

Article

Regulation of Gene Expression of phiEco32-like Bacteriophage 7-11

Daria Lavysh^{1*}, Vladimir Mekler², Evgeny Klimuk¹, and Konstantin Severinov^{1,2,3*}

¹ Institute of Molecular Genetics of National Research Centre «Kurchatov Institute», Moscow, Russia

² Waksman Institute for Microbiology and Department of Molecular Biology and Biochemistry, Rutgers, State University of New Jersey, Piscataway, USA

³ Skolkovo Institute of Science and Technology, Center for Life Sciences, Skolkovo, Russia

* Correspondence: Daria Lavysh (lavyshD@gmail.com) or Konstantin Severinov (severik@waksman.rutgers.edu), Waksman Institute, 190 Frelinghuysen Road, Piscataway, NJ 08854, Phone: (732) 445 6095, FAX: (732) 445 5735

Abstract: *Salmonella enterica* serovar Newport bacteriophage 7-11 shares 41 homologous ORFs with *Escherichia coli* phage phiEco32 and both phages encode a protein similar to bacterial RNA polymerase promoter specificity σ subunit. Here, we investigated the temporal pattern of 7-11 gene expression during the infection and compared it to the previously determined transcription strategy of phiEco32. Using primer extension and *in vitro* transcription assays we identified eight promoters recognized by host RNA polymerase holoenzyme containing 7-11 σ subunit SaPh711_gp47. These promoters are characterized by a bipartite consensus GTAA τ -(16)-aCTA and are located upstream of late phage genes. While dissimilar from single-element middle and late promoters of phiEco32 recognized by holoenzyme formed by the phi32_gp36 σ factor, the 7-11 late promoters are located at genome positions similar to those of phiEco32 middle and late promoters. Two early 7-11 promoters are recognized by RNA polymerase holoenzyme containing host primary σ^{70} factor. Unlike the case of phiEco32, no shut off of σ^{70} -dependent transcription is observed during 7-11 infection and there are no middle promoters. These differences can be explained by the fact that phage 7-11 does not encode a homologue of phi32_gp79, an inhibitor of host and early phage transcription and an activator of transcription by the phi32_gp36-holoenzyme.

Keywords: bacteriophage, alternative sigma factor, transcription regulation

1. Introduction

Salmonella enterica serovar Newport phage 7-11 is a podovirus with a distinctive strongly elongated virion head [1]. Phylogenetic analysis reveals that it is related to coliphage phiEco32, however it belongs to unclassified *Podoviridae*. The *E. coli* phage phiEco32 was considered to be unique when its genomic sequence was first determined [2]. However, subsequently, multiple related phages were isolated in Canada, China, South Korea and elsewhere [3–5]. PhiEco32 remains the best-studied member of *Kuravirus* genus to date. As most bacteriophages, it relies on host RNA polymerase (RNAP) for its development within the host cell [6–8]. Early phage genes are transcribed from four promoters located at the beginning of the early gene cluster. These promoters are recognized by the housekeeping σ^{70} RNAP holoenzyme of the host. All late genes and at least some of the middle genes of phiEco32 are transcribed from promoters recognized by an RNAP holoenzyme containing phage σ factor phi32_gp36. Six phi32_gp36-dependent middle and late promoters were identified experimentally. A small phage polypeptide, phi32_gp79, a product of a middle gene, inhibits transcription by the σ^{70} holoenzyme and stimulates phi32_gp36 dependent transcription *in vitro* [2,6]. Thus phi32_gp79 may be responsible for regulated expression of phage genes and host transcription shut-off.

In this report, we describe the use of bioinformatics and biochemical approaches to characterize the temporal patterns of phage and host genes expression during the

infection by phage 7-11, a distant relative of phiEco32, and compare and contrast its transcription strategy to phiEco32 and related phages.

2. Materials and Methods

Comparison of phiEco32-like genomes and prediction of their promoters

The comparison of phiEco32-like bacteriophages ORFs was done using BLASTP. The σ^{70} -promoters in phage 7-11 genome were searched for as described previously [6]. PhiEco32_gp36 promoter and SaPh711_gp47 promoter recognition profiles were constructed using SignalX [9]. Search of candidate promoters in the phage genome was done using the GenomeExplorer software package [9].

Bacterial strains, phage, and growth conditions

E. coli XL10-Gold (Δ (mcrA)183 Δ (mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac Hte [F' proAB lacIq Z Δ M15 Tn10 (Tetr) Amy Camr]) ultracompetent cells (Stratagene) were used for molecular cloning [10].

E. coli B BL21(DE3) (F - *dcm* *ompT* *hsdS(rB - mB -)* *gal* λ (DE3)) (Stratagene) was used for recombinant proteins overproduction.

All bacterial strains were grown in LB media (1% Bactotryptone, 1% NaCl, 0.5% yeast extract, with or without 1.5% Bactoagar) at 37 °C with appropriate antibiotics.

Bacteriophage 7-11 and its host *S. enterica* serovar Newport were obtained from the Félix d'Hérelle Reference Center for Bacterial Viruses (Université Laval, Quebec, Canada). Bacterial cultures were grown and phage infection performed in liquid LB media at 37 °C with orbital shaking. To prepare 7-11 lysates, one plaque and 100 μ l of overnight culture of *S. Newport* were added to 100 ml LB media and incubated with shaking at 37 °C for 4 hours. The culture was treated with 1% chloroform and centrifuged for 30 minutes at 5000 g. The titer of resulting phage lysate usually was \sim 4x10¹⁰ p.f.u./ml. Phage DNA was purified from 5 ml of lysate using Thermo Scientific column DNA extraction kit. For RNA purification, cells were grown to an OD₅₉₅=0.4 and infected with 7-11 phage at MOI (multiplicity of infection) of 10. At this multiplicity, the efficiency of infection of the host was above 80%. Infection was stopped at various time points (including a zero-point, which corresponds to uninfected cells) by rapid chilling and by the addition of rifampicin to a final concentration of 30 μ g/ml. 15 ml aliquots of infected cultures were collected for each time point, cells were collected by centrifugation at 5000 g for 15 min and used for total RNA purification.

RNA purification

Total RNA was isolated by two extractions with hot phenol (pH 4.5) followed by ethanol precipitation as previously described [8,11]. After the extraction, RNA was treated with 15 u of RNase-free DNase I (ThermoFisher) at 37 °C for 30 min. Reactions were terminated by the addition of EDTA to 5 mM and by heating at 65 °C for 10 min. Samples were phenol extracted and ethanol precipitated. RNA was dissolved in DEPC-treated water and stored at -70 °C.

