*Proc. Natl. Acad. Sci. USA*
Vol. 77, No. 12, pp. 7084–7088, December 1980
Biochemistry

# Feedback regulation of ribosomal protein gene expression in *Escherichia coli*: Structural homology of ribosomal RNA and ribosomal protein mRNA

(translational repressor/ribosome assembly/RNA–protein interaction/RNA secondary structure)

MASAYASU NOMURA, JOHN L. YATES, DENNIS DEAN, AND LEONARD E. POST*

Institute for Enzyme Research, and the Departments of Genetics and Biochemistry, University of Wisconsin, Madison, Wisconsin 53706

*Contributed by Masayasu Nomura, August 25, 1980*

**ABSTRACT** **Certain ribosomal proteins (r proteins) in *Escherichia coli*, such as S4 and S7, function as feedback repressors in the regulation of r-protein synthesis. These proteins inhibit the translation of their own mRNA. The repressor r proteins so far identified are also known to bind specifically to rRNA at an initial stage in ribosome assembly. We have found structural homology between the S7 binding region on 16S rRNA and a region of the mRNA where S7 acts as a translational repressor. Similarly, there is structural homology between one of the reported S4 binding regions on 16S rRNA and the mRNA target site for S4. The observed homology supports the concept that regulation by repressor r proteins is based on competition between rRNA and mRNA for these proteins and that the same structural features of the r proteins are used in their interactions with both rRNA and mRNA.**

In exponentially growing *Escherichia coli* cells, the rates of synthesis of most if not all ribosomal proteins (r proteins) are essentially identical to the rate of ribosome accumulation and coordinately regulated in response to environmental changes. A feedback model for translational regulation has been suggested to explain this balanced synthesis of r proteins (1), and there is experimental evidence that supports and refines this model (2–4). In this model, it is proposed that certain free r proteins act as feedback inhibitors of the translation of their own mRNAs and that, as long as the assembly of ribosomes removes r proteins, the corresponding mRNA escapes from the inhibition and continues to direct their synthesis. It has been shown *in vitro*, by using a protein synthesizing system that has various template DNAs carrying r-protein genes (2) and, *in vivo*, by examining the effect of overproduction of certain r proteins on the synthesis of other r proteins (3), that L1 (and not L11) is the translational repressor that regulates the synthesis of both L11 and L1 in the L11 operon (see Fig. 1). Similarly, S4 (2, 3), S8 (2), S7 (see below), L4 (4, 5), and L10 (6, 7; unpublished results) have been shown to be translational repressors involved in the regulation of the $\alpha$, *spc*, *str*, S10, and $\beta$ operons, respectively.

Repressor r proteins probably interact with a specific region of polycistronic mRNA and cause their inhibitory effects indirectly on distal genes through a "polar effect." For example, the *in vitro* synthesis of L1 from DNA templates that carry the L1 gene but lack the promoter and an adjacent segment of the L11 gene is not inhibited by L1. This and other experiments have shown that the site of action of L1 is localized in a region of the mRNA that includes the beginning part of the L11 message (unpublished data). One plausible way to explain these observations and the balanced synthesis of L11 and L1 is to assume that every ribosome initiated at the beginning of the L11 message continues to translate the L1 message and that no independent translational initiation takes place at the beginning of the L1 message. By applying this kind of model to other "units of regulation," which may or may not be identical to the units of transcription, one can formally explain the balanced and coordinately regulated synthesis of the r proteins through the direct regulation of the synthesis of a single ribosomal component. Presumably, the regulation of rRNA synthesis in most conditions can theoretically lead to the coordinate regulation of the synthesis of all the other components through the postulated translational feedback-regulation model.

One common feature of the repressor proteins so far identified is that they are among the "initial binding proteins" in the *in vitro* assembly of ribosomes and have strong and specific binding to rRNA. Thus, the regulation can be regarded as competition between rRNA and r-protein mRNA for repressor r proteins, and it is plausible that the affinity of the repressor r proteins for rRNA is much stronger than their affinity for the target sites on their mRNA. In fact, the translational repression observed in the *in vitro* experiments by L1 can be abolished by the presence of 23S rRNA in the incubation mixtures (unpublished results). Although other possibilities exist, a simple way to evolve such a regulatory mechanism would be to use the same region of pertinent repressor r protein to interact with both rRNA and r-protein mRNA. If this is so, one might find some structural homology between r-protein binding sites on rRNA and the putative target sites on r-protein mRNA. We have in fact found such homology in connection with r proteins S7 and S4.

