhance the possibility that transcription starts from the 3′ LTR promotor into adjacent cellular sequences. Hence, the intrinsic variability of the HIV-1 genome (28) may actually frequently result in a 3′ LTR promotor activity. Alternatively, when readthrough transcription occurs, the induction of chimeric transcripts would not rely on an activity in the 3′ LTR. Readthrough transcription occurs and depends on inefficent cleavage and polyadenylation of viral transcripts in the 3′ LTR as recently shown in avian leukosis virus infected cells (29). Although, HIV-1 has an efficient polyadenylation signal in the 3′ LTR, mutations in the signal can provoke readthrough transcription (30,31). Finally, another form of readthrough transcription may occur and involves alternative splicing of a viral splice donor to a host splice acceptor. Our used protocol precludes the finding of such transcripts, since the colony screening of cloned amplicons with different oligonucleotide probes excluded clones containing viral 5′ leader sequences (see Fig. 1).

Many of the isolated chimeric mRNAs seem 'sterile'. The viral part (R-U5) of the transcript lacks an initation codon and ORF (15). Therefore, translational regulatory sequences have to be derived from the host gene in a chimeric transcript. Only 1 out of 18 transcripts (b2/*Hin-1*) displayed such an ORF and provided a start codon that could allow efficient translational initation. This transcript was not detected in uninfected H9 cells and was induced specifically by insertional activation from the HIV-1 3′ LTR. Its potential to encode a gene was suggestive given a similarity with the ovarian tumor gene (otu) of the fruit fly *D.melanogaster*. Also, the results of Fickett's statistic method (32) that identifies protein coding sequences by measure of non-randomness of composition at every third base are in agreement with *Hin-1* being a putative cellular protein (data not shown). Computer analysis

for a variety of specific domain consensus sequences revealed only potential protein kinase C and casein kinase II phosphorylation sites in the C-terminal region of *Hin-1*. It should be noted that the 5′ boundary of the Hin-1 gene is not yet known. Hence, *Hin-1* might have been truncated by HIV-1 insertion.

Whether transcriptional activation and/or truncation of cellular gene products as a result of HIV-1 integration could contribute to pathogenesis, remains to be seen. Aberrant expression of cellular genes has been postulated as a factor that contributes to the complex mechanism by which HIV-1 causes pathogenesis (33–36). Abnormal expression of such proteins or protein epitopes might well contribute to immunological abnormalities in HIV-1 infected individuals (37,38), in particular, if a superantigen encoding protein would be activated.

In any case, our method allows a detailed survey of HIV-1 integration-mediated transcriptional activation events and this may help in elucidating some of the pathogenetic potential of HIV-1 insertional gene activation.

## ACKNOWLEDGEMENTS

We thank A.Wodnar-Filipowicz, H.van der Putten and H.P.Saluz for reviewing the manuscript, P.Erb, C.Moroni, A.Wallace and K.Wiebauer, for helpful discussions, C.Rudin and M.Soler for providing HIV-1 containing plasma and material of broncho-alveolar lavages, respectively, and F.Fäh for creating useful devices.

