Volume 7 Number 5 1979 Nucleic Acids Research

Homology of the 3' terminal sequences of the 18S rRNA of Bombyx mori and the 16S rRNA of Escherichia coli

David R.Samols*†, Otto Hagenbuchle+ and L.Patrick Gage*

*Department of Cell Biology, Roche Institute of Molecular Biology, Nutley, NJ 07110, and +Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520, USA

Received 18 September 1979

ABSTRACT

The terminal 220 base pairs (bp) of the gene for 18S rRNA and 18 bp of the adjoining spacer rDNA of the silkworm Bombyx mori have been sequenced. Comparison with the sequence of the 16S rRNA gene of Escherichia coli has shown that a region including 45 bp of the B. mori sequence at the 3' end is remarkably homologous with the 3' terminal E. coli sequence. Other homologies occur in the terminal regions of the 18S and 16S rRNAs, including a perfectly conserved stretch of 13 bp within a longer homology located 150-200 bp from the 3' termini. These homologies are the most extensive so far reported between prokaryotic and eukaryotic genomic DNA.

INTRODUCTION

The 3' terminal twenty nucleotides of a variety of eukaryotic 18S rRNA molecules are almost completely homologous to one another and to the 3' end of 16S rRNA from E. coli (1,2). Although the reason for this sequence conservation is unknown, it does imply that this region of the smaller rRNA functions in a process both critical and common to both prokaryotes and eukaryotes. A key function of this region in E. coli is in mRNA recognition and binding during the formation of the initiation complex for protein synthesis (3,4). However, the CCUCC pyrimidine tract that is involved in message recognition by E. coli 16S rRNA is specifically missing in eukaryotic 18S rRNAs (1-4). Instead a purine rich sequence in the 18s rRNA which is complementary to sequences within the 5' non-translated regions of some eukaryotic mRNA sequences may serve this function (2,5), although the absence of this complementarity in other mRNAs has brought this into question (6).

In an effort to understand more fully the important role that the 3' end of the smaller rRNA plays in the cell, we have further compared the terminal sequences of the E. coli 16S rRNA and the 18S rRNA of the silkworm, Bombyx mori. The sequence of the 220 bp comprising the 3' end of the B. mori 18S rRNA gene and an additional 18 bp of the flanking transcribed

spacer was determined using a silkworm rDNA recombinant plasmid previously described (7). The results reveal considerable homology between the 3' terminal 200 bp of the smaller rRNAs of E. coli and B. mori, the most extensive encompassing the very terminal 45 bp. The functional implications of these homologies are discussed.

MATERIALS AND METHODS

Restriction endonucleases HaeIII, HpaI, HincII, BamHI, AluI and AvaI were purchased from New England Biolabs. EcoRI was purchased from Miles and Sau3A from Bethesda Research Labs. T4 polynucleotide kinase was purchased from P-L Biochemicals or New England Biolabs. Bacterial alkaline phosphatase was the gift of Jan Chlebowski. γ-^{32}P-ATP was prepared at a specific activity of about 1500 Ci/mmole by the method of Glynn and Chappell (8), as described by Maxam and Gilbert (9).

We have used a small piece of DNA isolated from the recombinant plasmid pBml (7) for this study. pBml consists of a complete ribosomal DNA repeat from the silkworm Bombyx mori inserted at the EcoRI site of the plasmid RSF2124 (10). An 1100 bp segment bounded by sites for the restriction endonucleases EcoRI and HpaI contains the 3' end of the 18S rRNA gene, the 5' end of the 5.8S RNA gene and the intervening transcribed spacer (7). The 1100 bp segment was prepared by digesting 1 mg of pBml with EcoRI and HpaI followed by 2 rounds of sucrose gradient centrifugation. The gradients were 5-20% sucrose in 0.1 M NaCl, 10 mM Tris, pH 7.5, 2 mM EDTA and were spun in a SW40 at 39,000 rpm for 18 hr. Gradient fractions were analyzed by electrophoresis on a 1% agarose slab gel (15 x 18 cm), at 30 mA for 8 hr and stained with 0.5 µg/ml ethidium bromide. Fractions containing the EcoRI/HpaI fragment were pooled and precipitated. In some cases, the DNA from the slab gel was transferred to nitrocellulose paper by the Southern blot technique and hybridized with 18S[^{125}I] rRNA (11), or 5.8S[^{32}P] RNA (7). Preparatively isolated EcoRI/HpaI fragment DNA was subcloned as described in Results.

