MiniReview

Old phage, new insights: Two recently recognized mechanisms of transcriptional regulation in bacteriophage T4 development

E.N. Brody a, G.A. Kassavetis b, M. Ouhammouch a, G.M. Sanders b, R.L. Tinker b, E.P. Geiduschek b

a Department of Biological Sciences, State University of New York at Buffalo, Buffalo, NY 14260-1300, USA
b Department of Biology and Center for Molecular Genetics, University of California, San Diego, La Jolla, CA 92093-0634, USA

Received 1 February 1995; revised 20 February 1995; accepted 20 February 1995

Abstract

The regulation of bacteriophage T4 middle and late gene expression involves previously unrecognized mechanisms. Middle transcription requires a DNA-binding transcriptional activator and a σ 70-binding co-activator. The coupling of late transcription to DNA replication is effected by a DNA-tracking protein that is loaded onto DNA by an assembly factor at enhancer-like entry sites. Late transcription also requires an RNA polymerase core-binding co-activator. The co-activators of T4 middle and late transcription share the property of depressing unactivated, basal transcription.

Keywords: Transcriptional regulation; Bacteriophage T4; DNA-tracking proteins; Anti-sigma; Replication–transcription coupling; Co-activators

1. Introduction

Recent work on the long-standing problem of how expression of the bacteriophage T4 genome is regulated yields new and generalizable insights into transcription mechanisms. In this brief account, we present what has been found and comment on its implications. The programme of phage T4 gene expression (approx. 165 kb genome; > 200 genes) depends on the sequential utilization of three classes of promoters: early, middle and late (Table 1). ( Needless to say, other effectors also influence the total programme of gene expression during phage T4 development; the reader is referred to a recently published book [1] for other aspects of this subject.) We focus our attention on the newly understood mechanisms governing the efficient utilization of middle and late promoters.

2. Middle genes

2.1. Middle promoters and middle transcription: the MotA protein

Twenty-six T4 middle promoters [1] share the −10 element that is characteristic of strong E. coli promoters served by the σ 70 holoenzyme, but lack the conjugate −35 element (Table 1). Instead, centred half a turn forward along the DNA, near bp −30, each middle promoter has a site that binds the MotA protein [2–4]. Deletion of the MotA gene generates a failure of middle promoter utilization and is lethal for phage development, unless complemented by a plasmid copy of the gene. The C-terminal domain of the MotA protein (Mr 23.577) is involved in binding to the −30 DNA consensus [4]. NMR spectroscopic analysis suggests a structural
resemblance of this domain to the more-studied TATA binding protein (TBP) [5,6], which interacts with the minor groove of its target DNA, bending and unwinding the latter [7,8]. Inspection of amino acid residues at presumptively structurally equivalent positions in MotA and TBP re-inforces the idea that the two proteins bind to DNA in similar fashion. The introduction of a TBP-like DNA bend by MotA in the centre of its binding site (near bp −30) would have interesting implications for its mechanism of action, since certain promoter complexes of σ70-RNA polymerase holoenzyme (E.σ70; E = core enzyme) are known to sharply bend DNA between the −10 and −35 σ70-binding elements [9].

In a general sense, MotA protein functions as does the cII protein of bacteriophage λ, which binds to its conjugate DNA site centred on bp −35 of promoters and is thought to act by directly contacting either the α or σ70 subunits of E. coli RNA polymerase (or both) [10,11]. However, something more is required for middle promoter utilization: purified MotA protein only stimulates middle mode transcription in vitro in conjunction with RNA polymerase isolated from T4-infected cells and not with RNA polymerase from uninfected E. coli [4,12].

2.2. AsiA protein

The missing ingredient has only recently been shown to be the AsiA protein [13,14]. Phage T4 carrying an asIA amber mutation abortively infect non-suppressing E. coli and are specifically defective for middle transcription. RNA synthesis patterns in motA− and asIA− infections are indistinguishable, suggesting that both of the corresponding gene products function in the same process [13]. The asIA gene codes for a 90-amino acid protein (M, 10583) whose biochemical properties have been investigated, on and off, since the early 1970s [15]. This protein binds tightly and specifically to σ70. When added to E.σ70 holoenzyme, the AsiA protein inhibits σ70-dependent transcription from E. coli promoters and T4 early promoters. Thus, for many years, this protein’s role was postulated to be: (i) inhibition of host transcription and T4 early transcription; (ii) indirect stimulation of late transcription by allowing gp55, the T4-encoded sigma family protein that directs late transcription, to successfully compete with σ70 for RNA polymerase core (E). In fact, the name anti-sigma A derives from these postulated functions. Finding AsiA required for middle transcription suggests that this protein plays a role in all classes of T4 transcription, most directly as a positive effector of middle transcription.

