

COMMENTARY

Implementing a Rational and Consistent Nomenclature for Serine/Arginine-Rich Protein Splicing Factors (SR Proteins) in Plants

Andrea Barta,^{a,1} Maria Kalyna,^a and Anireddy S.N. Reddy^b

^a Max F. Perutz Laboratories, Medical University of Vienna, A-1030 Vienna, Austria

^b Department of Biology, Program in Molecular Plant Biology, Program in Cell and Molecular Biology, Colorado State University, Fort Collins, Colorado 80523

Growing interest in alternative splicing in plants and the extensive sequencing of new plant genomes necessitate more precise definition and classification of genes coding for splicing factors. SR proteins are a family of RNA binding proteins, which function as essential factors for constitutive and alternative splicing. We propose a unified nomenclature for plant SR proteins, taking into account the newly revised nomenclature of the mammalian SR proteins and a number of plant-specific properties of the plant proteins. We identify six subfamilies of SR proteins in *Arabidopsis thaliana* and rice (*Oryza sativa*), three of which are plant specific. The proposed subdivision of plant SR proteins into different subfamilies will allow grouping of paralogous proteins and simple assignment of newly discovered SR orthologs from other plant species and will promote functional comparisons in diverse plant species.

SR proteins are a family of important RNA binding proteins, which are conserved in higher eukaryotes and function as essential factors for constitutive and alternative splicing. They contain one or two N-terminal RNA binding domains (RBDs; also known as RNA recognition motifs [RRMs]) and an Arg/Ser-rich (RS) C-terminal region and as such contribute significantly to the proteome complexity of higher eukaryotes. Since their discovery ~20 years ago, they have been studied intensively in a number of organisms, specifically in mammals, *Drosophila melanogaster*, *Caenorhabditis elegans*, and plants. These studies also uncovered several other roles for SR proteins, such as their involvement in mRNA nuclear export, mRNA stability, translation, genome maintenance, and oncogenic transformation (Huang and Steitz, 2005; Long and Caceres, 2009; Zhong et al., 2009). Their multifunctional roles illustrate the importance of SR proteins in regulating gene expression at various levels.

The historical timeline of SR protein discovery and the imprecise definition of what constitutes a bona fide SR protein left the field with somewhat arbitrary classifications and nomenclature of proteins (dis-

cussed in Manley and Krainer, 2010). In particular, the existence of many additional proteins with RS domains that do not necessarily possess an RRM domain introduced further confusion. Due to the importance of SR proteins as regulators for proper gene expression and protein diversity, the splicing community recently proposed a more precise definition of SR proteins and a unified nomenclature for each SR protein (Manley and Krainer, 2010). To begin with, the approach was limited to the best investigated mammalian genes and proteins. Manley and Krainer (2010) proposed that SR proteins be defined solely according to their sequence properties: one or two N-terminal RRMs (RBDs; PF00076) followed by a downstream RS domain of at least 50 amino acids with >40% RS content characterized by consecutive RS or SR repeats. This definition allowed the identification of 12 SR proteins in humans (see Table 1 in Manley and Krainer, 2010).

We supported the initiative for the revised nomenclature of the mammalian SR proteins. However, adapting this system to SR proteins in plants has proved difficult as the plant proteins have certain peculiarities. Phylogenetic studies of SR proteins and the recent completion of several plant genomes revealed a larger number of diverse SR proteins in comparison to those

