Volume 15 Number 19 1987 Nucleic Acids Research

# Purification in an active form of the phage $\phi$29 protein p4 that controls the viral late transcription

Isabel Barthelemy, José M.Lázaro, Enrique Méndez[1], Rafael P.Mellado and Margarita Salas

Centro de Biología Molecular (CSIC-UAM), Universidad Autónoma, Canto Blanco, 28049 Madrid and [1]Servicio de Endocrinología, Centro Ramón y Cajal, Carretera de Colmenar Viejo, Km. 9,1, 28034 Madrid, Spain

Received August 5, 1987; Accepted September 4, 1987

ABSTRACT

The phage Ø29 protein p4, that controls viral late transcription, was highly purified from Escherichia coli cells harbouring a gene 4-containing plasmid. This protein, representing about 6% of the total cellular protein, was obtained in a highly purified form. The protein was characterized as p4 by amino acid analysis and $NH_2$-terminal sequence determination. The purified protein was active in an in vitro transcription assay, allowing specific initiation of transcription at the $\phi$29 A3 late promoter in the presence of Bacillus subtilis $\sigma^{43}$-RNA polymerase holoenzyme.

## INTRODUCTION

Bacillus subtilis phage Ø29 contains a linear, double-stranded DNA of 19,285 bp (1) with the viral protein p3 covalently linked at the two 5' ends (2). Phage Ø29 late genes are clustered in the middle of the genome, flanked by the early genes, localized at both ends on the viral genetic map (3). In vivo early transcription takes place from the light DNA strand from right to left, whereas the late transcription occurs from the heavy DNA strand in the opposite direction (4,5). Synthesis of early RNA is carried out by the host $\sigma^{43}$-RNA polymerase holoenzyme, the product of the viral gene 4 being required to transcribe the Ø29 late genes (6). The initiation sites of the RNA transcripts synthesized in vivo in Bacillus subtilis infected with bacteriophage Ø29 have been mapped by S1 protection experiments (7) and the early and late promoter sequences determined (8). The B. subtilis RNA polymerase containing the $\sigma^{43}$ subunit basically recognizes the same Ø29 promoters in vitro as those used in vivo (9). The Ø29 early promoters have sequences homologous to the consensus ones recognized by the $\sigma^{43}$-RNA polymerase from B. subtilis at their

-10 and -35 regions (8,9), whereas the main late promoter A3 only shares homology with the consensus sequence at its -10 region. Transcription from the late promoter can take place *in vitro* after prolonged incubation with the *B. subtilis* or *E. coli* RNA polymerase holoenzymes in the absence of protein p4 (9,10), although this transcription is negligible if the salt is increased in the *in vitro* assay (10,11).

The Ø29 gene 4 was cloned and its product, protein p4, overexpressed in *E. coli* (12,13). In this paper we report the purification of protein p4 in an active form from the *E. coli* cells. An *in vitro* assay has been deviced in which there is specific recognition of the Ø29 A3 late promoter by the *B. subtilis* $\sigma^{43}$-RNA polymerase in the presence of purified protein p4.

## MATERIALS AND METHODS

a) *Bacterial strains, plasmids, proteins and radioactive compounds.*

*B. subtilis* 110NA trp$^-$ *spoA* *su*$^-$ was used as a source of *B. subtilis* RNA polymerase and as a host for the bacteriophage Ø29. *E. coli* strain M72 Sm$^R$ *lacZam* *bio-uvrB* Δtrp EA2 (λ *Nam7Nam53* *cI857ΔH1*), here designated K12ΔH1Δtrp, was obtained from E. Remaut and used as a host for the propagation of plasmids pRMn25 (13), pRMcn385 (13), pRMw51 (12), pRMs463 (12) and pRMt121 (14).

