Monocistronic and Polycistronic Bacteriophage T4 Gene 23 Messages

ELTON T. YOUNG,* ROSEMARY CRONE MENARD, AND JOHN HARADA

Department of Biochemistry, University of Washington, Seattle, Washington 98195

Received 13 February 1981/Accepted 1 July 1981

We studied transcription of T4 late genes by *in vitro* translation of size-fractionated late RNA and by hybridization of T4 late RNA to plasmids containing identified T4 late genes. We identified mRNA species that coded for the late proteins gp10, gp18, gp21, gp22, gp23, and gp24. Functional mRNA's that coded for the early proteins gp32 and IPIII were also detected after fractionation of late RNA. As in preparations of early RNA, gene 32 message activity was present in two species of RNA, which had molecular weights of approximately 0.5×10^6 and 0.8×10^6 , and IPIII message activity was present in multiple species of RNA. Gene 24 and gene 10 message activities migrated as single species that had approximate molecular weights of 0.5×10^6 and 1.2×10^6 , respectively. mRNA activity for gp18 migrated heterogeneously. We detected multiple transcripts from gene 23 by *in vitro* translation and by hybridization of late RNA to plasmids containing genes 21 through 23. Both types of analysis indicated that the major gene 23 transcript had a molecular weight of 0.75×10^6 . In addition, two gene 23 transcripts having molecular weights of about 1.0×10^6 and 1.3×10^6 were present; these RNA species also coded for gp21 and gp22. Physical linkage of transcripts from genes 21, 22, and 23 was demonstrated by hybridization.

Bacteriophage T4 late gene expression appears to be unique among procaryotes in that DNA replication is required to achieve efficient late gene transcription. Late transcription also requires the polypeptides that are encoded by phage genes 33, 55, and 45. These polypeptides associate with and alter the transcriptional specificity of the host core RNA polymerase, which itself is modified covalently early in T4 infection. Despite considerable genetic, biochemical, and physiological knowledge of the requirements for late transcription, late transcription *in vitro* was not achieved until recently (10). This was due both to the complex requirements for late gene expression and to the large genome size of phage T4. Much of the T4 genome has now been cloned into phage and plasmid vectors (8, 16, 17). In particular, plasmids containing most of the genes in the late region have been identified, and the T4 restriction enzyme fragment present in the genetically identified plasmid has been characterized. These isolated restriction fragments and the plasmids containing them make it possible to study transcription of a single late gene. This should facilitate analysis of late gene transcription. Initial studies (5a, 5b; T. Mattson, personal communication) have suggested that late genes can be expressed from a plasmid under certain circumstances. However, it is still difficult to distinguish among vector-promoted T4 gene

expression, T4 DNA-mediated gene expression from a plasmid, and gene expression from a partial recombinant between an infecting phage genome and a plasmid containing T4 gene (5a, 5b; Mattson, personal communication). This difficulty emphasizes the need to locate the late transcription promoters or RNA transcription start sites utilized during normal T4 infections.

At present there is little evidence that can be used to identify late promoter sites. The T4 late genes are present in three major clusters (18), which are separated by early genes. Thus, there are presumably at least three T4 late promoters. Since most and probably all true late genes are transcribed from the strand opposite that used for transcribing T4 early genes, the late genes cannot be transcribed by readthrough from the early region, as appears to be the case in bacteriophage lambda (13). Within the clusters of T4 late genes, there is some indication of polarity, suggesting that the genes involved are cotranscribed. For example, some amber mutations in gene 9 cause a defect in gene activity such that phage containing a gene 9 amber mutation cannot be complemented by a mixed infection with a gene 10 amber mutant (10, 14). There is also some biochemical evidence for polarity. For example, mutations in genes 9 and 6 cause decreased synthesis of polypeptides from the adjacent genes (genes 10 and 7, respectively) (15).

So far, however, there is no physical evidence for cotranscription of T4 late genes and no indication of the location or nature of T4 late promoters.

In an initial attempt to locate transcription units within one of the T4 late regions, we measured the sizes of a few specific late mRNA's. Functional late messages were detected after T4 late RNA was fractionated by continuous-elution preparative gel electrophoresis and translated in a cell-free system. Specific late RNAs were also detected by hybridizing T4 late RNA to plasmids containing restriction enzyme fragments of known T4 genetic origin. We focused on the gene 23 message, which codes for the major capsid protein (the most abundant T4 late protein).

