
Translational regulation of expression of the bacteriophage T4 lysozyme gene

David S. McPheeters, Alan Christensen^{1*}, Elton T. Young¹, Gary Stormo and Larry Gold

Department of Molecular, Cellular and Developmental Biology, University of Colorado, Boulder, CO 80309 and ¹Department of Biochemistry, University of Washington, Seattle, WA 98195, USA

Received 6 March 1986; Revised and Accepted 24 June 1986

ABSTRACT

The bacteriophage T4 lysozyme gene is transcribed at early and late times after infection of *E. coli*, but the early mRNA is not translated. DNA sequence analysis and mapping of the 5' ends of the lysozyme transcripts produced at different times after T4 infection show that the early mRNA is initiated some distance upstream from the gene. The early mRNA is not translated because of a stable secondary structure which blocks the translational initiation site. The stable RNA structure has been demonstrated by nuclease protection *in vivo*. After DNA replication begins, two late promoters are activated; the late transcripts are initiated at sites such that the secondary structure can not form, and translation of the late messages occurs.

INTRODUCTION

Bacteriophage T4 infection of *E. coli* is accompanied by a large variety of mechanisms that lead to precise regulation of the output of specific genes (1, 2, 3, 4, 5). Regulation of transcriptional initiation yields early, middle, and late mRNAs. Among early and middle gene products are those required for cessation of host macromolecular metabolism and for phage DNA replication; late gene products include those found in the mature phage and others that are required to lyse the cell (6). Superimposed upon global regulation of transcriptional initiation are complex perturbations to termination of transcription (2), as well as translational regulation of the expression of a large subset of phage pre-replicative genes (4, 7, 8, 9).

Hybridization analyses of lysozyme transcription and measurements of lysozyme activity *in vivo* have shown that a pre-replicative lysozyme transcript, present at high concentrations, is not translated *in vivo* or *in vitro* (10, 11, 12, 13, 14, 15). The data suggested that the early lysozyme transcript was initiated upstream from lysozyme and that the late transcript was initiated at or near the N-terminal end of the gene (14). The T4 lysozyme gene was recently sequenced (16). When we inspected the sequence around the lysozyme initiation codon, a hypothesis was easily formed that could explain the lack of translation of the early lysozyme transcript. The published sequence predicts a very stable secondary structure in the mRNA that would occlude ribosomal inspection of the Shine and Dalgarno domain

Bacteria and Bacteriophage

NapIV, NapIV^{strR}, and NapIV^{supDstrR} are described in Nelson et al. (23). T4D⁺ and the gene 55 amber mutation BL292 were from the lab stock collection.

Nucleic acids

The synthetic oligonucleotide primers were made on an Applied Biosystems model 380A DNA synthesizer. The synthetic primer, lys1, complementary to nucleotides +92 to +115 of the T4 lysozyme sequence (16) was used in all experiments except where noted. The synthetic oligonucleotide, lys2, is complimentary to nucleotides -227 to -206 of the lysozyme sequence. The synthetic oligonucleotides were purified by preparative gel electrophoresis and 5' end-labeled with T4 polynucleotide kinase according to manufacturers instructions.

RNA from T4 infected cells was isolated using a method similar to one described by Hagen and Young (24). Briefly, cells were grown in M9 + 2ug/ml B1 at 30°C to a density of 3×10^8 cells/ml. The phage were then added to the cells at a multiplicity of 10. Aliquots of these infected cells (20 ml) were taken at various times after infection and placed in centrifuge tubes containing 10 grams of ice. Cells were pelleted by centrifugation at 10K rpm for 15 minutes at 4°C and resuspended in 0.45 ml ice-cold 0.15M NaCl containing 0.001% gelatin. Then 0.05 ml SDS-extraction buffer (0.5M tris-Cl pH 6.8, 0.02M EDTA, 10% SDS) was added and the mixture heated to 68°C for three minutes. Sodium acetate (pH 5.2) was added to a final concentration of 0.2M and the nucleic acids extracted four times with phenol:CHCl₃:isoamyl alcohol (25:24:1). Two ether extractions were used to remove residual phenol and the nucleic acids were precipitated with ethanol. The pellets were resuspended in 1 mM EDTA and stored at -20°C. Total nucleic acid concentration (RNA + DNA) was estimated assuming 1 O.D.₂₆₀ = 40 ug/ml.

Reverse transcription and RNA sequencing

Reverse transcription and RNA sequencing was done as described by Inoue and Cech (25) with the following changes: i) crude unpurified RNAs (30ug total nucleic acid per 10 ul annealing mix), rather than purified RNAs, were used in the annealing mixtures, ii) the annealings were done at 60°C for 3 minutes followed by immersion in a dry-ice/ethanol bath for 1 minute, iii) ddNTP's were used at 200 uM (except where noted), and iv) the sequencing and primer extension reactions were prewarmed to 48°C for 2 minutes before addition of reverse transcriptase. The sequencing and primer extension reactions were then incubated at 48°C for 30 minutes. Under these conditions, hybridization of the end-labeled primer was to RNA. Double-stranded T4 DNA, present in the crude RNA preparations does not serve as a template for reverse transcriptase under the conditions used (data not shown). For markers, dideoxy-sequencing reactions were done using nucleic acids extracted at 6 minutes from T4D⁺-infected NapIV. In all experiments, the 5' end-labeled primer was present in excess over the lysozyme transcripts as shown by a linear signal increase with increasing template concentration (data not shown).

