The EMBO Journal vol.3 no.12 pp.2863–2871, 1984

Regulation of a new bacteriophage T4 gene, *69*, that spans an origin of DNA replication

Paul M.Macdonald[1] and Gisela Mosig

Department of Molecular Biology, Vanderbilt University, Nashville, TN 37235, USA
[1]Present address: Department of Biochemistry and Molecular Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA 02138, USA

Communicated by P.Starlinger

We have determined the DNA sequence and transcription patterns in a 3-kb segment (between 15 and 18 kb on the standard phage T4 map) spanning an origin of DNA replication. A new gene, *69*, spans this origin. Gene *69* codes for two overlapping proteins that share a common C-terminal segment. Defective DNA replication in an appropriate amber mutant shows that at least the larger of the two proteins is required for efficient T4 DNA replication. The two proteins coded by gene *69* are expressed from different transcripts that are under different regulation. The smaller protein, gp*69, can be expressed immediately from an *Escherichia coli*-like promoter, whereas expression of the larger protein, gp*69*, must be delayed since its middle promoter requires T4 coded proteins, most likely gp *mot*, for activation. We discuss the possible significance of two overlapping proteins in the assembly of replisomes. Gene *69* is bracketed by the non-essential early gene *dam* (DNA adenine methylase) and the late gene *soc* (small outer capsid protein). Transcripts through this region are interdigitated in a complex pattern, which reveals all elements that are thought to be important in regulation of pre-replicative and post-replicative T4 genes.**

Key words: phage T4/gene *69*/DNA replication/transcription/translation

Introduction

Bacteriophage T4 DNA replication can be initiated at several different origins (for reviews, see Kozinski, 1983; Mosig, 1983a). One of these, *ori*A, in the region between 15 and 18 kb on the T4 map has been characterized in considerable detail (Mosig *et al.*, 1981; Macdonald *et al.*, 1983; Mosig, 1983a; Macdonald and Mosig, in preparation). Here we present the DNA sequence as well as the transcription pattern of a 3-kb segment of the T4 chromosome that contains this origin. We found a new DNA replication gene, *69*, spanning this origin. This gene codes for two overlapping proteins that share a common C-terminal segment.

Gene *69* is bracketed by the non-essential early gene *dam* (DNA adenine methylase) and the late gene *soc* (small outer capsid protein). The sequence presented here also contains the promoter distal part of the early gene *58-61* (DNA priming protein) upstream from *dam*, and an early promoter for an unidentified gene, ORF1, downstream from *soc*. Although the coding regions of these genes are clearly separated, the transcripts are interdigitated in a complex pattern.

This complex transcription pattern reveals all elements which are thought to be important in regulation of pre-replicative T4 genes (Pulitzer *et al.*, 1979; Linder and Sköld, 1980; Brody *et al.*, 1983). According to current models, immediate early (IE) genes are transcribed by *Escherichia coli* RNA polymerase without requiring prior expression of other T4 genes. They map proximal to early promoters, which resemble *E. coli* promoters. Delayed early (DE) genes can be transcribed in either or both of two ways. (i) They can be transcribed from early promoters that map distal to potential transcription termination sites. In that case, their expression requires prior expression of T4 coded anti-termination function(s). (ii) They can be transcribed from middle promoters which depend on prior expression of the T4 *mot* gene (Mattson, 1974). These middle promoters have good homology to the '−10' region, but no homology to the '−35' region of *E. coli* promoters. Instead, they share a consensus sequence, a 'mot box' which is thought to bind the positive activator gp *mot* (Uzan *et al.*, 1983).

Our results suggest that the two proteins coded by gene *69* are expressed from different transcripts that are under different regulation. The small 69* protein can be expressed immediately (IE) whereas the large 69 protein is DE and requires gp *mot* for efficient expression.

Results

DNA sequence

A replication origin is located within the *Eco*RI fragment spanning the *Xba*I site at 15.8 kb on the T4 physical map (Mosig *et al.*, 1981; Macdonald *et al.*, 1983). This *Eco*RI fragment is cloned in pMAC12 (Figure 1), and neighboring T4 restriction fragments covering most of the DNA segment between positions 8 and 18 kb have also been cloned (Macdonald and Mosig, 1984). To obtain the origin DNA sequence and to ensure that nearby genes, promoters and terminators potentially involved in initiation could be accurately mapped, we sequenced the T4 DNA segment from 15.0 to 18.1 kb using DNA from the cloned restriction fragments shown in Figure 1.

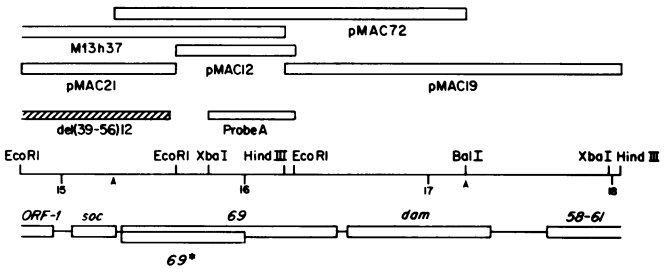
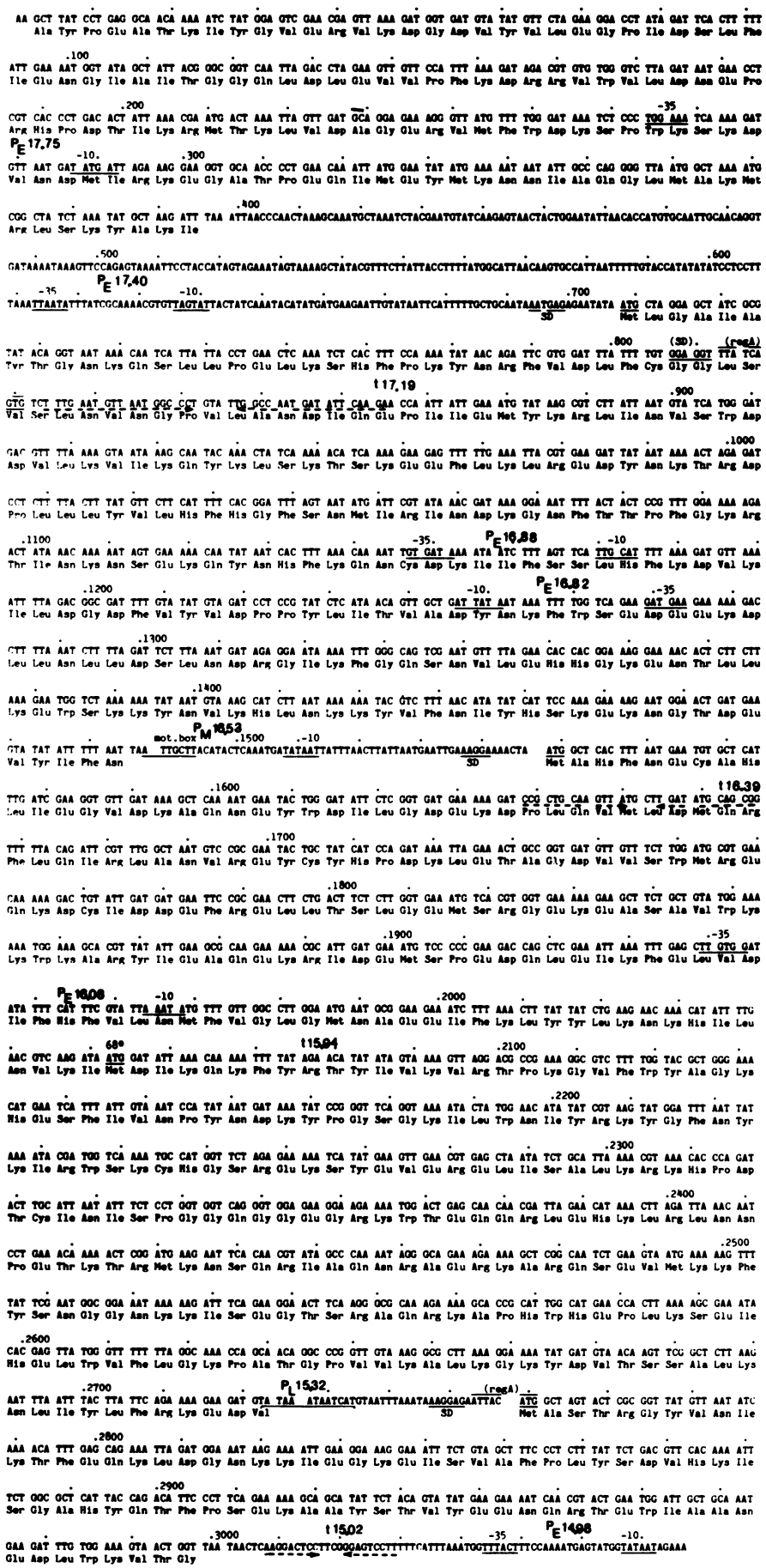