Primer extension analysis

Primers were 5'-end labeled with (γ -³²P)-ATP using phage T4 PNK (New England Biolabs) as recommended by the manufacturer. Annealing of 1 pmol of a mixture of (γ -³²P)-labeled primers with 10 μ g of total RNA was performed in 40 mM PIPES (pH 6.4), 400 mM NaCl, 1 mM EDTA and 80% formamide. Samples were denatured at 85 °C for 10

minutes and then cooled on ice. The annealing continued overnight at 0 °C. Next, samples were precipitated with ethanol, washed with 70% and 96% ethanol, dried and dissolved in DEPC-treated water. Primers were extended by Maxima reverse transcriptase (Fermentas) according to manufacturer's recommendations. Equal volume of formamide loading buffer (80% formamide, 10 mM EDTA, 0,05% bromphenol blue, 0,05% xylene cyanol) was added to the samples, after which they were heated for 5 min at 85 °C and cooled on ice. For primer extension reactions using *in vitro* synthesized RNA as templates, RNA was produced by *in vitro* transcription from appropriate PCR fragments by the *E. coli* core RNAP supplemented with recombinant SaPh711_gp47 in the presence of 100 µM of each NTPs at conditions described in the *in vitro* transcription section. RNA was cleaned up with QIAGEN RNeasy Mini Kit and used for primer extension. Sequencing reactions were set with USB Thermo Sequenase Cycle Sequencing Kit on PCR products synthesized with the same primer as the one used for the primer extension reaction. The reaction products were separated on 6 or 8% denaturing polyacrylamide sequencing gels followed by autoradiography and visualization using PhosphorImager.

7-11. gene 47 cloning and SaPh711_gp47 purification

Phage 7-11 gene 47 was cloned in between NdeI or BamHI sites of the pET18b vector. His-tagged *E. coli* RNAP core and recombinant σ^{70} subunit were prepared as previously described [8,12–15]. Phage phiEco32 proteins phi32_gp79 and phi32_gp36 as well as 7-11 protein SaPh711_gp47 were prepared as described [2].

In vitro transcription

Transcription reactions were performed in transcription buffer (40 mM Tris-HCl pH 7.9, 40 mM KCl, 10 mM MgCl₂, 5 mM DTT, 100 µg/ml BSA) and contained 150 nM *E. coli* RNAP core enzyme, 450 nM recombinant σ^{70} or 450 nM of recombinant SaPh711_gp47 protein and, if mentioned, 450 nM of recombinant phi32_gp79 and 20 nM of DNA template. Reactions were incubated for 10 min at 37 °C, followed by the addition of 100 µM ATP, CTP, GTP; 10 µM UTP and 0.4 µCi of [α -³²P]UTP. Reactions proceeded for 10 more minutes at 37 °C and were terminated by the addition of an equal volume of formamide loading buffer. Primers used to generate templates for transcription are listed in Table S1.

KMnO₄ probing

Promoter complexes were formed as described in the *in vitro* transcription section with the following modifications: 100 nM of (γ -³²P)-5'-end-labeled DNA fragment was used and transcription buffer did not contain DTT. After preincubation for 15 min at 37 °C, the reactions were probed with 2 mM KMnO₄ for 20 s at 37 °C. Reactions were terminated by the addition of 0.5 volume of stop solution containing 600 mM β -mercaptoethanol, followed by ethanol precipitation, and 10% piperidine treatment at 95 °C for 15 min. Samples were chloroform extracted and ethanol precipitated. Pellets were dissolved in formamide loading buffer and reaction products were analyzed on sequencing gels.

Fluorometric measurements

Fluorescence measurements were carried out at 25 °C using a QuantaMaster QM4 spectrofluorometer (PTI) in transcription buffer containing 0.02% Tween 20 [16]. The σ^{70} protein labeled at position 517 with fluorescein (σ^{70*}) was obtained as previously described [17]. Final assay mixtures (800 µl) contained 1.5 nM σ^{70*} , 1 or 3 nM RNAP core complexed with Rifampicin (Rif) and 12 nM gp47 (if indicated). The RNAP core-Rif complex was prepared by incubation of RNAP core (1 or 3 nM) and 1 µM Rif for 10 min at 25 °C [18]. The fluorescein fluorescence intensities were recorded with an excitation wavelength of 498 nm and an emission wavelength of 520 nm. Time-dependent fluorescence changes were monitored using manual mixing; the mixing dead time was 15 s.

3. Results

3.1. The phiEco32-like phages and their genomes

The genome of *E. coli* phage phiEco32 was determined in 2008 [2]. At the time, phiEco32 appeared to be a novel phage, i.e., no more than 40% of the products of its ORFs had homologues in public databases [2]. Subsequently, over hundred bacteriophages with similarities in genomic sequence and gene organization were submitted to public databases, including, among others *E. coli* (APEC) phages NJ01 (NC_018835), ECB2 (NC_018859), SU10 (KM044272), and 172-1 (KP308307) [3–5]. BLASTP searches identified 69 genes common to phiEco32 and these *E. coli* phages (Supplementary Table 2). These genes thus define core genes [19] within the *Podoviridae* family *Kuravirus* group that we will refer as “the phiEco32-subgroup”.

Salmonella enterica serovar Newport phage 7-11 (NC_015938) and *Cronobacter sakazakii* phage vB_CsaP_GAP52 (GAP52) (NC_019402) are also related to phiEco32-like phages. Comparison of phiEco32 and 7-11 genomes is schematically presented in Fig. 1. 27.15% of 7-11 and 32.03% of phiEco32 ORFs are homologous to each other [1]. The genome of phage GAP52 [20] encodes 79 proteins homologous to 7-11 gene products, which corresponds to ~52% of 7-11 ORFs and ~69% of GAP52 ORFs (Supplementary Table 2). Despite the relatively high similarity of many proteins, the genomes of GAP52 and 7-11 share only 10% of similarity at the level of DNA. Among the proteins common to 7-11 and GAP52, only 33 share common descent with phiEco32 proteins. Surprisingly, only 23 of these correspond to phiEco32-like phages core genes products. Since the amount of homologues between phiEco32 and 7-11/GAP52 is less than that shared by other *Kuravirus* phages, we consider phages 7-11 and GAP52 as members of a distinct “7-11 subgroup” of phiEco32-like bacteriophages. Very close relatives of 7-11 – *Salmonella* phage SE131 and pSal-SNUABM-01 and for GAP52 – *Cronobacter* phage A24 were sequenced. Since their DNA sequence similarity to, respectively, 7-11 and GAP52 is higher than 88%, we will not consider them separately from the original isolates.

Transcription of phiEco32 genome was studied by macroarray and primer extension analysis and conventional early, middle, and late expressed classes of phage genes were revealed [6]. Since phiEco32-like bacteriophages share a common overall genome architecture one can assume that genes of other phiEco32-like phages can be similarly divided into three expression classes. In the schematic representation of phiEco32 and 7-11 genomes (Fig. 1) extended clusters of genes that encode structural virion proteins and DNA packaging machinery components can be identified. These genes (colored red in Fig. 1 and shown to belong to late expression class in phiEco32) are transcribed in the rightward direction and occupy about one third of each genome. The second cluster, (colored green in Fig. 1 and shown to belong to middle expression class in phiEco32), includes shared genes coding for phage DNA polymerase, nucleotide metabolism enzymes, and a host RNA polymerase σ subunit. Finally, the cluster of genes colored blue in Fig. 1 (shown to belong to early expression class in phiEco32), is most diverse between the two phages and contains mostly genes of unknown function that likely play a role in host takeover [21]. The middle and early genes are transcribed in the direction opposite to that of late gene cluster transcription.

