Sequence Analysis of Specific Areas of the 16S and 23S Ribosomal RNAs

by
P. FELLNER
F. SANGER

MRC Laboratory of Molecular Biology,
Hills Road, Cambridge

These ribosomal RNAs seem to be largely homogeneous, at least with respect to the methylated sequences. Methylation is very specific and occurs at a small number of loci.

ALTHOUGH much has been learnt in recent years about the mechanism of protein synthesis, the means by which ribosomes carry out their multiple functions remain unclear. A fuller knowledge of their structure would seem to be necessary in order to understand how they work. Recently, the nucleotide sequence of the 5*S* ribosomal RNA has been determined[1,2]. The 5*S* RNA comprises about 2·5 per cent of the total RNA content of *E. coli* ribosomes, with the rest of the RNA being made up of equimolar amounts of the 16*S* and 23*S* species. Besides the four major mononucleotides, the latter RNAs also contain small amounts of methylated nucleotides[3–5] which can be labelled specifically *in vivo* with radioactive methyl-labelled methionine. As a preliminary approach to the problem of structural investigation of these larger RNAs, we have studied the nucleotide sequences around the methylated sites, using the methods which have been developed in this laboratory for fractionating radioactive oligonucleotides[6].

The principal part of this study was conducted with ribosomal RNA from *E. coli K*12 *W*6 thr⁻ leu⁻ met⁻ (strain *CB*3 of Dr S. Brenner). A more cursory examination of the methylated oligonucleotides from *E. coli* MRE 600 RNA has also been carried out, and these seem to be identical with those from the RNA of *E. coli* strain *CB*3. RNA was isotopically labelled by incorporation of ^{32}P-orthophosphate, and either ^{3}H or ^{14}C-(methyl)-methionine, into inocula growing in the appropriate carrier-free media. Incorporation was effected in conditions of stringent coupling of RNA and protein biosynthesis to avoid aberrant methylation of the RNA. The ribosomal RNA was extracted in the manner of Sanger *et al.*[6], and the 16*S* and 23*S* components were fractionated either by centrifugation through 15–30 per cent sucrose gradients[7] or by disc electrophoresis through 2·4 per cent polyacrylamide gels[8]. The RNAs purified by these procedures were utilized as substrates in digestions with *T*1 ribonuclease and bacterial alkaline phosphatase[9]. The resulting de-phosphorylated oligonucleotides were fractionated by electrophoresis in two dimensions, and detected by autoradiography.

In the first case, the methylated oligonucleotides were identified from among the large numbers of non-methylated products by employing RNA substrates labelled with both ^{32}P and ^{3}H-(methyl) groups, and subsequently searching for doubly labelled material among the fractionated oligonucleotides on fingerprints of the RNAs. The positions of the methylated products indicated by this double labelling method were confirmed by fingerprinting ^{14}C-(methyl)-RNAs, and detecting the products by direct autoradiography. Such autoradiographs of the 16*S* and 23*S* ^{14}C-(methyl)-RNA fingerprints are shown in Figs. 1 and 2.

Sequence analysis of these methylated oligonucleotides has been carried out, using both ^{32}P and ^{14}C-(methyl)-labelled material. The method of sequencing used most often has been to subject the oligonucleotides to partial digestion with exonucleases, obtaining a variety of degradation products resulting from the sequential removal of nucleotides by the action of the enzymes. Snake venom phosphodiesterase[9,10] has been used most frequently for this purpose. The sequences which have been determined are reported in Tables 1 and 2. The identities of the methylated mononucleotides contained within these sequences have been confirmed in most instances by co-chromatography of ^{14}C-(methyl)-nucleotides or the corresponding bases with authentic marker compounds. Two methylated nucleotides have been encountered which have not been identified so far. These

