Volume 2 · Number 1 · 1992
GENE EXPRESSION

MINI REVIEW

RNA binding proteins, splice site selection, and alternative pre-mRNA splicing

Donald C. Rio

Whitehead Institute for Biomedical Research and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts

Recent studies using both biochemical and genetic approaches have shown that a common class of RNA-binding proteins is involved both in general splice site selection and in several examples of alternative pre-mRNA splicing. These experiments suggest that combinations of RNA-binding proteins may interact with pre-mRNA, that these interactions may be stabilized via protein-protein interactions, and that this array of RNA-protein complexes may either direct or preclude binding of the spliceosomal small ribonucleoprotein particles (snRNPs) to the pre-mRNA during the initial phases of spliceosome assembly. This article will review our current understanding of the role RNA-binding proteins play in general and regulated pre-mRNA splicing.

The ribonucleoprotein consensus sequence (RNP-CS) RNA binding domain

A number of RNA-binding proteins, including the yeast poly(A) binding protein, mammalian heterogeneous nuclear ribonucleoprotein particle (hnRNP) proteins, and small nuclear ribonucleoprotein particle (snRNP) proteins, contain an 80–90 amino acid domain referred to as the RNP-CS motif, which is known to be responsible for mediating RNA binding (for reviews, see Bandziulis et al., 1989; Mattaj, 1989; Kenan et al., 1991). This motif has since been found in a large number of proteins with diverse functions, all of which are involved in (or by inference presumed to be involved in) RNA metabolism. Analysis of several members of this family, such as the U1 snRNP 70K and A proteins, has shown that the RNP-CS domain is indeed sufficient to bind RNA in the absence of other regions of the molecule (reviewed by Kenan et al., 1991).

A series of mutagenesis and RNA-binding experiments using the U1 snRNP 70K protein have indicated that the RNP-CS domain is necessary and sufficient for RNA binding to the stem-loop I of U1 snRNA (Query et al., 1989). Similar experiments with U1 snRNP A and U2 snRNP B″ proteins have indicated that the RNP-CS domain is responsible for RNA binding (Scherly et al., 1989; Scherly et al., 1990; Lutz-Freyermuth et al., 1990; Bentley and Keene, 1990). However, in other instances, such as with the La RNP protein and Ro RNP 60K proteins, additional amino acid sequences lying outside the RNP-CS domain are also required for specific RNA binding (reviewed by Kenan et al., 1991). Presumably, these additional residues might aid in the folding or stabilization of the RNP-CS domain within the context of the full protein, rather than directly contacting the RNA.

The ability of the isolated 90 amino acid RNP-CS domain of the U1 snRNP A protein to interact specifically with stem-loop II of U1 snRNA (Scherly et al., 1989; Lutz-Freyermuth et al., 1990) has allowed a detailed structural analysis of this peptide both by X-ray crystallography (Nagai et al., 1990) and nuclear magnetic resonance

Correspondence: Donald C. Rio, Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142 Tel (617) 258-5240 Fax (617) 258-5061. After May 1, 1992: Department of Molecular and Cell Biology, Divisions of Genetics/Biochemistry and Molecular Biology, 401 Barker Hall, University of California, Berkeley, CA 94120
 1052-2166/92/0201/001-05$2.00

(NMR) methods (Hoffman et al., 1991). Although the structures determined by the two methods differ in detail, the overall secondary structural motifs are identical, consisting of a four-stranded antiparallel β sheet with two α helices behind the sheet. The most highly conserved regions within the RNP-CS domain are designated RNP-1 (K/R G F/Y G/A F V X F) and RNP-2 (L/I F/Y V/I G/K N/G). The RNP-1 and RNP-2 motifs contain hydrophobic amino acids that are brought together in the two central strands of the β sheet. It is presumed that the RNA lies across the β sheet based on a variety of mutagenesis and RNA-binding experiments (Nagai et al., 1990; Jessen et al., 1991), and on the fact that two aromatic amino acid residues in RNP-1 and RNP-2 can be covalently crosslinked to RNA. Amino acids outside the β sheet are also involved in RNA binding because a number of basic amino acids, including a conserved arginine residue that is essential for RNA binding, are located in loops adjacent to the surface of the β sheet.