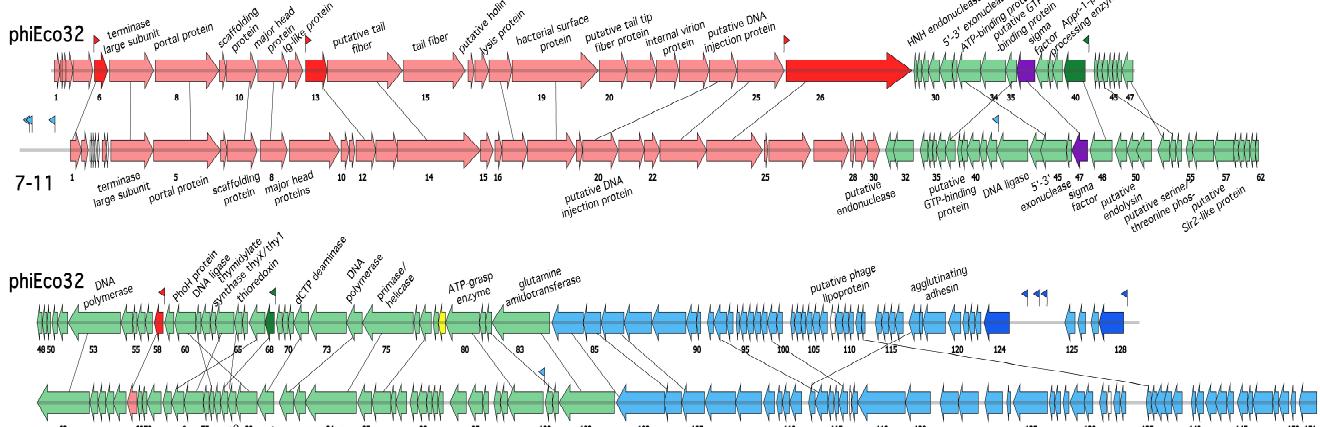


Figure 1. Comparison of phiEco32 and 7-11 genomes. The genomes of phages phiEco32 (NC_010324) and 7-11 (NC_015938) are schematically presented. Arrows represent annotated genes; the direction of an arrow indicates the direction of transcription. The ORFs are colored according to experimentally determined temporal classes of phiEco32: blue - early genes, green - middle genes, red - late genes. Dark shaded colors indicate phiEco32 genes in front of which promoters were located experimentally (shown as appropriately colored flags). ORFs encoding transcription factors are colored violet (sigma factors) and yellow (inhibitor of σ^{70} transcription phi32_gp79). Homologous genes in the two genomes are connected by lines. For 7-11, bioinformatically predicted σ^{70} promoters are indicated with colors matching expected gene expression class.

All phiEco32-like phages encode a σ factor (phi32_gp36 in phiEco32, a product of a middle gene). Previously, we identified six phi32_gp36 promoters located upstream of middle and late genes of the phage [6]. These promoters are characterized by a single highly conserved consensus sequence tAATGTAtA. Transcription start sites are located 6–8 bp downstream of phi32_gp36 holoenzyme promoter element. We searched for phi32_gp36-like promoters in every phiEco32-like phage. Sequences similar to phiEco32 middle/late promoter consensus elements were found in expected locations for every members of the phiEco32-subgroup phages (Supplementary Table 2). However, no matching sequences were found in the 7-11 and GAP52 genomes.

The putative σ factor of 7-11, SaPh711_gp47, is a distant homologue of phi32_gp36 and other phiEco32 subgroup σ factors, which likely explains our inability to find 7-11 promoters using the phiEco32 middle and late promoters consensus motif. We have searched all 7-11 intergenic regions (that were defined as DNA sequences between nucleotides positions “-100” and “+200” relative to predicted start codons) for the presence of overrepresented DNA motifs by MEME suite [22] and Align2N algorithm. The intergenic regions were also scanned for motifs similar to promoter consensus elements of known bacterial promoters using Prodigic algorithm [23]. No overrepresented motives or putative promoters were identified using this procedure.

Transcription factor phi32_gp79 inhibits σ^{70} -dependent host and early phage phiEco32 transcription but activates phi32_gp36-dependent transcription *in vitro* [6]. Phi32_gp79 homologues can be predicted in all phiEco32-like phages. No phi32_gp79 homologues are encoded by either the 7-11 or GAP52 genomes.

3.2. Identification of 7-11 late promoters

There shall be (at least) two types of promoters in the 7-11 genome: early promoters dependent on the host RNAP holoenzyme containing the primary σ^{70} factor and middle and/or late promoters dependent on the holoenzyme containing SaPh711_gp47. Since we could not predict SaPh711_gp47 promoters bioinformatically, we decided to proceed from the fact that out of six phiEco32 genes (6, 13, 26, 40, 58, 68) located downstream of phi32_gp36-dependent promoters (in Fig. 1, shown by dark shade or green (for middle)

and red (late) genes), five are present in the 7-11 genome. We reasoned that noncoding regions directly upstream of these genes or in front of the first genes of operons that contain these genes may house SaPh711_gp47-RNAP holoenzyme promoters. In order to confirm this hypothesis, total RNA was purified from 7-11 infected *S. enterica* serovar New-port cultures at different time points post-infection and analyzed by primer extension with primers annealing downstream of possible SaPh711_gp47-dependent promoters. Distinct primer extension products were detected in reactions with four primers annealing to intergenic regions of genes 1, 8, 16 and 48 in RNA samples prepared from cells collected 40 min post-infection (Fig. 2A). Since the infection cycle of 7-11 is ca. 60 minutes, late promoters were expected to be active at the chosen time-point.

The products of primer extension shown in Fig. 2A were loaded on sequencing gels alongside with marker sequencing reactions performed from same primers using PCR fragments carrying intergenic regions under study as templates. Alignment of sequences upstream of mapped primer extension products 5' ends is shown in Fig. 2B. As can be seen, a bipartite GTAAtg -(16)- aCTA consensus motif is present in all four sequences.

We next used the SignalX program [9] to create a pattern describing the consensus element based on four experimentally identified sequences and searched the 7-11 genome with this pattern. The search retrieved four additional matching sequences upstream of genes 12, 22, 28, and 69 (Fig. 2C). Primer extension analysis revealed the presence of expected products for primers annealing downstream of genes 22, 28, and 69 (Fig. 2D).

Time-resolved multiplex analysis showed that all primer extension products studied above became visible 20 min post-infection, reached a maximum at 50 min post-infection, and then their abundance decreased, likely due to cell lysis that became apparent 60 min post-infection (Fig. 2E). Thus, the kinetics of accumulation of primer extension products corresponded to the late expression class. We therefore conclude that the bipartite motif whose logo is presented in Fig. 2D defines the consensus of late promoter of phage 7-11.

