SUPPLEMENTARY DISCUSSION

Mathematical treatment of the model for transcription blockage by promoter-proximal R-loop formation.

Refer to the scheme shown in Fig.6 in the manuscript.

During one round of transcription, the probability of R-loop formation for an R-loop-prone substrate is p. If the R-loop doesn't form, RNAP synthesizes the full-size (run-off) transcript. If an R-loop is formed, then RNAP could either stall immediately, or continue transcription in the "R-loop mode", in which the newly synthesized RNA product continues to hybridize with the DNA template strand; consequently, the R-loop grows in size as long as transcription proceeds. In this mode, the RNAP is prone to stalling, somewhere within or not far downstream from the R-loop-prone sequence. However, in principle, there is a certain probability (designated as $(1-\theta)$), that the RNAP transcribing in the R-loop mode would still reach the end of the template to produce a full-size transcript. Note that the full-size transcript produced in the R-loop mode (as with all other transcripts produced in this mode) would remain bound to the DNA template during the experiment, because the RNA-DNA hybrid within the R-loop-prone sequence is much more stable than the DNA-DNA hybrid; so the RNA-DNA hybrid would not be displaced by a DNA-DNA hybrid. However, upon denaturation of the sample for analysis on the sequencing gel, this transcript would be released from the DNA and would contribute to the run-off product signal.

Thus, the general probability to produce a full-size transcript during one round is

$$1 - p + p(1 - \theta) = 1 - p\theta \quad (1)$$

We will approximate the transcription process as a single reaction with an apparent rate constant k, which increases with increase of RNAP concentration, and also depends upon other factors, like NTP concentration, ionic conditions, etc. In principle, it could change during the time of reaction, t (e.g. due to NTP depletion).

Let S be the number of DNA substrate molecules that doesn't contain an R-loop. If we assume that the R-loop is so stable that within the period of the transcription experiment its formation can be considered as irreversible, then

$$\frac{dS}{dt} = -pkS \quad (2)$$

Let F be the number of full-size transcripts. If we assume that an R-loop in the vicinity of promoter completely blocks the following rounds of transcription, and that consequently, only substrates without an R-loop could be transcribed, then (taking into account Eq.1)

$$\frac{dF}{dt} = (1 - p\theta)kS \quad (3)$$

From Eqs.2,3 we obtain

$$S = S_0\, e^{-p\tau} \quad (4)$$

and

$$F = S_0\,(1 - p\theta)\,\frac{1-e^{-p\tau}}{p} \tag{5}$$

where S_0 is the initial number of DNA substrate molecules, and

$$\tau = \int_0^t k\,dt \tag{6}$$

that could be interpreted as the number of transcription rounds during the period t. (If the apparent rate constant k doesn't change over time, then τ is simply equivalent to kt.) As mentioned above, k increases with increased RNAP concentration. Thus, τ could be increased or decreased by increasing or decreasing the reaction time, as well as by increasing or decreasing the RNAP concentration.

If τp is large (i.e. $\tau \gg 1/p$), then F (Eq.5) approaches maximal value

$$F_\infty = S_0\,\frac{1-p\theta}{p} = S_0\left(\frac{1}{p} - \theta\right) = S_0\left(\left(\frac{1}{p} - 1\right) + (1 - \theta)\right) \tag{7}$$

This equation has a simple interpretation: If the probability of R-loop formation during one round of transcription is p, then the average number of transcriptional rounds that occur within a given DNA substrate molecule before the R-loop is formed is $\frac{1}{p}$ (this includes the round of transcription at which the R-loop is formed); and consequently,

before the R-loop is formed, $\left(\frac{1}{p}-1\right)$ full-size transcripts would be produced from this DNA molecule. After R-loop formation, one more full-size transcript would be produced with a probability $(1-\theta)$, and with probability θ no full-sized transcripts would be produced. Thus, after R-loop formation, the average number of full-sized transcripts produced from the substrate molecule is $(1-\theta)$. Therefore, each of the S_0 substrate molecules would eventually produce on average $\left(\frac{1}{p}-1\right)+(1-\theta)$ full-sized transcripts, leading to Eq.7.

If τp is small (i.e. $\tau \ll 1/p$), then $1-e^{-p\tau} \approx p\tau$, and in this case Eq.5 can be approximated by

$$F \approx S_0\,(1-p\theta)\,\tau \qquad (8)$$

In the case of non-R-loop-prone substrates, the number of "active" substrate molecules remains the same, so that the yield of full-size molecules (designated as Φ) is simply

$$\Phi = S_0\,\tau \qquad (9)$$

Because with increasing τ, F (Eq.5) increases only up to some finite value (Eq.7), while Φ (Eq.9) increases without limitation, the ratio Φ/F would also increase without limitation upon increase of τ.

