Regulation of Two Nested Proteins From Gene *49* (Recombination Endonuclease VII) and of a λ RexA-Like Protein of Bacteriophage T4

Katrin A. Barth,[1] Doris Powell,[2] Michael Trupin[3] and Gisela Mosig

Department of Molecular Biology, Vanderbilt University, Nashville, Tennessee 37235

Manuscript received December 4, 1987
Revised copy accepted June 22, 1988

ABSTRACT

Phage T4 gene *49*, encoding recombination endonuclease VII, specifies, by initiation from an AUG and an internal GUG codon, two in-frame overlapping peptides (of 18 and 12 kD). The gene is transcribed early and late, albeit from different promoters. The sequence predicts that in long early transcripts, initiated far upstream of the coding sequence, the Shine-Dalgarno sequence of the first ribosome binding site can be sequestered in a hairpin and/or cleaved. These processes might reduce initiation from the first AUG and facilitate initiation of the 12-kD peptide from the internal GUG. The potential of this hairpin to participate in Y structures or cruciforms suggests possible autoregulation. Shorter, more stable late transcripts initiated from a late promoter immediately upstream of the first ribosome binding site cannot form this hairpin. More efficient translation of the longer 18-kD gene *49* peptide from these late transcripts accounts for the strong dependence of endonuclease VII activity on late gene expression. An ORF downstream from gene *49* can be translated from a *motA*-dependent transcript that starts inside gene *49* as well as from the gene *49* transcripts. Its initiation codon overlaps the stop codon of gene *49*, suggesting some coupling of translation. The deduced protein resembles, among others, the RexA protein of phage λ. Possible implications for T4 recombination and for the interference of λ lysogens with T4 gene *49* and *rII* mutants are discussed.

HOMOLOGOUS recombination generates branched intermediates whose heteroduplex regions (HOLLIDAY 1964; WHITEHOUSE 1963) are derived from different parental DNA molecules by strand exchange and branch migration. Cleavage of the branched intermediates at four-way "H" junctions or at simpler, three-way "Y" junctions is a critical step in resolving the intermediates (for reviews see STAHL 1979; DRESSLER and POTTER 1982; WHITEHOUSE 1982). Endonuclease VII (endo VII), the product of gene *49* of phage T4, is the first enzyme shown to catalyze cleavage of these intermediates (MINAGAWA and RYO 1978; KEMPER and GARABETT 1981; JENSCH and KEMPER 1986) and of cruciforms (MIZUUCHI *et al.* 1982; KEMPER *et al.* 1984). This enzyme also excises large heteroduplex loops *in vivo* and *in vitro* (MOSIG and POWELL 1985 and unpublished results; KLEFF and KEMPER 1988). Regulation of its synthesis is of considerable interest for recombination in general, and for T4 specifically, because recombination is essential for T4 DNA replication and growth (for review see MOSIG 1987).

Endo VII has been characterized as a late protein because it is required for packaging of DNA (KEMPER and BROWN 1976) and because its activity is greatly reduced in gene *55* mutants (KEMPER and GARABETT 1981) which are defective in the sigma factor for T4 late transcription (EPSTEIN *et al.* 1964; GEIDUSCHEK, ELLIOTT and KASSAVETIS 1983; RABUSSAY 1983; KASSAVETIS and GEIDUSCHEK 1984). On the other hand, the phenotype of gene *49* mutants, especially the timing of temperature sensitivity of a *ts* mutant, the poor growth of this *ts* mutant at an otherwise permissive temperature in λ lysogens, and the poor growth of *49-rII* double mutants in nonlysogens (MOSIG, SHAW and GARCIA 1984; G. MOSIG, unpublished data) suggested to us that gene *49* might also have some function early after infection.

We show here that gene *49* is transcribed both early and late. It encodes two peptides of different sizes in the same reading frame. Our results also suggest a ready explanation for the effects of λ lysogens and the strong enhancement of endo VII activity by late gene expression.

A preliminary account of this work was reported at the Cold Spring Harbor Meeting on Molecular Genetics of Bacteria and Phages, 1986.

MATERIALS AND METHODS

Phage: T4 wild-type and gene *49* mutants *amE727* and *tsC9*, purified from secondary mutations and the *motA* mutant *sip1* (HOMYK, RODRIGUEZ and WEIL 1976; HALL and SNYDER 1981), have been maintained in this laboratory

[1] Present address: Department of Human Genetics, Columbia University, New York, New York 10032.

[2] Present address: 1904 Woodhue Drive, Burlington, North Carolina 27215.

[3] Present address: Tulane University School of Public Health and Tropical Medicine, 1501 Canal Street, New Orleans, Louisiana 70112.

Abbreviations used: endo VII = T4 endonuclease VII; dC-DNA = T4 DNA containing cytosine instead of glucosylated hydroxymethylcytosine; gp = gene product.

(MOSIG, SHAW and GARCIA 1984). The *35⁻-55⁻* double mutant (*amN134-amBL292*) was constructed and described by LUDER and MOSIG (1982). The dC-DNA-producing strains GT7 (*56⁻-42⁻-denB⁻-alc⁻*) (WILSON *et al.* 1979) and JW800 (*56⁻-42⁻-denA⁻-denB⁻-alc⁻*) (CLARK, WEVER and WIBERG 1980) were kindly supplied by G. WILSON and J. WIBERG, respectively. Although both strains were thought to contain the same gene *56* mutation, we found in appropriate crossses that GT7 carries *amC153* and JW800 carries *amE51*. Phage T7 wild type and a gene *3* amber mutant were kindly supplied by R. WEISBERG.

Bacteria: *Escherichia coli* CR63 (*supD*) and B*suI* (*supD*) or S/6 (*sup⁰*), B (*sup⁰*) (from our collection) and TB-1 (*sup⁰*) from BRL were used as permissive or restrictive hosts, respectively, for routine plating of T4 or T7 phage. UT481 (*r⁻m⁻, met⁻, thy⁻Δ(pro lac) supD*:Tn*10*/F′ *traD36 proA⁺B⁺ lacI^q lacZΔM15*), permissive for *am* mutants (a gift from C. LARK) and TB-1 (*sup⁰*) (purchased from BRL), restrictive for *am* mutants, transformed with T4 insert-bearing plasmids or M13 derivatives, were used as hosts in marker rescue and complementation tests, respectively. *E. coli* M5219 [Sm^R, *am lacZ, am trpA* (λ *bio252, cI857,* Δ*H1*)] (REMAUT, STANSSENS and FIERS 1981) was a gift from E. REMAUT and W. FIERS. Phage and bacteria were grown and assayed by standard procedures (ADAMS 1959; MOSIG, BERQUIST and BOCK 1977). T4 phage containing dC-DNA were grown in *E. coli* B834 *galU sup⁰* (RUNNELS and SNYDER 1978). The *rnc⁻* (RNaseIII defective) strain AB105, originally isolated by P. H. HOFSCHNEIDER, was obtained from DAVID SCHLESSINGER.

Cloning vectors: M13mp19 (MESSING and VIEIRA 1982) was obtained from BRL. Riboprobe Gemini vectors pGEM1 through pGEM3 were purchased from Promega Biotech. The expression vector pPLc2833 (REMAUT, STANSSENS and FIERS 1981) was kindly supplied by E. REMAUT and W. FIERS.

Media: Media are described in ADAMS (1959), GRUIDL and MOSIG 1986) or MANIATIS, FRITSCH and SAMBROOK 1982). All plasmid-bearing bacteria were grown in the presence of 50 or 100 μg/ml ampicillin.

Chemicals and enzymes: Sequencing primers were single-stranded "SP6 Promoter" primer or "T7 Promoter" primer, from Promega Biotech, or custom-synthesized oligonucleotides purchased from the Vanderbilt University Toxicology Center. Reverse transcriptase from avian myeloblastosis virus (AMV) was purchased from Life Sciences; deoxynucleoside triphosphates (dNTP) and dideoxynucleoside triphosphates (ddNTP) were from Pharmacia. Sequencing reactions were carried out using [^{32}P]dCTP (~800 Ci/mmol) from Amersham or [α-thio-^{35}S]dCTP (~1000 Ci/mmol) from New England Nuclear. DNA sequencing gels contained 8 M urea and 4% or 6% acrylamide-bis in 1× TBE buffer (pH 8.3) (121.1 g ultra pure Tris base, 61.8 g boric acid and 7.44 g disodium EDTA·2H_2O per 1000 ml H_2O). They were fixed in 5% methanol and 5% acetic acid and dried prior to autoradiography using KODAK XAR5 film.

Preparation of DNA: Large-scale and small-scale plasmid preparations were done by standard procedures (MANIATIS, FRITSCH and SAMBROOK 1982). To obtain replicative form (RF) DNA of M13 mp19, infected cells were grown in 2× YT medium for 7 hr at 37° with high aeration and processed as in plasmid preparations. T4 dC-DNA was extracted with SDS and phenol from phage GT7 or JW800 particles purified through a CsCl step gradient.

Recombinant DNA procedures, restriction enzyme digests and purification of DNA fragments were as described (MANIATIS, FRITSCH and SAMBROOK 1982; MACDONALD and MOSIG 1984a,b).

Construction of clones: Most clones shown in Figure 1, as well as smaller subclones used for sequencing, were obtained by isolating a 1004-bp *RsaI* fragment (identified by hybridization with *Xba*I-10 T4 DNA) from a total T4 dC-DNA *Rsa*I digest and inserting it into the *Sma*I site of pGEM vectors. Most subclones were constructed after cutting this cloned insert with the restriction enzymes shown in Figure 1. The clones labeled "del" were constructed by Bal 31 digestion of this cloned *Rsa*I fragment. The T4 segments inserted into pGEM vectors are called pA*n* (where *n* stands for the numbers shown in Figure 2). The same T4 segments, when inserted downstream from the λp_L promoter of pPLc2833, are called pLA*n*. pG113, pMT3 and pA70 were subcloned from an *Xba*I-10 fragment originally cloned in M13 mp19. pG113 contains a *Rsa*I-*Pst*I fragment, pMT3 contains a *Rsa*I-*Rsa*I fragment and pA70 contains an *Xba*I-*Rsa*I fragment indicated in Figure 1, inserted into the *Sma*I site of pGEM vectors.

Marker rescue: Since both gene *49* mutants are leaky and accumulate pseudorevertants (CONKLING and DRAKE 1984), spot tests proved to be unreliable. *E. coli* UT481 with plasmids bearing different T4 inserts were grown at 30° with high aeration to 1 to 3 × 10^8 cells/ml in LB. Cells carrying M13 clones were grown in 2× YT medium at 37° to the same density. The cells (0.1 ml) were then infected with T4 gene-*49* mutants at a multiplicity of infection of 5 and incubated at 30° (*amE727*) or 25° (*tsC9*). After 40–90 min the cells were lysed by addition of chloroform. Appropriate dilutions of *amE727* lysates were plated on *E. coli* CR63 (34°) for total progeny and on *E. coli* S/6 (45°) for wild-type recombinants. Lysates of *tsC9* were plated on S/6 at 25° for total progeny and at 45° for wild-type recombinants, because most pseudorevertants do not grow at 45°. All clones negative for marker rescue gave proportions of wild-type progeny below 3 × 10^{-5}. Clones positive for marker rescue gave 10–100-fold higher proportions.

Complementation assays: *E. coli* TB-1 (*sup⁰*), bearing plasmids with or without T4 inserts, were grown to 1 × 10^8 cells/ml in HB or LB medium containing 50 μg/ml ampicillin and concentrated 10-fold by centrifugation. T4 *amE727* or *tsC9* were plated on these bacteria at 42° (to exclude pseudorevertants) and, for comparison, on the permissive host, CR63, at 34°. T7*am3* was plated at 34° and at 37°. The mutants plate on bacteria containing complementing inserts (filled boxes in Figure 1) under otherwise nonpermissive conditions with similar efficiencies and plaque sizes as under permissive conditions (with or without insert-bearing plasmids). Since the gene-*49* mutants are leaky, we also measured burst sizes after infection, at 42°, of TB-1 containing various T4-insert-bearing plasmids in one-step growth experiments (MOSIG, BERQUIST and BOCK 1977). The average burst size of *amE727* under such complementing conditions was approximately one-third (50–70) of the wild-type burst size (180–210). In contrast, the burst sizes of the gene-*49* mutants in hosts carrying plasmids without inserts or with inserts containing only part of the gene *49* coding region (open boxes in Figure 1) were between three and five.

