

Pulling on the Nascent RNA during Transcription Does Not Alter Kinetics of Elongation or Ubiquitous Pausing

Ravindra V. Dalal, Matthew H. Larson, Keir C. Neuman, Jeff Gelles, Robert Landick, and Steven M. Block

Supplemental Results and Discussion

U-Rich Tracts in the DNA Template

Intrinsic terminators contain a uridine-rich tract which is thought to be necessary for RNA release from the TEC (Martin and Tinoco, 1980; Platt, 1986; von Hippel, 1998; Yarnell and Roberts, 1999). As shown in Table S1, the template used in our experiments carried a single region containing 4 uridines (U's) in a row (and also 6 U's situated within a 15-nt window). Five other regions within the transcript contained 3 U's in a row (and also 6-7 U's within a 15-nt window). These sequences contain fewer U's than are found in the U-rich regions of most *E. coli* intrinsic termination sequences (de Hoon et al., 2005). Within our absolute positional uncertainty of ~300 nt, we found no strong peaks in the rupture probability near any of these regions, suggesting that these sequences were not sufficient to cause reliable RNA release, even when coupled with high force.

Table S1: U-Rich Sequences Encountered

Sequential U's	# U's in 8-nt window	# U's in 15-nt window	Sequence
4	5	6	AUCUUUUG
3	6	7	UUUGUUAU
3	6	7	UCUUCUUU
3	6	6	UGUUCUUU
3	6	7	UUAUCUUU

3	5	7	UGUAAUUU
---	---	---	----------

Modeling of Force Extension Curves (FECs)

To model the FEC of the stalled transcription complex shown in Figure 2B, we first treated this system as an elastic dsDNA ‘handle’ of fixed length (modeled as a worm-like chain, WLC) in series with a section of ssRNA (modeled as a freely-jointed chain, FJC), neglecting for the moment any contribution from RNA secondary structure (DNA and RNA stretching parameters are described in Experimental Procedures). The resulting curve agreed well with our data at both high and low forces, but failed to capture the broad plateau found at intermediate force. To model this plateau, we represented RNA secondary structure as a population of $N = 65$ hairpins with helix width of 2 nm, loop length of 4 nt, and random stem lengths, distributed as a Gaussian centered at 10.5 bp with a std.dev. of 3 bp. The opening force for each random hairpin was taken to be proportional to its stem length (1 pN/bp opening force), with a small amount of variability incorporated as a Gaussian noise of width 1 pN. For forces below the opening force of the most stable hairpin, this model allows for the formation of RNA structure, reducing the overall extension of the RNA. This model reproduced the broad force plateau seen for intermediate forces in the FECs, consistent with its assignment to RNA secondary structure that is eliminated at high forces. Above ~ 18 pN, the model is free of residual hairpins and reduces to the simple WLC + FJC limit.

Reanalysis of Pause Distribution from Neuman et al. (2003)

In a previous study, Neuman et al. (2003) reported the pause distribution in terms of the pause density (i.e., the spatial pause probability, measured in units of pauses per bp transcribed) and the pause strength, defined as the product of the pause density multiplied by the pause duration. They found that these quantities were largely independent of force, and concluded that backtracking of the polymerase was not likely to be responsible for ubiquitous pauses. We show below that the pause density and pause strength are functions of the forward elongation rate, which has been found to be force-dependent (Abbondanzieri et al., 2005). We derive here a new measure for the pause strength (defined as the product of the temporal pause probability, or pause frequency, measured in pauses per unit time, and the pause duration) and show that this metric is independent of the forward elongation rate under most conditions. Fitting the Neuman et al. (2003) data to this new expression confirmed that pausing is largely independent of backtracking.

Normal elongation is frequently interrupted by pausing. Based on the simple elongation pathway shown in Figure S1, we model forward elongation as a single step occurring at rate k_n , which may be force-dependent. Pausing is modeled as an off-pathway state, with entry into the pause occurring at rate k_p and exit at rate k_{-p} .

