Update TRENDS in Genetics Vol.xxx No.x

Full text provided by www.sciencedirect.com ScienceDirect

Research Focus

SR proteins: a foot on the exon before the transition from intron to exon definition

Oren Ram and Gil Ast

Department of Human Molecular Genetics and Biochemistry, Sackler Faculty of Medicine, Tel Aviv University, Ramat Aviv 69978, Israel

Two recent publications illuminate the evolution of alternative splicing, showing that a SR (serine–arginine-rich) protein that regulates alternative splicing in multicellular organisms is also found in a unicellular organism without alternative splicing, in which it can assist in the splicing of weak introns. Moreover, insertion of SR proteins into an organism lacking such proteins can restore the splicing of weak introns. These results imply that SR proteins had already facilitated the splicing of weak introns before the evolution of alternative splicing.

Introduction

Introns are found in all eukaryotic organisms. However, the number and size of introns and exons and their modes of recognition by the splicing machinery are different in unicellular compared with multicellular organisms, and also vary within these groups [1]. In the budding yeast *Saccharomyces cerevisiae*, only 3% of the genes contain introns; 99% of these contain only a single intron ~270 nt long. By contrast, in the fission yeast *Schizosaccharomyces pombe*, 45% of the genes contain introns, ranging from 40 to 70 nt long; half of these genes contain more than one intron [2,3]. However, in multicellular organisms, most of the genes contain several introns (8.4 on average in humans), and comparative analyses among multicellular organisms have revealed both a greater number of exons per gene and larger intronic sequences in primates and other mammals compared with all other organisms [4–6]. This indicates a great degree of variability in the intron–exon structure of genes among different eukaryotic organisms.

SR (serine–arginine-rich) proteins are splicing factors that regulate both alternative and constitutive splicing. They bind to short RNA sequences and mediate spliceosome assembly (Box 1). In metazoans, SR protein genes constitute nine families, of which six have two or more members in mammals. However, there are no SR proteins in *S. cerevisiae* and only two SR proteins in *S. pombe*. In other unicellular eukaryotes, there are one or two SR protein genes [7]. Thus, the diversity of SR proteins seems to have emerged with multicellularity.

There are two potential mechanisms for exon and intron selection by the splicing machinery, called intron and exon definition. These two models are still unproven, and all the indications for their existence are circumstantial. However, intron definition is presumably the ancient one, in which the splicing machinery recognizes an intronic unit and places the basal machinery across introns. Therefore, the size of the intron is under selection. Indeed in *S. cerevisiae* and *S. pombe*, almost all introns are less than 350 nt long, and all the information for accurate splicing is within the intron sequences [8,9]. This suggests that intron definition is the only system that directs the splicing machinery in these organisms [10]. In the second mechanism, exon definition, the basal splicing machinery is placed across exons. The length of exons must not exceed 300 nt. It was postulated that during evolution the enlargement of intronic sequences forced the splicing machinery to shift from the recognition of short intronic sequences to the selection of short exonic sequences – from intron to exon definition. This could explain the selective pressure to maintain short intronic sequences in yeast genes and short internal exons in the human genome (and other higher metazoans; see also Supplementary Data) [11].

To the best of our knowledge, there is no alternative splicing in *S. cerevisiae* and *S. pombe*, whereas alternative splicing is prevalent in multicellular organisms [2,3]. Therefore, one explanation might be that the ability to handle multi-intron genes and alternative splicing was lost in *S. cerevisiae* and *S. pombe*. Recent results that might support such a scenario imply a massive intron loss during the evolution of worms and flies, rather than intron gain in other organisms [12]. However, it is more likely that complex regulatory networks evolved from simple ones [13]. Pursuing this theme, it was estimated that the percentage of genes that undergo alternative splicing increases in higher metazoans compared with lower metazoans [4,14]. So, can we trace the steps leading to the appearance of alternative splicing (or to the loss of it in those yeasts)?

Srp2p in *S. pombe* supports splicing of weak introns

Recent results from the Wise laboratory suggest that the SR proteins were already involved in enhancing splicing of suboptimal introns in organisms that support only intron definition [15]. These authors found that one of the two SR proteins in *S. pombe*, Srp2p, improves the recognition of a suboptimal 3′ splice site (3′ss) and thus facilitates the splicing of the cognate intron. Webb *et al.* demonstrated that Srp2p binds to an exonic sequence that is rich in purines and is located downstream of an intron with a suboptimal 3′ss. The suboptimization of that 3′ss was achieved by insertion of adenosines upstream of the 3′ss that increased the distance between the branch site sequence and the 3′ss. Srp2p was shown to interact

Corresponding author: Ast, G. (gilast@post.tau.ac.il).
Available online xxxxxx.

