Roles of RNA : DNA hybrid stability, RNA structure, and active site conformation in pausing by human RNA polymerase II

ARTICLE *in* JOURNAL OF MOLECULAR BIOLOGY · SEPTEMBER 2001

Impact Factor: 4.33 · DOI: 10.1006/jmbi.2001.4842 · Source: PubMed

2 AUTHORS:

Murali Palangat
National Institutes of Health
28 PUBLICATIONS **932** CITATIONS
SEE PROFILE

Robert Landick
University of Wisconsin–Madison
164 PUBLICATIONS **9,752** CITATIONS
SEE PROFILE

doi:10.1006/jmbi.2001.4842 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (2001) 311, 265–282

JMB

Roles of RNA:DNA Hybrid Stability, RNA Structure, and Active Site Conformation in Pausing by Human RNA Polymerase II

Murali Palangat and Robert Landick*

Department of Bacteriology University of Wisconsin-Madison, 1550 Linden Dr Madison, WI 53706, USA

*Corresponding author

Human RNA polymerase II recognizes a strong transcriptional pause signal in the initially transcribed region of HIV-1. We report the use of a limited-step transcription assay to dissect the mechanism underlying recognition of and escape from this HIV-1 pause. Our results suggest that the primary determinant of transcriptional pausing is a relatively weak RNA:DNA hybrid that triggers backtracking of RNA polymerase II along the RNA and DNA chains and displaces the RNA 3′ OH from the active site. In contrast, two alternative RNA secondary structures, TAR and anti-TAR, are not required for pausing and affect it only indirectly, rather than through direct interaction with RNA polymerase II. TAR accelerates escape from the pause, but anti-TAR inhibits formation of TAR prior to pause escape. The behavior of RNA polymerase II at a mutant pause signal supports a two-step, non-equilibrium mechanism in which the rate-determining step is a conformational change in the enzyme, rather than the changes in nucleic-acid base-pairing that accompany backtracking.

Keywords: RNA polymerase; pausing; transcriptional regulation; RNA:DNA hybrid

Introduction

In eukaryotes, accessory proteins must convert transcription complexes containing RNA polymerase II (RNAPII) from an inefficient state that is sensitive to obstacles that block RNA chain elongation to a state that transcribes through these obstacles efficiently to allow expression of most genes.[1–3] These obstacles include nucleosomes,[4,5] cellular transcription inhibitors like DSIF and NELF,[6,7] and intrinsic pause and arrest signals encoded in the DNA.[8,9] To begin to understand the physical basis for this change in RNAPII's catalytic efficiency, we have characterized the mechanism by which human RNAPII recognizes a discrete pause signal at +62 in the single transcriptional unit of HIV-1.[10]

By analogy to similar pause sites in prokaryotes,[11] pausing at the HIV-1 site appears to occur by a two-step mechanism (Figure 1). Prior to pausing, RNAPII transcribes rapidly in an activated state, oscillating in each round of nucleotide addition between the post-translocated conformation, which binds NTP efficiently, and the pre-translocated conformation, which forms upon nucleotide addition. In the first pause step, RNAPII isomerizes to an unactivated state that binds NTP more weakly†. This sets up the second step in

Abbreviations used: RNAP, RNA polymerase; TEC, transcription elongation complex; oligo, oligonucleotide.

† Artsimovitch & Landick called the weak NTP-binding conformation a slow intermediate.[11] However, we favor the traditional name unactivated intermediate (as first applied to RNAP by Erie *et al.*)[12] because it is unclear that the weak NTP-binding conformation is capable of nucleotide addition. Unactivated intermediate can apply to either situation, whereas slow intermediate implies a second pathway of nucleotide addition is possible.

E-mail address of the corresponding author: landick@bact.wisc.edu

0022-2836/01/020265–18 $35.00/0

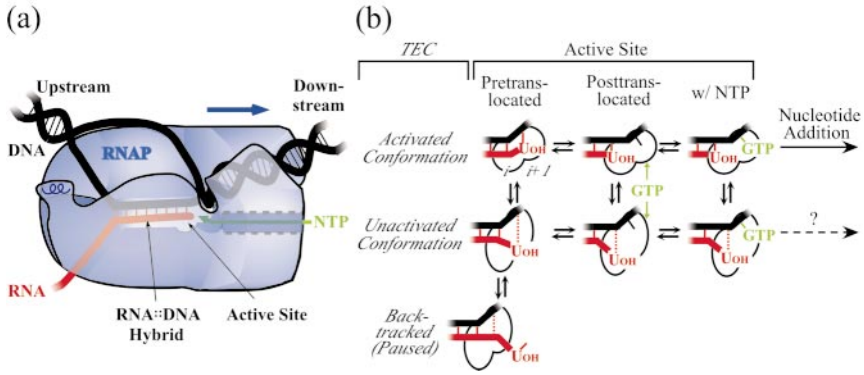


Figure 1. Possible mechanism of HIV-1 pausing. (a) Structure of the TEC. Recently reported structures of RNAPs suggest that NTP substrates enter the RNAP active site through a secondary channel that is located at the downstream end of the active-site cleft, which encircles the RNA:DNA hybrid, and that the secondary channel may become occupied by the extruded 3′-end of the RNA chain in an arrested TEC.[19,50,58,80,81] The RNA:DNA hybrid is 9 bp long in the pre-translocated conformation and 8 bp long after translocation opens the active site to NTP binding (the post-translocated conformation is depicted here).[19,25,26,50,51,82] (b) State of the RNAP active site during pausing. Nucleotide addition occurs in an activated TEC conformation in which closure of the two large RNAP subunits around the downstream DNA, the RNA:DNA hybrid, and the exiting RNA transcript restricts the RNA 3′ nucleotide to the i and $i+1$ sites (circles). The pause signal triggers formation of the unactivated intermediate in which a conformational change distorts the active site (ovals) and weakens proper alignment of the RNA 3′ nucleotide (indicated by dotted red line; see Discussion). The altered active site and 3′ nucleotide position in the unactivated intermediate reduce affinity for NTP substrate and allow backtracking of RNAP. Nucleotide addition occurs only in the activated conformation in the sequential (or essential activation) version of the mechanism, but can occur in the unactivated conformation (dotted line) in the parallel path (or non-essential activation) version of the mechanism.

which RNAPII reversibly backtracks ~1 bp along the DNA and RNA chains, displacing the RNA 3′ nucleotide from the active site†.

Several studies argue that interconversion of the activated and unactivated states of RNAP involves a conformational change in the enzyme.[11–16] Depending on the nature of this change, RNAP either might need to return to the activated state to catalyze nucleotide addition (a sequential mechanism called essential activation when it involves allosteric effectors),[11–13] or it might be capable of nucleotide addition in both the activated and unactivated states (a parallel-path mechanism called non-essential activation when it involves allosteric effectors).[14–16] It is likely, though not yet shown, that the inefficient form of RNAPII favors the unactivated conformation, whereas the efficient form generated by accessory proteins favors or is locked in the activated conformation. For bacterial RNAP, the difference between the activated and unactivated conformations has been proposed to involve partial opening of RNAP's active-site cleft,[11,17] which may involve movement of a mobile clamp domain that contacts RNA and DNA in the transcription elongation complex (TEC).[18–20]

Although the unactivated-intermediate mechanism can explain pausing by human RNAPII at the HIV-1 site, important aspects remain untested. First, although reversible backtracking has been proposed to cause pausing,[8,10,21–26] it is unclear what protein or nucleic-acid interactions govern backtracking at the pause site. One idea is that relative instability of the ~9-bp heteroduplex (or hybrid) between the 3′-proximal RNA and the template DNA strand in TECs (see Figure 1) causes the enzyme to slide backwards to form a more stable hybrid. This mechanism has been proposed to explain transcriptional arrest, in which RNAPII backtracks so far (⩾10-15 nt) that it is unable to slide forward to the RNA 3′ end unless the accessory protein TFIIS stimulates hydrolytic removal of 3′ single-stranded RNA, which extends from the backtracked RNA:DNA hybrid, to generate a new RNA 3′ end.[22,23,27–34] This relative-hybrid-stability model is supported by findings that antisense oligonucleotides (oligos) that anneal to the nascent RNA just upstream of the TEC inhibit pausing or arrest,[11,22,23] and is consistent with the effects on pausing and arrest of base analogs incorporated into the nascent RNA to stabilize or destabilize the RNA:DNA hybrid in different TEC registers.[23,25] However, a direct correlation between relative TEC

† We refer to the HIV-1 paused TEC as backtracked by 1 bp (or as being in the −1 backtracked state) because TFIIS-stimulated transcript cleavage removes 2 nt from the pause RNA 3′-end.[10] Transcript cleavage (like pyrophosphorolysis) should remove 1 nt from the pre-translocated conformation of the TEC (Figure 1); cleavage of 2 nt suggests that the paused TEC is backtracked 1 nt relative to the pre-translocated conformation. This definition differs from that used by Palangat *et al.*,[10] where we wrote that the HIV-1 paused TEC is backtracked by 2 nt. However, it would be misleading to number the pre-translocated conformation as the −1 backtracked conformation, because the TEC is pre-translocated immediately after every round of nucleotide addition, leading us to adopt a revised nomenclature.

stability and pausing has not been reported. Further, it is unclear if changes in RNA:DNA and DNA:DNA base-pairing in different TEC registers alone can explain backtracking and pausing, or if they also depend on RNAPII-nucleic acid interactions or on conformational changes in RNAPII.[35]

Second, there is no direct evidence that the unactivated and backtracked states of RNAPII are kinetically distinguishable intermediates in pausing. Conceivably, TECs in the active and backtracked registers could equilibrate rapidly at most template positions (called a positional equilibrium), as previously proposed for the pretranslocated and post-translocated configurations of the TEC.[24,36–38] If this were true, then pausing could simply reflect template positions at which the active state is especially disfavored thermodynamically. Alternatively, a significant energetic barrier to backtracking might exist so that even more stable backtracked TECs may not form before nucleotide addition causes forward TEC movement.

Finally, the role of nascent RNA structures, which can modulate the rate of HIV-1 pause escape by up to a factor of 3,[10] is unclear. RNA upstream from the pause can form two alternative RNA secondary structures, TAR and anti-TAR†. TAR, which is first able to form at the pause site, recruits HIV-1 Tat and cellular P-TEFb to trigger efficient transcription of HIV-1 DNA.[39–43] Anti-TAR forms from the first 39 nt of the HIV-1 transcript and prevents TAR formation prior to RNAPII arrival at the pause site. Increasing the stability of TAR decreases pausing, whereas increasing the stability of anti-TAR increases pausing.[10] These effects could be either direct, through interaction of the hairpins with RNAPII, or indirect, through effects on backtracking that are mediated by changes in the availability of single-stranded RNA for re-entry into RNAPII. Precedents exist for both possibilities.[11] At class I bacterial pause sites, an RNA hairpin stimulates pausing directly through an interaction with RNAP that alters its catalytic efficiency. At class II sites, backtrack-induced pausing can be inhibited by antisense oligos that sequester nascent RNA upstream of RNAP. TAR appears to act similarly to antisense oligos; that is, to inhibit pausing indirectly by sequestering upstream RNA.[10] However, it has been unclear whether the anti-TAR structure acts simply to prevent TAR formation and thus allow backtracking and pausing, or if it also is an allosteric effector of pausing like the bacterial pause RNA hairpin.[44]

† Previously, we referred to the anti-TAR structure as the HIV-1 pause RNA hairpin because it favored pausing.[10] In light of the finding that anti-TAR affects pausing only indirectly, referring to it as a pause hairpin is confusing, because the bacterial pause RNA hairpin acts through a direct interaction with RNAP. Thus, we have adopted the more appropriate name of anti-TAR for this RNA structure.

To study the kinetic mechanism of HIV-1 pausing, and to elucidate the roles of RNA structure and relative hybrid stability, we developed a simplified pause assay. We used the assay to examine pausing on different mutant templates and in conditions that could reveal the rate of paused TEC formation and the existence of kinetic intermediates.

