Inter-RNA Homology and Possible Roles of Small RNAs

Takashi Gojobori and Masatoshi Nei

Center for Demographic and Population Genetics, The University of Texas at Houston, PO Box 20334, Houston, Texas 77025, USA

Summary. The nucleotide sequence of a segment of U1 and U3b small RNAs (sRNAs) is shown to have a high complementarity with the nucleotide sequence of a part of the leader region of almost all eukaryotic genes studied so far. The complementary region of U3b is located in the unpaired segment of the secondary structure of U3b constructed by Reddy et al. (1979). A similar complementarity is also observed between these RNAs and the leader regions of eukaryotic viruses, but the complementary region is not always identical with that for eukaryotic genes. Complementarity is also observed between the 3' end of 18S rRNA and a segment of U1 or U3b which is almost contiguous to the region complementary with mRNA. These observations suggest that U1 and U3b may be involved in mRNA processing and transport in the nucleus or in translation in the cytoplasm. In addition to U1 and U3b, another sRNA, i.e., 4.5S RNA1, is shown to have segments which are homologous to the Hogness box of the flanking region of gene and the Proudfoot-Brownlee (PB) box of mRNA near the poly(A) attachment site. The two segments which are complementary with these boxes are located almost contiguously on a co-joined loop of the secondary structure of 4.5S RNA1 constructed by Ro-Choi et al. (1972). Since the Hogness box and PB box are both considered as a recognition site by the RNA polymerase, it is possible that 4.5S RNA1 is involved in mediating gene transcription.

Key words: Small RNAs — mRNA — 18S rRNA — Sequence homology — Transcription — mRNA processing

Introduction

Discrete, stable, small RNAs (sRNAs) are ubiquitous in eukaryotic cells (Weinberg and Penman 1968; Helling-Larsen and Frederiksen 1977). They are concentrated in the nucleus but seem to exist also in the cytoplasm (Zieve and Penman 1976; Jelinek and Leinwand 1978; Kano et al. 1978; Flytzanis et al. 1978). Although several species of sRNAs have been isolated, and U1, U2, U3b, and 4.5S sRNAs from rat hepatoma nuclei have been sequenced (Reddy et al. 1974; Ro-Choi and Henning 1977; Shibata et al. 1975; Reddy et al. 1979; Ro-Choi et al. 1972; Branlant et al. 1980), their functions are not well understood. Recently, one of them, i.e., U1, was shown to have a segment (near the 5' end) that is complementary with the splicing site (exon-intron junctions) of precursor mRNA, and this finding has led to the suggestion that U1 sRNA is the recognition site component of the nuclear RNA splicing enzyme (Lerner and Steitz 1979; Lerner et al. 1980; Rogers and Wall 1980; Ting et al. 1980).

With the aim of finding any other involvement of these sRNAs in the processing of mRNA or transcription, we examined the complementarity of their sequences with mRNA, rRNA, and the flanking regions of genes. This examination resulted in two interesting findings which suggest that sRNAs may in fact function as a mediator for various processes of transcription and translation. The purpose of this paper is to report these findings.

Homology Between sRNA and the Leader Region of mRNA

Our first finding is the homology between U1 and U3b sRNAs and the leader region of mRNA in eukaryotes.
In this study we examined the nucleotide sequence of 28 eukaryotic mRNAs reported so far (Fig. 1). The nucleotide sequence of U1 is that of Branlant et al. (1980). It is seen that the nucleotide sequence at positions 143 to 147 of U1 is identical with that of positions 58 to 62 of U3b (boxed in Fig. 1) and this sequence has a high complementarity with a segment of the leader region of mRNA in almost all genes examined. It is clear that the nucleotide sequence at this region of mRNA has been conserved in the evolutionary process in various genes in eukaryotes. The consensus sequence of mRNAs or the sequence that fits the complementary sequence in U1 and U3b optimally is 5' CAGUC3', which is identical with the sequence of rabbit α globin, rat insulin II, sea urchin histone IV, chick ovalbumin and mouse κ 41 immunoglobulin light chain. However, the number and positions of complementary base pairs vary with mRNA, and in many mRNAs even the neighboring regions have some complementarity with U1 and U3b. If we take into account the neighboring regions, some mRNAs (e.g., those for rabbit β globin and chicken conalbumin) have a higher complementarity with U1 than with U3b, whereas others (e.g., those of rat insulin II and yeast actin) have a higher
complementarity with U3b. The homology region of U1 is different from the region (near the 5' end) that is complementary with the splicing site of intron (Lerner and Steitz 1979; Lerner et al. 1980; Rogers and Wall 1980; Ting et al. 1980).

