Bacteriophage 4, e29399; January–December 2014; © 2014 Landes Bioscience

Molecular basis of RNA polymerase promoter specificity switch revealed through studies of *Thermus* bacteriophage transcription regulator

Konstantin Severinov,[1,2,3,] * Leonid Minakhin,[1] Shun-ichi Sekine,[4,5] Anna Lopatina,[2,6] and Shigeyuki Yokoyama[4,7]

[1]Waksman Institute; Rutgers; The State University of New Jersey; Piscataway, NJ USA; [2]St. Petersburg Polytechnical State University; St. Petersburg, Russia; [3]Skolkovo Institute of Science and Technology; Skolkovo, Russia; [4]RIKEN Systems and Structural Biology Center; Tsurumi-ku, Yokohama Japan; [5]Division of Structural and Synthetic Biology; RIKEN Center for Life Science Technologies; Tsurumi-ku, Yokohama Japan; [6]Institutes of Gene Biology and Molecular Genetics; Russian Academy of Sciences; Moscow, Russia; [7]RIKEN Structural Biology Laboratory; Tsurumi-ku, Yokohama Japan

Transcription initiation is the central point of gene expression regulation. Understanding of molecular mechanism of transcription regulation requires, ultimately, the structural understanding of consequences of transcription factors binding to DNA-dependent RNA polymerase (RNAP), the enzyme of transcription. We recently determined a structure of a complex between transcription factor gp39 encoded by a *Thermus* bacteriophage and *Thermus* RNAP holoenzyme. In this addendum to the original publication, we highlight structural insights that explain the ability of gp39 to act as an RNAP specificity switch which inhibits transcription initiation from a major class of bacterial promoters, while allowing transcription from a minor promoter class to continue.

Keywords: bacterial RNA polymerase, bacteriophage, inhibitor, sigma factor, transcription regulation

Submitted: 04/21/2014

Revised: 05/29/2014

Accepted: 05/29/2014

Published Online: 05/29/2014

Citation: Severinov K, Minakhin L, Sekine S, Lopatina A, Yokoyama S. Molecular basis of RNA polymerase promoter specificity switch revealed through studies of *Thermus* bacteriophage transcription regulator. Bacteriophage 2014; 4:e29399; http://dx.doi.org/10.4161/bact.29399

*Correspondence to: Konstantin Severinov; Email: severik@waksman.rutgers.edu; Shigeyuki Yokoyama; Email: yokoyama@riken.jp

Addendum to: Tagami S, Sekine S, Minakhin L, Esyunina D, Akasaka R, Shirouzu M, Kulbachinskiy A, Severinov K, Yokoyama S. Structural basis for promoter specificity switching of RNA polymerase by a phage factor. Genes Dev 2014; 28:521–31; http://dx.doi.org/10.1101/gad.233916.113; PMID:24589779

The first structure of a multisubunit DNA-dependent RNA polymerase (RNAP) core enzyme, from thermophilic bacterium *Thermus aquaticus*, was published in 1999[1] and heralded a new era in the studies of transcription mechanism. The structures of *Thermus* RNAP holoenzyme,[2,3] its promoter and transcription elongation complexes,[4-7] and complexes with antibiotics, substrates, and low-molecular weight regulators followed shortly.[8-12] Together, these structures provided a wealth of information about the mechanism of transcription but offered relatively less insight into the mechanism of transcription regulation.

Historically, phages provided some of the classical examples of transcription regulation and paradigms revealed during such studies are widely applicable to transcription regulation in other systems including higher organisms. The advantage of phages as a source of transcription regulators stems from the fact that all phages orchestrate expression of several temporal classes of their own genes and many phages shut down transcription of their hosts. This has to be achieved in a short period of time, which explains why many transcription regulators encoded by phages are very robust: they bind RNAP tightly and the consequences of their binding are generally strong. The best understood phage transcription regulators come, not surprisingly, from classical *Escherichia coli* phages such as λ, T4, and T7. Despite the wealth of genetic and biochemical data on the function of *E. coli* phage transcription factors, the structural understanding is lacking. The breakthroughs with *Thermus* RNAP structures were of little help, since known *E. coli* phage transcription regulators do not bind to, and therefore do not regulate the activity of the crystallizable *Thermus* enzyme. One strategy under the circumstances was to put effort into obtaining diffracting crystals of *E. coli* RNAP, first alone and then in complex with known phage-encoded transcription factors. Such a strategy bore fruit in late 2013 when a structure of the *E. coli* σ^{70} RNAP holoenzyme complexed with

