Biophysical Journal Volume 106 January 2014 55–64

Thermodynamic Modeling of Variations in the Rate of RNA Chain Elongation of *E. coli rrn* Operons

David Fange,[†] Harriet Mellenius,[†] Patrick P. Dennis,[‡] and Måns Ehrenberg[†*]

[†]Department of Cell and Molecular Biology, Uppsala University, Uppsala, Sweden; and [‡]Janelia Farm Research Campus, Howard Hughes Medical Institute, Ashburn, Virginia

ABSTRACT Previous electron-microscopic imaging has shown high RNA polymerase occupation densities in the 16S and 23S encoding regions and low occupation densities in the noncoding leader, spacer, and trailer regions of the rRNA (*rrn*) operons in *E. coli*. This indicates slower transcript elongation within the coding regions and faster elongation within the noncoding regions of the operon. Inactivation of four of the seven *rrn* operons increases the transcript initiation frequency at the promoters of the three intact operons and reduces the time for RNA polymerase to traverse the operon. We have used the DNA sequence-dependent standard free energy variation of the transcription complex to model the experimentally observed changes in the elongation rate along the *rrnB* operon. We also model the stimulation of the average transcription rate over the whole operon by increasing rate of transcript initiation. Monte Carlo simulations, taking into account initiation of transcription, translocation, and backward and forward tracking of RNA polymerase, partially reproduce the observed transcript elongation rate variations along the *rrn* operon and fully account for the increased average rate in response to increased frequency of transcript initiation.

INTRODUCTION

The bacterium *Escherichia coli* contains within its genome seven ribosomal RNA operons (*rrn*) with an average length of ~5500 nucleotides (nt). Each operon encodes in its proximal, middle, and distal positions the 16S, 23S, and 5S ribosomal RNAs (rRNAs), respectively. These operons are among the most highly transcribed regions of the *E. coli* chromosome. In wild-type strains, it takes the RNA polymerase 60 s to transcribe the entire *rrnC* operon (1), corresponding to an average rRNA chain elongation rate of ~91 nt s^{-1} (= 5500 nt/60 s; see Table 3 of Bremer and Dennis (2)). Deletion of individual *rrn* operons from the *E. coli* chromosome has little impact on the growth rate. This means that in deletion strains, the remaining intact operons are transcribed correspondingly more frequently to produce the required number of ribosomes to sustain growth. For example, in a strain with four of the seven *rrn* operons inactivated by partial deletion, the rate of transcript initiation per promoter at the intact *rrn* operons was 1.6-fold higher than that in wild-type bacteria growing under the same conditions (Table 1 of Ehrenberg et al. (3)). Surprisingly, the time to transcribe the intact *rrn* operons in the deletion strain was only 40 s (1), corresponding to an average rRNA chain elongation rate of 138 nt s^{-1}, 1.5-fold faster than the elongation rate in the wild-type strain.

For wild-type *E. coli* bacteria, electron-microscopic (EM) images have shown a nonuniform but reproducible distribution of numbers of RNA polymerase transcription complexes (TCs) in segments along the *rrn* operons (4) (here reproduced; see Fig. 2). This nonuniform density reflects an uneven elongation rate as the TCs traverse along the operon (5). Where the TCs move slowly along the operon, they clump together to produce a higher density; where they move more rapidly, they are more spread out on the template. The average rate by which a polymerase traverses such a segment is, more precisely, equal to the inverse of the polymerase density multiplied by the per operon initiation frequency (5). Based on these considerations, the following pattern emerges. As a TC leaves the promoter, it passes through an antitermination site, where it picks up the NusA, B, and G factors (6). Transcription in the leader region speeds up to an elongation rate of ~230 nt s^{-1} (the rates refer to averages within 20 sections of the operon, each ~275 nt long). Transcription then slows down to ~65 nt s^{-1} within the 16S gene, speeds up again to ~100 nt s^{-1} in the spacer between the 16S and 23S genes, slows down again to 65 nt s^{-1} in the 23S region, and finally accelerates to a rate of >400 nt s^{-1} near the end of the operon (5).

In contrast to this wild-type pattern, the EM images of *rrn* operons in the *rrn* deletion strain appear to show two queues of polymerases, one in the leader region in front of the 16S gene that extends all the way back to the promoter and a second of similar length in the spacer between the 16S and 23S regions (Fig. 5 of Condon et al. (1)). We have earlier interpreted these queues to be indicative of transcriptional pauses of ~0.56 s in duration at the beginning of the 16S and 23S regions (5). The pausing of the TC causes congestion behind the pause sites where polymerases are spaced along the DNA at a distance of ~60 nt. The first queue extending

Submitted July 16, 2013, and accepted for publication November 25, 2013.

*Correspondence: ehrenberg@xray.bmc.uu.se

Editor: Reka Albert.

0006-3495/14/01/0055/10 $2.00

http://dx.doi.org/10.1016/j.bpj.2013.11.4487

back to the *rrn* promoter appears to limit the rate of transcript initiation to 1/0.56 s, and corresponds to an *rrn*P1 promoter activity of 107 transcript initiations/min (5). Within these queues, the polymerases are predicted to move forward with a chain elongation rate of ~110 nt s^{-1}. This elongation rate is less than the rate of 230 nt s^{-1} observed in the leader region of *rrn* operons of wild-type bacteria (see above). This apparent slower chain elongation in the leader and spacer regions must then be compensated by much faster elongation in the 16S and 23S coding regions to produce an average chain elongation in the deletion strain that is 1.5-fold faster than in the wild-type.

These observations raise two fundamental questions: first, what causes the variation in the transcript elongation rate along the *rrn* operon, and second, why do the polymerases speed up when the rate of transcript initiation is increased (i.e., in the strain with four *rrn* operons inactivated by partial deletion)? In the following, these questions are addressed by modeling the transcription along *rrn* operons, based on DNA template-dependent standard free-energy variations of the transcript elongation complex (7,8). This approach has been employed previously to understand pausing in single-molecule in vitro RNA polymerase transcription (9–11). A similar approach has been used to show that interactions between polymerases can significantly accelerate transcription elongation over different short template sequences (12). Here, we use thermodynamic simulations of the transcription bubble to probe the mechanisms responsible for the variations of the rRNA chain elongation rate along the *rrn* operon and the observed stimulation of rRNA chain elongation in response to an enhanced frequency of transcript initiation.

MATERIALS AND METHODS

All computations were carried out for the *rrnB* operon sequence as downloaded from EcoGene (13). Free-energy calculations for the transcription bubble were carried out as described in previous work (8,9,14) and in the Supporting Material, Free energies of the elongation complex. Mean-time calculations were performed as described by Bilgin and co-workers (15) and carried out using MATLAB (for details of the computations, see the Supporting Material, Mean-time calculations. The stochastic simulations were carried out using an in-house computational tool written in C++. The stochastic simulations are based on Monte Carlo methods as described by von Kampen (16). Special care has been taken to avoid resampling of event times for nonmoving RNA polymerases. For details of the stochastic simulation algorithm, see the Supporting Material, Stochastic simulation algorithm.

