Nucleic Acids Research, 1998, Vol. 26, No. 14 **3433–3442**

Pausing of reverse transcriptase on retroviral RNA templates is influenced by secondary structures both 5′ and 3′ of the catalytic site

Geoffrey P. Harrison[+], Matthew S. Mayo[1], Eric Hunter and Andrew M. L. Lever[2,*]

Department of Microbiology, University of Alabama at Birmingham, 845 19th Street South, Birmingham, AL 35294, USA, [1]Kansas Cancer Institute and Department of Preventive Medicine, The Kansas University Medical Center, 3901 Rainbow Boulevard, Kansas City, KS 66160-7313, USA and [2]University of Cambridge Department of Medicine, Addenbrooke's Hospital, Hills Road, Cambridge CB2 2QQ, UK

Received February 2, 1998; Revised and Accepted June 3, 1998

ABSTRACT

In the most extensive examination to date of the relationship between the pausing of reverse transcriptase (RT) and RNA secondary structures, pause events were found to be correlated to inverted repeats both ahead of, and behind the catalytic site *in vitro*. In addition pausing events were strongly associated with polyadenosine sequences and to a lesser degree diadenosines and monoadenosine residues. Pausing was also inversely proportional to the potential bond strength between the nascent strand and the template at the point of termination, for both mono and dinucleotides. A run of five adenosine and four uridine residues caused most pausing on the HIV-1 template, a region which is the site of much sequence heterogeneity in HIV-1. We propose that homopolyadenosine tracts can act as termination signals for RT in the context of inverted repeats as they do for certain RNA polymerases.

INTRODUCTION

Retroviruses replicate their single-stranded (ss; + sense) RNA genomes into double-stranded (ds)DNA by a mechanism involving RNA and DNA templated polymerase and RNase H activities (1). This process is carried out by the virally encoded, multifunctional enzyme, reverse transcriptase (RT) (2–4) which initiates minus-strand DNA synthesis from a primer tRNA annealed to the viral genome at the primer binding site. It polymerizes a short distance until it reaches the 5′ end of the RNA template. The nascent DNA strand then transfers to the 3′ end of the genome and RT completes minus-strand DNA synthesis of the remaining 8.9 kb RNA template. RNase H activity selectively degrades the genomic RNA strand and creates ribonucleotide primers that are used for initiation of plus-strand DNA in two 5 kb segments resulting in plus-strand discontinuity in unintegrated linear DNA (5,6). Full length dsDNA is observed 4–8 h after infection (7,8). One explanation for high mutation rates in general in retroviruses and for the extensive heterogeneity exhibited by HIV-1 in particular is inaccurate replication by RT which is proof-reading deficient.

Retroviruses encapsidate two copies of the single-stranded + sense RNA genome which exist in mature virions as a stable dimer, linked near the 5′ end of the genome (9–11). Encapsidation of two different sequences as a heterodimer allows recombination during the process of replication as the nascent DNA strand can transfer to a similar or identical sequence on the other RNA template during extension. A high frequency of recombination has been observed from internal regions of the viral genome (12–17). This process, along with inaccurate replication, increases the genetic diversity in the virus population (9,18) allowing some viruses to evade the immune response and antiviral drugs. The factors influencing strand transfer are not understood. Analysis of polymerization products of RT shows reproducible patterns of early termination events (20,21). Pausing and premature termination of RT (hereafter referred to as pausing) have been associated with frameshifting (22) and template switching (23). Previous analyses of RT pausing have indicated that primary sequence context can influence pausing (21,24). Klarmann *et al.* (21) reported that homopolymeric runs of four or more cause most pausing. Ji *et al.* (25) reported that HIV-RT specifically paused at eight poly(rA) tracts in the *env* gene. In addition, pausing has been attributed to poor bonding between template and primer (25).

To investigate the factors that influence pausing of AMV RT, the frequency of pausing was recorded for cDNAs synthesized on RNAs encoding the HIV-1 and the Mason-Pfizer monkey virus (M-PMV) 5′ leader sequences. The RNA secondary structure of the TAR stem–loop of HIV-1 has been reliably modelled (26–30). The RNA secondary structure of the HIV-1 encapsidation signal region (Ψ) between the HIV-1 PBS and the beginning of the gag open reading frame has been investigated by ourselves (31) and others (32–38). The M-PMV Ψ has also been investigated by ourselves (44). This paper reports results of analysis of pausing of AMV RT over 476 nt in three retroviral RNA templates, with well-known RNA secondary structures. By offsetting the pause data for each of the three template sequences relative to the structural data, it was found that the enzyme pauses as it approaches RNA secondary structures. Potential structures behind the catalytic site either in the nascent strand or in the template were also shown to be associated with pausing of the enzyme. When the data were not offset, pausing of AMV RT occurred most frequently at adenosine residues within

*To whom correspondence should be addressed. Tel: +44 1223 336747; Fax: +44 1223 336846; Email: amll1@mole.bio.cam.ac.uk

[+]Present address: Ribotargets Ltd, Kett House, Station Road, Cambridge CB1 2JX, UK

homoadenosine tracts. Diadenosines and monoadenosines were also found to be associated with a high frequency of pausing. The frequency of pausing was associated with poor potential base pairing between the nascent cDNA and the template. Pausing patterns of AMV RT and HIV-1 RT show considerable similarities. Since RT is a major target for antiviral chemotherapy in HIV infection, information contained in this report contributes to understanding RT processivity which is of importance in the development and improvement of antiviral agents.

MATERIALS AND METHODS

The HIV-1 sequence was derived from pSVC21, an infectious proviral clone isolated originally from pHXB2 (39). The numbering of the HIV-1 sequence is that of the Los Alamos database (40). The numbering of M-PMV is that of Genbank accession no. M12349. Three templates were used for cDNA synthesis: (i) the HIV-1 TAR region from nucleotides 313 to 830, (ii) the HIV-1 Ψ region from bases 541 to 1086 and (iii) the M-PMV Ψ region from bases 567 to 1561.

The template for the expression of RNA encoding the HIV-1 TAR stem–loop was the construct pKSIIΨCS (41). This contains the *Sca*I–*Cla*I fragment of HIV-1 from nucleotides 313 to 830, ligated into the *Eco*RV and *Cla*I sites of the vector Bluescript KSII (Stratagene, UK). This plasmid was linearised with *Xho*I, and RNA was transcribed from the T7 promoter. RNA encoding the HIV-1 sequence between the primer binding site and the beginning of the *gag* open reading frame was transcribed from the plasmid pKSIIH3T3. This was constructed by excising the *Hin*dIII fragment of HIV-1 from bases 541 to 1086 and cloning it into the *Hin*dIII site of the vector Bluescript KSII (Stratagene). The plasmid was linearised with *Xba*I, and RNA was transcribed *in vitro* from the T3 promoter.

The 5′ leader region of M-PMV was excised from the proviral clone pSHRM15 (42) using the *Sph*I site at position 567 and the *Sac*I site at position 1561. This fragment was ligated into the same sites in the expression vector pGem-4Z (Promega, Southampton, UK) and RNA was transcribed from the T7 promoter.

Deoxyoligonucleotide primers for the HIV-1 template were: 5′-GTC ACA CAA CAG ACG GGC-3′ (nt 579–562), 5′-ACC AGT CGC CGC CCC-3′ (nt 744–730), 5′-ATC TCT CTC CTT CTA GCC-3′ (nt 790–773) and 5′-CTA ATT CTC CCC CGC-3′ (nt 828–814).

Deoxyoligonucleotide primers for the M-PMV template were: 5′-TTG CCC CAT ATC CGA GCG C-3′ (nt 898–880), 5′-CCC CGT GTC TTT AAA GCC-3′ (nt 957–939) and 5′-TAT GGT TCC CTC TTG CGG-3′ (nt 1042–1025).

Reverse transcription

cDNA synthesis reactions were carried out as described in Harrison and Lever (31). Briefly, *in vitro* transcribed RNAs were extracted with phenol/chloroform, and 10 ng of a synthetic oligonucleotide primer was added in avian myeloblastosis virus (AMV) RT buffer (Promega). The final concentrations in the reaction buffer supplied by Promega for AMV RT were 50 mM Tris–HC1 pH 8.3, l mM DTT, 7 mM $MgCl_2$, 40 mM KC1, 0.l mg/ml BSA. HIV-1 RT, (a p66/p51 heterodimer) was a kind gift from S.McPherson of the Center for AIDS research core facility at the University of Alabama at Birmingham. The final concentrations in the reaction buffer which were used for HIV-1 RT were 70 mM Tris–HCl pH 8.0, 2 mM DTT, l0 mM $MgCl_2$, 80 mM KCl. The RNAs were dissociated by heating them at 70°C for 5 min in their RT buffers. They were then allowed to cool to room temperature over a period of 20 min. cDNA synthesis reactions were carried out by adding 1 U of AMV RT (Promega) (or 1 U of HIV-1 RT), a final concentration of 1 mM of each dATP, dTTP and dGTP and 1 µl [^{32}P]dCTP. Reaction mixes were incubated at 42°C for 1 h. When experiments involved direct comparisons between HIV-1 and AMV RT the incubation temperature for both enzymes was 37°C to eliminate the effect of temperature. cDNAs were precipitated under ethanol and were re-dissolved in 1× TE buffer. An equal volume of formamide dye mix was added and approximately equal amounts of radioactivity were loaded onto 6% polyacrylamide–7 M urea gels along with dideoxy-sequencing ladders of template DNA, primed using the same oligonucleotide primer that had been used for cDNA synthesis (43).

