JOURNAL OF BACTERIOLOGY, Oct. 1992, p. 6539–6547
0021-9193/92/206539-09$02.00/0
Copyright © 1992, American Society for Microbiology
Vol. 174, No. 20

Sequence and Characterization of the Bacteriophage T4 *comC*α Gene Product, a Possible Transcription Antitermination Factor

BÉNÉDICTE SANSON AND MARC UZAN*
Institut de Biologie Physico-chimique, 13, Rue Pierre et Marie Curie, 75005 Paris, France

Received 12 June 1992/Accepted 15 August 1992

We have sequenced a 1,340-bp region of the bacteriophage T4 DNA spanning the *comC*α gene, a gene which has been implicated in transcription antitermination. We show that *comC*α, identified unambiguously by sequencing several missense and nonsense mutations within the gene, codes for an acidic polypeptide of 141 residues, with a predicted molecular weight of 16,680. We have identified its product on one- and two-dimensional gel systems and found that it migrates abnormally as a protein with a molecular weight of 22,000. One of the missense mutations (*comC*α803) is a glycine-to-arginine change, and the resulting protein exhibits a substantially faster electrophoretic mobility. The ComCα protein appears immediately after infection. Its rate of synthesis is maximum around 2 to 3 min postinfection (at 37°C) and then starts to decrease slowly. Some residual biosynthesis is still detectable during the late period of phage development.

The notion that termination and antitermination mechanisms may be operating in bacteriophage T4 development initially came from the observation of a strong polarity on early transcription when infection is carried out in the absence of translation. In vivo experiments that used inhibitors of protein synthesis, as well as in vitro studies, have shown that this polarity is due to the action of the host transcription terminator ρ (2). The question of whether translation of phage RNA is required just to mask potential *rho*-dependent termination sites and/or to allow the synthesis of viral antitermination proteins could not be answered on the basis of experiments that used inhibitors of protein synthesis. Furthermore, analysis of this problem is complicated by the fact that many genes, transcribed by elongation of transcripts initiated at distal early promoters, are also transcribed from proximal promoters (the middle promoters) activated soon after infection (2, 10, 13, 24, 30, 45, 46).

Another set of experiments showing the involvement of ρ in T4 development derives from the isolation of *rho* mutants unable to support growth of T4 wild-type phage. They were called *tabC* (5, 40, 41) or *hdf* (38). It has been suggested that the altered ρ proteins made in these mutants are insensitive to T4-induced antitermination factors (31, 38). Thus, these mutations can be considered as producing a "super-rho" phenotype. Further support for this view is provided by the observation that one of these *rho* mutations (*rho*026; also called *nusD*026) (37, 38) was shown to impede the antitermination activity of λN, specifically at *rho*-dependent terminators (8).

Phage mutants able to grow on *tabC*/*hdf rho* host mutants were isolated. These compensatory mutations, called *comC* or *goF*, were mapped mainly in two places on the chromosome: upstream of gene 39, in a nonessential region of the T4 genome, and between genes 55 and *e* (5, 7, 17, 36, 38, 42). The mutations located upstream of gene 39 define at least one gene called *comC*α. Other compensatory mutations were located in genes 45 (40) and 31 (36), that is, in genes involved in replication and in head assembly, respectively. Infection of a super-rho strain (*hdf*/*rho*026) with wild-type phage leads to an increased proportion of RNA ending at a specific site within gene 40 and to a decrease in the *uvsX*-40-41 polycistronic RNA (12). The presence of the *goF*1 mutation (located upstream of gene 39) in the infecting phage reverses these effects to a wild-type pattern of synthesis. One interpretation of these results is that a *rho*-dependent termination site lies within gene 40 (with the *rho*026 protein increasing termination) and that the gene *comC*α/*goF* encodes an antitermination function. Complete elimination of the *comC*α gene product (in a nonsense or a total deletion mutant of the gene) does not permit phage growth on *tabC*/*hdf* hosts (7, 38, 42). This observation suggests that the phage compensatory mutations correspond to an improved ability of the antitermination factors to antagonize ρ action. As in phage λ, antitermination in T4 (if it occurs) may recruit host factors in addition to viral factors. This is suggested by the finding that the *hdf*/*rho*026-induced restriction of T4 growth can be suppressed by an overproduction of the NusG host protein (39).

In contrast to the situation that prevails with phage λ, elimination of one of these putative antitermination factors by missense, nonsense, or deletion mutations does not affect T4 phage growth on wild-type (rho^+) hosts (14, 38, 40). Thus, despite the accumulation of evidence in favor of the existence of termination and antitermination in T4 transcription, their effective participation in normal phage development has not been established convincingly. It is possible that transcription antitermination is mediated by at least two interchangeable factors and that both need to be eliminated in order to affect T4 growth.

To get further insight into this problem, a molecular analysis of the putative antiterminator proteins and their corresponding genes has been undertaken. In this article, we present the primary sequence of the *comC*α gene (and surrounding regions). We have characterized its product on one- and two-dimensional gels and have determined its kinetics of biosynthesis.

* Corresponding author.