encoded by metazoan genomes. For example, according to the currently used nomenclature, *Arabidopsis thaliana* possesses 19 SR genes compared with 12 SR genes in humans or seven in *C. elegans*. The genomes of rice (*Oryza sativa*; Iida and Go, 2006; Isshiki et al., 2006) and *Brachypodium* (International Brachypodium Initiative, 2010) encode for 22 and 17 SR proteins, respectively. In addition to clear orthologs for some of the human SR proteins, such as SRSF1 (formerly SF2/ASF), SRSF2/SC35, and SRSF7/9G8, there are plant-specific SR proteins in *Arabidopsis* and other plants that are unusual in their domain structures. For instance, At-RSZ32 and At-RSZ33 have RRMs with significant similarity to human SRSF6/SRp55 and a canonical RS domain; however, they possess two zinc knuckles and have an acidic C-terminal extension rich in Ser and Pro residues. These features also distinguish these proteins from the mammalian SRSF7 and its three *Arabidopsis* orthologs containing one zinc knuckle (At-RSZp21/SRZ21, At-RSZp22/SRZ22, and At-RSZp22a). Four SC35-like (SCL) proteins (At-SCL28, At-SCL30, At-SCL30a, and At-SCL33/SR33) have canonical RRMs with sequence similarity to SRSF2; however, they possess an N-terminal extension rich in Arg, Pro, Ser, Gly, and Tyr residues, which places these proteins in

¹Address correspondence to andrea.barta@meduniwien.ac.at.
www.plantcell.org/cgi/doi/10.1105/tpc.110.078352

COMMENTARY

a plant-specific subfamily. The proteins of the RS subfamily lack the highly conserved SWQDLKD motif in their second RRM and possess an RS domain with many RS dipeptides. In general, although plant SR proteins possess highly conserved RRM, their RS domains are more divergent,

contain additional/novel domains, and seem to have evolved for more specific protein-protein interactions. For example, the RS domains of *Arabidopsis* orthologs of SRSF7 have an RS content below 40%.

Another peculiarity of plant SR proteins is their large and varying number. Both plant-specific SR proteins and those with clear orthologs in metazoan lineages are present in plants as multigene families. For example, 12 of the 19 *Arabidopsis* SR genes exist


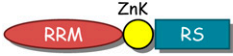




Subfamily name	Aliases	New protein /gene symbol	Accession	Reference [†]
SR subfamily 	At-SRp30	At-SR30	At1g09140	8
	At-SRp34, SR1	At-SR34	At1g02840	6, 8
	At-SRp34a	At-SR34a	At3g49430	11
	At-SRp34b	At-SR34b	At4g02430	11
	Os-SRp32	Os-SR32	Os03g22380	4
	Os-SRp33a	Os-SR33a	Os05g30140	4
	Os-SRp33b	Os-SR33	Os07g47630	4
	Os-SRp20*	Os-SR40	Os01g21420**	3*, 4**
RSZ subfamily 	At-RSZp21, SRZ21	At-RSZ21	At1g23860	1, 9
	At-RSZp22, SRZ22	At-RSZ22	At4g31580	1, 9
	At-RSZp22a	At-RSZ22a	At2g24590	9
	Os-RSZp21a	Os-RSZ21a	Os06g08840	4
	Os-RSZp21b	Os-RSZ21	Os02g54770	4
	Os-RSZp23	Os-RSZ23	Os02g39720	4
SC subfamily 	At-SC35	At-SC35	At5g64200	10
	Os-SC35a	Os-SC34	Os08g37960	4
	Os-SC35b	Os-SC32	Os07g43050	4
	Os-SC35c	Os-SC25	Os03g27030	4
SCL subfamily 	At-SCL28	At-SCL28	At5g18810	10
	At-SCL30	At-SCL30	At3g55460	10
	At-SCL30a	At-SCL30a	At3g13570	10
	At-SCL33, SR33	At-SCL33	At1g55310	2, 10
	Os-SCL25	Os-SCL25	Os07g43950	4
	Os-SCL26	Os-SCL26	Os03g25770	4
	Os-SCL30a	Os-SCL30a	Os02g15310	4
	Os-SCL30b	Os-SCL30	Os12g38430	4
	-	Os-SCL28	Os03g24890	3
-	Os-SCL57	Os11g47830	3	
RS2Z subfamily 	At-RSZ32	At-RS2Z32	At3g53500	10
	At-RSZ33	At-RS2Z33	At2g37340	10
	Os-RSZ36	Os-RS2Z36	Os05g02880	4
	Os-RSZ37a	Os-RS2Z37	Os01g06290	4
	Os-RSZ37b	Os-RS2Z38	Os03g17710	4
	Os-RSZ39	Os-RS2Z39	Os05g07000	4
RS subfamily 	At-RSp31a	At-RS31a	At2g46610	5
	At-RSp31	At-RS31	At3g61860	7
	At-RSp40, At-RSp35	At-RS40	At4g25500	7
	At-RSp41	At-RS41	At5g52040	7
	Os-RSp29	Os-RS29	Os04g02870	4
	Os-RSp33	Os-RS33	Os02g03040	4

