doi:10.1016/j.jmb.2012.01.002 J. Mol. Biol. (2012) 416, 389–399

Contents lists available at www.sciencedirect.com

Journal of Molecular Biology

journal homepage: http://ees.elsevier.com.jmb

Temporal Regulation of Gene Expression of the *Escherichia coli* Bacteriophage phiEco32

Olga Pavlova[1], Daria Lavysh[2,3], Evgeny Klimuk[2,3], Marko Djordjevic[4], Dmitry A. Ravcheev[5,6], Mikhail S. Gelfand[5,7], Konstantin Severinov[1,2,3]* and Natalja Akulenko[1,2,3]

[1]*Waksman Institute for Microbiology, Piscataway, NJ 08854, USA*
[2]*Institute of Molecular Genetics, Russian Academy of Sciences, Moscow 123182, Russia*
[3]*Institute of Gene Biology, Russian Academy of Sciences, Moscow 117312, Russia*
[4]*Institute of Physiology and Biochemistry, Faculty of Biology, University of Belgrade, Belgrade, Serbia*
[5]*Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow 127994, Russia*
[6]*Sanford–Burnham Medical Research Institute, La Jolla, CA 92037, USA*
[7]*Faculty of Bioengineering and Bioinformatics, M. V. Lomonosov Moscow State University, Moscow 119992, Russia*

Received 29 November 2011;
received in revised form
2 January 2012;
accepted 2 January 2012
Available online
10 January 2012

Edited by M. Gottesman

Keywords:
bacteriophage;
genome;
RNA polymerase;
σ factor;
transcription regulation

Escherichia coli phage phiEco32 encodes two proteins that bind to host RNA polymerase (RNAP): gp79, a novel protein, and gp36, a distant homolog of σ^{70} family proteins. Here, we investigated the temporal pattern of phiEco32 and host gene expression during infection. Host transcription shutoff and three distinct bacteriophage temporal gene classes (early, middle, and late) were revealed. A combination of bioinformatic and biochemical approaches allowed identification of phage promoters recognized by a host RNAP holoenzyme containing the σ^{70} factor. These promoters are located upstream of early phage genes. A combination of macroarray data, primer extension, and *in vitro* transcription analyses allowed identification of six promoters recognized by an RNAP holoenzyme containing gp36. These promoters are characterized by a single-consensus element tAATGTAtA and are located upstream of the middle and late phage genes. Curiously, gp79, an inhibitor of host and early phage transcription by σ^{70} holoenzyme, activated transcription by the gp36 holoenzyme *in vitro*.

Introduction

Phages are the most abundant and variable organisms on Earth.[1] Despite this abundance, it is rather uncommon to find a totally new phage infecting *Escherichia coli*, the best-studied model bacterium. Instead, most new isolates of *E. coli* phages tend to be close relatives of already known *E. coli* phages that were extensively studied during the second half of the 20th century. We described the isolation and genomic and proteomic characterization of a novel *E. coli* phage, phiEco32.[2] This phage was isolated in Tbilisi, Georgia, in 2004 and was the

**Corresponding author.* Waksman Institute for Microbiology, 190 Frelinghuysen Road, Piscataway, NJ 08854, USA. E-mail address: severik@waksman.rutgers.edu.

Present address: O. Pavlova, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA.

Abbreviations used: RNAP, RNA polymerase; *PSII*, photosystem II; MOI, multiplicity of infection; ORF, open reading frame; SSC, saline–sodium citrate.

only phage present in the large collection of *E. coli* and *Salmonella* phages from the Eliava Bacteriophage Institute in Tbilisi that appeared to be unrelated to previously reported viruses. PhiEco32 is a lytic phage belonging to the Podoviridae family. PhiEco32 was isolated on *E. coli 55*, a natural strain isolated from a cow affected by mastitis. Under laboratory conditions in rich media at 37 °C, phiEco32 lyses its host in 30–35 min.[2] While phiEco32 does not infect most laboratory strains of *E. coli*, it was found to lyse more than 95% of *E. coli* strains isolated from cows suffering from acute mastitis in Georgia and thus can potentially be used, together with other phages, in therapeutic phage cocktails.

When a complete linear 77,554-bp genome of phiEco32 was sequenced, the phage appeared to be unique.[2] However, since then, a DNA fragment with a sequence practically identical with that of a phiEco32 genomic fragment has serendipitously been found in a fecal sample collected from a fatal case of human gastroenteritis in Hungary (G. Reuter, personal communication). In addition, *Serratia marcescens* phage KSP100, which was recently isolated from environmental water in Kochi Prefecture, Japan;[3] *Salmonella enterica* phage 7–11;[4] and *Pseudomonas aeruginosa* phage PaP3 also contain genomic segments homologous to phiEco32 sequences.[5] These observations suggest that phages such as phiEco32 are widespread. However, nothing is known about the regulation of gene expression during infection by these phages.

Analysis of the gene expression strategies of various bacteriophages provided important insights into the basic mechanisms of genetic regulation. Since phiEco32 is like no other phage that has been studied to date, it is of considerable interest to determine how it controls the coordinated temporal expression of its genes and how it affects the gene expression of its host during infection. PhiEco32 has 128 annotated genes, with about one-third of the genome transcribed rightward and with the rest transcribed leftward. The phage genome does not encode an RNA polymerase (RNAP) but encodes a putative σ-factor gp36 and a small protein gp79 that inhibits transcription by the host σ^{70} RNAP holoenzyme *in vitro*.[2] These two proteins, as well as other yet unidentified transcription factors, may participate in the regulated expression of phiEco32 genes. In this report, we describe the combined use of bioinformatic and biochemical approaches to characterize the temporal patterns of phiEco32 and host gene expression during infection. We identify promoters recognized by the RNAP holoenzyme containing gp36 and reveal that gp79 is a dual-function regulator that inhibits transcription by the host σ^{70} holoenzyme while activating transcription by the gp36 RNAP holoenzyme.

Results

Analysis of phiEco32 gene expression using macroarray

A macroarray containing spots that corresponded to 24 representative phiEco32 genes was prepared. The genes were selected to cover most of the likely phage operons (defined here as phiEco32 genes transcribed in the same direction and separated by no more than 60 nt of noncoding DNA). In most cases, the most upstream genes in putative phage operons were selected for the macroarray (Fig. 1). Spots containing total phiEco32 DNA, total host DNA, and several host gene-specific probes (*rpoD*, *rpoB*, *uspA*, and *treB*) were also included in the array. As loading and normalization control, two spots containing a PCR fragment corresponding to a photosystem II (*PSII*) core subunit plant gene were used.