Theoretically, it is possible that two short segments of nucleotide sequences become identical by chance. However, the similarity mentioned above is unlikely to have occurred by chance alone, since all the genes examined display it though the degree of similarity varies with the gene. We believe that this similarity has been evolutionarily conserved because of the existence of some interaction between sRNA and mRNA. A rough idea about the stability of the complementary base pairing can be obtained by computing the base-pairing energy. We computed this energy by using Salser's (1977) method. The results obtained are presented in the last two columns of Fig. 1. The energy computed varies considerably with the gene and sRNA but suggests that most pairings are apparently quite stable. It is noted that except in six out of 28 genes examined, the base pairing energy for one of U1 and U3b sRNAs is less than $-8$ Kcal. Lomedico et al. (1979) computed the same energy for a complementary base-pairing between the leader region of mRNAs and a 3' end region of 18S rRNAs for five eukaryotic genes (see below) and obtained an average of $-8.8$ Kcal. Therefore, in most genes the base-pairing between sRNA and mRNA seems to be equally or even more stable than the similar base-pairing between mRNA and 18S rRNA.

It is known that in the leader regions of mRNA of rabbit $\alpha$- and $\beta$-globins, ovalbumin, fibroin, and preproinsulin, a secondary structure with a stem and a loop can be constructed, and one side of the loop has a region which is complementary with a 3' end region of 18S rRNAs (Salser 1977; Baralle and Brownlee 1978; Hagenbuchle et al. 1978; Lomedico et al. 1979). This secondary structure exists always in the region of mRNA between the cap site and the initiation codon AUG. Our examination of the nucleotide sequences of the other mRNAs has shown that most of them can form a similar secondary structure with a region complementary to 18S rRNA. Interestingly, this secondary structure has a close relationship with the region complementary with U1 and U3b. Namely, the latter complementary region is located on the same loop of the secondary structure but opposite to the side where the complementarity with 18S rRNA exists. Furthermore, there is complementarity between sRNA and 18S rRNA, so that a tripartite RNA complex may be formed (Fig. 3). This possible tripartite complex was observed for all mRNAs, where the secondary structure could be constructed. However, the degree of complementarity among the three RNAs involved varied from gene to gene, and in some cases the complementarity between sRNA and 18S rRNA appeared rather weak. Interestingly, the complementary region in U3b exists in the unpaired region (positions 56 to 70) of the secondary structure of this RNA constructed by Reddy et al. (1979) (see their Fig. 10). Unfortunately, the secondary structure of U1 has not been constructed. We have also noted that the complementary region of 18S rRNA has a homology with the 3' end region of 16S rRNA in prokaryotes, so that this region has been conserved in the evolutionary process (Hagenbuchle et al. 1978).

In addition to the eukaryotic genes mentioned above we also examined the leader regions of nine mRNAs transcribed by eukaryotic viruses. The leader regions of about a half of these mRNAs again showed complementarity with the same region of U1 and U3b. The other mRNAs did not show complementarity at this region, but complementarity was observed at the neighboring regions (Fig. 2). Furthermore, when a secondary structure can be constructed in the leader region of mRNA as in the cases of SV40 VPI and HBcAG, the same tripartite RNA complex as that for eukaryotic genes was observed (Fig. 3). However, since the complementary regions are not necessarily concentrated in the same segment of U1 and U3b as in the case of eukaryotic genes, the possibility that they occurred by

![Fig. 2.](image)

**Fig. 2.** Complementarity between sRNAs (U1 and U3b) and the leader regions of eukaryotic viral mRNAs. Some of these mRNAs have a complementary region (dotted and solid lines) outside the boxed sequence of sRNAs. Data source: 1, 3 -9 (Baralle et al. 1978), 2 (Pasek et al. 1979).
Fig. 3. Tripartite RNA complex formed by mRNA, 18S rRNA, and sRNA (U1 and U3b). This complex is formed always in the region between the cap site (double underlines) and the initiation codon (single underline) of mRNA. Essentially the same complex can be made for most of the eukaryotic genes examined and several eukaryotic viral genes.