RESULTS

Primer extension reveals the expected late transcripts

We mapped the 5' ends of the lysozyme transcripts present at different times after infection by primer extension with reverse transcriptase (Figure 2). Lysozyme transcription can be detected by this method beginning between 2 and 4 minutes post-infection. At both early and late times after infection, a transcript is present whose 5' end is at least 400 bp upstream from the lysozyme initiation codon. In fact, the distance is probably greater; the closest known early T4 promoter is over 3600 nucleotides upstream of lysozyme (2, 26). Other data support this location of the early lysozyme promoter (27). At late times after infection, we detect many shorter transcripts whose 5' ends are located just downstream from the two conserved late promoter sequences (Figure 2, lanes 10-30). In primer extension reactions, a band does not always indicate the presence of a 5' end; pausing and premature termination by AMV reverse transcriptase does occur (28). We have obtained similar results by mapping the 5' ends of the early and late lysozyme mRNAs using S1 nuclease (29). The bands at positions U(-23), C(-22), and A(-20), and the bands located at positions G(-9), G(-8) and A(-7) (Figures 2 and 7) probably represent the products of initiation from the two late T4 promoter sequences. The data of Christensen and Young (21, 22) show that late transcripts are usually initiated about five bases downstream from the conserved TATAAATA sequence. The reverse transcriptase stops corresponding to positions C(-13) and U(-11) are discussed below. Quantification of the late transcripts by densitometry of a lighter exposure (data not shown) shows that the transcripts initiated at eP1 are more abundant than transcripts initiated at eP2 (see below). The late lysozyme transcripts initiating at eP1 can be detected by approximately 10 minutes post-infection, which agrees with the first appearance of other late T4 transcripts (1, 30). The long, early transcript is still present at late times after infection, as was found by Kasai and Bautz (14).

In vivo structure mapping of the early transcript

The data in Figure 3a suggest that the proposed hairpin exists in vivo. The primer extension reactions reveal cleavages of the early transcript immediately adjacent to the base of the hairpin and opposite the bulge-loop; no other cleavages are seen in the loop or base-paired regions of the hairpin (Figure 3b). These bands probably reflect cleavages of the early transcript by ribonucleases in vivo. It is interesting that no cleavages in the hairpin loop which would generate translatable message are seen at early times.

Initiation versus degradation of early lysozyme transcripts

Most of the bands seen only at late times probably arise from initiation at the late promoter sequences and not from degradation of the early or late transcripts. If transcriptional initiation is inhibited by rifampicin at early times, the bands normally seen only at late times are much less abundant (Figure 4, lanes a-d). The appearance of the bands seen at late times is dependent on the product of gene 55 (data not shown), which is