E. coli 55 cultures at the midlog phase of growth were infected with phiEco32 at a high multiplicity of infection (MOI), and total RNA was extracted from aliquots of infected cultures 5, 10, 25, and 35 min postinfection (previous work indicated that phiEco32-infected cells lyse ~40 min postinfection[2]). As control, we used RNA prepared from cells collected immediately before the infection (a 0-min time point). Equal amounts of total RNA from each time point were combined with the *PSII* probe and used to generate radioactively labeled cDNA by random priming/reverse transcription, followed by hybridization to the array. Two independent infections were analyzed in this way.

The amount of radioactivity in each array spot reflects the abundance of the corresponding transcript at a certain time point through infection. The signals from spots corresponding to phage genes *124* and *127* were very close to background throughout the infection in both experiments and were consequently excluded from further analysis. Radioactive signals from each remaining spot were corrected for background and normalized based on the relative strength of the *PSII* spot signal. The normalized signals from the two individual infections were averaged. As can be seen from Fig. 2a, the amount of total *E. coli 55* transcripts decreased through infection, reaching the background level 25 min postinfection. In contrast, the amount of total phiEco32 transcripts increased at the beginning of infection (up until 25 min) but started to decrease later (from 25 to 35 min). Based on these results, we conclude that phiEco32 shuts off the transcription of host genes sometime around 10 min postinfection.

For an easier comparison of the temporal patterns of individual phiEco32 transcripts, the averaged signal intensities were scaled to make equal the mean transcript abundances for each gene. The scaled signal intensities of individual phiEco32 transcripts

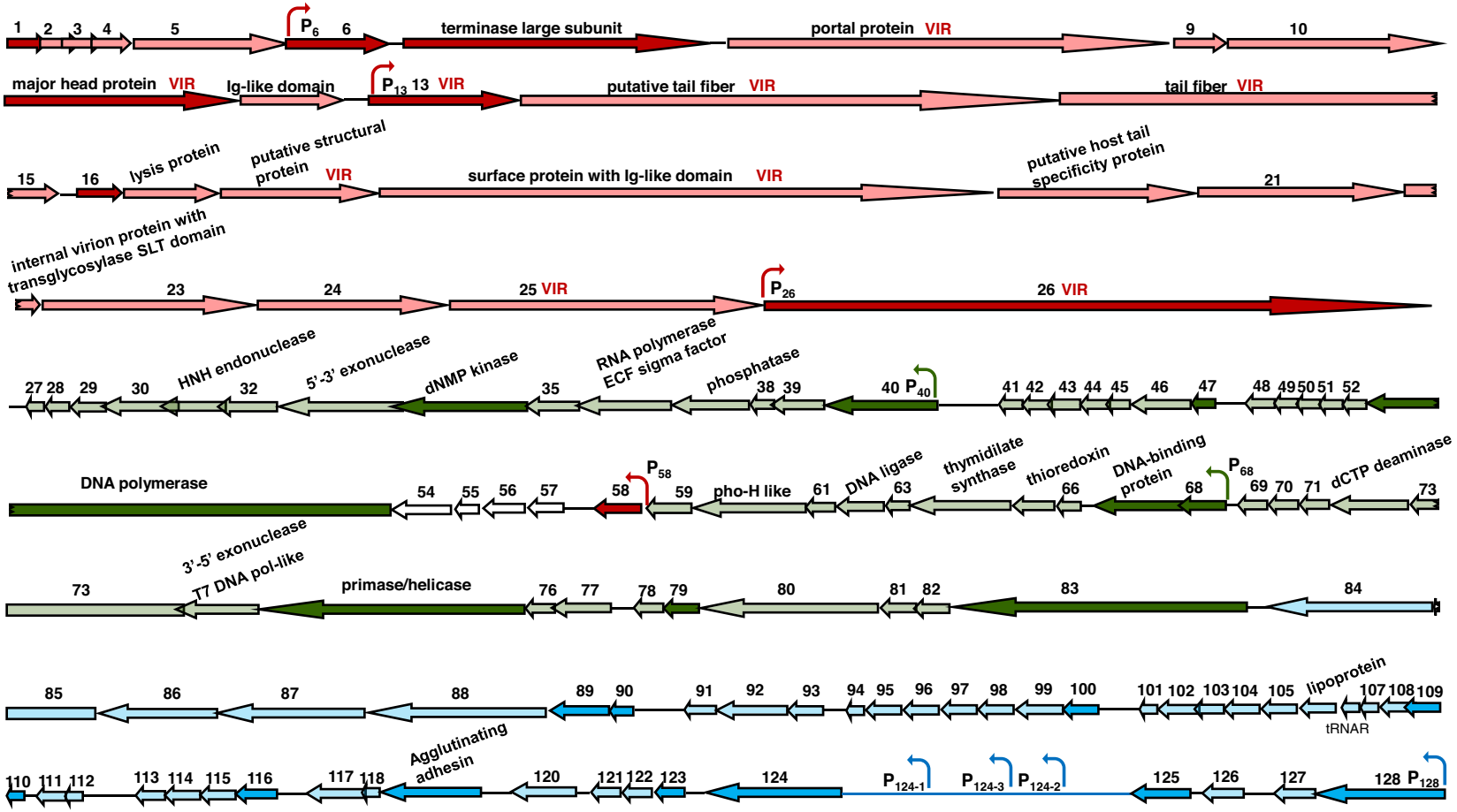
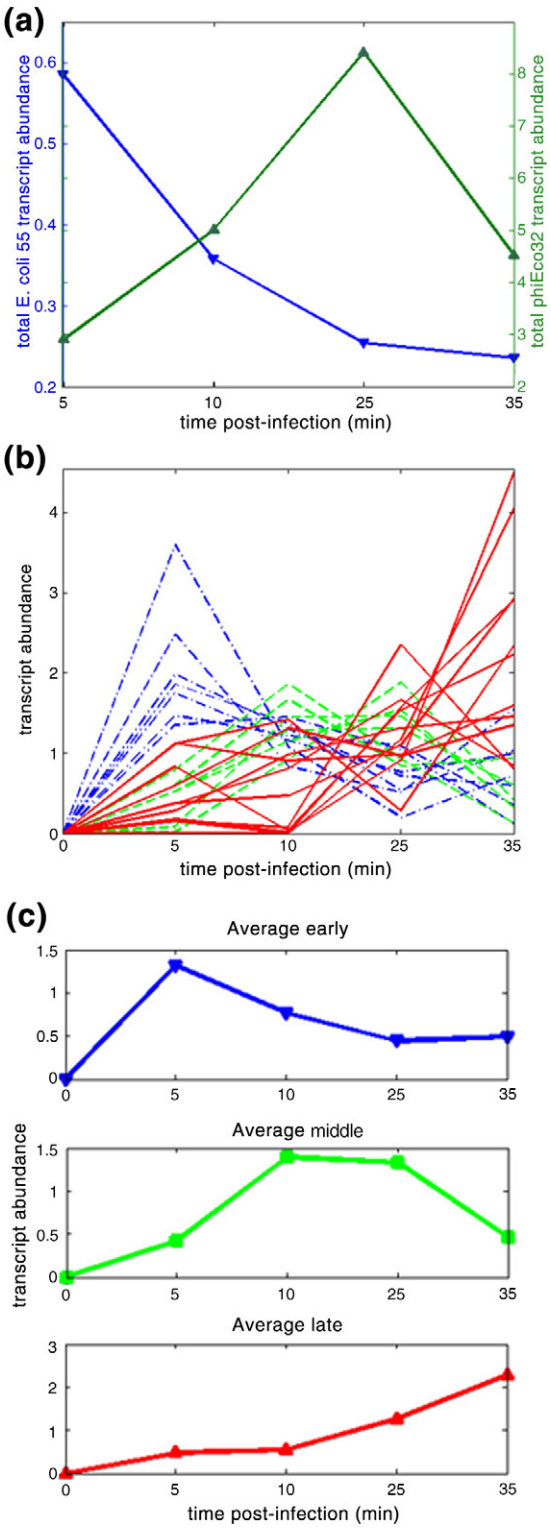


