Virology **299,** 182–191 (2002)
doi:10.1006/viro.2002.1409

Microarray Analysis of Gene Expression during Bacteriophage T4 Infection

Kimberly Luke,* Agnes Radek,* XiuPing Liu,* John Campbell,* Marc Uzan,† Robert Haselkorn,* and Yakov Kogan*[,1]

*Integrated Genomics, Inc., 2201 W. Campbell Park Drive, Chicago, Illinois 60612; and †Institut Jacques Monod, CNRS and Universites Paris 6 and 7, Place Jussieu, 75251 Paris Cedex, France

Received January 8, 2002; returned to author for revision January 24, 2002; accepted January 31, 2002

Genomic microarrays were used to examine the complex temporal program of gene expression exhibited by bacteriophage T4 during the course of development. The microarray data confirm the existence of distinct early, middle, and late transcriptional classes during the bacteriophage replicative cycle. This approach allows assignment of previously uncharacterized genes to specific temporal classes. The genomic expression data verify many promoter assignments and predict the existence of previously unidentified promoters. © 2002 Elsevier Science (USA)

Key Words: bacteriophage T4; microarray; transcription.

INTRODUCTION

Bacteriophage T4 infection of *Escherichia coli* has provided many milestones in the history of molecular biology. These include the demonstration that DNA is the genetic material, the discovery of the very nature of mutations, the discovery of messenger RNA, and the triplet nature of the genetic code (reviewed in Mathews, 1994). One of the most important concepts first studied in T4 infection was that of conditional lethal mutations (temperature sensitive and nonsense mutations), which allow examination of essential genes and gene products (Hinton, 1989; Horvitz, 1974; Karam and Bowles, 1974; Karam and O'Donnell, 1973; Wiberg *et al.,* 1962). This revolutionized bacteriophage biochemistry and physiology and facilitated new discoveries in nearly every area of molecular biology (Mathews, 1994; Wood and Revel, 1976). The introduction of microarray technology, allowing simultaneous genome-wide analysis of transcription, heralds a similar revolution in our understanding of cell and molecular biology (Schena *et al.,* 1995).

The bacteriophage T4 genome is a double-stranded DNA molecule of 169 kb encompassing about 300 open reading frames (ORFs), only half of which have known functions. The DNA molecule packaged in the virion is slightly longer due to a few percentages of terminal redundancy. About two-thirds of the genome is devoted to the early and middle genes transcribed from one DNA strand. The remaining third of the genome, carrying most of the late genes, is transcribed from the opposite strand. Several late genes are found embedded within early or middle gene clusters; these are transcribed in the same direction as the early and middle genes (Kutter *et al.,* 1994).

Bacteriophage T4 gene expression depends upon host transcription machinery. Phage-encoded transcriptional regulators modify and redirect the host RNA polymerase to recognize and bind certain promoter classes (reviewed in Mosig and Hall, 1994). The temporally regulated transcription program relies upon the sequential utilization of three classes of promoters: the early, middle, and late promoters, each characterized by a consensus sequence. As soon as T4 DNA is injected into the host cell, about 40 early promoters are recognized by unmodified host σ^{70} RNA polymerase holoenzyme (Wilkens and Ruger, 1994). Transcription from these promoters gives rise to polycistronic mRNAs, which can be very long (Brody *et al.,* 1983; Hsu and Karam, 1990; Sanson and Uzan, 1993). Considerable read-through, resulting in one gene being transcribed from both a single proximal and several more distal promoters, occurs during the prereplicative period. There are many factor-independent T4 terminators, which may terminate transcription initiated from several promoters, resulting in multiple transcripts covering the same gene. Further complications arise from rho-dependent termination, which does not occur at specific sites and is not 100% efficient.

After 2 to 3 min of infection (at 30°C) early promoters are abruptly shut off and middle promoters turned on (Pene and Uzan, 2000.). Thirty-three such promoters have been identified (Marshall *et al.,* 1999; Truncaite, 2001). Their activation requires two T4-encoded proteins. MotA binds to the −30 region of these promoters and AsiA associates with the σ^{70} RNA polymerase (Marshall *et al.,* 1999; Ouhammouch *et al.,* 1995). Genes transcribed from a proximal middle promoter are also fre-

[1] To whom correspondence and reprint requests should be addressed at Microarray Division, Integrated Genomics, Inc., 2201 West Campbell Park Drive, Chicago, IL 60612. Fax: (312) 491-0856. E-mail: yakov@integratedgenomics.com

0042-6822/02 $35.00

quently transcribed from distal early promoters, indicating an apparent lack of efficient transcription termination (Daegelen and Brody, 1990; Hinton, 1989; Hsu and Karam, 1990; Macdonald and Mosig, 1984b,c; Sanson and Uzan, 1993). Inefficient transcription termination may explain how RNA polymerase initiating at early or middle promoters can read through into late regions, to produce anti-late RNAs (Brody *et al.,* 1983.).

Late promoters are turned on about 7 min after infection. Late transcription predominates after 12 to 15 min, when middle transcription has started to decline. A specific T4-encoded sigma factor, the product of gene *55*, is required for late promoter recognition (Kassavetis and Geiduschek, 1984). Late transcription also depends on ongoing viral DNA replication and several T4 polypeptides are responsible for this coupling (Sanders *et al.,* 1997; Tinker *et al.,* 1994; Williams *et al.,* 1994).

Although transcriptional regulation is a predominant factor in T4 gene expression, mRNA processing and degradation also play pivotal roles in determining the abundance of a transcript. The rapid switch from early to late gene expression is likely to be facilitated by selective transcript degradation or stabilization. Both host and phage endo- and exoribonucleases are involved in processing transcripts (Young *et al.,* 1980; Ueno and Yonesaki, 2001; Uzan, 2001). *E. coli* RNase E was the first endonuclease shown to be involved in the processing of T4 mRNAs (Mudd *et al.,* 1988). This endonuclease appears to be responsible for general transcript turnover (Young *et al.,* 1980). The phage-encoded RegB endoribonuclease has been shown to contribute to this processing. This sequence-specific RNase only targets early mRNAs (Sanson *et al.,* 2000; Sanson and Uzan, 1993; Uzan, 2001). In addition to its role in rRNA maturation, *E. coli* RNase III is involved in the processing of at least one specific T4 transcript (Barth *et al.,* 1988). The product of T4 gene *61.5* (now called *dmd*) selectively stabilizes late transcripts and destabilizes middle transcripts via a mechanism that may be activated by MotA (Ueno and Yonesaki, 2001). In addition to controlling mRNA abundance by differential degradation, T4 gene expression involves a great deal of translational control. Many mRNAs contain alternate ribosome-binding sequences whose availability depends on where the transcript begins. These cases give rise to circumstances that appear to place a gene in more than one category, for example, the lysozyme gene (*e*), whose transcript is present both early and late but which is only translated late (McPheeters *et al.,* 1986).

