

RESEARCH ARTICLE

Open Access

The RST and PARP-like domain containing SRO protein family: analysis of protein structure, function and conservation in land plants

Pinja Jaspers[†], Kirk Overmyer[†], Michael Wrzaczek[†], Julia P Vainonen, Tiina Blomster, Jarkko Salojärvi, Ramesha A Reddy, Jaakko Kangasjärvi*

Abstract

Background: The SROs (SIMILAR TO RCD-ONE) are a group of plant-specific proteins which have important functions in stress adaptation and development. They contain the catalytic core of the poly(ADP-ribose) polymerase (PARP) domain and a C-terminal RST (RCD-SRO-TAF4) domain. In addition to these domains, several, but not all, SROs contain an N-terminal WWE domain.

Results: SROs are present in all analyzed land plants and sequence analysis differentiates between two structurally distinct groups; cryptogams and monocots possess only group I SROs whereas eudicots also contain group II. Group I SROs possess an N-terminal WWE domain (PS50918) but the WWE domain is lacking in group II SROs. Group I domain structure is widely represented in organisms as distant as humans (for example, HsPARP11). We propose a unified nomenclature for the SRO family. The SROs are able to interact with transcription factors through the C-terminal RST domain but themselves are generally not regulated at the transcriptional level. The most conserved feature of the SROs is the catalytic core of the poly(ADP-ribose) polymerase (PS51059) domain. However, bioinformatic analysis of the SRO PARP domain fold-structure and biochemical assays of AtRCD1 suggested that SROs do not possess ADP-ribosyl transferase activity.

Conclusions: The SROs are a highly conserved family of plant specific proteins. Sequence analysis of the RST domain implicates a highly preserved protein structure in that region. This might have implications for functional conservation. We suggest that, despite the presence of the catalytic core of the PARP domain, the SROs do not possess ADP-ribosyl transferase activity. Nevertheless, the function of SROs is critical for plants and might be related to transcription factor regulation and complex formation.

Background

The RCD1 (RADICAL-INDUCED CELL DEATH1) protein is an important regulator of plant stress and developmental responses [1,2]. In *Arabidopsis thaliana* it is a member of a small protein family consisting of RCD1 and five SROs (SIMILAR-TO-RCD-ONE). *RCD1* was first identified as a plant gene able to complement the oxidative stress sensitive phenotype of a yeast strain deficient in the YAP1 transcription factor [3]. Since then it has also been characterized as a major regulator of plant ozone (O₃) tolerance [4]. A loss-of-function

mutation in *RCD1* results in highly pleiotropic phenotypes including increased sensitivity to extracellular reactive oxygen species (ROS), resistance to chloroplastic ROS formation by paraquat (methyl viologen) and ultraviolet radiation, salt sensitivity, aberrant leaf and rosette morphology, early flowering, altered nitric oxide and hormone (jasmonic acid and ethylene) responses, as well as defects in developmental processes, such as root architecture and reproductive development [1,2,4-9]. While *rcd1* displays a vast range of well-characterized phenotypes, the function of its closest ortholog, *SRO1*, is dispensable for normal plant development and stress response [1]. Mutant *sro1* plants exhibit only very subtle phenotypes [2]. However, loss of a single *SRO1* allele in *rcd1* background results in severe developmental defects

* Correspondence: jaakko.kangasjarvi@helsinki.fi

† Contributed equally

Plant Biology Division, Department of Biosciences, University of Helsinki, Viikinkaari 1, FI-00014 Helsinki, Finland

with the *rcd1 sro1* double mutant displaying even more extreme phenotypes [1,2]. This demonstrates unequal genetic redundancy between *RCD1* and *SRO1* in *A. thaliana* [1,10]. In species other than *A. thaliana*, several studies, mostly based on gene expression analysis, suggest roles for *RCD1* and *SRO1* orthologs in hormone signaling, plant development and response to biotic and abiotic stresses [1,2,11-17]. However, the phylogenetic relationships of these proteins to the *RCD1/SRO* gene family members in *A. thaliana* has so far not been characterized.

Another member of the *A. thaliana* SRO family, *SRO5* (*At5g62520*), is transcriptionally induced by ROS in response to salt treatment and is required for the proper response to oxidative stress [18]. It forms an antisense overlapping gene pair with Δ^1 -pyrroline-5-carboxylate dehydrogenase (*P5CHD*). In the presence of both transcripts, a 24-nucleotide siRNA is formed, downregulating expression of *P5CDH* [18]. A salt stress responsive *SRO5* ortholog from tomato can functionally complement the *A. thaliana sro5* mutant [19]. The other members of the SRO protein family, *SRO2*, *SRO3* and *SRO4*, have not been functionally analyzed.

The domain composition of the SROs is unique within plants. While two *A. thaliana* SRO family members contain an N-terminal WWE domain (PS50918 [20]), all of them are characterized by the possession of the core of the poly(ADP-ribose) polymerase (PARP; PS51059) domain and a conserved C-terminal *RCD1-SRO-TAF4* domain (RST domain; PF12174) [1]. The combination of PARP and RST domain is specific to plants but the WWE-PARP domain architecture is widely conserved in organisms as distantly related as humans. The WWE domain is involved in protein-protein interactions and predicted to have a globular structure [20,21]. However, in SROs the function of the WWE domain in dimerization and other protein-protein interactions remains to be shown. The RST domain is a plant specific domain found in plant WWE-PARPs and TAF4s (TBP-Associated Factor 4), a component of the TFIID general transcription factor. The RST domain-bearing C-termini of *RCD1* and *SRO1* are suggested to be critical for the interaction with several, mostly plant specific transcription factors [1].