Figure 1. Domain Architecture of the *Arabidopsis* and Rice SR Protein Subfamilies and the Newly Proposed Protein/Gene Symbols.

The proteins of the SR subfamily (orthologs of mammalian SRSF1/SF2/ASF) possess an evolutionary conserved SWQDLKD motif in their second RRM followed by an RS domain with characteristic SR dipeptides. The RSZ subfamily (orthologs of mammalian SRSF7/9G8) consists of SR proteins with one Zn knuckle. The SC subfamily (orthologs of SRSF2/SC35) contains proteins with a single RRM followed by an RS domain. The plant-specific SCL subfamily (SC35-like) is similar to SRSF2 (RRM domain) but has an N-terminal charged extension. The proteins of the plant-specific RS2Z subfamily possess two Zn knuckles and have an additional SP-rich region following the RS domain. The plant-specific RS subfamily proteins contain two RRM (without the SWQDLKD motif) followed by the RS domain rich in RS dipeptides. *, truncated protein; **, full-length protein. References: 1, Golovkin and Reddy (1998); 2, Golovkin and Reddy (1999); 3, Iida and Go (2006); 4, Isshiki et al. (2006); 5, Kalyna and Barta (2004); 6, Lazar et al. (1995); 7, Lopato et al. (1996); 8, Lopato et al. (1999b); 9, Lopato et al. (1999a); 10, Lopato et al. (2002); 11, Lorković and Barta (2002).

COMMENTARY

as duplicated pairs of paralogs (Kalyna and Barta, 2004). The SR genes in rice and *Brachypodium* genomes are also extensively duplicated. Whole-genome and extensive segmental duplications are prominent features in the evolution of plant lineages. This amplification resulted in multiple paralogous genes coding for SR proteins in *Arabidopsis*, rice, and other plants, which created several subgroups. Sequence alignments support subdivision into several subfamilies, in which members of each group have similar domain organization and likely originated from a common ancestor (Kalyna and Barta, 2004; Iida and Go, 2006; Kalyna et al., 2006). However, each plant species has experienced whole-genome and segmental duplications to a different extent. The proposed subdivision of plant SR proteins into different subfamilies will allow grouping of paralogous proteins and simple assignment of newly discovered SR orthologs from other plant species. In addition, exon-intron structures of genes coding for SR proteins are conserved across plant species within each subfamily (Iida and Go, 2006; Kalyna et al., 2006) and can serve as an additional criterion for their assignment. Currently, six subfamilies of SR proteins can be identified in *Arabidopsis* on this basis (Figure 1).

As outlined above, the differences between the plant and metazoan SR proteins necessitate a different nomenclature system to account for the special requirements for defining and naming plant SR proteins. We propose the definition of a plant SR protein as follows: one or two N-terminal RRMs (RBDs; PF00076) followed by a downstream RS domain of at least 50 amino acids and a minimum of 20% RS or SR dipeptides. Furthermore, the high conservation of gene structures in plant SR protein subfamilies will allow newly discovered SR proteins to be placed in the proper subfamily.