Splice site selection

One of the most important questions in the study of pre-mRNA splicing is how 5′ and 3′ splice sites are recognized, chosen, and paired accurately to allow catalysis of intron removal and the generation of mature functional mRNA. A number of studies have shown that exon sequences, as well as the sequences and locations of the 5′ and 3′ splice sites in a given pre-mRNA, can influence splice site choice (for reviews, see Sharp, 1987; Andreadis et al., 1987; Krainer and Maniatis, 1988; Smith et al., 1989a). Recent biochemical studies reveal that RNA-binding proteins can influence splice site selection, are required for splicing, and interact intimately with the pre-mRNA to participate in spliceosome assembly (Zamore and Green, 1989; Garcia-Blanco et al., 1989; Fu and Maniatis, 1990; Ge and Manley, 1990; Krainer et al., 1990). It is likely that many different RNA-binding proteins interact with the pre-mRNA to dictate and facilitate splice site recognition and spliceosome assembly.

Studies in yeast and mammalian cells have indicated that both the intron 5′ splice site and branchpoint-polypyrimidine tract are involved in the formation of ribonucleoprotein complexes early in spliceosome assembly, and that these complexes play a role in splice site usage (for reviews, see Sharp, 1987; Guthrie, 1991; Ruby and Abelson, 1991; Rosbash and Seraphin, 1991). For example, a number of studies have shown that stable ATP-independent complexes form on the pre-mRNA 5′ splice site during the initial stages of spliceosome assembly and 5′ splice site recognition. Furthermore, it has been shown that juxtaposition of competing 5′ splice sites can influence 5′ splice site selection (Nelson and Green, 1988 and 1990). Many studies have shown that the 3′ splice site region can influence spliceosome assembly and the efficiency of splicing (Reed, 1989; Smith et al., 1989b; Goguel et al., 1991; Patterson and Guthrie, 1991). Furthermore, the "quality" of the intron polypyrimidine tract (i.e., its length and pyrimidine content) and its distance from the branchpoint sequence can affect the choice of 3′ splice sites (Reed, 1989; Smith et al., 1989b; Garcia-Blanco, 1989; Patterson and Guthrie, 1991; Mullen et al., 1991).

Proteins that recognize the intron polypyrimidine tract near the 3′ splice site also contain the RNP-CS domain. U2 snRNP auxilliary factor (U2AF), the splicing factor required in vitro for stable U2 snRNP binding to the intron branchpoint (Ruskin et al., 1988), was recently purified to homogeneity and found to consist of two subunits of 65kD and 35kD molecular weight (Zamore and Green, 1989). Recent biochemical (Zamore and Green, 1991) and gene cloning (M. Green, personal communication) experiments have indicated that the 65kD subunit is sufficient for all of the biochemical activities of authentic native U2AF, including binding to the intron polypyrimidine tract and facilitation of U2 snRNP binding to the intron branchpoint. This subunit contains three RNP domains and an R/S domain (M. Green, personal communication). The presence in spliceosomes of a second polypyrimidine tract binding protein, pPTB, suggests that this RNA-binding protein is also involved in splicing (Garcia-Blanco et al., 1989). The gene encoding pPTB has been cloned, and the sequence revealed the presence of an RNP domain and homology to hnRNP L and to the Drosophila elav RNP-CS-containing gene (Gil et al., 1991; Patterson et al., 1991).

A general mammalian splicing factor, SF-2/ASF, was recently identified biochemically by its ability to complement a cytoplasmic extract for splicing and to shift 5′ splice site usage from

a distal to a proximal site (Ge and Manley, 1990; Krainer et al., 1990a,b). Highly purified SF-2/ASF has been shown to possess an intrinsic RNA-binding activity, as well as the ability to anneal complementary single-stranded RNA in a sequence-independent manner (Krainer et al., 1990a). The gene encoding SF-2/ASF was isolated using microprotein sequence information, and the recombinant protein was shown to possess all of the properties of the mammalian factor (Krainer et al., 1991; Ge et al., 1991). Analysis of the gene sequence revealed that the protein contained an RNP domain and an arginine-serine rich (R/S) domain, both of which have been found in Drosophila splicing regulatory proteins (see below) and in the U1 snRNP 70K protein (Query et al., 1989). Another mammalian splicing factor was identified by using monoclonal antibodies raised to purified spliceosomes (Fu and Maniatis, 1990; Spector et al., 1991). This factor, called SC-35, is similar in molecular weight to SF2/ASF and also possesses RNP and R/S domains (T. Maniatis, personal communication). Both SF2/ASF and SC-35 are required for the earliest steps in splicing complex assembly. Thus, a number of RNA-binding proteins involved in splicing contain the RNP-CS domain and bind RNA; and at least one, SF-2/ASF, has the interesting property of altering splice site utilization in vitro. This observation raises the possibility that variations in the levels or activities of general splicing factors following a tissue-specific or developmental program may contribute to differential splice site selection.