Results

A simple assay of pause efficiency and escape

The proposed mechanism of HIV-1 pausing (see Introduction) predicts complicated kinetics. To simplify analysis, transcriptional pausing can be described by two empirically defined parameters (Figure 2(a)): E, the fraction of TECs arriving at the site that enter into a paused state (pause efficiency), and k_e^{obs}, the pseudo first-order rate constant for escape from the pause site by addition of the next nucleotide at a given NTP concentration (inversely proportional to the pause half-life).[45] E can be described as a competition between empirically defined rates of nucleotide addition for bypass of the pause site, determined by k_b^{obs}, and of isomerization into the paused state, determined by k_i^{obs} ($E = k_i^{obs}/(k_i^{obs} + k_b^{obs})$). The fraction E/k_e^{obs}, called τ_i is a useful measure of the overall strength of a pause signal.[45,46] These empirical rate constants do not account for the substrate-dependence of pausing and will not correspond to elementary rate constants in an unactivated-intermediate mechanism. However, they are useful to evaluate the effects of sequence changes on pausing by RNAPII.

To measure the E and k_e^{obs} for HIV-1 pausing, we needed to measure synchronous elongation of a relatively homogenous population of TECs through the pause site. To accomplish this, we used stepwise transcription to prepare TECs halted at position A60, 2 nt upstream of the HIV-1 pause site at U62 (Figure 2(a); see Materials and Methods). We then added CTP, UTP, and GTP to allow transcription to the pause site and, after the pause, to U66, at which point the TECs were trapped by the absence of ATP. Because G63 is the only G in this short sequence, variation of GTP concentration affects nucleotide addition only at the pause and at no other step. Further, omission of ATP prevents TECs halted upstream from A60 from reaching the pause site.

Upon addition of the three NTPs, TECs halted at A60 resumed elongation rapidly (Figure 2(b)). A fraction of the TECs recognized the HIV-1 pause at U62 as evidenced by the appearance of a strong 62 nt RNA band, and then escaped gradually from the pause at a pseudo first-order rate (at fixed GTP concentration). All TECs that originated at A60 accumulated at U66, except a small fraction that apparently misincorporated 1 nt to give a 67 nt band. To quantify pause recognition and escape, we calculated the relevant amount of RNA in three

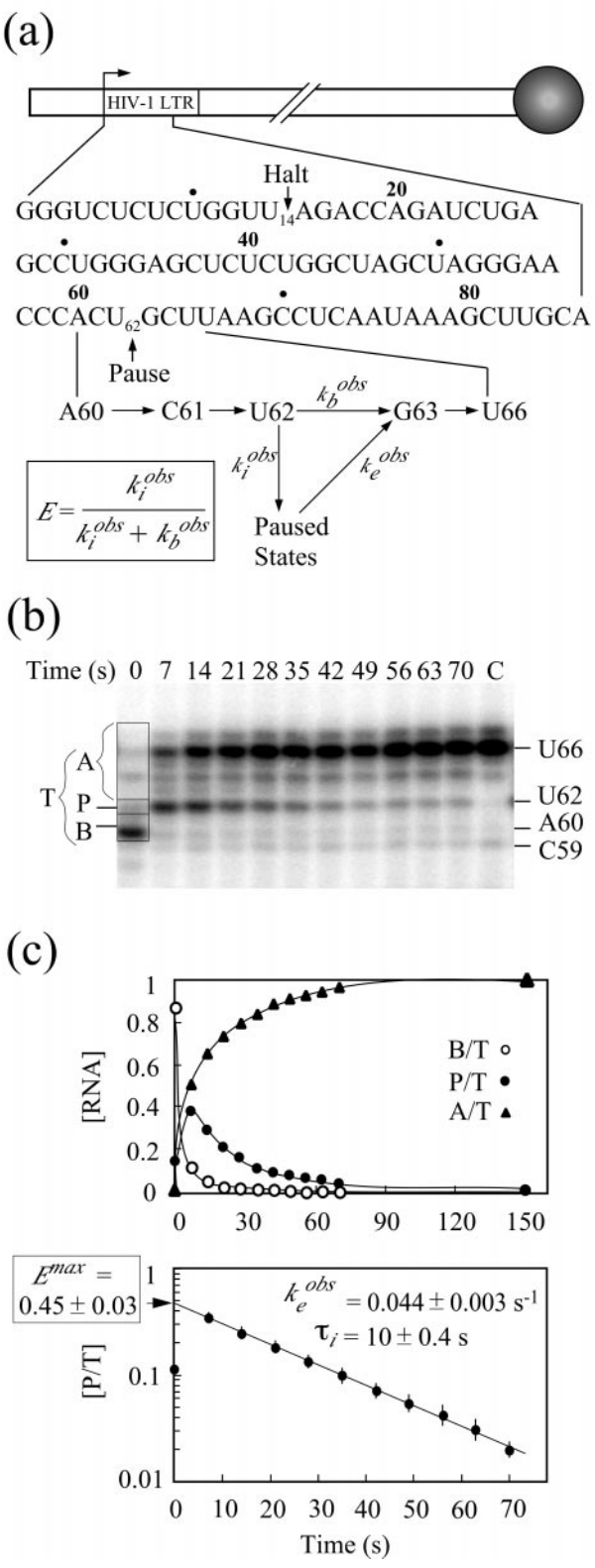


Figure 2. Assay to measure recognition of and escape from the HIV-1 pause site. (a) The transcription template. A 774 bp, linear DNA containing the HIV-1 promoter and early transcribed region (−138 to +85 with respect to HIV-1 start site),[10] immobilized on streptavidin-coated agarose beads through biotin attached to the downstream end of the DNA was used to form TECs (see Materials and Methods). The sequence of the RNA transcribed from the template and the key steps in transcription from A60 to U66 in the presence of CTP, GTP and UTP (but no ATP) are shown. Empirical parameters used to characterize pausing are indicated on the reaction scheme and in the side box. (b) Portion of a denaturing RNA gel showing transcription through the HIV-1 pause site. TECs halted at A60 were elongated in the presence of 1 mM each CTP, GTP and UTP, aliquots were removed at times indicated above each lane, processed and separated on a denaturing 8% polyacrylamide/urea gel (see Materials and Methods). The sample in lane C was incubated for an additional five minutes after increasing the [GTP] to 5 mM. The RNA species were grouped into three sets for quantification (boxed on lane 0): before pause, *B*; at the pause, *P*; after the pause, *A*. *B* and *P* were corrected by subtracting the amounts remaining in lane C; *A* was corrected by subtracting the amount present in lane 0. *T* was calculated from the sum of the corrected values ($T = B + P + A$). (c) Relative concentrations of RNA from the experiment shown in (b) are plotted as a function of time. Open circles, RNA before the pause (B/T). Filled circles, RNA at the pause (P/T). Triangles, RNA after the pause (A/T). The semilogarithmic plot of $[P/T]$ *versus* time in the lower panel is from five independent measurements with standard deviations as shown by the error bars. Back-extrapolation of P/T to time zero yields a maximum pause efficiency of 0.45 (E^{max}); non-linear regression of P/T as a function of time yields an observed rate constant for escape of 0.044 s^{-1} (k_e^{obs}). Pause strength τ_i is ~10 s ($\tau_i = E/k_e^{obs}$).

separate regions of each gel lane (Figure 2(b)): *B*, corresponding to before the pause (A60 + C61); *P*, the pause RNA (U62); and *A*, corresponding to after the pause (G63-U66, plus the misincorporated species at 67). By converting each region to a fraction of the total RNA, *T* ($T = B + P + A$; see Materials and Methods), we could determine k_e^{obs} from the slope of P/T *versus* time on a semilogarithmic plot and the maximum possible value of *E*, E^{max}, from the intercept at time zero (Figure 2(c)).[45] Modification of this assay by varying the time after addition of CTP and UTP and before addition of GTP can reveal k_i^{obs} (see below).

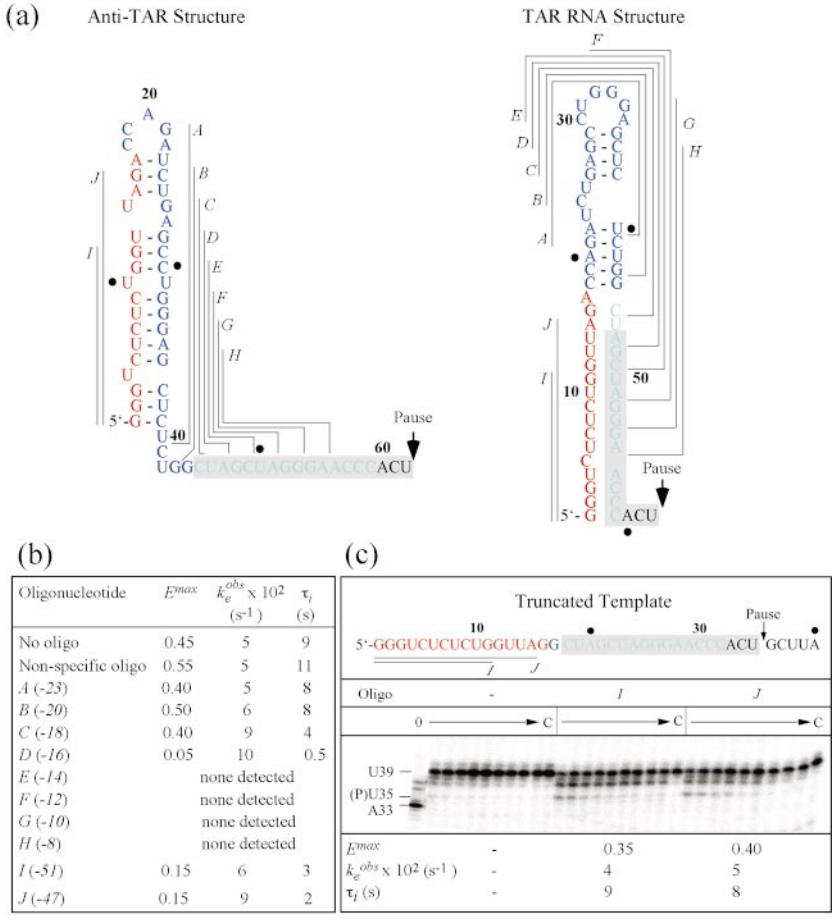
For the wild-type HIV-1 pause signal at 1 mM GTP, this assay yielded $k_e^{obs} = 0.044\ s^{-1}$, $E^{max} = 0.45$, and $\tau_i =$ ten seconds (Figure 2(c)). This rate of pause escape is in reasonable agreement with that determined previously using halted U14 complexes that had to transcribe 48 bp before reaching the pause site (0.031 s^{-1}).[10] However, E^{max} is significantly less than the previous estimate of 0.7, possibly reflecting the greater accuracy of the new assay, in which the time required to reach the pause site is negligible.

Upstream RNA structures affect HIV-1 pausing indirectly

We first used this assay to test whether the effects on pausing of the alternative anti-TAR and TAR RNA structures involved their direct inter-

action with RNAPII. To distinguish direct from indirect effects of RNA structure, we performed two types of experiments. First, we tested the effects on pausing of antisense oligonucleotides that anneal to different segments of the nascent transcript and thereby inhibit formation of the anti-TAR or TAR RNA structures. Second, we constructed a transcription template that lacked the region encoding nucleotides 17-43 of the nascent transcript; this deletion eliminated both anti-TAR and TAR from the RNA transcript. If the effects of RNA structure on HIV-1 pausing were indirect, we predicted that antisense oligo nucleotides would affect pausing only by altering the potential for TAR formation or by directly inhibiting backtracking. Further, the deletion template should allow wild-type levels of pausing by removing both the anti-TAR and TAR RNA structures.

We added antisense oligos separately to samples of TECs halted at A60 and then measured their effects on pause recognition and escape. Antisense oligos A-H all should block formation of both the anti-TAR and TAR structures; only the latter part of the series should prevent backtracking by annealing immediately behind RNAPII (Figure 3(a)).[11,22,24] Antisense oligos whose pairing extended to −23 or −20 (from the pause RNA 3′ end; oligos A and B, respectively; Figure 3(b)) had little effect on pausing. Extension of antisense oligo pairing to include the −18 nt accelerated pause escape twofold (oligo C); extension to include −16 additionally decreased pause recognition by a factor of 8 (oligo D); and extension to include the −14 nt or nucleotides closer to the RNA 3′-end abolished pausing completely (oligos E-H).



