Volume 13 Number 15 1985 Nucleic Acids Research

The sequence of the distal end of the *E. coli* ribosomal RNA rrnE operon indicates conserved features are shared by rrn operons

Hope Liebke and Graham Hatfull

Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06510, USA

Received 10 May 1985; Accepted 9 July 1985

ABSTRACT

The 1440 nucleotides of the distal region of the E.coli ribosomal RNA operon found on the λaroE transducing phage has been sequenced. We show that the λaroE hybrid rrn operon ends after a solitary 5S RNA gene with rrnE distal sequence. A single terminator structure of dyad symmetry followed by a run of six T's has been identified and compared to other sequenced rrn terminator hairpins. Immediately adjacent to the hairpin is a region of interrupted but conserved sequence that is shared by rrnE, rrnB and rrnD. An open reading frame of 127 amino acids abuts the terminator structure. Another open reading frame of 147 amino acids is found on the opposite strand several hundred nucleotides downstream.

INTRODUCTION

Transcriptional regulation of the E. coli ribosomal operons includes antitermination features which presumably allow an RNA polymerase complex to efficiently proceed through the large regions of secondary structure present in the rRNA (1). In view of this recently defined regulatory mechanism, it is not surprising that in vivo transcripts arising from ribosomal promoters can traverse very long distances through insertion sequences which, in other translated operons, are strongly polar (2-4). The sequences near the promoter that control antiterminated transcription have been localized to an 80 nucleotide region (5; Cashel, personal communication). The recognition of the rrn antitermination process clarifies findings from several laboratories that indicated that ribosomal promoters cloned without terminators were extremely unstable, but that the addition of a downstream ribosomal terminator stabilized the clones (6,7).

One might ask what features differentiate the terminators that stop transcripts originating from antiterminated promoters versus normal promoters. The recent description of a ribosomal terminator construction that is at least 90% efficient for transcription termination in vivo from a ribosomal promoter suggests that the features required are present in the 76 nucleotides distal to the second 5S RNA gene of rrnD (8,9). What these

features are unfortunately cannot be determined by sequence comparison with the analogous region from the other operon (rrn B) that also terminates transcription after 5S RNA because these regions differ by only 5 nucleotides (10,11).

We have sequenced 1440 nucleotides of the distal region of a hybrid rrn operon, which is found on the λaroE transducing phage (12), as an initial analysis of the requirements for E.coli antitermination control. We identify the terminus of this fused operon by hybridization of a terminator distal probe to E.coli genomic DNA and comparison of the labeled fragment with the known rrn operons visualized by a 5S gene probe. In this communication, we also identify a conserved interrupted sequence which is within the 76 nucleotides necessary for stopping antiterminated transcripts (8,9), and which is found immediately distal to a stem loop structure in both the rrnB T1 terminator region (10) and the rrnD terminator region (11). We compare the hairpin terminator structures of all the sequenced rrn operons and describe their common features.

EXPERIMENTAL PROCEDURES

Materials

CGSC 4401 and 4474 were obtained from B.Bachmann of the Coli Genetic Strain Center. JF962, a lysogen of λaroE and λCI857S7b515b519xis6 (13), was maintained on minimal media, diluted into broth, and temperature induced to produce phage. The unlysed cells were pelleted, diluted into TMG buffer (.01M Tris 7.4, .01 M $MgSO_4$, .01% gelatin) and lysed by the addition of $CHCl_3$. After clarification, the supernatant was applied to a CsCl step gradient. The phage band at the interface was diluted with CsCl to produce a 1.3840 refractive index and was subjected to equilibrium centrifugation. The lower phage band, containing λaroE phage, was collected. After dialysis to remove CsCl, the phage was phenol extracted and dialyzed to produce suitable DNA. Restriction enzymes were purchased from N.E.Biolabs and used according to the manufacturers recommendations. Klenow fragment was either purchased from Boehringer Mannheim or was the generous gift of C.Joyce (Yale). [^{35}S] dATP and [^{32}P] dXTPs were purchased from Amersham. Nitrocellulose was purchased from Schleicher and Schuell.

