Seminars in Cell & Developmental Biology xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Seminars in Cell & Developmental Biology

journal homepage: www.elsevier.com/locate/semcdb

Review

Regulation of gene expression programmes by serine–arginine rich splicing factors

Minna-Liisa Änkö*

Department of Anatomy and Developmental Biology, Monash University, Level 3, Building 75, Wellington Road, Clayton, Melbourne, VIC 3800, Australia

ARTICLE INFO

Article history:
Available online xxx

Keywords:
SR protein
Splicing
Gene expression
RNA processing
Development

ABSTRACT

Serine–arginine rich splicing factors (SR proteins) are a family of RNA binding proteins that are essential for development in various model organisms. Although SR proteins are necessary for pre-mRNA splicing in metazoans, their binding is not limited to pre-RNA. SR proteins associate with various classes of RNAs, including intronless transcripts and non-coding RNAs, and regulate many processes during the gene expression pathway. Recent studies taking advantage of high-throughput sequencing and other genome-wide approaches have started to shed light into the distinct and overlapping roles of SR proteins in the regulation of gene expression in cells and have led to the identification of endogenous gene targets. These studies together with animal models where individual SR proteins have been depleted in specific tissues suggest that SR proteins may regulate distinct gene expression programmes through their interactions with RNAs and provide crosstalk between splicing and other regulatory processes.

Contents

Abbreviations: ChIP, chromatin immunoprecipitation; CLIP-seq, UV crosslinking and immunoprecipitation followed by high-throughput sequencing; ESE, exonic splicing enhancer; EJC, exon junction complex; hnRNP, heterogenous nuclear ribonucleoprotein; ISE, intronic splicing enhancer; lncRNA, long non-conding RNA; ncRNA, non-coding RNA; NMD, nonsense-mediated decay; PTC, premature termination codon; RIP-chip, RNA immunoprecipitation followed by microarray analysis; RRM, RNA recognition motif; RRMH, RNA recognition motif homolog; RNP, ribonuclear protein; scaRNA, small Cajal body RNA; SELEX, systemic evolution of ligands by experimental enrichment; snoRNA, small nucleolar RNA; snRNA, small nuclear RNA; SRSF, serine–arginine rich splicing factor; UTR, untranslated region.

* Tel.: +61 3 9905 0622.
E-mail address: minni.anko@monash.edu

http://dx.doi.org/10.1016/j.semcdb.2014.03.011
1084-9521/

Please cite this article in press as: Änkö M-L. Regulation of gene expression programmes by serine–arginine rich splicing factors. Semin Cell Dev Biol (2014), http://dx.doi.org/10.1016/j.semcdb.2014.03.011

1. Introduction

Serine–arginine rich splicing factors (SR proteins) are a conserved family of RNA binding proteins essential for spliceosome assembly and cell survival (Fig. 1). The splicing of messenger RNA precursors (pre-mRNAs) is a critical process for the expression of almost all mammalian genes because the vast majority of genes contain introns. The catalytic molecular machine for this highly dynamic process of splice site selection, intron removal and exon ligation is the spliceosome. However, additional splicing factors, such as SR proteins, are required for the correct recognition of exon–intron boundaries (Fig. 2). Although SR proteins are indispensable for constitutive and alternative splicing in metazoans, their binding to nascent RNA is not limited to intron-containing genes [1–3], implying that SR proteins have functions outside pre-mRNA splicing. A number of studies have now shown that SR proteins play a role in the regulation of many RNA processing and gene expression steps [4]. Since their discovery, it has become clear that SR proteins are not just splicing factors but multifunctional proteins that may link together gene regulatory pathways from transcription to translation. The functional importance of SR proteins *in vivo* is highlighted by the fact that in the cases tested so far, individual SR proteins are required for development in mouse, fly and worm [5–12].

By definition, SR proteins can activate splicing *in vitro* in so-called S100 splicing deficient cell extracts [13]. The core group of 'classical' SR proteins comprises seven proteins with reactivity to mAb104 [14], (reviewed in Refs. [15,16]). The mAb104 antibody recognises a common epitope in the C-terminus of SR proteins, which is enriched in serine (S) and arginine (R) residues, hence the name SR protein. Individual proteins were originally named according to their apparent molecular weight on a SDS-PAGE gel (20 kDa, 30 kDa, 40 kDa, 55 kDa and 75 kDa proteins, Table 1). Each of these seven SR proteins is composed of one or two RNA recognition motif(s) (RRM) and an arginine–serine rich (RS) domain (Fig. 1). The RRM binds to RNA and appears to determine the binding specificity of SR proteins [1,17,18], whereas the RS domain is mainly involved in protein–protein interactions [19] and contains a nuclear localisation signal [20]. The serine residues of the RS domain are targeted for phosphorylation and this phosphorylation status is important for SR protein activity. However, the functional division between the domains may not be clear-cut as biochemical evidence indicates that the RS domain may also contact RNA (reviewed in Ref. [21]) and the RRM may participate in protein–protein interactions [22,23]. Interestingly, one of the SR proteins, SRSF7, contains a Zinc-finger domain which potentially contributes to RNA binding [24]. Recently, SR proteins were re-named in a systemic fashion (SRSF, SR Splicing Factor) and the definition of an SR protein was formalised to a protein with one or two N-terminal RRM(s), followed by a downstream RS domain of at least 50 amino acids with >40% RS content, characterised by consecutive RS or SR dipeptides [25]. Based on this definition, the SR protein family was expanded to twelve proteins with functions in pre-mRNA splicing (Table 1 and Fig. 1). In addition to these twelve proteins, there is a large number of SR-related proteins that share defined structural characteristics and may also interact with SR proteins [26,27].