## S7 sites

By using $\lambda fus3$ DNA as template, we have found that S7 specifically inhibits the synthesis of S7 *in vitro* but does not appear to inhibit the expression of the S12 gene, which precedes the S7 gene in the *str* operon. We have also shown that specific *in vitro* inhibitory effects by S7 on S7 synthesis take place even when template DNA that lacks the promoter as well as the beginning part of the S12 structural gene is used (Fig. 1; unpublished results). Thus, the mRNA target site where S7 acts probably involves the beginning of the S7 message, including the intercistronic region between the S12 and S7 genes. The

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

* Present address: Kovler Laboratories of Viral Oncology, University of Chicago, Chicago, IL 60637.

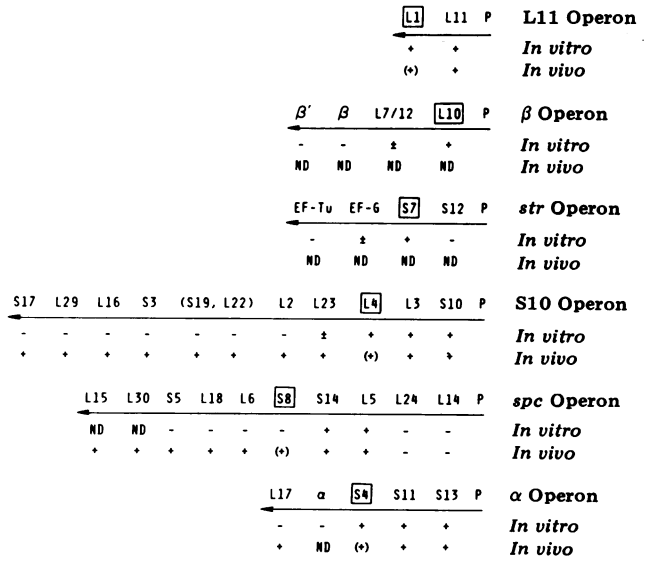


FIG. 1. Summary of experiments on identification of translational feedback-repressor r proteins. Genes in r-protein operons are indicated by the name of the protein product. The direction of transcription is shown by an arrow from the promoter (P). The boxed gene indicates the repressor r protein. *In vitro* indicates effects of purified repressor r proteins on DNA-directed *in vitro* protein synthesis; *in vivo* indicates effects of overproduction of repressor r proteins on r-protein synthesis rates *in vivo*. +, (+), ±, −, and ND indicate inhibition by the repressor protein of that operon; not determined, but believed to be inhibited; slight inhibition; no inhibition; and not determined, respectively. Identification of L4 as a repressor by *in vivo* experiments was by Zengel *et al.* (5). All other experiments were done in this laboratory (2–4; unpublished results). Inhibition of *in vitro* synthesis of L10 and L7/12 has also been observed by others (6, 7). It was previously thought that S8 inhibited *in vitro* synthesis of L24 (2); it has now been shown that the inhibited protein is S14 (which comigrated with L24 in the gel system used) and that S8 does not inhibit L24 synthesis (unpublished results). It is not known how the synthesis of L14 and L24 is regulated.

nucleotide sequence of this region has been determined, together with the entire mRNA sequence preceding this region (which includes the S12 coding sequence) (8, 9). We therefore compared (with the aid of a computer program; 10) the mRNA sequence with the nucleotide sequence of the S7 binding site on 16S rRNA.