In the opposite situation, in which τ is approaching zero, the ratio Φ/F approaches its minimal possible value, which is the ratio of Eqs.9 and 8 :

$$\lim_{\tau \to 0}(\Phi/F) = \frac{1}{1-p\theta} \quad (10)$$

which for sufficiently small p is close to unity. (As discussed in the main text, p is likely to be small even for very energetically favorable R-loops because of the significant kinetic barrier for R-loop formation.)

Note that Eq.10 is based upon the simplified assumption, that the only difference between R-loop-prone and non-R-loop-prone substrates is the propensity to form an R-loop, while there could be other, more subtle sequence-specific effects of promoter-proximal sequence that could require further corrections.

In the above model, we postulate irreversible R-loop formation, and complete blockage of following rounds of transcription after formation of the promoter-proximal R-loop. Below we will briefly consider consequences of small deviations from these postulates:

First, we assume that the R-loop, instead of being formed irreversibly, can dissociate, though very slowly in comparison with the rate of its formation. In this case, after a sufficiently long time period, the fraction of R-loop-free DNA substrates, instead of approaching zero, would approach a finite small value designated as ε.

Second, we assume that transcription from the R-loop-containing substrate, instead of being completely blocked, could occur though much more slowly than from the R-loop-free substrate (i.e. its rate constant would be equivalent to the rate constant of normal transcription times a small coefficient designated as ϵ.)

Under these assumptions, the steady-state rate of transcription would be $k\varepsilon S_0 + \epsilon k(1-\varepsilon)S_0$ (here the first and the second terms correspond to transcription from R-loop-

free and R-loop-containing substrates, respectively; also, we assumed that the rate constant k does not change over time). After a sufficiently long period, the amount of full-size transcripts could be approximated as the steady-state rate multiplied by time, i.e. $F \approx \varepsilon S_0 \tau + \epsilon(1-\varepsilon) S_0 \tau$.

Combining this result with Eq.9, we obtain: $F \approx \big(\varepsilon + \epsilon(1-\varepsilon)\big)\Phi$.

Thus, the ratio Φ/F, instead of increasing indefinitely with time, would approach a finite large number $1/\big(\varepsilon + \epsilon(1-\varepsilon)\big)$.

SUPPLEMENTARY FIGURE LEGENDS

Fig.S1. Ratio of distal to proximal run-off signals as a function of T7 RNAP concentration.

Substrates containing G-rich sequence were used in this experiment. Undiluted T7 RNAP corresponds to concentration 1.7 units/μl, that we refer to as a "high" concentration, while the maximal (1/30) dilution correspond to "low" concentration. It is seen that the ratio decreases with reduced concentration, approaching a value close to unity.

Fig.S2. Promoter-distal and promoter-proximal substrates do not affect each other when transcribed in the same mixture.

Substrates containing the G-rich sequence were used in this experiment.

A: Gel image. Lanes 1, 2, 3 contain both promoter-distal and promoter-proximal substrates, only promoter-distal substrate, and only promoter-proximal substrates,

respectively. All reagent concentrations and conditions were standard. **B:** Quantitation. The yields of transcription products for both substrates are practically the same, whether they are transcribed together, or separately.

Fig.S3. Transcription blockage is similar at increased concentration of potassium (K) and lithium (Li).

Substrates containing the G-rich sequence were used in this experiment.

To test effect of monovalent cations upon the transcription blockage, in the standard transcription buffer 8.3 mM of NaCl were replaced by 83 mM of either KCl or LiCl (further referred as K- or Li-buffers). This 10-fold increased salt concentration makes conclusions about effect of monovalent cations more reliable, because it strongly (~ 20-fold) exceeds the residual amount of potassium (~4 mM) that comes into the transcription mixture from RNAP and RNasin storage buffers. However, the increased salt concentration causes strong inhibitory effect upon transcription, and consequently, decreases the number of transcriptional rounds during a given time interval, which makes the observed blockage less pronounced in a similar way as decrease in RNAP concentration. To compensate for that, we used a "pre-transcription" protocol, that allows us to pre-form R-loops before producing radioactively labeled transcripts; thus making the blockage more pronounced (e.g., see Fig.4). In this protocol, transcription was first performed for 3 h at 37°C in 12 µl of either in K- or Li-buffers with standard concentration of RNAP, non-radioactive NTPs and other components, but without radioactive NTP. After that, 2 µl of mixture containing 15 µCi of (α-^{32}P) CTP and 0.5 mM of non-radioactive NTPs (the latter is to compensate possible NTP depletion during

