Evidence for a Mechanism of Recombination during Reverse Transcription Dependent on the Structure of the Acceptor RNA*⑤

Abdeladim Moumen‡, Lucette Polomack, Torsten Unge§, Michel Véron, Henri Buc¶, and Matteo Negroni‖

From the Unité de Regulation Enzymatique des Activités Cellulaires, CNRS-FRE 2364, Département de Biologie Structurale et Chimie and ¶CNRS-URA 1960, Institut Pasteur, 25–28 rue du Docteur Roux, 75724 Paris cedex 15, France and §Uppsala Universitet, Department of Cell and Molecular Biology Biomedical Center, SE-75124, Uppsala, Sweden

Genetic recombination is a major force driving retroviral evolution. In retroviruses, recombination proceeds mostly through copy choice during reverse transcription. Using a reconstituted *in vitro* system, we have studied the mechanism of strand transfer on a major recombination hot spot we previously identified within the genome of HIV-1. We show that on this model sequence the frequency of copy choice is strongly influenced by the folding of the RNA template, namely by the presence of a stable hairpin. This structure must be specifically present on the acceptor template. We previously proposed that strand transfer follows a two-step process: docking of the nascent DNA onto the acceptor RNA and strand invasion. The frequency of recombination under copy choice conditions was not dependent on the concentration of the acceptor RNA, in contrast with strand transfer occurring at strong arrests of reverse transcription. During copy choice strand transfer, the docking step is not rate limiting. We propose that the hairpin present on the acceptor RNA could mediate strand transfer following a mechanism reminiscent of branch migration during DNA recombination.

By reshuffling large regions of genetically distinct genomes, recombination speeds up the rate of evolution (1, 2). Recombination constitutes the most frequent genomic aberration in retroviruses; its frequency of occurrence is equal to the cumulative frequency of all the other types of mutations (3). The most intensively studied member of this group of viruses, the human immunodeficiency virus type 1 (HIV-1),[1] illustrates the impact of recombination on the dynamics of retroviral evolution. In this case, at the very least 10% of the circulating strains have been generated by genetic recombination among different HIV-1 subtypes (4). In retroviruses, recombination occurs mainly during reverse transcription (5). Each viral particle contains two copies of single-stranded positive genomic

RNA (6). If two different variants of a virus infect the same cell, as recently documented for HIV-1 (7), the viral progeny will be constituted by homozygous and heterozygous virions. Recombination can then occur when a heterozygous virus infects a new cell. Indeed, during synthesis of the (−) DNA strand the reverse transcriptase (RT) can switch template and, guided by the local sequence homology, transfer DNA synthesis from one genomic RNA molecule (the donor) onto the other (acceptor RNA). In a heterozygous virion this process, known as copy choice, leads to genetic recombination (8).

Despite the dramatic impact of recombination on the evolution of retroviruses, the underlying mechanisms are not yet understood. Based on the observation that the purification of viral RNA from retroviral particles led to the isolation of fragmented molecules, it was suggested that the genomic RNA is not intact within the viral particle. It was therefore proposed that the switch would be a consequence of a block of reverse transcription caused by a break on the RNA, a model called "forced copy choice" (9). In this case the stalling of the RT would constitute the crucial step of the process by allowing an extensive degradation of the RNA template by the RNase H activity carried by the RT itself, as demonstrated for (−)DNA strong stop strand transfer (10). The resulting single-stranded DNA would then be available for annealing onto the complementary sequence provided by the acceptor RNA. A similar situation can be encountered if stalling is generated by strong pause occurring during reverse transcription of an intact template. Indeed, a prominent pause site detected during *in vitro* reverse transcription of a stretch of the HIV-1 *nef* gene was shown to increase significantly the local frequency of strand transfer (11, 12). In these instances the stalling of reverse transcription is regarded as the limiting step for strand transfer.

This idea has, however, recently been challenged by increasing evidence demonstrating that pausing during reverse transcription and strand transfer are not necessarily coupled (13–15). In parallel, studies carried out on RNA templates containing hairpin regions have suggested that such structures could favor template switching by RTs (13, 16, 17). In these cases it was proposed that the hairpin structures enhance the probability of strand transfer by mediating an interaction between donor and acceptor RNA that increases their spatial proximity (13, 16, 17). Extensive random searches for the occurrence of recombination hot spots during *in vitro* reverse transcription by HIV-1 RT had revealed the correlation between the location of these hot spots and the presence of predicted hairpin regions in the RNA template (14, 18). Based on this observation, it was proposed that template switching proceeds through a two-step mechanism: docking of the acceptor RNA onto the nascent DNA and displacement of the donor RNA by the acceptor RNA (14). The latter step would be guided by

---

[1] The abbreviations used are: HIV, human immunodeficiency virus; nt, nucleotides; RT, reverse transcription; NC, nucleocapsid protein; SL, stem-loop.

the folding of the RNA. A recent study on the primer binding site of the equine infectious anemia virus has shown that the hairpin present in that region induces a strong pause of reverse transcription that increases the efficiency of the docking step (15). In addition, *in vitro* studies on the effect on strand transfer of the nucleocapsid protein (NC), a major co-factor of the reverse transcription complex (19), have suggested a mechanism of recombination governed by the structures of the RNA rather than by pausing of reverse transcription (reviewed in Refs. 3 and 20). Indeed, although NC enhances strand transfer *in vitro* (reviewed in Ref. 3) it does not lead to a parallel increase of pausing during reverse transcription, as predicted for a mechanism of template switching governed by pausing of DNA synthesis (14). Because the NC is a RNA chaperone (21, 22), it was suggested that the enhancement of strand transfer observed in its presence was because of its ability to modulate the structures of the RNA templates (14).

Among the several recombinant HIV-1 strains isolated to date, a well defined case is constituted by chimerical genomes between subgroups A and either C or D, generated by recombination on the region coding for the constant portion C2 of the envelope glycoprotein gp120 (23). Recombination on this segment of genome allows reshuffling of the portions of gp120 coding for the variable regions V1 and V2, relative to regions V3 through V5. The spatial arrangement of regions V2 and V3 with respect to the constant regions of the protein has been shown to be critical for allowing the virus to escape neutralizing antibodies raised by the immune system of the host (24). In a previous report we used several RNA sequences issued from the HIV-1 genome to investigate the mechanism of template switching by HIV-1 reverse transcriptase *in vitro*. We observed that the genomic sequence coding for the C2 region constituted, indeed, the most important hot spot we found during that work (18). Interestingly, a subsequent study on recombination during infection of cells in culture with different HIV-1 subtypes has also shown the occurrence of frequent recombination in the same region (25). In our previous study, the portion of 200 nt that constituted the hot spot within the C2 domain was called "Eb" and was initially included in a model template where it was surrounded by the sequences that flank it on the viral genome (Fig. 1*A*, *RNA E2*). It was subsequently observed that by changing the surrounding sequences (Fig. 1, *RNA G1Eb*) the frequency of strand transfer on Eb was decreased by a 4- to 5-fold factor (18). We referred to this effect as "context effect." In this work, we took advantage of the context effect to investigate the role of the RNA structure in the transfer process and to address the question of the mechanism responsible for copy choice by HIV-1 RT.