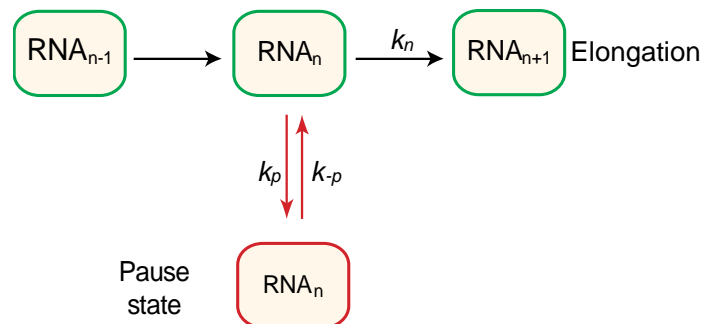


Figure S1. Elongation Reaction with an Off-Pathway Pause State

From this model, we can calculate several pause characteristics. Pause density (pauses per bp transcribed) is given by the ratio of rate of entry into the pause state to the forward elongation rate: *Pause density* = k_p / k_n . The density-dependent pause strength is equal to the pause density times the mean pause duration ($1/k_{-p}$):

$$\text{Density-dependent pause strength} = \frac{k_p}{k_{-p} k_n} \quad (1)$$

The pause frequency (pauses per second) results from a kinetic competition between entry into the pause state and further elongation. It is equal to the product of the efficiency for entering a pause [efficiency = $k_p / (k_p + k_n)$] and the rate that a decision to pause is encountered, which is the same as the forward elongation rate, *Pause frequency* = $k_p k_n / (k_p + k_n)$. We define the frequency-dependent pause strength to be the product of the pause frequency and the pause duration:

$$\text{Frequency-dependent pause strength} = \left(\frac{k_p}{k_p + k_n} \right) \frac{k_n}{k_{-p}} \quad (2)$$

In the limit $k_n \gg k_p$, the pause frequency reduces to simply k_p , and the frequency-dependent pause strength reduces to:

$$\text{Frequency-dependent pause strength} \approx k_p / k_{-p}. \quad (3)$$

The average rate of elongation measured in single molecule experiments is ~ 10 bp/sec, while the rate of entry into the pause state is $\sim 0.1 \text{ sec}^{-1}$ (Neuman et al., 2003); under these conditions the approximation in Eq. (3) holds. We note that Eq. (3) is independent of the forward elongation rate, whereas Eq. (1), used in (Neuman et al., 2003), is not.

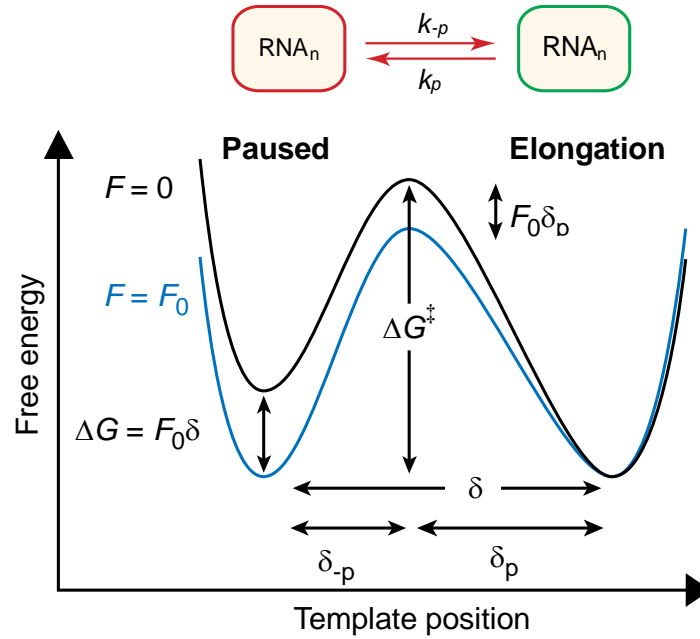


Figure S2. Energy Landscape for a Backtracking Model

To enter into the paused state (red) from the normal elongation state (green), the enzyme backtracks through an overall distance δ . In the absence of force, the free energy is shown by the black curve. When a force, F , is applied, the free energy landscape is tilted (blue).

