The structure of HIV-1 genomic RNA in the gp120 gene determines a recombination hot spot in vivo

Román Galetto, Abdeladim Moumen1, Véronique Giacomoni, Michel Véron, Pierre Charneau2, and Matteo Negroni3

Unite de Régulation Enzymatique des Activités Cellulaires CNRS-URA 2185, and 2 Groupe de Virologie Moléculaire et Vectorologie, Institut Pasteur, 25-28 rue du Dr. Roux, 75724 Paris, France
1 Present address: The Wellcome Trust, Tennis Court Road, Cambridge CB2 1QR, UK
3 Corresponding author

Corresponding author information:
Dr. Matteo Negroni
Unité de Régulation Enzymatique des Activités Cellulaires
Institut Pasteur
25, Rue du Dr Roux
75724, Paris, Cedex 15
France

Tel: (33) 145 688 505
Fax: (33) 145 688 399
e-mail: matteo@pasteur.fr

Running title: HIV-1 recombination in vivo

Abbreviations: RTP: reverse transcription products
Summary

By frequently rearranging large regions of the genome, genetic recombination is a major
determinant in the plasticity of the human immunodeficiency virus type I (HIV-1) population.
In retroviruses, recombination mostly occurs by template switching during reverse
transcription. The generation of retroviral vectors provides a means to study this process after
a single cycle of infection of cells in culture. Using HIV-1 derived vectors, we present here
the first characterisation and estimate of the strength of a recombination hot spot in HIV-1 in
vivo. In the hot spot region, located within the C2 portion of the gp120 envelope gene, the rate
of recombination is up to ten times higher than in the surrounding regions. The hot region
corresponds to a previously identified RNA hairpin structure. While recombination
breakpoints in vivo cluster in the top portion of the hairpin, the bias for template switching in
this same regions appears less marked in a cell-free system. By modulating the stability of this
hairpin we were able to affect the local recombination rate both in vitro and in infected cells,
indicating that the local folding of the genomic RNA is a major parameter in the
recombination process. This characterisation of reverse transcription products generated after
a single cycle of infection provides insights in the understanding of the mechanism of
recombination in vivo, and suggests that specific regions of the genome might be prompted to
different rates of evolution due to the presence of circumscribed recombination hot spots.

Introduction

In their extracellular life, retroviruses store genetic information as two copies of single
stranded RNA molecules. During synthesis of the complementary DNA strand, the reverse
transcriptase (RT) frequently switches template between these two molecules, a process
known as copy choice (1). If the two RNAs are not identical this process can lead to genetic
recombination. The diffusion of the Acquired Immunodeficiency Syndrome (AIDS) pandemic
has provided an alarming indication of the massive impact of recombination on the evolution
of retroviruses. Indeed, by scattering and rearranging mutations already existing along the
HIV-1 genome, recombination stands nowadays as a worldwide hindrance for molecular and serological diagnosis, vaccine development and treatment against AIDS (2,3). Features of the HIV-1 infectious cycle as the poor fidelity of the RT (4), that boosts the probability of co-packaging non-identical viral RNAs, and the dynamic nature of HIV-1 infection, where de novo infection of CD4+ T cells amplifies the possibility of copy choice to occur during the repeated cycles of reverse transcription (5), can explain the extremely high level of genetic recombination documented (6).

Dissecting the mechanisms underlying copy choice in vivo will contribute to the understanding of how recombination drives HIV-1 genome evolution. Several efforts have been made in this sense during the last decade, and various mechanisms have been proposed, based either on the infection of cells in culture (ex vivo systems) (7,8), or on the reconstitution of the process of reverse transcription with purified proteins and nucleic acids (in vitro systems) (9-12). Some hard facts have been jointly established by these two approaches, as the enhancement of template switching observed by decreasing the rate of DNA synthesis (13,14) and the importance of a temporal coupling of RT-encoded polymerase and RNaseH activities (8,15). However, detailed mechanistic models, mostly proposed on the sole basis of in vitro studies, still await an evaluation of their physiological relevance.

In HIV-1 template switching appears as a frequent process. Indeed, ex vivo experiments in HeLa CD4+ infected cells estimated that three events of template switching occur, on average, per replication cycle (16). Furthermore, it has recently been shown that the recombination rates are influenced by the type of cell infected, with the highest rates observed in macrophages (17). A crucial issue to address now is whether this high rate results from the presence of recombination hot spots interspersed among sequences yielding low recombination rates, or if it reflects a nearly constant frequency of strand transfer along the whole genome. Determining the existence of preferential sites for template switching is important since they would constitute ideal sites for the study of the mechanisms of recombination, and they would pinpoint blocks of sequences with relatively independent evolutionary history along the genome. At present, indications on this issue are extremely
scarce. An analysis of ten intra-subtype B recombinant clones generated in cell culture highlighted the presence of a few putative hot spots along the genome (18), but no molecular characterization of these hot spots was carried out. On the other hand, in vitro studies have highlighted the existence of local fluctuations in the frequency of strand transfer on different sequences leading to the proposal of various mechanistic sketches (10-12,19,20). However, no indication concerning the behaviour of these sequences in vivo is available.

Using a cell-free system we previously identified a sequence within the C2 region of the envelope glycoprotein gp120 where strand transfer was up to five times more frequent than in the surrounding regions (21). This behaviour was correlated to the folding of this portion of RNA into a large hairpin, since destabilising this hairpin triggered a fourfold decrease in the rate of template switching (10). Here we have developed an experimental system to study recombination in HIV-1 after a single cycle of infection of cells in culture, and used it to study recombination on the C2 region.