MATERIALS AND METHODS

Bacterial strains, phage, and plasmids. We used *Escherichia coli* B^F, a nonsuppressing host for amber mutations, for all T4 infections. Bacteriophage T4D⁺ was the wild-type phage. Various amber mutants of T4 were obtained from G. Doermann. The plasmids used have been described previously (8, 20) and were maintained in *E. coli* 802 (suII⁺r_k⁻m_k⁺). Plasmid p652A contains a 1.0-kilobase *EcoRI*-*HindIII* restriction enzyme fragment of T4 DNA that includes all known genetic markers in gene 21, the 5'-proximal portion of gene 22, and no genetic markers in gene 20. Plasmid p652B contains a 1.1-kilobase *HindIII* restriction enzyme fragment of T4 DNA that includes the 3' end of gene 22 and the 5' end of gene 23. Plasmid p652C contains a 1.4-kilobase *HindIII*-*EcoRI* restriction enzyme fragment of T4 DNA that includes all of the gene 23 amber alleles not in p652B; it probably also contains some non-gene 23 DNA, but does not contain any known gene 24 amber alleles. The plasmid vector in which these T4 DNA fragments were cloned was pBR313.

Media. The media used are described in the accompanying paper (20).

Chemicals. Most reagents were from standard sources. *m*-Nitrobenzyl alcohol was obtained from Aldrich Chemical Co. 1-(*m*-Nitrobenzyloxy)methyl pyridinium chloride was prepared as described previously (1) by Mark Rixon under the guidance of Scott Chilton.

Preparation of derivatized cellulose paper and transfer of RNA. We used the procedure of Alwine et al. (1), with a few modifications. The amount of 1-(*m*-nitrobenzyloxy)methyl pyridinium chloride per square centimeter of paper was doubled.

Agarose gel electrophoresis of glyoxalated RNA. RNA was glyoxalated by incubating the RNA in 50% dimethyl sulfoxide-1 M glyoxal-10 mM KPO₄ (pH 7.0) for 60 min at 50°C (9). The treated RNA was loaded onto a slot in a slab gel containing 1.5% agarose in 10 mM KPO₄ (pH 7.0) and electrophoresed for 16 h at 2 V/cm. The buffer was circulated to avoid ion depletion.

Nick-translation of plasmid and phage DNAs. For nick-translation we used the procedure of Rigby

et al. (12); we used [³²P]dTTP and *E. coli* DNA polymerase I (Boehringer Mannheim), but pancreatic DNase I was not added. After incubation at 20°C for 60 min, the non-incorporated nucleoside triphosphates were removed by filtration through Sephadex G-100. DNA with a specific activity of about 10⁶ cpm/μg was obtained.

Hybridization. Cellulose sheets containing covalently linked RNA were hybridized to nick-translated DNA probes as described by Alwine et al. (1). Usually, 0.5 × 10⁶ to 1 × 10⁶ cpm of DNA was used. The other hybridization procedures used are described in the accompanying paper (20).

Other experimental procedures. Fractionation of RNA by continuous-elution preparative electrophoresis, cell-free protein synthesis, analysis of radioactive polypeptides, and other techniques are described in the accompanying paper (20).

RESULTS

Sizes of functional late transcripts. First, we measured the sizes of specific late transcripts by fractionating T4 late RNA on polyacrylamide gels and then translating the eluted RNA in a cell-free protein-synthesizing system. The radioactive polypeptides were separated by electrophoresis on sodium dodecyl sulfate-polyacrylamide gels, and the specific polypeptides were identified by comparison with gels containing radioactive proteins derived from infections of *E. coli* with T4 amber mutants. This assay for late RNA revealed only the most prominent message species and could identify unambiguously those messages which coded for polypeptides that migrated in unique regions of the sodium dodecyl sulfate-polyacrylamide gels. It was also limited in its usefulness to the messages that coded for proteins smaller than 100,000 daltons, since larger polypeptides were not synthesized efficiently in the cell-free system.

Figure 1 shows an autoradiogram of the ³⁵S-labeled proteins that were synthesized in response to fractionated T4 late RNA. Two polypeptides were synthesized in large amounts. The smaller of these had a molecular weight of 35,000 and was gp32 (gp32 in the polypeptide product of gene 32), the T4 helix-destabilizing protein which is identified in the accompanying paper (20). gp32 is made from a stable, self-regulating mRNA (6, 7). The larger conspicuous polypeptide had a molecular weight of about 58,000 and comigrated with uncleaved gp23, the precursor of the major capsid protein of the phage. This identification was supported by translating RNA obtained by infecting *E. coli* with an amber mutant in gene 23. As shown in Fig. 1, slots A through C, the RNA obtained from 23amB17-infected cells coded for much less of this polypeptide. (Amber mutants were designated by gene number, followed by allele designation; for