Alternative pre-mRNA splicing

In Drosophila, genetic analysis has identified a number of loci capable of altering splicing patterns of a variety of genes. For example, control of somatic sexual differentiation depends on a cascade of alternative splicing events (for review, see Baker, 1989). A number of gene products involved in these splicing decisions have been identified and shown to contain RNP-CS domains and/or R/S domains similar to the mammalian splicing regulators described above (Amrein et al., 1990; Goralski et al., 1989; Bell et al., 1988; Boggs et al., 1987; Chou et al., 1987). Clearly, a common theme in the control of both splice site choice and alternative splicing reactions is the involvement of RNA-binding proteins carrying the RNP-CS motif and/or the arginine-serine rich (R/S) domain.

Three examples of negative control of splice site choice have been described in Drosophila. The *Sex-lethal* (*Sxl*) protein, which contains two RNP-CS motifs (Bell et al., 1988), appears to repress the use of a male-specific 3′ splice site in the *Sxl* gene itself in an autoregulatory mode (Bell et al., 1991) and also acts to repress the use of a non-sex specific 3′ splice site in the *transformer* (*tra*) gene, a gene downstream in the pathway of sexual differentiation (Sosnowski et al., 1989). Both molecular genetic and biochemical experiments indicate that *Sxl* acts by binding to the polypyrimidine tracts of the introns it controls (Sosnowski et al., 1989; Inoue et al., 1990). Another example of negative splicing control is the germline-specific splicing of the P transposable element third intron (IVS3; for reviews, see Rio, 1990 and 1991). Genetic and biochemical experiments have indicated that at least one aspect of this control is repression of this splicing event in somatic cells, resulting in the retention of the third intron in the mature somatic mRNA (Laski and Rubin, 1989; Chain et al., 1991; Tseng et al., 1991; Siebel and Rio, 1990). Competition between the accurate 5′ splice site and pseudo-5′ splice sites in the adjacent 5′ exon seems to play a role in somatic inhibition of IVS3 splicing (Siebel and Rio, 1990). Mutations in the 5′ exon that disrupt these 5′ splice site-like sequences activate third intron splicing in vivo in somatic cells of transgenic Drosophila (Chain et al., 1991) and in somatic cell extracts in vitro (Siebel and Rio, 1990; Tseng et al., 1991). An excess of this inhibitory exon RNA sequence relieves the inhibitory effect observed with Drosophila somatic cell nuclear extracts in vitro (Siebel and Rio, 1990; Tseng et al., 1991), suggesting that this exon RNA fragment titrates inhibitory factors away from the pre-mRNA. Furthermore, the inhibitory activity correlates with the binding of several somatic Drosophila RNA-binding proteins to the inhibitory RNA target site in the 5′ exon (Siebel and Rio, 1990; Chain et al., 1991). These RNA-binding proteins appear to act by influencing U1 snRNP binding in vitro (Siebel et al., personal communication). The further characterization and identification of these RNA-binding proteins should help to determine their tissue distribution and reveal which ones might be responsible for the different splicing patterns

of this intron in the germline and soma. A third example of negative splicing control in Drosophila involves the *suppressor-of-white-apricot* (su^{wa}) gene. The su^{wa} protein, which contains an R/S domain but no RNP-CS motif, autoregulates the splicing of the su^{wa} pre-mRNA to prevent synthesis of the protein by blocking removal of one of the su^{wa} introns (Chou et al., 1987). Thus, several examples of splicing control in Drosophila involve the repression of splice site use in alternative splicing decisions.

One example of positive control of pre-mRNA splicing in Drosophila has been described. Genetic experiments have shown that the female-specific 3′ splice site of the *double-sex* (*dsx*) gene, another in the sex-determination pathway, is activated by the products of the *transformer* (*tra*) and *transformer-2* (*tra-2*) genes (Burtis and Baker, 1989). It is the female product of the *dsx* gene that is responsible for directing further female development (Nagoshi et al., 1988). Sequencing of the *tra* and *tra-2* genes has revealed that the *tra* protein possesses an R/S domain (Boggs et al., 1987), and that the *tra-2* protein has RNP-CS and R/S domains (Amrein et al., 1988; Goralski et al., 1989). A *tra-2* product can also act to autoregulate splicing of the *tra-2* gene itself, and it is clear that the *tra-2* primary transcript is subject to alternative RNA splicing in a tissue-specific manner (Amrein et al., 1990; Mattox et al., 1990; Mattox et al., 1991). Genetic experiments in Drosophila have identified a region near the *dsx* female 3′ splice site as the target of *tra* and *tra-2* action (Nagoshi and Baker, 1990). A series of tissue culture transfection experiments have indicated that the *tra*/*tra-2* products activate the female *dsx* 3′ splice site by interacting with the regulatory site in the downstream 3′ exon consisting of six repeats of a thirteen nucleotide sequence. These repeats are necessary for positive regulation (Hedley and Maniatis, 1991; Hoshijima et al., 1991; Ryner and Baker, 1991) and interact directly with *tra-2* protein in vitro (Hedley and Maniatis, 1991). *Tra* and *tra-2* also activate female-specific polyadenylation in the absence of the regulated 3′ splice site (Hedley and Maniatis, 1991). It is possible that *tra*/*tra-2* activate the *dsx* female-specific 3′ splice site by stabilizing an otherwise weak interaction between U2AF and the rather poor polypyrimidine tract of the *dsx* female-specific intron. Alternatively, *tra* and *tra-2* may simply alter the conformation of the *dsx* pre-mRNA to improve its recognition as a substrate for splicing in females. Both of these ideas are consistent with an early role for these regulatory proteins in intron recognition, splice site selection, and spliceosome assembly, and with mammalian biochemical studies of the early steps of spliceosome assembly.