Hybridization Probes

Appropriate M13 phage were incubated in a sequencing reaction containing three radioactive deoxynucleotides in the absence of dideoxynucleotides. Following heat inactivation of the Klenow fragment enzyme, EcoRI and NaCl

were added. After digestion, the samples were desalted, EtOH precipitated, applied to a 5% acrylamide 7M urea gel, and electrophoresed. The single stranded fragment was eluted from the gel and hybridized to the filter in buffer (5xSSC, 50 mM PO_4, .1% SDS, 200 ug/ml salmon sperm DNA) for 18 hrs at 48°C. The filter was washed successively with 2xSSC at room temperature.

RESULTS AND DISCUSSION

Defining the terminator fragment

The λaroE transducing phage was used previously as a source of DNA for determining the sequence of the rrnD promoter, RNAse III processing sites and 5S RNA gene sequence (12, 14-17). The terminator region following the 5S RNA gene was located on an AvaI/BamHI fragment that was approximately 2.5 KB long (R. Bram, personal communication). In order to identify the specific fragment of the several in that size range, an AvaI/BamHI double digest of λaroE DNA was transferred to DBM paper and probed with *in vivo* $^{32}PO_4$ labeled 5S RNA. The fragment hybridizing to 5S RNA was purified from a preparative sample of DNA and restriction sites were mapped according to the method of Smith and Birnstiel (18). All restriction sites verified those previously determined for that region of λaroE DNA (15) and an additional PvuII site was located 1600 nucleotides downstream from the AvaI site (data not shown).

From which rrn operon does the terminator fragment originate?

The rrn operon found on λaroE is a hybrid rrn operon of unknown origin. Boros et al. (19) suggest that the rrn operon on λaroE152 phage is a hybrid of rrnD and rrnG (or F). However, we have noted (unpublished results) that it shares distal restriction sites with DNA that originates from rrnE (M. Ellwood, personal communication). Our phage were grown with selection for complementation of an aroE deficient strain by the aroE marker which is carried in the proximal region of the chromosomal DNA carried on the transducing phage, but no rrn terminal markers were known that would allow selection for and maintenance of the terminal region.

To define the distal DNA carried by our λaroE phage, we chose to identify the rrn operon in chromosomal DNA that hybridized with a probe containing only the distal region, and then to compare that to all the rrn operons in chromosomal DNA that hybridized to a probe containing the 5S RNA gene. Therefore, we digested total chromosomal DNA from two *E.coli* strains (CGSC 4474 and 4401) that differ in the ribosomal fragments they contain because of a chromosomal inversion between two ribosomal operons (20,21). 4474 is an auxotrophic W3110 strain that contains an inversion of the

chromosome between rrnD and rrnE creating two hybrid operons: rrnD/E and rrnE/D. 4401, one of the early K12 strains, lacks that inversion. The DNA from each strain was digested with either SalI or a double digest of BamHI/PstI because these restriction enzymes have been used to identify the fragments that correspond to the rrn operons (20,22). SalI cuts internal to the rrn operons and upstream of the 5S gene; downstream sites vary depending on the specific rrn operon. SalI digests can therefore define the terminal rrn operon fragments if appropriate probes are used. In addition, SalI digests indicate the origin of the terminal rrn operon fragment irrespective of whether it arose from a hybrid operon. Since the BamHI and PstI enzymes cut outside of the rrn operons, fragments characteristic of the hybrid operons vary enough in size from those of the wild type operons to allow their identification. The BamHI/PstI double digest comparison should, therefore, indicate whether the fragment originated from a hybrid operon. As a control, λaroE DNA was digested with a combination of EcoRI and BamHI to produce a 4 kb fragment that is recognized by both probes.

After electrophoresis the DNA digests were transferred to nitrocellulose according to the method of Southern (23) as modified by Thomas (24). The filter was hybridized first with a single stranded M13 probe that contained distal sequence approximately 1 kb from the end of the 5S gene (initiating at nucleotide 1024 in the sequence presented below) in order to identify the specific fragment in each chromosomal digest that corresponds to the distal rrn sequence carried by λaroE as is shown in Figure 1A. In the second hybridization of the filter, all seven ribosomal operons were detected using a single stranded M13 probe containing the 5S gene as shown in Figure 1B.