DNA sequencing: Since cloning, in M13 derivatives, of *Xba*I restriction fragments 10 and 16 and subfragments derived from them resulted in frequent spontaneous deletions of both insert and vector sequences, all sequencing was carried out using the modified chain-termination (dideoxy) technique of SANGER, NICKLEN and COULSON (1977) on double-stranded templates (CHEN and SEEBURG 1985) with reverse transcriptase. DNA sequences were determined at least twice from both strands, exactly as described by THOMPSON and MOSIG (1987), directly from various T4 inserts cloned in pGEM plasmids, as indicated in Figure 1.

Sequences were analyzed with the DNA Inspector II Program from Textco Co. on a Macintosh computer. Critical sequences (*e.g.*, the overlap of gene *49* and ORF 49.-1) and sequences spanning the gene *49* mutants were also determined by primer extension sequencing on RNA templates as described by MCPHEETERS *et al.* (1986), except that the RNA was isolated and purified through CsCl centrifugation as described (MACDONALD and MOSIG 1984b).

Analysis of RNA synthesized *in vivo* by Northern blot analysis: RNA made *in vivo* was isolated from infected and uninfected *E. coli* B as described (MACDONALD and MOSIG 1984b), denatured by glyoxalation and separated by electrophoresis (MANIATIS, FRITSCH and SAMBROOK 1982). DNA size standards, generated from restriction digests of pBR322, were glyoxalated and size-fractionated in parallel with the RNA samples. The nucleic acids were blotted to Biodyne nylon membrane (Pall Ultrafine Filtration Corp.). The blots were probed with strand-specific RNA probes synthesized *in vitro* as described below. Hybridization conditions and autoradiography were as previously described (MACDONALD and MOSIG 1984b) except that the post-hybridization washes were done at 65° in 15 mM NaCl, 1.5 mM sodium citrate and 0.1% SDS.

S1 nuclease mapping of the gene-*49* transcript ends: Hybridizations, subsequent S1 nuclease digestions and gel analyses of the protected fragments were done as described by GRUIDL and MOSIG (1986).

***In vitro* preparation of strand-specific RNA probes for hybridization:** pGEM derivatives containing the desired T4 insert were linearized with appropriate restriction enzymes, extracted with phenol, precipitated with ethanol and transcribed with T7 or SP6 RNA polymerase (MELTON *et al.* 1984) as described by the vendor.

Protein analysis: Plasmid bearing cells were grown at 25° in M9 medium (ADAMS 1959) to 3–5 × 10^8/ml. They were shifted to 42° and, 5 min later, ^{3}H- labeled amino acid mixtue (5 μCi/ml) (ICN) was added. After 30 min at 42° or 37°, the cells were harvested by centrifugation and resuspended in lysis buffer. The proteins were electrophoretically separated in denaturing polyacrylamide gels and autoradiographed (LAEMMLI 1970).

RESULTS

Identification of gene *49*: Gene *49* maps (MOSIG 1983b) between the early genes *nrdC* (thioredoxin, LEMASTER 1986) and *55* (sigma factor for late T4 transcription, GEIDUSCHEK, ELLIOTT and KASSAVETIS 1983). Marker rescue experiments (see MATERIALS AND METHODS) map gene *49* to an *Xba*I-10 fragment cloned into M13 mp19, to a *Rsa*I fragment (pA2) nested in *XbaI*-10, and, more precisely, the gene-*49* mutations *tsC9* to a 76-nucleotide *Taq*I-*Pst*I fragment and *amE727* to a 137-nucleotide *Hin*fI fragment (Figure 1).

The DNA sequence (Figure 2) predicts at least two ORFs (facing in the leftward direction in Figure 1). No ORF of significant length is found in the opposite direction. The marker rescue results and sequence changes in the two mutants show the first ORF to be the coding sequence of gene *49*: the glutamine codon CAG is changed at position 583 to a termination codon UAG in *amE727* (Figure 3) and a glycine codon GCA is changed at position 295 to AGA (arginine) in *tsC9* (data not shown).

Transcription: To distinguish between early, middle and late transcripts we compared RNA isolated at different times after wild-type or gene *49* mutant infections with RNA from a *motA* mutant, *sip1* (HOMYK, RODRIGUEZ and WEIL 1976), deficient in middle-mode transcription (BRODY, RABUSSAY and HALL 1983), and from a *33*$^-$-*55*$^-$ double mutant which is defective in late transcription and continues early and middle-mode transcription at late times (GEIDUSCHEK, ELLIOTT and KASSAVETIS 1983). Numerous experiments (most of them not shown) demonstrate that gene *49* transcription and the processing of these transcripts is unusually complex. For reasons of simplicity we show here only a few of the many results that established the transcript pattern shown in Figure 1. All other results (not shown) are completely consistent with these assignments. Also to facilitate orientation within the sequence, we show in Figure 4 some possible secondary structures of the RNA near the landmarks described below.

All gene *49* transcripts are synthesized *in vivo* in the counterclockwise (leftward) direction of the standard T4 map. No transcripts were detected with probes specific to the opposite strand (data not shown). Overlapping gene *49* transcripts are initiated at different times after infection from different promoters (Figure 1). These transcripts terminate at (or are digested up to) different positions, either at *t*, downstream of gene *49*, or farther downstream from the sequenced region. The early primary transcripts from which endo VII is presumably poorly translated are apparently unstable and subject to site-specific cleavages (Figures 1 and 5 through 8) and to subsequent degradation.

Early transcripts: These are initiated upstream of the sequenced region. Northern blots of early RNA (examples probed with pA22- and pA16-derived labeled strands are shown in Figure 5, A and B, respectively, lanes 1, 2 and 4) show a whole size range of prereplicative RNAs. The longest early RNA, approximately 2300 nucleotides, traverses the entire sequenced *Rsa*I fragment (see S1 protections, Figure 8 below). Other RNA species have a 3′ end at or near *t* (position 757). Either they were prematurely terminated or they were degraded in the 3′ to 5′ direction up to the potential hairpin (Figure 4B).

Although the early transcripts are processed and degraded, degradation is not entirely random. Weak bands appear over the diffuse background of random degradation products both in the Northern blots (Figure 5) and in S1 protection experiments (Figure 8, below).

Consistent with the lengths of these RNA species, primer extension analyses of early or *33*$^-$-*55*$^-$ double mutant RNA with primers 102 and 135 (marked in Figure 2) (data not shown) reveal several 5′ ends upstream of the region sequenced by us. The main extension product of primer "102," of approximately

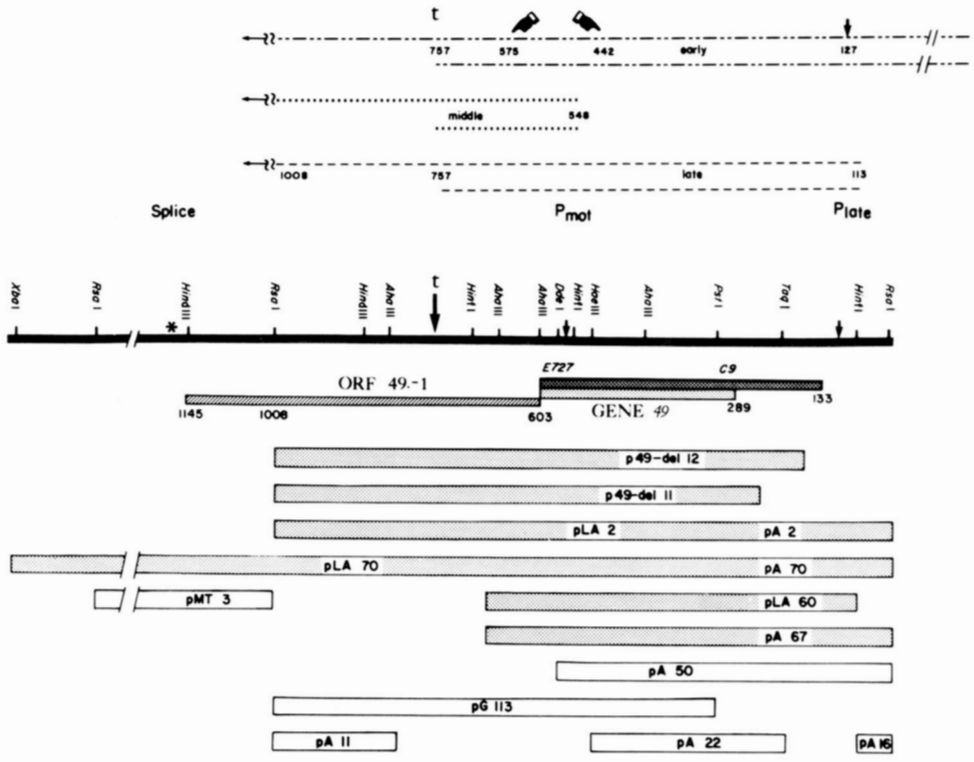


FIGURE 1.—A map of the T4 gene *49* region. The central solid line shows the restriction sites used in cloning and sequencing, the positions of the late and *motA*-dependent promoters (arrows), and an internal 3′ RNA end, *t*. The class I splice site (position 1140) of SHUB *et al.* (1988) is labeled with an asterisk. The coding regions for the 18-kD and the 12-kD gene-*49* proteins and for ORF 49.-1 are marked as differently shaded rectangles below the restriction map. ORF 49.-1 corresponds to ORF 55.13 of TOMASCHEWSKI and RÜGER (1987) and to part of *sunY* (SHUB *et al.* 1988). We have called it ORF 49.-1 because RNA splicing downstream of this ORF (SHUB *et al.* 1988; G. MOSIG, unpublished) makes the numbers of proteins encoded between genes *55* and *49* uncertain (see also Figure 9). Numbers refer to positions in the sequence shown in Figure 2. Positions 3 through 1006 bracket the *Rsa*I fragment cloned in pA2. The upper portion shows the major primary and processed transcripts identified in this work. The RNaseIII-independent cleavage site is marked by an arrow and the RNaseIII-dependent cleavage sites are marked with hands. The lower part of the figure shows major clones used for sequencing, to generate hybridization probes, and in the complementation studies. Clones pA2, pA70, pA60, pA67, p49-del11 and p49-del12 (shaded rectangles) complement gene *49* mutations. The other clones fail to complement. Construction and nomenclature of the clones is described in MATERIALS AND METHODS.

440 bases, maps a 5′ transcript end immediately downstream from the postulated (early) promoter for *pin* (SKORUPSKI *et al.* 1988). Within the sequence shown in Figure 2, primer extensions from primers 103 (Figure 6A, lanes 1 and 3, and Figure 6B), 134 (Figure 7) and 135 (data not shown) map major 5′ RNA ends at or near positions 127 to 129 and positions 442 and/or 575 (arrows or hands, respectively, in Figure 1). There is no obvious consensus sequence for early or middle T4 promoters upstream of position 127, but position 129 corresponds exactly to the preferred cleavage site UU↓AU found by CANNISTRARO, SUBBARAO and KENNELL 1986) in *E. coli* RNA. We therefore consider it likely to be a cleavage site. There is also no consensus promoter sequence upstream of 5′ ends near position 442. Since these ends are greatly reduced in T4 RNA isolated from RNaseIII-deficient cells (compare Figure 6B, lanes 1 and 2), they are also likely cleavage ends. Overexposure of the autoradiograms reveals many additional bands of varying intensities, consistent with the extensive degradation also evident in Figures 5 and 8.

In complete agreement with these assignments, summarized in Figures 1 and 2, the prereplicative transcripts, when hybridized to pA2-derived single-stranded probes, protect fragments of the expected lengths (Figure 8, lane 3) from S1 digestion. Again, the longest (1004-base) protected fragment is the one expected from unprocessed early transcripts.