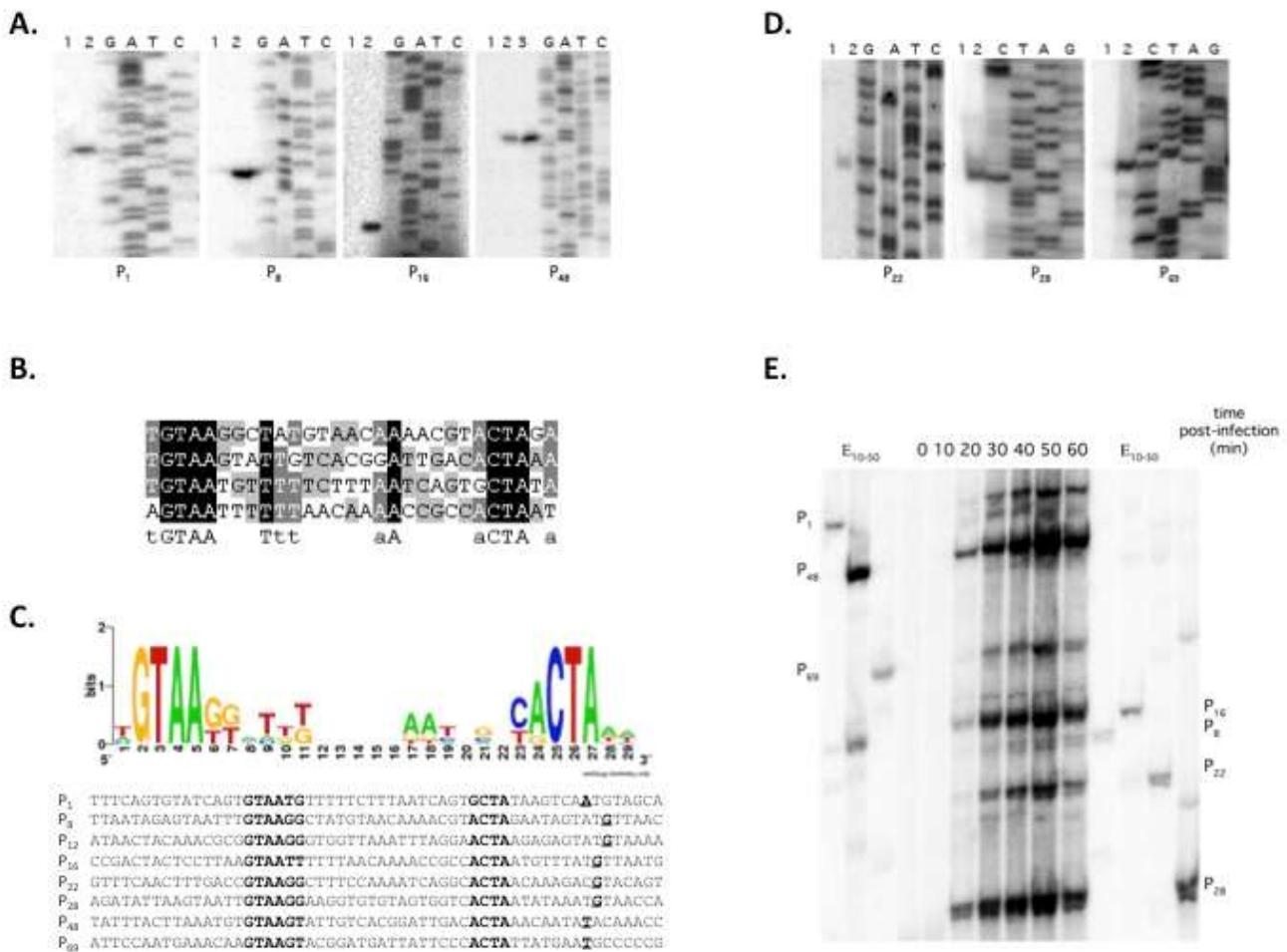


Figure 2. Identification of bacteriophage 7-11 late promoters. A. Results of primer extension analysis using RNA prepared from cells before (lanes 1) or 40 minutes after (lanes 2) 7-11 infection using primers annealing upstream of 7-11 genes 1, 8, 16, and 48 corresponding to phiEco32 genes proximal to its late promoters. Lane 3 for P_{48} panel shows the results of primer extension with RNA synthesized *in vitro* by the SaPh711_gp47-RNAP holoenzyme. The reaction products were separated in 6% denaturing polyacrylamide gels and visualized using PhosphorImager. B. The alignment of P_1 , P_8 , P_{16} , and P_{48} promoters validated in panel A. C. A Logo representation of phage 7-11 late promoters. The alignment below shows, in addition to P_1 , P_8 , P_{16} , and P_{48} promoters validated in panel A, promoters retrieved by an additional targeted search using the 7-11 late promoter consensus. Conserved elements are shown in bold typeface. Transcription start points are shown in bold and are underlined. D. Validation of P_{22} , P_{28} , and P_{69} promoters predicted bioinformatically in C. See panel A legend for details. E. Kinetics of accumulation of late transcripts during 7-11 infection. Results of multiplex primer extension analysis of total RNA extracted from phage-infected cells collected at various time points post-infection using primers designed to reveal transcription from indicated late phage promoters. On the left and on the right products of individual primer extension reaction on RNA pooled from samples from each time point are shown. The reaction products were separated in 6% denaturing polyacrylamide gel and visualized using PhosphorImager.

3.3. SaPh711_gp47 as a late sigma factor

To determine whether SaPh711_gp47 is responsible for late transcription of bacteriophage 7-11, the SaPh711_gp47 gene was cloned into an *E. coli* expression vector. Recombinant protein was purified and tested for its ability to direct transcription by the *E. coli* RNAP core (99% identical to *S. enterica* enzyme) from DNA fragments that contained identified 7-11 late promoter sequences. Efficient *in vitro* RNA synthesis from each late

promoter detected *in vivo* as well as for P12, whose activity was not observed *in vivo*, was revealed (Fig. 3A). The sizes of transcripts detected by primer extension with gene-specific primers using RNA purified from infected cells matched those detected with *in vitro* transcribed RNA (data shown for P₄₈ in Fig. 2A).