Homology Between sRNA and the Flanking Regions of Genes

Our second finding is the homology between the 4.5S sRNA1 and the flanking gene regions. It is known that the 5' flanking regions of eukaryotic genes have the so-called Hogness box (van Ooyen et al. 1979), which consists of a sequence of TATAAG or its minor variation and is located 23 ± 1 nucleotides upstream from the cap site. This box has been conserved evolutionarily and is considered as a recognition site by the RNA polymerase. Deletion of this box does not abolish transcription but generates novel mRNA 5' termini (Grosschedl and Birnstiel 1980). Near the end of the 3'
non-coding region of mRNA there is another segment which has a highly conserved nucleotide sequence of 5' AATAAAA 3' (Proudfoot and Brownlee 1976). It is located about 10 ~ 20 nucleotides upstream from the poly(A) attachment site. This is again regarded as a recognition site. We call it the Proudfoot-Brownlee box (PB box), using the discovery of the discoverers. When we examined the nucleotide sequence of 4.5S snRNA1 (Ro-Choi et al. 1972), we discovered that one segment of this RNA is homologous with the Hogness box, the sequence being 5' UAAUAG 3' at positions 49 to 54 (Fig. 4). Obviously, this is complementary with the other strand of DNA at the Hogness box region. Furthermore, this segment is located on one side of a co-joined loop of the secondary structure of this RNA constructed by Ro-Choi et al. (1972). Interestingly enough, we found that the other side of the co-joined loop has a sequence (5' AAUAA 3' at positions 40 to 44), which is identical with the first five of the six bases of the PB box when U is replaced by T.

Discussion

Earlier authors (Jelinek and Leinwand 1978; Flytzanis et al. 1978; Ro-Choi et al. 1977; Busch 1980) have suggested that snRNAs have a role in processing and transporting mRNAs and ribosomes, since they are concentrated in the nucleus and generally exist associated with ribonucleoproteins. If the tripartite RNA complex is really formed, it would strengthen this view. Of course, formation of the tripartite structure is not a requirement for involvement of snRNAs in the RNA processing and transport. It is possible that base pairing occurs only between two types of RNAs at a time yet this may still play an important role. In addition to RNA processing, snRNAs may also be involved in translation. Particularly, if the tripartite structure is really formed, it is possible that U1 and U3b have a function to stabilize mRNA and modulate translation. If this is true, the complementarity of 18S rRNA with the leader region of mRNA and sRNA may be useful for recognition of the
translation initiation site by ribosomes. We note that some sRNAs are hydrogen-bonded to hnRNA (Flytzanis et al. 1978) and poly(A)-terminating nuclear and cytoplasmic RNA (Jelinek and Leinwand 1978). However, the second possibility is weaker than the first, since sRNAs apparently exist in a high concentration in the nucleus rather than in the cytoplasm.

The homology between 4.5S sRNA1 and the flanking regions of eukaryotic genes suggests that 4.5S sRNA1 may be involved in transcription. It may help the two strands of DNA to separate at the position of the Hoggness box or it may stabilize the separated strands, so that the RNA polymerase starts to function. Of course, the actual transcription starts from the cap site rather than from the Hoggness box, which is 23 ± 1 nucleotides upstream from the cap site, but this distance may be required because of the size of the RNA polymerase involved. On the other hand, the homology of the AAUAAA sequence of 4.5S sRNA1 with the PB box may be important for terminating transcription. It is possible that 4.5S sRNA1 is bound to the RNA polymerase and moves along the DNA sequence in the transcriptional process. Another possibility is that at the time of transcription DNA forms a loop and the Hoggness box and the PB box are connected through 4.5S sRNA1, which is possibly bound to some proteins. We note that 4.5S sRNA1 is localized in the chromatin fraction of the nucleus in the form of ribonucleoprotein (Ro-Choi et al. 1972; Kanehisa et al. 1971). For this reason Ro-Choi et al. (1972) stated that this RNA might have a role in gene transcription.

The above findings together with the homology between a segment of Ul and the splicing site of intron recently reported suggest that small RNAs have a general function of mediating and facilitating various processes of transcription and translation. If this is true, they may be called mediator RNA.

Acknowledgment: We thank Drs. H. Busch, G. Saunders, M.-I. Tsai, R. Reddy, and A. C. Ting for their helpful discussions. This study was supported by research grants from the U.S. National Institute of Health and U.S. National Science Foundation.

References

Busch H, personal communication
Lerner MR, Steitz JA (1979) Proc Natl Acad Sci USA 76: 5495–5499
Roksam WG, Rougeon F (1979) Nucleic Acids Res 7:305–320
Ting AC, Tsai MJ, O'Malley BW (1980) In press

Received September 25, 1980 / Revised Jan. 5, 1981