FIG. 1. Autoradiogram of polypeptides synthesized by fractionated wild-type T4 RNA. A 400- μ g sample of RNA extracted 20 min after infection of *E. coli* B^E by wild-type T4 was fractionated on a 5.0-cm 2.25% polyacrylamide gel. The eluted RNA was concentrated and translated *in vitro* in the presence of [³⁵S]-methionine. The radioactive polypeptides were separated on a 10% sodium dodecyl sulfate-polyacrylamide gel, and an autoradiogram of the dried gel was made. The molecular weight scale at the top was obtained from the elution rates of marker RNAs run on an identical gel. The three slots on the right show the polypeptides synthesized by unfractionated T4 23amB17 RNA (slot A), unfractionated wild-type T4 RNA (slot B), and endogenous RNA (slot C). The protein molecular weight scale on the left was obtained from marker proteins run on the same gel. Identifications of T4 polypeptides corresponding to the genes shown on the right were made by comparison with extracts from cells infected with phages containing amber mutations in genes 32, 10, 18, 24, 23, and 22. The identifications of gp21 and IPIII were based on comparison with published data for IPIII (3) and the molecular weight of gp21 (18; Showe, personal communication) and were considered tentative.

example, the amber mutation B17 is in gene 23, and the mutant containing this mutation was designated 23amB17.) The radioactivity that remained at the position of gp23 in the slot containing the translation products of the mutant RNA was probably gp12 and, perhaps, gp-wac, which have molecular weights similar to the molecular weight of gp23.

To confirm the identification of gene 23 message activity, RNA extracted from cells infected with 23amB17 was fractionated and analyzed as described above (Fig. 2). The 58,000-dalton band was missing except for a faint band, which probably corresponded to gp12 or gp-wac or both. Thus, we concluded that gp23 comprised the majority of the radioactivity at this position in Fig. 1.

Several other T4 late proteins were also identified in Fig. 1 and 2. The two polypeptides larger than gp23 were gp10 and gp18. The polypeptide with a molecular weight of about 45,000 was gp24. The two polypeptides below gp32 on the gels were gp22 and, probably, gp21. The prominent polypeptide with a molecular weight of about 25,000 was probably IPIII. The identifications of gp21 and IPIII were not confirmed in an analysis of extracts from cells infected with the appropriate amber and deletion mutants. The identification of IPIII was based on a com-

parison with previously published data (3). For gp21 the identification was less certain. M. Showe (personal communication) has reported that the molecular weight of gp21 is 27,500 and the molecular weight of gp22 is 32,000 (18). We were not able to confirm the identification of gp21 by analyzing extracts from our gene 21 amber mutant.

The sizes of the mRNA activities that coded for these polypeptides were determined by a comparison with molecular weight standards run on an identical gel (20). The positions of the peak mRNA activities were determined either by inspection or by densitometry of the autoradiograms. Figure 3 shows the results of densitometry for gp23. There was a major peak of gene 23 message activity at a molecular weight of about 0.75×10^6 and another region of gp23-coding activity with a higher molecular weight. There was a suggestion that two mRNA species migrated at positions corresponding to molecular weights of 1.0×10^6 and 1.3×10^6 . These two higher-molecular-weight gene 23 message activities comigrated with message activities for gp21 and gp22. The activities of the latter two polypeptides and their separation into two peaks are shown most clearly in Fig. 2, which shows the analysis of 23amB17 RNA. Message activity for gp24 had a molecular weight of 0.5×10^6 to 0.6

$\times 10^6$. The gene 32 message activity was present in two peaks that represented two RNA species with molecular weights of about 0.45×10^6 and 0.8×10^6 . These same two species of gene 32 message activity were observed in our analysis of T4 early RNA (20). Message activity for IPIII was also present in both early and late RNA preparations (2). The RNA activity for IPIII was present in several molecular weight species of early RNA (20). This was also true for T4 late RNA. Four distinct peaks of IPIII mRNA activity were present in both Fig. 1 and Fig. 2. These species had molecular weights of 0.3×10^6 , 0.5×10^6 , 0.75×10^6 , and 1.0×10^6 . As stated above, we did not verify that any or all of these species were indeed IPIII messages by using amber or deletion mutants of IPIII. Thus, these identifications were tentative.

The late messages that coded for polypeptides larger than gp23 were more difficult to identify. The only two that we reproducibly detected coded for gp10 and gp18. The functional message for gp10 migrated at a rate corresponding to a molecular weight of 1.2×10^6 and formed a sharp peak on the autoradiogram. The mRNA activity for gp18 was spread through a molecular weight range from 0.7×10^6 to 1.4×10^6 , with possible peaks at 0.8×10^6 and 1.1×10^6 .