Ribosomal protein S7 has been shown to bind specifically to 16S rRNA in the absence of any other proteins under conditions of *in vitro* ribosome reconstitution (11, 12). Brimacombe and his coworkers have recently demonstrated that S7 can be crosslinked to U at position 1240 of 16S rRNA by irradiation of ribosomes with UV light (13, 14). Similarly, Ehresmann and his coworkers have presented evidence that S7 can be crosslinked to a site in the region 1261–1266 (A-C-C-U-C-G in Fig. 2*A*) (16). Therefore, we first compared the nucleotide sequence of this region (1220–1340; see refs. 15, 17, 18) with the nucleotide sequence of a segment of S12–S7 mRNA that includes the intercistronic region (between 480 and 630 in the numbering system given in Fig. 2 of ref. 9). We have found that there are four sets of five- or six-nucleotide-long sequences common to the 16S rRNA and mRNA regions (A-A-U-G-G, U-C-G-U-A-G, U-U-G-G-A, and U-C-C-A-U-G; these are boxed in Fig. 2) and that the mRNA sequence can take a secondary structure similar to that proposed for the 16S rRNA in this region (15), as shown in Fig. 2.

The four common sequences make two sets of base-paired "stem-root" structures. One of these includes the $U_{1240}$ of 16S rRNA, which is crosslinked to S7 in the ribosomes, and the other includes the initiation codon AUG (at 601 in Fig. 2*B*) of the S7 gene on mRNA. Occurrence of such homology by chance is highly unlikely.¶ In addition, the two helical structures in 16S rRNA relevant to the homology (1241–1247/1290–1296 and 1308–1314/1323–1329) have been proved by Woese, Noller, and their coworkers, based on extensive comparison of sequence data from various bacterial species (15). Thus, we conclude that the observed homology, including these two sets of stem-root structures, is significant in relation to S7-rRNA and S7-mRNA interactions.

In addition to the above stem-root structures, we also found another stem-root structure (G-U-G-N-A-G-C paired with G-C-U, which is located near the A-A-U-G-G, 1238–1242, of 16S rRNA). This homology is perhaps less striking than the first two stem-root structures. However, the helical structure, 946–955/1225–1235, in 16S rRNA has also been proven, based on comparative phylogenetic sequence data (15), and all three stem-root structures exist close to each other in 16S rRNA. Therefore, it is possible that the observed homology involving the third stem-root structure could also be related to the interaction of S7 with mRNA.

The postulated secondary structures in the mRNA (as shown in Fig. 2*B*) are all unstable or at least weaker than the corresponding structures in 16S rRNA (see ref. 19). This feature might be related to the (postulated) stronger affinity of S7 for rRNA than for mRNA. In addition, the mRNA is normally translated, and stable secondary structures involving the initiation site might hinder efficient translation initiation. We postulate that the secondary structure that effectively blocks the initiator AUG is stabilized by the binding of S7.

## S4 sites

Ribosomal protein S4 regulates the synthesis of r proteins encoded in the α operon by translational feedback inhibition (2, 3; Fig. 1). This inhibition probably takes place by binding of S4 at a beginning region of the mRNA transcribed from the promoter for this operon. We found that S4 inhibited the synthesis of S4 *in vitro* when the template DNA contained the intact promoter but failed to do so when the DNA template used lacked the promoter and a beginning part of the structural gene for S13 (unpublished results). Thus, we infer that the target for S4 is located within the first 202 nucleotides of the polycistronic mRNA transcribed from the promoter for the α operon (from 557 to 758, according to the numbering system given in ref. 20 and Fig. 3).

Like that of S7, specific binding of S4 to 16S rRNA was originally shown by incubating S4 with 16S rRNA under reconstitution conditions in the absence of other r proteins (11, 12); since then, there have been many studies by many investigators, and various approaches have been used (reviewed in ref. 21). Although RNase protection analysis of the S4–16S rRNA complex suggests that a wide segment of the 5′-proximal region of 16S rRNA (positions 10–580) is related to the S4

¶ Assume that the four oligonucleotides (A-A-U-G-G, U-C-G-U-A-G, U-U-G-G-A, and U-C-C-A-U-G) are recognition signals that are arranged in this order and separated from each other at distances such that (*i*) there are at least six nucleotides between the first and the second and between the third and the fourth (to make reasonably stable hairpin structures) and (*ii*) there are at least four nucleotides between the second and the third. The probability of occurrence of such recognition oligonucleotides within a 200-nucleotide-long region (with random base composition) has been calculated by Millard Susman to be $1.56 \times 10^{-4}$, which gives a rough idea of the likelihood of occurrence of the homology in both the S7 binding region of 16S rRNA (at some specific place such as U-1240) and a mRNA segment (about 150–200 nucleotides long), where S7 very likely acts as a translational repressor.