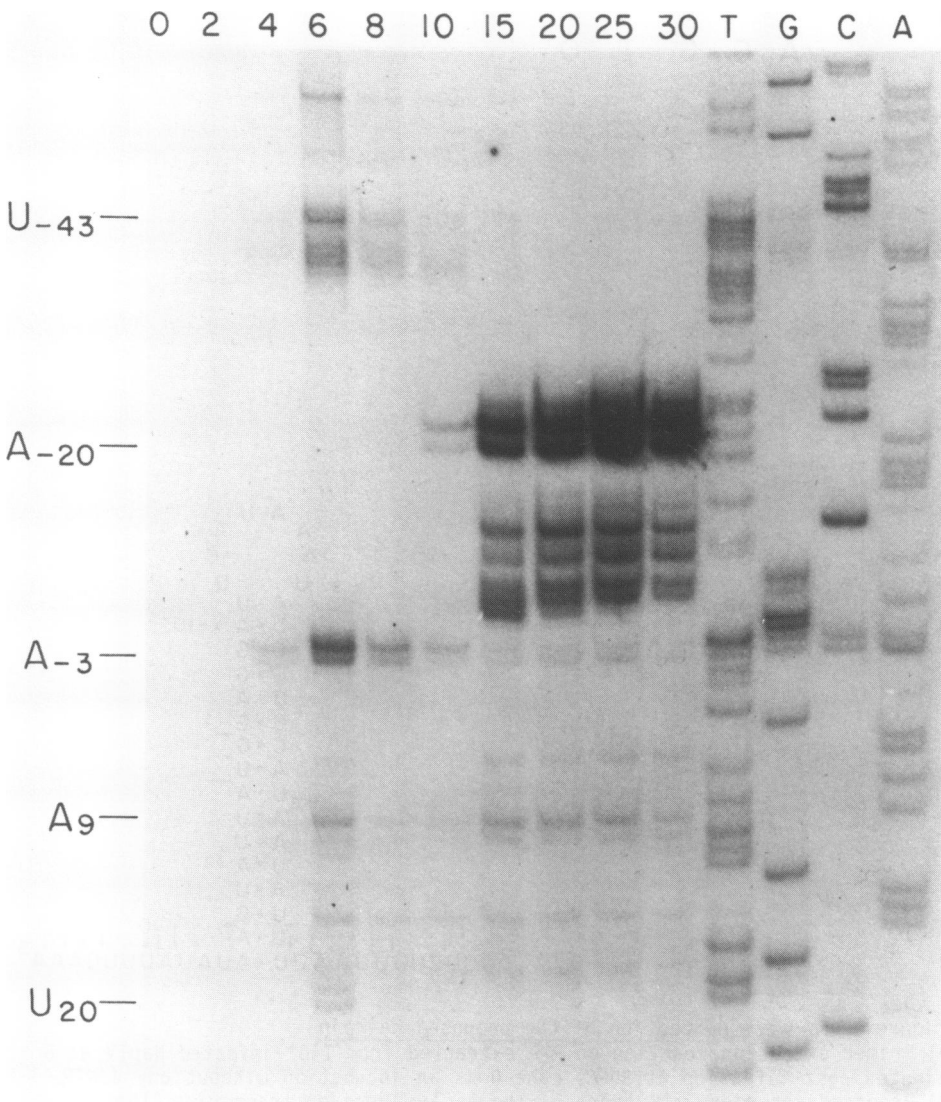


Figure 2. Time course of lysozyme transcription

Lanes 0 through 30 are primer extension reactions using RNA harvested from T4⁺-infected NapIV at 0, 2, 4, 6, 8, 10, 15, 20, 25 and 30 minutes post-infection at 30°C. Lanes T,G,C and A are primer extension reactions done using 6 minute RNA in the presence of ddATP, ddCTP, ddGTP and ddTTP, respectively; one may read the sequences directly.

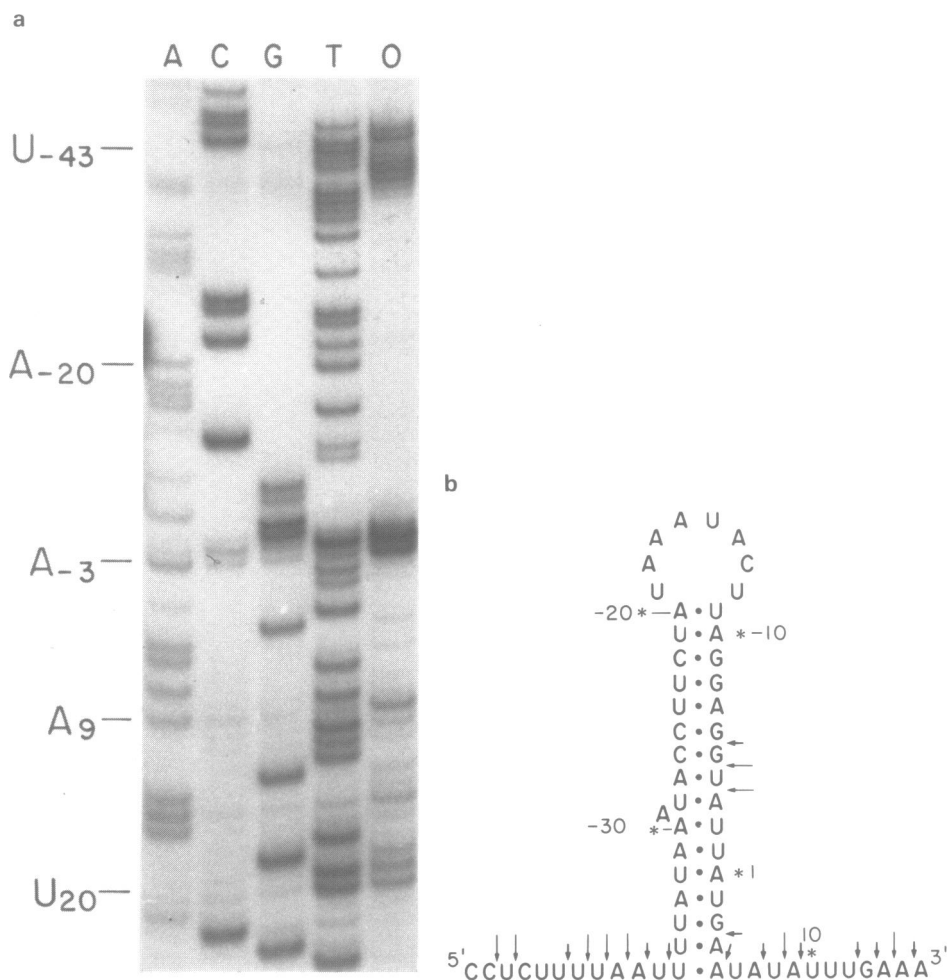


Figure 3. In vivo protection of the proposed hairpin
 a) Primer extension reaction on RNA extracted from T4D⁺-infected NapIV at 5 minutes post-infection at 30°C. Lane O is an incubation without any ddNTP.
 b) Proposed secondary structure of the early lysozyme transcript. The positions of the major and minor cleavages seen in (a) are indicated by long and short arrows, respectively, placed 5' to last nucleotide present in the template RNA.

required for transcriptional initiation at late T4 promoters (31, 3). These results strongly suggest that most of the 5' ends arise directly from a transcriptional initiation event.

