J Stat Phys (2011) 142: 1252–1267
DOI 10.1007/s10955-011-0120-3

# Pausing and Backtracking in Transcription Under Dense Traffic Conditions

**Stefan Klumpp**

Received: 8 November 2010 / Accepted: 6 January 2011 / Published online: 15 January 2011


**Abstract** RNA polymerases transcribe the genetic information from DNA to RNA. They move along the DNA by stochastic single-nucleotide steps that are interrupted by pauses. Here we use a variant of driven lattice gas models or exclusion processes to study the effects of these pauses under conditions, where many RNA polymerases transcribe the same gene. We consider elemental pauses, where RNA polymerases are inactive and immobile, and backtracking pauses, during which RNA polymerases translocate backwards in a diffusive fashion. Under single-molecule conditions, backtracking can lead to complex dynamics due to a power-law distribution of the pause durations. Under conditions of dense RNA polymerase traffic, as in the highly transcribed genes coding for ribosomal RNA and transfer RNA, backtracking pauses are strongly suppressed because the trailing active RNA polymerase restricts the space available for backward translocation and ratchets the leading backtracked RNA polymerase forward. We characterize this effect quantitatively using extensive computer simulations. Furthermore, we show that such suppression of pauses may have a regulatory role and lead to highly cooperative control functions when coupled to transcription termination.

**Keywords** Transcription · RNA polymerase · Traffic models · Exclusion process · Pausing · Backtracking · Cooperativity

## 1 Introduction

The genetic information stored in the DNA of a cell is read out in a multi-step process that starts with transcription, the copying of the nucleotide sequence of a gene from the DNA into a complementary RNA molecule. This process is carried out by RNA polymerases (RNAPs) that bind to DNA at start sequences for transcription, the promoters, and then move along the gene in a sequence of single-nucleotide steps that are coupled to nucleotide-by-nucleotide synthesis of the RNA transcript, until they reach a termination site at the end of a gene or

S. Klumpp (✉)
Max Planck Institute of Colloids and Interfaces, Science Park Golm, 14424 Potsdam, Germany
e-mail: klumpp@mpikg.mpg.de

an operon, a group of genes that are transcribed together [48]. Transcription is also a major point of control and the read-out of a gene is regulated mainly by controlling the rate at which RNAPs bind to the promoter, but there are also regulatory mechanisms that act on the level of transcript elongation [48].

The process of transcript elongation is not a continuous process, but rather interspersed with pauses, as has been observed directly in single-molecule experiments [2, 10, 37]. Several mechanisms for pausing have been distinguished by how they are affected by external force, by the effects of elongation-modulating factors, and by the required DNA sequences [3, 37]. The two types of pauses that appear to be most frequent are backtracking pauses, during which an RNAP translocates backwards on the DNA, and short 'ubiquitous' or 'elemental' pauses without such backward movement [21, 37]. The short pauses are also believed to represent an intermediate state during the establishment of stable pause types such as backtracking, but also pauses stabilized by protein-DNA contacts or by structures in the nascent RNA [31]. During backtracking pauses, RNAPs move in a diffusive fashion with (on average) equal rates for forward and backwards steps [48]. During these steps, the transcript is neither elongated nor shortened, therefore the end of a nascent RNA is not correctly aligned with the active site of the RNAP during backtracking. To reactivate, RNAPs either have to return to the position where backtracking was initiated or cleave the transcript at the active site [48].

In the cell, specific pauses may have particular regulatory roles [41], but the main role of pausing in general appears to be setting the overall speed of transcript elongation, which in bacteria is adapted to the speed of translation [40], but may also need to be coordinated with other processes such as RNA processing and folding [32].

If the rate of transcription initiation is sufficiently high, multiple RNAPs can transcribe the same gene at the same time, a situation reminiscent of dense traffic on a highway. Dense traffic of RNAPs can be visualized by electron microscopy (Miller spreads) [8, 36] or by chromatin immunoprecipitation [19]. The similarities to highway traffic have triggered a number of theoretical studies [13, 24, 45] that addressed the dynamics of transcription using traffic models based on exclusion processes from non-equilibrium statistical physics [42]. In addition to their central role in the study of non-equilibrium phase transitions [12, 29, 43] and in physical traffic modeling [7], such models had previously also been used to study other biological processes such as translation [6, 14, 35, 44] and transport along cytoskeletal filaments [22, 34].

It is however important to note that in bacterial transcription, most genes are actually transcribed at rather low rates, and RNAP densities on these genes are therefore also low [4, 33]. An important exception are the genes encoding ribosomal RNA (rRNA) and transfer RNA (tRNA), which are transcribed at high densities ($\approx$50% coverage has been observed for rRNA transcription in fast growing wild type Escherichia coli cells and even higher densities in strains with reduced number of rRNA genes [8]). The high transcription rates of rRNA and tRNA are physiologically required because the RNA products of these genes are not translated and therefore the amount of product of the gene is not amplified by the translation process, as it is for protein-coding genes. Furthermore, fast growing cells require large concentrations of the products of these genes, in particular of ribosomes, as the ribosomes synthesize protein, which accounts for about 50 percent of the biomass produced by growing bacterial cells [5]. One constraint that arises from the requirement for high transcription rates on these genes, is that transcription should not be limited by transcript elongation to rates less than the physiologically required ones. Our previous theoretical analysis indicated that this constraint requires that transcriptional pausing is suppressed in transcription of rRNA [24], which indeed happens in bacterial cells through the activity of a so-called antitermination mechanism that speeds up transcript elongation of these genes [46]. It is also

noteworthy that, even though rRNA and tRNA genes are exceptional in the sense that they account only for a small fraction of all the genes in the cell, their transcription actually dominates the overall cellular transcription and accounts for more than 70 percent of the total cellular transcription in a rapidly growing bacterium [25].

In this article, we are interested in the effects of pausing on the transcription rate as well as in the effect of the transcription rate on pauses. In particular, backtracking pauses may be influenced by the presence of a trailing RNAP that restricts how far a paused RNAP may backtrack. This has been demonstrated experimentally, both in vitro and in vivo [16, 17]. Also, our previous theoretical study has indicated that in a simple exclusion model, backtracking is almost completely suppressed under dense (saturated) traffic conditions, while pauses without backtracking may have a dominating effect on the transcription rate [24]. Here we report on a simulation study that extends the previous study and addresses the mechanism of the suppression of backtracking as well as sequence effects and the variability of the results resulting from those sequence effects. In addition, we use these results to explore how the suppression of backtracking can be used to sense the rate of transcript initiation and how this signal may be processed in control elements regulating transcription.