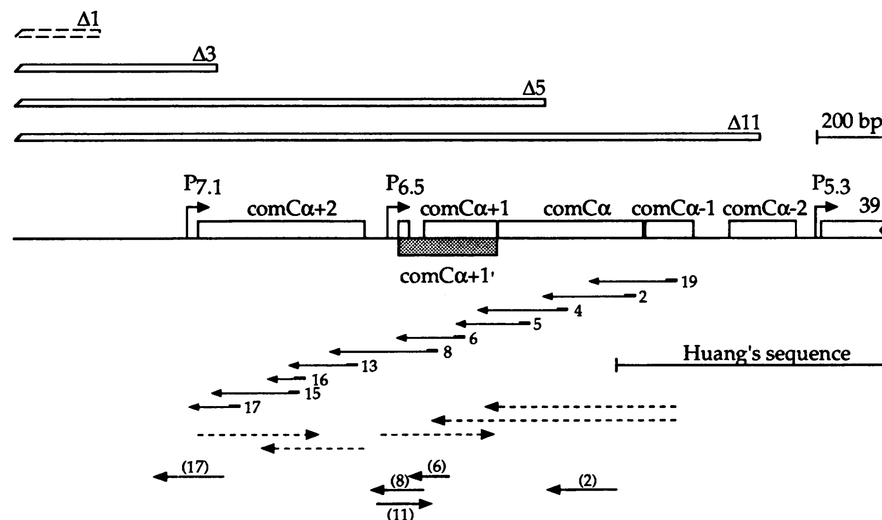

FIG. 1. Sequencing strategy for the *comCα* region. The ORFs are represented by open boxes. The filled box represents a hypothetical ORF (see the text). Promoters are shown by right-angled arrows. The promoters are numbered by reference to the distance in kilobases from the origin of the physical map to the transcription start site (21). The arrows below the DNA show the sequencing strategy. Sequences derived from RNA are indicated by solid arrows with a thicker end representing the oligonucleotide primer used in the extension reactions. The numbers indicate the different oligonucleotides. Sequences obtained from DNA fragments cloned in phage M13 are shown by dotted arrows. The direct genomic sequencing technique was also used to sequence the regions indicated by thick solid arrows. With this technique, the first 20 nucleotides from the primer could not be read; the arrows show the location of the actual sequence read; the primers used are noted above the arrows. Part of the sequence already published by Huang (15) is indicated by a solid bar. The gene 39 proximal ends of deletions del(39,56)-3, del(39,56)-5, and del(39,56)-11, sequenced in this work, are represented above the DNA by solid lines. They are labeled Δ3, Δ5, and Δ11, respectively. The gene 39 proximal end of del(39,56)-1 deletion (Δ1), shown by a box in dotted lines, is adapted from the work of Homyk and Weil (14).

MATERIALS AND METHODS

Bacteria and bacteriophages. B^E (sup^0), an *Escherichia coli* B strain, is our standard bacterium to grow phage T4. CT3-*tabC*803, a derivative of the K-12 strain CT3 [F^- *lac galE galK*(Am) *trp*(Am) *ara*(Am) *tsx thy* sup^0] (41), contains a mutation in the *rho* gene which makes the strain restrictive for wild-type T4 growth (40). It has been used to control the phenotype of the different T4 *comCα/goF* mutants (see below).

Our wild-type bacteriophage T4 is T4D. T4*comCα*803, T4*comCα*55.6, and T4*goF*1 contain mutations that allow the phage to grow on *tabC*803, *tabC*5521, and *hdf* hosts, respectively (5, 7, 38, 40). *hdf* mutations, like *tabC*, are mutations in *rho* that strongly restrict growth of T4 wild-type (38). T4*comCα*803*amB*58 contains an additional amber mutation within the same gene ([42] and this study). This phage does not grow on a *tabC* strain. Phages del(39,56)-3, del(39,56)-5, and del(39,56)-11 contain large deletions in a nonessential region of the T4 genome located between genes 39 and 56 (14).

RNA sequencing. RNA extraction and sequencing were performed under conditions similar to those described previously (47). RNAs were usually extracted after 3 to 4 min of infection by T4. In some cases, RNA from a T4*regB* mutant infection was used as a template in order to sequence through major processing sites generated by the RegB endonuclease (32, 47). Primer extension reactions using avian myeloblastosis virus reverse transcriptase were carried out in the presence of the dideoxyribonucleotides. The ratios of deoxyribonucleotides to dideoxyribonucleotides were 2, 5, or 10, according to the distance from the oligonucleotide that we wanted to read.

DNA cloning and sequencing. Sequencing the RNA as described above avoids the cloning of T4 fragments that may contain lethal functions and therefore ensures that we are producing a wild-type sequence. Nevertheless, because of the strong secondary structures that RNA can adopt, some parts of our gels were not well resolved. After we obtained a first sequence of the *comCα* region on the RNA, any ambiguities were resolved by sequencing DNA fragments of the region (shown in Fig. 1) after their cloning into M13 phages mp18 and mp19 (49). Sequences were determined by the chain termination method (33), with the T7 DNA polymerase (Sequenase from United States Biochemical Corp.). The direct genomic DNA sequencing technique was also used in some cases (see Fig. 1) (20).

Phage infection and protein labeling. Labeling experiments were carried out essentially as described previously (46, 48). Briefly, cells were grown in MOPS (morpholine propanesulfonic acid)-Tricine medium (25) supplemented with glucose and all amino acids except methionine. When the cells reached a density of 5×10^8 per ml, they were UV irradiated (to stop host gene expression) and then infected with phages at a multiplicity of infection of 7 to 10. Proteins were labeled by the addition to the medium of L-[^{35}S]methionine (>1,000 Ci/mmol; Amersham) at concentrations and for periods specified in the legends of Fig. 3, 4, and 5.

Protein analysis. For one-dimensional electrophoresis, the infected cells were centrifuged and resuspended in 1/20 the original volume of sodium dodecyl sulfate (SDS)-EDTA sample buffer (4) and then boiled for 3 min. The electrophoresis was carried out on 15% polyacrylamide gels. The gel was prepared as described previously (4), except that a unique concentration of acrylamide was used instead of a step gradient. For two-dimensional analyses, the nonequilibrium pH gradient electrophoresis technique (27), followed

by an SDS-polyacrylamide gel electrophoresis, was used. The labeled infected cells were centrifuged and treated according to the method of O'Farrell et al. (27). The first dimension was a pH gradient electrophoresis on gels containing Ampholines (pH 3.5 to 10). This was performed in tubes (0.16-cm diameter and 7.5-cm length). The second dimension was an SDS-EDTA-15% polyacrylamide electrophoresis, as described above. General conditions of electrophoresis and treatment of the gels were as described previously (27, 48).

Computer analyses. Computer analyses of the nucleic acid and protein sequences were performed by using facilities offered by the Base Informatique sur les Séquences d'Acides Nucléiques pour les Chercheurs Européens at the Centre Inter-Universitaire d'Informatique à Orientation Biomédicale, Paris, France (9).

Determination of the ability of a DNA sequence to code for proteins was carried out by using two codon usage matrices. One matrix was made up on the basis of the codon frequency found in a subset of 34 prereplicative genes: *αgt*, *alc*, *βgt*, *cd*, *denV*, *dexA*, *frd*, *imm*, *ip*I, *ip*III, *pseT*, *regA*, *rpbA*, *r*IIA, *td*, *tk*, *uvsY*, 1, 30, 32, 33, 42, 43, 44, 45, 46, 47, 49, 52, 55, 59, 62, 63, and 69. Altogether, they represent 9,169 codons. Similarly, a codon usage matrix for the late genes was built from the sequences of 20 well-characterized late genes: *e*, *soc*, 6, 7, 8, 10, 11, 12, 16, 17, 20, 21, 23, 26, 27, 36, 37, 51, 67, and 68. They represent 7,429 codons. All the sequences were extracted from the GenBank data base.