Microarray technology makes it possible to follow each stage of T4 development on a gene-by-gene basis. Bacteriophage systems present unique technical challenges to conventional microarray methods. Not only are the mechanisms of transcript production, maintenance, and degradation complex, but the lack of a reference state complicates data analysis. The standard approach to genomic expression profiling emphasizes deviation from a control condition (Schena *et al.,* 1995). The present study used microarrays to identify temporal classes of RNA produced during T4 infection of *E. coli*. These data can be compared with the 2D gel electrophoresis data of Cowan *et al.* (1994) to identify differences that could indicate possible posttranscriptional and translational regulation. The genomic array data also allow assignment of previously uncharacterized ORFs to the various temporal classes of T4 gene expression and suggest new locations for promoters of some genes.

RESULTS

Temporal patterns of T4 transcription

Total RNA from bacteriophage T4-infected *E. coli* B was isolated 0, 1, 3, 5, 9, 15, 20, and 25 min after infection and the abundance of RNA from each ORF in the genome was measured. A plot of the relative abundance of RNA from each ORF over the time of infection reveals how well the temporal classification scheme matches gene expression on a global level. Figure 1 shows four distinct classes of genes: immediate early, delayed early, middle, and late. For the most part, the data closely match previously published assignments. In addition, there is at least one other pattern: a few genes exhibiting high abundance early in infection disappear during the middle stages and then reappear late.

The classification we used assigns an ORF as immediate early, delayed early, middle, or late based on the time it is first detected within the total RNA pool. The 114 genes comprising the immediate early class are those detected in the first three time samples (0–3 min). Transcripts of another 30 genes identified as delayed early genes first appear between 3 and 5 min. Fifty-seven middle gene transcripts are initially detected between 5 and 8 min. Sixty-two late genes have transcripts only appearing after 8 min postinfection. Discrepancies between this temporal classification and other systems can arise from several sources. We classify a gene based on the time a transcript first appears, rather than when the protein product becomes most abundant. Many published temporal assignments are based upon initial or maximal appearance of a protein product. Large lags between the appearance of a transcript containing a gene and the corresponding protein product infer the presence of a posttranscriptional regulatory system. Finally, some predictions of temporal class are based on promoter studies that do not reflect the complex transcriptional regulatory mechanisms such as pausing or antitermination that may occur *in vivo*.

Immediate early

The transcriptional profiles of a number of immediate early genes exhibit interesting features. Transcripts of

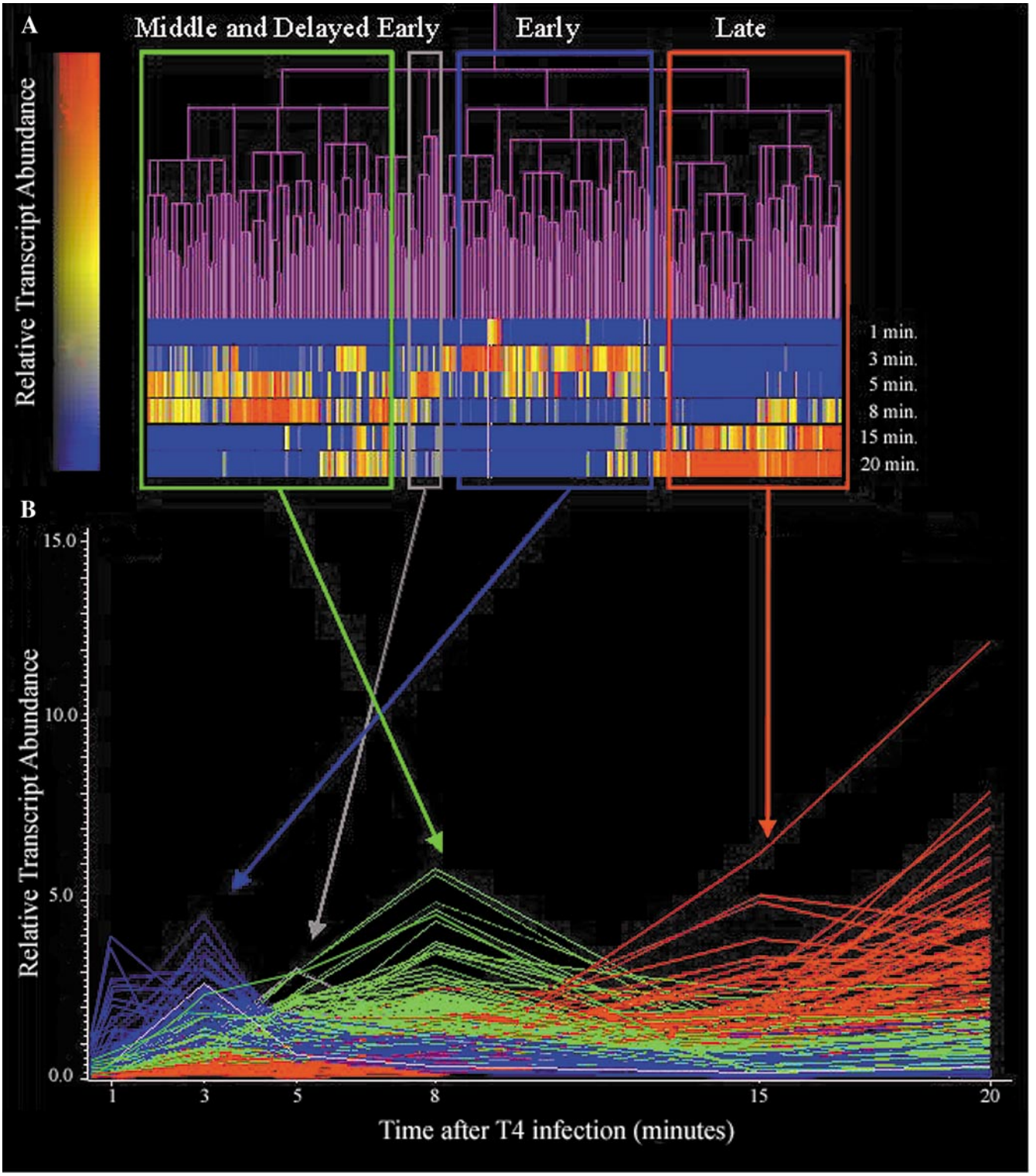
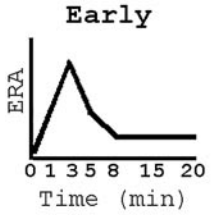
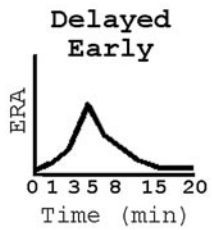