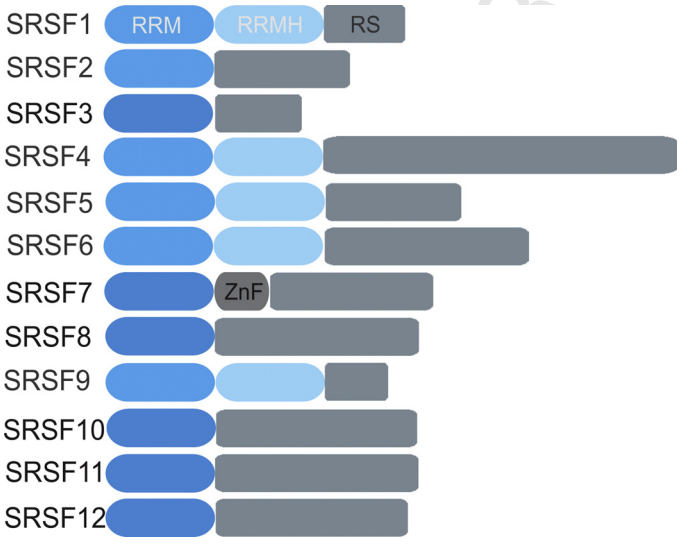

Fig. 1. SR protein family. SR proteins have one or two N-terminal RRM(s), followed by a downstream RS domain of at least 50 amino acids with >40% RS content.

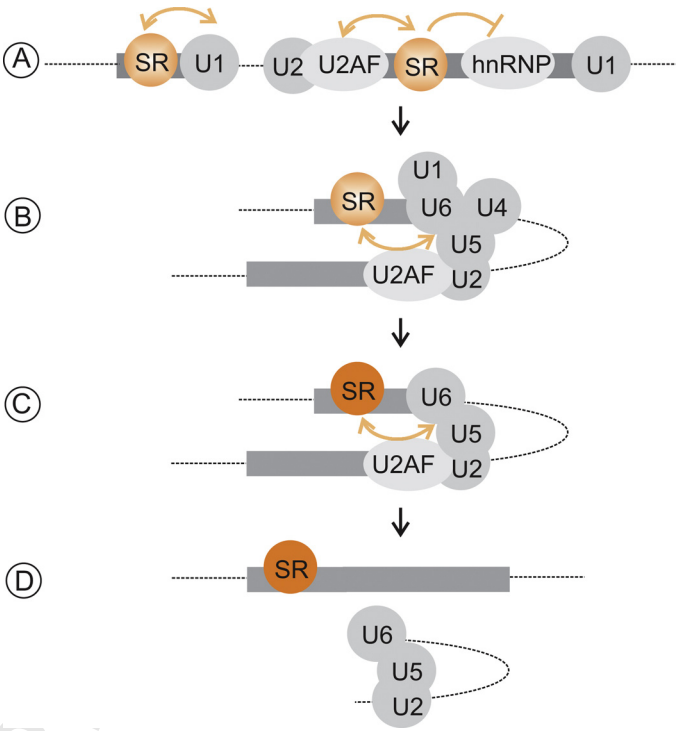

Fig. 2. SR proteins can enhance pre-mRNA splicing by several mechanisms. (A) SR proteins can help the recruitment of spliceosome and exon definition by interacting with U1 snRNP and U2AF at early steps of splicing. They can also antagonise the functions of hnRNP splicing factors. (B) During later steps SR proteins can enhance the recruitment of tri-snRNP (U4/U6-U5) to the splice site. (C and D) SR proteins can remain associated with the spliced mRNA after splicing is complete and/or associate with mature mRNAs.

This review summarises the functions of SR proteins, with an emphasis on recent attempts to determine genes regulated by SR proteins in cells. The focus will be on the 'classical' members of the SR proteins family although the other family members will be included when considered relevant. Determination of SR protein gene targets has led to the identification of novel cellular functions of SR proteins and provided insights into how SR proteins may regulate gene expression programmes in cells.

2. RNA binding specificity of SR proteins

Biochemical and structural studies have shown that both of the RRM domains of SR proteins have the capacity to bind RNA and contribute to the binding specificity of SR proteins [18]. The RS domain is reported to contact RNA at the branchpoint and 5′ splice site [22,28] but the sequence-specificity and potential effect on SR protein binding remains unclear. Before the emergence of genome-wide methods, the study of SR protein interactions with RNA was limited to *in vitro* assays. For instance, mutagenesis studies and

Please cite this article in press as: Änkö M-L. Regulation of gene expression programmes by serine–arginine rich splicing factors. Semin Cell Dev Biol (2014), http://dx.doi.org/10.1016/j.semcdb.2014.03.011