Nucleotide sequence accession number. The EMBL/GenBank/DDBJ accession number for the primary nucleotide sequence presented in this article is M89919.

RESULTS AND DISCUSSION

Sequence of the *comC*α region. Some of the mutations which permit T4 growth on *tabC*/*hdf* hosts have been mapped just upstream of gene 39 (7, 38, 40, 42). Preliminary results of S1 mapping and primer extension analyses (10, 43, 44) indicate that gene 39 is transcribed from both early and middle promoters and that the early promoter(s) maps approximately 1.5 kb upstream of gene 39. To avoid the problems arising from the cloning of strong early promoters (22, 45), we decided to sequence this region by walking on the RNA, using a set of oligonucleotides as primers in a series of extension reactions that used avian myeloblastosis virus reverse transcriptase in the presence of dideoxyribonucleotides. The first primer used (no. 19 in Fig. 1) was complementary to a region toward the 5′ end of the sequence published by Huang for gene 39 (15). Putative promoter regions were further sequenced at the level of the DNA by the direct genomic sequencing technique (20). The sequencing strategy is summarized in Fig. 1. In Fig. 2, the sequence obtained (nucleotides 1 to 1340) has been fused to the beginning of the sequence published by Huang (15) for gene 39. Partial resequencing of Huang's sequence with oligonucleotides 19 and 2 (see Fig. 1) allowed us to find two differences: an insertion of an A at position 1408 and an A instead of a G at position 1475 (Fig. 2).

Five and possibly six open reading frames (ORFs) can be found in the direction of early transcription. One of them is the *comC*α gene (see below). The locations of their start and termination codons, the predicted molecular weights and isoelectric points (pI) of the products encoded, and their putative ribosomal binding sites are shown in Table 1. We found that all these ORFs are located in a region presenting a high probability of coding for proteins (see Materials and Methods). In contrast, this probability is very weak when the analysis is performed for the "late strand."

The ORF that starts at nucleotide 711 is prematurely interrupted by a UGA stop codon (Fig. 2). Forty-two nucleotides farther on is a start site for another ORF (*comC*α.+1). Both ORFs are preceded by good Shine-Dalgarno sequences; furthermore, the two ORFs are in phase, offering the possibility of reading a larger ORF (*comC*α.+1′) if the UGA codon is suppressed. The same sequence was found in our T4 wild-type phage and in two different T4 phages carrying unrelated mutations. Therefore, we believe that the presence of the UGA codon does not result from a mutation. In Table 1, we show the properties of the *comC*α.+1′-encoded polypeptide when the UGA codon is read as a tryptophan (see also Fig. 1). Some examples of natural UGA suppression have been described for *E. coli* and its phages (1, 28). We do not know whether this is happening here.

We found no strong homology of the ORF-encoded polypeptides described in Table 1 with any of the polypeptides contained in the National Biomedical Research Foundation data base (release no. 31).

Identification and mapping of the transcription start points in this region will be presented in detail in a following paper (34). We found that P7.1 and P6.5 are early promoters, while P5.3 is a *motA*-dependent (middle) promoter (10).

Identification of the *comC*α gene. The *comC*α gene was first defined by mutations that allow T4 to grow on *E. coli tabC* super-rho mutants. These mutations were mapped just upstream of gene 39 (5, 42) (see the introduction). Sequencing of RNA made after infection of *E. coli* by *comC*α mutants allowed us to identify unambiguously the *comC*α gene among the six ORFs. The two mutations, *comC*α803 and *comC*α55.6, isolated independently (5, 7, 40), correspond to the same nucleotide change, resulting in a glycine-to-arginine substitution at position 84 of the protein (Fig. 2). We found that the *goF*1 mutation (38), conferring the same phenotype as the *comC*α mutations, also lies within *comC*α. Here, the aspartic acid at position 25 is replaced by a tyrosine. The *amB*58 amber mutation is thought to map in the *comC*α cistron for the following reasons. The double mutant *amB*58*comC*α803 was isolated as a phage that does not grow on *tabC* sup^0 strains but grows on *tabC* sup^+ and tab^+ sup^+ strains. This phage does not complement wild-type T4 growth on *tabC* sup^0 bacteria. The frequency of recombination between the two markers is very low (42). Sequencing of the RNA made after infection by the double mutant showed that *amB*58 interrupts the ComCα polypeptide after the 100th amino acid (Fig. 2).

A number of large deletions located in a nonessential region of the T4 genome, between genes 56 and 39 (14), have been used to map several loci upstream of gene 39. We have sequenced three of the deletions. The del(39,56)-11 deletion removes most of the region studied in this work; its rightmost end is located approximately 170 bp upstream of gene 39. The rightmost end of del(39,56)-5 is in the middle of *comC*α, while that of del(39,56)-3 lies within *comC*α.+2 (Fig. 1 and 2). Genetic studies that used these three deleted phages to map the different *comC*α and *goF* mutations (7, 38, 42) are in full agreement with the physical map of the three deletions.

The *motC* gene has been located in a region between the gene 39 proximal ends of del(39,56)-3 and del(39,56)-5 (30). Effects of *motC* deletions are subtle and were observed only on transcription of the tRNA genes in combination with *motB* deletion mutations (*motB* has been assigned to a region