S1 protection of a 498-base probe, derived from a *Rsa*I-*Hae*III fragment (Figure 8, lanes 5 and 6), confirms that, in addition to the RNA traversing the entire probe and some random degradation products, there is one major 3′ end in both prereplicative and late RNA, at or near position 757.

GOTT, SHUB and BELFORT (1986) have found, by labeling with [α^{32}P]GTP, a self-splicing group I intron which they believed to map in the T4 thioredoxin gene. Comparison of their results with ours (G. MOSIG and D. POWELL, unpublished results) indicated, however, that the 5′ splice site of this intron maps downstream of gene *49*; it is now mapped at position 1140 in Figure 2 (SHUB *et al.* 1988). Under conditions which label the group I splice site characterized by GOTT,

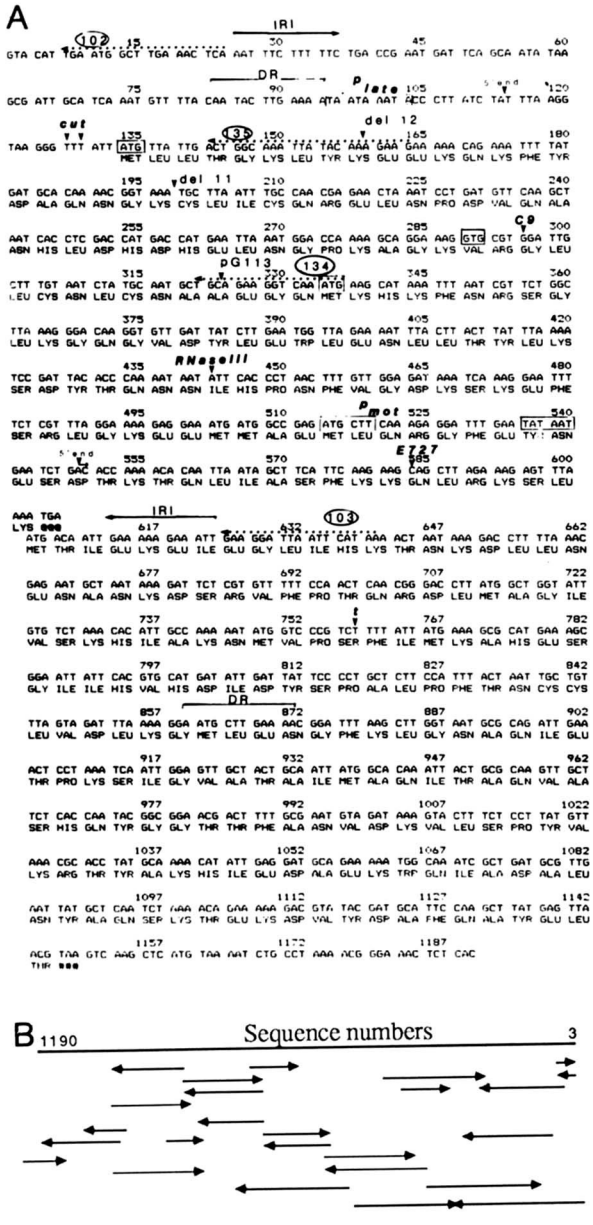


FIGURE 2.—A, The DNA sequence of gene *49* and of ORF 49.-1. The sequence was determined repeatedly from both strands, from the clones shown in Figure 1 and from additional subclones. Critical bases, around position 600 and in several other segments, were confirmed by sequencing from mRNA (Figure 3 and data not shown). The sequence agrees with that reported by TOMASCHEWSKI and RÜGER (1987) after our sequence was communicated to them. The indicated ends of deletions, discussed in the text, were confirmed by sequencing across the novel junctions. Major direct repeats (DR) and inverted repeats (IR) (in addition to those shown in Figure 4), promoters, 5′ and 3′ ends of transcripts and translation start sites, mentioned in the text, are also marked. Synthetic oligonucleotides, 102, 103, 134 and 135, used for primer extensions and sequencing are marked with dots. B, The individual sequenced segments. Sequence numbers refer to positions in A.

SHUB and BELFORT (1986), neither *in vivo* RNA hybrid selected with pA67 DNA nor RNA synthesized *in vitro* from pA67 DNA are labeled with GTP (data not shown), indicating that there is no additional group I intron in gene *49*.

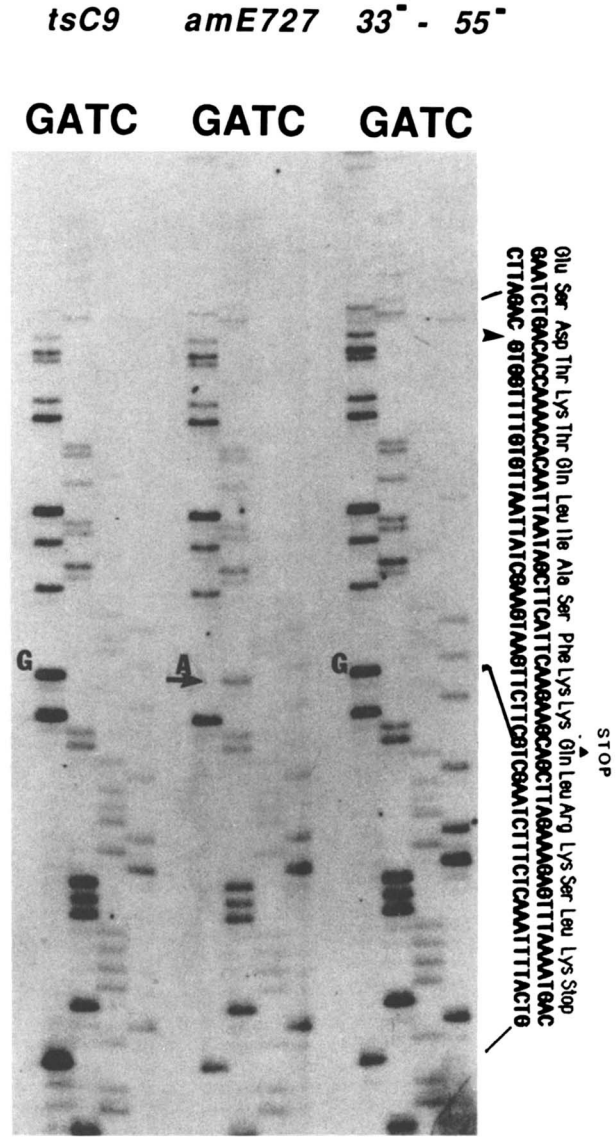


FIGURE 3.—DNA sequence obtained from *in vivo* mRNA templates, by *in vitro* extension from primer 103 (see Figure 2), showing the position of the *am* mutation *E727* and the gene-*49* stop codon. Panel 1, late *C9* RNA; panel 2, late *E727* RNA; panel 3, early $E727^{+}$ 33^{-}-55^{-} RNA. The sequence obtained from wild-type RNA or from late 33^{-}-55^{-} RNA (not shown) is indistinguishable from the $E727^{+}$ sequences shown here. The gap at position 548 (arrowhead) indicates the 5′ end of the *motA*-dependent transcripts discussed in the text. It is seen as a stop in all lanes in longer exposures (not shown).

Late transcripts: Two major, more stable gene *49* transcripts appear late after infection (Figure 5A, lanes 3, 5 and 6, open triangles) and only under 33^{+}-55^{+} conditions (compare Figure 5A, lanes 3 and 4). Primer extensions with primers 103 (Figure 6A, lane 2), 134 (Figure 7, lane 3) and 135 (data not shown) reveal a major late 5′ end at or near position 113, downstream from a late promoter (Figure 1 and underlined in Figure 4A1). These late RNAs are more stable than the early RNAs, similar to late gene-*32* RNA (BELIN *et al.* 1987). We consider it unlikely that the late transcripts are processed from the longer

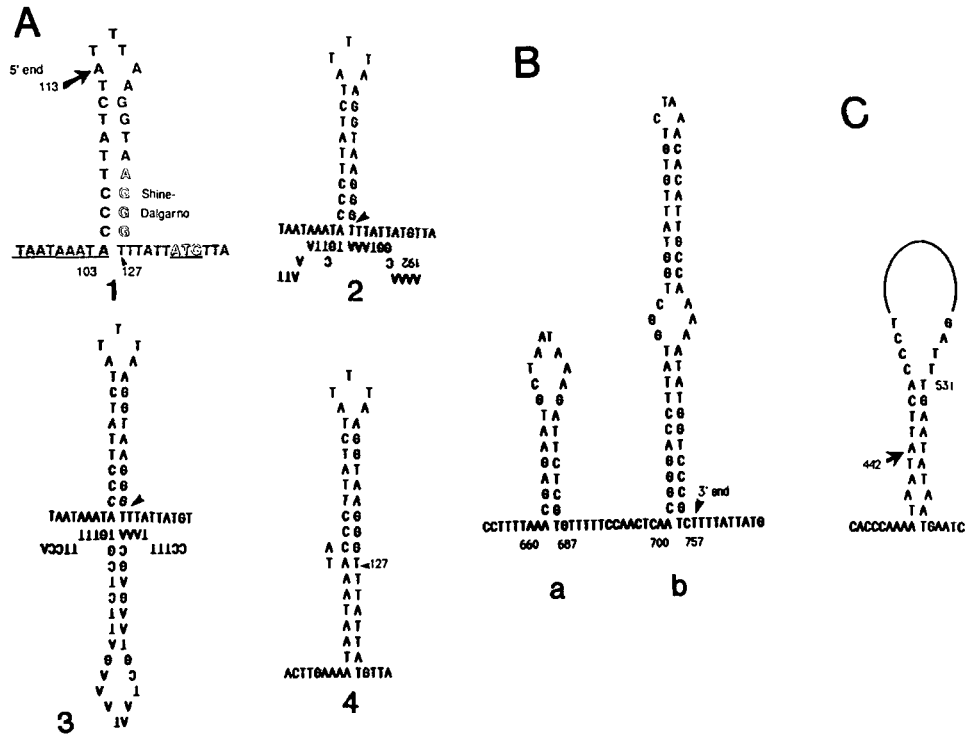


FIGURE 4.—Some predicted secondary structures in gene *49* RNA. Numbers correspond to those of the DNA sequence in Figure 2. To facilitate comparisons with Figure 2, U's are shown as T's. **A**, 1–4 Possible alternative pairings around a stem-loop that would sequester the SHINE-DALGARNO sequence of the first ribosome binding site and that might be important in generating the cleavages at positions 127 to 129. In A1, the late promoter is underlined and the 5′ end of the late transcript (which cannot form the hairpin) is indicated. **B**, Two possible hairpins that precede a strong 3′ end of early, middle and late transcripts. The position of this 3′ end is labeled "t" (for terminator) in Figures 1 and 2. From the results shown in Figure 8, we consider it likely that formation of the hairpin also stops random degradation of longer transcripts in the 3′ to 5′ direction (SCHMIDT and APIRION 1983; KING, SIRDESKMUKH and SCHLESSINGER 1986). Note that hairpin (a) is also part of structure A3. **C**, A possible stem-loop structure encompassing the RNaseIII-dependent cleavage site at position 442. No similar structure has been found surrounding nucleotides 570–575.

prereplicative transcripts, since they are not found in *33*$^-$-*55*$^-$ mutants either early or late, and since extensions from primer 102 (not shown) do not detect late (*33*$^+$-*55*$^+$-dependent) transcripts upstream of the sequenced region.

As expected, the two late transcriptions are not detected on northern blots with pA16-derived probes (Figure 5B), because pA16 maps upstream of position 113 (compare Figure 1). Like some of the early RNA species, the shorter of the two late transcripts has its 3′ end at or near position 757, at *t*, *i.e.*, after palindrome b in Figure 4B. The S1 protection of the 1004-nucleotide *Rsa*I fragment with late RNA (Figure 8, lane 4) shows the expected 895 and 644 nucleotide fragments. In addition, the 630-nucleotide fragment suggests that the 5′ end of the early RNA cleaved at position 127 is stable and/or that the late RNA can also be cleaved at the same or adjacent positions. Protection of the 498-nucleotide *Rsa*I-*Hae*III RNA by late T4 RNA (Figure 8, lane 6) confirms these assignments.

