Sequence homologies between eukaryotic 5.8S rRNA and the 5' end of prokaryotic 23S rRNA: evidences for a common evolutionary origin

Bernard Jacq

Centre de Biochimie et de Biologie Moléculaire, C.N.R.S., 31 chemin Joseph Aiguier, B.P. 71, 13277 Marseille Cedex 9, France

Received 24 March 1981

ABSTRACT

The question of the evolutionary origin of eukaryotic 5.8S rRNA was re-examined after the recent publication of the E. coli 23S rRNA sequence (26,40). A region of the 23S RNA located at its 5' end was found to be approximately 50 % homologous to four different eukaryotic 5.8S rRNAs. A computer comparison analysis indicates that no other region of the E. coli ribosomal transcription unit (> 5 000 nucleotides in length) shares a comparable homology with 5.8S rRNA. Homology between the 5' end of E. coli 23S and four different eukaryotic 5.8S rRNAs falls within the same range as that between E. coli 5S RNA and 5S RNA from the same four eukaryotic species. All these data strongly suggest that the 5' end of prokaryotic 23S rRNA and eukaryotic 5.8S RNA have a common evolutionary origin. Secondary structure models are proposed for the 5' region of E. coli 23S RNA.

INTRODUCTION

Ribosomes from all living cells are composed of two ribonucleoprotein subunits differing in size, each containing a high molecular weight RNA. The small subunit contains a 16S or a 18S RNA (in prokaryotes and eukaryotes respectively) and the large one, a 23S or a 26S-28S RNA. In addition, this latter subunit contains a low molecular weight species, the 5S RNA.

Another small RNA, known as 5.8S RNA (1) is present only in eukaryotic cells. Unlike 5S RNA, 5.8S RNA is hydrogen-bonded to 26S-28S and arises from a late maturation step of the 28S precursor. 5.8S RNA has been found in cytoplasmic ribosomes of all plants and animals investigated and its size is fairly constant (156-162 nucleotides), except in Drosophila melanogaster and Sciara coprophila. In these two latter cases, the 5.8S RNA is only 123 nucleotides long and there is an additional 2S RNA of 30 nucleotides which is homologous to the 3' part of 5.8S RNAs from other species (2,3,4,5). Approximately, ten different 5.8S RNA have now been sequenced (for a review, see 6) and different secondary structure models (7,8,9,10) have been proposed for this molecule. Comparisons among 5.8S RNAs (3,11) and among eukaryotic 5S RNAs (12,13) show that 5.8S RNAs are more conservative than 5S RNA to

evolutionary changes (3,11,14). Therefore, considering : i) that all known 5.8S RNAs share important homologies and ii) that the existence of a 5.8S-like molecule has never been reported in prokaryotic organisms, a question immediately arises : what is the evolutionary origin of 5.8S RNA ?

Ten years ago, Doolittle and Pace proposed that 5.8S might be the evolutionary counterpart of bacterial 5S rRNA (15). This assumption was essentially founded on a comparison between the prokaryotic and eukaryotic RTU organization schemes : One gene coding for a small RNA species is present in addition to the genes coding for the two large rRNAs, both in pr and eu RTU. In prokaryotic organisms this small RNA is a 5S RNA while in eukaryotes it is a 5.8S RNA but both are co-transcribed with the large rRNAs and both are components of the large ribosomal subunit. The differences between pr and eu 5S RNAs and the homology found between pr 5S RNA and eu 5.8S RNA (16), further supported the idea that eu 5.8S RNA -and not eu 5S RNA- is related to pr 5S RNA :

- In eukaryotes the genes coding for 5.8S RNA are linked to the 18S and 28S genes whereas the 5S genes are not.
- Eu 5S RNA is a primary transcription product in contrast to pr 5S and eu 5.8S. Moreover distinct polymerases are responsible for transcription of eu 5S and 5.8S (17).
- Pr 5S and eu 5.8S possess one or more CGAAY sequences supposed to interact with the GTψCR sequences of tRNA (16).
- Reconstitution experiments have shown that eu 5S cannot be incorporated into bacterial ribosomes (18,19). Moreover E. coli 5S binding proteins interact with yeast 5.8S but not with yeast 5S RNA (20).

However, other observations argue against the hypothesis of Doolittle and Pace :

- With the exception of that in Drosophila (see above), all 5.8S RNAs are longer than pr or eu 5S RNAs (160 instead of 120 nucleotides).
- 5.8S RNA generally contains 4 or 5 nucleotides which are modified at specific sites whereas pr and eu 5S RNAs do not (yeast 5S RNA is an exception).
- Even though the 5.8S gene is part of the eu RTU as is the 5S RNA gene in the pr RTU, their respective localizations in the transcription unit are different, the order being 5'-16S 23S 5S-3' in prokaryotes and 5'-18S 5.8S 28S-3' in eukaryotes.
- 5.8S RNA is hydrogen-bonded to 28S RNA, in contrast to eu 5S RNA, for which

the formation of an in vitro complex with eu 18S has been reported (21).

- Finally, if eu 5.8S RNAs and pr 5S RNAs were evolutionarily linked, one would expect their sequences to be related. Cedergren and Sankoff have clearly demonstrated that the degree of homology between eu or pr 5S RNA and 5.8S RNA is not higher than that between two random sequences. The same comparison test showed that although eu 5S RNA and pr 5S RNA sequences are very different, they are indeed evolutionarily related. Consequently, they proposed that 5.8S may have evolved from the intercistronic region of the bacterial genome(14).

Evidence is presented here that significant homology (50 %) exists between a unique region located at the 5' end of the pr 23S RNA and eu 5.8S RNAs, suggesting that 5.8S RNA has evolved from the 5' end of the prokaryotic 23S RNA. In the course of writing this paper, R. Nazar (61) reported homology between trout 5.8S RNA and the 5' end of E. coli 23S RNA.

SEQUENCE DATA AND COMPARISON PROCEDURES

The bacterial ribosomal sequence

It seems now established that the E. coli genome contains seven operons coding for the ribosomal RNAs (22). In the past few years, the sequence of different regions of these operons have been determined but unfortunately no complete sequence of any one of them is as yet available. It is however possible to artificially construct an "hybrid" uninterrupted sequence representative of an almost complete RTU by the compilation of different results. In this work, we have constructed such a composite sequence, with different regions taken from the rrnB and rrnD operons. The entire sequence we used is 5 166 nucleotides long (Fig. 1) and is made up of six different contiguous sequences in the following order :

- the 5' part of the sequence starts in the promotor region of rrnD (23), 72 bp downstream from the beginning of the 16S gene ;
- the 16S rrnB gene : 1 541 bp (24) ;
- the 16S-23S spacer region of rrnD : 437 bp (25) ;
- the 23S rrnB gene : 2 904 bp (26) ;
- the 23S-5S intergenic region of rrnD : 92 bp (27) ;
- the 3' end is the 5S sequence previously determined at the RNA level : 120 bp (28).