Restriction endonucleases were from New England Biolabs, nuclease S1 from PL Biochemicals, calf intestinal alkaline phosphatase, bacteriophage T4 polynucleotide kinase and the large fragment of the *E. coli* DNA polymerase I (Klenow enzyme) from Boehringer Mannheim. [$\gamma$-$^{32}$P]ATP (~3000 Ci/mmol), [$\alpha$-$^{32}$P]dATP (~400 Ci/mmol), [$\alpha$-$^{32}$P]UTP (~400 Ci/ mmol) and [$^{35}$S]-methionine (1200 Ci/mmol) were from Amersham International plc.

b) *Preparation and labelling of nucleic acids. S1 mapping and DNA sequence analysis.*

Ø29 DNA was prepared as described before (6). Total RNA was isolated from Ø29 infected cells either in the presence or absence of chloramphenicol as described (7). Labelling of DNA fragments at their 5′ termini with polynucleotide kinase and [$\gamma$-$^{32}$P]ATP and at their 3′ ends with [$\alpha$-$^{32}$P]dATP and the Klenow enzyme was as described (15). Double and single-stranded DNAs were purified by diffusion from polyacrylamide gels. Conditions

for the protection of 5' end-labelled DNA or RNA to the S1 digestion were as described (7). DNA sequencing reactions were as described (16).

c) Protein analysis and purification. In vitro transcription assay.

E. coli cells harbouring the p4-overproducer plasmid pRMn25 or the plasmid pRMs463, containing a nonsense mutation in the gene 4 cloned sequences, were grown at 30°C, induced at 42°C for 2 h and labelled for 10 min with [$^{35}$S]-methionine (80 $\mu$Ci/ml) as described (12). Fifteen g of E. coli K12ΔH1Δtrp cells transformed with the plasmid pRMn25 and induced for 2 h at 42°C were mixed with $^{35}$S-labelled cells, ground with alumina (30 g) and extracted with buffer A (50 mM Tris-HCl, pH 7.0, 5% glycerol) containing 1 M NaCl. The lysate was centrifuged for 10 min at 20,000 x g and the pellet reextracted with the same buffer. The two supernatants were pooled, diluted with buffer A to bring the NaCl concentration down to 0.5 M and passed through a phosphocellulose column (13 cm x 3.6 cm) equilibrated with buffer A containing 0.5 M NaCl. The column was extensivelly washed with the same buffer and protein p4 was recovered by addition of buffer A containing 1 M NaCl. The p4-containing fractions were pooled and made 0.5 M NaCl with buffer A before loading them in a second phosphocellulose column (5 cm x 1.4 cm) equilibrated with buffer A containing 0.5 M NaCl. After extensive washing, protein p4 was eluted with the same buffer containing 0.6 M NaCl. The equivalent fractions from a parallel purification of cells containing the plasmid pRMs463, carrying the nonsense mutation in gene 4, were used to prepare rabbit antibodies. The purified IgGs (17) were then covalently bound to CNBr-activated Sepharose 4B (Pharmacia), according to the supplier, to prepare an affinity column equilibrated with PBS buffer (8.1 mM $Na_2HPO_4$, 1.47 mM $KH_2PO_4$, pH 7.4, 0.13 M NaCl, 2.68 mM KCl)) containing 0.01% Nonidet-P40. The pooled p4-containing fractions (half of the sample) were passed through the affinity column (5.5 cm x 1.5 cm), protein p4 being recovered in the flow-through fractions. The protein was finally concentrated in a phosphocellulose column (0.7 cm x 0.8 cm), and protein p4 was eluted with buffer A containing 1 M NaCl.

