

## Quantitative analysis of a virulent bacteriophage transcription strategy

Marko Djordjevic<sup>a,b,\*</sup>, Ekaterina Semenova<sup>c</sup>, Boris Shraiman<sup>d</sup>, Konstantin Severinov<sup>c,e</sup>

<sup>a</sup> Department of Physics, Columbia University, New York, NY 10027, USA

<sup>b</sup> Mathematical Biosciences Institute, The Ohio State University, Columbus, OH 43210, USA

<sup>c</sup> Waksman Institute, Piscataway, NJ 08854, USA

<sup>d</sup> Kavli Institute for Theoretical Physics, University of California, Santa Barbara, CA 93106, USA

<sup>e</sup> Department of Molecular Biology and Biochemistry, Rutgers University, Piscataway, NJ 08854, USA

Received 12 January 2006; returned to author for revision 8 March 2006; accepted 29 May 2006

Available online 2 August 2006

### Abstract

An increasingly large number of bacteriophage genomes are being sequenced each year. What is an efficient experimental and computational procedure to analyze transcription strategies of newly sequenced novel bacteriophages? We address this issue using an example of bacteriophage Xp10, which infects rice pathogen *Xanthomonas oryzae*. This phage is particularly challenging for analysis, since part of its genome is jointly transcribed by two (host and viral) RNA polymerases. To understand the roles played by the two RNA polymerases, we developed a novel method of data analysis which combines quantitative analysis of Xp10 global gene expression data and kinetic modeling of the infection process. To generalize our approach, we discuss how our method can be applied to other systems and argue that genomic array experiments combined with the methods of data analysis that we present provide an efficient way to analyze gene expression strategies of novel bacteriophages.

© 2006 Elsevier Inc. All rights reserved.

**Keywords:** Bacteriophages; Xp10 phage; Gene expression; Genomic array data; Kinetic modeling; T7-like RNAP

### Introduction

Recently, a significant number of phage genome sequences have become available (currently around 230). The pace of phage genome sequencing is expected to quicken in the future, and thousands of phage genomes should be available in a very few years (Brussow and Hendrix, 2002; Hendrix, 2003). However, this avalanche of bacteriophage genome sequences has resulted in a situation where functional analysis of new, potentially very interesting genomes is lagging behind. Our motivation is to propose an efficient experimental and data analysis method that would help bridge this gap.

The approach presented below is particularly well-suited for analysis of gene expression of phages that encode their own RNA polymerases. The already known examples of such phages include a large family of T7-like phages, including highly divergent phages such as SP6 and phiKMV (Molineux,

2005), the N4 phage (Zivin et al., 1981), as well as the Xp10 bacteriophage (Yuzenkova et al., 2003).

We will demonstrate our method on the example of bacteriophage Xp10, which infects *Xanthomonas oryzae*, a rice pathogen. Xp10 appears to combine transcription control strategies characteristic of two distinct viral families: the virulent phages related to T7 phage and the temperate phages related to phage  $\lambda$  (Yuzenkova et al., 2003; Semenova et al., 2005). A scheme of Xp10 genome organization is shown in Fig. 1. Half of the Xp10 genome contains genes coding for proteins involved in host shut-off, enzymes of viral genome replication, and a T7-like RNA polymerase (RNAP); the other half of the genome contains genes coding for structural proteins and host lysis proteins, in an arrangement typical for many lambdoid phages. The two groups of genes are divergently transcribed and are separated by a regulatory region, which contains divergent promoters for both host RNAP and phage RNAP (Semenova et al., 2005). Further in the text, we will refer to the leftward-transcribed genes as L genes and to the rightward-transcribed genes as R genes.

As a starting point, we use experimental data published in our previous paper (Semenova et al., 2005). There, a

\* Corresponding author. Mathematical Biosciences Institute, The Ohio State University, Columbus, OH 43210, USA. Fax: +1 614 2476643.

E-mail address: [mdjordjevic@math.ohio-state.edu](mailto:mdjordjevic@math.ohio-state.edu) (M. Djordjevic).

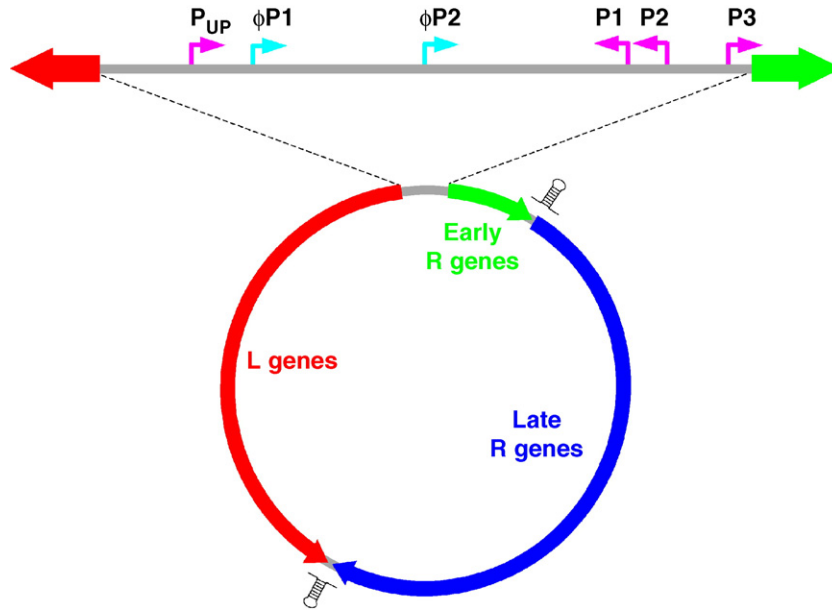


Fig. 1. Organization of Xp10 genome. In the top of the figure, promoters in the intergenic region that separates leftward and rightward transcribed genes are shown. Host and phage promoters are shown in magenta and cyan respectively. Below, circular Xp10 genome is shown. Groups of genes that belong to different temporal classes, together with the corresponding directions of transcription, are indicated. Putative transcription terminators that separate the two temporal classes of R genes are indicated by hairpins.

macroarray of Xp10 genes was used to measure transcript abundance at different times (1, 3, 5, 7, 10, 15, 20, 25, 40, and 60 min) post-infection. The results of those measurements, used in the present analysis, are listed in Table S1. Based on these measurements, Xp10 genes were shown to cluster into three different classes according to the pattern of their transcript abundance versus time after the infection. All L transcripts belong to a single temporal class: their transcript abundance rapidly increases until about 15 min post-infection and then starts to decrease (a representative behavior of an L gene transcript abundance versus time post-infection is shown in Fig. 2). R transcripts are separated into two different temporal classes. For most R transcripts, the increase of abundance is very slow early in the infection, but becomes rapid later, at about 10 min post-infection. We call such transcripts ‘late’ R, and a representative behavior of a late R gene transcript abundance is shown in Fig. 2. However, some R transcripts show rapid increase of abundance early in the infection (similarly to that of the L transcripts) with their levels remaining constant (high) late in the infection. We call these transcripts ‘early’ R and a representative behavior of an early R transcript abundance is also shown in Fig. 2. The genomic positions of Xp10 genes belonging to the three temporal classes are illustrated in Fig. 1.