MATERIALS AND METHODS

Labeling of RNA and Determination of the Secondary Structures—To determine their folding, the various RNA templates were labeled at their 3′-end as follows. A 21-mer oligonucleotide with a sequence complementary to the 3′-end of the RNA to label was annealed, at a molar ratio of 4:1 (oligo:RNA) and in a buffer containing 10 mM Tris-HCl, pH 7.5, 50 mM NaCl, 1 mM EDTA, by heating the mixture for 30 s at 95 °C followed by a slow cooling to 30 °C. The oligonucleotide carries at its 5′-end a non-hybridizing tail of six nucleotides constituted by the sequence 5′-CTTTTT-3′. The annealed RNA was then incubated for 20 min at 37 °C in a final volume of 55 µl in a buffer containing 7 mM dithiothreitol, 20 mM Tris-HCl, pH 7.5, 12.5 mM MgCl₂, 20 mM NaCl, 40 units of RNasin (Promega), 5 units of Sequenase (United States Biochemicals), and 0.05 mCi of [α³²P]dideoxyATP (Amersham Biosciences). Labeled RNAs were purified on 7% polyacrylamide gel and eluted by passive diffusion at 4 °C in a buffer containing 10 mM magnesium acetate, 500 mM ammonium acetate, 0.1% SDS, and 1 mM EDTA. The eluted RNA was extracted with phenol-chloroform and precipitated in ethanol; the dried pellet was conserved under ethanol at −20 °C. For determination of the structure of the RNA, 8 pmol of the labeled RNA were heated in the reverse transcription buffer (50 mM Tris-HCl, pH

7.8, 75 mM KCl, 7 mM MgCl₂) at 65 °C for 5 min, slowly cooled to 40 °C, and transferred on ice. The RNA was then treated with either T1 (0.3 and 0.15 units) or T2 (0.04 and 0.02 units) RNases for 5 min at 37 °C. T1 and T2 RNases cleave single-stranded RNA molecules with preference for guanine residues for T1 and adenine residues for T2 (26). The reaction was stopped by phenol-chloroform extraction followed by ethanol precipitation. The products were analyzed by autoradiography after electrophoresis on 7% polyacrylamide gels (see Fig. 1). Quantification was performed using phosphorimaging apparatus (Molecular Dynamics). The positions of enzymatic cleavage were identified by reference to a ladder generated by extensive T1 digestion of the same RNA molecule. The residues identified as single-stranded in four independent experiments were introduced as constraints in the structure prediction analysis by the m-fold program (27).

RNA Synthesis and Recombination Assays—The various constructs used for RNA synthesis were generated following standard cloning procedures (28). Each construct was systematically sequenced before its use in RNA synthesis. RNA synthesis was performed as previously described (29). Reverse transcription of the donor RNA was carried out in the presence of the acceptor RNA (at a final concentration of 100 nM each, unless otherwise stated) after annealing an oligonucleotide specifically onto the donor template (Fig. 2*A*). Annealing was performed at a molar ratio of primer to donor RNA of 10:1 in 50 mM Tris-HCl (pH 7.8), 75 mM KCl, 7 mM MgCl₂ at 65 °C for 5 min followed by a slow cooling to 40 °C. Dithiothreitol (1 mM final concentration), the four dNTPs (1 mM each), and RNasin (100 units; Promega) were added after incubation for 5 min on ice. For the experiments with NC (55 amino acids, synthesized as described in Ref. 30), the protein was added at this step at a ratio of 1 molecule of NC/8 nt of total RNA and incubated for 10 min 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 90 min. The reaction was stopped by extraction with phenol-chloroform. The samples containing NC were treated, before phenol-chloroform extraction, for 1 h at 56 °C with proteinase K (8 mg/ml), 0.4% (w/v) SDS, and 50 mM EDTA (pH 8.0). The phenol-chloroform-extracted samples were then submitted to RNase treatment. Purification of the reverse transcription product and synthesis of the second DNA strand were performed as previously described (14). *Bam*HI and *Pst*I digestion, ligation, and *Escherichia coli* transformation were also carried out as previously described (18) and as shown in Fig. 2. For the experiments where the concentration of acceptor template was varied, the procedure of reverse transcription of a fixed amount (100 nM) of donor RNA was identical to the one described above, but the concentration of acceptor included in the assay was varied as detailed in Fig. 5. In all cases HIV-1 RT was used at a concentration of 400 nM.

Estimation of Recombination Rates per Nucleotide—The recombination frequency in the various intervals of the model templates was calculated as follows. In the case of recombination between dE2 and aE2 RNAs, as an example, a *Nco*I restriction site was present at the boundary between Ea and Eb on the donor RNA, and an *Apa*LI site marked the transition between Eb and Ec on the acceptor RNA (Fig. 3*A*, *column I*). All recombinant molecules *Nco*I⁺/*Apa*LI⁺ were considered to issue from template switching within Eb. We define *F* as the overall frequency of recombination observed in the experiment, as given in Fig. 2, *left panel*. If *b* is the number of blue colonies whose restriction pattern has been analyzed and *c* is the number of recombinant colonies *Nco*I⁺/*Apa*LI⁺, the frequency of recombination within Eb (*f*) is given by $F(c/b)$. Recombination rates per nt were calculated by dividing the frequency of recombination within a given region by its size in nt (in the example given here, Eb is 200 nt long, and the recombination rate is therefore given by $f/200$).

