Volume 9 Number 22 1981 Nucleic Acids Research

Nucleotide sequence of a Bacillus subtilis promoter recognized by Bacillus subtilis RNA polymerase containing σ^{37}

Charles P.Moran, Jr., Naomi Lang and Richard Losick

The Biological Laboratories, Harvard University, Cambridge, MA 02138, USA

Received 21 July 1981

ABSTRACT

We report the nucleotide sequence of the promoter for a Bacillus subtilis gene (spoVC) whose transcription is controlled by a 37,000 dalton species of B. subtilis sigma factor known as σ^{37} but not by the principal sigma factor of 55,000 daltons (σ^{55}). Using S1 nuclease mapping we show that the startpoint for σ^{37}-directed transcription of the spoVC gene in vitro corresponded closely to the 5' terminus of in vivo synthesized spoVC RNA. The binding site for σ^{37}-containing RNA polymerase extended from 43 bp to 51 bp (positions -43 to -51) upstream from the transcription startpoint to 22 bp (position +22) downstream from the startpoint. The nucleotide sequence of the spoVC promoter differed significantly from promoters whose recognition is controlled by σ^{55} but was similar to other σ^{37}-controlled promoters in regions known to be important in promoter recognition. Our results are consistent with the hypothesis (Lee and Pero, J. Mol. Biol., in press) that sigma factors work by contacting specific bases in both the -35 and -10 regions of promoters.

INTRODUCTION

Promoters are the sites at which RNA polymerase initiates transcription. Among procaryotes these sites are highly conserved in evolution. Promoters from distantly-related eubacteria exhibit hexanucleotide homologies centered at positions 35 ("-35") and 10 ("-10") base pairs (bp) upstream from the startpoint of transcription (1-3). Conversely, RNA polymerases from a variety of eubacteria recognize and initiate transcription from identical promoters on well-defined phage DNA templates (4,5). An important exception to the conservation of bacterial promoters occurs, however, in the Gram-positive bacterium Bacillus subtilis. In uninfected cells and in cells infected with B. subtilis phage SP01 a variety of modified RNA polymerases are observed that differ strikingly in their

promoter recognition specificities from the unmodified form of the bacterial RNA polymerase (for a review see ref. 6). In these modified RNA polymerases the usual sigma subunit of 55,000 daltons (σ^{55}), which dictates promoter selection, has been replaced by a variety of bacterial- and phage-coded sigma factors, which dictate novel promoter specificities. One such modified RNA polymerase containing an SP01-coded sigma factor (the product of gene 28 or σ^{gp28}) recognizes phage promoters that differ substantially at both the -10 and -35 regions from the conserved eubacterial promoters (2,3). Here we report the nucleotide sequence of a chromosomal promoter (the spoVC gene promoter) whose recognition is dictated by a B. subtilis sigma factor of 37,000 daltons (σ^{37}) (7,8). In an accompanying communication, Gilman, Wiggs and Chamberlin (9) report on the structure of a promoter recognized by a 28,000 dalton species of B. subtilis sigma factor (σ^{28}) (10).

RNA polymerase containing σ^{37} ("Eσ^{37}") directs the transcription of at least two B. subtilis genes whose promoters are not utilized by enzyme containing σ^{55} ("Eσ^{55}"). These are the sporulation genes spoVG (or "0.4 kb") and spoVC (or "ctc") from the purA-cysA region of the chromosome (11,12). In other work we (13) have determined the nucleotide sequence of the promoter for the spoVG gene. However, since the initiation site for this developmentally-regulated gene is a relatively weak promoter in vitro (8) and since its transcription in vivo is positively controlled by a class of regulatory genes known as the spo0 loci (14,15) the spoVG gene promoter may not exhibit the preferred sequences for recognition by Eσ^{37}. The spoVC gene initiation site, on the other hand, is a relatively strong promoter (as measured both by transcription and by promoter binding in competition experiments) and does not require the spo0 regulatory products for efficient transcription in vivo. We reasoned, therefore, that the nucleotide sequence of the spoVC gene promoter might be representative of the preferred recognition signals for Eσ^{37}.