The RNA sequence was then inferred from this DNA sequence and for practical reasons it was divided into eight consecutive subsequences of 640 nucleotides (A to H in Fig. 1), the last subsequence (H) being 686 nucleo-

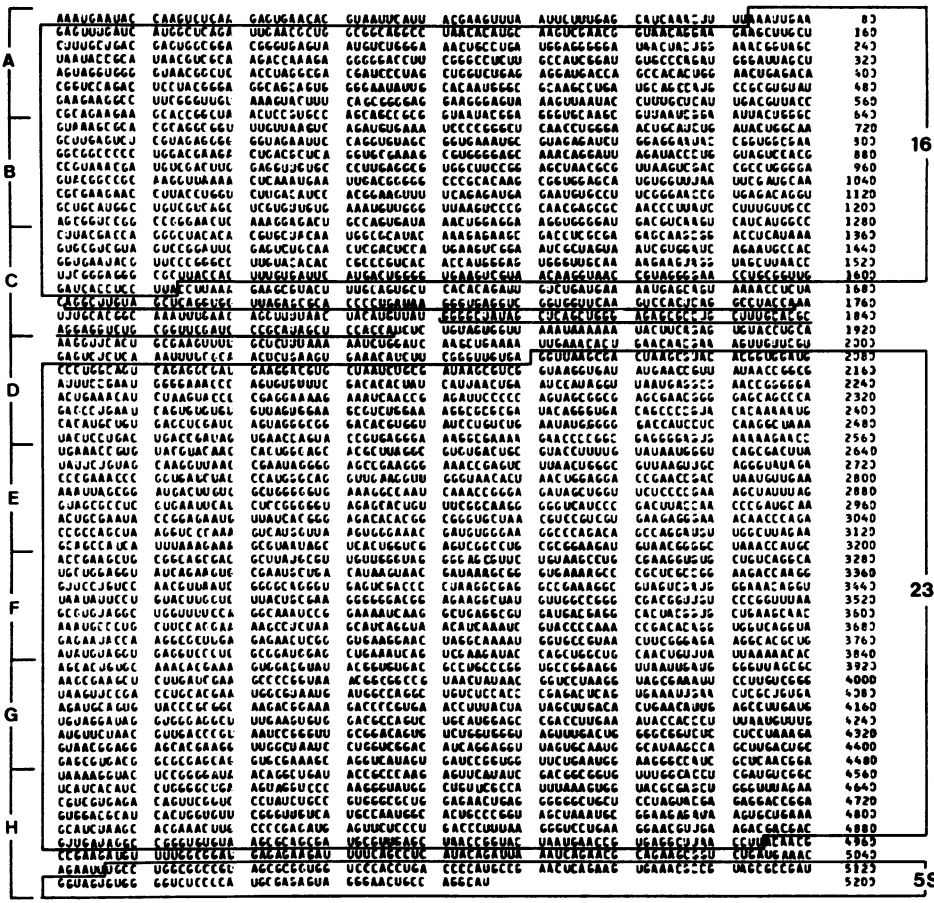

Figure 1 : The hybrid ribosomal bacterial RNA sequence.

This composite sequence was constructed with different regions of rrnB and rrnD E. coli operons as described in the sequence data section. The sequence is 5 166 nucleotides long and is divided into 8 subsequences (A to H) of 640 nucleotides (686 nucleotides for sequence H). The 16S, 23S and 5S sequences are boxed. The 16S-23S spacer region (24) contains two tRNA sequences ($tRNA_1^{Ile}$ and $tRNA_{1B}^{Ala}$) which are underlined.

tides long.

5.8S and 5S sequences :

Amongst ten known 5.8S sequences (see (3) and (6)) we have chosen four 5.8S sequences as different as possible and representative of evolutionarily distant species such as yeast (7), Drosophila (3), Xenopus (10) and HeLa cells (29). For control comparisons we have used the 5S sequences of the four

eu species described above (i.e. yeast (30), Drosophila (31), Xenopus somatic 5S RNA (32), HeLa cells (33)) and the E. coli 5S sequence (28).

Comparison procedures :

For comparison between the different 5.8S RNAs and the 5' 170 nucleotides of E. coli 23S RNA, the four sequences were aligned as described in Pavlakis et al. (3) and the bacterial sequence was then compared (Fig. 3).

Alignment of the different 5S molecules (Fig. 6) was done by first matching the GAAC or GAUC sequences at position 40-45 and then arranging the remaining sequences to obtain the maximum number of identities (13).

When necessary, deletions were introduced to maximize homology between different sequences. The percentage of homology was obtained by dividing the number of identical residues found in the same position in two sequences by the total number of compared residues (insertions were counted as residues). When two insertions were introduced at the same place in two different sequences (23S and 5.8S, pr 5S and eu 5S) to maximize homology with a third one, they were counted as identities (positions identified by a star in Fig. 3 and 6). When no insertion is introduced, two random sequences are expected to give 25 % homology by this method.

The percentage of insertions-deletions was obtained by dividing the number of insertions introduced in the eu sequence by the total number of compared residues. The number of insertions introduced in the bacterial sequence was not taken into account since it is considered as the reference

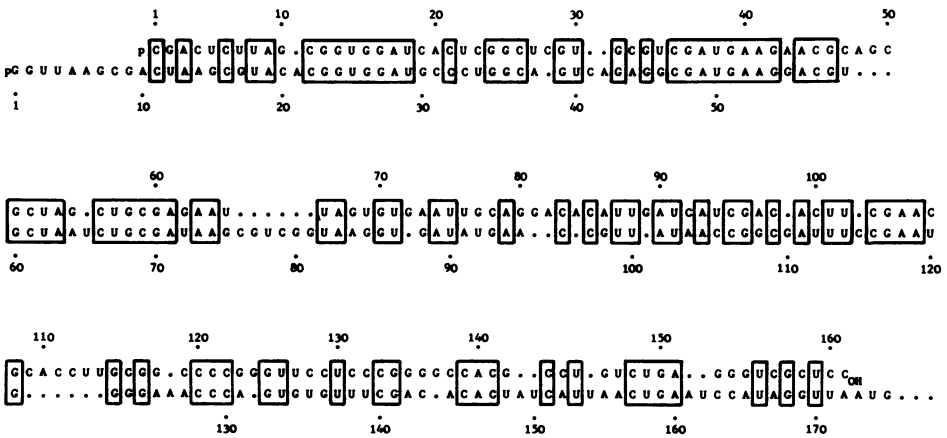

Figure 2 : Comparison of the nucleotide sequence of Xenopus laevis 5.8S RNA (upper line) with the 5' terminal sequence of E. coli 23S rRNA (lower line). Identical nucleotides in the two sequences are boxed.

in every composition test. This leads to an insertion percentage about half of that expected if the insertions-deletions in the bacterial sequence were also counted. However, our method points out more significantly the differences between eu sequences with respect to pr sequences.