Experimental procedures

DNA constructs
HIV-1 cis-acting regulatory elements (HXB2 strain) were obtained by PCR amplification from pTRIP-GFP plasmid (22) and cloned in a pKS-derived Kn'/origin of replication backbone plasmid (21) following standard cloning techniques. For the study of recombination we have designed two types of plasmids differing in the genetic marker present upstream the sequence where recombination is studied. This marker is therefore transcribed after the studied sequence during reverse transcription. pLac' and pLac' plasmids carry, as genetic markers, either a functional LacZ' gene or a sequence complementary to a portion of the mRNA coding for E. coli malT gene, respectively. These two genetic markers are indicated as lac' and lac' in the schematic representation of these plasmids shown in figure 1B. Sequences used as region of homology were derived from in vitro vectors described in (10), and inserted
after XhoI/BamHI, or XhoI/BspHI restriction. The sequences previously called E2 and G1Eb generated here C2 and Delhp genomic plasmids, respectively. All constructions were verified by sequencing. The transcomplementation plasmids used were pCMVΔR8.2 (23), coding for HIV-1 gag, pol and accessory proteins; and pHCMV-G (24) that carries the VSV envelope protein gene.

Cells

293T cells were grown in Dulbecco's modified Eagle's medium supplemented with 10% foetal calf serum (FCS), penicillin and streptomycin (from Invitrogen), and maintained at 37°C with 10% CO₂. MT4 cells were maintained in RPMI 1640 medium supplemented with 10% FCS and antibiotics at 37°C with 5% CO₂.

Vector particles production and transductions of cells in culture

HIV-1 based vectors were produced by transient transfection of 293T cells with the genomic plasmids, an HIV-1 encapsidation plasmid (pCMVΔR8.2) (23) and a VSV envelope expression plasmid (pHCMV-G) (24) using the calcium phosphate method. Cells were plated at a density of 3.5 x 10⁶ per 100-mm-diameter dish and transfected 16-20 h later. The medium was replaced 8 hours after transfection and the vector supernatants recovered 36 hours later. Non-internalized DNA was removed by treatment of the vector supernatants with DNaseI (1 g/mL in the presence of 1 M MgCl₂) for 30 min at 37°C. The amount of p24 present in supernatants was determined by using the HIV-1 p24 ELISA Kit (Perkin Elmer) and, when necessary, vector supernatants were concentrated by using Centricon® YM-50 centrifugal filter devices (Amicon-Millipore) before transduction. MT4 cells were transduced with 200 ng of p24 antigen per 10⁶ cells (an approximate multiplicity of infection of 20) in 35 mm dishes in a 500μL volume. Two hours post-transduction, the cells were diluted up to a 4 mL volume with supplemented RPMI medium and maintained at 37°C in a 5% CO₂ incubator for 40 h.
Purification, cloning and analysis of HIV-1 reverse transcription products

Reverse transcription products (RTP) were extracted 40 h after transduction by following the method described by Hirt (25): cells were lysed by incubation for 10 min at room temperature in a buffer containing 10 mM TrisHCl pH 8.0, 10 mM EDTA and 0.6% SDS. High molecular weight DNA was removed by precipitation at a high salt concentration (NaCl 1M) during 12-18 h on ice. The lysates were centrifuged at 30,000 rpm for 1 h and the supernatants treated with 100 μg/mL RNaseA for 1 h at 37°C and 100 μg/mL Proteinase K for 3 h at 50°C. After phenol/chloroform extraction, DNA was ethanol-precipitated and purified using the NucleoSpin® Extract clean-up kit (Macherey-Nagel). The purified double stranded DNA was digested with DpnI for 2 h at 37°C prior to PCR amplification (20 cycles) with primers BH and SH (figure 1). The amplified product was purified on agarose gel, digested with SacII and BamHI, ligated into an appropriate plasmid vector and used for transformation of E.coli.

Plating on IPTG/X-Gal containing dishes allowed blue/white screening of recombinant and parental colonies, respectively (10,21). In order to determine the regions where strand transfer occurred, 48 recombinant clones were analyzed by restriction mapping for each assay, as described in the "Results" section.

Estimation of the frequency of recombination

For the generation of heterozygous particles, equal amounts of pLac+ and pLac− plasmids were used in the transfection step. Since the same promoter drives the expression of both types of genomic RNAs, and since the regions involved in their dimerization are strictly homologous between lac+ and lac− RNAs, encapsidation of the two genomic RNA moieties is expected to be random. The presence of equivalent amounts of genomic RNAs was confirmed by slot blot analysis on the viral RNA using specific probes for lac+ and lac− RNA (not shown). As currently assumed, the viral population is predicted to be constituted by 25% lac+/− homozygous vectors, 25% lac−/− homozygous vectors, and 50% heterozygous vectors (lac+/−), according to the Hardy-Weinberg equation (7). Our experimental strategy allows the cloning of products issued from reverse transcription in lac+/− as well as in lac+/- vectors.
Assuming that only one double stranded DNA molecule is produced per each vector particle, one third of the bacterial colonies will be generated by reverse transcription products issued from lac<sup>-</sup> vectors, and will have a white phenotype, introducing a bias in the estimation of the frequency of recombination. Therefore, for the calculation of the frequency of recombination in heterozygous particles, the total number of colonies is multiplied by 2/3. In order to accurately estimate the frequency of recombination, another factor to take into account is the background among the white colonies derived from cloning of cellular DNA co-purified with the reverse transcription products. These colonies are distinguished from those issued from the cloning of RTP by the size of the cloned insert and by their restriction pattern. Typically, 48 white colonies are analyzed in each assay and a correction factor is established, given by \( n/48 \), where \( n \) is the number of colonies resulting from cloning of RTP. In all experiments \( n \) corresponded to approximately half of the colonies analyzed. The frequency of recombination (\( F \)) is therefore given by \( F = b / (2/3[N(n/48)+b]) \), where \( N \) and \( b \) are the total number of white and blue colonies, respectively. The recombination rates per nucleotide (\( f \)) within a given interval (\( i \)) is given by \( f = F(x_i/X)/z \), where \( F \) is as above, \( x_i \) is the number of colonies analyzed where recombination was identified to have occurred within the interval considered, \( X \) is the total number of colonies on which mapping was performed, and \( z \) is the size in nt of the interval. To calculate the frequency of recombination in control samples ("homozygous", see Results section), we only corrected for the occurrence of white colonies derived from the cloning of cellular DNA.