RESULTS

We (8) have previously shown that Eσ^{37} binds tightly to an

endonuclease restriction fragment that contains the spoVC promoter. To determine the exact location of this binding site we employed the "DNase footprinting" procedure of Galas and Schmitz (16). A promoter-containing HindIII-HpaII fragment of 520 bp (Fig. 1) that had been labeled with ^{32}P-phosphate at the HindIII terminus was treated with DNAse I in the presence and absence of prebound Eσ^{37}. DNA digested in the absence of Eσ^{37} yielded an uninterrupted ladder of fragments representing cleavage at most sites in the DNA (Fig. 2A, tracks b and d). However, when the end-labeled DNA was first bound with RNA polymerase a stretch of 70 to 80 bp was protected from DNase I action although a stretch of about 5 bp (TATTC) within the protected region exhibited normal or enhanced sensitivity to endonucleolytic digestion (track c). The underlining in Fig. 1 shows our interpretation of the end points of the "footprint"; the downstream end point (344 bp from the HpaII site) was a clear discontinuity in protection (indicated with an arrowhead in Fig. 2A) whereas the upstream end point (408 to 416 bp from the HpaII site) occurred in a region of only partial protection from DNase I action.

As a first step in mapping the startpoint of transcription within the Eσ^{37} "footprint", we measured the size of "run-off" RNA transcribed from a DNA template that had been cut at the

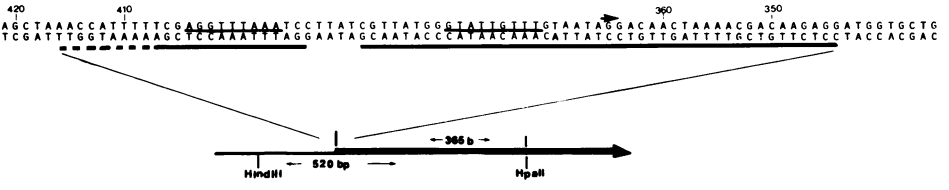

Fig. 1. Structure of the spoVC promoter. The thick arrow at the bottom of the figure shows the location and orientation of the spoVC transcription unit [Ollington et al. (15); N. Lang, unpublished results]. The nucleotide sequence of the spoVC promoter was determined by the procedure of Maxam and Gilbert (18) with DNA fragments end-labeled at either the HindIII or the HpaII sites shown in the figure. The base pairs underlined with a thick line identify regions protected (continuous line) or partially protected (interrupted line) by Eσ^{37} from the action of DNase I as determined in Fig. 2A.