## 2 Model for Transcript Elongation

### 2.1 Dynamics of a Single RNAP

We model the transcription process by stochastic stepping of particles representing the RNAPs on a one-dimensional lattice (Fig. 1). Each lattice site corresponds to a particular nucleotide position on a gene or operon and RNAPs move along this lattice in single nucleotide steps, which occur with rate $\epsilon$ and describe forward translocation by the step size $s = 1$ nt, coupled to transcript elongation. The stepping rate $\epsilon$ is also referred to as the elongation attempt rate, because stepping may not succeed in the presence of exclusion interactions (Sect. 2.3). In addition to the forward steps, RNAPs can also undergo transitions to a paused state with rate $f$, the pause frequency. This transition is taken to occur only at specific pause sites, as we will discuss in Sect. 2.2 and the rate $f$ may have different values at pause site

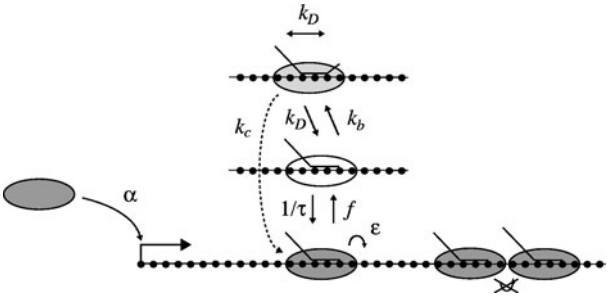


**Fig. 1** Model for RNAP traffic: RNAPs are described as extended particles moving on a one-dimensional lattice representing the DNA template. Active RNAPs (*dark grey*) enter the system with rate $\alpha$ (provided the first 50 sites are vacant) and step forward, elongating the transcript with rate $\epsilon$, unless the site ahead is occupied by another RNAP. At pause sites, RNAP enter the inactive paused state (*white*) with rate $f$, from which they return to the active state with rate $1/\tau$. At backtracking sites, they may also enter the backtracked state (*light grey*) by making a backward step with rate $k_b$. Backtracked RNAPs make forward or backward steps that are not coupled to transcript elongation or shortening with rate $k_D$. They may also cleave the transcript with rate $k_c$ and return to the active state at their current position

without backtracking and backtracking sites. From the paused state, an RNAP reactivates with rate $1/\tau$, where $\tau$ is the average pause duration. At some pause sites, the backtracking sites, paused RNAPs may also start backtracking by making a backward step with rate $k_b$. Once in the backtracked state, RNAPs make forward and backward steps with equal rates, $k_D$, which corresponds to a diffusion coefficient for the diffusive movement in the backtracked state. We note that these steps are not coupled to elongation or shortening of the transcript, therefore the end of the transcript is not aligned with the active site of the RNAP during backtracking, but is protruded from the RNAP, as indicated by the piece of RNA that sticks out of the RNAP in Fig. 1. An RNAP can return to its active state by returning to the position where it initiated backtracking. In that case, the RNA is again correctly aligned. In addition, with rate $k_c$, it may return to the active state directly from the backtracked state at any position by cleaving the protruded part of the transcript, such that it is again correctly aligned. In this scenario, the RNAP has effectively moved backwards and shortened the transcript during the backtracking pause and needs to re-synthesize the cleaved part of the transcript. In vitro, transcript cleavage is very slow and may be neglected, but it is likely to be the dominant pathway for RNAP reactivation in vivo, where cleavage is stimulated by additional protein factors (the Gre proteins) [48].

We note that in principle all rates may depend on the local DNA sequence, which influences the stepping rates through the binding energies between RNAP, DNA, and RNA [48]. This sequence-dependence is neglected in our model, except for the fact that pauses occur at specific pause sites (Sect. 2.2). Likewise, we neglect effects of the chemomechanical cycle of the individual step. In principle such effects could be important if, e.g., the rate-limiting step of that cycle is different under different traffic densities, and if the concentration of nucleotides is considered as another control parameter of the system. In our case, these simplifications are based on the observations that the stochasticity of RNAP movements is dominated by pausing [24], so the details of the steps should be less important.

The rates for the reactions used in our simulations are summarized in Table 1. How these parameters were estimated is described in detail in our previous study [24]. Briefly, some of these parameters have been measured directly in single-molecule experiments ($f$, $\tau$) [2, 37] and are taken to also apply to the in vivo situation, as the measured average elongation speed is consistent with them. The stepping rate under single molecule conditions is typically much smaller than in vivo, and we use an estimate based on the elongation speed measured under conditions where pausing is suppressed [24, 46]. The stepping rate for steps of backtracked RNAPs is taken to be approximately ten-fold smaller than the stepping rate

**Table 1** Parameters of the model and values used in the simulations

| Parameter | Symbol | Value | Reference |
|---|---|---|---|
| Stepping (or elongation attempt) rate | $\epsilon$ | 100 $s^{-1}$ | [46] |
| Step size | $s$ | 1 nt | [1] |
| Pause frequency (at pause sites) | $f$ | 10 $s^{-1}$ | [2, 37] |
| Pause frequency (at backtracking sites) | $f_{\mathrm{bt}}$ | 50 $s^{-1}$ | [24] |
| Pause density | $\rho_p$ | 0.01 | [2, 21, 37] |
| Pause duration (without backtracking) | $\tau$ | 1 s | [2, 21, 37] |
| First backtracking stepping rate | $k_b$ | 10 $s^{-1}$ | see text |
| Backtracked stepping rate | $k_D$ | 10 $s^{-1}$ | [1] |
| Transcript cleavage rate | $k_c$ | 0.05 $s^{-1}$ | see text |
| Size of RNAP footprint | L | 50 nt | [24] |

of elongating RNAPs as observed in single-molecule experiments [1], in addition, we take the first backtracking step (when allowed to occur, i.e. at backtracking sites) to occur with the same rate as other backtracking steps by setting $k_b = k_D$. The cleavage rate is chosen such that RNAPs backtrack on average over $\sqrt{k_D/k_c} \approx 10$–$15$ nt before transcript cleavage, which corresponds to backtracking pause of 10–20 s.

### 2.2 Sequence-Dependence of Pauses

Whether and where an RNAP pauses depends on the local sequence of the transcribed DNA. While backtracking pauses are strongly sequence-dependent, short pauses without backtracking have been observed to occur almost randomly with a constant density along the DNA [2, 37]. More detailed analysis of such pauses showed however that also these pauses occur at specific positions [21], but that RNAPs pause only with a certain probability (the pause efficiency) at any such position. In contrast to our previous study, where we treated short pauses as occurring randomly at any position and backtracking pauses as restricted to backtracking sites [24], here we treat all pauses as sequence-dependent.

At the beginning of each simulation, we generate a sequence by designating each position along the lattice as a pause site with probability $\rho_p$, the pause density, and each pause site as a backtracking site with probability $\rho_{\mathrm{bt}}$. While we allow the number of pause sites to vary between simulations for different sequences, keeping only its average value fixed by the pause density, we impose a constraint that fixes the number of backtracking sites to a defined small integer value. In simulations where we allow pausing at all positions, we use a pause frequency that is related to the rate for entry to the paused state at a pause site by $f_{\mathrm{rand}} = f\rho_p$.