  doi:10.1016/j.tig.2006.10.002

Please cite this article in press as: Ram, O., Ast, G., SR proteins: a foot on the exon before the transition from intron to exon definition, Trends Genet. (2006), doi:10.1016/j.tig.2006.10.002

Box 1. The role of SR proteins in mRNA splicing

Four splice signals are essential for accurate splicing of each intron: the 5′ and 3′ splice sites (5′ss and 3′ss) located at the 5′ and 3′ ends of introns; the polypyrimidine tract (PPT) located upstream of the 3′ss; and the branch site sequence located upstream of the PPT. The mRNA precursor is assembled into a complex, the spliceosome, composed of basal machinery and regulatory proteins. The basal machinery is greatly conserved from yeast to human and consists of proteins and five complexes of small nuclear RNAs assembled with proteins (U1, U2, U4, U5 and U6 snRNPs). These snRNPs form a dynamic network of interactions among themselves and with the conserved splice signals, such as U1–5′ss, U2–branch site and U2–U6 pairing. This network of interactions is required for folding the intron correctly, splicing out each intron, and ligating the flanking exons. The regulatory system is divided into two types of proteins that assist the basal machinery in locating the correct splice junctions. The first type consists of the heterogeneous nuclear ribonucleoprotein (hnRNP) proteins, which are found in the nucleus and cytoplasm and are involved in processes including chromatin remodeling and transcription, in addition to mRNA splicing, exporting and translation [21]. The second type comprises the SR proteins, which also influence mRNA processing, including splicing, exporting and translation. SR proteins bind to relatively short exonic and intronic sequences, usually 4–18 nt, which are generally found up to 150 bases from the regulated splice site [22]. The RS domains of SR proteins are phosphorylated by several different kinases. The phosphorylation modulates protein–protein interactions within the spliceosome, thereby contributing to dynamic structural reorganization during splicing [23]. The binding of these proteins by means of their RNA recognition motif to the exonic and intronic sequences facilitates the recruitment of the basal splicing machinery to the splice junctions [17,24].

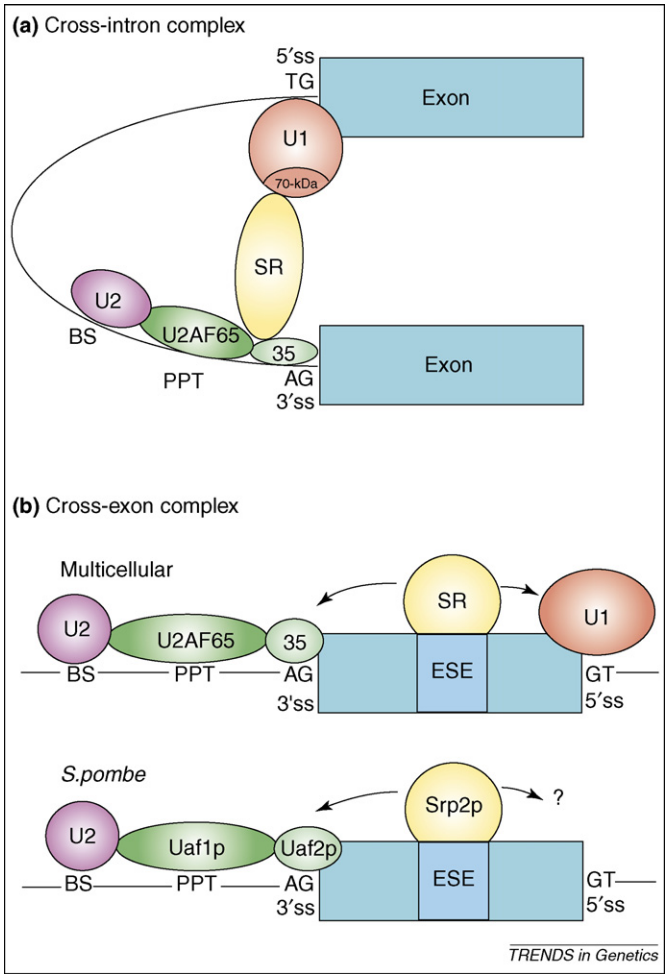

Figure 1. Cross-intron versus cross-exon complexes. **(a)** SR proteins function in a 'cross-intron' recognition complex by bridging between the interactions of U1 snRNP bound to the upstream 5′ss and U2AF65 and 35-kDa subunits bound to the polypyrimidine tract (PPT) and the AG dinucleotide of the downstream 3′ss, respectively. Then, U2AF65 recruits U2 to the branch site sequence (BS). **(b)** Upper panel: SR proteins also facilitate a 'cross-exon' recognition complex (see Supplementary Data). The exons contain exonic splicing enhancers (ESEs) that are binding sites for SR proteins. When an SR protein binds to an ESE, the SR protein recruits U1 snRNP to the downstream 5′ss, and U2AF65 and 35-kDa subunits to the PPT and the 3′ss-AG dinucleotide, respectively. In turn, U2AF65 recruits U2 snRNP to the BS. Lower panel: in *S. pombe*, the SR protein, Srp2p, binds to an ESE and recruits U2AF orthologs (Uaf1p and Uaf2p) to the PPT and 3′ss-AG located on the upstream intron. Then, Uaf1p directs U2 binding to the branch site.