Perspective

One common theme emerging from studies on basic splicing factors in mammalian cells and both genetic and biochemical studies on Drosophila is that many of the protein factors involved in splicing contain RNA-binding domains of the RNP-CS type and interact with the pre-mRNA early in the pathway of spliceosome assembly. A number of proteins involved in both general and regulated splicing also carry an arginine-serine rich (R/S) domain. Determining the role of this domain in splicing and the targets (protein and/or RNA) with which it interacts should shed light on basic mechanisms of splicing and its control. So little is known about how these RNA-binding proteins interact with the pre-mRNA substrate, and how these interactions facilitate splice site recognition, selection, and the early steps in spliceosome assembly, that important and fundamental insights are certain to emerge by using molecules already available as probes. Finally, as more insight is gained into how RNA-binding proteins of the RNP-CS family recognize their target RNAs, it should be possible to engineer and design RNA-binding and splicing factors of predetermined specificities and thereby alter patterns of splice site selection in predictable ways.

Acknowledgments

I thank Chris Siebel for critical review of the manuscript. Work in my laboratory is supported by the NIH (R01-HD28063-01), NSF (DMB-8857176), and Lucille P. Markey Charitable Trust. I am supported by a Lucille P. Markey Scholar Award.

References

H. Amrein, M. Gorman, and R. Nöthinger (1988), Cell 55, 1025–1035.

H. Amrein, T. Maniatis, and R. Nöthinger (1990), EMBO J 9, 3619–3629.

A. Andreadis, M. E. Callego, and B. Nadal-Ginard (1987), Annu Rev Cell Biol 3, 207–242.

B. S. Baker (1989), Nature 340, 521–524.

R. J. Bandziulis, M. S. Swanson, and G. Dreyfuss (1989), Genes Dev 3, 431–437.
L.R. Bell, E. M. Maine, P. Schedl, and T. W. Cline (1988), Cell 55, 1037–1046.
L. R. Bell, J. I. Horabin, P. Schedl, and T. W. Cline (1991), Cell 65, 229–239.
R. C. Bentley and J. D. Keene (1991), Mol Cell Biol 11, 1829–1839.
R. T. Boggs, P. Gregor, S. Idriss, J. M. Belote, and M. McKeown (1987), Cell 50, 739–747.
K. C. Burtis and B. S. Baker (1989), Cell 56, 997–1010.
A. C. Chain, S. Zollman, J. C. Tseng, and F. A. Laski (1991), Mol Cell Biol 11, 1538–1546.
T. B. Chou, Z. Zachar, and P. M. Bingham (1987), EMBO J 6, 4095–4104.
X.-D. Fu and T. Maniatis (1990), Nature 343, 437–441.
M. A. Garcia-Blanco, G. J. Anderson, J. Beggs, and P. A. Sharp (1989), Genes Dev 3, 1874–1886.
H. Ge and J. L. Manley (1990), Cell 62, 25–34.
H. Ge, P. Zuo, and J. L. Manley (1991), Cell 66, 373–382.
A. Gil, P. A. Sharp, S. F. Jamison, and M. A. Garcia-Blanco (1991), Genes Dev 5, 1224–1236.
V. Goguel, X. Liao, B. C. Rymond, and M. Rosbash (1991), Genes Dev 5, 1430–1438.
T. J. Goralski, J.-E. Edström, and B. S. Baker (1989), Cell 56, 1011–1018.
C. Guthrie (1991), Science 253, 157–163.
M. L. Hedley and T. Maniatis (1991), Cell 65, 579–586.
