Hypothesis

MECHANISM FOR RNA SPLICING OF GENE TRANSCRIPTS

Vincent MURRAY and Robin HOLLIDAY
National Institute for Medical Research, The Ridgeway, Mill Hill, London, NW7 1AA, England

Received 9 July 1979

1. Introduction

It has been demonstrated that structural genes in eukaryotes contain intervening sequences (introns) which are absent from expressed sequences (exons) [1–6]. There is evidence that intervening sequences are transcribed [7], but subsequently the introns are removed by a process which is now known as RNA splicing [8,9].

RNA splicing must be a very precise process since it results in the joining of successive coding sequences to produce an unambiguous message. We wish to propose a mechanism for RNA splicing which depends on intermolecular hybridisation and which is closely analogous to the genetic recombination of DNA molecules. In both processes the point at which the breakage and reunion of polynucleotide genes occurs is determined by the base pairing of complementary sequences (hybrid DNA or hybrid RNA). In this way the addition of extraneous nucleotides or the deletion of essential nucleotides is prevented.

2. Mechanism for RNA splicing

The primary transcript of DNA is assumed to be hnRNA which contains both intron and exon sequences in a colinear copy of DNA. We assume that in cells in which the gene product is present, a small RNA molecule known as splicer RNA is transcribed from a genetic locus which is distinct from the structural gene. The base sequence of this RNA is such that it could hybridise with hnRNA either to adjacent exon sequences, the ends of intron sequences or to both exons and introns. This is illustrated in fig.1. Consider first the case in fig.1b. The configuration is the same as in the likely intermediate in DNA recombination [10–13]. Hydrogen bonding is possible on each side of the point where the polynucleotide genes exchange pairing partners. We suppose that there is an RNA endonuclease and an RNA ligase which breaks genes at the point indicated and rejoins the ends of the exon sequences. In DNA recombination it is believed that either pair of strands of like polarity can be cut and rejoined, but in fig.1b only the hnRNA strand must be susceptible to enzymic cleavage. It should be noted that in fig.1b, the two splicer RNA sequences may form part of one RNA molecule (see also fig.2).

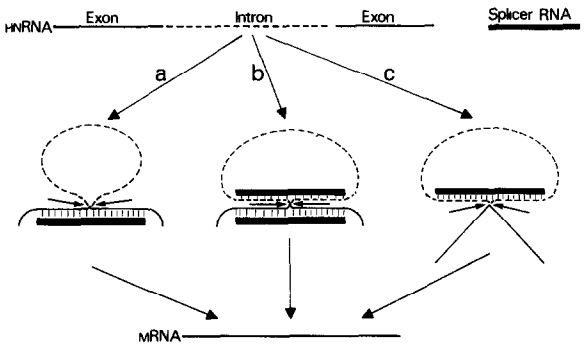

Fig.1. Three ways that splicer RNA can bind across a splice point: (a) hybridised to exon sequences; (b) hybridised to intron and exon sequences; (c) hybridised to intron sequences. Splice points are indicated by short arrows.

In fig.1a the splicer RNA is hybridised only to the exon sequences. This structure is the same as DNA heteroduplexes containing a deletion (or addition). Such structures arise in genetic recombination and can be repaired by removal of the loop and ligation to produce normal hydrogen bonded DNA. The

possibility that a heteroduplex might be an intermediate in RNA splicing has been mentioned recently [14]. In fig.1c splicer RNA is hybridised only to intron sequences. In this case the endonuclease and ligase action would have to be in some way coordinated to achieve end-to-end joining sequences. The important points are first that each of the three structures in fig.1 provides the necessary specificity for the precise joining of exon sequences, and secondly, that exons in different gene transcripts are spliced by a common enzymic mechanism. The model also implies that in evolutionary terms, there would be conservation of exon sequences (fig.1a,1b) or intron sequences (fig.1b,1c) which hybridised to splicer RNA.

3. Other possible mechanisms

Alternative suggestions for splicing mechanisms are not completely convincing. One possibility is that the enzyme recognises the sequences at intron/exon junctions and carries out the splicing reaction. However, there are only short 'consensus' sequences at these junctions [1,2,5,6] and it is apparent that these cannot be the signal for an enzyme since the same sequences occur elsewhere in mRNA [2]. If longer sequences are recognised, then there must be a very large family of splicing enzymes and this seems implausible.

It has also been proposed that intramolecular RNA hybridisation occurs at the splice point (but with non-perfect base pairing) and that this structure can be recognised by the splicing enzymes [3–5]. The main drawback of this mechanism is the exons are not in the correct stereochemical position for ligation to occur, and the same kind of structures could not be formed from the many known intron/exon junctions [2].

4. Experimental evidence

In an adenovirus 2 infection a number of small RNAs are produced which could function as splicer RNAs. In particular VA (virus associated) RNA_1 is produced in relatively large amounts late in infection and has been sequenced [15,16]. Three intron/exon

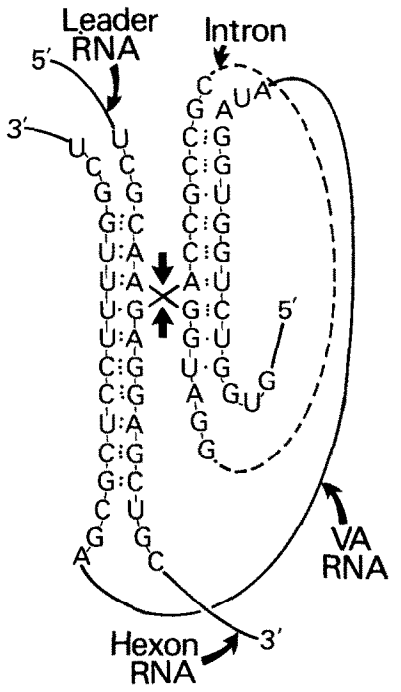

Fig.2. Structure formed from adenovirus 2 VA RNA_1 and the third splice point (from the 5′ end) of adenovirus 2 hexon pre-mRNA. Arrows indicate splice point.

junctions present in the adenovirus 2 hexon gene (expressed late in infection) have been sequenced [4]. In accordance with our mechanism, 19 bases of VA RNA_1 can hybridise across the third splice point (from the 5′ end) of the hexon gene (see fig.2). This stable structure is achieved by base-pairing with both intron and exon sequences. The probability of these hybridisable sequences occurring by change is quite remote. We feel that this is extremely strong supportive evidence for our mechanism.