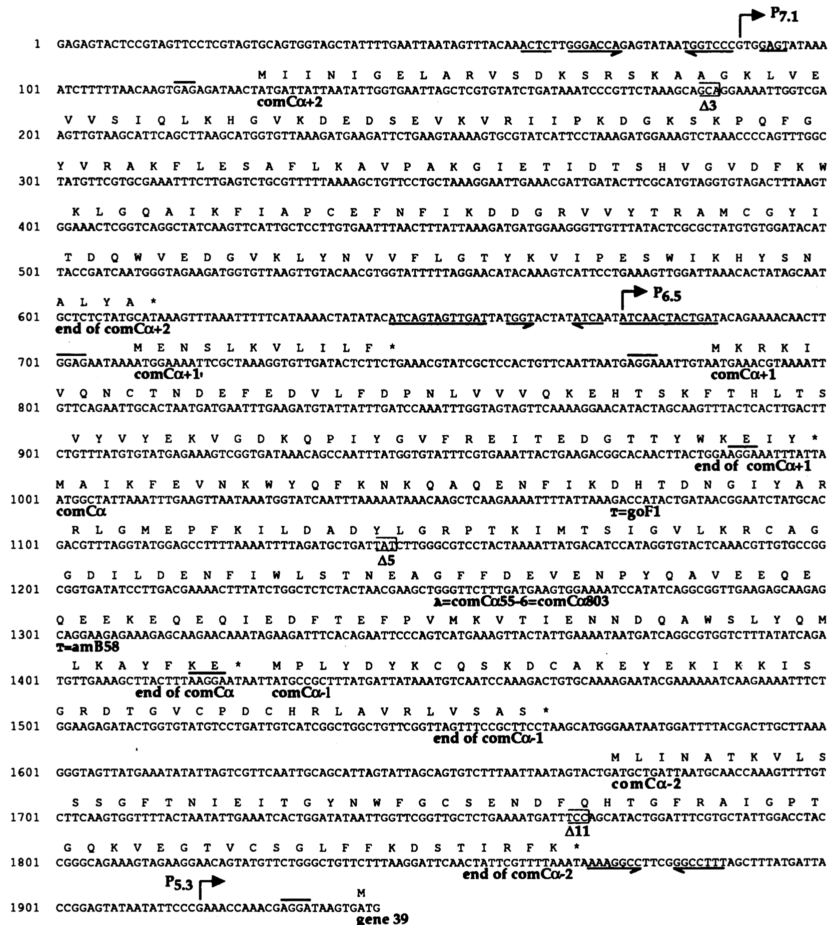

FIG. 2. Sequence of the *comCα* gene and surrounding regions. Translation of the ORFs is given in the one-letter code. For reasons of clarity, translation of ORF *comCα.+1'* is given only to the UGA stop codon (see the text). The presumed Shine-Dalgarno sequences of the ORFs are indicated by bars above the DNA sequence. Right-angled arrows show the transcription start sites found in this region. Sequences showing dyad symmetry are indicated by convergent arrows below the sequence. The right-most ends of deletions del(39,56)-3, del(39,56)-5, and del(39,56)-11 (labeled Δ3, Δ5, and Δ11, respectively, in the figure) are indicated by brackets in the sequence. Nucleotide changes corresponding to *goF*1, *comCα*803, *comCα*55.6, and *comCαamB*58 mutations are shown below the sequence.

upstream of del(39,56)-1 [see Fig. 1]). On the basis of the genetic data, *motC* could be *comCα.+*1 or *comCα* itself.

Can we learn something about the function of the ComCα protein by analyzing the type of substitutions found in the compensatory mutants? The two different mutations analyzed produce additional pairs of aromatic and basic residues in the protein chain. The wild-type *comCα* polypeptide contains six examples of aromatic and basic amino acid pairs (seven when a space of one residue is allowed between them). Aromatic amino acids, when in close proximity to basic residues, have been implicated in the interaction of certain proteins with single-stranded nucleic acids (6, 18, 19, 29, 35). Hence, we suggest that ComCα is a single-stranded nucleic acid-binding protein (RNA rather than DNA could be its ligand). Thus, the *goF*1 or *comCα*803 mutation would lead to a compensation of the host *rho* mutant effect by increasing the number of protein-RNA interactions.

Identification of the *comCα* gene product. Identification of the *comCα* gene product is an important prerequisite for in vitro experiments aimed at studying its mechanism of action.

TABLE 1. Characteristics of ORFs found in the *comCα* region

ORF	Nucleotides	Ribosomal binding site[a]	Termination codon	Number of codons	Predicted mol wt	Theoretical pI
comCα.+2	127–612	ACAAGTGAGAGATAACTATG	TAA	162	18,217	9.3
comCα.+1'	711–998	ACAACTTGGAGAATAAAATG	TAA	96	11,378	5.3
comCα.+1	786–998	TTAATGAGGAAATTGTAATG	TAA	71	8,464	5.5
comCα	1,001–1,423	TGGAAGGAAATTTATTAATG	TAA	141	16,681	4.7
comCα.−1	1,429–1,563	TACTTTAAGGAATAATTATG	TAA	45	5,106	8.8
comCα.−2	1,673–1,867	TTTAATTAATAGTACTGATG	TAA	65	7,165	8.5

[a] Putative Shine-Dalgarno sequences and initiation codons are underlined.

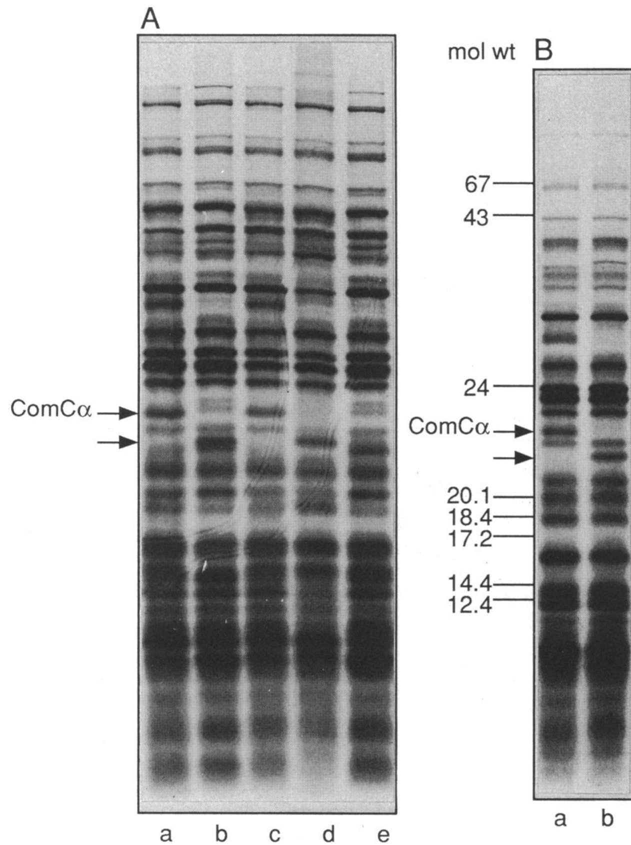

FIG. 3. Identification of the *comC*α gene product on one-dimensional gel. Phage-infected *E. coli* B^E cells were labeled 0.5 to 5 min after infection at 37°C with 60 μCi of [^{35}S]methionine per ml of culture. Pulse-labeling was stopped by the addition of cold L-methionine to 400 μg per ml and sodium azide to 2 mM. The samples were then divided into two parts. One part, shown in this figure, was treated as described in Materials and Methods for electrophoresis on SDS-polyacrylamide gels. The other part was treated for electrophoresis on two-dimensional gels (see Fig. 4). The same volume of phage-infected cells was loaded onto the gel from each infection. This corresponded roughly to the loading of the same number of counts in each slot. Electrophoresis was carried out on SDS-EDTA-15% polyacrylamide gels. (A) Infections were carried out with the following phages: T4$^+$ (lane a), *comC*α803 (lane b), *goF*1 (lane c), *comC*α55.6 (lane d), and *comC*α803*amB*58 (lane e). Positions of the *comC*α gene product in the wild type and in the *comC*α803 and *comC*α55.6 mutants are indicated by arrows. (B) Two of the samples used in panel A, corresponding to infection by T4$^+$ (lane a) and *comC*α803 (lane b), were electrophoresed with the following molecular weight markers (sizes given in thousands): bovin serum albumin (67), alkaline phosphatase (43), trypsinogen (24), soybean trypsin inhibitor (20.1), β-lactoglobulin (18.4), myoglobin (17.2), lysozyme (14.4), and cytochrome *c* (12.4), the positions of which are indicated to the left of the gel.