Protein ADP-ribosylation is a post-translational modification catalyzed by ADP-ribosyl transferases (ARTs) that are present in all eukaryotes except yeast [22]. The major classes of ARTs are the PARPs and mono(ADP-ribosyl) transferases (mARTs). PARPs attach single ADP-ribose units to proteins and catalyze the elongation and branching of long poly(ADP-ribose) chains. PARPs have roles in many processes, including cell death, DNA repair, telomere stability, chromatin remodeling, transcription, and memory [23]. *A. thaliana* has three

PARPs which most closely resemble classical DNA dependent PARPs <http://www.arabidopsis.org/>. The presence of the catalytic core of the PARP domain in *RCD1* and *SROs* suggests an ART or related activity.

The mARTs attach a single ADP-ribose unit to protein substrates. Humans possess both ectoenzymes and intracellular endogenous mARTs [24,25]. To date, in plants no mARTs have been isolated or predicted by bioinformatics [26,27]. Most known human intracellular mARTs resemble PARPs [26] and have until recently been classified as PARPs [24]. There are 11 such human PARPs with various domain structures. HsPARP7, HsPARP12, HsPARP13 and HsPARP14 contain the WWE and PARP domain together with other domains. HsPARP11, with only WWE and PARP but no other conserved domains, is the human protein most similar in domain architecture to *A. thaliana* *RCD1* and *SRO1* and has currently no known function. Given the evolutionary distance between plants and humans it is not clear which, if any, of these proteins are functionally similar to the SROs.

The identification of *RCD1* orthologs in several plant species prompted us to investigate the SRO protein family in a comparative manner. The availability of sequenced and annotated genomes allows the analysis of full gene families *in silico*. We compared the SRO family in several species from evolutionarily divergent branches of the plant kingdom showing a different composition of the family in different plants. In addition, we suggest a naming convention for the family members. We identified the RST domain as a protein-protein interaction domain, analyzed the predicted function of the PARP domain and studied the transcriptional regulation of the gene family members in *A. thaliana*. Based on our findings we propose that, while SROs contain a highly conserved PARP domain, at least *RCD1* does not possess ADP-ribosyl transferase activity. Bioinformatic comparisons suggest this is likely to also apply to several other SROs.

Results and Discussion

Based on their domain composition the *A. thaliana* SROs could be divided into two structural types (Figure 1A). Type A SROs contain an N-terminal WWE domain (PS50918) [20], the catalytic core of the poly(ADP-ribose) polymerase (PARP; PS51059) domain and a C-terminal *RCD1-SRO-TAF4* (RST; PF12174) [1] domain. The type B SROs lack the WWE domain but possess the PARP and RST domains.

The *A. thaliana* SRO protein family consists of six members (Figure 1B), *AtRCD1* and *AtSRO1* to *AtSRO5*. Based on a Neighbour-joining tree using full length protein sequences, they formed distinct groups: *AtRCD1* and *AtSRO1* belong to group I while the others form