B. subtilis $\sigma^{43}$-RNA polymerase was purified as described (6). Protein concentration was determined by the method of

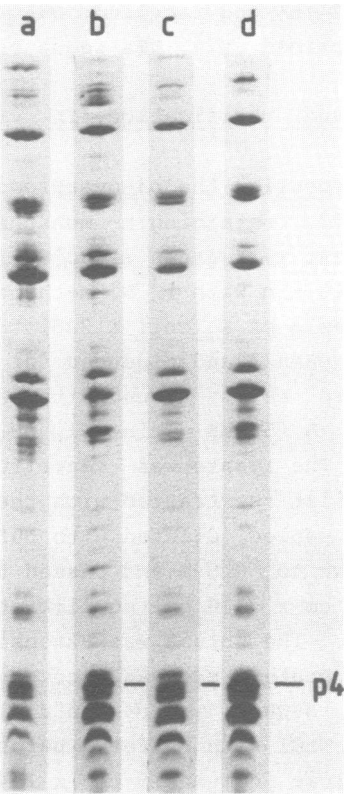


Figure 1. Overproduction of the Ø29 protein p4 in E. coli. The proteins contained in E. coli cells carrying the indicated recombinant plasmids were analyzed in 10-20% polyacrylamide gels in the presence of SDS after 2 h of induction at 42°C, as indicated in Materials and Methods, section c. a) Plasmid pRMs463; b) plasmid pRMt121; c) plasmid pRMw51; d) plasmid pRMn25. The position of protein p4 is indicated.

Bradford (18). Electrophoresis of proteins was carried out in slab gels containing 10-20% acrylamide gradients in the presence of 0.1% SDS (19). After electrophoresis, the proteins were stained as described by Fairbanks et al. (20).

The in vitro transcription assay was as described by Sogo et al. (6) modified as indicated in Mellado et al. (9).

d) Amino acid analysis and protein sequence determination.

Purified protein p4 was hydrolyzed with 100 µl of 5.7 M HCl containing 0.05% (v/v) 2-mercaptoethanol in evacuated and sealed tubes at 110° for 24, 48 and 72 h. The analysis were performed

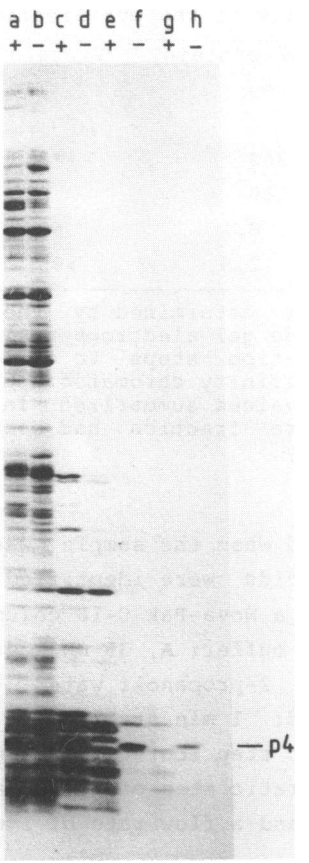


Figure 2. SDS-polyacrylamide gel electrophoresis of purified protein p4. Proteins at various purification steps were subjected to SDS-electrophoresis in slab gels containing a 10-20% polyacrylamide gradient. After electrophoresis the proteins were stained as indicated in Materials and Methods, section c. Lanes: a,b) extract (25.5 and 38.7 μg); c,d) first phosphocellulose column, 1 M NaCl eluate (23.6 and 12.5 μg); e,f) second phosphocellulose column, 0.6 M NaCl eluate (2.7 and 1.4 μg); g,h) affinity column (0.9 and 0.03 μg). + and - symbolize the use of cells harbouring plasmid pRMn25 and pRMs463, respectively. The position of protein p4 is indicated.

on a Beckman 121-MB Analyzer equipped with a Shimadzu Integrator model CR-3A.