Comparison of the proteins made after infection of *E. coli* B^E by T4 wild type and a phage containing a nonsense mutation in *comC*α (*comC*α803*amB*58) on one-dimensional gels reveals a few differences, among which is the absence in the infection with the amber mutant, of a band with an apparent molecular weight of approximately 22,000 (Fig. 3A, lanes a and e, and B, lane a). This result was unexpected since the primary sequence of the gene (see above) predicts a molecular weight of approximately 16,700. Moreover, analysis of the proteins synthesized after *comC*α803 or *comC*α55.6 infection (the two mutations are identical; see above) shows, in both cases, a shift of the 22,000 band to a position corresponding to an apparent molecular weight of approximately 21,000 (Fig. 3A, lanes b and d, and B, lane b). After infection by T4*goF*1, which contains a different mutation in *comC*α, the position of the 22,000 band is unaffected (Fig. 3A, lane c). Besides the 22,000 protein band, several differences can be seen between the proteins synthesized by the different mutant phages (Fig. 3A). Although all the mutations have been isolated in nominally the same parental phage (T4D), the phages come from different laboratories and have evolved independently. We consider the observed differences to be the result of silent mutations that are not, a priori, related to the phenomenon which we are studying.

To confirm the identification of ComCα, the proteins were analyzed on two-dimensional gels by using the nonequilibrium pH gradient electrophoresis system developed by O'Farrell et al. (27) in the first dimension, combined with SDS-polyacrylamide gel electrophoresis in the second dimension. This technique gives a good resolution of the T4 early proteins, many of which are basic (3, 48). By using different lengths of time of electrophoresis in the first dimension, it is possible to distinguish the acidic proteins, which stop moving after a short time, from the more basic ones which continue migrating as long as the tension is maintained. In Fig. 4A and B, the proteins synthesized after wild-type and *comC*α803*amB*58 infections are compared. A spot migrating as an acidic protein with an apparent molecular weight of 22,000 is absent in the *comC*α amber infection. Comparison of the phage proteins made in wild-type and *comC*α803 infections (Fig. 4C and D) shows that the mutation leads to a displacement of the spot in both dimensions: it has shifted towards less-acidic pH and appears clearly as a protein with a lower molecular weight. In a similar analysis, the *goF*1 polypeptide was found to migrate exactly as did the wild-type protein (data not shown). These results identify unambiguously this protein with an apparent molecular weight of 22,000 as the *comC*α gene product.

We conclude from these studies that the *comC*α gene product has an electrophoretic mobility on SDS-polyacrylamide gels lower than that predicted by its calculated molecular weight. This abnormal behavior may be correlated to the rather acidic pI of the protein (calculated pI = 4.7). Several acidic proteins, the σ factors (11) and NusA protein of *E. coli* (16), for example, have been shown to migrate more slowly than expected from their primary sequences. We have shown that in *comC*α803, an arginine replaces a glycine at position 84 of the protein. This change adds a positive charge, which is expected to increase the electrophoretic mobility on SDS-polyacrylamide gels. Nevertheless, this cannot be the only explanation for the observed dramatic increase in mobility since in the *goF*1 mutant, an acidic amino acid, aspartic acid, is replaced by a tyrosine without any change in the electrophoretic mobility of the corresponding protein.

Kinetics of biosynthesis of the *comC*α gene product. Since the ComCα protein can be resolved from the other T4 proteins on a one-dimensional 15% polyacrylamide gel, it is easy to follow its kinetics of synthesis during phage development. Figure 5 shows the result of an experiment involving pulse-labeling of the T4 proteins from the beginning of infection to 20 min later at 37°C. The protein ComCα starts to be synthesized immediately after infection. This result was expected since the gene is preceded by two early promoters (Fig. 2) (34). The rate of synthesis reaches a maximum between 1 and 3 min after infection and decreases