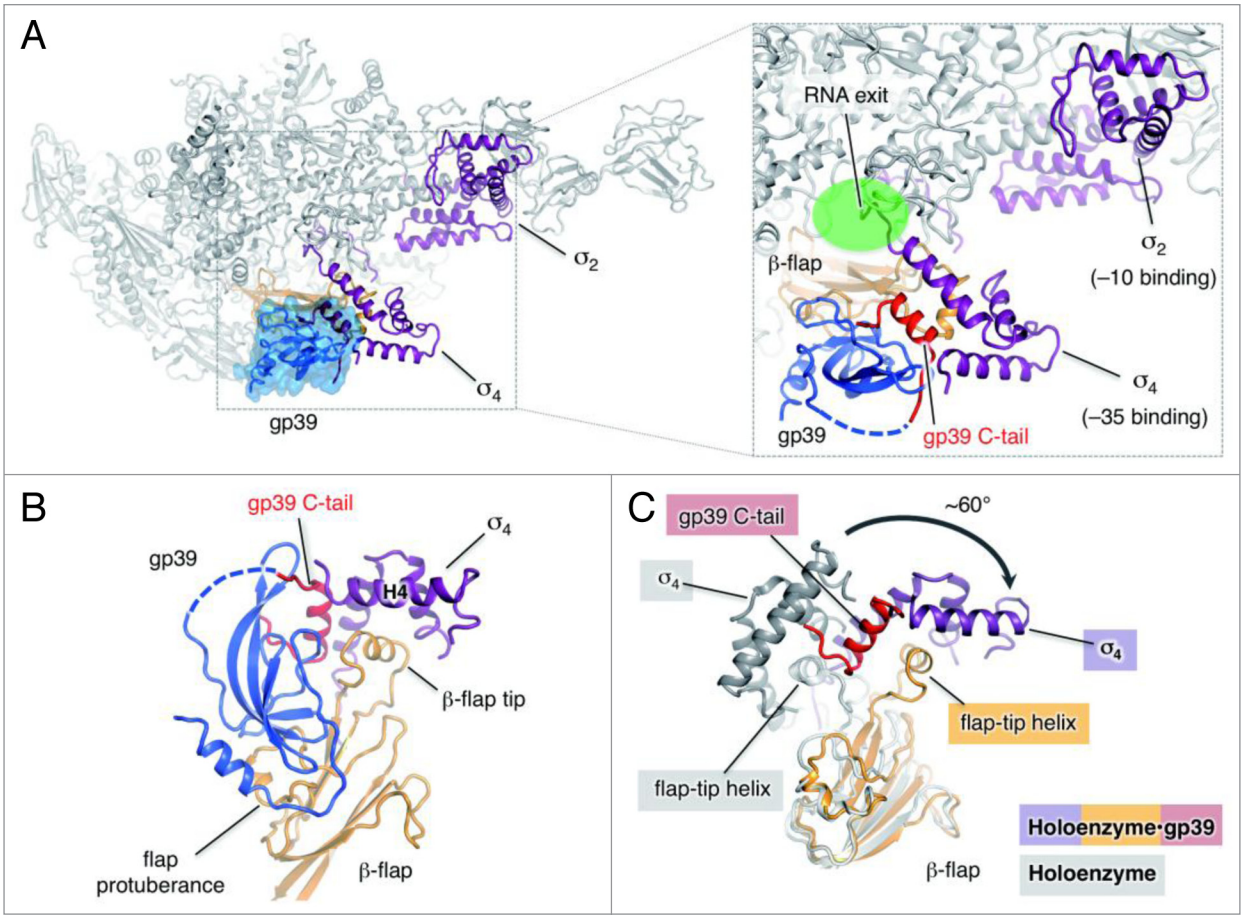

Figure 1. The structure of the *Thermus thermophilus* σ^A holoenzyme complex with P23-45 phage gp39 protein. (**A**) The overall structure of the complex is shown on the left. The RNAP core is shown in gray except for the β-flap shown in golden color. The σ^A subunit DNA binding domains 2 and 4 are shown in magenta; gp39 is shown in blue. On the righthand-side of the panel, the part of the structure containing gp39 is expanded with elements of gp39 structure discussed in the text (the core and the C-terminal tail) highlighted in, respectively, blue and red. The exit point of nascent RNA from the complex is shown as a green oval. (**B**) A close-up view of gp39-RNAP interaction site. See text for details. (**C**) A superposition of the β-flap•σ_4 areas in the free RNAP holoenzyme (gray) and in the gp39•RNAP holoenzyme structures. The arrow indicates the extent of gp39-induced rotation of DNA-binding helix of σ_4.

bacteriophage T7-encoded transcription initiation inhibitor gp2 was published.[13] The structure confirmed many of the inferences obtained by earlier functional analyses (i.e., the location of inhibitor binding site, the main mode of promoter binding inhibition, etc.[14-16]) and also provided non-trivial insights into the finer details of the inhibition. The second strategy was to try to isolate phages infecting *Thermus,* identify transcription factors encoded by these phages and characterize them both functionally and structurally. This line of research came as a close second to the "*E. coli*-centric" approach with a structure of the *Thermus thermophilus* σ^A RNAP holoenzyme complexed with bacteriophage P23-45 dual function transcription regulator gp39.[17]

