Volume 8 Number 23 1980 Nucleic Acids Research

Characterization of the Escherichia coli 23S ribosomal RNA region associated with ribosomal protein L1. Evidence for homologies with the 5'-end region of the L11 operon

Christiane Branlant+, Alain Krol and Jean-Pierre Ebel

Institut de Biologie Moléculaire et Cellulaire du CNRS, Université Louis Pasteur de Strasbourg, 15, rue René Descartes, 67084 Strasbourg Cédex, France

Received 29 September 1980

SUMMARY

The results previously obtained upon studying the L1-23S RNA complex by the fingerprint technique have been reexamined in the light of new data on 23S RNA primary structure. The 23S RNA region that remains associated with the L1 ribosomal protein after RNase digestion of the synthetic complex lies between nucleotides 2067 and 2235 from the 5'-end of the molecule. This region contains a m^7G near to the 5'-end and possesses a high degree of mutability in *E.coli*. Three different sequences were observed in *E.coli* MRE 600. All three sequences differ in two positions relative to the corresponding sequence in rrnB cistron from *E.coli* K12.
Striking homology is observed between the 23S RNA region associated with protein L1 and the 5'-part of L11 operon. This observation supports the model of feedback regulation of r-proteins synthesis proposed by Yates *et al.* (PNAS, 77, 1837) and strongly suggests that the region of 23S RNA located between positions 2155 and 2202 is essential for the binding of protein L1.

INTRODUCTION

The 50S ribosomal protein L1 is assumed to play an important role in ribosome function (1,2). Yates *et al.* (3) recently showed that protein L1 is the translational repressor that regulates the synthesis of both protein L1 and L11 in the L11 operon. From their proposed explanation, every ribosome initiated at the beginning of the L11 message would continue to translate the L1 message and no independant translational initiation would take place at the beginning of the L1 message. Furthermore, initiation at the beginning of the L11 message would be repressed by the binding of protein L1 to the mRNA region containing the initiation site. Since protein L1 binds specifically to 23S RNA under conditions of *in vitro* ribosome reconstitution (4,5), this model of regulation of L11 operon by protein L1 can be regarded as competition between 23S RNA and protein's L11, L1 mRNA. If this is so, one might find some structural homology between this mRNA and the region of 23S RNA associated with protein L1. Since, this 23S RNA region was only partially characterized using the fingerprint technique (6,7,8), it was therefore of great interest to complete

its identification and to compare its sequence to that of the postulated mRNA binding region.

We recently determined 70 % of the 23S RNA nucleotide sequence using the methods for RNA sequencing based on polyacrylamide gel electrophoresis (9) and Brosius *et al.* (10) determined the complete nucleotide sequence of the 23S RNA gene from *E.coli* rrnB cistron present in the transducing phage λ rifd[18].

In this paper, our previous results obtained on studying protein L1-23S RNA complex are re-examined in the light of these new sequencing data. This investigation completes the identification of the region of *E.coli* 23S RNA associated with protein L1. It reveals the presence of several microheterogeneities and shows the existence of sequence homologies with the region of L11 operon corresponding to the initiation site for translation.

METHODS

This paper deals with experimental results already described in previous papers.

1. Preparation and T1 RNase digestion of the L1-23S RNA complex (6).
2. Fingerprint analysis of the T1-resistant ribonucleoprotein complex (6,8).
3. Localization of the RNA region contained in the ribonucleoprotein complex within 23S RNA molecule (11).
4. Sequence analysis of the 5'-part of this RNA region by the gel sequencing technique (9).
5. Obtention of this RNA region upon digestion of the 50S subunits (7).

RESULTS AND DISCUSSION

T1 RNase digestion of the synthetic complex formed between *E.coli* 23S RNA and protein L1, released three T1-resistant ribonucleoprotein complexes which were previously denoted 1, 2 and 3. Each of them contained a series of RNA subfragments denoted 1 to 13 and 1'. These subfragments were previously analyzed by the fingerprint technique, and the sequences of the products obtained by T1 and pancreatic RNases digestion were determined (6,8). These RNA subfragments were found to be located in a 175 nucleotide long region, located between nucleotides 700 and 900 from the 3'-end of 23S RNA, known as section S (11).