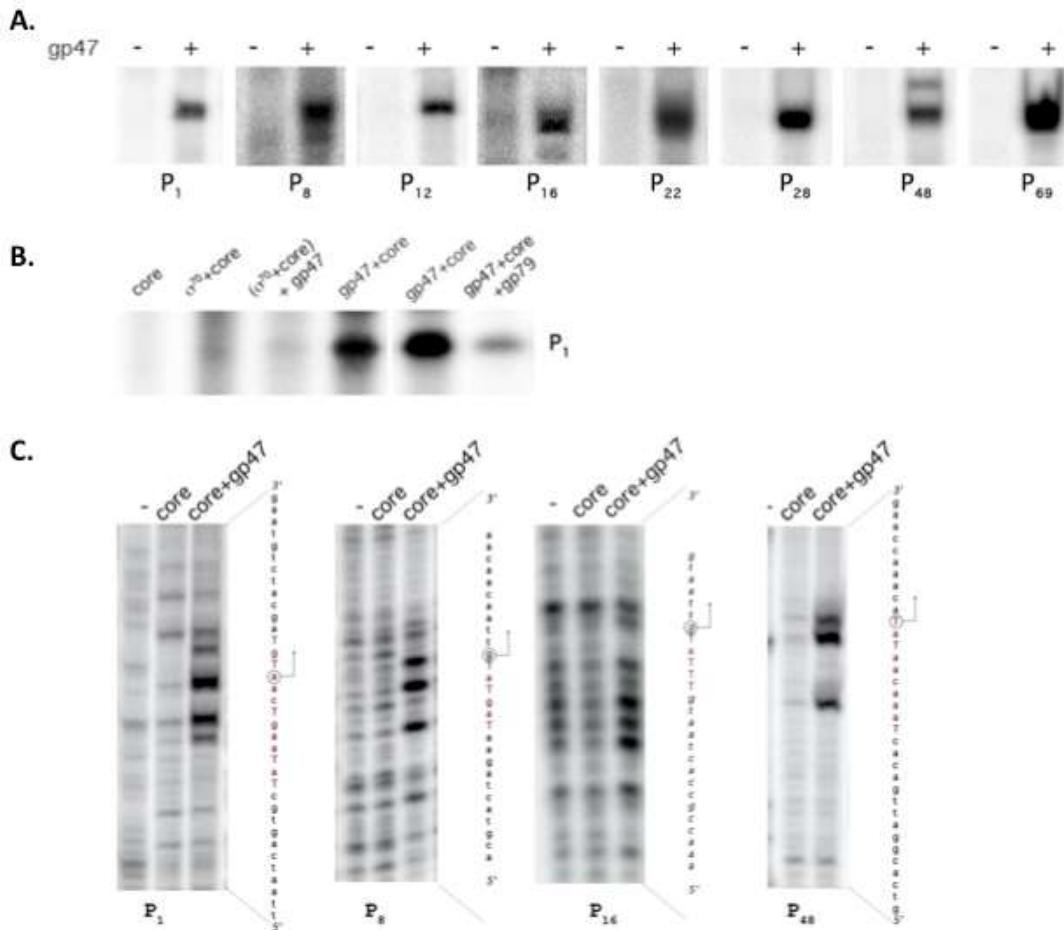


Figure 3. SaPh711_gp47 is a sigma factor required for late promoter transcription *in vitro*. A. The products of *in vitro* transcription reaction by *E. coli* RNAP core enzyme in the presence or in the absence of 3-fold molar excess of recombinant SaPh711_gp47 from DNA fragments corresponding to intergenic regions in front of 7-11 genes 1, 8, 12, 16, 22, 28, 48, and 69 are shown. The mapping of the product of P₄₈ transcription by primer extension is shown in Fig. 2A. B. Analysis of *in vitro* transcription products by *E. coli* RNAP core in the presence of σ^{70} , SaPh711_gp47, or both and with the addition of phiEco32 inhibitor of σ^{70} -transcription gp79 from a DNA fragment corresponding to intergenic regions in front of 7-11 gene 1. The reaction products were separated in a 10% denaturing polyacrylamide gel and visualized using PhosphorImager. C. DNA fragments harboring indicated late 7-11 promoters terminally labeled at the template strand were probed with KMnO₄ in the absence or in the presence of core RNAP with or without SaPh711_gp47. The reaction products were separated by 6% denaturing PAGE and revealed using PhosphorImager. Mapping of permanganate sensitive bands to DNA sequence is shown at the right of each panel.

As expected, RNAP holoenzyme containing the σ^{70} subunit did not produce transcripts from DNA fragments containing late phage promoters *in vitro* (Fig. 3B, data shown for P₁). Conversely, the SaPh711_gp47-holoenzyme did not transcribe from a DNA template containing a strong σ^{70} -promoter T7 A1 [24]. The addition of SaPh711_gp47 to reactions containing σ^{70} RNAP holoenzyme did not lead to the appearance of bands corresponding to late phage promoter transcripts, indicating that the two σ factors compete for the same RNAP core binding site. Surprisingly, dual transcription regulator of phi32_gp79

that inhibits σ^{70} -transcription *in vitro* and activates phi32_gp36-dependent transcription [6] inhibited SaPh711_gp47 transcription *in vitro* (Fig. 3B).

Promoter melting at several late 7-11 promoters was analyzed by KMnO₄ probing. As can be seen, the region of localized DNA melting included the TSS and the -10 promoter element (Fig. 3C) [25]. Thus, SaPh711_gp47 is a σ factor that directs host RNAP core to transcribe late 7-11 promoters located in front of genes 1, 8, 12, 16, 22, 28, 48, and 69.

3.4. *SaPh711_gp47* sigma factor forms a stable complex with host core RNAP

The affinity of *SaPh711_gp47* to core-RNAP was analyzed in competition experiments with σ^{70} subunit fluorescently labeled at amino acid 517 (σ^{70*}) (Fig. 4).

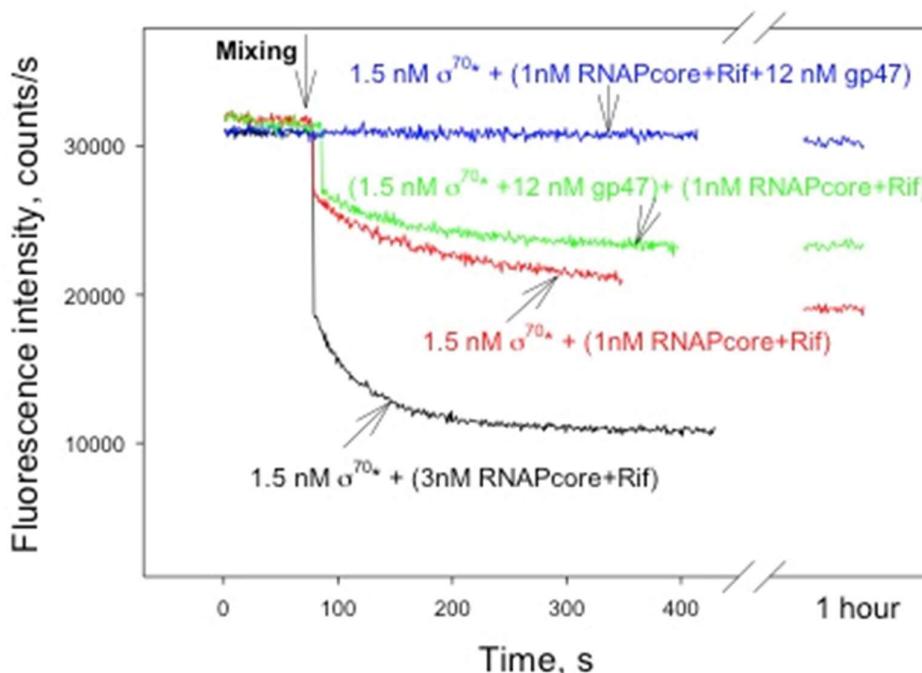


Figure 4. Binding of *SaPh711_gp47* to *E. coli* RNA core enzyme. The affinity of gp47 for *E. coli* RNAP core enzyme was evaluated by a real time fluorometric competition binding assay using a derivative of the *E. coli* RNAP σ^{70} subunit labeled with fluorescein (σ^{70*}) as reference competitor. Shown are time dependences of the decrease in fluorescence upon the RNAP core-Rif complex upon the addition to σ^{70*} in the absence (black and red curves) and presence of gp47 (blue and green curves). Black curve shows decrease of σ^{70*} fluorescence upon the formation of the holoenzyme, since rifampicin is quenching the fluorophore. Red curve shows that σ^{70*} fluorescence quenching depends on the ratio of core and σ^{70*} . The blue curve shows the lack of decrease in σ^{70*} fluorescence upon the addition of RNAP core preincubated with gp47 to σ^{70*} . The green curve shows time dependence of the decrease in fluorescence upon the addition of RNAP core to premixed σ^{70*} and gp47. The concentrations of σ^{70*} , gp47 and RNAP core were 1.5 nM, 12 nM and 1 or 3 nM, respectively, unless otherwise indicated.