RESULTS

Transcription model and stochastic simulation of *rrn* operon transcription

Transcription of an *rrn* operon occurs in four major stages: 1), transcript initiation, including formation of the first internucleotide phosphodiester bond; 2), promoter regeneration, where the polymerase moves sufficiently far away from the promoter to allow for promoter binding of the next polymerase; 3), further transcript elongation along the operon; and 4), transcript termination and polymerase dissociation at the end of the operon.

Transcript initiation

Each *rrn* operon has two tandem promoters, *rrn*P_1 and *rrn*P_2, ~120 basepairs apart (Fig. 1 *A*). The P_1 promoter is regulated by the Fis activator protein, which binds to sites upstream of P_1, and by the effector molecule ppGpp, which interacts with RNA polymerase and decreases its affinity for the P_1 promoter (17,18). The P_2 promoter has been suggested to be either constitutive (17,18) or regulated by an unknown mechanism (19,20). During rapid growth at 37°C, initiation of transcription occurs almost exclusively from the P_1 promoter, since access to the downstream P_2 promoter is blocked by occlusion. The per operon initiation frequency in rapidly growing *E. coli* is about 1 s^{-1} and the average distance between transcribing RNA polymerases is ~100 basepairs (21). Since the experiments on local polymerase density variation and shifting transit times were carried out in rich medium at 37°C (1,4), our model considers only initiation of transcription from the P_1 promoter. The multistep initiation process (17,18) is here approximated as a single step with rate constant k_I, bringing free RNA polymerase and free P_1 promoter to promoter-bound, elongation-primed RNA polymerase (Fig. 1 *B*). Our model does not discriminate between initiation frequency variation due to changing free RNA polymerase concentration (17,18) or promoter regulation (19).

Promoter regeneration

In line with previous models (3,17,20), we propose that for a next polymerase to bind to the promoter, the previous polymerase must have moved at least 50 bases downstream from its initial binding site (Fig. 1 *B*). Moreover, we use a minimal center-to-center distance of 50 basepairs in considering interactions between adjacent transcribing RNA polymerases. Downstream from the P_2 promoter, the polymerase passes through the antitermination sites (Fig. 1 *A*) which, together with Nus-factors (6), speeds up transcription of *rrn* operons compared to that of mRNA operons (6,22). In our simulations, we have calibrated the average transcription rate to experimental data. The Nus-factor effect, which results in a more rapid rate of RNA chain elongation in *rrn* operons compared to that in mRNA operons, is taken into account by this calibration procedure.

Transcript elongation

The time required for polymerase movement from the beginning to the end of the *rrn* operon is much longer than the time required for transcription initiation or termination. At each step of elongation, the polymerase alternates

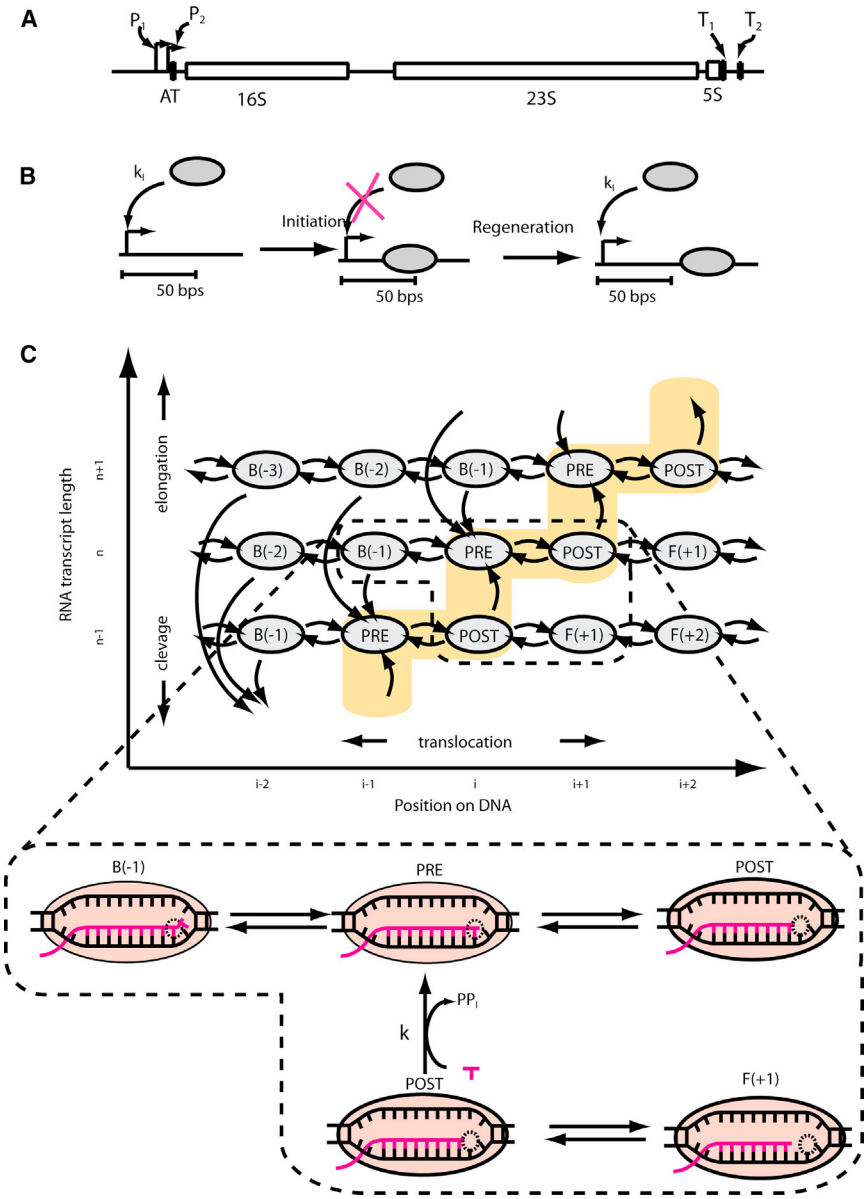

FIGURE 1 Transcription of the *rrn* operon. (*A*) The typical *rrn* operon is ~5500 nt in length, and contains two promoters (P_1 and P_2), an antitermination site (*AT*) and two terminators (T_1 and T_2). (*B*) RNA polymerase binds to the free promoter and initiates transcription with a forward rate constant of k_1. Immediately after initiation, the promoter is closed and only becomes again free for initiation after the transcribing polymerase has elongated a distance of ~50 nt. (*C*) The positions and movement of the transcribing RNA polymerase around position *i* on the DNA template are illustrated. Productive elongation from position *i*−*1*, through position *i*, to position *i*+*1* is shown in the yellow highlight. The transcription bubble, the RNA transcript, and the RNA-DNA hybrid are shown for the pretranslocation (PRE) and posttranslocation (POST) states and for the single backward-track (B(−1)) and forward-track (F(+1)) states. Unidirectional movement between POST and PRE occurs upon nucleotide incorporation at the 3′ end of the RNA transcript with a rate constant of *k*. Reversible movement along the productive elongation pathway occurs between the PRE and POST states. Movement to nonproductive states occurs by forward (*F*) or backward (*B*) tracking.