A single unique fragment is detected when the control λaroE DNA is hybridized with either probe (lane 1 in figure 1A and B). The rrn terminal fragments of chromosomal DNA that are detected by the first probe are the same size in the SalI digests of either 4474 or 4401 (lane 2 and 3, figure 1A) and correspond to the rrnE operon fragment (figure 1B) as previously defined (22). However, the rrn terminal fragments differ in size for the BamHI/PstI double digest of 4401 as compared to 4474 (lane 4 and 5, figure 1A and 1B). This result would be expected if the probe hybridized to a fragment originating from a hybrid operon carried by 4474. The decrease in size of the fragment detected with the terminal probe for 4474 in lane 5 as compared to 4401 in lane 4 indicates that it is the hybrid rrnD/E operon that is being detected (figure 1A compared to 1B). These results prove that the distal DNA of the hybrid rrn operon found on our λaroE transducing phage originates from the DNA adjacent to the end of the rrnE operon.

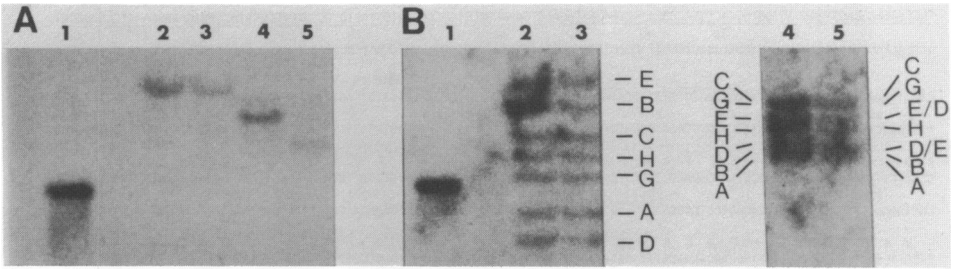

Figure 1. Panel A represents the result of hybridization with the distal sequence M13 clone. Panel B is the autoradiogram of the same filter which, after washing in .1X SSC at 100°C, was probed with the 5S RNA gene probe. 1, λaroE DNA, EcoRI/BamHI; 2, 4401 DNA, SalI; 3, 4474 DNA, SalI; 4, 4401 DNA, BamHI/PstI; and 5, 4474 DNA, BamHI/PstI. The letter for each rrn operon is adjacent to the fragment that defines it. The BamHI/PstI fragments are not well resolved but consist of a doublet, another doublet, and a triplet of fragments whose order is represented by the letters of the rrn operons.

Sequence of the DNA distal to the rrnE 5S gene

The PvuII fragment containing the 5S gene was purified from λaroE DNA, sonicated (25), and ligated into the M13 mp8 vector (26). The random clones that were generated by the sonication were sequenced by the dideoxy chain termination method (27) using [^{35}S] dATP and buffer gradient electrophoresis (28). The sequencing data was compiled in a DEC VAX computer using the programs of Staden (29). Each nucleotide of the sequence presented in Figure 2 was determined from at least 3 and not more than 8 independently isolated overlapping M13 clones. The sequence was analyzed by computer using the Analseq program of Staden (30) to predict a) open reading frames by start and stop codon assignment and positional base preference: b) promoter -10 and -35 regions based on consensus comparisons; c) ribosome binding sites by matrix algorithm comparison to the 101 nucleotides surrounding the AUG start codons of 124 known prokaryotic proteins; and d) secondary structures that might function as terminator regions.

The sequence initiates at the nucleotide corresponding to the mature 5' end of the 5S RNA gene (nucleotide 1 in figure 2) and extends for 1440 nucleotides. The sequence of the 5S RNA gene of rrnE is identical to the sequence of 5S RNA found in the rrnB operon (10). The sequence CAAAT (marked with superscripted dots in both figure 2 and 3) is found at the end of the 5S gene in rrnE just as it is in each of the other operons, including rrnF(or G) and rrnC (31, 32). The CAAAT sequence could pair with the upstream conserved AUUUG, which includes two nucleotides of the 5' end of 5S RNA (the U and G

