Volume 9 Number 16 1981 Nucleic Acids Research

Organization and nucleotide sequence of a new ribosomal operon in Escherichia coli containing the genes for ribosomal protein S2 and elongation factor Ts

Gynheung An[1], Dirk S.Bendiak[2], Linda A.Mamelak[1] and James D.Friesen[1]

Department of Biology, York University, 4700 Keele St., Downsview, Ontario M3J 1P3, Canada

Received 5 May 1981

ABSTRACT

We report the nucleotide sequence of the four min region of the Escherichia coli genetic map that includes the genes for ribosomal protein S2 (rpsB) and translation elongation factor EF-Ts (tsf), and the possible location of regulatory sites within this two gene operon. The data indicate that the gene order is: rpsBp-rpsB-tsf-tsft. One potential regulatory site is a 16 nucleotide sequence in the rpsB leader region encompassing the ribosome binding site and the translation initiation codon. This has a high degree of homology with nucleotides 8 through 23 on the 5' end of 16S ribosomal RNA, and might signify a sequence that is necessary for post-transcriptional control of rpsB expression. The data allow one to infer the amino acid sequences of S2 and EF-Ts.

INTRODUCTION

Studies of the organization of essential genes in Escherichia coli, especially those for components of the transcription and translation apparatus, have revealed the expected close linkage of related functions, such as among the structural genes for the ribosomal proteins (r-proteins) (3). Rather more unexpected is the location within the same operon of genes whose functions are associated, but not closely related. Thus the genes for β and β' subunits of RNA polymerase, rpoB and rpoC, are located in the same operon with the genes for r-proteins L10 and L12, rplJ and rplL (4-6). rpoA, the gene for RNA polymerase subunit α, is surrounded by r-protein structural genes (7). Among the genes for translation elongation factors one finds a similar situation; fus and tufA, genes for elongation factor G (EF-G) and elongation factor Tu (EF-Tu), are located as the distal-most two genes of a four gene operon that also includes two r-protein genes (8). We have recently reported the analysis of a group of genes in the four minute region of the E. coli linkage map (9). Among these were the genes for r-protein S2 (rpsB) and elongation factor EF-Ts (tsf) (10). The evidence suggested that these two genes lie side-by-side in a single transcriptional unit. As an extension of this study we report in this

communication the nucleotide sequence of the region and the location of possible regulatory sites within it.

MATERIALS AND METHODS

The E. coli strain used as a recipient for plasmids was MC1000 (araD139, ara(ABOIC), leu7697, lac(IPOYZ-X74), galU, galK, strA) (11) . The structure of plasmids pDB9 (10) and pGA39 (12) have been described earlier. The method of determining resistance to tetracycline on L-broth agar has been described (13).

Restriction endonuclease digestion, ligation with T4 DNA ligase, analysis of DNA fragment size by gel electrophoresis and preparation of plasmid DNA have been described previously (12).

The positions of the restriction endonuclease sites, determined in preparation for nucleotide sequence analysis, were on the basis of analyzing the size of DNA fragments from singly- or doubly-digested pDB9 DNA, or on the basis of radioactively end-labelling a DNA fragment and measuring by gel electrophoresis and autoradiography the fragment size after partial digestion.

DNA fragments were radioactively end-labelled with ^{32}P-deoxyribonucleotide triphosphates using the large fragment of E. coli DNA polymerase I (14). DNA fragments were purified by electrophoresis on 5 % polyacrylamide gels followed by electroelution. Nucleotide sequence analysis was carried out by the method of Maxam and Gilbert (14).