Oligonucleotide	E^{max}	k_e^{obs} x 10^2 (s^{-1})	τ_i (s)
No oligo	0.45	5	9
Non-specific oligo	0.55	5	11
A (-23)	0.40	5	8
B (-20)	0.50	6	8
C (-18)	0.40	9	4
D (-16)	0.05	10	0.5
E (-14)		none detected	
F (-12)		none detected	
G (-10)		none detected	
H (-8)		none detected	
I (-51)	0.15	6	3
J (-47)	0.15	9	2

E^{max}	-	0.35	0.40
k_e^{obs} x 10^2 (s^{-1})	-	4	5
τ_i (s)	-	9	8

Figure 3. Upstream RNA secondary structures play only an indirect role in HIV-1 pausing. (a) Pause and TAR RNA hairpin structures. Oligonucleotides A-I that anneal to different regions of the pause and TAR RNA hairpin structures are indicated by lines. The numbers in bold are positions on the transcript relative to the 5′-end of the pause RNA. The gray box represents the footprint of RNAPII on the RNA in a TEC.[83] (b) Effect of annealing oligonucleotides to the upstream RNA on pause efficiency and pause escape. The italicized numbers in parentheses indicate the position of the 3′-most RNA base that is sequestered by the oligonucleotide. TECs halted at A60 were elongated to U66 in the presence of oligonucleotides (100 nM), and pause efficiency and pause escape rate calculated as in Figure 2. (c) Effect of truncation of the upstream RNA on pause recognition. TECs halted at position A33 (analogous to A60 on wild-type template) were elongated through the pause site at U35 (analogous to U62 on wild-type template) in the presence of CTP, GTP and UTP, and with or without the indicated oligonucleotide. The gels were quantified and pause efficiency and pause escape rate determined as for Figure 2.

These results suggest strongly that RNA structure plays no direct role in HIV-1 pausing, since elimination of both the anti-TAR and TAR structures by oligo A or B had no effect on pausing. The results also are consistent with backtracking of the paused RNAP, because inhibition of backtracking with oligos E-H abolished pausing. However, the paused RNAP must be backtracked by no more than about 2 nt because at least 14 nt are required to fill RNA-binding sites within RNAP,[47] yet sequestration of nucleotides more than 16 from the RNA 3′ end had only modest effects on pausing. Finally, these results show that formation of TAR can inhibit HIV-1 pause recognition as well as accelerate pause escape: oligos I and J, which interfere with anti-TAR formation and thus stabilize TAR, reduced E^{max} by a factor of 3, and oligo J, which should completely disrupt anti-TAR, additionally accelerated pause escape twofold.

Based on the effects of antisense oligos, we were surprised to find that deletion of template base-pairs that encoded HIV-1 transcript nucleotides 17 to 43 also abolished pausing (Figure 3(c)). Although neither anti-TAR nor TAR can form on this template, we considered the possibility that an RNA structure that is not predicted might still form in the 5′-proximal RNA and inhibit pausing (some RNA structures are notoriously difficult to predict from sequence).[48] To test this idea, we added antisense oligo I or J to TECs halted at A33 on the deletion template. Both oligos restored pausing to near wild-type levels (Figure 3(c)). Thus, oligos I and J must block formation of RNA structure near RNAPII on the deletion template and thereby promote pausing, even though they inhibit pausing on the wild-type template by blocking formation of anti-TAR and therefore promoting formation of TAR. This interpretation is consistent with the additional observation that complete removal of upstream RNA from wild-type TECs by addition of RNaseH and antisense oligos had no effect on pausing (e.g. with oligo C and RNase H; data not shown). Therefore, we concluded that the only effects of RNA secondary structure on HIV-1 pausing are indirect. Stabilization of TAR (possibly by binding of Tat)[10] promotes escape from the pause by sequestering upstream RNA and driving forward translocation of RNAPII, but the anti-TAR RNA structure must act simply by inhibiting TAR formation and allowing backtracking.

Predicted RNA:DNA stability correlates with HIV-1 pausing by human RNAPII

If HIV-1 pausing depends principally on a positional equilibrium between the active TEC and more stable backtracked TECs, then it should be possible to predict pausing by calculating the relative thermodynamic stability of the TEC at various template positions. A method for this calculation was first described by Yager & von Hippel.[49] We have updated their approach to include the more accurate estimate of a constant-length, 9 bp RNA:DNA hybrid (in the post-translocated conformation)[19,25,26,50,51] and a formula for calculating RNA:DNA hybrid stability based on thermal denaturation experiments.[52] We constructed ten variants of the HIV-1 pause signal that changed the relative strength of active and backtracked RNA:DNA hybrids, and then measured E^{max} and k_e^{obs} for each (Table 1). To evaluate the positional equilibrium model of pausing, we then compared these measurements to the predicted ΔG° of either the RNA:DNA hybrid alone (in the active, pretranslocated conformation *versus* backtracked conformations), or the same conformations of the TEC itself by including the contributions of the DNA bubble and RNAP-nucleic acid interaction as estimated by Yager & von Hippel[49] (see Materials and Methods). The predicted ΔG° values are useful to compare relative stabilities of hybrids and TECs (e.g. to calculate $\Delta\Delta G^\circ_p$, the free energy change for formation of the paused TEC), rather than as absolute measures of TEC stability.

We found that significant pausing occurred at the HIV-1 pause site whenever the RNA:DNA hybrid in the active (pretranslocated) conformation was predicted to be less stable than −11.5 kcal/mol (1 cal = 4.184 J) and backtracking could increase the predicted stability of the TEC ($-\Delta\Delta G^\circ_p > 0.5$ kcal/mol). In other words, pausing required that backtracking be energetically favorable, and that the hybrid be relatively weak. This is most evident by comparing templates 4, 5, and 7 to templates 1 and 2 (Table 1). Templates 4, 5, and 7 all yield predicted TEC stabilization upon backtracking at the pause site of greater than 2 kcal/mol, but do not trigger pausing. In contrast, templates 1 and 2 both support pausing, even though they yield a predicted TEC stabilization upon backtracking of less than 2 kcal/mol. The difference appears to be the absolute stability of the hybrid in the active conformation. The templates that do not support pausing specify RNA:DNA hybrids that are more stable than −12.5 kcal/mol, whereas templates that support pausing specify less-stable hybrids. However, an energetic advantage upon backtracking also is required: template 8 specifies a weak RNA:DNA hybrid, but does not support pausing, apparently because backtracking does not significantly increase TEC stability ($-\Delta\Delta G^\circ_p < 0.5$ kcal/mol).

Our conclusions about the relative stability of the paused TEC relied on an assumption that it is backtracked 1 bp relative to the pretranslocated conformation (Figure 1). Although this is consistent with available data,[10] we wanted to examine the correlation among predicted TEC stability, backtracking, and pausing more broadly. Therefore, we calculated the thermodynamic stability of paused TECs in different backtracked registers on the set of mutant templates shown in Table 1, and compared these predictions to the susceptibility of the paused TECs to TFIIS-stimulated transcript cleavage. The rate of transcript cleavage is known

Table 1. TEC stability and observed kinetic parameters

Templates		$\Delta G^{\circ}_{RNA:DNA}$	$\Delta\Delta G^{\circ}_{p,RNA:DNA}$	$\Delta\Delta G^{\circ}_{p,TEC}$	E^{max}	k_e^{obs} x 10^2	τ_i
		(kcal/mol)				(s^{-1})	(s)
	50 60 P 66						
WT	GCUAGGGAACCCACU▼GCUUAA	-9.4	-2.0	-3.1	0.45	4.4	10
#1	GCGGGUAAACCCACU GCUUAA	-9.1	0.3	-0.7	0.15	9.0	1.5
#2	GCUGGGAAACCCACU GCUUAA	-9.1	-0.4	-1.5	0.25	5.5	4.5
#3	GCUAAGGGACCCACU GCUUAA	-11.3	-2.0	-3.1	0.55	16	3.5
#4	GCUAAACGGGCCACU GCUUAA	-13.1	-1.2	-2.3	0.02	13	0.15
#5	GCUAAACCCGGGACU GCUUAA	-12.9	-1.2	-2.3	-	-	<0.1[a]
#6	GCUAGGGAAGGGACU GCUUAA	-11.1	-2.0	-3.1	0.25	4.0	6.0
#7	GCUAAACCCACGGGU GCUUAA	-12.7	-1.0	-2.1	-	-	<0.1
#8	GCUAGUUAACCCACU GCUUAA	-8.7	0.7	-0.4	-	-	<0.1
#9	GCUAAGGCCGGGACU GCUUAA	-12.3	0.7	-0.3	-	-	<0.1
#10	GCUAAUUCCGGGACU GCUUAA	-12.3	0.7	-0.3	-	-	<0.1

Pretranslocated
Paused (-1 backtracked)
RNA:DNA hybrid

The gray boxes indicate base changes in the nascent RNA relative to the wild-type sequence. The stability of the 9 bp RNA:DNA hybrid alone and of the TEC were calculated as described in Materials and Methods for the active conformation (pre-translocated, $\Delta G^{\circ}_{RNA:DNA}$ and ΔG°_{TEC}) and for the paused conformation (−1 backtracked; reported as $\Delta\Delta G^{\circ}_{P,\ RNA:DNA}$ and $\Delta\Delta G^{\circ}_{p,TEC}$, the predicted relative stabilities of the RNA:DNA hybrid and the TEC in the paused *versus* active conformations). Pause efficiency and pause escape rate measurements were performed at 1 mM GTP as described in Materials and Methods and the legend to Figure 2. These values reported for E^{max}, k_e^{obs}, and τ_i are the average of two independent measurements that varied by no more that 15%. E^{max} was rounded to the nearest 0.05, except for mutant 4; other measurements are reported to two significant figures.

[a] Based on the lowest efficiency and highest escape rates that we were able to detect, we estimate that pause signals with τ_i values up to ~0.1 would not be detectable in our assay.

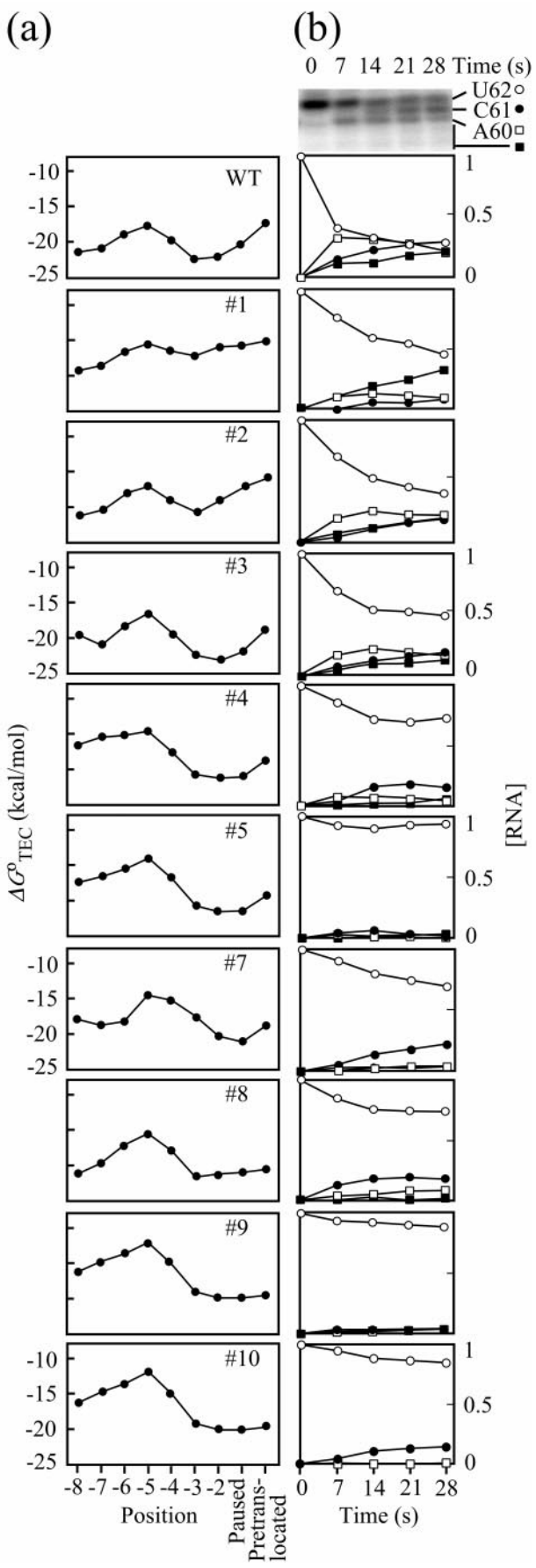
to correlate with the propensity of TECs to backtrack.[23,24]

Comparison of predicted ΔG°_{TEC} in different backtrack registers to rates of transcript cleavage supports the idea that an energetic potential for backtracking correlates with pausing (Figure 4). For all templates that supported pausing, backtracking by 1-3 bp was predicted to yield more stable TECs and more than half of the paused (U62) transcripts were cleaved in 30 seconds. The wild-type paused TEC, which was the strongest (Table 1), gave the fastest cleavage rate and greatest energetic advantage upon backtracking (template 6 was not tested). In contrast, templates that did not support pausing (Table 1) allowed less than half of the U62 RNA to be cleaved in 30 seconds and yielded less or no predicted stabilization upon backtracking. However, it was notable that some templates (e.g. template 5) supported neither pausing nor transcript cleavage, even though they were predicted to yield stabilized TECs upon backtracking. Apparently, transcript cleavage, like pausing, requires a relatively weak RNA:DNA hybrid in addition to energetic advantage upon backtracking.