**In vitro recombination assays**

*In vitro* recombination assays were done using the reconstituted system previously developed in our laboratory (21). RNA synthesis was performed as previously described (26). RT purification and activity tests were carried out as described by Canard and colleagues (27). Constructs used for RNA synthesis were generated following standard cloning procedures. Reverse transcription was carried out on the donor RNA (100 mM) in the presence of an equimolar amount of acceptor RNA after annealing an oligonucleotide specifically onto the
donor template. For the experiments with NC (55 aminoacids), the protein was added at a ratio of 1 molecule of NC for 8 nt of total RNA, and incubated for 10 minutes at 37°C. Reverse transcription was started by the addition of HIV-1 RT at a final concentration of 400 nM and carried out for 60 minutes. Synthesis of the second DNA strand, *Bam*HI and *Pst*I digestion, ligation, and *E.coli* transformation were carried out as previously described (21).

**Results**

**Recombination in HIV-1 after a single cycle of infection**

We have developed an experimental system to study HIV-1 copy choice recombination after a single cycle of infection of human cells in culture. VSV envelope-pseudotyped HIV-1 particles are produced by transient co-transfection of the human fibroblast cell line 293T, with two transcomplementation and two genomic plasmids (figure 1A). These particles are perfectly equivalent from the standpoint of the structure of the viral capsid and of the reverse transcription process to wild type HIV-1 viruses (28). Genomic plasmids (figure 1B) will lead to the synthesis of two types of RNAs (lac⁺ and lac⁻ RNAs), both containing the sequences required in cis for their dimerisation, encapsidation and reverse transcription. These RNAs contain a region of homology on which recombination is studied. To this aim, this region is surrounded by genetic and biochemical markers different on the two RNAs: a lac⁺ or lac⁻ marker at the 5' border and the presence or absence of a *Bam*HI recognition site at the 3' end (figure 1B). The co-transfection procedure leads to the generation of homozygous lac⁺/⁺ and lac⁺ vectors as well as heterozygous vectors lac⁺/⁻. To study recombinant products generated in the heterozygous population, the vectors are collected and used to transduce the human T lymphoid cell line MT4. Analysis of the reverse transcription products (RTP) found within these cells will provide a snapshot of the recombinant population generated in the absence of any type of selection. Indeed, this approach limits replication to a single cycle since the RTP will carry a non-functional copy of U3 (∆U3, figure 1C), and will therefore lack a functional
promoter to drive transcription of new genomic RNAs, hampering the generation of viral progeny. The RTP are recovered using the Hirt technique (25) that allows isolation of low molecular weight DNA. Since our genomic RNAs are devoid of FLAP sequence, which enhances nuclear import and integration (ΔFLAP, figure 1B) (22), most RTP will be present in the unintegrated low molecular weight fraction. The RTP are analysed after PCR amplification using primers SH and BH, which allow the simultaneous amplification of parental and recombinant RTP (figure 1C). Prior to cloning in E. coli, the amplified products are digested with BamHI, whose target sequence is present only on lac− RNAs, and with SacII, whose recognition site is carried by SH primer (figure 1C). This procedure allows cloning only of parental lac− products and of recombinant lac+ products shown in figure 1C. The number of lac+ bacterial colonies over the total number of colonies leads to an estimate of the frequency of recombination, as detailed in Experimental Procedures. For each experiment a control sample is run where homozygous lac++ and lac− vectors are produced separately by transfection of 293T cells with transcomplementation plasmids and either pLac− or pLac+ genomic plasmids ("homozygous sample", Table I). In this sample only parental RTP can be generated and, after restriction with BamHI and SacII, only white bacterial colonies should be found (figure 1C). The frequency of blue colonies recovered provides an estimate of the background of artfactual recombinant molecules generated during the experimental procedure.

The potential problem of carrying over the genomic plasmids used for transfection, inherent to the production of vectors by transient transfection, was considered. In fact these plasmids could then be amplified and, in the case of pLac− that contains the BamHI site used for cloning, yield white colonies indistinguishable from those issued from transformation with lac− parental RTP. To circumvent this problem, vector preparations were treated with DNaseI before the transduction step, and the RTP preparation was digested with DpnI before PCR. DpnI is a frequent cutter that selectively digests DNA of bacterial origin. Furthermore, amplification products from RTP and from contaminating plasmids yield DNA molecules of
different size (figure 1D). This difference was exploited to remove potential contaminating products by agarose gel purification prior to restriction with BamHI and SacII.