RESULTS AND DISCUSSION

Restriction Endonuclease Map of the rpsB-tsf Region

An earlier publication reported a genetic analysis of the four min region of the E. coli genetic map (15,16). Recently, we have cloned restriction endonuclease fragments of a specialized transducing bacteriophage, λpolC-9 (15), on a number of different plasmid cloning vehicles (10). A determination of which proteins were encoded on these plasmids enabled us to locate with a fair degree of precision rpsB and tsf, the structural genes for r-protein S2 and for EF-Ts. One of the plasmids that was analyzed is shown in Fig. 1. Plasmid pDB9 (10) carries a 3.3 Kb (kilobase) PstI fragment that was derived from λpolC-9 and which was cloned on pGA46 (12). Fig. 1 shows a detailed restriction endonuclease map of this plasmid along with the location of the two genes that concern us here. It has previously been shown (10) that this plasmid is capable of expressing rpsB and tsf, and that these two genes were probably in a single transcription unit whose promoter (rpsBp) lies to the left of rpsB, and whose direction of transcription is from left to right, as is indicated in Fig. 1

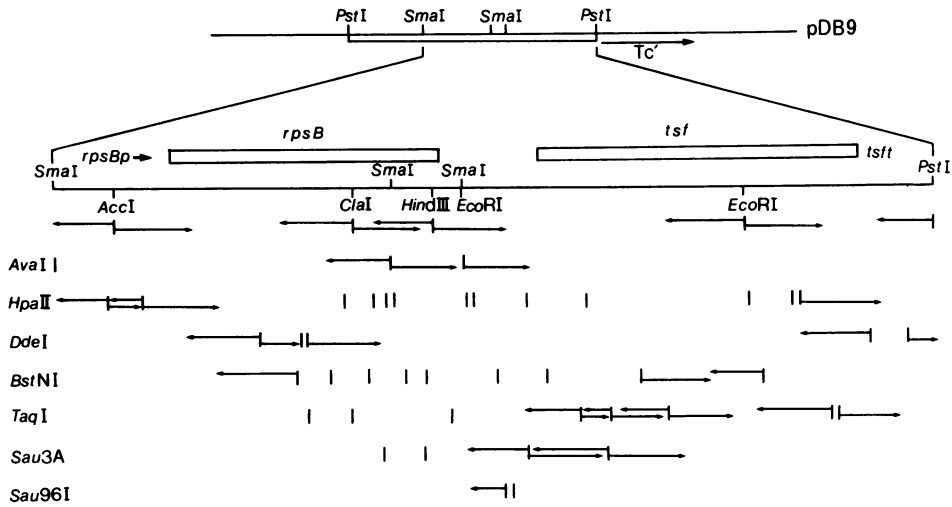

Fig. 1. A restriction map of pDB9 and of the rpsB-tsf region, including the tactics for obtaining its nucleotide sequence. The 3.3 Kb PstI fragment (double holizontal lines on pDB9) was derived directly from λpolC-9 and was cloned on plasmid pGA46 (single horizontal line on pDB9). pDB9 is shown in linear form as if it were opened at the EcoRI site in the chloramphenicol acetylase gene. The positions and sizes of the genes indicated by the open rectangles in the middle region of the figure are based on the nucleotide sequence data shown in Fig. 2. rpsBp symbolizes a promoter and tsft symbolizes a transcription termination site for the rpsB-tsf operon. The horizontal arrows represent the site of radioactive end-labelling and the direction and extent of reading in individual sequence analysis experiments.

Nucleotide Sequence of rpsB and tsf

In order to obtain a better picture of the genetic organization and of the control mechanisms that operate in this gene cluster we have determined the nucleotide sequence of the rpsB-tsf region. Fig. 1 shows the tactics used in this study. The sequence itself is shown in Fig. 2.

Several points are noteworthy:

1) The amino acid sequence of S2 as inferred from the nucleotide sequence of rpsB matches the complete amino sequence determined by Liebold-Wittmann and Wittmann (personal communication). There are 240 amino acids in S2 and its calculated molecular weight is 26,582. The inferred amino acid sequence of EF-Ts contains matches the sequence of three peptide fragments (including the N-terminal peptide of 18 amino acids) that have been determined by A. Wittinghofer (personal communication). There are 282 amino acids in EF-Ts and its calculated molecular weight is 30,257.