Computer procedures :

Comparisons between different eu 5.8S RNA and the bacterial RTU sequence (or one of the subsequences derived from it, see above and Fig. 1) were carried out with the comparison program "CLARA" of J. Ninio (34). In its actual form, this program allows the comparison of two nucleic acid sequences of any size (up to 12 000 nucleotides each) and searchs for all strict identities $\geq$ 6 nucleotides wherever their location in the sequences. The program assigns an index to each hexamer and then determines in one pass all hexamers common to the two sequences. Finally, it makes pairwise comparisons to check whether or not the homology extends further in the sequences. The program was run either on a CDC 6600 computer (IN_2P_3, Université de Paris VI - Jussieu) or on an IRIS 80 (Centre de Calcul du Pharo, Université d'Aix-Marseille II).

The strict homologous sequences which were found were then plotted on a two dimensional matrix with the bacterial RTU subsequence on the horizontal line and the 5.8S sequence on the vertical one. On such a diagram, with two molecules of the same size displaying a high homology, the identical regions should appear as dots ranging on the main diagonal of the matrix. In the case of two molecules of different size, the dots corresponding to the homologous sequences will be aligned either on the main diagonal or on a parallel to it, whether these sequences are located at the same distance or at unequal distances respectively from the 5' ends of both molecules. In any event, only the dots which are strictly arranged on a line extending at a 45° angle will be considered.

It has to be noted that the program does not handle insertions and deletions. When three (or more) points are found aligned as described above, the corresponding sequences have to be aligned manually to check whether insertions have to be introduced.

The program "CLARA" was also used to make comparisons between the sequence of the bacterial RTU and random sequences and finally to search for possible base-paired regions in the 23S rRNA sequence.

RESULTS AND DISCUSSION

Is there a 5.8S equivalent in prokaryotes ?

As discussed in the introduction, the question of 5.8S evolutionary

origin is still a matter of debate. Comparisons of the known sequences of eukaryotic 5.8S RNAs reveal a large extent of conservation, even greater than that observed between eukaryotic 5S RNAs. This may indicate an important structural and/or functional role for 5.8S RNA (14). It is then intriguing that a rRNA molecule equivalent to eukaryotic 5.8S RNA has never been described in prokaryotes and this could be explained in two different ways :

(i) After divergence of prokaryotes and eukaryotes from a common hypothetical ancestor, the sequence of the rRNAs has dramatically diverged during evolution, such that it is no longer possible to detect clear homologies between eu 5.8S and its bacterial counterpart. However, because of the higher degree of sequence conservation between eu 5.8S rRNAs than between eu 5S RNAs, this hypothesis can probably be ruled out.

(ii) All attempts to find a 5.8S equivalent in prokaryotes have so far failed because a RNA molecule of approximately 160 nucleotides was searched for. A molecule with a 5.8S function may however exist in prokaryotes, but could either have a smaller size, or be covalently linked to another RNA component, thus displaying a greater apparent molecular weight than 5.8S RNA.

In the event the "bacterial 5.8S RNA" has a reduced size, say 70-80 nucleotides, it could be indistinguishable from the tRNA group, when analysed on polyacrylamide gels, and therefore could be overlooked.

Alternatively, in bacteria, 5.8S RNA could be part of a larger molecule, in the same manner that 5S RNA is linked to 16S and 23S RNAs within a 30S molecule in RNase III^- strains of E. coli (36). More simply, 5.8S RNA, though present in the bacterial ribosomal precursor could remain uncleaved from it, by contrast to what happens in eu cells. In these latter cells, it has clearly been shown that the sequence of 5.8S RNA is present in the direct precursor of 28S RNA close to its 5' end (37,38).

Assuming that a similar situation may occur in prokaryotes, we have compared the sequence of the 5' end of E. coli 23S RNA and its adjacent flanking region with different 5.8S sequences in order to look for eventual homologies.

Comparison between the 5' end of prokaryotic 23S RNA and eukaryotic 5.8S RNA sequences.

When the 5' end sequence of 23S rRNA from E. coli is compared to any 5.8S RNA sequence, it is clear that a portion of the 23S sequence, between nucleotides 13 and 170, shares a significant homology with 5.8S RNA. An example is shown in Fig. 2 where the sequences of Xenopus 5.8S RNA and of

the 5' end of E. coli 23S RNA are compared. The results of comparisons carried out between four different eu 5.8S RNAs and the 5' end of E. coli 23S RNA are shown in Fig. 3. The percentages of sequence homology presented in Table I were calculated from these alignements. These percentages are slightly lower than those obtained from the alignment of only two sequences (the homology calculated from Fig. 2 is 50.8 % whereas it is 47.7 % in Table I). They are however more significant since on one hand all comparisons have been made under the same alignment conditions and on the other hand, the number of insertions-deletions introduced is slightly lower : it has been previously suggested that alignments implying relatively few insertions seem more plausible than those implying many (35).

47-50 % homology is found with the different 5.8S sequences. When the comparisons are restricted to the 5' half of 5.8S RNA, this percentage rises to 55-60 % (Table I). As seen in Fig. 2 and 3, the region of homology

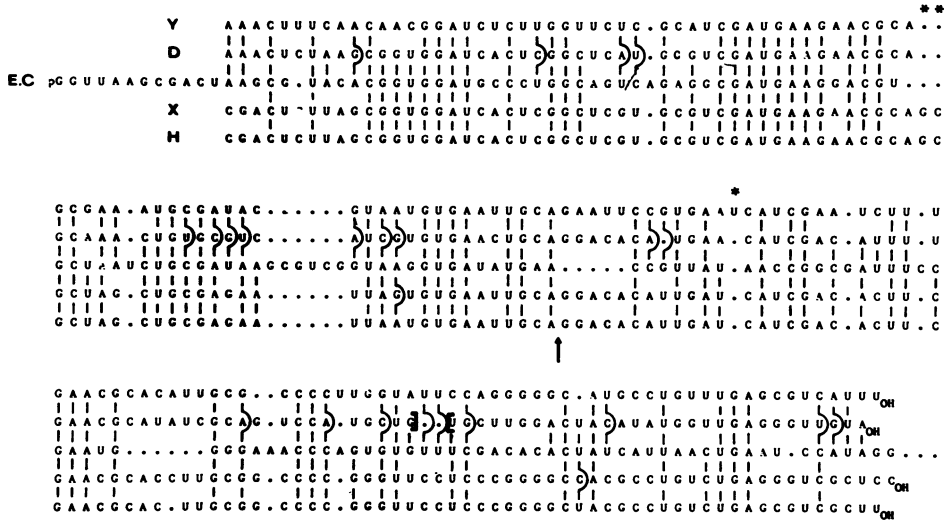

Figure 3 : Comparison of the 5' terminal sequence of E. coli 23S rRNA (middle line) with four eukaryotic 5.8S RNA sequences.