Using these criteria, *Arabidopsis* has 18 SR proteins. There are two cases of previously annotated *Arabidopsis* SR proteins that must be reconsidered and now fall outside the proposed definition of SR proteins: SR45 and SR45a. SR45 has been regarded as a classical SR protein because it could complement an animal in vitro splicing extract deficient in SR pro-

teins (Ali et al., 2007). However, this criterion was excluded in the recent proposal for mammalian SR protein nomenclature (Manley and Krainer, 2010). In addition, this protein possesses two RS domains (N- and C-terminal), and its closest homolog in humans is RNPS1, an RNA binding protein initially identified as part of the exon junction complex and later found to be involved in posttranscriptional processing and mRNA export (Lykke-Andersen et al., 2001). The second *Arabidopsis* protein, SR45a (Tanabe et al., 2007), is a homolog of metazoan Tra-2 and does not qualify as an SR protein according to the criteria proposed by Manley and Krainer (2010).

We suggest a standardized nomenclature for plant SR proteins, which consist of the following components: (1) a species identifier based on the Latin binomial (e.g., *At* for *Arabidopsis thaliana*; *Os* for *Oryza sativa*); three-letter prefixes can be used in ambiguous cases; (2) an abbreviation of the subfamily (Figure 1); (3) a calculated molecular weight of the longest protein isoform; and (4) a suffix (a, b, c, etc.) where required to distinguish paralogous proteins with the same calculated molecular weight belonging to the same subfamily.

We hope that the unified nomenclature proposed here will facilitate assignment of new plant SR proteins as they are being discovered and will promote functional comparisons in diverse plant species. Although the extensive sequencing of new plant genomes might necessitate definition of additional SR protein families, we believe that the initiative for a clear classification of SR proteins will provide benefits both for established researchers and scientists becoming involved in the field of RNA binding proteins and their functions, especially in an era of growing interest in alternative splicing in plants.

NOTES

The following researchers have endorsed the proposed definition of plant SR proteins and their nomenclature: Keith Adams, Gul Shad Ali, Volker Brendel, John WS Brown, Paula Duque, Sergei A. Filichkin, Robert Fluhr, Maxim Golovkin, Kei Iida, Artur Jarmolowski, Adrian Krainer, Shailesh Lal, Alex Lyznik, Jim Manley, Karen M. McGinnis,

Todd C. Mockler, Patrick Motte, Steve Mount, Shigeru Shigeoka, Ko Shimamoto, and Kazuo Shinozaki.

ACKNOWLEDGMENTS

We thank Yamile Marquez for help with bioinformatics analysis and would like to acknowledge J.W.S. Brown, S. Filichkin, A. Krainer, T. Mockler, J. Manley, and S. Mount for valuable comments and suggestions on the manuscript. A.B. acknowledges funding from the Austrian National Science Foundation (FWF: SFB1711, W1207), EU FP6 Programme Network of Excellence on Alternative Splicing EURASNET and GENAU (ncRNA). A.S.N.R. acknowledges funding from the National Science Foundation, the Department of Energy, and the Office of Naval Research.

Received July 28, 2010; revised September 15, 2010; accepted September 16, 2010; published September 30, 2010.