Labeling and Preparation of DNA Segments

5-10 µg of a subcloned 1050 base pair insert, either full length or restriction enzyme cut, was treated with ~0.5 µg bacterial alkaline phosphatase (BAP) in 50 µl of 10 mM Tris, pH 8.1, for 1 hr at 37°C. The BAP was removed after the digestion by phenol extraction and contaminating phenol was removed by several ether extractions. 5' end labeling with ^{32}P was carried out using T4 polynucleotide kinase as described by Maxam and Gilbert

(9). Typically, 10 µg of BAP treated DNA in a reaction mix of 50 µl containing 5 mM Tris, pH 9.5, 0.1 mM EDTA, 0.1 mM dithiothrietol, 20 µM [γ-^{32}P]ATP of specific activity ~ 1500 Ci/mmole, was incubated with 1 unit T4 polynucleotide kinase for 1 hr at 37°C. The reaction mixture was then passed over a Sephadex G-50 column (1 x 12 cm) in 10 mM Tris, pH 7.9 and the voided ^{32}P-labeled DNA was precipitated in ethanol. To prepare DNA fragments which were ^{32}P labeled on only one end, kinased fragments were digested with a second restriction endonuclease and separated by electrophoresis on either 5 or 7% acrylamide (40:1, acrylamide:bis) gels in 50 mM Tris Borate, pH 8.3, 5 mM $MgCl_2$. In some cases, DNA fragments were strand separated according to Maxam and Gilbert (9) on 8% acrylamide gels (30:1; acrylamide:bis) containing 7 M urea. Samples were layered on thin gels in small volumes of NaOH. Gels were quickly autographed (15-30 min), labeled DNA bands cut out and the DNA electroeluted out of the gel slices and into a dialysis bag in 50 mM Tris-acetate, pH 7.9. Electroeluted fragments were extracted once with phenol:chloroform (1:1), precipitated with EtOH and dissolved in H_2O. Sequencing of these end labeled fragments was done according to Maxam and Gilbert (9). In a typical sequence run, two gels were used, a 20% 20 x 40 cm slab and a 12.5% 20 x 80 cm slab. 100-150 bases could usually be read from both gels in a single run.

RESULTS

An 1100 base pair DNA segment is cleaved from the B. mori rDNA recombinant plasmid, pBm1, by the restriction endonucleases EcoRI and HpaI. This segment contains 220 base pairs of the 3'-terminal sequence of the 18S rRNA gene at its EcoRI end, 810 base pairs of transcribed spacer and 70 base pairs of the 5'-terminal sequence of the 5.8S RNA gene at its HpaI end (Figure 1) (7). In order to isolate and analyze this segment more easily, we sought to subclone it in the plasmid pBR322. Restriction mapping (12) revealed six Sau3A sites (↓GATC) within the segment, and one of particular utility within the 5.8S gene 20 nucleotides from its 5' end (Figure 1). Sequencing showed that this Sau3A site had a context (G↓GATCA) such that when cleaved, one of the products had the same staggered end as that produced by BamHI digestion. Hence, following cleavage of the EcoRI/HpaI segment at this Sau3A site, the EcoRI/Sau3A segment (1050 bp) could be inserted in place of the small EcoRI/BamHI segment of pBR322, and could be recovered by digestion of this subclone with EcoRI and BamHI.