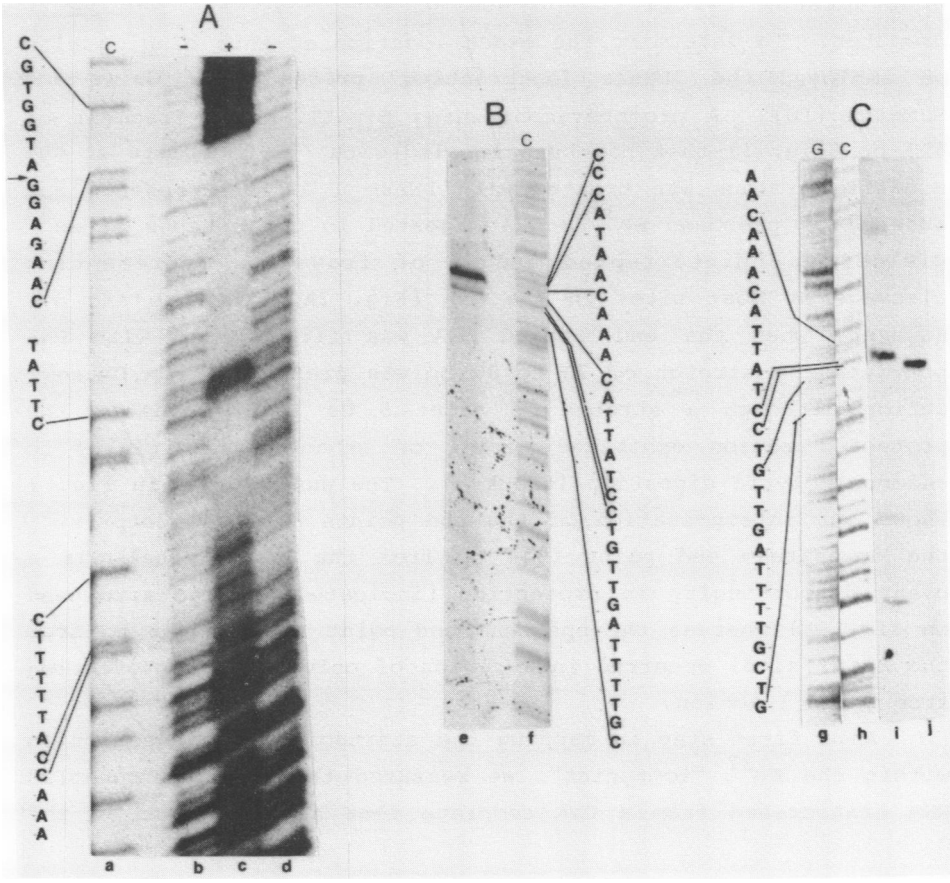

Fig. 2. RNA polymerase binding site and transcription startpoint on spoVC DNA. Panel A, DNase I footprint. A HindIII-HpaII-580 bp fragment of spoVC DNA (Fig. 1) that had been labeled with ^{32}P-phosphate at the HindIII terminus was treated with DNase I in the presence ("+"; track c) and absence ("-"; tracks b and d) of prebound $E\sigma^{37}$ as modified from the procedure of Galas and Schmitz (16) by Talkington and Pero (2). The fragments were denatured at 90° and subjected to electrophoresis through an 8% polyacrylamide/8.3 M urea slab gel (twice as much radioactive fragment was loaded onto track c as in tracks b and d) together with base-specific fragments [Maxam and Gilbert (18)] of the end-labeled 580 bp fragment. Only the C-specific fragments are shown (track a). The nucleotide sequence is of the non-coding strand. Panel B, Run-off transcription. RNA was generated in vitro by $E\sigma^{37}$ from HpaII-cut template as described previously [Ollington et al. (15)] and subjected to electrophoresis (track e) on a 6% polyacrylamide/8.3 M urea slab gel together with base-specific fragments of HindIII-HpaII-580 bp DNA selectively labeled with ^{32}P-phosphate at the HpaII site. Only C-specific

fragments (track f) are shown. The nucleotide sequence is of the DNA coding strand. Panel C, S1-nuclease mapping of the 5' termini of in vitro and in vivo synthesized RNA. spoVC RNA generated in vitro (track i) from run-off transcription at a HaeIII site located downstream from the HpaII site (420 bases from the transcription startpoint) and in vivo synthesized RNA (track j) extracted from Spo0A cells harvested at the fourth hour of stationary-phase were incubated under hybridization conditions with denatured HindIII-HpaII-580 bp DNA that had been end labeled at the HpaII site. After treatment with S1 nuclease the remaining DNA-RNA hybrids were denatured in alkali and subjected to electrophoresis on a 6% polyacrylamide/8.3 M urea slab gel together with base-specific fragments of 580 bp DNA end-labeled at the HpaII site. Only G-specific (track g) and C-specific (track h) fragments are shown. The nucleotide sequence is of the DNA coding strand. Detailed procedures for the hybridization reactions and S1 nuclease treatment are described in Moran et al. (13).

HpaII site (Fig. 1) within the spoVC gene. Radioactive RNA generated by $E\sigma^{37}$ from the truncated DNA template was subjected to electrophoresis on a "DNA sequencing" gel together with base-specific fragments of spoVC DNA labeled uniquely at the HpaII terminus of the coding strand. High resolution gel electrophoresis revealed a predominant transcript migrating at 375 bases from the HpaII site (position 375) and small amounts of a slightly shorter RNA (Fig. 2B, track e). This shows that the transcription startpoint for the spoVC gene was well within the region of close contact with RNA polymerase. However, because RNA migrates somewhat slower than DNA of complementary sequence (J. Pero, personal communication), the "run-off" transcription experiment of Fig. 2B could not provide a precise location for the $E\sigma^{37}$ initiation site.