A simple explanation for our results is that a strong RNA:DNA hybrid creates kinetic barriers to backtracking and pausing that prevent them from occurring on the same time scale as transcription (i.e. in less than one second). This explanation contradicts the equilibrium model of transcriptional pausing, which requires that more stable backtracked states be accessed more rapidly than nucleotide addition occurs.

RNA:DNA hybrid stability affects the rate of isomerization into the paused conformation

To test the idea that the barrier to pausing is kinetic rather than an energetic disadvantage in a positional equilibrium, we studied a pause-resistant template (template 5) more carefully. Active TECs on template 5 are predicted to be slightly more stable than wild-type prior to backtracking, but to exhibit similar relative changes in hybrid and TEC stability upon backtracking by 1 bp (Table 1). To ask if paused complexes could form on template 5 if given time to overcome a kinetic barrier, we withheld GTP for increasing amounts of time after CTP and UTP were added to A60 TECs (Figure 5(a)). For the wild-type template, E^{max} rose from 0.45 to 0.60 in less than five seconds, with no change in k_e^{obs} (0.044 s^{-1}; Figure 5(b), (c), and (d)). However, for template 5, paused TECs formed slowly, to a final E^{max} of ~0.2 after 120 seconds (Figure 5(b), (c), and (d);

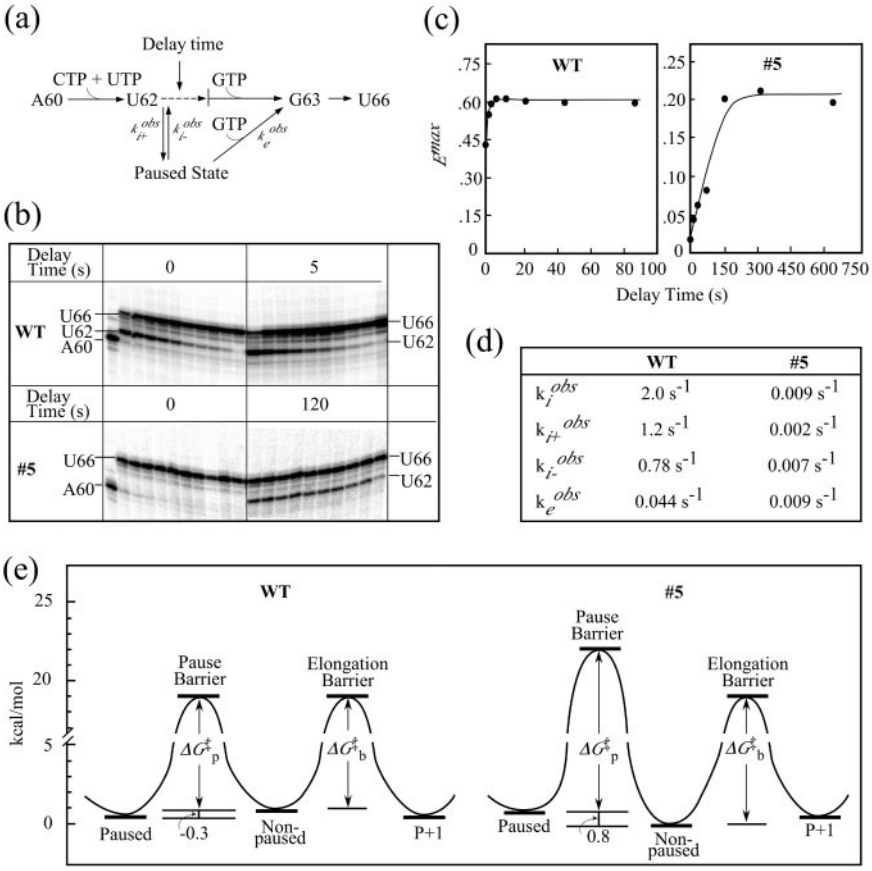


$k_e^{obs} = 0.009\ s^{-1}$). Assuming the final E^{max} values reflect an equilibrium between nonpaused and paused TECs, they correspond to $\Delta\Delta G^\circ_p$ of −0.3 kcal/mol for wild-type and +0.8 kcal/mol for template 5 (*versus* predicted $\Delta\Delta G^\circ_p$ of −3.1 kcal/mol and −2.3 kcal/mol, respectively; Table 1). From the time-dependence of paused TEC appearance and the final E^{max}, we estimated k_i^{obs}, k_{i+}, and k_{i-} for both the wild-type and 5 paused TECs (Figure 5(d); see Materials and Methods).

These results support the view that stronger base-pairing in the RNA:DNA hybrid generates a kinetic barrier to paused TEC formation, but that paused complexes can still form if the dwell time at the pause site is increased enough to allow isomerization. In other words, a strong hybrid decreases k_i^{obs}. Further, the slower k_e^{obs} on template 5 suggests that the stable hybrid may also decrease the rate of pause escape.

The effect of the RNA:DNA hybrid on paused TEC formation can be interpreted as a change in the activation energy required for paused TEC formation (Figure 5(e); see von Hippel[35] for a complete description of this approach). To obtain hypothetical energy diagrams for pausing, we assumed that k_b^{obs} for GMP addition in the active conformation at 1 mM GTP was approximately equal to our estimate of k_{i+} on the wild-type template because E^{max} is about 50 % on this template at 1 mM GTP. We then used k_{i+}, k_{i-}, and k_b^{obs} to calculate $\Delta G^\ddagger$ for the wild-type and template 5 U62 TECs (plotted relative to arbitrary assignment of zero kcal/mol for the active TEC conformation on template 5 in Figure 5(e); see Materials and Methods). These diagrams illustrate why paused complexes do not form on template 5 when nucleotide addition is possible: the energetic barrier to paused TEC formation is greater than the barrier to nucleotide addition. When GTP is withheld, paused TECs form slowly, reaching equilibrium at 20 % of the population because they are less stable than non-paused complexes. Interestingly, on template 5 the barrier to reforming the non-paused

Figure 4. RNA:DNA hybrids that support pausing also favor TFIIS-mediated transcript cleavage in paused TECs. (a) Predicted stability of U62 TECs (ΔG°_{TEC}) on wild-type and mutant templates plotted as a function of TEC register (see Materials and Methods). The −1 backtracked TEC corresponds to the proposed conformation of the paused TEC (see Figure 1). The templates are the same set shown in Table 1. (b) Quantitative analysis of TFIIS-mediated transcript cleavage of paused transcription complexes. TECs halted at A60 were elongated with CTP and UTP (100 μM each) for one minute at 30 °C (time 0), washed rapidly with transcription buffer and incubated with TFIIS (1.25 nM) at 30 °C for indicated times. The U62 pause (open circle), C61 (filled circle), A60 (open square) and shorter than A60 (filled square) RNA species were quantified and are shown as a fraction of the total RNA in each lane.



	WT	#5
k_i^{obs}	2.0 s^{-1}	0.009 s^{-1}
k_{i+}^{obs}	1.2 s^{-1}	0.002 s^{-1}
k_{i-}^{obs}	0.78 s^{-1}	0.007 s^{-1}
k_e^{obs}	0.044 s^{-1}	0.009 s^{-1}

Figure 5. A weak RNA:DNA hybrid induces pausing by enhancing the rate of isomerization into the paused conformation. (a) A representation of the kinetic assay with delayed addition of GTP. (b) Time-course of elongation through the HIV-1 pause site on the wild-type and template 5 under the standard (left panel) and delayed (right panel) assay conditions. (c) Pause efficiency and pause escape rate measurements were performed as for Figure 2. Pause efficiency is plotted as a function of delay time. (d) k_i^{obs} and k_e^{obs} measurements for the wild-type and variant 5 determined as described in Materials and Methods. (e) Energy diagram illustrating the activation energy barrier to pausing on wild-type and mutant templates. The energy barriers to pausing ($\Delta G_p^{\ddagger}$) and forward elongation ($\Delta G_b^{\ddagger}$) were calculated from k_{i+}^{obs}, k_{i-}^{obs} and E^{max} as described in the text and in Materials and Methods. The energy differences between the non-paused and paused TECs should correspond to $\Delta\Delta G_p^{\circ}$ (compare to Table 1).

conformation is greater than the barrier to nucleotide addition. This predicts that isomerization will be rate-limiting for pause escape on template 5, and that the rate of pause escape will be insensitive to changes in the concentration of GTP (near 1 mM, where effects are seen on the wild-type). In agreement with this prediction, paused TECs formed on template 5 by incubation of A60 TECs with UTP and CTP but no GTP displayed the same rate of GMP incorporation upon addition of 1 mM to 25 μM GTP (see below).

In contrast, for the wild-type, the activation energy barrier to pausing is comparable to the barrier to elongation, allowing pause complex formation to compete with nucleotide addition. Once formed, the wild-type paused complexes face a slightly greater barrier to reforming a non-paused conformation, but the lower barrier still allows equilibrium to be reached in a few seconds. Because the TECs can pass back and forth over this barrier on the same time-scale as nucleotide addition, the rate of pause escape depends on substrate binding to the non-paused portion of the population, and thus on the concentration of NTP.

Paused TEC formation occurs *via* initial formation of an unactivated intermediate

Our experiments with the mutant template 5 also yielded evidence for an unactivated intermediate that is distinct from the active and paused species and that is consistent with the unactivated state described by Erie *et al.*[12] We detected this intermediate when we examined the [GTP]-dependence of pausing on template 5. Two different slow kinetic species, neither of which could be the active TEC, were detected depending on whether or not RNAPII was delayed at the pause site by temporary omission of GTP. When RNAPII was not held at U62, no pause occurred with 1 mM

GTP (Figure 5(b)); however, at or below 100 μM GTP, a fraction of U62 TEC ($E^{max} = \sim 0.25$) added GMP slowly, with k^{obs} ranging from 0.14 s^{-1} (at 100 μM GTP) to 0.04 s^{-1} (at 25 μM GTP; Figure 6(c)). The properties of this [GTP]-sensitive species suggest it could be the unactivated intermediate. It cannot be the active species, because it represents only ~25 % of the U62 TEC. It cannot be the paused TEC, because the paused TEC, when generated by withholding GTP for 120 seconds, added GMP at a rate insensitive to the concentration of GTP ($k_e^{obs} = 0.009$ from 1 mM to 25 μM GTP; Figure 6(d)). This contrasts with the response of the wild-type paused species formed in the delay assay, which exhibited the same [GTP]-dependent rates of pause escape as observed in the non-delay assay (compare Figure 6(a) and (b) to (c) and (d)). As described above, this difference in the responses of the wild-type and mutant 5 paused TECs to GTP concentration is consistent with a greater energy barrier to formation and escape of the mutant paused TEC.