If transcriptional initiation is inhibited at late times, the intensity of all the late bands subsequently decreases (Figure 4, lanes e-h).

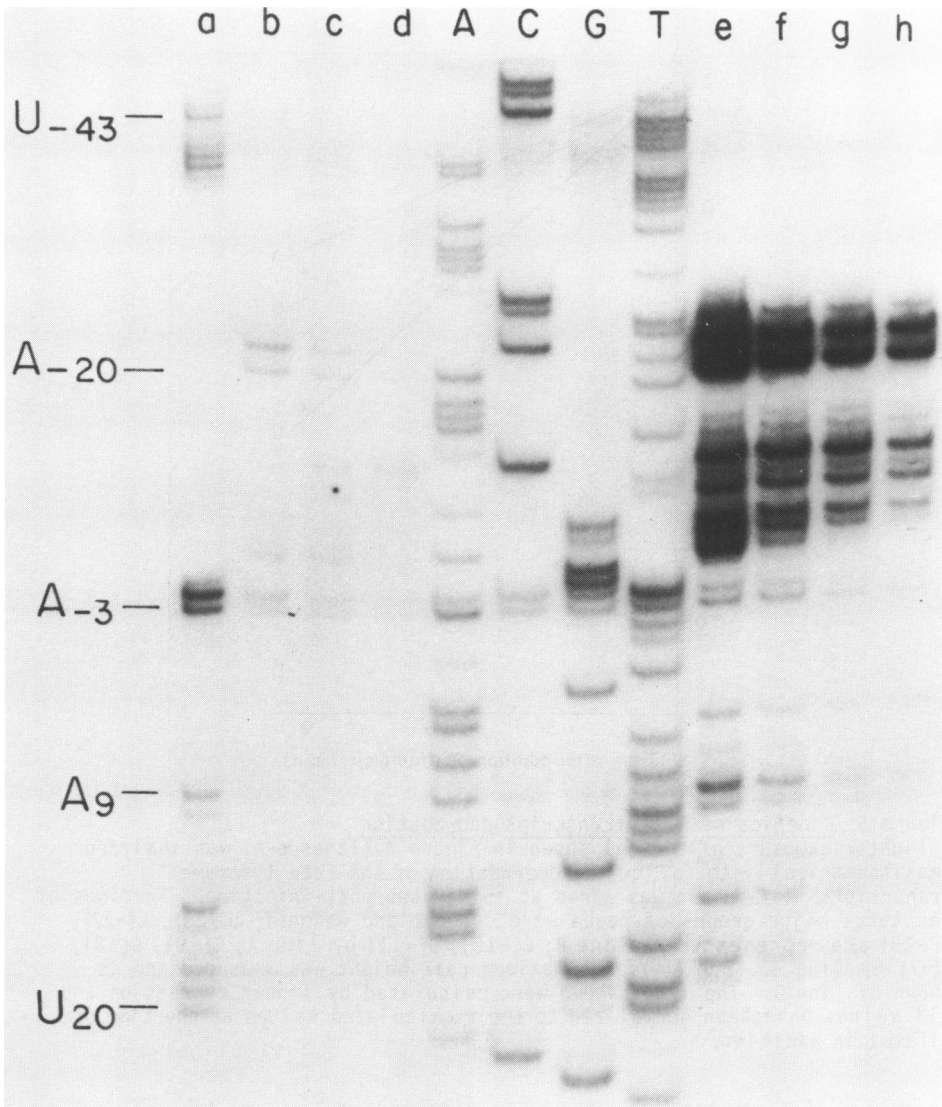


Figure 4. Degradation of early and late lysozyme transcripts

The 5' end labeled primer was used in primer extension reactions as described in Experimental Procedures. Lanes a-d: RNA was extracted from T4⁺-infected NapIV at 8 minutes post-infection at 30°C (lane a), rifampicin was then added immediately to the infected cells at a final concentration of 200 µg/ml, and samples were taken at 11 (lane b), 14 (lane c) and 17 minutes (lane d). Lanes e-h: RNA was extracted from T4⁺-infected NapIV at 15 minutes post-infection (lane e), rifampicin was then added immediately to the infected cells and samples taken at 18 (lane f), 21 (lane g) and 24 minutes (lane h).