Protein P4 was sequenced in a Beckman sequencer (model 890 M) according to the method of Edman and Begg (21). The sequence of the protein was determined using the Beckman protein/peptide/micro/macrosequencing program. To reduce peptide washout, 1 mg

Table 1. Summary of the purification of protein p4

| | Total protein, mg | Protein p4, % | Total protein p4, mg |
|---|---|---|---|
| Extract | 1360 | 3.3 | 45.5 |
| Phosphocellulose 1 | 180 | 8.8 | 15.8 |
| Phosphocellulose 2 | 8.1 | 69 | 5.6 |
| Affinity chromat. | 2.5 | >95 | 2.5 |

Total protein was determined by the method of Bradford (18). SDS-polyacrylamide gel electrophoresis was carried out at the different purification steps to follow the presence of protein p4. For the affinity chromatography step, half of the sample was used. The values summarized in this step are those expected if the entire fraction had been subjected to the affinity chromatography.

of polybrene was added when the sample was applied on the cup (22). The PTH-amino acids were identified on a reverse-phase HPLC system based upon a Nova-Pak C-18 column (Waters) and eluted with the following buffer: A, 35 mM sodium acetate, pH 5.0: acetonitrile (5:1); B, 2-propanol: water (3:2). The following Waters program was used: 1 min at 0% B; 3.5 min, from 0% B to 40% B with a hyperbolic step (curve 4 in a Waters M-680 controller) followed by isocratic step of 3 min at 40% B. The column was mantained at 38°C and a flow rate of 1 ml/min was used.

## RESULTS AND DISCUSSION

### 1. Synthesis and purification of protein p4.

*E. coli* K12ΔH1Δtrp cells, carrying different p4-overproducer plasmids, were grown and induced for 2 h as indicated in Materials and Methods, section c, and the synthesized polypeptides analyzed on SDS-polyacrylamide gels. Figure 1 shows that plasmid pRMn25 (lane d) is the higher producer, being the amount of protein p4 (Mr 15,000) synthesized approximately 6% of the total cellular protein. Therefore, *E. coli* cells carrying plasmid pRMn25 were used as a source of the ϕ29 protein p4. The cells were grown and lysed as indicated in Materials and Methods, section c, and the extracted proteins passed sequentially through two phosphocellulose columns, being protein p4 finally purified by an affinity column made with rabbit IgGs against the peptides present in a parallel

Table 2. Amino acid composition of purified protein p4

| Residue | Predicted[a] | Observed[b] |
|---|---|---|
| Asp+Asn | 13 | 12.2 |
| Thr | 4 | 5.1 |
| Ser | 7 | 7.2 |
| Glu+Gln | 12 | 12.9 |
| Pro | 4 | 3.5 |
| Gly | 4 | 8.4 |
| Ala | 6 | 7.1 |
| Val | 8 | 7.7 |
| Met | 4 | 2.5 |
| Ile | 5 | 5.0 |
| Leu | 9 | 8.9 |
| Tyr | 7 | 6.3 |
| Phe | 7 | 6.6 |
| Trp | 4 | N.D. |
| Lys | 12 | 11.0 |
| His | 3 | 3.0 |
| Arg | 11 | 10.0 |

[a]Amino acid composition predicted from the nucleotide sequence (23,24). The amino-terminal Met, which is removed by processing, has not been included.

[b]Amino acid composition of purified protein p4. Each value represents the average after 24, 48 and 72 h hydrolysis except the values for Thr and Ser that were extrapolated to zero hours of hydrolysis. Values represent residues per molecule. Trp was not determined (N.D.).

purification of cells carrying plasmid pRMs463 that contains a nonsense mutation in the cloned gene 4. Figure 2 shows the analysis by SDS-polyacrylamide gel electrophoresis of the different fractions of the purification. Densitometric analysis of the protein at the final step indicated that p4 was more than 95% pure. Table 1 gives a summary of the purification. About 1.3 mg of protein p4 were obtained from 15 g of *E. coli* cells harbouring the recombinant plasmid pRMn25.