Sizes of gene 21, 22, and 23 transcripts by Northern gel analysis. Hybridization of denatured, size-fractionated RNA to specific DNA probes after transfer to a solid support (1) has been used extensively to determine the sizes of eucaryotic mRNA's. We used RNA transfer hy-

bridization to confirm the existence and identities of the gene 21, 22, and 23 messages revealed by in vitro translation (Fig. 4). T4 late RNA was denatured with dimethyl sulfoxide and glyoxal (9) and subjected to electrophoresis on a 1.5% agarose gel. The RNA was transferred to derivatized cellulose paper (1) and hybridized to nick-translated DNA. When total T4 phage DNA was nick-translated and used as a probe, a broad smear of radioactivity was observed (Fig. 4A and B). Most of the RNAs detected by this probe had molecular weights in the range from 0.02×10^6 to 0.5×10^6 . A significant amount of RNA was detected between the positions of 18S RNA and 23S RNA. When a plasmid (p652B) containing the 3' end of gene 22 and the 5' end of gene 23 was nick-translated and used as a probe, distinct RNA species were produced. These migrated with apparent molecular weights of 0.75×10^6 , 1.0×10^6 , and 1.3×10^6 . The band at the position of 18S RNA (0.75×10^6 daltons) was the most abundant. These molecular weights were very similar if not identical to the molecular weights of the functional gene 23 messages shown in Fig. 3. Even the relative amounts of the three species agreed. In addition, there was a considerable amount of lower-molecular-weight RNA that hybridized to the probes. To prove that hybridization was specific to the T4 DNA sequences on the plasmid, the vector pBR322 was nick-translated and hybridized to a blot of the same RNA. As shown in Fig. 4A, there was no hybridization of pBR322 sequences to the RNA on the blot.

FIG. 2. Autoradiogram of polypeptides synthesized by fractionated T4 23amB17 RNA. A 400- μ g sample of RNA extracted from *E. coli* B⁸ infected with T4 23amB17 was analyzed as described in the legend to Fig. 1 and in the text. The three slots on the right show the polypeptides synthesized in vitro in response to no added RNA (endogenous), unfractionated wild-type T4 RNA (T4 wt) and unfractionated T4 23amB17 RNA. The open circle on the autoradiogram indicates the expected position of gp23. The protein molecular weights on the left were obtained by fractionating standard proteins with known molecular weights on the same gel. The T4 polypeptides indicated on the right were identified as described in the legend to Fig. 1.

FIG. 3. Densitometric analysis of *gp23* mRNA activity. An autoradiogram of the *gp23* polypeptide synthesized *in vitro* in response to fractionated wild-type T4 RNA was scanned with a densitometer equipped with an integrator. The molecular weight scale at the top was taken from the original autoradiogram (Fig. 1).

Another preparation of T4 late RNA was analyzed in the same manner (Fig. 4B). This time the nick-translated DNA probes used for hybridization were T4 phage DNA, p652A, and p652C. Plasmid p652A contains gene 21 and the 5'-terminal end of gene 22. It hybridized to RNAs having a broad range of molecular weights (Fig. 4B), including two regions representing RNA species with molecular weights of about 1.0×10^6 and 1.3×10^6 . It did not hybridize specifically to RNA in the 18S region. Plasmid p652C contains the 3'-terminal two-thirds of gene 23. Its hybridization pattern was very similar to that of p652B, which contains parts of both gene 22 and gene 23. Thus, we concluded that the prominent band on the autoradiogram at the 18S position represented gene 23 transcripts and that the two

minor bands at about 23S and 25S represented transcripts that included genes 21, 22, and 23.

Figure 5 shows an analysis of these autoradiograms by densitometry. T4 phage DNA annealed to RNA in the same molecular weight range as the gene 23 transcripts even though most of the hybridization was to lower-molecular-weight RNA. The distribution of gene 23 transcripts detected by hybridization was strikingly similar to the distribution of the functional gene 23 transcripts shown in Fig. 3 if the lower-molecular-weight, apparently degraded RNA was subtracted from the profile.

We applied this technique to other late genes with minimal success. In most cases no distinct bands were found, suggesting that very little RNA of discrete size was present.

Pulse-labeling of gene 23 transcripts. There were several possible explanations for the existence of multiple gene 23 transcripts. The smaller, presumably monocistronic RNA could have been derived by processing of a larger polycistronic RNA. This could have explained the greater abundance of the smaller RNA. Another possibility was that there were two or more promoters in this region. One promoter could have been used for transcribing the smaller gene 23 message, and another one (or two) could have been used for transcribing the polycistronic RNAs. Other alternatives were also possible.

If the large RNA were a precursor of the smaller, most of the gene 23-specific radioactivity incorporated in a short pulse would be expected to accumulate in the larger RNA. To investigate this possibility, RNA was labeled from 12 to 15 min after T4 infection, extracted, and fractionated on a preparative polyacrylamide gel. Portions of each fraction were hybridized to filters containing various plasmid DNAs. Figure 6 shows the distribution of the radioactive RNA that hybridized to the plasmids containing genes 22 and 23 (p652B) and gene 23 (p652C). The distribution of the gene 23-specific RNA was very similar to the distribution revealed by hybridization of nick-translated gene 23 DNA to a Northern blot of late T4 RNA (Fig. 5). A peak of radioactivity that represented RNA with an approximate molecular weight of 0.75×10^6 was the most conspicuous feature of the distribution, which also had a shoulder that suggested an RNA species with a molecular weight of about 1.1×10^6 . A small amount of higher-molecular-weight RNA was also present. We concluded that if the large RNA identified by both translational assays and by Northern gel analysis is a precursor of the smaller RNA, it must have a short half-life.

