Supporting Material for "Development of a 'modular' scheme to describe the kinetics of transcription by RNA polymerase".

Simple Transcription Module

Simple transcription over three nucleotide positions on an idealized template is described in Fig. S1A, eqs. 1-3 and the associated text of the paper and was performed using the following Berkeley Madonna Script with n=2.

```
d/dt(S[0]) = - kf*S[0] + kr*S[1]
d/dt(S[1..n-1]) = kf*S[i-1] + kr*S[i+1] - kf*S[i] - kr*S[i]
d/dt(S[n]) = kf*S[n-1] - kr*S[n]

INIT S[0] = 1
INIT S[1..n] = 0

kf =20
kr = 1e-4

n=2

m= S[0]
t = S[n]
```

S[0] represented TECi in Fig. 1A and n = the number of nucleotide addition events from this position. m and t correspond to the outputs of a TEC located at a defined template position as a function of time. Simulations from this module are shown in Figure 1B. This model for transcription over 21 positions using an idealized template (Fig. S1) is described by eqs. 1, 4 and 5 and the associated text in the paper.

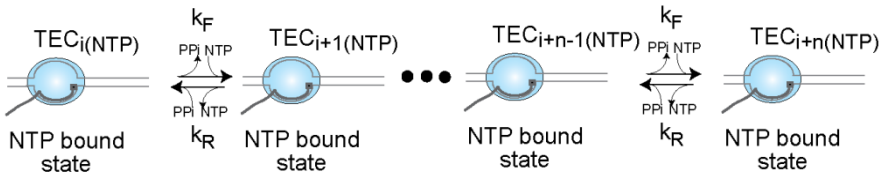

Fig. S1. ***Model for transcript elongation by a TEC over a number of template positions, starting with the NTP-bound state at position (i) and ending at position (i+n).*** The color-coding in the schematic is as follows: RNA polymerase (blue oval), template DNA (fine lines), nascent RNA (dark line), bound NTP (grey square) and the black circles represent repeat reactions between template positions (i+1) and (i+n-1). The events that occur during the forward reaction (NTP catalysis, pyrophosphate release, TEC translocation and binding of the next templated NTP) or the reverse reaction (pyrophosphoroylsis) are lumped together single rate constants for the forward (k_F) and reverse (k_R) reactions between each position, and are represented by the reaction arrows.

The Berkeley Madonna script above was adjusted to n=20 and includes additional outputs g, h, and j, representing TECs at the noted positions as a function of time.

```
n=20

m= S[0]
g= S[5]
h=S[10]
j=S[15]
t = S[n]
```

Simulations obtained using this adjusted elongation model are shown in Fig. 2 of the paper.

Pause Module

A simple pause was introduced into the basic transcription elongation model described above, as shown in Fig. S2 below:

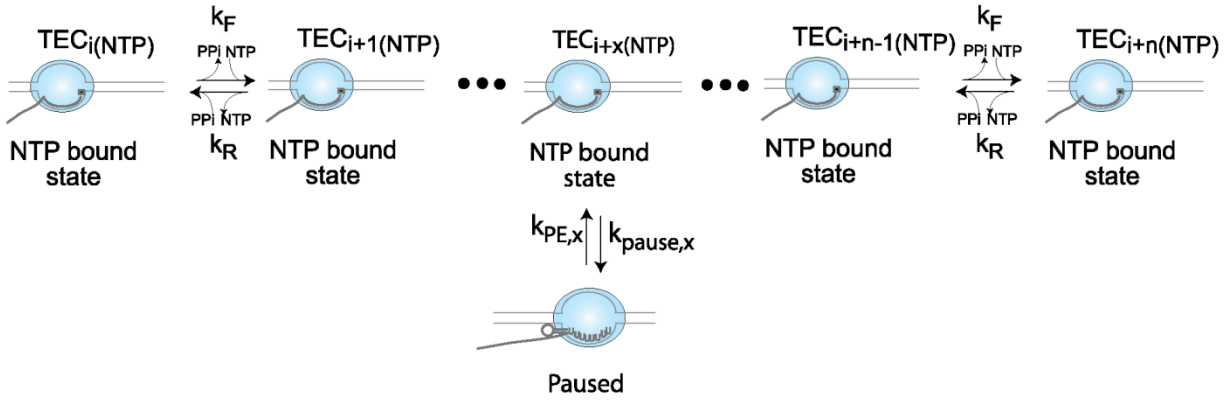

Fig. S2. ***Model for transcription by a TEC over many template positions, as described for Fig. S1 above, with the addition of an off-pathway alternative paused state that can be accessed at template position (i+x).*** The rate constants for entry into, and exit from, this state are $k_{pause,x}$ and $k_{PE,x}$, respectively. The possibility of accessing either a Class I or a Class II pause are represented by the hairpin in the nascent RNA (Class I) or a weak RNA-DNA hybrid (Class II).