A ribonucleoprotein complex containing proteins L1 and L9 was released upon T1 RNase digestion of the 50S subunit. This complex contained RNA subfragments similar to those observed in ribonucleoprotein 2 (7).

1. Characterization of the 23S RNA region remaining associated with protein L1 after digestion of the synthetic complex.

1.1 Identification of nucleotide sequence in the 23S RNA gene studied by Brosius *et al.* (10) which corresponds to the L1-associated region of 23S RNA. We compared the T1-oligonucleotides from ribonucleoprotein complexes 1 and 2 (6,7), with those expected from the various regions of the 23S RNA according to the DNA sequence established by Brosius *et al.* We found that the RNA in ribonucleoprotein L1 corresponds to the 23S RNA region lying between positions 2067 and 2235 and that in ribonucleoprotein 2 to the 23S RNA region located between positions 2083 and 2235. Figure 1 shows that most of the T1 and pancreatic RNAases digestion products expected to be released by these RNA regions according to the DNA structure were previously observed upon digestion of ribonucleoprotein 1 and 2 with these enzymes (6,7). There is nevertheless a few discrepancies. T1 oligonucleotides U-U-G {10} is expected twice instead of once as previously described (6), and oligonucleotide U-G {17} 5 times instead of 6. These discrepancies can be explained by a base difference at position 2203. Indeed, comparison of the fingerprints of the various RNA subfragments contained in ribonucleoprotein 2 shows that our RNA had a guanine at this position, whereas Brosius *et al.* observed a uracil. In addition to this difference; T1 oligonucleotide U-A-A-U-C-C-G is expected from the DNA sequence, instead of it we found the oligonucleotide (A-Up, 2Cp)G {13}. This oligonucleotide can be explained by the presence in the RNA studied of a guanine at position 2211 instead of the adenine as observed in the DNA sequence. These two changes explain the discrepancies also observed between the pancreatic RNase digestion products previously obtained (7) and those expected from the DNA structure : oligonucleotides G-G-A-Cp {11}, A-A-Up {14}, G-A-Cp {15} and A-Up {17} were found with molar frequencies of 1, 1, 1 and 3 respectively instead of 2.
The existence of differencies between our RNA sequence and the DNA sequence of Brosius *et al.* is not surprising since the RNA was obtained from *E.coli* MRE 600 which derives from an *E.coli* B strain, whereas they used phage λrif d^{18} which was constructed from the *E.coli* K12 strain (12).

1.2 Cistron heterogeneities in the *E.coli* MRE 600 23S RNA region associated with protein L1. In addition to the sequence variations mentioned above, which are probably due to strain differences, the 23S RNA region associated with protein L1 displays cistronic heterogeneities within the *E.coli* MRE 600 strain. One of them was clearly identified previously at position