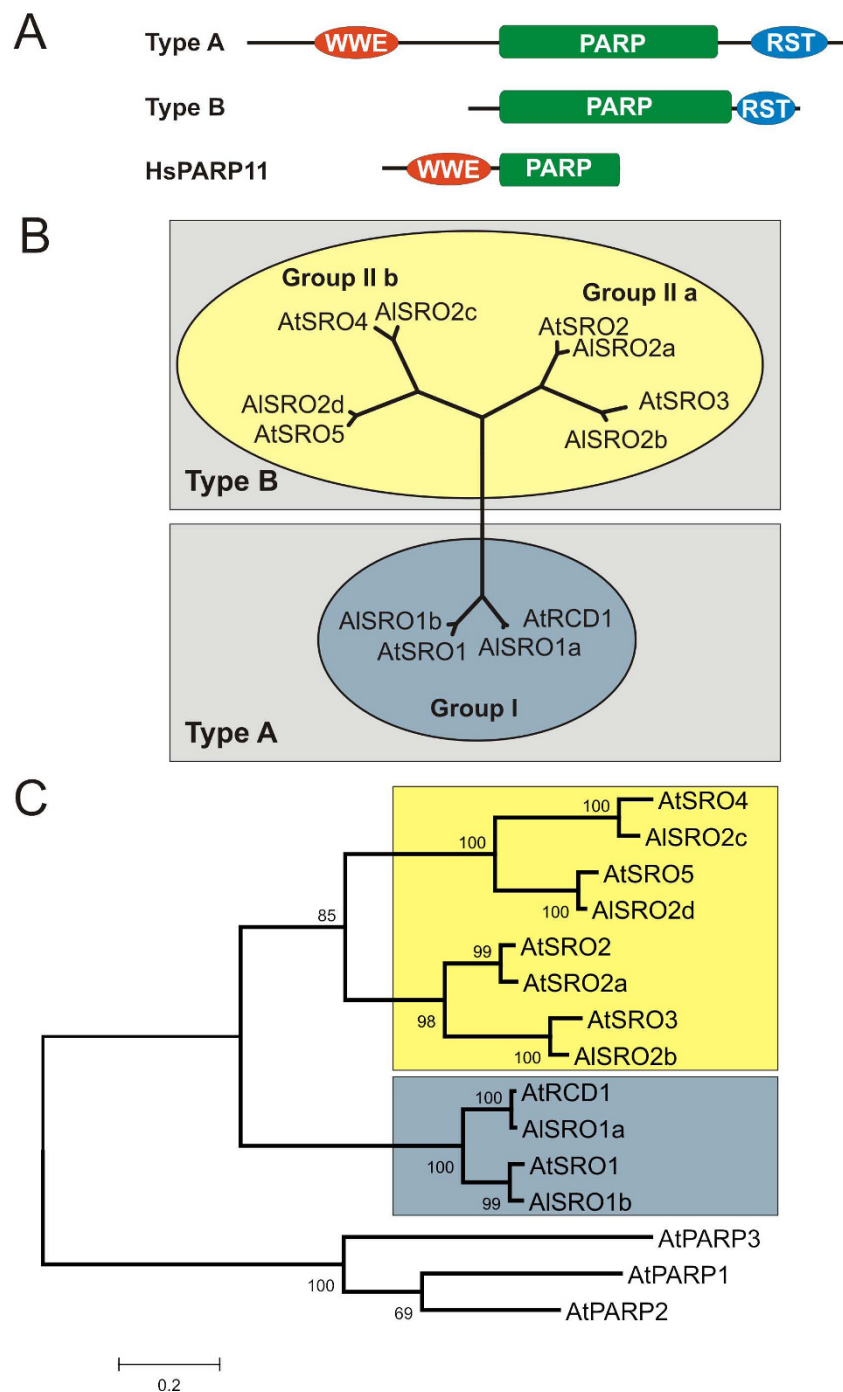


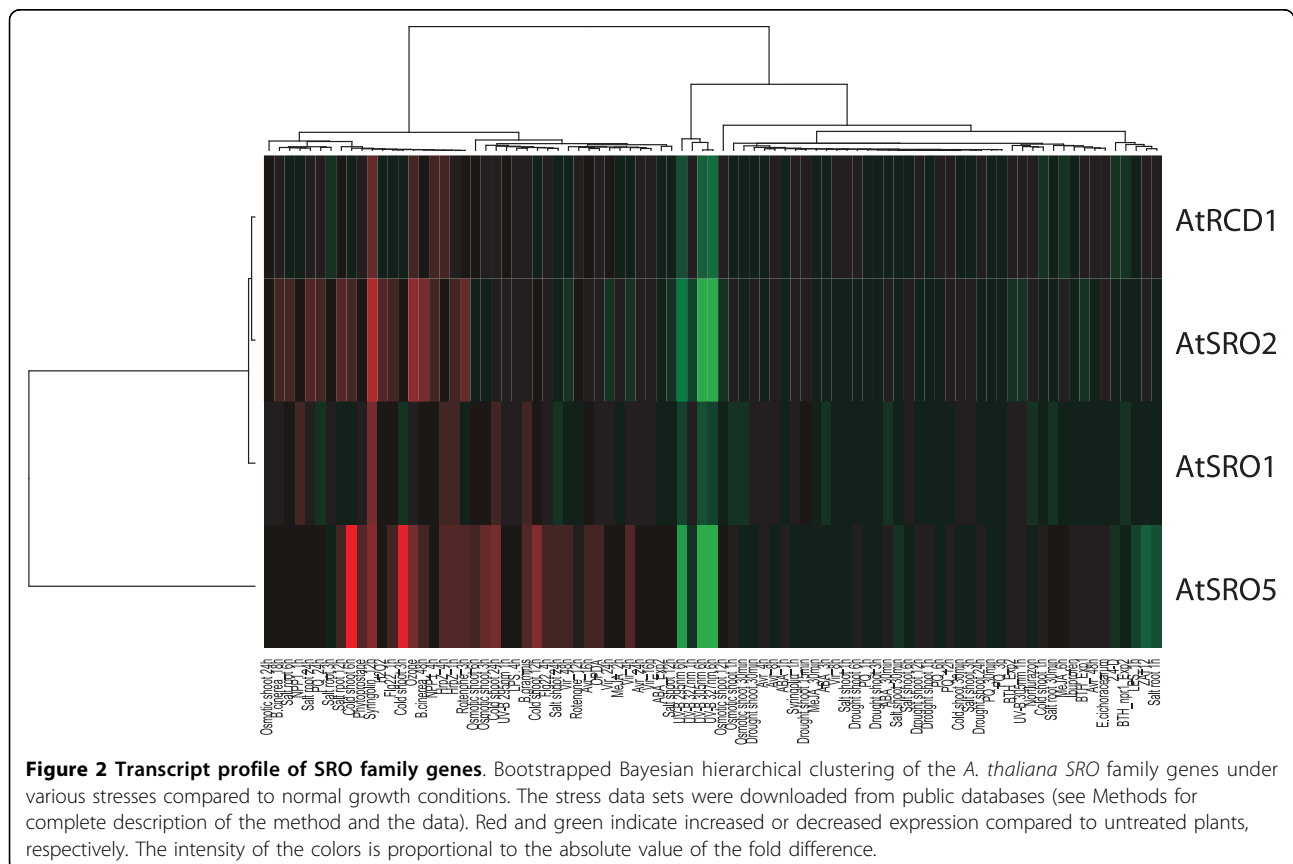
Figure 1 Structural classes and evolutionary relationships of RCD1 and RCD1-like proteins. (A) Schematic diagram depicting the protein structure of *A. thaliana* RCD1/SRO protein family members representing the two structural classes, type A (AtRCD1) and B (AtSRO2). RCD1 and all SROs possess a poly(ADP-ribose) polymerase (PARP) catalytic region (PS51059) and a C-terminal RST (RCD1-SRO-TAF4; PF12174) domain. Additionally, the presence (type A) or absence (type B) of a WWE domain (PS50918) differentiates between the two structural classes. Human HsPARP11 exhibits similar domain composition to the *Arabidopsis* SROs containing a WWE domain and the PARP domain. It is representative of the five human WWE-PARPs although the remaining four have additional conserved domains. (B) *A. thaliana* and *A. lyrata* SROs clustered in three groups in an unrooted Neighbour-joining tree. AtRCD1 and AtSRO1 and the *A. lyrata* orthologs AISRO1a and AISRO1b formed group I, which structurally belongs to type A. AtSRO2 and AtSRO3 and the *A. lyrata* orthologs AISRO2a and AISRO2b formed group IIa. AtSRO4 and AtSRO5 and the *A. lyrata* orthologs AISRO2c and AISRO2d form group IIb. All members of group II belong to structural type B. (C) Neighbour-joining tree of the *A. thaliana* and *A. lyrata* SROs rooted the *A. thaliana* PARPs (AtPARP1, 2 and 3). The SRO proteins clustered together and form a monophyletic group while AtPARP1, 2 and 3 clustered together to form a single outgroup for the SRO protein family.