This model is described by the coupled differential eqs. 1 and 5-7 in the text and was simulated using the following Berkeley Madonna script.

```
d/dt(S[0]) = - kf*S[0] + kr*S[1]
d/dt(S[1..g-1]) = kf*S[i-1] + kr*S[i+1] - kf*S[i] - kr*S[i]
d/dt(S[g]) = kf*S[g-1] - kr*S[g] - kf*S[g] + kr*S[g+1] - k1f*S[g] + k1r*P
d/dt(P)= k1f*S[g] - k1r*P
d/dt(S[g+1..n-1]) = kf*S[i-1] + kr*S[i+1] - kf*S[i] - kr*S[i]
d/dt(S[n]) = kf*S[n-1] - kr*S[n]

INIT S[0] = 1
INIT S[1..n] = 0
INIT P = 0

kf =20
kr = 1e-4
k1f = 10
k1r = 0.1
```

```
n=20
g=10

m= S[0]
t = S[n]
h= S[g]
j = S[g] + P
```

S[0] represents TECi in Figs. 1 and Figs. S1 and S2, while n = the number of nt addition events from this position, n = 20. G, represented by x in Fig. S2 and also in eqs. 6 and 7, represents the number of nucleotide addition events between the first position (i) and the final position (n). P represents the TECs in the paused state at template position i+x in Fig. S2 and Fig. 4A, as well as in eqs. 6-8, with entry and exit rate constants k_{1f} and k_{1r}, $k_{pause,x}$ and $k_{PE,x}$. j describes the sum of total fraction of TECs in both paused and elongation competent states at template position x (g in the Berkeley Madonna model). Simulations from this model are displayed in Fig. 4B of the paper.

Termination Module

The simple pause model was adjusted to include a terminated state at position (TECi+x), as depicted in Fig 3, in addition to the previous paused and elongation states.

This model is described by coupled differential eqs. 1, 5-6 and 8-9 in the paper and was simulated using the following Berkeley Madonna script.

```
d/dt(S[0]) = - kf*S[0] + kr*S[1]
d/dt(S[1..g-1]) = kf*S[i-1] + kr*S[i+1] - kf*S[i] - kr*S[i]
d/dt(S[g]) = kf*S[g-1] - kr*S[g] - kf*S[g] + kr*S[g+1] - k1f*S[g] + k1r*P
d/dt(P)= k1f*S[g] - k1r*P – k2f * P
d/dt(T) = k2f * P
d/dt(S[g+1..n-1]) = kf*S[i-1] + kr*S[i+1] - kf*S[i] - kr*S[i]
d/dt(S[n]) = kf*S[n-1] - kr*S[n]

INIT S[0] = 1
INIT S[1..n] = 0
INIT P = 0
INIT T = 0

kf =20
kr = 1e-4
k1f = 10
k1r = 0.1
k2f = 1

n=20
g=10

m= S[0]
t = S[n]
h= S[g]
j = S[g] + P
```

S[0] represents TECi in Fig. S2 above and n = the number of nucleotide addition events from this position to n = 20. g, equal to x in Fig. 4A and in eqs. 8-9, represents the number of nucleotide additions between the first position (i) and the termination position. P was used to describe the TECs in the paused state at template position i+x in Figs. S2 and 4, as well as in eqs. 6-9, with rate constants for entry and exit of k_{1f} and k_{1r}, $k_{pause,x}$ and $k_{PE,x}$, respectively. j describes the sum of the TECs in both paused and elongation competent states at template position x (g in the Berkeley Madonna model). T represents the fraction of terminated TECs with an entry rate constant of k_{2f} or $k_{release}$. Because this is an irreversible reaction the reverse rate constant was set to zero.

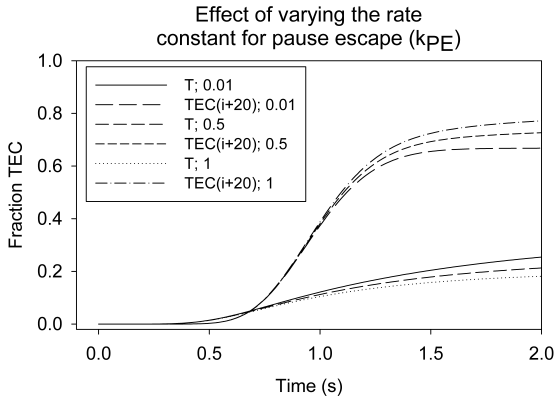

Figure S3 Effect on termination efficiency of varying the rate of pause escape (k_{PE}). The fraction of total TECs terminated (T) or run-off (TEC(i+20)), when the rate constant for pause escape (k_{PE}) was varied. Units for the rate constants in the legend are s^{-1}. The rates for pause entry (k_{pause}) and termination release ($k_{release}$) were set to 10 and 1 s^{-1}, respectively.