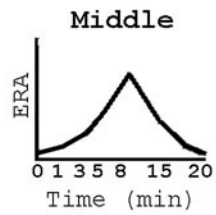
FIG. 1. Clustering based on gene expression. (A) T4 genes were randomly divided into distinct groups based on similarities in expression pattern via K-means clustering. Genes are initially divided into five equal groups and the average expression behavior of each group calculated. Individual genes are then reassigned to the group best representing their expression profile. Group averages are then recalculated and the process reiterated until group compositions converge. This results in four main groups. The dendrogram depicts the relative abundance of each ORF throughout the course of T4 development. The color bar to the left indicates relative expression levels with warm colors denoting increased expression relative to the control pool while cooler colors indicate decreasing levels of expression. The horizontal bars at the bottom of the panel represent the relative level of gene expression at times indicated to the left. (B) Time course of the relative abundance of each T4 gene with relative expression level of each ORF on the abscissa and time after infection along the ordinate. Early genes are blue; delayed early genes are gray; middle genes are green, and late genes are red.



24.1	31.2	55.4	arn.2	dda.1	frd	modB	nrdC.1	pseT.3	tk.3
24.2	33	55.5	arn.3	dda.2	frd.1	motA	nrdC.2	regA	trna.2
24.3	39.1	55.6	arn.4	denA	frd.2	motA.1	nrdC.3	regB	trna.3
30	39.2	55.8	asiA	dsbA	frd.3	motB	nrdC.4	repEA	trna.4
30.2	42	58=61	asiA.1a	e.1	imm	motB.1	nrdC.5	repEB	vs
30.3'	45	61.4	Bgt	e.2	imm.1	motB.2	nrdC.8	rnh	vs.1
30.4	52	61.5	cd	e.3	I-TevI	mrh	nrdC.9	sp	
30.6	52.1	ac	cd.1	e.4	modA	mrh.2	nrdC.10	stp	
30.7	55	agt.2	cd.2	e.5	modA.2	ndd.1	nrdC.11	td	
30.8	55.1	alc	cef	e.6	modA.3	ndd.2a	pseT	tk	
30.9	55.2	arn	comC-A	e.7	modA.4	ndd.4	pseT.1	tk.1	
31.1	55.3	arn.1	dda	e.8	modA.5	nrdB+	pseT.2	tk.2	



30.1	49.1	nrdC
30.5	49.2	nrdC.6
31	49.3	nrdC.7
32.1	62	nrdD
43	cd	vs.7
44	cd.3	vs.8
45.2	cd.5	
46	denV	
46.1	ndd	
46.2	ndd.2	
49	ndd.3	
49'	ndd.5	



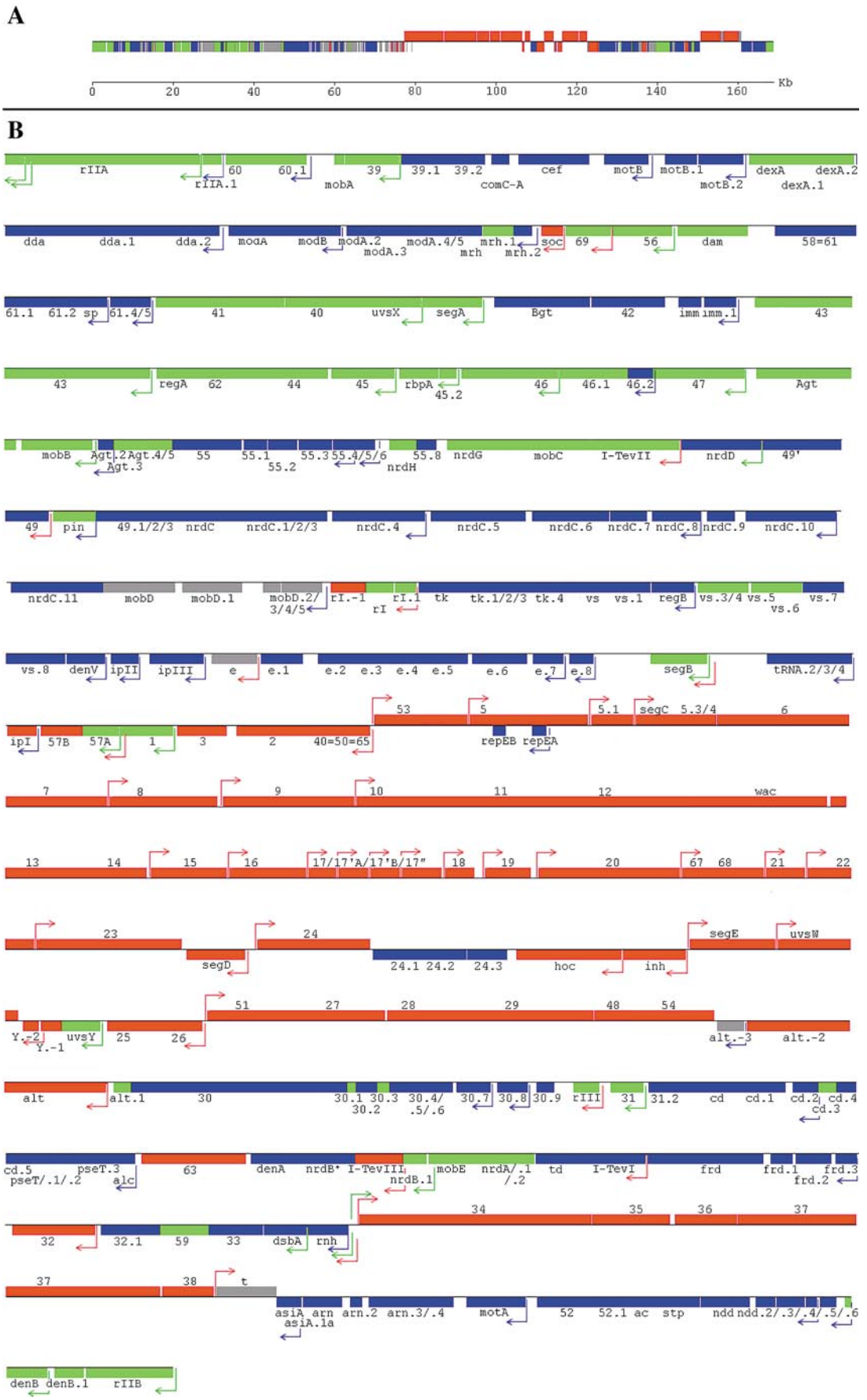
1	69	dexA.2	nrdA.1	rpbA
30.3	agt	ipII	nrdA.2	segA
39	agt.3	ipIII	nrdB.1	segB
40	agt.4	I-TevII	nrdG.1	uvsX
41	agt.5	I-TevIII	nrdH	uvsY
43	alt.1	mobA	pin	vs.3
47	cd.4	mobB	rI	vs.4
56	dam	mobC	rI.1	vs.5
57A	denB	mobE	rIIA	vs.6
59	denB.1	meh.1	rIIA.1	
60	dexA	ndd.6	rIIB	
60.1	dexA.1	nrdA	rIII	