Fig. 1. Temporal classes of bacteriophage phiEco32 genes. The genome of phiEco32 phage is schematically presented. Arrows represent annotated genes; the direction of the arrow indicates the direction of transcription. Intergenic regions of 60 bp or more are shown as a plain line. Genes belonging to different temporal classes are indicated in blue (early genes), green (middle genes), and red (late genes). Genes assigned to specific temporal classes based on experimental data (primer extension and/or macroarray analysis) are shown in dark colors. Genes assigned to specific temporal classes based on a predicted operonic structure of the phiEco32 genome are shown in light colors. Genes for which no assignments can be made based on data available are shown in white. Promoters are shown as bent arrows; the color of a promoter arrow indicates its temporal class. The putative functions of phiEco32 genes, when known, are indicated. "Vir" indicates proteins present in the phiEco32 virion.[2]

were then plotted as a function of time postinfection (Fig. 2b). Temporal patterns of phiEco32 genes were systematically clustered in three different classes as described in Materials and Methods. PhiEco32 transcripts corresponding to these classes are indicated by different colors in Fig. 2b. For clarity, the averages of scaled abundances calculated for each of the three temporal classes of phage genes are also shown as separate panels in Fig. 2c. As can be seen, the three classes are clearly distinguished by the time interval during which the transcript abundance reaches the maximal level. For the first (early) class, most of the transcripts are accumulated during the first 10 min of infection. Transcripts of the second (middle) class reach the maximal level between 10 and 25 min postinfection. Finally, the abundance of transcripts from the third temporal class, which we classified as phiEco32 late transcripts, is highest between 25 and 35 min postinfection.



The genomic positions of phiEco32 genes, which belong to different temporal classes, are shown in Fig. 1. Although the clustering of phage transcripts did not use any information on gene position or gene functional annotation, these features clearly correlate with the clustering results. Genes *1–26*, which comprise all rightward-transcribed genes, belong to the late class. Late genes with known functions encode exclusively structural and DNA-packaging proteins of the phage. A stretch of phiEco32 genes from gene *27* to gene *83* includes all genes whose products are predicted to be involved in DNA replication, recombination, and nucleotide metabolism. Most genes from this stretch that are present in the array belong to the middle class. However, four genes—*47*, *67*, *68*, and *36* (encodes gp36, the putative σ factor of the phage)—behaved as late genes based on macroarray analysis. For several reasons, including primer extension data presented in the following sections, we nevertheless consider these genes as middle, and they are labeled as such in Fig. 1. The apparent "late" behavior of these gene transcripts could be due to low rates of transcript degradation at the late stages of infection, rather than the late onset of their transcription.

The remaining leftward-transcribed genes of phiEco32 (*84–128*) belong to the early class. This class includes most phiEco32 genes of unknown function. The early transcribed region of the phiEco32 genome also includes the 1945-bp-long noncoding region between genes *124* and *125*. We investigated the transcription of this region in more detail. A small array containing a spot with a product

Fig. 2. Macroarray analysis of host and viral gene expression during phiEco32 infection. (a) The abundance of total phiEco32-encoded transcripts (green line) is shown together with the abundance of total transcripts encoded by *E. coli 55* (red line). Since the two abundances have very different ranges, a double-axis plot is used. (b) Transcript abundances are presented for individual phiEco32 transcripts as a function of time. Colors match those used in Fig. 1 (red lines, late genes; green lines, middle genes; blue lines, early genes). (c) Average transcript abundances corresponding to the three temporal classes or phiEco32.

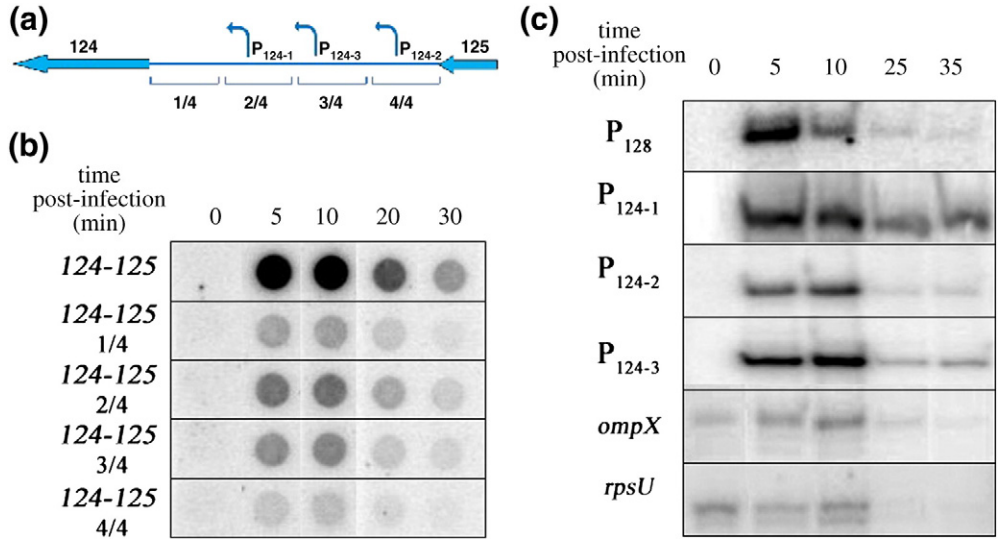


Fig. 3. Mapping of early phiEco32 promoters. (a) Schematic presentation of the *124–125* intergenic region. The relative positions of fragments 1/4, 2/4, 3/4, and 4/4 used as probes on the array are shown below. Bent leftward arrows represent early promoters P_{124-1}, P_{124-2}, and P_{124-3}. (b) The representative result of an experiment with an array of fragments of the *124–125* intergenic region hybridizing to the total RNA prepared from cells collected at various time points postinfection. (c) The results of the primer extension analysis of total RNA prepared from cells collected at various time points postinfection using primers designed to reveal transcription from individual phiEco32 early promoters and selected host promoters.

of the PCR amplification of the entire region, four spots each roughly corresponding to a quarter of the intergenic region (1/4 through 4/4, where 1/4 corresponds to the part closest to gene *124* and 4/4 corresponds to the part closest to gene *125*; Fig. 3a), and two control spots corresponding to *PSII* plant gene was prepared and analyzed. The results are presented in Fig. 3b. As can be seen, the entire *124–125* intergenic region is transcribed, and the transcript pattern matches that of early phage genes.