FIG. 4. Autoradiogram of nick-translated DNA hybridized to fractionated bacteriophage T4 RNA. RNA was extracted and purified from phage-infected cells, denatured with dimethyl sulfoxide and glyoxal, fractionated on a 1.5% agarose gel, and transferred to derivatized cellulose. The hybridization with nick-translated probes (about 10^6 cpm) was at 67°C , as described in the text. The migration positions of rRNA markers fractionated on the same gel are shown (dots next to each slot). The marker RNAs were identified in each gel before transfer by staining the NaOH-treated and neutralized gel with ethidium bromide. (A) Bacteriophage T4 RNA was extracted 18 min after wild-type T4 infection and was probed with nick-translated total T4 DNA, pBR322 DNA, and p652B (genes 22 and 23) DNA. Duplicate strips were incubated for 18 h (lanes a) and 72 h (lane b) for hybridization. (B) Bacteriophage T4 RNA was extracted 25 min after wild-type T4 infection and was probed with nick-translated T4 phage DNA, p652A (genes 21 and 22), or p652C (gene 23). Hybridization was for 3 days.

Cotranscription of genes 21, 22, and 23. Cotranscription of genes 21, 22, and 23 should be detectable by hybridization to filters containing the appropriate plasmid DNAs. If some RNA molecules contained both gene 21-specific sequences and gene 23-specific sequences, then hybridization to a plasmid containing gene 21 DNA followed by elution of the RNA and hybridization a second time to a plasmid containing gene 23 DNA sequences should detect such molecules. By performing the hybridization reactions in the reverse order, the approximate amounts of the polycistronic and monocistronic RNAs could be measured (Table 1). [^3H]RNA was labeled continuously from 0 to 18 min after T4 infection in order to reach a steady-state distribution of monocistronic and polycistronic gene 23 messages. This radioactive RNA was

hybridized first to filters containing either p652A DNA or p652C DNA. Plasmid p652A contains all or most of gene 21 and the 5' end of gene 22, whereas plasmid P652C contains the 3' two-thirds of gene 23. The T4 DNA in these plasmids is separated in the T4 genome by 1.1 kilobases, which is present as a *Hind*III restriction fragment in p652B. After hybridization to filters containing either p652A DNA or p652C DNA, the RNA was eluted as described above. Next, each eluted RNA was hybridized in a vial containing either a filter charged with p652A DNA or a filter charged with p652C DNA. A control filter containing only vector DNA was also present in each vial. The RNAs eluted from these filters showed some cross-hybridization, indicating that RNA molecules initially annealed to p652A were able to anneal to p652C and that the

FIG. 5. Densitometric analysis of T4 mRNA fractionated on an agarose gel. The autoradiograms shown in Fig. 4 were scanned with a densitometer. The following nick-translated probes were used during hybridization: T4 phage DNA (line A), p652B (genes 22 and 23) DNA (line B), p652A (genes 21 and 22) DNA (line C), and pBR322 DNA (line D). The open circles indicate the positions of artifactual peaks produced by spurious background regions in the autoradiogram.

converse was also true. Since these sequences were 1.1 kilobases apart in the T4 genome, some of the mRNA molecules must have spanned this region. The amounts of hybridization transfer were different for the RNA species eluted from p652A and p652C. About 31% of the RNA eluted from p652A was able to hybridize to p652C, whereas only about 7% of the RNA eluted from p652C was able to hybridize to p652A. These values suggested that most gene 23 transcripts were not linked to gene 21 and 22 transcripts, whereas many gene 21 and 22 transcripts were linked to gene 23 transcripts. This was consistent with the observations that most functional gene 23 transcripts were approximately monocis-

tronic in size but a minority (about 10 to 20%) were polycistronic and also encoded gp21 and gp22 and that no other gene 21 or gene 22 transcripts were present (Fig. 1). A more quantitative analysis of the amount of cotranscription is not reasonable at this time since RNA degradation would separate gene 21 and gene 23 sequences on polycistronic RNA.

An alternative explanation is that the T4 DNA present in p652A was partially homologous to the T4 DNA present in p652C. However, if the hybrids were treated with RNase and then the annealed RNA was eluted, very little if any cross-hybridization was detected (data not shown). This suggested that genes 21, 22, and 23 are not sufficiently homologous to cross-hybridize.

FIG. 6. Fractionation of pulse-labeled late T4 RNA and hybridization to gene 23-containing plasmids. RNA was labeled from 15 to 18 min after infection with wild-type T4, rapidly extracted by a hot sodium dodecyl sulfate lysis technique (4), and fractionated by electrophoresis as described in the legend to Fig. 1 and in the text. About 10^6 cpm was put on the gel. One-half of each 1-ml fraction was hybridized with a filter containing pBR322 DNA, p652B DNA, or p652C DNA. The radioactivity that annealed to pBR322 (vector DNA) was subtracted from all data. The migration positions of the *E. coli* 16S and 23S rRNA's are shown at the top.