Pausing was statistically analysed on the basis of presence or absence of a pause site. Pausing at dinucleotides was analysed by summing the number of pausing events within the pair. The frequencies of pauses at dinucleotides were analysed by allocating a score of zero to those nucleotides at which no pausing at any level was observed, those dinucleotides in which a pause occurred at either of the two nucleotides were given a frequency score of one, and those dinucleotides at which a pause was observed at both nucleotides were allocated the frequency score of two. The expected number of pauses in dinucleotides was calculated thus: for the dinucleotide AA of which there were 42 in the templates, 16 pairs had no pauses at all (16×0), two pairs had one pause between them (2×1) and 24 pairs had pauses at both nucleotides (24×2); therefore, the mean score was determined to be $(0 + 2 + 48)/42 = 1.19$. The enzyme proceeds 3′→5′ along the RNA template, therefore dinucleotides are written 3′→5′, where the 5′ nucleotide is the position of the termination event. We note that Shelness and Williams (45) wrote dinucleotides 5′→3′, where the 3′ base was the point of termination. In this study, the term homopolymeric refers to a run of three or more of the same nucleotide.

The data were sorted according to the potential hydrogen bond strength between the nascent strand and the template at each mononucleotide and each dinucleotide. Thus, the scores for the mononucleotides adenosine and uridine were grouped against the cytosine and guanosine scores. The relative potential bond strength given to adenosines and uridines was two and that for cytosines and guanosines was three. For dinucleotides the relative bond strengths for each nucleotide in the pair were summed, creating three categories, with combined potential bond strengths of four, five and six.

In order to analyze the effect of RNA secondary structure on the processivity of RT, nucleotides were given one of seven codes according to their predicted position within RNA structures. Structures were determined by considering the phylogenetic and biochemical data for these sequences and the results of free energy minimization computations which we and others have reported previously (27,31,32,35,44). The seven categories into which nucleotides (and their pausing) were sorted were as follows (see also Fig. 9A): 1, unpaired and not within any local structure; 2, the 3′ side of a loop at the end of a double-stranded stem; 3, the 5′ side of a loop at the end of a double-stranded stem; 4, a bulge on the 3′ side of a double-stranded stem; 5, a bulge on the 5′ side of a double-stranded stem; 6, the 3′ side of a base-paired stem; 7, the 5′ side of a base-paired stem.

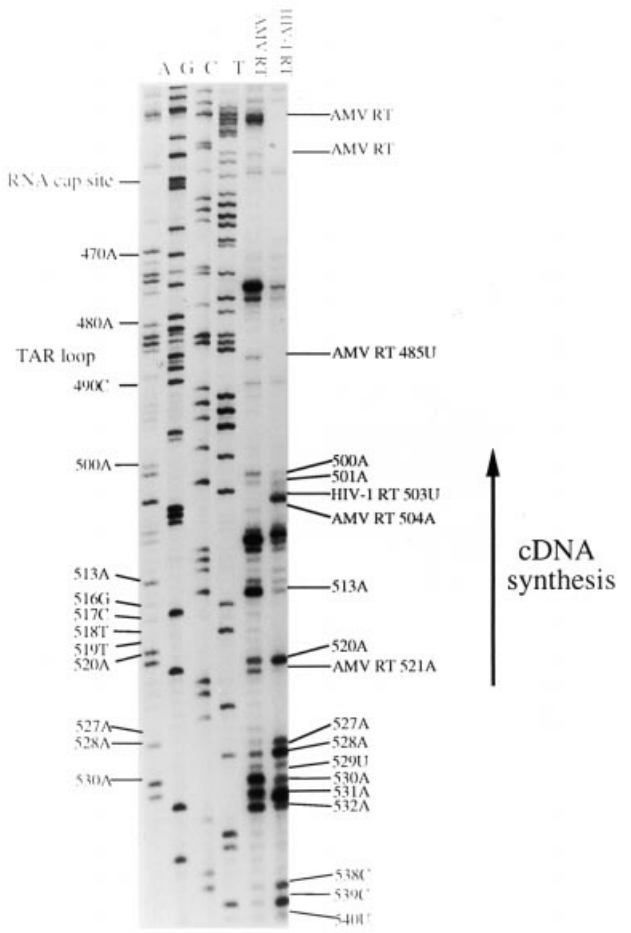

Figure 1. Comparison between pausing of AMV RT and pausing of HIV-1 RT on an *in vitro* transcribed RNA template. Autoradiograph of a gel comparing pausing by AMV RT and HIV-1 RT during cDNA synthesis on an HIV-1 RNA template (isolate HXBc2). cDNAs were electrophoresed against a dideoxysequencing ladder generated on ssDNA template of the same sequence. The sequence of the primer used in all lanes was 5′-GTCACACAACA-GACGGGC-3′ (nt 579–562). Reverse transcription gives rise to cDNAs 1 nt shorter than their RNA template. The dideoxysequencing ladder is labeled on the left. Pausing events unique to either RT are labeled on the right.

In addition to comparing the frequency of pausing within the seven groupings of RNA structures, groups were combined and compared with one another thus: groups 1–4 (unpaired nucleotides) were compared with groups 6 and 7 (paired nucleotides), and groups 2, 4 and 6 (3′ sides of structures) were compared with groups 3, 5 and 7 (5′ sides of structures).

Offsetting of pause data

In order to investigate the effects of RNA secondary structures at several different positions ahead of and behind the enzyme, the pause data were offset by 2, 4, 6, 8, 10 and 12 nt relative to the structural code (given above) which had been given to each nucleotide. The sample size consequently decreased as some of the pause data were offset past the ends of each of template. The pause data were then sorted according to the structural code to which they had been re-aligned.

The relationship between sequence variations and pausing was investigated by comparing the number of variations in HIV-1 isolates from clades A, B and D (40) with pausing events on the HIV-1 template, and by comparing pausing on the M-PMV template with variations between M-PMV and the simian retroviruses 1 and 2 (SRV1 and SRV2).

Statistical analysis

Statistical analyses were performed on a Gateway P5-120 PC using the Statistical Analysis System, SAS Windows version 4.0.950 release 6.11. Comparisons of the locations of pausing in various groupings of the data (nucleotide, bond strength, etc.) were compared using the Chi-squared test, except in cases when expected cell counts were small, then Fisher's exact test was performed, using the frequency procedure.

RESULTS

Comparison of pausing of AMV RT with HIV RT

cDNAs were synthesized on an *in vitro* transcribed HIV-1 RNA template from nucleotides 313 to 830 using an oligonucleotide primer from nucleotides 579 to 562 (5′→3′). Pausing events are shown for the two RTs side by side in Figure 1, and observational scoring of pausing intensities are shown in Figures 2–4. The pausing patterns of the cDNAs were broadly similar with certain exceptions in site and intensity.

The locations of pauses during cDNA synthesis were analysed over 476 nucleotides on three *in vitro* transcribed RNA templates. Representative examples of pausing events from two out of the 60 gels analysed are presented in Figures 1 and 5. Other figures showing AMV RT pausing events on two of these templates have been published previously (31,44). The locations of pausing events on predicted RNA secondary structures are represented in Figures 2–4 with an indication of their intensities. Results of analyses of the location of pausing are shown in Figures 6 and 7.

The effect of primary sequence

Effect of homopolymeric runs. Pausing occurred within different homopolymeric runs on these templates. For example, on one of the HIV-1 templates, four Us from bases 762 to 759 and five As from bases 758 to 754, constitute a reproducible high frequency pause site. On the M-PMV template, three Gs from bases 847 to 845 do likewise. Pausing was more frequent within homopolymeric runs than within heteropolymeric runs (Figs 6H and 7G). The most frequent pauses of all occurred within homopolymeric runs of adenosines ($P < 0.0005$) (Figs 6I and 7H). Ten out of 24 homopolymeric runs were not associated with pausing, two others contained very weak pauses. There were a large number of pauses on the HIV-1 template at the sequence 3′-AGUUUUAAA-AA-5′ (nt 764–754). However, a relatively similar sequence on the M-PMV template 3′-AGUUUUUUAAA-5′ (nt 996–989) only resulted in pausing at the A residues (nt 996–986) and not at the following four U residues. The M-PMV sequence from 828 to 814 contains three runs of three adenosines, but very little pausing occurred at these A residues in this context.