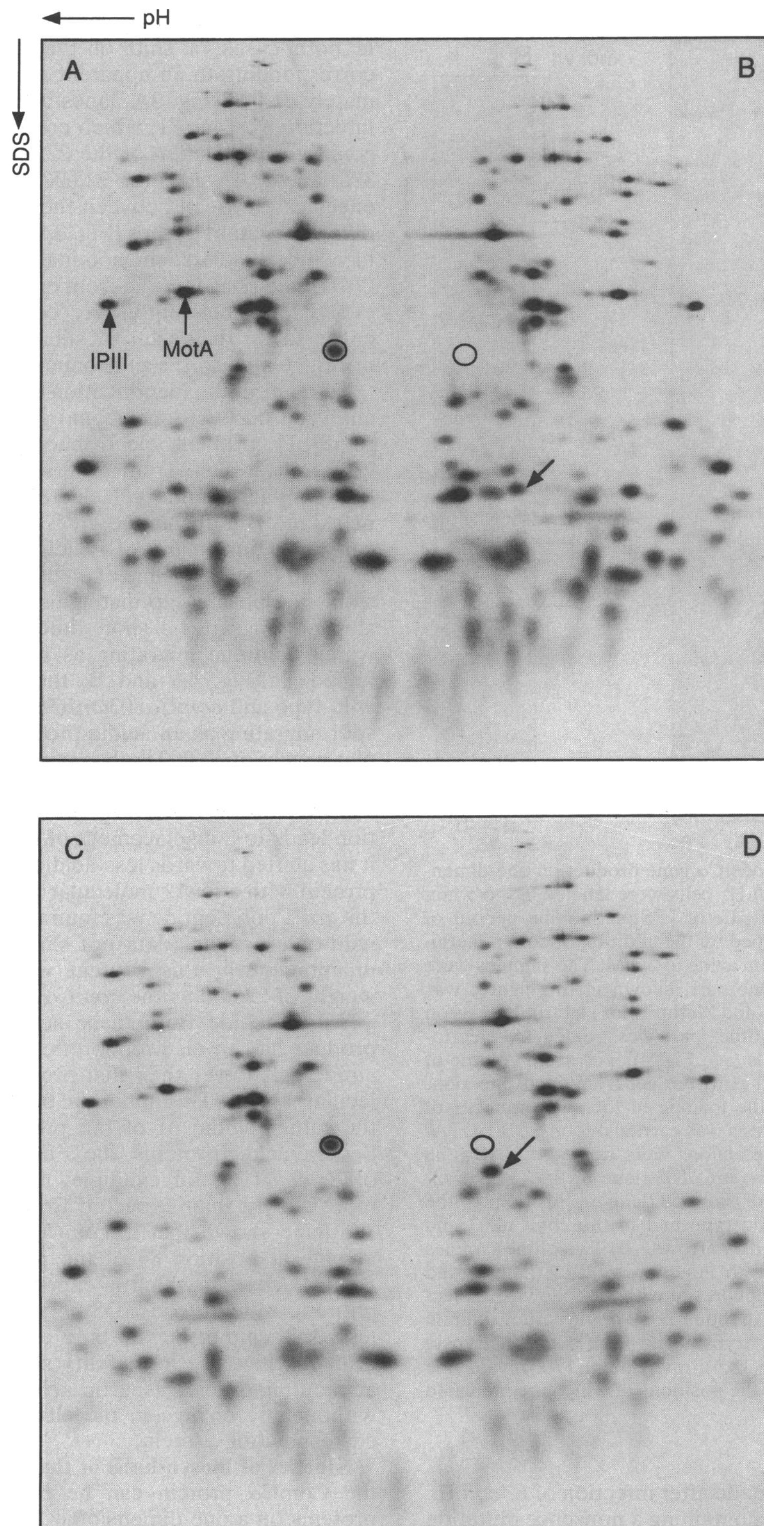

FIG. 4. Two-dimensional gel analysis. Infection and labeling were performed as described in the legend to Fig. 3. Treatment of the infected cells to be analyzed on Ampholine-containing gels was done as described previously (27, 48). In the first dimension, a pH gradient electrophoresis, the four samples were run simultaneously for 1 3/4 h at 500 V. The second dimension separates the proteins according to their sizes. Two cylindrical gels to be compared (A and B; C and D) were placed on top of the same slab SDS-EDTA-15% polyacrylamide gel, with the acidic ends facing each other at the center of the slab. Thus, one protein distribution becomes the mirror image of the other. Infection with T4^{+} (A and C), T4*comC*α803*amB*58 (B), and T4*comC*α803 (D) is shown. The circles mark the positions of the wild-type *comC*α protein. In panel A, the MotA and IpIII proteins are shown for reference (48). The arrows show the presumed ComCα amber fragment (panel B) and the position of the ComCα polypeptide made by the *comC*α803 mutant (panel D).

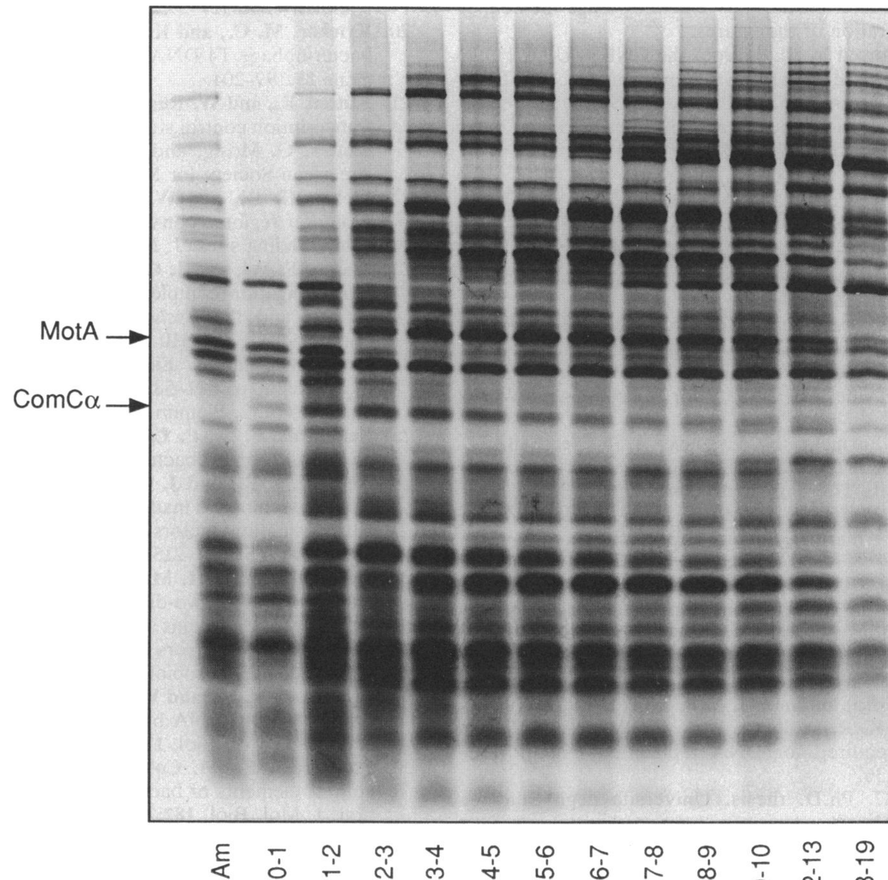

FIG. 5. Kinetics of ComCα biosynthesis. T4$^+$-infected *E. coli* B^E cells were pulse-labeled at 37°C for 1 min with 25 μCi of [^{35}S]methionine per ml of culture. The times (minutes) of labeling are indicated at the bottom of the figure. Pulse-labelings were stopped as described in the legend to Fig. 3. The same number of infected cells (5×10^7) was loaded onto the gel for each pulse-labeling. The lane labeled Am shows the 0.5- to 5-min labeled T4*comC*α*803amB58*-infected cell extract described in the legend to Fig. 3. Conditions of electrophoresis were the same as those described in the legend to Fig. 3. Arrows indicate the positions of the ComCα and MotA polypeptides.

slowly thereafter; some ComCα synthesis is still detectable at least 13 min postinfection. Thus, ComCα is an early protein whose synthesis is prolonged past the prereplicative period. Such kinetics might explain why ComCα appears as an abundant protein with respect to the total T4 prereplicative proteins (Fig. 3 and 4). Synthesis of other T4 early proteins is inhibited soon after infection, with different kinetics. This type of behavior is exemplified by the transcription activator MotA whose synthesis is shut off much more dramatically than that of ComCα (Fig. 5) (45, 47, 48).