In the context of the holoenzyme, the labeled amino acid of σ^{70*} is adjacent to the binding site of rifampicin (Rif), an inhibitor of transcription elongation that interacts with RNAP core [26]. Upon formation of the holoenzyme in the presence of Rif, fluorescence of σ^{70*} decreases, due to quenching of the σ^{70*} fluorophore by Rif via FRET mechanism (black curve) [17]. The degree of fluorescence quenching depends on the ratio of core and σ^{70} (red curve). When RNAP core bound to rifampicin was pre-incubated with *SaPh711_gp47* followed by the addition of fluorescently labeled σ^{70*} , no quenching was observed even after a 1-hour incubation with σ^{70*} (blue curve), presumably because the σ^{70} binding site is occupied by *SaPh711_gp47*. When σ^{70} and *SaPh711_gp47* were added to

Rif-bound RNAP core simultaneously, fluorescence quenching (green curve) was only moderately less than that observed in the same ratio of RNAP core/ σ^{70} in the absence of SaPh711_gp47 (red curve). Since this experiment was performed under conditions when the concentration of SaPh711_gp47 was 8 times greater than that of σ^{70} , it follows that the formation of holoenzyme with σ^{70} is faster than the formation of the SaPh711_gp47 holoenzyme. However, if the complex with SaPh711_gp47 is already formed, σ^{70} is unable to displace bound SaPh711_gp47.

3.5. Identification of 7-11 σ^{70} -dependent promoters

Five putative σ^{70} -dependent promoters were predicted in phage 7-11 genome (Fig. 5A) [1]. In order to confirm this prediction and monitor the behavior of early transcripts during the infection, primer extension reactions with primers annealing downstream of predicted σ^{70} -dependent promoters were performed on total RNA purified from 7-11 infected cultures collected at different time points post-infection. Out of five predicted early promoters, only two were validated *in vivo* by primer extension (Fig. 5B). These promoters are located upstream of gene 151, at the very right end of the genome. Fig. 3B also shows changes in abundance of primer extension products for viral RNA transcribed from these promoters as well as for a σ^{70} -dependent host *ompX* transcript throughout the infection. Primer extension products corresponding to viral σ^{70} -dependent promoters appear 5 minutes post-infection and their abundance remains constant throughout the infection. Likewise, the amount of primer extension product corresponding to host *ompX* gene promoter remains constant throughout the infection. It thus follows that there is no mechanism for σ^{70} -dependent transcription shut off during phage 7-11 infection.

For phage phiEco32, a bioinformatic search resulted in prediction of a putative σ^{70} -dependent promoter within gene 37, which could be used for transcription of downstream phiEco32 sigma factor gene 36 [6]. For phage 7-11, we and others failed to predict a σ^{70} -promoter in the intergenic region between genes 47 and 48 or within gene 48. Nevertheless, we performed primer extension reactions using RNA from infected cells and a primer annealing downstream of gene 47. A distinct primer extension product was detected, however, its accumulation pattern corresponded to late transcript class (Fig. 5B). No SaPh711_gp47 promoter consensus was located upstream of the mapped position of primer extension end point, which mapped in the intergenic region between genes 48 and 47. This intergenic region contains a sequence that can form a stem-loop structure with a GTTCGG loop (Fig. 5C). The stem-loop structure is also found in the corresponding position of phage GAP52 genome. We assume that the primer extension product that maps upstream of the 7-11 σ gene arises not due to transcription initiation but is either a result of *in vivo* processing or *in vitro* reverse transcription stalling on a hairpin formed on a phage transcript.

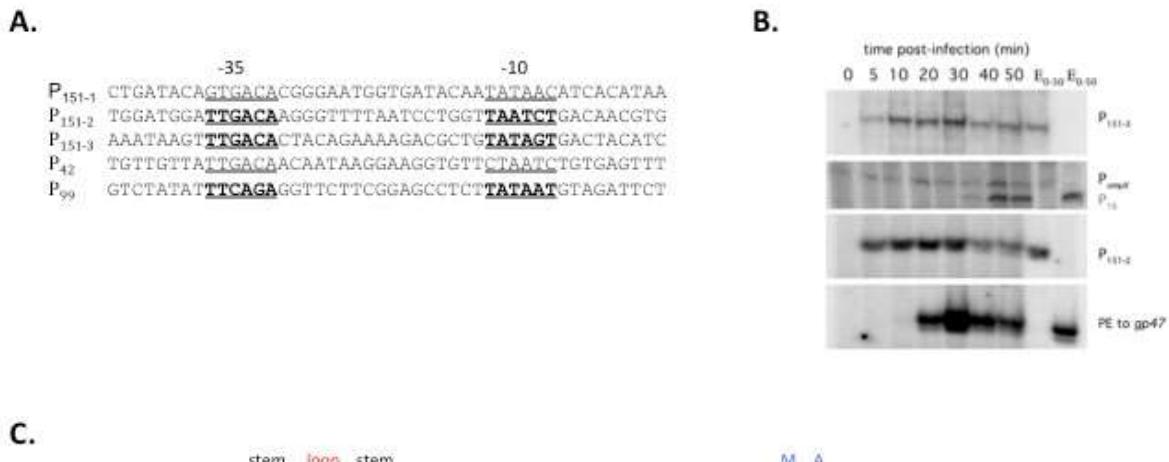


Figure 5. Kinetics of accumulation of σ^{70} -dependent transcripts during 7-11 infection and identification of a potential regulatory hairpin in front of gene 47. A. Sequences of all predicted σ^{70} -promoters of bacteriophage 7-11 (from NC_015938 file) are shown; putative -35 and -10 promoter consensus elements are underlined/indicated by bold typeface. B. Results of multiplex primer extension analysis with total RNA extracted from phage-infected cells collected at various time points post-infection using primers designed to reveal transcription from bioinformatically predicted σ^{70} -promoters shown in panel A. Only promoters in front of gene 151 (151-2 and 151-3) were validated. A primer annealing downstream of late promoter P16 was used as a control to indicate late gene expression. A primer annealing to *Salmonella* Newport gene *ompX* mRNA was used to follow the fate of a host transcript. Results of reaction of extension of a primer annealing downstream of gene 47 are shown at the bottom. On the right, products of extension reactions with individual primers on RNA sample prepared by mixing RNA samples from each time point (E0-50) are shown (a control to establish the identity of primer extension products). The reaction products were separated in 6% denaturing polyacrylamide gel and visualized using PhosphorImager. C. Identification of type (X)₅GTTCGG(Y)₅ hairpins in the intergenic region in front of SaPh711_gp47 and phage GAP52 σ factor gp73 genes. Conserved nucleotides are shown in boldface, protein-coding sequence – in blue, loop of the hairpin – in red.

4. Discussion

In this work we report analysis of temporal gene expression regulation of phiEco32-like bacteriophage 7-11. Unlike phiEco32 and most other lytic phages, 7-11 genes belong to just two expression classes: early and late. Early transcription initiates from two strong σ^{70} -dependent promoters located at the right end of the genome and proceeds leftward,

resulting in long polycistronic transcripts covering almost half of the genome. The first late promoter is located upstream of gene 69. Gene 47, coding for 7-11 σ factor responsible for late transcription, shall be transcribed, at least initially, by RNAP molecules that initiated transcription from early σ^{70} -dependent promoters. Since 47 is located deep in the leftward transcribed cluster of late phage genes, there shall be some transcription read-through from early promoters to allow late transcription to occur. It is highly likely that 7-11 employs some strategy to antiterminate transcription that remains to be defined. The amount of gene 47 transcripts can subsequently strongly increase through a positive feedback loop due to activity of late P₄₈ promoter and thus orchestrating a switch from early to late phage transcription.