Almost ten years ago, a paper describing a collection of diverse phages infecting bacteria of the *Thermus* genome was published.[18] Initially, *Thermus* YS40 phage, a myophage the size of the *E. coli* T4 phage known to be obsessive-compulsive about transcription regulation of its genes,[19] was deemed to be the best source of *Thermus* RNAP interacting regulators. However, its analysis yielded no transcription factors.[20,21] Despite this setback, the search for *Thermus* phage-encoded transcription factors continued with P23-45, a phage isolated from a Kamchatka Geyser valley hot spring. P23-45 is a siphovirus with an extremely (1 micron) long tail and a ~80 kbp double-stranded DNA genome. The early genes of P23-45 are likely transcribed by a highly unusual RNAP encoded by the phage genome[22] while a very long late gene cluster containing viral structural gene (39 genes constituting more that 45 kbp of DNA, including the 15,009 bp-long tape-measure protein gene, responsible for the unusually long tail of the virus) is transcribed by host RNAP from a late promoter that belongs to the extended –10 bacterial promoter class. Unlike the major –10/–35 bacterial RNAP promoters, extended –10 class promoters activity depends on RNAP interaction, through σ subunit region 2 (σ_2), with just one promoter consensus element. The second interaction, between σ subunit region 4 (σ_4) and the –35 promoter element is not required for recognition of promoters of this class. A 141 aminoacid-long P23-45 protein gp39 was identified as a component of host RNAP affinity