We conclude that the behavior of the slow species observed on template 5 at or below 100 μM GTP is consistent with the expected properties of the unactivated intermediate. We cannot distinguish whether this unactivated intermediate becomes detectable because lowering the GTP concentration increases the dwell time of RNAPII at U62 and allows it to form or because it always forms, but binds >100 μM GTP so fast that it is

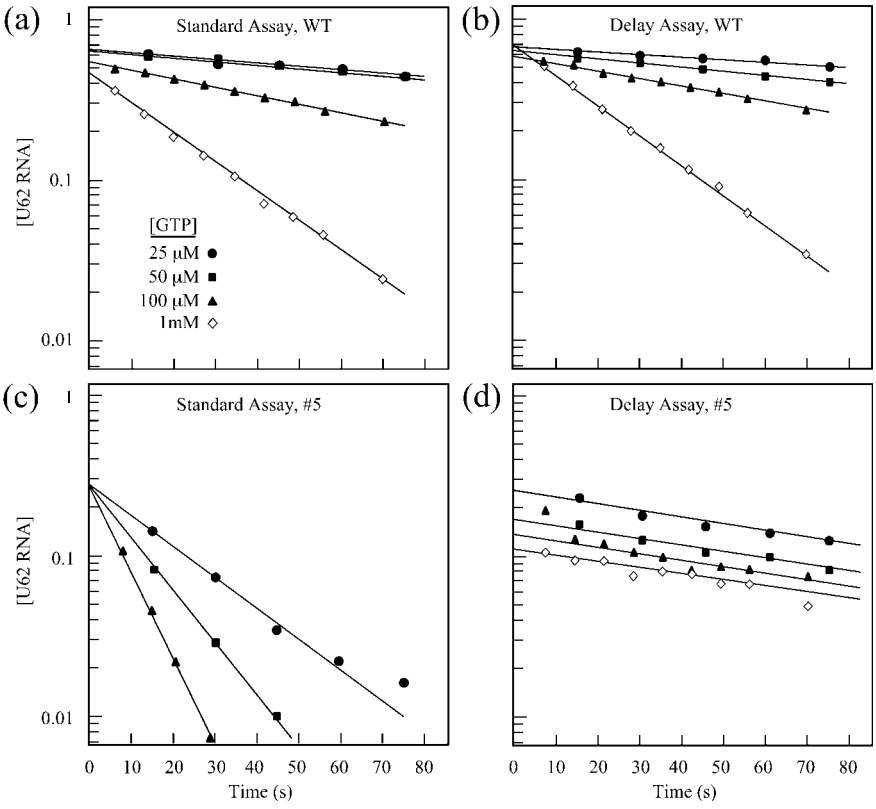


Figure 6. Evidence for an unactivated intermediate in pause recognition. Effect of [GTP] on pause recognition and escape on the wild-type and mutant 5 templates. TECs halted at A60 were elongated in the presence of CTP, UTP (1 mM each), and varying concentrations of GTP (25 μM, circle; 50 μM, square; 100 μM, triangle; 1 mM, diamond), either using the standard assay conditions (Figure 2) or by delaying the addition of GTP until paused TEC formation reached equilibrium (Figure 5(c)). (a) Relative concentration of U62 RNA after addition of CTP, GTP, and UTP to wild-type A60 TECs (standard assay, Figure 2). k^{obs} for GMP addition increased with increasing [GTP]: 25 μM, 0.004 s^{-1}; 50 μM, 0.006 s^{-1}; 100 μM, 0.012 s^{-1}; 1 mM, 0.043 s^{-1}. (b) Relative concentration of U62 RNA when GTP was added 30 seconds after addition of CTP and UTP to wild-type A60 TECs (delay assay, Figure 5(a)). k^{obs} for GMP addition increased with increasing [GTP]: 25 μM, 0.004 s^{-1}; 50 μM, 0.007 s^{-1}; 100 μM, 0.011 s^{-1}; 1 mM, 0.042 s^{-1}, identical within error with those observed in standard assay and consistent with the formation of a single paused species in both conditions. (c) Relative concentration of U62 RNA after addition of CTP, GTP, and UTP to mutant 5 A60 TECs (standard assay, Figure 2). An intermediate species ($E^{max} = \sim 0.25$) was detectable at ⩽100 μM GTP. k^{obs} for GMP addition increased with increasing [GTP]: 25 μM, 0.04 s^{-1}; 50 μM, 0.077 s^{-1}; 100 μM, 0.14 s^{-1}. The rate of GMP addition on the mutant 5 template was too fast to measure at 1 mM GTP. (d) Relative concentration of U62 RNA when GTP was added 120 seconds after addition of CTP and UTP to mutant 5 A60 TECs (delay assay, Figure 5(a)). k^{obs} remained unchanged (0.009 s^{-1}) at all GTP concentrations tested; E^{max} increased from 0.1 at 1 mM GTP to 0.25 at 25 μM GTP. These responses of the paused TEC to changes in GTP concentration suggest that the species detectable at low [GTP] in the standard assay (see (c)) is the unactivated intermediate, and not the paused TEC.

detectable only at lower concentrations (see Discussion).

Discussion

The results from this study lead to three major conclusions. First, although TAR hairpin formation can inhibit pausing indirectly, upstream RNA secondary structures play no direct role in HIV-1 pausing. Rather, pausing appears to be determined principally by a relatively weak RNA:DNA hybrid coupled with energetically favorable backtracking of the TEC. Second, a kinetic barrier, rather than an energetic disadvantage in a positional equilibrium, prevents paused TEC formation at many template positions. A weak RNA:DNA hybrid enhances the rate of isomerization into the paused conformation; a strong hybrid reduces this rate several hundred-fold. Third, a distinct unactivated intermediate appears to form prior to backtracking and paused TEC formation.

RNA structures affect the HIV-1 pause indirectly

Artsimovitch & Landick[11] recently defined two classes of bacterial pause signals, one of which involves an allosteric interaction of a nascent RNA secondary structure with RNAP, and the second of which involves backtracking, but not RNA secondary structure. The HIV-1 pause corresponds to the second class of pause signal but, interestingly, it can be affected indirectly by the TAR RNA structure if TAR can form at the pause site and inhibit backtracking (Figure 3). Base changes that allow TAR to outcompete anti-TAR increase pause escape (up to threefold).[10] Therefore, stabilization of TAR through interactions of elongation factors like Tat and P-TEFb also could accelerate pause escape. A possible biological role of the HIV-1 pause could be to delay RNAPII near the promoter and allow productive interactions of these factors, after which pause escape could be stimulated.[10]

Recognition of the HIV-1 pause site requires a non-equilibrium transition

Our finding that the barrier to HIV-1 paused TEC formation is kinetic is consistent with models of both hairpin-dependent and hairpin-independent pausing derived from studies of bacterial RNAP.[11,53] However, it is inconsistent with the idea that RNAP establishes a positional equilibrium at most template positions that includes conformations other than the pre- and post-translocated species (see Introduction). Instead, transcriptional arrest resulting from backtracking of TECs, which has been detected when either bacterial RNAP or mammalian RNAPII are halted by NTP deprivation,[22–25] is unlikely to occur during active chain elongation except at special sequences that lower the kinetic barrier to backtracking. This conclusion is consistent with a previous study by Komissarova & Kashlev,[24] who showed backtracking does not occur immediately upon halting TECs by NTP deprivation, and Gu and Reines,[21] who showed that transcriptional arrest occurs relatively slowly after an initial event puts RNAPII in a paused state.

Paused TEC formation requires a conformational change that affects RNAPII's active site

What is the physical basis of this relatively slow isomerization to a paused TEC? We propose that the isomerization reflects a conformational change in RNAP, most likely involving movement of RNAPII's clamp domain[18] (see Figure 7(a)) to open the active-site cleft, slightly shift the position of the RNA:DNA hybrid, and allow backtracking to occur.[11,20] We base this postulate on the following arguments.

First, some templates that do not support pausing (e.g. templates 4, 5, and 7) yield the same or nearly the same predicted stabilization of the TEC upon backtracking as does template 1, which supports pausing (Table 1). Although in many cases, efficiency and longevity of pausing correlated with the energetic predictions, a weak RNA:DNA hybrid in the active conformation (predicted $-\Delta G^{\circ}_{\mathrm{RNA:DNA}} < 11.5$ kcal/mol) appears to be a prerequisite for pausing to occur at saturating NTP concentrations, irrespective of the predicted energetic advantage of backtracking.

Second, formation and breaking of base-pairs in nucleic acids occur on a much more rapid timescale than the observed rate of isomerization to the paused conformation. Although the 3.5 kcal/mol difference between predicted hybrid stabilities of the wild-type and mutant 5 U62 TECs is sufficient to account for the difference in their rates of paused TEC formation (see Figure 5(e)), it seems unlikely that the ~20 kcal/mol energy barrier required to account for the slow rate of isomerization at even the wild-type pause site reflects only changes in base-pairing. For instance, at 25 °C, formation of a single base-pair at the end of a duplex DNA occurs at 8×10^6 s^{-1}; dissociation occurs at 2×10^6 s^{-1}.[54] These rates are many orders of magnitude faster than even wild-type paused TEC formation and suggest that translocation of a constant-length duplex and bubble within a rigid RNAP cannot explain the observed activation energy barrier for pausing. Rather, the magnitude of this energy barrier is more consistent with a protein conformational change, whose rates can range from fast to at least as slow as 0.003 s^{-1}.[55]

Third, single-molecule measurements of the force required to stall bacterial RNAP reveal that the rate of nucleotide addition does not slow gradually as a force that resists forward translocation is increased.[56,57] Rather, at most template positions, RNAP maintains a near-constant reaction velocity until the force reaches close to 30 pN, at which point transcription stops abruptly and coincident with reverse translocation over ~10 bp.

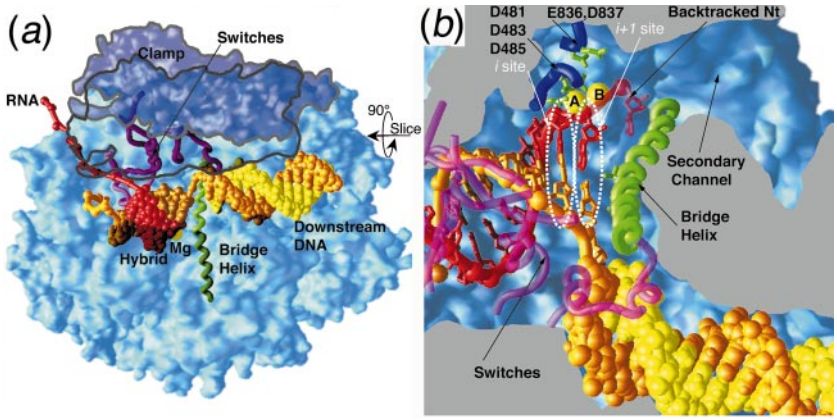


Figure 7. Model of paused RNAPII. The yeast RNAPII TEC structure reported by Gnatt *et al*. (PDB coordinates 1I6H)[19] is shown in the same orientation as the TEC diagram in Figure 1(a), but with some RNA, DNA and protein loops modeled on the basis of proposed locations.[18,19,50] RNAPII is blue and semitransparent, the RNA transcript is red, the template DNA is gold, the non-template DNA is yellow, the catalytic magnesium ions are yellow, and the bridge helix is green. (a) Clamp movement in RNAPII. The downstream DNA and RNA:DNA hybrid are shown in CPK spacefill; for clarity other nucleic acid segments are shown only as a backbone are removed entirely (non-template DNA strand and all upstream DNA). The clamp position in free RNAP (PDB coordinates 1I50)[18] is shown in dark blue with gray outline. Five RBP1 and RBP2 segments (called switches)[18] that fold upon clamp closure and interact with RNA and DNA in the TEC are magneta. The closed position of the clamp is outlined in dark gray, (b) The active-site channel, secondary channel, and key features are revealed by slicing the view in (a) at the arrow shown, rotating 90° toward the viewer to remove the clamp and other obscuring protein segments, and enlarging. In this view, the non-backbone parts of the RNA:DNA hybrid are rendered as bonds to reveal the underlying, RPB2-portion of the active-site channel (he RPB2 D loop II would project upwards toward the base pair in the $i+1$ site; see the text). Magnesium ion A, the 3′ nucleotide (in the $i+1$ site), and other key features are positioned as in the yeast TEC structure.[19] Magnesium ion B is positioned relative to A based on the locations of catalytic magnesium ions in the *Thermus aquaticus* replication complex (PDB coordinates 3KTQ).[84] Note its displacement from E836 and D837, which are thought to chelate magnesium ion B in an active TEC.[18,19] Upon backtracking in the paused conformation, the 3′ nucleotide enters the secondary channel (shown as a semitransparent red nucleotide).