TABLE 1. *Cotranscription of genes 21, 22, and 23*

First hybridization ^a			Second hybridization ^b				
Plasmid DNA	Amt an-nealed (cpm)	% An-nealed	Plasmid DNA	Input (cpm)	Amt an-nealed (cpm)	% An-nealed ^c	Normal-ized % an-nealed ^d
p652A (genes 21 and 22)	3,170	0.31	p652A	1,427	253	14	100
			pBR322	1,427	47		
			p652C	1,427	92	4.4	31
			pBR322	1,427	29		
p652C (gene 23)	48,300	4.3	p652C	21,700	4,158	19	100
			pBR322	21,700	103		
			p652A	21,700	336	1.4	7.4
			pBR322	21,700	39		

^a The initial input of radioactive RNA was 1.1×10^6 cpm of RNA labeled from 0 to 18 min after wild-type T4 infection. The hybridization reaction lasted 5 days at 42°C in a final volume of 1.0 ml containing 50% formamide, 0.03 M sodium citrate, and 0.3 M sodium chloride. The hybrids were not treated with RNase.

^b The second hybridization reaction was performed under the conditions described in footnote *a* and lasted 5 days. The hybrids were not treated with RNase.

^c The background radioactivity that annealed to a filter containing pBR322 was subtracted from all data.

^d The percent annealed in the second hybridization to the same DNA from which the radioactive RNA was eluted after the first hybridization was defined as 100%.

DISCUSSION

In our analysis of T4 late transcripts, we concentrated on identifying mRNA from the gene 23 region. Gene 23 codes for the most abundant late protein in cells, the major capsid protein. This protein is synthesized as a polypeptide which is about 550 amino acids long (gp23); 65 amino acid residues are cleaved from the amino terminus of this polypeptide by the gene 21 protease during head maturation to produce gp23* (5; A. Boosman, personal communication). The synthesis of gp23 and gp23* constitutes 20 to 30% of the protein synthesis during the late phase of T4 infections, an amount which is consistent with the stoichiometry of these proteins in phage particles. The neighboring genes (genes 21, 22, and 24) are also involved in head morphogenesis. Gene 21 codes for a protease which is required for head assembly; gene 22 codes for an internal scaffolding protein that is cleaved into small pieces by gp21 during head maturation; and gene 24 codes for the capsid structural protein gp 24, which forms the vertex of the viral capsid. Like gp23, gp24 is synthesized as a precursor and is cleaved by gp21 to produce the polypeptide found in mature capsids (gp24*). gp21, gp22, and gp24 are synthesized at much lower rates than gp23, although the rates are difficult to measure because the complex head maturation pathway requires cleavage of gp22, gp23, and gp24.

Gene 23 expression is regulated in the same manner as the expression of other late genes; that is, it requires DNA replication, gp33, gp55, and gp45. The mRNA of gene 23 is transcribed

at a rate that is typical for many late genes (19), but the gene 23 RNA appears to be somewhat more stable than most late RNAs (A. Christensen and E. T. Young, unpublished data). In particular, the rates of transcription measured by hybridizing pulse-labeled RNA to p652A (genes 21 and 22), p652B (genes 22 and 23), and p652C (gene 23) are about the same (19). However, when the radioactive RNAs that accumulate during infections are hybridized to the same plasmids, there is a distinct difference (Table 1). The plasmids containing gene 23 anneal with much more of the radioactive RNA than the plasmids containing genes 21 and 22.

The abundance of gene 23 transcripts and polypeptides and the role of gene 23 in head assembly make gene 23 an obvious model for the study of late gene regulation in vitro. Recent studies (5a, 5b; Mattson, personal communication) have suggested that gene 23 can be expressed from a bacterial plasmid. We determined the nucleotide sequence of T4 DNA present in p652B. The 5' end of gene 23 and the presumed 3' end of gene 22 have been identified, and we are currently mapping the 5' ends of the gene 23 messages (Christensen and Young, unpublished data).