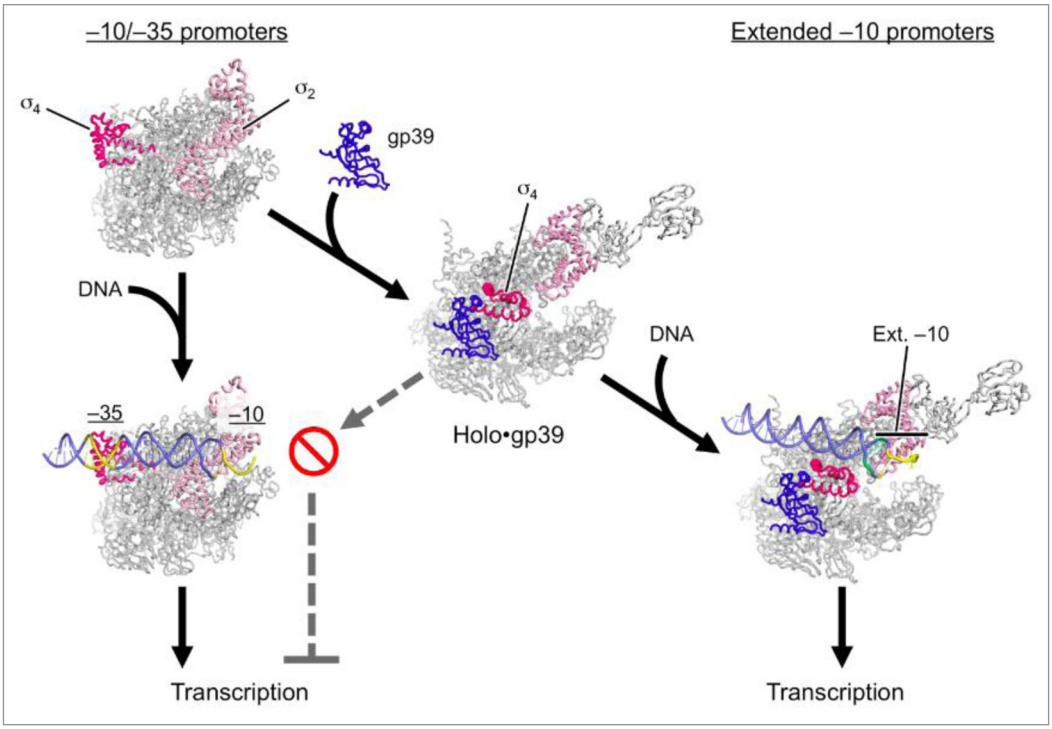

Figure 2. A structural model of gp39-controlled RNAP promoter specificity switch. On the left, the interaction of RNAP holoenzyme with a –10/–35 promoter is shown. Notice that the distance between the DNA binding regions of σ_2 and σ_4 in free holoenzyme (top) matches the distance between the –10 and –35 promoter consensus elements (below), allowing promoter complex formation. The binding of gp39 (middle) decreases the distance between σ_2 and σ_4 making productive interaction with –10/–35 promoters impossible. Promoter complex formation with extended –10 promoters can still occur (right).

purified from P23-45-infected *Thermus* cultures.[22] Recombinant gp39 was shown to strongly inhibit transcription from –10/–35 promoters but not from extended –10 promoters.[22] Therefore, gp39 could execute an RNAP specificity switch to direct host RNAP from host genes transcription to phage late gene transcription.

The structure of gp39 complex with the σ^A RNAP holoenzyme[17] is shown in **Figure 1A**. The gp39 structure consists of 1) a core (residues 1–107) comprising the central β sheet and the N-terminal α helix (indicated by blue color in **Figures 1A–C**), and 2) the C-terminal tail (residues 118–141) (indicated in red); these two parts are connected by a linker that is mostly disordered. The primary site of gp39 binding is the RNAP β subunit flap domain (**Fig. 1B**). The gp39 core binds primarily to flap protuberance (β subunit residues 723–740) that protrudes from a three-strand β sheet that forms the main part of the β-flap (**Fig. 1B**). The interaction relies on shape complementarity and extensive hydrophobic/hydrophilic interactions. Disruption of these interactions by point mutations prevents gp39-RNAP interaction.[17] The secondary site of interaction is afforded by gp39 α-helix of C-terminal tail, which is fixed through interactions with gp39 core, the tip of the β-flap, and σ_4.

In the holo•gp39 complex structure, the C-terminal helix of gp39 displaces an α-helix of σ_4 that normally interacts with the β-flap tip (**Fig. 1C**). Consequently, σ_4 and the β-flap tip are rotated by ~60° relative to the rest of the flap domain. The relative positions of σ_4 and the β-flap tip are not changed during this motion but, as the end result, the position of σ_4 is shifted by ~45 Å. As a result of gp39-induced rotation the simultaneous interaction of σ_2 and σ_4 with, respectively, –10 and –35 promoter consensus elements, becomes impossible (**Fig. 2**). In contrast, the holo•gp39 complex structure is compatible with promoter complex formation on extended –10 type promoters (**Fig. 2**). Thus, the new structure explains, at the molecular level, the mechanism of promoter-specific inhibition of promoter binding by gp39. A gp39 mutant lacking the C-terminal tail binds RNAP normally but is severely defective in transcription initiation inhibition, underscoring the importance of the secondary interaction between gp39 C-terminal helix and σ_4 for transcription regulation mechanism.