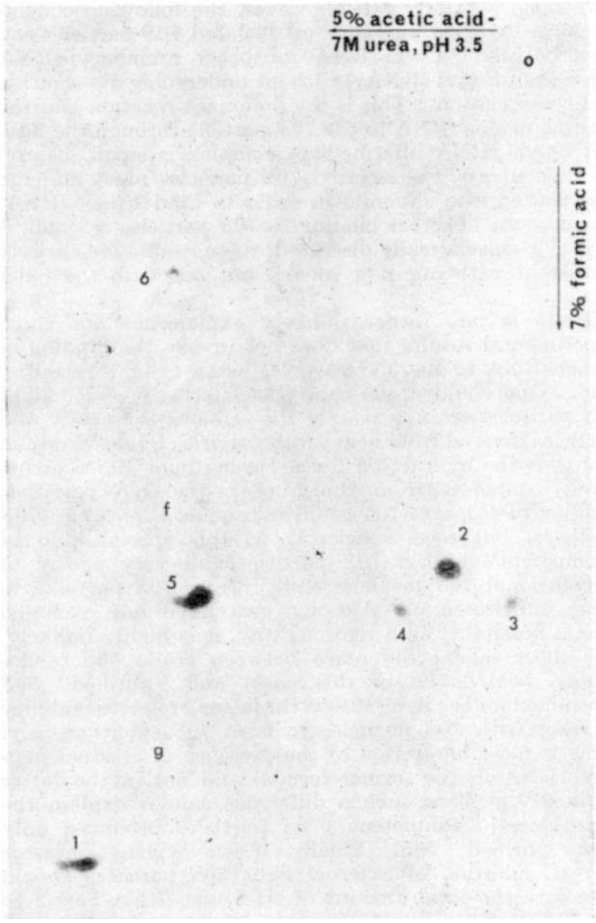

Fig. 1. Fingerprint of ^{14}C-(methyl) 16*S* RNA. Electrophoresis in the first dimension (*p*H 3·5) was carried out on cellulose acetate, and in the second dimension on DEAE-paper, according to Sanger *et al.*[6].

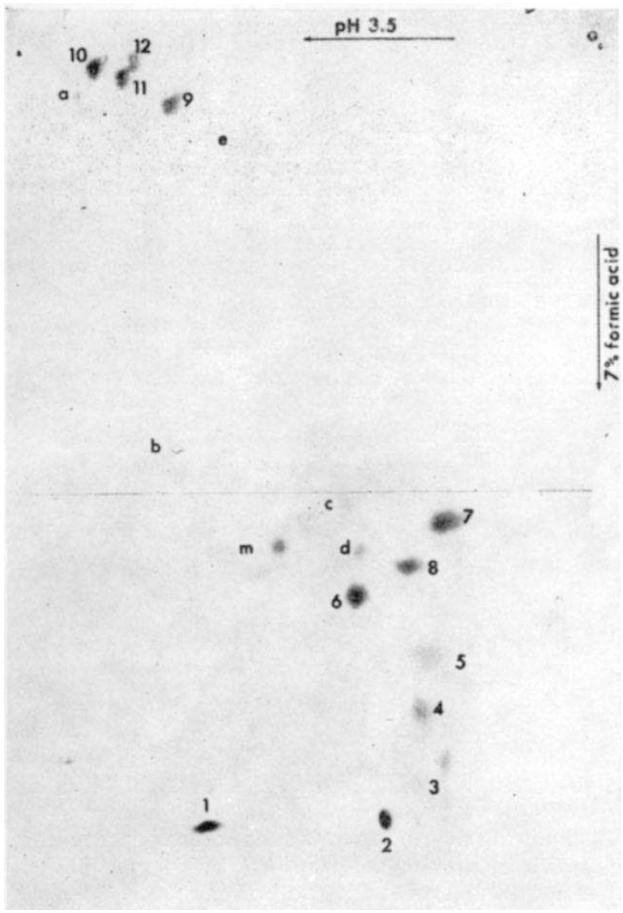

Fig. 2. Fingerprint of ^{14}C-(methyl) 23*S* RNA. Fractionation was carried out as described for Fig. 1. Spots *a*, *b* and *e* are minor components believed to result from further cleavage of spot 10. Spot *m* results from some cross-contamination of this preparation of the 23*S* RNA with 16*S* RNA, and is caused by the presence of the oligonucleotide m‡A–M‡A–C–C–U–G (spot 5 on Fig. 1).

possess properties which do not yet correspond with those reported for any other methylated nucleotides (see footnotes to Tables 1 and 2).

The molar ratios of the methylated sequences were directly determined from the relative amounts of the products arising on fingerprinting of the ^{14}C-(methyl)-labelled RNAs. Results are reported in Tables 1 and 2.