between nucleotide incorporation into the transcript and one-base forward translocation along the DNA template (8,23,24). We approximate the multiple steps on the pathway from free NTP to NMP incorporation by phosphodiester bond formation (25) by a single step with rate constant *k*, normally taken to be the same for all different positions along the operon. The effects of varying *k* values are discussed in the Supporting Material (Explicit model of nucleotide addition). Nucleotide incorporation brings the polymerase from the post- to the pretranslocation state, and subsequent translocation brings it from the pre- to the posttranslocation state (Fig. 1 *C*). The polymerase may also backward track from the pretranslocation state to a B state (Fig. 1 *C*) in which the nascent transcript protrudes into the catalytic site and temporarily shuts down base incorporation (26). The polymerase may return from the backward-tracked to the pretranslocation state either by forward tracking or by GreA-and-GreB-facilitated cleavage of the protruding mRNA (27,28). The polymerase may also forward track from the posttranslocation state to the F state. In this state, the 3′ end of the transcript has lost contact with the nucleotide addition site, which temporarily shuts down further nucleotide incorporation (9) and may induce transcription termination (29).

Normal polymerase translocation from template base *i* to *i*+*1* (Fig. 1 *C*) is described by a first-order rate constant $k_T^{i,\,i+1}$ (9–11). Its magnitude is determined by the standard free energies of the transcription elongation complex (TEC) in position *i* in pretranslocation state ($\Delta G^0_{pre,i}$), in position *i*+*1* in posttranslocation state ($\Delta G^0_{post,i+1}$), a

translocation barrier ($\Delta G^0_{barrier}$), and a posttranslocation bias (ΔG^0_{bias}) through the relationships:

$$k_T^{i,i+1} = \begin{cases} k_0 e^{-\frac{\left(\Delta G^0_{barrier} + \left(\Delta G^0_{post,i+1} - \Delta G^0_{bias} - \Delta G^0_{pre,i}\right)\right)}{RT}} & \text{when } \Delta G^0_{post,i+1} > \Delta G^0_{bias} + \Delta G^0_{pre,i} \\ k_0 e^{-\frac{\Delta G^0_{barrier}}{RT}} & \text{when } \Delta G^0_{post,i+1} < \Delta G^0_{bias} + \Delta G^0_{pre,i} \end{cases}. \quad (1)$$

Polymerase movements associated with backward tracking or forward tracking are also described as single-step reactions (see Supporting Material). The equilibrium constant, $K_T^{i,i+1}$, connecting pretranslocation state i with posttranslocation state $i+1$ is given by

$$K_T^{i,i+1} = \frac{k_T^{i,i+1}}{k_T^{i+1,i}} = e^{-\frac{\left(\Delta G^0_{post,i+1} - \Delta G^0_{bias} - \Delta G^0_{pre,i}\right)}{RT}}. \quad (2)$$

Formation of the transcription bubble from a DNA double helix (Fig. 1 *C*) requires breaking of 12 DNA-DNA basepairs (8,9) and formation of either nine RNA-DNA basepairs (pretranslocated or backward-tracked states) or eight RNA-DNA basepairs (posttranslocated states) (Fig. 1 *C*) (23,25). The $\Delta G^0_{pre,i}$ and $\Delta G^0_{post,i+1}$ free energies were estimated as the sum of three terms (7,8). The first, $\Delta G^0_{DNA-DNA}$, accounts for the varying free energy change associated with opening of the DNA-DNA double helix. The second, $\Delta G^0_{RNA-DNA}$, accounts for the varying free energy of the RNA-DNA double helix. These estimates were obtained from the nearest-neighbor model (7,14,30) and supported by experimental data (14,30) (see Supporting Material). The third term, $\Delta G^0_{polymerase}$, accounts for the standard free energy of RNA polymerase interaction with nucleotides. This was assumed to be sequence-independent and biased in favor of the posttranslocation state with the term $\Delta G^0_{forward-bias}$. This term greatly increases the efficiency of transcription in the presence of extensive backward tracking. Given all the constraints imposed by the sequence of the DNA template, our parameter sets are optimized to fit both the polymerase density from Quan et al. (4) and the decreasing transit time from Condon et al. (1).

Transcript termination

For termination at the *rrn* operons we considered the two rho-independent terminators, T_1 and T_2 (31), where T_1 is less efficient than T_2 (32). We have ascribed 10% ($P_{term,\ T1} = 0.1$) termination probability to T_1 and 100% to T_2 ($P_{term,\ T2} = 1.0$) (see the Supporting Material for details).

Sequence-dependent elongation rate variation during *rrn* operon transcription

The GC contents of *rrn* operons have maxima in the 16S and 23S coding regions of *rrn* operons, as illustrated for the *rrnB* operon in Fig. 2 *A*. This suggested to us that the DNA-sequence-dependent variation of the standard free energy of the transcription bubble might contribute to the experimentally observed RNA polymerase density variation during *rrn* operon transcription (Fig. 2 *A*). This approach is also in line with in vitro experiments showing that the transcription rate varies considerably along a template sequence (33). Exploring this possibility, we first modeled the movement of polymerases over the *rrn* operon, neglecting interpolymerase interference, backward tracking, and forward tracking and only taking productive nucleotide incorporation and movement between pre- and posttranslocation states (Fig. 1 *C*) into account. We divided the *rrn* operon into 20 equal segments, each 280 basepairs in length and computed the average time, τ_i, that the polymerases spend in each segment, i (see the Supporting Material, Mean-time calculations, for details). Using the steady-state relation $\tau_i = N_i/v$ (where v is the rate of initiation per promoter and N_i is the polymerase density in segment i, defined as the average number of polymerases in this segment (3)), our simulation generated N_i maxima in the 16S and 23S regions of the *rrn* operons (Fig. 2 *B*), in line with experimental data (4). However, the relative amplitudes of the modeled density variation are much smaller than those observed experimentally.

To improve the correspondence between theory and experiment, we introduced backward tracking, which is known to cause transcriptional pausing in vitro (34). Initially, only single-step backward tracking was allowed, in line with structural data suggesting that the first backward-tracked state is energetically favored in relation to subsequent states (35,36). Through this modification, the relative amplitudes of the density variation increased (Fig. 2, *B* and *C*), suggesting that backward tracking may contribute significantly to the experimentally observed polymerase density variation.