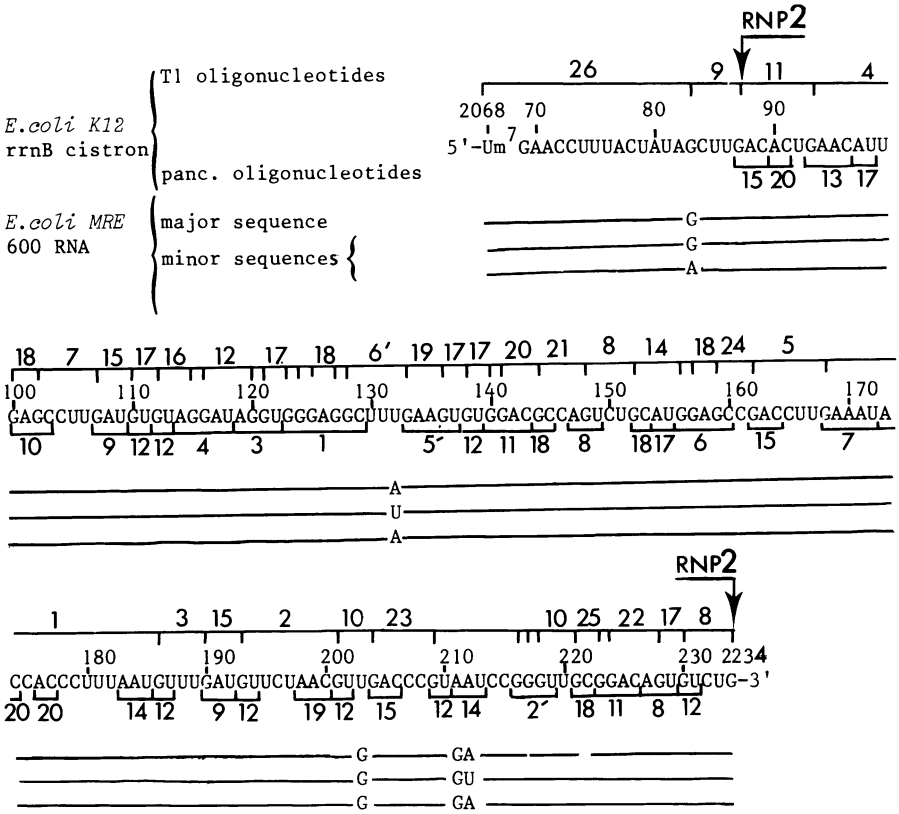

Figure 1 : The *E.coli* 23S RNA region associated with protein L1. (a) The DNA sequence encoding for this RNA region in the rrnB cistron of *E.coli* K12 (10), T1 and pancreatic digestion products expected from digestion of the corresponding RNA are indicated with the numbering previously adopted for the oligonucleotides of ribonucleoproteins 1 and 2 (6,7). Hyphens representing phosphate linkages are omitted for reason of space. (b) The three possible sequences observed for the corresponding RNA region of *E.coli* MRE 600, only observed changes are represented.

2132 which was found to be either adenine or uracil. In the first case, T2 oligonucleotide 6 was observed, it was replaced by oligonucleotide 6' in the second case (6). This mutation was previously observed to be linked to another one. Indeed, the presence of T1 oligonucleotide 6' was accompanied with the appearance of a new T1 -oligonucleotide with the following composition : (2Up, 2Cp)G. Re-interpretation of the fingerprints and comparison of the DNA sequence reveal that this mutation occurs at position 2212, where the adenine

is replaced by uracil, leading to T1 oligonucleotide U-U-C-C-G instead of A-U-C-C-G {13}. Thus a cistronic heterogeneity exists in *E.coli* MRE 600 23S RNA at position 2212, adjacent to the position 2211 where a difference is observed between the RNA we studied and the DNA sequenced by Brosius *et al.* It should be pointed out that in *E.coli* MRE 600 those cistrons having U at position 2132 also have U at position 2312, which is not the case for the gene sequence where there is U at position 2132 but A at position 2312.
Finally re-examination of our previous results reveals the existence of a third cistronic heterogeneity in *E.coli* MRE 600. This one previously escaped notice : the guanine 2083 can be replaced by another nucleotide (X), thus T1 oligonucleotide U-m^7G-A-A-C-C-U-U-U-A-C-U-A-U-A-G, which is the oligonucleotide 2a from the 18S part of 23S RNA (11), is replaced by U-m^7G-A-A-C-C-U-U-U-A-C-U-A-U-A-X-C-U-U-G {26}. Due to their close properties, these two oligonucleotides were previously confused. Re-examination of the pancreatic RNase digestion products of oligonucleotide {26} indicates that X should be an adenine. This heterogeneity explains the existence of the larger ribonucleoprotein 1. Replacement of guanine by adenine at position 2083 prevents T1 RNase cleavage at this position and therefore a longer RNA region remains associated with protein L1 after digestion of the synthetic complex. As ribonucleoprotein 1 released T1 oligonucleotides 6 and 13 but not T1 oligonucleotides 6' and U-U-C-C-G, even in trace amount, the cistrons bearing an adenine at position 2083 also have adenine at position 2132 and 2312.
Therefore, 23S RNA molecules of *E.coli* MRE 600 present a high degree of cistronic heterogeneity (1.8 %) in the region that binds protein L1. Three different sequences were observed (Fig. 1). One of them was present in larger amount relative to the other two. Therefore, either it is encoded by several cistrons, or the corresponding cistrons is expressed more frequently than the others. Taking into account both the RNA sequences and the DNA sequence the L1 associated region of *E.coli* 23S RNA present a significant degree of variability (3 %).
This region of 23S RNA is probably not the only one exhibiting such variability. Brosius *et al.* reported the sequence A-C-C-C-U-U-U-A-A-G-G-G between positions 2792 and 2803, whereas we found the sequence A-C-U-C-C-U-U-G-A-G-A-G. Bram *et al.* (13) recently sequenced the 5'-end of the 23S RNA genes of the rrnD and rrnX cistrons and found the latter sequence in both cistrons.