REFERENCES

- Ali, G.S., Palusa, S.G., Golovkin, M., Prasad, J., Manley, J.L., and Reddy, A.S. (2007). Regulation of plant developmental processes by a novel splicing factor. *PLoS ONE* **2**: e471.
- Golovkin, M., and Reddy, A.S. (1998). The plant U1 small nuclear ribonucleoprotein particle 70K protein interacts with two novel serine/arginine-rich proteins. *Plant Cell* **10**: 1637–1648.
- Golovkin, M., and Reddy, A.S. (1999). An SC35-like protein and a novel serine/arginine-rich protein interact with *Arabidopsis* U1-70K protein. *J. Biol. Chem.* **274**: 36428–36438.
- Huang, Y., and Steitz, J.A. (2005). SRprises along a messenger's journey. *Mol. Cell* **17**: 613–615.
- Iida, K., and Go, M. (2006). Survey of conserved alternative splicing events of mRNAs encoding SR proteins in land plants. *Mol. Biol. Evol.* **23**: 1085–1094.
- International *Brachypodium* Initiative (2010). Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* **463**: 763–768.
- Isshiki, M., Tsumoto, A., and Shimamoto, K. (2006). The serine/arginine-rich protein family in rice plays important roles in constitutive and alternative splicing of pre-mRNA. *Plant Cell* **18**: 146–158.
- Kalyna, M., and Barta, A. (2004). A plethora of plant serine/arginine-rich proteins: Redundancy or evolution of novel gene functions? *Biochem. Soc. Trans.* **32**: 561–564.

COMMENTARY

- Kalyna, M., Lopato, S., Voronin, V., and Barta, A.** (2006). Evolutionary conservation and regulation of particular alternative splicing events in plant SR proteins. *Nucleic Acids Res.* **34**: 4395–4405.
- Lazar, G., Schaal, T., Maniatis, T., and Goodman, H.M.** (1995). Identification of a plant serine-arginine-rich protein similar to the mammalian splicing factor SF2/ASF. *Proc. Natl. Acad. Sci. USA* **92**: 7672–7676.
- Long, J.C., and Caceres, J.F.** (2009). The SR protein family of splicing factors: master regulators of gene expression. *Biochem. J.* **417**: 15–27.
- Lopato, S., Forstner, C., Kalyna, M., Hilscher, J., Langhammer, U., Indrapichate, K., Lorković, Z.J., and Barta, A.** (2002). Network of interactions of a novel plant-specific Arg/Ser-rich protein, atRSZ33, with atSC35-like splicing factors. *J. Biol. Chem.* **277**: 39989–39998.
- Lopato, S., Gattoni, R., Fabini, G., Stevenin, J., and Barta, A.** (1999a). A novel family of plant splicing factors with a Zn knuckle motif: Examination of RNA binding and splicing activities. *Plant Mol. Biol.* **39**: 761–773.
- Lopato, S., Kalyna, M., Dorner, S., Kobayashi, R., Krainer, A.R., and Barta, A.** (1999b). atSRp30, one of two SF2/ASF-like proteins from *Arabidopsis thaliana*, regulates splicing of specific plant genes. *Genes Dev.* **13**: 987–1001.
- Lopato, S., Waigmann, E., and Barta, A.** (1996). Characterization of a novel arginine/serine-rich splicing factor in *Arabidopsis*. *Plant Cell* **8**: 2255–2264.
- Lorković, Z.J., and Barta, A.** (2002). Genome analysis: RNA recognition motif (RRM) and K homology (KH) domain RNA-binding proteins from the flowering plant *Arabidopsis thaliana*. *Nucleic Acids Res.* **30**: 623–635.
- Lykke-Andersen, J., Shu, M.D., and Steitz, J.A.** (2001). Communication of the position of exon-exon junctions to the mRNA surveillance machinery by the protein RNPS1. *Science* **293**: 1836–1839.
- Manley, J.L., and Krainer, A.R.** (2010). A rational nomenclature for serine/arginine-rich protein splicing factors (SR proteins). *Genes Dev.* **24**: 1073–1074.
- Tanabe, N., Yoshimura, K., Kimura, A., Yabuta, Y., and Shigeoka, S.** (2007). Differential expression of alternatively spliced mRNAs of Arabidopsis SR protein homologs, atSR30 and atSR45a, in response to environmental stress. *Plant Cell Physiol.* **48**: 1036–1049.
- Zhong, X.Y., Wang, P., Han, J., Rosenfeld, M.G., and Fu, X.D.** (2009). SR proteins in vertical integration of gene expression from transcription to RNA processing to translation. *Mol. Cell* **35**: 1–10.