Primer Extension Assays—Reverse transcription was primed using a 5′-terminal-labeled deoxyoligonucleotide and carried out under the same buffer and RT conditions as above. The templates used in these assays consisted in truncated versions of dG1Eb or dE2, devoid of the sequences coding for the reporter gene and therefore including only the viral sequences. Similarly when the experiments were performed in the presence of the acceptor RNAs, modified versions of aG1Eb and of aE2 RNAs were used, constituted only by the viral sequence. The complex between HIV-1 RT and the primer/template was pre-formed by incubation for 10 min in the same reaction buffer as described above, devoid of dNTPs and MgCl₂. The reaction was started by the addition of dNTPs and MgCl₂ and stopped at various time intervals by addition of EDTA to a final concentration of 15 mM. All samples were ethanol-precipitated before electrophoresis on 8% (w/v) polyacrylamide gel containing 8 M urea in a loading buffer containing formamide at a final concentration of 22.5%. The intensity of each band was estimated by phosphorimaging as described above.
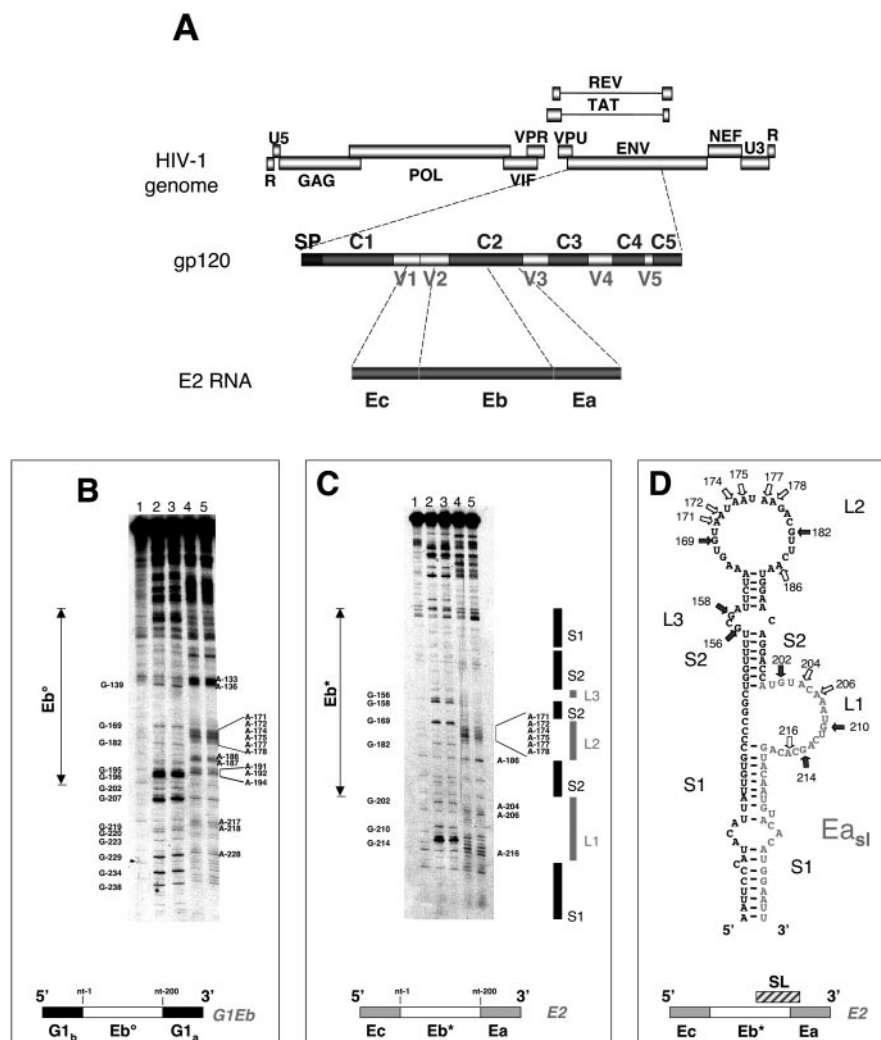
FIG. 1. *A*, location of the main sequence studied in this work, E2, within the HIV-1 genome. The three portions constituting the E2 RNA (*bottom drawing*: *Ec*, *Eb*, and *Ea*) are aligned with respect to the primary structure of the glycoprotein gp120 (*middle drawing: SP*, signal peptide; *C1-C5*, constant regions; *V1-V5*, variable regions). *B–D*, folding of the conditional hot spot Eb. In each panel the drawings at the *bottom* indicate the model templates used. The names of the templates are given in *gray* and *italic* on the *right* of the drawing and those of the subregions that constitute them *below* the drawings. *B* and *C*, analysis by sequencing gel electrophoresis of the cleavage pattern, by RNases T1 and T2, of the Eb region either in G1Eb (*B*) or in E2 (*C*) RNAs. Individual residues are numbered starting from the 5′-border of the region Eb. Consequently, Eb spans from 1 to 200 nt on both RNAs (see *diagrams*). The letters and numbers indicate the bases sensitive to cleavage by T1 (*left*) or T2 (*right*). *B*, *lane 1*, mock sample. *Lanes 2* and *3*, samples treated with 0.3 and 0.15 units of T1, respectively. *Lanes 4* and *5*, samples treated with 0.04 and 0.02 units of T2, respectively. *C*, *lane 1*, mock sample. *Lanes 2* and *3*, samples treated with 0.3 and 0.15 units of T1, respectively. *Lanes 4* and *5*, samples treated with 0.02 and 0.04 units of T2, respectively. *C*, the *bars* on the *right* indicate the participation of the corresponding bases to a stem (*S*) or loop (*L*) region as detailed in *panel D*. *D*, folding of the 3′-portion of the Eb region in the model template E2 (conformation Eb* in the text). *White* and *black arrows* indicate the residues whose conformation has been determined empirically by T2 and T1 RNases, respectively, as shown in *panel C*. The sequence Eb is given in *black letters*; *gray letters* indicate residues belonging to Ea, the region preceding Eb on E2 RNA, following the sense of reverse transcription. *L1–L3*, loop regions; *S1* and *S2*, stem portions. The *dashed box SL* in the bottom drawing of *panel D* indicates the location of the hairpin structure depicted above.

RESULTS

Role of the Secondary Structure of the Templates—To study the potential correlation between the frequency of recombination on the Eb region and its structure, we determined its folding in solution when it is included either in E2 RNA (condition under which Eb is a recombination hot spot) or in G1Eb RNA (where Eb was not a recombination hot spot). In these two cases we call this region Eb* and Eb°, respectively. The determination of the folding of the RNAs was performed under the same buffer and temperature conditions as the recombination assay (see "Materials and Methods"). These two RNAs were labeled at their 3′-end, subjected to enzymatic probing (Fig. 1, *panels B* and *C*), and their secondary structures were assessed by including the constraints revealed by these probing assays into the m-fold program (27). The most significant difference

found between the folding of Eb* and Eb° was in the 3′-terminal region of the sequences. In this part, we identified a stem-loop motif, called here SL (Fig. 1*D*), present exclusively in the case of Eb*. As shown in Fig. 1, the SL hairpin is constituted at its 3′-end by a portion of the region Ea, the sequence downstream Eb* in E2 RNA. When Eb is part of G1Eb RNA (Eb°), Ea is replaced by the region G1a and the formation of the SL hairpin is no longer possible (Fig. 1, *panels B* and *C*, *bottom*). No other stable hairpins were found within Eb°.

Which Template Drives the Switch?—The influence of these structural differences in the strand transfer reaction was then investigated by using a recombination assay previously described (18) and outlined in Fig. 2, *left panel*. Strand transfer involves two types of RNA templates, the donor and the acceptor. The correlation between the folding of the Eb region
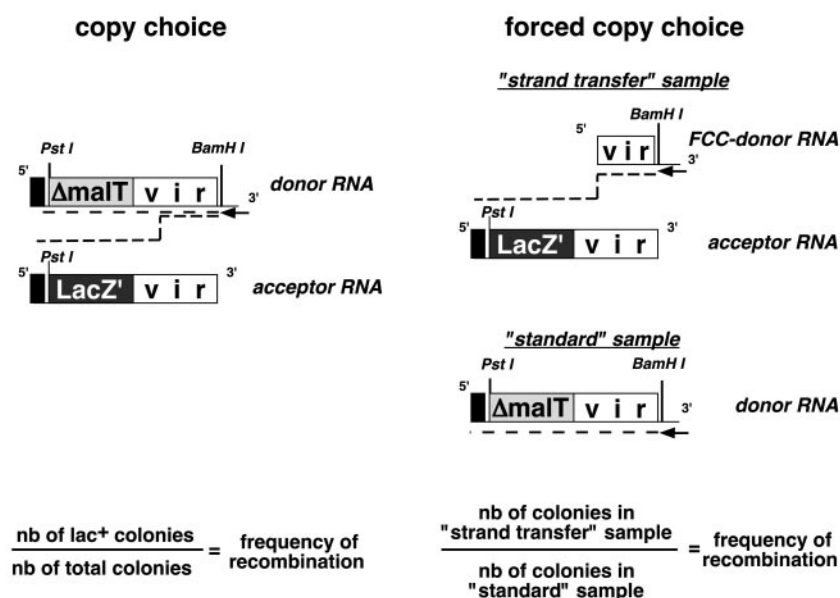
FIG. 2. **Experimental systems used to study copy choice or forced copy choice.** *Left panel*, copy choice. Reverse transcription is selectively primed on the donor RNA in the presence of the acceptor RNA. The two RNAs share a region of homology constituted by a viral sequence (*vir*), followed, in the sense of reverse transcription, by a genetic marker, different on the two RNAs. Although both templates carry a *Pst*I site near their 5′-terminus, only the donor template possesses a *Bam*HI site at the 3′-end. Processive reverse transcription of the donor RNA leads to the synthesis of *lac⁻* molecules, whereas template switching on the region of homology generates molecules carrying the sequence of a functional *lacZ'* gene. The single-stranded DNAs possess the same sequence at the 3′-end (*black box*) that will be used to prime synthesis of the second DNA strand using *Taq* polymerase (this is not a PCR reaction). The resulting double-stranded products are cloned in *E. coli* using the *Bam*HI and *Pst*I sites, which will be present on both parental and recombinant molecules. On the appropriate dishes, recombinant and parental molecules will generate blue (*lac⁺*) and white (*lac⁻*) colonies, respectively. As a control, an equivalent amount of plasmid vector used for the cloning of the reverse transcription products was ligated and used for transformation, providing an estimate of the background of the white colonies resulting from transformation with circularized vectors. The background value never exceeded 10% of the white colonies recovered from the recombination samples and was systematically subtracted for computation of the frequency of recombination. The ratio of blue colonies to the sum of blue plus white colonies, corrected for the background value, gives the frequency of copy choice recombination. When comparing regions of different size the frequency of recombination was measured in rate (per nt), following the procedure described under "Materials and Methods." *Right panel*, the system used to study forced copy choice. The symbols used are the same as for the *panel* on the *left*. See "Estimation of Recombination Rates per Nucleotide" under "Materials and Methods" for details.