2	11	20	35	alt	Y.-1
3	12	21	36	alt.-1	Y.-2
4=50=65	13	22	37	alt.-2	
5	14	23	38	hoc	
5.1	15	24	48	inh	
5.3	16	25	51	rI.-1	
5.4	17	26	53	segC	
6	17"	27	54	segD	
7	17'A	28	57B	segE	
8	17'B	29	63	soc	
9	18	32	67	uvsW	
10	19	34	68	wac	

FIG. 2. Model expression profiles of individual representative genes from each temporal class of T4 are shown in the figures at the left of the table. Next to each figure is a list of individual genes displaying similar expression profiles. Genes that do not fit any of the temporal classes such as *e, alt.-3, t,* and the six genes of the *mobD* cluster are discussed individually in the text. ERA, Estimated Relative, Transcript Abundance.

e.8, *e.7*, *mobD.5*, and *frd.3* are most abundant in the 1-min samples and entirely absent by 5 min. Comparison with the data of Cowan *et al.* (Table 1 of Cowan *et al.,* 1994) suggests that some of these genes may encode the seven unknown proteins that exhibit similar temporal profiles on 2D gels. The 2D gel data list another eight proteins including MotA and Alc as immediate early products. These genes are also detected by the microarray as immediate early. In addition, genes closely linked to both *alc* and *motA* are detected as immediate early products by microarray analysis. These include *pseT.3*, *pseT.2*, *pseT.1*, and *pseT* which are found immediately downstream of *alc* and *arn.4*, *arn.3*, *arn.2*, *arn.1*, and *arn* downstream of *motA*. A complete list of immediate early genes is shown in Fig. 2. Many of these genes exist in operons downstream of immediate early promoters such as the *frd.3-frd* and *e.8-e.1* clusters, revealing the extent of the transcriptional unit associated with each promoter. A graphic representation of the temporal expression pattern for each T4 gene is shown in Fig. 3.

Identification of an immediate early cluster of genes comprising *24.3*, *24.2*, and *24.1* within a large cluster of genes encoding structural proteins (between *24* and *hoc* on the T4 map) suggests the presence of an immediate early promoter in this region. A very good match to the *E. coli* σ^{70} −10 consensus is found 30 bases upstream of the start codon of *24.3*, but there is no obvious −35 element associated with this sequence. It is possible that the actual promoter for this region lies farther upstream within the *hoc* gene. The PCR product used to identify *hoc* on the microarray is roughly centered within *hoc* (from 110,988 to 110,655 in the GenBank file). An


A
0
20
40
60
80
100
120
140
160
Kb
B
rIIA
rIIA.1
60
60.1
mobA
39
39.1
39.2
comC-A
cef
motB
motB.1
motB.2
dexA
dexA.2
dexA.1
dda
dda.1
dda.2
modA
modB
modA.2
modA.3
modA.4/5
mrh
mrh.1
mrh.2
soc
69
56
dam
58=61
61.1
61.2
sp
61.4/5
41
40
uvsX
segA
Bgt
42
imm
imm.1
43
regA
62
44
45
rbpA
45.2
46
46.1
46.2
47
Agt
mobB
Agt.2
Agt.4/5
Agt.3
55
55.1
55.2
55.3
55.4/5/6
55.8
nrdH
nrdG
mobC
I-TevII
nrdD
49'
49
pin
49.1/2/3
nrdC
nrdC.1/2/3
nrdC.4
nrdC.5
nrdC.6
nrdC.7
nrdC.8
nrdC.9
nrdC.10
nrdC.11
mobD
mobD.1
mobD.2/
3/4/5
rI.-1
rI.1
rI
tk
tk.1/2/3
tk.4
vs
vs.1
regB
vs.3/4
vs.5
vs.6
vs.7
vs.8
denV
ipII
ipIII
e
e.1
e.2
e.3
e.4
e.5
e.6
e.7
e.8
segB
tRNA.2/3/4
53
5
5.1
segC
5.3/4
6
ipI
57B
57A
1
3
2
40=50=65
repEB
repEA
7
8
9
10
11
12
wac
13
14
15
16
17/17'A/17'B/17"
18
19
20
67
68
21
22
23
segD
24
24.1
24.2
24.3
hoc
inh
segE
uvsW
51
27
28
29
48
54
Y.-2
Y.-1
uvsY
25
26
alt.-3
alt.-2
alt
alt.1
30
30.1
30.2
30.3
30.4/
.5/.6
30.7
30.8
30.9
rIII
31
31.2
cd
cd.1
cd.2
cd.4
cd.3
cd.5
pseT.3
pseT/.1/.2
alc
63
denA
nrdB*
I-TevIII
nrdB.1
mobE
nrdA/.1
/.2
td
I-TevI
frd
frd.1
frd.3
frd.2
34
35
36
37
32
32.1
59
33
dsbA
rnh
37
38
t
asiA
arn
asiA.1a
arn.2
arn.3/.4
motA
52
52.1
ac
stp
ndd
ndd.2/.3/.4/.5/.6
denB
denB.1
rIIB

early promoter found in the distal portion of *hoc* would first be detected by the microarray as *24.3* (from 110,041 to 109,915, GenBank). However, sequence inspection does not reveal any potential promoters within the *hoc* gene. A similar situation occurs with *nrdB*$^+$ and *denA* which are detected early in infection. These genes are associated with late genes (*63* and *I-TevIII*) and the nearest known early promoter that could transcribe *nrdB*$^+$ and *denA* is associated with *frd.3*, almost 9 kb away. The microarray data suggest the presence of an early promoter, in close proximity to the beginning of *nrdB*$^+$.

Delayed early

Although early and middle gene promoters are highly conserved and can be distinguished at the sequence level, delayed early genes are not so easily identified. We classify a delayed early gene as one that first appears between 3 and 5 min postinfection. A distinct delayed early class of genes is visible in Fig. 1. A complete list of the 30 delayed early genes detected by microarray analysis is shown in the second panel of Fig. 2. Many genes that are physically linked to immediate early genes fall into this class. In some cases, delayed early and middle gene expression may result from read-through from early promoters. This may explain the appearance of *nrdA.2*, *nrdA.1*, *nrdA*, and *mobE*, which are presumably transcribed from the immediate early promoter responsible for expression of the *frd*-associated genes (see Fig. 3).