Y : yeast 5.8S (Saccharomyces cerevisiae), D : Drosophila melanogaster 5.8S + 2S, E.C : Escherichia coli 23S, X : Xenopus laevis 5.8S, H : HeLa cell 5.8S. The sequences were arranged as described in the comparison procedures section. Nucleotides common to one or more 5.8S sequences and to the 5' end of 23S RNA are joined by a | (or a) for drawing convenience). Identities between 5.8S RNAs only are not shown. Brackets indicate the 3' end of the Drosophila 5.8S sequence and the 5' end of the 2S sequence. The arrow separates the 5' and 3' halves of the 5.8S sequences. Positions denoted by a * : see the comparison procedures section.

Table I : Comparison of four eukaryotic 5.8S RNA sequences with the 5' end of E. coli 23S rRNA.

RNA	Number of residues compared (a)	Identical residues (b)	Insertions-deletions	% Insertions-deletions(c)	% Sequence homology	% Sequence homology in the 5' region (d)
Y 5.8S	173	86	15	8.7	49.7	59.8
D 5.8S + 2S	172	80	18	10.4	46.5	55.2
X 5.8S	174	83	13	7.5	47.7	55.2
H 5.8S	173	84	14	8.1	48.5	56.3
Y 5.8S/H 5.8S	173	125	14	8.1	72.3	74.7

Comparisons were made according to alignments shown in Fig. 3. The last line of the table is the comparison of two eukaryotic 5.8S RNA sequences. Abbreviations are as in Fig. 3.

(a) Insertions, when introduced, were counted as residues.

(b) Identical residues between the 23S sequence and each 5.8S sequence were counted. In some cases, denoted by a * in Fig. 3, two insertions were counted as an identity (see comparison procedures).

(c) Only insertions introduced in the eu sequences were counted (see comparison procedures).

(d) The end of the 5' region (the first half of 5.8S molecules) is shown by an arrow in Fig. 3.

emcompasses three consecutive blocks of 8, 8 and 6-7 nucleotides located in the 5' half of the 5.8S molecules. It has been previously noted that the 3' part of 5.8S has diverged faster than the 5' part during evolution (3). It is noteworthy that eu 5.8S RNA and the 5' end of pr 23S RNA also share a lower extent of homology in the 3' region.

Sequence homologies between the bacterial ribosomal transcription unit and eu 5.8S RNA.

The sequence homology described above, although interesting, may not be unique and other regions of the 23S RNA or of the bacterial ribosomal transcription unit may share equivalent or greater homologies with 5.8S RNA.

To check this point, different 5.8S sequences have been compared to an hybrid sequence constructed with different regions of the rrnB and rrnD E. coli operons which encompasses nearly all the transcribed region of an E. coli rRNA gene (see Sequence data). The degree of sequence heterogeneity between the 16S, 23S and 5S genes of different E. coli operons is very low (39,40,41) and apparently the same is true for most of the transcribed spacer sequences from operons carrying the same tRNA genes (23,25,27,42). It is therefore assumed that the sequence used is representative of a typical E. coli RTU.

The bacterial sequence was divided into eight consecutive subsequen-

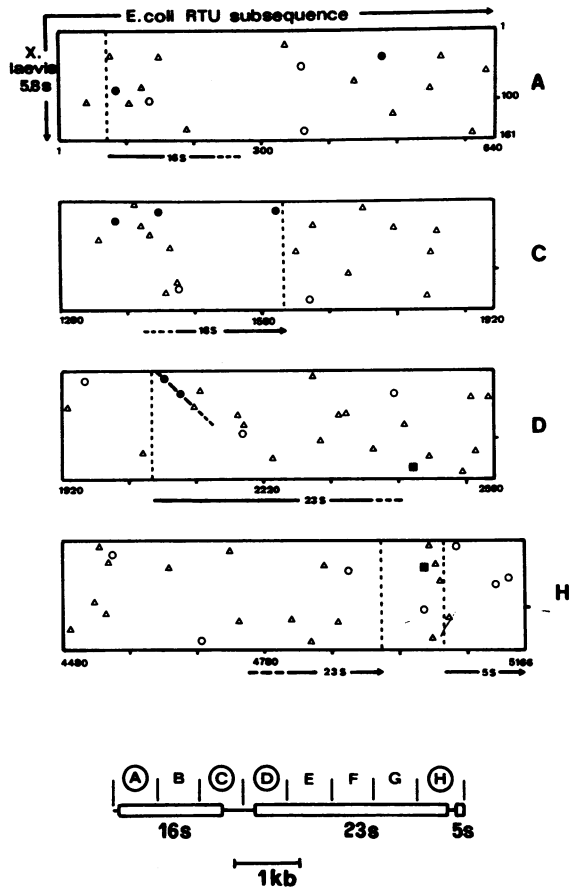

Figure 4 : Matrix representation of the homologies between 5.8S RNAs and the bacterial RTU.

Strict homologies (⩾ 6 nucleotides in length) between Xenopus laevis 5.8S RNA and different sections of E. coli ribosomal genes were detected by a computer analysis and plotted on a two-dimensional matrix as described in the text. The letters A,C,D,H refer to sections of the bacterial sequences shown at the bottom of the figure (see also Fig. 1) and the positions of the ribosomal genes are : 16S : 73-1613, 23S : 2051-4954, 5S : 5047-5166 (Fig. 1). On the matrix, the sequence of the E. coli RTU is on the horizontal line and the X. laevis 5.8S RNA is on the vertical line. The 5' ends of both sequences are located in the upper left corner of the matrix. Each region of homology between the two sequences is represented by a symbol whose abcissa and ordinate give its position in the E. coli RTU and X. laevis 5.8S RNA sequences respectively. Symbols are as follows :

△ : Strict homology of 6 nucleotides
○ : " " 7 "
● : " " 8 "
■ : " " 9 "

ces and each of them (A to H in Fig. 1) was then compared to a 5.8S sequence with the aid of the computer by using the comparison program "CLARA" (34).

