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. Reddyb

a Department of Biology, Program in Molecular Plant Biology, Program in Cell and Molecular Biology, Colorado State University, Fort Collins, Colorado 80523
b Max F. Perutz Laboratories, Medical University of Vienna, A-1030 Vienna, Austria

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 (discussed 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) 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.

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.
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 RRMs, 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%.

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 RSZ2 subfamily possess two Zn knuckles and have an additional SP-rich region following the RS domain. The plant-specific RS subfamily proteins contain two RRMs (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. (1999a); 9, Lopato et al. (1999b); 10, Lopato et al. (2002); 11, Lorkovic and Barta (2002).
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 proteins (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


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


Implementing a Rational and Consistent Nomenclature for Serine/Arginine-Rich Protein Splicing Factors (SR Proteins) in Plants
Andrea Barta, Maria Kalyna and Anireddy S.N. Reddy
*Plant Cell;* originally published online September 30, 2010;
DOI 10.1105/tpc.110.078352

This information is current as of June 20, 2017

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>eTOCs</td>
<td>Sign up for eTOCs at: <a href="http://www.plantcell.org/cgi/alerts/ctmain">http://www.plantcell.org/cgi/alerts/ctmain</a></td>
</tr>
<tr>
<td>CiteTrack Alerts</td>
<td>Sign up for CiteTrack Alerts at: <a href="http://www.plantcell.org/cgi/alerts/ctmain">http://www.plantcell.org/cgi/alerts/ctmain</a></td>
</tr>
<tr>
<td>Subscription Information</td>
<td>Subscription Information for <em>The Plant Cell</em> and <em>Plant Physiology</em> is available at: <a href="http://www.aspbo.org/publications/subscriptions.cfm">http://www.aspbo.org/publications/subscriptions.cfm</a></td>
</tr>
</tbody>
</table>

© American Society of Plant Biologists
ADVANCING THE SCIENCE OF PLANT BIOLOGY