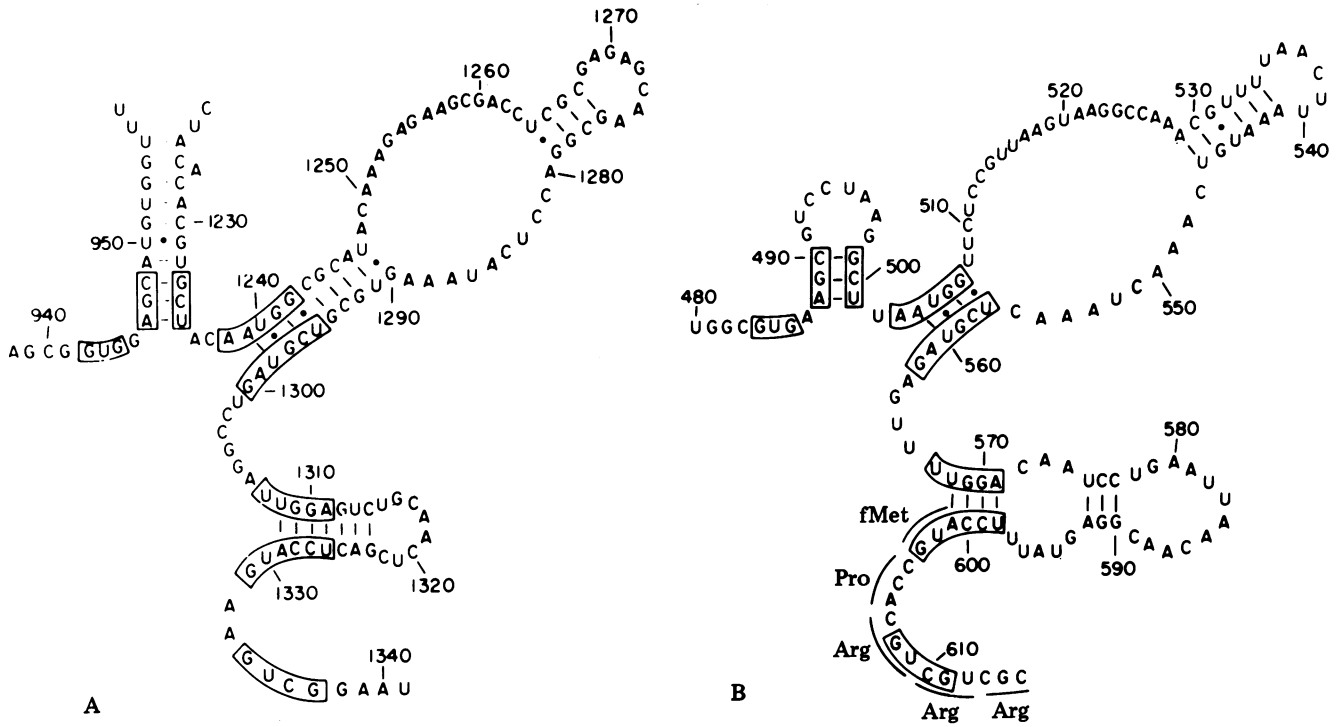


FIG. 2. Model of secondary structure of S7 binding sites on 16S rRNA (*A*) and on mRNA (*B*). Boxed sequences indicate homology. The structure shown in *A* is taken from ref. 15. In *B*, the S12 coding region ends at 501 and the S7 coding region begins at 604.

binding sites (21), the observed protection of such a large area is undoubtedly due to the presence of secondary and tertiary structures of rRNA and S4 may be directly interacting with only small specific segments of 16S rRNA. Two different kinds of experiments have suggested locations for these specific segments. First, three oligonucleotide fragments crosslinked to S4 were isolated after UV irradiation of the S4–16S rRNA complex (22). These are 303–324, 423–505, and 528–557, according to the established 16S rRNA sequence (15, 17, 18) and the numbering system in ref. 15 (22). Second, electron microscopic examination of formaldehyde-fixed S4–16S rRNA complexes suggests three widely separated specific binding sites (23). Based on the Woese–Noller secondary structure model of 16S rRNA (15), the results can be interpreted as suggesting that, under the spreading conditions for electronmicroscopy, the first and second binding sites are responsible for stabilizing the 27–37/547–556 stem structure and the second and third sites are responsible for stabilizing the 563–569/879–885 stem structure. Therefore, we compared the nucleotide sequences in these possible S4 binding regions on 16S rRNA with the first 202 nucleotides of mRNA. Fig. 3 shows the significant homology we found.