Fig. 1. Map of T4 DNA segment including *ori*A and the gene *69*. Map co-ordinates represent the distance in kb clockwise from the rIIA/rIIB junction on the standard T4 map (Kutter and Rüger, 1983). Cloned T4 DNA segments (open boxes) are shown above the map. The striated box indicates one end of the deletion del(39−56)12 (Homyk and Weil, 1974). Evidence leading to the gene assignments shown below the map is presented in the text. Probe A was used in the Northern blot analysis of RNA synthesized *in vivo* (Figure 5).

a



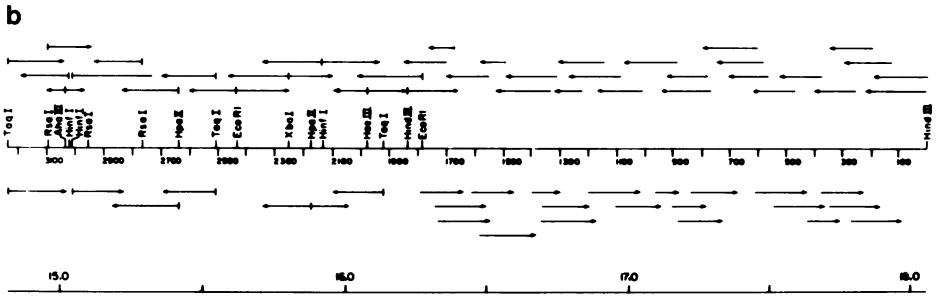


Fig. 2. **(a)** DNA sequence of a 3.1-kb T4 segment which contains genes *dam*, *69* and *soc*. The sequence is presented 5′ to 3′, beginning at the *Hind*III site at map co-ordinate 18.1 (see Figure 1). Below the DNA sequence are the predicted amino acid sequences of the genes specifically identified in Figure 1. Where appropriate, initiation codons and Shine-Dalgarno sequences are underlined, the latter indicated by SD. Potential RegA protein target sequences (Wiberg and Karam, 1983) at the 5′ end of *soc* and the second possible *dam* ribosome binding site are overlined and indicated by (regA). Gene locations are: *58-61* (identified by B.Alberts and M.Nakanishi, personal communication) from upstream of the sequenced region to nucleotide 395; *dam*, from 708 or 819 to 1484; *69*, from 1548 to 2723; *69**, from 2052 to 2723; *soc*, from 2759 to 2998. Probable −10 and −35 regions for promoters mapped as described in the text are underlined and identified as such. Terminators (t) are also marked. Large palindromes that may be involved in termination of transcription are underlined (dashed). In addition, the 'mot box' consensus sequence of $P_M16.53$ is underlined and identified. **(b)** Sequence strategy. End-labeled restriction fragments sequenced by the Maxam-Gilbert technique are shown with vertical bars at the site of end labeling. Fragments that were labeled at the 5′ end are shown above the restriction map. 3′ end-labeled fragments are shown below the map. Segments sequenced by the dideoxy technique are shown above or below the map in accordance with the orientation of the sequence.

The 1.8-kb *Hind*III T4 DNA insert from pMAC19 was sequenced by the chain termination method. The other segment of the sequence was determined by the partial chemical degradation method, using restriction fragments from the cloned segments shown in Figure 1. The individual DNA segments, whose sequences were determined, are shown in Figure 2B. The total DNA sequence is presented in Figure 2A, together with important transcription and translation signals, which are discussed below. Note that this sequence is presented in the direction of transcription of early genes, i.e., from co-ordinate 18.1 towards 15.0 (Figure 1) on the T4 map (Kutter and Rüger, 1983; Gram *et al.*, 1984).

Genes

Inspection of the DNA sequence, gene fusion experiments, and subsequent *in vivo* tests (see below) identified several significant open reading frames (ORFs) and genes which are summarized in Figure 1. The predicted amino acid sequences of the gene products are shown in Figure 2.

We focus first on the ORF that spans the replication origin (which maps in the *Eco*RI fragment cloned in pMAC12). From marker rescue experiments (one set of data is summarized in Table I) and the position of the cloned segments shown in Figure 1, we conclude that *am*HL627-2 is a mutation in a new gene *69*. HL627-2 has a more severe DNA-delay phenotype than the gene *58-61* mutant E219-10 (Figure 3). The first ATG codon (at position 1548) in gene *69* is preceded by a Shine-Dalgarno sequence at an appropriate distance. The DNA sequence predicts that gp*69*, initiated at position 1548, is a basic protein with a mol. wt. of 46 kd.

There is a second functional initiation site in gene *69* at position 2052 (see below). The putative protein initiated at this site, gp*69**, is read in the same frame as gene *69*, but has a predicted size of only 224 amino acid residues. In contrast to gp*69* which is mainly expressed from a middle promoter, gp*69** is potentially expressed as an early protein, since its ribosome binding site maps directly downstream from an early promoter, $P_E16.08$ that exists within the complete gene *69* (see below).

The functioning of an internal translation initiation site within gene *69*, to produce gp*69**, was shown in the following manner. We inserted the internal *Eco*RI fragment from gene *69* into M13mp7 (to construct M13r1 and M13r21) (Figure 4). This T4 *Eco*RI fragment contains the promoter $P_E16.08$ within gene *69*, and ~470 bp of downstream sequence, but lacks the 5′ end of gene *69*. In the desired orientation (M13r1) this construct creates an in-phase fusion between the gene *69** reading frame (which is also the gene *69* reading frame) and the lacZ′ reading frame at the lac promoter-distal *Eco*RI site of M13mp7. M13r1 forms blue plaques when plated on JM103 in the presence of IPTG and XGal, indicating that translation is initiated from a ribosome binding site within the inserted DNA and that the resulting fusion protein has lacZ′ activity. To define further the location of the translation initiation site, two derivatives of M13r1 were constructed. M13r1X° has 4 bp inserted at the single *Xba*I site of M13r1, and M13r1H° has 4 bp inserted at the single *Hind*III site of M13r1 (arrowheads in Figure 4). Both create +1 frameshift mutations and should eliminate β-galactosidase expression if translation responsible for the blue plaque phenotype is initiated upstream from the two different 4 bp insertion sites. M13r1H° retains β-galactosidase activity, whereas M13r1X° does not. Therefore, the translation initiation site must be located between the *Xba*I and *Hind*III sites. Of five potential initiator codons in that segment (four ATG and one GTG), the ATG starting at nucleotide 2050 is defined by the computer program of Stormo *et al.* (1982) to be the most likely candidate.