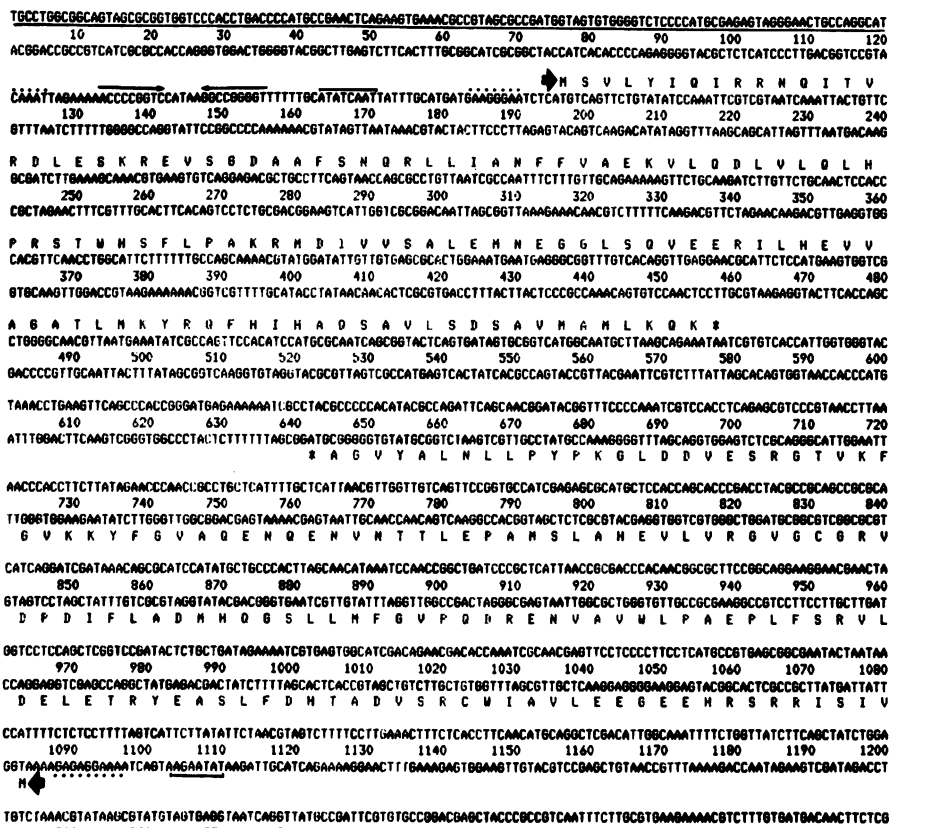

Figure 2. The sequence of the DNA downstream of the rrnE 5S RNA gene. The 5S RNA gene initiates at nucleotide 1 and is underlined. The conserved CAAAT sequence found adjacent to every known 5S gene is highlighted with superscripted dots. The juxtaposed arrows indicate the dyad symmetry of the putative rrnE terminator structure. The start and stop of each open reading frame are indicated by the arrows and asterisks adjacent to the proper codons. The ribosome binding site consensus predicted by the Analseq program is underscored with a dotted line. The putative -10 promoter region predicted by the Analseq program is underscored with a bracket.

are nucleotides 1 and 2 respectively in the sequence in figure 2), to form a stem structure as previously noted (11). This structure may be recognized in the generation of the mature 5' end and precursor 3' end of the 5S RNA by RNAse E (9).

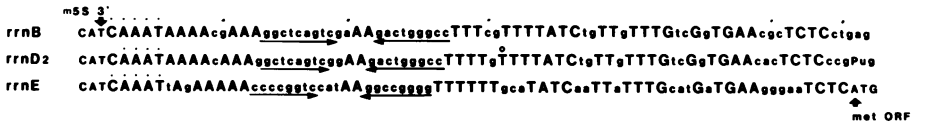

Figure 3. Comparison of the sequences in the terminator regions of rrnB, rrnD2 and rrnE. The first nucleotide of the rrnE sequence corresponds to nucleotide 118 in figure 2 and the last corresponds to nucleotide 207. Uppercase characters represent conserved sequences. The asterisks indicate sequence differences between rrnB and rrnD2. The dyad symmetries of the terminator hairpin are indicated by juxtaposed arrows. The ^{o} is over the T nucleotide where transcription terminates in rrnD2.