2. Amino acid analysis of purified protein p4.

Table 2 shows that the amino acid analysis of the purified protein p4 gives values very similar to those deduced from the DNA sequence (23,24). Moreover, the $NH_2$-terminal sequence of

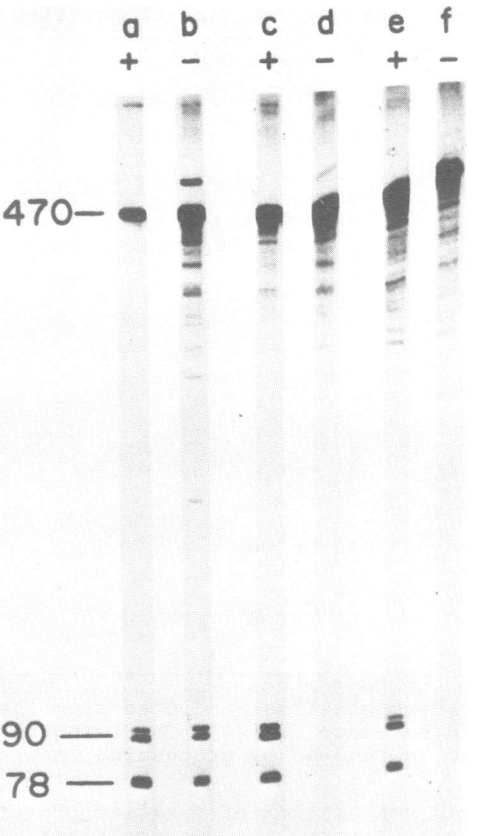


Figure 3. Activity of protein p4 at different salt concentrations. The Ø29 DNA HindIII H fragment (~80 ng) was incubated with B. subtilis $\sigma^{43}$-RNA polymerase (2.4 μg) in the presence of 1 μg of protein p4 purified through the second phosphocellulose column (+) or the equivalent volume from the parallel purification of cell extracts containing the plasmid pRMs463 (−). The in vitro run-off assay was performed as indicated in Materials and Methods, section c. The transcripts were analyzed on a denaturing 4% polyacrylamide gel containing 7 M urea. The transcripts sizes in nucleotides are indicated at the left margin. The 470 nt and the 78 nt ones correspond to transcription initiated at the A2c and A3 promoters, respectively. The different ammonium sulfate concentrations present in the incubation mixture were 72 mM (a,b), 90 mM (c,d) and 108 mM (e,f).

purified protein p4 showed the sequence Pro-Lys-Thr-Gln-Arg-Gly-Ile-Tyr-His, in agreement with the one predicted from the nucleotide sequence if the first Met residue is processed in E. coli.