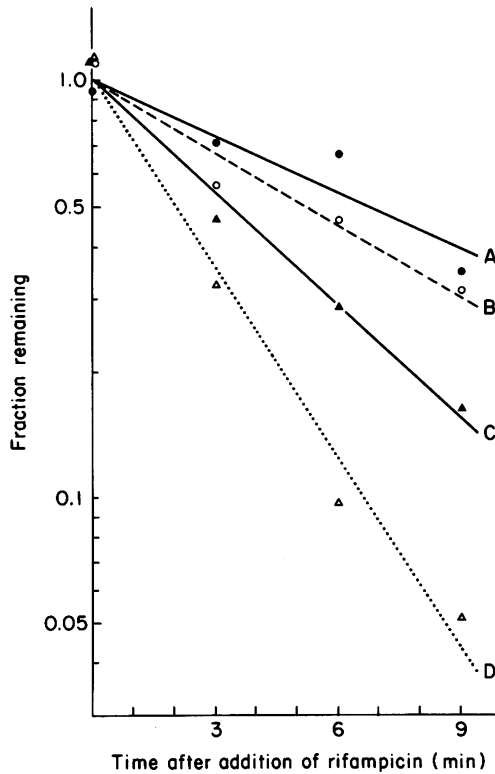


Figure 5. Kinetics of late transcript degradation

A lighter exposure of the gel shown in Figure 4 (lanes e-h) was analyzed densitometrically to follow the degradation of the late lysozyme transcripts. Rifampicin was added at 15 minutes post-infection. Tracings of the three major groups of peaks were cut-out and weighed; U(-23), C(-22), A(-20) are represented by line B; C(-13), U(-11) by line A; G(-9), G(-8), A(-7) by line C. For A(-7) the maximum peak height was measured and is shown by line D. The lines shown were calculated by linear regression and all values have been normalized to their calculated values at the time of rifampicin addition.

Quantification of these late transcripts by densitometry show that the transcripts initiating from eP1 have a half-life of over 5 minutes, while those initiating from eP2 have a half-life of approximately 3 minutes (Figure 5). The transcript initiating at A(-7), from eP2, is particularly unstable, having a half-life of only 2 minutes. These differences in half-lives may account totally for the levels of late transcripts seen in Figure 2, although late T4 promoters may have different strengths (32). The reverse transcriptase stops corresponding to C(-13) and U(-11) (Figure 2) are probably not initiation sites, since they are inappropriately spaced