### 2.3 RNAP Traffic

Finally, interactions between RNAPs are described by a simple exclusion rule that forbids stepping to sites occupied by another RNAP. RNAPs are taken to occupy $L = 50$ nt on the DNA template, i.e. the presence of one RNAP makes 50 lattice sites unavailable to other RNAPs [24]. This exclusion rule applies to active elongation, where it tends to slow down RNAPs in dense traffic, as well as to backtracking, where it tends to restrict how far RNAPs may backtrack and thereby acts to suppress backtracking pauses. This exclusion rule is the simplest possible way to model RNAP-RNAP interactions, and it is possible that other interactions exist. For example, an elongating RNAP may actively push forward a backtracked RNAP, which will further increase the effect of suppression of backtracking. As in our previous study [24], we also assume that short pauses without backtracking are unaffected by the presence of a trailing RNAP that might push on the paused RNAP, because these pauses were unaffected by mechanical force in single-molecule experiments [37].

In the simulations, RNAPs enter the lattice with rate $\alpha$, provided that the first 50 sites are vacant. $\alpha$ is called the initiation attempt rate, as it describes the rate at which RNAPs attempt to initiate transcription. The real initiation rate is smaller, because some of these attempts are unsuccessful due to elongating RNAPs blocking the access to the first 50 sites. When an RNAP reaches the end of the system (gene or operon), which in all our simulations has a size of 3000 nt, it leaves the systems quickly, as there is no evidence that this step may be limiting under realistic conditions.

The RNAP traffic is simulated with the following method [26]: At each basic time step, corresponding to $5 \times 10^{-3}$ s, we perform $m$ Monte Carlo moves, where $m$ is the number of lattice sites in the simulated system. In each move a lattice site is selected randomly. If the

site is occupied by the active site position of a particle, it is updated according to the rates described above and the corresponding transition is performed unless the exclusion rule does not allow it. If the chosen site is the first site of the lattice, a new particle is introduced with rate $\alpha$ if the first 50 sites are vacant. If the last site is chosen and a particle is present there, it is removed from the system. If the chosen site is empty or merely blocked by a particle, i.e. it is covered by an RNAP, but not the position of the active site of that RNAP, nothing happens. These simulations are typically run for $5 \times 10^6$ time steps.

## 3 Backtracking of RNAPs Under Single-Molecule Conditions

Before we start to discuss backtracking in dense RNAP traffic, we make a few comments on backtracking of individual RNAPs. The fact that backtracking is a diffusive process with on average equal stepping rates for forward and backward steps [48] is intriguing from the theoretical point of view, in particular in the in vitro situation, where transcript cleavage may be neglected. In that case, reactivation of an RNAP is dependent on its return to the position where it started backtracking, which from a theoretical point of view corresponds to the return of a random walker to the origin in one dimension, a classical problem in stochastic dynamics [39, 49]. As a consequence of this type of reactivation dynamics, the distribution of the durations $\tau_{\rm bt}$ of a backtracking pause has a power-law tail,

$$P(\tau_{\rm bt}) \sim \tau_{\rm bt}^{-3/2}, \tag{1}$$

and the average duration of a backtracking pause diverges.

Such a dependence, which has been pointed out by several groups [11, 18, 47], has indeed been observed in single-molecule experiments, but the origin of that dependence is still controversial. In one case (with eukaryotic RNAP II), this observation has been taken as evidence for backtracking as the dominant mechanism of pausing [18], while in another case (with bacterial RNAP), this distribution was interpreted as arising from pauses without backtracking due to averaging over pause sites with different pause durations. Indeed, this averaging complicates the interpretation of the experimental distributions, and data for pause time distributions for specific pauses sites seem to be consistent with exponential distributions [21, 37]. However, available data for specific pause sites is very limited and some of the arguments of Ref. [37] have recently been challenged and it has been proposed that at least some of the short pauses may in fact be backtracking pauses where RNAPs only translocate backwards over very short distances, e.g. 1 or 2 nt [11]. This question will be discussed further elsewhere (Klumpp and Hwa, in preparation). Here we simply note that, while there is agreement that pauses without backtracking exist, it remains controversial what their relative frequency is.

Another consequence of the diffusive dynamics is that, for long times, elongation will proceed slower than linearly with time. This can be seen using a simple scaling argument: Let us assume that RNAPs initiate transcription at time $t = 0$. At time $t$, RNAPs typically have backtracked only over a distance $\sim \sqrt{k_D t}$. The reactivation rate, which is proportional to the probability that the RNAP is at the origin of backtracking, is therefore $\approx 1/(\tau\sqrt{k_D t})$. Using this estimate of the reactivation rate, we can estimate the effective elongation speed from the probability that an RNAP is active as

$$u \simeq \frac{\epsilon s}{1 + f_{\rm bt}\tau\sqrt{k_D t}} \sim 1/\sqrt{t}. \tag{2}$$

As a consequence the average position of the RNAP will increase with time as $x \sim \sqrt{t}$ for long times. We note that essentially the same sublinear dynamics has been predicted for the combination of diffusive and active movements of cytoskeletal molecular motors in quasi-two dimensional geometries where however the diffusive excursions take place along the direction perpendicular to the active directed movements [34, 38].

This sublinear behavior should be observed for elongation times longer than a cross-over time $t_*$, which can be calculated by equating the short-time and long-time limits of (2). This calculation leads to

$$t_* = \frac{1}{(f_{\mathrm{bt}}\tau)^2 k_D} \tag{3}$$

and shows that $t_*$ is proportional to the square of the frequency $f_{\mathrm{bt}}$ of backtracking pauses. Whether the sublinear behavior is observable in vitro therefore depends strongly on the value of this parameter, which is subject to controversies as discussed above. If most pauses seen in single-molecule experiments are indeed backtracking pauses [11], we can estimate $t_*$ to be approximately 100 s (using $\tau = 1$ s, $k_D = 1\ \mathrm{s}^{-1}$, $f_{\mathrm{bt}} = 0.1\ \mathrm{s}^{-1}$ [1, 37]—note that these values are for the in vitro situation and different from those given in Table 1). During that time, an RNAP transcribes about 1000 nt, so a transcript length of a few 1000 nt should be sufficient to observe the sublinear dynamics. These time and length scales are accessible to such experiments. On the other hand, if backtracking pauses are rare [31, 37] (with $f_{\mathrm{bt}} = 0.01\ \mathrm{s}^{-1}$), $t_*$ is $10^4$ s and an unrealistically long transcript ($10^5$ nt) would be required for the observation of this effect. It may however be possible to increase the frequency of backtracking pauses by tuning experimental conditions or by using mutant RNAPs that pause more frequently. Finally, we want to emphasize again that these dynamic phenomena, although appealing from a statistical physics point of view and possibly observable under single-molecule conditions, are likely to have limited relevance for transcription in its cellular context, where extended backtracking pauses are typically terminated by transcript cleavage.