Does the prolonged synthesis of ComCα mean that this protein is required for a long period and/or that it acts stoichiometrically rather than catalytically? If this protein does act as an antitermination factor, then it might be associated either with elongating RNA polymerase in complexes of the kind described for bacteriophage λ antitermination (23, 26) or directly with mRNA, thus competing with ρ binding. If this protein is not made in excess, then whatever its mode of action, newly synthesized antitermination factor could be titrated (or "trapped") by its nucleic acid ligand or within transcription elongation complexes. Any new wave of transcription would require more ComCα protein. Since there is evidence that ρ acts throughout the T4 cycle (38, 40), it should not be surprising that antiterminator synthesis is required until late in the cycle. Dominance/recessivity tests support the notion that ComCα is limiting in the antitermination process. Mixed infections of a super-rho host (*hdf*12.5) by T4 *goF*1 and T4$^+$ give a burst size of intermediate value (0.8 phage per bacterium), compared with the yield obtained by either phage separately (0.07 with T4$^+$ and 12.3 with *goF*1) (38). Similar results were obtained with the *comC*α*55.6* allele (7). These results could mean that the ComCα protein acts as an oligomer. Nevertheless, this explanation does not hold when one considers the result of the mixed infection with *goF*1 and T4del(39,56)-12, a phage which contains a large deletion encompassing the *comC*α gene. Surprisingly, in this case, a burst size (1.9) similar to that obtained in the *goF*1/wild type experiment was found (38). One interpretation of these results is that the ComCα protein is limiting with respect to the antitermination process and that the altered antitermination factor produced by one phage is not sufficient, in the presence of a *tabC/hdf* super-rho, to overcome all the termination events.

ACKNOWLEDGMENTS

We are indebted to J. Pulitzer, M. Chiurazzi, and C. Thermes for communicating their unpublished data. We thank J. Plumbridge and C. Chapon for critical reading of the manuscript and for discussions, C. Olsson for discussion and advice, and J. Pulitzer and B. Stitt for

the gift of phages and bacterial strains. We acknowledge the help of B. Savelli in the preparation of the figures.

This work was supported by funds from the CNRS (URA 1139), CEE (no. SC1*/0194-C[AM]) to M. Grunberg-Manago, INSERM (Contrat de recherche externe no. 891017) to M. Springer, Fondation pour la Recherche Médicale to M. Grunberg-Manago, and Université Paris 7 (Chapitre 66-71, Soutien de Programme) to P. Régnier.