Gene *69* is flanked by two non-essential genes, *dam* and *soc*. The *dam* gene, upstream from gene *69*, codes for a DNA adenine methylase. Schlagman and Hattman (1983) have reported the presence of *dam* in the 1.8-kb *Hind*III fragment. They also demonstrated that the gene is inactivated by cleavage at the single *Bal*I site in this restriction fragment. Thus, this *Bal*I site at position 847 identifies the overlapping ORF as the *dam* gene. Two potential translation initiation sites at positions 708 and 819 would yield proteins of 30 or 26 kd, respectively. A protein in this size range has been detected by maxicell experiments using a plasmid containing the 1.8-kb *Hind*III fragment (Hattman *et al.*, personal communication). Wiberg and Karam (1983) have stated that gp *dam* of T4 is under control of the T4 *reg*A protein, a translational repressor. A *reg*A consensus binding sequence (Karam *et al.*, 1982) at position 813−821 overlaps the second, downstream

Table I. Proportions of wild-type particles among the progeny of the second-cycle marker-rescue experiments described in Materials and methods

T4		Host:	UT481				MHI	
Gene	Mutant	Plasmid:	None	pMAC21	pMAC12	pMAC19	None	pMAC72
ORF1	E219-12		3×10^{-5}	3.2×10^{-1}	3.1×10^{-4}	2×10^{-5}	9.3×10^{-4}	2.5×10^{-4}
69	HL627-2		3.2×10^{-5}	1.5×10^{-2}	2.5×10^{-5}	2.9×10^{-1}	6.3×10^{-4}	3.2×10^{-2}
58-61	E219-10		1.6×10^{-5}	6.8×10^{-4}	1.3×10^{-4}	2.6×10^{-1}	4.8×10^{-3}	1.5×10^{-4}
56	E56		0	0	0	5×10^{-1}	0	6×10^{-1}
17	N56		N.T.[a]	5.7×10^{-4}	2.9×10^{-5}	6.9×10^{-3}	N.T.	N.T.

[a]N.T. = not tested.

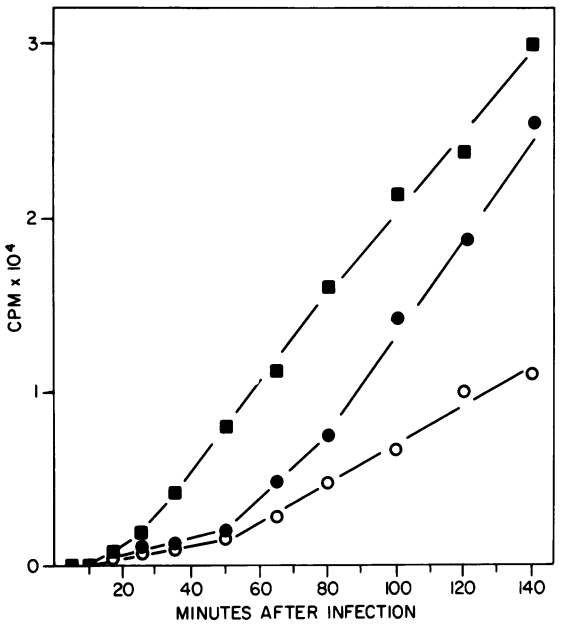


Fig. 3. Net T4 DNA synthesis, measured by continuous [^{3}H]thymidine incorporation in *E. coli* B at 25°C. The T4 strains used were: wt (■), E219-10, gene *58-61* (●), HL627-2, gene *69* (○).

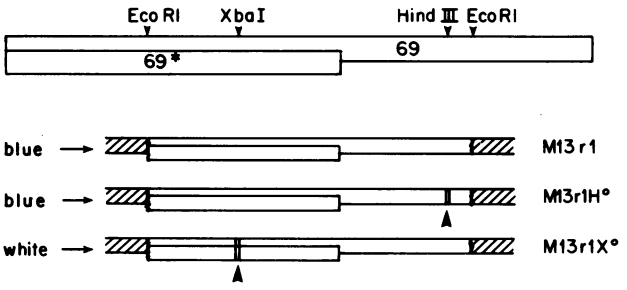


Fig. 4. Constructions that demonstrate an internal, in-phase initiation codon within gene *69*. Key restriction sites in gene *69* are marked in the upper line. M13r1 was constructed by inserting the internal gene *69 Eco*RI fragment into the large *Eco*RI fragment of M13mp7, creating an in-frame fusion between the gene *69* and *lacZ* (striated) coding sequences at the 3′ proximal end of the gene *69* segment. The other fusion site is not in frame. Paired vertical lines indicate the sites of 4-bp insertions, which create +1 frameshift mutations, in M13r1H° and M13r1X°.

putative initiation site.

The structure and regulation of *soc* (small outer capsid protein) have been described (Macdonald *et al.*, 1984). A *reg*A binding sequence also overlaps the ribosome binding site of this late gene, although *reg*A protein is thought to regulate expression of pre-replicative early genes.

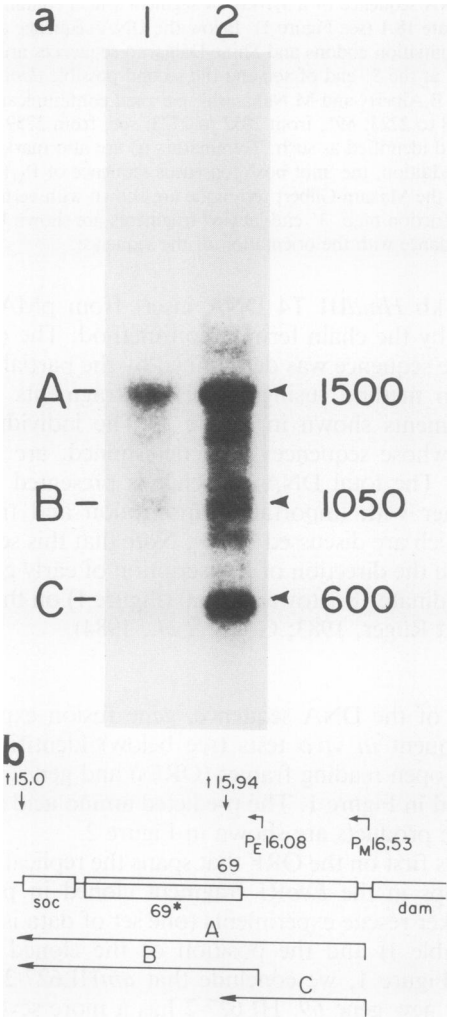


Fig. 5. Northern blot analysis of gene *69* transcripts. **(a)** Autoradiograms in **lanes 1** and **2** are from Northern blots of two different preparations of T4 RNA isolated 8 min after infection at 30°C. In each case the probe was ^{32}P-labeled *Eco*RI-*Xba*I restriction fragment shown in Figure 1 as probe A. Transcript lengths were estimated by comparison with single-stranded DNA fragments of known sizes that were electrophoresed in adjacent lanes of the same gels. **(b)** The deduced transcriptional map. In addition to the three transcripts marked, A, B and C, several minor RNA species can be seen in **(a)**. These are probably initiated at early promoters further upstream (see Figure 6) and terminated at either t15.94 or t15.0.

Transcription through gene 69

We have mapped the transcripts that traverse gene *69* and their promoters and terminators by (i) Northern blot analysis of *in vivo* RNA and (ii) *in vitro* run-off transcription from purified templates. Here we emphasize transcripts that are important for expression of gene *69*. Origin promoters which are thought to be involved in initiating DNA replication at

*ori*A will be discussed in a subsequent paper.