Robust production of SaPh711_gp47 in amounts necessary to compete with σ^{70} may be further stimulated by a sequence in the intergenic region between genes 47 and 48 that can form a stem-loop structure with a GTTCGG loop (Fig. 5C). An identical sequence is found in the corresponding place of GAP45, a 7-11 relative that has only 10% identity with 7-11 genome wide. Similar sequences (X)₂₋₈CUUCGG(Y)₂₋₈ (where X is complement to Y, G-U pairs allowed) are commonly found, predominantly in intergenic regions, of many phages [27]. They are especially abundant in T4 phage and it has been shown that primer extension reactions by reverse transcriptase terminate at T4 CUUCGG hairpins [27]. The known functions of such hairpins vary. In T4, a 5' CUUCGG hairpin confers stability to T4 gene 32 (DNA polymerase) transcript [28,29]. The highly structured single-stranded RNA genome of the MS2 phage contains a CUUCGG hairpin at its 3'-end, where it was shown to be involved in temporally regulated rounds of translation and replication of the phage genome [30]. We propose that a similar mechanism may be operational in the case of the 7-11 and GAP45 late σ transcripts.

There does not appear to be specific shut off of either early phage or host transcription during 7-11 development. Presumably, SaPh711_gp47, which binds RNAP core better than σ^{70} , is alone able to efficiently compete for the binding such that enough SaPh711_gp47 holoenzyme to serve the needs of the virus is formed. The lack of host transcription shut off and the absence of the middle expression class of phage genes may be due to the absence of a homologue of phi32_gp79, a middle gene product that inhibits σ^{70} -transcription and activates phi32_gp36-dependent transcription [6].

Promoters recognized by SaPh711_gp47 consist of two consensus elements: GTAA₁₆-aCTA. Promoters recognized by phi32_gp36 consist of a single consensus element tAATGTAtA. For both cases transcription start sites are located 6-9 bp upstream relatively to the promoter sequence. The absence of the -35 element in phi32_gp36 promoter can be explained by the absence of homology with region 4 of other sigma factors (region 4 is responsible for the recognition of -35 promoter elements) in phi32_gp36. Conversely, amino acid sequence in the C-terminal part of SaPh711_gp47 is homologous to region 4 of known σ factors and should be therefore capable of sequence-specific recognition of DNA.

The 7-11 σ factor, SaPh711_gp47, is a distant homologue of phi32_gp36 and other phiEco32-like phages σ factors. The closest homologues of SaPh711_gp47 are σ^{24} from *Winogradskyella* sp. PG-2 (WP_045474069.1), a σ^{70} family factor from *Aeribacillus pallidus* CIC9 (WP_094245485.1) and a hypothetical protein from *Desulfuromonadales* bacterium (HJ97285.1). Promoter specificity for proteins most closely related to SaPh711_gp47 is unknown. Given the evolutionary distance between phiEco32 and 7-11 σ factors, it is clear that the ancestors of these phages must have acquired their σ factor genes independently and from different sources. Yet, the locations (but not the sequences) of promoters recognized by phi32_gp36 and SaPh711_gp47 holoenzymes in their respective genomes are conserved, providing an interesting evolutionary example of how functional constraints on the expression strategy of phage genes lead to common but independent solutions. Given that organized transcription of non-early phage genes in phi32 and 7-11 group phages depends on independently acquired σ factors recognizing multiple distinct promoters, we predict that further analysis of related phages shall either reveal interesting

hybrid variants or conversely, additional phages with other, yet to be described, pairs of σ -factor genes and promoters they recognize.

Supplementary Materials: The following supporting information can be downloaded at: www.mdpi.com/xxx/s1, Figure S1: title; Table S1: title; Video S1: title.

Author Contributions: Conceptualization: D.L. and K.S.; methodology: D.L. and K.S.; investigation: D.L., V.M.; writing—original draft preparation: D.L.; writing—review and editing: E.K., K.S. All authors have read and agreed to the published version of the manuscript.

Funding: The work was supported by a grant from the Ministry of Science and Higher Education of Russian Federation (agreement No. 075-10-2021-114 from 11 October 2021).

Acknowledgments: EK wishes to acknowledge Philip Morris International Systems Biology Fellowship Program.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Kropinski, A. M.; Lingohr, E. J.; Ackermann, H. W. The genome sequence of enterobacterial phage 7-11, which possesses an unusually elongated head. *Arch. Virol.* **2010**, *156*, 149–151, doi:10.1007/S00705-010-0835-5.
2. Savalia, D.; Westblade, L. F.; Goel, M.; Florens, L.; Kemp, P.; Akulenko, N.; Pavlova, O.; Padovan, J. C.; Chait, B. T.; Washburn, M. P.; Ackermann, H. W.; Mushegian, A.; Gabisonia, T.; Molineux, I.; Severinov, K. Genomic and Proteomic Analysis of phiEco32, a Novel *Escherichia coli* Phage. *J. Mol. Biol.* **2008**, *377*, 774, doi:10.1016/J.JMB.2007.12.077.
3. Mirzaei, M. K.; Eriksson, H.; Kasuga, K.; Haggard-Ljungquist, E.; Nilsson, A. S. Genomic, Proteomic, Morphological, and Phylogenetic Analyses of vB_EcoP_SU10, a Podoviridae Phage with C3 Morphology. *PLoS One* **2014**, *9*, doi:10.1371/JOURNAL.PONE.0116294.
4. Li, Y.; Chen, M.; Tang, F.; Yao, H.; Lu, C.; Zhang, W. Complete Genome Sequence of the Novel Lytic Avian Pathogenic Coliphage NJ01. *J. Virol.* **2012**, *86*, 13874, doi:10.1128/JVI.02727-12.
5. Nho, S.-W.; Ha, M.-A.; Kim, K.-S.; Kim, T.-H.; Jang, H.-B.; Cha, I.-S.; Park, S.-B.; Kim, Y.-K.; Jung, T.-S. Complete Genome Sequence of the Bacteriophages ECBP1 and ECBP2 Isolated from Two Different *Escherichia coli* Strains. *J. Virol.* **2012**, *86*, 12439, doi:10.1128/JVI.02141-12.
6. Pavlova, O.; Lavysh, D.; Klimuk, E.; Djordjevic, M.; Ravcheev, D. A.; Gelfand, M. S.; Severinov, K.; Akulenko, N. Temporal regulation of gene expression of the *Escherichia coli* bacteriophage phiEco32. *J. Mol. Biol.* **2012**, *416*, 389–399, doi:10.1016/j.jmb.2012.01.002.
7. Klimuk, E.; Akulenko, N.; Makarova, K. S.; Ceyssens, P. J.; Volchenkov, I.; Lavigne, R.; Severinov, K. Host RNA polymerase inhibitors encoded by φ KMV-like phages of *pseudomonas*. *Virology* **2013**, *436*, 67–74, doi:10.1016/j.virol.2012.10.021.
8. Klimuk, E.; Mekler, V.; Lavysh, D.; Serebryakova, M.; Akulenko, N.; Severinov, K. Novel *Escherichia coli* RNA Polymerase Binding Protein Encoded by Bacteriophage T5. *Viruses* **2020**, *12*, doi:10.3390/V12080807.
9. Mironov, A. A.; Vinokurova, N. P.; Gel'fand, M. S. [Software for analyzing bacterial genomes]. *Mol. Biol. (Mosk.)* **2000**, *34*, 253–262.
10. Sambrook, J.; Fritsch, E. F.; Maniatis, T. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor laboratory press; 1989; ISBN 13: 978-0879693091.
11. Decker, K.; Krauel, V.; Meesmann, A.; Heller, K. J. Lytic conversion of *Escherichia coli* by bacteriophage T5: blocking of the FhuA receptor protein by a lipoprotein expressed early during infection. *Mol. Microbiol.* **1994**, *12*, 321–332, doi:10.1111/j.1365-2958.1994.tb01020.x.
12. Kashlev, M.; Nudler, E.; Severinov, K.; Borukhov, S.; Komissarova, N.; Goldfarb, A. Histidine-tagged RNA polymerase of *Escherichia coli* and transcription in solid phase. *Methods Enzymol.* **1996**, *274*, 326–334, doi:10.1016/S0076-6879(96)74028-4.
13. Severinov, K.; Darst, S. A. A mutant RNA polymerase that forms unusual open promoter complexes. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 13481–13486, doi:10.1073/pnas.94.25.13481.
14. Frange, P.; Leruez-Ville, M. Maribavir, brincidofovir and letermovir: Efficacy and safety of new antiviral drugs for treating cytomegalovirus infections. *Médecine Mal. Infect.* **2018**, *48*, 495–502, doi:10.1016/J.MEDMAL.2018.03.006.
15. Klimuk, E.; Bogdanova, E.; Nagornyykh, M.; Rodic, A.; Djordjevic, M.; Medvedeva, S.; Pavlova, O.; Severinov, K. Controller protein of restriction-modification system Kpn2I affects transcription of its gene by acting as a transcription elongation roadblock. *Nucleic Acids Res.* **2018**, doi:10.1093/nar/gky880.
16. Mekler, V.; Severinov, K. RNA polymerase molecular beacon as tool for studies of RNA polymerase-promoter interactions. *Methods* **2015**, *86*, 19–26, doi:10.1016/j.ymeth.2015.04.033.