Table 1. SEQUENCES AND AMOUNTS OF METHYLATED OLIGONUCLEOTIDES FROM THE 16*S* RNA

Spot	Sequence*	Molar ratio†	Suggested frequency (moles/mole of RNA)
1‡	G–m^{2}G	3·62	4
2§	G–m^{4}C^m–C–m$_2$C–C–G	1·00	1
3	G–C–C–m^{7}G–C–G	1·93	2
4	G–m^{6}C–A–A–C–G	1·33	1
5	G–m‡A–m‡A–C–C–U–G	2·00	2
6	G–U–m^{5}C–A–C–A–C–C–U–A–G	1·04	1
f¶		0–1·08	
g‖		0–1·06	

* From the known specificity of *T*1 RNase action, a Gp residue can be placed at the 5′-terminus of each of the oligonucleotide sequences, as shown here (however, the presence of some or all of the m^{2}G residues in these positions cannot be excluded). The following symbols are used to represent the modified components: m^{2}G, N^2-methylguanosine; m^{5}C, 5-methylcytidine; m^{4}C , N^4-methyl-(2′-O-methyl)cytidine; m^{7}G, N^7-methylguanosine; m‡A, N^6-dimethyladenosine.

† The molar ratios were determined from two to five sets of observations.

‡ In spot 1, m^{2}G was liberated as the 2′ : 3′ cyclic phosphate.

§ The component m$_2$C is not fully identified, but displays properties which would suggest that it is a di-methylated derivative of cytidine.

¶ The minor spot *f* has been encountered in variable amounts in the 16*S* RNA fingerprints. No sequence analysis has been conducted on this oligonucleotide, but it is believed to contain T (ribothymidine).

‖ The minor spot *g* has also been found to occur in variable amounts, and is thought to contain m^{2}G.

Table 2. SEQUENCES AND AMOUNTS OF THE METHYLATED OLIGONUCLEOTIDES FROM THE 23*S* RNA

Spot	Sequence*	Molar ratio†	Suggested frequency (moles/mole of RNA)
1‡	G–m^{2}G	1·21	2
2	G–C–U^m–G	0·89	2
4	G–C(C, U)m^{2}G	1·12	2
5	G–C–m^{2}A–U–G	1·05	2
6	G–A–U–m^{6}C–C–G	1·23	2
7	G–C–m^{6}A–A–G	1·16	2
8	G–A–C–C–m^{6}A–G	0·89	2
9§	G–ψ–A–A–C–A–mU–A–ψ–C–G	0·93	2
10	G–A–C–A–U–A–U–m^{1}G–ψ–T–G	0·99	2
11	G–A–A–A–T–U–C–C–U–U–G	1·00	2
12¶	. . . m^{7}G . . .	0·85	2
c‖	G–C^m–C . . .	0·34	
d‖	G–C^m–C . . .	0·24	
3‖	G–C^m–C^m–U–G	0·35	

* † ‡ See appropriate footnotes to Table 1; additional abbreviations for modified nucleosides employed here are: m^{2}A, 2-methyladenosine; m^{6}A, N^6-methyladenosine; m^{1}G, N^1-methylguanosine; ψ, pseudo-uridine; C^m, 2′-O-methylcytidine; U^m, 2′-O-methyluridine.

§ mU is a derivative of uridine, as yet unidentified, displaying properties very similar to T.

¶ Spot 5 contains m^{7}G, and has not been sequenced in this study. This oligonucleotide is thought to contain at least fifteen nucleotides, from its fractionation properties.

‖ These oligonucleotides are always found in low amounts on fingerprints of the 23*S* RNA.

To obtain the absolute frequencies of occurrence of these sequences in the ribosomal RNA molecules, it is necessary to determine the total numbers of methyl groups which are introduced into the RNA molecules. These values were determined by comparing the ^{32}P/^{14}C ratios of doubly labelled RNAs with the corresponding ratios in methylated oligonucleotides of defined structure, arising from the fingerprinting of such RNAs. The results indicate that the 16*S* and 23*S* RNA molecules contain 21·9 and 27·5 methyl groups respectively. This result is in good agreement with the findings of Dubin and Gunalp[5] from a similar type of estimation, but roughly 50 per cent greater than the values of Hayashi *et al.*[4] derived from a direct measurement of the specific activities of ^{14}C-(methyl)-labelled RNAs. On the basis of these findings, the mean frequencies of occurrence of the methylated sequences have been suggested in Tables 1 and 2.