A 73-amino-acid-long Xp10 protein p7 binds to and inhibits *X. oryzae* host RNAP (Nechaev et al., 2002). The in vitro activities of p7 suggest that it may act as a key regulator of Xp10 development. P7 strongly inhibits *X. oryzae* RNAP transcription from the major  $-10/-35$  class promoters but transcription from promoters of the minor extended  $-10$  class is partially resistant to p7. For example, in vitro experiments showed that leftward-oriented host RNAP promoters present in the intergenic

regulatory region of Xp10 are effectively inhibited by p7, while the rightward extended  $-10$  class host RNAP promoter P3 is partially resistant (Yuzenkova et al., 2003) (positions of the promoters are indicated in Fig. 1). Thus, p7 may execute host transcription shut-off and may act as a switch from L genes promoters to R genes promoters. In addition to regulating promoter utilization, in vitro, p7 prevents transcription termination by host RNAP at all intrinsic terminators tested (Nechaev et al., 2002). Thus, different temporal patterns of early R and late R genes expression may be explained by the presence of a

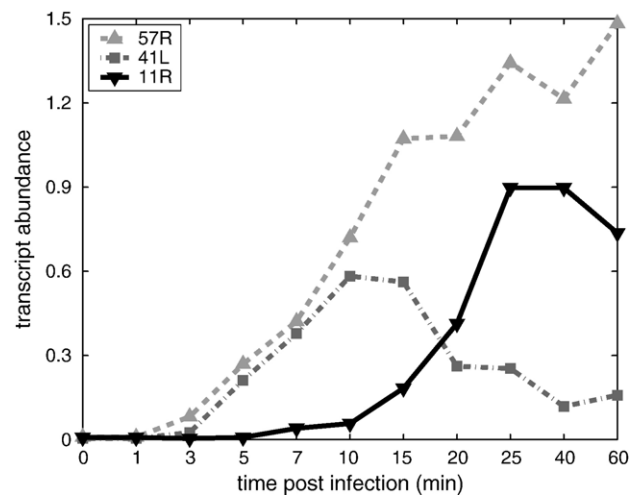


Fig. 2. Examples of Xp10 transcripts that belong to three distinct temporal classes. Transcript abundances of 41L (an L gene), 57R (an early R gene) and 11R (a late R gene), during the course of the Xp10 infection, are shown. Transcript abundances are obtained from the gene expression measurements (in the absence of Rif) and are given in arbitrary units.

terminator that separates the early and late R genes, and by transcription anti-termination caused by accumulation of p7. Indeed, from our bioinformatic search we identified several putative intrinsic terminators between rightward-transcribed genes, including one between the early and late R genes (Semenova et al., 2005).

In addition to host RNAP promoters, two rightward phage RNAP promoters,  $\phi$ P1 and  $\phi$ P2, are also located in the intergenic region (see Fig. 1), indicating that phage RNAP transcribes Xp10 R genes. The existence of rightward-oriented phage RNAP promoters appears to make the existence of p7-resistant P3 promoter and p7-induced anti-termination redundant, provided that Xp10 RNAP does not recognize (or responds poorly to) predicted termination signals. This possibility is plausible since T7 RNAP, which is related to Xp10 RNAP, recognizes only some strongest host RNAP-like termination signals and responds to those signals much less efficiently than the bacterial enzyme (Jeng et al., 1990, Macdonald et al., 1994, Molineux, 2005). Conversely, if the host RNAP is capable of transcribing all Xp10 genes, there appears to be no role left for Xp10 RNAP. Therefore, in addition to presenting the data analysis approach of general importance, another (more specific) goal of this paper is to obtain quantitative understanding of the Xp10 gene expression strategy, and in particular to determine the contributions of both enzymes to viral gene expression and phage progeny development.

The outline of the paper is as follows. In Methods and results, we will introduce our method and apply it to analyze the Xp10 gene expression strategy. First, we will use genomic array data to determine which Xp10 genes are transcribed exclusively by host RNAP and for which genes phage RNAP has to be involved in transcription. Next, we will estimate *in vivo* rates of phage transcript decay from gene expression data. Subsequently, we will combine genomic array data, results from primer extension experiments, and transcript kinetic modeling to determine contributions of two RNA polymerases to transcription rates of all three classes of Xp10 genes throughout infection. As an example of qualitative predictions that can be obtained from the estimated transcription rates, we will find that (1) the absence of transcription of Xp10 structural genes by either of the two RNA polymerases would severely decrease phage progeny numbers, and (2) that there is a highly nonlinear relation between the amounts of synthesized phage structural proteins and the phage progeny production. We will then summarize the understanding of Xp10 transcription strategy that follows from our analysis. Finally, in Discussion, we will address how our analysis can be generalized to study other bacteriophages and identify some potential limitations of our approach.

## Discussion

### *Xp10 transcription strategy*

The following view of the Xp10 gene expression strategy emerges from quantitative analysis performed here and from the previous experimental biochemical studies of the phage and the process of its infection. Early in infection, host RNAP

transcribes both the L genes and the early R genes. Transcription of the late R genes is prevented due to the action of a terminator located downstream of the early R gene cluster. Transcription of the L genes leads to accumulation of viral RNAP and the p7 protein. Accumulation of increasing amounts of viral RNAP leads to proportional and large increase in both early and late R gene transcription. The activity of viral RNAP declines late in infection through a mechanism that remains undefined. Accumulation of p7 switches off the L genes transcription, along with the majority of host genes transcription by  $\sim 15$  min post-infection. P7 also leads to  $\sim 4$ -fold decline in a rightward transcription initiation by host RNAP, through (i) complete inhibition of the  $-10/-35$  promoter  $P_{UP}$  and (ii) partial inhibition of the extended  $-10$  P3 promoter. However, transcription initiated at P3 can now proceed into the late R genes due to p7-induced transcription anti-termination (host RNAP transcription of the late R genes occurs with a  $\sim 10$ -min delay compared to the late R gene transcription by viral RNAP). For times later than 20 min post-infection, the contribution of host RNAP to late R genes transcription is  $\sim 25\%$ . Thus, host RNAP contributes in a significant way to R gene transcription and to phage progeny production, particularly since a relationship between the amounts of Xp10 structural proteins synthesized and the number of phage progeny generated appears to be highly nonlinear.

Importantly, the quantitative method presented here allowed to reach biological insights that could not have been made from experimental observations of Semenova et al. (2005). First, our kinetic model was used to show that L genes are transcribed exclusively by host RNAP. Next, by using the calculation of transcript decay rates and the model for transcript generation/decay, we obtained that transcription of L genes ceases after about 15 min from start of the infection. Finally, our analysis allowed to compute relative contributions of the two RNA polymerases to the R genes transcription, which led to the following novel qualitative conclusions: (1) bacterial RNAP transcribes *all* R genes even late in the infection, (2) absence of transcription of structural (late R) genes by either of the two RNA polymerases would severely reduce the phage progeny numbers, (3) the relationship between the phage protein amounts and the numbers of phage progeny is highly non linear.

For a successful method of quantitative analysis of gene expression by new phages, it is important to understand to what extent is the method based on additional experimental knowledge about the phage. The experimental inputs used in our analysis to obtain contributions of both RNA polymerases to transcription rates of all Xp10 genes throughout the infection were (1) genomic array data (with and without Rif), primer extension data, and annotated Xp10 genome sequence, and (2) the fact that an anti-terminator, the p7 protein, belongs to the L genes. We note that one could quite safely assume the latter even without prior experimental knowledge, since (1) from the Xp10 genome sequence it follows that R genes encode structural proteins and host lysis proteins, and (2) many lambdoid phages, to which Xp10 belongs based the morphology of its virions, rely on transcription anti-termination by early genes products to express their structural genes. The additional experimental

knowledge (e.g., about the ability of p7 or other phage-encoded proteins to inhibit host RNAP) was not used as an input in our analysis. On the other hand, prior knowledge about the mechanism of host RNAP transcription shut-off by p7 provided a much more detailed understanding of results obtained from our analysis. For example, the experimental fact that p7 *in vitro* inhibits transcription of  $-10/-35$  class of the promoters (Yuzenkova et al., 2003), provided a plausible explanation for the result that the transcription of L genes ceases later in the infection. Conversely, our analysis strongly indicates that p7 also plays the same inhibitory function *in vivo*, which was not known before.

#### *Application of the method to other systems*

The kinetic modeling based method of data analysis given in this paper, together with the experimental and sequence analysis methods presented in our previous paper (Semenova et al., 2005), provide an efficient method to study transcription strategies of novel bacteriophages. To our knowledge, an integrated experimental and computational method allowing efficient analysis of transcription strategy by a novel virulent bacteriophage was not proposed before. The experimental method consists of measurements of transcript abundances of bacteriophage genes, as well as the amounts of transcripts generated from individual promoters at several time points post-infection. Further, available sequence analysis methods (see, e.g., Lawrence et al., 1993; Djordjevic et al., 2003; Ermolaeva et al., 2000) combined with genomic array data can be used to predict, quite reliably, positions of phage promoters and terminators. Predicted promoters can be experimentally confirmed through standard experimental techniques (see, e.g., Semenova et al., 2005).

As shown here, once phage genes are clustered in different temporal classes, transcription rates corresponding to each class can be determined, provided that transcript decay rates can be calculated (as was done in our analysis) or are already known (see the discussion below with regard to this issue). Similarly, the kinetic model allows us to calculate transcription activities corresponding to individual promoters from primer extension data. The central part of our method is to equate transcription rates corresponding to a given temporal class, with either linear or nonlinear superposition of transcription rates of upstream promoters. Nonlinear superposition has to be used whenever a time-dependent anti-termination mechanism is suspected. Fitting appropriate parameters then allows one to determine contribution of individual promoters to the transcription rates. It is evident that our approach is of particular importance in the case when two RNAPs jointly transcribe the viral genome.