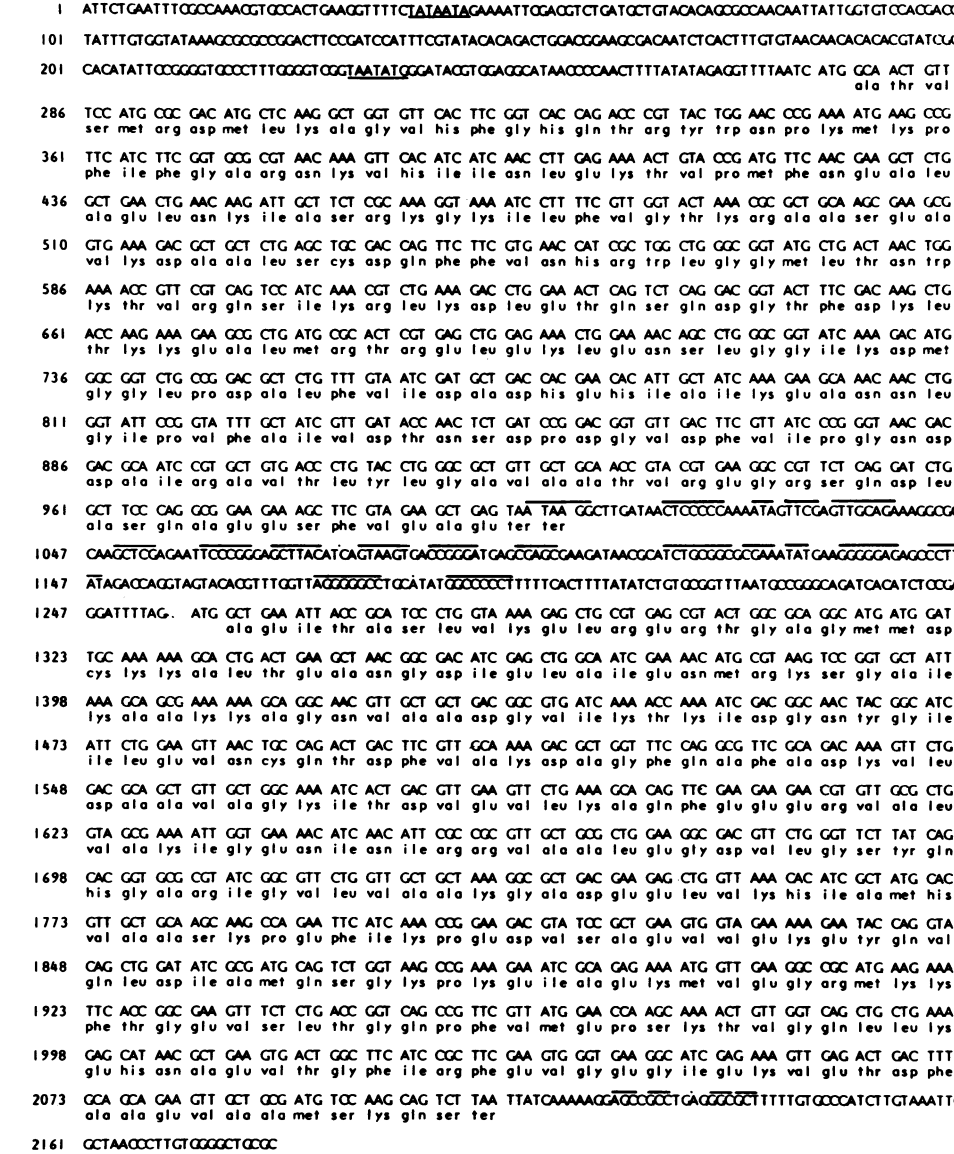

Figure 2. The DNA nucleotide sequence of the rpsB-tsf region. As is indicated in Fig. 1, for almost the entire region both DNA strands were analyzed. Possible RNA polymerase binding sites are underlined. Regions of inverted repeat are indicated by horizontal lines immediately above the relevant sequences.

2) The codon useage for both rpsB and tsf (Fig. 3) is typical of other genes for translation-associated functions (17). Whether this is related to the fact that most of these proteins are relatively abundant in the cell (18-20) or whether there are other more subtle reasons for components of the translation apparatus having a peculiarly skewed codon useage is not known.