## 4 Backtracking in Dense RNAP Traffic

### 4.1 Effects of Initiation Rate and Pauses on RNAP Traffic

We now turn to transcription under conditions of high transcription rates or dense RNAP traffic. To address this issue, we first used the model as described above to determine how the main characteristics of RNAP traffic, i.e. the transcription rate and the elongation speed depend on the initiation attempt rate. The results are shown in Fig. 2A and B, respectively. These results are very similar to those of our previous work [24], where we have treated pauses without backtracking as sequence-independent. There are two regimes, an initiation-limited regime, where the transcription rate $J$ is determined by the initiation attempt rate $\alpha$, and an elongation-limited regime where the elongation process, i.e. stepping and pausing determines the transcription rate, independent of $\alpha$. In that regime, the transcription rate has the maximal value that can be achieved. The main topic of our previous study [24] was how the cell increases this maximal value in the case of rRNA transcription where large transcription rates are required to allow for rapid cell growth. The existence of the two regimes is typical for this type of models and it is well-established through a number of heuristic and exact methods for the asymmetric simple exclusion process, the paradigm for this class of models, that the two regimes are separated by a boundary-induced non-equilibrium phase

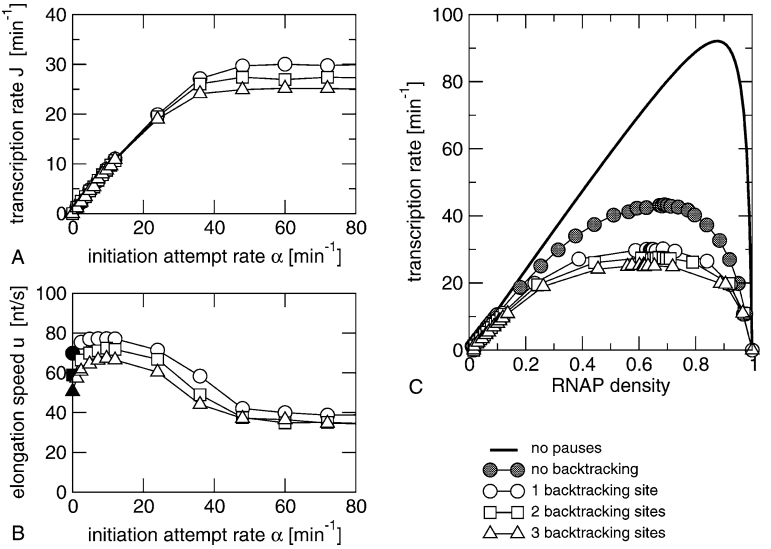


**Fig. 2** (**A**) Transcription rate and (**B**) elongation speed as functions of the initiation attempt rate for systems with 1, 2, or 3 backtracking sites (in a 3000 nt sequence). The *black symbols* at $\alpha = 0$ in (**B**) indicate the elongation speed for a single RNAP. (**C**) Current-density plots or fundamental diagram of RNAP traffic: The transcription rate as a function of the RNAP density for different pausing scenarios. For comparison, the corresponding curves for the case without backtracking pauses and without any pauses are also shown

transition [12, 28, 29, 43]. In the exclusion process literature the two regime are usually referred to as the low-density phase and the maximal-current phase. A third phase, the high-density phase is found when the current is limited by the exit from the system, i.e. in our case by the termination of transcription. There is however no evidence that this case might occur in transcription and we do not consider it here.

Figure 2A also shows that increasing the number of backtracking sites has no effects on the transcription rate at small $\alpha$, where transcription is initiation-limited, but decreases the maximal transcription rate in the elongation-limited regime. In that respect, the effect of additional backtracking pauses is the same as that of additional pauses without backtracking.

The elongation rate (Fig. 2B) exhibits non-monotonic behavior. It increases for small initiation attempt rates, but for larger initiation attempt rates it is reduced. The reduction at large $\alpha$ is due to jamming of RNAPs behind a paused RNAP which increases the effect of pauses [24]. The speed-up for small initiation attempt rates, on the other hand, is a result of shortened backtracking pauses, as backtracking becomes increasingly restricted due to the presence of a trailing RNAP. This speed-up corresponds to the increased elongation speed observed by Epshtein et al. [16]. In our model, the trailing RNAP does not actively push the leading backtracked RNAP forward, but rather progressively restricts the distance the leading RNAP may backtrack, thus 'ratcheting' it forward. Increasing the number of backtracking pause sites decreases the elongation speed, however the effect is stronger in the single-RNAP limit ($\alpha \approx 0$) than around the maximal elongation speed. The ratio of these two quantities, i.e. the relative speed-up is therefore larger with a larger number of backtracking pauses. This speed-up can be estimated in the following way: If backtracking is completely suppressed, the RNAP needs the time $t_0$ to transcribe the full sequence. Each backtracking site adds a time $T$, the duration of the backtracking pause, which need to be weighted with the probability $p$ that each RNAP backtracks at such a site. The relative

speed-up is then given by $t_0/(t_0 + npT)$. In our case, $t_0 \simeq 39$ s, $p \simeq 0.5$, and $T \simeq 10$ s, so this estimate predicts an increase of the speed of $n \times 13$ percent, slightly higher than what is seen in the simulation data, probably because backtracking is not completely suppressed (see Sect. 4.2).

The properties of any traffic system can be characterized by the current-density relation, the dependence of the current or flux on the density of particles. In our case, this relation is given by the density dependence of the transcription rate. This dependence is shown in Fig. 2C for the three systems of Fig. 2A and B as well as for cases without any pauses and without backtracking pauses (but with short 'ubiquitous' pauses). We note that the density $\rho$ used here is the coverage density, defined here as the fraction of sites that are unaccessible due to the presence of an RNAP. The density of RNAPs (particles/length of sequence) is then $\rho/L$. We also note that the simulation data for densities larger than the density at which the current has its maximum have been obtained by varying the rate at which particles leave the system at the right boundary. This case is not realistic for transcription, but is included here for completeness.

In the absence of any pausing, our model is reduced to the usual totally asymmetric simple exclusion process with extended particles [24]. For this case, many quantities are known analytically [30, 35, 44]. In particular, the current-density relation is given by

$$J(\rho) = \epsilon \frac{\rho(1-\rho)}{L(1-\rho+\rho/L)}, \tag{4}$$

which is plotted as the solid line in Fig. 2C. This curve shows that the maximal transcription rate is quite high (approximately 92 $\text{min}^{-1}$) and that this maximal transcription rate occurs for a high density, close to full coverage ($\rho \simeq 0.88$). The presence of pauses with or without backtracking reduces this maximal transcription rate and in our case, where there are more pause sites (without backtracking) than backtracking sites, the main reduction is due to pauses without backtracking. In addition, pausing also shifts the maximum towards lower densities, so that saturated RNAP traffic for different sequences corresponds to different RNAP densities, depending on the frequency of pauses for each sequence. It is also worth noting that saturated traffic in these cases is obtained for only about 60–70 percent coverage of the DNA.