The finding that the major methylated oligonucleotides all occur twice in the 23*S* RNA suggests that the molecule is made up of two sections, which may be identical or else display considerable homologies of nucleotide sequence, for it seems extremely unlikely that the duplications encountered could have arisen by chance or be confined only to the sequences studied. This might suggest that the 23*S* RNA is composed of two identical or similar polynucleotide chains, or that it consists of a single chain with a duplicated sequence, which could have arisen by a "gene-doubling" mechanism. The results of end-group determinations[11,12] support the former suggestion, whereas the fact that a dissociated molecule cannot readily be obtained in conditions rigorously excluding nuclease activity[13] supports the latter possibility.

In the 16*S* RNA certain of the methylated sequences are repeated, suggesting that some more generalized repetition may also occur in it. Such a partial duplication was found in the 5*S* RNA from *E. coli*[1] and may be a general property of ribosomal RNAs.

These results suggest that each of the ribosomal RNAs is largely homogeneous, at least with respect to the methylated sequences, and that the methylation is a highly specific process, occurring at a small number of loci, encompassed by specific nucleotide sequences. It can be suggested that such sequences are specifically positioned within the RNA molecules, in terms of either primary or secondary structure, or both; but that all specific nucleotide sequences fulfilling these restrictions of placement will become modified in the mature RNA. This proposal is supported by the finding that all of the sequence G–A–A–A–A–T–U–C–C–U–U–G is present in the fully modified condition. It may be that methylases recognize stretches of primary sequence sufficiently extensive to exclude all except the correct sequences from being subjected to modification processes. Such a mechanism would allow

the smaller methylated fragments encountered in this study to form part of larger specific nucleotide sequences recognized by the methylases, which have been cleaved in the course of digestion with *T*1 RNase. If, however, the primary structure acts as the sole determinant of methylase specificity, it is necessary to postulate the existence of a rather large number of methylases. For example, it is found that m^5C occurs in three different sequences, lacking any apparent similarities, so that three distinct methylases would be needed to carry out this type of methylation in ribosomal RNA maturation. Generally, it is found that no characteristic sequences are associated exclusively with particular types of modification, nor do any such sequences occur in association with a variety of types of modification.

A feature of the modified sequences is the frequency with which multiple modifications are encountered, very close to each other. It is remarkable that such a property has become evident even from examination of these short stretches of the molecules, and it seems very likely that some special functional significance attaches to such a distribution of modified nucleotides.

Received June 26, 1968.

1 Brownlee, G. G., Sanger, F., and Barrell, B. G., *Nature*, **215**, 735 (1967).
2 Forget, B. G., and Weissman, S. M., *Science*, **158**, 1695 (1967).
3 Starr, J. L., and Fefferman, R., *J. Biol. Chem.*, **239**, 3457 (1964).
4 Hayashi, Y., Osawa, S., and Miura, K., *Biochim. Biophys. Acta*, **129**, 519 (1966).
5 Dubin, D. T., and Gunalp, A., *Biochim. Biophys. Acta*, **134**, 106 (1967).
6 Sanger, F., Brownlee, G. G., and Barrell, B. G., *J. Mol. Biol.*, **13**, 373 (1965).
7 Scherrer, K., and Darnell, J. E., *Biochem. Biophys. Res. Commun.*, **7**, 486 (1962).
8 Loening, U. E., *Biochem. J.*, **102**, 251 (1967).
9 Brownlee, G. G., and Sanger, F., *J. Mol. Biol.*, **23**, 337 (1967).
10 Holley, R. W., Madison, J. T., and Zamir, A., *Biochem. Biophys. Res. Commun.*, **17**, 389 (1964).
11 Nichols, J. L., and Lane, B. G., *Canad. J. Biochem.*, **45**, 937 (1967).
12 McIlreavy, D. J., and Midgley, J. E. M., *Biochim. Biophys. Acta*, **142**, 47 (1967).
13 Spirin, A. S., in *Progress in Nucleic Acid Research*, **1**, 301 (Academic Press, New York, 1963).