In order to construct the desired subclone, the 1100 bp segment was

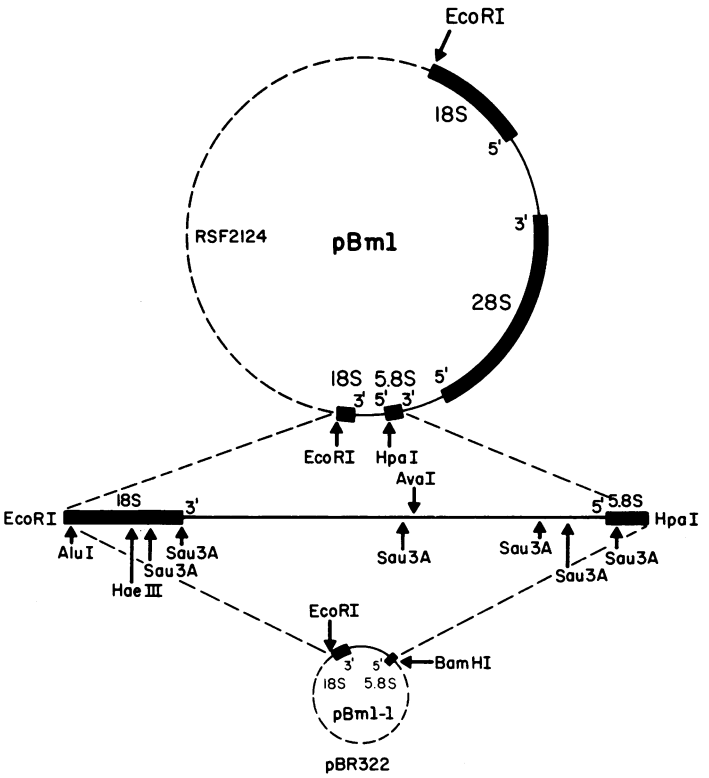

Fig. 1. Map of B. mori rDNA plasmids. Recombinant pBml contains a complete silkworm rDNA repeat (10.5 Kb) inserted at the EcoRI site of the vector RSF2124 (7). The restriction map of the 1100 bp segment removed from pBml by the restriction enzyme EcoRI and HpaI is shown at increased resolution. The subclone, pBml-1, contains a 1050 bp segment derived from the EcoRI/HpaI segment by cleavage at the Sau3A site in the 5.8S RNA gene.

cleaved from pBml, purified by sucrose gradient fractionation, and digested with Sau3A so that less than one cleavage occurred per molecule. This partial digestion was necessary to guarantee that some molecules were cleaved at the desired site in the 5.8S RNA gene only, thus maintaining the integrity of the remainder of the segment. After ligation of the Sau3A digested segment with pBR322 which had been cleaved by EcoRI and BamHI, and transformation of HB101 (13) only colonies which hybridized both labeled 18S rRNA and 5.8S RNA were selected (Figure 1). Two of these, out of six examined, contained the desired 1050 base pair insert, and one of these, pBml-1, was shown to have a restriction map colinear with the parental 1100 base pair segment. This

1050 bp segment contains 220 bp from the 3' end of the 18S rRNA gene, 810 bp of transcribed spacer and 20 bp from the 5' end of the 5.8S RNA gene (Figure 1).

The 18S rRNA gene sequence contained on the segment was determined according to Maxam and Gilbert (9). Sequencing was performed using molecules labeled at sites for the enzymes AluI, HaeIII and Sau3A in the 18S gene region of Figure 1. Eighty percent of this sequence has been confirmed by sequencing of the "opposite" strand and all of the sequence has been determined from more than one labeled restriction site. Figure 2 shows a typical sequencing gel of the region 5' of the HaeIII site (Figure 1). Figure 3 shows the sequence of the 220 bp at the 3' end of the 18S RNA