Predictions of putative σ^{70}-dependent promoters in the phiEco32 genome

PhiEco32 encodes a putative σ-factor gp36 but does not encode a recognizable RNAP.[2] We hypothesized that there are (at least) two types of promoters in the phage genome: early promoters dependent on the *E. coli* RNAP holoenzyme containing the primary σ^{70} factor and middle and/or late promoters dependent on the holoenzyme containing gp36. To identify σ^{70}-dependent promoters, we used a positional weight matrix (pattern) describing the −10 and −35 elements of *E. coli* σ^{70} promoters. The matrix was constructed using the DPInteract database collection of known *E. coli* σ^{70} promoters.[6] The promoter pattern was used to search the phiEco32 genome with the GenomeExplorer program package.[7] We used the following search parameters: (i) intergenic regions extending from −200 to +25 with respect to an annotated translation start point of predicted phiEco32 genes were analyzed; (ii) the spacer length between the −10 and the −35 promoter elements was allowed to vary from 16 to 19 bp; (iii) the sequence and the length of the spacer did not influence the search; and (iv) the search was bidirectional. A conservative score cutoff was chosen, since promoter prediction methods typically lead to a large number of false positives.[8]

Only four σ^{70}-dependent promoters were predicted by the search with the specified parameters. All four predicted promoters are located at the right end of the genome and display a leftward orientation. One of these promoters is located upstream of gene *128*, which is the upstreammost gene in the leftward-transcribed gene cluster. Three additional closely located putative promoters were predicted in the long intergenic region between genes *124* and *125*. Based on array data, one can assume that predicted σ^{70} promoters are responsible for the transcription of early phage genes.

Since our bioinformatic search resulted in only four predicted σ^{70}-dependent promoters, we searched for additional promoters by extending the search region from −200 to +75 (so that a 75-bp overlap with coding sequence is allowed). This search resulted in five intragenic putative promoters, with two predictions appearing particularly significant: (i) a predicted promoter within gene *37*, which could be used for the transcription of the phiEco32 putative σ-factor gene *36*; and (ii) a predicted promoter within gene *53*, which has the highest score among all predicted σ^{70}-dependent promoters of the phage.

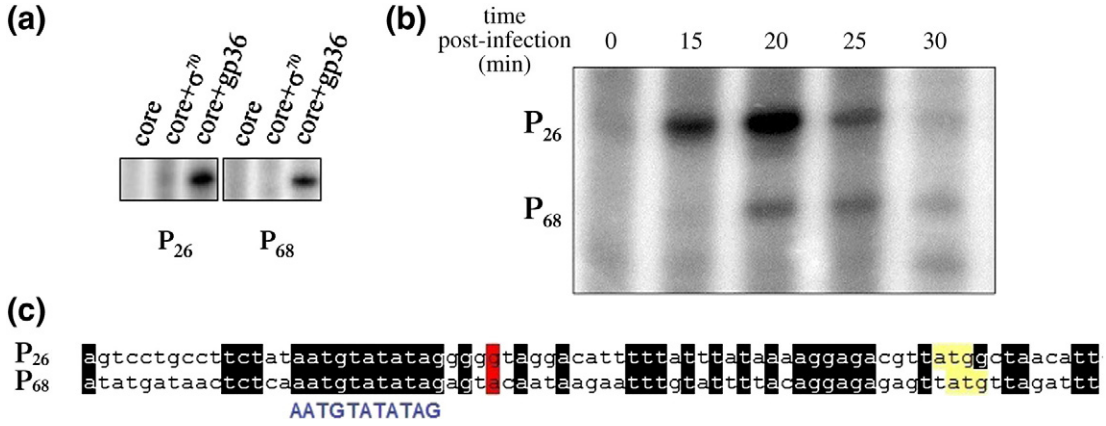


Fig. 4. Experimental identification of gp36-dependent phiEco32 promoters. (a) The results of *in vitro* transcription by the *E. coli* RNAP core enzyme in the presence of σ^{70} or gp36 from DNA fragments corresponding to intergenic regions from phiEco32 genes *26* and *68*. (b) Total RNA isolated from cells collected at various time points postinfection and subjected to primer extension analysis with primers annealing within genes *26* and *68*. The reaction products were separated in a 6% denaturing polyacrylamide gel and visualized using PhosphorImager. (c) Comparison of DNA sequences upstream of experimentally determined transcription start points for P_{26} and P_{68}. The transcription start point is shown in red; annotated translation start codons are highlighted in yellow. Identical bases in compared sequences are shown against a black background. A common motif upstream of transcription start points used for bioinformatic searches is shown below the alignment.

Analysis of predicted σ^{70}-dependent promoters

Bioinformatically predicted σ^{70}-dependent promoters of phiEco32 were validated by primer extension analysis, with total RNA samples used for macroarray experiments and with primers annealing downstream of predicted σ^{70}-dependent promoters. All four initial predictions turned out to be correct: primer extension products whose 5′ ends corresponded to transcription start points matching the ones expected from the prediction were observed. Figure 3c shows changes in the abundance of primer extension products from four phage σ^{70}-dependent promoters and for two host genes throughout the infection. As can be seen, primer extension products corresponding to viral σ^{70}-dependent promoters appear 5 min postinfection, and their abundance decreases by 25 min postinfection. This dynamics is similar to that of early phage gene transcripts revealed by macroarray, indicating that σ^{70}-dependent promoters are indeed early phiEco32 promoters. The abundance of two host transcripts (*ompX* and *rpsU*) strongly decreased ~10 min postinfection, appearing to coincide with the decrease in early phage gene transcript abundance. Therefore, the same mechanism may be responsible for the shutoff of both early viral and host genes.