Binding of Formylmethionyl-*t*RNA and Aminoacyl-*t*RNA to Ribosomes

by
TAKAHISA OHTA
ROBERT E. THACH

Department of Chemistry and
Department of Biochemistry and Molecular Biology,
Harvard University,
Cambridge, Massachusetts

Inhibition of aminoacyl-*t*RNA binding by the purified initiation factors f_1 and f_2 could be regarded as a mechanism for ensuring that only F-met-*t*RNA binds initially to the entry site, thereby preventing random or non-specific binding by aminoacyl-*t*RNA. This inhibition would be relieved by the hydrolysis of the GTP molecule normally involved in the binding of F-met-*t*RNA.

GTP is required for the initial binding of *N*-formylmethionyl-*t*RNA (F-met-*t*RNA) to ribosomes[1–4], and for the subsequent incorporation of the F-met moiety into a dipeptide[5,6]. Hydrolysis of the GTP molecule does not seem to occur during the binding step, but rather at some intermediate stage, before the actual formation of the first peptide bond[2,4]. These facts led to the "single entry site" model for the 70*S* ribosome, in which it was proposed that in normal conditions there might be only one ribosomal site (the A site or entry site) available for the initial binding of *t*RNA (ref. 5). In the process of initiation, F-met-*t*RNA would bind first at this entry site and then be translocated to a second site, the peptidyl or "P" site, in which state it would be reactive with puromycin, or a second aminoacyl-*t*RNA which could now bind to the entry site.

An alternative to this model was a hypothetical scheme in which the F-met-*t*RNA was thought to bind directly to the P site, without having first gone to the A site. In this position it would be available for peptide bond formation without the need for the hydrolysis of GTP. Monro[7] has shown that this sequence of events can, in fact, be observed in non-physiological conditions.

Evidence has recently been obtained which suggests that neither of these fairly simple models can be entirely correct, for it now seems that in the normal initiation process F-met-*t*RNA binds first to the 30*S* subunit to form an "initiation complex"[4,8–11]. This complex is later joined by the 50*S* subunits to form a 70*S* particle, which is now capable of forming the first peptide bond[10–12].

Sequences Coding for F-met and Val

In order to obtain further insight into this extremely complicated situation, we have studied the translation of the first two adjacent codons in the synthetic message ApUpGpGpUpUpU . . . [abbreviated: $ApU(pG)_2(pU)_n$]. These two sequences code for F-met and val, respectively. We have tried to answer the following question. Can both F-met-*t*RNA and val-*t*RNA bind simultaneously to the same ribosome (whether it be 30*S* or 70*S*) when hydrolysis of GTP is prevented, that is, in the presence of GMP-PCP ? As we shall show, the answer is no. In the presence of purified initiation factors (f_1 and f_2), only F-met-*t*RNA binds[13]; in the presence of T-factor[14], only val-*t*RNA (and phe-*t*RNA) binds. When all three factors are present (f_1, f_2 and T), only F-met-*t*RNA binds. These results lead to a second question: what is the physical basis for the inhibition of val-*t*RNA binding by the addition of f_1 and f_2 ? The answer to this question is not known. It is suggested, however, that initiation factor may alter the specificity of the ribosome so that it can bind only F-met-*t*RNA, and not aminoacyl-*t*RNA. This alteration might occur through the attachment of factors to the entry site, which is thought to be on the 30*S* particle[11]. This inhibition does not require the simultaneous binding of F-met-*t*RNA.

A prerequisite for the successful execution of these experiments was a thorough understanding of the enzymatic requirements for the binding of F-met-*t*RNA and aminoacyl-*t*RNA to ribosomes. At the time this investigation was begun, it was not clear whether an enzyme was required for the normal, A-site binding of aminoacyl-*t*RNA. Similarly, it was uncertain whether the initiation factors alone were sufficient for the binding of F-met-*t*RNA. To answer these questions a study of the requirements for the binding of F-met and aminoacyl-*t*RNA to ribosomes at low Mg^{++} concentration was undertaken. The results of this study are presented in the first part of this communication; subsequently, binding experiments using both GTP and GMP-PCP are described.