Ribonucleoprotein complexes containing hnRNA exist in the nuclei of every eukaryote examined [17]. In rat liver these complexes also contain small nuclear RNA (snRNA) which is base-paired to hnRNA [18]. The length of this hybridisation is a minimum of 15–25 nucleotides. snRNA is 150–250 nucleotides in length, highly methylated and with a lifetime >48 h. This stable snRNA is similar in properties to that proposed for splicer RNA. The methylation of snRNA could be the method by which splicer RNA and hnRNA strands are distinguished. The protein

complexes, with which snRNA is associated, might contain the splicing enzymes and also possibly enzymes for the further processing of hnRNA to mRNA.

The prediction that splicer RNA hybridising to intron or exon sequences (see fig.1) should result in the evolutionary conservation of these sequences is confirmed in two genes. Mouse, rabbit and human β globin mRNA sequences show most sequence homology in all three codon positions in exons neighbouring the two splice points [6,19]. In mouse and rabbit β globin intron sequences are also most highly homologous next to the two splice points [20]. In the small intervening sequence of immunoglobulin λ_I and λ_{II} light chains the introns are as highly conserved as the exons [21].

5. Post-transcriptional control of gene expression

The mechanism for RNA splicing immediately suggests a means for controlling which genes are expressed in a cell. By regulating the synthesis of splicer RNAs, the removal of introns from individual genes is also directly controlled. Thus, if the removal of all intervening sequences from an hnRNA molecule is a necessary pre-requisite for transportation of that RNA to the cytoplasm, then gene expression can be controlled by splicing.

The innumerable combinations of expressed genes found in differentiated tissues can be accomplished by our splicing mechanism. Because the specificity is provided by RNA, splicer RNA hybridisation of 20 bases in length is capable of sufficient variation to account for the necessary number of controlling elements in a multicellular organism. The further possibility that splicer RNA is transferable between cells and able to influence gene expression in neighbouring cells could be important in differentiation and development [22,23].

References

[1] Breathneck, R., Benoist, C., O'Hara, K., Gannon, F. and Chambon, P. (1978) Proc. Natl. Acad. Sci. USA 75, 4853–4857.
[2] Catterall, J. F., O'Malley, B. W., Robertson, M. A., Staden, R., Tanaka, Y. and Brownlee, G. G. (1978) Nature 275, 510–515.
[3] Tonegawa, S., Maxam, A. M., Tizard, R., Bernard, O. and Gilbert, W. (1978) Proc. Natl. Acad. Sci. USA 75, 1485–1489.
[4] Akusjarvi, G. and Pettersson, U. (1979) Cell 16, 841–850.
[5] Ghosh, P. K., Reddy, V. B., Swinscoe, J., Lebowitz, P. and Weissman, S. M. (1978) J. Mol. Biol. 126, 813–846.
[6] Konkel, D. A., Tilghman, S. M. and Leder, P. (1978) Cell 15, 1125–1132.
[7] Tilghman, S. M., Curtis, P. J., Tiemeier, D. C., Leder, P. and Weissman, C. (1978) Proc. Natl. Acad. Sci. USA 75, 1309–1313.
[8] Gilbert, W. (1978) Nature 271, 501.
[9] Crick, F. (1979) Science, 204, 264–271.
[10] Sobell, H. M. (1974) in: Mechanisms in Recombination (Grell, R. F. ed) pp. 433–438, Plenum Press, New York.
[11] Potter, H. and Dressler, D. (1976) Proc. Natl. Acad. Sci. USA 73, 3000–3004.
[12] Sigal, N. and Alberts, B. (1972) J. Mol. Biol. 71, 789–793.
[13] Holliday, R. (1964) Genet. Res. 5, 282–304.
[14] Davidson, E. H. and Britten, R. J. (1979) Science 204, 1052–1059.
[15] Ohe, K. and Weissman, S. M. (1971) J. Biol. Chem. 246, 6991–7001.
[16] Harris, B. and Roeder, R. G. (1978) J. Biol. Chem. 253, 4120–4127.
[17] Preobrazhensky, A. A. and Spirin, A. S. (1978) Prog. Nucleic Acid Res. Mol. Biol. 21, 1–38.
[18] Flytzanis, C., Alonso, A., Louis, C., Krieg, L. and Sekeris, C. E. (1978) FEBS Lett. 96, 201–206.
[19] Kafatos, F. C., Efstradiadis, A., Forget, B. A. and Weissman, S. M. (1977) Proc. Natl. Acad. Sci. USA 74, 5618–5622.
[20] Van den Berg, J., Van Ooyen, A., Mantei, N., Schambock, A., Grosveld, G., Flavell, R. A. and Weissman, C. (1978) Nature 276, 37–44.
[21] Bernard, O., Hozumi, N. and Tonegawa, S. (1978) Cell 15, 1133–1144.
[22] Kolodny, G. M. (1971) Exp. Cell Res. 65, 313–324.
[23] Temin, H. M. (1974) Ann. Rev. Genet. 8, 155–174.