1.3 Occurrence of post-transcriptional modifications in or close to the protein L1 binding region of 23S RNA. rRNAs contain a limited number of post-transcriptionally modified nucleotides (14). It is nevertheless impor-

tant to determine their localization in the molecule since they might have an important role. They can only be detected by direct analysis of the RNA. Due to the cistronic heterogeneities described above it was not possible to analyse the whole 23S RNA region associated with protein L1 using new methods for RNA sequencing. Nevertheless, using this technique we studied the nucleotide sequence of the section S which contains the 60 nucleotides at the 5'-end of the L1 associated region. The sequence obtained for these 60 nucleotides which was reported in a short communication (9), is identical to that found by Brosius *et al.* But, according to the RNA analysis, guanine 2069 is methylated on position 7. Figure 2b and 2c give the demonstration of the existence of this m^7Gp. A nucleotide of section S, which corresponds to the second nucleotide of the L1 associated region (starting from the 5'-end), was recognized by none of the enzymes employed excepted *Phy* I RNase (Fig. 2b). This nucleotide creates on the sequencing gel a band deplacement characteristic of the addition of a guanine (Fig. 2b). Finally, according to two-dimensional analysis on polyacrylamide gel, this nucleotide was positiveley charged at pH 3.5, and negatively at pH 8.3 (Fig. 2c) as should be expected for m^7Gp whose N7 is protonated at pH 3.5 but not at pH 8.3.

We would like to mention, that the part of section S preceding the L1 associated region of 23S RNA, should also contain a modified nucleotide. At position 2030, Brosius *et al.* reported an adenine whereas our study shows that the corresponding nucleotide is cleaved by none of the RNases employed (Fig. 2a). In particular, it is not cut by U_2 RNase as are the other 25 adenines present in this RNA fragment. The presence of a m^6Ap at this position would explain both the absence of U_2 RNase cleavage on the sequencing gel and an observation we previously made on examining the fingerprint obtained after T1 RNase digestion of the 18S fragment of 23S RNA (15). This fragment which contains section S has an unusual T1 RNase digestion product, that we denoted c and whose electrophoretic mobility on DEAE paper is close to that of A-A-G.

Therefore, in addition to the m^7Gp at position 2069, a second modified nucleotide, likely m^6Ap, is present near to the L1 binding region of 23S RNA at position 2030.

Re-examination of the previous results allowed us to achieve a complete characterization of the region of 23S RNA which remains associated with protein L1 after digestion of the synthetic complex. This re-examination can also provide information on the nucleotide sequences from the L1 associated region which are in close contact with the protein.

Figure 2 : Sequencing gels of 23S RNA section S

(a,b) Mono-dimensional fractionation of partial digests with T1(G), *Phy* I (-C), U2 (A) and pancreatic (C+U) RNases and with boiling water (L) (9) on 20 % (a) and 15 % (b) polyacrylamide gels. (c) Two-dimensional fractionation of a digest with boiling water (9).

2. The nucleotide sequences of the L1 associated region of 23S RNA which are in close contact with the protein.

Figure 3 represents the major RNA subfragment 1 of ribonucleoprotein 1 and all the RNA subfragments of ribonucleoprotein 2. These subfragments extend between positions 2067 and 2235 for ribonucleoprotein 1 and between positions 2083 and 2235 for ribonucleoprotein 2. As mentioned above, the presence of an additional fragment in ribonucleoprotein 1 results from the impossibility of T1 RNase cleavage at position 2083 when the nucleotide at this position is Ap instead of Gp. So that this additional fragment should not interact with protein L1. The protein L1 binding site should be located between nucleotides 2083 and 2235. The guanines 2201 and 2211 are very sensitive to T1 RNase attack (Fig. 3). These guanines and the surrounding nucleotides should not be in close contact with the protein.
Experiments previously made on the 50S subunits show that the interaction between protein L1 and 23S RNA should be identical in the synthetic complex and in the subunit. Indeed, limited digestion of the 50S subunit released a ribonucleoprotein complex , previously denoted II, which contains proteins L1 and L9 and a 23S RNA region slightly larger than the L1 associated region of the synthetic complex. The RNA subfragments corresponding to the region 2083-2235 were identical to those of ribonucleoprotein 2 from the synthetic complex (7). The additional nucleotide sequence might be related to the presence of protein L9.