To identify the rrn terminator, the entire sequence was scanned for regions of dyad symmetry and a single hairpin loop structure was identified. It occurs approximately fifteen nucleotides from the end of the mature 5S RNA gene and is followed by a run of T's providing features extremely similar to the terminators recognized for the rrnB, rrnD, rrnF(G) and rrnC operons (10, 11, 31, 32). A region of interrupted homology that is also shared by rrnB and rrnD follows the T run as described in more detail in the next section.

The sequence originating at nucleotide 190 is unique and is not found in other rrn operons. An open reading frame is found on the same strand as the ribosomal operon and begins at the methionine located at position 191. This open reading frame is preceded by an excellent ribosome binding site centering on nucleotide 184 that is predicted by matrix comparison to the sequences surrounding known protein start sites using the Analseq program of Staden (30). In addition, a very good -10 promoter consensus sequence can be located immediately upstream of the ribosomal binding site. The -35 region for the promoter, however, is poorly conserved. The positional base preference analysis of the Analseq program for this open reading frame predicts that it has a good coding capability. If translated, a small protein of 127 amino acids would be produced from this reading frame. A search of the NBRF protein sequence library has not revealed absolute identity with any known protein.

Another open reading frame can be identified on the opposite strand initiating at the methionine at nucleotide 1080. This open reading frame also has a very good ribosome binding site, as well as an excellent -10 consensus promoter. The -35 region for this promoter is also poorly conserved. If translated, as is indicated by the positional base preference analysis, a protein of 147 amino acids would be made. This protein also has

not been previously identified as determined by comparison with the NBRF protein sequence library. It is interesting to note that no terminator structure of dyad symmetry can be found between the two ORF's suggesting that the rrnE terminator may function as a bidirectional terminator as proposed for rrnD (33). Moreover, regulation of translation of either protein could therefore occur if RNA coding for one of the proteins was more abundant and hybridized to the other blocking its translational recognition sites.

Comparison of the rrn terminator regions

The three rrn operons that end in 5S RNA share several features that become apparent upon comparison of the 80 nucleotides just distal to their 5S genes as shown in Figure 3. Although the sequence following the 5S gene of rrnB and rrnD2 [D1 refers to the first 5S gene in rrnD and D2 to the second (11)] is the same for all but five nucleotides (which are marked with an asterisk in the figure), the sequence of the rrnE distal region is different. The several conserved features are highlighted in the figure by large print.

Just adjacent to the CAAAT sequence is the start of the terminator region which begins with a region of A richness. The span of A's immediately precedes a region of dyad symmetry in rrnE, rrnD2 and rrnB which can form a nine bp G/C rich stem. The loop of each hairpin structure in the three operons has two conserved adenine nucleotides, while a run of downstream T's abuts the hairpin. This hairpin loop structure has been defined as the T1 terminator in rrnB (10), based originally on similarities to other rho independent terminator structures (33-36), and more recently by determination of the 3' end by S1 nuclease mapping (37). Thus in rrnB, termination occurs *in vivo* at various T's within the 20 nucleotides adjacent to the dyad symmetry stem (37). However, transcriptional termination in rrnD2 occurs *in vivo* at the first T in the second T cluster in the run (8, 9) and is marked in the figure (^{o}). We have noted that even though the sequence of the rrnE terminator region differs from the similar sequence exhibited by both rrnB and rrnD2, there is a region of interrupted homology that is striking. Immediately following the run of T's is an unevenly interrupted conserved sequence that encompasses 18 nucleotides of 31:

TATCxPuTTPuTTTGxx(x)GPuTGAAxxx(xx)TCTC.