group II, which is further divided into two subgroups. AtSRO2 and AtSRO3 belong to group IIa and AtSRO4 and AtSRO5 to group IIb. AtRCD1 and AtSRO1 have an identical protein domain structure and belong to structural type A (Figure 1A). HsPARP11 (Figure 1A) and a few other human PARPs possess similar domain structure with an N-terminal WWE domain and a PARP domain as the structural type A SROs, but lack the C-terminal RST domain. Group II SROs (both subgroups) form the structural type B. The closest sequenced relative of *A. thaliana*, *Arabidopsis lyrata*, possesses the same complement of SRO proteins. Orthologs from the two species clustered together in the phylogenetic trees based on the full length protein sequences (Figure 1B) and the PARP domain (Figure 1C).

Transcriptional regulation of the *A. thaliana* SROs

Three *A. thaliana* SROs, *AtRCD1*, *AtSRO1* and *AtSRO5* have previously been functionally characterized. Several studies suggest that the expression of *AtRCD1* and *AtSRO1* is developmentally regulated and only slightly stress responsive [1,2,9], whereas *AtSRO5* has previously been indicated as common stress response gene [28]. To probe transcriptional regulation of the *AtSRO* gene

family, we mined publicly available Affymetrix microarray chip data (see Methods; *AtSRO3* and *AtSRO4* are not represented on the Affymetrix arrays). These results confirmed that *AtSRO5* was the transcriptionally most responsive member of the *SRO* family (Figure 2). In order to verify and complement the microarray data, quantitative real-time RT-PCR (qPCR) analyses indicated that *AtRCD1* and *AtSRO1* exhibited only subtle regulation in response to stress treatments (Figures 2 and 3). Low variation in transcript abundance in response to stress conditions suggests that these proteins are regulated primarily at the post-translational level under our conditions. This is consistent with the observed low and tightly controlled amounts of RCD1 protein [1]. In contrast to our results, Bechtold *et al.* [16] reported a strong increase in *AtRCD1* transcript abundance in response to excess light stress. This difference is most likely due to the intensity and quality of the light used. *AtSRO2*, *AtSRO3* and *AtSRO5* showed changes in transcript levels in response to light stress, salt treatment and exposure to O₃ (Figure 3). *AtSRO5* showed the clearest transcriptional responses to the stress treatments also in the qPCR analysis. No reproducible results were obtained for *AtSRO4* but its presence in EST databases suggests that it is expressed in plants.