To identify the spoVC startpoint directly, we employed the mapping procedure of Berk and Sharp (17) to generate an S1 resistant DNA complement of spoVC RNA. RNA synthesized in vitro from spoVC-containing DNA was hybridized to denatured HindIII-HpaII 520 bp fragment that had been end-labeled at the HpaII terminus, thereby exclusively labeling the DNA coding strand. Single-stranded DNA and RNA were then digested with nuclease S1. The remaining DNA-RNA hybrids were denatured and subjected to electrophoresis through a DNA sequencing gel to measure the size of the S1-protected DNA strand (Fig. 2C). The

position of G- and C-specific fragments generated by the chemical cleavage reactions of Maxam and Gilbert (18) of DNA end-labeled at the HpaII site are indicated in tracks g and h. (Note that in contrast to Fig. 2A, the indicated nucleotide sequence is of the coding strand.) The S1-protected DNA strand co-migrated with DNA sequencing fragments of 365 to 367 bases, but we subtract a base from the apparent length of the S1-protected DNA in interpreting this gel because the chemical cleavage reactions of Maxam and Gilbert (18) remove 3' nucleotides but retain 3' phosphates whereas S1 nuclease removes 3' phosphates but retains 3' nucleotides (19). Thus the startpoint for the spoVC gene is probably an A or a G within the sequence TAGG (arrow in Fig. 1).

As confirmation of this startpoint we examined the pattern of transcription priming by synthetic dinucleotides. Dinucleotides are known to prime RNA synthesis at or very near transcription startpoints with which they are homologous (20,21). $E\sigma^{37}$ was allowed to bind to HpaII-cut template in the presence of various synthetic dinucleotides. RNA synthesis was then initiated at a concentration of nucleoside triphosphates that was sufficiently low to prevent de novo initiation of spoVC RNA synthesis but high enough to allow elongation of primed transcription. spoVC RNA was then visualized as a 365 base RNA generated by "run-off" transcription from the truncated DNA template (Fig. 1). Of the 16 possible combinations of dinucleotides only UA strongly primed (AA also primed but weakly) the synthesis of 365 base run-off transcript (Fig. 3 and data not shown), a finding which is consistent with a startpoint within the sequence TAGG.

Finally, we employed S1 nuclease mapping to compare the startpoint of in vitro transcription with the 5' terminus of spoVC RNA synthesized in vivo. Unlabeled RNA purified from cells harvested in stationary-phase was hybridized as described above to denatured DNA labeled at the HpaII site within the spoVC gene. After digestion with S1 nuclease, the remaining RNA-DNA hybrids were denatured and subjected to electrophoresis in an adjacent track to S1-protected DNA from the mapping experiment of in vitro-synthesized RNA. As shown in Fig. 2B in