3) There are at least two possible sites for promoters in the region upstream from S2. The two acceptable Pribnow sequences (21) are located at postions 38-44 and 231-237. However, neither of these has the typical -35 region that is often associated with the RNA polymerase recognition site (22). Nor is either of the Pribnow sequences followed by a CG-rich region, which has been proposed as necessary for stringent control of expression (23), even though tsf (and one assumes also rpsB) is regulated by the stringent mechanism (24). Nevertheless, data presented below indicate that the promoter for the operon, rpsBp, lies within 300 bp upstream from rpsB and thus presumably involves one of the Pribnow sequences that we have noted.

4) There is a stem-loop structure at 2121-2137 (ΔG=-12 Kcal/mol,approx.) (25) that might signify a rho-independent transcription termination sequence (tsft) following tsf (22). The data presented below also strongly suggest the

Second base

First base	U	a,b	C	a,b	A	a,b	G	a,b	Third base
U	Phe	2,1	Ser	4,4	Tyr	0,1	Cys	0,0	U
		10,9		3,4		2,2		1,2	C
	Leu	0,0		0,0	term	1,1	term	0,0	A
		0,0		0,0		0,0	Trp	3,0	G
C	Leu	2,0	Pro	0,0	His	1,1	Arg	8,5	U
		1,0		0,0		5,3		5,4	C
		0,0		0,2	Gln	0,0		0,0	A
		18,16		7,3		7,10		0,0	G
A	Ile	3,5	Thr	7,7	Asn	0,0	Ser	0,0	U
		11,14		6,4		12,8		4,2	C
	Met	0,0		0,0	Lys	14,23	Arg	0,0	A
		7,9		0,0		5,5		0,0	G
G	Val	9,20	Ala	16,17	Asp	4,2	Gly	12,9	U
		0,0		0,0		13,12		5,15	C
		5,5		5,14	Glu	12,24		0,0	A
		3,4		4,8		4,8		0,0	G

Figure 3. Frequency of codon usage in rpsB (a) and tsf (b). The number of occurences of each codon is indicated, including the termination codon, UAA.

presence of an active terminator within 200 bp downstream from tsf.

5) There are several possible internal regulatory sites in the operon. A typical rho-independent termination or attenuation site (22) lies in the intergenic region (1175-1195) between rpsB and rts (Fig. 2). This is preceded by a very large and extremely stable (ΔG=-80 Kcal/mol, approx.) (25) stem-loop structure (999-1148) that, in its RNA form, is reminiscent of an RNaseIII site that occurs at gene 1.0 in the T7 bacteriophage early transcript (26). It is interesting to note that the structures found in the region between rpsB and tsf are similar to those found between the first two and the last two genes of the rplJL-rpoBC operon (27). There is another interesting similarity between these two operons: in both cases distal genes (tsf on the one hand and rpoBC on the other) are expressed about one-third to one-fifth as strongly as the promoter-proximal gene(s) (19,28,29). In the case of the rplJL-rpoBC there is a transcription attenuation of about five-fold (19,29) and an RNaseIII recognition site (27) between the promoter-proximal and promoter-distal genes. We do not yet know whether this holds in the rpsB-tsf operon. The intergenic sequence suggests that it might, however the data presented below failed to reveal any evidence for an attenuator or a terminator between rpsB and tsf. The intergenic region between rpsB and tsf has at least two sequences that could correspond to Pribnow boxes. However, as discussed below, we do not think that either of these is active under normal conditions.

(6) There is a significant base sequence homology between the rpsB leader messenger RNA (mRNA) and the 5' end of 16S ribosomal RNA (rRNA) (Fig. 4). The region of homology encompasses the Shine and Dalgarno sequence (30) through the first five nucleotides of rpsB; this region of rspB leader mRNA had no significant homology with any other region of the 16S rRNA. This might be related to the presence of homologies between rRNA and other r-protein mRNAs, such as rpsD (r-protein S4) (31), rpsG (r-protein S7) (31), rpsT (r-protein S20) (G. Mackie, personal communication), rplJ (r-protein L10) (unpublished observations), and rplK (r-protein L11) (32). However, unlike most of the other cases cited (31,32), there is no significant secondary structure in either the

```
   U A U A U A G A G G U U U U A A U C A U G G C A A C   rpsB mRNA
       *   *   * *   * * *   * * * * * * * *
5'pA A A U U G A A G A G U U U G A U C A U G G C U C A   16S rRNA
```