	Light stress				Salt stress				O ₃		<i>rcd1-2</i> vs. Col-0
	1 h	2 h	4 h	8 h ^a	2 h	4 h	8 h	12 h	1 h	8 h	
<i>AtRCD1</i>	1.0	1.0	0.6	1.0	1.6	2.7	3	1.1	1.3	2.6	1.1
<i>AtSRO1</i>	1.1	1.4	0.9	0.9	1.1	2.6	2.8	1.9	1.7	2	1.3
<i>AtSRO2</i>	2.7	2.8	4.8	2.4	1.4	6.5	6.4	0.7	1.3	2.7	2.9
<i>AtSRO3</i>	1.0	0.8	0.6	0.1	2.7	8.5	5.6	1.3	2.9	4.2	2.7
<i>AtSRO4</i>	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
<i>AtSRO5</i>	2.7	0.4	0.8	1.7	3.1	8.7	6.5	0.4	0.6	3.5	2.9

Figure 3 qPCR analysis of SRO family genes. Steady state transcript levels of *A. thaliana* SRO family genes were investigated by qPCR. Relative gene expression under light stress, salt stress, and exposure to O₃ and in the *rcd1-2* mutant is shown compared to Col-0 wildtype plants grown under normal conditions. Red indicates elevated and green decreased expression. Black indicates unaltered transcript levels or in the case of *AtSRO4* not reproducible (NR). Numbers indicate relative fold-change ratios. All experiments were repeated three times, one representative experiment is shown.

Expression of the SRO genes was analyzed by qPCR also in the *rcd1-2* mutant. SRO2, 3, and 5 exhibited higher transcript accumulation in *rcd1-2*, suggesting that RCD1 acts as a negative regulator of these other gene family members. The effect could be indirect and due to the *rcd1* mutant being primed for stress responses [1].

The *AtSRO5* gene forms a natural siRNA pair with its neighbouring gene *P5CDH* in *A. thaliana* where they participate in a regulatory network during ROS-mediated salt responses [18]. Interestingly, in *A. lyrata*, grapevine, or poplar the *P5CDH* gene is not located next to the orthologs of *AtSRO5*; none of the *AtSRO5* orthologs overlap with their respective neighbouring genes http://gbrowse.arabidopsis.org/cgi-bin/gbrowse_syn/arabidopsis/. This suggests that the system of *P5CDH* transcript regulation by natural siRNA formation with SROs is specific to *A. thaliana*. In order to address *AtSRO5* gene function in transcriptional regulation, we performed microarray analysis of unstressed *sro5-2* plants. The *sro5-2* allele (GABI-325B05) used in our study carries a T-DNA insertion in the second exon and expresses a truncated transcript [19]. Microarray results revealed several genes with altered expression according to the fold-change ratio (data not shown). However, these differences were not supported as significant by statistical analysis. To verify the array results with an independent method, we analyzed the expression of the genes with the clearest fold-changes by qPCR (Figure 4). Similar to Babajani *et al.* [19], *AtSRO5* itself had increased expression in the *sro5-2* mutant (Figure 4). Only one other gene, *At3g30720*, encoding *QUA-QUINE-STARCH* [29], exhibited reproducible changes of expression levels in the *sro5-2* mutant. The expression of *P5CDH*, the *AtSRO5* cis-antisense gene pair, was not altered according to our results and the study by Babajani *et al.* [19], suggesting that natural siRNA formation might not be the primary regulatory

<i>sro5-2</i> vs. Col-0	
<i>At1g12970</i>	1.1
<i>At1g50890</i>	1.1
<i>At1g56720</i>	0.8
<i>At1g62422</i>	1.0
<i>At1g69680</i>	1.2
<i>At1g79410</i>	1.4
<i>At3g14300</i>	1.0
<i>At3g30720</i>	8.2
<i>At3g42860</i>	1.1
<i>At3g43290</i>	1.1
<i>At3g47060</i>	1.1
<i>At5g14600</i>	1.1
<i>At5g62520</i> (<i>AtSRO5</i>)	532.5

Figure 4 Real-time quantitative PCR analysis of the *sro5-2* mutant. The expression of 13 genes which were most differentially expressed in non-stressed *sro5-2* mutant plants according to microarray results (data not shown) was re-examined by qPCR. Red indicates elevated and black unaltered transcript levels compared to Col-0 wildtype plants. Numbers indicate relative fold-change ratios. All experiments were repeated three times, one representative experiment is shown.

mechanism in unstressed plants despite the elevated *AtSRO5* transcript.

SRO conservation and nomenclature in land plants

To better understand the structure of the SRO protein family in plants, the *A. thaliana* protein sequences were used to identify and analyze the sequences of SROs in several fully sequenced plant genomes (see Table 1 for list of species names and abbreviations). No SRO protein orthologs were found in the sequenced genomes of algae or photosynthetic bacteria (see Methods). Because no sequence data or EST information is available for any of the streptophyte algae, we cannot exclude the possibility that SROs are present in this group. However, the SRO family was present in all land plant genomes analyzed and showed considerable variation in its composition between plant species (Figures 5 and 6).

The lack of clear one-on-one orthology outside the Brassicaceae (see below) rendered naming conventions based on *A. thaliana* impractical for most plant species. Therefore, a new unified nomenclature system is proposed (Figure 5). The key features of this system are: i) *A. thaliana* proteins retain their current names. ii) All RCD1/SRO family members in other species are named SROs and prefixed with a two letter abbreviation of the species scientific name. iii) All SROs are assigned a number designation; i.e. all SRO1s are in group I and SRO2s in group II. iv) Multiple proteins within one group are then assigned an arbitrary letter designation in the order of their discovery to identify them individually. This nomenclature system allows the differentiation between group I and II SROs and will facilitate the comparison of related SROs between species. All proteins

used in this study have been named according to these conventions (Figure 5).