We now introduce a two state model for pausing induced by enzyme backtracking (Figure S2). The backtracking model assumes that a transient pause occurs when RNAP slides backward through one or more bases along the DNA over a distance, δ . The resulting movement of the 3'-end of the nascent RNA relative to the polymerase active site prevents further transcription until this motion is reversed. The corresponding free energy diagram is a single energy barrier between two potential wells separated by δ . The application of force $F = F_0$ alters the relative free energy of the states and the respective transition rates by a factor $\exp(F_0 \cdot d/k_B T)$, where d is the distance to the transition state (which is necessarily less than the full distance of the translocation step). The rate of entry into the pause state equals $k_p = k_{p,0} \exp(-F_0 \cdot \delta_p/k_B T)$, while the rate of exit from the pause state equals $k_{-p} = k_{-p,0} \exp(+F_0 \cdot \delta_{-p}/k_B T)$. Note that the two exponential factors have opposite signs, corresponding to opposite directions of motion. Inserting these transition rates into Eq. (3), we obtain the following expression:

$$\text{Frequency-dependent pause strength} = \frac{k_{p,o}}{k_{-p,o}} \exp\left(-\frac{F_0 \delta}{k_B T}\right) \quad (4)$$

Fitting the frequency-dependent pause strength data (Figure S3) to Eq. (4) yields a backtracking step of 0.15 ± 0.05 bp (blue curve, $\chi_v^2 = 0.1$; $\nu = 9$; $p(\chi_v^2) = 0.99$; 2 parameters).. This finding ($\delta \ll 1$ bp) indicates that ubiquitous pausing is largely independent of force, and thus confirms the previous conclusion that ubiquitous pausing is largely independent of backtracking.

To explain the small nonzero value of the step size, we hypothesize that a minority subset of the pauses may still be caused by backtracking. For example, rare, longer duration pauses ($t > 20$ s; average frequency $\sim 10^{-3}$ bp $^{-1}$) have been shown to be caused by misincorporation of a nucleotide, causing the enzyme to backtrack (Shaevitz et al., 2003). Some of these misincorporation-induced pauses may be mixed with the population of shorter duration, backtracking-independent ubiquitous pauses. To model this possibility, we fit the frequency-dependent pause strength to the sum of a constant term plus an exponential with fixed distance parameter of 1 bp. The fit (green curve, $\chi_v^2 = 0.1$; $\nu = 9$; $p(\chi_v^2) = 0.99$; 2 parameters) yields a constant offset of 0.25 ± 0.04 and an amplitude for the exponential of 0.017 ± 0.005 . At the lowest force point (-26 pN, hindering load geometry), this indicates that $\sim 37\%$ of the pauses detected may be due to backtracking; the fraction of backtracking-related pauses decreases exponentially with force. The χ_v^2 value obtained for this fit was not significantly better than the value obtained from a single exponential without a constant offset. Nevertheless, we prefer the model with an offset because it provides a straightforward explanation for the weak force dependence seen in the pause data.

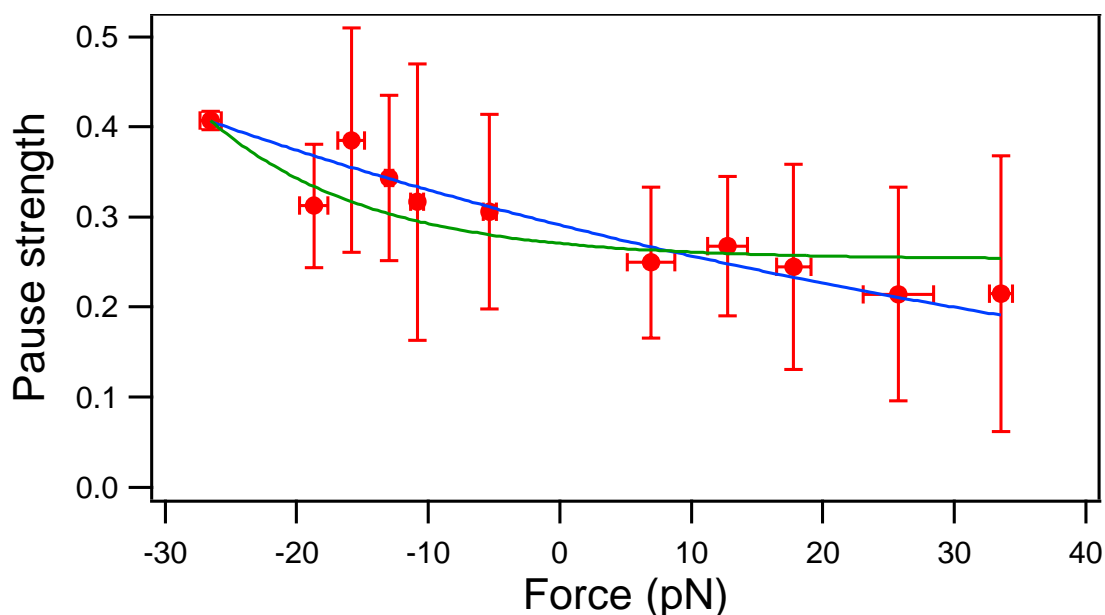


Figure S3. Frequency-Dependent Pause Strength (Pause Frequency Multiplied by Duration) versus Force Calculated from the DNA-Pulling Data of Neuman et al. (2003), Plotted as Mean \pm Standard Deviation