Such behavior suggests that RNAP ordinarily resides in a stable conformation that resists backtracking even when significant force is applied. In other words, the elongating TEC exists in a backtrack-resistant conformation that must undergo a significant conformational change to allow direct coupling between an externally applied force and translocation of RNAP. Pause and arrest signals may make this conformational change more favorable and also explain why the stall force is much lower at certain template positions (around 12 pN).[56]

Fourth, neither yeast RNAPII nor bacterial RNAP recognize the HIV-1 pause signal even when the GTP concentration is lowered to the point that pausing at other sites becomes evident (M. P., unpublished results). Therefore, pausing cannot be a simple consequence of changes in nucleic-acid base-pairing, since all multisubunit RNAPs appear to have the same RNA and DNA geometry in a TEC. Instead, some feature of RNAP itself must contribute to paused TEC formation. One simple explanation is that the propensity for a pause-inducing conformational change differs in a template-sequence-dependent fashion among various multisubunit RNAPs.

Fifth, in the absence of RNA and DNA, both bacterial and eukaryotic RNAPs crystallize in a relatively open conformation.[18,58] In these structures, the active-site cleft (which surrounds the RNA:DNA hybrid) forms a larger volume than required to enclose a 9 bp RNA:DNA hybrid because a clamp-like domain rotates away from the cleft.[18] The clamp closes over the RNA:DNA hybrid when RNAP forms a TEC.[19] In the bacterial TEC, the active-site cleft encloses the RNA:DNA hybrid to the extent that even an OH radical cannot penetrate and react with the hybrid backbone.[59] Closing movements of other portions of the enzyme around the downstream DNA and the exiting RNA are either well documented or highly likely.[18,19,50,58,60,61] Thus, models of pausing based on opening and closing movements of parts of RNAP are plausible, given the structure of the enzyme.

Sixth, the 3′ nucleotide in a halted bacterial TEC appears to move away from the i or $i+1$ site even before backtracking occurs. Markovstov *et al*. reported crosslinking of 3′-terminal 4-thiouridine to β residues 1097-1107 in a non-backtracked TEC.[62] This segment of β is >20 Å from the pre-translocated 3′ nucleotide position predicted by the TEC model described by Korzheva *et al*.[50] and observed in a yeast RNAPII TEC.[19] Thus, some change in a halted TEC appears to allow repositioning of the 3′-terminal nucleotide. This change is consistent with a possible loss of proper 3′-end

alignment in the unactivated intermediate (Figure 1(b)).

Finally, simple consideration of the reaction cycle of the TEC seems inconsistent with the view that the energetics of base-pairing alone generates the kinetic barrier to formation of paused (backtracked) TECs. To move forward, RNAP must shift from the pre-translocated to the post-translocated conformation (Figure 1). This shift requires melting a base-pair at the front edge of the transcription bubble, re-annealing a base-pair at the back edge of the transcription bubble, and shifting the RNA:DNA hybrid by 1 bp along the template DNA strand. These changes in nucleic-acid base-pairing must happen rapidly during transcription, yet there is no evidence that a strong RNA:DNA hybrid slows forward translocation, which itself would create a type of transcriptional pausing.[63] If nucleic-acid base-pairing can shift rapidly in the forward direction, there is no reason *a priori* to believe it would be significantly slower in the reverse direction. Rather, it makes sense that the structure of the TEC itself limits such backward movements in the absence of a conformational rearrangement.

Possible structural basis for paused TEC formation

Taken together, our results and the arguments listed above strongly favor the idea that a conformational change in RNAPII is required for HIV-1 pausing. The results are consistent with either the essential or non-essential activation versions of unactivated-intermediate-type mechanisms (see Introduction and Figure 1).[11–16,64] Further, our results with the mutant template 5 directly support the existence of an unactivated intermediate (Figure 6).

What is the nature of this conformational change and at what point in formation of the paused TEC does it occur? To us, the most appealing view is that a significant conformational change that partially opens the active-site cleft occurs between the activated and unactivated intermediates, and that paused states of the TEC arise from the unactivated intermediate subsequently by nucleic acid rearrangements that lessen the probability of proper RNA 3′-end alignment in the active site (reverse translocation of the RNA:DNA hybrid or, in bacteria, nascent RNA hairpin formation). In this view, movement of the clamp domain away from the active-site cleft would shift side-chains in the active site away or the position of the RNA:DNA hybrid and allow RNA 3′ nucleotide to assume locations other than the i and $i+1$ site.[20] The likely candidates for active-site movements are the bridge helix, which contacts the downstream end of the RNA:DNA hybrid,[18,19] the acidic side-chains that chelate the catalytic magnesium ions, and a loop, called βD loop II in bacterial RNAP,[50] that extends from RPB2 toward the base-pair in the $i+1$ site (Figure 7(b); the RPB2 D loop II is not shown in the Figure). A homologue of the bridge helix in DNA polymerases, called the O helix, is thought to oscillate toward and away from the catalytic center in each round of nucleotide addition to align the reactive groups and to release pyrophosphate; the magnesium ions coordinately bond the reactive groups in the transition state.[65–67] In the yeast RNAPII TEC, the bridge helix is connected to the clamp through salt-bridges.[19] Therefore, if the clamp opens slightly upon unactivated intermediate formation, the base-pair in the active site may no longer be fixed in place between the clamp and RPB2 D loop II. This could allow the bridge helix and RNA:DNA hybrid to shift position, perhaps to the locations observed in the yeast TEC crystal structure. In this structure, the hybrid appears to be aligned with the secondary channel for backtracking, and the side-chains that chelate Mg^{2+} B are too far from the hybrid to facilitate nucleotide addition (Figure 7(b)).[19]

How can we reconcile a required conformational change in RNAP for pausing with our finding that kinetic accessibility of pausing depends on a relatively weak RNA:DNA hybrid? The recent RNAPII[18] and RNAPII TEC[19] structures support the idea that the stability of the hybrid may be coupled to conformational opening or closing of RNAP by an induced-fit mechanism.[17,20] In induced fit, structural changes that accompany interaction create complementary surfaces.[68] Induced-fit protein-nucleic acid interactions are widespread in biology and frequently involve co-dependent changes in the structures of both the protein and nucleic acid.[69–71] In such interactions, each partner stabilizes the conformation of the other, thermodynamically linking their free energies of formation. When the clamp closes in the TEC, five ''switch'' or hinge regions that are unfolded or disordered in the open yRNAPII structure fold upon interaction with the hybrid and downstream DNA (magenta segments; Figure 7). Because the RNA:DNA hybrid is thermodynamically linked to clamp movement through the folding of these switches, a less stable hybrid may weaken this induced fit, allowing the conformational change that partially opens the active-site cleft in the unactivated intermediate (Figure 1). Partial opening of the active-site cleft by clamp movement could be coupled to separation of the parts of the subunits surrounding the downstream DNA and exiting RNA in cases where upstream RNA structure or downstream DNA sequence affect paused TEC formation directly.

It is likely that an additional conformational change in RNAPII is required to allow backtracking after the unactivated intermediate forms. Based on our dissection of pause kinetics on template 5, the slowest step in paused TEC formation, and thus highest energetic barrier to a structural rearrangement in RNAPII that is required for backtracking, appears to occur after unactivated intermediate formation (Figures 5 and 6). At low [GTP] we could detect the unactivated intermediate even

without halting RNAPII at U62 (Figure 6(c)), suggesting it forms at >0.01 s^{-1} (~25% of TECs forming the unactivated state in five seconds or less). However, the paused TEC forms more slowly from the unactivated intermediate ($k_{i+}^{obs} = 0.002\ s^{-1}$), and then escapes at a rate insensitive to [GTP] at and above 25 μM ($k_{e}^{obs} = 0.009\ s^{-1}$; Figures 5 and 6(d)). This may suggest that only partial clamp, hybrid, or bridge helix movements are required to create the unactivated intermediate and that more substantial movements are required before backtracking is possible. It is unclear whether the halted yRNAPII TEC trapped in the crystal structure is closer to the proposed unactivated or paused conformations.

Implications of an unactivated-intermediate mechanism for elongation factor action

The unactivated-intermediate model suggests that pausing and arrest are off-line states that arise from an elongation-deficient conformation of RNAPII. This idea has important implications for the mechanism of action of elongation factors such as TFIIF, elongin, and ELL, which are known to accelerate the rate of nucleotide addition by suppressing pausing.[2,72] If pausing depends on a conformational opening of RNAP, then it seems probable that these factors could act on RNAPII by stabilizing the elongation-proficient conformation and reducing the propensity to enter the off-line pathway. The assays we report here should allow us to test this idea by direct measurement of the effects of elongation factors on steps in the pause mechanism.

Materials and Methods

Template DNA

The 774 bp DNA fragment used for transcription of the wild-type HIV-1 pause signal was PCR-amplified from the plasmid pLL283 using the universal M13 forward and biotinylated reverse primers.[10] Nucleotide changes in pLL283 were made by PCR-amplifying a DNA fragment from pLL283 using the M13 reverse primer and a second primer that specified the change and that overlapped the unique *Nhe*I site in the plasmid. After digestion with *Nhe*I and *Xho*I, the fragment was ligated between the appropriate sites in pLL283. The deletion template that removed anti-TAR and TAR coding regions (Figure 3(c)) was constructed by PCR-amplifying a DNA fragment from the plasmid pLL283 using the M13 forward primer and a reverse primer that specified the *Nhe*I site and a deletion of the DNA fragment between +17 and +43 (both positions included) relative to the transcription start site. An *Eco*RI-*Nhe*I fragment from the PCR product was ligated between these two sites in pLL283 and the deletion template was prepared by PCR from this new plasmid. Transcription from the deletion template yields a pause RNA 35 nt long. All transcription templates were immobilized on streptavidin-agarose beads (Sigma, St. Louis, MO; 125 μl suspension) by washing beads with 10 mM Tris-HCl (pH 7.5), 1 M NaCl, 0.25 mM Na_2EDTA and then incubating with 50 μg of DNA template in 125 μl of the same buffer at 22 °C for 60 minutes with intermittent mixing. The beads were then washed with the same buffer and 0.15 M NaCl, stored in suspension at 4 °C with 0.01% (w/v) sodium azide. Prior to use, the beads were washed with water and suspended in transcription buffer (see below).

Proteins and substrates

Nuclear extracts[73] were prepared from HeLa cells procured from Cell Culture Center (Connrapids, MN). Transcription factor TFIIS was overexpressed in *E. coli* and purified as described by Yoo *et al*.[74] FPLC-purified NTPs and 2′ dNTPs were obtained from Pharmacia (Piscataway, NJ). HPLC-purified oligonucleotides were from Operon Technologies (Alameda, CA).

