Hypothesis

MATURE 23 S rRNA OF PROKARYOTES APPEARS HOMOLOGOUS WITH THE PRECURSOR OF 25–28 rRNA OF EUKARYOTES

Comments on the evolution of 23–28 rRNA

Robert A. COX and John M. KELLY
National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, England

Received 20 May 1981

1. Introduction

Although ribosomes, whatever their origin, fulfil the same role in mRNA-directed protein biosynthesis, there is an apparent evolutionary discontinuity separating prokaryotic ribosomes and eukaryotic cytoplasmic ribosomes. Whereas the larger subparticle of cytoplasmic ribosomes comprises ~42 proteins and 3 RNA species known as 5 S, 5.8 S and L-rRNA (25–28 S rRNA), the prokaryotic larger subparticle comprises ~34 proteins and only 2 RNA species, 5 S and 23 S rRNA [1–3]. The absence of 5.8 S rRNA from prokaryotes has provoked speculation of the functions of 5 S rRNA in prokaryotes and of 5 S rRNA and 5.8 S rRNA in eukaryotes. It has been suggested [4,5] that 5.8 S rRNA is the analogue of prokaryotic 5 S rRNA; and also that the roles of 5 S rRNA in prokaryotes have become specialised in eukaryotes so that 2 species (namely 5 S rRNA and 5.8 S rRNA) are needed to do the same job.

The appearance of 5.8 S rRNA as a separate species and the variations in the mass and nucleotide composition of L-rRNA can both be explained in the light of several recent developments. We present evidence that there are extensive homologies between mature L-rRNA of prokaryotes (e.g., *Escherichia coli* L-rRNA) and the precursor L-rRNA (pre L-rRNA) of eukaryotes (e.g., *Xenopus laevis* pre L-rRNA). These homologies suggest that the apparent evolutionary discontinuity separating prokaryotes and eukaryotes has arisen as a result of mutations altering the processing of pre L-rRNA.

2. Sequence homologies between *Neurospora crassa* 5.8 S rRNA and the 5′-terminal sequences of *Escherichia coli* L-rRNA

The 5.8 S rRNA gene of cytoplasmic ribosomes lies upstream from the 5′-end of the L-rRNA gene [6,11] and 5.8 S rRNA appears as a separate entity only at a late stage in the processing of pre L-rRNA [12]. Nazar [13] has suggested that sequences near to the 5′-terminus of L-rRNA (23 S rRNA) of *E. coli* are homologous with trout 5.8 S rRNA. We confirm this notion since sequences homologous with *N. crassa* [14,15] and human 5.8 S rRNA [16] can be identified close to the 5′-terminus of *E. coli* L-rRNA. The sequences of 5.8 S rRNA that have been most conserved during evolution, and so form a characteristic feature of this rRNA species, lie within a stretch of ~20 nucleotides located at ~40–60 nucleotides from the 5′-end (see fig.1a). The homologous sequence is located within 50–70 nucleotides of the 5′-end of *E. coli* L-rRNA (see fig.1b).

3. Presence at the 3′-end of *Escherichia coli* L-rRNA of a binding site for 5.8 S rRNA-like sequences

Within the ribosome, cytoplasmic L-rRNA and 5.8 S rRNA form a specific complex. We have shown [15] that a binding site for 5.8 S rRNA is located in the region 30–80 nucleotides from the 3′-OH end of *N. crassa* L-rRNA (see fig.1a). This is likely to be a general model for eukaryotic 5.8 S rRNA–L-rRNA interactions, since complex formation between rabbit 5.8 S rRNA and an oligonucleotide fragment from the 3′-OH end of rabbit L-rRNA was demonstrated