Pauses between 1-25 s in duration were included in the analysis. The sign convention for the force is the reverse of Neuman et al. The fit curve in blue corresponds to a 0.15 ± 0.05 bp backtracking step. The fit curve in green is the sum of a constant background of ubiquitous pausing (amplitude 0.250 ± 0.040) and an exponential term due to backtracking by ~ 1 bp with (amplitude 0.017 ± 0.005).

Opening Force of the *his* Hairpin

To calculate the expected opening force for the *his* terminator hairpin (Yin et al., 1999), we implemented an approach developed for a recent study of DNA hairpins measured under load (Woodside et al., 2006). For the given hairpin, we first calculated the energy landscape using MFOLD version 2.3 at $T = 21^\circ\text{C}$ (Zuker, 2003), which includes nearest-neighbor stacking and base pairing interactions. We then incorporated the entropic effects of extending the single stranded bases in the loop section of the hairpin using the RNA stretching parameters supplied in Experimental Procedures (main text). The application of force was modeled by tilting the energy landscape (Liphardt et al., 2001); this tilt was increased until the density of the open and closed

states equalized. This procedure yielded the predicted force required to open the RNA hairpin at equilibrium.

Supplemental Experimental Procedures

Experimental Assays

TECs stalled at position A29 were produced as described (Neuman et al., 2003). Of the 29 bases of nascent RNA present in the stalled complex, we estimate that ~15 bases were accessible outside of the polymerase prior to addition of NTPs (Komissarova and Kashlev, 1997), sufficient to form a stable hybrid with the 25 nt overhang present on the DNA handle. The force required to break the bead tether increased somewhat after complexes were restarted, suggesting that the number of bases hybridized between the DNA handle and nascent RNA may have increased somewhat after complexes were restarted.

An oxygen-scavenging system was introduced into the transcription assay buffers to minimize photochemical damage (Neuman et al., 1999). Both transcription buffers and oxygen-scavenger components were tested for RNase activity using an Ambion RNaseAlert Lab Test Kit. RNase contamination was occasionally detected in glucose oxidase solutions, in which case the stocks were subsequently purified by FPLC (Amersham, model ÄKTA) on a Superdex 200 10/300 GL size-exclusion column (Amersham). Purified glucose oxidase was eluted with 2X transcription buffer and retested to ensure removal of contamination.

Bead-linked TECs attached to the coverglass surface of the flow cell were pre-screened prior to restarting transcription for the correct contour length ($L_c \sim 3$ kb) and persistence length

($L_p = 35\text{--}45$ nm for dsDNA) to identify candidates bearing a single DNA tether, as opposed to multiple tethers. After NTPs were introduced and transcription restarted, the apparent persistence length decreased due to the growing length of ssRNA. However, bead-attached TECs tethered to a surface by multiple DNA molecules also yielded a lower persistence length (typically $L_p < 25$ nm). To distinguish these possibilities, it was necessary to check the persistence lengths of tethers prior to NTP addition. An NTP mix (1 mM of each ribonucleotide triphosphate, Roche Molecular Biochemicals, along with the oxygen scavenger system in transcription buffer) was introduced into the flow cell to restart the stalled complexes and begin an experiment.