D. W. Hoffman, C. C. Query, B. L. Golden, S. W. White, and J. D. Keene (1991), Proc Natl Acad Sci USA 88, 2495–2499.
K. Hoshijima, K. Inoue, I. Higuchi, H. Sakamoto, and Y. Shimura (1991), Science 252, 833–836.
K. Inoue, K. Hoshijima, H. Sakamoto, and Y. Shimura (1990), Nature 344, 461–463.
T.-H. Jessen, C. Oubridge, C. H. Teo, C. Pritchard, and K. Nagai (1991), EMBO J 10, 3447–3456.
D. J. Kenan, C. C. Query, and J. D. Keene (1991), Trends Biochem Sci 16, 214–220.
A. R. Krainer and T. Maniatis (1988), in Molecular Biology: Transcription and Splicing (B. D. Hames and D. M. Glover, eds.), IRL Press, Oxford, pp. 131–206.
A. R. Krainer, G. C. Conway, and D. Kozak (1990a), Genes Dev 4, 1158–1171.
A. R. Krainer, G. C. Conway, and D. Kozak (1990b), Cell 62, 35–42.
A. R. Krainer, A. Mayeda, D. Kozak, and G. Binns (1991), Cell 66, 813–833.
F. A. Laski and G. M. Rubin (1989), Genes Dev 3, 720–728.
C. Lutz-Freyermuth, C. C. Query, and J. D. Keene (1990), Proc Natl Acad Sci USA 87, 6393–6397.
I. W. Mattaj (1989), Cell 57, 1–3.
W. Mattox and B. S. Baker (1991), Genes Dev 5, 786–796.
W. Mattox, M. J. Palmer, and B. S. Baker (1990), Genes Dev 4, 789–805.
M. P. Mullen, C. W. J. Smith, J. G. Patton, and B. Nadal-Ginard (1991), Genes Dev 5, 642–655.
K. Nagai, C. Oubridge, T. H. Jessen, J. Li, and P. R. Evans (1990), Nature 348, 515–520.
R. N. Nagoshi, M. McKeown, K. C. Burtis, J. M. Belote, and B. S. Baker (1988), Cell 53, 229–236.
R. N. Nagoshi and B. S. Baker (1990), Genes Dev 4, 89–97.
K. K. Nelson and M. R. Green (1988), Genes Dev 2, 319–329.
K. K. Nelson and M. R. Green (1990), Proc Natl Acad Sci USA 87, 6253–6257.
B. Patterson and C. Guthrie (1991), Cell 64, 181–187.
J. G. Patton, S. A. Mayer, P. Tempst, and B. Nadal-Ginard (1991), Genes Dev 5, 1237–1251.
C. C. Query, R. C. Bentley, and J. D. Keene (1989), Cell 57, 89–101.
R. Reed (1989), Genes Dev 3, 2113–2123.
D. C. Rio (1990), Annu Rev Genet 24, 543–578.
D. C. Rio (1991), Trends Genet 7, 282–287.
M. Rosbash and B. Seraphin (1991), Trend Biochem Sci 16, 187–190.
S. Ruby and J Abelson (1991), Trends Genet 7, 79–85.
B. Ruskin, P. D. Zamore, and M. R. Green (1989), Cell 52, 207–219.
L. C. Ryner and B. S. Baker (1991), Genes Dev 5, 2071–2085.
D. Scherly, W. Boelens, W. J. van Venrooij, N. A. Dathan, J. Hamm, and I. W. Mattaj (1989), EMBO J 8, 4163–4170.
D. Scherly, W. Boelens, N. A. Dathan, W. J. van Venrooij, and I. W. Mattaj (1990), Nature 345, 502–506.
P. Sharp (1987), Science 235, 766–771.
C. W. Siebel and D. C. Rio (1990), Science 248, 1200–1208.
C. W. J. Smith, J. G. Patton, and B. Nadal-Ginard (1989a), Annu Rev Genet 23, 527–577.
C. W. J. Smith, E. B. Porro, J. G. Patton, and B. Nadal-Ginard (1989b), Nature 342, 243–247.
B. A. Sosnowski, J. M. Belote, and M. McKeown (1989), Cell 58, 449–459.
D. L. Spector, X.-D. Fu, and T. Maniatis (1991), EMBO J 10, 3467–3481.
J. C. Tseng, S. Zollman, A. C. Chain, and F. A. Laski (1991), Mech Dev 35, 65–72.
P. D. Zamore and M. R. Green (1989), Proc Natl Acad Sci USA 86, 9243–9247.
P. D. Zamore and M.R. Green (1991), EMBO J 10, 207–214.