In vivo, three predominant gene *69* transcripts, marked as A, B and C in Figure 5 (detected by probe A, see Figure 1), are initiated from two distinct promoters, P_M16.53 and P_E16.08. All three transcripts are oriented in the 'early' direction, since a probe labeled only in the anti-'early' strand, but otherwise identical to the probe A (Figure 1) used in Figure 5, gave an identical pattern after hybridization (Macdonald, 1983). Both the 1050 and 1550 nucleotide transcripts (A and B) end at a previously identified terminator, t15.0, downstream from *soc*; probes upstream from the terminator detect both transcripts (Macdonald *et al.*, 1984) while probes downstream from the terminator do not (Macdonald, 1983; and data not shown). Therefore the terminator-proximal (early) promoter, (P_E16.08) at position 1939–1970, maps at co-ordinate 16.08 within gene *69* but upstream from the gene *69** initiation site (at position 2042), while the terminator-distal middle promoter, P_M16.53, with a 'mot box' (Brody *et al.*, 1983) at position 1487–1493, maps at co-ordinate 16.53 just upstream from gene *69*. The short 600 nucleotide transcript is most likely initiated at P_M16.53 and terminated 600 bp downstream exactly within an *in vitro* termination site, t15.94 (Macdonald, 1983). In contrast to the longer transcripts, the 600 nucleotide transcript does not hybridize to any probe from the region downstream from probe A. Therefore its promoter must be located at least 600 bp upstream from the *Xba*I site of probe A and no more than 600 bp upstream from the *Eco*RI site of probe A. Only the promoter responsible for synthesis of the 1550 nucleotide transcript, P_M16.53, is found within this segment *in vitro* (see below).

In vitro 'run off' transcripts (Figure 6) were used to map more precisely the promoters deduced from analysis of *in vivo* RNAs, including those that are probably responsible for the longer minor transcripts seen in Figure 5. Since the *in vitro* transcription reaction mixture did not include T4 Mot protein (Uzan *et al.*, 1983), or other potential T4 coded regulator proteins (Rabussay, 1983), early promoters should be active but transcription from T4 middle promoters must be greatly reduced, and potential termination-anti-termination must be different from that *in vivo* (Brody *et al.*, 1983).

Five master templates, derived from the 1.8-kb *Hind*III fragment of pMAC19, were transcribed directly or after digestion with additional restriction enzymes (Figure 6a). The promoters identified by the individual 'run-off' transcripts are also shown in Figure 6a. Nucleotide positions used in describing templates, promoters and transcripts are the same as in Figure 2, the numbers following the promoters refer to the co-ordinates on the T4 map. To map precisely P_E16.08, additional templates were derived from pMAC12 or from *Bal*31-derived deletion mutants of pMAC12.

P_E16.08

Mapping of the 1050 nulceotide transcript shown in Figure 5 indicated that it was initiated at or near position 1980 in the nucleotide sequence, i.e., near map co-ordinate 16.08. *In vitro* transcription products from templates A (~142 nucleotides, Figure 6B, lane 2), B (~232 nucleotides, lane 3), and C (~220 nucleotides, lane 1) confirm the map location. The high level of *in vitro* transcription from this promoter identifies it as an early promoter (P_E16.08).

P_M16.53

Close examination of lanes 4–6 in Figure 6B reveals minor transcripts of 320, 260 and 213 nucleotides (marked with arrowheads) from templates D (850–1834) (lane 4), E (850–1782) (lane 5) and *Hpa*II cleaved D (lane 6), respectively. These transcripts must initiate at or near position 1520, the estimated location of P_M16.53. Even though transcripts initiated from P_M16.53 were the predominant gene *69* transcripts *in vivo* (Figure 5), P_M16.53 appeared extremely weak *in vitro*, as expected, because gp *mot* was not present in the transcription reaction mixture. The consensus 'mot box' sequence, ATTGCTT (Brody *et al.*, 1983), is present near the −35 position (which does not resemble an *E. coli* −35 consensus sequence) of P_M16.53. As in several other putative T4 middle promoters (Brody *et al.*, 1983), the −10 region perfectly matches the *E. coli* consensus sequence. These results, taken together, identify P_M16.53 as a middle promoter.

P_E17.75

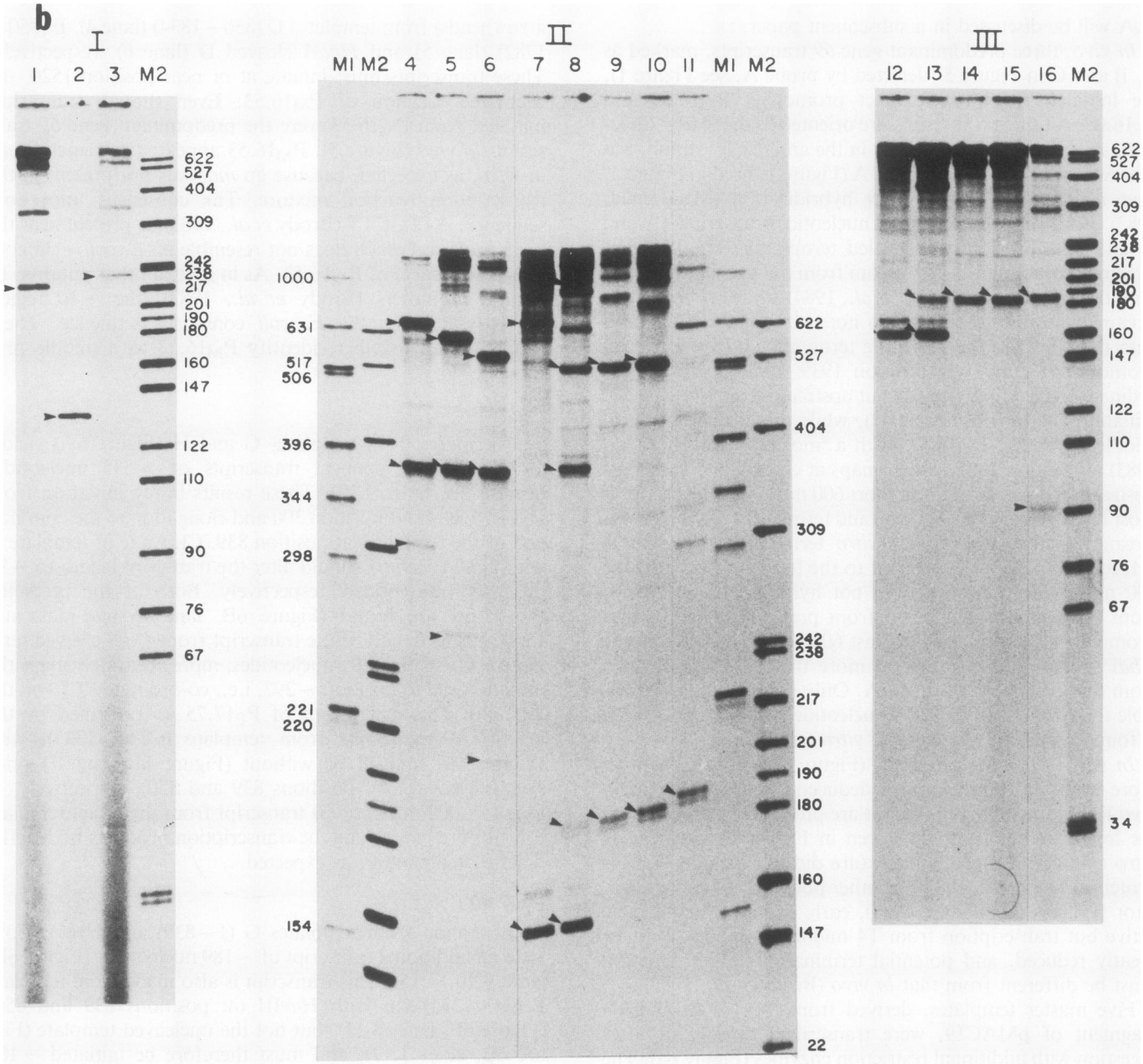
Transcription from templates G and H (Figure 6A) yields, among others, identical transcripts of ~535 nucleotides (Figure 6B, lanes 9,10). These results imply initiation from P_E17.75 at about position 300 and elongation to the common end of the templates at position 839. Cleavage of template G with *Dde*I or *Hinf*I should alter the transcript length to ~90 or ~485 nucleotides, respectively. Both of the predicted transcripts are found (Figure 6B, lane 16 and data not shown). The length of the transcript from *Dde*I cleaved template G (1–839), ~91 nucleotides, more accurately maps the initiation site to position ~297, i.e., co-ordinate 17.75 on the T4 map. This assignment of P_E17.75 is confirmed by the lengths of transcripts from template F (34–1335), with (Figure 6B, lane 8) or without (Figure 6B, lane 7) prior *Hae*III cleavage (at positions 839 and 850), respectively. A long (~1000 nucleotides) transcript from the intact template F is not present among the transcription products of *Hae*III-digested template F, as expected.