We feel it is possible that this conserved homology may be peculiar to the class of terminators required for stopping antiterminated transcripts at the ends of rrn operons ending in 5S genes. Immediately downstream of the homology, the sequences diverge completely. The conserved homology is

included in the only region of the entire distal rrnE sequence that has known termination features. Although there is no experimental evidence to date that indicates that this region in rrnE is involved in termination, the very similar regions of rrnD and rrnB do function as specific terminators for antiterminated promoted transcripts. For example, in rrnD2 the CCCGPuG adjacent to the end of this conserved region is the AvaI site at the 3' end of the fragment used in the construction of a "minioperon" that has allowed an analysis of the maturation (9) and expression of the 5S RNA (8). The rrnD2 minioperon construct is 85-90% efficient for in vivo termination of an rrnA promoted transcript in a galK assay system (8) or by fingerprint analysis of the shortened primary product of the minioperon (9). Similar results are obtained for a construct (37) containing the rrnB terminator downstream from the rrnA promoter where S1 mapping has been used to characterize the 3' end of the RNA. Because the rrnA promoter region (38) also contains the antitermination signal defined for the rrnG promoter (5), transcripts originating from this promoter are antiterminated (M. Cashel, personal communication). The fact that antiterminated transcripts stop at a similar site in these two minioperon constructs suggests that the 76 nucleotides distal to the 5S RNA gene and before the AvaI site of rrnD must have features necessary for stopping antiterminated transcription. Although strong dyad symmetries found in the 16S RNA coding regions are sufficient to terminate transcripts from non-rrn promoters in vivo (1), they are read through by RNA polymerase that initiates upstream of the antitermination region of the rrn promoters (5). This result suggests that dyad symmetries are not sufficient to stop antiterminated transcripts.

The need for special dyad symmetries in stopping antiterminated transcripts can not be ignored however. We recognize that the conserved sequence of rrnE, rrnD and rrnB is not found adjacent to the dyad symmetry terminator stem of the operons ending in tRNAs [rrnC and rrnF(G)] as might be expected if it were a general feature required for stopping all rrn transcripts. We have therefore compared (figure 4) the dyad symmetry terminator structures known for rrnD, rrnB, rrnC and rrnF(G) with that predicted for rrnE to determine if they have special features not apparent in other rho independent terminators (see 34 for examples). The following characteristics can be noted. The 9 bp G/C rich stem in each structure can be extended with A/U pairing to form a 14-20 bp hairpin of essentially perfect complementarity as previously noted for rrnD (10). The rrnC structure is unique in having a single mismatched pair. The base of each

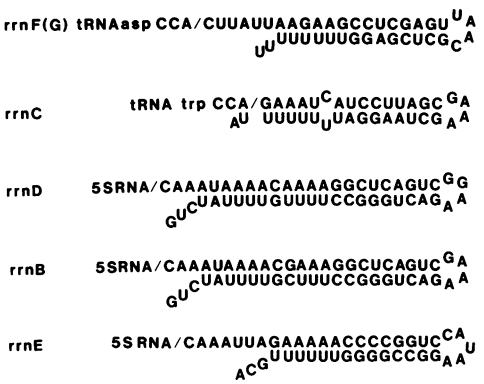

Figure 4. Comparison of the terminator dyad symmetry structures of the sequenced rrn operons. The / indicates the mature 3' end of either 5S RNA or the tRNA which precedes the terminator.

ascending stem is at least 75% A which pairs with the descending stem U rich bottom portion. This would allow each structure to function bidirectionally since rho independent terminator dyad symmetries are usually followed by U rich regions. The resulting structures are bimodal hairpins that are very strong at the apex and weaker at the base. The rrn hairpins are always within 6-7 nucleotides of a strong upstream secondary structure imposed by the 5S or tRNA sequence. Each of the 4 or 5 nucleotide loops at the hairpin apex has at least two adenines adjacent to the 3' side of the loop. Finally there is only weak purine versus pyrimidine preference on the ascending side of the stem when all the structures are compared. It should be recognized that transcription termination for rrnC and rrnF(G) has only been demonstrated *in vitro* from transcripts originating at ends of restriction fragments. Whether these *in vitro* terminator structures found at the ends of rrn operons ending in tRNA's (as opposed to 5S RNA) function to stop antiterminated transcripts *in vivo* remains to be determined, as does the possible role of the conserved sequence. Once these questions have been answered, the contribution of the features we have noted to the process of termination of antiterminated transcripts can then be fully assessed.

ACKNOWLEDGEMENTS

We are grateful to Joan Steitz and Bart Barrell for allowing us the opportunity to complete this work in their laboratories. This work was supported by a grant from the National Institutes of Health to H.L. and a training fellowship to G.H. from the MRC. We thank Jill Galloway, Terry Platt and Joan Steitz for their helpful comments on the manuscript. We thank Roger Staden for providing us with his very useful Analseq computer program.