The region of 16S rRNA shown in Fig. 3*A* represents the S4 binding sites that include one of the three oligonucleotides crosslinked to S4 by UV irradiation (528–557), as well as a portion of the second (493–505), and the first and second sites (the 27–37/547–556 stem structure) suggested by the electron microscopic experiments. The homologous structure consists of two sets of stem-root structures: One set includes an identical four-base-long segment (C-C-G-U/A-C-G-G) in the stem and an identical five-base-long single-stranded segment (A-G-C-A-G) close to this stem and the other is a stem-root structure that is similar but not identical in 16S rRNA and mRNA. The first common stem structure in 16S rRNA is a part of a helical structure, 511–516/535–540, that is phylogenetically conserved and has been proved by Noller, Woese, and their coworkers (18); the corresponding helical structure in mRNA (Fig. 3*B*) is very close to (or includes a part of) the initiation codon (GUG) for the first structural gene (the gene for S13). If the r protein S4 interacts with the two stem-root structures in mRNA as it appears to do in S4-16S rRNA, the function of S4 as a translational repressor could be explained on the basis of stabilization of the helical structures by S4 binding. From these considerations, it appears likely that the structural homology shown in Fig. 3 is significant. Again, we note that one of the two helical structures of the mRNA discussed above, 624–627/644–647, is unstable and, in the absence of free S4, would not hinder the initiation of translation of S13 mRNA, which in turn would prevent formation of the more stable helical structure, 603–609/653–659.

## Discussion

The striking homology found between the r-protein binding regions of rRNA and the presumptive regulatory regions of mRNA described above gives further support to the model of translational feedback regulation of r-protein synthesis. Specifically, this homology supports the concept that regulation by repressor r proteins is based on competition between rRNA and mRNA for them and that this competition for r-protein binding involves the same structural features of r proteins. We suspect that this concept applies not only to the S7 and S4 sites discussed here but also to other repressor r proteins. In fact, our preliminary analyses suggest that this is also the case for L1 and S8 (unpublished results). Thus, the pertinent nucleotide sequences of rRNA and mRNA (and the structural features of the repressor r proteins) could have evolved to optimize the efficiency of the regulatory function relative to the ribosome assembly reaction.

A certain aspect of the regulatory features in the system described here resembles that of the phage T4 gene 32 system; the gene 32 protein normally binds to single-stranded DNA and, when produced in excess, inhibits the translation of its own mRNA (24). However, the gene 32 protein binds to any kind of single-stranded DNA and shows specificity only in the in-