For many novel or even known phages, only annotated phage genome sequence and gene expression and/or primer extension data are available for analysis. As we showed on the example of Xp10 (see previous subsection), and as discussed above for a general case, such relatively scarce knowledge is sufficient to make biologically significant and nontrivial predictions using our method. We therefore expect that our

analysis can be successfully applied to newly sequenced bacteriophages. However, in the absence of additional data, our method does not identify specific phage (or bacterial) proteins and their interactions that are mechanistically responsible for predicted patterns of phage transcription. Nevertheless, the method allows making testable predictions about types of transcription regulators that may be employed by the phage during its development and these regulators can then be specifically searched for. If available (as was mostly the case for Xp10 phage), this additional data can be combined with the results of our method to give a reasonably complete understanding of the transcription strategy of a novel phage. Some additional (potential) limitations of our method are discussed below.

In our work, we differentiated between the classes of genes transcribed by different polymerases using a (known) drug, which allowed only one RNAP to remain active and shut down the transcription of the other (host) RNAP virtually instantaneously (less than 1–2 min, as per Bernstein et al., 2002; O'Hara et al., 1995). In the absence of such a drug, the analysis would be significantly more complicated. However, a relatively large number of low-molecular compounds targeting bacterial RNAP are available, and since almost all phages (with the exception of N4 phage) that encode their own RNAP have only one phage RNAP, the above limitation should likely not appear in most cases. Further, it would be significantly more complicated to determine phage transcript decay rates (than in our method), if there would *not* be a portion of phage genome where the transcription could be completely stopped (by a known drug(s)). In all the examples of phages studied so far, there is a portion of phage genome transcribed exclusively by bacterial RNAP (Molineux, 2005; Zivin et al., 1981; Yuzenkova et al., 2003) but, in principle, this may not happen for all novel bacteriophages.

An additional issue relevant to the application of our method is related to the genome entry dynamics. We note that the data analysis itself will not be affected by the dynamics of genome internalization, i.e., the determined promoter transcription activities and the gene transcription rates will simply give zero values until the corresponding part of the genome has entered the cell. It is, however, evident that the potential delay in transcription associated with genome entry should be kept in mind in order to properly interpret the results. This could be particularly significant when analyzing and/or comparing a gene expression strategy of a phage such as phiKMV, in whose genome the gene coding for the viral RNAP is centrally located, to T7-like phages where the corresponding genes are located close to the genome end that is first inserted in the cell. A related issue is associated with the latency of transcription as a result of loading of RNAP onto DNA. As with the genome entry dynamics, this will not affect the data analysis procedure, i.e., the transcription activities will appear with zero (small) values until RNAP molecules have located phage promoters. This issue is, however, relevant for the result interpretation, e.g., in the case of Xp10 phage the initial lag in the transcription of L genes may be due to the finite time needed for the promoter location.

Finally, we note that although the analysis presented here was discussed in the context of bacteriophages, it is interesting to note that chloroplast genes are simultaneously transcribed by two types of RNA polymerases (Allison et al., 1986; Hajdukiewicz et al., 1997). This suggests that the analysis developed here can also be used for analysis of transcription in these organelles.

## Methods and results

### Which genes are transcribed by what RNAP?

Analysis of transcript abundances measured at different times post-infection leads to the conclusion that all L genes belong to a single temporal class, while the R genes are divided into two different classes (Semenova et al., 2005). To determine which RNA polymerase(s) is responsible for transcription of genes belonging to different temporal classes, Semenova et al. (2005) performed macroarray analysis of Xp10 transcripts at conditions when Rifampicin (Rif), a drug that inhibits host RNAP, was added at 0, 1, 3, 5, 7, 10, 15, 20, 25, 40, and 60 min post-infection, followed by additional 20-min incubation of infected cells in the presence of the drug (Semenova et al., 2005). Representative Xp10 transcripts were measured at the times of Rif addition and after the 20-min incubation with Rif. The measured transcript abundances at the time of Rif addition are listed in Table S1, while the measured transcript abundances after the 20-min incubation with Rif are listed in Table S2. Appendix S3 gives explanation of the experimental data used in our analysis. Note that, while, in principle, macroarray hybridizations could also detect nonfunctional mRNA, we here assume that most of the RNA hybridizing to the macroarray spots represents functional transcripts. This assumption is plausible based on the previous microarray studies (see Bernstein et al., 2002 and references therein).

For the purposes of our analysis, it is suitable to define, for each representative Xp10 transcript species  $i$ , quantity  $\Delta_i$  equal to the difference in the transcript abundance before and after Rif addition:

$$\Delta_i(t) = c_i^{\text{Rif}}(t, \Delta t) - c_i(t). \quad (1)$$

Here  $c_i^{\text{Rif}}(t, \Delta t)$  is the measured amount of transcript  $i$  at time  $\Delta t$  after Rif addition at time  $t$ , while  $c_i(t)$  is the measured amount of transcript  $i$  at the time  $t$  of Rif addition (in our case,  $\Delta t = 20$  min). For example, in order to calculate the value of  $\Delta$  for a given transcript at 10 min post-infection, one should take the measured value of transcript abundance at 10 min post-infection (see Table S1), and subtract this from the value of transcript abundance measured 20 min after Rif was added at 10 min post-infection (see Table S2).

We first look at transcript generation when Rif is added at time  $t$ . Previous experimental studies showed that host RNAP is inhibited in less than 1–2 min from Rif addition (see, e.g., O'Hara et al., 1995; Bernstein et al., 2002). Consequently, we assume that Rif acts on a time scale significantly less than 20-min interval in which the measurement of  $\Delta$  is performed. The

amount of transcript  $i$ , at time  $t + t'$ , can be then described by the following differential equation:

$$\frac{dc_i^{\text{Rif}}(t, t')}{dt'} = -\lambda_i c_i^{\text{Rif}}(t, t') + \gamma_p^{\text{Rif}}(t, t'). \quad (2)$$

Here,  $\lambda_i$  denotes decay constants for transcripts  $i$ , which are assumed to be time-independent.  $\gamma_p^{\text{Rif}}(t, t')$  is the rate of transcription of gene  $i$  by phage RNAP at time  $t + t'$  post-infection, provided that Rif was added at time  $t$ .  $\gamma_p^{\text{Rif}}(t, t')$  is, in general, smaller than or equal to the transcription rate by phage RNAP when no Rif is added (i.e.,  $\gamma_p^{\text{Rif}}(t, t') \leq \gamma_p(t + t')$ ). This is because inhibition of host RNAP by Rif at time  $t$  may reduce (or completely stop), generation of transcripts of Xp10 RNAP gene.

After solving Eq. (2), with the initial condition  $c_i^{\text{Rif}}(t, t' = 0) = c_i(t)$  and using Eq. (1), we obtain that:

$$\Delta_i(t) = \exp(-\lambda_i \Delta t) \int_0^{\Delta t} \gamma_p^{\text{Rif}}(t, \xi) \exp(\lambda_i \xi) d\xi - c_i(t)(1 - \exp(-\lambda_i \Delta t)), \quad (3)$$

where  $\xi$  is an integration variable. As can be seen, the value of  $\Delta_i$  equals the amount of transcript that was generated during the period when infected cells were incubated with Rif (the first term on the right side of Eq. (3)) less the transcript amount present at  $t$  that decayed during the same period (the second term on the right-hand side of Eq. (3)). When  $\Delta_i(t)$  is plotted for all representative transcripts  $i$  (we use the experimental measurements from Table S1 and Table S2), it can be seen that it takes (1) positive values for all R transcripts, at least during one part of the infection, and (2) exclusively negative values for all L transcripts (see Fig. 3A). Further, from Fig. 3A can be noticed that  $\Delta$  values for late R genes are exclusively positive, while the values for early R genes change sign, being positive in the beginning of infection and negative later.