17. Knight, J. L.; Mekler, V.; Mukhopadhyay, J.; Ebright, R. H.; Levy, R. M. Distance-Restrained Docking of Rifampicin and Rifamycin SV to RNA Polymerase Using Systematic FRET Measurements: Developing Benchmarks of Model Quality and Reliability. *Biophys. J.* **2005**, *88*, 925, doi:10.1529/BIOPHYSJ.104.050187.
18. James, E.; Liu, M.; Sheppard, C.; Mekler, V.; Cámaras, B.; Liu, B.; Simpson, P.; Cota, E.; Severinov, K.; Matthews, S.; Wigneshweraraj, S. Structural and Mechanistic Basis for the Inhibition of Escherichia coli RNA Polymerase by T7 Gp2. *Mol. Cell* **2012**, *47*, 755, doi:10.1016/J.MOLCEL.2012.06.013.
19. Krisch, H. M.; Comeau, A. M. The immense journey of bacteriophage T4—From d'Hérelle to Delbrück and then to Darwin and beyond. *Res. Microbiol.* **2008**, *159*, 314–324, doi:10.1016/J.RESMIC.2008.04.014.
20. Abbasifar, R.; Griffiths, M. W.; Sabour, P. M.; Ackermann, H. W.; Vandersteegen, K.; Lavigne, R.; Noben, J. P.; Alanis Villa, A.; Abbasifar, A.; Nash, J. H. E.; Kropinski, A. M. Supersize me: Cronobacter sakazakii phage GAP32. *Virology* **2014**, *460–461*, 138–146, doi:10.1016/J.VIROL.2014.05.003.
21. Roucourt, B.; Lavigne, R. The role of interactions between phage and bacterial proteins within the infected cell: a diverse and puzzling interactome. *Environ. Microbiol.* **2009**, *11*, 2789–2805, doi:10.1111/J.1462-2920.2009.02029.X.
22. Bailey, T. L.; Boden, M.; Buske, F. A.; Frith, M.; Grant, C. E.; Clementi, L.; Ren, J.; Li, W. W.; Noble, W. S. MEME Suite: tools for motif discovery and searching. *Nucleic Acids Res.* **2009**, *37*, W202, doi:10.1093/NAR/GKP335.
23. Münch, R.; Hiller, K.; Barg, H.; Heldt, D.; Linz, S.; Wingender, E.; Jahn, D. PRODORIC: prokaryotic database of gene regulation. *Nucleic Acids Res.* **2003**, *31*, 266, doi:10.1093/NAR/GKG037.
24. Kadesch, T. R.; Rosenberg, S.; Chamberlin, M. J. Binding of Escherichia coli RNA Polymerase Holoenzyme to Bacteriophage T7 DNA Measurements of Binding at Bacteriophage T7 Promoter A1 Using a Template Competition Assay. *J. Mol. Biol.* **1982**, *155*, 1–29.
25. Hawley, D. K.; McClure, W. R. Compilation and analysis of Escherichia coli promoter DNA sequences. *Nucleic Acids Res.* **1983**, *11*, 2237–2255.
26. Campbell, E. A.; Korzheva, N.; Mustaev, A.; Murakami, K.; Nair, S.; Goldfarb, A.; Darst, S. A. Structural Mechanism for Rifampicin Inhibition of Bacterial RNA Polymerase. *Cell* **2001**, *104*, 901–912, doi:10.1016/S0092-8674(01)00286-0.
27. Tuerk, C.; Gauss, P.; Thermes, C.; Groebe, D. R.; Gayle, M.; Guild, N.; Stormo, G.; d'Aubenton-Carafa, Y.; Uhlenbeck, O. C.; Tinoco, I. CUUCGG hairpins: extraordinarily stable RNA secondary structures associated with various biochemical processes. *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85*, 1364, doi:10.1073/PNAS.85.5.1364.
28. Gorski, K.; Roch, J. M.; Prentki, P.; Krisch, H. M. The stability of bacteriophage T4 gene 32 mRNA: a 5' leader sequence that can stabilize mRNA transcripts. *Cell* **1985**, *43*, 461–469, doi:10.1016/0092-8674(85)90176-X.
29. Loayza, D.; Carpousis, A. J.; Krisch, H. M. Gene 32 transcription and mRNA processing in T4-related bacteriophages. *Mol. Microbiol.* **1991**, *5*, 715–725, doi:10.1111/J.1365-2958.1991.TB00742.X.
30. Vandenberghe, A.; Min Jou, W.; Fiers, W. 3'-Terminal nucleotide sequence (n equals 361) of bacteriophage MS2 RNA. *Proc. Natl. Acad. Sci. U. S. A.* **1975**, *72*, 2559, doi:10.1073/PNAS.72.7.2559.