3. Comparison between the L1 associated region of 23S RNA and other RNA regions found to or expected to bind this protein.

As stated in the introduction, Yates *et al.* proposed a model of regulation of L11 operon by protein L1 in which 23S RNA would compete with protein's L11, L1 mRNA for the binding of protein L1. We, indeed, observed interesting homologies between the nucleotide sequence preceding L11 message in the L11 operon (16) and the sequence of 23S RNA found associated with protein L1 (Fig. 4). Stanley *et al.* (17) observed that *Bacillus stearothermophilus* 23S RNA can bind ribosomal protein L1 from *E.coli*. They established a partial nucleotide sequence of the interacting region. In Figure 4, we compare the part of this sequence of *Bacillus stearothermophilus* 23S RNA which corresponds to the *E.coli* 23S RNA region displaying an homology with L11 mRNA. Similitudes are observed between *E.coli* and *B.stearothermophilus* 23S RNAs at the same place where homologies are observed between *E.coli* 23S RNA and the L11 operon and the two 23S RNAs display different sequences at the places

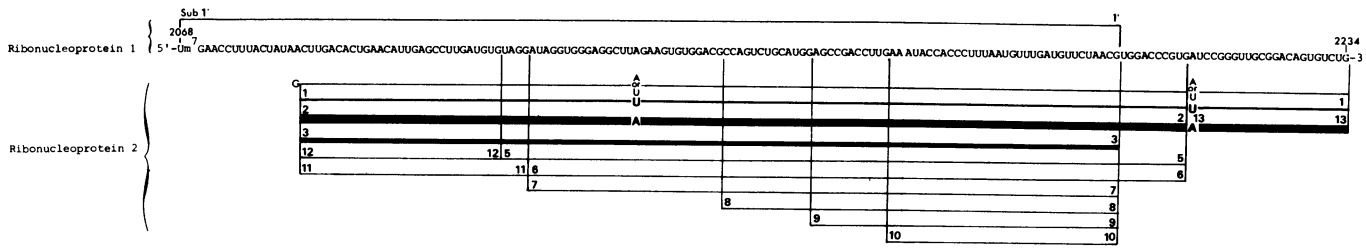

Figure 3 : RNA composition of ribonucleoproteins 1 and 2 released upon digestion of the L1-23S RNA complex. Subfragment 1' of ribonucleoprotein 1 and all the subfragments of ribonucleoprotein 2 are indicated. The thickness of the lines represent their relative yields. The nucleotides are numbered according to Brosius *et al.*(10).

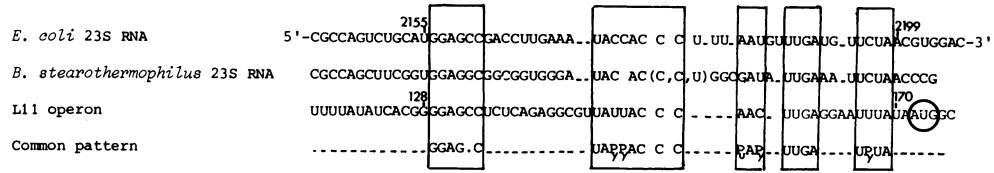

Figure 4 : Homologies between the region of 23S RNA associated with protein L1 and the region preceding the L11 message in the L11 operon (14). The sequence of the L11 operon has been numbered according to Post *et al.* The initiator A-U-G has been circled.

where *E.coli* 23S RNA and the L11 operon have different sequences. In other words, these three sequences have a common pattern which is represented in Figure 4. The occurrence of such homologies by chance in three different sequences is highly unlikely and probably reflects the common affinity of these three RNA regions for protein L1. The homologous region of mRNA is located just before the initiator A-U-G. Binding of protein L1 to this region will certainly block initiation. Therefore, these observations support the model of translational feedback regulation of r-proteins synthesis proposed by Yates *et al.* (3).