The first and the second terms in the right-hand side of Eq. (3) give, respectively, positive and negative contributions to  $\Delta$ . Positive  $\Delta(t)$  then necessarily implies that the first term, which gives the amount of transcript generated by phage RNAP from  $t$  to  $t + \Delta t$ , is nonzero. Therefore, since  $\Delta$  is positive for all R genes (at least during the part of the infection) phage RNAP necessarily has to be involved in the R genes transcription. We suggested, in our previous paper (Semenova et al., 2005), that the unusual sign change for  $\Delta$  values of the early R genes may be due to kinetics of the early R transcripts decay. This explanation is consistent with our kinetic model of Xp10 gene expression. The second term in Eq. (3) gives a negative contribution and is directly proportional to transcript abundance  $c_i(t)$ . In the beginning of infection,  $c_i(t)$  is small and this 'decay' term takes a small absolute value, which results in a positive value of  $\Delta_i(t)$  for both early R and late R transcripts. However, since transcript abundance of the early R genes rapidly increases in the beginning of infection, the 'decay' term in Eq. (3) becomes large, which results in a negative value for  $\Delta_i(t)$ . On the other hand, late R transcript abundances increase very slowly early in the infection, so later in the infection  $c_i(t)$  is significantly less for the late R genes than for the early R genes (see Fig. 2), which results in positive values of  $\Delta_i(t)$  for late R genes throughout the infection.

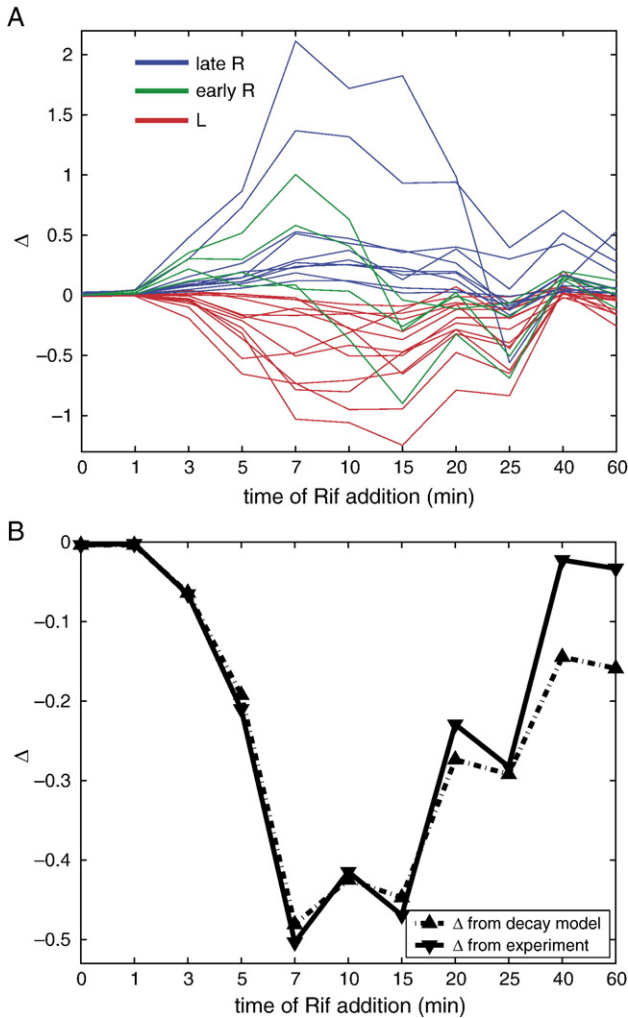


Fig. 3. Effect of host RNAP inhibition on the generation of Xp10 transcripts. (A) Value of  $\Delta$  is presented for individual Xp10 transcripts as a function of time. The values are calculated from the experimentally measured transcript abundances after 20 min of incubation with Rif, and from the measured transcript abundances at the time when Rif was added. Transcripts that belong to different temporal classes are shown in different colors. (B) Solid line shows  $\Delta$  for 31L transcript determined from the experimental data. Dash-dotted line is the predicted value of  $\Delta$  for 31L transcript, obtained from the model which assumes that L genes are not transcribed by phage RNAP.

For the L genes,  $\Delta_i(t)$  takes negative values throughout the infection, suggesting (but not proving) that they may be transcribed exclusively by host RNAP. If host RNAP were solely responsible for L gene transcription, L transcripts generation should cease upon addition of Rif reducing Eq. (3) to:

$$\Delta_i(t) = -\alpha_i c_i(t). \tag{4}$$

Here,  $\alpha_i$  is a positive, time-independent constant:

$$\alpha_i = [1 - \exp(-\lambda_i \Delta t)]. \tag{5}$$

According to Eq. (4),  $\alpha_i$  can be determined by calculating the average of  $\Delta_i(t_k)/c_i(t_k)$  over all measured points:

$$\alpha_i = \langle -\Delta_i(t_k)/c_i(t_k) \rangle_k \equiv -\frac{1}{11} \sum_{k=1}^{11} \Delta_i(t_k)/c_i(t_k). \tag{6}$$

From Eq. (4), we see that  $\Delta_i(t)$  is directly proportional to  $-c_i(t)$ . Therefore, if the hypothesis that L genes are transcribed exclusively by host RNAP holds, measured  $\Delta_i(t_k)$  should be given by Eq. (4) for all time points  $t_k$ . To test this, we compare measured  $\Delta_i(t_k)$  to  $-\alpha_i c_i(t_k)$ . The example of such a comparison, for the 31L transcript, is shown in Fig. 3B. As can be seen, there is a very good correspondence between  $\Delta_i(t_k)$  and  $-\alpha_i c_i(t_k)$  for the first nine time points, i.e., for times less than 40 min post-infection. For the last two time points (40 and 60 min),  $\Delta_i(t_k)$  has a smaller absolute value than  $-\alpha_i c_i(t_k)$ . Similar behavior is observed for all L transcripts  $i$  (results not shown).

From the good agreement of  $\Delta_i(t_k)$  with  $-\alpha_i c_i(t_k)$  we conclude that the L genes are transcribed exclusively by host RNAP. To further support this conclusion, we tried to fit  $\Delta_i(t)$  assuming a nonzero rate of L transcript synthesis by phage RNAP in Eq. (3). For simplicity, we assumed that  $\gamma_p^{\text{Rif}} \approx \gamma_p$  and that synthesized phage RNAP molecules are stable on the time scale of infection. We also assumed that the translation rate does not depend on time. Under these assumptions, we obtained the concentration of phage RNAP molecules by integrating the experimentally measured amounts of phage RNAP transcripts. We further assumed that the transcription rate is proportional to the concentration of phage RNAP molecules, i.e., that there is no saturation in the binding of phage RNAP to the promoters and that promoter clearance happens as soon as RNAP locates a promoter. The results of the fit indicated a negligible contribution of the first term on the right-hand side of Eq. (3), which further supports the conclusion that there is no synthesis of L transcripts by phage RNAP.

The systematically more negative values of  $-\alpha_i c_i(t_k)$  compared to  $\Delta_i(t_k)$  for the 40 and 60-min time points (see Fig. 3B), probably indicate that when infection approaches the stage of host cell lysis, the decay constants  $\lambda_i$  decrease such that our assumption that  $\lambda_i$  is constant is no longer valid. A possible explanation for observed decrease of decay constants late in infection may be that the cell machinery that processes transcripts deteriorates at this stage.

Finally, one can notice that  $\Delta$  values for all three classes of transcripts take values close to zero toward the end of the infection (see Fig. 3A at 40 min and 60 min). For L transcripts, this is because  $\Delta$  is directly proportional to the transcript abundance, and since the transcript abundance for L transcripts becomes small towards the end of the infection, so does  $\Delta$ . In the case of R transcripts,  $\Delta$  takes small values late in the infection due to the following reason: the amounts of R transcripts stop to significantly change later in the infection (see Fig. 2), which means that the rate of transcript generation becomes approximately equal to the rate of transcript decay. We will show below that later in the infection generation of R transcripts is mostly due to phage RNA polymerase. Therefore, late in infection the rate of R transcripts generation by phage RNAP has to be similar to the rate of R transcript decay. Since  $\Delta$  is determined by the difference of these two rates (see Eqs. (2) and (3)), it follows that for R transcripts late in the infection  $\Delta$  has to take values close to zero.