Thermus phages related to P23-45 appear to be globally distributed around the world. We recently isolated one such phage, named phiFa, from around mount Etna and determined its genome sequence (**Fig. 3A**). While standard bioinformatics analysis failed to identify a homolog of gp39 in phiFA, manual analysis revealed that a product of phiFa gene *15* contains counterparts of gp39 residues important for primary interaction with the β-flap protuberance (**Fig. 3B**). No similarity to gp39 outside this area could be detected, however. The phiFa gene *15* is located at the boundary of the early and middle genes clusters, i.e., in the same position as gene *39* in P23-45 (**Fig. 3A**). Like gp39, gp15 inhibits transcription initiation by *Thermus* RNAP from –10/–35 class

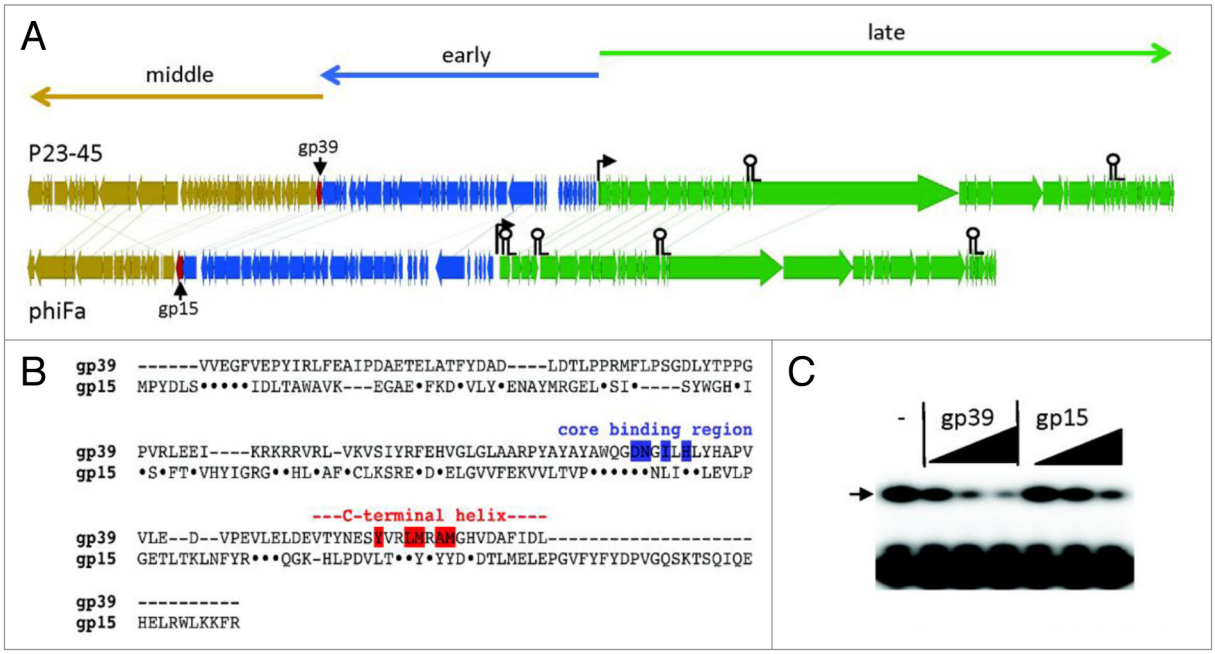

Figure 3. A distant gp39 homolog encoded by the phiFA *Thermus* phage. (**A**) Schematic comparison of P23-45 and phiFA phage genomes. Genes belonging to different temporal expression classes are color-coded. Arrows indicate the direction of transcription. Homologous genes in two genomes are connected by thin punctuated lines. A P23-45 gene coding for gp39, at the boundary of the early and middle gene clusters, is indicated. A rightward arrow indicates the P23-45 late promoter. Two validated transcription terminators in the late gene cluster are indicated. (**B**) Sequence alignment of P23-45 gp39 and phiFA gp15 (single aminoacid code). Dots indicate identities, hyphens – gaps. The regions of gp39 that participate in the primary interaction with the RNAP core (through the β-flap protuberance) and that form the C-terminal helix are highlighted in blue and red, respectively. Individual amino acids involved in interactions with the β-flap and σ_4 are indicated by a correspondingly colored background in the alignment. (**C**) The results of abortive in vitro transcription by *Thermus* RNAP σ^A holoenzyme from a –10/–35 class T7 A1 promoter in the absence, or in the presence of increasing concentrations of P23-45 gp39 or phiFA gp15 are shown. The transcript CpApU is indicated by a horizontal arrow. An autoradiogram of a denaturing polyacrylamide gel is shown. A strong radioactive band present at the bottom of each lane is unincorporated ^{32}P labeled UTP substrate.

promoters (**Fig. 3C**). While the primary RNAP binding site of gp39 and gp15 is likely identical, the phiFa protein does not have residues matching P23-45 gp39 amino acids forming the C-terminal tail helix that packs against σ_4 and its C-terminal tail is considerably longer than that of gp39 (**Fig. 3B**). Thus, the mechanism of transcription initiation inhibition may be different for different gp39-like proteins. Structure determination of gp15, alone and in complex with *Thermus* RNAP holoenzyme, will allow us to test this conjecture in the near future.