**Recombination on the C2 region of gp120**

We first produced vector particles where the region of homology between the two genomic RNAs was constituted by the 400 nt sequence depicted in figure 2A. This sequence spans nt 6,639-7,039 numbered according to HXB2 proviral DNA, and includes the portion coding for the C2 region of the gp120 we previously identified as a recombination hot spot in a cell-free system (10). In three independent experiments the average frequency of recombination was 13.4 % (standard deviation: ± 0.4) in "heterozygous" samples, compared to 0.5 % (± 0.2) observed in "homozygous" samples (Table I). An extrapolation of this estimate of 0.13 recombination events to the full-length genome (nearly 10,000 nt long in contrast to the 400 nt region used here) predicts 3.25 recombination events per each infection cycle. This value is in agreement with the estimate of three recombination events per infectious cycle made by Jetzt and colleagues (16).

The possibility that the recombinant molecules identified were generated during the transfection step was ruled out by a control experiment with C2 RNA where lac^- viral particles were produced during co-transfection with an equimolar amount of pCDNA3-derived expression vector carrying the cassette containing the viral region of homology and the LacZ' gene, as in Lac+ plasmids (pCDNA3-vir-Lac+, figure 1E). The rationale of this control is that, if artfactual lac^+ RNAs were generated through recombination at the DNA or at the RNA level within the producer cells, they would be identified by the presence of blue bacterial colonies after transduction of MT4 cells with these viral particles. The vectors generated in this control sample, together with lac^± homozygous particles, were used to co-transduce MT4 cells. This control accounts for artfactual recombination within the producer cells as well as for PCR-driven recombination. The frequency of recombination detected in this control in three independent experiments was 0.37 ± 0.26, a value at least 20 times lower than the frequency found in the heterozygous samples (Table I) indicating that
recombination within the producer cells either at the DNA or at the RNA level does not significantly contribute to the generation of artfactual recombinant molecules.

To map the positions of strand transfer along the region of homology, point mutations generating specific restriction sites were introduced in the region of homology either on the donor or on the acceptor RNA, defining sub-regions named R1 to R5 following the sense of minus DNA strand synthesis during reverse transcription (figure 2A). This allows determining a recombination rate per nt for each individual region after restriction analysis of the cloned RTP, as described in Experimental Procedures. Regions R1, R2, and R3 participate in the formation of the hairpin previously identified by in vitro probing (figure 2B). Three independent experiments yielded consistent results (figure 2C) with the upper part of the C2 hairpin (R2) standing out with respect to the other regions. The recombination rate per nt in R2 was $13.3 \times 10^{-4}$ ($\pm 2.6 \times 10^{-4}$), a value significantly higher than those found in the other regions (chi-squared test: $p \leq 0.001$). In contrast, when the few blue colonies obtained in the homozygous control samples were analysed, the distribution of breakpoints appeared random (Table II).

Copy choice and structure of the genomic RNA

In order to check whether the high recombination rate found in this region was due to its folding, a series of genomic RNAs were constructed where the structure of the region embedding R2 was changed without altering the primary structure of R2 itself (figure 3A). As for the experiments with C2 RNA, point mutations distinguishing the lac- from lac+ RNAs allow mapping the position of strand transfer, and determining the recombination rate within each of the regions. We previously showed that strand transfer in vitro was decreased by a fourfold factor in the hairpin when regions R1 and R5 were replaced by two sequences derived from the gag gene of the LAI isolate (10). This was attributed to the lack of complementarity between the sequences that constituted the base of the C2 hairpin that, as determined by in vitro probing experiments, abolished the formation of this hairpin (10). Here, we followed the same strategy generating Delhp genomic RNA (for "deleted hairpin",...
figure 3B), where the C2 hairpin is no longer present. In vectors carrying Delhp genomic RNA, R2 was no longer a hot spot (figure 3C, chi-squared test: \( p \leq 0.95 \)), indicating that the presence of the hairpin is important for recombination also in infectious particles. To test this hypothesis we replaced the 25 nt of R3 with the 25 nt in blue in figure 3B, which are expected to regenerate a structure similar to the C2 hairpin (SL RNA, for "Stem Loop", figure 3B). On SL RNA, the frequency of recombination in R2 was again significantly higher than in all the other regions (figure 3C, chi-squared test: \( p \leq 0.001 \)), strongly suggesting that the restoration of the hairpin re-established the presence of the hot spot. The predicted stability of the hairpin in SL is significantly higher than in C2 (\( \Delta G = -46.1 \) kcal/mol for SL, and \( \Delta G = -25.0 \) kcal/mol for C2). To address the question of the influence of the stability of this hairpin on recombination, we constructed 2b RNA (for "2 bulges") by introducing in SL the mutations shown in orange in figure 3B. These base substitutions are expected to generate a hairpin with a lower stability than SL (\( \Delta G = -28.2 \) kcal/mol), and similar to the one of C2. Furthermore, also the shape of the 2b hairpin better resembles the one of C2 with the presence of bulges in the lower part of the stem. With this new RNA, recombination in R2 almost doubled with respect to SL RNA, with more than 50% of the total recombinant molecules mapping in R2 (figure 3C, chi-squared test: \( p \leq 0.001 \)).