pre-transcription) were added to the transcription mixture and incubation was continued for another 50 min. The reaction was stopped by adding 0.4 µl of 0.5 M EDTA, supplemented with spiking transcript, and the products were analyzed on denatured gel as in the standard protocol (see Materials and Methods, the main text). Experiments were repeated twice. The molar transcription yields were normalized to the molar yield for promoter-distal substrate in the presence of potassium (the first column in the top chart). It is seen that in the presence of K or Li, for promoter-proximal substrate the yield of transcription is strongly decreases indicating strong transcription blockage in both cases. For Li this blockage appears to be 2-fold greater than for K (14-fold versus 7-fold), however, it could be due to the fact that in Li the rate of transcription per se (judging from non-R-loop-prone promoter-distal substrate) appears to be about 2-fold greater, thus more R-loops could be accumulated during pre-transcription. In any case, at least the blockage in Li is not smaller, than in K, arguing against quadruplex contribution to the blockage. Note that the minor "repeat-exiting" blockage signals (small white block arrows) are also similar in K and in Li.

SUPPLEMENTARY FIGURES

Fig.S1

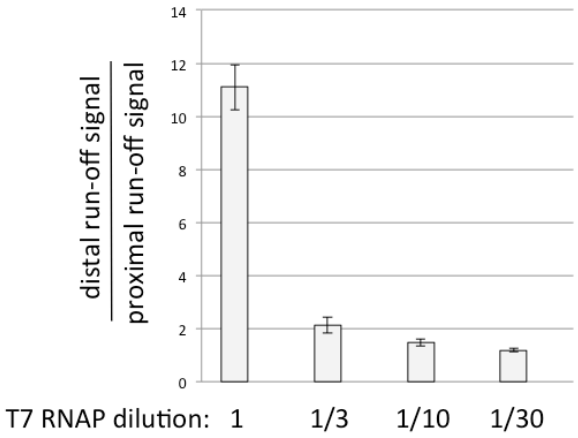

distal run-off signal
proximal run-off signal
14
12
10
8
6
4
2
0
T7 RNAP dilution: 1 1/3 1/10 1/30