Inspection of Fig. 3 reveals many instances for which no known middle promoters are closely associated with middle genes and the nearest early promoters are a significant distance upstream. Delay in the transcription of many genes from early promoters may simply reflect the time it takes RNA polymerase to produce long messages (Truncaite, 2001). Classification of delayed early genes on the basis of the appearance of protein does not directly measure the appearance of the RNA. An example of this is gp55, which is classified as a middle gene based on its appearance on 2D gels at 5 to 7 min (Cowan *et al.,* 1994), yet RNA is first detected between 3 and 5 min by microarray. In this case, *55* is a delayed early gene, but gp55 appears as a middle protein.

Middle transcripts

Figure 1 clearly shows two sets of genes that are best described as delayed early and middle. The 57 genes first appearing between 5 and 8 min are listed in Fig. 2 as middle genes. Many middle promoters require activation by MotA (Adelman *et al.,* 1998; Hinton, 1989). Twelve genes with MotA-dependent promoters are detected by the microarrays, including *1, 39, 43, 47, 56, 57A, αgt, rIIA, rIIB, segA, uvsX,* and *uvsY*. Many of the remaining middle genes are associated with this group. For example, *40* and *41* are immediately downstream of *segA* and *uvsX*. Others are found some distance downstream from early promoters, such as *nrdA.2, nrdA.1, nrdA,* and *mobE,* which occur at the end of the early *frd* cluster.

Late transcripts

Most of the 62 ORFs identified as late genes cluster into one or two large contiguous regions on the chromosome. Most late genes are transcribed in the clockwise direction on the circular T4 map, left to right on the linearized version in Fig. 3. Most of these genes encode structural proteins that comprise the virus head, tail, and tail fibers. Figure 3 clearly shows additional isolated examples of late genes around the chromosome, including *soc*, *rI.-1*, *63,* and *I-TevIII.* These encode a nonessential small outer capsid protein, an uncharacterized ORF adjacent to *rI*, an RNA ligase/tail fiber attachment protein, and an ORF within the *nrdB* intron, respectively. In general, the microarray data are in agreement with previously published late gene assignments.

Complex transcription patterns

A small number of genes including *t*, *alt.-3*, *e*, and the *mobD* cluster are not easily classified as early, middle, or late genes. These transcripts appear early, disappear during the middle periods, and reappear late. There are a number of possible explanations for this behavior. A given sequence may be present on transcripts generated from different promoters. The six genes comprising the *mobD* cluster could be expressed from either of two promoters. A proximal early promoter at 58,730 and a more distal late promoter between *tk* and *rI.1* could produce two overlapping transcripts containing the *mobD* cluster. Another explanation for dual peaks of gene expression early and late invokes sequence-specific endonuclease activity. *RegB* has been shown to specifically degrade early, but not middle or late, mRNAs (Sanson *et al.,* 2000; Sanson and Uzan, 1993, 1995). However, the extent of RegB posttranscriptional regulation in these cases is unknown.

FIG. 3. The expression pattern of phage genes superimposed onto the physical map of bacteriophage T4. The map displays each ORF as defined in GenBank file (NC 000866). Colored boxes on the corresponding strand of the chromosome represent each ORF. (A) Gene coordinates (in kb) are shown below the figure. (B) ORFs are labeled with the corresponding gene names. Promoter locations, indicated by arrows, are those of Kutter *et al.* (1994) and E. Kutter, personal communication. Blue boxes represent both immediate early and delayed early genes. These genes are detectable on the microarrays between 0 and 3 min and 1 and 5 min postinfection, respectively. Middle genes are shown in green and are observed between 5 and 8 min after infection. The red boxes represent late genes, appearing after 8 min postinfection. Similarly, the color of the arrows signifies early, middle, and late gene promoters. Gray boxes represent ORFs with more complex expression patterns.

The *e* gene product is an endolysin, required for host cell lysis. The *e* gene is transcribed both early and late in the infection cycle, but the protein is predominately expressed late (McPheeters *et al.,* 1986). Transcription from the early promoter produces a transcript with a highly stable RNA hairpin that sequesters the ribosome binding site and start codon of the *e* gene. Transcription from the late promoter produces a shorter transcript with no potential secondary structure within the ribosomal loading sequences. Genes *soc*, *49,* and the intron endonucleases are regulated by similar transcription-translation relations (Macdonald and Mosig, 1984a; Barth *et al.,* 1988).

The transcription of both *alt.3* and *t* may be simpler than the microarray data suggest. Known early and late promoters flank both genes. In the case of *alt.3*, a known early promoter exists immediately upstream of the gene, which is likely to account for its transcript early in the infection. The *alt.3* gene has late promoters on either side on both the sense and antisense strands. Either of these would be capable of producing RNA that could hybridize to the *alt.3*-specific double-stranded DNA features on the microarrays. These promoters are associated with the *alt* transcript on the sense strand, or the *51-54* transcript on the antisense strand. It seems likely that *alt.3* is an early gene and that detection of this gene late in infection is an artifact of the double-stranded DNA construction of the microarrays and inefficient termination of the *alt* transcript. In the case of the *t* gene, there is a known late promoter immediately upstream of the gene, implying that this gene should be regarded primarily as a late gene. Detection of *t* as an early gene is probably due to hybridization of sequences from the early *asiA* transcript on the non-sense strand of the target feature on the arrays.

DISCUSSION

The results of the microarray experiments reported here support all the previously observed features of the bacteriophage T4 transcriptional program. At a gross level, the microarrays reflect the temporal organization of transcription into immediate early, delayed early, middle, and late classes. This is dramatically illustrated in Fig. 1. The fact that most of the ORFs comprising the T4 genome can be neatly categorized as early, delayed early, middle, or late demonstrates that these categories do not merely reflect the behavior of a small subset of well-studied T4 genes.

The microarray data further reveal the relative abundance of each gene within the pool of transcripts. In particular, the rapid appearance and subsequent disappearance of some of the immediate early genes such as *e.8, e.7, frd.3,* and *mobD.5* are striking. With the exception of *mobD.5,* all of these genes could produce reasonably sized proteins. These proteins are likely to play a crucial role in the early events of phage infection. There is no genetic or 2D gel information available for these proteins (Williams *et al.,* 1994), while previous low-resolution array studies lack the ability to detect gene transcription at the level of individual ORFs (Young *et al.,* 1980).

Data from these microarrays are displayed with the promoter assignments of Kutter *et al.* (1994) in Fig. 3. This map shows only a handful of discrepancies between promoters and gene assignments. For example, an early promoter is predicted immediately upstream of *rIIA.1* but this gene appears as a middle gene in the microarray data. A late promoter is assigned to the sequences immediately upstream of *69* (Macdonald and Mosig, 1984b), but the gene is identified as a middle gene in the microarray experiments. Similarly, late promoters are found associated with *I-TevII*, *rI.1,* and *rIII*, although all are classified as middle genes in the microarray experiments. Gene *49′*, classified as an early gene, is associated with a late promoter as well. These discrepancies may be due to small differences in the criteria used for calling the exact temporal class of a particular gene or promoter.