Stall Escape Module

TECs, including the polymerase and the associated moveable features of the nucleic acid scaffold (the open transcription bubble and the RNA-DNA hybrid), can 'detach' from the template position at which the 3'-end of the nascent RNA is in the product binding sub-site of the polymerase. Such detached TECs can then move upstream along the template strand 'through' the DNA duplex in a thermally-driven random walk, moving in a 'zipper-like' isoenergetic manner to (potentially) within 8-9 nts of the 5' end of the nascent RNA, because such movement results in no net gain or loss of duplex DNA bps (1-2). This diffusion of such detached or decoupled TECs not only provides a mechanism for the formation of backtracked pauses and arrested states, but – in the form of a single bp translocation -- also can be used to allow undetached TECs to move rapidly between the pre-translocated state (in which the 3'-terminus of the nascent RNA is in the substrate binding site) and the post-translocated state (in which the TEC is positioned one template bp further downstream with the 3'-end of the RNA located in the product binding site) (3-5). Binding of the next required NTP within the active site is thought to 'pin' the TEC in the latter position, preventing movement back into the pre-translocated state by a mechanism that can be described by the Brownian ratchet model (5 , 6).

Stalled TECs may therefore occupy a number of positions along the DNA template upstream of the next nucleotide addition position (Fig. 6), with these positions including the post-translocated state, the pre-translocated state and various backtracked states (depending on sequence context), as has been shown experimentally in numerous DNA foot-printing studies (5, 7). Finally,

backtracked TECs may enter into 'arrested' states, which effectively comprise dead-end complexes in the absence of the transcription co-factors (for example GreA or GreB) that facilitate transcript cleavage and removal of sequences from the 3'-end of the transcript, thus promoting the rebinding of the shortened transcript to the active site of the RNAP. This model for stall escape is depicted in Fig. 6 and in the simplified version described by eqs. 10-12 in the paper.

Because the experimental transcription elongation data were generated with TECs stalled in the absence of NTPs, which could thus be in a number of alternative states in addition to the elongation competent mode, the model used to fit such data must include these alternative states at the first elongation position (i).

This model for stall escape was scripted for the Berkeley Madonna kinetic simulation/fitting program as follows:

```
d/dt(S[0]) = - k1f * S[0] + k1r * X - kf * S[0] + kr * S[1]
d/dt(X) = - k1r * X + k1f * S[0] - k2f * X
d/dt(Y) =  k2f * X

Init X = 0.25* I
Init Y = 0.00001
Init S[0] = 0.75*I

C = X + Y + S[0]
limit C <= I

I = 0.16

kf =15
kr = 0.00001
k1f = 20
k1r = 0.1

u = S[0]
```

Here S[0], X and Y represent the pre/post translocated, backtracked and arrested states at the stall position I, corresponding to TECi(NTP), TECi(B) and TECi(A) in Fig. 6. Rate constants k_{1F}, k_{1R} and k_2 correspond to k_{BF}, k_{BR} and k_A. C represents the sum of all states at template position i, while l was the initial value of C at time zero and was determined as the average from three experiments for the fraction of TECs present at position i at time zero. u is the output data for the fraction of active, elongation competent TEC in position S[0] or i. The data and resulting fits using this model are discussed in (8).

Reference for Supporting Material

1. Greive, S. J., and P. H. von Hippel. 2005. Thinking quantitatively about transcriptional regulation. *Nat Rev Mol Cell Biol* 6:221-232.
2. Landick, R. 2006. The regulatory roles and mechanism of transcriptional pausing. *Biochem Soc Trans* 34:1062-1066.

3. Toulme, F., M. Guerin, N. Robichon, M. Leng *et. al.* 1999. In vivo evidence for back and forth oscillations of the transcription elongation complex. *Embo J* 18:5052-5060.
4. Komissarova, N., and M. Kashlev. 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.
5. Bar-Nahum, G., and E. Nudler. 2001. Isolation and characterization of sigma(70)-retaining transcription elongation complexes from Escherichia coli. *Cell* 106:443-451.
6. Guajardo, R., and R. Sousa. 1997. A model for the mechanism of polymerase translocation. *J Mol Biol* 265:8-19.
7. Toulokhonov, I., J. Zhang, M. Palangat, and R. Landick. 2007. A central role of the RNA polymerase trigger loop in active-site rearrangement during transcriptional pausing. *Mol Cell* 27:406-419.
8. Greive, S. J., B. A. Dyer, S. E. Weitzel, J. P. Goodarzi *et. al.* 2011. Fitting experimental transcription data with a comprehensive template-dependent modular kinetic model. *Biophys J* Submitted.