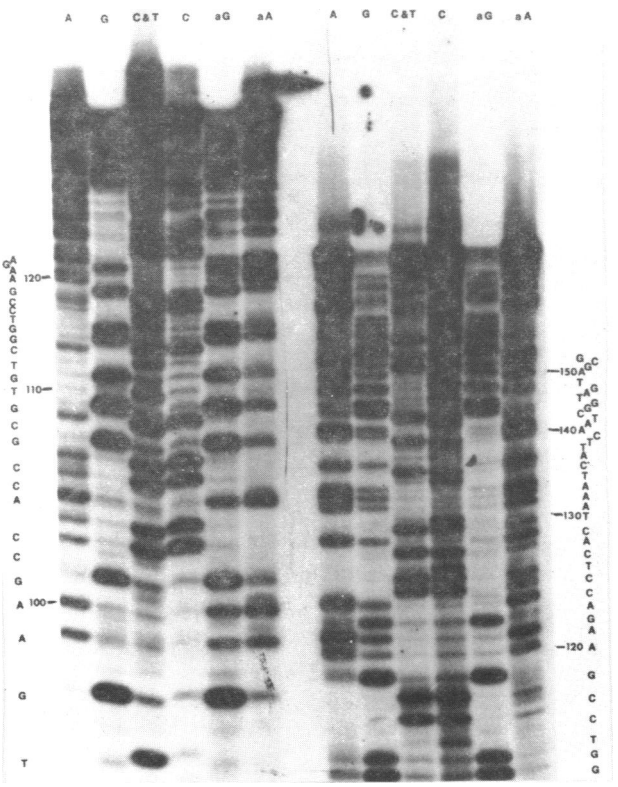

Fig. 2. Sequencing gel autoradiograph of the region of the 18S gene 5' to the HaeIII site shown in Figure 1. The sequence is numbered beginning at the 3' end of the 18S rRNA gene proceeding 5' as in Figure 3. The lanes are labeled according to the nomenclature of Maxam and Gilbert (9).

220	210	200	190	180	170
TTCCAGTAAG	CGCGAGTCAT	AAGCTCGCGT	TGATTACGTC	CCTGCCCGTT	GTAACCACAC
AAGGTCATTC	GCGCTCAGTA	TTCGAGCGCA	ACTAATGCAG	GGACGGGCAA	CATTGGTGTG
160	**150**	**140**	**130**	**120**	**110**
CGCCCGTCGC	TACTACCGAT	TGAATGATTT	AGTGAGGTCT	TCGGACCGAC	ACGCGGTGGC
GCGGGCAGCG	ATGATGGCTA	ACTTACTAAA	TCACTCCAGA	AGCCTGGCTG	TGCGCCACCG
100	**90**	**80**	**70**	**60**	**50**
TTCACGGCCG	TCGGCGTTGG	GAAGTTGACC	AAACTTGATC	ATTTAGAGGA	AGTAAAAGTC
AAGTGCCGGC	AGCCGCAACC	CTTCAACTGG	TTTGAACTAG	TAAATCTCCT	TCATTTTCAG
40	**30**	**20**	**10**	**-1**	**-10**	**-18**
GTAACAAGGT	TTCCGTAGGT	AACCTGCGGA	AGGATCATTA	ACGGGTGATG	GGAAGAAA---
CATTGTTCCA	AAGGCATCCA	TTGGACGCCT	TCCTAGTAAT	TGCCCACTAC	CCTTCTTT---

Fig. 3. Sequence of the 3' end of the 18S rRNA gene of Bombyx mori and a portion of its adjacent spacer region. The sequence is numbered beginning at the 3' end of the gene proceeding 5' into the gene. The lower strand corresponds to the coding strand.

gene and 18 bp of the adjacent transcribed spacer.

DISCUSSION

The 3' terminal 18 bases of a representative group of eukaryotic 18S rRNAs are identical to one another and are for the most part homologous with the 3' sequence of E. coli 16S rRNA (1,2). Alberty et al. (14) sequenced a 32 nucleotide RNA fragment derived from this region of rat 18S rRNA and showed that this homology with E. coli extends at least 35 nucleotides in from the 3' end. The B. mori sequence presented here, which was determined from a cloned rDNA repeat, encompasses 220 bases of the 3' end of the 18S gene and 18 nucleotides of the 3' adjacent transcribed spacer. The corresponding B. mori 18S rRNA sequence is aligned so as to yield maximum homology with the E. coli 16S rRNA sequence (15,16) as well as the known rat sequence (14) (Figure 4). This comparison reveals that the terminal homology with E. coli includes a total of 45 nucleotides of the 18S rRNA and contains a perfectly conserved stretch of 15 residues.