specifically with the U2 small nuclear ribonucleoprotein (snRNP) auxiliary factor (Uaf2p), an ortholog of U2AF35 [14]. In multicellular organisms U2AF35 interacts with the 3′ss-AG and with U2AF65 [16]. The latter binds the polypyrimidine tract (PPT) and directs U2 binding to the branch site sequence (Figure 1). The U2AF65 protein is part of the basal splicing machinery, whereas the U2AF35 ortholog is found in *S. pombe* and not in *S. cerevisiae* [17]. Webb *et al.* also demonstrated that the dependency on Srp2p for suboptimal intron splicing is abolished when the intron contains a strong 3′ss, and the enhancement effect is detected when the exonic binding site is up to 100 nt downstream of the 3′ss [15]. Therefore, the type of interaction found in *S. pombe* is consistent with a model in which the binding of an SR protein to an exonic sequence recruits U2AF65 and influences the splicing of an upstream intron. The distance between the SR-binding site and the PPT-3′ss sequences is important for the ability to maintain the enhancement effect (see Supplementary Data). Also, these results suggest that before the transition from intron to exon definition, SR proteins already supported the basal machinery in the splicing of suboptimal introns.

SR proteins facilitate splicing of weak introns in an SR-free organism

How might the evolution of SR proteins facilitate alternative splicing? The transition from intron to exon definition, probably occurring between *S. pombe* and multicellular organisms in the course of evolution, was the major selective pressure leading to the proliferation of SR genes in multicellular organisms. Proliferation of such genes can give an advantage in assisting the basal splicing machinery in finding short exons in large intronic sequences. The transition from intron to exon definition is also linked to a shift in mechanism – from one that relies only on four conserved splicing signals to direct the basal splicing machinery to the correct exon–intron junctions (as in *S. cerevisiae*), to a system that includes other sequences (and proteins that bind to them) located outside of the four splice signals and providing additional information to help direct the basal machinery to the splice junctions (as in higher eukaryotes).

Investigating this further, Shen and Green [18] found that directing mammalian SR proteins to bind to an exonic sequence of mRNA precursors, in which the 5′ss of that exon was mutated to a suboptimized site, or a mutated branch point sequence, rescued splicing. Remarkably, these findings were obtained in an organism (*S. cerevisiae*)

Please cite this article in press as: Ram, O., Ast, G., SR proteins: a foot on the exon before the transition from intron to exon definition, Trends Genet. (2006), doi:10.1016/j.tig.2006.10.002

that lacks SR proteins. Therefore, these results indicate that SR proteins can support mRNA splicing activity even in an SR-free organism. Also, this finding explains why the 5′ss and the branch site sequence motif are highly conserved in *S. cerevisiae* – they cannot afford to accumulate mutations, because they do not have the backup system provided by SR proteins.

How do SR proteins relate to the evolution of alternative splicing?

According to these findings, we can add another layer to the hypothesis regarding the origin of alternative splicing [3]. The proliferation of SR proteins during the evolution of multicellular organisms from unicellular organisms released the burden from the basal machinery of having to bind efficiently to the four splice signals. Consequently, natural selection permitted mutations that result in suboptimization of certain splice sites (such as U1 base pairing to the 5′ss). This causes the skipping of suboptimal exons on several splicing events and the ligation of the flanking exons, which is the prevalent form of alternative splicing. Thus, SR proteins that bind to exonic sequences were already in existence before the appearance of alternative splicing, and the ability of SR proteins to support recognition of suboptimal introns was, likewise, in existence before the transition from intron to exon definition. Hence, SR proteins shifted from assisting the splicing machinery in the recognition of suboptimal introns to assisting in the placement of the basal machinery across exons. The proliferation of SR proteins during evolution from unicellular to multicellular organisms could be related directly to the abundance of alternative splicing.

A correlation between intron size and exon skipping

The appearance of alternative splicing is presumably related to the expansion of intronic sequences beyond the maximal recognition length of introns by the intron definition system. This could explain the selective pressure leading to shorter internal exons in higher compared with lower eukaryotes [10,11].