and the frequency of strand transfer observed during its reverse transcription was exploited to assess the role played by each of these templates. The experiments were first performed on naked RNA templates. The rationale of the experiment is outlined in Fig. 3*A*. In the two cases outlined in Fig. 3*A*, *columns I* and *IV*, the donor and the acceptor RNAs share complete sequence homology on the viral sequence, apart from the presence of the restriction sites indicated in the figure. We refer to these conditions as "symmetric," because the folding of the region Eb is the same on the donor and the acceptor RNAs (either Eb* on both or Eb° on both). In contrast, under the conditions depicted in columns II and III ("asymmetric conditions") Eb adopts a different folding between the donor and the acceptor RNAs. Furthermore, under the asymmetric conditions the region Eb constitutes the only region of homology between the two RNAs (in black in the figure). The frequency of template switching within Eb was determined by restriction analysis of the recombinant products, as detailed under "Materials and Methods." When Eb was in the Eb* conformation on the acceptor RNA, strikingly close frequencies of recombination were observed on the Eb interval, independent of which donor RNA was used (Fig. 3*A*, *columns I* and *II*, *naked RNAs*). Conversely, the frequency of strand transfer on Eb was similar when it was in the Eb° conformation on the acceptor RNA, independent from the use of dE2 or of dG1Eb as donor RNA (Fig. 3*A*, *columns III* and *IV*). Therefore, it was evident that the conformation of Eb on the acceptor RNA determined the frequency of template switching. We then tested whether these conclusions reached on experiments on naked RNAs could apply to recombination occurring in the presence of the NC protein. Also, in this case

the frequency of recombination on Eb was found to depend on the folding of the acceptor RNA (Fig. 3*A*, *RNA·NC* complex).

Strand Transfer within SL—To document the role of the SL hairpin in recombination taking place within Eb, we decided to distinguish between the events of strand transfer occurring within SL itself from those taking place outside the hairpin in the Eb interval. We employed aE2 as the acceptor RNA and either dE2 or dG1Eb as donor templates, the conditions under which the highest frequency of transfer was observed on Eb (Fig. 3*A*, *columns I* and *II*). For these experiments, a point mutation generating an *Eco*RI site was introduced on both types of donor RNAs, dE2 and dG1Eb, creating dE2-eco and dG1Eb-eco RNAs, respectively (Fig. 3*B*). This *Eco*RI site is located immediately 5′ with respect to the base of the SL hairpin and allows mapping of strand transfer within Eb. As shown in Fig. 3*B* (*naked RNA*), the use of either dE2-eco or dG1Eb-eco as donor RNAs yielded a recombination rate higher within the hairpin itself than in the downstream portion. Furthermore the rates of recombination were very close when dE2-eco or dG1Eb-eco was used (15.1 and 14.0 × 10⁻⁴ per nt), confirming that the type of donor RNA used does not influence the distribution of the positions of strand transfer within Eb. We then evaluated whether the same conclusions can be drawn from experiments performed in the presence of the NC protein by performing the recombination assays on RNA·NC complexes. Also, in this case strand transfer occurred at high rates on SL, regardless of the type of donor template employed.

Analysis of Pausing Pattern during Reverse Transcription—The observation that the type of donor RNA used does not modify the frequency of template switching strongly suggests that pausing of reverse transcription on the donor RNA
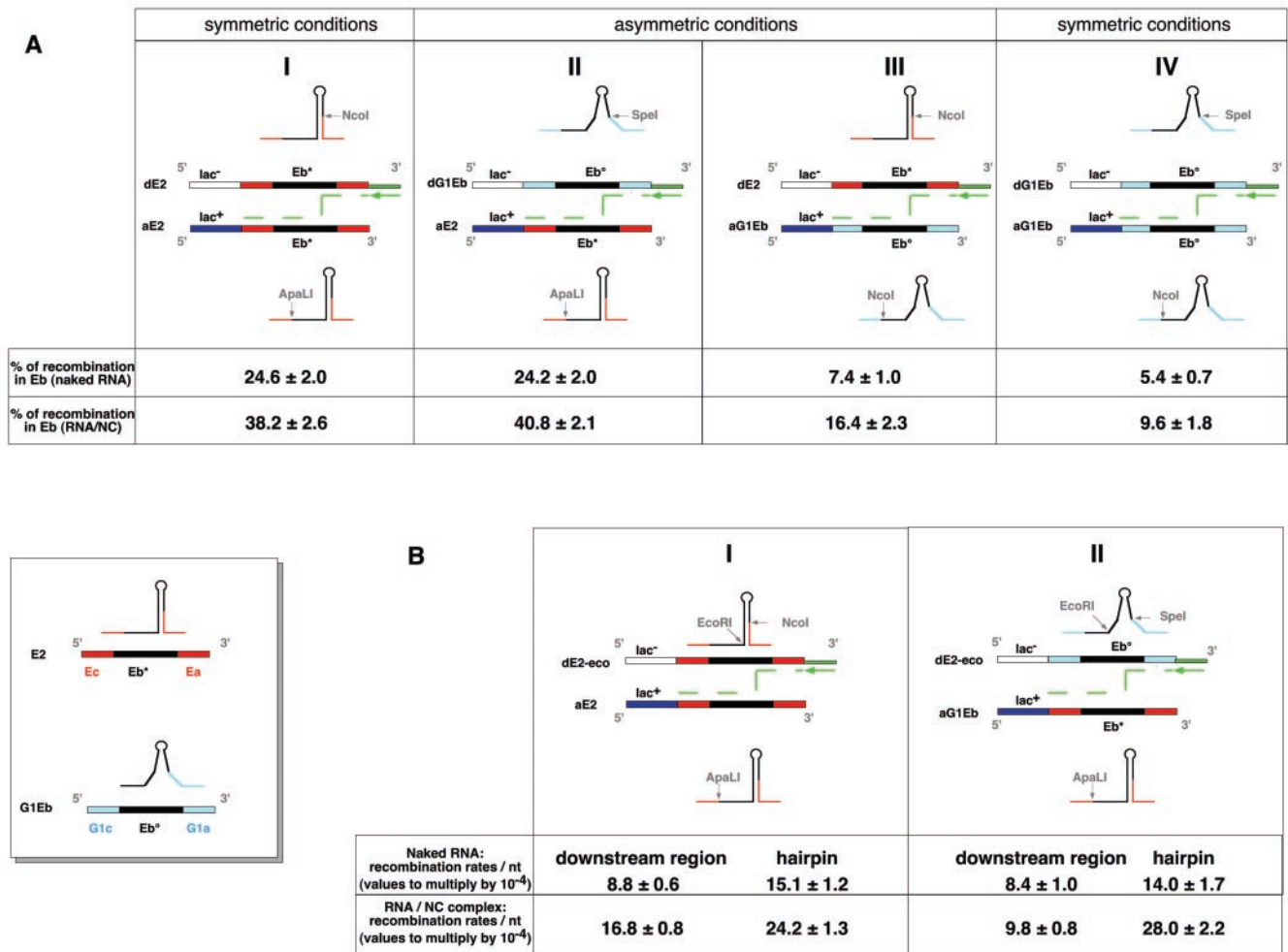
FIG. 3. **Role of the acceptor RNA in template switching.** *Inset*, schematic depiction of the conformations of the viral sequence referred to in text as Eb* and Eb°. The Eb region, drawn in *black*, is encompassed by *red* sequences in the E2 template and by *light blue* ones in G1Eb RNA. The sizes of the different regions are: *Ea*, 100 nt; *Eb** (or Eb°) 200 nt; *Ec*, 100 nt; *G1a*, 150 nt, *G1b*, 150 nt. The *green arrow* shows the direction of reverse transcription on the donor template. *A* and *B*, the letters "a" and "d" preceding the name of the RNA template stand for acceptor and donor RNAs, respectively. *A*, symmetric (*columns I* and *IV*) and asymmetric (*columns II* and *III*) experimental conditions. The folding of the donor and the acceptor RNAs in the SL region is schematically indicated *above* each model template. *Gray arrows* indicate the location of specific restriction sites that, after restriction analysis of the recombinant DNA molecules, allow mapping of strand transfer events along the sequence of homology as described under "Materials and Methods." These sites were generated by the introduction of point mutations. The frequency of recombination observed under the different experimental conditions is given below. *B*, mapping of strand transfer events at the level of the SL region under the two experimental conditions that yield a high degree of transfer (*panel A*, *columns I* and *II*) was performed by introducing the additional restriction site *Eco*RI on the donor RNA. This mutation did not affect the overall frequency of recombination. The region Eb is therefore subdivided into the hairpin and the downstream region. To normalize for the different size of the two regions, the frequency of recombination is given as rates per nt.