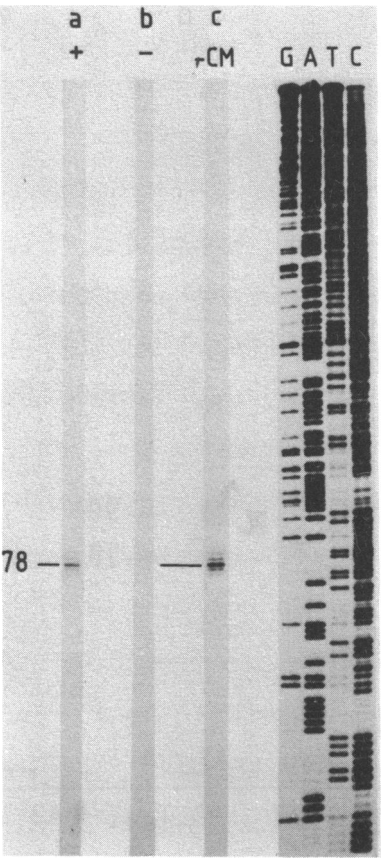


Figure 4. Specific transcription initiation at the A3 promoter in the presence of protein p4. Transcripts made in the in vitro run-off assay (see legend to Figure 3) by the B. subtilis $\sigma^{43}$-RNA polymerase (a) in the presence of 1 $\mu$g of protein p4 purified through the second phosphocellulose step (+) or (b) in the presence of the equivalent volume from the parallel purification of cell extracts containing the plasmid pRMs463 (-) were hybridized to the 5' end-labelled late strand from the Ø29 DNA HindIII H fragment and digested with the S1 nuclease as indicated in Materials and Methods, section b. Twenty five $\mu$g of RNA made in cells infected with Ø29 wild type in the absence of chloramphenicol (-CM) were equally processed (c). The protected oligodeoxinucleotides were analyzed on a denaturing 8% polyacrylamide gel containing 7 M urea. Sequencing reactions (G,A,T,C) were from the plasmid pAZa305 of known sequence (A. Zaballos, unpublished results). The size of the protected oligonucleotide is indicated in the left margin.

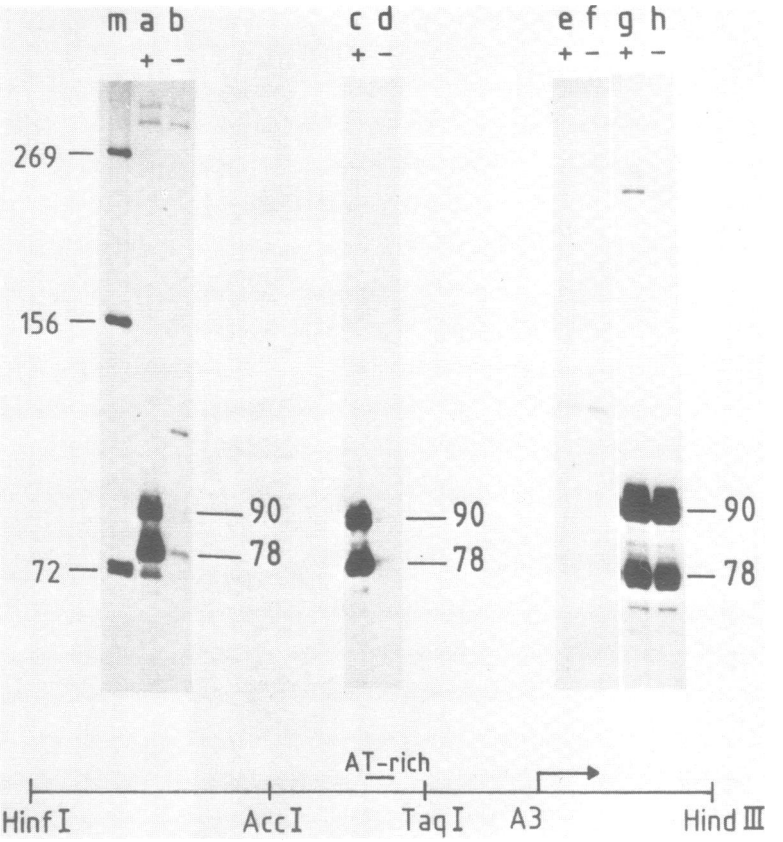


Figure 5. Template requirements for the in vitro transcription mediated by protein p4. The Ø29 DNA HindIII H fragment was shortened with the indicated endonucleases and the different fragments were used as templates in the in vitro run-off assay. The incubation mixtures were as in the legend of Figure 3, but containing 84 mM (a-f) or 36 mM (g,h) ammonium sulfate. +, indicates the presence of protein p4 purified through the affinity column (0.25 μg); -, indicates the presence of the equivalent volume from the parallel purification of cell extracts containing the plasmid pRMs463. Templates: (a,b) HinfI-HindIII fragment, 1308 bp long (28 ng); (c,d) AccI-HindIII fragment, 198 bp long (40 ng); (e-h) TaqI-HindIII fragment, 129 bp long (12 ng). The transcripts were analyzed on a denaturing 4% polyacrylamide gel containing 7 M urea. Figures at the left margin are the sizes of the Ø29 DNA HindIII L (269), M (156) and N (72) fragments carried as markers (m). Sizes of the run-off transcripts are indicated at the right margins. The arrow in the map indicates the start point and direction of transcription from the A3 late promoter.