Representation of SRO groups and structural types in land plants

ScanProsite [30,31] and SMART [32,33] were used to identify conserved domains in the SRO protein sequences. The catalytic core of the PARP domain was the most consistently conserved feature of all identified SRO proteins. Therefore, the PARP domains were used for the construction of a phylogenetic Neighbour-joining tree (Figure 6). The tree was rooted using *A. thaliana* classical PARP proteins (AtPARP1, 2 and 3) as an out-group. AtRCD1/ALSRO1a and also AtSRO1/ALSRO1b from both *Arabidopsis* species grouped tightly and, along with SROs from grapevine, poplar, castor bean, rice, and *Brachypodium distachyon*, formed the subgroup Ia. These proteins are of the structural type A containing WWE, PARP, and RST domains (Figure 1A). The second subgroup Ib contains only proteins from the grasses rice and *B. distachyon*. This subgroup includes only structural type A proteins (Figure 1A). The orthologs from the moss *Physcomitrella patens* and the representative of basal vascular plants *Selaginella moellendorffi*, together with sequences from castor bean, poplar, rice, and *B. distachyon*, formed the subgroup Ic. These proteins retain the PARP and RST domains while the WWE domain is only present in PtSRO1c, OsSRO1c, and BdsSRO1c. Of the group Ic members in which no WWE domain was detected, only PpSRO1c appears to be a full-length sequence (Figure 5).

The group IIa (Figure 6) contains AtSRO2 and 3, which grouped with their orthologs from *A. lyrata*

Table 1 Species of sequenced plant genomes used in this study

Genomes					
Species	Abbr.	Common Name	Clade	Ref.	Data Source
<i>Arabidopsis thaliana</i>	At	Thale cress	Dicot/Eurosid II	[65]	TAIR
<i>Arabidopsis lyrata</i>	Al	Lyrate rock cress	Dicot/Eurosid II	-	TAIR
<i>Brachypodium distachyon</i>	Bd	Purple false-brome	Monocot/Poales	[66]	BDB
<i>Oryza sativa ssp. japonica</i>	Os	Rice	Monocot/Poales	[67]	RGADB
<i>Physcomitrella patens</i>	Pp	Club moss	Bryophyte	[68]	JGI
<i>Populus trichocarpa</i>	Pt	Poplar	Dicot/Eurosid I	[69]	JGI
<i>Ricinus communis</i>	Rc	Castor bean	Dicot/Eurosid I	-	CVI
<i>Selaginella moellendorffi</i>	Sm	Spikemoss	Lycophyte	-	JGI
<i>Vitis vinifera</i>	Vv	Grapevine	Dicot (basal Rosid)	[70]	VGC

List of the species used in this analysis including their abbreviation (abbr.), common name, phylogenetic classification (clade), and sources of data used. Analysis of all species utilized resources of the NCBI: National Center for Bioinformatics <http://www.ncbi.nlm.nih.gov/>. Additional web resources are listed below according to the abbreviations: PZ, Phytozome <http://www.phytozome.net/>; TAIR, the Arabidopsis Information Resource <http://www.arabidopsis.org/>; JGI, Joint Genome Initiative (Poplar: <http://genome.jgi-psf.org/Poptr11/Poptr11.home.html>; *Physcomitrella*: <http://genome.jgi-psf.org/Phypa11/Phypa11.home.html>; *Selaginella*: <http://genome.jgi-psf.org/Selmo1/Selmo1.home.html>); CVI, Craig Venter Insititute <http://castorbean.jcvi.org/>; BDB, *Brachypodium* database <http://www.brachypodium.org/>; RGADB, Rice Genome Annotation database <http://rice.plantbiology.msu.edu/>; VGC (Vitis Genome Consortium), The French-Italian Public Consortium for Grapevine Genome Characterization <http://www.cns.fr/spip/vitis-vinifera-e.html>.