We were unable to detect any primer extension products corresponding to transcription initiation events from the predicted intragenic σ^{70} promoters mentioned above, leading us to conclude that these predicted promoters are not functional. To experimentally detect additional early promoters, we screened phiEco32 intergenic regions for promoter activity *in vitro* using a reconstituted σ^{70} holoenzyme. We selected 12 phiEco32 intergenic regions that (i) were more than 50 bp in length and (ii) did not separate convergently transcribed genes. The long noncoding region was also included in the analysis as positive control. Since this region is quite

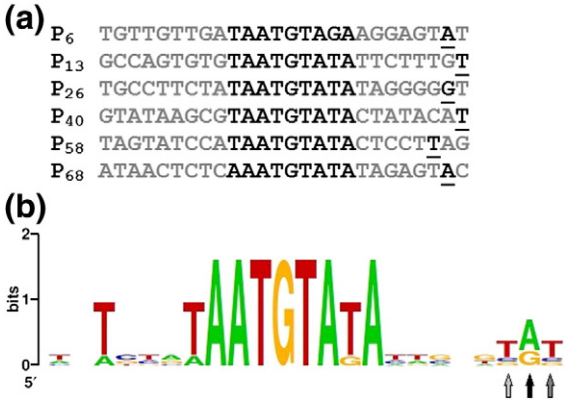


Fig. 5. Bioinformatic search for additional gp36-dependent promoters. Sequences upstream of phiEco32 genes *6*, *13*, *26*, *40*, *58*, and *68* are presented with a shared motif indicated in black. Experimentally determined *in vivo* transcription initiation start points (as judged by primer extension) are underlined. Below: A logo representation of the gp36 promoters of the phage is shown. The height of a letter indicates the degree of conservation. Upward vertical arrows indicate the positions of transcription start points; the color of the arrow indicates transcription initiation frequency, with black being the highest, white - the lowest.

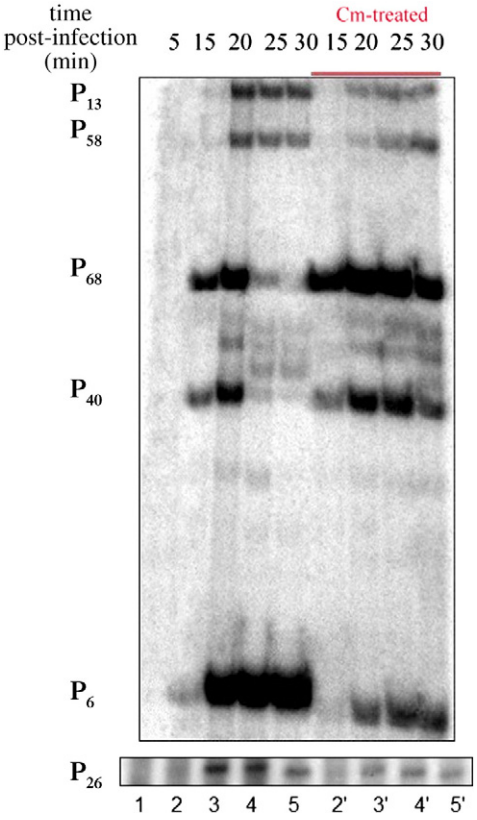


Fig. 6. Kinetics of accumulation of gp36-dependent transcripts during phiEco32 infection. Results of the multiplex primer extension analysis of total RNA extracted from phage-infected cells collected at various time points postinfection are shown. Primer extension reactions on the right-hand side of the figure (lanes 2′–5′) were conducted with RNA prepared from infected cells that were subjected to chloramphenicol treatment 12 min postinfection. The reaction products were separated in a 6% denaturing polyacrylamide gel and visualized using PhosphorImager.

long, we tested four partially overlapping PCR fragments that together covered the entire noncoding region. As expected, a fragment corresponding to a region upstream of gene *128* and fragments of a noncoding region containing early promoters identified *in vivo* supported robust transcription by the σ^{70} holoenzyme (data not shown). However, no σ^{70} holoenzyme transcription products were detected in reactions containing other intergenic regions (data not shown). We conclude that the four predicted and validated promoters are very likely the only functional phiEco32 σ^{70}-dependent promoters. These promoters belong to the early class.

Identification of phiEco32 promoters recognized by the gp36 RNAP holoenzyme

To identify gp36-dependent phiEco32 promoters and to prove that gp36 is a *bona fide* σ factor, we screened all phiEco32 intergenic regions that were no less than 50 bp and located upstream of the middle and late phiEco32 genes for promoter activity *in vitro* in the presence of *E. coli* RNAP core enzyme and recombinant gp36. Robust gp36-dependent transcription was observed, with two fragments corresponding to intergenic regions upstream of genes *26* and *68* (Fig. 4a). Thus, gp36 alone is able to direct specific transcription by the RNAP core and is therefore indeed a σ factor.

Analysis of RNA prepared from phiEco32-infected cells revealed primer extension products matching those identified for both gp36-dependent promoters *in vitro*, confirming that gp36-dependent promoters are active *in vivo*. The kinetics of accumulation of primer extension products from gp36-dependent promoters differed from that observed for σ^{70}-dependent early phage promoters. A primer extension product for P_{68} became visible 15 min postinfection and reached a maximum 20 min postinfection, and then its abundance decreased (Fig. 4b). A primer extension product for P_{26} became visible 20 min postinfection and remained at a steady level thereafter (Fig. 4b). We classify P_{68} as a middle phage promoter and P_{26} as a late promoter.

Identification of additional gp36-dependent promoters

To identify additional gp36-dependent phage promoters, we used the SignalX program[7] to create a pattern describing the gp36-dependent promoter consensus element based on two identified promoter sequences and searched the phiEco32 intergenic regions with this pattern. The search retrieved four additional matching sequences upstream of genes *6*, *13*, *40*, and *58*. The sequences of all putative gp36 promoters are shown in Fig. 5a.

Total RNA from cells infected with phiEco32 and collected at different time points postinfection was analyzed by primer extension, with primers annealing downstream of predicted gp36-dependent promoters. In each case, a primer extension product with a 5′ end located 5–7 bp downstream of a

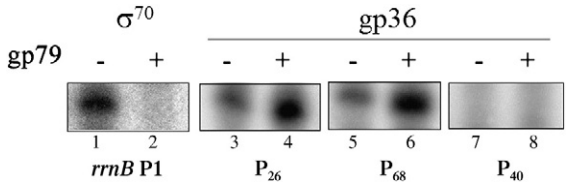


Fig. 7. PhiEco32 gp79 activates gp36-dependent transcription *in vitro*. The results of *in vitro* transcription by indicated *E. coli* RNAP holoenzymes in the presence or in the absence of gp79 are shown. In lanes 1 and 2, a DNA fragment containing a strong σ^{70}-dependent *rrnB* P1 promoter was used as template.

predicted gp36 promoter element was observed (data not shown; experimentally determined primer extension end points are underlined and highlighted in boldface in the alignment shown in Fig. 5a). We therefore conclude that all predicted gp36-dependent promoters are indeed functional *in vivo*. Analysis of the kinetics of transcript accumulation revealed that P_6, P_{13}, and P_{58} belong to the late phage promoter class; P_{40} is a middle phage promoter (Fig. 5b, lanes 1–5). The temporal classification of gp36 promoters generally matches the behavior of downstream phage genes revealed by the macroarray experiment.