Both late transcripts contain the entire coding sequence of gene *49*. Since they begin at position 113, they cannot fold into any of the putative complex structures shown in Figure 4A, 1 through 4.

Middle transcripts. There is a sequence inside gene *49* (boxed P_{mot} in Figure 2) that resembles the consensus sequence for *motA*-dependent promoters (BRODY, RABUSSAY and HALL 1983; GUILD *et al.* 1988). Primer extensions and sequencing from primer 103 detected a major 5′ RNA end downstream of this sequence at position 548 (Figures 3 and 6). This 5′ end is absent in *motA*$^-$ infections (compare Figure 6C, lanes 1 and 2 with 3 and 5), strongly suggesting that transcripts are initiated from P_{mot}. Possibly, some 5′ ends could be also derived by cleavage from the larger early transcripts, but none of these are greatly reduced in *motA*$^-$ infections (data not shown). The 5′ ends at position 548 are unstable, *i.e.*, they are barely detectable at late times (Figures 6A, lane 2; 6C, lane 4; and 7, lane 4), when early and middle mode transcription has ceased due to phage-evoked changes in RNA polymerase (RABUSSAY 1983). This is in contrast to the *uvsY* transcript ends, which are also *motA*-dependent but are stable (GRUIDL and MOSIG 1986).

Proteins: The gene-*49* product has not been detectable as a protein band in denaturing polyacrylamide gels, presumably because it is present in T4-infected cells in very low amounts (KEMPER and GARABETT 1981). Therefore, we looked for gene-*49* prod-

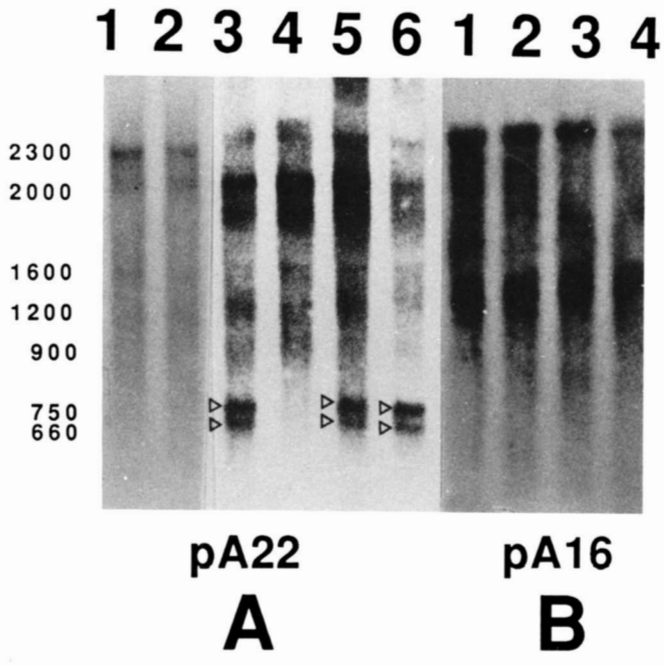


FIGURE 5.—Northern blots of unlabeled *in vivo* RNA isolated from infected cells. "Early" RNA was isolated 10, "late" RNA 30 min after infection of *E. coli* B at 30°. The RNA blots were then probed with ^{32}P-labeled "antisense" RNA synthesized with T7 RNA polymerase *in vitro*, in panel A from pA22, in panel B from pA16. Lane 1, wild-type early RNA. Lane 2, *33*$^-$-*55*$^-$ early RNA. Lane 3, wild-type late RNA. Lane 4, *33*$^-$-*55*$^-$ late RNA. Lane 5, *E727* late RNA. Lane 6, *C9* late RNA. Panel B is overexposed to show that the two late transcripts (open triangles) which start at position 113, *i.e.*, downstream of pA16, are not recognized by this probe. DNA sizes (leftmost lane) were estimated from DNA size standards run in parallel.

ucts in appropriate plasmid-bearing cells. If protein synthesis starts at the first AUG codon (position 133 in Figure 2) preceded by a SHINE-DALGARNO (1974) sequence, gene *49* is predicted to code for a peptide of 157 amino acids, or 18.1 kD, referred to as the 18-kD peptide. This first ribosome binding site is, however, part of a palindrome which could fold, in gene *49* transcripts initiated from vector promoters as well as in the long early gene *49* transcripts, in several possible ways (Figure 4A). Such folding is likely to inhibit or reduce translation initiation, as shown in other similar situations (ISERENTANT and FIERS 1980; MACDONALD, KUTTER and MOSIG 1984; MCPHEETERS *et al.* 1986). Analogous to restarts after premature termination due to nonsense mutations, an internal GUG codon at position 289 could initiate a smaller peptide of 105 residues in the same reading frame (Figures 1 and 2), referred to as the 12-kD peptide below.

Figure 9 shows that plasmid constructs containing the entire *49* coding and complementing sequence, inserted downstream from the same vector promoter, produce the two expected peptides (arrowheads) of approximately 18 and 12 kD. In addition, a 20-kD peptide (dot), predicted from the downstream ORF 49.-1 (Figures 1 and 2), can be seen. Relative proportions of the two gene-*49* peptides vary in different clones and under different growth conditions, consistent with the idea that initiation from the first AUG would override initiation from the internal GUG site (GOLD *et al.* 1981; STORMO 1986; compare lanes 1 and 3 in Figure 5). The ratio of 18-kD to 12-kD peptides is higher at 37° than at 25° (data not shown), consistent with estimated stabilities of the secondary structures (Figure 4A). At 42° both peptides are rapidly degraded (data not shown), precluding a more quantitative analysis of hairpin formation and initiation from the first AUG.

In addition, initiation from the first AUG is probably enhanced by components encoded downstream of gene *49*. Transcripts from pLA2 and pLA70 both contain identical (vector and T4) sequences upstream of gene *49*, but different downstream sequences. They produce different proportions of the two peptides (compare lanes 1 and 3 of Figure 8). This aspect will be considered further in the Discussion.

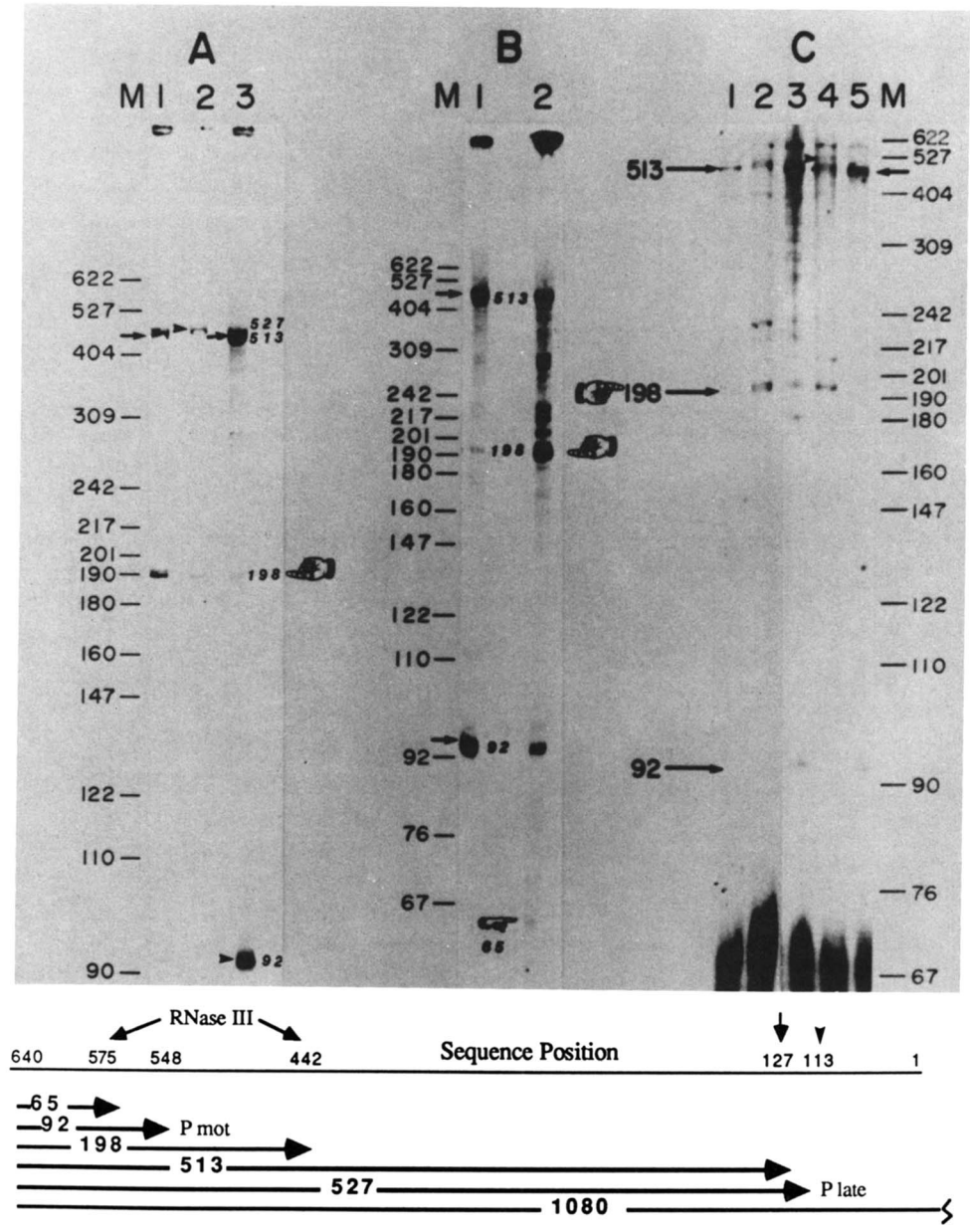
Complementation: To eliminate the possibility that fusion with a downstream sequence (*e.g.*, by RNA splicing or ribosomal frameshifting) is required for endo VII function, we deleted various segments from the cloned T4 *Rsa*I fragment by cutting with restriction enzymes or by *Bal*31 digestion (Figure 1), confirmed the deletion end points by sequencing and tested the resulting clones for complementation of gene *49* mutants and for synthesis of expected proteins.

The T4 inserts in pA60, pA67, pA70 and pA2 (shaded in Figure 1) complement both gene-*49* mutants at temperatures between 25° and 45° and, at temperatures between 30° and 40° only, a phage T7 gene-*3* amber mutant which is defective in the equivalent T7 enzyme, endonuclease I (DEMASSY *et al.* 1984; DEMASSY, WEISBERG and STUDIER 1987; DICKIE, MCFADDEN and MORGAN 1987). Since the pA60 insert can code for both the 18-kD and the 12-kD peptides but contains few additional sequences, it defines gene *49*.

Plasmids p49-*del11* and p49-*del12* (Figure 1, deletion end points indicated in Figure 2), which cannot synthesize the 18-kDa peptide but can program the GUG-initiated 12-kD peptide, only partially complement both T4 gene-*49* mutants. (They do not complement the T7 gene-*3* mutant.) We have excluded the possibility that this is false complementation due to marker rescue, because plasmid pG113, from which the wild-type allele of *amE727* is rescued as well as from p49-*del*11, does not complement. This aspect is further discussed below.