We previously observed that T1 oligonucleotide A-A-A-U-A-C-U-A-C-C-A-C-C-C-U-U-U-A-A-U-G {1} can bind 30S ribosomal protein S1 (18). This oligonucleotide belongs to the L1 associated region, displaying an homology with the nucleotide sequence preceding the L11 message (Fig. 4). It would be interesting to check if protein S1 can bind to the mRNA and therefore can control the synthesis of proteins L11 and L1.

The affinity of protein L1 for 23S RNA is expected to be stronger than that for the mRNA. Thus within the L1-23S RNA complex, protein L1 probably interacts with a larger number of nucleotides than within the postulated L1-mRNA complex. This could explain why the part of 23S RNA displaying an homology with mRNA does not encompass the whole 23S RNA region associated with protein L1. On the other hand,it is possible that some of the RNA subfragments present in the ribonucleoprotein are held by RNA-RNA interactions and therefore do not interact directly with the protein. One point is obvious : the set of short 23S RNA sequences located between positions 2155 and 2199 which present an homology with the sequence preceding the L11 message in the L11 operon should be very important for the binding of protein L1 to 23S RNA. Examination of the pattern common to both RNAs reveals that U can be replaced by C in many places, which is not the case for A and G. It is relevant that no heterogeneity was observed in the region of 23S RNA displaying the homology with the mRNA and that this area of 23S RNA is highly protected against ribonuclease action within the synthetic complex whereas position 2201 which is very sensitive to T1 RNase, is located next to the 3'-end of the homologous region.

ACKNOWLEDGMENTS

We wish to thank Dr. M. NOMURA for helpful discussions. This work was supported by the 'Centre National de la Recherche Scientifique' and the 'Délégation Générale à la Recherche Scientifique et Technique'. Dr. E. CROUSE is acknowledged for careful reading of the manuscript.

REFERENCES

1. KAZEMIE, M., (1976) Eur. J. Biochem. 67, 373-378.
2. HERNANDEZ, F., VAZQUEZ, D. and BALLESTA, J., (1977) Eur. J. Biochem. 78, 267-272.
3. YATES, J., ARFSTEN, A. and NOMURA, M., (1980) Proc. Nat. Acad. Sci. US, 77, 1837-1841.
4. MIZUSHIMA, S. and NOMURA, M., (1970) Nature, 226, 1214-1218.
5. GARRETT, R., MULLER, S., SPIERER, P. and ZIMMERMANN, R., (1974), J. Mol. Biol. 88, 553-557.
6. BRANLANT, C., KROL, A., SRIWIDADA, J., EBEL, J.P., SLOOF, P. and GARRET, R. (1976) Eur. J. Biochem. 70, 457-469.
7. BRANLANT, C., KROL, A., SRIWIDADA, J. and BRIMACOMBE, R., (1976) Eur. J. Biochem. 70, 483-492.
8. BRANLANT, C., KOROBKO, V. and EBEL, J.P. (1976) Eur. J. Biochem. 70, 471-482.
9. BRANLANT, C., KROL, A., MACHATT, M. and EBEL, J.P., (1979) FEBS Lett. 107, 177-181.
10. BROSIUS, J., DULL, T.G. and NOLLER, H., (1980) Proc. Nat. Acad. Sci. US, 77, 201-204.
11. BRANLANT, C., SRIWIDADA, J., KROL, A. and EBEL, J.P. (1977) Nucleic Acids Res. 4, 4323-4346.
12. KONRAD, D., KIRSCHBAUM, J. and AUSTIN, S., (1973) J. Bacteriol. 116, 511-516.
13. BRAM, R., YOUNG, R. and STEITZ, J., (1980) Cell., 19, 393-401.
14. FELLNER, P., (1969) Eur. J. Biochem. 11, 12-18.
15. SRIWIDADA, J., BRANLANT, C. and EBEL, J.P., (1979) Biochimie 61, 869-876.
16. POST. L., STRYCHARZ, G., NOMURA, M. LEWIS, H. and DENNIS, D., (1979) Proc. Nat. Acad. Sci. US, 76, 1697-1701.
17. STANLEY, J., SLOOF, P. and EBEL, J.P., (1978), Eur. J. Biochem. 85, 309-313.
18. KROL, A., BRANLANT, C., EBEL, J.P. and VISENTIN, L., (1977), FEBS Lett. 80, 255-260.