is not the trigger for strand transfer. It cannot be ruled out, though, that the pausing pattern on the donor templates is modified when reverse transcription is performed in the presence of aG1Eb or aE2 as acceptor templates. To investigate this point we first carried out a labeled primer extension analysis on dG1Eb and dE2 RNAs (Fig. 4*A*, *lanes 1–4*, and *B*, *lanes 1–5*). Despite the different folding of the SL region on these RNAs, a prominent pause site was found in both cases at the same position, corresponding to a stretch of four uridine residues. To check whether the presence of an acceptor RNA could induce a change in the pausing pattern on the donor template, reverse transcription of dG1Eb (Fig. 4*A*) or of dE2 (Fig. 4*B*) was then performed in the presence of either aG1Eb or aE2. As indicated at the bottom of Fig. 4, these conditions reproduce those employed for the recombination assays depicted in Fig. 3*A*. In no case could a significant change in the pausing pattern be detected. The 4- to 5-fold difference in the frequency of recombination observed, depending on the use of E2 or G1Eb as acceptor RNA (Fig. 3*A*),

was therefore not associated with an increased stalling of the reverse transcription.

Setting Up a System to Investigate Copy Choice and Forced Copy Choice in Parallel—To better evaluate the effect of strong arrests of DNA synthesis on template switching, we have developed an experimental system that reproduces the situation described in the forced copy choice model (Fig. 2, *right panel*). In this system reverse transcription was performed in parallel under two different experimental conditions, referred to in Fig. 2 as "strand transfer" and "standard" samples. In the strand transfer sample, a donor template ("FCC donor RNA," for forced copy choice) is reverse-transcribed in the presence of an acceptor RNA. The FCC donor RNA is truncated within the region of homology with the acceptor RNA. This system allows strand transfer to occur either from internal positions of the region of homology or at the very 5′-end of the donor template. The reverse transcription products are then treated as for the copy choice experiments and cloned in *E. coli* (see the Fig. 2 legend and "Materials and Methods"). Because only the reverse

Mechanism of Copy Choice in Retroviruses
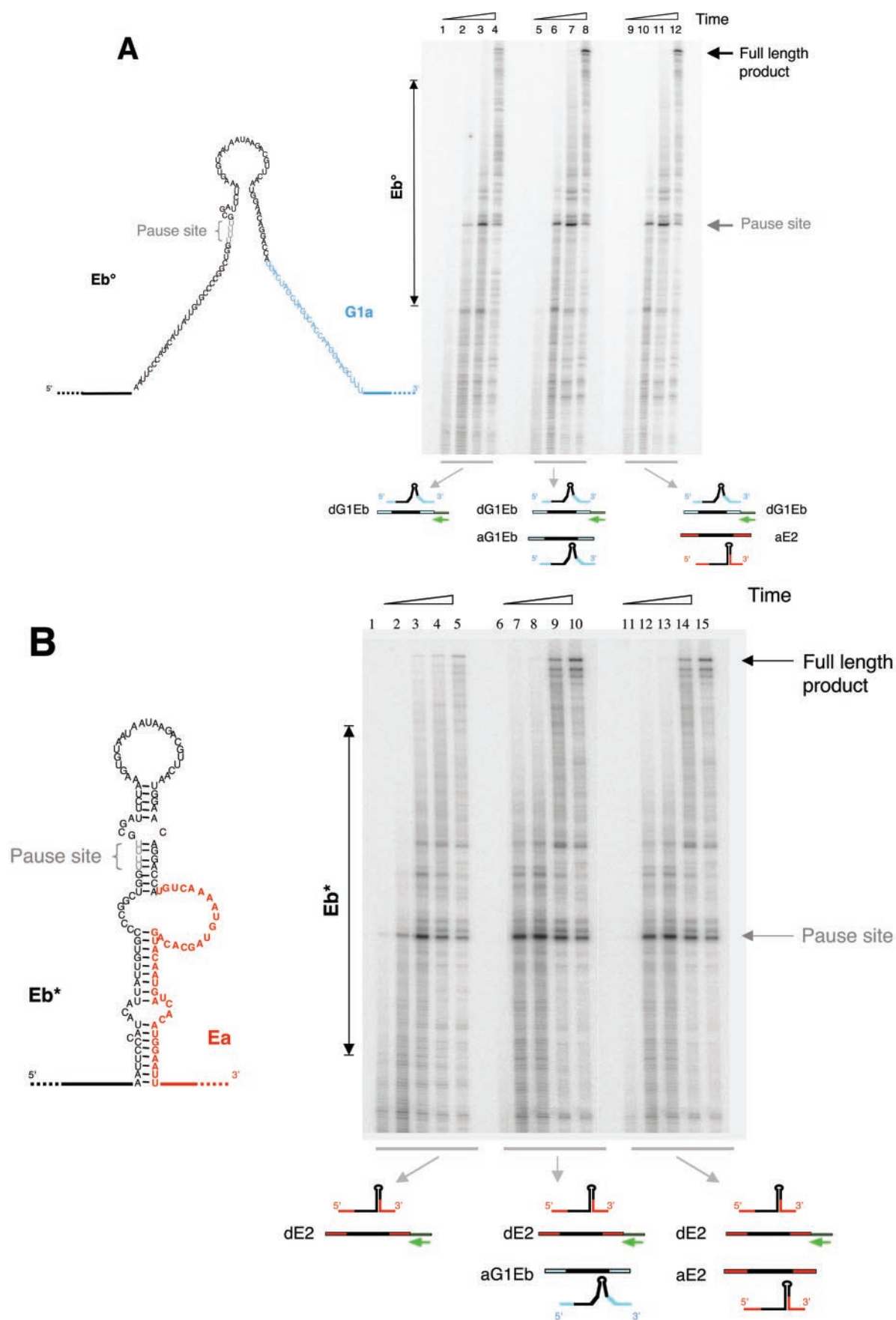
FIG. 4. **Primer extension assays.** Denaturing polyacrylamide gel analysis of primer extension products. *A*, synthesis of DNA was initiated specifically on the dG1Eb RNA in the absence (*1–4*) or in the presence of an equimolar amount of either aG1Eb (*5–8*) or aE2 templates (*9–12*) as acceptor RNA. *B*, DNA synthesis was primed on dE2 in the absence (*1–5*) or the presence of either aG1Eb (*6–10*) or aE2 (*11–15*). The reactions were