Next, we allowed for interpolymerase interference by postulating that for two polymerases edge to edge, the trailing partner cannot move forward and the leading partner cannot move backward. We implemented Monte Carlo simulations of the movement of each one of the operon-bound polymerases (~50) at an initiation frequency per empty promoter, k_I, of 1.5 s^{-1}. Apart from initiation, the model includes normal translocation, one backward-tracked step, and termination at the T_1 and T_2 terminators (see above, Transcription model and stochastic simulation). To compare the Monte Carlo calculations predicted with the

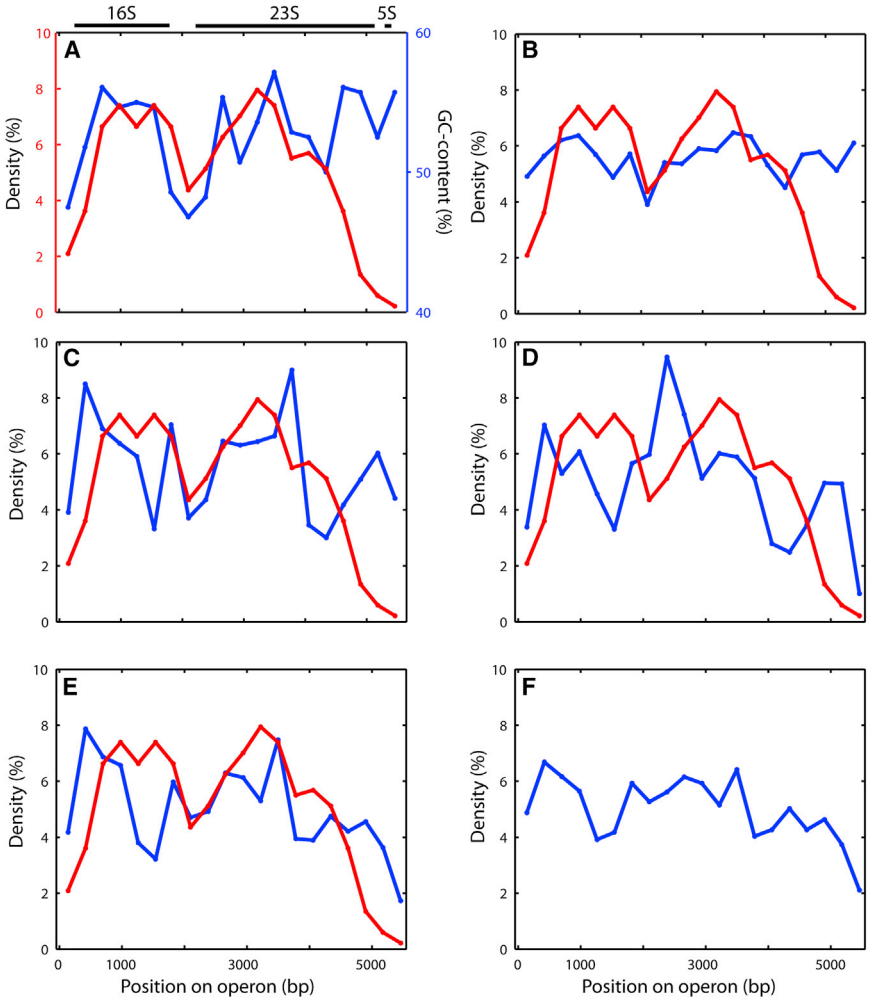

FIGURE 2 Simulation of *rrn* operon transcription. Experimental measurement of polymerase density along the 5500-nt-long *rrn* operon (4) (*red curves* in *A–E*) serves as a comparison to the different simulations (*blue curves*). The approximate positions of the 16S and 23S rRNA genes within the operons are indicated at the top of *A*. All computations and simulations use the parameters listed in Table 1, when applicable. (*A*) The variation in the GC content (*rrn B* operon) and the predicted local variation in the RNA chain elongation rate along the *rrn* operon are illustrated. (*B*) The modeled polymerase density in the absence of backward tracking was obtained by computing the mean times the polymerase spends in each segment (τ_i) (see Supporting Material for details) and converting them to polymerase density in each segment using the relationship $(1/v)(\tau_i/\sum_j \tau_j) = N_i/\sum_j N_j$, where $v = 0.9\ s^{-1}$. The inverse of the density (N_i, number of polymerases per segment) is proportional to the average velocity ($1/\tau_i$) with which a polymerase transverses the segment. (*C*) The computation was the same as in *B* except that a single backward-track step between the PRE and B(−1) states (see Fig. 1) was allowed. (*D*) The polymerase density was obtained by monitoring the Monte Carlo simulated movement and positions of multiple polymerases and allowing one backward-track step between the PRE and B(−1) states; the initiation rate per empty promoter, k_I, was set at $1.5\ s^{-1}$. (*E*) Simulation was the same as that in *D* except that unrestricted backward and forward tracking was allowed. (*F*) Simulation was the same as in *E* except that k_I was set at $10\ s^{-1}$. A quantified comparison of model data and the experimental polymerase density shows that the model in *E* has the closest fit, whereas the model in *B* is better after penalization of free parameters. The models in *B* and *C* are unrealistic, however, as they do not include polymerase interactions and therefore cannot explain the observed *rrn* operon transit time decrease with increasing operon initiation frequency (Fig. 3).

experimentally observed RNA polymerase density profile (4), we binned the numbers of polymerases in the 20 evenly spaced sections along the *rrnB* operon from the P_1 promoter to the T_2 terminator and sampled the polymerase positions every 10 s over transcription periods of 10,000 s (Fig. 2 *D*).

Finally, we also allowed for multiple forward-track and backward-track steps, where all backward-track steps except the first were penalized by the free energy increase 3RT. The polymerase density at k_I values of 1.5 and $10\ s^{-1}$ are displayed in Fig. 2, *E* and *F*, respectively. We note that the closest correspondence between the experimentally estimated polymerase density profile in wild-type cells and that modeled in simulations including interactions between polymerases is obtained for the model in Fig. 2 *E* with multiple backward-track steps (Table S3). At the same time, there are deviations between model and experiment, in particular in the distal portion of the 16S rRNA gene. More detailed models, including a varying rate of nucleotide addition, give very similar results (see Supporting Material). In the next section, we discuss what the multiple backward- and forward-track model has to say about the polymerase transit time as a function of varying rate of transcriptional initiation (Table 1).

RNA polymerase interference reduces the *rrn* operon transit time

To clarify how the frequency of transcription initiation affects *rrn* operon polymerase density and operon transit time, we expanded our multiple-step forward- and backward-track simulations (Figs. 2, *E* and *F*) to cover a large range of initiation rate values, k_I (Fig. 3, *black curves*). The actual initiation rate per promoter, v, increases almost linearly with increasing k_I at small k_I values and then saturates in a Michaelis-Menten-like manner at large k_I values (Fig. 3 *A*). The ultimate cause of saturation is the finite promoter regeneration time, which may vary with the initiation frequency due to RNA polymerase congestion and

TABLE 1 Parameters used in simulations

Parameter:	Value:
k_I[a]	As stated
k[b]	2000 s^{-1}
k_{cut}[c]	1 s^{-1}
$\Delta G_{barrier}$[d]	2RT
$\Delta G_{forward\text{-}bias}$[e]	7RT
$\Delta G_{backtrack\text{-}bias}$[f]	3RT
k_0[g]	10^5 s^{-1}