### Decay rates of L transcripts

The analysis presented above allows us to determine the decay constants for L transcripts. Having determined  $\alpha_i$ 's from Eq. (6), we can determine  $\lambda_i$ 's from Eq. (5). Values of half-lives  $T_i$ , defined as the time it takes for half of the initial transcript amount to decay, are directly connected to  $\lambda_i$ 's through:  $T_i = \ln(2)/\lambda_i$ . Using Eq. (5), we obtain:

$$T_i = -\frac{\ln(2)\Delta t}{\ln(1 - \alpha_i)}. \quad (7)$$

In practice, we determine half-lives  $T_i$  (that is, decay constants  $\alpha_i$ ) from Eq. (6) by summing not over all 11 measurements, but instead over 7 measurements at  $t_k = 3, 5, 7, 10, 15, 20,$  and  $25$  min. Measurements at 0 and 1 min are not included in order to avoid division by 'zero' (note that transcript abundances are very low in the beginning of infection), while measurements at 40 and 60 min are excluded because  $\lambda_i$  ceases to be constant at times larger than 40 min post-infection (see the discussion in the previous subsection). Values of  $T_i$  for all L transcripts are given in Table 1. Since we have multiple measurements (7 in total), standard deviations for every  $T_i$  value can be determined. The results are shown in Table 1. Since all differences between  $T_i$  for different transcripts  $i$  are within the variations in the experimental measurements used in our modeling, we conclude that there are no significant differences between the half-lives for different L transcripts. As can be seen, for seven out of twelve L transcripts analyzed (31L, 32L, 36L, 41L, 45L, 49L, and 56L),  $T_i$ 's are determined quite precisely by our method. To determine mean half-life for L transcripts, we averaged over those seven values. The mean half-life obtained in this way is about 13 min, which is significantly longer than half-lives of bacterial transcripts that are typically around 5 min (Bernstein et al., 2002). For a given rate of transcription, transcripts with longer half-lives will accumulate faster and

reach higher steady state levels, so the observed stability may be due to the need to rapidly achieve high levels of viral transcripts during the infection.

We finally note that while our results indicate that transcript decay rates are constant through much of the Xp10 infection, this may not be the case for all other phages. For example, T7 and T4 are known to modulate RNase activity through the synthesis of phage-encoded proteins (Marchand et al., 2001; Ueno and Yonesaki, 2004). A change in transcript decay rates will result in a change of the proportionality constant between  $\Delta$  and transcript abundance  $c$  (see Eq. (4)), starting from a certain time post-infection. Therefore, it is straightforward to detect the change in transcript stability and to determine the corresponding decay rates (before and after the modulation) by plotting  $\Delta_i(t)/c_i(t)$  and inferring the corresponding proportionality constants. The appropriate equations later in the text, specifically Eq. (8) and the equations in Appendix S2, should then be modified, so that the correct decay rates are used in each time interval. Thus, the general applicability of our method does not depend of phage-induced alterations in transcript decay rates.

### Transcription of L genes by host RNAP

We have established that L genes are transcribed exclusively by host RNAP. From macroarray data (in the absence of Rif), we can extract the host RNAP transcription rates for L genes. It is easy to see that from the equation, which describes the transcript generation process (see, for example, Eq. (2)), the transcription rate of gene  $i$ ,  $\gamma_i(t)$ , can be expressed in the following way:

$$\gamma_i(t) = \frac{dc_i(t)}{dt} + \lambda_i c_i(t). \quad (8)$$

Here,  $c_i(t)$  is the measured transcript abundance and  $\lambda_i$  is the decay constant. Note that the transcription rate of a gene  $\gamma_i(t)$  (or equivalently the rate of transcript generation) is a different quantity from the rate of transcript accumulation. The later is given by the first term on the right-hand side of Eq. (8), and it depends on both the transcription rate and the transcript stability. Decay constants  $\lambda_i$  for all L transcripts were determined in the previous section. We calculate the derivative  $dc(t)/dt|_{t_k}$  at the time point  $t_k$  as a mean value of the left and right derivatives (left derivative being the slope corresponding to transcript abundance measurements at  $t_{k-1}$  and  $t_k$ , and right derivative being the slope corresponding to measurements at  $t_k$  and  $t_{k+1}$ ). Transcription rates of L gene transcripts,  $\gamma_i(t_k)$ , can be immediately determined (at time points  $t_k$ ) by using Eq. (8) and the data listed in Table S1. The transcription rates obtained in this way are expressed in arbitrary units, since the measured radioactivity signal (Table S1) is proportional, but not directly equal, to the number of transcripts of species  $i$ . The average transcription rate of the L genes (average of  $\gamma_i(t_k)$  being taken over all  $i$  for individual L genes) is shown in Fig. 4.

Table 1  
Estimates of half-lives for L transcripts

ORF number	Functional annotation	Half-life (min)	Standard deviation <sup>a</sup> (min)
31L	ATP-dependent DNA ligase, minimal catalytic domain	10	2
32L	DNA-dependent RNA polymerase, phage-type monosubunit	23	13
33L	Unknown	13	9
35L	Exonuclease VII	32	37
36L	5'-Exonuclease most similar to N-terminal domains of DNA polymerase I	24	10
38L	DNA polymerase lacking N-terminal 5' exonuclease domain	49	31
41L	DNA primase of DnaG family	8	2
45L	The 7 k protein, inhibitor of transcription initiation and anti-terminator	11	3
48L	Unknown	26	42
49L	HNH family endonuclease with predicted AP2-like DNA-binding domain	10	3
53L	Unknown	56	58
56L	Unknown	8	2

<sup>a</sup> Standard deviation gives an estimate of error for the calculated half-lives.

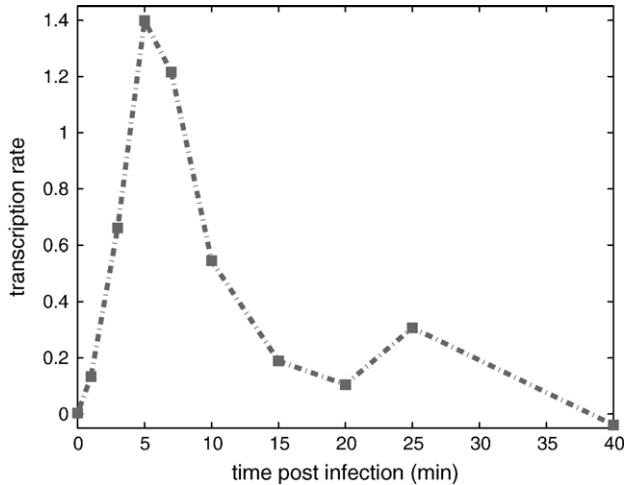


Fig. 4. Total transcription rate of L genes is shown as a function of time post-infection. The transcription rate is obtained from the kinetic model for transcript generation/decay, and from the experimental measurements of transcript abundance.

As can be seen, the L genes are transcribed at a very small rate after 15 min post-infection. This is most probably due to accumulation of p7, since leftward host RNAP promoters are inhibited by p7 in vitro (Yuzenkova et al., 2003). The L gene transcription rate reaches its maximal value at around 5 min post-infection. It is interesting that the maximal value is not reached immediately at the start of the infection. This is surprising, since host RNAP is present at this time, and there is no p7. The ‘delay’ in maximal transcription rate could be a consequence of several effects. First, a certain time may be needed for the part of the phage genome containing the L genes and their promoters to enter the cell and/or a finite time is needed for host RNAP to locate viral promoters (Gerland et al., 2002). Second, a phage-encoded activator may be needed to maximize the L genes transcription.

#### Contributions of the two RNA polymerases to transcription of early R genes

In the macroarray experiment (Semenova et al., 2005), transcript abundances of representative early R genes were determined as a function of time post-infection (Table S1). Since there are no significant differences in the estimated half-lives of individual L transcripts, we assume that decay constants of all Xp10 transcripts also take similar values. Using Eq. (8) and the data listed in Table S1, we can determine total transcription rates for all early R genes in the same way as in the previous section, using the mean of decay constants values listed in Table 1. As expected, transcription rates of individual early R genes show similar dependence on time. Average total transcription rate of early R transcripts  $T_{ER}(t)$  is shown in Fig. 5 (dashed line).

In the experiment by Semenova et al. (2005), the contribution of individual rightward promoters to transcript accumulation was measured by primer extension (Kassavetis and Geiduschek, 1982) at 0, 1, 3, 5, 7, 10, 20, 25, and 60 min

post-infection, using primers that anneal a short distance downstream from these promoters (Semenova et al., 2005). We here quantitated the primer extension signals as described in Appendix S3, and the obtained data are listed in Table S3. Using Eq. (8), the method described in the previous section, and the data in Table S3, we can calculate transcription activities of host RNAP promoters P3,  $h_{P3}(t)$ , and  $P_{UP}$ ,  $h_{PUP}(t)$ , and of phage promoters,  $p(t)$  (transcription activity of a promoter is defined as the number of transcripts initiated from this promoter per unit time). We note that since a substantial amount of transcripts initiated at the  $\phi P1$  promoter is cleaved fairly close to the 5' end (see Semenova et al., 2005) accumulation of the  $\phi P2$  promoter transcript (Table S3) was used to obtain  $p(t)$ .