FIG. 5. **Effect of the concentration of the acceptor RNA on copy choice and forced copy choice.** *A*, copy choice; *B*, forced copy choice. Only the viral portion ("vir" in Fig. 2) of the model templates used in these assays is shown in each panel. A detailed description of these templates is given in Ref. 18. The recombination rates refer to recombination occurring on the regions shown in *gray* on the donor RNAs (the size of which is given in the table). *Error bars* were calculated as $(b^{1/2}/b)$, where b is the number of blue colonies, and r is the recombination rate. The *standard sample* is the one indicated in Fig. 2 for the forced copy choice conditions.

Ratio donor/acceptor	Eb (200 nt)			R (97 nt)			gag1 (400 nt)		
	Blue colonies	Total colonies	Rate per nt (x10⁻⁴)	Blue colonies	Total colonies	Rate per nt (x10⁻⁴)	Blue colonies	Total colonies	Rate per nt (x10⁻⁴)
0.2	49	657	3.7 ± 0.5	43	456	9.8 ± 1.5	68	1986	0.9 ± 0.1
0.5	77	528	7.3 ± 0.8	78	712	11.3 ± 1.3	66	1884	0.9 ± 0.1
1	58	152	19.1 ± 2.5	74	575	13.2 ± 1.5	68	1506	1.1 ± 0.1
2	59	158	18.7 ± 2.4	148	1046	14.6 ± 1.2	75	1567	1.2 ± 0.1
5	112	268	20.1 ± 2.0	81	502	16.6 ± 1.8	52	1043	1.2 ± 0.2
10	61	156	19.6 ± 2.5	80	473	17.4 ± 1.9	48	794	1.5 ± 0.2

Ratio donor/acceptor	Eb (131 nt)		R (97 nt)		gag1 (231 nt)	
	Blue colonies	Rate per nt (x10⁻⁴)	Blue colonies	Rate per nt (x10⁻⁴)	Blue colonies	Rate per nt (x10⁻⁴)
0.5	67	2.8 ± 0.3	24	7.2 ± 1.5	26	1.5 ± 0.3
1	116	4.8 ± 0.4	42	12.7 ± 2.0	39	2.2 ± 04
2	192	8.0 ± 0.6	77	23.2 ± 2.6	55	3.2 ± 0.4
3.5	335	13.9 ± 0.8	103	31.0 ± 3.1	85	4.9 ± 0.5
5	504	20.9 ± 0.9	120	36.1 ± 3.3	106	6.1 ± 0.6
10	1146	47.6 ± 1.4	243	73.2 ± 4.7	189	10.9 ± 0.8
Colonies in "standard sample"*	1839		332		753	

transcription products that have reached the 5′-end of the acceptor RNA will be converted into double-stranded molecules, this assay allows cloning specifically of the products of strand transfer (see the Fig. 2 legend). The number of blue colonies generated is therefore proportional to the number of samples that underwent strand transfer. However, to obtain a frequency of occurrence of strand transfer under these conditions one needs to determine the amount of full-length molecules that would have been obtained if no obligatory strand transfer were required. For this reason, and with the aim of comparing these frequencies with those found under copy choice conditions, reverse transcription was performed in parallel on the same donor RNA used for the copy choice experiments (Fig. 2, *right panel*, *standard sample*). The resulting reverse transcription products were then treated as described above and cloned in *E. coli*. Because the amount of RNA employed in this sample was the same as the forced copy choice (FCC) donor RNA used in the *strand transfer*

sample, the ratio of colonies found in the *strand transfer* sample to colonies found in the *standard* sample gives the frequency of strand transfer under forced copy choice conditions (Fig. 2, *right panel*). This system allows a strict comparison with the frequencies found for copy choice because the same recombinant and parental molecules, respectively, are generated in the two cases.

Effect of the Concentration of the Acceptor RNA—The influence of the concentration of the acceptor RNA on the strand transfer reaction was then studied in parallel under the copy choice and the forced copy choice conditions. This study was carried out not only on E2 RNA but also on two other sequences we previously studied for their ability to promote strand transfer *in vitro*: the region R and a 400-nt segment of the *gag* gene we called "gag1" (18). These two sequences were chosen as representative of another recombination hot spot (the sequence R) and of a region where strand transfer is a rare event (the *gag* region). The recombination assays were performed at a con-

terminated at different incubation times: 1, 3, 10, and 30 min (*A*) and 1, 3, 5, 10, and 30 min (*B*). The position of the pause site on the sequence corresponding to SL is shown in *gray*. The drawings at the bottom schematically indicate which donor and acceptor RNA templates were employed in the assays shown above, using the same representation as in Fig. 3.

stant concentration of donor template (100 nM), varying the concentration of acceptor RNA from 25 nM to 1 μM (Fig. 5). Under the copy choice conditions, for all the three sequences studied the frequency of strand transfer remained constant for ratios higher than 1:1 (Fig. 5*A*). This result is in sharp contrast with the one found under forced copy choice conditions where, for the three sequences studied, the frequency of transfer was clearly dependent on the concentration of acceptor template in the whole range of concentrations tested (Fig. 5*B*).

DISCUSSION

We have studied here the mechanism of copy choice recombination during *in vitro* reverse transcription by HIV-1 RT. In a previous work we conducted a random search for recombination hot spots among various sequences of the HIV-1 genome. That analysis revealed the presence of two strong hot spots, one constituted by the transactivation response element (TAR) hairpin region, the other by a 200-nt-long region coding for the C2 domain of the glycoprotein gp120. Here we have mostly focused on the latter region. We have first determined the folding of this hot spot in solution and shown the direct correlation existing between the high frequency of recombination on this region and the presence of a stable hairpin structure. The ability of the same sequence to promote recombination at different rates depending on the structure adopted has provided the means to assess, for the first time to our knowledge, the respective roles played by stem-loop structures on the donor and on the acceptor RNAs. Indeed, the comparison of the results obtained in the symmetric and asymmetric experiments (Fig. 3) demonstrates that on this RNA sequence the presence of the stem-loop structure on the acceptor RNA appears to be the most important structural element for this region to constitute a recombination hot spot. Strand transfer occurred at high and perfectly comparable frequencies when the hairpin was present on the acceptor RNA regardless of its presence or absence on the donor template (Fig. 3*A*, *columns I* and *II*). It is noteworthy that this observation stands true for naked RNA templates, conditions under which we have determined the folding of the RNA in solution, as well as in the presence of the NC protein, conditions closer to the situation found *in vivo*. The finding that the folding of the acceptor RNA constitutes the crucial parameter for strand transfer is in line with previous observations on the role of NC in the process. In fact, it was shown that the NC, which modulates the folding of the RNA (21, 22), enhances copy choice through its binding onto the acceptor RNA (14).