Tether Length Determination

Determination of RNA transcript length (in nt) was accomplished by subtracting the known extension (x_{handle}) of the dsDNA handle (contour length, $L_{\text{handle}} = 3,057$ bp) at the given force from the total tether extension (x_{tether}). The handle extension at force F and temperature $T = 21^\circ\text{C}$ was calculated from Odijk's approximation to the worm-like chain (WLC) (Wang et al., 1997):

$$x_{\text{handle}} = L_{\text{handle}} \left[1 - \frac{1}{2} \left(\frac{k_B T}{F L_{\text{ds}}} \right)^{1/2} + \frac{F}{K_{\text{ds}}} \right],$$

with a dsDNA persistence length $L_{\text{ds}} = 41$ nm and bulk stiffness modulus $K_{\text{ds}} = 1200$ pN; k_B is Boltzmann's constant. The remainder ($x_{\text{ssRNA}} = x_{\text{tether}} - x_{\text{handle}}$) was fit to a freely-jointed chain (FJC) model for ssRNA to obtain the contour length:

$$L_{\text{ssRNA}} = \frac{x_{\text{ssRNA}}}{\left[\coth \left(\frac{2FL_{\text{ss}}}{k_B T} \right) - \frac{k_B T}{2FL_{\text{ss}}} \right] \left(1 + \frac{F}{K_{\text{ss}}} \right)}.$$

RNA stretching parameters were estimated by comparisons with FEC's in the high and low force regimes (Figure 2B); secondary structure prevented fits over the full range. Estimated parameters were $L_{ss} = 1.3$ nm, $K_{ss} = 130$ pN (assuming $N_{ss} = 2,200$ bases and an interphosphate distance of 0.59 nm/base for ssRNA), which are similar to values determined previously for poly(rU) (Seol et al., 2004). Systematic errors attributable to uncertainties in stretching parameters may affect the computed lengths by $\sim 5\%$ and may help to explain small differences in velocity between RNA-pulling and DNA-pulling data.

Supplemental References

Abbondanzieri, E. A., Greenleaf, W. J., Shaevitz, J. W., Landick, R., and Block, S. M. (2005). Direct observation of base-pair stepping by RNA polymerase. *Nature* *438*, 460-465.

de Hoon, M. J., Makita, Y., Nakai, K., and Miyano, S. (2005). Prediction of Transcriptional Terminators in *Bacillus subtilis* and Related Species. *PLoS Comput Biol* *1*, e25.

Komissarova, N., and 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 U S A* *94*, 1755-1760.

Liphardt, J., Onoa, B., Smith, S. B., Tinoco, I. J., and Bustamante, C. (2001). Reversible unfolding of single RNA molecules by mechanical force. *Science* *292*, 733-737.

Martin, F. H., and Tinoco, I., Jr. (1980). DNA-RNA hybrid duplexes containing oligo(dA:rU) sequences are exceptionally unstable and may facilitate termination of transcription. *Nucleic Acids Res* *8*, 2295-2299.

Neuman, K. C., Abbondanzieri, E. A., Landick, R., Gelles, J., and Block, S. M. (2003). Ubiquitous transcriptional pausing is independent of RNA polymerase backtracking. *Cell* *115*, 437-447.

Neuman, K. C., Chadd, E. H., Liou, G. F., Bergman, K., and Block, S. M. (1999). Characterization of photodamage to *Escherichia coli* in optical traps. *Biophys J* *77*, 2856-2863.

Platt, T. (1986). Transcription termination and the regulation of gene expression. *Annu Rev Biochem* *55*, 339-372.

Seol, Y., Skinner, G. M., and Visscher, K. (2004). Elastic properties of a single-stranded charged homopolymeric ribonucleotide. *Phys Rev Lett* *93*, 118102.

Shaevitz, J. W., Abbondanzieri, E. A., Landick, R., and Block, S. M. (2003). Backtracking by single RNA polymerase molecules observed at near-base-pair resolution. *Nature* *426*, 684-687.

von Hippel, P. H. (1998). An integrated model of the transcription complex in elongation, termination, and editing. *Science* *281*, 660-665.

Wang, M. D., Yin, H., Landick, R., Gelles, J., and Block, S. M. (1997). Stretching DNA with optical tweezers. *Biophys J* *72*, 1335-1346.

Woodside, M. T., Behnke-Parks, W. M., Larizadeh, K., Travers, K., Herschlag, D., and Block, S. M. (2006). Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins. *Proc Natl Acad Sci U S A* *103*, 6190-6195.

Yarnell, W. S., and Roberts, J. W. (1999). Mechanism of intrinsic transcription termination and antitermination. *Science* *284*, 611-615.

Yin, H., Artsimovitch, I., Landick, R., and Gelles, J. (1999). Nonequilibrium mechanism of transcription termination from observations of single RNA polymerase molecules. *Proc Natl Acad Sci U S A* *96*, 13124-13129.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* *31*, 3406-3415.