We have classified both *regA* and *45* as early genes, while the 2D gel data from Cowan *et al.* (1994) list these as middle genes, with protein abundance peaking around 3 to 5 min. Promoters upstream of this region are classified as middle gene promoters (Kutter *et al.,* 1994). Both proteins are present on the 2D gels as faint bands very early in the course of infection, which indicates that message RNA must be present immediately after infection. By our criteria, *regA* and *45* are early genes.

Genes *24.3*, *24.2*, and *24.1* are identified as early genes despite the fact that late genes surround them on the chromosome. Both *24.2* and *24.3* are very small ORFs, 92 and 56 amino acids, respectively, and may not produce functional products. However, *24.1* encodes an approximately 35-kDa protein that appears to have significant similarity to RNA ligases (Altschul *et al.,* 1997). No genetic data concerning these genes are available.

A single gene, *ndd.2*, is detected at high a level in the 0-min sample, decreases in the 1-min sample, and is maximally expressed by 5 min. This gene is similar (54 of 108 nucleotides conserved) to an uncharacterized *E. coli* gene *yhfC*. The T4 microarray may cross-hybridize with cDNA produced from *E. coli* message RNA encoding this gene. Host transcripts disappear rapidly after infection and the decrease in *ndd.2*-specific cDNA may reflect the decrease in *yhfC* abundance. The increase in signal at 5 min is likely to represent *ndd.2*-specific RNA, so this gene is best considered a delayed early gene.

Internal discrepancies within the microarray data include *mrh.1*, identified as a middle gene, while all the surrounding genes are clearly expressed early. The presence of an early promoter just upstream of the ORF preceding *mrh.1* also suggests it is likely to be an early, rather than middle, gene. The transcriptional activity of

the early promoter may be below our threshold of detection but has been demonstrated by others (Frazier and Mosig, 1988). A similar situation exists with *nrdH*, *55.8*, and *nrdD*. The latter two are categorized as early genes, while *nrdH* is classified as a middle gene. The final anomalous assignment is *59*, which is flanked by early genes that are probably transcribed from a complex set of promoters of all classes immediately upstream of *rnh*. Throughout the entire array, only a few genes are classified in a way that cannot be explained from the known biology or as semantic differences in defining early, middle, and late transcripts. This reflects how well the T4 developmental program is described by this classification system. Moreover, it also dramatically demonstrates the utility of microarray-based analysis in defining complex transcriptional events.

In general the microarray experiments have proven useful for confirming and extending our knowledge of the T4 transcriptional program. The ability to identify genes expressed early in the course of infection allows us to focus on those genes most likely to inhibit critical host functions. The protein products of these immediate early genes might provide useful antimicrobial compounds.

MATERIALS AND METHODS

Array design and fabrication

Microarrays were designed using the public annotated T4 genome (GenBank/NCBI Accession No.: NC 000866). Custom software was used to design primers specific to 300-bp regions within each ORF exhibiting the least sequence similarity to the rest of the genome. Primers were synthesized by MWG Biotech (High Point, NC).

Each 300-bp target feature was originally amplified from genomic DNA prepared by phenol extraction from T4 (provided by Peter Geiduschek). Amplification reactions used MBI Fermentas (Hanover, MD) enzymes and buffers. The thermal profile was 95°C, 2 min followed by 25 cycles of 95°C, 30 s; 64°C, 30 s; 72°C, 1 min; followed by a single 72°C, 10-min incubation. A second round of amplification to increase the purity and concentration of the PCR products used the same conditions, with first-round product diluted 1:50 as template. Second-round products were purified using the TeleChem ArrayIt (Sunnyvale, CA) purification system. The length of each product was verified by gel electrophoresis and each was sequenced to ensure its identity. The concentration of each PCR product was determined by absorbance at 260 nm, dried, and resuspended in 50% DMSO to a concentration of 0.25 μg/μl.

For 29 ORFs, 60-base single-strand oligomers were obtained from MWG Biotech. Oligos for ORFs that failed PCR amplification were made for genes *49*, *mobD.2*, *32*, and *ndd.6*. Strand-specific oligos were also made for *rI*, *denV*, and *ipIII*, which are surrounded by late genes but transcribed in the "early" direction and for *soc*, which is a late gene surrounded by early genes. Likewise, three strand-specific oligos were made for early genes *24.1*, *24.2*, and *24.3* in addition to the gene-specific double-stranded PCR products. Strand-specific oligos were also used to resolve overlapping genes: *nrdD* and *I-TEVII*, *repEB* and *repEA* and *5*, and *30.3* and *30.3′*. Strand-specific positive controls were used for genes *30*, *31*, *51*, *54*, and *e* such that the oligos would map to the region covered by the PCR products. Three pairs of oligonucleotides were also designed to examine transcription through intergenic regions between *dam* and *58 = 61*, *e.8* and *segB*, and *segB* and *tRNA.2*.

PCR products and oligonucleotides were spotted onto amino-silane coated glass slides from Corning (Corning, NY) using a Flexys Arrayer manufactured by Genomic Solutions, Inc. (Ann Arbor, MI). Each ORF was spotted four times per slide. A set of positive controls was included on each slide to allow normalization of data between and across slides and experiments. Ninety-six *E. coli* stress-related genes were also included on the array. Details of the layout of the arrays as well as raw expression data are available online at http://www.integratedgenomics.com/T4/index.html.

Bacterial and phage growth

E. coli B^E was grown at 30°C in MOPS-Tricine medium (Neidhardt *et al.,* 1974) supplemented with 0.4% glucose and 0.4 mM of each amino acid. Bacteriophage T4D was added to give a multiplicity of infection of 7–10 to cells in log-phase growth. Three different cultures (A, B, and C) were infected and processed separately.

RNA sampling and cDNA probe synthesis

At the desired time, aliquots of the cultures were removed and added to an equal volume of lysis solution (SDS 2%, 4 mM EDTA) maintained in boiling water. After a few minutes at 100°C, the samples were stored at −80°C until processing. Cell samples were processed by addition of an equal volume of water-saturated phenol and RNA was extracted at 67°C twice, then extracted with phenol/chloroform, and precipitated with ethanol. Samples were treated with RNase-free DNase (Qiagen, Valencia, CA), extracted with phenol/chloroform, and precipitated with ethanol. The resulting RNA preparations were quantified by UV absorbance at 260 nm and examined by agarose gel electrophoresis.