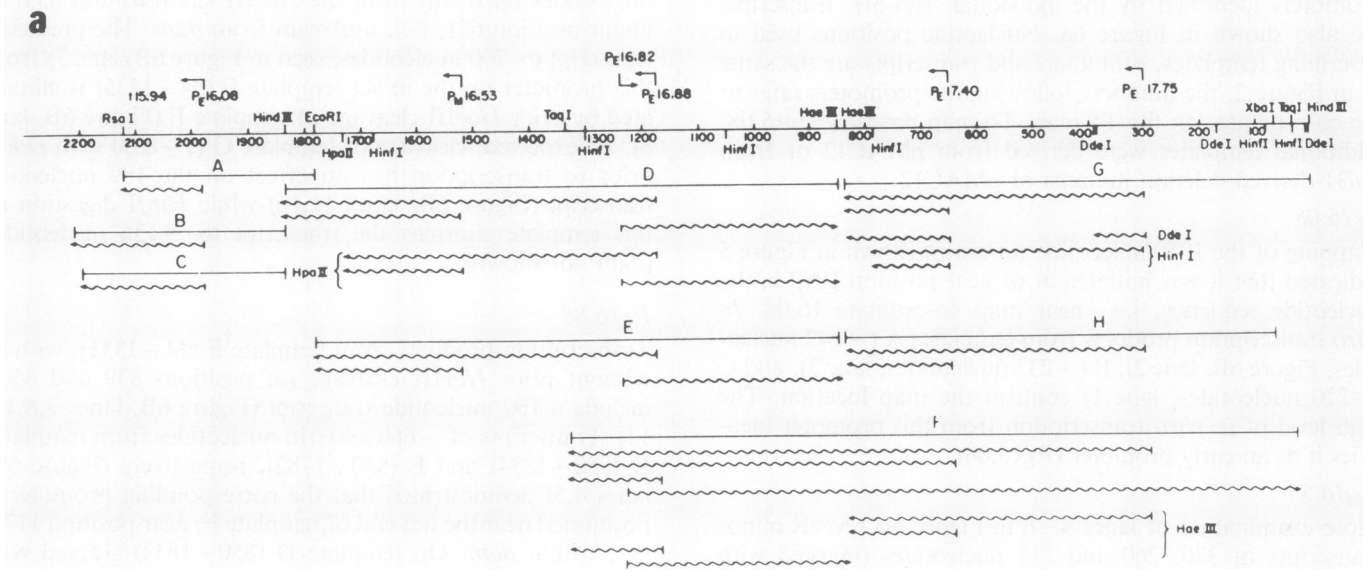
P_E17.40

Transcription from templates G (1–839) and H (65–839) yield an additional transcript of ~189 nucleotides (Figure 6B, lanes 9,10,14,15). This transcript is also made from template F (34–1335) cut with *Hae*III (at positions 839 and 850) (Figure 6B, lanes 8,13), but not the uncleaved template (Figure 6B, lanes 7,12), and must therefore be initiated ~189 nucleotides upstream from the *Hae*III site (position 839) at about position 651, i.e., upstream from *dam*. The predicted transcript (~700 nucleotides, seen in Figure 6B, lane 7) from this promoter on the intact template F (34–1335) is eliminated by prior *Hae*III cleavage of template F (Figure 6B, lane 8). As expected, cleavage of template G (1–839) with *Dde*I prior to transcription has no effect on the 189 nucleotide transcript (Figure 6B, lanes 11,16) while *Hinf*I digestion of this template shortens the transcript to ~136 nucleotides (data not shown).

P_E16.88

Transcription products from template F (34–1335), with or without prior *Hae*III cleavage (at positions 839 and 850), include a 160 nucleotide transcript (Figure 6B, lanes 7,8,12, 13). Transcripts of ~660 and 610 nucleotides from templates D (850–1834) and E (850–1782), respectively (Figure 6B, lanes 4,5) demonstrated that the corresponding promoter is positioned near the left end of template F, near position 1176, i.e., within *dam*. On template D (850–1834) cleaved with


b
I
1 2 3 M2
622
527
404
309
242
238
217
201
190
180
160
147
122
110
90
76
67
II
M1 M2 4 5 6 7 8 9 10 11 M1 M2
1000
631
517
506
396
344
298
221
220
154
622
527
404
309
242
238
217
201
190
180
160
147
22
III
12 13 14 15 16 M2
622
527
404
309
242
238
217
201
190
180
160
147
122
110
90
76
67
34


a
115.94
PE16.08
PM16.53
PE16.82
PE16.88
PE17.40
PE17.75
Rsa I
Hind III
EcoRI
Taq I
Hae III
Hae III
Xba I
Taq I
Hind III
HpaII
Hinf I
Hinf I
Hinf I
Hinf I
Dde I
Dde I
Hinf I
Dde I
A
B
C
D
E
F
G
H
HpaII
Dde I
Hinf I
Hae III

*Hpa*II, transcripts from this promoter are shortened to 560 nucleotides, as expected (Figure 6B, lane 6).

P_E14.98

One other strong *in vitro* promoter was mapped downstream from *soc*, within the *Eco*RI fragment cloned in pMAC21. Transcripts from this promoter are initiated at or near position 3075 (P_E14.98). They are probably responsible for expression of ORF1, downstream from *soc* (data not shown).

Four additional weak promoters are located within gene *69*, downstream from the weak terminator t15.95 mentioned above. They appear to function in the initiation of DNA replication and will be discussed in that context (Macdonald and Mosig, in preparation). A late promoter, P_L15.03, which initiates transcription at or near position 2722, responsible for expression of *soc* (Macdonald *et al.*, 1984), was not detected *in vitro* since late promoters require modification of RNA polymerase by T4 gp*55* for proper functioning (Kassavetis *et al.*, 1983; Elliott and Geiduschek, 1984).

P_E16.82

In addition to the transcripts in the 'early' direction, a strong early promoter was detected *in vitro* that initiated transcription in the opposite direction. Templates D (850−1834), E (850−1782) and *Hpa*II digested template D all direct synthesis of a ~388 nucleotide transcript (Figure 6B, lanes 4−6). Their identical size suggests that the transcripts are all initiated at P_E16.82 (position 1238) and extend to the common end of all three templates at position 850. This assignment of P_E16.82 is confirmed by comparison of transcripts from template F without (Figure 6B, lane 7) or with (lane 8) *Hae*III digestion, respectively. A long transcript from template F (better resolved in a shorter exposure of the gel shown in Figure 6B, lane 7) is shortened to a ~388 nucleotide transcript when the template has been cut with *Hae*III (lane 8). Additional confirmation of the promoter assignment and the direction of transcription is provided by the appearance of a predicted 186 nucleotide transcript from *Hinf*I digested template D (850−1834) (data not shown).

Discussion

We have determined the DNA sequence (Figure 2) and transcription pattern (Figures 5,6,7) of a segment of the T4 chromosome containing a new gene, *69*, which overlaps a replication origin, *ori*A. Upstream from gene *69* we find an open reading frame that codes for the *dam* methylase (Schlagman and Hattman, 1983). Overlapping one of the two potential start codons of *dam* is a sequence similar to the proposed consensus sequence for action of gp *reg*A, a T4 regulatory protein that has been reported to control translation of many early genes, including *dam* (Wiberg and Karam, 1983).