The demonstration that both MotA and AsiA are necessary for middle transcription in vivo is complemented by the recent demonstration that they are also sufficient in vitro: E. coli RNA polymerase holoenzyme (E.σ70), a middle promoter, and the two purified proteins MotA and AsiA, are all that is needed for the initiation of middle transcription. At saturating levels of both proteins, no T4 early promoters are recognized; all middle promoters (five have been tested) are activated. Although the AsiA protein alone is sufficient to inhibit early promoter recognition, it must cooperate with MotA and with σ70 for open promoter complex formation at middle promoters. Thus, σ70 is not merely an inhibitor of middle mode transcription, and the role of AsiA is not simply to remove that inhibition.

We are left with a picture of open complex formation at middle promoters that is best appreciated by

<table>
<thead>
<tr>
<th>Promoter recognition sequences [1]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Promoter</td>
</tr>
<tr>
<td>E. coli strong promoters</td>
</tr>
<tr>
<td>T4 early</td>
</tr>
<tr>
<td>(−30)</td>
</tr>
<tr>
<td>T4 middle</td>
</tr>
<tr>
<td>T4 late</td>
</tr>
</tbody>
</table>

* W, A or T; Y, T or C
reference to the current understanding of the structure and functions of the \( \sigma^{70} \) initiator subunit of RNA polymerase holoenzyme. Comparative analysis identifies four segments of evolutionarily conserved sequence in \( \sigma \)-family proteins that are further sub- 
visible (i.e. regions 1a, 1b, 2.1, 2.2, 2.3, 2.4, 3, 4.1 and 4.2). Key functions, such as interactions with two promoter elements (\(-10\) and \(-35\), see Table 1), with RNA polymerase core, and with certain transcriptional activators, are associated with certain conserved segments [41]. Conserved region 2.4 of \( \sigma^{70} \) interacts with the \(-10\) consensus sequence of \( E. \ coli \) promoters [16] and the corresponding region of gp55 should likewise interact with the identically placed \(-10\) sequence of T4 middle promoters. Conserved region 4.2 of \( \sigma^{70} \) ordinarily interacts with the \(-35\) consensus sequence in \( E. \ coli \) promoters, which is lacking in T4 middle promoters. However, region 4.2 of \( \sigma^{70} \) is included in the carboxy-terminal domain that interacts with the so-called class II transcriptional activators, PhoB, CRP (at the P1 gal promoter) and \( \lambda \) cl [17]. When MotA binds in the vicinity of bp \(-30\) of middle promoters, probably in the minor groove, it may lie close to, and on the same side of, the DNA helix as the carboxy-terminal domain of \( \sigma^{70} \). This leaves a number of possible roles for AsiA to play in middle promoter opening. Although no experimental evidence has yet been gathered, a first hypothesis to test is that DNA-bound MotA contacts \( \sigma^{70} \)-bound AsiA during promoter activation, in other words, that AsiA serves as the co-activator (in the conventional eukaryotic-transcription sense) of MotA.

2.3. Is AsiA the central T4 transcriptional regulator?

Saturating quantities of AsiA and MotA proteins completely re-programme \( E. \ coli \) RNA polymerase in vitro: transcription from early promoters is abolished and recognition of middle promoters is complete. As already mentioned, AsiA is thought to play yet another role in late RNA synthesis, although this remains to be proven: it is postulated that AsiA binding to \( \sigma^{70} \) allows gp55 to successfully compete for RNA polymerase core, E. Thus, AsiA positively controls middle transcription, by itself inhibits early transcription, and possibly is indirectly necessary for effective late transcription. It appears at first glance that more protein would be better for each of these important roles, but AsiA is not produced in vast quantities after T4 infection. In fact, its synthesis rate is limited by three factors: (i) the ribosome binding site (GGUG) controlling its translation is weak. (ii) The asiA gene is transcribed from an early promoter, and is therefore transcriptionally autoregulated. (iii) An internal GGAG of AsiA mRNA is a substrate for the T4-cncoded RegB nuclease, further limiting its availability for translation. It is clear that T4 has evolved refined mechanisms to limit the quantity of AsiA protein in the cell. Further work is needed to see which of the two proven and one postulated functions of this protein need to be so tightly controlled.