The RNA preparations were labeled using a standard cDNA protocol. Briefly, 10 μg of purified RNA and 5 μg of random hexamer oligonucleotides (Invitrogen, Carlsbad, CA) were heated to 70°C for 10 min and then placed on ice. RNA was reverse-transcribed with 400 units of SuperScript II (Invitrogen) in the presence of 0.5 mM dATP, dCTP, dGTP, 0.3 mM dTTP, and 0.2 mM 5-(3-aminoallyl) dUTP (Sigma, St. Louis, MO) and 40 U of RNasin (Promega, Madison, WI), at 25°C for 10 min followed by 2.5 h

at 42°C, in a total volume of 15 μl. The reaction was terminated with 0.1 M EDTA, and RNA was hydrolyzed with 0.2 M NaOH at 65°C for 15 min. The solution was neutralized with 0.3 M Tris–HCl, pH 7.4. Unincorporated nucleotides were removed by filtration on a Microcon YM30 membrane (Millipore Corp., Bedford, MA), and cDNA was vacuum-dried. For Cy3 or Cy5 coupling, cDNA samples were resuspended in 9 μl 0.1 M carbonate buffer (pH 8.5–9.0), added to dried aliquots of monofunctional Cy-dyes (Amersham-Pharmacia, Piscataway, NJ), and incubated in the dark at room temperature for 1 h. Unreacted dye was quenched with 4.5 μl of 4 M hydroxylamine (pH 8.5–9.0) and removed using QiaQuick PCR purification columns (Qiagen).

Hybridization conditions

Hybridizations were conducted on a GeneTAC Hybridization Station (Genomic Solutions, Inc.). Slides were incubated in a prehybridization solution consisting of 5× SSC, 0.1% SDS, and 1% BSA for 45 min at 42°C, rinsed with water, and dried by centrifugation. Labeled cDNA probes were resuspended in UltraHyb hybridization solution (Ambion, Austin, TX) and loaded into hybridization chambers. The slides were incubated at 65°C for 2 h and then shifted to 42°C for at least 14 h. Slides were removed from the hybridization chambers and washed manually to remove unhybridized probe. An initial room temperature wash of 5 min in 1× SSC, 0.1% SDS was followed by 5 min in 0.1× SSC, 0.1% SDS, and then 0.1× SSC for 1 min. The slides were then rinsed with deionized H_2O and dried by centrifugation at 1000 rpm for 3 min. Washed slides were scanned in a GenePix 4000B scanner (Molecular Dynamics, Sunnyvale, CA).

Data analysis

Microarray images were processed with GenePix Pro 3.0 software (Axon Instruments, Foster City, CA). The 635/532 nm ratio was calculated using GenePix Pro 3.0 and normalized results were imported into GeneSpring 4.0 (Silicon Genetics, Redwood City, CA) for further analysis. Each of the 21 RNA time samples was analyzed a minimum of four times, resulting in four independent measurements of each feature in quadruplicate.

Images produced by comparing differentially labeled time points revealed many features with purely green or purely red signals. Strong signals in one channel with no detectable signal in the other fail the automated quality control filters built into the software package. To avoid this problem, each sample was compared to a universal control sample. Early experiments showed that a mixture of RNA harvested 5 and 20 min after infection produced a pool of transcripts representing every gene in the T4 chromosome. Equal quantities of total RNA from each sample were combined and diluted to a working concentration of 3 μg/μl. All the relative abundance measurements reported here are normalized against the extension products from the control pool.

ACKNOWLEDGMENTS

We thank Tiffany Sharma for excellent technical assistance and Elizabeth Kutter for encouragement and critical discussion throughout this work. We also thank Gisela Mosig and Peter Geiduscheck for helpful comments on the manuscript.

REFERENCES

Adelman, K., Brody, E. N., and Buckle, M. (1998). Stimulation of bacteriophage T4 middle transcription by the T4 proteins MotA and AsiA occurs at two distinct steps in the transcription cycle. *Proc. Natl. Acad. Sci. USA* **95,** 15247–15252.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402.

Barth, K. A., Powell, D., Trupin, M., and Mosig, G. (1988). Regulation of two nested proteins from gene *49* (recombination endonuclease VII) and of a lambda RexA-like protein of bacteriophage T4. *Genetics* **120,** 329–343.

Brody, E. N., Rabussay, D., and Hall, D. (1983). Regulation of transcription of prereplicative genes. *In* "Bacteriophage T4" (C. K. Matthews, E. M. Kutter, G. Mosig, and P. B. Berget, Eds.), American Society of Microbiology, Washington, DC.

Cowan, J., d'Acci, K., Guttman, B., and Kutter, E. (1994). Appendix B: Gel analysis of T4 prereplicative proteins, *In* "Molecular Biology of Bacteriophage T4" (J. Karam, J. Drake, K. Kreuzer, G. Mosig, D. Hall, F. Eiserling, L. Black, E. Spicer, E. Kutter, K. Carlson, and E. Miller, Eds.), pp. 520–527. American Society of Microbiology, Washington, DC.

Daegelen, P., and Brody, E. N. (1990). The rIIA gene of bacteriophage T4. II. Regulation of its messenger RNA synthesis. *Genetics* **125,** 249–260.

Frazier, M. W., and Mosig, G. (1988). Roles of the *Escherichia coli* heat shock sigma factor 32 in early and late gene expression of bacteriophage T4. *J. Bacteriol.* **170,** 1384–1388.

Georgopoulos, C., Georgiou, M., Selzer, G., and Eisen, H. (1977). Bacteriophage T4 mutants which propagate on *E. coli* K12 but not on *E. coli* B. *Experientia* **33,** 1157–1158.

Hinton, D. M. (1989). Altered expression of the bacteriophage T4 gene *41* (Primase-helicase) in an *Escherichia coli rho* mutant. *J. Biol. Chem.* **264,** 14440–14446.

Hinton, D. M. (1991). Transcription from a bacteriophage T4 middle promoter using T4 MotA protein and phage-modified RNA polymerase. *J. Biol. Chem.* **266,** 18034–18044.

Horvitz, H. R. (1974). Bacteriophage T4 mutants deficient in alteration and modification of the *Escherichia coli* RNA polymerase. *J. Mol. Biol.* **90,** 739–750.

Hsu, T., and Karam, J. D. (1990). Transcriptional mapping of a DNA replication gene cluster in bacteriophage T4. *J. Biol. Chem.* **265,** 5303–5316.

Karam, J. D., and Bowles, M. G. (1974). Mutation to overproduction of bacteriophage T4 gene products. *J. Virol.* **13,** 428–438.

Karam, J. D., and O'Donnell, P. V. (1973). Suppression of amber mutations of bacteriophage T4 gene 43 (DNA polymerase) by translational ambiguity. *J. Virol.* **11,** 933–945.