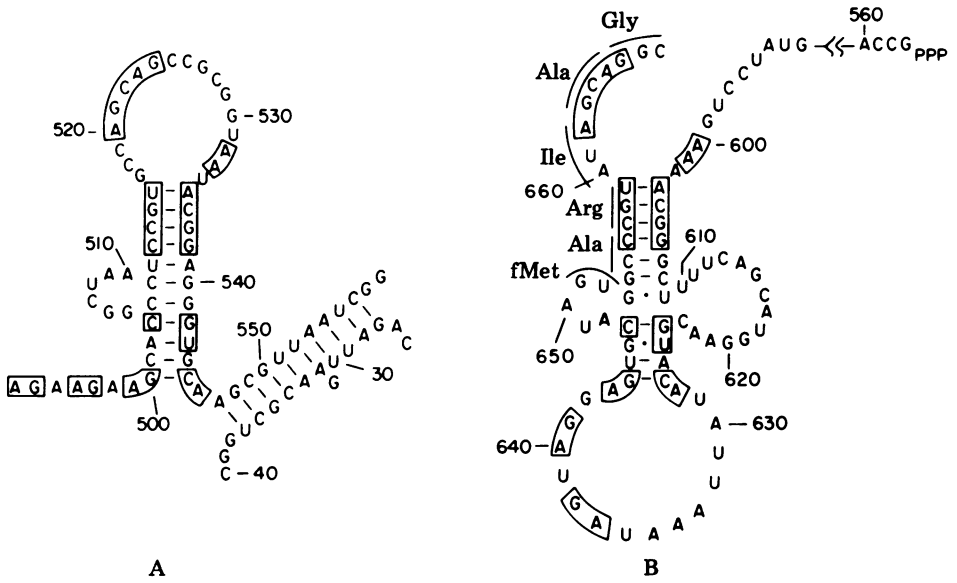


FIG. 3. Model of secondary structure of S4 binding sites on 16S rRNA (*A*) and on mRNA (*B*). Boxed sequences indicate homologies. The structure shown in *A* is taken from ref. 15. In *B*, the S13 coding region begins at 654.

teraction with mRNA, and it has been suggested that the specific interaction of the gene 32 protein with its own mRNA involves a structural element that is not essential for the DNA binding function of the protein (24). In contrast, for the repressor r proteins, the same structural features appear to be responsible for the specific interaction with both rRNA and their own mRNA. The only known mutations affecting the repressor function of r proteins are those isolated as suppressors of streptomycin-dependent mutations and involve structural alterations of r protein S4. Analysis of a few such S4 mutations suggests that the mutations affect the ability of the protein both to bind to 16S rRNA and to function as a repressor (25; unpublished results).

The importance of the secondary structure of mRNA in determining the efficiency of initiation of translation has been pointed out by several investigators (e.g., see ref. 26). It has been suggested, for example, that the mechanism of inhibition of the translation of RNA phage R17 replicase gene by phage coat protein involves stabilization of a preformed RNA helical structure that sequesters the replicase initiation codon AUG (and the Shine–Dalgarno sequence for this initiation site) (27). The mechanism of translational repression of r-protein synthesis suggested in this paper is similar in principle.

The model of the S4 binding site on the mRNA shown in Fig. 3*B* suggests that the nucleotide sequences in some protein coding regions might also be used in forming unique structures that are responsible for the binding of repressor r proteins. Thus, we suspect that there might have been constraints on the codon use (or on the use of amino acids or both) during the evolution of the present ribosome system and that this might be reflected in the pattern of codon use in these "regulatory" regions relative to the average codon use pattern seen in r-protein genes. We have indeed found that this appears to be the case. One such example is the Ile codon (660–662 in Fig. 3*B*) at the beginning of the S13 gene. It has previously been found that r-protein genes preferentially use codons that are recognized by the most abundant tRNA species (9, 28). Regarding Ile codons, 53 out of 54 sequenced Ile codons in r-protein genes are either AUU or AUC (9), both of which are recognized by the major isoleucine tRNA in *E. coli* (29). The codon AUA, which is recognized by a minor isoleucine tRNA, composing <5% of the total isoleucine tRNA (30), is used only once in the sequences determined so far (9)—i.e., at 660–662 in Fig. 3*B*. The use of the rare Ile codon at this position could be related to the common sequence A-G-C-A-G, which might be functionally important as a site involved in S4 binding.

Extensive studies have been carried out to identify r-protein binding sites on rRNA, and some information is now available regarding the approximate locations of the binding sites. However, exactly what kinds of structural features of these binding sites are responsible for the specificity of the interactions between rRNA and r proteins has not been determined (see ref. 21). The homology found thus far suggests that not only secondary structural features but also some of the primary sequences in both helical and single-stranded regions, as exemplified by Figs. 2*A* and 3*A*, are important for the specificity of RNA–protein interaction in the ribosome assembly. Such specific suggestions can be tested experimentally.

We thank many of our colleagues, especially Drs. J. J. Hopfield, Millard Susman, S. R. Fahnestock, and P. Olins, for useful discussion. This work was supported in part by the College of Agricultural and Life Sciences, University of Wisconsin-Madison, by Grant GM-20427 from the National Institutes of Health, and by Grant PCM79-10616 from the National Science Foundation. D.D. is supported by National Institutes of Health Postdoctoral Fellowship GMO 7553. This is paper no. 2456 from the Laboratory of Genetics, University of Wisconsin-Madison.