[a]k_I is a first-order rate constant approximation for the multistep initiation process, bringing free RNA polymerase and free P_1 promoter to promoter-bound, elongation-primed RNA polymerase.
[b]k is a rate constant for nucleotide incorporation into the nascent transcript. It is normally sequence-independent, but see the Supporting Material for instances where this condition is relaxed.
[c]k_{cut} is a sequence independent rate constant for transcript cleavage in a backward-tracked state of the polymerase.
[d]$\Delta G_{barrier}$ is a sequence-independent standard free-energy barrier between the translocation states of the polymerase.
[e]$\Delta G_{forward\text{-}bias}$ is sequence-independent free energy favoring the POST in relation to the PRE translocation state.
[f]$\Delta G_{backtrack\text{-}bias}$ is a sequence-independent free-energy penalty for all backward-tracked states except the first.
[g]k_0 is a sequence-independent frequency factor multiplied with the sequence-dependent part of the rate constants in the model.

interference on the operons. The simulated operon transit time is approximately constant at small k_I values, decreases sharply at intermediate k_I values and eventually reaches an asymptote at large k_I values (Fig. 3 *B*). In the steady state, the number, N, of polymerases per operon is equal to the operon transit time, τ, multiplied by the actual initiation frequency, v (5). The simulated value of N increases linearly with k_I at small k_I values and asymptotically approaches $N = 80$ at large k_I values (Fig. 3 *C*). As the k_I value increases from 1.5 to 10 s^{-1}, the τ value decreases from 60 to 40 s and N increases from 50 to 70 (Fig. 3, *B* and *C*). These simulated changes in the values of τ and N are close to their experimentally estimated values (1). We note that the decrease in the value of τ with increasing initiation frequency, depends on multiple-step backward tracking, since a single backward-tracking mechanism responds with an increasing, rather than decreasing, value of τ as k_I increases (Fig. 3 *B*, *red curve*). Moreover, we note that with a single backtracking step only, the initiation frequency (v) reaches a plateau at 1.0 s^{-1}, and the number of polymerases per operon, N, saturates when the value of initiation rate k_I exceeds 1.5 s^{-1} per free promoter (Fig. 3, *A* and *C*, *red curve*).

To understand better the gradual saturation of the actual initiation frequency, v, and the monotonically decreasing transit time, τ, with increasing k_I, we plotted the probability that a nascent transcript of any length is elongated by one nucleotide at a certain average time or dwell time after its formation (Fig. 4 *A*). It is seen that when the value of k_I increases from 1.5 to 10 s^{-1}, the frequency of long dwell times decreases, whereas the frequency of short dwell times

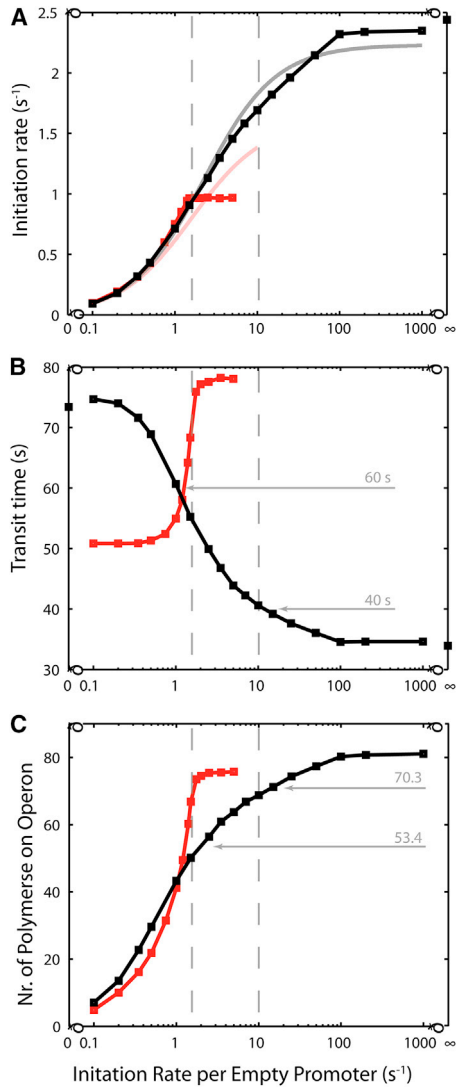

FIGURE 3 Simulated relationships between the initiation rate per empty promoter (k_I) and the steady-state rate of initiation (v), operon transit time (τ), and number of polymerases per *rrn* operon (N). The stippled vertical lines represent the transcription initiation rates per empty promoter (k_I) of 1.5 s^{-1} and 10 s^{-1} used in the simulations in Fig. 2. The red curves were generated for a model allowing only one back-tracking step from the PRE to the B(−1) states and the black curves were generated for a model with unrestricted forward and backward tracking. (*A*) The two simulations of the mean initiation rate (v; *solid lines with squares*) is fitted onto a Michalis-Menten equation describing the polymerase/promoter interaction and given by $v = k_{cat}k_I/(k_{cat} + k_I)$, where $k_I = [RNAP]k_{cat}/K_M$, is the initiation rate per empty promoter (*light gray* and *light red lines*). The fitted value of k_{cat} is 2.23 s^{-1}. (*B*) The simulations of the mean operon transit time (*solid lines with squares*) are illustrated; the horizontal lines indicate the operon transit times (60 s and 40 s) measured in vivo (1,4) for the wild-type and the four *rrn* operon deletion strain, respectively. (*C*) Simulation of the mean number of polymerases on the *rrn* operon is illustrated; the horizontal lines indicate the number of polymerases per operon (53.4 and 70.3) measured in vivo in the wild-type and the four *rrn* operon deletion strains, respectively.

remains almost unaltered. Furthermore, the sum of all dwell times >0.1 s^{-1} decreases greatly with increasing k_I value, showing that this decrease is the ultimate cause of the