Species	Group I						Group IIa						Group IIb					
	Name	Identifier	size	WWE	PARP	RST	Name	Identifier	size	WWE	PARP	RST	Name	Identifier	size	WWE	PARP	RST
<i>Arabidopsis thaliana</i>	AtRCD1	At1g32230	589	+	+	+	AtSRO2	At1g23550	323	-	+	+	AtSRO4	At3g47720	316	-	+	+
	AtSRO1	At2g35510	568	+	+	+	AtSRO3	At1g70440	305	-	+	+	AtSRO5	At5g62520	309	-	+	+
<i>Arabidopsis lyrata</i>	AlSRO1a	fmRNAsh2_kg.1_3367_AT1G32230.1	585	+	+	+	AlSRO2a	Al_scaffold_0001_2476	317	-	?/-	+	AlSRO2c	fmRNAsh1_pm.C_scaff_ old_5000924	331	-	?/-	+
	AlSRO1b	fmRNAsh2_kg.4_1571_AT2G35510.1	596	+	+	+	AlSRO2b	fmRNAsh2_kg.2_1294_AT1G70440.1	304	-	+	+	AlSRO2d	jgiAraly1496430(fgene sh2_kg.8_2246_AT5 G62520.1	312	-	+	+
<i>Brachypodium distachyon</i>	BdSRO1a	Bd3g34250.1	615	+	+	+												
	BdSRO1b	Bd1g01170.1	562	+	+	+												
	BdSRO1c	Bd1g69140.1	418	+	+	+												
	BdSRO1d	Bd5g25800.1	432	+	+	+												
	BdSRO1e	Bd1g63360.1	435	+	+	+												
<i>Oryza sativa ssp. japonica</i>	OsSRO1a	Os10g42710	499	+	+	+												
	OsSRO1b	Os03g63770	481	+	+	+												
	OsSRO1c	Os03g12820	470	+	+	+												
	OsSRO1d	Os06g13860	507	+	+	+												
	OsSRO1e	Os04g57640	499	+	+	+												
<i>Physcomitrella patens</i>	PpSRO1a	Pp1s341 26V2.1	202	#	+	+												
	PpSRO1b	Pp1s215 51V2.1	253	#	+	+												
	PpSRO1c	Pp1s30 92V2.1	733	-	+	+												
<i>Populus trichocarpa</i>	PtSRO1a	eugene3.00030694	615	+	+	+	PtSRO2a	fgenes4_pg.C_LG_XV_ III000151	335	-	+	+	PtSRO2e	eugene3.00120745	347	-	+	+
	PtSRO1b	eugene3.00290239	623	+	+	+	PtSRO2b	fgenes4_pg.C_LG_VI_ 001588	386	-	+	+	PtSRO2f	eugene3.00150557	351	-	+	+
	PtSRO1c	fgenes4_pg.C_LG_II0_ 01036	464	+	+	+	PtSRO2c	eugene3.00061639	347	-	+	+						
							PtSRO2d	fgenes4_pg.C_LG_VI_ 001587	330	-	+	+						
<i>Ricinus communis</i>	RcSRO1a	gi 223541972	536	+	+	+	RcSRO2a	gi 223535337	327	-	+	+	RcSRO2b	gi 223550512	374	-	+	+
	RcSRO1b	gi 223550716	395	#	+	+												
<i>Selaginella moellendorffii</i>	SmSRO1a	C_860065	424	#	+	-												
<i>Vitis vinifera</i>	VvSRO1a	LOC100245552	610	‡	+	+	VvSRO2a	LOC100260173	325	-	+	+	VvSRO2b	LOC100265712	310	-	+	+
	VvSRO1b	LOC100260925	604	‡	+	+												

Figure 5 SRO Orthologs in Sequenced Plant Genomes. All SRO sequences used for analyses are listed with names according to the proposed nomenclature and their original identifiers. The length of the proteins in amino acids (AAs; size) and the presence (+) or absence (-) of potential conserved domains (WWE P550918, PARP P551059, RST PF12174) are indicated. Proteins predicted to lack domains because they are not full length are indicated (#). Domains present but with low statistical support are indicated with (‡). Data source: NCBI (National Center for Bioinformatics, <http://www.ncbi.nlm.nih.gov/>). Additional web resources are listed below: PZ, Phytozome <http://www.phytozome.net/>; TAIR, the Arabidopsis Information Resource <http://www.arabidopsis.org/>; JGI, Joint Genome Initiative (Poplar: http://genome.jgi-psf.org/Poptr1_1/Poptr1_1.home.html); *Physcomitrella*: http://genome.jgi-psf.org/Phypa1_1/Phypa1_1.home.html; *Selaginella*: <http://genome.jgi-psf.org/Selmo1/Selmo1.home.html>; CVI, Craig Venter Institute <http://castorbean.jcvi.org/>; BDB, Brachypodium database <http://www.brachypodium.org/>; RGADB, Rice Genome Annotation database <http://rice.plantbiology.msu.edu/>. Two SROs from *Brachypodium*, Bradi2g10720.1 and Bradi1g01340.1, were only present as very short and incomplete predictions, and thus could not be assigned to any group.

(AlSRO2a and b, respectively). The other members of group II are VvSRO2a, RcSRO2a, and a group of 4 closely related orthologs from poplar (PtSRO2a, b, c and d). The group IIb (Figure 6) contains AtSRO4 and 5, which clustered together with AlSRO2c and d; as well as VvSRO2b, RcSRO2b, and PtSRO2e and f. The group II (IIa and b) contains only SRO members with domain structure of type B (PARP and RST domain). Strikingly, *P. patens*, *S. moellendorffii*, rice, and *B. distachyon* do not contain proteins that cluster together with group II (Figure 6) suggesting that this group is specific for eudicots.