Figure 4. Sequence homology between the leader region of rpsB-tsf mRNA and 16S rRNA of E. coli. The asterisks indicate homologous nucleotides. The Shine-Dalgarno sequence and the translation start codon for rpsB are underlined.

rpsB leader or the 16S rRNA, in the region of base sequence homology. It has been demonstrated that S4, S7, S8, L1, L4 and L10 are key proteins in the autoregulation at the post-transcriptional level not only of their own synthesis, but also of other genes in the same operon (31,33-38). Moreover, S4, S7, S20, L1 and L10 (as the L10/L7/L12 complex) bind directly to rRNA, as does S2 (weakly) (39,40). It is not yet known whether S20 and S2 are subject to feedback regulation. However, it should be noted that biosynthesis of S2 is subject to regulation, since cells that harbour a rpsB multi-copy plasmid do not over-produce S2 (10). These results fit an emerging picture of autogenous regulation of r-protein production that is governed at the post-transcriptional level by competition between a sequence in its own leader mRNA and a homologous sequence in rRNA for binding of a particular key r-protein.

Location of the Promoter and Terminator for the rpsB-tsf Operon

We have made use of the plasmid cloning system devised by An and Friesen (12) for a further analysis of possible regulatory regions in the rpsB-tsf operon. These plasmids carry the genes for tetracycline resistance (Tc) without their normal promoter; thus Tc can be activated by insertion of a foreign promoter. Plasmid pDB9, shown in Fig.1, is already such a construction. This plasmid was intially constructed by cloning the 3.3 Kb PstI fragment originally derived from λpolC-9 into the recipient plasmid vehicle, pGA46 (12). A promoter on this fragment activated Tc to the extent that the host strain that carried pDB9 became resistant to tetracycline at a concentration of 3 µg/ml (Fig. 5). However, this is almost 10-fold lower that we would have predicted on the basis of previous results with other r-protein promoters (13). From the structure of pDB9 (Fig. 5) it is not immediately obvious that the low expression of Tc is being driven by the rpsB promoter, rpsBp, or by some other unidentified promoter. This was clarified by the in vitro removal from pDB9 of one or both adjacent SmaI (XmaI) restriction fragments, to yield plasmids pGA207 and pGA209 (Fig. 5). In both cases host cells that carry either of these two derivative plasmids were resistant to tetracycline at a concentration of 1 µg/ml, suggesting that the promoter had been removed and thus is located in the position indicated in Fig. 1. A corroboration of the inferred promoter position was obtained by cloning the XmaI fragment that contains the proximal region of rpsB plus 300 bp to the 5' side of it. For this purpose another member of the promoter-cloning plasmid set (12), pGA39, was used. The resulting plasmid, designated pGA214 (Fig.5), endowed the host cell with resistance to 24 µg/ml tetracylcine. We conclude that rpsBp lies within the larger of the two SmaI fragments, as is shown in Fig. 1. These results confirm the position of