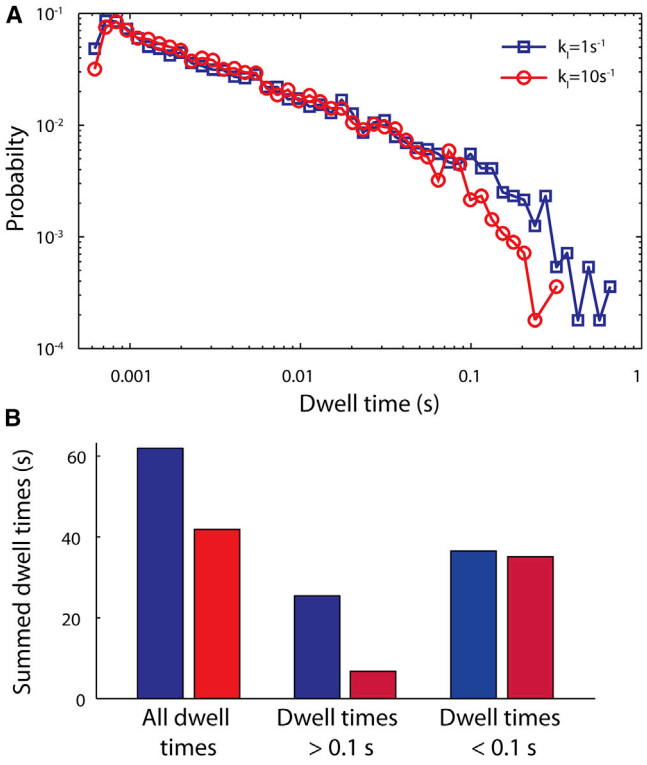

FIGURE 4 Relationship between backtracking and the *rrn* transit time. Simulations involving unrestricted forward and backward tracking and carried out at initiation frequencies per empty promoter (k_I) of 1 s^{-1} and 10 s^{-1} produce operon transit times (τ) of 60 s and 40 s, respectively. (*A*) Probability distribution of the RNA polymerase dwell time (τ_i) at all nucleotide positions along the transcript length is illustrated. The sum of the dwell times across the entire operon is equivalent to τ, the operon transit time. (*B*) The total dwell time (operon transit time τ), and the sums of dwell times above and below 0.1 s are shown for 1 s^{-1} (*blue bars*) and 10 s^{-1} (*red bars*) frequency of initiation of transcription. The figure illustrates that at low initiation frequency (1 s^{-1}), the sum of dwell times >0.1 s is much larger than at high initiation frequency (10 s^{-1}) and that this leads to a shorter operon transit time at high than at low initiation frequency.

reduced overall transit time τ (Figs. 3 *B* and 4 *B*). In this model, the ultimate cause of the monotonic transit time decrease with increasing initiation frequency is the restriction of long backward-tracking excursions of forward polymerases by the edge-to-edge presence of trailing polymerases. A similar conclusion was reached from recent modeling of transcription of a few comparatively short DNA sequences (12).

In contrast, in a model allowing for only a single backtracking step, the dwell time increases abruptly when the initiation rate increases, and this leads to RNA polymerase queues reaching all the way back to the promoter (compare Fig. 3, *A* and *B*, *black* and *red curves*). Here, the dominating effect is restriction of forward movement of trailing polymerases by the edge-to-edge presence of forward polymerases. In the multiple-step backward-tracking case, it is the interplay between the great transcription rate increase by restriction of long backward-tracking excursions and the comparatively small transcription rate decrease by restriction of forward movement by trailing polymerases that leads to the gradual saturation of initiation of transcription (Fig. 3 *A*, *black curve*). In the single-step backward-tracking case, it is the transcription rate decrease by restriction of forward movement of trailing polymerases that leads to the abrupt saturation of initiation (Fig. 3 *A*, *red curve*). Interestingly, in the simplistic backward-tracking model used by Klumpp and Hwa, described in their Supporting Material (20), increasing initiation frequency first reduces and then greatly increases the transit time. This, we suggest, is caused by reduced backward tracking as the initiation frequency increases at small values followed by greatly reduced forward movement of the polymerases at further initiation rate increase.

Hairpin-induced transcriptional pauses

Transcriptional pausing is sometimes associated with hairpinlike structural elements in nascent transcripts (34). We have studied the putative role of hairpins in the generation of the polymerase density pattern observed for *rrn* operon transcription. A problem here is the lack of theory to relate the thermodynamic properties of hairpinlike nascent RNA transcript structures to transcriptional kinetics. Our preliminary analysis, based on an empirical relation (37), suggests that hairpin-based transcriptional pausing cannot explain either the experimentally observed polymerase density patterns (4) or the transit time variation (1) (see Discussion and Supporting Material).

DISCUSSION

We have addressed two remarkable observations regarding transcription of *rrn* operons in growing *E. coli* bacteria. The first is the large variation in density of TCs along the *rrn* operons (4), which we have previously ascribed to a correspondingly large variation in transcription elongation rate (5). The second is the greatly reduced time for RNA polymerase to transit the operon in response to inactivation of four of the seven *rrn* operons in the *E. coli* genome (1). With the starting observation that the GC content peaks in the 16S and 23S encoding regions of the *rrn* operons (Fig. 2 *A*), we searched for an explanation for the polymerase density variation (4) within the physical properties of the DNA template sequence. For this, we designed a scheme for Monte Carlo simulations of transcribing RNA polymerases with thermodynamic tools, initially formulated in the pioneering work by Yager and von Hippel (7) and subsequently refined and adapted to transcriptional kinetics (9,10).

The simulations include varying the rate of initiation at the P1 promoter, polymerase backward and forward tracking, and interpolymerase interference by inhibition of backward movement of a leading polymerase and forward movement of a trailing polymerase for any two edge-to-edge-positioned polymerases. We found the positions of the polymerase

density peaks and valleys to be robust, but the peak heights and valley depths to be sensitive to parameter choice. A major discovery is that our Monte Carlo modeling, in conjunction with multiple step backward tracking, can account not only for the experimentally observed large polymerase density variation (Fig. 2), but also for the observed polymerase transit time reduction in response to the initiation frequency increase in the *E. coli* strain with only three of the seven *rrn* operons intact (Fig. 3 *B*). In our modeling scenario, the transit time reduction is due to inhibition of long backtracking excursions by the leading polymerase in edge-to-edge-positioned RNA polymerases: the higher the initiation frequency, the more edge-to-edge polymerases on the *rrn* operons and the more effective the inhibition of backtracking events that lead to longer dwell times (Fig. 4 *A*). However, in the case of short backtracking excursions, it is the inhibition of forward movement of the trailing polymerase in edge-to-edge polymerase couples that dominates. The net result of these opposing effects is a monotonically decreasing operon transit time with increasing initiation frequency (Fig. 3 *B*), although the actual initiation frequency saturates gradually (Fig. 3 *A*). Similar reductions in transit time have been obtained in recent transcription simulations based on other and much shorter templates (12).

Previous simulations of *rrn* operon transcription, based on simple kinetics with ad hoc backward tracking at a small number of sites, show a small transit time reduction in response to increasing initiation frequency, rapidly overshadowed by a transit time increase upon further increase in initiation rate (20). In the Klumpp and Hwa model, the main effect of suppressed backward tracking appears to be reduction of multiple visits to backtracked states with the same transcript length (38). At the same time, the inhibition of forward movement of trailing polymerases plays a much more dominating role in the Klumpp and Hwa model than in our template-sequence-based Monte Carlo simulations.

Approach to the maximal rate of *rrn* operon initiation

During growth in rich medium, the initiation frequency at the P_1 promoter may be controlled by the varying concentration of free RNA polymerase, $[RNAP_f]$ (17), by the varying strength (k_{cat}/K_m) of the P_1 promoter at a constant concentration of free RNA polymerase according to an unknown mechanism (19), or by a combination of these two effects. The model presented here simulates the rapid growth in rich medium, where only the P_1 promoter is active and where the initiation frequency is controlled by the initiation rate per empty promoter, k_I. Since $k_I = [RNAP_f](k_{cat}/K_m)$, this model is neutral to the choice of control mechanism.