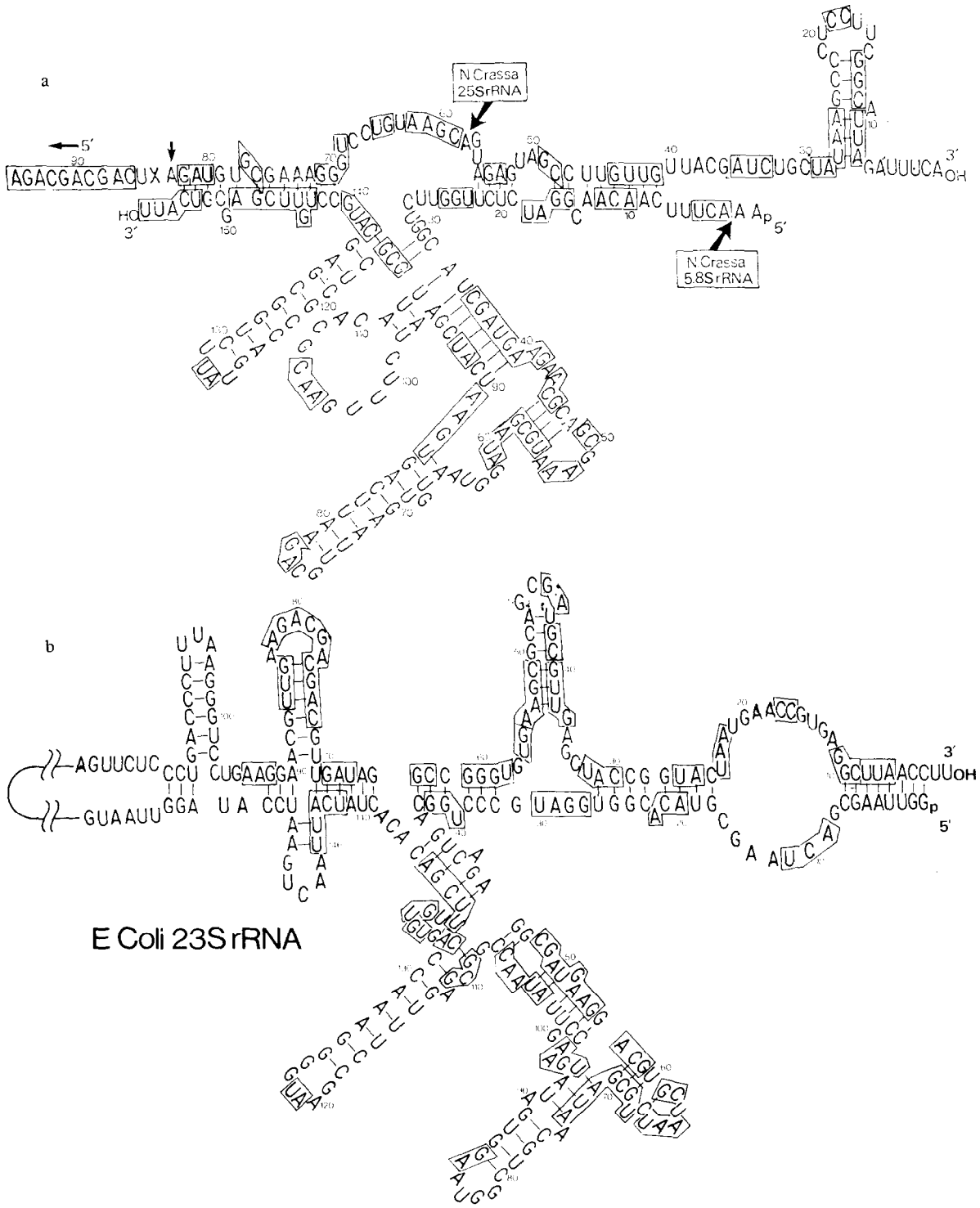

Fig.1. Possible binding sites for 5.8 S rRNA or 5.8 S rRNA-like sequences at the 3'-ends of *N. crassa* L-rRNA and *E. coli* L-rRNA. (a) Location of a binding site for *N. crassa* 5.8 S rRNA at the 3'-end of *N. crassa* L-rRNA [15]. Independent studies support the involvement of both the 3'-OH [17] and 5'-terminal [18] sequences of 5.8 S rRNA in binding to L-rRNA. (b) A possible interaction between 5.8 S rRNA-like sequences at the 5'-end of *E. coli* L-rRNA with a binding site at the 3'-end (based on [19]). In both (a) and (b) sequences common to *N. crassa* and *E. coli* rRNA are boxed. Residues of 5.8 S rRNA and 5.8 rRNA-like sequences are numbered from the 5'-end, whereas the residues at the binding sites are numbered from the 3'-end.

(A. J. Colbeck, J. M. K., R. A. C., unpublished). Further, we have shown that hamster 5.8 S rRNA, but not *N. crassa* 5.8 S rRNA, will interact with this 3′-OH fragment from rabbit L-rRNA. Sequence data for 5.8 S rRNA and for the 3′-OH terminal sequences of yeast [11] and *X. laevis* [10] L-rRNA suggest that a complex similar to that illustrated in fig.1a can also be formed in these two species.