1. Fallon, A. M., Jinks, C. S., Strycharz, G. D. & Nomura, M. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 3411–3415.
2. Yates, J. L., Arfsten, A. E. & Nomura, M. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 1837–1841.
3. Dean, D. & Nomura, M. (1980) *Proc. Natl. Acad. Sci.* **77**, 3590–3594.
4. Yates, J. L. & Nomura, M. (1980) *Cell*, **21**, 517–522.
5. Zengel, J. M., Mueckle, D. & Lindahl, L. (1980) *Cell*, **21**, 523–535.
6. Fukuda, R. (1980) *Mol. Gen. Genet.* **178**, 483–486.
7. Brot, N., Caldwell, P. & Weissbach, H. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 2592–2595.
8. Post, L. E., Arfsten, A. E., Reusser, F. & Nomura, M. (1978) *Cell* **15**, 215–229.
9. Post, L. E. & Nomura, M. (1980) *J. Biol. Chem.* **255**, 4660–4666.
10. Korn, L. J., Queen, C. L. & Wegman, M. N. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 4401–4405.

11. Mizushima, S. & Nomura, M. (1970) *Nature (London)* **226,** 1214–1218.
12. Schaup, H. W., Green, M. & Kurland, C. G. (1970) *Mol. Gen. Genet.* **109,** 193–205.
13. Moller, K., Zwieb, C. & Brimacombe, R. (1979) *J. Mol. Biol.* **126,** 489–506.
14. Zwieb, C. & Brimacombe, R. (1979) *Nucleic Acids Res.* **6,** 1775–1790.
15. Woese, C. R., Magrum, L. J., Gupta, R., Siegel, R. B., Stahl, D. A., Kop, J., Crawford, N., Brosius, J., Gutell, R., Hogan, J. J. & Noller, H. F. (1980) *Nucleic Acids Res.* **8,** 2275–2293.
16. Ehresmann, B., Bachendorf, C., Ehresmann, C., Millon, R. & Ebel, J.-P. (1980) *Eur. J. Biochem.* **104,** 255–262.
17. Brosius, J., Palmer, M. L., Kennedy, P. J. & Noller, H. F. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 4801–4805.
18. Carbon, P., Ehresmann, C., Ehresmann, B. & Ebel, J.-P. (1978) *FEBS Lett.* **94,** 152–156.
19. Tinoco, I., Jr., Borer, P. N., Dengler, B., Levine, M. D., Uhlenbeck, O. C., Crothers, D. M. & Gralla, J. (1973) *Nature (London) New Biol.* **246,** 40–41.
20. Post, L. E., Arfsten, A. E., Davis, G. R. & Nomura, M. (1980) *J. Biol. Chem.* **255,** 4653–4659.
21. Zimmerman, R. A. (1980) in *Ribosomes: Structure, Function, and Genetics,* eds. Chambliss, G., Craven, G. R., Davies, J., Davis, K., Kahan, L. & Nomura, M. (Univ. Park, Baltimore), pp. 135–169.
22. Ehresmann, B., Backendorf, C., Ehresmann, C. & Ebel, J.-P. (1977) *FEBS Lett.* **78,** 261–266.
23. Cole, M. D., Beer, M., Koler, Th., Strycharz, W. A. & Nomura, M. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 270–274.
24. Lemaire, G., Gold, L. & Yarus, M. (1978) *J. Mol. Biol.* **126,** 73–90.
25. Olsson, M. O. & Isaksson, L. A. (1979) *Mol. Gen. Genet.* **169,** 271–278.
26. Steitz, J. A. (1979) in *Biological Regulation and Development,* ed. Goldberger, R. F. (Plenum, New York), Vol. 1, pp. 349–399.
27. Gralla, J., Steitz, J. A. & Crothers, D. M. (1974) *Nature (London)* **248,** 204–208.
28. Post, L. E., Strycharz, G. D., Nomura, M., Lewis, H. and Dennis, P. P. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 1697–1701.
29. Yarus, M. & Barrell, B. G. (1971) *Biochem. Biophys. Res. Commun.* **43,** 729–734.
30. Harada, F. & Nishimura, S. (1974) *Biochemistry* **13,** 300–307.