Other \( \sigma \)-binding proteins are known to inhibit transcription: the \( S. \ typhimurium \) FlgM binds to and inhibits the flagellar gene-specific \( \sigma^{F} \) [18]; in \( B. \ subtilis \) sporulation, the SpoIIB protein binds to and inhibits the forespore-specific \( \sigma^{F} \) [19,20]; the \( B. \ subtilis \) RbsW protein binds to and blocks the RNA polymerase core association of the secondary \( \sigma^{B} \) [21]. None of these three anti-\( \sigma \) proteins has any discernible amino acid sequence homology with AsiA; neither does any of them act as a transcriptional activator.

3. Late genes

3.1. Late promoters and the rules of gene expression

The approximately 40 T4 late promoters, which generate transcription of approximately 40% of the T4 genome are extremely simple, consisting merely of TATAAATA (or, in a few cases, variants of TATAAATA), placed approximately 10 bp upstream of a transcriptional start site (Table 1). The T4 gene 55 protein (gp55), a \( \sigma \)-family protein, confers the ability to recognize these promoters upon the RNA polymerase core (E), and E.gp55 holoenzyme accurately initiates late transcription in vitro in negatively supercoiled DNA. More complex requirements for T4 late transcription prevail in vivo, where viral DNA is not, on average, supercoiled. The RNA polymerase-binding gene 33 protein (gp33; 112 amino acids) is required, and so is concurrent DNA replication (reviewed with references in [22]).
3.2. Effects of the DNA polymerase accessory proteins on transcription

Recent biochemical analysis shows how three of the phage T4 replication proteins activate T4 late transcription, and leads to a plausible mechanism for coupling late gene expression to DNA replication. The three accessory proteins of the T4 DNA polymerase, gp45, gp44 and gp62, which confer processivity on replicative DNA chain elongation, activate transcription by a mechanism that has not been previously described. Transcriptional activation requires the ATPase activity of these proteins and a site, such as a primer-template junction or a simple nick, at which they can assemble on DNA. This assembly site can be located at varying distances from the promoter and accordingly has some of the properties of an enhancer, but two properties set it apart from conventional enhancers. First, the assembly site/enhancer and the late promoter must be connected by a continuous and unobstructed path along DNA. Second, there is a polarity constraint for

Fig. 1. A model of transcriptional activation by a phage T4 DNA polymerase-associated DNA-tracking protein. Top line: the DNA nick serves as the binding site for the assembly factor (gp44/62 complex, drawn in black and white) and as the loading site for the DNA-tracking and transcription-activating gp45 trimer (drawn as a ring around DNA). Middle line: The orientation of gp45 loading (red lateral face to the right) is determined by the orientation of the gp44/62 complex at the nick. Bottom line: ATP hydrolysis frees gp45 to track freely along DNA. The orientation of gp45 that is shown is compatible with transcription-activating RNA polymerase interactions only at promoter Y. The representation of the gp45-RNA polymerase interaction at P_Y reflects information referred to in the text, shows the core enzyme (blue) with a long cleft [39], and represents the α subunit as two domains, one capable of DNA-binding, connected by a proteolysis-prone linker [31,40]; only one α subunit is shown. Gp55 (purple) is shown interacting with gp45 [30]; the suggested gp33 (yellow-orange) interaction with gp45 is speculative.
transcriptional activation (Fig. 1). For example, a DNA nick serving as the enhancer must be in the non-transcribed strand of its target transcription unit. These properties of the enhancer suggest that it is an entry site for at least one protein and that transcriptional enhancement requires protein to track along DNA from the entry site to the promoter [23–25].