REFERENCES

1. Aksoy, S. and Squires, C.L. (1984) J. Bacteriol. 159, 260-264.
2. Morgan, E. (1980) Cell 21, 257-265.
3. Brewster, J. and Morgan, E. (1981) J. Bacteriol. 148, 897-903.
4. Siehnel, R. and Morgan, E. (1983) J. Bacteriol. 153, 672-684.
5. Li, S., Squires, C.L. and Squires, C. (1984) Cell 38, 851-860.
6. Brosius, J. (1984) Gene 27, 151-160.
7. Sarmientos, P. and Cashel, M. (1983) Proc. Natl. Acad. Sci. USA 80, 7010-7013
8. Elford, R. and Holmes, W. (1983) J. Mol. Biol. 168, 557-561.
9. Szeberenyi, J. and Apirion, D. (1983) J. Mol. Biol. 168, 525-561.
10. Brosius, J., Dull, T., Sleeter, D., and Noller, H. (1981) J. Mol. Biol. 147, 107-127.
11. Duester, G. and Holmes, W. (1980) Nucl. Acids Res. 8, 3793-3807.
12. Young, R. and Steitz, J. (1978) Proc. Natl. Acad. Sci. USA 75, 3593-3597.
13. Jorgensen, P. (1976) Mol. Gen. Genet. 146, 303-307.
14. Young, R., Macklis, R. and Steitz, J. (1979) J. Biol. Chem. 254, 3264-3271.
15. Bram, R., Young, R. and Steitz, J. (1980) Cell 19, 393-401.
16. Young, R., Bram, R. and Steitz, J. (1980) tRNA: Biological Aspects, Soll et al., eds., Cold Spring Harbor Laboratory 99-106.
17. Young, R. and Steitz, J. (1979) Cell 17, 225-234.
18. Smith, H. and Birnstiel, M. (1976) Nucl. Acids Res. 3, 2387-2397.
19. Boros, I., Kiss, A. and Venetianer, P. (1979) Nucl. Acids Res. 6, 1817-1830.
20. Hill, C. and Harnish, B. (1981) Proc. Natl. Acad. Sci. USA 78, 7069-7072.
21. Hill, C. and Harnish, B. (1982) J. Bacteriol. 149, 449-457.
22. Ellwood, M. and Nomura, M. (1982) J. Bacteriol. 149, 458-468.
23. Southern, E. (1975) J. Mol. Biol. 98, 503-517.
24. Thomas, P. (1980) Proc. Natl. Acad. Sci. USA 77, 5201-5206.
25. Deininger, P. (1983) Analyt. Biochem. 129, 216-223.
26. Messing, J. and Vieira, J. (1982) Gene 19, 269-276.
27. Sanger, F., Nicklen, S. and Coulson, A. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
28. Biggin, M., Gibson, T. and Hong, G. (1983) Proc. Natl. Acad. Sci. USA 80, 3963-3965.
29. Staden, R. (1982) Nucl. Acids Res. 10, 4731-4751.
30. Staden, R. (1984) Nucl. Acids Res. 12, 505-520.
31. Sekiya, T., Mori, M., Takahashi, N. and Nishimura, S. (1980) Nucl. Acids Res. 8, 3809-3827.
32. Young, R. (1979) J. Biol. Chem. 254, 12725-12731.
33. Holmes, W., Platt, T. and Rosenberg, M. (1983) Cell 32, 1029-1032.
34. Platt, T. and Bear, D. (1983) Gene Function in Prokaryotes, Beckwith et al.. eds., Cold Spring Harbor Laboratory, 123-181.
35. Rosenberg, M., Chepelinsky, A. and McKenney, K. (1983) Science 222, 734-739.
36. Rosenberg, M. and Court, D. (1979) Ann. Rev. Genet. 13, 319-353.
37. Sarmientos, P., Sylvester, J., Contente, S. and Cashel, M. (1983) Cell 32, 1337-1346.
38. DeBoer, H., Gilbert, S. and Nomura, M. (1979) Cell 17, 201-209.