DISCUSSION

Complementation, marker rescue and sequencing of two gene-*49* mutations establish that gene *49* (endo VII) corresponds to the first of two ORFs that we



Predicted Primer Extension Products

FIGURE 6.—Extensions from primer 103 (starting at position 640) to map 5′ ends in gene *49* transcripts. (Sequences determined by extensions from the same primer are shown in Figure 3.) M, size markers (in nucleotides). As indicated in the lower diagram, all 5′ transcript ends expected from the promoters and RNA cleavage sites mentioned in the text and in Figure 1 can be seen: 92-nucleotide products are expected for the 5′ starts of the *motA*-dependent transcript, 65–70- and 198-nucleotide products for the RNaseIII-dependent cleavage sites at positions 570–575 and 442 respectively, 513-nucleotide products (arrows) for the RNaseIII-independent cleavage site at position 127, and 527-nucleotide products (arrowheads) for the 5′ starts of the late transcript. The extension product corresponding to the early *pin* promoter remains close to the well. The ratio of the 65–70- and 198-nucleotide extension products corresponding to the RNase III-dependent cleavages varies with time after infection. This variation is still under investigation. A, Extension products from early wild-type RNA (lane 1), late wild-type RNA (lane 2) and late *33*$^-$-*55*$^-$ RNA (lane 3). Since a 5′ end (revealed by a 92-nucleotide extension product) maps downstream from a consensus *mot* box (BRODY, RABUSSAY and HALL 1983; see Figure 2) and is not found in T4 *motA*$^-$ infections (C, lanes 1 and 2), it is probably the start of a middle transcript. The 527-nucleotide extension products (arrowheads) corresponding to the late promoter are seen only in late *33*$^+$-*55*$^+$ RNA (lane 2; see also C, lane 4). In contrast, the putative cleavage site preceding the first AUG at position 127 exists in early RNA and is enhanced in *33*$^-$-*55*$^-$ mutants which continue to synthesize prereplicative RNA. Extension products from primer 134 and 135 (not shown) confirm these early and late 5′ ends. B, Comparison of extension products from late *33*$^-$-*55*$^-$ RNA isolated after infection of RNaseIII-deficient (lane 1) or RNase III-proficient (lane 2) bacteria. The cleavage site at position 127 (513-base extension product) is unaffected. The 5′ start of the *motA* transcript appears elevated in RNaseIII$^-$ hosts, probably because degradation farther downstream (positions 570–575) is reduced (C and data not shown). Many additional minor bands, also reduced in this host, are probably due to further, nonrandom degradation (CANNISTRARO, SUBBARAO and KENNELL 1986). C, Extension products from early *motA*$^-$ (lane 1), late *motA*$^-$ (lane 2), early *C9 motA*$^+$ (lane 3), late *C9 motA*$^+$ (lane 4) and early *33*$^-$-*55*$^-$ *motA*$^+$ (lane 5) RNA. Early wild-type *motA*$^+$ RNA from the same experiment (not shown) gave the same pattern as lane 5. Compare also panel A, lane 1.

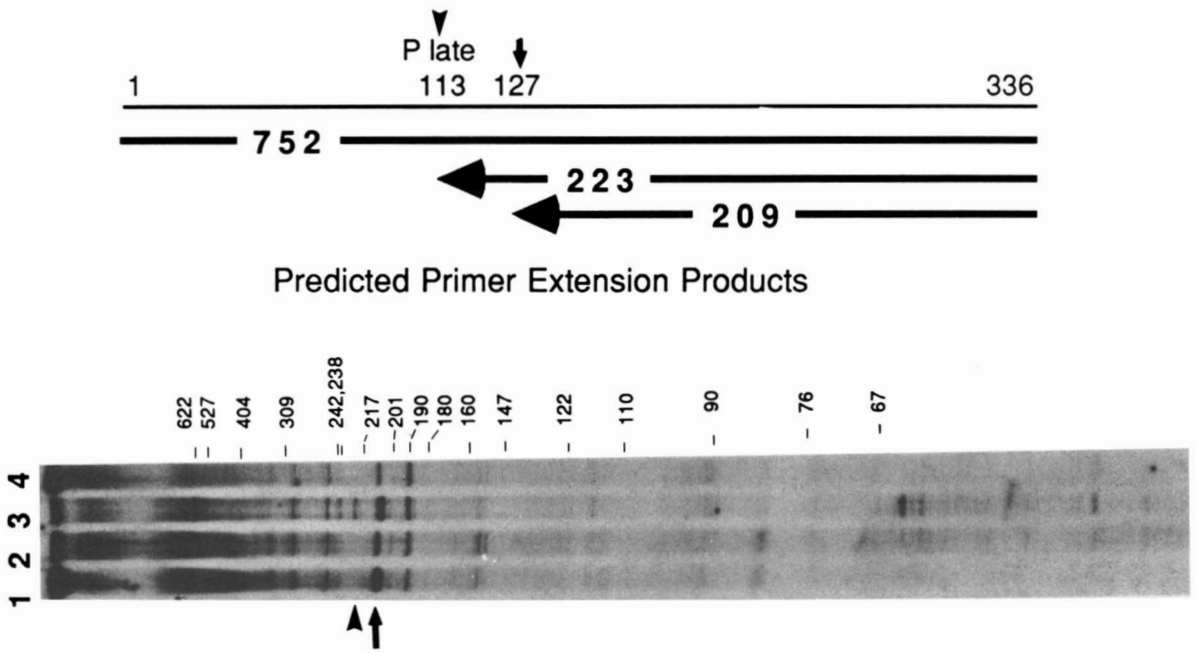


FIGURE 7.—Extensions from primer 134 (starting at position 336) to map 5′ ends in gene-*49* transcripts. Lane 1, early wild-type RNA; lane 2, early *33⁻-55⁻* RNA; lane 3, late wild-type RNA; lane 4, late *33⁻-55⁻* RNA. As expected, the 223 nucleotide-extension product predicted for late RNA (arrowhead) is seen only in lane 3. (The band above it, while enhanced in late RNA, is also present in the three other lanes.) The 207–209-nucleotide extension products predicted for cleaved transcripts (arrow) are seen in all lanes. The data suggest that during infection, cleavages shift from position 129 to 127 of the DNA sequence of Figure 2. Positions on the gel and corresponding lengths of size standards are marked. The right panel shows a diagram of the primer extensions.

have sequenced between 45 and 46 kb on the standard T4 map (Figures 1 and 2). We notice that this gene is flanked by direct and inverted repeats (Figure 2), which could suggest a transpositional origin.

Deletion analyses together with complementation studies establish that gene *49* specifies at least two in-frame overlapping peptides of 18 and 12 kD, respectively. Since the apparent molecular mass of native endo VII is estimated to be between 40 and 46 kD (KEMPER and GARABETT 1981; B. KEMPER, personal communication), the enzyme must be multimeric and might exist in different forms. The complementation results suggested the possibility that the 12-kD protein has endo VII activity. However, in collaboration with RUTH EHRING and BÖRRIS KEMPER (University of Köln) we have recently synthesized these products *in vitro* in a coupled transcription-translation system and tested them for endo VII activity as defined by cleavage of very fast sedimenting T4 DNA (KEMPER and GARABETT 1981). The results, to be reported in a forthcoming paper, show that DNA from pLA60 (as well as pLA70 and pLA2) can program functional endo VII activity, but that DNA from p49*del11*, which can only program the 12-kD peptide, does not. These findings imply that the 12-kD protein alone does not suffice for endo VII activity and/or that the 12-kD product may have a yet unknown additional function. One intriguing possibility is that the 12-kD protein can bind to cruciforms without cleaving them. Such an activity has been isolated recently from rat liver nuclei (BIANCHI 1988).

Gene *49* is transcribed both early (prereplicative) and late (postreplicative), in the same (counterclockwise) direction, albeit from different promoters. Long, early transcripts, initiated in T4 upstream of gene *pin* (SKORUPSKI *et al.* 1988) or, in plasmid clones, from upstream vector promoters, are apparently inefficient templates for endo VII, whose activity is greatly reduced in gene-*55* mutants (KEMPER and GARABETT 1981) and shows little accumulation in overexpression vectors (Figure 9), even when the transcripts were grossly overproduced (data not shown). We speculate that the long transcripts can fold into several alternative structures (some shown in Figure 4A) which can sequester, perhaps with the help of proteins, the first SHINE-DALGARNO sequence (ISERENTANT and FIERS 1980). Translation of two other "late" T4 genes (*soc* = small outer capsid protein and *e* = lysozyme) which are transcribed early and late, albeit from different promoters, is inhibited in the early messages by a similar mechanism (MACDONALD, KUTTER and MOSIG 1984; MCPHEETERS *et al.* 1986).

In addition, and perhaps related to the poor translation, these early transcripts are unstable (Figures 5 through 8). They are susceptible to cleavages at or near positions 442 and 575, probably by RNaseIII acting directly or indirectly. Position 442 can exist in a double-stranded structure (Figure 4C), often formed at RNaseIII cleavage sites. As in many other prokaryotic transcripts (SCHMIDT and APIRION 1983; KING, SIRDESKMUKH and SCHLESSINGER 1986; CANNISTRARO, SUBBARAO and KENNELL 1986; KENNELL

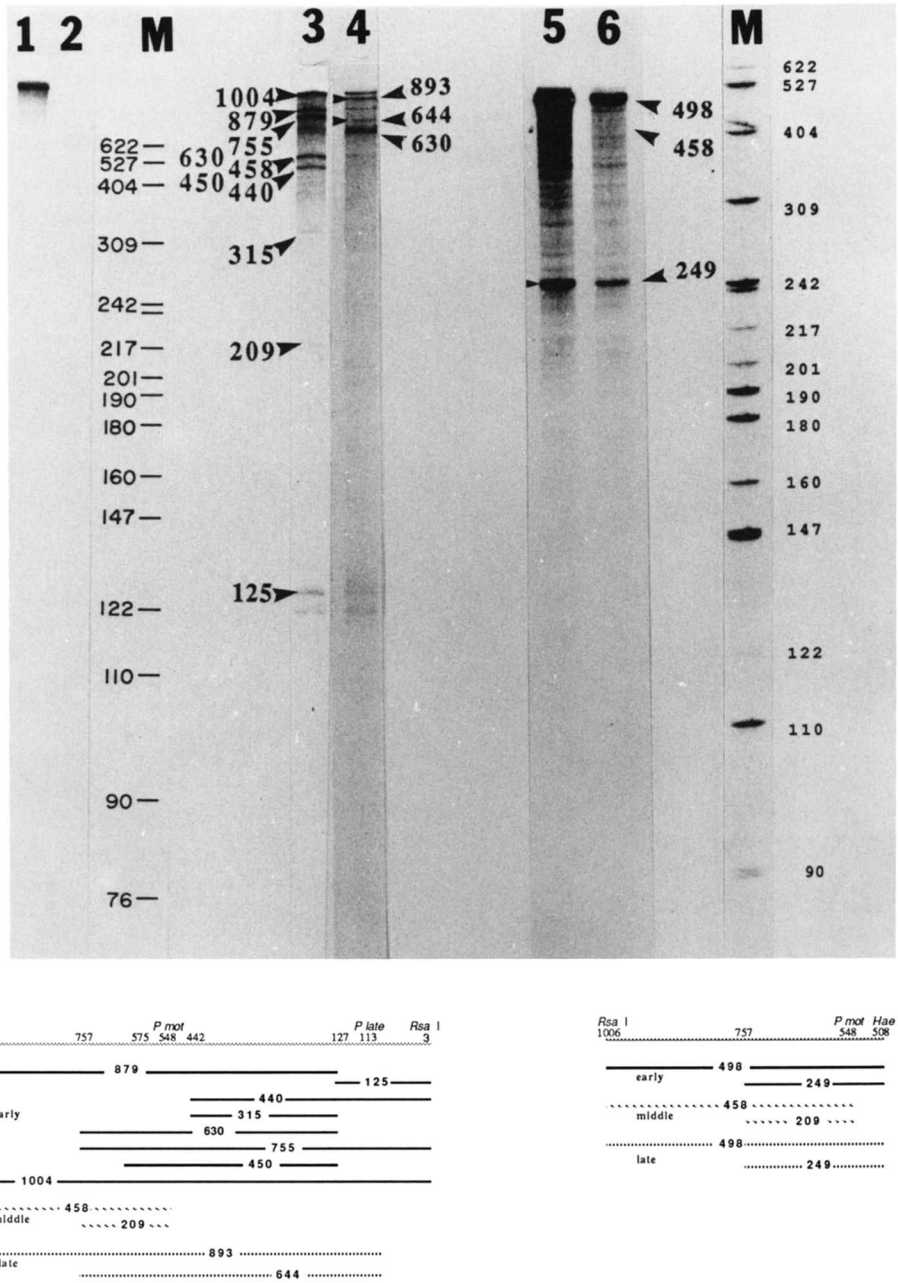


FIGURE 8.—Protection against S1 nuclease digestion of ^{32}P-labeled antisense RNA, synthesized *in vitro*, by T4 RNA isolated from infected cells. The diagrams at the bottom show the positions of start, termination and cleavage sites in the sequence (thin numbers) and the expected sizes (bold numbers) of protected fragments seen in lanes 3 through 6. Lane 1, the 1036-nucleotide probe (the 1004-nucleotide *Rsa*I fragment + 32 vector nucleotides); lane 2, the same probe after S1 digestion; lane 3, probe fragments protected from S1 digestion by early wild-type RNA. Lane 4, fragments protected by late wild-type RNA. The 893- and 644-nucleotide fragments are expected and found only in late RNA. The top band in lane 3, the fully protected probe, is slightly smaller than the probe before hybridization in lane 1, since only 1004 bases correspond to T4 DNA. Lane 5, fragments from a 498-base RNA probe corresponding to a *Rsa*I-*Hae*III fragment (positions 1006 to 508) protected from S1 digestion by early T4 RNA; lane 6, fragments from the same probe protected by late T4 RNA. The weak band corresponding to 209 bases from a *motA*-dependent transcript is visible only in longer exposures, consistent with the instability of these transcripts. M, size markers.