The σ_4•β-flap interface emerges as a critical point accepting regulatory inputs from the "upstream" side of RNAP. The case of *E. coli* bacteriophage T4 AsiA protein provides another example. AsiA is responsible for a switch from host and early phage promoters (belong to the –10/–35 class) to middle phage promoters (belong to the extended –10 class). Unlike gp39, AsiA binds to σ^{70} region 4. While the structural information is lacking, available biochemical and biophysical data suggest that the AsiA binding disrupts the σ_4•β-flap interaction and also likely reorients both the flap and σ_4, thus achieving the same effect as gp39. Unlike AsiA,[23] gp39 is also a transcription elongation factor, strongly antiterminating transcription by *Thermus* RNAP on all intrinsic transcription terminators tested.[24] This function is likely biologically relevant, allowing gp39 help host RNAP transcribe through a very long cluster of late phage genes separated by several terminators (**Fig. 3A**). The transcription antitermination function of gp39 must be σ-independent, as σ dissociates from RNAP core upon promoter escape. Indeed, the gp39 mutant lacking the C-terminal helix needed for transcription initiation inhibition mentioned above is fully capable of antiterminating transcription. The location of gp39 primary binding site, close to the nascent RNA exit pathway (**Fig. 1A**), suggests that gp39 may affect transcription termination properties by altering RNAP interactions with exiting RNA. Interestingly, phage λ transcription antitermination protein Q also may affect transcription termination properties of *E. coli* RNAP through the flap domain.[25,26] Irrespective of which—*E. coli* RNAP•λ Q or *Thermus* RNAP•P23-45 gp39 elongation complex—structure is determined first, comparative analyses of complexes between *E. coli* and *Thermus* phage-encoded transcription regulators with cognate RNAPs will continue to provide insights into molecular mechanisms of transcription regulators for years to come.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

This work was supported in part by a Japan Society for the Promotion of Science Grant-in-Aid for Scientific Research (to SS and SY); the Targeted

Proteins Research Program, the Ministry of Education, Culture, Sports, Science and Technology of Japan (to SY), the National Institutes of Health grant R01 59295 to K.S.; the Russian Academy of Sciences Presidium Program in Molecular and Cellular Biology (to KS) and the Ministry of Education and Science of Russia project 14.B25.31.0004 (to KS).