Fig. 4 shows the results obtained with Xenopus laevis 5.8S RNA and 4 of the 8 subsequences from the bacterial RTU encompassing the 5' and 3' end of the 16S and 23S genes. As expected from the direct comparison, (see Fig. 2), in the case of the subsequence containing the 5' end of the 23S gene, three points aligned on a parallel to the diagonal and corresponding to homologies of 8, 8 and 6 nucleotides were found (Fig. 4D). However, due to the background caused by non-statistically significant homologies, some other regions appeared, in which three shorter homologous sequences (6 nucleotides) could be aligned in a similar way. In these cases, the sequences located between the homologous regions were manually aligned to detect smaller identical oligonucleotides (1-5 nucleotides) that the program did not take into account. Such comparisons, also carried out with the three other 5.8S sequences (Drosophila, Yeast, HeLa cell) and the different bacterial subsequences never led to an extent of homology approaching that shown in Fig. 2 even when large insertions-deletions (> 10 nucleotides) were introduced (data

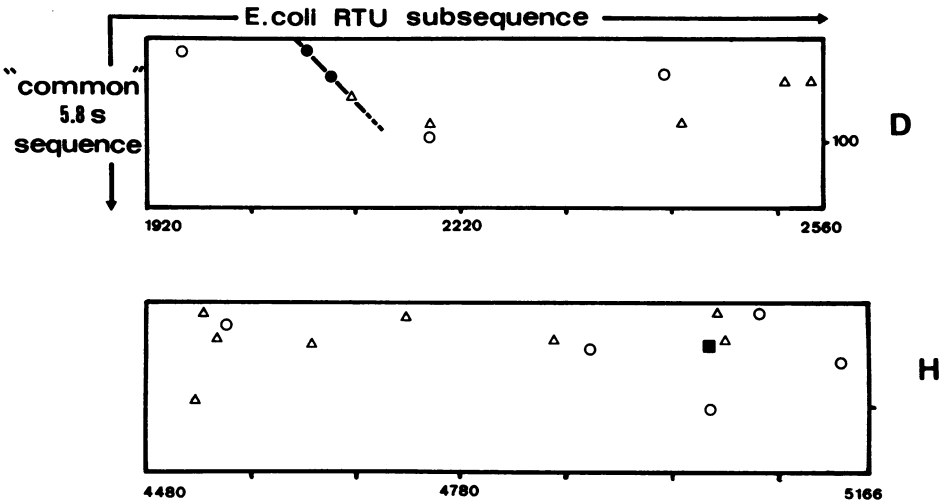

Figure 5 : "Filtered" matrix representation of the homologies between different 5.8S RNA, sequences and the bacterial RTU.

Regions D and H of the bacterial RTU were compared to each of the four 5.8S sequences (see Sequence data) with the aid of the computer. Homologous oligonucleotides (length ≥ 6) with identical sequences and positions present in at least 3 comparison tests are represented. Symbols are as in Fig. 4.

not shown).

In order to "filter out" the background, a given bacterial subsequence was successively compared to each of the four 5.8S sequences with the aid of the computer. Only oligonucleotides with identical sequences and positions which were found at least in three out of four comparison tests were taken into account. They were then plotted on a two dimensional matrix as described above. The regions A, C, D and H of the bacterial RNA sequence were examined and Fig. 5 depicts the results obtained with the subsequences D and H (respectively, the 5' and 3' end of the 23S gene). The homologies between the 5' end of 23S RNA and 5.8S RNA appear clearly by comparing Fig. 4 and 5 (D and H in both) and this was the only case where an alignement of three blocks of homologous residues could be found.

We have compared the sequence of the bacterial RTU with different 5.8S RNA sequences read in the 3'→5' sense. These sequences, having exactly the same nucleotide composition and length than 5.8S RNAs, but a different primary structure, can be considered as random sequences. The comparisons which were made with all eight E. coli subsequences never gave an homology as important as that in Fig. 4D (Data not shown).

Control comparisons with eukaryotic and prokaryotic 5S RNA sequences.

The homology between the 5' end of E. coli 23S RNA and 5.8S RNAs from four different eu species was compared to that between E. coli 5S RNA and 5S RNAs from the same four eukaryotic species.

5S RNA sequences were aligned for maximal homologies (see comparison procedures) and the percentages of homologies, calculated as in Table I, are reported in Fig. 6 and Table II.

The percentage of homology between the first half of 5.8S RNA and the corresponding region of 23S RNA is significantly higher than that found between 5S RNAs (55-59 % instead of 51-53 %, Tables I and II), and still falls within the same range when the whole sequence of 5.8S is considered (47-50 %, Table I). The percentage of insertions-deletions which were introduced was similar in both cases.

Secondary structure models for the 5' end of 23S RNA.

Different secondary structure models for isolated 5.8S RNA have been proposed (7,8,9,10) and it was tempting to check whether the 5' region of the 23S RNA chain could be folded according to one of these models.

All our numerous attempts have failed to fit with one of the models mentioned above for 5.8S RNA. However, it must be kept in mind that, in the ribosome, 5.8S RNA is always hydrogen-bonded to 28S RNA. In this respect, it

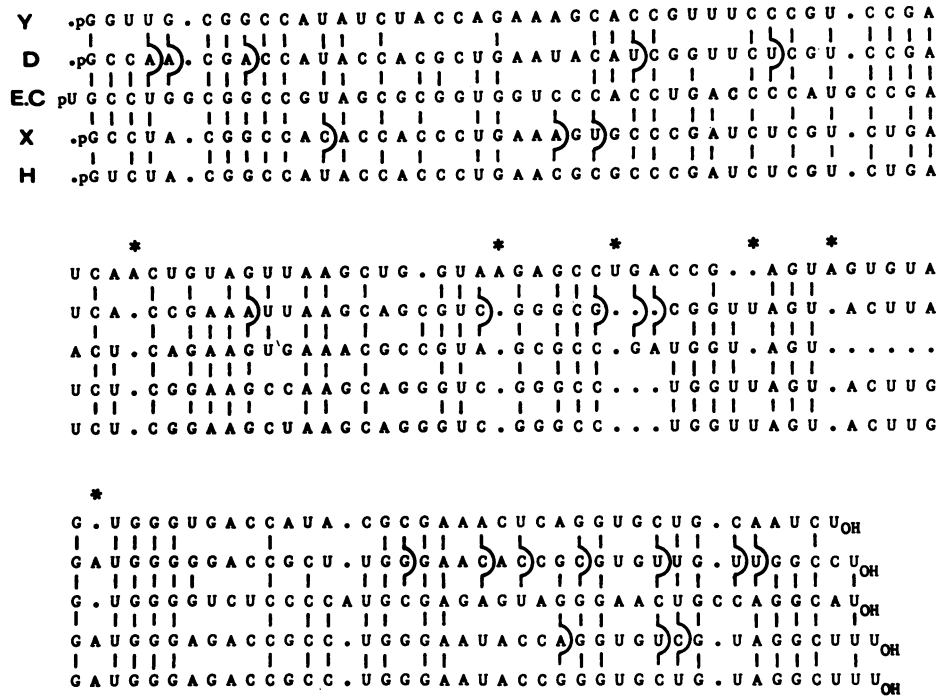

Figure 6 : Comparison of the E. coli 5S RNA sequence (middle line) with four eukaryotic 5S sequences.
Conventions and abbreviations as in Fig. 3.

has been proposed that the 3' end of the 5.8S molecule (43,44,3) and possibly its 5' end (3,45) are hydrogen-bonded to 28S RNA .