### 4.2 Suppression of Backtracking in Dense Traffic

To study the suppression of backtracking pauses by trailing RNAPs in more detail, we next determined several quantities that characterize backtracking pauses. In these simulations we used a system that contained only a single backtracking site, such that only a single RNAP can be backtracked at any time. We varied again the initiation attempt rate and first determined the duration of the backtracking pause, from the first arrival of an RNAP at the backtracking site to when it steps beyond that site. As observed previously in [24], the duration of the backtracking pause is strongly reduced in dense traffic (Fig. 3A). For the parameters used here, the overall duration of that pause in saturated RNAP traffic is shorter by almost a factor 10 than under single RNAP conditions. Likewise we find a substantial reduction of the average backtracking distance (the average distance of a backtracked RNAP from the site where it entered the backtracked state) with increasing initiation attempt rate (Fig. 3B).

It is worth noting that the pause duration exhibits a very steep decrease as a function of $\alpha$. A reduction to 50 percent of the duration under single RNAP conditions is obtained already at initiation attempt rate of $<5$ $\text{min}^{-1}$. This is well within the linear initiation-limited

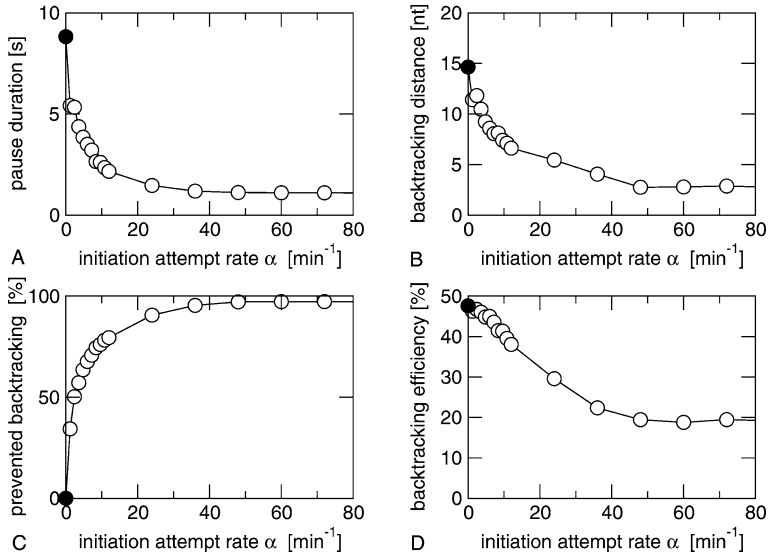


**Fig. 3** Characteristics of a backtracking pause: (**A**) The duration of a backtracking pause, (**B**) the average distance of a backtracked RNAP from the backtracking site, (**C**) the fraction of unsuccessful backtracking attempts and (**D**) the fraction of RNAPs that backtrack upon first arrival at the backtracking site as functions of the initiation attempt rate $\alpha$. Simulations were performed for a system with a single backtracking site. The *filled circles* at $\alpha = 0$ indicate the values of the pause characteristics in simulations of single-RNAP transcription

regime (Fig. 2), in which $J \approx \alpha$. Therefore substantial suppression of backtracking can not only be expected for genes characterized by dense traffic transcription such rRNA and tRNA genes, but also for mRNA genes with typical transcription rates of a few transcripts per minute. Indeed the experiment of Epshtein et al. [17] that demonstrated the suppression of backtracking in vivo was done with an mRNA promoter that, although it is strong compared to promoters of other mRNAs, should be weaker than rRNA promoters.

We next asked what the main mechanism for the shortened backtracking pauses is. In principle there are two non-exclusive possibilities: Backtracking attempts could be prevented by a trailing RNAP that occupies the site upstream of the RNAP attempting to enter the backtracked state. Alternatively, the RNAP may be able to enter the backtracked state, but remains there for a short time period, because it is ratcheted back to the active state by a trailing RNAP that catches up. We therefore determined the fraction of backtracking attempts that are prevented by the presence of a trailing RNAP (Fig. 3C). In the limit of a single RNAP this fraction vanishes, but it increases quickly with increasing initiation attempt rate. In dense traffic almost all attempts to enter the backtracked state are prevented.

In contrast to the strong effect on the fraction of successful backtracking attempts, we found however that the efficiency of the backtracking site, which is defined as the probability that an RNAP enters the backtracked state upon its first arrival at the backtracking site, is reduced only about 2.5-fold (Fig. 3D) in saturated RNAP traffic. The apparent inconsistency of these results can be resolved by noting that backtracked RNAPs may switch repeatedly between the backtracked state, the non-backtracked paused state, and the active state before escaping from the backtracking site. This switching between state is often denoted as 'oscillations' in the experimental literature and can include diffusive movements back and forth between the state as well as cycles of backtracking, transcript cleavage, and transcript

re-elongation. The observation that attempts to enter the backtracked state are strongly suppressed, while there is only a weak effect on the backtracking efficiency can then be understood as indicating that the main effect of the trailing RNAP is to prevent multiple cycles of backtracking and re-elongation or diffusive return to the backtracking site.

### 4.3 Sequence Effects and Variability of the Results

Next, we asked to what extent these results are affected by the particular sequence used in the simulation, i.e. by the particular location of pause sites with and without backtracking. To answer this question, we generated a large number of sequences with an average of 30 pause sites without backtracking and 1, 2 or 3 backtracking sites (208 sequences each) and simulated their transcription under saturated RNAP traffic conditions. Figure 4 shows the results. There is indeed considerable variability in the maximal transcription rate (Fig. 4A) and an even more pronounced variability in the elongation speed (Fig. 4B). The distribution of transcription rates is given by an almost symmetrical peak, which is shifted towards smaller values for increasing number of backtracking sites, in agreement with the data in Fig. 2. The smaller peak marked by the arrowhead will be discussed later.