References

1. Zhang G, Campbell EA, Minakhin L, Richter C, Severinov K, Darst SA. Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 A resolution. Cell 1999; 98:811-24; PMID:10499798; http://dx.doi.org/10.1016/S0092-8674(00)81515-9
2. Vassylyev DG, Sekine S, Laptenko O, Lee J, Vassylyeva MN, Borukhov S, Yokoyama S. Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 A resolution. Nature 2002; 417:712-9; PMID:12000971; http://dx.doi.org/10.1038/nature752
3. Murakami KS, Masuda S, Darst SA. Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 A resolution. Science 2002; 296:1280-4; PMID:12016306; http://dx.doi.org/10.1126/science.1069594
4. Zhang Y, Feng Y, Chatterjee S, Tuske S, Ho MX, Arnold E, Ebright RH. Structural basis of transcription initiation. Science 2012; 338:1076-80; PMID:23086998; http://dx.doi.org/10.1126/science.1227786
5. Vassylyev DG, Vassylyeva MN, Perederina A, Tahirov TH, Artsimovitch I. Structural basis for transcription elongation by bacterial RNA polymerase. Nature 2007; 448:157-62; PMID:17581590; http://dx.doi.org/10.1038/nature05932
6. Weixlbaumer A, Leon K, Landick R, Darst SA. Structural basis of transcriptional pausing in bacteria. Cell 2013; 152:431-41; PMID:23374340; http://dx.doi.org/10.1016/j.cell.2012.12.020
7. Murakami KS, Masuda S, Campbell EA, Muzzin O, Darst SA. Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. Science 2002; 296:1285-90; PMID:12016307; http://dx.doi.org/10.1126/science.1069595
8. Campbell EA, Korzheva N, Mustaev A, Murakami K, Nair S, Goldfarb A, Darst SA. Structural mechanism for rifampicin inhibition of bacterial rna polymerase. Cell 2001; 104:901-12; PMID:11290327; http://dx.doi.org/10.1016/S0092-8674(01)00286-0
9. Temiakov D, Zenkin N, Vassylyeva MN, Perederina A, Tahirov TH, Kashkina E, Savkina M, Zorov S, Nikiforov V, Igarashi N, et al. Structural basis of transcription inhibition by antibiotic streptolydigin. Mol Cell 2005; 19:655-66; PMID:16167380; http://dx.doi.org/10.1016/j.molcel.2005.07.020
10. Vassylyev DG, Vassylyeva MN, Zhang J, Palangat M, Artsimovitch I, Landick R. Structural basis for substrate loading in bacterial RNA polymerase. Nature 2007; 448:163-8; PMID:17581591; http://dx.doi.org/10.1038/nature05931
11. Vassylyev DG, Svetlov V, Vassylyeva MN, Perederina A, Igarashi N, Matsugaki N, Wakatsuki S, Artsimovitch I. Structural basis for transcription inhibition by tagetitoxin. Nat Struct Mol Biol 2005; 12:1086-93; PMID:16273103; http://dx.doi.org/10.1038/nsmb1015
12. Campbell EA, Pavlova O, Zenkin N, Leon F, Irschik H, Jansen R, Severinov K, Darst SA. Structural, functional, and genetic analysis of sorangicin inhibition of bacterial RNA polymerase. EMBO J 2005; 24:674-82; PMID:15692574; http://dx.doi.org/10.1038/sj.emboj.7600499
13. Bae B, Davis E, Brown D, Campbell EA, Wigneshweraraj S, Darst SA. Phage T7 Gp2 inhibition of Escherichia coli RNA polymerase involves misappropriation of σ70 domain 1.1. Proc Natl Acad Sci U S A 2013; 110:19772-7; PMID:24218560; http://dx.doi.org/10.1073/pnas.1314576110
14. Nechaev S, Severinov K. Inhibition of Escherichia coli RNA polymerase by bacteriophage T7 gene 2 protein. J Mol Biol 1999; 289:815-26; PMID:10369763; http://dx.doi.org/10.1006/jmbi.1999.2782