Further analysis distinguishes two functions of the DNA polymerase accessory proteins: gp45 is the actual activator of transcription, and tracks from its DNA-loading site to the promoter; the gp44/62 complex is the assembly factor that loads gp45 onto DNA, but has no further essential role in transcriptional activation. A plausible model of the mode of action of the assembly factor, based on extensive analysis of T4 DNA replication (summarized in [26,27]) proposes that ATP binding to gp44 promotes the attachment of the gp44/62 complex and gp45 to the enhancer, and ATP hydrolysis releases gp45 from this assembly for tracking along DNA. The polarity of DNA strands determines the orientation of the asymmetric gp44/62–45 complex and imposes a particular orientation on gp45 as it tracks along DNA. The orientation of gp45 in turn determines a compatible orientation for interacting RNA polymerase and, through that interaction, determines the polarity of the promoter that can be transcriptionally activated (Fig. 1, top and middle lines) ([28] and G.M.S., unpublished observations).

Support for the proposal that gp45 is the primary transcriptional activator comes from the observation that participation of the gp44/62 complex is not required for activation of T4 late transcription under conditions of macromolecular crowding (created by adding polyethylene glycol to the reaction medium). Transcriptional activation in such conditions is not ATP hydrolysis-dependent, does not require a special DNA entry site such as a nick or a primer-template junction, and is not restricted with respect to polarity. This supports the conclusion that the gp44/62 complex imposes polarity through its mechanism for loading gp45 onto DNA [29].

3.3. The co-activator and the constitution of open promoter complexes

Transcriptional activation by gp45 requires one additional partner, gp33, which serves as a co-activator. Gp33 also depresses unenhanced basal transcription by E.gp55 alone. In combination, these two properties of this small (112 amino acid) RNA polymerase-binding protein enforce the dependence of T4 late transcription on activation by replication proteins [24].

Photocrosslinking analysis of enhanced, open transcription initiation complexes at T4 late promoters locates gp45 and gp33 in close vicinity to DNA and to each other at the upstream end of the enhanced promoter complex. One might therefore imagine that gp45, the transcriptional activator, and gp33, the co-activator, interact directly (Fig. 1, bottom line), but experimental evidence for such an interaction has not yet been obtained. On the other hand, genetic and biochemical evidence does point to a direct gp45–gp55 interaction ([30] and unpublished observations by R.L.T. and J. Winkelman). It is, of course, possible that gp33 might exercise its co-activator function indirectly, for example, by changing the structure of the E.gp55 holoenzyme in such a way that the gp45–gp55 interaction generates promoter opening. Further experiments identifying the sites of protein–protein interaction in transcriptional activation will certainly be required to understand co-activator function. It is also useful to bear in mind that no reaction pathway to promoter opening has yet been completely delineated, even for the case of the most-studied transcriptional activators, such as CRP [31,32].

3.4. Gp45 and its cellular homologues track along DNA

The T4 DNA polymerase accessory proteins have functional homologues in the DNA replication machineries of prokaryotic and eukaryotic cells. The 5-subunit γ complex of E. coli DNA polymerase III holoenzyme and the human 5-subunit replication factor RF-C are homologues of gp44/62 complex (reviewed in [33]). The homologues of gp45 are the dimeric β subassembly of E. coli DNA polymerase III holoenzyme and the eukaryotic PCNA (proliferating cell nuclear antigen). E. coli β2 has been shown to track along DNA and the γ complex has been identified as its assembly factor, loading β2 onto DNA in an ATP hydrolysis-requiring process [34].
How \( \beta_2 \) tracks along DNA is vividly suggested by its structure, a ring with pseudo 6-fold symmetry contributed by the three structurally similar domains of nearly equal size of each \( \beta \) monomer [35]. Presumably, loading \( \beta_2 \) onto DNA requires ring opening (or some equivalent process) that is facilitated by an ATP hydrolysis-dependent structural transition of the \( \gamma \) complex.