Mononucleotide type. When locations of pausing were analysed across all the data from both viral templates, AMV RT was found to pause more frequently at monoadenosines than at any other nucleotide (Figs 6A and 7A). The differences between the frequency of pausing at A residues and the frequency of pausing at G and U

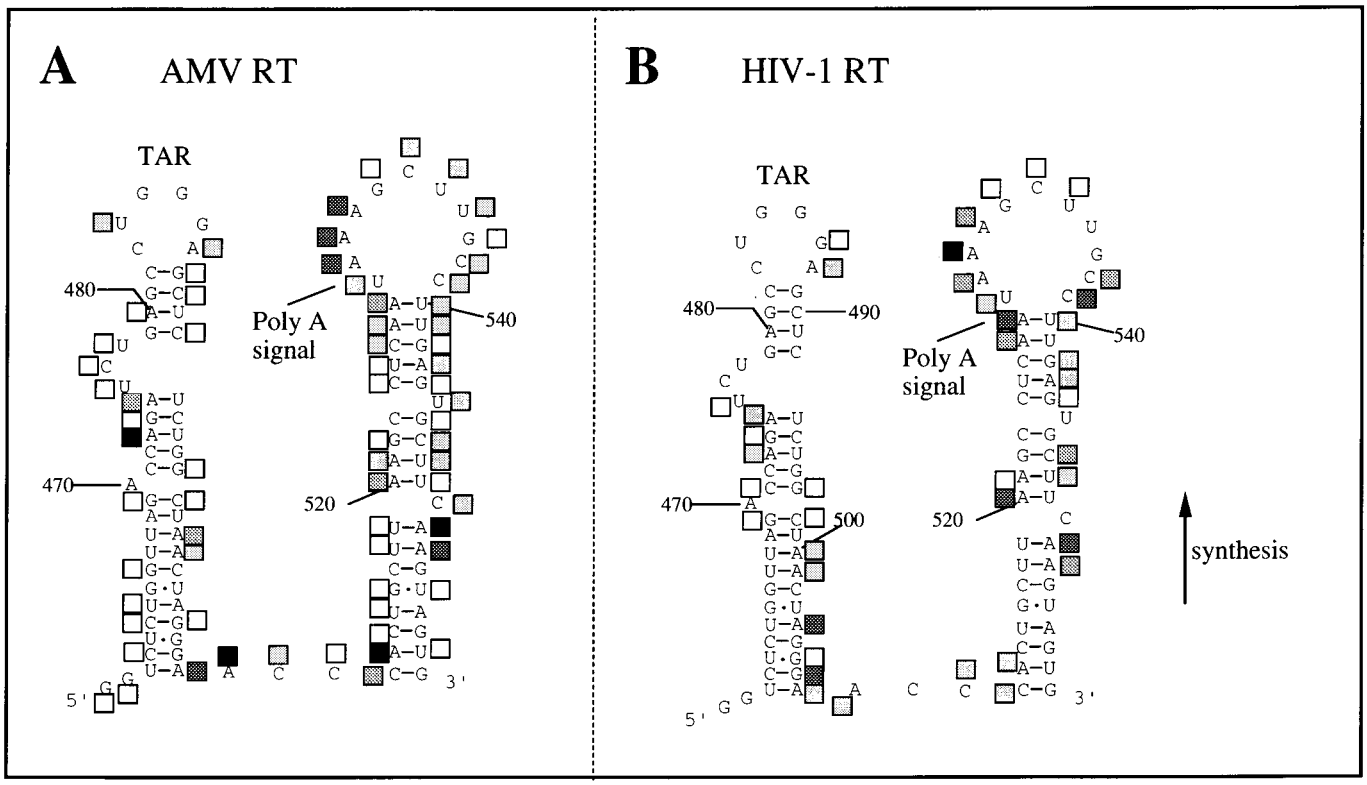

Figure 2. The locations of RT pausing events on the HIV-1 TAR region from nucleotides 457 to 548. Boxes (□) indicate the positions and approximate intensities of pausing events. The more intense the pausing, the darker the shading.

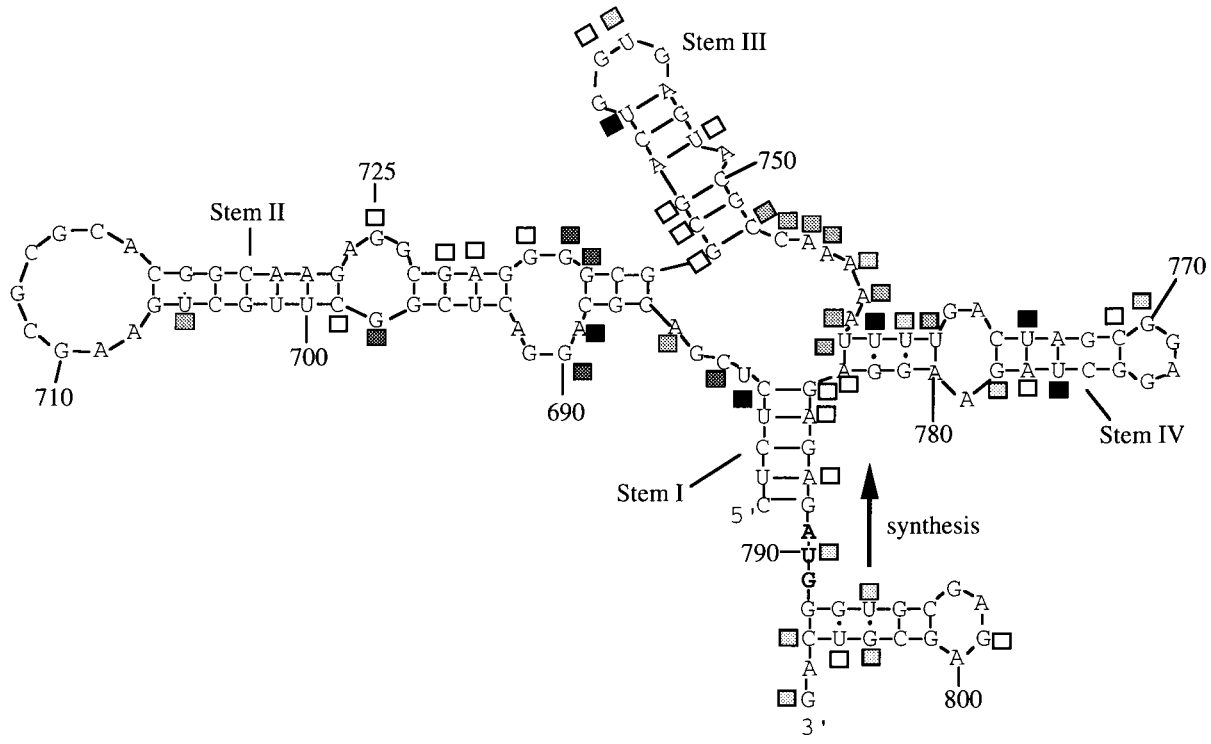

Figure 3. The locations of RT pausing events on the HIV-1 Ψ region. Boxes (□) indicate the positions and approximate intensities of pausing events. The more intense the pausing, the darker the shading.