Preparation of TECs and stepwise transcription

Preinitiation complexes were formed by incubating immobilized template (0.5 μg) with 2 μl of nuclear extracts for 20 minutes at 30 °C in 16 μl of transcription buffer (10 mM Hepes (pH 7.9), 33 mM KCl, 8 mM $MgCl_2$, 1 mM DTT, 100 μM Na_2EDTA, 100 μM dATP, 6% (v/v) glycerol) and 20 units of Prime RNase Inhibitor (Eppendorf, Westbury, NY). Transcription was initiated for one minute at 30 °C by the addition of 4 μl of initiating NTPs (100 μM CTP, 100 μM GTP and 0.5 μM [α-^{32}P]UTP (10 μCi)) in transcription buffer. The reaction was diluted tenfold with 1% (w/v) Sarkosyl in 20 mM Tris-HCl (pH 8.0), 1 M KCl, 10 mM β-mercaptoethanol, 0.2 mM Na_2EDTA, 20% (v/v) glycerol. The TECs were collected by centrifugation at 2000 rpm for two minutes at 4 °C, washed with the above buffer without Sarkosyl and containing 100 mM KCl, then washed with transcription buffer containing 200 μg of acetylated bovine serum albumin/ml, and suspended in transcription buffer. TECs were then elongated in a stepwise manner to position A60, 2 nt upstream from the HIV-1 pause site at U62, as described.[10] For pausing measurements, TECs halted at position A60 were elongated through the pause site to position U66 in the presence of chase NTPs (CTP, GTP and UTP, 1 mM each) at 30 °C. Aliquots were removed at indicated times and terminated into stop buffer (125 mM Tris-HCl (pH 8.0), 15 mM Na_2EDTA, 333 mM NaCl, 1.25% (w/v) SDS, 175 μg tRNA/ml). The RNA was extracted with phenol/chloroform (1:1, v/v), ethanol-precipitated, dissolved in formamide loading dye (95% (v/v) formamide, 15 mM Tris-HCl (pH 7.9), 5 mM EDTA, 0.1% (w/v) each bromophenol blue and xylene cyanol) at 90 °C for three minutes and then separated on an 8% (w/v) polyacrylamide gel (19:1 (w/w) acrylamide/bis-acrylamide) containing 8 M urea, 44 mM Tris-borate (pH 8.3) and 1.25 mM EDTA. When oligonucleotides were annealed to the nascent transcript (Figure 3), the halted A60 complexes were incubated with oligonucleotides (100 nM final) at 30 °C for five minutes before addition of NTPs. Gels were analyzed using a Phosphorimager (Molecular Dynamics, Sunnyvale, CA) and quantified using the Image QuaNT software (Molecular Dynamics, Sunnyvale, CA) as shown in Figure 2(b) and as described.[45]

Calculation of relative TEC stability

The predicted free energy of formation of the RNA:DNA hybrid ($\Delta G^{\circ}_{RNA:DNA}$) and the TEC ($\Delta G^{\circ}_{TEC}$) were calculated as described by Yager & von Hippel[35,49] using

equation (1), with appropriate modifications:

$$\Delta G^{\circ}_{\mathrm{TEC}} = \Delta G^{\circ}_{\mathrm{DNA\ bubble}} + \Delta G^{\circ}_{\mathrm{RNA:DNA}} + \Delta G^{\circ}_{\mathrm{RNAP\ binding}} \quad (1)$$

$\Delta G^{\circ}_{\mathrm{DNA\ bubble}}$ is the energetic cost of melting the DNA bubble, $\Delta G^{\circ}_{\mathrm{RNAP\ binding}}$ is the binding energy of RNAP to the nucleic acid framework (RNA + DNA) in the TEC. The length of the RNA:DNA hybrid was fixed at 9 bp in all TECs, with the 3′-most paired RNA nucleotide occupying the $i+1$ site. Note that our definition of the i and $i+1$ sites differs from the definition used by Yager & von Hippel (see Figure 1).[75] Thus, for the active TEC, the 3′-terminal RNA nucleotide is in the $i+1$ site; for the −1 backtracked TEC, the penultimate RNA nucleotide is in the $i+1$ site (the presumed conformation of the HIV-1 paused TEC); for the −2 backtracked TEC, the −3 nucleotide of RNA is in $i+1$ site, and so on. $\Delta G^{\circ}_{\mathrm{DNA\ bubble}}$ was calculated assuming that the downstream-most melted base-pair was 1 nt past the template nucleotide in the $i+1$ site, based on the TEC model described by Korzheva *et al.*;[50] this is 1 bp further downstream than assumed by Yager & von Hippel.[49] The number of melted base-pairs in the bubble was fixed at 17.[76–78] We assigned the entropic cost of RNA:DNA and DNA:DNA base-pairing at 3.1 kcal/mol *versus* 3.0 kcal/mol as estimated by Yager & von Hippel,[49] based on measurements made by Sugimoto *et al.*[52] We set $\Delta G^{\circ}_{\mathrm{RNAP\ binding}}$ at −36.4 kcal/mol *versus* −32 kcal/mol estimated by von Hippel & Yager,[78] an increase of −4.4 kcal/mol, because we assumed that interactions of RNAP with nascent RNA in the exit channel should compensate energetically for the loss of 3 bp of RNA:DNA relative to the 12 bp RNA:DNA hybrid assumed by Yager & von Hippel.[49] These modifications resulted in an average value of $\Delta G^{\circ}_{\mathrm{TEC}}$ of −17 kcal/mol for the TECs we studied, *versus* −18 kcal/mol estimated by von Hippel & Yager[78] for the average $\Delta G^{\circ}_{\mathrm{TEC}}$ of a different population of TECs.

Transcript cleavage reactions

TECs halted at position A60 were elongated in the presence of CTP and UTP for one minute at 30 °C to form paused complexes at U62. The TECs were then washed rapidly with transcription buffer and suspended in the same buffer (80 μl). TFIIS (1.25 nM final) was added to the reaction mix, the volume was adjusted to 100 μl with transcription buffer and incubated at 30 °C. Aliquots were removed at indicated times, terminated, processed and separated on 8% polyacrylamide/urea gel as described above.

Estimation of k_i^{obs} and the activation energy of pausing

The observed rate of paused complex formation, k_i^{obs}, was estimated by non-linear regression of E^{max} as a function of time:

$$E^{max} = 1 - e^{(-kt)} \quad (2)$$

where $k = k_i^{obs}$, and t = time.

Because E^{max} was less than 1 even after prolonged delay to allow equilibration, we inferred that formation of the paused complex must be reversible. Although we cannot relate this to a fundamental mechanism, we can extract apparent forward and reverse isomerization constants, k_{i+}^{obs} and k_{i-}^{obs} from these data using equations (3) and (4):

$$k_{i+}^{obs} = k_i^{obs} E^{max} \quad (3)$$

$$k_{i-}^{obs} = k_i^{obs}(1 - E^{max}) \quad (4)$$

The activation free energy barrier for these steps is related to the rate constants by equation (5):

$$\Delta G_i^{\ddagger} = -RT \ln(k_i(h/\kappa kT)) \quad (5)$$

where R is the gas constant, T is temperature (Kelvin), k_i is the rate constant, h is Plank's constant, κ is the transmission coefficient for isomerization, and k is Boltzmann's constant.

Although the transmission coefficient for isomerization cannot be calculated, we assume it is near 1 for this simple conformational change.[79] Regardless, it should be the same for forward and reverse isomerization and for paused TEC formation on different templates, making comparisons of the $\Delta G^{\ddagger}$ appropriate.

Acknowledgments

We thank members of the Landick laboratory for helpful discussions and comments on the manuscript, and S. Darst, D. Erie, and R. Kornberg for helpful discussions and sharing results prior to publication. This work was supported by the UW-Madison Graduate School and a grant from the NIH (GM 38660).

References

1. Bentley, D. L. (1995). Regulation of transcriptional elongation by RNA polymerase II. *Curr. Opin. Genet. Dev.* **5**, 210-216.
2. Conaway, J. W. & Conaway, R. C. (1999). Transcription elongation and human disease. *Annu. Rev. Biochem.* **68**, 301-319.
3. Garber, M. & Jones, K. (1999). HIV-1 Tat: coping with negative elongation factors. *Curr. Opin. Immunol.* **11**, 460-465.
4. Izban, M. & Luse, D. S. (1991). Transcription on nucleosomal templates by RNA polymerase II in vitro: inhibition of elongation with enhancement of sequence-specific pausing. *Genes Dev.* **5**, 683-696.
5. Orphanides, G., LeRoy, G., Chang, C. H., Luse, D. S. & Reinberg, D. (1998). FACT, a factor that facilitates transcript elongation through nucleosomes. *Cell,* **92**, 105-116.
6. Wada, T., Takagi, T., Yamaguchi, Y., Ferdous, A., Imai, T. & Hirose, S. *et al.* (1998). DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes Dev.* **12**, 343-356.
7. Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A. & Sugimoto, S. *et al.* (1999). NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell,* **97**, 41-51.
8. Hawley, D. K., Wiest, D. K., Holtz, M. S. & Wang, D. (1993). Transcriptional pausing, arrest, and read-through at the adenovirus major late attenuation site. *Cell Mol. Biol. Res.* **39**, 339-348.