With the aid of the computer, local and long-range interactions of the 5' end of 23S RNA either with itself or with the whole 23S molecule have been searched for, by pairing together complementary regions of six or more consecutive nucleotides and then looking for smaller regions of complementarity. This led to an important number of models and only two of the most stable models are presented in Fig. 7, and will be discussed below.

. Model A exhibits some interesting features :

- On the basis of a computer search for long complementarities, the 5' first 50 nucleotides of 23S RNA have been arranged to form 3 helical regions with other parts of the 23S molecule (A,B,C in model A). This is in good agreement with previous results. The stem between the 5' and 3' end of the molecule (region A of the model) has been suggested by Branlant et al.

Table II : Comparison of four eukaryotic 5S RNA sequences with the E. coli 5S RNA sequence.

RNA	Number of residues compared (a)	Identical residues (b)	Insertions-deletions(c)	% Insertions-deletions	% Sequence homology
Y 5S	131	67	10	7.6	51.2
D 5S	131	70	11	8.4	53.4
X 5S	132	71	11	8.3	53.8
H 5S	132	70	11	8.3	53
Y 5S/H 5S	131	79	10	7.6	60.3

The four eukaryotic species are the same as in the 5.8S RNA comparisons to give an internal control. Comparisons were made according to the alignment shown in Fig. 6. The last line of the table is the comparison of two eukaryotic 5S RNA sequences (see Fig. 3 for abbreviations).

(a) Same as in Table I.

(b) Identical residues between the E. coli 5S sequence and each eukaryotic 5S sequence were counted. In some cases (denoted by a * in Fig. 6) two insertions were counted as an identity (see comparison procedures).

(c) Same as Table I.

(46). RNA fragments encompassing the helical regions B and C of the model have previously been found associated within a complex between 23S RNA and protein L24 which is resistant to ribonuclease T1 digestion. Moreover our pairing scheme is in complete agreement with that proposed by Krol et al. for this same region (47).

- The 58-176 region of 23S RNA can be folded into an internal secondary structure involving five stable helical regions (D,E,F,G,H in model A). It must be emphasized that none of the proposed pairing scheme for this region fits with corresponding 5.8S secondary structure models (data not shown).
- Highly conserved sequences in RNA molecules are very often unpaired or only partially paired (48). This is also the case for the three regions of 23S RNA sharing important homologies with 5.8S RNAs (boxed areas in Fig. 7A).
- Two kethoxal-reactive sites (corresponding to single-stranded guanines accessible to chemical modification in 50S subunits) are present in the 5' 200 first nucleotides of the 23S sequence (49,26). In model A, these residues, G_{93} and G_{141} (identified by a letter K in Fig. 7A) are found in hairpin loops, suggesting their accessibility to chemical modification.

. The most striking feature of model B (Fig. 7) is a very long interaction between two widely separated regions of the 23S RNA molecule. It may be interesting to point out that this possible base-pairing involves a

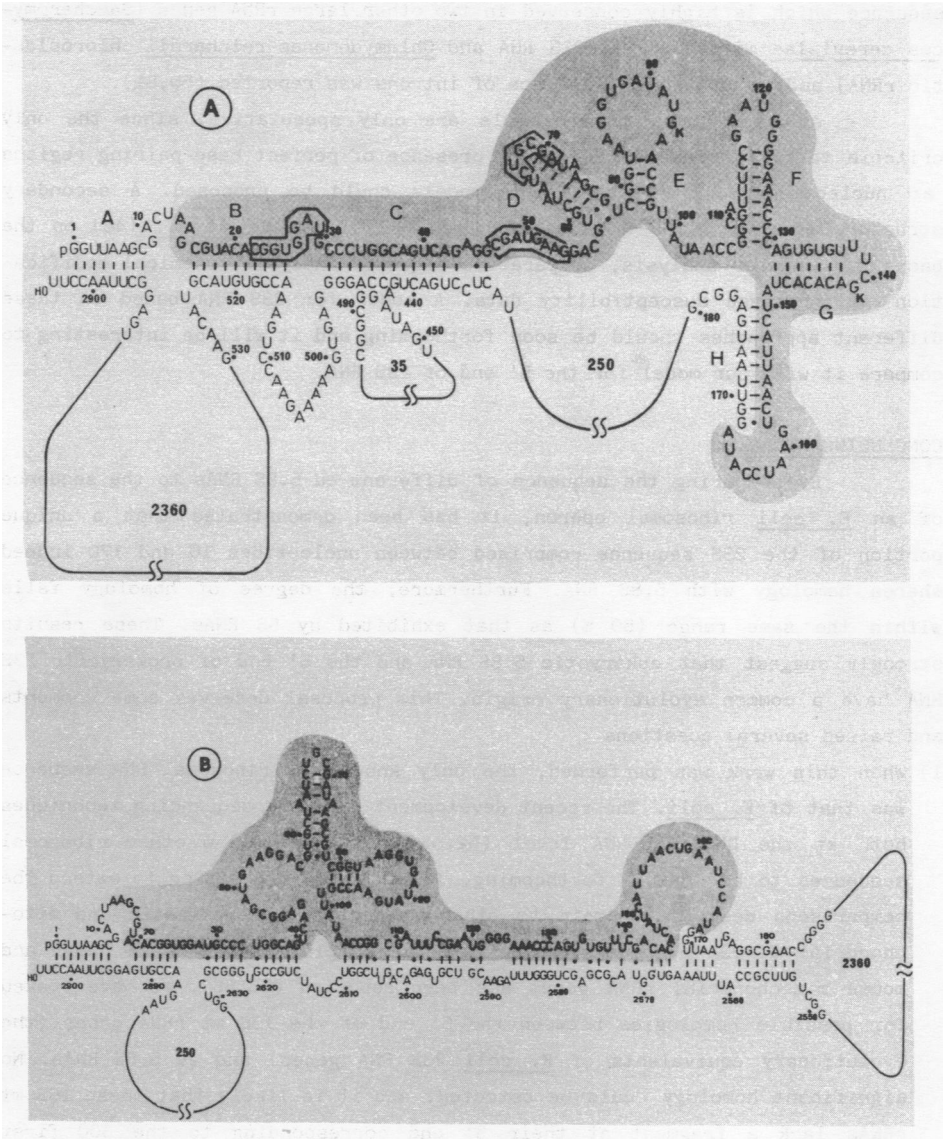

Figure 7 : Two possible secondary structure models for the 5' region of E. coli 23S RNA.

Numbering of the RNA chain is that of Brosius et al. (25). The bold-faced numbers represent the length of loops, the sequence of which is not shown.