## REFERENCES

1. Van Lohuizen, M. and Berns A. (1990) *Biochim Biophys Acta* **1032**, 213–235.
2. Peter G. (1990) *Cell Growth Diff.* **1**, 503–510.
3. Kung, H.-J., Boerkoel, C. and Carter T.H. (1991) *Current Topics Microbio. & Immunol.* **171**, 1–25 .
4. Teich, N., Wyke, J., Mak, T., (1982) In Weiss, R., Teich, N., Varmus, H. and Coffin, J. (eds.), RNA tumor viruses. Cold Spring Harbor Laboratory, New York, pp 785–998.
5. Tsichlis, P. N. and Lazo, P.A. (1991) *Current Topics Microbiol. & Immunol.* **171**, 95–172.
6. Dhar, R., McClements, W.L., Enquist, L.W. and Vande Woude, G.F. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 3937–3941.
7. Mooslehner, K., Karls, U. and Harbers, K. (1990) *J. Virol.* **64**, 3056–3058.
8. Rohdewohld, H., Weiher, H., Reik, W., Jaenisch, R. and Breindl, M. (1987) *J. Virol.* **61**, 336–343.
9. King, W., Patel, M.D., Lobel, L., Goff, S.P. and Nguyen-Huu, M.C. (1985) *Science* **228**, 554–558.
10. Shih, C.C., Stoye, J.P. and Coffin, J.M. (1988) *Cell* **53**, 531–537.
11. Hayward, W.S., Neel, B.G. and Astrin, S.M. (1981) *Nature* **290**, 475–480.
12. Pauza, C.D. and Galindo, J. (1989) *J. Virol.* **63**, 3700–3707.
13. Popovic, M., Sarngadharan, G., Read, E. and Gallo R.C. (1984) *Science* **224**, 497–500.
14. Raineri, I., Moroni, C. and Senn, H.-P. (1991) *Nucleic Acids Res.* **19**, 4010.
15. Muesing, M.A., Smith, D.H., Cabradilla, C.D., Benton, C.V., Lasky, L.A. and Capon, D.J. (1985) *Nature* **313**, 450–58.
16. Schwartz, S., Felber, B.K., Benko, D.M., Fenyö, E.-M. & Pavlakis, G.N. (1990) *J.Virol.* **64**, 2519–2529.
17. Mullis, K.B. (1991) *PCR Methods Applic.* **1**, 1–4.
18. Kulkowsky, J. and Skalka, A.M. (1990) *J. AIDS* **3**, 839–851.
19. Kozak, M. (1991) *J. Cell Biol.* **115**, 887–903.
20. Pearson, W.R. and Lipman, D. J. (1988) *Proc. Natl.. Acad. Sci. U.S.A.* **85**, 2444–2448.
21. Steinhauer, W.R., Walsh, R.C. and Kalfayan, L.J. (1989) *Mol. Cell. Biol.* **8**, 1481–88.
22. Weiner A.M., Deininger P.L. and Efstratiadis A. (1986) *Ann. Rev. Biochem.* **55**, 631–661.
23. Meltzer, M.S., Skillman, D.R., Hoover, D.L., Hanson, B.D., Turpin, J.A., Kalter, D.C. and Gendelman, H.E. (1990) *Immunol. Today* **11**, 217–223.
24. Zucker, M. and Stiegler, P. (1981) *Nucl. Acids Res.* **9**, 133–148.
25. Kulkosky, J., Katz, R. and Skalka, A.M. (1990) *J. AIDS* **3**, 852–858.
26. Wickens, M. and Stephenson, P. (1984) *Science* **226**, 1045–1051.
27. Boerkoel, C.F. and Kung, H.-J. (1992) *J. Virol.* **66**, 4814–4823.
28. Hahn, B.H., Shaw, G.M., Taylor, M.E., Redfield, R.R., Markham, P.D., Salahuddin, S.Z., Wong-Staal, F., Gallo, R.C., Parks, E.S. and Parks W.P. (1986) *Science*, **232**, 1548–1553.
29. Swain, A. and Coffin, J.M. (1992) *Science* **255**, 841–845.
30. Cherrington, J. and Ganem D. (1992) *EMBO J.* **11**, 1513–1524.
31. Weichs an der Glon, C., Monks, J., Proudfoot, N.J. (1991) *Genes & Dev.* **5**, 244–253.
32. Fickett, J. (1982) *Nucleic Acids Res.* **10**, 5303–5318.
33. Nakamura, S., Salahuddin, S.Z., Biberfeld, P., Ensoli, B., Markham, P.D., Wong-Staal, F. and Gallo, R.C. (1988) *Science* **242**, 426–430.
34. Pulliam, L., Herndier, B. G., Tang, N.M. and McGrath, M.S. (1991) *J. Clin. Invest.* **87**, 503–512.
35. Laurence, J., Kulkosky, J., Dong, B., Early, E., Snyderman, R. and Cianciolo, G. (1990) *Cell. Immunol.* **128**, 337–352.
36. Locksly , R.M., Crowe, S., Sadick, M.D., (1988) *J. Clin. Invest.* **82**, 2097–2105.
37. Stricker, R.B., Abrams, D.I., Corash, L. and Shuman, M.A. (1985) *N. Eng. J. Med.* **313**, 1375–1380.
38. Via, C.S. and Shearer, G.M. (1989) Curr. Opin. Immunol.1, 753–756.
39. Honoré, B., Madsen, P., Basse, B., Andersen, A., Walbum, E., Celis, J.E. and Leffers, H. (1990) *Nucleic Acids Res.* **18**, 6692.