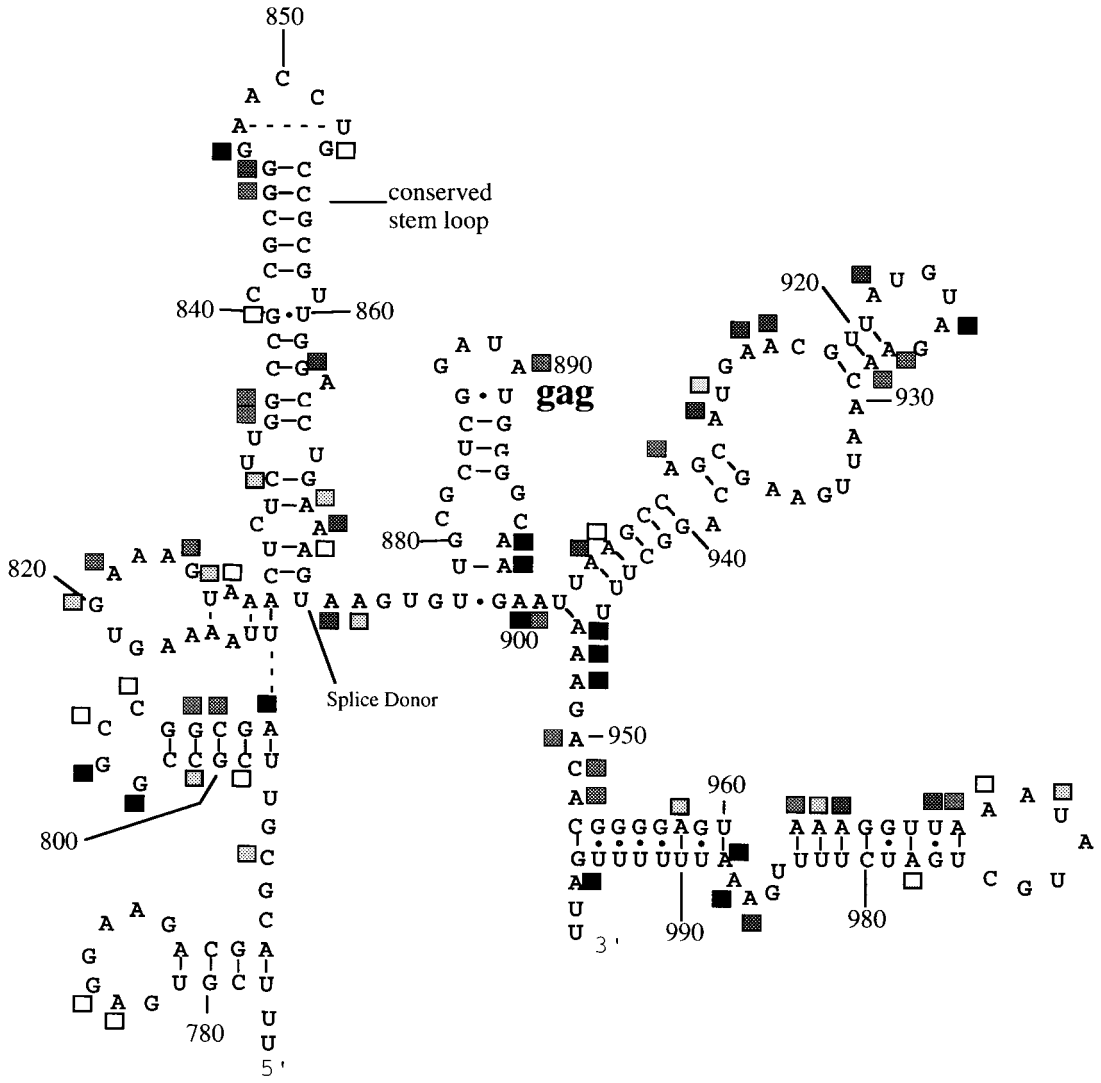

Figure 4. The locations of AMV RT pausing events on the M-PMV Ψ region. Boxes (□) indicate the positions and approximate intensities of pausing events. The more intense the pausing, the darker the shading.

residues were highly significant ($P < 0.00l$), whereas the differences between the frequency of pausing at C, G and U residues were not.

Monopurines versus monopyrimidines. When mononucleotide pause scores were analysed by comparing the frequency of pausing at monopurines versus monopyrimidines, the frequency of pausing at monopurines was significantly higher ($P = 0.028$) than the frequency of pausing at monopyrimidines (Figs 6B and 7B). The mean frequency of pausing for monopurines was also significantly higher than that for monopyrimidines (Fig. 7B). These results were influenced by the dominance of pausing at adenosine residues.

Effect of the preceding nucleotide on mononucleotide pausing. When the score at the point of termination (the 5′ nucleotide) was analysed in the context of the 3′ nucleotide (which the RT encounters first) it was found that the frequency of pausing events was higher ($P = 0.022$) at the dinucleotide 3′-AN-5′ (Fig. 6C). The scores for AN and AC were significantly different ($P < 0.05$).

Effect of mononucleotide bond strength. When the potential bond strength between the template and the nascent strand was analysed, mononucleotides with the lower bond strength were found to have a significantly higher frequency of pausing events ($P = 0.016$) (Figs 6D and 7C). This result was influenced by the dominance of pausing at adenosine residues.

Effect of dinucleotide type. There were significant differences between the amount of pausing at different dinucleotide pairs ($P = 0.000001$). A large number of pausing events occurred at the dinucleotide 5′-AA-3′ followed by AC, and UA (Figs 6E and 7D). All but one of the dinucleotides most frequently associated with pauses contained an adenosine residue. Over all the data, pausing was higher at dinucleotides containing an adenosine than at those which did not. The lowest score for the frequency of pausing was for the dinucleotide GC.

Effect of pyrimidines and purines on pausing at dinucleotides. There were significant differences between the amount of pausing at the four combinations of purines and pyrimidines in dinucleotide pairs ($P = 0.013$). Pausing occurred most frequently at double purines, and least frequently at double pyrimidines (Figs 6F and 7E). The effect is probably influenced by the dominance of

pausing at dinucleotides containing an adenosine compared with the frequency of pausing at all those containing a guanosine.

Effect of dinucleotide bond strength. When the dinucleotides were sorted according to their combined relative hydrogen bond strengths (for bonding to the nascent strand), significant differences were found between the levels of pausing at the three potential H bond strengths (P = 0.00015). Increased frequency of pausing was significantly correlated to lower bond strength (Figs 6G and 7F). These results were influenced by the frequency of pausing at adenosine residues, since there was a large difference between the pausing at the dinucleotides AA and UU and the homopolymeric sequences poly(A) and poly(U).

Relationship between sequence heterogeneity and pausing

The locations of pausing events were examined in the light of known sequence heterogeneity in HIV-1 isolates, and differences between M-PMV and the closely related D-type retroviruses SRVl and SRV2. When the relationship between the positions of AMV RT pause sites and variations between the non-coding regions of published HIV-1 sequences were examined, the site with the most striking pausing (of all three templates) was between HIV-1 bases U761 and G751 which includes a run of four Us and five As. This is also a site of much sequence heterogeneity amongst HIV-1 isolates (Fig. 8). Clusters of variations occur in the HIV-l sequence from nucleotides 751 to 763 and 815 to 822, which coincide with homopolymeric runs. A run of five Gs from 820 to 816 directly coincides with eight variations in sequences. Six of these are changes to A, and the other two are changes to C. Just 3′ to this homopolymeric run at 822G, 32 out of 59 isolates from clades A and B have an adenosine and two have a cytosine, the consensus nucleotide 822 is an adenosine for clade A. At HIV-l nucleotide 815C, just 5′ to the homopolymeric run of Gs at HIV-1 HXB2, 11 out of 59 isolates from clades A and B have a cytosine and 17 have a uridine. However, many locations of sequence variations were not sites of pauses, and vice versa.

The effect of secondary structure in the template

It is generally assumed that some RT pausing is attributable to the enzyme encountering RNA secondary structure in a template. This is the first study in which this possibility has been closely studied at well-known RNA structures. These were HIV-1 TAR, from bases 459 to 508, (26–30,46), the RNA secondary structure of the HIV-1Ψ (31–36) and the M-PMV Ψ from nucleotides 842 to 858 (44). The existence of RNA secondary structures in stem II (bases 701–723) and stem IV (bases 765–778) of the HIV-1 HXB2 Ψ region have been confirmed by the introduction of disruptive and compensatory mutations into proviruses, by ourselves using Jurkat tat cells (46), and others, in non-host cells (37,38). In Figure 3 we present the sequence of the infectious clone HXB2 (38). Stem loop IV of the HIV-1 (from bases 782 to 765) has been widely predicted and extra weight is given to these predictions by the fact that stem–loops II (G778 to C764) and IV (G723 to C701) can be formed with compensatory mutations at this region in Cameroonian isolates of HIV clade 0 (48,49). In the M-PMV RNA template, the stem–loop from bases 840 to 860 can be formed in the equivalent region of squirrel monkey retrovirus despite only 42% primary sequence homology (44). This phylogenetic evidence gives strong support to the existence and functional importance of this stem *in vivo*.