1986; PORTIER *et al.* 1987), RNaseIII-dependent cleavages could facilitate subsequent degradation in the 3′ or 5′ direction by unknown nucleases.

In addition, there are RNase III-independent cleavages between the SHINE-DALGARNO sequence and the first AUG codon at or near positions 127 to 129. One of these corresponds to the preferred UU↓AU cleavage site in *E. coli lac* transcripts found by CANNISTRARO, SUBBARAO and KENNELL (1986). These authors have presented a statistical argument why such cleavages occur frequently near ribosome binding sites. Phage-encoded nuclease(s) or other proteins might be responsible for the apparent 2-nucleotide shift of the cleavage site at later times (Figure 7).

We consider the additional possibility that these cleavages might also be related to the putative com-

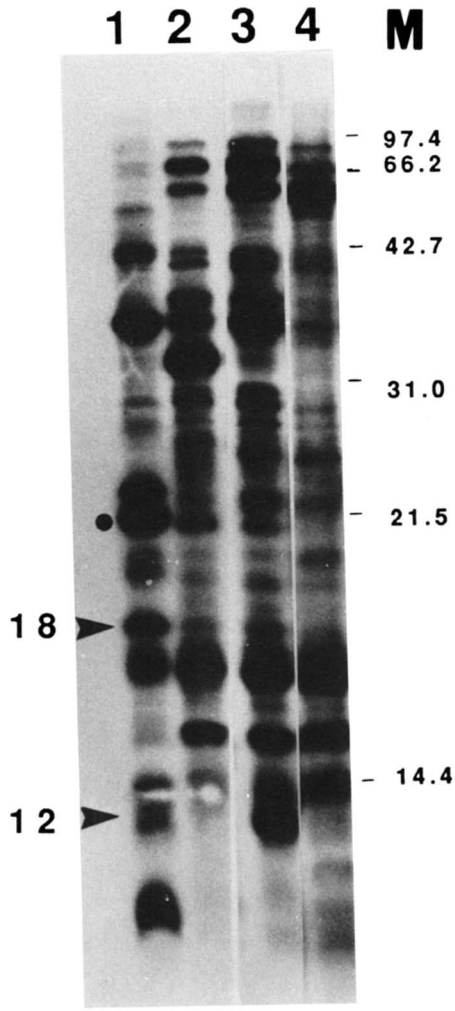


FIGURE 9.—Autoradiograms of SDS-urea 18% polyacrylamide gels (NASH *et al.* 1987) of proteins synthesized in M5219 cells. Size standards in kD (M) were stained with Coomassie blue. Lane 1, 30 min after induction at 42° of pLA70; lane 2, pLA70 grown at 27° (uninduced); lane 3, pLA2 30 min after induction at 42°. (In pLA2, a part of ORF 49.-1 is fused with vector sequences to give a peptide of similar size but less stability as compared with the 20-kD peptide of pLA70.) Lane 4, 30 min after induction of pPLc2833 without T4 inserts. Differences in larger proteins, due to different sequences downstream from gene *49* or to different temperatures, are presently under investigation.

plex folding patterns drawn in Figure 4A. The stem-loop that sequesters the ribosome binding site might become part of a Y-like structure (Figure 4A, 2) or, alternatively, a cruciform (Figure 4A, 3) by pairing with the small stem-loop (a), preceding the putative termination stem-loop (b), shown in Figure 4B. If endo VII (or the 12-kDa protein) binds to RNA, as it does to DNA, and facilitates their cleavage, it might autoregulate translation of the 18-kD peptide. Secondary RNA structures may also act as ribozymes (CECH and BASS 1986), facilitating the extensive cleavage and degradation that we find. These possibilities are being further investigated.

In contrast, the late transcripts whose 5′ ends are at position 113 cannot form the critical hairpin (Figure 4A) and are, therefore, free to initiate the 18-kD peptide. These late transcripts are more stable than the early ones, perhaps, in part, because they are more efficiently translated. In addition, a T4-encoded factor may protect them from degradation, as has been postulated for late gene *32* transcripts (BELIN *et al.* 1987).

The relative proportions of the two gene-*49* peptides made from different clones (Figure 9, compare lanes 1 and 3; and data not shown) are consistent with current ideas that initiation from the internal GUG site is more frequent when initiation from sequences further upstream is inhibited (ISERENTANT and FIERS 1980; GOLD *et al.* 1981; STORMO 1986). Our preliminary results also suggest that regulation of translation at the first gene-*49* ribosome binding site may involve other components in addition to the palindrome mentioned before. Transcripts from pLA2 and pLA70 both contain the same upstream sequences, but pLA70 produces more of the 18-kD protein than pLA2 or pLA60 (which behaves like pLA2, data not shown). In contrast to pLA2 and pLA60, pLA70 contains a complete ORF 49.-1. If, as we proposed, one or both products of gene *49* affect folding and/or cleavage near the first ribosome binding site, then interactions with ORF 49.-1 (discussed below) might alter this function.

The late gene-*49* promoter TAATAAATA (Figures 3 and 4A) upstream of the 5′ end of the late transcripts (Figures 6 and 7) deviates from the consensus sequence TATAAATA previously considered to be invariant (CHRISTENSEN and YOUNG 1983; ELLIOTT and GEIDUSCHEK 1984). Similar deviations have now been found in other late T4 promoters, *i.e.*, in a minor promoter upstream of gene *e* (MCPHEETERS *et al.* 1986), in promoters for genes *16* and *17* (D. POWELL, J. FRANKLIN, F. ARISAKA and G. MOSIG, in preparation), in a postulated promoter for the lysis gene *t* (RIEDE 1987) and in synthetic promoters that function with late modified T4 RNA polymerase *in vitro* (KASSAVETIS, ZENTNER and GEIDUSCHEK 1986).

The distance between the promoter sequence and the 5′ end is larger than in the late T4 promoters described at first (CHRISTENSEN and YOUNG 1983). Possibly, a few nucleotides are cleaved *in vivo* from the late transcript. Alternatively, the 5′ ends are true start sites and the distance between promoters and transcript starts is more variable in T4 late promoters than previously thought.

We searched the Protein Identification Resource (PIR) data bank with several programs (DOOLITTLE 1986), expecting some similarity with T7 endonuclease I (gp3), because T7 gp3 and T4 gp49 complement each other (DEMASSY *et al.* 1984 and our present results). Only the Fast P program (LIPMAN and PEARSON 1985) found any similarity with T7 endonuclease I (DUNN and STUDIER 1981).

This similarity is, however, limited to the central

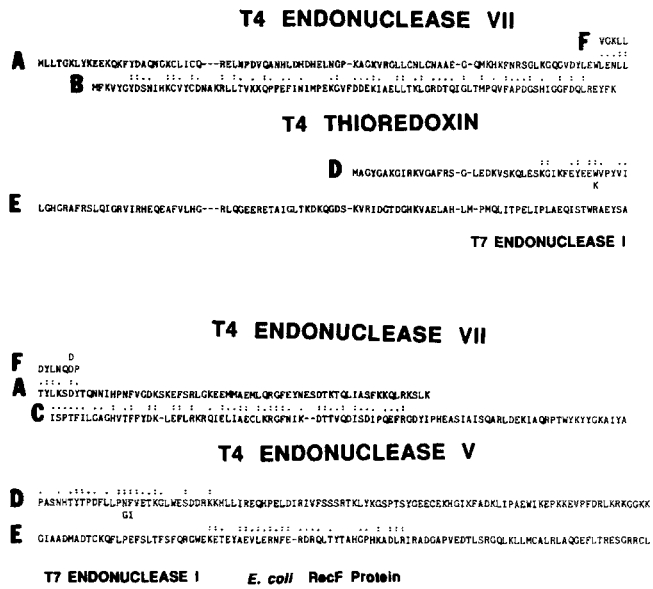


FIGURE 10.—Alignment of the deduced amino acid sequence of T4 endo VII (A) with that of other proteins found with the "Fast P" program of LIPMAN and PEARSON (1985). B = complete T4 thioredoxin; C = complete T4 endonuclease V; D = N-terminal segment of T7 endonuclease I. E = internal segment of *E. coli RecF* protein. F = internal segment of bovine endonuclease. The sequences are aligned (colon = identity, period = similarity) only with the endo VII sequence.

segments that give similar predicted hydropathicity profiles (not shown). Interestingly, there is also a short span of similarity with bovine endonuclease in this region. In the crystal structure of the latter enzyme (OEFNER and SUCK 1986), this region corresponds to an α-helix underneath the active site.

The "Fast P" search also revealed intriguing similarities with other T4 proteins. As shown in Figure 10, the N-terminal portion of endo VII resembles the T4 thioredoxin (LEMASTER 1986) and the C-terminal portion resembles T4 endonuclease V (VALERIE *et al.* 1986). The "thioredoxin domain" might fold into a metal-binding finger (MILLER, MCLACHLAN and KLUG 1985; BERG 1986). Since thioredoxin of *E. coli* serves in assembly of T7 DNA polymerase (MODRICH and RICHARDSON 1975) and of filamentous phages (FULFORD, RUSSEL and MODEL 1986), it is tempting to speculate that this domain might facilitate interactions of endo VII with itself or other proteins or membrane adhesion sites (LUNN and PIGIET 1986).

Additional similarities suggest that the C-terminal "domain" is important in binding to DNA. It resembles a flexible region in the DNA binding cleft of *E. coli* DNA polymerase I (OLLIS *et al.* 1985), which, in turn, resembles a segment of the DNA binding subunit of T4 topoisomerase (HUANG 1986) and a segment of the *E. coli* recombination RecF protein (BLANAR *et al.* 1984). It remains to be seen whether the latter similarity gives a clue to the unknown *RecF* function. Another segment of RecF protein resembles ORF 49.-1, discussed below.

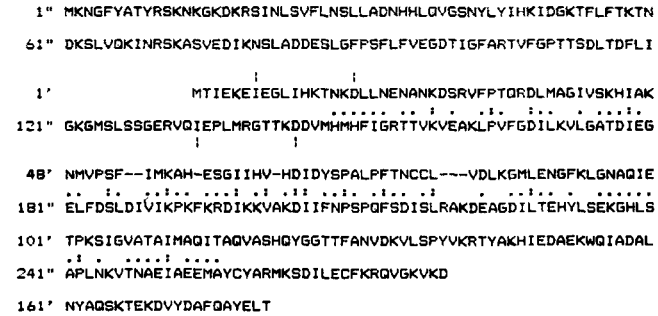


FIGURE 11.—Similarity of the deduced amino acid sequence of ORF 49.-1 with the phage λ *RexA* protein, determined as in Figure 10.

The termination codon of gene *49* overlaps the initiation codon of a downstream ORF, 49.-1. It is preceded by a (poor) ribosome binding site which might be activated by coupling to translation of gp49. Similar coupling has been observed or postulated in other transcription units (OPPENHEIM and YANOFSKY 1980; GRAM *et al.* 1984; GRAM and RÜGER 1985; BROIDA and ABELSON 1985; VALERIE *et al.* 1986; MCGRAW, MINDICH and FRANGIONE 1986; DUNN and STUDIER 1981; for review see STORMO 1986) and in some cases has been interpreted in terms of physiological interactions of the corresponding proteins. Since ORF 49.-1 might also be translated from a *motA*-dependent transcript (Figure 1), its expression might be semi-independent of that of gene *49*. RNA splicing near its termination codon makes this ORF also part of a larger ORF, called *sunY*, with an unknown function (SHUB *et al.* 1988).