Sequence analysis fails to pinpoint the reasons responsible for the distinct temporal activity patterns of middle and late phiEco32 promoters. To determine if an additional transcription factor may be responsible, we treated half of the cell culture used to monitor the temporal pattern of expression of gp36-dependent promoters with chloramphenicol, an inhibitor of translation, 12 min postinfection, just before the onset of middle gene transcription. Comparison of the abundances of gp36-dependent promoter products in treated and untreated cells revealed a different response of the middle and late promoters to chloramphenicol (Fig. 6). Primer extension products corresponding to middle promoters P_{40} and P_{68} continuously accumulated in cells treated with chloramphenicol (lanes 2′–5′). We therefore infer that phiEco32 encodes an inhibitor of middle transcription. This inhibitor must be synthesized after the first 12 min of infection, so it is itself a product of a middle phage gene. In contrast, the abundance of primer extension products corresponding to late promoters P_6, P_{13}, P_{26}, and P_{58} decreased after chloramphenicol treatment (the effect was particularly strong for P_6). The behavior of late transcripts in chloramphenicol-treated cells suggests that phiEco32 may also encode an activator of late transcription (however, we cannot exclude the indirect effects of phage genome replication).

Dual regulatory role of gp79

It is curious that our *in vitro* analysis revealed *in vitro* transcription by the gp36 RNAP holoenzyme from only two promoters, P_{26} and P_{68}, belonging to the late and middle temporal classes, respectively. Since our *in vivo* analysis suggested that phiEco32 may encode additional transcription factors, we tested the effect of gp79 (a small phiEco32 protein that binds host RNAP and inhibits σ^{70} transcription)[2] on gp36-dependent transcription *in vitro*. As can be seen from Fig. 7, gp79 inhibited σ^{70} transcription (compare lanes 1 and 2) but activated transcription from both P_{26} and P_{68} (compare lanes 3 and 4, and lanes 5 and 6, respectively). However, gp79 was unable to activate other middle and late promoters of the phage that were not functional *in vitro* (a representative result for P_{40} is shown in lanes 7 and 8). We conclude that gp79 is a specific activator of at least some middle and late phiEco32 promoters. The molecular mechanism of the dual regulatory function of gp79 (inhibition of σ^{70}-dependent transcription and activation of gp36-dependent transcription) is currently under investigation in our laboratory.

Discussion

In this work, we report an analysis of the temporal gene expression regulation of bacteriophage phiEco32 that infects *E. coli*. PhiEco32 appeared to be a truly novel *E. coli* phage [i.e., no more than 40% of the predicted open reading frame (ORF) had sequence homology in databases; the best matches of 54 conserved ORFs belonged to diverse phage and bacteria]. Because the phage encodes its own σ factor and possibly exploits some new strategies of gene expression control, we were interested in investigating the pattern of phiEco32 gene temporal expression.

Macroarray analysis and primer extension analysis proved that phiEco32 executes host transcription shutoff and identified three distinct temporal groups of phage genes (early, middle, and late genes), as is common for other phages. Overall, temporal clustering correlates well with annotated gene product functions,[2] and genes belonging to the same temporal class cluster in the phage genome. Late genes are transcribed in one direction; all annotated genes belonging to this class code for structural proteins and DNA packaging proteins. Late genes are located in the first third of the phiEco32 genome. Middle genes are transcribed in the opposite direction; they are also located together on the genome map. Genes whose products are involved in phage DNA replication, as well as transcription factors gp36 and gp79, belong to this class. Early genes are all located at the right end of the genome and are transcribed in the same direction as middle genes. Most of the early genes products do not have functional annotations. Their products may be involved in host acquisition by the phage and are therefore interesting subjects of further studies.

While there was a very good correlation between macroarray and primer extension results, the data for one gene, *58*, did not match. This gene behaved as a middle gene in the macroarray experiment and is in fact located in the middle gene cluster, away from late phage genes. However, it is preceded by a late promoter P_{58}. The reasons for such unusual behavior require further investigation.

Early phage genes are transcribed from four promoters located at the beginning of the early gene cluster. These promoters are recognized by the σ^{70} holoenzyme. All four early promoters were

predicted *in silico* and validated *in vivo*. We used *in vitro* transcription assays to screen intergenic regions for additional promoters recognized by the σ^{70} holoenzyme but did not find additional functional promoters. It is therefore likely that the four early promoters identified in our work are the only functional σ^{70} promoters in the genome. Therefore, the entire cluster of early genes, which constitutes ~30% or 20,951 bp of the phiEco32 genome, is likely transcribed from these promoters. It appears not to be unusual for bacteriophages to have long stretches of genes transcribed from very distant promoters,[9,10] and this feature may be used for temporal regulation of gene expression (see below).

All late genes and at least some middle genes of phiEco32 are transcribed from promoters recognized by the gp36 RNAP holoenzyme. Six gp36-dependent promoters belonging to both middle and late temporal classes were identified. These promoters are characterized by a highly conserved consensus sequence tAATGTAtA, with positions indicated by upper-case letters being strictly conserved while lower-case "t's" allow variations. We did not find any additional conserved motifs upstream or downstream of this consensus. The absence of upstream consensus (analogs of the −35 boxes in σ^{70} promoters) corresponds to the fact that gp36 has no homology with region 4 of other σ factors (region 4 is responsible for the −35 promoter element recognition). In contrast, gp36's similarity to σ^{70} region 2 (responsible for binding to the RNAP core and recognition of the −10 promoter element) is relatively strong.

The gp36 promoters are not recognized by the host σ^{70} holoenzyme *in vitro*, and the consensus element of these promoters is unrelated to the consensus elements of other six σ factors encoded by the *E. coli* genome. PhiEco32 gene *36* is a middle gene, and there are no recognizable gp36 promoters immediately upstream of it. From general considerations, it follows that gene *36* should be transcribed, at least initially, from a σ^{70} promoter. Such a promoter may be a weak one, thus explaining our inability to identify it either bioinformatically or experimentally. Alternatively, read-through transcription from strong early phage promoters may be responsible for the appearance of the first gene *36* transcripts. If so, the delayed appearance of gene *36* transcripts may be due to the time it takes for transcribing RNAP to reach gene *36*, which is over 35 kbp away from the closest validated σ^{70} promoter of the phage. Note that transcribing RNAP needs ~10 min to cover 35 kbp, given its average speed of ~50 bp/s; this time roughly matches the observed delay in the appearance of gp36 transcripts. Additional synthesis of gp36 can occur from RNA transcribed from P_{40} and P_{68} middle promoters, which are gp36-dependent. A positive feedback loop ensuring the synthesis of gene *36* transcripts by the gp36 holoenzyme is similar to the situation described for several extracytoplasmic function family σ factors, except that, in characterized cases, extracytoplasmic function σ holoenzyme promoters are located immediately upstream of the σ genes and/or cognate anti-σ genes located nearby.[11]