Copy choice in vitro

A salient feature of our cell culture experimental system is to allow a strict comparison with the system we previously developed for the study of recombination in vitro (21). In the cell free assay, outlined in figure 4A, reverse transcription is primed on the donor RNA in the presence of the acceptor RNA. As for the assay in cell cultures, template switching on the region of homology is monitored by a genetic screening after cloning the products of reverse transcription, as described in the legend of figure 4A. Analysis of recombinant products is also carried out as for the ex vivo approach. We first used the same region of homology of C2 genomic RNA. The presence of distinctive point mutations on the donor and acceptor RNAs in this region of homology allows a more precise mapping of the hot region than the one we
previously performed on C2 in vitro (10). As shown in figure 4B (pale blue) and in Table III, R2 constituted the preferential region for strand transfer also in this case, although at a lesser extent than ex vivo (chi-squared test: p ≤ 0.05). When the SL sequence was used, instead, a manifest inhibition was observed at the level of R2 (figure 4B, deep blue). To investigate whether the low rate observed in this region resulted from the high stability of the SL hairpin, we ran the assay in the presence of the RNA chaperone NC protein, known to favour breathing of highly stable RNA hairpins (29,30). The global frequency of strand transfer was increased in this case, with the most pronounced enhancement observed for R2 (figure 4B, yellow), although this region still clearly did not constitute a hot spot (chi-squared test: p ≤ 0.95). Similarly, with 2b RNA, where the stability of the hairpin was decreased by means of the introduction of point mutations, a manifest enhancement was observed in R2 (figure 4B, red), although the recombination rate in R2 was not significantly higher than in the other regions (chi-squared test: p ≤ 0.95).

Discussion

Using a new system to study recombination in HIV-1 based vector particles after a single infectious cycle, we identify and characterise a recombination hot spot (R2) within the C2 region of the gp120 gene of HIV-1. The rate of recombination in R2 is up to ten times higher than in the surrounding regions, providing the first estimate of the strength of a recombination hot spot in HIV-1 infected cells and suggesting that different portions of the genome can be prompted to different rates of evolution.

The hot region maps in the upper portion of a hairpin structure we previously identified in vitro (10). By using a series of genomic RNAs where the stability of this hairpin was varied without altering the R2 sequence, we show that the folding of the genomic RNA is crucial for the existence of this hot spot. In fact, in all cases where a large hairpin was expected to be present (C2, SL and 2b RNAs), the rate of recombination in R2 clearly stood
out from that of the other regions (figures 2C and 3C). In contrast, on an RNA devoid of a stable hairpin in R2 (Delhp RNA), the rate of transfer in this region fell almost into the background of the other sequences (figure 3C). The wild type (C2) and 2b RNAs, which display a similar stability in the hairpin region, were the optimal substrates for recombination. Limited base substitutions expected either to reduce or to increase the stability of the 2b hairpin, as the seven bases changed in R1 to generate SL hairpin or the replacement of the R3 region in Delhp RNA, both led to a significant decrease in the efficiency of strand transfer specifically in R2 (figure 3B). Altogether these results strongly suggest the existence of an optimal window of stability for R2 to constitute a hot spot, as illustrated by the plot presented in figure 5. Obviously other factors might also contribute to the high rate of recombination observed in this region, as the presence, location and size of bulges in the stem portion of the hairpin, the presence of pause sites of reverse transcription, or other features of the R2 sequence itself. These factors might also explain the slight residual preference for recombination in R2 on Delhp RNA (figure 3C).

This work also provides the first comparison of recombination products generated on the same sequences in a reconstituted in vitro system and after infection of cells in culture. The overall comparison of the results obtained with the two systems suggests that the results on copy choice obtained in cell free systems should be transposed to the situation in vivo with extreme caution. Indeed, significant discrepancies are observed between the two systems. A most striking one is constituted by the distribution of breakpoints, predominating in R2 in infected cells, while much less localized in the in vitro assays. Pausing of reverse transcription increases the efficiency of strand transfer (9,31), and RTs are known to have a very low processivity in vitro. This feature has thereby often been considered as responsible for the high frequency of strand transfer observed in the cell free system. On the basis of the data presented here, it is tempting to speculate that the divergence in the distribution of the breakpoints observed in vitro and ex vivo, could be at least partially explained if reverse transcription in vivo were more processive than in vitro. Another important difference concerns the inhibition of strand transfer in R2 with SL RNA in vitro, while in infected cells
the same region is clearly a preferential site for template switching. However, significant similarities are still found between the two systems. One is the observation that in both cases R2 is the hottest region on C2 RNA, even if the preference for R2 is clearly more marked in infected cells (figures 2C and 4B); another is that strand transfer in R2 is reduced on Delhp RNA with respect to C2 RNA, as previously observed in vitro (10,21); and the last one, that decreasing the stability of the hairpin going from SL to 2b RNAs, leads to an enhancement of strand transfer in R2 (figure 3C and figure 4B, deep blue and red).

The work described here was carried out on sequences from the LAI isolate. The C2 hairpin, however, does not seem to be restricted to this isolate, nor to subtype B, since a comparison of the C2 region between LAI and more than two hundred isolates from all subtypes of the M group indicated that a comparable stem-loop structure is potentially present in all cases (not shown). An implication of the C2 region in the generation of recombinant forms has also been suggested by an analysis of recombination junctions generated after multiple cycles of infection of cells in culture between A/D and B/E subtypes (32). Interestingly, recombination breakpoints in the C2 region were also found in intersubtype recombinant forms isolated from patients (ISR), although they seemed to be lost in the endemic circulating recombinant forms (CRF) (32). Understanding the bases of these differences will certainly constitute an attractive area of research in the future.