3. Protein p4 stimulates in vitro transcription initiation at the Ø29 A3 late promoter.

The Ø29 A3 late promoter is recognized in vitro by the B. subtilis $\sigma^{43}$-RNA polymerase at low (10) but not at high salt concentration (10,11). To assay the activity of the purified protein p4 on directing the transcription from the A3 promoter, the $\sigma^{43}$-RNA polymerase was incubated with the 763 bp long HindIII H fragment from Ø29 DNA, at high salt, in the presence or absence of protein p4 and [$\alpha$-$^{32}$P]UTP. The HindIII H fragment contains the early promoters A2b and A2c, where the B. subtilis $\sigma^{43}$-RNA polymerase mainly initiates transcription in vivo and in vitro, respectively, and the main viral late promoter A3 (8,9). The radioactively labelled transcripts were analyzed by electrophoresis in denaturing polyacrylamide gels. Figure 3 shows that, at 72 mM $(NH_4)_2SO_4$, transcription initiation at the A3 promoter, indicated by a run-off transcript 78 nt long, took place either in the presence of p4 (lane a) or when the equivalent fraction from the parallel purification of cells carrying the plasmid containing a nonsense mutation in gene 4 was used (lane b). When the salt concentration of the incubation mixture was increased to 90-108 mM $(NH_4)_2SO_4$, transcription initiation at the A3 promoter was dependent on the presence of purified protein p4 (lanes c-f). In addition to the 78 nt long transcript, a band around 90 nt long was also observed. In all cases, a run-off transcript 470 nt long, corresponding to initiation at the A2c early promoter was seen, being independent on the salt or on protein p4.

When the corresponding S1 mapping experiment was done with the RNA synthesized in vitro at 90 mM $(NH_4)_2SO_4$, hybridizing the unlabelled RNA to the 5'end-labelled late strand of the HindIII H fragment, only the 78 nt long transcript was protected from the S1 digestion (Fig. 4a). The length of the transcript was identical to that of the RNA made in vivo in cells infected with Ø29 in the absence of chloramphenicol (Fig. 4c). No transcript was protected from the S1 digestion when the fraction from the purification of cell extracts containing the plasmid carrying a nonsense mutation in gene 4 was used in the in vitro assay (Fig. 4b). The 90 nt long transcript was not protected from the S1 digestion, suggesting that it was due to an initiation and

termination event within the HindIII H fragment. Truncated template studies confirmed this possibility (not shown).

To determine the DNA sequence requirements for the *in vitro* stimulation of transcription at the A3 promoter by protein p4, different subfragments from the HindIII H DNA fragment have been used as templates in the run-off transcription assay. As shown in Fig. 5, in the presence of the *B. subtilis* $\sigma^{43}$-RNA polymerase and purified protein p4 at 90 mM $(NH_4)_2SO_4$, both, the 308 bp long HinfI-HindIII and the 198 bp long AccI-HindIII subfragments showed to be good templates for the p4-dependent A3 transcription (Fig. 5, lanes a-d), whereas the 129 bp long TaqI-HindIII subfragment was not (Fig. 5, lanes e and f). When low salt was used in the assay, the TaqI-HindIII subfragment showed to be a good template for the p4-independent transcription initiated at the A3 promoter (Fig. 5, lanes g and h). Therefore, deletion of the DNA sequences comprised between the AccI and TaqI sites in the HindIII H fragment resulted in the lost of p4-dependance for the A3 transcription. The deleted region includes an AT-rich sequence upstream the A3 promoter and a correlation among the strength of promoters and the presence of upstream AT-rich regions has been proposed (25). Protein p4 could be a DNA-binding protein since it contains the helix-turn-helix motif (26) present in DNA binding proteins (27). Alternatively, the size of the TaqI-HindIII fragment could be not long enough for the p4-dependent transcription to initiate at the A3 promoter. Experiments aiming to clarify this point are in progress.