REFERENCES

1. **Böck, A., K. Forchhammer, J. Heider, and C. Baron.** 1991. Selenoprotein synthesis: an expansion of the genetic code. Trends Biochem. Sci. **16:**463–467.
2. **Brody, E. N., D. Rabussay, and D. H. Hall.** 1983. Regulation of transcription of prereplicative genes, p. 174–183. *In* C. K. Mathews, E. M. Kutter, G. Mosig, and P. B. Berget (ed.), Bacteriophage T4. American Society for Microbiology, Washington, D.C.
3. **Burke, R. L., T. Formosa, K. S. Cook, A. F. Seasholtz, J. Hosoda, and H. Moise.** 1983. Use of two-dimensional polyacrylamide gels to identify T4 prereplicative proteins, p. 321–326. *In* C. K. Mathews, E. M. Kutter, G. Mosig, and P. B. Berget (ed.), Bacteriophage T4. American Society for Microbiology, Washington, D.C.
4. **Cardillo, T. S., E. F. Landry, and J. S. Wiberg.** 1979. *regA* protein of bacteriophage T4D: identification, schedule of synthesis, and autogenous regulation. J. Virol. **32:**905–916.
5. **Caruso, M., A. Coppo, A. Manzi, and J. F. Pulitzer.** 1979. Host-virus interactions in the control of T4 prereplicative transcription. I. *tabC* (*rho*) mutants. J. Mol. Biol. **135:**959–977.
6. **Chase, J. W., and K. R. Williams.** 1986. Single-stranded DNA binding proteins required for DNA replication. Annu. Rev. Biochem. **55:**103–136.
7. **Chiurazzi, M.** 1987. Ph.D. thesis. Universita degli Studi di Napoli, Naples, Italy.
8. **Das, A., M. E. Gottesman, J. Wardwell, P. Trisler, and S. Gottesman.** 1983. A mutation in the *Escherichia coli rho* gene that inhibits the N protein activity of phage lambda. Proc. Natl. Acad. Sci. USA **80:**5530–5534.
9. **Dessen, P., C. Fondrat, C. Valencien, and C. Mugnier.** 1990. BISANCE: a French service for access to biomolecular sequence databases. Comput. Appl. Biosci. **6:**355–356.
10. **Guild, N., M. Gayle, R. Sweeney, T. M. Hollingsworth, T. Modeer, and L. Gold.** 1988. Transcriptional activation of bacteriophage T4 middle promoters by the MotA protein. J. Mol. Biol. **199:**241–258.
11. **Helmann, J. D., and M. J. Chamberlin.** 1988. Structure and function of bacterial sigma factors. Annu. Rev. Biochem. **57:**839–872.
12. **Hinton, D. M.** 1989. Altered expression of the bacteriophage T4 gene 41 (primase-helicase) in an *Escherichia coli rho* mutant. J. Biol. Chem. **264:**14440–14446.
13. **Hinton, D. M.** 1991. Transcription from a bacteriophage T4 middle promoter using T4 motA protein and phage-modified RNA polymerase. J. Biol. Chem. **266:**18034–18044.
14. **Homyk, T. J. R., and J. Weil.** 1974. Deletion analysis of two nonessential regions of the T4 genome. Virology **61:**505–523.
15. **Huang, W. M.** 1986. Nucleotide sequence of a type II DNA topoisomerase gene, bacteriophage T4 gene 39. Nucleic Acids Res. **14:**7751–7765.
16. **Ishii, S., M. Thara, T. Maekawa, Y. Nakamura, H. Ushida, and F. Imamoto.** 1984. The nucleotide sequence of the cloned *nusA* gene and its flanking region of *Escherichia coli*. Nucleic Acids Res. **12:**3333–3342.
17. **Jensen, J. L., and M. Susman.** 1980. A mutant of *E. coli* that restricts growth of bacteriophage T4 at elevated temperatures. Genetics **94:**301–325.
18. **Kenan, D. J., C. C. Query, and J. D. Keene.** 1991. RNA recognition: towards identifying determinants of specificity. Trends Biochem. Sci. **16:**214–220.
19. **King, G. C., and J. E. Coleman.** 1987. Two-dimensional 1H NMR of gene 5 protein indicates that only two aromatic rings interact significantly with oligodeoxynucleotide bases. Biochemistry **26:**2929–2937.
20. **Kricker, M. C., and K. R. Tindall.** 1989. Direct sequencing of bacteriophage T4 DNA with a thermostable DNA polymerase. Gene **85:**199–204.
21. **Kutter, E., and W. Rüger.** 1983. Map of the T4 genome and its transcription control sites, p. 277–290. *In* C. K. Mathews, E. M. Kutter, G. Mosig, and P. B. Berget (ed.), Bacteriophage T4. American Society for Microbiology, Washington, D.C.
22. **Liebig, H. D., and W. Rüger.** 1989. Bacteriophage T4 early promoter regions: consensus sequences of promoters and ribosome-binding sites. J. Mol. Biol. **208:**517–536.
23. **Mason, S. W., and J. Greenblatt.** 1991. Assembly of transcription elongation complexes containing the N protein of phage lambda and the *Escherichia coli* elongation factors NusA, NusB, NusG, and S10. Genes Dev. **5:**1504–1512.
24. **Mosig, G., and F. Eiserling.** 1988. Phage T4 structure and metabolism, p. 521–589. *In* R. Calendar (ed.), The bacteriophages, vol. 2. Plenum Press, New York.
25. **Neidhardt, F. C., P. L. Bloch, and D. F. Smith.** 1974. Culture medium for enterobacteria. J. Bacteriol. **119:**736–747.
26. **Nodwell, J. R., and J. Greenblatt.** 1991. The *nut* site of bacteriophage lambda is made of RNA and is bound by transcription antitermination factors on the surface of RNA polymerase. Genes Dev. **5:**2141–2151.
27. **O'Farrell, P. Z., H. M. Goodman, and P. H. O'Farrell.** 1977. High resolution two-dimensional electrophoresis of basic as well as acidic proteins. Cell **12:**1133–1142.
28. **Parker, J.** 1989. Errors and alternatives in reading the universal genetic code. Microbiol. Rev. **53:**273–298.
29. **Prasad, B. V. V., and W. Chiu.** 1987. Sequence comparison of single-stranded DNA binding proteins and its structural implications. J. Mol. Biol. **193:**579–584.
30. **Pulitzer, J. F., M. Colombo, and M. Ciaramella.** 1985. New control elements of bacteriophage T4 pre-replicative transcription. J. Mol. Biol. **182:**249–263.
31. **Pulitzer, J. F., A. Coppo, and M. Caruso.** 1979. Host-virus interactions in the control of T4 prereplicative transcription. II. Interaction between *tabC* (*rho*) mutants and T4 *mot* mutants. J. Mol. Biol. **135:**979–997.
32. **Ruckman, J., D. Parma, C. Tuerk, D. H. Hall, and L. Gold.** 1989. Identification of a T4 gene required for bacteriophage mRNA processing. New Biol. **1:**54–65.
33. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
34. **Sanson, B., and M. Uzan.** Unpublished data.
35. **Shamoo, Y., L. R. Ghosaini, K. M. Keating, K. R. Williams, J. M. Sturtevant, and W. H. Konigsberg.** 1989. Site-specific mutagenesis of T4 gene 32: the role of tyrosine residues in protein-nucleic acid interactions. Biochemistry **28:**7409–7417.
36. **Simon, L. D., D. Snover, and A. H. Doermann.** 1974. Bacterial mutation affecting T4 phage DNA synthesis and tail production. Nature (London) **252:**451–455.
37. **Stitt, B. L., and G. Mosig.** 1989. Impaired expression of certain prereplicative bacteriophage T4 genes explains impaired T4 DNA synthesis in *Escherichia coli rho* (*nusD*) mutants. J. Bacteriol. **171:**3872–3880.
38. **Stitt, B. L., H. R. Revel, I. Lielausis, and W. B. Wood.** 1980. Role of the host cell in bacteriophage T4 development. II. Characterization of host mutants that have pleiotropic effects on T4 growth. J. Virol. **35:**775–789.
39. **Sullivan, S. L., D. F. Ward, and M. E. Gottesman.** 1992. Effect of *Escherichia coli nusG* function on lambda N-mediated transcription antitermination. J. Bacteriol. **174:**1339–1344.
40. **Takahashi, H.** 1978. Genetic and physiological characterization of *Escherichia coli K12* mutants (*tabC*) which induce the abortive infection of bacteriophage T4. Virology **87:**256–265.
41. **Takahashi, H., A. Coppo, A. Manzi, G. Martire, and J. F. Pulitzer.** 1975. Design of a system of conditional mutations (tab/k/com) affecting protein-protein interactions in bacteriophage infected *Escherichia coli*. J. Mol. Biol. **96:**563–578.
42. **Takahashi, H., and H. Yoshikawa.** 1979. Genetic study of a new early gene, *comCα*, of bacteriophage T4. Virology **95:**215–217.

43. **Thermes, C. (Centre de Genetique Moleculaire, Gif-sur-Yvette, France).** 1984. Personal communication.
44. **Uzan, M.** Unpublished data.
45. **Uzan, M., E. Brody, and R. Favre.** 1990. Nucleotide sequence and control of transcription of the bacteriophage T4 *motA* regulatory gene. Mol. Microbiol. **4:**1487–1496.
46. **Uzan, M., Y. d'Aubenton-Carafa, R. Favre, V. de Franciscis, and E. Brody.** 1985. The T4 mot protein functions as part of a prereplicative DNA-protein complex. J. Biol. Chem. **260:**633–639.
47. **Uzan, M., R. Favre, and E. Brody.** 1988. A nuclease that cuts specifically in the ribosome binding site of some T4 mRNAs. Proc. Natl. Acad. Sci. USA **85:**8895–8899.
48. **Uzan, M., J. Leautey, Y. d'Aubenton-Carafa, and E. Brody.** 1983. Identification and biosynthesis of the bacteriophage T4 mot regulatory protein. EMBO J. **2:**1207–1212.
49. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33:**103–119.