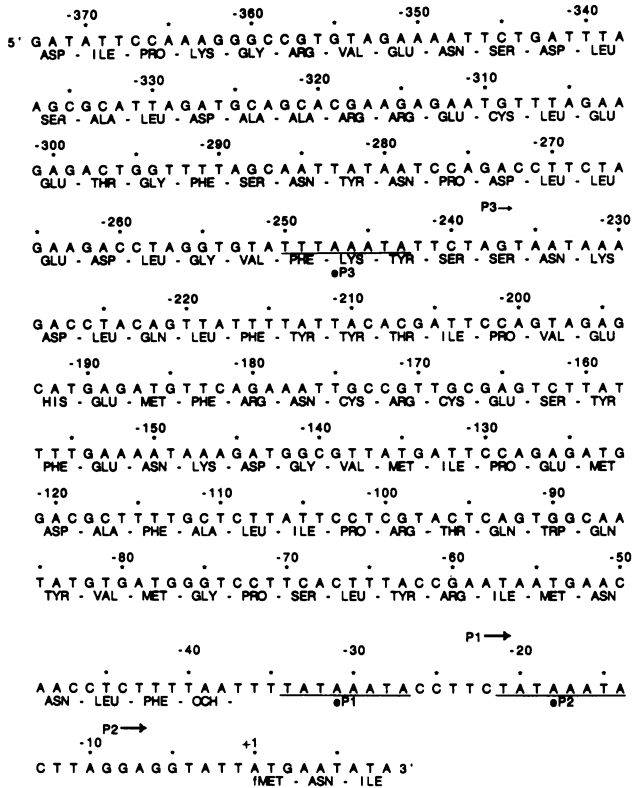


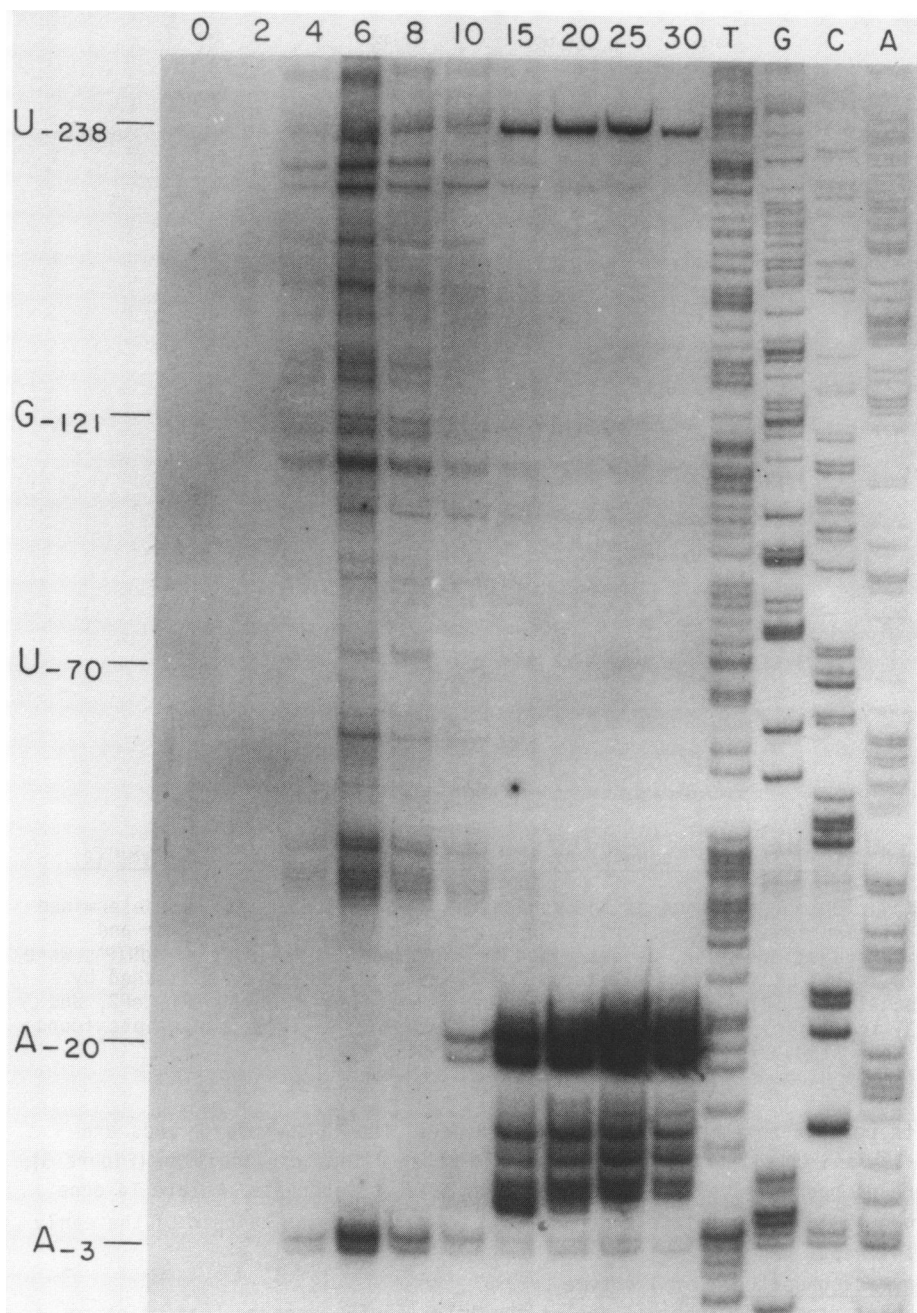
Figure 6. Nucleotide sequence of the region upstream of the lysozyme (e) gene

The nucleotide sequence of the region upstream from lysozyme was determined by dideoxy-sequencing using the long, early transcript as template and primers lys1 and lys2, as described in Experimental Procedures; ddNTP's were used at 100 μ M. The sequence from -134 to +9 was previously published by Owen et al. (16). The three late promoter sequences (labeled eP1, eP2, and eP3) are underlined. The approximate 5' ends of the late transcripts found in this study are labeled P1, P2 and P3.

with respect to the late promoter sequences. The RNAs whose 5' ends are C(-13) and U(-11) are lost very slowly after rifampicin addition (Figure 5), perhaps because they are generated from other transcripts. A late T4 gene product might facilitate endonucleolytic attack in the hairpin of the early transcript.

Sequencing upstream of lysozyme

We have used dideoxysequencing of the long, early lysozyme transcript to extend the known sequence upstream of lysozyme to -373 (Figure 6). The nucleotide sequence of the upstream region contains one open reading frame



encoding at least 111 amino acids. Codon usage analysis of this open reading frame indicates it is likely to code for a T4 protein (data not shown). However, there are no known T4 genes for over 3800 nucleotides upstream from lysozyme (26).

There is a minor, third late promoter (eP3) located at -245 (Figures 6 and 7). The transcript initiated at this promoter is sensitive to rifampicin and is dependent on the gene 55 product (data not shown). The sequence of this third late promoter includes TTAAATA, differing by one nucleotide from the late T4 promoter consensus sequence TATAAATA (21, 22).

DISCUSSION

We conclude that early lysozyme mRNA is translationally inactive due to the presence of a specific secondary structure, and that the late lysozyme mRNAs are translatable because the locations of the late promoters yield transcripts that do not contain the inhibitory secondary structure. Our data rationalize most of the data heretofore published for lysozyme transcription and translation (10, 11, 12, 13, 14, 15). Our data strongly suggest that the late RNAs arise from transcriptional initiation. The presumptive 5' ends of the late messages are just downstream from DNA sequences known to serve as late promoters for other T4 genes (21, 22, 28, 33) and their appearance is dependent on the gene 55 product. Lysozyme expression under laboratory conditions depends on the T4 gene 55 product (34) which directly participates in initiation of late transcription (31, 3, 1). The early, but not the late transcripts, are seen with RNA prepared from su^- cells infected with a gene 55 amber mutant.