It is important to note that the values of transcription activities ( $h_{P3}(t)$ ,  $h_{PUP}(t)$ , and  $p(t)$ ) cannot be directly compared to each other, because there are different (initially unknown) time-independent multiplicative constants associated with them. The differences in the constants are due to different primers being used in primer extension experiments to visualize activity of different promoters.

Our goal is to determine relative contributions of host RNAP  $H_{ER}(t)$  and phage RNAP  $P_{ER}(t)$  to the early R gene transcription rate. The  $H_{ER}(t)$  is equal to the sum of transcription activities of two host promoters (P3 and  $P_{UP}$ ).

Therefore,

$$H_{ER}(t) = \varphi_{P3} h_{P3}(t) + \varphi_{PUP} h_{PUP}(t), \quad (9)$$

and

$$P_{ER}(t) = \varphi_p p(t). \quad (10)$$

Here,  $\varphi_{P3}$ ,  $\varphi_{PUP}$ , and  $\varphi_p$  are unknown ‘scaling’ constants that have to be determined so that activities of different promoters can be compared with each other (see the above discussion). These constants also allow us to connect transcription activities of individual promoters (inferred from primer extension measurements in Table S3), with the total transcription activity

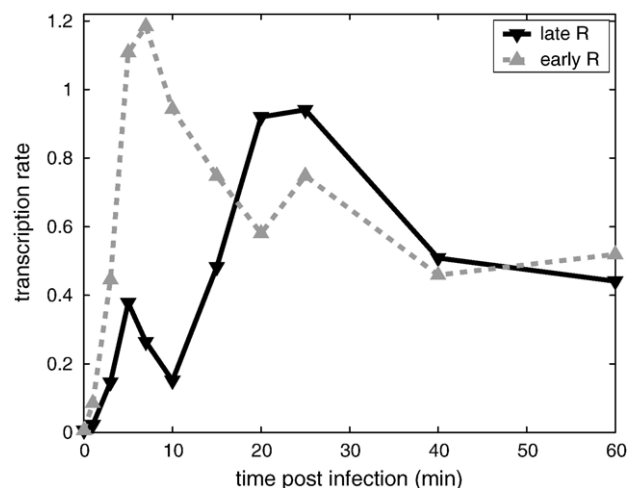


Fig. 5. Total transcription rates of early R and late R genes are shown as a function of time post-infection. The transcription rates are obtained similarly as in Fig. 4.



of early R genes (inferred from the genomic array measurements in Table S1).

To determine  $\varphi_{P_3}$ ,  $\varphi_{P_{up}}$ , and  $\varphi_p$ , we assume that the total transcription rate of early R transcripts,  $T_{ER}(t)$ , has to be a linear superposition of transcription activities of individual rightward promoters

$$T_{ER}(t) = \varphi_{P_3} h_{P_3}(t) + \varphi_{P_{up}} h_{P_{up}}(t) + \varphi_p p(t). \quad (11)$$

The linear superposition used above implies that the amount of transcripts that are (potentially) terminated before the early R operon does not depend on time. Since  $T_{ER}(t)$ ,  $h_{P_3}(t)$ ,  $h_{P_{up}}(t)$ , and  $p(t)$  are known, we can determine the unknown constants from Eq. (11) through a linear fit. The procedure used to fit the parameters is described in Appendix S1. The  $H_{ER}(t)$  and  $P_{ER}(t)$ , determined through Eqs. (9) and (10), are shown in Fig. 6A.

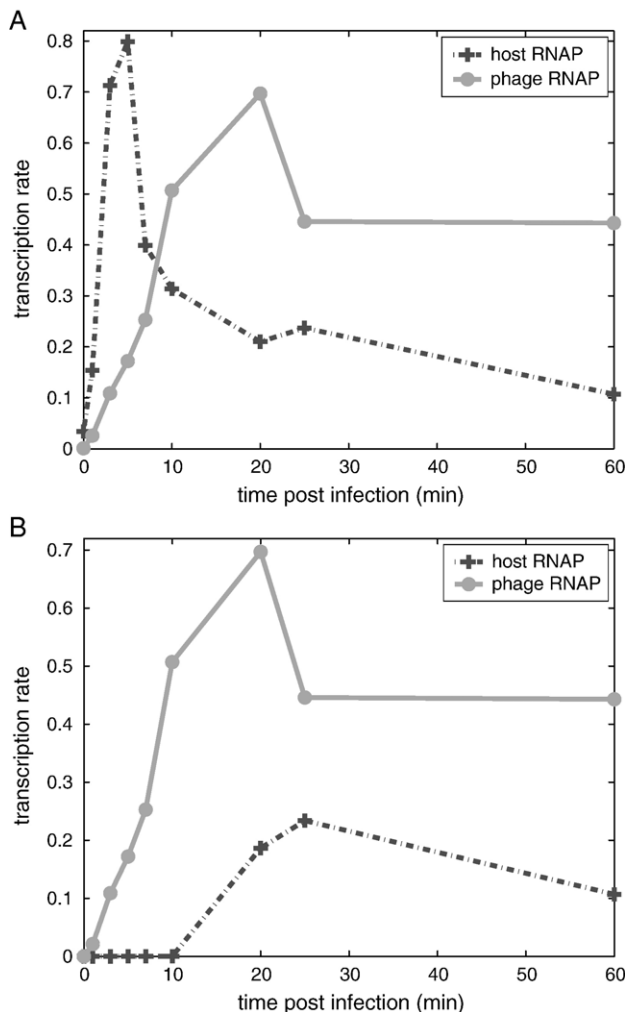


Fig. 6. Contributions of two RNA polymerases to transcription of early R and late R genes. (A) Contributions of host RNAP and phage RNAP to total early R transcription rate are shown as a function of time post-infection. The contributions are obtained by using promoter transcription activities (calculated by using the kinetic model and primer extension data) together with the results shown in Fig. 5. (B) Contributions of host RNAP and phage RNAP to total late R transcription rate are shown as a function of time post-infection. The contributions are computed in a similar way as for early R genes, except that we used the model which accounts for time-dependent anti-termination.

Contrary to the case of the L genes, host RNAP does not stop transcribing early R genes late in infection. However, compared to the maximum value reached 5 min post-infection, host RNAP transcription of early R genes decreases later in infection. This is most likely due to accumulation of p7, since p7 partially inhibits transcription from P3 in vitro (Yuzenkova et al., 2003). More precisely, the in vitro measurements showed that p7 reduces transcription from P3 for about 60%, which is roughly in agreement with our in vivo prediction shown in Fig. 6A.

In the beginning of infection, transcription rate by phage RNAP has a much smaller value than transcription rate by host RNAP, but keeps increasing so that it reaches a value comparable to transcription rate by host RNAP at around 7 min post-infection and becomes significantly larger 10 min post-infection and later. Transcription rate by phage RNAP reaches its maximum at around 20 min post-infection and remains constant after that time. Increase of transcription rate by phage RNAP in the first 20 min of infection must result from the increase of the amount of phage RNAP in the infected cell.

The fact that the predicted transcription rate by phage RNAP remains constant after 20 min post-infection is surprising, since even though transcription of the Xp10 RNAP gene (an L gene) becomes very small after 15 min post-infection, translation of its mRNA should continue, resulting in continued increase (albeit at a decreased rate) of the amount of phage RNAP even after 15 min post-infection. Therefore, the absence of increase of phage RNAP transcription rate after 20 min post-infection may be due to an inhibitory mechanism similar, for example, to the one described for T7 phage, where phage-encoded lysozyme (gp 3.5) binds to and attenuates transcription by phage RNAP (Molineux, 2005). Finally, the potential interference of late transcription (by either RNAP) by DNA replication may also account for the observed reduction in late transcription rates.