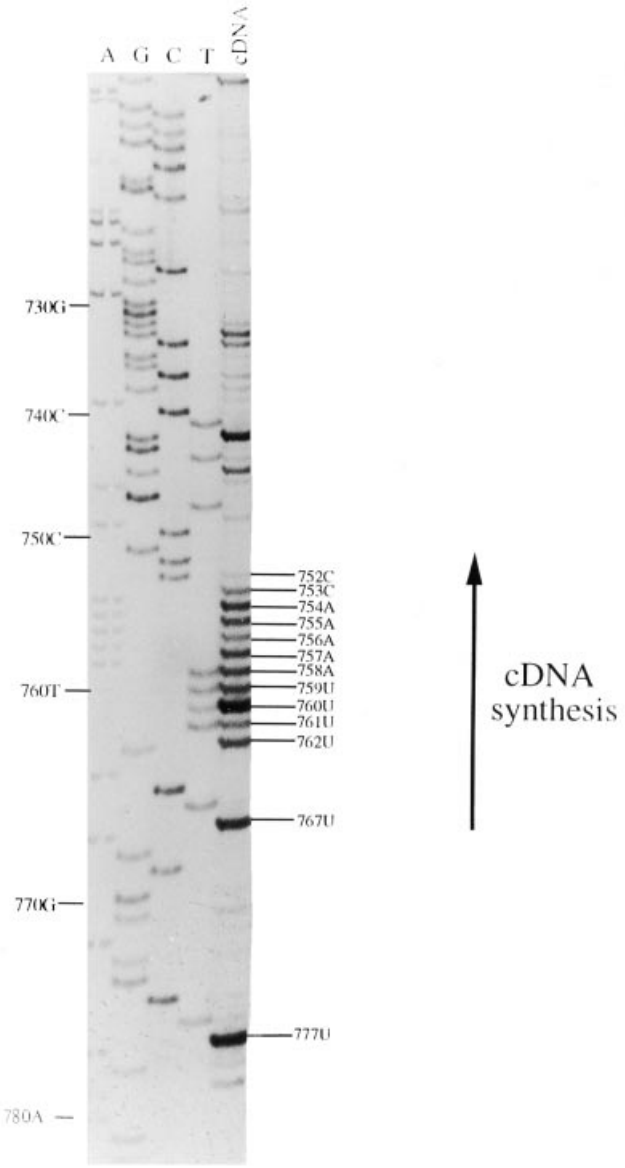

Figure 5. Pausing of AMV RT on an *in vitro* transcribed RNA template. Autoradiograph of a gel showing pausing events during cDNA synthesis by AMV RT (right-hand lane) on an HIV-1 RNA template (isolate HXBc2). cDNAs were electrophoresed against dideoxysequencing ladders generated on a ssDNA template of the same sequence. The sequence of the primer used in all lanes was 5′-CTAATTCTCCCCCGC-3′ (nt 828–814). Pauses were scored visually on a scale of 1–6, a score of 6 represents the most intense bands. Reverse transcription gives rise to cDNAs 1 nt shorter than their RNA template.

When all the nucleotides in all three of the templates were sorted into seven groups according to their location within predicted RNA structures (Fig. 9A) no association was found between the position within predicted RNA structures and the frequency of pausing events or their intensity (Fig. 10, position 0). The enzyme first encounters template secondary structure not at the catalytic site, but in the finger domain. This domain in HIV-1 RT (p66) is in intimate contact with its template up to six nucleotide positions ahead of the catalytic site (50,51); therefore, the effect of RNA secondary structure ahead of the enzyme on pausing might be offset 5′ on the template by ~6 nt. In order to investigate this, the locations

A Mononucleotide scores

	n	$\bar{x}$	s.e.
A	124	1.8	0.2
U	105	0.6	0.1
C	94	0.7	0.1
G	152	0.7	0.1
totals	475	1.0	

B Mononucleotide scores grouped by **nucleotide type**

	n	$\bar{x}$	s.e.
Pu	276	1.2	0.1
Py	199	0.6	0.1

C 3' nucleotide score sorted by **5' nucleotide**.

3'→5'	n	$\bar{x}$	s.e.
AN	124	1.3	0.2
GN	151	0.9	0.1
UN	106	0.9	0.1
CN	94	0.7	0.1

D Mononucleotide scores grouped by **bond strength**

Bond	n	$\bar{x}$	s.e.
A+U 2	229	1.3	0.1
G+C 3	246	0.7	0.1

E Dinucleotide score by pair

3'→5'	n	$\bar{x}$	s.e.
AA	42	4.2	0.6
AC	15	3.7	0.7
UA	19	2.8	0.5
AU	21	2.2	0.5
AG	46	2.1	0.3
GG	43	2.0	0.4
GA	43	2.0	0.3
CU	16	1.7	0.4
CA	20	1.7	0.5
UC	32	1.5	0.3
GU	35	1.3	0.3
CG	41	1.2	0.3
CC	17	1.2	0.4
UU	33	1.1	0.4
GC	30	1.0	0.3
UG	22	1.0	0.3

F Dinucleotide scores sorted by **nucleotide type**.

3'→5'	n	$\bar{x}$	s.e.
Purines	174	2.6	0.2
Pyrimidines	98	1.4	0.2
Pu/Py	101	1.8	0.2
Py/Pu	102	1.6	0.2

G Dinucleotides grouped by **bond strength**

	bond	n	$\bar{x}$	s.e.
AA, AU, UA, UU.	4	115	2.7	0.3
AC, AG, CA, CU GA, GU, UC, UG	5	229	1.8	0.1
GG, CC,CG, GC	6	131	1.4	0.2

H Homopolymeric *versus* non-homopolymeric

	n	%	s.e.
n<2	386	0.9	0.1
n>2	88	1.5	0.2

I Mean scores within **homopolymeric runs**

	n	$\bar{x}$	s.e.
A	30	2.5	0.4
G	29	1.0	0.3
C	6	1.0	0.5
U	21	0.7	0.3

J Percentage **at each scoring level**
0= no pauses; 6= strongest pauses

Adenine Score	%	Uridine Score	%
6	3	6	1
5	9	5	2
4	11	4	4
3	15	3	3
2	8	2	13
1	10	1	12
0	44	0	65

Guanine Score	%	Cytosine Score	%
6	0	6	1
5	3	5	0
4	4	4	1
3	4	3	4
2	6	2	13
1	16	1	16
0	68	0	65

Figure 6. Analyses of locations of pausing events, independent of intensity. Pausing for all three templates. Pause intensity was not included in these calculations. The expected number of pauses in dinucleotides was calculated thus: for the dinucleotide AA, of which there were 42 in the templates, 16 pairs of AA had no pauses at all (16 × 0), two pairs had one pause between them (2 × 1) and 24 pairs had pauses at both nucleotides (24 × 2); therefore, the mean score was determined to be (0 + 2 + 48)/42 = 1.19.

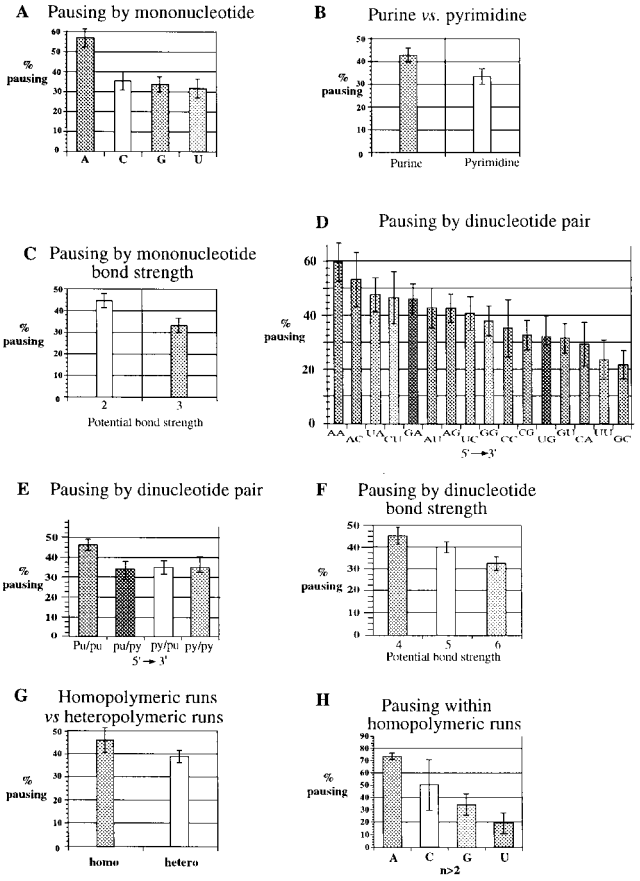

Figure 7. Analyses of locations of pausing events, independent of intensity. Mean percentage pausing for all three templates.

AAAAAUUUU	Clade B consensus
AAAAAAUUUUU	YU2/10
AUUUAUUUU	CAM1
AAAAAUU**C**UU	MN
AAAAAUUUUU	OYI/3202A21
GAAAAUUUU	RF
AAAAUUUU	HAN
GAAAUUUU	JRCSF/WEAU160
AAAACUUU	3202A12
AAAUUUUU	P896
AAUUUUU	SF2
UAAAAUUUUU	ELI Clade D
UAAAAAUUUU	Z2Z6 Clade D
UGAAAUUUUU	NDK Clade D
CAAUUUUU	MAL Clade U
CUAAAAUUUUUUU	U455 Clade A
UAAAAUUUU	IBNG Clade A

Figure 8. Variations in the HIV-1 sequence at a region of frequent and intense pausing. The consensus sequence at HXBc2 nucleotides 763–754 aligned to variations of the sequence within clade B and also to isolates from other clades.

of pauses were offset from the structural data on each template both 5′ and 3′. Since the length of template held within the primer grip domain of AMV RT can only be inferred from the data available for HIV-1 RT, the pause data were offset by 2, 4, 6, 8, l0 and 12 nt 5′ and 3′, to look for associations with pausing ahead of and/or behind the catalytic site. These data are represented in Figure 10, and some results of frequency of pauses are plotted in Figure 11A–C. Position zero (on the *x*-axes in Fig. 11) represents the ribonucleotide on the template at which RT ceased to make cDNA, i.e. the position of catalytic site on the template.