Extensive homology between the smaller rRNAs of E. coli and B. mori exists throughout the entire 3' terminal sequences (Figure 4). The region

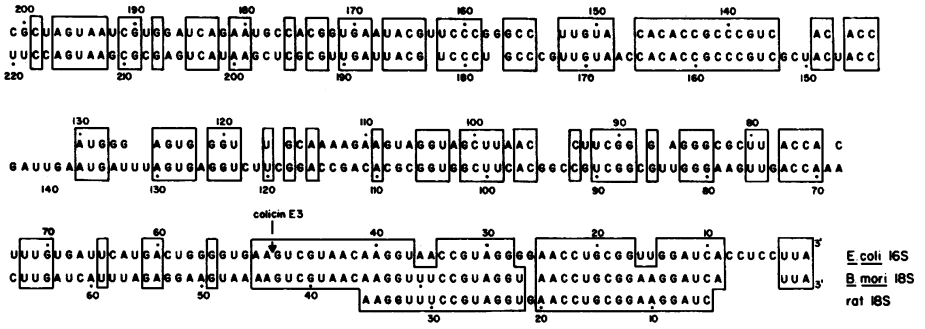

Figure 4. The homologous sequences at the 3' ends of B. mori, rat (14) and E. coli (15,16) small subunit rRNAs. The sequences are aligned for maximum homology, and the nucleotides of rat and E. coli which are homologous with the silkworm sequence are in boxes. The location of the colicin E3 cleavage is noted by an arrow (37). The sequences are numbered 3'→5' beginning at the 3' terminus.

of the E. coli sequence from positions 70-105 is well conserved in the B. mori sequence with only 9 of the 35 residues not shared in common, while all the E. coli sequence from 128-156 is present in the B. mori sequence and this region includes one perfectly homologous stretch of 13 residues. In total, 139 of the 201 3' terminal residues of the E. coli 16S rRNA sequence are found in the B. mori sequence, 114 residues in blocks of 3 nucleotides or larger. That these sequences have a common function and had a common origin in E. coli and B. mori is suggested by their similar location in both small rRNA molecules as well as by their homology. These highly conserved regions of the small rRNAs are the longest sequences common to a prokaryotic and eukaryotic RNA or DNA molecules so far reported.

The few differences between the E. coli and B. mori sequences within the 3' terminal conserved region make strong predictions about the structure of this part of the small rRNA in the ribosome. Figure 5 shows stem and loop configurations which can be derived from the E. coli, B. mori and rat sequences. At the base of the stem, the UU to AA difference at position 15-16 (E. coli) between prokaryotes and eukaryotes has an exactly reciprocal AA to UU difference at position 35-36 (E. coli). This reciprocal difference first noted by Alberty et al. (14) in the rat rRNA sequence, perfectly preserves the base pairing of the stem and therefore strongly implicates this stem in an important function in vivo. All other eukaryotic rRNAs examined have the same sequence at these two sites as B. mori 18S rRNA,

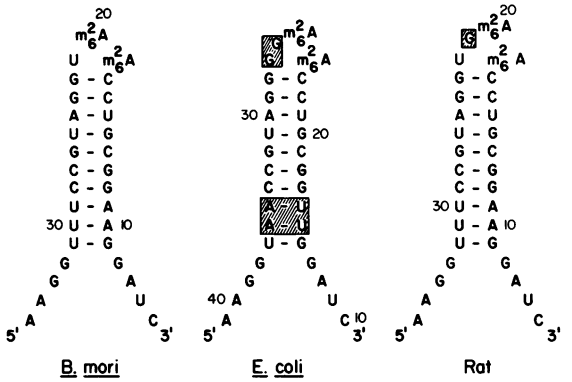

Fig. 5. Secondary structure of the small subunit rRNAs in their 3' terminal homologous regions. Each sequence is numbered beginning from the 3' terminal nucleotide. The nucleotides of the rat (14) and E. coli (15,16) sequence which differ with the silkworm sequence are in shaded boxes.

except for Dictyostelium discoideum in which positions 15, 16 (E. coli) are UA rather than AA (1,2). In order to maintain the apparently critical base pairing of the stem, we would predict that Dictyostelium rRNA should have UA at positions 35-36 (E. coli) (Figure 5). The loop sequence centered on adjacent m^2_6A residues contains the most sequence divergence between rat, silkworm and E. coli, and therefore it is likely that the loop has a less critical function than the more highly conserved sequences comprising the remainder of the terminal homologous region. Although two m^2_6A residues are shown in all three sequences in Figure 5, such modifications have not been demonstrated in the silkworm, and only one m^2_6A residue has been unequivocally identified in the rat (14). However, an oligonucleotide containing adjacent m^2_6A residues is common to many eukaryotic 18S rRNAs (1,17,18) and it is likely to be present at this same position since the 18S rRNAs share with E. coli 16S rRNA the property that reverse transcription is blocked at this site in the sequence (2).