By contrast, the distribution of the elongation speed (Fig. 4B) is only very weakly dependent on the number of backtracking sites. Although there is a shift towards lower speeds for large numbers of backtracking sites, this shift is very small compared to the width of the distribution, which is approximately the same for the different numbers of backtracking sites. The weak dependence of the elongation speed distribution on the number of backtracking sites indicates that the speed under dense traffic conditions is dominated by the short pauses without backtracking, as also indicated by the dependence of the elongation speed $u_{\min}$ on the initiation attempt rate (see Fig. 2 and also Ref. [24]). The variability of

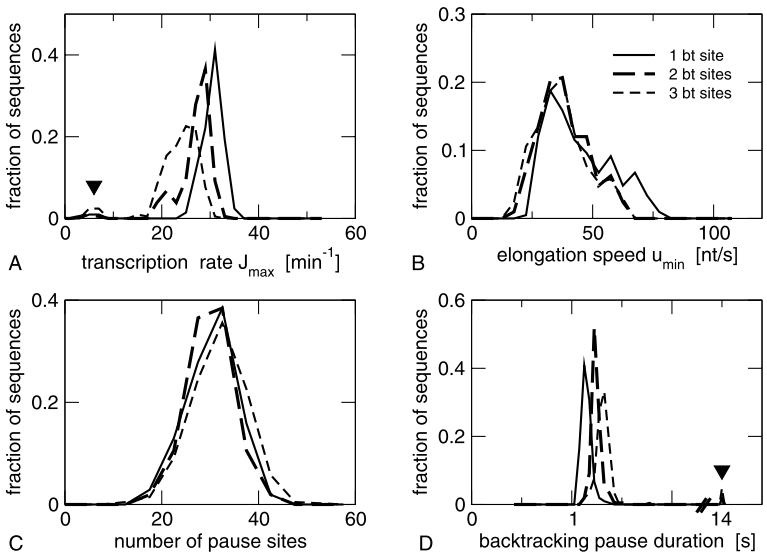


**Fig. 4** Sequence-to-sequence variation of quantities characterizing transcription: Histograms of (**A**) the maximal transcription rate, (**B**) the elongation speed, (**C**) the number of pause sites, and (**D**) the average duration of a backtracking pause under conditions of saturated traffic. The *arrowheads* mark cases with backtracking sites overlapping the promoter, i.e. backtracking sites located within the first 50 sites

the speed is a consequence of how we constructed the simulated sequences with randomly placed pause sites, which results in a broad distribution of the total number of pause sites along the simulated sequences (Fig. 4C).

Figure 4D shows the distribution of the durations of pauses at the backtracking sites. Setting aside the small peak marked by the arrowhead, one can clearly see that these distributions are dominated by a relatively narrow peak at a pause duration slightly above 1 s. Based on the characteristics of backtracking in dense traffic determined in the previous section, we can identify this peak as corresponding to the duration of a non-backtracked pause plus a small amount of backtracking. Figure 4D also shows that this peak is shifted towards slightly longer pause durations with increasing number of backtracking sites. This shift indicates that the suppression of backtracking becomes less efficient as the number of backtracking pauses is increased. Indeed if there are additional backtracking sites upstream of a backtracked RNAP, the trailing RNAP may be stuck there for some time and will therefore need a longer time to catch up with the leading RNAP.

Finally, let us address the small peaks marked by arrowheads in the distributions of the maximal transcription rate $J_{\max}$ (Fig. 4A) and of the pause duration (Fig. 4D). Inspection of the sequences used for the simulations that contribute to these peaks shows that these peaks are due to backtracking sites located within the first 50 sites of the system. If an RNAP pauses at such a site, it prevents initiation of transcription by other RNAPs as it blocks the promoter, which in our simulation is represented by the first 50 lattice sites. Backtracking can therefore not be suppressed by trailing RNAPs. In that case, the transcription rate is limited by this single long pause. A backtracking site overlapping the promoter should therefore be considered as a special case. However this special case may not be so rare in real biological situations. Promoter-proximal pausing is typical for transcription in the fruitfly, Drosophila [9] and there are known cases in bacterial transcription. The best studied such case is the late operon of phage lambda where the promoter-proximal pause site has a regulatory role in loading the antitermination factor Q [41], but additional cases have recently been found by systematic search for genes with such pauses [20].

## 5 Regulatory Roles

The suppression of backtracking due to the pushing of a trailing RNAP may simply be an unavoidable consequence of dense RNAP traffic at high transcription rates. In some cases, where very high transcription rates are physiologically required such as in the transcription of rRNA and tRNA, this effect may have a role in increasing the transcription rate, but in that case, pauses are also suppressed by an antitermination complex in order to avoid a limitation of the transcription rate by slow elongation [24, 25]. However even when transcription rates are smaller, the suppression of specific backtracking pauses may have a biological function. As an example, we consider a backtracking pause that is coupled to transcription termination [16], such that a conformational transition that happens while the RNAP is pausing tags it for termination at a transcription terminator. This transition may either be a step in the termination pathway itself or the binding of a protein that influences the termination probability at a downstream terminator. If this process occurs with rate $k$, the probability that the RNAP does not terminate, but reads through the terminator depends exponentially on the pause duration and is given by

$$P_{\mathrm{R}} = \exp(-kT(\alpha)), \tag{5}$$

where $T(\alpha)$ is the duration of the backtracking pause and depends on the initiation attempt rate. The transcription rate of a gene downstream of the terminator is then given by $J \times P_{\mathrm{R}}$,

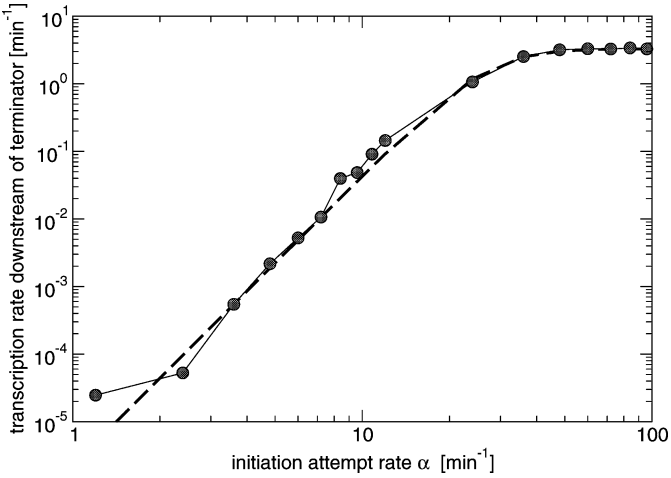


**Fig. 5** Regulatory role of the suppression of backtracking: Transcription rate downstream of a terminator coupled to a backtracking pause in such a way that a transition during the pause (with rate $k = 2\ \mathrm{s}^{-1}$) decides about the fate of the RNAP at the terminator (see text). The transcription rate is calculated from (5) with the pause duration taken from Fig. 3A and plotted as a function of the initiation attempt rate (*circles*). The *dashed line* shows a fit to the data points by a Hill function with Hill coefficient $\simeq 4.3$

which may exhibit a strongly non-linear dependence on $\alpha$. This is shown in Fig. 5 for the backtracking pause from the simulation of Fig. 3 and a rate $k = 2\ \mathrm{s}^{-1}$. In that case, the downstream transcription rate is very sensitive to changes in $\alpha$, which may either reflect changes in gene regulation or in the availability of RNAPs [23, 27]. In the case shown in Fig. 5, we fitted the data point with a Hill function (dashed line) and obtained a Hill coefficient of $\simeq 4.3$. This parameter characterizes the sensitivity with respect to the initiation attempt rate as given by the slope of the curve on a double-logarithmic plot. The high Hill coefficient indicates strong cooperativity between RNAPs in overcoming the terminator. This example demonstrates that the combination of a backtracking pause and a terminator results in a termination site that amplifies differences in the initiation (attempt) rate.