Several in vitro works have suggested an implication of hairpin structures in promoting copy choice by bringing donor and acceptor RNAs into close proximity (33,34), or by favouring pausing of reverse transcription at the base of the hairpin followed by strand transfer within the hairpin region (20). Based on an in vitro study on the same C2 sequence studied here, we previously proposed that strand transfer would begin by the docking of the nascent DNA onto the acceptor RNA in the loop of the hairpin, and would then proceed through a process of strand exchange similar to branch migration occurring during DNA-DNA recombination (10). For this process to be efficient it is predicted that the hairpin must be stable enough as to resist the dynamic changes that the genomic RNA is expected to undergo during reverse transcription, but not too stable in order to allow its opening during
the process of branch migration. The indication of the existence of a window of optimal stability for strand transfer to occur ex vivo supports this idea. Along the same lines, the difficulty in melting the stable hairpin may be also reflected by the inhibition of \textit{in vitro} strand transfer in R2 on SL RNA, since it was relieved by the RNA chaperon NC, known to favour breathing of hairpins. Furthermore, with 2b RNA, where the hairpin is significantly less stable than in SL RNA, the recombination rate \textit{in vitro} in R2 was ten times higher than in SL, despite the close sequence similarity. The deleterious effect of too stable hairpins on template switching was also previously demonstrated for terminal strand transfer \textit{in vitro} on the TAR hairpin (29,30). Noteworthy, the possibility that the role of the C2 hairpin is just to promote strand transfer at the base of the hairpin by enhancing pausing of reverse transcription is not supported by our results, since the hot region (R2) is located at least 40 nt downstream the base of the C2 hairpin, in the sense of reverse transcription. Furthermore, in that case, one would expect that increasing the stability of the hairpin would enhance the efficiency of strand transfer in R1, a prediction disproved by the data given in figure 5.

In conclusion this work shows that recombination in a portion of the \textit{env} gene of HIV-1 does not occur at constant rates but rather presents a clear hot spot region, and underscores the role of the structure of the genomic RNA as an important parameter in the process of copy choice. Determining whether hot spots are frequent along the genome, and progressing in their molecular characterisation will certainly improve the understanding of the dynamics leading to the generation of the recombinant forms of HIV-1 that challenge immune control \textit{in vivo}.

\textbf{Aknowledgemets}

We thank Henri Buc for critical comments on the manuscript, Sarah Gallois-Montbrun for expertise and assistance during the purification of HIV-1 RT, and Bernard Roques for the generous gift of the NC protein. This work was supported by grant 02172 from the Agence
Nationale de Recherches sur le SIDA (ANRS) to M.N. R.G. is a recipient of a postdoctoral fellowship from ANRS.

References

1. Weiss, R. A., Mason, W. S., and Vogt, P. K. (1973) Virology 52, 535-552.
2. Peeters, M., Toure-Kane, C., and Nkengasong, J. N. (2003) Aids 17, 2547-2560
3. Rambaut, A., Posada, D., Crandall, K. A., and Holmes, E. C. (2004) Nat Rev Genet 5, 52-61
4. Telesnitsky, A., and Goff, S. P. (1997) in Retroviruses (Coffin, J. M., Hughes, S. H., and Varmus, H. E., eds), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
5. Douek, D. C., Picker, L. J., and Koup, R. A. (2003) Annu Rev Immunol 21, 265-304
6. Peeters, M., and Sharp, P. M. (2000) Aids 14, S129-140
7. Onafuwa, A., An, W., Robson, N. D., and Telesnitsky, A. (2003) J Virol 77, 4577-4587
8. Hwang, C. K., Svarovskaia, E. S., and Pathak, V. K. (2001) Proc Natl Acad Sci U S A 98, 12209-12214.
9. DeStefano, J. J., Bambara, R. A., and Fay, P. J. (1994) J. Biol. Chem. 269, 161-168
10. Moumen, A., Polomack, L., Unge, T., Veron, M., Buc, H., and Negroni, M. (2003) J Biol Chem 278, 15973-15982
11. Raja, A., and DeStefano, J. J. (2003) J Biol Chem 278, 10102-10111
12. Balakrishnan, M., Roques, B. P., Fay, P. J., and Bambara, R. A. (2003) J Virol 77, 4710-4721
13. Pfeiffer, J. K., Topping, R. S., Shin, N. H., and Telesnitsky, A. (1999) J Virol 73, 8441-8447
14. Rodriguez-Rodriguez, L., Tsuchihashi, Z., Fuentes, G. M., Bambara, R. A., and Fay, P. J. (1995) J Biol Chem 270, 15005-15011
15. Wisniewski, M., Balakrishnan, M., Palaniappan, C., Fay, P. J., and Bambara, R. A. (2000) Proc Natl Acad Sci U S A 97, 11978-11983
16. Jetzt, A. E., Yu, H., Klarmann, G. J., Ron, Y., Preston, B. D., and Dougherty, J. P. (2000) J Virol 74, 1234-1240
17. Levy, D. N., Aldrovandi, G. M., Kutsch, O., and Shaw, G. M. (2004) Proc Natl Acad Sci U S A 101, 4204-4209
18. Zhuang, J., Jetzt, A. E., Sun, G., Yu, H., Klarmann, G., Ron, Y., Preston, B. D., and Dougherty, J. P. (2002) J Virol 76, 11273-11282
19. Negroni, M., and Buc, H. (2000) Proc Natl Acad Sci U S A 97, 6385-6390
20. Roda, R. H., Balakrishnan, M., Kim, J. K., Roques, B. P., Fay, P. J., and Bambara, R. A. (2002) J Biol Chem 277, 46900-46911
21. Moumen, A., Polomack, L., Roques, B., Buc, H., and Negroni, M. (2001) Nucleic Acids Res 29, 3814-3821
22. Zennou, V., Petit, C., Guetard, D., Nerhbass, U., Montagnier, L., and Charnay, P. (2000) Cell 101, 173-185.
23. Naldini, L., Blomer, U., Gallay, P., Ory, D., Mulligan, R., Gage, F. H., Verma, I. M., and Trono, D. (1996) *Science* **272**, 263-267
24. Yee, J. K., Miyashara, A., LaPorte, P., Bouic, K., Burns, J. C., and Friedmann, T. (1994) *Proc Natl Acad Sci U S A* **91**, 9564-9568
25. Hirt, B. (1967) *J Mol Biol* **26**, 365-369
26. Negroni, M., Ricchetti, M., Nouvel, P., and Buc, H. (1995) *Proc Natl Acad Sci U S A* **92**, 6971-6975
27. Canard, B., Sarfati, S. R., and Richardson, C. C. (1998) *J Biol Chem* **273**, 14596-14604
28. Poznansky, M., Lever, A., Bergeron, L., Haseltine, W., and Sodroski, J. (1991) *J Virol* **65**, 532-536
29. Hong, M. K., Harbron, E. J., O'Connor, D. B., Guo, J., Barbara, P. F., Levin, J. G., and Musier-Forsyth, K. (2003) *J Mol Biol* **325**, 1-10
30. Klasens, B. I., Huthoff, H. T., Das, A. T., Jeeninga, R. E., and Berkhout, B. (1999) *Biochim Biophys Acta* **1444**, 355-370
31. Wu, W., Blumberg, B. M., Fay, P. J., and Bambara, R. A. (1995) *J Biol Chem* **270**, 325-332
32. Quinones-Mateu, M. E., Gao, Y., Ball, S. C., Marozsan, A. J., Abraha, A., and Arts, E. J. (2002) *J Virol* **76**, 9600-9613.
33. Balakrishnan, M., Fay, P. J., and Bambara, R. A. (2001) *J Biol Chem* **276**, 36482-36492.
34. Berkhout, B., Vastenhouw, N. L., Klasens, B. I., and Huthoff, H. (2001) *Rna* **7**, 1097-1114
35. Zuker, M. (2003) *Nucleic Acids Res* **31**, 3406-3415