A *reg*A consensus sequence (Karam *et al.*, 1982) is also found at the ribosome binding site of the late gene *soc*, which maps downstream from gene *69*. The significance of this site in *soc* is not understood, since gp *reg*A is thought to repress early genes at late times after infection, but the *reg*A binding site is present in the two early transcripts, as well as in the late transcript traversing *soc* (Macdonald *et al.*, 1984). The late transcript and only this transcript is translated into Soc protein, probably because the ribosome binding site for gp *soc* in the long early transcripts originating from P_E16.08 and from P_M16.53 is sequestered in a hairpin (Macdonald *et al.*, 1984).

The genes identified in the DNA sequence would saturate the simple coding capacity for the region between 15 and 18 kb. Gene *56* (dCTPase), which apparently maps in this region (for review, see Mosig, 1983b), has, however, not yet been precisely located. The marker rescue data of the gene-*56 am* mutation E56 (Table I) suggest that at least part of this gene is located in the sequenced region. Although we found no ORF corresponding to the published size and amino acid composition of T4 dCTPase (Price and Warner, 1969; Snustad *et al.*, 1983), it is possible that this gene is split, like the gene for T4 thymidilate synthetase (Chu *et al.*, 1984). It is also possible that the *am* mutation that we re-isolated from E56 was a secondary mutation in the original gene *56* mutant (see Materials and methods). We are testing this possibility.

Gene *69* transcripts are primarily initiated at three different promoters P_E16.88, P_M16.53 and P_E16.08 (Figure 7). Transcripts initiated at P_E16.08 are peculiar in that they can code for gp*69** but not the entire gp*69*. P_E16.88 and P_M16.53 both lie upstream from the start of gene *69* and are responsible for delayed early expression of the complete gene *69*. P_E16.88, together with a putative factor-dependent termination site (t16.39, see Figure 2), and P_M16.53 constitute some of the typical transcriptional control elements of T4 delayed early genes (Brody *et al.*, 1983). Immediate early transcription from P_E16.88 is probably terminated at t16.39 during the early phase of transcription but is allowed to continue at later times, when T4 directed anti-termination occurs (Pulitzer *et al.*, 1979; Stitt *et al.*, 1980). Relatively few gene *69* transcripts are initiated from this promoter (Figure 5). After T4 gp *mot* has been synthesized, the gene *69* middle promoter (P_M16.53) is activated. As seen in Figure 5, the majority of gene *69* transcripts are initiated at this promoter. Few, if any, of the *in vivo* gene *69* transcripts appear to be initiated from the strong *in vitro* promoters at map co-ordinates 19.7, 20.2 and 26.4 described by Gram *et al.* (1984).

Gene *69* protein synthesis can be initiated at two widely separated positions in the same reading frame. Normark *et al.* (1983) have summarized the known genes coding for two overlapping proteins which are translated in the same reading frame. Some examples come from origin specific proteins, e.g., gp*A* of phages φX174 and G4, and gp*II* of phages f1 and M13 (for review, see Kornberg, 1980, 1982). Other examples are morphogenetic genes of several viruses, e.g., gp*C* and gp*Nu3* of phage lambda. Shaw and Murialdo (1980) have suggested that such overlaps could be a general strategy in

Fig. 6. Mapping promoters by *in vitro* run-off transcription. **(a)** A schematic representation of the results shown in **(b)**. Above the restriction map (numbered as in Figure 2) are shown the locations of all promoters mapped here, with one exception (P_E14.98). Each purified template is shown as a straight line, labeled alphabetically. Transcripts [as seen in **(b)**] are shown as wavy lines. Some of these templates were digeted with additional restriction enzymes immediately prior to transcription. Transcripts from such templates are marked at the side to indicate which secondary restriction enzyme was used. The templates A through H extend through the following nucleotide positions (numbered as in Figure 2): A, 1782−2121; B, 1834−2211; C, 1834−2198; D, 850−1834; E, 850−1782; F, 34−1335; G, 1−839; H, 65−839. **(b) Panels I** and **III** are autoradiographs of transcription products separated on 8% acrylamide urea gels, while a 4% acrylamide-urea gel was used for **panel II**. Templates used were: A, **lane 2**; B, **lane 3**; C, **lane 1**; D, **lane 4**; D cut with *Hpa*II, **lane 6**; E, **lane 5**; F, **lanes 7** and **12**; F cut with *Hae*III, **lanes 8** and **13**; G, **lanes 9** and **14**; G cut with *Dde*I, **lanes 11** and **16**; H, **lanes 10** and **15**; Transcription reactions were performed as described in Materials and methods. Identical results were obtained for all transcripts when the reactions included KCl at 200 mM instead of 100 mM. Size markers were ^{32}P end-labeled single-stranded DNA fragments, produced by *Hinf*I and *Eco*RI digestion of pBR322 (M1) or *Hpa*II digestion of pBR322 (M2). The transcription products mentioned in the text are marked with arrowheads.

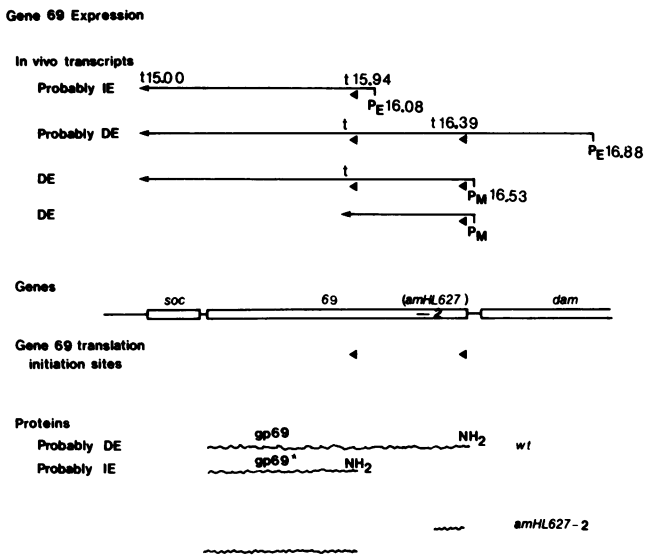


Fig. 7. Summary of transcription and translation through gene *69*, described in the text. The mutation HL627-2 does not eliminate gp*69**. Thus, it is not known whether this protein is essential for T4 growth.

virus assembly to generate a pair of proteins with regions of structural homology which may result in interaction through the shared structural domains, while regions of non-homology would allow separate specific interactions. Smith and Parkinson (1980), using similar arguments, suggested that the two overlapping proteins from the *che*A locus may promote proper assembly of a protein complex involved in chemotaxis. Since origins of DNA replication contain not only starting points of DNA chains, but also assembly sites of replisomes, it is plausible to extend this argument and to suggest that the two proteins from T4 gene *69* may play a similar role in assembly of the T4 replication machine (Alberts, 1984). An additional feature of the two proteins from T4 gene *69* not found in the overlapping proteins compiled by Normark *et al.* (1983) is that the two ribosome binding sites can be found on transcripts originating from different promoters under different control. As mentioned above, the smaller gp69* can potentially be expressed early from a transcript initiated at P_E16.08. The larger gp*69* must be predominantly a delayed early protein.