For a next RNA polymerase to bind to the promoter, the previous polymerase must have moved ~50 basepairs from the promoter to leave sufficient space for another binding event. When there is no interference between RNA polymerases downstream from the promoter, the current initiation rate saturates with increasing k_I value in a Michaelis-Menten-like fashion (17). However, when polymerase queuing is taken into account in a simple stochastic model, the promoter saturates abruptly in a fashion that greatly deviates from Michaelis-Menten behavior (20). Our model, in contrast, suggests that initiation of transcription at *rrn* operons saturates in a gradual manner also in the presence of extensive polymerase interference. The gradual saturation behavior is in line with previous suggestions, but for reasons more complex than the simple Michaelis-Menten kinetics envisaged at the time (17,18).

Why do RNA polymerases backward track?

At first sight, backward tracking of RNA polymerases would appear to be detrimental to bacterial growth. It slows down transcription and leads to excess dissipation of energy by transcript cleavage and degradation (27,28). However, backtracking provides a mechanism for synchronization of the movements of RNA polymerases and ribosomes during transcription/translation of messenger RNAs (39). Moreover, it is not only trailing polymerases that can enhance the rate of transcription by blocking of polymerase backward tracking. During transcription of protein genes, trailing ribosomes can also provide a transcription-rate-enhancing function (39). Indeed, it has been suggested that blocking of polymerase backward tracking by trailing ribosomes provides a mechanism by which RNA polymerases and ribosomes move with equal speed during bacterial messenger RNA transcription and cotranscriptional translation. Although antitermination by Nus factors (6,22) accelerates *rrn* operon in relation to mRNA gene transcription, it is possible that backward tracking during *rrn* operon transcription spuriously reflects design principles that have evolved to optimize mRNA gene transcription.

There is yet another role for backward tracking, of perhaps even greater importance for mRNA and *rrn* operon transcription. It follows from the observation that cleavage of transcripts containing an erroneously incorporated base is favored over cleavage of correct transcripts. This makes backward tracking the vehicle for detecting erroneously inserted nucleotides in transcription, thereby significantly enhancing the accuracy of base selection (28). Indeed, in our model, frequent backward tracking greatly increases the base selectivity in relation to the accuracy provided by initial nucleotide selection (40; H. Mellenius, unpublished results).

Other putative causes of transcriptional rate variations

Nascent RNA chains may form hair pin structures which induce pausing of transcribing RNA polymerases (34).

Such structures are essential for regulation of amino acid synthesis by ribosome-dependent attenuation of transcription (41,42). To examine the possibility that hairpin structures contribute to the experimentally observed transcription rate variations (4), we searched the *rrn* operons for RNA hairpin structures, similar to those used in attenuation of transcription (37). These transcriptional pauses depend on the distance between the hairpin loop and the catalytic site of the polymerase, the nucleotide to be incorporated next, and also the downstream DNA context. Using the first two criteria, we identified a set of putative transcriptional pause sites (see Supporting Material for details).

The lengths of the pauses were assigned as proportional to the free energy of the stem of the hairpin loop (37). We found that the presence of these hairpins does not reproduce the local rate variation of *rrn* operon transcription at the relatively low initiation frequency of the wild-type strain. Moreover, the inclusion of hairpin-induced pausing leads to greatly increased operon transit time upon increasing initiation frequency, in contrast to experimental observation (4). We therefore conclude that structures similar to those employed in the mechanism of attenuation of transcription cannot account for the experimental data on local transcription rate variation (1) and decreased operon transit time in response to increased initiation frequency (4). Explaining the experimental data by hairpin-induced pauses would require a number of ad hoc assumptions with little experimental support.

SUMMARY

The thermodynamic modeling of transcription of *rrn* operons presented here is a first attempt to explain some of the unusual features that emerged from inspection of the electron microscopic images of *rrn* operons in wild-type bacteria and bacteria with four of the seven *rrn* operons inactivated by deletion (1,3–5). The stochastic simulations, based on standard free-energy variation of the transcription bubble as it moves along the DNA template (7), reproduced major parts of the RNA polymerase density profile observed along *rrn* operons as well as the decrease in the operon transit time observed at increased frequency in transcription initiation.

The model did not reproduce entirely the varying pattern of transcription elongation across the operon; the most notable deviations are in the central to distal region of the 16S gene and near the end of the operon where elongation is predicted to accelerate to a rate of over 400 nt s^{-1} (Fig. 2 *E*). Moreover, in the deletion strain, there are discrepancies between the predicted (Fig. 2 *F*) and the observed patterns (3,4) of polymerase density across the operon. Specifically, the simulations in their current state of sophistication have been unable to reproduce the pattern of polymerase queuing and congestion that occurs in the leader and intergenic regions of the operon in the mutant strain with only three active *rrn* operons. The queues are presumably the result of long transcriptional pauses (with durations of ~0.56 s) that occur near the beginning of the 16S and 23S RNA genes (3). The nature and significance of these pauses with respect to *rrn* operon transcription and ribosome assembly are unknown and at present, there is no theory that allows for their inclusion in a thermodynamic framework. Further enhancement and application of this model has the potential to promote a deeper understanding DNA sequence evolution and transcription rate variation along the DNA template.

SUPPORTING MATERIAL

Three tables, 12 figures, and details of the models and calculations are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(13)05746-9.

We thank Hans Bremer for valuable discussions and comments.

This work was supported by the Swedish Research Council, Vinnova, and the Knut and Alice Wallenberg foundation through the RiboCORE consortium.

REFERENCES

1. Condon, C., S. French, …, C. L. Squires. 1993. Depletion of functional ribosomal RNA operons in *Escherichia coli* causes increased expression of the remaining intact copies. *EMBO J.* 12:4305–4315.
2. Bremer, H., and P. P. Dennis. 1996. Modulation of chemical composition and other parameters of the cell at different exponential growth rates. *In Escherichia coli* and *Salmonella*: Cellular and Molecular Biology. F. C. Neidhardt, editor. ASM Press, Washington, D.C., pp. 1553–1569.
3. Ehrenberg, M., P. P. Dennis, and H. Bremer. 2010. Maximum *rrn* promoter activity in *Escherichia coli* at saturating concentrations of free RNA polymerase. *Biochimie.* 92:12–20.
4. Quan, S., N. Zhang, …, C. L. Squires. 2005. Transcriptional polarity in rRNA operons of *Escherichia coli* nusA and nusB mutant strains. *J. Bacteriol.* 187:1632–1638.
5. Dennis, P. P., M. Ehrenberg, …, H. Bremer. 2009. Varying rate of RNA chain elongation during rrn transcription in *Escherichia coli*. *J. Bacteriol.* 191:3740–3746.
6. Zellars, M., and C. L. Squires. 1999. Antiterminator-dependent modulation of transcription elongation rates by NusB and NusG. *Mol. Microbiol.* 32:1296–1304.
7. Yager, T. D., and P. H. von Hippel. 1991. A thermodynamic analysis of RNA transcript elongation and termination in *Escherichia coli*. *Biochemistry.* 30:1097–1118.
8. Greive, S. J., and P. H. von Hippel. 2005. Thinking quantitatively about transcriptional regulation. *Nat. Rev. Mol. Cell Biol.* 6:221–232.
9. Bai, L., A. Shundrovsky, and M. D. Wang. 2004. Sequence-dependent kinetic model for transcription elongation by RNA polymerase. *J. Mol. Biol.* 344:335–349.
10. Tadigotla, V. R., D. O'Maoileidigh, …, A. E. Ruckenstein. 2006. Thermodynamic and kinetic modeling of transcriptional pausing. *Proc. Natl. Acad. Sci. USA.* 103:4439–4444 (Erratum in *Proc. Natl. Acad. Sci. USA*. 2006. 203:7198).
11. Maoiléidigh, D. O., V. R. Tadigotla, …, A. E. Ruckenstein. 2011. A unified model of transcription elongation: what have we learned from single-molecule experiments? *Biophys. J.* 100:1157–1166.