By analogy, we sought and found a binding site close to the 3′-end of *E. coli* L-rRNA for the 5.8 S rRNA -like sequences (fig.1b) that are located near to the 5′-end. The extensive homologies between the two systems (cf. fig.1a,b) extend to sequences close to the 3′-ends of *N. crassa* and *E. coli* L-rRNA. Our hypothesis leads to the prediction that within the ribosome the 5′- and 3′-terminal regions of *E. coli* L-rRNA are in close proximity forming a (5.8 S rRNA-like–L-rRNA) complex. We infer that because the (5.8 S rRNA-like–L-rRNA) complex proposed for *E. coli* is homologous with the 5.8 S rRNA–L-rRNA *N. crassa* complex, both complexes are also homologous in their functions. The finding by Wrede and Erdmann [5] that certain *E. coli* L-subparticle proteins can bind yeast 5.8 S rRNA provides circumstantial evidence that these proteins interact with 5.8 S rRNA-like sequences in *E. coli* and that the binding sites of the proteins are sufficiently conserved to interact with yeast 5.8 S rRNA.

4. The region of *Escherichia coli* L-rRNA (nucleotides 157–290) is homologous to sequences at the 5′-end of yeast and *Xenopus laevis* L-rRNA

The hundred or so nucleotides at the 5′-end of yeast [20] and *X. laevis* [21] L-rRNA share a high degree of homology (see fig.2). The sequences common to both species have their counterpart in *E. coli* L-rRNA, but in the region of nucleotides 157–290 that follow immediately after the 5.8 S rRNA-like sequences (see fig.2).

5. Eukaryotic L-rRNA have additional sequences that have diverged rapidly during evolution

The mass of all bacterial L-rRNA species so far examined lies close to 1.05×10^6 M_r. In contrast, the mass of eukaryotic L-rRNA is greater by at least 0.2×10^6 M_r and is species specific ranging from 1.3×10^6 M_r in lower eukaryotes to 1.7×10^6 M_r in mammals [1–3]. However, bacterial (e.g., *Bacillus stearothermophilus*) L-rRNA and cytoplasmic (e.g., rabbit or *X. laevis*) L-rRNA appear to have extensive features of secondary structure in common [22]. This observation and also the finding that distantly related cytoplasmic L-rRNA species share common sequences [23,24] led to the suggestion that L-rRNA comprises

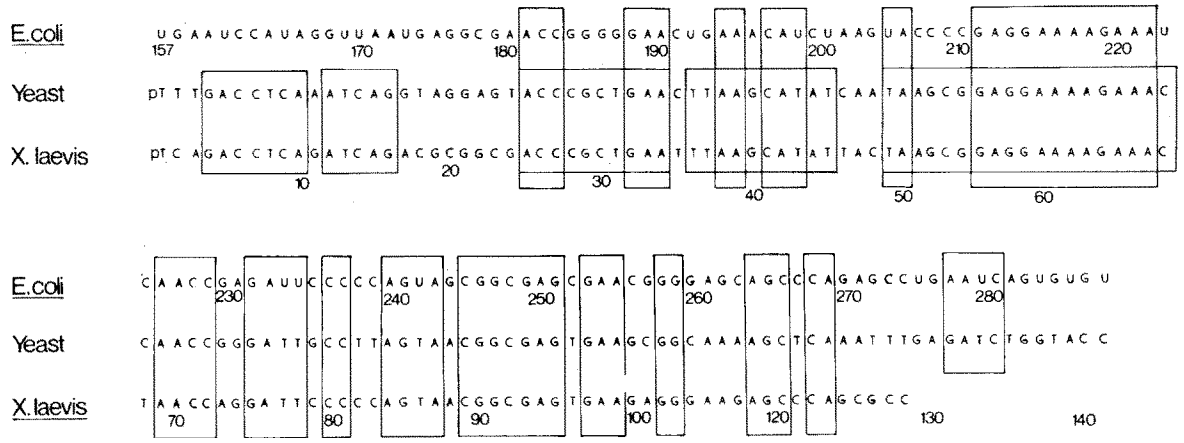

Fig.2. Sequences common to the 5′-ends of yeast and *X. laevis* L-rRNA and to residues 157–296 of *E. coli* L-rRNA. The boxes indicate regions of homology. The sequences of *E. coli* refer to L-rRNA [19], whereas the sequences of the non-coding strand of the L-rRNA gene are presented for yeast [20] and *X. laevis* [21].

a core that has been conserved throughout evolution to which highly divergent sequences have been added [22,25–28]. The conserved core of cytoplasmic L-rRNA comprises at least 1000 nucleotides [27,29]. The divergent sequences range from large tracts (up to 1000 nucleotides) of mainly guanine and cytosine residues (~80% G + C) found in *X. laevis* and rabbit L-rRNA [22] to large tracts of mainly adenine and uracil residues (~24% G + C) found in *Drosophila melanogaster* L-rRNA [26]. The tracts of divergent and conserved sequences are intermingled [27,29]. We infer that those parts of cytoplasmic L-rRNA whose primary sequence has been conserved [27,29] give rise to secondary structures that are homologous with particular regions of native bacterial L-rRNA.

6. Bridging the evolutionary discontinuity

The organisation of *E. coli* L-rRNA and *X. laevis* pre L-rRNA are compared in fig.3 and serve to illustrate a scheme for the evolution of L-rRNA. The two species differ by the presence in *X. laevis* pre L-rRNA of sequences analogous to intervening sequences in eukaryotic genes. We propose that these features arose during evolution as the result of the insertion of additional sequences into L-rRNA genes and subsequent mutations at one or more processing sites leading to the incorporation of inserted sequences into mature L-rRNA and generation of a 5.8 S rRNA species. These divergent (non-conserved) sequences of cytoplasmic L-rRNA have no counterpart in bacterial L-rRNA, but they appear to resemble the spacer sequences of the rRNA gene cluster in their overall nucleotide composition. In *X. laevis* and rabbit L-rRNA the divergent sequences are G + C rich [22] and so are the spacer sequences [21] and in *Drosophila melanogaster* L-rRNA the divergent sequences are A + U rich [26] and apparently so are the spacer sequences [31].

In *X. laevis* pre L-rRNA, it appears as though 5.8 S rRNA and L-rRNA are separated by an intervening sequence (the so-called internal transcribed spacer-2 [ITS-2]). In this case, the ITS-2 sequence is excised, although the two L-rRNA fragments remain unligated. The divergent sequences within L-rRNA may be

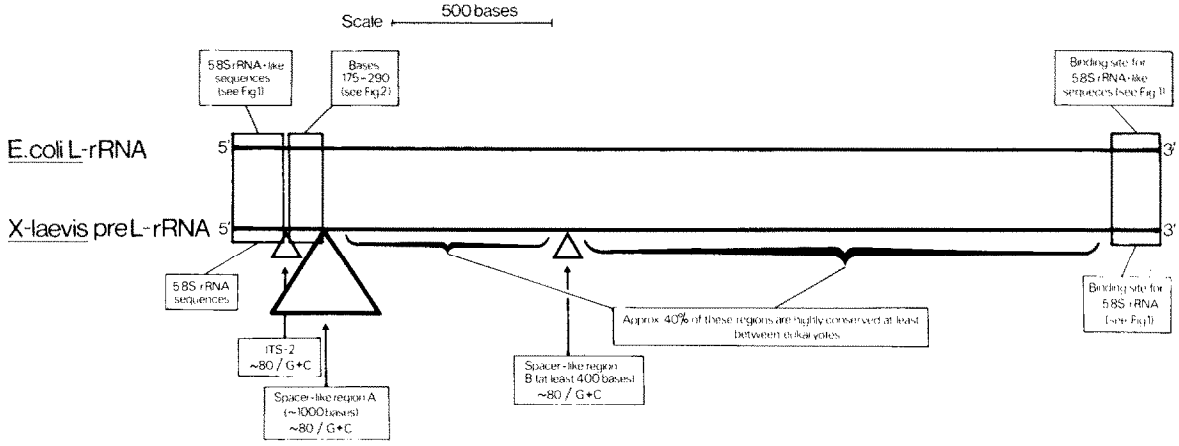

Fig.3. Comparison of the organisation of *E. coli* L-rRNA and *X. laevis* pre L-rRNA. Known regions of sequence homology (see fig.1,2) are boxed. Sequences amounting to ~0.4 × 10^6 M_r are tightly conserved among eukaryotes, as shown by analysis with restriction endonucleases and hybridisation with heterologous L-rRNA probes [27,29]. About 0.4 × 10^6 M_r of L-rRNA of rabbit, *X. laevis*, *E. coli* and *B. stearothermophilus* have features of secondary structure in common [22]. It is inferrred that those sequences common to eukaryotes are also present, but less stringently conserved, in bacterial L-rRNA. Tracts of eukaryotic L-rRNA have diverged very rapidly. In *X. laevis* these divergent sequences are G + C rich [22] and in this respect they resemble spacer sequences [21]. The positions of these tracts have been identified by electron microscopy using the technique of denaturation mapping [30]. These locations have been related to the restriction endonuclease maps of *X. laevis* L-rRNA gene [27]. The position of the internal transcribed spacer region-2 (ITS-2) that is removed on processing to generate a separate 5.8 S rRNA species is indicated. This region is G + C rich [21]. The approximate locations of major tracts of divergent sequences are also indicated and are designated spacer-like region A and spacer-like region B. They are considered as insertions into the L-rRNA, although they are not removed in processing. It is proposed that the evolutionary diversity apparent in eukaryotic L-rRNA [1,22,26–28] is generated by variations in the mass and nucleotide composition of the spacer-like insertions.

regarded as insertions that are no longer excised.

There is ample evidence for intervening sequences in L-rRNA. Chloroplast L-rRNA *Chlamydomonas rheinhardii* contains an intervening sequence that is removed on processing and yields normal L-rRNA as a result of a ligation [32]. In some cases, e.g., tobacco chloroplast L-rRNA, it appears ligation is not normal so that a 4.5 S RNA fragment homologous with the 3′-terminus of *E. coli* L-rRNA is generated after the intervening sequence is excised [33]. Intervening sequences are also known to be present in mitochondrial L-rRNA [34–36] and in cytoplasmic L-rRNA [37–42]. Therefore, the suggestion that mature L-rRNA has retained additonal sequences that resemble intervening sequences involves no new principles.

The addition of non-conserved (possibly spacer-like) sequences to a conserved L-rRNA core [22,25–27] accounts for the elements for diversity [1,25–28] that are found in this component of the ribosome. We also infer that as a consequence of the increase in the mass of L-rRNA additional binding sites for basic proteins may be created, thereby increasing the number of L-subparticle protein subunits. These additional protein subunits could be species specific in view of the divergent properties of these particular RNA sequences.

Our suggestion that the evolutionary discontinuity separating prokaryotic and eukaryotic ribosomes is attributable to modifications in the structure of rRNA genes that alter processing of pre-rRNA, also implies that the basic partial reactions of mRNA-directed protein biosynthesis, especially those involving 5 S rRNA and 5.8 S rRNA, are the same in both prokaryotes and eukaryotes. Thus, we propose that 5.8 S rRNA in eukaryotes and the 5.8 S rRNA-like sequences in prokaryotes are not only homologous in structure but may also fulfil the same functional role. We also suggest that 5 S rRNA of eukaryotes which are homologous in structure with 5 S rRNA of prokaryotes [1,2] are probably also homologous in function, contrary to previous suggestions [4,5].

While it is evident that the putative alterations in rRNA gene structure and pre-rRNA processing have become fixed, the advantages to the cell of a larger L-rRNA species and a separate 5.8 S rRNA entity are not yet apparent. The enlarged ribosome found in the cytoplasm of eukaryotes contrasts with the trend leading to the evolution of the mitochondrial ribosome that results in the miniribosome of amphibian and mammalian mitochondria [43].

References

[1] Cox, R. A. (1977) Prog. Biophys. Mol. Biol. 32, 193.
[2] Bielka, H. and Stahl, J. (1978) Int. Rev. Biochem. 18, 79.
[3] Wool, I. G. (1979) Annu. Rev. Biochem. 48, 719.
[4] Erdmann, V. A. (1976) Prog. Nucleic Acids Res. Mol. Biol. 18, 45.
[5] Wrede, P. and Erdmann, V. A. (1977) Proc. Natl. Acad. Sci. USA 74, 74.
[6] Pellegrini, M., Manning J. and Davidson, N. (1977) Cell 10, 213.
[7] Boseley, P. G., Tuyns, A. and Birnstiel, M. L. (1978) Nucleic Acids Res. 5, 1121.
[8] Free, S. J., Rice, P. W. and Metzenberg, R. L. (1979) J. Bacteriol. 137, 1219.
[9] Nath, K. and Bollon, A. P. (1978) Mol. Gen. Genet. 160, 235.
[10] Sollner-Webb, B. and Reeder, R. H. (1979) Cell 18, 485.
[11] Veldman, G. M., De Jonge Klootwijk, P., Leer, R. J. and Planta, R. J. (1980) Nucleic Acids Res. 8, 5179.
[12] Hadjiolov, A. A. and Nikolaev, N. (1976) Prog. Biophys. Mol. Biol. 31, 95.
[13] Nazar, R. N. (1980) FEBS Lett. 119, 212.
[14] Selker, E. and Yanofsky, C. (1979) Nucleic Acids Res. 6, 2561.
[15] Kelly, J. M. and Cox, R. A. (1981) Nucleic Acids Res. 9, 1111.
[16] Maden, B. E. H. and Robertson, J. S. (1974) J. Mol. Biol. 87, 227.
[17] Pace, N. R., Walker, T. A. and Schroeder, E. (1977) Biochemistry 16, 5321.
[18] Nazar, R. N. and Sitz, T. O. (1980) FEBS Lett. 115, 71.
[19] Brosius, J., Dull, T. J. and Noller, H. F. (1980) Proc. Natl. Acad. Sci. USA 77, 201.
[20] Veldman, G. M., De Jonge Klootwijk, P. and Planta, R. J. (1981) Biochem. Soc. Trans. 9, 170P.
[21] Hall, L. M. C. and Maden, B. E. H. (1980) Nucleic Acids Res. 8, 5993.
[22] Cox, R. A., Godwin, E. A. and Huvos, P. (1973) Isr. J. Chem. 11, 407.
[23] Sinclair, J. H. and Brown, D. D. (1971) Biochemistry 10, 2761.
[24] Birnstiel, M. L. and Grunstein, M. (1972) FEBS Symp. 23, 349.
[25] Godwin, E. A., Cox, R. A. and Huvos, P. (1974) Acta Biol. Med. Ger. 33, 733.
[26] Cox, R. A., Godwin, E. A. and Hastings, J. R. B. (1976) Biochem. J. 155, 465.
[27] Cox, R. A. and Thompson, R. D. (1980) Biochem. J. 187, 75.
[28] Cammarano, P., Londei, P. and Mazzei, F. (1980) Biochem. J. 189, 313.
[29] Gourse, R. L. and Gerbi, S. A. (1980) J. Mol. Biol. 140, 321.
[30] Wellauer, P. and Dawid, I. (1974) J. Mol. Biol. 89, 379.
[31] Jordan, B. R., Latil-Damotte, M. and Jourdan, R. (1980) FEBS Lett. 117, 227.
[32] Rochaix, J. D. and Malnoe, P. (1978) Cell 15, 661.
[33] MacKay, R. M. (1981) FEBS Lett. 123, 17.

[34] Bos, J. L., Heyting, G., Borst, P., Arnberg, A. C. and Van Bruggen, E. F. J. (1978) Nature 275, 275.
[35] Heckman, J. E. and RajBhandary, U. L. (1979) Cell 17, 583.
[36] Lazarus, C. M., Lunsdorf, H., Hahn, U., Stepien P. P. and Kuntzel, H. (1980) Mol. Gen. Genet. 177, 389.
[37] Glover, D. M. and Hogness, D. S. (1977) Cell 10, 167.
[38] Wellauer, P. K. and Dawid, I. B. (1977) Cell 10, 193.
[39] Cech, T. R. and Rio, D. C. (1979) Proc. Natl. Acad. Sci. USA 76, 5051.
[40] Din, N., Engberg, J., Kaffenberger, W. and Eckert, W. A. (1979) Cell 18, 525.
[41] Wild, M. A. and Gall, J. G. (1979) Cell 16, 565.
[42] Gubler, U., Wyler, T. and Braun, R. (1979) FEBS Lett. 100, 347.
[43] Chua, N. H. and Luck, D. (1974) in: Ribosomes (Nomura, M. et al. eds) p. 519, Cold Spring Harbor Labs. New York.