Fig.S2

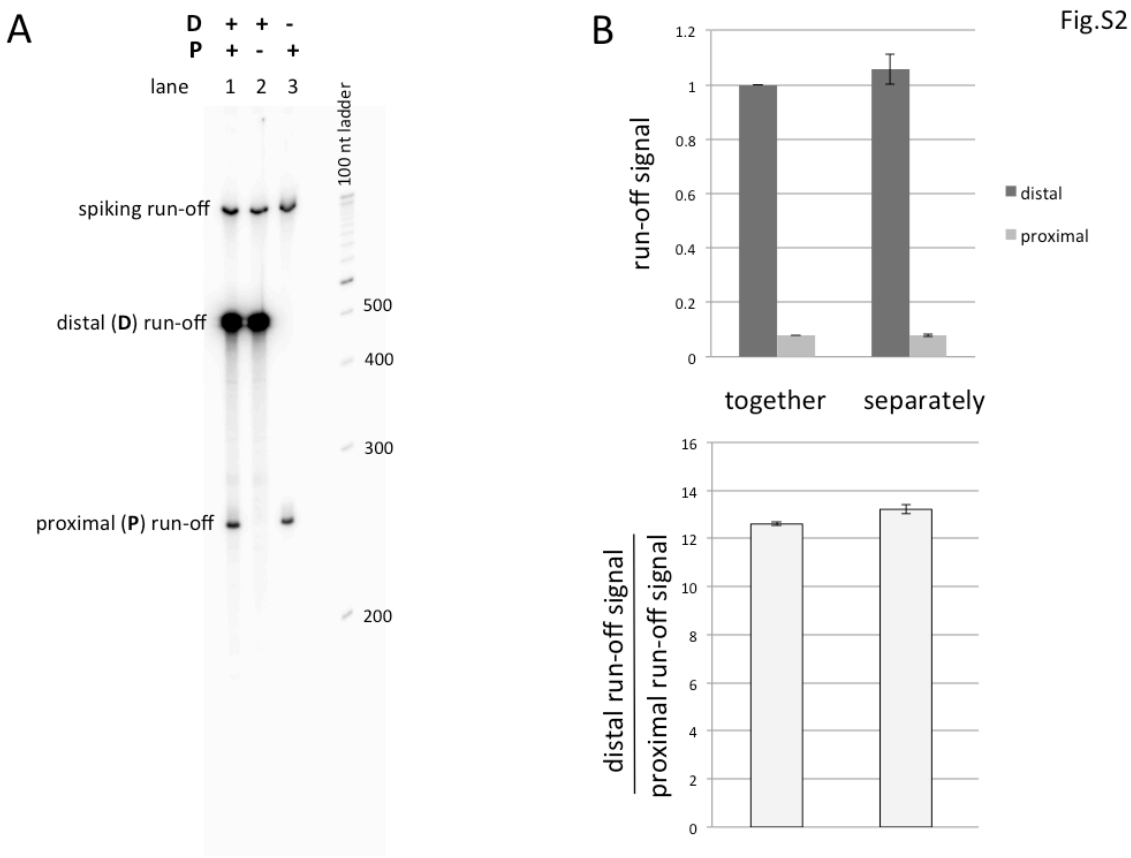
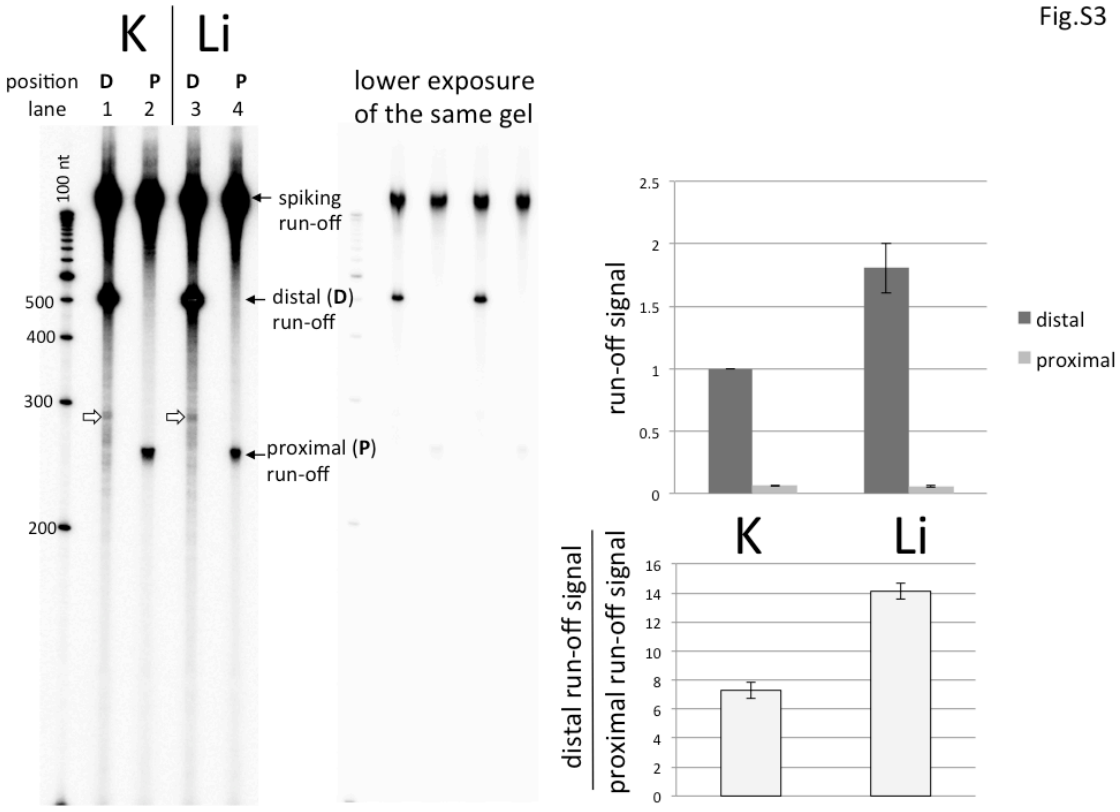

A
D + + -
P + - +
lane 1 2 3
100 nt ladder
spiking run-off
distal (D) run-off
proximal (P) run-off
500
400
300
200
B
run-off signal
1.2
1
0.8
0.6
0.4
0.2
0
distal
proximal
together
separately
distal run-off signal
proximal run-off signal
16
14
12
10
8
6
4
2
0

Fig.S3


K
Li
position D P D P
lane 1 2 3 4
lower exposure of the same gel
100 nt
500
400
300
200
spiking run-off
distal (D) run-off
proximal (P) run-off
run-off signal
2.5
2
1.5
1
0.5
0
distal
proximal
distal run-off signal / proximal run-off signal
K
Li
16
14
12
10
8
6
4
2
0