Initiation of DNA synthesis at many other replication origins requires the protein product of a gene located either adjacent to, or overlapping the origin (for reviews, see Kornberg, 1980, 1982; Sternberg and Hoess, 1983; Scott, 1984). Replicons with more than one replication origin appear to code for separate proteins involved in initiation at any one of the distinct origins, e.g., the E gene of the *E. coli* F plasmid codes for a protein involved in initiation at the adjacent *ori*S, but not at *ori*V (Komai *et al.*, 1982). Gene *69* is defined by a conditional lethal mutation, *am*HL627-2, that results in a deficiency in DNA replication (Figure 3). Like the gene for the origin-binding O-protein of phage lambda, gene *69* spans a replication origin. Furthermore, the lambda O-protein interacts with the host primosome. Genetic evidence suggests that gp*69* also interacts with DNA priming proteins: the gene *69* mutation *am*HL627-2 was present as an apparent second site mutation in a T4 priming protein mutant. In support of this interaction, DNA synthesis in the double mutant was less reduced than in the single gene *69* mutant (data not shown).

It is tempting to speculate that the controlled synthesis of two different proteins from gene *69* plays a role in regulating T4 DNA replication at *ori*A, which is located downstream from the ribosome binding site for gp69* (Macdonald and Mosig, in preparation). Experiments to test this idea are in progress.

Materials and methods

Bacteria and bacteriophage

E. coli B, S/6 (Su^+) and CR63 (SuI) were from our collection. UT481 (SuI), an rK^- host for M13mp9 (Messing and Vieira, 1982), was from C.Lark and MH1 (Su^+), a m^+r^- host was from B.Alberts. T4 mutants *am*E219 (gene *58-61*), *am*HL627 (gene *58-61*), *am*N56 (gene *17*), *am*E56 (gene *56*), *am*N134 (gene *33*) and *am*HL292 (gene *55*) were originally obtained from R.S.Edgar, backcrossed to wild-type and maintained in this laboratory. Double mutants were isolated from the progeny of appropriate crosses. The original *am*E219 and *am*HL627 strains contained multiple mutations. We isolated strains with single *am* mutations which could rescue their wild-type allele from single clones (see Table I). We call these *am*HL627-2 (gene *69*) and *am*E219-10 (gene *58-61*), respectively. Yegian *et al.* (1971) demonstrated that two mutations, isolated from *am*HL627 and *am*E219, do not complement one another and, thus, are in a single gene, *58-61*. Gp 58-61, which together with gp41 constitutes the T4 priming protein (Nossal and Alberts, 1983), is encoded by the gene immediately upstream from the *dam* sequence. The sequence from positions 1 to 399 in Figure 2 determined by us constitutes the 3′ end of gene *58-61* (B.Alberts and M.Nakanishi, personal communication). *am*E219-12 maps most likely in ORF1 downstream from *soc*. del(39-56)12 (Homyk and Weil, 1974) was from J.Weil.

Marker rescue

E. coli (with or without plasmid-bearing T4 inserts) were grown to a titer of $1-3 \times 10^8$/ml, infected with a multiplicity of 3–9 T4 particles per bacterium and incubated for 15 min at 30°C. Infected bacteria (10 μl) were spotted onto plates seeded with *E. coli* S/6. After overnight incubation aliquots from the spots were picked with capillary tubes, lysed with chloroform and assayed for total progeny on *E. coli* CR63 and for am^+ progeny on *E. coli* S/6. Results are expressed as proportions of am^+ progeny. This procedure avoids complications due to leakiness of certain T4 mutants and amplifies the marker rescue by selection for wild-type recombinants and by permitting potential non-viable progency particles that are formed in the infected clones (Mattson *et al.*, 1983), to contribute their rescued alleles. Typical differences between positive results (underlined in Table I) and negative results obtained with the same bacterial culture are between one and two orders of magnitude, although results obtained with different bacterial cultures showed considerable background variation.

[^{3}H]Thymidine incorporation

This was done as described by Mosig and Bock (1976). *E. coli* B were grown at 25°C in M9 medium to a titer of 1×10^8/ml, concentrated by centrifugation to 1×10^9/ml, infected at time 0 with an average of five particles per bacterium, and incubated for different times at 25°C. [methyl-^{3}H]Thymidine ICN, 73.5 Ci/mM was added 3 min after infection to a final concentration of 10 μCi/ml. (Note that both host and phage thymidilate synthetases are functional, thus the intracellular thymidine concentration is unknown.)

Plasmids, DNA preparation and manipulations

Purification of plasmid and M13 RF DNAs, gel electrophoresis, restriction enzyme digestions and other recombinant DNA techniques were done as described in Maniatis *et al.* (1982). pMAC12, with a 653 bp T4 *Eco*RI fragment, and pMAC19, with a 1834 bp T4 *Hind*III fragment (see Figure 1), were constructed by ligating the T4 DNA inserts of M13r1 and M13h1 (Macdonald and Mosig, (1984), respectively, into pUC8 (Messing, 1983). pMAC21 and M13h37 have been described (Macdonald and Mosig, 1984). pMAC72 was constructed in the following manner. A *Sca*I-*Hind*III fragment from M13h37 and a *Hae*III-*Hind*III fragment from pMAC19 were isolated and ligated to one another in the presence of *Bam*HI linkers. After *Bam*HI digestion the DNA was ligated to *Bam*HI-digested pUC12 (Messing, 1983). M13r1X° and M13r1H° were constructed by digesting M13r1 RF DNA with *Xba*I or *Hind*III, respectively, filling the overhanging 5′ ends with *E. coli* polymerase I large fragment, and re-ligating the ends.

DNA sequencing

The 1.8-kb *Hind*III T4 insert of pMAC19 was self-ligated, randomly fragmented by sonication and the fragments were subcloned (Deininger, 1983) into the *Sma*I site of M13mp9. These clones were sequenced by the dideoxy method (Sanger *et al.*, 1977) using AMV reverse transcriptase (Smith, 1980).

All other DNA sequencing was done by the partial chemical cleavage method (Maxam and Gilbert, 1980). DNA sequences were compiled and analyzed using the computer programs of Isono (1982).

Northern blot analysis

RNA preparation, gel electrophoresis, transfer to nitrocellulose, DNA probe labeling and hybridization have been described (Macdonald *et al.*, 1984). RNA shown in Figure 5 was isolated 8 min after infection of *E. coli* B at 30°C with the T4 double mutant, *am*N134-*am*HL292 (genes *33-55*), which prevents late T4 transcription in this host.

In vitro transcription

Template DNAs were purified from agarose gels by binding to DEAE membrane (Schleicher and Schuell). DNA was eluted from the membrane using the protocols provided by the manufacturer. Transcription reactions were done in 20 μl containing 40 mM Tris-Cl (pH 8.0), 100 mM KCl, 10 mM $MgCl_2$, 10 mM β-mercaptoethanol, 250 μM each of ATP, CTP and GTP, 10 μCi [α-^{32}P]UTP (10 Ci/mmol; ICN), 10 nM template DNA and 0.4 units *E. coli* RNA polymerase. The mixture was incubated at 37°C in the absence of nucleotides for 10 min. Then the nucleotides were added along with heparin (final concentration 25 μg/ml) and incubation at 37°C was continued for an additional 15 min. The reaction was terminated by adding 30 μl of 15 mM EDTA. After extraction with an equal volume of phenol:chloroform (1:1), 3 μg of tRNA were added and the nucleic acids were ethanol precipitated. Lyophilized pellets were re-suspended in loading buffer [90% formamide, 100 mM Tris-borate (pH 8.3), 2.5 mM EDTA, 0.25% bromophenol blue and 0.25% xylene cyanol] and electrophoresed in 4% or 8% urea-acrylamide sequencing gels (Maniatis *et al.*, 1982). Results identical to those shown in Figure 6 were obtained when 200 mM KCl was used in the transcription reactions.

Enzymes

Restriction enzymes were purchased from New England Biolabs, Bethesda Research Laboratories or Boehringer Mannheim. T4 DNA ligase and DNA polymerase were purified by Lee Rowen in this laboratory and T4 polynucleotide kinase was provided by Barbara Walker. Bacterial alkaline phosphatase, *E. coli* RNA polymerase, and AMV reverse transcriptase were purchased from Bethesda Research Laboratories, New England Biolabs and Life Sciences, respectively.