G-C and A-U pairs are joined by a | and G-U pairs by a • . The region of the 23S RNA chain which is compared to 5.8S RNA in Fig. 2 and 3 is shaded and regions of strong homology are boxed. Guanines residues modified by kethoxal in active 50S subunits are identified by the letter K.

sequence which is highly conserved in two other large rRNA genes (Saccharomyces cerevisiae mitochondrial 21S RNA and Chlamydomonas reinhardii chloroplastic rRNA) and in which the existence of introns was reported (50,51).

As said above these models are only speculative, since the only criteria taken into account were the presence of perfect base pairing regions ≥ 6 nucleotides, and several other models could be proposed. A secondary structure model for 16S RNA has been proposed by Woese et al. (48) on the basis of computer analysis, comparative sequence analysis, chemical modification and nuclease susceptibility data. A model for 23S RNA based on these different approaches should be soon forthcoming and it will be interesting to compare it with our model for the 5' end of 23S RNA.

CONCLUDING REMARKS.

By comparing the sequence of different eu 5.8S RNAs to the sequence of an E. coli ribosomal operon, it has been demonstrated that a unique portion of the 23S sequence comprised between nucleotides 10 and 170 indeed shares homology with 5.8S RNA. Furthermore, the degree of homology falls within the same range (50 %) as that exhibited by 5S RNAs. These results strongly suggest that eukaryotic 5.8S RNA and the 5' end of prokaryotic 23S RNA have a common evolutionary origin. This proposal deserves some comments and raises several questions :

1) When this work was performed, the only known 23S ribosomal RNA sequence was that of E. coli. The recent development of rapid sequencing techniques both at the DNA and RNA level (52,53,54) should allow other ribosomal sequences to be rapidly forthcoming. It would be interesting to extend the comparisons described herein to other bacterial, chloroplastic and mitochondrial 21-23S sequences. Recently, the complete sequence of human and mouse mitochondrial rRNA genes has been reported (55,56). We have looked for possible homologies between the 5' end of the 16S mt rRNA genes (the evolutionary equivalents of E. coli 23S RNA genes) and eu 5.8S RNAs. No significant homology could be detected, and it is likely that these 16S mt rRNAs lack a fragment at their 5' end corresponding to the 300 first nucleotides of E. coli 23S RNA (56).
2) A high degree of sequence conservation throughout evolution is usually related to an important structural and/or functional role. In this respect, it has been proposed that the 3' end of pr 16S RNA and eu 18S RNA which share important homologies, may perform the same function (57,58). The same might be true for eu 5.8S RNAs and the 5' end of pr 23S RNA.

Structural comparisons between 5.8S-28S and 23S secondary structure models as well as studies of 5.8S/proteins interaction should lead to a better understanding of the function of 5.8S RNA which is still largely unknown.

3) The homology between the 5' end of pr 23S RNA and eu 5.8S RNA suggests that the bacterial counterpart of eu 5.8S RNA is part of the 23S molecule, whereas in eukaryotic cells, a 200-300 bp transcribed spacer (removed during processing of the precursor) separates the 5.8S sequence from the 5' end of 28S RNA (Fig. 8). Different molecular mechanisms which might have occured after the divergence of prokaryotes and eukaryotes, and led to a different organization near the 5' end of the large rRNA gene can be proposed :

- A deletion in the large rRNA gene of prokaryotes could have removed a non functional region (or alternatively an internal spacer) near the 5' end of the gene.
- An insertion in the large rRNA gene of eukaryotes could have separated its 5' end from the rest of the molecule.
- A duplication of the 5' end of the eukaryotic large rRNA gene could have been followed by an insertion event between the two duplicated regions.

Comparisons between the 5' ends of prokaryotic 23S RNA and eukaryotic 28S RNA sequences should help to choose between these different hypotheses.

While this work was being completed, Hall and Maden (60) reported the sequence of the 18S-28S intergenic spacer region of Xenopus laevis, extending from the 3' end of the 18S gene to the beginning of the 28S gene. We have compared the sequence of the 120 first nucleotides of Xenopus 28S RNA

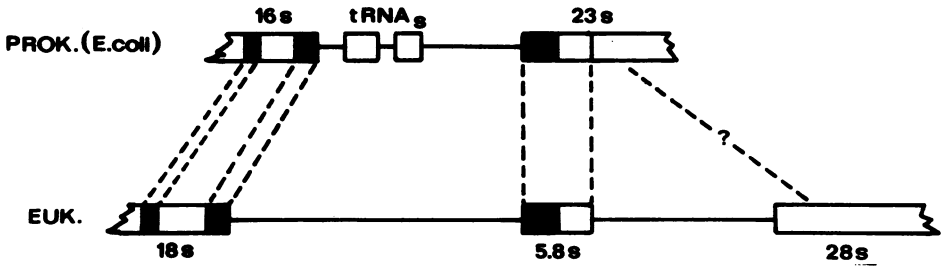

Figure 8 : Schematic comparison of the ribosomal internal spacer regions in eukaryotes and prokaryotes.

Black areas are regions where sequence homology is $\geq$ 50 %. The regions of homology in the last 200 nucleotides of 16S and 18S RNA were taken from B.R. Jordan et al. (59).

with the 5' end of E. coli 23S RNA. As expected from our work, no homology was found with the 170 first nucleotides of 23S RNA. However, when nucleotides 177-275 of E. coli 23S RNA are aligned with nucleotides 21-118 of Xenopus 28S RNA, the two sequences display very strong homology (73 %). Preliminary results indicate that a similar situation occurs in Drosophila (B. Jacq, in preparation). These observations fit very well our finding of homology between nucleotides 10-173 of E. coli 23S RNA and 5.8S RNA, and suggests that in eukaryotes an insertion event has separated the 5' end of the large rRNA from the rest of the molecule.

A preliminary report of this work has been presented at the EMBO workshop on "Molecular biology of Drosophila" held at Kolymbari (Crète) in August 1980.

ABBREVIATIONS :

eu : eukaryotic, mt : mitochondrial, pr : prokaryotic, rrn operon : ribosomal ribonucleic operon, RTU : ribosomal transcription unit, R : A or G, Y : C or T.

ACKNOWLEDGMENTS :

I would like to thank M. Demeestere and A.J. Valleron (Université de Paris VII) for their help in the initiation of the computer work. I am very grateful to J. Ninio (IRBM, Université de Paris VI) for his stimulating discussions and for making his comparison program "CLARA" available to me. I am indebted to B.R. Jordan, M. Latil, M. Laval and R. Rosset for helpful ideas and critical discussions and to D. Gratecos for her help in the preparation of the manuscript. I also thank B. Beurton for the verification of the sequences and the typing of the manuscript.

This work was supported by grants from the "Centre National de la Recherche Scientifique" (ATP CNRS 4257) and from the "Délégation Générale à la Recherche Scientifique et Technique" (contrat DGRST 78.7.0338).