We previously proposed that the transfer process follows two steps, docking of the nascent DNA onto the acceptor template and displacement by the acceptor RNA of the donor RNA from the 3′-end of the nascent DNA (14). The first step (docking) is expected to depend on the probability of encounter of the acceptor RNA and of the nascent DNA, which is a function of the concentration of these two moieties in solution. The displacement step, in contrast, would involve an acceptor RNA already docked onto the reverse transcription complex and would therefore not depend on the concentration of the acceptor RNA in solution. To distinguish between these two steps we have investigated the response curves of recombination as a function of the concentration of the acceptor RNA on three different sequences. For ratios of acceptor to donor RNAs higher than one, no increase in the frequency of copy choice was observed (Fig. 5*A*). The fact that recombination on the gag sequence, even for high concentrations of acceptor RNA, never reached the values found for Eb or R (Fig. 5*A*) indicates that the efficiency of the process depends on intrinsic features of the RNA considered rather than on the concentration of the acceptor RNA. Because docking is expected to be sensitive to the con-
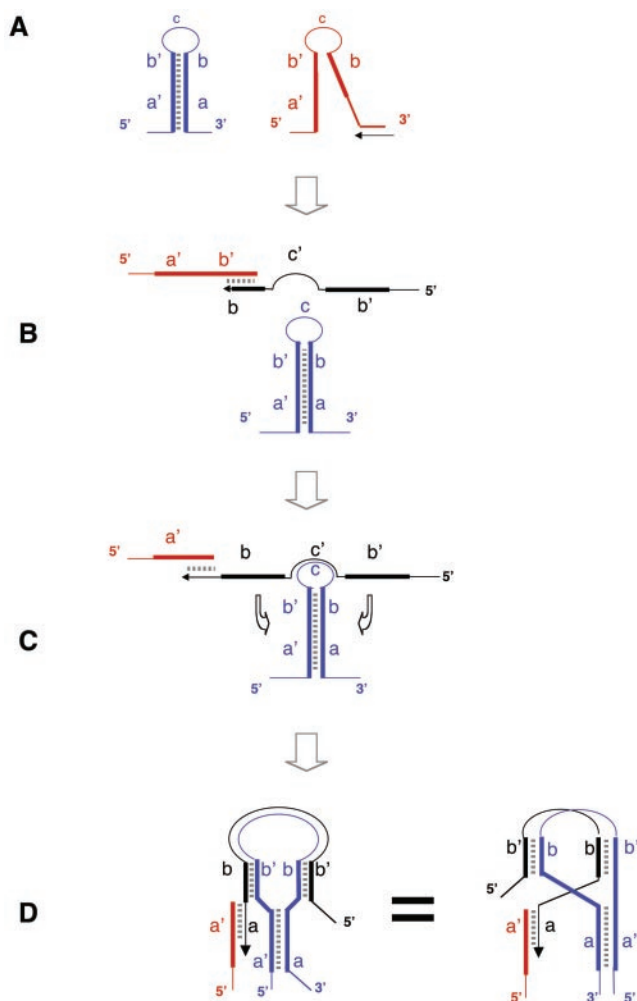

FIG. 6. **The hairpin-mediated strand transfer model for copy choice by RTs.** *A*, the case presented is the one where the donor template is dG1Eb and the acceptor is aE2 (see Fig. 3*A*, *column II*). Donor RNA, *red*; acceptor RNA, *blue*; nascent DNA, *black* (the *arrow* indicates the direction of reverse transcription). *A*, the stem of the SL hairpin on the acceptor RNA is represented as constituted by a lower and an upper part, *a* and *b*, respectively, annealed to their complementary sequences, *a'* and *b'*. The loop region is indicated as *c*. *Region a* corresponds to the Ea sequence in Fig. 1*D* and is therefore absent from the donor RNA. *B*, reverse transcription progresses on the donor RNA. *C*, because annealing of the nascent DNA is likely favored on a single-stranded region of the acceptor RNA, docking might occur once the loop region *c* is reverse-transcribed (generating the complementary sequence *c'*). *D*, the hybridization of the nascent DNA onto the acceptor RNA then invades the stem region of the hairpin on the acceptor RNA, generating the maximum possible extent of double-stranded regions. The resulting intermediate structure redrawn on the *right* of the panel resembles an intermediate of branch migration occurring during DNA recombination (see supplementary information). A migration downward of the position of the crossover would lead to the transfer of the 3′-end of the nascent DNA on the acceptor RNA.

centration of RNA in solution, we reason that the different efficiencies of strand transfer are likely not because of the efficiency of docking. The parallel observation, made in this work, that the frequency of recombination observed on the same sequence (Eb) varies as a function of its folding strongly suggests that the structure of the RNA template is rather responsible for the efficiency of copy choice.

The involvement of hairpin structures in strand transfer has been the object of several studies. It was first suggested that the hairpins formed by these regions could favor template switching by increasing the spatial proximity between the donor and the acceptor RNAs through an intermolecular interac-

tion occurring within the hairpin region (13, 16, 17). Our results do not support such an interpretation because here the hairpin structure is not required on the donor RNA and, therefore, cannot act by mediating an interaction with the acceptor RNA. Furthermore, the response curves given in Fig. 5*A* for copy choice suggest that the spatial proximity of donor and acceptor RNAs is not rate limiting. A recent work on strand transfer on a hairpin region of the genomic RNA of the equine infectious anemia virus has suggested an implication for hairpin structures in the two-step model discussed above (15). In this new model ("dock and lock" model) the docking of the nascent DNA on the acceptor RNA constitutes the limiting step. In this case stalling of reverse transcriptase at the base of the hairpin on the donor RNA improved its degradation by the RNase H activity, thereby increasing the accessibility of the nascent DNA to the annealing for the docking step. In the present work, the hairpin cannot act in a similar way because, if this were the case, its presence would be required on the donor RNA and not on the acceptor. However, the analysis of the pausing pattern during reverse transcription of the SL hairpin has highlighted the presence of a strong pause site in the descending portion of the stem (Fig. 4). This pause might, therefore, act in a similar way by favoring the annealing of the nascent DNA onto the acceptor RNA. The contribution of slowing down reverse transcription in the enhancement of strand transfer both *in vitro* and *in vivo* has, in fact, been previously described (31–33). However, although the presence of this pause site most likely favors strand transfer, it is unlikely that it constitutes the crucial step for the reaction. In fact, the observation that the intensity of pausing on a given donor RNA is not affected by the type of acceptor used (Fig. 4), although the frequency of recombination is clearly different (Fig. 3*A*), manifestly argues against this idea. Furthermore, if stalling of reverse transcription constituted the trigger for the transfer event during copy choice, the response curves as a function of the concentration of acceptor RNA would be expected to be similar to those found under forced copy choice conditions, where a manifest strong pause site is present. Fig. 5 shows instead that this is not the case for any of the sequences studied. We conclude therefore that a pause site is not sufficient to generate recombination hot spots and that it is consequently impossible to predict recombination hot spots simply on the analysis of the pausing pattern found during reverse transcription on the donor RNA.