15. Cámara B, Liu M, Reynolds J, Shadrin A, Liu B, Kwok K, Simpson P, Weinzierl R, Severinov K, Cota E, et al. T7 phage protein Gp2 inhibits the Escherichia coli RNA polymerase by antagonizing stable DNA strand separation near the transcription start site. Proc Natl Acad Sci U S A 2010; 107:2247-52; PMID:20133868; http://dx.doi.org/10.1073/pnas.0907908107
16. Mekler V, Minakhin L, Sheppard C, Wigneshweraraj S, Severinov K. Molecular mechanism of transcription inhibition by phage T7 gp2 protein. J Mol Biol 2011; 413:1016-27; PMID:21963987; http://dx.doi.org/10.1016/j.jmb.2011.09.029
17. Tagami S, Sekine S, Minakhin L, Esyunina D, Akasaka R, Shirouzu M, Kulbachinskiy A, Severinov K, Yokoyama S. Structural basis for promoter specificity switching of RNA polymerase by a phage factor. Genes Dev 2014; 28:521-31; PMID:24589779; http://dx.doi.org/10.1101/gad.233916.113
18. Yu MX, Slater MR, Ackermann HW. Isolation and characterization of Thermus bacteriophages. Arch Virol 2006; 151:663-79; PMID:16308675; http://dx.doi.org/10.1007/s00705-005-0667-x
19. Miller ES, Kutter E, Mosig G, Arisaka F, Kunisawa T, Rüger W. Bacteriophage T4 genome. Microbiol Mol Biol Rev 2003; 67:86-156; PMID:12626685; http://dx.doi.org/10.1128/MMBR.67.1.86-156.2003
20. Naryshkina T, Liu J, Florens L, Swanson SK, Pavlov AR, Pavlova NV, Inman R, Minakhin L, Kozyavkin SA, Washburn M, et al. Thermus thermophilus bacteriophage phiYS40 genome and proteomic characterization of virions. J Mol Biol 2006; 364:667-77; PMID:17027029; http://dx.doi.org/10.1016/j.jmb.2006.08.087
21. Sevostyanova A, Djordjevic M, Kuznedelov K, Naryshkina T, Gelfand MS, Severinov K, Minakhin L. Temporal regulation of viral transcription during development of Thermus thermophilus bacteriophage phiYS40. J Mol Biol 2007; 366:420-35; PMID:17187825; http://dx.doi.org/10.1016/j.jmb.2006.11.050
22. Berdygulova Z, Westblade LF, Florens L, Koonin EV, Chait BT, Ramanculov E, Washburn MP, Darst SA, Severinov K, Minakhin L. Temporal regulation of gene expression of the Thermus thermophilus bacteriophage P23-45. J Mol Biol 2011; 405:125-42; PMID:21050864; http://dx.doi.org/10.1016/j.jmb.2010.10.049
23. Severinova E, Severinov K, Darst SA. Inhibition of Escherichia coli RNA polymerase by bacteriophage T4 AsiA. J Mol Biol 1998; 279:9-18; PMID:9636696; http://dx.doi.org/10.1006/jmbi.1998.1742
24. Berdygulova Z, Esyunina D, Miropolskaya N, Mukhamedyarov D, Kuznedelov K, Nickels BE, Severinov K, Kulbachinskiy A, Minakhin L. A novel phage-encoded transcription antiterminator acts by suppressing bacterial RNA polymerase pausing. Nucleic Acids Res 2012; 40:4052-63; PMID:22238378; http://dx.doi.org/10.1093/nar/gkr1285
25. Deighan P, Diez CM, Leibman M, Hochschild A, Nickels BE. The bacteriophage lambda Q antiterminator protein contacts the beta-flap domain of RNA polymerase. Proc Natl Acad Sci U S A 2008; 105:15305-10; PMID:18832144; http://dx.doi.org/10.1073/pnas.0805757105
26. Vorobiev SM, Gensler Y, Vahedian-Movahed H, Seetharaman J, Su M, Huang JY, Xiao R, Kornhaber G, Montelione GT, Tong L, et al. Structure of the DNA-binding and RNA-polymerase-binding region of transcription antitermination factor λQ. Structure 2014; 22:488-95; PMID:24440517; http://dx.doi.org/10.1016/j.str.2013.12.010