**Legend of Tables and Figures**

**Table I.** *Recombination assays in the four genomic RNAs used in this study.* The number of total and of recombinant (blue) colonies (corrected as described in Experimental Procedures) obtained in control ("homozygous") and recombination ("heterozygous") assays is shown for three independent assays. F is the global frequency of recombination expressed as percentage.

**Table II.** *Distribution of breakpoints for C2 RNA in recombinant colonies from the "homozygous" control samples.* The 11 colonies available from three independent assays (Exp. 1 to 3) were analyzed by restriction mapping, and the number of colonies found to recombine in each of the regions are shown in the table.
**Table III.** *Recombination rates per nucleotide observed in vitro.* The values given constitute the data used to generate the graphic on figure 4B, and are the average of three to six independent experiments (mean ± standard deviation).

**Figure 1.** *Experimental system for the study of recombination ex vivo.* Panel A, outline of the experimental procedure. Panels B and C, structure of the genomic plasmids and RTP, respectively. The position of the *Bam*HI site used for cloning of the RTP is shown (Ba). In panel B, the schematic positions of the ∆FLAP and of the encapsidation sequence (∈) are shown. In panel C, the positions of hybridisation of primers SH and BH (see Experimental Procedures) are indicated. The *Sac*II site carried by the non-annealing tail of SH primer is shown. Panel D, agarose gel showing the difference in size of PCR products from plasmids used for transfection (lane 1), or from RTP (lane 2). MW: molecular weight marker. This difference is due to the presence of a full-length 5' U3 region in the plasmids, and a deleted form in the RTP (top and bottom drawings, respectively). The sizes of the amplified products are 2,982 and 2,591 base pairs, respectively. Panel E, schematic representation of the RNAs generated after transfection with pLac⁻ and pCDNA3-vir-Lac⁺ for the control of recombination during transfection of the producer cells.

**Figure 2.** *Recombination on C2 RNA in transduced cells.* Panel A, schematic representation of the gp120 protein with the signal peptide (SP) and constant (C1-C5) and variable (V1-V5) regions. The location of the sequence used as region of homology in C2 RNA is shown, subdivided into the five regions R1 to R5. Panel B, folding in solution of the C2 hairpin as determined in (10). The colour code is the same as for panel A. Only the 39 residues of R1 involved in the formation of the hairpin are shown, while the whole R2 and R3 regions are given. Panel C, recombination rates in the 5 regions for three independent experiments, after analysis of 48 recombinant colonies from each assay.
Figure 3. Recombination efficiency and stability of genomic RNA structures. Panel A, schematic representation of the region of homology of the different genomic RNAs used. C2 RNA (drawn as in figure 2) is given as a reference. Different colours indicate different sequences. The vertical orange bars in R1 in 2b RNA represent the base substitutions shown in orange in panel B. Delhp, SL and 2b: region of homology 500 nt; C2: 400 nt. Panel B, folding of Delhp RNA in solution as determined in (10) and folding predictions of SL and 2b RNAs. The colour code is the same as in panel A. Delhp RNA is schematically drawn completely open even if these sequences could likely be engaged in interactions with other portions of the RNA. Panel C, recombination rates in the five regions. The values are the average of three independent experiments for each RNA after analysis of 48 recombinant colonies per assay.