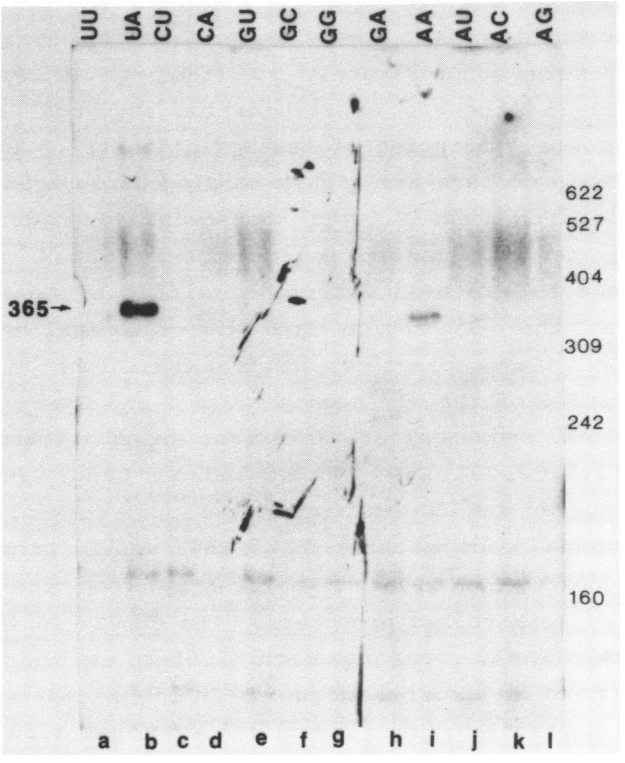

Fig. 3. Dinucleotide primed transcription. Radioactive RNA was transcribed from spoVC-containing template (plasmid p63-1 described in Ollington et al. (15) cleaved with the restriction endonuclease HpaII in reaction mixtures containing $E\sigma^{37}$, 2 μM ATP, GTP, UTP, 0.5 μM (10 μCi) α^{32}P-CTP and the indicated dinucleotide (150 μM) as described by Moran et al. (13). The products of in vitro transcription were subjected to electrophoresis on a 5% polyacrylamide/7 M urea slab gel. The arrow on the left shows the position of 365 base run off RNA synthesized at high nucleotide triphosphate concentration and the numbers on the right are the sizes in bases of denatured pBR322 HpaII fragments.

vivo-synthesized spoVC RNA was only a few nucleotides shorter than the corresponding product generated in vitro.

DISCUSSION

The rules for promoter recognition by B. subtilis RNA polymerase containing σ^{55} ($E\sigma^{55}$) appear to conform closely to

those established for E. coli RNA polymerase. In E. coli promoter recognition is largely determined by hexanucleotide sequences centered approximately 10 (the -10 or Pribnow box sequence) and 35 (the -35 or recognition sequence) base pairs upstream from the startpoint of transcription (1). B. subtilis σ^{55}-controlled promoters are highly homologous in both their -10 and -35 regions to these E. coli consensus sequences. Indeed, none of the ten σ^{55}-controlled promoters whose nucleotide sequences have so far been determined differ by more than one nucleotide at the -35 region from the canonical recognition sequence TTGACA and nucleotides at four positions (A and T at positions 4 and 6 of the -10 sequence and T and G at positions 1 and 3 of the -35 sequence) are invariant among all ten Bacillus promoters (3,5,22,23, C.P. Moran and N. Lang, unpublished results; C. Murray and J. Rabinowitz, personal communication; I. Smith, personal communication). The spoVC promoter, in contrast, appears to differ significantly from the conserved features of σ^{55}-controlled promoters. The spoVC -10 (GTATTGTTT) and -35 (AGGTTTAAA) regions each differ at least at two positions from the corresonding σ^{55} canonical sequences, lacking, in particular, the invariant A at position 4 of the Pribnow box and the invariant G at position 3 of the recognition sequence. These differences from the preferred sequences for σ^{55} may be a sufficient explanation for the failure of $E\sigma^{55}$ to recognize the spoVC promoter. In this regard we note that the E. coli lac UV5 gene, which is an extremely poor template for B. subtilis $E\sigma^{55}$ enzyme (5), lacks the conserved G in the third position of its -35 promoter sequence (24).

What features of the spoVC promoter determine its recognition by $E\sigma^{37}$? Although $E\sigma^{55}$ is unable to initiate transcription at the spoVC promoter, $E\sigma^{37}$ transcribes, albeit weakly, several genes that are efficiently utilized by $E\sigma^{55}$ (e.g. veg in ref. 25, and penP, S. Chang and R. Doi, personal communication). A comparison of five such $E\sigma^{37}$-recognized promoters reveals two regions of approximate homology to the spoVC promoter (C.P. Moran and N. Lang, unpublished results). These regions are located approximately 35 (A--AGGT) and 10 (GTA-TPu-T) base pairs upstream from the startpoints of

transcription. The functional significance of these sequences must be regarded as uncertain until their role in promoter recognition can be tested by genetic and chemical modification studies. Nevertheless, in the case of spoVC the conserved sequences are well within the regions (position -43 through -26 and position -22 through +22, Fig. 1) of close contact with $E\sigma^{37}$. It is therefore tempting to suppose that, like other forms of bacterial RNA polymerase, $E\sigma^{37}$ recognizes distinctive nucleotide sequences in both the recognition and Pribnow box regions.