The Hertel laboratory found a correlation between the size of the flanking introns and the ability of the internal exon to undergo alternative splicing [19]. They demonstrated that in the fruit fly *Drosophila melanogaster* the extension of an intron beyond 350 nt abrogates splicing. Thus, 350 nt is the maximum length for recognition by intron definition, at least in that organism. The same authors then demonstrated that most of the introns flanking alternatively spliced exons in *D. melanogaster* are larger than 350 nt, whereas constitutively spliced exons are flanked by short intronic sequences [19]. However, such a correlation was less significant in human, raising the possibility that exon definition in human is a mechanism that depends on additional factors that regulate splicing in mammals but are absent in lower organisms. There are also rare occasions of alternative splicing of the exon-skipping type in *Cryptococcus neoformans, Plasmodium flaciparum* and *D. discoideum*, although these organisms possess short introns [20]. This raises an intriguing question about the similarities and differences of alternative splicing regulation in higher eukaryotes compared with these organisms. Namely, it is unclear whether SR proteins regulate exon skipping in these organisms, and whether exon definition is involved in the selection of these exons.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.tig.2006.10.002.

References

1 Graveley, B.R. (2001) Alternative splicing: increasing diversity in the proteomic world. *Trends Genet.* 17, 100–107
2 Barrass, J.D. and Beggs, J.D. (2003) Splicing goes global. *Trends Genet.* 19, 295–298
3 Ast, G. (2004) How did alternative splicing evolve? *Nat. Rev. Genet.* 5, 773–782
4 Kim, H. *et al.* (2004) Estimating rates of alternative splicing in mammals and invertebrates. *Nat. Genet.* 36, 915–916
5 Lander, E.S. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860–921
6 Yeo, G. *et al.* (2004) Variation in sequence and organization of splicing regulatory elements in vertebrate genes. *Proc. Natl. Acad. Sci. U. S. A.* 101, 15700–15705
7 Barbosa-Morais, N.L. *et al.* (2006) Systematic genome-wide annotation of spliceosomal proteins reveals differential gene family expansion. *Genome Res.* 16, 66–77
8 Goguel, V. *et al.* (1991) U1 snRNP can influence 3′-splice site selection as well as 5′-splice site selection. *Genes Dev.* 5, 1430–1438
9 Guthrie, C. (1991) Messenger RNA splicing in yeast: clues to why the spliceosome is a ribonucleoprotein. *Science* 253, 157–163
10 Berget, S.M. (1995) Exon recognition in vertebrate splicing. *J. Biol. Chem.* 270, 2411–2414
11 Robberson, B.L. *et al.* (1990) Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol. Cell. Biol.* 10, 84–94
12 Raible, F. *et al.* (2005) Vertebrate-type intron-rich genes in the marine annelid *Platynereis dumerilii*. *Science* 310, 1325–1326
13 Martin, W. and Koonin, E.V. (2006) Introns and the origin of nucleus-cytosol compartmentalization. *Nature* 440, 41–45
14 Xing, Y. and Lee, C. (2006) Alternative splicing and RNA selection pressure – evolutionary consequences for eukaryotic genomes. *Nat. Rev. Genet.* 7, 499–509
15 Webb, C.J. *et al.* (2005) Exonic splicing enhancers in fission yeast: functional conservation demonstrates an early evolutionary origin. *Genes Dev.* 19, 242–254
16 Forch, P. *et al.* (2003) U2 small nuclear ribonucleoprotein particle (snRNP) auxiliary factor of 65 kDa, U2AF65, can promote U1 snRNP recruitment to 5′ splice sites. *Biochem. J.* 372, 235–240
17 Smith, C.W. and Valcarcel, J. (2000) Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem. Sci.* 25, 381–388
18 Shen, H. and Green, M.R. (2006) RS domains contact splicing signals and promote splicing by a common mechanism in yeast through humans. *Genes Dev.* 20, 1755–1765
19 Fox-Walsh, K.L. *et al.* (2005) The architecture of pre-mRNAs affects mechanisms of splice-site pairing. *Proc. Natl. Acad. Sci. U. S. A.* 102, 16176–16181
20 Collins, L. and Penny, D. (2006) Investigating the intron recognition mechanism in eukaryotes. *Mol. Biol. Evol.* 23, 901–910
21 Caceres, J.F. and Kornblihtt, A.R. (2002) Alternative splicing: multiple control mechanisms and involvement in human disease. *Trends Genet.* 18, 186–193
22 Blencowe, B.J. (2000) Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. *Trends Biochem. Sci.* 25, 106–110
23 Misteli, T. (1999) RNA splicing: What has phosphorylation got to do with it? *Curr. Biol.* 9, R198–R200
24 Cartegni, L. *et al.* (2002) Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat. Rev. Genet.* 3, 285–298

Please cite this article in press as: Ram, O., Ast, G., SR proteins: a foot on the exon before the transition from intron to exon definition, Trends Genet. (2006), doi:10.1016/j.tig.2006.10.002