## 6 Concluding Remarks

In this study, we have addressed some aspects of pausing and backtracking of RNAPs under conditions of dense-traffic transcription. At high transcription initiation rates, transcription may become limited by elongation rather than initiation, which is the typical case under cellular conditions with the notable exceptions of transcription of rRNA and tRNA [25]. In elongation-limited transcription, pausing has a strong effect on the transcription rate, as any increase or decrease in pausing directly changes the limiting step in the transcription pathway.

Under these conditions, backtracking, the backward translocation of paused RNAPs, is strongly suppressed, because the trailing RNAP (i) restricts how far the paused RNAP may move backwards and (ii) effectively pushes the backtracked RNAP forward by a ratchet mechanism, as each forward step of the actively elongating trailing RNAP restricts the space available for backtracking further. This suppression of backtracking by the trailing RNAP has been observed experimentally in vitro and in vivo [16, 17]. Recently, Ehrenberg et al. [15] have invoked this effect to explain the puzzling observations that bacteria

with reduced number of rRNA genes compensate for that loss with an increased elongation speed [8]. In their model, the speed-up is a direct consequence of suppressed backtracking, as increased RNAP availability increases the initiation attempt rate. In our model, the elongation-speed is increased at intermediate transcription rates, but not at high transcription rates, where backtracking is almost completely suppressed and elongation is limited by short pauses without backtracking that are not suppressed by trailing RNAPs and are not taken into account in Ref. [15]. As discussed above, it is controversial whether backtracking occurs at only a small fraction of pauses, as we have assumed in most of our simulations here, based on single-molecule experiments [37] or whether most (but not all) pauses invoke backtracking. In our view, the observation of increased elongation speed in the bacterial strains with reduced number of rRNA genes therefore provides strong support for the existence of elongation-limited transcription, but its mechanistic interpretation remains unclear (as does its physiological significance) [25].

Our analysis also indicates that the suppression of backtracking already comes into play at relatively small transcription rates, far below the limit where elongation becomes limiting. For that reason, the suppression of backtracking may be relevant even for weak promoters if the backtracking pause is sufficiently long. A simple estimate can be made by comparing the time it takes a trailing RNAP to catch up with a paused RNAP. At a transcription rate $J$, the average distance between two consecutive RNAPs is $\Delta = u/J - L$. In the linear initiation-limited regime, the transcription rate $J$ and the elongation speed $u$ are given by $J = \alpha$ and $u = \epsilon s$. When the leading RNAP stops, the RNAP following directly behind it, needs a time $T^* = \Delta/(\epsilon s) = 1/J - L/(\epsilon s)$ to catch up. Any pause longer than $T^*$ will then lead to an interaction between the two RNAPs and potentially to suppressed backtracking. With a modest transcription rate of 5 $\text{min}^{-1}$, we can estimate $T^*$ as $T^* \simeq 10$ s, which is comparable to the duration of backtracking pauses observed in vitro [2, 10, 37], so that at such intermediate transcription rates, suppression of backtracking can be expected to be effective.

Finally, we have considered potential regulatory consequences of the suppression of backtracking using an example where a backtracking pause is coupled to transcription termination. This example demonstrates that the cooperation of RNAPs to overcome backtracking sites may be used to achieve cooperative transcription regulation. In such a scenario, transcription is strongly non-linear in any parameter that controls the initiation attempt rate such as the availability of RNAPs or the concentration of a transcription factor that recruits RNAPs to the promoter.

**Acknowledgements** During the initial phase of this work, SK was supported by Deutsche Forschungsgemeinschaft (grants KL818/1-1 and KL818/1-2) and by the NSF through the Center for Theoretical Biological Physics (grant PHY-0822283) at UC San Diego.

## References

1. Abbondanzieri, E.A., Greenleaf, W.J., Shaevitz, J.W., Landick, R., Block, S.M.: Direct observation of base-pair stepping by RNA polymerase. Nature **438**, 460–465 (2005)
2. Adelman, K., La Porta, A., Santangelo, T.J., Lis, J.T., Roberts, J.W., Wang, M.D.: Single molecule analysis of RNA polymerase elongation reveals uniform kinetic behavior. Proc. Natl. Acad. Sci. USA **99**, 13538–13543 (2002)
3. Artsimovitch, I., Landick, R.: Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. Proc. Natl. Acad. Sci. USA **97**, 7090–1095 (2000)
4. Bon, M., McGowan, S.J., Cook, P.R.: Many expressed genes in bacteria and yeast are transcribed only once per cell cycle. FASEB J. **20**, 1721–1723 (2006)