This pattern of distinctive and conserved nucleotide sequences in the -10 and -35 positions is emerging as a general feature of promoters whose recognition is controlled by one of a variety of sigma factors. Pero and her coworkers (2,3) previously observed that the promoters for five B. subtilis phage SP01 middle genes, which are transcribed by RNA polymerase containing a phage-coded sigma factor (σ^{gp28}) in place of σ^{55}, exhibit the canonical sequences AGGAGA and TTT-TTT in the -35 and -10 positions, respectively. In an accompanying communication, Gilman, Wiggs and Chamberlin (9) show that two B. subtilis promoters whose recognition is controlled by the B. subtilis sigma factor σ^{28} share identical pentamers (CTAAA) and heptamers (CCGATAT) in the -35 and -10 positions, respectively. Thus, the promoters controlled by four species of B. subtilis sigma factor (σ^{55}, σ^{37}, σ^{28} and the SP01 factor σ^{gp28}) exhibit conserved but distinctive nucleotide sequences in homologous positions. We also note that the DNA binding sites (as determined in DNase I footprinting experiments) for $E\sigma^{55}$, $E\sigma^{37}$ and $E\sigma^{gp28}$ are similar to each other (ref. 2; J. Pero, personal communication; Fig. 2A) both in length (roughly 70 bp) and in location (from about position -50 to about +20 relative to the transcription startpoints). Thus, the detailed promoter contacts for each form of RNA polymerase may differ but the overall conformations of the RNA polymerase-promoter complexes are likely to be very similar. In other words, each form of RNA polymerase may differ in what it sees but not in how it sees promoters.

If these conserved -35 and -10 nucleotide sequences are

functionally important, then each species of sigma factor must somehow determine the recognition of nucleotide sequences separated by about 16 to 19 bp of helix. As discussed previously (3,6), this probably occurs by sigma contacting specific bases in both the -35 and -10 regions during the formation of the open promoter complex. In fact, through cross-linking experiments Simpson (26) and Chenchick et al. (27) have shown that E. coli sigma contacts bases at both the -10 and -35 regions of the lacUV5 promoter.

If, as we speculate, the -10 and -35 sequences GTA-TPu-TT and A--AGGT are preferred signals for promoter selection by σ^{37}, then what features of the relatively-weak spoVG (0.4 kb) gene promoter determine its recognition by $E\sigma^{37}$? [The spoVG promoter is actually composed of overlapping promoters from which RNA polymerase initiates at two sites separated by about 10 base pairs (13). For the purposes of this discussion, we only consider the upstream promoter since transcription from the downstream initiation site requires an as yet unidentified transcription factor in addition to $E\sigma^{37}$.] Centered about 10 base pairs upstream from the spoVG transcription startpoint is an octanucleotide sequence (GAATTGAT; ref. 13) which conforms in five out of six positions to the presumed preferred -10 bases (GTATTGTT) in spoVC. At 35 bp preceding the startpoint, however, the spoVG promoter bears little homology to the -35 sequences (AGGTTT) in spoVC, although at three bases out of register (three bases closer to the -10 sequence) is a hexanucleotide sequence (AGGATT) in the spoVG promoter which conforms in five positions to the spoVC -35 sequence AGGTTT. We propose therefore that the spoVG promoter either lacks a -35 sequence or lacks a -35 sequence in the optimum position for efficient utilization by $E\sigma^{37}$. This could explain why in promoter competition experiments, spoVG behaves as a much weaker binding (8) and transcription initiation (15) site than the spoVC promoter. If our interpretation is correct, then one or more of the spo0 regulatory proteins, which are required for spoVG but not spoVC RNA synthesis in vivo (15), could act by compensating for the missing or improperly positioned recognition sequence at the -35 region of the spoVG promoter.

[As a precedent for this idea, the absence of a -35 sequence in the phage λ p_E promoter is thought to be compensated for by the action of the λ cII regulatory protein which promotes transcription from this site (28).] Our proposal for spo0 action could be tested by determining whether the insertion of three base pairs between the -35 and -10 regions strengthens the spoVG promoter in vitro and relieves spo0 dependence in vivo.