Gene *79* is also a middle gene whose product may be transcribed from upstream early promoters. Based on the action of gp79 *in vitro*, during infection, gp79 may be responsible for the shutoff of the σ^{70}-dependent host and early phage transcription. An unexpected aspect of gp79 involvement in phiEco32 gene expression regulation is its stimulation of gp36 transcription from at least some middle and late promoters of the phage. The mechanism of such unusual dual-action involvement of gp79 in transcription regulation is unknown. It is somewhat surprising that, among the six middle and late phiEco32 promoters identified, only two (the middle promoter P_{68} and the late promoter P_{26}) are efficiently recognized by the gp36 holoenzyme, alone or in the presence of gp79. Yet, the remaining four promoters are clearly highly active *in vivo*—in fact more active than P_{26} or P_{68}, as judged by the abundance of their transcripts. The reasons that make P_{68} and P_{26} stand out are not clear. It is possible that factors such as DNA sequences outside of the gp36 consensus elements, transcription factors other than gp79, DNA topology, or processes such as the replication of phage genome contribute to the activity of these promoters *in vivo*. We are currently investigating the molecular details of gp79 regulatory functions and its role in infection, as well as identifying additional transcription factors encoded by phiEco32.

Materials and Methods

Bacterial strains, phage and growth conditions, and RNA purification

Bacteriophage phiEco32 and its host *E. coli 55* were grown in standard LB media at 37 °C, with shaking. A single plaque and 200 μl of an overnight culture of *E. coli 55* were added to 20 ml of LB and incubated (with shaking) overnight at 37 °C to prepare phiEco32 lysates. Cells were disrupted by the addition of 1% chloroform, brief shaking, and centrifugation at 7500*g* at +4 °C for 30 min. The resulting lysate usually had a titer of 5×10^{11} to 9×10^{11} plaque-forming units/ml and was stored at +4 °C. Fifty milliliters of such lysate was used to purify total phage DNA. DNA was purified using the Qiagen Lambda Midi Kit according to the manufacturer's instructions. For RNA purification, *E. coli* culture was grown at 37 °C to midlog phase ($A600 = 0.4$) and infected with the lysate prepared as described above at an MOI of 10. At MOI = 10 used throughout the work, the efficiency of infection of the host was always greater than 90% (less

than 10% of the host "survivors" were detected). Infection was stopped at various times by rapid chilling and addition of rifampicin. The culture was centrifuged at 5000*g* for 5 min and used for total RNA purification with the Qiagen RNeasy Mini Kit according to the manufacturer's instructions.

Macroarray membrane preparation and hybridization

Membrane preparation and hybridization were carried out as described previously,[12] with minor modifications. PCR products, corresponding to each of the selected phiEco32 ORFs and to selected *E. coli* housekeeping genes, were synthesized from phiEco32 DNA and total *E. coli* DNA. Fragments were purified with the QIAquick Gel Extraction Kit (Qiagen). Concentrations of the fragments were determined by measuring absorbance at 260 nm and by agarose gel electrophoresis. Each membrane contained 100 ng of each PCR product, 10 ng of phiEco32 total DNA, 10 ng of *E. coli 55* total DNA, and 10 ng of PCR fragment corresponding to plant *PSII* gene as controls. PCR products and controls were denatured by alkali/heat treatment [0.4 M NaOH and 10 mM ethylenediaminetetraacetic acid, 10 min at 100 °C] and spotted onto a positively charged nylon membrane Immobilon-Ny$^+$ (Millipore) using S&S Filtration manifold for dot-blot assay. After applying the samples to the membrane, we fixed the DNA with UV cross-linking. Each membrane was used for hybridization once.

The cDNA probes for array hybridization were synthesized from total RNA purified from infected cells, as described above. Reactions were performed using the SuperScript III First-Strand synthesis system for RT-PCR (Invitrogen). Five micrograms of RNA and 50 ng of random hexamer primer were denatured at 65 °C for 5 min and chilled on ice, and the primers were annealed at 25 °C for 10 min. Synthesis was performed for 50 min at 50 °C with 200 U of SuperScript III; 0.5 mM dCTP, dGTP, and dTTP; 20 μCi of [α-^{32}P]ATP (4000 Ci/mmol); and 40 U of RNaseOUT. Reaction was terminated at 85 °C for 5 min. RNA was digested with RNaseH at 37 °C for 30 min. cDNA probes for control *PSII* gene were synthesized from the total *Arabidopsis thaliana* RNA purified with the RNeasy Plant Mini Kit (Qiagen). Reaction was performed the same way as described above, but with a gene-specific anti-sense primer. All labeled cDNA probes were purified with the QIAquick PCR Purification Kit (Qiagen).

Hybridization was performed in roller bottles in a hybridization oven. The membranes with spotted DNA samples were prehybridized for 2 h at 65 °C in 10 ml of a hybridization solution 5XSSC (5× saline–sodiumcitrate) is 0.75 M NaCl, 75 mM sodium citrate (pH 7.0), 0.1% wt/vol SDS, 5× Denhardt's solution, and 100 μg/ml sheared sonicated calf thymus DNA]. Labeled cDNA probes were denatured at 100 °C for 7 min, chilled on ice, and loaded into the rolling bottles containing prehybridized membranes and 2–3 ml of hybridization solution. After 12–18 h of hybridization at 65 °C, the membranes were washed twice with 20 ml of 2× SSC and 0.1% SDS for 5 min at room temperature and washed twice with 0.2× SSC and 0.1%SDS for 20 min at 65 °C. Membranes were air-dried and analyzed by PhosphorImager. The signal intensities for each spot were quantified using ImageQuant software (Molecular Dynamics). Each value was background-corrected (the background signal value was subtracted) and normalized relative to the *PSII* signal, which allowed a comparison of the signals from different membranes.

Clustering of phiEco32 temporal patterns

The transcripts were clustered according to the following procedure. The averages of the measured gene expression levels between 5 and 10 min, 10 and 25 min, and 25 and 35 min are calculated. It is then determined which of the three averages has the largest value. If the largest value corresponds to the average between 5 and 10 min, the gene is classified as early. If the largest value corresponds to the average between 10 and 25 min, the gene is classified as middle. If the largest value corresponds to the average calculated between 25 and 35 min, the gene is classified as late.

