Intervening sequences excised in vitro

from Richard J. Roberts

SINCE the initial discovery of bizarre happenings during eukaryotic mRNA biosynthesis there has been considerable discussion of the mechanism by which intervening sequences are removed. The bulk of circumstantial evidence pointed to a mechanism which operates on an initial RNA transcript containing a linear representation of the genome. By folding this transcript into an appropriate conformation, it was imagined that the sequences to be joined could be brought together and a concerted cleavage-ligation reaction accomplished. Among the many genes which are now known to contain intervening sequences, the yeast tRNAs for tyrosine and phenylalanine are unusual in that the extra sequences which occur within their anticodon loops are rather small—14 and 18 nucleotides long, respectively. Their existence was deduced by direct DNA sequence analysis of cloned segments of yeast DNA encoding these tRNA genes (Goodman *et al. Proc. natn. Acad. Sci. U.S.A.* **74,** 5453; 1977; Valenzuela *et al. Proc. natn. Acad. Sci. U.S.A.* **75,** 190; 1978). Two recent papers (Knapp *et al. Cell* **14,** 221; 1978; O'Farrell *et al. Nature* **274,** xxx; 1978) now report that corresponding precursors have been isolated which are longer than the mature tRNAs by 14 and 18 nucleotides. Direct RNA sequence analysis has shown that they contain the mature 5′ and 3′ terminal sequences and that the extra nucleotides correspond to the intervening sequence present in the genome. Their isolation (Hopper *et al. Cell* **14,** 211; 1978) was possible by virtue of a mutant strain of yeast, *ts*136 (Hutchinson *et al. J. Bact.* **99,** 807; 1969) which is defective in the transport of RNA from the nucleus to the cytoplasm, although the precise lesion is unknown. Now that these specific precursors are available, it is possible to assay for an enzyme activity able to remove the intervening sequences, and indeed such an activity has been detected in crude extracts of yeast cells (Knapp *et al.*, Valenzuela *et al. op. cit.*). Preliminary evidence suggests that it requires only Mg^{2+} and ATP in order to function but more detailed parameters, as well as the number of activities, must await further purification. It is present in the ribosomal wash fraction and if this reflects its intracellular location the accumulation of precursors would be explained by the failure of the mutant strain to transport them out of the nucleus. This might in turn suggest that this enzyme is quite distinct from the enzymes required to process nuclear RNAs into mRNAs since the latter must surely reside within the nucleus or its membrane. The remarkably short intervening sequences found in the tRNA genes also serve to differentiate them, and hence possibly the mechanisms for their removal, from their mRNA counterparts.

While the fine details of the mechanism must await the further purification of the splicing enzyme, it should already be possible to determine whether the intervening sequences are removed as a linear or a circular oligonucleotide. This distinction could be important, for the latter finding would imply a reciprocal mechanism. Splicing would then lead not only to the joining of non-contiguous sequences to form functional mRNAs, but also to the joining of non-contiguous sequences from the intervening regions. The resulting RNA sequences might have interesting properties, both as a result of their novel joints and of their circular nature. The only circular RNA species known at present is the genome of the potato spindle tuber viroid (Gross *et al. Nature* **237,** 203; 1978) which certainly invites intense speculation as to its mode of infection. No gene products have been identified either *in vivo* or *in vitro* and since it lacks both a functional 5′-terminus and an AUG codon, this is perhaps not too surprising. Nevertheless, it plays such havoc within infected cells that it seems reasonable to believe that it may act on some regulatory system. Could it be that the RNA itself is directly responsible? If so, then RNA may have a crucial role in gene control—a theoretical possibility that may provide some experimental ideas about the role of intervening sequences and RNA splicing. □

Richard J. Roberts is a Senior Staff Investigator at the Cold Spring Harbor Laboratory.

late genes is both rapid and precise. However, fibre mRNA seems to be an exception: it often contains supernumerary leader sequences derived from the early genes which lie just distal to the fibre structural gene (Chow). The presence or absence of these extra leaders does not seem to affect the translation of mRNAs (at least in cell-free systems) because mRNAs containing four leaders work as well *in vitro* as mRNAs with only three (A. Dunn, Cold Spring Harbor). However, there are indications that sequences contained within the conventional tripartite leaders may be necessary for efficient translation of late mRNAs. First, the activity of several late mRNAs in cell free systems is partially suppressed by the presence of DNA sequences complementary to the leaders (Paterson, NIH). Furthermore the first leader contains a sequence which, being complementary to the sequence at the 3′ end of 18S ribosomal RNA may be involved in a Shine/Dalgarno fashion in binding of mRNAs to 40S ribosomal subunits (Ziff).

Nomenclature

In addition to scientific matters, problems of nomenclature were also discussed at the conference. The participants considered it inadvisable at present to subscribe to the use of new terms that attempt to label those parts of a primary RNA transcript that succeed in reaching the cytoplasm compared with those parts which are presumably destroyed in the nucleus. This decision rests on the fact that from both early and late adenovirus transcription units—probably the same is true for papovaviruses, retroviruses and eventually may be shown for cellular transcription units as well—multiple spliced mRNA arrangements are possible. Thus what is conserved in one primary RNA transcript may be discarded in another and *vice versa*. The terms 'intron' and 'exon' which have been suggested to describe regions of DNA from which discarded or conserved RNA sequences are transcribed were thought particularly inadvisable as they name regions of DNA whereas the molecules on which the choice for conservation or destruction is exercised are primary RNA transcripts. Perhaps the simplest course at the moment is to refer to regions of a primary transcript that never reach the cytoplasm as 'spacers' (in line with common practice for pre-rRNA) or 'intervening sequences' and to regions that may reach the cytoplasm as 'mRNA sequences.'

Splicing mechanism

Clearly, further progress in elucidating the mechanism of splicing of late adenoviral RNA will depend on the isolation of the enzyme(s) responsible as well as the determination of the

0028–0836/78/0274—0530$01.00