#### Contributions of the two RNA polymerases to transcription of late R genes

The average (total) transcription rate of late R genes,  $T_{LR}(t)$ , can be determined from measured transcript abundances (Table S1) using Eq. (8) and the method described for early R genes. The results are shown in Fig. 5 (solid line). The differences in total transcription rates for the two classes of R genes could be attributed to the presence of bioinformatically predicted terminator that separates the late R genes from the early R genes and by transcription anti-termination due to accumulation of p7, a product of an L gene. Since accumulation of p7 should allow the increasing number of RNAP molecules to transcribe late R genes, the observed amount of late R transcripts is not equal to a linear combination of transcription activities of rightward promoters. Therefore, in order to determine host RNAP  $H_{LR}(t)$  and phage RNAP  $P_{LR}(t)$  contribution to transcription rate of late R genes, we need to introduce a model that takes into account initial transcription termination and subsequent anti-termination by p7 protein. We define ‘anti-termination efficiency’ as a fraction of transcripts that read

through a terminator at a given time post-infection. We model anti-termination efficiency  $A$ , by a sigmoidal function:

$$A(C_{p7}(t)) = \frac{1}{1 + (C_0/C_{p7}(t))^v} \tag{12}$$

Here,  $C_{p7}$  is the concentration of the p7 protein (which increases during infection).  $C_0$  and  $v$  are initially unknown constants that describe different levels of sensitivity of  $A$  to  $C_{p7}$ . Note that we indicate protein concentrations by capital letters. Also, note that anti-termination efficiencies for phage and host RNAPs do not have to be the same. Hence, two separate expressions,  $A_H$  (for host RNAP) and  $A_P$  (for phage RNAP), have to be introduced. These expressions are characterized by different values of constants  $C_0$  and  $v$ .

According to our model,  $H_{LR}(t)$  and  $P_{LR}(t)$  on the one side, and  $H_{ER}(t)$  and  $P_{ER}(t)$  on the other side are related through the following expressions:

$$H_{LR}(t) = H_{ER}(t) \frac{1}{1 + (C_{0,H}/C_{p7}(t))^{v_H}} \tag{13}$$

and

$$P_{LR}(t) = P_{ER}(t) \frac{1}{1 + (C_{0,P}/C_{p7}(t))^{v_P}} \tag{14}$$

The dependence of  $C_{p7}(t)$  on time can be easily determined from experimentally measured amounts of the p7 transcript  $c_{p7}(t)$ . If we assume that the p7 protein remains stable during infection, we have that:

$$C_{p7}(t) = k_t \int_0^t c_{p7}(\xi) d\xi, \tag{15}$$

where  $k_t$  is the (unknown) p7 translation rate.

Next, we use that the total transcription rate of late R genes  $T_{LR}(t)$  has to be equal to the sum of  $H_{LR}(t)$  and  $P_{LR}(t)$ :

$$\begin{aligned} T_{LR}(t) &= H_{LR}(t) + P_{LR}(t) \\ &= H_{ER}(t) \frac{1}{1 + [C_{0,H}/C_{p7}(t)]^{v_H}} \\ &\quad + P_{ER}(t) \frac{1}{1 + [C_{0,P}/C_{p7}(t)]^{v_P}} \end{aligned} \tag{16}$$

Since it should take less than 1 min for both Xp10 and host RNAP to transcribe the ~3 kbp that separate host and phage RNAP promoters from late R genes (based on the fact that T7 RNAP and *E. coli* RNAP synthesize RNA at about 250 bp/s and 50 bp/s, respectively (Molineux, 2005)), we neglect transcriptional delay in Eq. (16).

We have already determined  $T_{LR}(t)$ ,  $H_{ER}(t)$ ,  $P_{ER}(t)$ , and  $C_{p7}(t)$ , so Eq. (16) can be used to fit unknown parameters  $C_{0,H}$ ,  $v_H$ ,  $C_{0,P}$  and  $v_P$  (note that in Eq. (16)  $C_{0,H}$  and  $C_{0,P}$  can be re-scaled by  $k_t$ ). Since there is a total of eight data points, it is possible to determine four unknown parameters through nonlinear fit. Since phage RNAP is relatively insensitive to bacterial RNAP termination signals (Molineux, 2005), we search for a fixed point (i.e., for a solution) that corresponds to only one of the two RNA

polymerases (either host or phage) being terminated by the termination signal. The procedure used to fit the parameters is described in Appendix S1. The contributions of host and phage RNAP to late R gene transcription rates ( $H_{LR}(t)$  and  $P_{LR}(t)$ ) that correspond to the obtained fixed point are shown in Fig. 6B. Corresponding anti-termination efficiencies for host RNAP and phage RNAP ( $A_H(t)$  and  $A_P(t)$ ) are shown in Fig. 7. From our analysis (see Fig. 7), it follows that the terminator is efficient in terminating host RNAP, while phage RNAP is not sensitive to the termination signal. We point out that transcription rates shown in Fig. 6A can be directly compared to those shown in Fig. 6B. We finally note that the apparent small lag needed for phage RNAP to reach 100% anti-termination efficiency shown in Fig. 7 is the consequence of the fact that in the first 3 min of infection phage RNAP transcription activity is close to zero. Therefore, the small lag does not have biological significance, since any value of anti-termination efficiency would lead to zero transcripts reading through the terminator in the first 3 min post-infection.

The results shown in Fig. 6B indicate that transcription termination completely prevents expression of late R genes by host RNAP in the beginning of the infection, but that accumulation of an anti-terminator (most probably p7, though we cannot exclude that additional L gene products may also have such a function) allows host RNAP transcription of late R genes to occur after 10 min post-infection (Fig. 7). Later in infection, transcription rates by host RNAP for early and late R genes become equal (see Figs. 6A and B).

*Relation between the amounts of synthesized phage proteins and phage progeny number*

Yuzenkova et al. (2003) performed an experiment in which Rif was added at various times post-infection and the resulting phage yield was measured. The predicted transcription rate of the L genes (see Fig. 4) indicates that they are transcribed at a

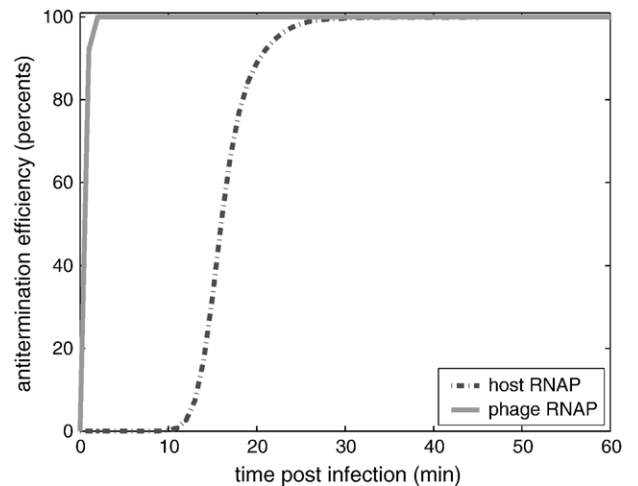


Fig. 7. Anti-termination efficiencies, corresponding to the terminator that separates early R and late R genes, are shown as a function of time post-infection. The two curves correspond to anti-termination efficiencies for host and phage RNAP. The anti-termination efficiencies are obtained from the same computational analysis as in Fig. 6B.

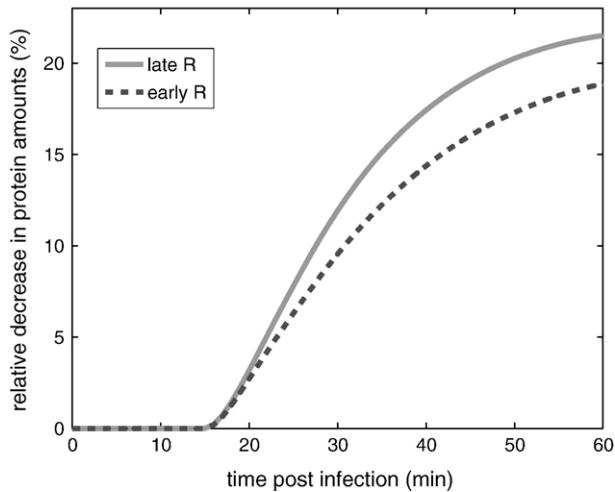


Fig. 8. Predicted relative decreases in protein amounts, resulting from the inhibition of host RNAP at times later than 15 min post-infection, are shown. The two curves correspond to relative decreases in protein amounts for early R and late R genes. The predictions are obtained from the results shown in Fig. 6 and from the kinetic model of phage protein synthesis.

very small rate at times later than 15 min post-infection, so the addition of Rif after the 15-min time point should influence only the expression of the R genes. Yuzenkova et al. (2003) reported that the addition of Rif between 17 and 40 min post-infection leads to significant reduction in phage progeny. In particular, the addition of Rif at 17 min reduces phage yield by 70%. It therefore necessarily follows that transcription of the R genes by host RNAP must significantly contribute to phage fitness. Note that in this argument we assumed that the addition of Rif at 17 min post-infection does not effect transcription of R genes by phage RNAP. This is implicitly based on an assumption that interactions/collisions between transcribing phage RNA polymerases and host RNAP which is either transcribing DNA (in the absence of Rif) or trapped in promoter complexes by Rif do not significantly impair phage RNAP transcription. Note that, since promoter complexes are much less stable than elongating transcription complexes (Saecker et al., 2002), potential inhibitory interactions would be expected to decrease in the presence of Rif. Therefore, even if the inhibitory interactions would exist, it is plausible to expect that in the presence of Rif transcription rate by phage RNAP would effectively increase. Consequently, our argument that the transcription by host RNAP must significantly contribute to phage fitness would be even more valid in such case.