These analyses showed that pausing was significantly associated with the presence of secondary structures both ahead of and behind the enzyme. Pausing occurred particularly frequently when the template between 6 and 10 bases ahead of the enzyme was predicted to be base paired. Pausing occurred very much less frequently when the bases 6 and 10 nt ahead of the catalytic site were predicted to be unpaired (groups 1–5 versus groups 6 and 7, Figs 9, 10 and 11). There was also a significant increase in pausing at paired nucleotides 6 bases behind (3′ to) the catalytic site. Pausing occurred significantly less frequently when the catalytic site was 8 bases ahead of (5′ to) nucleotides in group 1, (nucleotides not in a local structure) and when nucleotides in group 1 were 6 bases ahead of (5′ to) it (Fig 11B). When the catalytic site was between 4 and 8 nt ahead of (3′ to) the 5′ sides of RNA secondary structures (groups 3, 5 and 7), pausing was more frequent than when it was the same distances from the 3′ sides of RNA structures (groups 2, 4 and 6). This effect was largely due to the effect of group 7.

DISCUSSION

This is the first report that finds association between RT pausing and RNA secondary structure, although the association has been

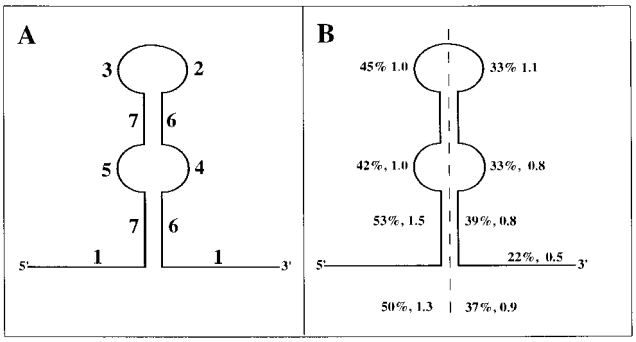

Figure 9. Frequency of pausing on RNA secondary structures. (**A**) The locations of the seven groups into which the individuals were divided on the basis of their positions within predicted RNA structures are represented on the model. (**B**) The frequency (%) of pausing by structure grouping are shown at their respective positions on a model RNA structure for the pause data set which was offset 5′ by 6 nt. The dashed line indicates the division between groups on the 5′ side of RNA structures (groups 3, 5 and 7) and those on the 3′ side of RNA structures (groups 2, 4 and 6).

assumed for a considerable time it has not previously been demonstrated. In this study, we considered that the effect of RNA secondary structures in templates for RT might be offset from the catalytic site. No relationship was found between well known RNA secondary structures and pausing events until the pausing events were offset from the structural data. This was done for all three templates over all the data. The frequencies of pausing events were sorted according to the seven structural groupings that we identified. It was only then that clear patterns of association between pausing and structures were found to occur. Initially, we had speculated that there could be an association between RNA structures ahead of the enzyme at the point where the leading edge of the RT (the finger domain) came into contact with hairpin structures. In HIV-1 RT the number of template nucleotides protected by the finger domain of the enzyme is approximately six (50,51). Since the distance between the catalytic site of AMV RT and the leading edge of the enzyme has not been determined, we offset the data by 2–12 nt to investigate the possible effect of structures ahead of the enzyme, and as a control, we offset the data behind the catalytic site as well. The association between pausing and structures behind the catalytic site was unexpected, but the large sample sizes of the entire data set are compelling. The peak association between pausing of AMV RT events and RNA secondary structures occurs when the catalytic site is 8 nt ahead of paired nucleotides in the template. It is reasonable to speculate from this that the distance of template protected by the finger domain of AMV RT is 8 nt.

The effects of potential structures behind the catalytic site bear some relationship to the termination mechanisms of other polymerases which have similar morphology and sequence to RTs.

It is likely that all nucleic acid polymerases have a common ancestor since nucleic acid polymerases have homologous dNTP binding motifs (60–63) and have similar global structures. The T7 RNAP has recently been shown to be related to other nucleotide polymerases such as DNA polymerases, RNA-dependent RNA polymerases and RTs (63). It is probable that all RNA polymerases, both bacterial and eukaryotic, use similar mechanisms to ensure efficient transcript termination (66).

In RNA polymerases (RNAPs) RNA secondary structures in the nascent strand behind the catalytic site contribute to dissociation of the polymerase from the template at termination signals. The pausing of AMV RT when the catalytic site is 6 nt past a 5′ side of an RNA secondary structure (consistently over the entire data set) could be due to the formation of secondary structure in the nascent cDNA behind the catalytic site. Enzymatic probing (51) has shown that nucleotides up to 25 bases behind the catalytic site are within the HIV-RT complex. Therefore, if secondary structures are forming

A % pausing frequency for each **folding state** of RNA

			Distance offset 5' (nucleotides)												
RNA structure	group #	n	-12	-10	-8	-6	-4	-2	0	+2	+4	+6	+8	+10	+12
Unstructured	1	41	34	24	20	32	41	51	51	37	33	22	32	35	46
3' side of a loop	2	39	57	49	32	43	43	33	38	38	45	33	43	35	35
5' side of a loop	3	40	48	58	58	50	36	41	39	36	35	45	50	32	25
3' bulge	4	27	19	22	26	19	26	22	37	37	26	33	19	27	42
5' bulge	5	26	46	38	62	41	41	48	44	33	41	42	41	26	54
3' side of a stem	6	138	35	35	38	44	35	41	39	41	34	39	44	42	45
5' side of a stem	7	126	45	48	46	57	50	41	39	45	50	53	57	48	38
all unpaired bases	1-5	174	41	39	38	38	38	40	42	36	36	35	3	32	39
all paired bases	6 & 7	264	40	41	42	50	42	41	39	43	42	46	50	45	41
3' sides of structures	2,4 & 6	204	37	36	35	40	35	37	39	40	35	37	40	38	42
5' sides of structures	3,5, & 7	192	47	44	43	53	43	42	40	41	46	50	53	41	37

B Mean scores for each **folding state** of RNA

			Distance offset 5' (nucleotides)												
	group #	n	-12	-10	-8	-6	-4	-2	0	+2	+4	+6	+8	+10	+12
Unstructured	1	41	0.8	0.7	0.5	0.9	1.2	1.7	1.5	1.1	1.0	0.5	0.7	0.9	1.2
3' side of a loop	2	39	1.3	1.2	0.9	0.7	1.0	0.8	0.8	0.7	1.0	1.1	1.1	0.7	0.7
5' side of a loop	3	40	1.2	1.3	1.3	0.8	0.9	1.2	1.1	0.8	0.9	1.0	1.0	0.8	0.7
3' bulge	4	27	0.6	0.6	0.6	1.0	0.6	1.0	0.9	1.2	0.6	0.8	0.5	1.0	1.2
5' bulge	5	26	1.3	1.2	1.2	0.7	1.3	0.9	1.3	0.9	1.3	1.0	1.0	0.7	1.5
3' side of a stem	6	138	0.8	0.8	0.8	1.0	0.8	0.6	0.9	1.0	0.8	0.8	1.1	1.0	1.0
5' side of a stem	7	126	1.2	1.3	1.3	1.2	1.0	1.3	0.9	1.1	1.2	1.5	1.3	1.2	1.0
all unpaired bases	1-5	174	1.0	1.0	1.0	0.9	1.1	0.8	1.1	0.9	1.0	0.9	0.9	0.8	1.0
all paired bases	6 & 7	264	1.0	1.0	1.0	1.2	0.9	1.2	0.9	1.1	1.0	1.1	1.2	1.1	1.0
3' sides of structures	2,4 & 6	204	0.9	0.8	0.8	1.0	0.8	1.0	0.8	1.0	0.8	0.9	1.1	1.0	0.9
5' sides of structures	3,5, & 7	192	1.2	1.3	1.3	1.2	1.1	1.0	1.1	1.0	1.2	1.3	1.2	1.0	1.0

Figure 10. The relationship between RNA structure and RT pausing when the pausing data were offset. The frequency of pausing at the different RNA structure groupings when the pause data were offset by different distances relative to the position at which the transcript terminated.