REFERENCES

1. Pene, J.J., Knight, E. and Darnell, J.E. (1968) J. Mol. Biol. 33, 609-623
2. Jordan, B.R., Jourdan, R. and Jacq, B. (1976) J. Mol. Biol. 101, 85-105.
3. Pavlakis, G.N., Jordan, B.R., Wurst, R. and Vournakis, J.N. (1979) Nucl. Acids Res. 7, 2213-2238.
4. Jordan, B.R., Latil-Damotte, M. and Jourdan, R. (1980) Nucl. Acids Res. 8, 3565-3573.
5. Renkawitz, R. and Gerbi, S.A. (1979) Molec. Gen. Genet. 173, 1-13.
6. Erdmann, V.A. (1980) Nucl. Acids Res. 8, r31-r47.

7. Rubin, G.M. (1973) J. Biol. Chem. 248, 3860-3875.
8. Nazar, R.N., Sitz, T.O. and Busch, H. (1975) J. Biol. Chem. 250, 8591-8597.
9. Luoma, G.A. and Marshall, A.G. (1978) Proc. Natl. Acad. Sci. (USA) 75, 4901-4905.
10. Ford, P.J. and Mathieson, T. (1978) Eur. J. Biochem. 87, 199-214.
11. Nazar, R.N. and Roy, K.L. (1978) J. Biol. Chem. 253, 395-399.
12. Brownlee, G.G. and Cartwright, E. (1975) Nucl. Acids Res. 2, 2279-2288.
13. Hori, H. (1975) J. Mol. Evol. 7, 75-86.
14. Cedergren, R.J. and Sankoff, D. (1976) Nature 260, 74-75.
15. Doolittle, W.F. and Pace, N.R. (1971) Proc. Natl. Acad. Sci. (USA) 68, 1786-1790.
16. Erdmann, V.A. (1975) in Progress in nucleic acid research and molecular biology, Cohn, W.E. Editor, Vol. 18, pp. 45-90, Academic Press, New-York.
17. Price, R. and Penman, S. (1972) J. Mol. Biol. 70, 435-450.
18. Wrede, P. and Erdmann, V.A. (1973) FEBS Lett. 33, 315-319.
19. Bellemare, G., Vigne, R. and Jordan, B.R. (1973) Biochimie 55, 29-35.
20. Wrede, P. and Erdmann, V.A. (1977) Proc. Natl. Acad. Sci. (USA) 74, 2706-2709.
21. Azad, A.A. and Lane, B.G. (1973) Can. J. Biochem. 51, 1669-1672.
22. Kiss, A., Sain, B. and Venetianer, P. (1977) FEBS Lett. 79, 77-79.
23. Young, R.A. and Steitz, J.A. (1979) Cell 17, 225-234.
24. Brosius, J., Palmer, M.L., Kennedy, P.J. and Noller, H.F. (1978) Proc. Natl. Acad. Sci. (USA) 75, 4801-4805.
25. Young, R.A., Macklis, R. and Steitz, J.A. (1979) J. Biol. Chem. 254, 3264-3271.
26. Brosius, J., Dull, T.J. and Noller, H.F. (1980) Proc. Natl. Acad. Sci. (USA) 77, 201-204.
27. Bram, R.J., Young, R.A. and Steitz, J.A. (1980) Cell 19, 393-401.
28. Brownlee, G.G., Sanger, F. and Barrell, B.G. (1967) Nature 215, 735-736.
29. Khan, M.S.N. and Maden, B.E.H. (1977) Nucl. Acids Res. 4, 2495-2505.
30. Miyazaki, M. (1974) J. Biochem. 75, 1407-1407.
31. Benhamou, J. and Jordan, B.R. (1976) FEBS Lett. 62, 146-149.
32. Brownlee, G.G., Cartwright, E., Mac Shane, T. and Williamson, R. (1972) FEBS Lett. 25, 8-12.
33. Hatlen, L.E., Amaldi, F. and Attardi, G. (1969) Biochemistry 8, 4989-5005
34. J. Ninio, unpublished.
35. Sankoff, D., Cedergren, R.J., Lapalme, G. (1976) J. Mol. Evol. 7, 133-149.
36. Ginsburg, D. and Steitz, J.A. (1975) J. Biol. Chem. 250, 5647-5654.
37. Maden, B.E.H. and Robertson, J.S. (1974) J. Mol. Biol. 87, 227-235.
38. Speirs, J. and Birnstiel, M. (1974) J. Mol. Biol. 87, 237-256.
39. Ehresmann, C., Stiegler, P., Carbon, P. and Ebel, J.P. (1977) FEBS Lett. 84, 337-341.
40. Branlant, C., Krol, A., Machatt, M.A. and Ebel, J.P. (1979) FEBS Lett. 107, 177-181.
41. Jarry, B. and Rosset, R. (1971) Molec. Gen. Genet. 113, 43-50.
42. De Boer, H.A., Gilbert, S.F. and Nomura, M. (1979) Cell 17, 201-209.
43. Pace, N.R., Walker, T.A. and Schroeder, E. (1977) Biochemistry 16, 5321-5328.
44. Boseley, P.G., Tuyns, A., Birnstiel, M.L. (1978) Nucl. Acids Res. 5, 1121-1137.
45. Nazar, R.N. and Sitz, T.O. (1980) FEBS Lett. 115, 71-76.
46. Branlant, C., Sri Widada, J., Krol, A. and Ebel, J.P. (1980) Nucl. Acids Res. 3, 1671-1687.
47. Krol, A., Machatt, M.A., Branlant, C. and Ebel, J.P. (1978) Nucl. Acids Res. 5, 4933-4947.

48. 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. and Noller, H.F. (1980) Nucl. Acids Res. 8, 2275-2293.
49. Herr, W. and Noller, H.F. (1975) FEBS Lett. 53, 248-252.
50. Dujon, B. (1980) Cell 20, 185-197.
51. Allet, B. and Rochaix, J.D. (1979) Cell 18, 55-60.
52. Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. (USA) 74, 560-564.
53. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. (USA) 74, 5463-5467.
54. Donis-Keller, H., Maxam, A. and Gilbert, W. (1977) Nucl. Acids Res. 4, 2527-2538.
55. Eperon, I.C., Anderson, S. and Nierlich, D.P. (1980) Nature 286, 460-466.
56. Van Etten, R.A., Walberg, M.W. and Clayton, D.A. (1980) Cell 22, 157-170.
57. Shine, J. and Dalgarno, L. (1974) Biochem. J. 141, 609-615.
58. Hagenbüchle, O., Santer, M., Steitz, J.A. and Mans, R.J. (1978) Cell 13, 551-563.
59. Jordan, B.R., Latil-Damotte, M. and Jourdan, R. (1980) FEBS Lett. 117, 227-231.
60. Hall, M.C.H. and Maden, B.E.H. (1980) Nucl. Acids Res. 8, 5993-6005.
61. Nazar, R.N. (1980) FEBS Lett. 119, 212-214.