Figure 4. Recombination in the cell-free system. Panel A, the experimental system used to study copy choice in vitro. The model RNA templates are shown. The sequences in the lac-, lac+, and "homology" boxes are the same as in figure 1B. Processive reverse transcription of the donor RNA leads to the synthesis of lac- molecules, whereas template switching on the region of homology generates lac+ molecules. Single stranded DNAs posses the same sequence at the 3' end (black box) that is used to prime synthesis of the second DNA strand using Taq polymerase. The resulting double stranded products are cloned in E.coli using the BamHI and PstI sites present on both, parental and recombinant molecules. The ratio of the blue colonies (recombinant) to the sum of blue plus white (parental) colonies gives the frequency of copy choice recombination. The recombination rate per nt is calculated as described under the "Experimental Procedures" section for the ex vivo system. Panel B, in vitro recombination rates per nt on C2 (pale blue), SL (deep blue), SL in the presence of NC (yellow), and 2b (red) RNAs. The values are the average of at least three independent experiments after analysis of 48 recombinant colonies per assay, and are given in Table III.
Figure 5. Correlation between recombination rates and stability of the hairpin. The recombination rates per nucleotide in regions R1 (white squares), R2 (black triangles) and R3 (grey circles) are represented as a function of the free energy ($\Delta G$) predicted by the m-fold program (35) for each of the structures: Delhp, C2, 2b and SL. Only the regions that participate in the formation of the hairpin are shown.
Table I. Recombination assays in the four genomic RNAs.

| Sample | Experiment | Blue | Total | F (%) |
|--------|------------|------|-------|-------|
|        | Homozygous 1 | 4 | 675 | 0.59 |
|        | 2 | 5 | 901 | 0.55 |
|        | 3 | 2 | 633 | 0.32 |
|        | Heterozygous 1 | 78 | 596 | 13.09 |
|        | 2 | 70 | 529 | 13.23 |
|        | 3 | 99 | 719 | 13.77 |
|        | Homozygous 1 | 3 | 825 | 0.37 |
|        | 2 | 1 | 621 | 0.16 |
|        | 3 | 1 | 725 | 0.14 |
|        | Heterozygous 1 | 97 | 1054 | 9.20 |
|        | 2 | 73 | 694 | 10.52 |
|        | 3 | 93 | 882 | 10.54 |
|        | Homozygous 1 | 3 | 698 | 0.43 |
|        | 2 | 1 | 639 | 0.16 |
|        | 3 | 3 | 435 | 0.69 |
|        | Heterozygous 1 | 76 | 553 | 13.74 |
|        | 2 | 74 | 701 | 10.56 |
|        | 3 | 85 | 913 | 9.31 |
|        | Homozygous 1 | 1 | 503 | 0.20 |
|        | 2 | 2 | 538 | 0.37 |
|        | 3 | 2 | 734 | 0.27 |
|        | Heterozygous 1 | 68 | 390 | 17.44 |
|        | 2 | 111 | 636 | 17.45 |
|        | 3 | 94 | 580 | 16.20 |
**Table II.** Distribution of breakpoints for C2 RNA in recombinant colonies from the "homozygous" control samples.

| Region | Exp.1 | 2 | 3 |
|--------|-------|---|---|
| 1      | 0     | 2 | 0 |
| 2      | 1     | 1 | 1 |
| 3      | 0     | 1 | 0 |
| 4      | 2     | 0 | 1 |
| 5      | 1     | 1 | 0 |
| total  | 4     | 5 | 2 |
Table III. Recombination rates per nucleotide observed in vitro.

|   | C2   | SL   | SL + NC | 2b   |
|---|------|------|---------|------|
| 1 | 5.0 ± 0.9 | 2.8 ± 1.4 | 5.5 ± 1.6 | 6.8 ± 1.4 |
| 2 | 15.5 ± 2.1 | 1.7 ± 0.8 | 6.2 ± 2.8 | 11.5 ± 3.1 |
| 3 | 3.1 ± 1.4 | 0.5 ± 0.2 | 0.5 ± 0.1 | 1.7 ± 0.7 |
| 4 | 5.3 ± 0.8 | 5.5 ± 1.4 | 8.8 ± 2.3 | 8.1 ± 1.7 |
| 5 | 8.4 ± 1.1 | 7.7 ± 0.8 | 6.6 ± 1.7 | 5.8 ± 1.3 |
Figure 1
Figure 2
Figure 3
Figure 4

A

\[ \frac{\text{nb of lac}^+ \text{ colonies}}{\text{nb of total colonies}} = \text{frequency of recombination} \]

B

Recombination rate (x 10^-4)

|    | R5 | R4 | R3 | R2 | R1 |
|----|----|----|----|----|----|
| 0  |    |    |    |    |    |
| 2  |    |    |    |    |    |
| 4  |    |    |    |    |    |
| 6  |    |    |    |    |    |
| 8  |    |    |    |    |    |
| 10 |    |    |    |    |    |
| 12 |    |    |    |    |    |
| 14 |    |    |    |    |    |
The structure of HIV-1 genomic RNA in the gp120 gene determines a recombination hot spot in vivo

Román Galetto, Abdeladim Moumen, Véronique Giacomoni, Michel Véron, Pierre Charneau and Matteo Negroni

J. Biol. Chem. published online June 24, 2004

Access the most updated version of this article at doi: 10.1074/jbc.M405476200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts