# STRUCTURAL STUDY OF RIBOSOMAL 23 S RNA FROM *ESCHERICHIA COLI*

C. BRANLANT, A. KROL, M. A. MACHATT and J. P. EBEL
*Institut de Biologie Moléculaire et Cellulaire du CNRS, 15, rue René Descartes, 67084 Strasbourg Cédex, France*

Received 20 August 1979

## 1. Introduction

Here we report the sequence of 80% of the ribosomal 23 S RNA of *Escherichia coli* and describe the cistron heterogeneities and base modifications detected in this sequence.

RNA sequencing techniques have been tremendously improved recently [1–3]. Nevertheless, even with these new tools, it is extremely difficult to determine the complete nucleotide sequence of a molecule as large as the ribosomal 23 S RNA from *E. coli* (3000 nucleotides). The difficulty lies in obtaining overlapping fragments encompassing the whole molecule. Obviously, the primary structure of this molecule can be obtained by sequencing one of its genes, as has been achieved by Brosius et al. (personal communication). But a structural study of the RNA itself is very useful to provide information on cistron heterogeneities and on base modifications, and data necessary for the study of the secondary structure of the molecule. For these reasons we have studied a large number of 23 S RNA fragments, comprising 80% of the molecule. As explained above, this approach has reached a limit since it is very difficult to get fragments from the remaining 20%.

## 2. Materials and methods

### 2.1. *Preparation of RNA fragments*

23 S RNA from 50 S subunits, intact or treated with LiCl, was prepared and digested in the conditions of [4]. Two new conditions of hydrolysis were improved in the case of 50 S subunits: they were incubated 20 min at 45°C in 10 mM $MgCl_2$, 10 mM Tris–HCl to pH 7.5 buffer and digested either 4 h at 0°C with *Naja oxiana* venom ribonuclease [5] or 30 min at 0°C with pancreatic ribonuclease at an enzyme/substrate ratio of 1/1000 (w/w). The digestion was stopped by phenol extraction in the presence of SDS. The resulting fragments were purified by electrophoresis on polyacrylamide slab gels, and were eluted according to [1].

### 2.2. *Labelling of the RNA fragments*

Labelling at the 5′-end was performed as in [6], with [$\gamma$-$^{32}$P]ATP and T4 polynucleotide kinase. Fragments resulting from venom ribonuclease digestion were dephosphorylated in the conditions of [7]. The 3′-ends were labelled overnight in the reaction mixture of [8]. Three units of ligase and 25 $\mu$Ci [5′-$^{32}$P]-pCp were used for each assay.

### 2.3. *Sequence analysis of the RNA fragments*

Partial digestions of the labelled fragments, with T1, U2, pancreatic and *Phy* I RNases, were carried out as in [6]. Instead of the partial alkaline digestion, we used a digestion in boiling bi-distilled water. Two times of incubation were used for each sample (15 and 30 min) and the two digests were mixed.

To fractionate the digests we used the classical one-dimensional fractionation on thin polyacrylamide gel slabs (900 × 300 × 0.5 $mm^3$) in Tris–borate buffer [9] and a two-dimensional system of electrophoresis similar to that in [3]. RNA digested in boiling water was first fractionated by electrophoresis on an 8% polyacrylamide slab gel (300 × 400 × 0.5 $mm^3$), made up with 25 mM citric acid, 8 M urea, pH 3.5 buffer. The buffer in the tank was 25 mM Tris–citrate (pH 3.5); electrophoresis was performed overnight at 400 V. The strip of gel containing the fractionated fragments was then included at the top of a 20%

polyacrylamide slab gel (300 × 900 × 0.5 $mm^3$) containing Tris–borate buffer. Electrophoresis was carried out for 48 h at 1200 V.

### 3. Results and discussion

50 S ribosomal subunits from *E. coli,* or the 23 S RNA extracted from these subunits, were digested in various conditions described [4] and herein. The resulting fragments were purified by electrophoresis on polyacrylamide slab gels. Those resulting from T1 or pancreatic ribonuclease digestions had free 5′-ends and were directly labelled with [$\gamma$-$^{32}$P]ATP and T4 polynucleotide kinase. Those resulting from digestion with *Naja oxiana* venom enzyme carried a phosphate at their 5′-end. They were either labelled at this end after dephosphorylation with alkaline phosphatase or labelled at the 3′-end using [5′-$^{32}$P]pCp and T4 RNA ligase.

Each labelled fragment was studied in two ways:

(i) With a technique adapted from that in [1] and [2];

(ii) With a two-dimensional electrophoresis technique similar to that in [3].

Figure 1 displays the autoradiographs of the gels obtained with both methods for one of the fragments studied. The positions of adenines, guanines and pyrimidines were deduced from the gel made according to [1] and the uridines and cytidines were distinguished by the two-dimensional electrophoresis.

A large number of fragments was sequenced in this way. The identification of the characteristic T1 RNase digestion products they contained allowed them to be located in the different sections of 23 S RNA characterized [4]. The results obtained are given in fig.2.

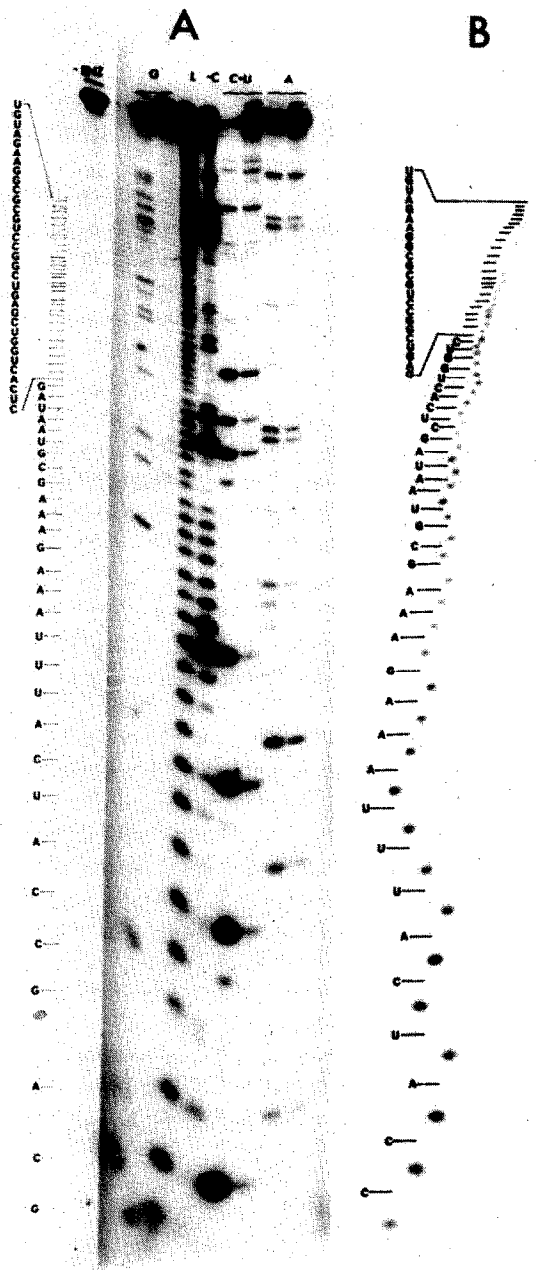
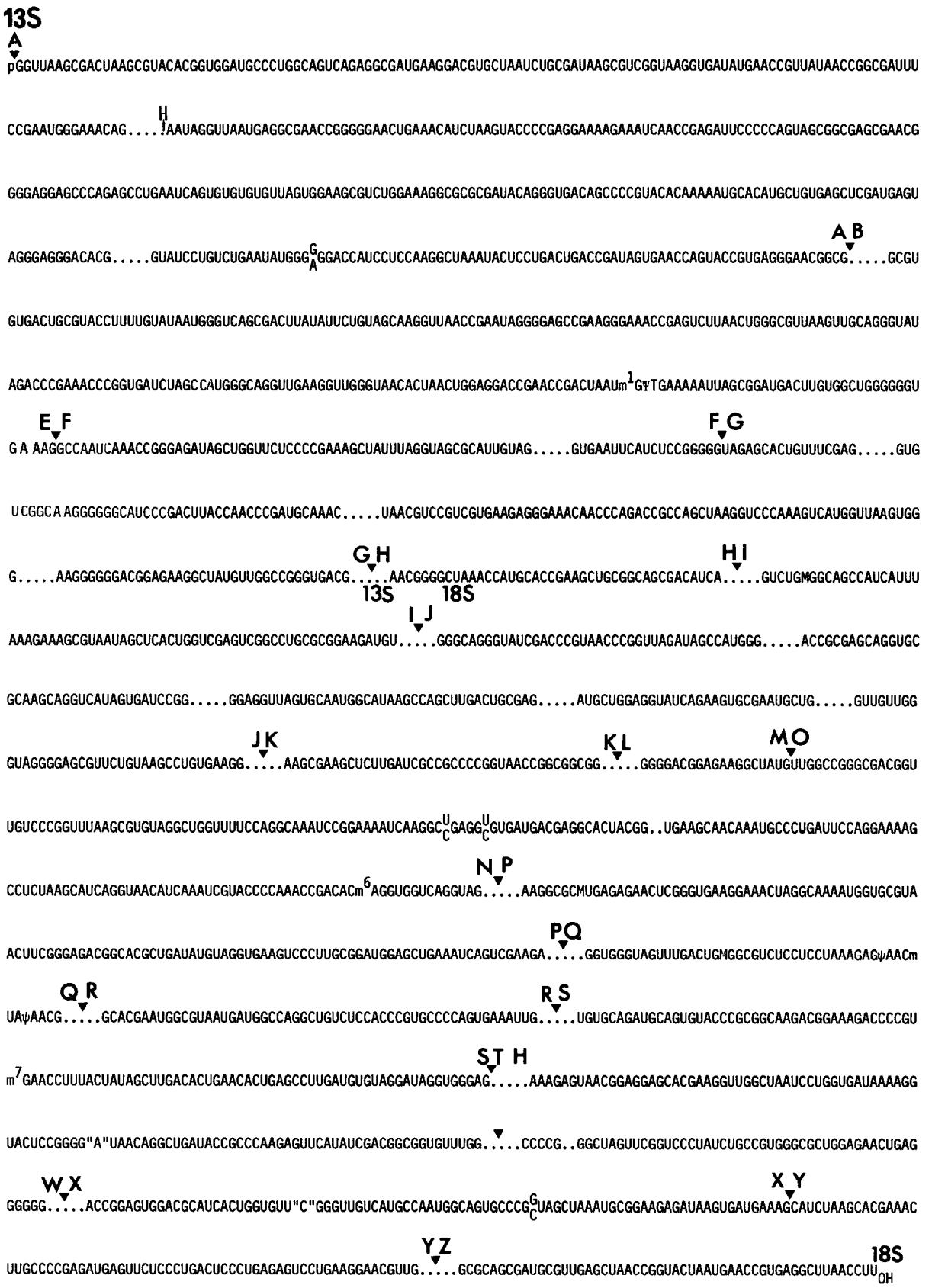



Fig.1. Autoradiograph of the gels obtained on analysing a RNA fragment: (a) with a technique similar to that in [1]; (b) with the technique of two-dimensional electrophoresis on polyacrylamide gel.

Fig.2. Nucleotide sequence of 80% of the molecule of 23 S RNA of *Escherichia coli.* The various sections are indicated by capital letters. The order of all the sections has not been verified and some inversions may still exist. Positions where 2 or 3 nucleotides are given correspond to heterogeneities. Those that were previously encountered in sequences still not verified by this technique are indicated by the letter H. Unidentified modified nucleotides are represented by M. Hyphens between nucleotides have been omitted for reason of space.


13S
A
pGGUUAAGCGACUAAGCGUACACGGUGGAUGCCCUGGCAGUCAGAGGCGAUGAAGGACGUGCUAAUCUGCGAUAAGCGUCGGUAAGGUGAUAUGAACCGUUAUAACCGGCGAUUU
H
CCGAAUGGGAAACAG....IAAUAGGUUAAUGAGGCGAACCGGGGGAACUGAAACAUCUAAGUACCCCGAGGAAAAGAAAUCAACCGAGAUUCCCCCAGUAGCGGCGAGCGAACG
GGGAGGAGCCCAGAGCCUGAAUCAGUGUGUGUGUUAGUGGAAGCGUCUGGAAAGGCGCGCGAUACAGGGUGACAGCCCCGUACACAAAAAUGCACAUGCUGUGAGCUCGAUGAGU
A B
AGGGAGGGACACG.....GUAUCCUGUCUGAAUAUGGGGAGGACCAUCCUCCAAGGCUAAAUACUCCUGACUGACCGAUAGUGAACCAGUACCGUGAGGGAACGGCG.....GCGU
GUGACUGCGUACCUUUUGUAUAAUGGGUCAGCGACUUAUAUUCUGUAGCAAGGUUAACCGAAUAGGGGAGCCGAAGGGAAACCGAGUCUUAACUGGGCGUUAAGUUGCAGGGUAU
AGACCCGAAACCCGGUGAUCUAGCCAUGGGCAGGUUGAAGGUUGGGUAACACUAACUGGAGGACCGAACCGACUAAUm1GΨTGAAAAAUUAGCGGAUGACUUGUGGCUGGGGGGU
E F
F G
GA AAGGCCAAUCAAACCGGGAGAUAGCUGGUUCUCCCCGAAAGCUAUUUAGGUAGCGCAUUGUAG.....GUGAAUUCAUCUCCGGGGGUAGAGCACUGUUUCGAG.....GUG
UCGGCA AGGGGGGCAUCCCGACUUACCAACCCGAUGCAAAC.....UAACGUCCGUCGUGAAGAGGGAAACAACCCAGACCGCCAGCUAAGGUCCCAAAGUCAUGGUUAAGUGG
G H
H I
G.....AAGGGGGGACGGAGAAGGCUAUGUUGGCCGGGUGACG.....AACGGGGCUAAACCAUGCACCGAAGCUGCGGCAGCGACAUCA.....GUCUGMGGCAGCCAUCAUUU
13S
18S
I J
AAAGAAAGCGUAAUAGCUCACUGGUCGAGUCGGCCUGCGCGGAAGAUGU.....GGGCAGGGUAUCGACCCGUAACCCGGUUAGAUAGCCAUGGG.....ACCGCGAGCAGGUGC
GCAAGCAGGUCAUAGUGAUCCGG.....GGAGGUUAGUGCAAUGGCAUAAGCCAGCUUGACUGCGAG.....AUGCUGGAGGUAUCAGAAGUGCGAAUGCUG.....GUUGUUGG
J K
K L
M O
GUAGGGGAGCGUUCUGUAAGCCUGUGAAGG.....AAGCGAAGCUCUUGAUCGCCGCCCCGGUAACCGGCGGCGG.....GGGGACGGAGAAGGCUAUGUUGGCCGGGCGACGGU
UGUCCCGGUUUAAGCGUGUAGGCUGGUUUUCCAGGCAAAUCCGGAAAAUCAAGGCUCGAGGUCGUGAUGACGAGGCACUACGG..UGAAGCAACAAAUGCCCUGAUUCCAGGAAAAG
N P
CCUCUAAGCAUCAGGUAACAUCAAAUCGUACCCCAAACCGACACm6AGGUGGUCAGGUAG.....AAGGCGCMUGAGAGAACUCGGGUGAAGGAAACUAGGCAAAAUGGUGCGUA
P Q
ACUUCGGGAGACGGCACGCUGAUAUGUAGGUGAAGUCCCUUGCGGAUGGAGCUGAAAUCAGUCGAAGA.....GGUGGGUAGUUUGACUGMGGCGUCUCCUCCUAAAGAGψAACm
Q R
R S
UAψAACG.....GCACGAAUGGCGUAAUGAUGGCCAGGCUGUCUCCACCCGUGCCCCAGUGAAAUUG.....UGUGCAGAUGCAGUGUACCCGCGGCAAGACGGAAAGACCCCGU
S T H
m7GAACCUUUACUAUAGCUUGACACUGAACACUGAGCCUUGAUGUGUAGGAUAGGUGGGAG.....AAAGAGUAACGGAGGAGCACGAAGGUUGGCUAAUCCUGGUGAUAAAAGG
UACUCCGGGG"A"UAACAGGCUGAUACCGCCCAAGAGUUCAUAUCGACGGCGGUGUUUGG.....CCCCG..GGCUAGUUCGGUCCCUAUCUGCCGUGGGCGCUGGAGAACUGAG
W X
X Y
GGGGG.....ACCGGAGUGGACGCAUCACUGGUGUU"C"GGGUUGUCAUGCCAAUGGCAGUGCCCGGCUAGCUAAAUGCGGAAGAGAUAAGUGAUGAAAGCAUCUAAGCACGAAAC
Y Z
18S
UUGCCCCGAGAUGAGUUCUCCCUGACUCCCUGAGAGUCCUGAAGGAACGUUG.....GCGCAGCGAUGCGUUGAGCUAACCGGUACUAAUGAACCGUGAGGCUUAACCUUOH

These results make some corrections to data obtained using the fingerprinting technique. A very limited number of errors was observed in the sequences of the characteristic T1 RNase digestion products containing more than one uridine [10]. In addition to the two errors already corrected in oligonucleotides 7b and 16 from the 13 S fragment [6], we found errors in oligonucleotides 2b, 5b, 10b, 12 and 24 from the 18 S fragment.

Concerning the ordering of these characteristic T1 products [4], we also observed a limited number of inversions in sections B to D and in sections P to Q, and section N was previously positioned at the left hand of section O while it is actually located on its right. The order of all the sections could not be verified by the present results. An inversion of two oligonucleotides was observed in the published sequence of section S [11] and a larger number of inversions was observed in the ordering of the T1 digestion products of section A [12]. Finally, a typing mistake was made in the sequence given in our last paper [6]: a U was omitted at the 5′-end of subfragment 6B, though it is mentioned in the legend to this figure.

The sequence obtained for the first 108 nucleotides at the 5′-end of the molecule is in good agreement with that obtained [13] by sequencing regions of the two cistrons: *rrnD* and *rrnX*. Figure 2 indicates the 5 positions where sequence heterogeneities were detected in the course of this study. In each case we observed substitutions. In section O the two mutations are simultaneous and those in section X are such that three different sequences were observed:

G–C–A–C–U–G–C–C–C–G–G–U

G–C–A–C–U–G–C–C–C–C–G–U

G–C–A–G–U–G–C–C–G–G–G–U

Altogether in our study of 23 S RNA we have observed 9 heterogeneities [11,12]. The four others are located in sequences which have not been checked in the present study. This represents a very low percentage, 0.3%, for a molecule encoded by 7 cistrons [14].

We detected 13 modified nucleotides. Eight of them can be identified without ambiguity as methylated nucleotides or pseudouridines described [15]. They belong to the T1 digestion products U–m$^7$G–A–A–C–C–U–U–U–A–C–U–A–U–A–G, A–C–U–A–A–U–m$^1$G–$\psi$–T–G, $\psi$–A–A–C–mU–A–$\psi$–A–A–C–G and A–C–A–C–m$^6$A–G. An error was previously made in the positioning of the m$^7$G in the first oligonucleotide [11]. We found the following modified nucleotides in addition to the 8 already characterized: a modified adenine, described in section V [4], which is not cleaved by RNase U2; an abnormal cytidine, already described in section X [16], and which was cleaved by *Phy* I enzyme in the present study; a uridine located in section P in the T1 digestion product C–'U'–U–G and which has an unusual property in the two-dimensional electrophoresis; and finally two nucleotides with very similar behavior, that they are cleaved neither by enzyme nor by water and that they are both preceded by the sequence C–U–G and followed by the sequence G–G–C. They may be methylated on the ribose. It is noticeable that 9 of the modified nucleotides are located in the 3′-half of the molecule.

In the course of this study, we have accumulated a large amount of information that will be very useful for the study of the secondary structure of the molecule and that will be given in a more detailed paper. Here we point out only that the already-mentioned [17] possible base-pairings between the two ends of 23 S RNA and between the 3′-end of this RNA and 16 S RNA are compatible with the complete sequence (fig.3).

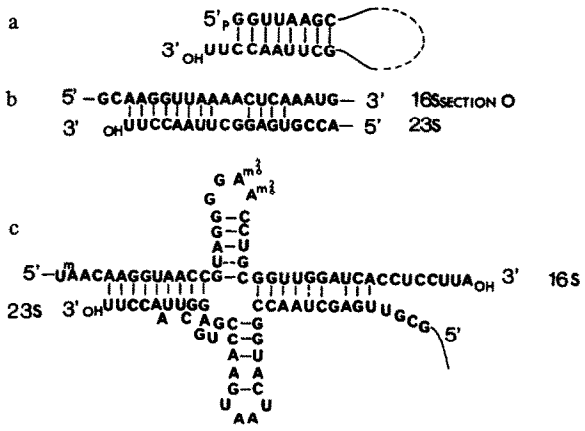



Fig.3. Possible base-pairings between: (a) the two ends of 23 S RNA; (b) the 3′-end of 23 S RNA and section O of 16 S RNA; (c) the 3′-end of 23 S RNA and the 3′-end of 16 S RNA.

At the end of this study, we learned that Brosius et al. have completely determined the sequence of the *rrnB* cistron. It will be interesting to compare their results with the present data.

## Acknowledgements

We are indebted to Dr G. Keith for his generous gift of T4 polynucleotide kinase, to Dr J. P. Bargetzi for providing us with *Phy* I RNase, to Dr S. K. Vassilenko who gave us the *Naja oxiana* venom RNase, and to the skilled technical assistance of Miss B. Muller. This work was financially supported by the Centre National de la Recherche Scientifique and the Délégation Générale à la Recherche Scientifique et Technique.

## References

[1] Donis-Keller, H., Maxam, A. and Gilbert, W. (1977) Nucleic Acids Res. 4, 2527–2538.
[2] Simoncsits, A., Brownlee, G. G., Brown, R., Rubin, J. and Guillez, H. (1977) Nature 269, 833–836.
[3] Lockard, R., Alzner-Deweerd, B., Heckman, J., Mac Gee, J., Tabor, M. and RajBhandary, U. (1978) Nucleic Acids Res. 5, 37–55.
[4] Branlant, C., Sriwidada, J., Krol, A. and Ebel, J. P. (1977) Nucleic Acids Res. 4, 4323–4346.
[5] Vassilenko, S. K. and Ryte, V. (1975) Biokhimia 40, 578–583.
[6] Krol, A., Machatt, M. A., Branlant, C. and Ebel, J. P. (1978) Nucleic Acids Res. 5, 4933–4948.
[7] Carbon, P., Ehresmann, C., Ehresmann, B. and Ebel, J. P. (1978) FEBS Lett. 94, 152–156.
[8] Bruce, A. and Uhlenbeck, O. (1978) Nucleic Acids Res. 5, 3665–3677.
[9] Sanger, F. and Coulson, A. R. (1978) FEBS Lett. 87, 107–110.
[10] Branlant, C. and Ebel, J. P. (1977) J. Mol. Biol. 111, 215–256.
[11] Branlant, C., Krobko, V. and Ebel, J. P. (1976) Eur. J. Biochem. 70, 471–482.
[12] Branlant, C., Sriwidada, J., Krol, A. and Ebel, J. P. (1977) Eur. J. Biochem. 74, 155–170.
[13] Young, R., Macklis, R. and Steitz, J. (1979) J. Biol. Chem. 254, 3264–3271.
[14] Kenerly, M., Morgan, E., Posyl-Lindah, L. and Nomura, M. (1977) J. Bacteriol. 132, 931–949.
[15] Fellner, P. (1969) Eur. J. Biochem. 11, 12–27.
[16] Branlant, C., Sriwidada, J., Krol, A., Fellner, P. and Ebel, J. P. (1975) Biochimie 175–225.
[17] Branlant, C., Sriwidada, J., Krol, A. and Ebel, J. P. (1976) Nucleic Acids Res. 3, 1671–1688.