Both the spoVG and spoVC promoters are recognized by an additional species of B. subtilis sigma factor known as σ^{29} (25). This transcriptional modifier, which is induced during the course of sporulation, differs from σ^{37}, however, in that it recognizes a B. subtilis promoter designated as 1 and a fortuitous promoter in the E. coli plasmid pMB9 that are not recognized by σ^{37} or σ^{55}. Thus, σ^{29} may be broader (less stringent) than, although overlapping with, σ^{37} in its promoter recognition specificity. It will be of interest to determine whether or not promoter recognition by σ^{29} requires only a subset of the nucleotide contacts that signal recognition by σ^{37}.

ACKNOWLEDGMENTS

We thank J. Pero and W.G. Haldenwang for helpful discussions and W.G. Haldenwang and W.C. Johnson for generously providing Eσ^{37}. The work was supported by a grant from the National Institutes of Health (GM 18568).

REFERENCES

1. Rosenberg, M. and Court, D. (1979) Ann. Rev. Genet. 13, 319-353
2. Talkington, C. and Pero, J. (1979) Proc. Nat. Acad. Sci. U.S.A. 76, 5465-5469
3. Lee, G. and Pero, J. (1981) J. Mol. Biol. 151, in press
4. Wiggs, J.L., Bush, J.W. and Chamberlin, M.J. (1979) Cell 16, 97-109
5. Lee, G., Talkington, C. and Pero, J. (1980) Mol. gen. Genet. 180, 57-65
6. Losick, R. and Pero, J. (1981) Cell 25, 585-587
7. Haldenwang, W.G. and Losick, R. (1979) Nature 282, 256-260
8. Haldenwang, W.G. and Losick, R. (1980) Proc. Nat. Acad. Sci. U.S.A. 77, 7000-7004
9. Gilman, M.Z., Wiggs, J.L. and Chamberlin, M.J. (1981) Nuc. Acids Res., this issue.
10. Wiggs, J.L., Gilman, M.Z. and Chamberlin, M.J. (1981) Proc. Nat. Acad. Sci. U.S.A. 78, 2762-2766

11. Moran, C.P., Jr., Losick, R. and Sonenshein, A.L. (1980) J. Bacteriol. 142, 331-334
12. Rosenbluh, A., Banner, C.D.B., Losick, R. and Fitz-James, P.C. (1981) J. Bacteriol. 148, in press
13. Moran, C.P., Jr., Lang, N., Banner, C.D.B., Haldenwang, W.G. and Losick, R. (1981) Cell 25, 783-791
14. Segall, J. and Losick, R. (1977) Cell 11, 751-761
15. Ollington, J.F., Haldenwang, W.G., Huynh, T.V. and Losick, R. (1981) J. Bacteriol. 147, 432-442
16. Galas, D.J. and Schmitz, A. (1978) Nucleic Acids Res. 5, 3157-3170
17. Berk, A.J. and Sharp, P.A. (1977) Cell 12, 721-732
18. Maxam, A.M. and Gilbert, W. (1980) in Methods in Enzymology, Grossman, L. and Moldave, K. Eds., Vol 65, pp. 499-559, Academic Press, New York.
19. Hentschel, C., Irminger, J.-C., Bucher, P. and Birnstiel, M.L. (1980) Nature 285, 147-151
20. Hoffman, D. and Niyogi, S. (1973) Proc. Nat. Acad. Sci. U.S.A. 70, 574-578
21. Minkley, E.G. and Pribnow, D. (1973) J. Mol. Biol. 77, 255-277
22. Kroyer, J. and Chang, S. (1981) Gene, in press
23. Doi, R.H., Kudo, T., Dickel, C.D., Sharp-Hayes, M. and Chang, S. (1981) in Microbiology 1981, in press.
24. Johnsrud, L. (1978). Proc. Nat. Acad. Sci. U.S.A> 75, 5314-5318
25. Haldenwang, W.G., Lang, N. and Losick, R. (1981) Cell 23, 615-624
26. Simpson, R.B. (1979) Cell 18, 277-285
27. Chenchick, A., Beabealashvilli, R. and Mirzabekov, A. (1981) FEBS Letters 128, 46-50
28. Schmeissner, U., Court, D., Shimatake, H. and Rosenberg, M. (1980) Proc. Nat. Acad. Sci. U.S.A. 77, 3191-3195