Kassavetis, G. A., and Geiduschek, E. P. (1984). Defining a bacteriophage T4 late promoter: Bacteriophage T4 gene 55 protein suffices for directing late promoter recognition. *Proc. Natl. Acad. Sci. USA* **81,** 5101–5105.

Kutter, E., Stidham, T., Guttman, B., Batts, D., Peterson, S., Djavakhishvili, T., Arisaka, F., Mesyanzhinov, V., Ruger, W., and Mosig, G. (1994).

Appendix A: Genomic map of bacteriophage T4. *In* "Molecular Biology of Bacteriophage T4" (J. Karam, J. Drake, K. Kreuzer, G. Mosig, D. Hall, F. Eiserling, L. Black, E. Spicer, E. Kutter, K. Carlson, and E. Miller, Eds.), pp. 491–519. American Society of Microbiology, Washington, DC.

Macdonald, P. M., and Mosig, G. (1984a). Regulation of a bacteriophage T4 gene, *soc*, which maps in an early region. *Genetics* **106,** 17–27.

Macdonald, P. M., and Mosig, G. (1984b). Regulation of a new bacteriophage T4 gene, *69*, that spans an origin of DNA replication. *EMBO J.* **3,** 2863–2871.

Macdonald, P. M., and Mosig, G. (1984c). Cloning and physical mapping of an early region of the bacteriophage T4 genome. *Genetics* **106,** 1–16.

Marshall, P., Sharma, M., and Hinton, D. M. (1999). The bacteriophage T4 transcriptional activator MotA accepts various base-pair changes within its binding sequence. *J. Mol. Biol.* **285,** 931–944.

Mathews, C. K. (1994). An overview of the T4 developmental program. *In* "Molecular Biology of Bacteriophage T4" (J. Karam, J. Drake, K. Kreuzer, G. Mosig, D. Hall, F. Eiserling, L. Black, E. Spicer, E. Kutter, K. Carlson, and E. Miller, Eds.), pp. 1–8. American Society of Microbiology, Washington, DC.

McPheeters, D. S., Christensen, A., Young, E. T., Stormo, G., and Gold, L. (1986). Translational regulation of expression of the bacteriophage T4 lysozyme gene. *Nucleic Acids Res.* **14,** 5813–5826.

Mosig, G., and Hall, D. H. (1994). Gene expression: A paradigm of integrated circuits. *In* "Molecular Biology of Bacteriophage T4" (J. D. Karam, Ed.), pp. 127–131. American Society for Microbiology, Washington, DC.

Mudd, E. A., and Carpousis, A. J. (1990). *Escherichia coli* RNase E has a role in the decay of bacteriophage T4 mRNA. *Genes Dev.* **4,** 873–881.

Mudd, E. A., Prentki, P., Belin D., and Krisch, H. M. (1988). Processing of unstable bacteriophage T4 gene 32 mRNAs into a stable species requires *Escherichia coli* ribonuclease E. *EMBO J.* **7,** 3601–3607.

Neidhardt, F. C., Bloch, P. L., and Smith, D. F. (1974). Culture medium for enterobacteria. *J. Bacteriol.* **119,** 736–747.

Ouhammouch, M., Adelman, K., Harvey, S. R., Orsini, G., and Brody, E. N. (1995). Bacteriophage T4 MotA and AsiA proteins suffice to direct *Escherichia coli* RNA polymerase to initiate transcription at T4 middle promoters. *Proc. Natl. Acad. Sci. USA* **92,** 1451–1455.

Pene, C., and Uzan, M. (2000). The bacteriophage T4 anti-sigma factor AsiA is not necessary for the inhibition of early promoters in vivo. *Mol. Microbiol.* **35,** 1180–1191.

Sanders, G. M., Kassavetis, G. A., and Geiduschek, E. P. (1997). Dual targets of a transcriptional activators that tracks on DNA. *EMBO J.* **16,** 3124–3132.

Sanson, B., Hu, R. M., Troitskayadagger, E., Mathy, N., and Uzan, M. (2000). Endoribonuclease RegB from bacteriophage T4 is necessary for the degradation of early but not middle or late mRNAs. *J. Mol. Biol.* **297,** 1063–1074.

Sanson, B., and Uzan, M. (1993). Dual role of the sequence-specific bacteriophage T4 endoribonuclease RegB mRNA inactivation and mRNA destabilization. *J. Mol. Biol.* **233,** 429–446.

Sanson, B., and Uzan, M. (1995). Post-transcriptional controls in bacteriophage T4: Roles of the sequence-specific endoribonuclease RegB. *FEMS Microbiol. Rev.* **17,** 141–150.

Schena, M., Shalon, D., Davis R. W., and Brown, P. O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270,** 467–470.

Tinker, R. L., Williams, K. P., Kassavetis, G. A., and Geiduschek, E. P. (1994). Transcription activation by a DNA-tracking protein: Structural consequences of enhancement at the T4 late promoter. *Cell* **77,** 225–237.

Truncaite, L. (2001). Deletional and transcriptional analysis of the genomic region between genes *30* and *31* of bacteriophage T4. Ph.D. thesis, Vilnius, Lithuania.

Ueno, H., and Yonesaki, T. (2001). Recognition and specific degradation of bacteriophage T4 mRNAs. *Genetics* **158,** 7–17.

Uzan, M. (2001). The bacteriophage T4 RegB endoribonuclease. *Methods Enzymol. B* **342,** 467–480.

Wiberg, J., Dirksen, M., Epstein, R., Luria, S., and Buchanan, J. (1962). Early enzyme synthesis and its control in *E. coli* infected with some amber mutants of bacteriophage T4. *Biochemistry* **48,** 293–302.

Wilkens, K., and Ruger, W. (1994). Transcription from early promoters. *In* "Molecular Biology of Bacteriophage T4" (J. Karam, J. Drake, K. Kreuzer, G. Mosig, D. Hall, F. Eiserling, L. Black, E. Spicer, E. Kutter, K. Carlson, and E. Miller, Eds.), pp. 132–141. American Society of Microbiology, Washington, DC.

Williams, K., Kassavetis, G., Herendeen, D., and Geiduschek, E. P. (1994). Regulation of late-gene expression. *In* "Molecular Biology of Bacteriophage T4" (J. Karam, J. Drake, K. Kreuzer, G. Mosig, D. Hall, F. Eiserling, L. Black, E. Spicer, E. Kutter, K. Carlson, and E. Miller, Eds.), pp. 161–175. American Society of Microbiology, Washington, DC.

Wood, W. B., and Revel, H. R. (1976). The genome of bacteriophage T4. *Bacteriol. Rev.* **40,** 847–868.

Young, E. T., Mattson, T., Selzer, G., Van Houwe, G., Bolle, A., and Epstein, R. (1980). Bacteriophage T4 gene transcription studied by hybridization to cloned restriction fragments. *J. Mol. Biol.* **138,** 423–445.