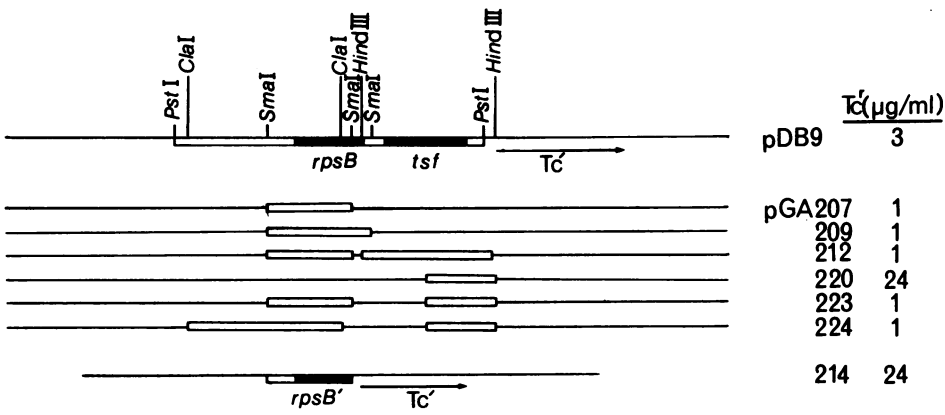

Figure 5. Structure of pDB9 and its derivatives. Bacterial DNA from λpolC-9 is indicated by double horizontal lines on pDB9. The plasmid is shown in linear form derived by opening at the EcoRI site in the chloramphenicol acetylase gene. The locations of rpsB and tsf are indicated by the filled boxes. The structures of several deletion plasmids are also illustrated. The deletions in these plasmids are represented by open boxes. The level of drug resistance bestowed on the host cell by each plasmid is given on the right-hand side of the diagram.

rpsBp that was inferred from the sequence analysis. However, we do not yet know the precise transcription start site. No functional promoter is able to read through from a position upstream from rpsB since plasmid pGA212, from which the entire rpsB-tsf operon has been excised, bestowed only 1 μg/ml tetracycline resistance on the host cell (Fig. 5).

From the characteristics of three other plasmids we conclude that a strong transcription terminator (tsft) lies distal to tsf. pGA220 is a spontaneous in vivo deletion that was selected as a mutant of pDB9 that bestowed 24 μg/ml resistance to tetracycline (Fig. 5). We infer that this deletion removes tsft. The nucleotide sequence indicates a possible stem-loop structure between rpsB and tsf that might correspond to the site of a transcription attenuator (Fig.2). However, we were unable to observe an attenuator effect since both pGA214 (which carries only rpsBp) and pGA220 (which carries rpsBp and the intergenic region) drive Tc to a level of 24 μg/ml.

Two in vitro derivatives were constructed from pGA220: pGA223 lacks the XmaI fragment that carries the major promoter of the operon (Fig. 5). This plasmid bestowed tetracylcine resistance at 1 μg/ml, indicating that there is no significant promoter activity downstream from rpsB, particularly in the intergenic region. Similarly, deletion of the ClaI fragment (pGA224) reduced

the tetracycline resistant level to 1 μg/ml (Fig. 5); this observation supports our earlier conclusion as to the position of the promoter for the operon.

In conclusion these deletion analyses indicate that there is only one major promoter for the rpsB-tsf operon, and it is located immediately upstream from rpsB. A strong transcription terminator is located downstream from tsf. We were unable to obtain evidence for the action of a possible transcription attenuator that on the basis of sequence analysis (Fig. 2) one might have expected to lie between rpsB and tsf.

ACKNOWLEDGEMENTS

We thank Pat Dennis for instructions in DNA sequencing, Philippe Marlière and Eric Kofoid for helping us with the computerized sequence comparisons, Kyungsook An for editional assistance, and Isobel Matthews and Pam Ohashi for technical assistance. This work was supported by The National Science and Engineering Research Council Canada and the National Cancer Institute of Canada.