In light of these results, a possible mechanism accounting for template switching on hairpin regions of the template is proposed in Fig. 6 ("hairpin-mediated strand transfer"). Although this model is elaborated focusing on the most stringent conditions under which Eb constituted a hot spot, the case where the SL hairpin is present only on the acceptor RNA (Fig. 3*B*, *column II*), obviously it also applies when both donor and acceptor RNAs contain a hairpin region. Under the conditions of Fig. 3*B*, *column II*, because the homology between donor and acceptor RNAs begins in the mounting portion of the stem of the SL hairpin, docking must necessarily occur within SL. Furthermore, under these conditions the frequency of strand transfer within SL was as high as when an extended homology between donor and acceptor RNAs was present even before reverse transcription reached the SL hairpin (compare frequency in the hairpin region in Fig. 3*B*, *columns I* and *II*). This observation indicates that the SL region is sufficient to allow efficient docking. As mentioned above, this step could be favored by stalling of the polymerase at the pause site in the descending portion of the stem (Fig. 4). Once docking is achieved, the 3′-end of the nascent DNA has to be transferred onto the acceptor RNA. How the hairpin on the acceptor template is opened to accept the invading strand of the nascent DNA constitutes the main problem in understanding the mechanism of template switching in hairpin regions. Even in the dock and lock model recently proposed, it is difficult to see how the hairpin on the acceptor RNA could be opened and, especially, why invasion should be favored within hairpin regions rather than on poorly structured templates. In this regard, in our model the generation of a structure equivalent to the intermediate that facilitates branch migration during DNA recombination (Fig. 6*D*) offers a plausible solution to this problem. As in that well established case (34), here the gradual opening of the stem portion of the hairpin on the acceptor would be favored by the concomitant formation of alternative double-stranded structures, a feature only possible when a hairpin is present on the acceptor RNA.

In conclusion, this study provides direct evidence for the correlation between recombination hot spots and the structure of the viral genomic RNA. Which specific features of the hairpin structure described here are crucial remains an open question. However, the model proposed here could account for recombination occurring on most hairpin regions, including strand transfer within the transactivation response element hairpin. Obviously this does not exclude that template switching also follows alternative mechanistic paths as shown, for instance, by the residual frequency of recombination found within Eb when both templates are devoid of the SL hairpin (Fig. 3, *column IV*). However, the observation that the main recombination hot spots issued from random searches correspond to hairpin regions (14, 18) suggests that such structures constitute preferential sites for frequent template switching. This study provides a basis for the dissection of the mechanism of template switching at such sites along the HIV-1 genome.

REFERENCES

1. Coffin, J. M. (1992) *Curr. Top. Microbiol. Immunol.* **176,** 143–164
2. Temin, H. M. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90,** 6900–6903
3. Negroni, M., and Buc, H. (2001) in *Annual Review of Genetics* (Campbell, A., ed) Vol. 35, pp. 275–302, Annual Reviews, Palo Alto, CA
4. Sharp, P. M., Bailes, E., Robertson, D. L., Gao, F., and Hahn, B. H. (1999) *Biol. Bull.* **196,** 338–342
5. Hu, W. S., and Temin, H. M. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87,** 1556–1560
6. Vogt, V. M. (1997) in *Retroviruses* (Coffin, J. M., Hughes, S. H., and Varmus, H. E., eds) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
7. Jung, A., Maier, R., Vartanian, J. P., Bocharov, G., Jung, V., Fischer, U., Meese, E., Wain-Hobson, S., and Meyerhans, A. (2002) *Nature* **418,** 144
8. Vogt, P. K. (1973) in *Possible Episomes in Eukaryotes* (Silvestri, L., ed) pp. 35–41, North Holland
9. Coffin, J. M. (1979) *J. Gen. Virol.* **42,** 1–26
10. Telesnitsky, A., and Goff, S. P. (1993) *EMBO J.* **12,** 4433–4438
11. DeStefano, J. J., Mallaber, L. M., Rodriguez-Rodriguez, L., Fay, P. J., and Bambara, R. A. (1992) *J. Virol.* **66,** 6370–6378
12. Wu, W., Blumberg, B. M., Fay, P. J., and Bambara, R. A. (1995) *J. Biol. Chem.* **270,** 325–332
13. Kim, J. K., Palaniappan, C., Wu, W., Fay, P. J., and Bambara, R. A. (1997) *J. Biol. Chem.* **272,** 16769–16777
14. Negroni, M., and Buc, H. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97,** 6385–6390
15. 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
16. Berkhout, B., Vastenhouw, N. L., Klasens, B. I., and Huthoff, H. (2001) *RNA* **7,** 1097–1114
17. Balakrishnan, M., Fay, P. J., and Bambara, R. A. (2001) *J. Biol. Chem.* **276,** 36482–36492
18. Moumen, A., Polomack, L., Roques, B., Buc, H., and Negroni, M. (2001) *Nucleic Acids Res.* **29,** 3814–3821
19. Welker, R., Hohenberg, H., Tessmer, U., Huckhagel, C., and Krausslich, H. G. (2000) *J. Virol.* **74,** 1168–1177
20. Negroni, M., and Buc, H. (2001) *Nat. Rev. Mol. Cell. Biol.* **2,** 151–155
21. Clodi, E., Semrad, K., and Schroeder, R. (1999) *EMBO J.* **18,** 3776–3782
22. Williams, M. C., Rouzina, I., Wenner, J. R., Gorelick, R. J., Musier-Forsyth, K., and Bloomfield, V. A. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98,** 6121–6126
23. Robertson, D. L., Sharp, P. M., McCutchan, F. E., and Hahn, B. H. (1995) *Nature* **374,** 124–126

24. Ye, Y., Si, Z. H., Moore, J. P., and Sodroski, J. (2000) *J. Virol.* **74,** 11955–11962
25. Quinones-Mateu, M. E., Gao, Y., Ball, S. C., Marozsan, A. J., Abraha, A., and Arts, E. J. (2002) *J. Virol.* **76,** 9600–9613
26. Ehresmann, C., Baudin, F., Mougel, M., Romby, P., Ebel, J. P., and Ehresmann, B. (1987) *Nucleic Acids Res.* **15,** 9109–9128
27. Zuker, M. (1989) *Science* **244,** 48–52
28. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
29. Negroni, M., Ricchetti, M., Nouvel, P., and Buc, H. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92,** 6971–6975
30. De Rocquigny, H., Gabus, C., Vincent, A., Fournie-Zaluski, M. C., Roques, B., and Darlix, J. L. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89,** 6472–6476
31. DeStefano, J. J., Buiser, R. G., Mallaber, L. M., Fay, P. J., and Bambara, R. A. (1992) *Biochim. Biophys. Acta* **1131,** 270–280
32. Svarovskaia, E. S., Delviks, K. A., Hwang, C. K., and Pathak, V. K. (2000) *J. Virol.* **74,** 7171–7178
33. Hwang, C. K., Svarovskaia, E. S., and Pathak, V. K. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98,** 12209–12214
34. Hiom, K. (2001) *Curr. Biol.* **11,** 278–280

Evidence for a Mechanism of Recombination during Reverse Transcription Dependent on the Structure of the Acceptor RNA

Abdeladim Moumen, Lucette Polomack, Torsten Unge, Michel Véron, Henri Buc and
Matteo Negroni

Access the most updated version of this article at doi: 10.1074/jbc.M212306200

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

Supplemental material:

http://www.jbc.org/content/suppl/2003/03/21/M212306200.DC1

This article cites 31 references, 19 of which can be accessed free at
http://www.jbc.org/content/278/18/15973.full.html#ref-list-1