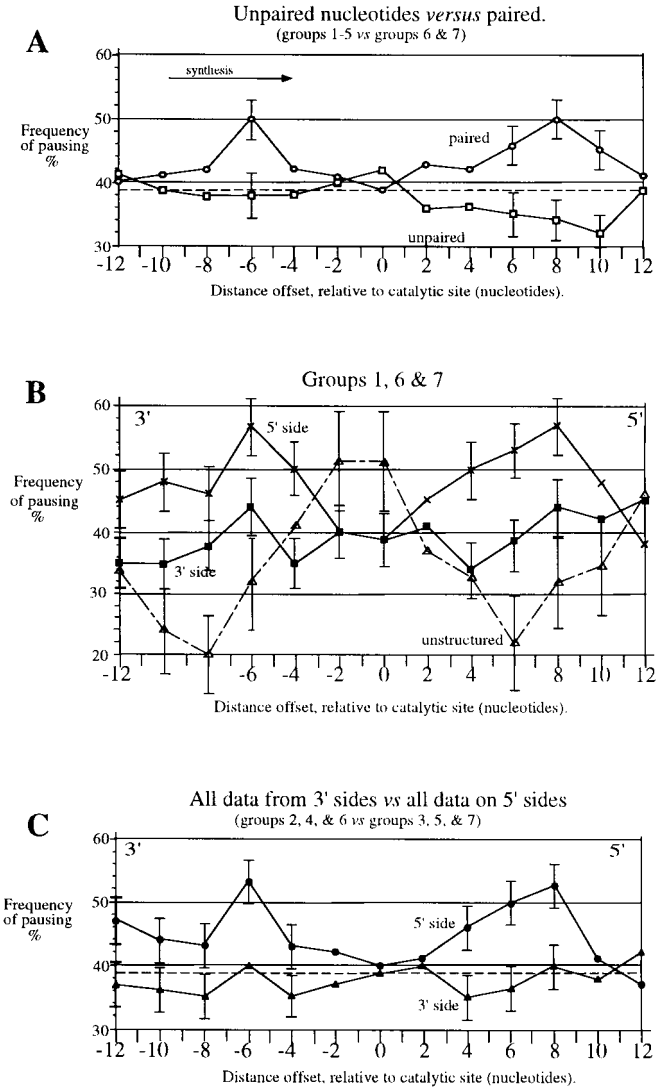

Figure 11. The effect of RNA structure is offset from the point of termination. (**A**) Frequency of pausing at nucleotides predicted to be paired (groups 1–5; ○) and unpaired (groups 6 and 7; □) plotted against the distance which the pausing data were offset 5′ and 3′ from the RNA structure at the point of termination. (**B**) Frequency of pausing at nucleotides predicted to be unpaired and not within a local RNA structure (group 1; Δ), paired and on the 3′ side of RNA structures (group 6; ■), and paired and on the 5′ side of RNA structures (group 7, ×) plotted against the distance which the pausing data were offset. (**C**) Frequency of pausing at all nucleotides predicted to be on the 3′ side of RNA structures (groups 2, 4 and 6; ▲), and paired and on the 5′ side of RNA structures (groups 3, 5 and 7; ●) plotted against the distance which the pausing data were offset. The mean score for all nucleotides is represented in (A) and (C) by a dashed line. Standard error bars are indicated where there was little or no overlap.

behind the catalytic site, they will be within the enzyme complex. This would be similar to the formation of the stem–loop that causes the dissociation of RNAP II from the major late arrest site of adenoviruses ahead of a run of adenosines (71). We and others (25,52,53) have shown that pausing occurs most frequently at homopolymeric runs of adenosines. Further support for the cooperation between secondary structure and primary sequence in termination events comes from RNAP III, which terminates transcription of a mouse 5S gene *in vitro* in a cluster of 23 A residues. In addition, 80% of the *in vitro* transcripts derived from an Alu repeat terminate within a stretch of nine A residues preceded by a hairpin (72) and in *Escherichia coli* RNAP where the RNA secondary structure 7–10 nt from the 3′ end of the nascent strand interacts directly with the polymerase causing it to pause. A recent review of RNA polymerase, in which a 'sliding clamp' model for processivity of RNA polymerase (73) was proposed, has some resonance to our findings. In that scenario, RNA secondary structures forming in the nascent strand are envisaged as triggering a conformation change in the enzyme, initiating a pause. Similarly, the strength of the RNA–DNA hybrid determines the likelihood that a pause leads on to a termination. Our data and a previous report (25) support the role of low bond strength at the catalytic site predisposing to termination. The finding that lower hydrogen bond strength between the template and the nascent strand increase pausing is consistent with this, but the influence of pausing at adenosines cannot be eliminated from the analysis of the effect of H bond strength. Wu *et al.* (23) found that the strength of H bonding was correlated with sequence variations, and Ji *et al.* (25) reported that an AT-rich region was particularly prone to misincorporation due to poor H bonding between template and primer. However, our study found least pausing within poly(U) tracts.

Once the polymerase and the nascent strand dissociate from the template, they might reassociate at the same position or on a homologous sequence. In the case of a homopolymeric run, reassociation may occur 1 or 2 bases out of correct alignment with the template sequence, (54,22). Ji *et al.* (25) found that HIV-RT paused at eight poly(dA) templates within the sequence of the HIV-l Bru *env* at the hypervariable V1 domain. They found that template sequence AGATAAAAAACTG (*env* nt 476–486) was associated with numerous insertion and deletion mutations in an *in vitro* assay. The striking sequence heterogeneity of the sequence 3′-UUUAAAAA-5′ within the HIV-1Ψ (Fig. 8) is feasibly a result of similar processes at both these sites.

This study finds that poly(rA) tracts are pausing signals for RT in certain, as yet undefined, contexts. Pausing at adenosines occurred most frequently at those within homopolymeric tracts (73%). Pausing at diadenosines was slightly less frequent (expected score of 1.19, out of a possible 2) and less still at monoadenosines (56%). It is notable that poly(dA) tracts are a termination signal for RNAP II (55–59) and we speculate that a similar mechanism is responsible for pausing of both RNAPs and RT. Pausing has been causally linked to the high frequency of recombination in HIV, a process contributing to the genetic diversity in the virus population (18,19) allowing some viruses to evade the immune response and antiviral drugs. Nucleotide misincorporation and pausing are likely associated at the HIV-1 sequence 3′-UUUAAAAA-5′, supporting the work of Bebenek *et al.* (22,54) and Wu *et al.* (23), who found a positive correlation between pausing probability and frameshift error rates and misincorporation at pause sites.

These assays were carried out in the absence of cellular proteins, which may function to overcome RT pausing during virus replication, as in other polymerases (75). However, the absence of cellular factors has enabled us to investigate more clearly the relationship between RT and its RNA template.

Similar patterns of pausing of HIV-l, AMV RT and MMLV RT on DNA templates have been reported previously (25,52) although the relative intensities varied. In previous studies, HIV-1 RT was found to be almost completely blocked at oligo (dA) tracts, whereas AMV RT and MMLV RT were able to traverse them at higher RT concentrations. Similarities between the pausing patterns of AMV RT and HIV-1 RT suggest AMV RT pausing has considerable relevance to the processivity of HIV-1

RT. A greater understanding of the mechanism of RT activity will be an asset for rational design of antiviral drugs.

ACKNOWLEDGEMENTS

We thank C.Turnbough, M.Sakalian and B.Browne for helpful comments and discussions. HIV-RT was supplied by S.McPherson of the Center for AIDS Research, Molecular Biology Core Facility at the University of Alabama at Birmingham supported by grant P30 AI 27767 from the NIAID NIH. This research was also supported by the Medical Research Council of the UK

NOTE ADDED IN PROOF

During the preparation of this manuscript, we became aware of work by Suo and Johnson (*Biochemistry*, 1997, **36**, 12459–12467) in which the kinetics of HIV-1 reverse transcriptase on a synthetic RNA template with a predicted secondary structure was studied.