Gene 23 is transcribed into multiple RNA species, including a presumed monocistronic RNA and a polycistronic RNA which also encodes gp21 and gp22. The evidence for this was derived from four different types of experiments. A strongly hybridizing band with a molecular weight of about 0.75×10^6 is present when nonradioactive, denatured RNA is transferred from an agarose gel to derivatized cellulose and

hybridized to a plasmid containing either genes 22 and 23 or gene 23 alone. Two other minor RNA bands are also present, which migrate with mobilities corresponding to molecular weights of 1.0×10^6 and 1.3×10^6 . When the same plasmids are used to analyze pulse-labeled, fractionated T4 late RNA, a radioactive RNA with a similar molecular weight is detected. In the pulse-labeled RNA the most prominent RNA species that hybridizes to gene 23 DNA has a molecular weight of about 0.7×10^6 , as observed in Northern gel analysis. The fact that in both total, nonradioactive late RNA and pulse-labeled late RNA the same gene 23 transcripts are observed and are present in about the same relative proportions suggests that the larger RNA is not a precursor of the smaller RNA. When a plasmid that contains genes 21 and 22 is used as a probe, the prominent band with a molecular weight of 0.75×10^6 is absent, but the two bands of higher-molecular-weight RNA are present, and these comigrate with the RNA that hybridizes to gene 23.

Presumptive monocistronic and polycistronic gene 23 messages were first observed by *in vitro* translation of fractionated late RNA. The most abundant functional gene 23 message has a molecular weight of about 0.75×10^6 . An RNA of this size has a coding capacity sufficient for gp23 (58,000 daltons) with about 300 nucleotides left over. The excess RNA sequences could encode a small protein, or more likely, they could be in noncoding regions. There are numerous small polypeptides whose mRNA's migrate in the same molecular weight range as the major gene 23 message, but we have not attempted to correlate any of these polypeptides with the major gene 23 RNA activity. Alternatively, our estimate of molecular weight could be wrong.

Putative polycistronic gene 23 message activities have molecular weights of about 1.1×10^6 and 1.3×10^6 . In this molecular weight range preparative RNA gels do not provide sufficient resolution to determine whether there is one message or two. However, a Northern gel hybridization transfer analysis, which also detected total steady-state gene 23 RNA, suggested the presence of two high-molecular-weight gene 23 RNAs in addition to the abundant lower-molecular-weight gene 23 RNA. Message activity for gp21 and gp22 comigrated with the higher-molecular-weight message activity coding for gene 23, suggesting that a polycistronic mRNA or mRNA's encode genes 21, 22, and 23. The identification of gp23 was unambiguous. RNA that was isolated from cells infected with an amber mutant in gene 23 (23amB17) and was fractionated by preparative gel electrophoresis did not

make the polypeptide previously identified as gp23. It did make the two polypeptides identified as gp21 and gp22, and the mRNA activities for these two polypeptides migrated at rates corresponding to molecular weights of about 1×10^6 and 1.3×10^6 , as they did when RNAs extracted from cells infected with wild-type T4 were analyzed.

Final proof for the cotranscription of genes 21 through 23 was provided by eluting RNA annealed to genes 21 and 22 to genes 23 sequences and vice versa. About 14% of the RNA annealed to genes 21 and 22 could be eluted and reannealed to genes 21 and 22, and about 4.4% annealed to gene 23, suggesting that approximately 30% of the gene 21 and 22 transcripts extend to at least the middle of gene 23. On the other hand, only 1.4% of the RNA originally annealed to gene 23 could anneal to gene 21 and 22 DNA, whereas 19% of the same RNA could reanneal to gene 23, suggesting that most gene 23 transcripts do not include gene 21 and 22 sequences. Since p652A contains parts of both gene 21 and gene 22, we concluded that at least one of these genes is cotranscribed with gene 23. The evidence from *in vitro* translation indicates that gp21 and gp22 are translated from messages of the same size, suggesting that they are cotranscribed and hence that all three genes can be represented on the same RNA molecule.

The origin of the multiple gene 23 transcripts has not been determined. As with the multiple early T4 transcripts described in the accompanying paper (20), these transcripts could arise by processing of the largest transcript or they could be transcribed from multiple promoters. We prefer the latter hypothesis since the same distribution of gene 23 RNA sizes was observed in steady-state RNA and in pulse-labeled RNA. Moreover, recent experiments (5a, 5b; Mattson, personal communication) have suggested that p652B (genes 22 and 23) contains a promoter that can function during T4 infection to allow gene 23 expression. It is unlikely that the multiple gene 23 RNAs arise by transcription from a single promoter followed by multiple termination sites since the direction of transcription is from gene 21 to gene 23, putting gene 23 sequences at the 3' end of the polycistronic RNA.

It is interesting that an analogous arrangement of genes with similar functions exists in bacteriophage T7 and that the T7 genes are transcribed from multiple promoter sites to a common termination signal at the end of gene 10 (the major capsid protein). The adjacent T7 gene 9 is a scaffolding protein whose function is analogous to the function of T4 gene 22.

The other T4 late messages revealed by translation of fractionated T4 late RNA include several messages that appear to be encoded in multiple mRNA species. One of these may be the message for IPIII, which is not a true late transcript. IPIII also appears to be encoded in multiple mRNA species early in infection (20).