9. Uptain, S., Kane, C. & Chamberlin, M. (1997). Basic mechanisms of transcript elongation and its regulation. *Annu. Rev. Biochem.* **66**, 117-172.
10. Palangat, M., Meier, T., Keene, R. & Landick, R. (1998). Transcriptional pausing at +62 of the HIV-1 nascent RNA modulates formation of the TAR RNA structure. *Mol. Cell*, **1**, 1033-1042.
11. Artsimovitch, I. & Landick, R. (2000). Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc. Natl Acad. Sci. USA,* **97**, 7090-7095.
12. Erie, D. A., Hajiseyedjavadi, O., Young, M. C. & von Hippel, P. H. (1993). Multiple RNA polymerase conformations and GreA: control of fidelity of transcription. *Science,* **262**, 867-873.
13. Matsuzaki, H., Kassavetis, G. A. & Geiduschek, E. P. (1994). Analysis of RNA chain elongation and termination by *Saccharomyces cerevisiae* RNA polymerase III. *J. Mol. Biol.* **235**, 1173-1192.
14. de Mercoyrol, L., Soulie, J. M., Job, C., Job, D., Dussert, C. & Palmari, J. *et al*. (1990). Abortive intermediates in transcription by wheat-germ RNA polymerase II. Dynamic aspects of enzyme/template interactions in selection of the enzyme synthetic mode. *Biochem. J.* **269**, 651-658.
15. Lyakhov, D. L., He, B., Zhang, X., Studier, F. W., Dunn, J. J. & McAllister, W. T. (1998). Pausing and termination by bacteriophage T7 RNA polymerase. *J. Mol. Biol.* **280**, 201-213.
16. Davenport, R. J., Wuite, G. J., Landick, R. & Bustamante, C. (2000). Single-molecule study of transcriptional pausing and arrest by *E. coli* RNA polymerase. *Science,* **287**, 2497-2500.
17. Mooney, R. A. & Landick, R. (1999). RNA polymerase unveiled. *Cell,* **98**, 687-690.
18. Cramer, P., Bushnell, D. & Kornberg, R. (2001). Structural basis of transcription: RNA polymerase II at 2.8 Å resolution. *Science,* **292**, 1863-1876.
19. Gnatt, A., Cramer, P., Fu, J., Bushnell, D. & Kornberg, R. D. (2001). Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science,* **292**, 1876-1882.
20. Landick, R. (2001). RNA polymerase clamps down. *Cell,* **105**, 567-570.
21. Gu, W. & Reines, D. (1995). Identification of a decay in transcription potential that results in elongation factor dependence of RNA polymerase II. *J. Biol. Chem.* **270**, 11238-11244.
22. Reeder, T. & Hawley, D. (1996). Promoter proximal sequences modulate RNA polymerase II elongation by a novel mechanism. *Cell,* **87**, 767-777.
23. Komissarova, N. & Kashlev, M. (1997). Transcriptional arrest: *Escherichia coli* RNA polymerase translocates backward, leaving the 3′ end of the RNA intact and extruded. *Proc. Natl Acad. Sci. USA,* **94**, 1755-1760.
24. Komissarova, N. & Kashlev, M. (1997). RNA polymerase switches between inactivated and activated states by translocating back and forth along the DNA and the RNA. *J. Biol. Chem.* **272**, 15329-15338.
25. Nudler, E., Mustaev, A., Lukhtanov, E. & Goldfarb, A. (1997). The RNA:DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase. *Cell,* **89**, 33-41.
26. Kireeva, M. L., Komissarova, N., Waugh, D. S. & Kashlev, M. (2000). The 8-nucleotide-long RNA:DNA hybrid is a primary stability determinant of the RNA polymerase II elongation complex. *J. Biol. Chem.* **275**, 6530-6536.
27. Krummel, B. & Chamberlin, M. J. (1992). Structural analysis of ternary complexes of *Escherichia coli* RNA polymerase. Deoxyribonuclease I footprinting of defined complexes. *J. Mol. Biol.* **225**, 239-250.
28. Wiest, D. K., Wang, D. & Hawley, D. K. (1992). Mechanistic studies of transcription arrest at the adenovirus major late attenuation site. Comparison of purified RNA polymerase II and washed elongation complexes. *J. Biol. Chem.* **267**, 7733-7744.
29. Gu, W., Powell, W., Mote, J., Jr & Reines, D. (1993). Nascent RNA cleavage by arrested RNA polymerase II does not require upstream translocation of the elongation complex on DNA. *J. Biol. Chem.* **268**, 25604-25616.
30. Izban, M. & Luse, D. (1993). The increment of SII facilitated transcript cleavage varies dramatically between competent and incompetent RNA polymerase II ternary complexes. *J. Biol. Chem.* **268**, 12874-12885.
31. Rudd, M. D., Izban, M. G. & Luse, D. S. (1994). The active site of RNA polymerase II participates in transcript cleavage within arrested ternary complexes. *Proc. Natl Acad. Sci. USA,* **91**, 8057-8061.
32. Nudler, E., Goldfarb, A. & Kashlev, M. (1994). Discontinuous mechanism of transcription elongation. *Science,* **265**, 793-796.
33. Izban, M. & Luse, D. (1992). The RNA polymerase II ternary complex cleaves the nascent transcript in a 3′ → 5′ direction in the presence of elongation factor SII. *Genes Dev.* **6**, 1342-1356.
34. Borukhov, S., Sagitov, V. & Goldfarb, A. (1993). Transcript cleavage factors from *E. coli*. *Cell,* **72**, 459-466.
35. von Hippel, P. H. (1998). An integrated model of the transcription complex in elongation, termination, and editing. *Science,* **281**, 660-665.
36. Guajardo, R. & Sousa, R. (1997). A model for the mechanism of polymerase translocation. *J. Mol. Biol.* **265**, 8-19.
37. Landick, R. (1997). RNA polymerase slides home: pause and termination site recognition. *Cell,* **88**, 741-744.
38. Guajardo, R., Lopez, P., Dreyfus, M. & Sousa, R. (1998). NTP concentration effects on initial transcription by T7 RNAP indicate that translocation occurs through passive sliding and reveal that divergent promoters have distinct NTP concentration requirements for productive initiation. *J. Mol. Biol.* **281**, 777-792.
39. Marciniak, R. A., Calnan, B. J., Frankel, A. D. & Sharp, P. A. (1990). HIV-1 Tat protein trans-activates transcription *in vitro*. *Cell,* **63**, 791-802.
40. Dingwall, C., Ernberg, I., Gait, M. J., Green, S. M., Heaphy, S. & Karn, J. *et al*. (1990). HIV-1 tat protein stimulates transcription by binding to a U-rich bulge in the stem of the TAR RNA structure. *EMBO J.* **9**, 4145-4153.
41. Zhu, Y., Pe'ery, T., Peng, J., Ramanathan, Y., Marshall, N. & Marshall, T. *et al*. (1997). Transcription elongation factor P-TEFb is required for HIV-1 tat transactivation *in vitro*. *Genes Dev.* **11**, 2622-2632.
42. Mancebo, H. S., Lee, G., Flygare, J., Tomassini, J., Luu, P. & Zhu, Y. *et al*. (1997). P-TEFb kinase is required for HIV Tat transcriptional activation *in vivo* and *in vitro*. *Genes Dev.* **11**, 2633-2644.
43. Karn, J. (1999). Tackling Tat. *J. Mol. Biol.* **293**, 235-254.
44. Artsimovitch, I. & Landick, R. (1998). Interaction of a nascent RNA structure with RNA polymerase is

required for hairpin-dependent transcriptional pausing but not for transcript release. *Genes Dev.* **12**, 3110-3122.
45. Landick, R., Wang, D. & Chan, C. (1996). Quantitative analysis of transcriptional pausing by RNA polymerase: the *his* leader pause site as a paradigm. *Methods Enzymol.* **274**, 334-352.
46. Theissen, G., Pardon, B. & Wagner, R. (1990). A quantitative assessment for transcriptional pausing of DNA-dependent RNA polymerases *in vitro*. *Anal. Biochem.* **189**, 254-261.
47. Komissarova, N. & Kashlev, M. (1998). Functional topography of nascent RNA in elongation intermediates of RNA polymerase. *Proc. Natl Acad. Sci. USA,* **95**, 14699-14704.
48. Moore, P. (1999). The RNA folding problem. In *The RNA World* (Gesteland, R., Cech, T. & Atkins, J., eds), 2nd edit., pp. 381-401, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
49. Yager, T. D. & von Hippel, P. H. (1991). A thermodynamic analysis of RNA transcript elongation and termination in *Escherichia coli*. *Biochemistry,* **30**, 1097-1118.
50. Korzheva, N., Mustaev, A., Kozlov, M., Malhotra, A., Nikiforov, V., Goldfarb, A. & Darst, S. A. (2000). A structural model of transcription elongation. *Science,* **289**, 619-625.
51. Sidorenkov, I., Komissarova, N. & Kashlev, M. (1998). Crucial role of the RNA:DNA hybrid in the processivity of transcription. *Mol. Cell,* **1**, 55-64.
52. Sugimoto, N., Nakano, S., Katoh, M., Matsumura, A., Nakamuta, H. & Ohmichi, T. *et al*. (1995). Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes. *Biochemistry,* **34**, 11211-11216.
53. Kassavetis, G. A. & Chamberlin, M. J. (1981). Pausing and termination of transcription within the early region of bacteriophage T7 DNA *in vitro*. *J. Biol. Chem.* **256**, 2777-2786.
54. Porschke, D. (1977). Elementary steps of base recognition and helix-coil transitions in nucleic acids. In *Molecular Biology Biochemistry and Biophysics* (Pecht, I. & Rigler, R., eds), vol. 24, pp. 191-218, Springer-Verlag, New York.
55. Ramaiah, A. & Tejwani, G. A. (1973). Interconvertible forms of phosphofructokinase of rabbit liver. The role of effectors on the interconversion. *Eur. J. Biochem.* **39**, 183-192.
56. Yin, H., Wang, M., Svoboda, K., Landick, R., Block, S. & Gelles, J. (1995). Transcription against an applied force. *Science,* **270**, 1653-1657.
57. Wang, M., Schnitzer, M., Yin, H., Landick, R., Gelles, J. & Block, S. (1998). Force and velocity measured for single molecules of RNA polymerase. *Science,* **282**, 902-907.
58. Zhang, G., Campbell, E. A., Minakhin, L., Richter, C., Severinov, K. & Darst, S. A. (1999). Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell,* **98**, 811-824.
59. Zaychikov, E., Denissova, L. & Heumann, H. (1995). Translocation of the Escherichia coli transcription complex observed in the registers 11 to 20: ''jumping'' of RNA polymerase and asymmetric expansion and contraction of the transcription bubble. *Proc. Natl Acad. Sci. USA,* **92**, 1739-1743.
60. Nechaev, S., Chlenov, M. & Severinov, K. (2000). Dissection of two hallmarks of the open promoter complex by mutation in an RNA polymerase core subunit. *J. Biol. Chem.* **275**, 25516-25522.
61. Finn, R. D., Orlova, E. V., Gowen, B., Buck, M. & van Heel, M. (2000). *Escherichia coli* RNA polymerase core and holoenzyme structures. *EMBO J.* **19**, 6833-6844.
62. Markovtsov, V., Mustaev, A. & Goldfarb, A. (1996). Protein-RNA interactions in the active center of the transcription elongation complex. *Proc. Natl Acad. Sci. USA,* **93**, 3221-3226.
63. Gilbert, W. J. (1976). Starting and stopping sequences of the RNA polymerase. In *RNA Polymerase* (Losick, R. & Chamberlin, M. J., eds), pp. 193-205, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
64. Landick, R. (1999). Switching RNA polymerase into overdrive. *Science,* **284**, 598-599.
65. Beese, L. S., Friedman, J. M. & Steitz, T. A. (1993). Crystal structures of the Klenow fragment of DNA polymerase I complexed with deoxynucleoside triphosphate and pyrophosphate. *Biochemistry,* **32**, 14095-14101.
66. Doublié, S. & Ellenberger, T. (1998). The mechanism of action of T7 DNA polymerase. *Curr. Opin. Struct. Biol.* **8**, 704-712.
67. Steitz, T. A. (1998). A mechanism for all polymerases. *Nature,* **391**, 231-232.
68. Koshland, D., Jr (1958). Application of a theory of enzyme specificity to protein synthesis. *Proc. Natl Acad. Sci. USA,* **44**, 98-104.
69. Frankel, A. D. & Smith, C. A. (1998). Induced folding in RNA-protein recognition: more than a simple molecular handshake. *Cell,* **92**, 149-151.
70. Spolar, R. S. & Record, M. T., Jr (1994). Coupling of local folding to site-specific binding of proteins to DNA. *Science,* **263**, 777-784.
71. Williamson, J. (2000). Induced fit in RNA-protein recognition. *Nature Struct. Biol.* **7**, 834-837.
72. Shilatifard, A. (1998). Factors regulating the transcriptional elongation activity of RNA polymerase II. *FASEB J.* **12**, 1437-1446.
73. Shapiro, D., Sharp, P., Wahli, W. & Keller, M. (1988). A high-efficiency HeLa cell nuclear transcription extract. *DNA,* **7**, 47-55.
74. Yoo, O., Yoon, H., Baek, K., Jeon, C., Miyamoto, K., Ueno, A. & Agarwal, K. (1991). Cloning, expression and characterization of the human transcription elongation factor, TFIIS. *Nucl. Acids Res.* **19**, 1073-1079.
75. Mooney, R., Artsimovitch, I. & Landick, R. (1998). Information processing by RNA polymerase: recognition of regulatory signals during RNA chain elongation. *J. Bacteriol.* **180**, 3265-3275.
76. Gamper, H. B. & Hearst, J. E. (1982). A topological model for transcription based on unwinding angle analysis of *E. coli* RNA polymerase binary, initiation and ternary complexes. *Cell,* **29**, 81-90.
77. Gamper, H. & Hearst, J. (1983). Size of the unwound region of DNA in *Escherichia coli* RNA polymerase and calf thymus RNA polymerase II ternary complexes. *Cold Spring Harbor Symp. Quant. Biol.* **47**, 455-461.
78. von Hippel, P. H. & Yager, T. D. (1991). Transcript elongation and termination are competitive kinetic processes. *Proc. Natl Acad. Sci. USA,* **88**, 2307-2311.
79. Fersht, A. (1998). *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*, W.H. Freeman and Company, New York.
80. Fu, J., Gnatt, A. L., Bushnell, D. A., Jensen, G. J., Thompson, N. E. & Burgess, R. R. *et al*. (1999). Yeast

RNA polymerase II at 5 Å resolution. *Cell,* **98**, 799-810.
81. Cramer, P., Bushnell, D., Fu, J., Gnatt, A., Maier-Davis, B. & Thompson, N. *et al*. (2000). Architecture of RNA polymerase II and implications for the transcription mechanism. *Science,* **288**, 640-649.
82. Lee, D. N. & Landick, R. (1992). Structure of RNA and DNA chains in paused transcription complexes containing *Escherichia coli* RNA polymerase. *J. Mol. Biol.* **228**, 759-777.
83. Gu, W., Wind, M. & Reines, D. (1996). Increased accommodation of nascent RNA in a product site on RNA polymerase II during arrest. *Proc. Natl Acad. Sci. USA,* **93**, 6935-6940.
84. Li, Y., Korolev, S. & Waksman, G. (1998). Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of *Thermus aquaticus* DNA polymerase I: structural basis for nucleotide incorporation. *EMBO J.* **17**, 7514-7525.

Edited by K. Yamamoto

(Received 17 January 2001; received in revised form 24 May 2001; accepted 7 June 2001)