REFERENCES

1 Coffin,J.M. (1990) In Fields,B.N., Knope,D.M., Chnock,R.M., Hirsch,M.S., Meenick,J.L., Monath,T.P. and Roizman,B. (eds), *Virology—Retroviridae and Their Replication*. Raven Press Ltd, New York, pp 1437–1499.
2 Skalka,A.M. and Goff,S.P. (1993) *Reverse Transcriptase*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
3 Starnes,M.C. and Cheng,Y. (1989) *J. Biol. Chem.* **264**, 7073–7077.
4 Hansen,J., Schulze,T. and Moelling,K. (1987) *J. Biol. Chem.* **262**, 12393–12396.
5 Charneau,P. and Clavel,F. (1991) *J. Virol*. **65**, 2415–2421.
6 Hungnes,O., Tjøtta,E. and Grinde,B. (1991) *Virology* **190**, 440–442.
7 Kim,S., Byrn,R., Groopman,J. and Baltimore,D. (1989) *J. Virol*. **63**, 3708–3713.
8 Sato,H., Orenstein,J., Dimitrov,D. and Martin,M. (1992) *Virology* **186**, 712–724.
9 Bender,W., Chien,Y.H., Chattopadhyay,S., Vogt,P.K., Gardner,M.B. and Davidson,N. (1978) *J. Virol*. **25**, 888–896.
10 Murti,K.G., Bondurant,M. and Tereba,A (1981) *J. Virol*. **37**, 411–419.
11 Weiss,R., Teich,N., Varmus,H. and Coffin,J.M. (1985) *RNA Tumor Viruses*. 2nd Ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
12 Srinivasan,A., York,D., Jannoun-Nasr,R., Kalyanaraman,S., Swan,D., Benson,J., Bohan,C., Luciw,P.A., Schnoll,S., Robinson,R.A., Desai,S.M. and Devare,S.G. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 6388–6392.
13 Clavel,F., Hoggan,M.D., Willey,R.L., Strebel,K., Martin,M.A. and Repaske,R. (1989) *J. Virol*. **63**, 1455–1459.
14 Hu,W.S. and Temin,H.M. (1990a) *Proc. Natl. Acad. Sci. USA* **87**, 1556–1560.
15 Hu,W.S. and Temin,H.M. (1990b) *Science* **250**, 1227–1233.
16 Goodrich,D.W. and Duesburg,P.H. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 2052–2056.
17 Howell,R.M., Fitzgibbon,J.E., Noe,M., Ren,Z.J., Gucke,D.J., Schwartzer,T.A. and Dubin,D.T. (1991) *AIDS. Res. Hum. Retroviruses* **7**, 869–876.
18 Katz,R.A. and Skalka,A.M. (1990) *Annu. Rev. Genet.* **24**, 409–445.
19 Temin,H.M. (1991) *Trends Genet*. **7**, 71–74.
20 Xu,H. and Boeke,J.D. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 8553–8557.
21 Klarmann,G.J., Schauber,C.A. and Preston,B.D. (1993) *J. Biol. Chem*. **268**, 9793–9802.
22 Bebenek,K., Abbotts,J., Wilson,S.H. and Kunkel,T.A. (1993) *J. Biol. Chem.* **268**, 10324–10334.
23 Wu,W., Blumberg,B.M., Fay,P.J. and Bambara,R.A. (1995) *J. Biol. Chem.* **270**, 325–332.
24 Abbotts,J., Bebenek,K., Kunkel,T.A. and Wilson,S.H. (1993) *J. Biol. Chem.* **268**, 10312–10323.
25 Ji,J., Hoffmann,J.-S. and Loeb,L. (1994) *Nucleic Acids Res*. **22**, 47–52.
26 Roy,S., Delling,U., Chen,C.H., Rosen,C.A. and Sonenberg,N. (1990) *Genes Dev.* **4**, 1365–1374.
27 Dingwall,C., Ernberg,I., Gait,M.J., Green,S.M., Heaphy,S., Karn.J., Lowe,A.D., Singh,M. and Skinner,M.A. (1990) *EMBO J*. **9**, 4145–4153.
28 Feng,S. and Holland,E.C. (1988) *Nature* **334**, 165–167.
29 Muesing,M.A., Sith,D. and Capon,D.J. (1987) *Cell* **48**, 691–701.
30 Harper,J.W. and Logsdon,N.J. (1991) *Biochemistry* **30**, 8060–8066.
31 Harrison,G.P. and Lever,A.M.L. (1992) *J. Virol*. **66**, 4144–4152.
32 Baudin,F., Marquet,R., Isel,C., Darlix,J.-L., Ehresmann,B. and Ehresmann,C. (1993) *J. Mol. Biol*. **229**, 383–397.
33 Hayashi,T., Ueno,Y. and Okamoto,K. (1993) *FEBS Lett*. **327**, 213–218.
34 Sakaguchi,K., Zambrano,N., Baldwin,E.T., Shapiro,B.A., Erickson,J.W., Omichinski,J.G., Clore,G.M., Gronenborn,A.M. and Appella,E. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 5219–5223.
35 Luban,J. and Goff,S. (1994) *J. Virol*. **68**, 3784–3793.
36 Kim,H.J., Lee,K. and O'Rear,J.J. (1994) *Virology* **198**, 336–340.
37 McBride,M.S. and Panganiban,A.T. (1996) *J. Virol*. **70**, 2963–2973.
38 Paillart,J.-C., Berthoux,L., Ottmann,M., Darlix,J.-L., Marquet,R., Ehresmann,B. and Ehresmann,C. (1996) *J. Virol*. **70**, 8348–8354.
39 Fisher,A.G., Collati,E., Ratner,L., Gallo,R.C. and Wong-Staal,F. (1985) *Nature* (*London*), **316**, 262–265.
40 Myers,G., Hahn,B.H., Mellors,J.W., Henderson,L.E., Korber,B., Jeang,K.-T., McCutchan,F.E. and Pavlakis,G.N. (1996) *Human Retroviruses and AIDS*. Los Alamos National Laboratory, NM.
41 Kaye,J.F. and Lever,A.M.L. (1996) *J. Virol*. **70**, 880–886.
42 Rhee,S.S., Hui,H. and Hunter,E. (1990) *J. Virol*. **64**, 3844–3852.
43 Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
44 Harrison,G.P., Hunter,E. and Lever,A.M.L. (1995) *J. Virol*. **69**, 2175–2186.
45 Shelness,G.S. and Williams,D.L. (1985) *J. Biol. Chem*. **260**, 8637–8646.
46 Harrison,G.P., Miele,G., Hunter,E. and Lever,A.M.L. (1998) *J. Virol*., in press.
47 Aboul-ela,F., Karn,J. and Varani,G. (1995) *J. Mol. Biol*. **253**, 313–332.
48 Gurtler,L.G., Hauser,P.H., Eberle,J., von Brunn,A., Knapp,S., Zekeng,L., Tsague,J.M. and Kaptue,L. (1994) *J. Virol*. **68**, 1581–1585.
49 Haesevelde,M.V., Decourt,J.L., DeLeys,R.J., Vanderborght,B., der Groen,G., van Heuverswijn,H. and Saman,E. (1994) *J. Virol*. **68**, 1586–1596.
50 Boyer,P.A., Tantillo,C., Jacobo-Molina,A., Nanni,R.G., Ding,J., Arnold,E. and Hughes,S.H. (1994) *Proc. Natl. Acad. Sci. USA*, 4882–4886.
51 Wöhrl,B., Tantillo,C., Arnold,E. and Le Grice,S.F.J. (1995) *Biochemistry* **34**, 5343–5350.
52 Huber,H.E., McCoy,J.M., Seehra,J.S. and Richardson,C.C. (1989) *J. Biol. Chem*. **264**, 4669–4678.
53 Williams,K.J., Loeb,L.A. and Fry,M. (1990) *J. Biol. Chem*. **265**, 18682–18689.
54 Bebenek,K., Abbotts,J., Roberts,J.D., Wilson,S.H. and Kunkel,T.A. (1989) *J. Biol. Chem*. **264**, 16948–16956.
55 Reines,D., Wells,D., Chamberlin,M.J. and Kane,C.M. (1987) *J. Mol. Biol*. **196**, 299–312.
56 Dedrick,R.L., Kane,C.M. and Chamberlin,M.J. (1987) *J. Biol. Chem.* **262**, 9098–9108.
57 Kerppola,T.K. and Kane,C.M. (1990) *Biochemistry* **29**, 269–278.
58 Kerppola,T.K. and Kane,C.M. (1991) *FASEB* **5**, 2833–2842.
59 Enriquez-Harris,P., Levitt,N., Briggs,D. and Proudfoot,N.J. (1991) *EMBO J*. **10**, 1833–1842.
60 Ollis,D.L., Kline,C. and Steitz,T.A. (1985) *Nature* **313**, 765–769.
61 Sousa,R., Chong,Y.J., Rose,J.P. and Wang,B.-C. (1993) *Nature* **364**, 593–599.
62 Delarue,M., Poch,O., Tordo,N., Moras,D. and Argos,P. (1990) *Prot. Engng.* **6**, 461–467.
63 McAllister,W.T. and Raskin,C.A. (1993) *Mol. Microbiol*. **10**, 1–6.
64 Kohlstaedt,L.A., Wang,J., Friedman,M., Rice,P.A. and Steitz,T.A. (1992) *Science* **256**, 1783–1790.
65 Ruttimann,C., Cotoras,M., Zaldivar,J. and Vicuna,R. (1985) *Eur. J. Biochem*. **149**, 41–46.
66 Reeder,R.H. and Lang,W. (1994) *Mol. Microbiol*. **12**, 11–15.
67 Platt,T. (1986) *Annu. Rev. Biochem*. **55**, 339–372.
68 Richardson,J.P. (1993) *CRC Crit. Rev. Bioch. Mol. Biol*. **28**, 1–30.
69 Chan,C.L. and Landick,R. (1993) *J. Mol. Biol*. **233**, 25–42.
70 Wiest,D.K., Wang,D. and Hawley,D.K. (1992) *J. Biol. Chem*. **267**, 7733–7744.
71 Reeder,T.C. and Hawley,D.K. (1996) *Cell* **87**, 767–777.
72 Hess,J., Perez-Stable,C., Wu,G.-J., Weir,B., Tiroco,I. and Shen,C.-K.J. (1985) *J. Mol. Biol*. **184**, 7–21.
73 Landick,R. (1997) *Cell* **88**, 741–744.
74 Jonckheere,H., Taymans,J.-M., Balzarini,J., Velazquez,S., Camarasa,M.-J., Desmyter,J., de Clereq,E. and Anne,J. (1994) *J. Biol. Chem*. **269**, 25255–25258.
75 Kornberg,A. and Baker,T.A. (1992) *DNA Replication*. 2nd Ed., W.H. Freeman and Co., New York, NY.
76 Filler,A.G. and Lever,A.M.L. (1997) *AIDS Research and Human Retroviruses* **13**, 291–299.