The specificity of protein p4 on the *in vitro* initiation of transcription from the Ø29 A3 late promoter will be an useful tool to gain further insight on the understanding of the Ø29 late transcriptional control.

ACKNOWLEDGEMENTS

This investigation has been aided by research grant 2 R01 GM27242-08 from the National Institutes of Health, by Grant n° 3325 from the Comisión Asesora para el Desarrollo de la Investigación Científica y Técnica and by a Grant from Fondo de Investigaciones Sanitarias. I.B. was recipient of a fellowship from the Spanish Research Council.

REFERENCES


1. Vlček, C. and Paces, V. (1986) Gene 46, 215-225.
2. Salas, M., Mellado, R.P., Viñuela, E. and Sogo, J.M. (1978) J. Mol. Biol. 119, 269-291.
3. Mellado, R.P., Moreno, F., Viñuela, E., Salas, M., Reilly, B.E. and Anderson, D.L. (1976) J. Virol. 19, 495-500.
4. Carrascosa, J.L., Camacho, A., Moreno, F., Jiménez, F., Mellado, R.P., Viñuela, E. and Salas, M. (1976) Eur. J. Biochem. 66, 229-241.
5. Schachtele, C.F., de Sain, C.V. and Anderson, D.L. (1973) J. Virol. 11, 9-16.
6. Sogo, J.M., Inciarte, M.R., Corral, E., Viñuela, E. and Salas, M. (1979) J. Mol. Biol. 127, 411-436.
7. Barthelemy, I., Salas, M. and Mellado, R.P. (1986) J. Virol. 60, 874-879.
8. Mellado, R.P., Barthelemy, I. and Salas, M. (1986) J. Mol. Biol. 191, 191-197.
9. Mellado, R.P., Barthelemy, I. and Salas, M. (1986) Nucl. Acids Res. 14, 4731-4741.
10. Sogo, J.M., Lozano, M. and Salas, M. (1984) Nucl. Acids Res. 12, 1943-1960.
11. Davison, B.L., Murray, C.L. and Rabinowitz, J.C. (1980) J. Biol. Chem. 255, 8819-8830.
12. Mellado, R.P. and Salas, M. (1982) Nucl. Acids Res. 10, 5773-5783.
13. Zaballos, A., Salas, M. and Mellado, R.P. (1986) Gene, 43, 103-110.
14. Mellado, R.P. and Salas, M. (1983) Nucl. Acids Res. 11, 7397-7407.
15. Escarmís, C. and Salas, M. (1981) Proc. Natl. Acad. Sci. USA 78, 1446-1450.
16. Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499-560.
17. Hjelm, H., Hjelm, K. and Sjöquist, J. (1972) FEBS Lett. 28, 73-76.
18. Bradford, M.M. (1976) Anal. Biochem. 72, 248,254.
19. Laemmli, U.K. (1970) Nature 227, 680-685.
20. Fairbanks, G., Steck, T.L. and Wallach, D.F.H. (1971) Biochemistry 10, 2606-2617.
21. Edman, P. and Begg, G. (1967) Eur. J. Biochem. 1, 80-90.
22. Tarr, G., Beecher, J.F., Bell, M. and McKean, D.L. (1978) Anal. Biochem. 84, 622-627.
23. Escarmís, C. and Salas, M. (1982) Nucl. Acids Res. 10, 5785-5798.
24. Yoshikawa, H. and Ito, J. (1982) Gene 17, 323-335.
25. Nishi, T. and Itoh, S. (1986) Gene 44, 29-36.
26. Miyukami, Y., Sekiya, T. and Hirokawa, H. (1986) Gene 42, 231-235.
27. Pabo, C.O. and Sauer, R.T. (1984) Ann. Rev. Biochem. 53, 293-320.