Conservation of sequence over so vast an evolutionary span indicates that the homologous regions of the small rRNAs must serve the same or a similar critical function(s) in all of these organisms. However, the nature of this common function(s) is readily apparent neither from the sequence and structure of the homologous region nor from our knowledge of protein synthesis. The best defined function of the 3' terminal region of E. coli 16S rRNA is in mRNA recognition and ribosome binding. The CCUCC pyrimidine

tract (Figure 4; E. coli positions 4-9) which is complementary to the ribosome binding sites of prokaryotic mRNAs (3) and which has been shown to interact specifically with mRNA binding sites (4,19) is specifically missing from eukaryotic 18S rRNAs. Hagenbuchle et al. (2) postulated an alternative interaction between eukaryotic mRNAs and a purine tract (GGAAGGA) located at positions 7-13 of the eukaryotic rRNA sequence (Figure 4). Certain mRNAs could interact there quite strongly, as in the case of late adenovirus 2 mRNA (5), resulting in the opening of the stem structure described above (Figure 5). The universality of this proposed interaction of the 18S rRNA with eukaryotic mRNAs has been seriously questioned (6). While direct proof of the involvement of such an mRNA:rRNA duplex in translation initiation complex formation in eukaryotes is lacking (6,20), it was found recently that UV photoreaction of initiation complexes with 4'-aminomethyl-4,5',8-trimethylpsoralen cross-links the 5'-terminal portion of reovirus mRNA specifically to 18S rRNA, presumably in the 3'-region (K. Nakashima and A. J. Shatkin, personal communication).

Functional differences in initiation specificity include the fact that prokaryotes have rather strict interspecies translational specificities while eukaryotes do not (20). Only one clear example of the proper functioning of a eukaryotic viral mRNA in an E. coli extract has been reported (20,21), however prokaryotic mRNAs, when "capped", translate as well or better than eukaryotic mRNAs in a eukaryotic cell free extract (22,23). Therefore, the essential features of the small ribosomal subunit involved in mRNA recognition, other than the CCUCC sequence of the 16S rRNA, must have been strongly conserved in prokaryotes and eukaryotes. The proposed interactions with the mRNAs could involve the entire 3' terminal homologous region which extends 45 bp into the 18S rRNA. One clue to the critical nature of this region is suggested by colicin E3 cleavage of the terminal homologous sequence in E. coli 16S rRNA (Figure 4), which destroys the ribosome's ability to elongate peptides (24), although the ribosome is still able to form initiation complexes (25). Studies using colicin E3 have shown that the 3' terminal 49 nucleotides of the 16S rRNA are not strongly associated with any ribosomal proteins (26,27). The terminal region has also been shown to be highly susceptible to RNase digestion (28,29) and sensitive to chemical modification (30) in the 30S ribosome again implying that proteins are not associated with the RNA there. However, formation of the 70S ribosome, protects the 3' terminal region from these agents (29,30). Recent studies by immunoelectron microscopy (31,32) suggest

that the 3' end of E. coli 16S rRNA lies in the platform region of the 30S subunit (33,34), at the site of mRNA decoding.

The 3' terminal one third of the E. coli 16S rRNA including all the homologous regions described here, but excluding the terminal colicin E3 segment, is also intimately involved with translation. This was shown by mRNA dependent, photochemical crosslinking of tRNA derivatives directly to E. coli 16S rRNA at the P site of 70S tight couple ribosomes (35,36). Strongly conserved sequences, like those we have described here, may take part in intermolecular interactions with tRNA and mRNA which are critical to the translation process in all organisms.

Present Address: †Institute of Molecular Biology, University of Oregon, OR 97403, USA
+ISREC, Ch. des Boveresses, 1066 Epa Pinges, Lausanne, Suisse.