Prediction of promoters in the phiEco32 genome

The σ^{70} promoter recognition matrix was constructed by applying SignalX[7] to the DPInteract database collection of known *E. coli* σ^{70} promoters.[6] PhiEco32 late gene promoter recognition profiles were constructed using the SignalX program. Identification of candidate promoters in the phage genome was performed using the GenomeExplorer software package.[7]

Primer extension analysis

For *in vivo* primer extension reaction, 5 μg of total RNA purified from cells infected with phiEco32, as described above, was reverse-transcribed in the presence of 1 pmol of γ-^{32}P end-labeled primer with 100 U of SuperScript III enzyme from the First-Strand Synthesis Kit (Invitrogen) according to the manufacturer's protocol. The reactions were treated with RNaseH, ethanol-precipitated, and dissolved in formamide loading buffer. A 0-min time point stands for total RNA from *E. coli* cells that were not infected with phiEco32. A PCR fragment corresponding to the predicted transcription start area was synthesized from phiEco32 DNA. A sequencing reaction (fmol DNA Cycle Sequencing Kit; Promega) of this PCR product, with the primer used for primer extension reaction, was used as marker.

Analysis of transcription by primer extension was performed as described previously.[13] The primer was labeled with [γ-^{32}P]ATP by phage T4 polynucleotide kinase (New England Biolaboratories), as recommended by the manufacturer. One picomole of a mixture of γ-^{32}P-labeled primers and 10 μg of RNA in 40 mM Pipes (pH 6.4), 400 mM NaCl, 1 mM ethylenediaminetetraacetic acid, and 80% formamide was heated at 85 °C for 10 min then incubated overnight at 0 °C. Then RNA annealed with the primer was precipitated with ethanol, washed, dried, and dissolved in water. RNA was reverse-transcribed using M-MuLV (SibEnzyme) according to the manufacturer's recommendations. The reaction products of sequencing and reverse transcription were separated by electrophoresis, followed by visualization using PhosphorImager.

Proteins

His-tagged *E. coli* RNAP core and σ^{70} subunit were prepared as described previously.[14] Phage phiEco32 proteins gp79 and gp36 were prepared as described previously.[2]

In vitro transcription

Transcription reactions were performed in 10 μl of transcription buffer (40 mM Tris–HCl, 40 mM KCl, 10 mM $MgCl_2$, 5 mM DTT, and 100 μg/ml bovine serum albumin) and contained 150 nM *E. coli* RNAP core enzyme, 450 nM recombinant σ^{70} or 450 nM of the recombinant phiEco32 protein gp36, and, if mentioned, 450 nM gp79. Reactions were incubated for 10 min at 37 °C, followed by the addition of 20 nM of a PCR fragment containing promoter *rrnB* P1; P_{26}, P_{68}, or P_{40}; 100 μM each of ATP, CTP, and GTP; 10 μM UTP; and 0.4 μCi of [α-^{32}P]UTP. Reactions proceeded for 10 min at 37 °C and were terminated by the addition of an equal volume of denaturing loading buffer. The reaction products were resolved upon denaturing 6 M urea and 20% (wt/vol) polyacrylamide gels and visualized using PhosphorImager.

PCR fragments containing phiEco32 promoters were amplified from phiEco32 DNA and purified with QIAquick Kit (Qiagen).

Acknowledgements

Work in the K.S. laboratories was supported by National Institutes of Health grant GM59295, by a grant from the "Molecular and Cellular Biology" Program of the Russian Academy of Sciences Presidium, and Federal Program "Scientific and scientific -pedagogical personnel of innovative Russia 2009-2012", state contract 02.740.11.0771. N.A. was supported by a grant from the Russian Foundation for Basic Research 08-04-00968-a, and State Contract P1166. M.S.G. and D.A.R. were partially supported by grants from the Russian Academy of Sciences (via the "Molecular and Cellular Biology" and "Biodiversity" programs), the Russian Foundation of Basic Research (09-04-92745 and 10-04-00431), and state contract 07.514.11.4007. O.P. was partially supported by a Charles and Johanna Busch Biomedical Fund postdoctoral fellowship and Burroughs Wellcome fund. M.D. was partially supported by a Marie Curie International Reintegration Grant within the 7th European Communion Framework Program (PIRG08-GA-2010-276996) and by the Ministry of Education and Science, Republic of Serbia, under project number ON173052.

References

1. Hendrix, R. W., Smith, M. C., Burns, R. N., Ford, M. E. & Hatfull, G. F. (1999). Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. *Proc. Natl Acad. Sci. USA*, **96**, 2192–2197.
2. Savalia, D., Westblade, L. F., Goel, M., Florens, L., Kemp, P., Akulenko, N. *et al.* (2008). Genomic and proteomic analysis of phiEco32, a novel *Escherichia coli* bacteriophage. *J. Mol. Biol.* **377**, 774–789.
3. Matsushita, K., Uchiyama, J., Kato, S., Ujihara, T., Hoshiba, H., Sugihara, S. *et al.* (2009). Morphological and genetic analysis of three bacteriophages of *Serratia marcescens* isolated from environmental water. *FEMS Microbiol. Lett.* **291**, 201–208.
4. Kropinski, A., Lingohr, E. & Ackermann, H. (2010). The genome sequence of enterobacterial phage 7–11, which possesses an unusually elongated head. *Arch. Virol.* **156**, 149–151.
5. Tan, Y., Zhang, K., Rao, X., Jin, X., Huang, J., Zhu, J. *et al.* (2007). Whole genome sequencing of a novel temperate bacteriophage of *P. aeruginosa*: evidence of tRNA gene mediating integration of the phage genome into the host bacterial chromosome. *Cell. Microbiol.* **9**, 479–491.
6. Robison, K., McGuire, A. M. & Church, G. M. (1998). A comprehensive library of DNA-binding site matrices for 55 proteins applied to the complete *Escherichia coli* K-12 genome. *J. Mol. Biol.* **284**, 241–254.
7. Mironov, A. A., Vinokurova, N. P. & Gel'fand, M. S. (2000). Software for analyzing bacterial genomes. *Mol. Biol. (Moscow)*, **34**, 253–262.
8. Huerta, A. M. & Collado-Vides, J. (2003). Sigma 70 promoters in *Escherichia coli*: specific transcription in dense regions of overlapping promoter-like signals. *J. Mol. Biol.* **333**, 261–278.
9. Djordjevic, M., Semenova, E., Shraiman, B. & Severinov, K. (2006). Quantitative analysis of transcription strategy by a virulent bacteriophage. *Virology*, **354**, 240.
10. 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.
11. Raivio, T. & Silhavy, T. (2001). Periplasmic stress and ECF sigma factors. *Annu. Rev. Microbiol.* **55**, 591–624.
12. Minakhin, L., Semenova, E., Liu, J., Vasilov, A., Severinova, E., Gabisonia, T. *et al.* (2005). Genome sequence and gene expression of *Bacillus anthracis* bacteriophage Fah. *J. Mol. Biol.* **354**, 1–15.
13. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*, 2nd edit. Cold Spring Harbor Laboratory, New York, NY.
14. Kashlev, M., Nudler, E., Severinov, K., Borukhov, S., Komissarova, N. & Goldfarb, A. (1996). Histidine-tagged RNA polymerase of *Escherichia coli* and transcription in solid phase. *Methods Enzymol.* **274**, 326–334.