Further, since Figs. 6A and B show that phage RNAP is responsible for most of transcription after 15 min, it follows that inhibition of phage RNAP would have a severe effect on phage progeny production. That is, the measured ~70% decrease in the number of phage progeny presents a lower bound estimate for expected decrease in phage yield if phage RNAP did not transcribe R genes. In summary, if either of the two RNA polymerases would not transcribe R genes, a severe reduction in the phage progeny numbers will be observed.

We next ask what is the connection between the synthesized amount of proteins and the number of produced phage progeny particles. Since we have previously estimated the relative

contributions of host and phage RNAP to R gene transcription, we can predict, for both classes of R genes, the relative decrease in the encoded proteins amounts  $\chi_{\text{Rif}}(t)$  caused by the addition of Rif 15 min post-infection. Note that the 15-min time point corresponds to the time after which the L genes are transcribed at a very low rate, and that the effect of Rif addition on phage progeny amount was experimentally measured at a time close to 15 min (i.e., at 17 min). Calculation of  $\chi_{\text{Rif}}(t)$  for early R and late R genes is given in Appendix S2, and the results are shown in Fig. 8. The results indicate that connection between the decrease in phage progeny numbers and the predicted decrease in the amounts of phage structural proteins is highly nonlinear. That is, a less than 20% decrease in protein amounts (see Fig. 8) leads to as much as a ~70% decrease in phage yield. Note that, similarly as above, potential interference of host RNAP with phage RNAP transcription would make this conclusion even more drastic. It is therefore likely that at least for some structural phage proteins, protein amounts have to reach certain ‘soft’ threshold in order for phage particles to be formed. Similar results were obtained during the in vitro study of assembly kinetics of coat and scaffolding protein subunits in procapsid shells of bacteriophage P22 (Prevelige et al., 1993).

#### Acknowledgments

This work was supported by NIH Grants GM67794 (to B.S.) and GM59295 (to K.S.). Final parts of this work were supported from NSF under Agreement No. 112050 and NSF grant MCB-0418891. E.S. was partially supported by Charles and Johanna Busch postdoctoral fellowship. We are grateful to Drs. Sergei Nechaev, Harmen Bussemaker and Istok Mendas for critical reading of the manuscript. M.D. acknowledges the hospitality of the Kavli Institute for Theoretical Physics, where part of this work was done.

#### Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.virol.2006.05.038.

#### References

- Allison, L.A., Simon, L.D., Maliga, P., 1986. Deletion of rpoB reveals a second distinct transcription system in plastids of higher plants. *EMBO J.* 15, 2802–2809.
- Bernstein, J.A., Khodursky, A.B., Lin, P.H., Lin-Chao, S., Cohen, S.N., 2002. Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. *Proc. Natl. Acad. Sci.* 99, 9697–9702.
- Brussow, H., Hendrix, R.W., 2002. Phage genomics: small is beautiful. *Cell* 108, 13–16.
- Djordjevic, M., Sengupta, A.M., Shraiman, B.I., 2003. A biophysical approach to transcription factor binding site discovery. *Genome Res.* 13, 2381–2390.
- Ermolaeva, M.D., Khalak, H.G., White, O., Smith, H.O., Salzberg, S.L., 2000. Prediction of transcription terminators in bacterial genomes. *J. Mol. Biol.* 301, 27–33.
- Gerland, U., Moroz, D.J., Hwa, T., 2002. Physical constraints and functional characteristics of transcription factor–DNA interaction. *Proc. Natl. Acad. Sci.* 99, 12015–12020.
- Hajdukiewicz, T.J., Allison, L.A., Maliga, P., 1997. The two RNA polymerases

- encoded by the nuclear and the plastid compartments transcribe distinct groups of genes in tobacco plastids. *EMBO J.* 16, 4041–4048.
- Hendrix, R.W., 2003. Bacteriophage genetics. *Curr. Opin. Microbiol.* 6, 506–511.
- Jeng, S.T., Gardner, J.F., Gumpert, R.I., 1990. Transcription termination by bacteriophage T7 RNA polymerase at rho-independent terminators. *J. Biol. Chem.* 265, 3823–3830.
- Kassavetis, G.A., Geiduschek, E.P., 1982. Bacteriophage T4 late promoters: mapping 5' ends of T4 gene 23 mRNAs. *EMBO J.* 1, 107–114.
- Lawrence, C.E., Altschul, S.F., Bogouski, M.S., Liu, J.S., Neuwald, A.F., Wooten, J.C., 1993. Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignment. *Science* 262, 208–214.
- Macdonald, L.E., Durbin, R.K., Dunn, J.J., McAllister, W.T., 1994. Characterization of two types of termination signal for bacteriophage T7 RNA polymerase. *J. Mol. Biol.* 238, 145–158.
- Marchand, I., Nicholson, A.W., Dreyfus, M., 2001. Bacteriophage T7 protein kinase phosphorylates RNase E and stabilizes mRNAs synthesized by T7 RNA polymerase. *Mol. Microbiol.* 42, 767–776.
- Molineux, I., 2005. The T7 group. In: Calendar, R. (Ed.), *Bacteriophages*. Oxford Univ. Press.
- Nechaev, S., Yuzenkova, Y., Niedziela-Majka, A., Heyduk, T., Severinov, K., 2002. A novel bacteriophage-encoded RNA polymerase binding protein inhibits transcription initiation and abolishes transcription termination by host RNA polymerase. *J. Mol. Biol.* 320, 11–22.
- O'Hara, E.B., Chekanova, J.A., Ingle, C.A., Kushner, Z.R., Peters, E., Kushner, S.R., 1995. Polyadenylation helps regulate mRNA decay in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* 92, 1807–1811.
- Prevelige Jr, P.E., Thomas, D., King, J., 1993. Nucleation and growth phases in the polymerization of coat and scaffolding subunits into icosahedral procapsid shells. *Biophys. J.* 64, 824–835.
- Saecker, R.M., Tsodikov, O.V., McQuade, K.L., Schlax, P.E., Capp, M.W., Record Jr., M.T., 2002. Abstract kinetic studies and structural models of the association of *E. coli* sigma(70) RNA polymerase with the lambdaP(R) promoter: large scale conformational changes in forming the kinetically significant intermediates. *J. Mol. Biol.* 319, 649–671.
- Semenova, E., Djordjevic, M., Shraiman, B., Severinov, K., 2005. The tale of two RNA polymerases: transcription profiling and gene expression strategy of bacteriophage Xp10. *Mol. Microbiol.* 55, 764–777.
- Ueno, H., Yonesaki, T., 2004. Phage-induced change in the stability of mRNAs. *Virology* 10, 134–141.
- Yuzenkova, Y., Nechaev, S., Berlin, J., Rogulja, D., Kuznedelov, K., Schloss, M., Inman, R., Mushegian, R., Severinov, K., 2003. Genome of *Xanthomonas oryzae* bacteriophage Xp10: an odd T-odd phage. *J. Mol. Biol.* 330, 735–748.
- Zivin, R., Zehring, W., Rothman-Denes, L.B., 1981. Transcriptional map of bacteriophage N4. Location and polarity of N4 RNAs. *J. Mol. Biol.* 152, 335–356.

### Further reading

MATLAB (R14), 2005. The MathWorks, Inc.