12. Costa, P. R., M. L. Acencio, and N. Lemke. 2013. Cooperative RNA polymerase molecules behavior on a stochastic sequence-dependent model for transcription elongation. *PLoS ONE.* 8:e57328.

13. Rudd, K. E. 2000. EcoGene: a genome sequence database for *Escherichia coli* K-12. *Nucleic Acids Res.* 28:60–64.

14. SantaLucia, Jr., J., H. T. Allawi, and P. A. Seneviratne. 1996. Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry.* 35:3555–3562.

15. Bilgin, N., F. Claesens, …, M. Ehrenberg. 1992. Kinetic properties of *Escherichia coli* ribosomes with altered forms of S12. *J. Mol. Biol.* 224:1011–1027.

16. Kampen, N. G. v. 1992. Stochastic Processes in Physics and Chemistry. North-Holland, Amsterdam, New York.

17. Dennis, P. P., M. Ehrenberg, and H. Bremer. 2004. Control of rRNA synthesis in *Escherichia coli*: a systems biology approach. *Microbiol. Mol. Biol. Rev.* 68:639–668.

18. Record, M. T. J., W. S. Reznikoff, …, P. J. Schlax. 1996. *Escherichia coli* RNA polymerase (Eσ70), promoters, and the kinetics of the steps of transcription initiation. *In Escherichia coli* and *Salmonella*: Cellular and Molecular Biology. F. C. Neidhardt, editor. ASM Press, Washington, D.C., pp. 792–821.

19. Klumpp, S., and T. Hwa. 2008. Growth-rate-dependent partitioning of RNA polymerases in bacteria. *Proc. Natl. Acad. Sci. USA.* 105: 20245–20250.

20. Klumpp, S., and T. Hwa. 2008. Stochasticity and traffic jams in the transcription of ribosomal RNA: intriguing role of termination and antitermination. *Proc. Natl. Acad. Sci. USA.* 105:18159–18164.

21. Zhang, X., and H. Bremer. 1996. Effects of Fis on ribosome synthesis and activity and on rRNA promoter activities in *Escherichia coli*. *J. Mol. Biol.* 259:27–40.

22. Vogel, U., and K. F. Jensen. 1995. Effects of the antiterminator BoxA on transcription elongation kinetics and ppGpp inhibition of transcription elongation in *Escherichia coli*. *J. Biol. Chem.* 270:18335–18340.

23. Nudler, E. 2009. RNA polymerase active center: the molecular engine of transcription. *Annu. Rev. Biochem.* 78:335–361.

24. Svetlov, V., and E. Nudler. 2009. Macromolecular micromovements: how RNA polymerase translocates. *Curr. Opin. Struct. Biol.* 19:701–707.

25. Vassylyev, D. G., M. N. Vassylyeva, …, I. Artsimovitch. 2007. Structural basis for transcription elongation by bacterial RNA polymerase. *Nature.* 448:157–162.

26. Komissarova, N., and M. Kashlev. 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.

27. Laptenko, O., J. Lee, …, S. Borukhov. 2003. Transcript cleavage factors GreA and GreB act as transient catalytic components of RNA polymerase. *EMBO J.* 22:6322–6334.

28. Toulmé, F., C. Mosrin-Huaman, …, A. R. Rahmouni. 2000. GreA and GreB proteins revive backtracked RNA polymerase in vivo by promoting transcript trimming. *EMBO J.* 19:6853–6859.

29. Yarnell, W. S., and J. W. Roberts. 1999. Mechanism of intrinsic transcription termination and antitermination. *Science.* 284:611–615.

30. Sugimoto, N., S. Nakano, …, M. Sasaki. 1995. Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes. *Biochemistry.* 34:11211–11216.

31. Lesnik, E. A., R. Sampath, …, D. J. Ecker. 2001. Prediction of rho-independent transcriptional terminators in *Escherichia coli*. *Nucleic Acids Res.* 29:3583–3594.

32. Orosz, A., I. Boros, and P. Venetianer. 1991. Analysis of the complex transcription termination region of the *Escherichia coli* rrnB gene. *Eur. J. Biochem.* 201:653–659.

33. Adelman, K., A. La Porta, …, M. D. Wang. 2002. Single molecule analysis of RNA polymerase elongation reveals uniform kinetic behavior. *Proc. Natl. Acad. Sci. USA.* 99:13538–13543.

34. Artsimovitch, I., and R. Landick. 2000. Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc. Natl. Acad. Sci. USA.* 97:7090–7095.

35. Sydow, J. F., F. Brueckner, …, P. Cramer. 2009. Structural basis of transcription: mismatch-specific fidelity mechanisms and paused RNA polymerase II with frayed RNA. *Mol. Cell.* 34:710–721.

36. Wang, D., D. A. Bushnell, …, R. D. Kornberg. 2009. Structural basis of transcription: backtracked RNA polymerase II at 3.4 Å resolution. *Science.* 324:1203–1206.

37. Chan, C. L., and R. Landick. 1993. Dissection of the his leader pause site by base substitution reveals a multipartite signal that includes a pause RNA hairpin. *J. Mol. Biol.* 233:25–42.

38. Klumpp, S. 2011. Pausing and backtracking in transcription under dense traffic conditions. *J. Stat. Phys.* 142:1252–1267.

39. Proshkin, S., A. R. Rahmouni, …, E. Nudler. 2010. Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science.* 328:504–508.

40. Mellenius, H., and M. Ehrenberg. 2012. Large DNA template dependent error variation during transcription. *In* Biophysics and Structure to Counter Threats and Challenges. J. D. Puglisi and M. V. Margaris, editors. Springer, New York, pp. 45–66.

41. Landick, R., C. L. Turnbough, and C. Yanofsky. 1996. Transcription attenuation. *In Escherichia coli* and *Salmonella*: Cellular and Molecular Biology. F. C. Neidhardt, editor. ASM Press, Washington, D.C., pp. 1263–1286.

42. Elf, J., and M. Ehrenberg. 2005. What makes ribosome-mediated transcriptional attenuation sensitive to amino acid limitation? *PLOS Comput. Biol.* 1:e2.