The extent of translational inhibition by the secondary structure (Figure 1) is considerable; Kasai and Bautz (14) showed that the late lysozyme transcript is translated in vivo more than one hundred-fold more efficiently than the early transcript. Recently, Perry et al. (35) cloned the T4 lysozyme gene, along with about 100 bases of the upstream sequence, under the control of the trp promoter. This construct expressed only low levels of lysozyme. However, when the untranslated sequences immediately upstream from lysozyme were removed, a higher level of lysozyme synthesis was found. Although secondary structures are usually assumed to limit translational initiation (36), in only a few cases have quantitative measurements been made (18, 19). When the early upstream promoter(s) is used to direct coupled lysozyme transcription and translation in vitro, the long, early mRNA is translated (37, 38). The conditions of coupled cell-

Figure 7. Upstream late lysozyme promoter

Lanes 0 through 30 are primer extension reactions using RNA harvested from T4⁺-infected NapIV at 0, 2, 4, 6, 8, 10, 15, 20, 25 and 30 minutes post-infection at 30°C. The 5' ends of the transcript initiated at the minor upstream late promoter are located at nucleotides A(-238) and G(-237).

free expression (and the methods of making cell-free extracts) might facilitate nucleolytic attack in the loop of the hairpin. Such attack does not activate early lysozyme mRNA that is added to a cell-free translation system (11), perhaps because the early transcripts are cut 3' to the hairpin more rapidly than at the hairpin. Coupled systems could present the hairpin to hypothetical nucleases prior to the moment of transcription of the coding portion of the gene.

T4 late promoters may be identified by the conserved sequence TATAAATA positioned about five bases upstream of the initiating base (21, 22). Over 25% of the T4 genome has now been sequenced, and everywhere tested the sequence TATAAATA is found associated with a late transcript (Stormo, unpublished data); the consensus sequence is probably sufficient for gene 55-mediated transcription (33). The lysozyme gene is preceded by two such sequences (Figure 1). The transcripts from eP2, which initiate within the lysozyme Shine and Dalgarno domain, are less stable than those initiating just upstream (Figure 5). Presumably these eP2 transcripts are translated less well than transcripts from eP1; ribosomes and inactivating nucleases may compete for the 5' ends of the late lysozyme mRNAs (39).

We would like to know why the lysozyme gene is transcribed at early times. Early lysozyme transcription is not a means to achieve expression of genes downstream, since the next gene, ipIII (16), is served by its own early promoter (14, 27). The regulation of lysozyme translation by an mRNA secondary structure is similar to the regulation of Tn10 transposase expression (40). Tn10 transposase expression is tightly regulated; translation of transposase from transcripts initiated in host sequences outside of Tn10 is prevented by the formation of a stable secondary structure which sequesters the transposase initiation codon. A Tn10 promoter located in a position analogous to the locations of eP1 and eP2 is used to initiate transcripts which are unable to form the inhibitory secondary structure. If expression of T4 lysozyme and the Tn10 transposase from promoters located far upstream is undesirable, we wonder why transcriptional terminators, located immediately upstream from these genes did not evolve first. In the case of Tn10, perhaps the structural elements involved in transposition would be destroyed by such a transcriptional terminator. In the case of T4 lysozyme, there may be conditions in which lysozyme expression from the long early transcript is necessary for survival of the phage. Interestingly, the T4 soc gene, which codes for a late, non-essential, T4 capsid protein, is transcribed at both early and late times, but is translated only at late times. The inhibition of early soc translation, like that of lysozyme, can be explained by an occluding secondary structure in the early, but not the late mRNA (41). Translation of lysozyme and soc from the early transcripts may insure survival of the phage under some conditions; if so, a mechanism must be found to activate these mRNAs for translation. We are currently testing for such mechanisms.

ACKNOWLEDGEMENTS

We are grateful to Gerry Smith, Ronald Wetzel, Jeanne Perry and Herbert Heyneker for communicating results before publication; Tan Inoue and Thomas Schneider for help with experiments; and Michael Parker, Duane Mooney, and Gus Doermann for helpful discussions. The lysozyme primers were made by Margit Gayle and Roger Saunders. This work was supported by Public Health Service Grants AI-09456, GM28755, and GM19963 from the National Institutes of Health.