5. Bremer, H., Dennis, P.P.: Modulation of chemical composition and other parameters of the cell by growth rate. In: Neidhardt, F.C. (ed.) Escherichia coli and Salmonella, 2nd edn., pp. 1553–1569. ASM Press, Washington (1996)
6. Chou, T., Lakatos, G.: Clustered bottlenecks in mRNA translation and protein synthesis. Phys. Rev. Lett. **93**, 198101 (2004)
7. Chowdhury, D., Santen, L., Schadschneider, A.: Statistical physics of vehicular traffic and some related systems. Phys. Rep. **329**, 199–329 (2000)
8. Condon, C., French, S., Squires, C., Squires, C.L.: Depletion of functional ribosomal RNA operons in Escherichia coli causes increased expression of the remaining intact copies. EMBO J. **12**, 4305–4315 (1993)
9. Core, L.J., Lis, J.T.: Transcription regulation through promoter-proximal pausing of RNA polymerase II. Science **319**, 1791–1792 (2008)
10. Davenport, R.J., Wuite, G.J., Landick, R., Bustamante, C.: Single-molecule study of transcriptional pausing and arrest by E. coli RNA polymerase. Science **287**, 2497–2500 (2000)
11. Depken, M., Galburt, E.A., Grill, S.W.: The origin of short transcriptional pauses. Biophys. J. **96**, 2189–2193 (2009)
12. Derrida, B., Evans, M.R., Hakim, V., Pasquier, V.: Exact solution of a 1D asymmetric exclusion model using a matrix formulation. J. Phys. A, Math. Gen. **26**, 1493–1517 (1993)
13. Dobrzynski, M., Bruggeman, F.J.: Elongation dynamics shape bursty transcription and translation. Proc. Natl. Acad. Sci. USA **106**, 2583–2588 (2009)
14. Dong, J.J., Schmittmann, B., Zia, R.K.P.: Towards a model for protein production rates. J. Stat. Phys. **128**, 21–34 (2007)
15. Ehrenberg, M., Dennis, P.P., Bremer, H.: Maximum rrn promoter activity in Escherichia coli at saturating concentrations of free RNA polymerase. Biochimie **92**, 12–20 (2010)
16. Epshtein, V., Nudler, E.: Cooperation between RNA polymerase molecules in transcription elongation. Science **300**, 801–805 (2003)
17. Epshtein, V., Toulmé, F., Rahmouni, A.R., Borukhov, S., Nudler, E.: Transcription through the roadblocks: the role of RNA polymerase cooperation. EMBO J. **22**, 4719–4727 (2003)
18. Galburt, E.A., Grill, S.W., Wiedmann, A., Lubkowska, L., Choy, J., Nogales, E., Kashlev, M., Bustamante, C.: Backtracking determines the force sensitivity of RNAP II in a factor-dependent manner. Nature **446**, 820–823 (2007)
19. Grainger, D.C., Hurd, D., Harrison, M., Holdstock, J., Busby, S.J.W.: Studies of the distribution of Escherichia coli cAMP-receptor protein and RNA polymerase along the E. coli chromosome. Proc. Natl. Acad. Sci. USA **102**, 17693–17698 (2005)
20. Hatoum, A., Roberts, J.: Prevalence of RNA polymerase stalling at Escherichia coli promoters after open complex formation. Mol. Microbiol. **68**, 17–28 (2008)
21. Herbert, K.M., La Porta, A., Wong, B.J., Mooney, R.A., Neuman, K.C., Landick, R., Block, S.M.: Sequence-resolved detection of pausing by single RNA polymerase molecules. Cell **125**, 1083–1094 (2006)
22. Klumpp, S., Chai, Y., Lipowsky, R.: Effects of the chemomechanical stepping cycle on the traffic of molecular motors. Phys. Rev. E **78**, 041909 (2008)
23. Klumpp, S., Hwa, T.: Growth-rate-dependent partitioning of RNA polymerases in bacteria. Proc. Natl. Acad. Sci. USA **105**, 20245–20250 (2008)
24. Klumpp, S., Hwa, T.: Stochasticity and traffic jams in the transcription of ribosomal RNA: Intriguing role of termination and antitermination. Proc. Natl. Acad. Sci. USA **105**, 18159–18164 (2008)
25. Klumpp, S., Hwa, T.: Traffic patrol in the transcription of ribosomal RNA. RNA Biol. **6**, 392–394 (2009)
26. Klumpp, S., Nieuwenhuizen, T.M., Lipowsky, R.: Self-organized density patterns of molecular motors in arrays of cytoskeletal filaments. Biophys. J. **88**, 3118–3132 (2005)
27. Klumpp, S., Zhang, Z., Hwa, T.: Growth rate-dependent global effects on gene expression in bacteria. Cell **139**, 1366–1375 (2009)
28. Kolomeisky, A.B., Schütz, G.M., Kolomeisky, E.B., Straley, J.P.: Phase diagram of one-dimensional driven lattice gases with open boundaries. J. Phys. A, Math. Gen. **31**, 6911–6919 (1998)
29. Krug, J.: Boundary-induced phase transitions in driven diffusive systems. Phys. Rev. Lett. **67**, 1882–1885 (1991)
30. Lakatos, G., Chou, T.: Totally asymmetric exclusion process with particles of arbitrary size. J. Phys. A, Math. Gen. **36**, 2027–2041 (2003)
31. Landick, R.: The regulatory roles and mechanism of transcriptional pausing. Biochem. Soc. Trans. **34**, 1062–1066 (2006)
32. Lewicki, B.T., Margus, T., Remme, J., Nierhaus, K.H.: Coupling of rRNA transcription and ribosomal assembly in vivo. Formation of active ribosomal subunits in Escherichia coli requires transcription of rRNA genes by host RNA polymerase which cannot be replaced by bacteriophage T7 RNA polymerase. J. Mol. Biol. **231**, 581–593 (1993)

33. Liang, S., Bipatnath, M., Xu, Y., Chen, S., Dennis, P., Ehrenberg, M., Bremer, H.: Activities of constitutive promoters in Escherichia coli. J. Mol. Biol. **292**, 19–37 (1999)
34. Lipowsky, R., Klumpp, S., Nieuwenhuizen, T.M.: Random walks of cytoskeletal motors in open and closed compartments. Phys. Rev. Lett. **87**, 108101 (2001)
35. MacDonald, C.T., Gibbs, J.H., Pipkin, A.C.: Kinetics of biopolymerization on nucleic acid templates. Biopolymers **6**, 1–25 (1968)
36. Miller, O.L. Jr, Beatty, B.R.: Visualization of nucleolar genes. Science **164**, 955–957 (1969)
37. Neuman, K.C., Abbondanzieri, E.A., Landick, R., Gelles, J., Block, S.M.: Ubiquitous transcriptional pausing is independent of RNA polymerase backtracking. Cell **115**, 437–447 (2003)
38. Nieuwenhuizen, T.M., Klumpp, S., Lipowsky, R.: Random walks of molecular motors arising from diffusional encounters with immobilized filaments. Phys. Rev. E **69**, 061911 (2004)
39. Polya, G.: Über eine Aufgabe der Wahrscheinlichkeitsrechnung betreffend die Irrfahrt im Straßennetz. Math. Ann. **84**, 149–160 (1921)
40. Proshkin, S., Rahmouni, A.R., Mironov, A., Nudler, E.: Cooperation between translating ribosomes and RNA polymerase in transcription elongation. Science **328**, 504–508 (2010)
41. Roberts, J.W., Yarnell, W., Bartlett, E., Guo, J., Marr, M., Ko, D.C., Sun, H., Roberts, C.W.: Antitermination by bacteriophage lambda Q protein. Cold Spring Harbor Symp. Quant. Biol. **63**, 319–325 (1998)
42. Schütz, G.M.: Exactly solvable models for many-body systems far from equilibrium. In: Domb, C., Lebowitz, J. (eds.) Phase Transitions and Critical Phenomena, vol. 19, pp. 1–251. Academic Press, San Diego (2001)
43. Schütz, G., Domany, E.: Phase transitions in an exactly soluble one-dimensional exclusion process. J. Stat. Phys. **72**, 277–296 (1993)
44. Shaw, L.B., Zia, R.K.P., Lee, K.H.: Totally asymmetric exclusion process with extended objects: A model for protein synthesis. Phys. Rev. E **68**, 021910 (2003)
45. Tripathi, T., Chowdhury, D.: Interacting RNA polymerase motors on a DNA track: effects of traffic congestion and intrinsic noise on RNA synthesis. Phys. Rev. E **77**, 011921 (2008)
46. Vogel, U., Jensen, K.F.: Effects of the antiterminator boxA on transcription elongation kinetics and ppGpp inhibition of transcription elongation in Escherichia coli. J. Biol. Chem. **270**, 18335–18340 (1995)
47. Voliotis, M., Cohen, N., Molina-París, C., Liverpool, T.B.: Backtracking and proofreading in DNA transcription. Phys. Rev. Lett. **102**, 258101 (2009)
48. von Hippel, P.H.: An integrated model of the transcription complex in elongation, termination, and editing. Science **281**, 660–665 (1998)
49. Weiss, G.: Aspects and Applications of the Random Walk. North-Holland, Amsterdam (1994)