The inner diameter of the dimeric \( \beta \) ring is approx. 35 Å, ample for sliding along a DNA double helix. The \( \beta_2 \) homologues, gp45 and PCNA, are smaller and trimeric rather than dimeric. Indeed, PCNA trimer has now been shown to be a ring, like \( \beta \), with a pseudo 6-fold symmetry contributed by the two structurally similar domains of each PCNA monomer, but with a slightly smaller central hole [36]. It has been suggested that gp45 trimer will turn out to have a similar structure [33]. One readily sees how these proteins slide along DNA and why forming the protein catenane on DNA might require a chaperone-like function. Little is known about the stability of these catenanes, beyond the fact that the DNA-threaded form of \( \beta \) is stable for minutes at 5\(^\circ\) C [34], but that the DNA-tracking state of gp45, and of \( \beta \), is considerably less stable at 37\(^\circ\) C, so that the maintenance of traffic of these sliding proteins on DNA requires continuous ATP hydrolysis-dependent (re)loading [37].

It is simple to see how catenated \( \beta_2 \) would confer processivity on DNA polymerase, tethering it to DNA by direct protein–protein interactions without constraining it to any fixed site. In fact, DNA chain elongation by the tethered polymerase may not be, in chemical kinetic terms, strictly processive; the enzyme may merely be limited to dissociation from its primer-template junction into a one-dimensional rather than three-dimensional space, and therefore able to return to that substrate very rapidly (and, incidentally, able to elude the experimenter’s standard criterion of processivity, which consists of testing for insensitivity to dilution; see also [38]).

### 3.5. Coupling DNA replication to late transcription

The just-summarized experimental evidence accounts for the involvement of replication proteins in positive control of transcription, but it fails to explain why replication is required, and does not account for the normally striking influence of concurrent replication on transcription. Concurrent rolling circle replication of nicked DNA in vitro only increases transcriptional activation modestly, probably due to conversion of DNA nicks into more efficient DNA-loading sites for gp45 (G.M.S. unpublished observations). Rather, we guess that discontinuous DNA chain elongation on the lagging strand of the replication fork continuously generates junctions of Okazaki fragments that serve as transient enhancers of T4 late transcription (transient because they are continuously joined together). The supply of these enhancers is augmented by concurrent and extremely active recombination, which generates structures containing primer-template junctions. In the space between two divergent replication forks, these transient enhancers generate both polarities of gp45 loading capable of activating distal promoters of either polarity. Coupling to concurrent replication must also be attributed to the instability of the presumably catenated tracking state of gp45. Thus, when DNA replication is interrupted, the remaining discontinuous DNA junctions are largely and rapidly converted into continuous strands, depleting the supply of enhancers, and lowering the rate of gp45 assembly onto DNA. As the density of gp45 tracking along DNA declines, so does the rate of gp45-dependent activation of transcriptional initiation.

Implicit in the preceding hypothetical mechanism is the notion that the ‘classical’ phenotypes of T4 late transcription – replication-dependence and replication-coupling – are not intrinsic and inexorable markers of transcriptional activation by DNA-tracking proteins, but are generated by quantitative particulars: coupling arises because both the enhancer and the postulated protein catenane have a limited lifetime. Thus, in thinking about how one might recognize other instances in which similar mechanisms are at work, it is important to consider that the catenanes of different DNA-tracking proteins differ in stability, and that ligands may further affect the stabilities of such catenanes.

### 4. Two concluding comments

AsiA and gp33 are not discernibly related in regard to amino acid sequence, yet they are quite
similar in their effects on transcription, functioning as transcriptional co-activators and, at the same time, as depressors of unactivated (middle) or unenhanced (late) transcription. The two proteins may generate their effects in very similar ways, perhaps by pushing out of the way some domain of the holoenzyme that participates in basal transcription (e.g. the C-terminal domain of α [31]). By whatever mechanism, each protein enforces the requirement of its respective promoters for its function as a transcriptional activator. We would not be surprised if that proved to be a more general property of eukaryotic transcriptional co-activators than is currently realized.

Although catenated DNA-tethering proteins are important for generating processive DNA chain elongation, the intrinsic processivity of the RNA polymerases makes a comparable role in transcription gratuitous. Instead, gp45 works on transcriptional initiation. Tethering RNA polymerase to DNA might make DNA scanning for promoter finding more processive, and should also increase the affinity of RNA polymerase for the closed promoter. If DNA-tethering could substitute for DNA-binding in facilitating promoter opening, that would further imply that contorting DNA so as to expose the template strand at the transcriptional start site does not absolutely require a fixed grip on the upstream end of the promoter complex.

References