*Present address: Synergen Associates, 1885 33rd Street, Boulder, CO 80301, USA

REFERENCES

1. Rabussay, D. (1983) In Matthews, C.K., Kutter, E.M., Mosig, G. and Berget, P.B. (eds). Bacteriophage T4, American Society for Microbiology, Washington, D.C., pp. 167-173.
2. Brody, E., Rabussay, D., and Hall, D.H. (1983) In Matthews, C.K., Kutter, E.M., Mosig, G. and Berget, P.B. (eds), Bacteriophage T4. American Society for Microbiology, Washington, D.C., pp. 174-183.
3. Geiduschek, E.P., Elliott, T., and Kassavetis, G.A. (1983) In Matthews, C.K., Kutter, E.M., Mosig, G. and Berget, P.B. (eds). Bacteriophage T4, American Society for Microbiology, Washington, D.C., pp. 189-192.
4. Wiberg, J.S., and Karam, J.D. (1983) Matthews, C.K., Kutter, E.M., Mosig, G. and Berget, P.B. (eds), Bacteriophage T4, American Society for Microbiology, Washington, D.C., pp. 193-201.
5. Gold, L., Inman, M., Miller, E., Pribnow, D., Schneider, T., Shinedling, S., and Stormo, G. (1984). Clark, B.F.C. and Petersen, H.U. (eds). Gene Expression: The Translational Step and its Control, Munksgaard, Copenhagen, pp. 379-394.
6. Rabussay, D., and Geiduschek, E.P. (1977) In H. Fraenkel-Conrat and R.R. Wagner (eds), Comprehensive Virology. (New York: Plenum Press), Vol. 8, p. 1-196.
7. Karam, J., Gold, L., Singer, B.S., and Dawson, M. (1981) Proc. Nat. Acad. Sci. USA 78, 4669-4673.
8. Lemaire, G., Gold, L., and Yarus, M. (1978) J. Mol. Biol. 125, 73-90.
9. von Hippel, P.H., Kowalczykowski, S.C., Lonberg, N., Newport, J.N., Paul, L.S., Stormo, G.D., and Gold, L. (1982) J. Mol. Biol. 162, 795-818.
10. Bautz, E.K.F., Kasai, T., Reilly, E., and Bautz, F.A. (1969) Proc. Nat. Acad. Sci. 55, 1081-1088.
11. Gesteland, R.F., and SaTser, W. (1969) Genetics 61, supplement 1, 429-437.
12. SaTser, W., Gesteland, R.F., and Bolle, A. (1967) Nature 215, 588-591.
13. Kasai, T.F., Bautz, E.K.F., Guha, A., and Szybalski, W. (1968) J. Mol. Biol. 34, 709-712.
14. Kasai, T.F., and Bautz, E.K.F. (1969) J. Mol. Biol. 41, 401-417.
15. Young, E.T., and VanHouwe, G. (1970) J. Mol. Biol. 51, 605-619.
16. Owen, J.E., Schultz, D.W., Taylor, A., and Smith, G.R. (1983) J. Mol. Biol. 165, 229-248.
17. Shine, J., and Dalgarno, L. (1974) Proc. Nat. Acad. Sci. USA 71, 1342-1346.

18. Hall, M.N., Gabay, J., Debarbouille, M., and Schwartz, M. (1982) *Nature* 295, 616-618.
19. Munson, L.M., Stormo, G.D., Niece, R.L., and Reznikoff, W.S. (1984) *J. Mol. Biol.* 177, 663-683.
20. Halling, S.M., Simons, R.W., Way, J.C., Walsh, R.B., and Kleckner, N. (1982) *Proc. Nat. Acad. Sci. USA* 79, 2608-2612.
21. Christensen, A.C., and Young, E.T. (1982) *Nature* 299, 369-371.
22. Christensen, A.C., and Young, E.T. (1983) In Matthews, C.K., Kutter, E.M., Mosig, G. and Berget, P.B. (eds), *Bacteriophage T4*. Society for Microbiology, Washington, D.C., pp. 184-188.
23. Nelson, M.A., Ericson, M., Gold, L., and Pulitzer, J.F. (1982) *Mol. Gen. Genet.* 188, 60-68.
24. Hagen, F.S., and Young, E.T. (1978) *J. Virol.* 26, 793-804.
25. Inoue, T., and Cech, T.R. (1985) *Proc. Nat. Acad. Sci. USA* 82, 648-652.
26. Gram, H., Liebig, H.-D., Hack, A., Niggemann, E., and Rueger, W. (1984) *Mol. Gen. Genet.* 194, 232-240.
27. Black, L.W., and Gold, L.M. (1971) *J. Mol. Biol.* 60, 365-388.
28. Kassavetis, G.A., and Geiduschek, E.P. (1982) *EMBO J.* 1, 107-114.
29. Christensen, A.C. (1982) Doctoral thesis, University of Washington, Seattle, Washington.
30. O'Farrell, P.Z., and Gold, L.M. (1973) *J. Biol. Chem.* 248, 5502-5511.
31. Kassavetis, G.A., and Geiduschek, E.P. (1984) *Proc. Nat. Acad. Sci. USA* 81, 5101-5105.
32. Elliott, T., Kassavetis, G.A., and Geiduschek, E.P. (1984) *Virology* 139, 260-282.
33. Elliott, T., and Geiduschek, E.P. (1984) *Cell* 36, 211-219.
34. Pulitzer, J.F., and Geiduschek, E.P. (1970) *J. Mol. Biol.* 49, 489-507.
35. Perry, L.J., Heyneker, H.L., and Wetzel, R. (1985) *Gene* 38, 259-264.
36. Gold, L., Pribnow, D., Schneider, T., Shinedling, S., Singer, B.S., and Stormo, G. (1981) *Ann. Rev. Micro.* 35, 365-403.
37. Schweiger, M., and Gold, L. (1969) *Proc. Nat. Acad. Sci.* 63, 1351-1358.
38. Gold, L., and Schweiger, M. (1970) *J. Biol. Chem.* 245, 2255-2258.
39. Kennell, D.E. (1986) Reznikoff, W.S. and Gold, L. (eds). *From Gene to Protein: Steps Dictating the Maximal Level of Gene Expression*, Benjamin/Cummings: New York, in press.
40. Davis, M.A., Simons, R.W., and Kleckner, N. (1985) *Cell* 43, 379-387.
41. MacDonald, P.M., Kutter, E., and Mosig, G. (1984) *Genetics* 106, 17-27.
42. Tinoco, T., Boere, P.N., Dengler, B., Levine, M.D., Ulhenbeck, O.L., Crothers, D.M., and Gralla, J. (1973) *Nat. New Biol.* 246, 40-41.