The deduced amino acid sequence of ORF 49.-1 resembles, among others, the phage λ RexA protein (Figure 11), suggesting that the two proteins might have functional similarity. We find this noteworthy, because our previous results (MOSIG, SHAW and GARCIA 1984) suggested that λ lysogens interfere with T4 *rII* mutants during the recombination-dependent mode of DNA replication and that the interference involves gp49 (see Introduction), and because the *rII* proteins function in DNA replication (SINGER, SHINEDLING and GOLD 1983). The λ Rex proteins are through to function in DNA replication and recombination of lambdoid phages (TOOTHMAN and HERSKOWITZ 1980). They and the λ *cI* repressor are the only λ-encoded proteins expressed in the lysogenic state. The λ *rex* genes exhibit a complex control (LANDSMANN, KROGER and HOBOM 1982) that has some aspects in common with the control of the gene *49* region described here. They are responsible (GUSSIN and PETERSON 1972) for the classical yet poorly understood interference of λ lysogens with growth of T4 *rII* mutants (BENZER 1957).

We speculate that the λ RexA protein might compete with the T4 ORF 49.-1 (or *sunY*) product and sequester gp49 in a nonproductive complex, and that this sequestering becomes limiting for growth of the

C9 mutant because endo VII is partially defective in this mutant at permissive temperatures (KEMPER and GARABETT 1981; MOSIG, SHAW and GARCIA 1984). Since *C9* also interacts with *rII* mutations and both corresponding proteins are thought to be part of a larger membrane-associated recombination-replication complex (MOSIG 1983a; MOSIG *et al.* 1979; MOSIG, SHAW and GARCIA 1984; MOSIG 1987), such competitive abortive interactions could explain the complex phenotypes of these mutations. Competition of λ RexA proteins with the putative T4 analogue ORF 49.-1 (or gp*sunY*) could also explain the lethality of wild-type T4 growth when the λ Rex proteins are overexpressed (SHINEDLING, PARMA and GOLD 1987).

Three other observations made during the course of these studies are relevant to this working hypothesis for the role of ORF 49.1-: (1) induction of gene *49* in expression vectors is lethal to the host cells only when a complete ORF 49.-1 is present in the same clone; (2) clones in which ORF 49.-1 is expressed are extremely unstable; and (3) on such strains, wild-type T4 has an RII plaque type.

Since the ORF 49.-1 protein also shows similarity with T7 genes *5.5* and *5.7*, which are required for growth of T7 in λ lysogens, we consider the more general implication that mutual exclusion of viruses (and perhaps plasmids) can occur by competitive interference with recombination complexes (TOOTHMAN and HERSKOWITZ 1980), which are probably of general importance in the initiation of DNA replication in many systems (for review see MOSIG 1983a; MOSIG 1987). Experiments are in progress to define the function of ORF 49.-1 and to test these more general implications.

In any case, we consider it likely that the complex interwoven transcriptional and translational controls of gene *49* and ORF 49.-1 are important to achieve an optimal balance of DNA replication, recombination and packaging during the development of this virus (MOSIG 1987; MOSIG and EISERLING 1988).

This work was supported by U.S. 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. We thank RUTH EHRING, PETER STARLINGER, FRANK STAHL and JEFF FRANKLIN for comments and suggestions for the manuscript, CINDY YOUNG for her expertise and patience in processing it and NANCY CANAN for photography. We thank STEPHEN LLOYD and the Vanderbilt Toxicology Center for synthesis of oligonucleotides, WALTER FIERS and ERIC REMAUT for their expression system, BOB WEISBERG for the T7 strains and the T7 gene-*3* clone, DAVID SCHLESSINGER for the *rnc*⁻ mutant, the National Biomedical Research Foundation of Georgetown University for use of their databank and computer resources, and JEFF FRANKLIN for help with the search.

These data will appear in the EMBL/GenBank/DDBJ Nucleotide Sequence Databases under the accession number X12629.

LITERATURE CITED

ADAMS, M. H., 1959 *Bacteriophages.* Interscience Publishers, New York.

BELIN, E. A., P. MUDD, Y. Y.-Y. PRENTKI and H. M. KRISCH, 1987 Sense and antisense transcription of bacteriophage T4 gene *32*. J. Mol. Biol. **194:** 231–243.

BENZER, S., 1957 The elementary units of heredity. pp. 70–93. In: *The Chemical Basis of Heredity*, Edited by B. GLASS. Johns Hopkins Press, Baltimore.

BERG, J. M., 1986 Potential metal-binding domains in nucleic acid binding proteins. Science **232:** 485–487.

BIANCHI, M. E., 1988 Interaction of a protein from rat liver nuclei with cruciform DNA. EMBO J. **7:** 843–849.

BLANAR, M. A., S. J. SANDLER, M.-E. ARMENGOD, L. W. REAM and A. J. CLARK, 1984 Molecular analysis of the *recF* gene of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **81:** 4622–4626.

BRODY, E., D. RABUSSAY and D. H. HALL, 1983 Regulation of transcription of prereplicative genes. pp. 174–183. In: *Bacteriophage T4*, Edited by C. K. MATHEWS, C. K. KUTTER, G. MOSIG and P. B. BERGET. American Society for Microbiology, Washington, D.C.

BROIDA, J., AND J. ABELSON, 1985 Sequence organization and control of transcription in the bacteriophage T4 tRNA region. J. Mol. Biol. **185:** 545–563.

CANNISTRARO, V. J., M. N. SUBBARAO and D. KENNELL, 1986 Specific endonucleolytic cleavage sites for decay of *Escherichia coli* mRNA. J. Mol. Biol. **192:** 257–274.

CECH, T. R., and B. L. BASS, 1986 Biological catalysis by RNA. Annu. Rev. Biochem. **55:** 599–629.

CHEN, E., and P. H. SEEBURG, 1985 Supercoil sequencing: a fast and simple method for sequencing plasmid DNA. DNA **4:** 165–170.

CHRISTENSEN, A. C., and E. T. YOUNG, 1983 Characterization of T4 transcripts. pp. 184–188. In: *Bacteriophage T4*. Edited by C. K. MATHEWS, E. M. KUTTER, G. MOSIG and P. B. BERGET. American Society for Microbiology, Washington, D.C.

CLARK, R. W., G. H. WEVER and J. S. WIBERG, 1980 High-molecular-weight DNA and the sedimentation coefficient: a new perspective based on DNA from T7 bacteriophage and two novel forms of T4 bacteriophage. J. Virol. **33:** 438–448.

CONKLING, M. A., AND J. W. DRAKE, 1984 Isolation and characterization of conditional alleles of bacteriophage T4 genes *uvsX* and *uvsY*. Genetics **107:** 505–523.

DEMASSY, B., F. W. STUDIER, L. DORGAIL, E. APPELBAUM and R. A. WEISBERG, 1984 Enzymes and sites of genetic recombination: studies with gene-*3* endonuclease of phage T7 and with site affinity mutants of phage lambda. Cold Spring Harbor Symp. Quant. Biol. **48:** 715–726.

DEMASSY, B., R. A. WEISBERG and F. W. STUDIER, 1987 Gene *3* endonuclease of bacteriophage T7 resolves conformationally branched structures in double-stranded DNA. J. Mol. Biol. **193:** 359–376.

DICKIE, P., G. MCFADDEN and A. R. MORGAN, 1987 The site-specific cleavage of synthetic Holliday junction analogs and related branched DNA structures by bacteriophage T7 endonuclease I. J. Biol. Chem. **262:** 14826–14836.

DOOLITTLE, R. F., 1986 *Of URFs and ORFs: A Primer on How to Analyze Derived Amino Acid Sequences.* University Science Books, Mill Valley, Calif.

DRESSLER, D., and H. POTTER, 1982 Molecular mechanisms in genetic recombination. Annu. Rev. Biochem. **51:** 727–761.

DUNN, J. J., and F. W. STUDIER, 1981 Nucleotide sequences from the genetic left end of bacteriophage T7 DNA to the beginning of gene *4*. J. Mol. Biol. **148:** 303–330.

ELLIOTT, T., and E. P. GEIDUSCHEK, 1984 Defining a bacteriophage T4 late promoter: absence of a "−35" region. Cell **36:** 211–219.

EPSTEIN, R. H., A. BOLLE, C. M. STEINBERG, E. KELLENBERGER, E. BOY DE LA TOUR, R. CHEVALLEY, R. S. EDGAR, M. SUSMAN, G. H. DENHARDT and A. LIELAUSIS, 1964 Physiological studies of conditional lethal mutants of bacteriophage T4D. Cold Spring Harbor Symp. Quant. Biol. **28:** 375–392.

FULFORD, W., M. RUSSEL and P. MODEL, 1986 Aspects of the growth and regulation of the filamentous phages. Progr. Nucleic Acid Res. Mol. Biol. **33:** 141–168.

GEIDUSCHEK, E. P., T. ELLIOTT and G. A. KASSAVETIS, 1983 Regulation of late gene expression. pp. 189–192. In: *Bacteriophage T4*, Edited by C. K. MATHEWS, E. M. KUTTER, G. MOSIG and P. B. BERGET. American Society for Microbiology, Washington, D.C.

GOLD. L., D. PRIBNOW, T. SCHNEIDER, S. SHINEDLING, B. S. SINGER and G. STORMO, 1981 Translational initiation in prokaryotes. Annu. Rev. Microbiol. **35:** 365–403.

GOTT, J., D. SHUB and M. BELFORT, 1986 Multiple self-splicing introns in bacteriophage T4: evidence from autocatalytic GTP. Cell **47:** 81–87.

GRAM, H., AND W. RÜGER, 1985 Genes *55*, *agt*, *47* and *46* of bacteriophage T4: the genomic organization as deduced by sequence analysis. EMBO J. **4:** 257–264.

GRAM, H., H.-D. LIEBIG, A. HACK, E. NIGGEMAN and W. RÜGER, 1984 A physical map of bacteriophage T4 including the positions of strong promoters and terminators recognized *in vitro*. Mol. Gen. Genet. **194:** 232–240.

GRUIDL, M. E., and G. MOSIG, 1986 Sequence and transcripts of the bacteriophage T4 DNA repair gene *uvsY*. Genetics **114:** 1061–1079.

GUILD, N., M. GAYLE, R. SWEENEY, T. HOLLINGSWORTH, T. MODEER and L. GOLD, 1988 Transcriptional activation of bacteriophage T4 middle promoters by the *motA* protein. J. Mol. Biol. **199:** 241–258.

GUSSIN, G. N., and V. PETERSON, 1972 Isolation and properties of *rex*$^-$ mutants of bacteriophage lambda. J. Virol. **10:** 760–765.

HALL, D. H., and R. D. SNYDER, 1981 Suppressors of mutations in the *rII* gene of bacteriophage T4 affect promoter utilization. Genetics **97:** 1–9.

HOLLIDAY, R., 1964 A mechanism for gene conversion in fungi. Genet. Res. **5:** 282–304.

HOMYK, T. JR., A. RODRIGUEZ and J. WEIL, 1976 Characterization of T4 mutants that partially suppress the ability of T4 *rII* to grow in lambda lysogens. Genetics **83:** 477–487.

HUANG, W. M., 1986 Nucleotide sequence of a type II DNA topoisomerase gene: bacteriophage T4 gene *39*. Nucleic Acids Res. **14:** 7751–7765.

ISERENTANT, D., and W. FIERS, 1980 Secondary structure of messenger RNA efficiency of translation initiation. Gene **9:** 1–12.

JENSCH, F., and B. KEMPER, 1986 Endonuclease VII resolves Y-junctions in branched DNA in vitro. EMBO J. **5:** 181–189.

KASSAVETIS, G. A., and E. P. GEIDUSCHEK, 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.