REFERENCES

1. Present address: Department of Medical Genetics, University of Toronto, Toronto, Ontario, Canada M5S 1A8.
2. Present address: Department of Genetics, University of Alberta, Edmonton, Alberta, Canada T6G 2E9.
3. Nomura, M. (1976) Cell 9, 633-644.
4. Fiil, N. P., Bendiak, D., Collins, J. and Friesen, J. D. (1979) Molec. gen. Genet. 173, 39-50.
5. Linn, T. and Scaife, J. (1978) Nature 276, 33-37.
6. Yamamoto, M. and Nomura, M. (1978) Proc. Natl. Acad. Sci. U. S. A. 75, 3891-3895.
7. Jaskunas, S. R., Burgess, R. R. and Nomura, M. (1975) Proc. Natl. Acad. Sci. U. S. A. 72, 5036-5040.
8. Jaskunas, S. R., Lindahl, L., Nomura, M. and Burgess, R. R. (1975) Nature 257, 458-462.
9. Bachmann, B. J. and Low, K. L. (1980) Microbiol. Rev. 44, 1-56.
10. Bendiak, D. and Friesen, J. D. Molec. gen. Genet. in the press.
11. Casadaban, M. J. and Cohen, S. N. (1980) J. Mol. Biol. 138, 179-207.
12. An, G. and Friesen, J. D. (1979) J. Bacteriol. 140, 400-407.
13. An, G. and Friesen, J. D. (1980) J. Bacteriol. 144, 904-916.
14. Maxam, A. M. and Gilbert, W. (1980) Methods in Enzymol. 65, 499-580.
15. Friesen, J. D., Parker, J., Watson, R. J., Bendiak, D., Reeh, S. V., Pedersen, S. and Fiil, N. P. (1976) Molec. gen. Genet. 148. 93-98.
16. Yamamoto, M., Strycharz, W. A. and Nomura, M. (1976) Cell 8, 129-138.
17. Post, L. E. and Nomura, M. (1980) J. Biol. Chem. 255, 4660-4666.
18. Schleif, R. (1967) J. Mol. Biol. 27, 41-55.
19. Pedersen, S., Bloch, P. L., Reeh, S. and Neidhardt, F. (1978) Cell 14, 179-186.
20. Gaussing, K. (1980) in Ribosomes: structure, function, and genetics, Chambliss, G. Craven, G. R., Davies, J., Davis, K., Kahan, L. and Nomura, M. Eds., pp. 693-718, University Park Press, Baltimore.

21. Pribnow, M. (1975) Proc. Natl. Acad. Sci. U. S. A. 72, 784-789.
22. Rosenberg, M. and Court, D. (1979) Annu. Rev. Genet. 13, 319-353.
23. Travers, A. A. (1980) J. Bacteriol. 141, 973-976.
24. Reeh, S., Pedersen, S. and Friesen, J. D. (1976) Molec. gen. Genet. 149, 279-289.
25. Uhlenbeck, O. C., Clothers, D. M. and Gralla, J. (1973) Nature New Biol. 246,40-41.
26. McConnell, D. J. (1979) Nucleic Acid Res. 6, 3491-3503.
27. Barry, G., Squires, C. and Squires, C. L. (1980) Proc. Natl. Acad. Sci. U. S. A. 77, 3331-3335.
28. Lindahl, L., Yamamoto, M., Nomura, M., Kirschbaum, J. B., Allet, B. and Rochaix, J. -D. (1977) J. Mol. Biol. 109, 23-47.
29. Dennis, P. P. (1977) J. Mol. Biol. 115, 603-625.
30. Shine, J. and Dalgarno, L. (1974) Proc. Natl. Acad. Sci. U. S. A. 71, 1342-1346.
31. Branlant, C., Krol, A., Machatt, A. and Ebel, J. -P. (1981) Nucleic Acid Res. 9, 293-307.
32. Nomura, M., Yates, J. L., Dean, D. and Post, L. E. (1980) Proc. Natl. Acad. Sci. U. S. A. 77,7084-7088.
33. Fiil, N. P., Friesen, J. D., Downing, W. L. and Dennis, P. P. (1980) Cell 19, 837-844.
34. Brot, N., Caldwell, P. and Weissbach, H. (1980) Proc. Natl. Acad. Sci. U. S. A. 77, 2592-2595.
35. Yates, J., Arfsten, A. and Nomura, M. (1980) Proc. Natl. Acad. Sci. U. S. A. 77, 1837-18.
36. Dean, D. and Nomura, M. (1980) Proc. Natl. Acad. Sci. U. S. A. 77, 3590-3594.
37. Yates, J. L. and Nomura, M. (1980) Cell 21, 517-522.
38. Zengel, J. M., Muckle, D. and Lindahl, L. (1980) Cell 21, 522-535.
39. Zimmermann, R. A. (1980) in Ribosomes: structure, function, and genetics, Chambliss, G. Craven, G. R., Davies, J., Davis, K., Kahan, L. and Nomura, M. Eds., pp. 135-169, University Park Press, Baltimore.
40. Littlechild, J., Dijk, J. and Garrett, R. A. (1977) FEBS Lett. 74, 292-294.