ACKNOWLEDGMENTS

We are grateful to Tom Mattson for his generosity in providing the initial plasmids used in this work and to Gerald Selzer, with whom the initial gene 21, 22, and 23 plasmids were isolated. We thank Mark Rixon and Scott Chilton for supplying 1-(*m*-nitrobenzyloxy)methyl pyridinium chloride. We also thank Lenore Bemmels for her help in preparing the manuscript.

This research was supported by Public Health Service grant AI-09456 from the National Institute of Allergy and Infectious Diseases.

LITERATURE CITED

1. Alwine, J. C., D. J. Kemp, and G. R. Stark. 1977. Method for the detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proc. Natl. Acad. Sci. U.S.A.* **74**:5350-5354.
2. Black, L., and L. Gold. 1971. Pre-replicative development of the bacteriophage T4. RNA:protein synthesis in vivo and in vitro. *J. Mol. Biol.* **60**:365-388.
3. Black, L. W. 1974. Bacteriophage T4 internal protein mutants: isolation and properties. *Virology* **60**:166-179.
4. Hagen, F., and E. T. Young. 1979. Effect of RNase III on the size of bacteriophage T7 lysozyme mRNA. *J. Virol.* **26**:783-792.
5. Isobe, T., M. Yanagida, A. Boosman, and A. Tsugita. 1978. Characterization of the morphogenetic-dependent cleavage region of the major capsid protein (P23) of bacteriophage T4; sequence of an amber fragment of P23. *J. Mol. Biol.* **125**:339-356.
- 5a. Jacobs, K. A., L. M. Albright, D. K. Shibata, and E. P. Geiduschek. 1981. Genetic complementation by cloned bacteriophage T4 late genes. *J. Virol.* **39**:31-45.
- 5b. Jacobs, K. A., and E. P. Geiduschek. 1981. Regulation of the expression of the cloned bacteriophage T4 late gene 23. *J. Virol.* **39**:46-59.
6. Krisch, H. M., A. Bolle, and R. H. Epstein. 1974. Regulation of the synthesis of bacteriophage T4 gene 32 protein. *J. Mol. Biol.* **88**:89-104.
7. Lemaire, G., L. Gold, and M. Yarus. 1978. Autogenous translational repression of bacteriophage T4 gene 32 expression in vitro. *J. Mol. Biol.* **126**:73-90.
8. Mattson, T., G. Van Houwe, A. Bolle, G. Selzer, and R. Epstein. 1977. Genetic identification of cloned fragments of T4 DNA and complementation of some clones containing early T4 genes. *Mol. Gen. Genet.* **154**:319-326.
9. McMaster, G. K., and G. G. Carmichael. 1977. Analysis of single- and double-stranded nucleic acids on polyacrylamide and agarose gels by using glyoxal and acridine orange. *Proc. Natl. Acad. Sci. U.S.A.* **74**:4835-4838.
10. Nakata, A., and F. W. Stahl. 1967. Further evidence for polarity mutations in bacteriophage T4. *Genetics* **55**:585-590.
11. Rabussay, D., and E. P. Geiduschek. 1977. Phage T4-modified RNA polymerase transcribes T4 late genes *in vitro*. *Proc. Natl. Acad. Sci. U.S.A.* **74**:5305-5309.
12. Rigby, P. W. J., M. Dieckmann, C. Rhodes, and P. Berg. 1977. Labeling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. *J. Mol. Biol.* **113**:237-251.
13. Roberts, J. W. 1975. Transcription termination and late control in phage lambda. *Proc. Natl. Acad. Sci. U.S.A.* **72**:3300-3304.
14. Stahl, F. W., J. M. Crasemann, C. D. Yegian, M. M. Stahl, and A. Nakata. 1970. Co-transcribed cistrons in bacteriophage T4. *Genetics* **64**:157-172.
15. Vanderslice, R. W., and C. D. Yegian. 1974. The identification of late bacteriophage T4 proteins on sodium dodecyl sulfate polyacrylamide gels. *Virology* **60**:265-275.
16. Velten, J., and J. Abelson. 1980. The generation and analysis of clones containing bacteriophage T4 DNA fragments. *J. Mol. Biol.* **137**:235-248.
17. Wilson, G. G., V. I. Tanyashin, and N. E. Murray. 1977. Molecular cloning of fragments of bacteriophage T4 DNA. *Mol. Gen. Genet.* **156**:203-214.
18. Wood, W. B., and H. Revel. 1976. The genome of bacteriophage T4. *Bacteriol. Rev.* **40**:847-868.
19. Young, E. T., T. Mattson, G. Selzer, G. Van Houwe, A. Bolle, and R. Epstein. 1980. Bacteriophage T4 gene transcription studied by hybridization to cloned restriction fragments. *J. Mol. Biol.* **138**:423-445.
20. Young, E. T., and R. C. Menard. 1981. Size of bacteriophage T4 early mRNA's separated by preparative polyacrylamide gel electrophoresis and identified by in vitro translation and by hybridization to recombinant T4 plasmids. *J. Virol.* **40**:772-789.