REFERENCES

1. de Jonge, R., Klootwijk, J. and Planta, R. J. (1977) Nucleic Acids Res. 4, 3655-3662.
2. Hagenbuchle, O., Santer, M., Steitz, J. A. and Mans, R. J. (1978) Cell 13, 551-563.
3. Shine, J. and Dalgarno, L. (1974) Biochem. J. 141, 609-615.
4. Steitz, J. A. and Jakes, K. (1975) Proc. Nat. Acad. Sci. USA 72, 4734-4738.
5. Ziff, E. B. and Evans, R. M. (1978) Cell 15, 1463-1465.
6. Barallee, F. E. and Brownlee, G. G. (1978) Nature 274, 84-87.
7. Manning, R. F., Samols, D. R. and Gage, L. P. (1978) Gene 4, 153-166.
8. Glynn, I. M. and Chappell, J. B. (1964) Biochem. J. 90, 147-149.
9. Maxam, A. M. and Gilbert, W. (1977) Proc. Nat. Acad. Sci. USA 74, 560-564.
10. So, M., Gill, R. and Falkow, S. (1975) Mol. Gen. Genet. 142, 239-249.
11. Gage, L. P. and Manning, R. F. (1976) J. Mol. Biol. 101, 327-348.
12. Smith, H. O. and Birnstiel, M. L. (1976) Nucleic Acids Res. 3, 2387-2398.
13. Cohen, S. N., Chang, A. C. Y. and Hsu, L. (1972) Proc. Nat. Acad. Sci. USA 69, 2110-2114.
14. Alberty, H., Raba, M. and Gross, H. J. (1978) Nucleic Acids Res. 5, 425-434.
15. Young, R. A. and Steitz, J. A. (1978) Proc. Nat. Acad. Sci. USA 75, 3593-3597.
16. Brosius, J., Palmer, M. L., Kennedy, P. J. and Noller, H. F. (1978) Proc. Nat. Acad. Sci. USA 75, 4801-4805.
17. Klootwijk, J., VandenBos, R. C. and Planta, R. J. (1972) FEBS Lett. 72, 361-369.
18. Khan, M. S. N. and Maden, B. E. H. (1976) J. Mol. Biol. 101, 235-254.
19. Steitz, J. A. and Steege, D. A. (1977) J. Mol. Biol. 114, 545-558.
20. Kozak, M. (1978) Cell 15, 1109-1124.
21. Klein, W. H., Nolan, C., Lazar, J. M. and Clark, J. M. (1972) Biochemistry 11, 2009-2014.
22. Patterson, B. M. and Rosenberg, M. (1979) Nature 279, 692-696.
23. Rosenberg, M. and Patterson, B. M. (1979) Nature 279, 696-701.

24. Tai, P. C. and Davis, B. D. (1974) Proc. Nat. Acad. Sci. USA 71, 1021-1025.
25. Ravetch, J. V. and Jakes, K. S. (1976) Nature 262, 150-153.
26. Turnowsky, F. and Hogenauer, G. (1973) Biochem. Biophys. Res. Commun. 55, 1246-1254.
27. Laughrea, M. and Moore, P. B. (1978) J. Mol. Biol. 121, 411-430.
28. Rinke, J., Ross, A. and Brimacombe, R. (1977) Eur. J. Biochem. 76, 189-196.
29. Santer, M. and Shane, S. (1977) J. Bact. 130, 900-910.
30. Chapman, N. M. and Noller, H. F. (1977) J. Mol. Biol. 109, 131-149.
31. Olson, H. M. and Glitz, D. G. (1979) Proc. Nat. Acad. Sci. USA 76, 3769-3773.
32. Keren-Zur, M., Boublik, M. and Ofengand, J. (1979) Proc. Nat. Acad. Sci. USA 76, 1054-1058.
33. Lake, J. A. (1976) J. Mol. Biol. 105, 131-159.
34. Boublik, M. and Hellman, W. (1978) Proc. Nat. Acad. Sci. USA 75, 2829-2833.
35. Ofengand, J., Liou, R., Kohut, J., Schwartz, I. and Zimmerman, R. A. (1979) Biochemistry 18, in press.
36. Zimmerman, R. A., Gates, S. M., Schwartz, I. and Ofengand, J. (1979) Biochemistry 18, in press.
37. Nomura, M., Sidikaro, J., Jakes, K. and Zinder, N. (1974) in The Ribosomes, Nomura, M., Tissieres, A. and Lengyel, P., Eds., ;;. 805-814 Cold Spring Harbor, New York.