KASSAVETIS, G. A., P. G. ZENTNER and E. P. GEIDUSCHEK, 1986 Transcription at bacteriophage T4 variant late promoters: an application of a newly devised promoter-mapping method involving RNA chain retraction. J. Biol. Chem. **261:** 14256–14265.

KEMPER, B., and D. T. BROWN, 1976 Function of gene *49* of bacteriophage T4. II. Analysis of intracellular development and the structure of very fast-sedimentating DNA. J. Virol. **18:** 1000–1015.

KEMPER, B., and M. GARABETT, 1981 Studies on T4-head maturation. I. Purification and characterization of gene-*49*-controlled endonuclease. Eur. J. Biochem. **115:** 123–131.

KEMPER, B., F. JENSCH, M. V. DEPKA-PRONDZYNSKI, H.-J. FRITZ, U. BORGMEYER and K. MIZUUCHI, 1984 Resolution of Holliday structures by endonuclease VII as observed in interactions with cruciform DNA. Cold Spring Harbor Symp. Quant. Biol. **48:** 815–825.

KENNELL, D. E., 1986 The instability of messenger RNA in bacteria. pp. 101–142. In: *Maximizing Gene Expression*, Edited by W. REZNIKOFF and L. GOLD. Butterworth, London.

KING, C. T., R. SIRDESKMUKH and D. SCHLESSINGER, 1986 Nucleoytic processing of ribonucleic acid transcripts in procaryotes. Microbiol. Rev. **50:** 428–451.

KLEFF, S., and B. KEMPER, 1988 Initiation of heteroduplex-loop repair by T4-encoded endonuclease VII *in vitro*. EMBO. J. **7:** 1527–1535.

LAEMMLI, U. K., 1970 Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature **227:** 680–685.

LANDSMANN, J., M. KROGER and A. HOBOM, 1982 The *rex* region of bacteriophage lambda: two genes under three-way control. Gene **20:** 11–24.

LEMASTER, D. M., 1986 Nucleotide sequence and protein overproduction of bacteriophage T4 thioredoxin. J. Virol. **59:** 759–760.

LIPMAN, D. J., and W. R. PEARSON, 1985 Rapid and sensitive protein similarity searches. Science **227:** 1435–1441.

LUDER, A., and G. MOSIG, 1982 Two alternative mechanisms for initiation of DNA replication forks in bacteriophage T4: priming by RNA polymerase and by recombination. Proc. Natl. Acad. Sci. USA **79:** 1101–1105.

LUNN, C. A., and V. P. PIGIET, 1986 Chemical cross-linking of thioredoxin to hybrid membrane fraction in *Escherichia coli*. J. Biol. Chem. **261:** 832–838.

MACDONALD, P. M., and G. MOSIG, 1984a Cloning and physical mapping of an early region of the bacteriophage T4 genome. Genetics **106:** 1–16.

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

MACDONALD, P. M., E. KUTTER and G. MOSIG, 1984 Regulation of a bacteriophage T4 late gene, *soc*, which maps in an early region. Genetics **106:** 17–27.

MANIATIS, T., E. F. FRITSCH and J. SAMBROOK, *1982 Molecular Cloning. A Laboratory Manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

MCGRAW, T., L. MINDICH and B. FRANGIONE, 1986 Nucleotide sequence of the small double-stranded RNA segment of bacteriophage ϕ6: novel mechanism of natural translational control. J. Virol. **58:** 142–151.

MCPHEETERS, D. S., A. CHRISTENSEN, E. G. YOUNG, G. STORMO and L. GOLD, 1986 Translational regulation of expression of the bacteriophage T4 lysozyme gene. Nucleic Acids Res. **14:** 5813–5826.

MELTON, D. A., P. A. KRIEG, M. R. REBAGLIATI, T. MANIATIS, K. ZINN and M. R. GREEN, 1984 Efficient *in vitro* synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucleic Acids Res. **12:** 7035–7056.

MESSING, J., and J. VIEIRA, 1982 A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. Gene **19:** 269–276.

MILLER, J., A. D. MCLACHLAN and A. KLUG, 1985 Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. EMBO J. **4:** 1609–1614.

MINAGAWA, T., and Y. RYO, 1978 Substrate specificity of gene *49*-controlled deoxyribonuclease of bacteriophage T4: special reference to DNA packaging. Virology **91:** 222–233.

MIZUUCHI, K., B. KEMPER, J. HAYS and R. A. WEISBERG, 1982 T4 endonuclease VII cleaves Holliday structures. Cell **29:** 357–365.

MODRICH, P., and C. C. RICHARDSON, 1975 Bacteriophage T7 deoxyribonucleic acid replication *in vitro*. A protein of *Escherichia coli* required for bacteriophage T7 DNA polymerase. J. Biol. Chem. **250:** 5508–5513.

MOSIG, G. 1983a Relationship of T4 DNA replication and recombination. pp. 120–130. In: *Bacteriophage T4*, Edited by C. K.

MATHEWS, E. M. KUTTER, G. MOSIG and P. B. BERGET. American Society for Microbiology, Washington, D.C.

MOSIG, G., 1983b Appendix: T4 genes and gene products. pp. 362–374. In: *Bacteriophage T4*, Edited by C. K. MATHEWS, E. M. KUTTER, G. MOSIG and P. B. BERGET. American Society for Microbiology, Washington, D.C.

MOSIG, G., 1987 The essential role of recombination in phage T4 growth. Annu. Rev. Genetics **21:** 347–371.

MOSIG, G., and F. EISERLING, 1988 Phage T4 Structure and Metabolism. pp. 521–606. In: *The Bacteriophages*, Vol. 2, Edited by R. CALENDAR. Plenum, New York.

MOSIG, G., and D. POWELL, 1985 Heteroduplex loops are packaged in gene *49* (endonuclease VII) mutants of bacteriophage T4. American Society for Microbiology, 85th Annual Meeting, **M4:** 209 (Abstr.)

MOSIG, G., W. BERQUIST and S. BOCK, 1977 Multiple interactions of a DNA-binding protein *in vivo*. III. Phage T4 gene-*32*-mutations differentially affect insertion-type recombination and membrane properties. Genetics **86:** 5–23.

MOSIG, G., M. SHAW and G. M. GARCIA, 1984 On the role of DNA replication, endonuclease VII, and *r*II proteins in processing of recombinational intermediates in phage T4. Cold Spring Harbor Symp. Quant. Biol. **49:** 371–382.

MOSIG, G., R. DANNENBERG, D. GHOSAL, A. LUDER, S. BENEDICT and S. BOCK, 1979 General genetic recombination in bacteriophage T4. Stadler Gamet. Symp. **11:** 31–56.

NASH, H. A., C. A. ROBERTSON, E. FLAMM, R. A. WEISBERG and H. I. MILLER, 1987 Overproduction of *Escherichia coli* Integration Host Factor, a protein with nonidentical subunits. J. Bacteriol. **169:** 4124–4127.

OEFNER, C., and D. SUCK, 1986 Crystallographic refinement and structure of DNase I at 2Å resolution. J. Mol. Biol. **192:** 605–632.

OLLIS, D., P. BRICK, R. HAMLIN, N. XUONG and T. STEITZ, 1985 Structure of large fragment of *Escherichia coli* DNA polymerase I complexed with dTMP. Nature **313:** 762–766.

OPPENHEIM, A., and C. YANOFSKY, 1980 Translational coupling during expression of the tryptophan operon of *Escherichia coli*. *Genetics* **95:** 785–795.

PORTIER, C., L. DONDON, M. GRUNBERG-MANAGO and P. REGNIER, 1987 The first step in the functional inactivation of the *Escherichia coli* phosphorylase messenger is a ribonuclease III processing at the 5′ end. *EMBO J.* **6:** 2165–2170.

RABUSSAY, D., 1983 Phage-evoked changes in RNA polymerase. pp. 167–173. In: *Bacteriophage T4*, Edited by C. K. MATHEWS, E. M. KUTTER, G. MOSIG and P. B. BERGET. American Society for Microbiology, Washington, D.C.

REMAUT, E., P. STANSSENS and W. FIERS, 1981 Plasmid vectors for high-efficiency expression controlled by the P_L promoter of coli phage lambda. Gene **15:** 81–93.

RIEDE, I., 1987 Lysis gene *t* of T-even bacteriophages: evidence that colicins and bacteriophage genes have common ancestors. J. Bacteriol. **169:** 2956–2961.

RUNNELS, J., AND L. SNYDER, 1978 Isolation of a bacterial host selective for bacteriophage T4 containing cytosine in its DNA. J. Virol. **27:** 815–818.

SANGER, F., S. NICKLEN and A. R. COULSON, 1977 DNA sequencing with chain termination inhibitors. Proc. Natl. Acad. Sci. USA **74:** 5463–5467.

SCHMIDT, F. J., and D. APIRION, 1983 T4 transfer RNAs: paradigmatic system for the study of RNA processing. pp. 208–217. In: *Bacteriophage T4*, Edited by C. K. MATHEWS, E. M. KUTTER, G. MOSIG and P. B. BERGET. American Society for Microbiology, Washington, D.C.

SHINE, J., and L. DALGARNO, 1974 The 3′-terminal sequence of *Escherichia coli* 16s ribosomal RNA: complementary to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA **71:** 1342–1346.

SHINEDLING, S., D. PARMA and L. GOLD, 1987 Wild-type bacteriophage T4 is restricted by the lambda *rex* genes. J. Virol. **61:** 3790–3794.

SHUB, D. A., J. M. GOTT, M.-Q. XU, B. F. LANG, F. MICHEL, J. TOMASCHEWSKI, J. PEDERSEN-LANE and M. BELFORT, 1988 Structural conservation among three homologous introns of bacteriophage T4 and the group I introns of eukaryotes. Proc. Natl. Acad. Sci. USA **85:** 1151–1155.

SINGER, B. S., S. T. SHINEDLING and L. GOLD, 1983 The *rII* genes: a history and a prospectus. pp. 327–333. In: *Bacteriophage T4*, Edited by C. K. MATHEWS, E. M. KUTTER, G. MOSIG and P. B. BERGET. American Society for Microbiology, Washington, D.C.

SKORUPSKI, K., J. TOMASCHEWSKI, W. RÜGER and L. D. SIMON, 1988 A bacteriophage T4 gene which functions to inhibit *Escherichia coli* lon protease. J. Bacteriol. **170:** 3016–3024.

STAHL, F. W., 1979 *Genetic Recombination: Thinking About It in Phage and Fungi*. W. H. Freeman, San Francisco.

STORMO, G. D., 1986 Translation initiation. pp. 195–224. In: *Maximizing Gene Expression*, Edited by W. REZNIKOFF and L. GOLD. Butterworth, Boston.

THOMPSON, R. J., and G. MOSIG, 1987 Stimulation of a *Chlamydomonas* chloroplast promoter by novobiocin in situ and in *E. coli* implies regulation by torsional stress in the chloroplast DNA. Cell **48:** 281–287.

TOMASCHEWSKI, J., and W. RÜGER, 1987 Nucleotide sequence and primary structures of gene products coded for by the T4 genome between map positions 48.266 Kb and 39.166 Kb. Nucleic Acids Res. **15:** 3632–3633.

TOOTHMAN, P., and I. HERSKOWITZ, 1980 Rex-dependent exclusion of lambdoid phages. III. Physiology of the abortive infection. Virology **102:** 161–171.

VALERIE, K., J. STEVENS, M. LYNCH, E. HENDERSON and J. DERIEL, 1986 Nucleotide sequence and analysis of the 58.3 to 65.5-kb early region of bacteriophage T4. Nucleic Acids Res. **14:** 8637–8654.

WHITEHOUSE, H. L. K., 1963 A theory of cross-over by means of hybrid deoxyribonucleic acid. Nature **199:** 1034–1040.

WHITEHOUSE, H. L. K., 1982 *Genetic Recombination: Understanding the Mechanisms*. John Wiley & Sons, New York.

WILSON, G. G., K. K. Y. YOUNG, G. J. EDLIN and W. KONIGSBERG, 1979 High-frequency generalized transduction by bacteriophage T4. Nature **280:** 80–82.

Communicating editor: J. W. DRAKE