Acknowledgements

We thank Lee Rowen and Robert Bird for advice during the initial stages of this work, Gary Stormo for a computer search to find the most likely ribosome binding site for gp69*, Robert Thompson and Gene Lin for discussions and Cindy Young for typing the manuscript. This work was supported by Public Health Service grant GM 13221, Biomedical Research grant RR07201 from the National Institutes of Health, and the Natural Science Fund of Vanderbilt to G.M. P.M. was supported by National Institutes of Health Cellular/Molecular Biology Graduate Training grant T32 GM07319.

References

Alberts,B.M. (1984) *Cold Spring Harbor Symp. Quant. Biol.*, in press.

Brody,E., Rabussay,D. and Hall,D.H. (1983) in Mathews,C.K., Kutter, E.M., Mosig,G. and Berget,P.B. (eds.), *Bacteriophage T4,* American Society for Microbiology, Washington, D.C., pp. 174-183.

Deininger,P.L. (1983) *Anal. Biochem.*, **129**, 216-223.

Chu,F.K., Maley,G.F., Maley,F. and Belfort,M. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 3049-3053.

Elliott,T. and Geiduschek,E.P. (1984) *Cell*, **36**, 211-219.

Gram,H., Liebig,H.-D., Hack,A., Niggemann,E. and Rüger,W. (1984) *Mol. Gen. Genet.*, **194**, 232-240.

Homyk,T., Jr. and Weil,J. (1974) *Virology*, **61**, 505-523.

Isono,K. (1982) *Nucleic Acids Res.*, **10**, 85-89.

Karam,J.D., Dawson,M., Gerald,W., Trojanowska,M. and Alford,C. (1982) in Grunberg-Manago,M. and Safer,B. (eds.), *Interaction of Translational and Transcriptional Controls in the Regulation of Gene Expression,* Elsevier Biomedical, NY, pp. 83-89.

Kassavetis,G.A., Elliott,T., Rabussay,D.P. and Geiduschek,E.P. (1983) *Cell*, **33**, 887-897.

Komai,N., Nishizawa,T., Hayakawa,Y., Murotsu,T. and Matsubara,K. (1982) *Mol. Gen. Genet.*, **186**, 193-203.

Kornberg,A. (1980) *DNA Replication,* published by W.H.Freeman and Co., San Francisco.

Kornberg,A. (1982) *Supplement to DNA Replication,* published by W.H. Freeman and Co., San Francisco.

Kozinski,A.W. (1983) in Mathews,C.K., Kutter,E.M., Mosig,G. and Berget, P.B. (eds.), *Bacteriophage T4,* American Society for Microbiology, Washington, D.C., pp. 111-119.

Kutter,E. and Rüger,W. (1983) in Mathews,C.K., Kutter,E.M., Mosig,G. and Berget,P.B. (eds.), *Bacteriophage T4,* American Society for Microbiology, Washington, D.C., pp. 277-290.

Linder,C.H. and Sköld,O. (1980) *J. Virol.*, **33**, 724-732.

Macdonald,P.M. (1983) Ph.D. Thesis, Vanderbilt University, Nashville, TN.

Macdonald,P.M. and Mosig,G. (1984) *Genetics*, **106**, 1-16.

Macdonald,P.M., Seaby,R.M., Brown,W. and Mosig,G. (1983) in Schlessinger,D. (ed.), *Microbiology – 1983,* American Society for Microbiology, Washington, D.C., pp. 111-116;

Macdonald,P.M., Kutter,E. and Mosig,G. (1984) *Genetics*, **106**, 17-27.

Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning. A Laboratory Manual,* published by Cold Spring Harbor Laboratory Press, NY.

Mattson,T. (1974) *Nature*, **250**, 48-50.

Mattson,T., Van Houwe,G. and Epstein,R. (1983) *J. Mol. Biol.*, **170**, 357-379.

Maxam,A. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499-560.

Messing,J. and Vieira,J. (1982) *Gene*, **19**, 269-276.

Messing,J. (1983) *Methods Enzymol.*, **101**, 20-78.

Mosig,G. (1983a) in Mathews,C.K., Kutter,E.M., Mosig,G. and Berget,P.B. (eds.), *Bacteriophage T4,* American Society for Microbiology, Washington, D.C., pp. 120-130.

Mosig,G. (1983b) in Mathews,C.K., Kutter,E.M., Mosig,G. and Berget,P.B. (eds.), *Bacteriophage T4,* American Society for Microbiology, Washington, pp. 362-374.

Mosig,G. and Bock,S. (1976) *J. Virol.*, **17**, 756-761.

Mosig,G., Luder,A., Rowen,L., Macdonald,P. and Bock,S. (1981) in Ray,D. (ed.), *The Initiation of DNA Replication,* Academic Press, NY, pp. 277-295.

Normark,S., Bergstrom,S., Edlund,T., Grundstrom,T., Jaurin,B., Lindberg, F.P. and Olsson,O. (1983) *Annu. Rev. Genet.*, **17**, 499-525.

Nossal,N.G. and Alberts,B.M. (1983) in Mathews,C.K., Kutter,E.M., Mosig, G. and Berget,P.B. (eds.), *Bacteriophage T4,* American Society for Microbiology, Washington, D.C., pp. 71-81.

Rabussay,D. (1983) in Mathews,C.K., Kutter,E.M., Mosig,G. and Berget, P.B. (eds.), *Bacteriophage T4,* American Society for Microbiology, Washington, D.C., pp. 167-173.

Price,A.R. and Warner,H.R. (1969) *Virology*, **37**, 882-892.

Pulitzer,J.F., Coppo,A. and Caruso,M. (1979) *J. Mol. Biol.*, **135**, 979-997.

Sanger,F., Air,G.M., Barrell,B.G., Brown,N.L., Coulson,A.R., Fiddes, J.C., Hutchison,C.A., Slocombe,P.M. and Smith,M. (1977) *Nature*, **265**, 687-695.

Schlagman,S. and Hattman,S. (1983) *Gene*, **22**, 139-156.

Scott,J.R. (1984) *Microbiol. Rev.*, **48**, 1-23.

Shaw,J.E. and Murialdo,H. (1980) *Nature*, **283**, 30-35.

Smith,A.J.H. (1980) *Methods Enzymol.*, **65**, 560-580.

Smith,R.A. and Parkinson,J.S. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 5370-5374.

Snustad,D.P., Snyder,L. and Kutter,E. (1983) in Mathews,C.K., Kutter, E.M., Mosig,G. and Berget,P.B. (eds.), *Bacteriophage T4,* American Society for Microbiology, Washington, D.C., pp. 40-55.

Sternberg,N. and Hoess,R. (1983) *Annu. Rev. Genet.*, **17**, 123-154.

Stitt,B., Revel,H., Lielausis,I. and Wood,W.B. (1980) *J. Virol.*, **35**, 775-789.

Stormo,G.D., Schneider,T.D., Gold,L. and Ehrenfeucht,A. (1982) *Nucleic Acids Res.*, **10**, 2997-3011.

Uzan,M., Leautey,J., D'Aubenton-Carafa,Y. and Brody,E. (1983) *EMBO J.*, **2**, 1207-1212.

Wiberg,J.S. and Karam,J.D. (1983) in Mathews,C.K., Kutter,E.M., Mosig, G. and Berget,P.B. (eds.), *Bacteriophage T4,* American Society for Microbiology, Washington, D.C., pp. 193-201.

Yegian,C.D., Mueller,M., Selzer,G., Russo,V. and Stahl,F.W. (1971) *Virology*, **46**, 900-919.

Received on 21 August 1984

Note added in proof

A protein of the predicted gp69 size is synthesized during the first 11 min after infection of *E. coli* B with wild-type T4, but not after infection with the gene *69 am* mutant HL627-2.