As described before, one to one orthology exists between the SROs from *Arabidopsis* species *A. thaliana* and *A. lyrata*, as evidenced by the tight clustering in cladograms, (Figure 1B and 1C; Figure 6). In *Arabidopsis*, SROs were always present in pairs consistent with a previously proposed duplication event ([2,34], Plant Genome Duplication Database <http://chibba.agtec.uga.edu/>

duplication/index/home. Similar duplications are documented for several other gene families, e.g. the B3 DNA-binding superfamily [35]. SRO group I members of other, more distantly related plant species lacked such pairing and bore no greater similarity to either AtRCD1 or AtSRO1 but rather formed a sister branch within group I. This raises the question of when the duplications occurred. An analysis of available expressed sequence tags (ESTs) from *Brassica rapa*, *Brassica oleracea* and *Brassica napus* revealed the presence of distinguishable orthologs for AtRCD1 and AtSRO1 in *Brassica* species (Additional file 1). This suggests that the split between AtRCD1 and AtSRO1 might have occurred during the diversification of the Brassicaceae, while other plant species retained so-called “co-orthologs” to AtRCD1/AtSRO1 [36]. These refer to sister groups related equally to both proteins, which are derived from the expansion of paralogous genes in the individual species. The situation was similar for group

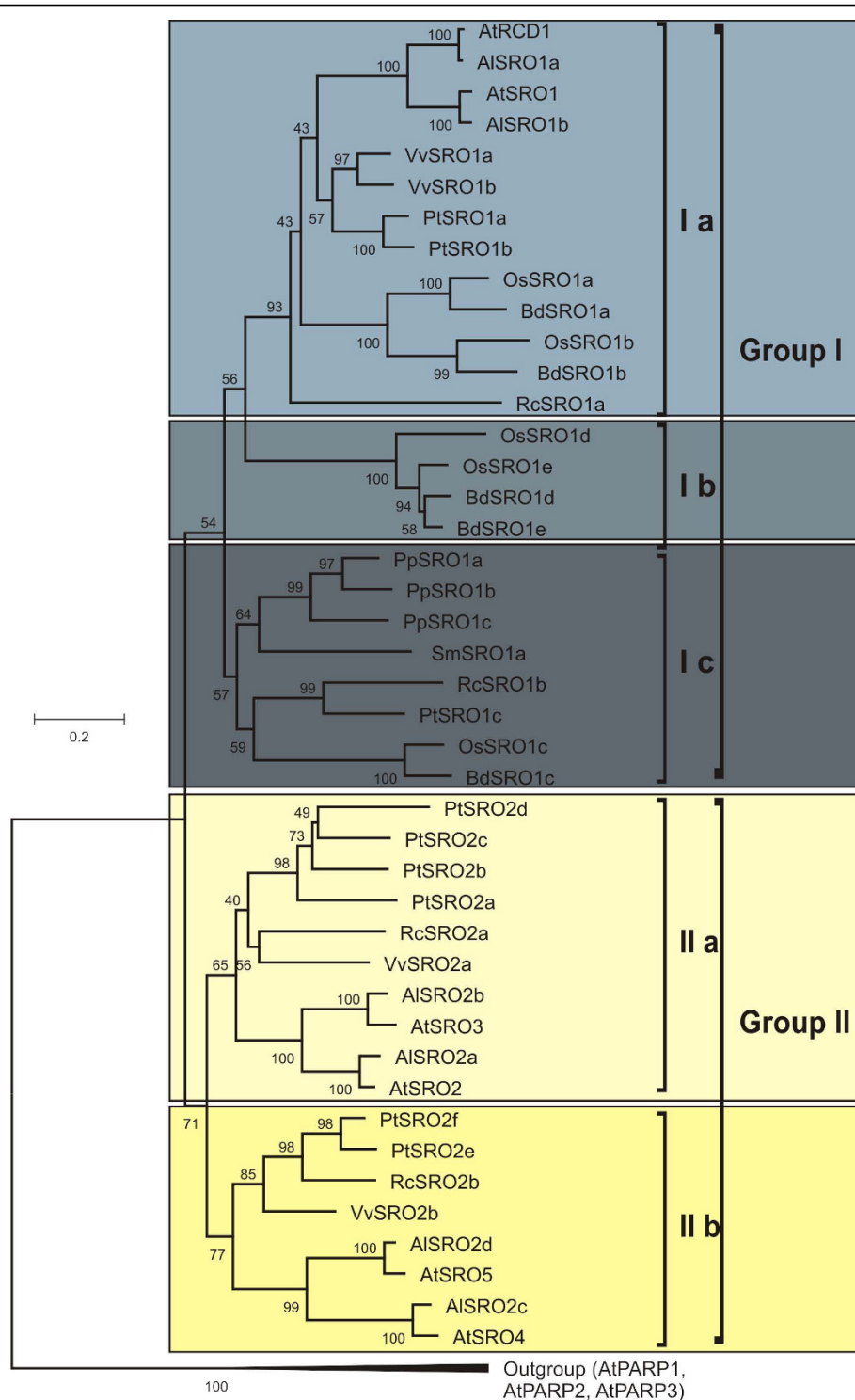


Figure 6 Neighbour-joining tree of the PARP domains of the plant SRO protein family. The PARP domains of the SRO proteins from the sequenced genomes of *A. thaliana*, *A. lyrata*, *Vitis vinifera*, *Ricinus communis*, *Populus trichocarpa*, *Oryza sativa ssp. japonica*, *Brachypodium distachyon*, *Physcomitrella patens* and *Selaginella moellendorffii* were identified and aligned. Subsequently, a Neighbour-joining phylogenetic tree was constructed using MEGA4. AtPARP1, 2 and 3 were used as outgroups. Plant SROs could be classified into two groups. Group I contained SROs from all included plant species and could be further divided into three subgroups (Ia, Ib and Ic) according to the C-terminal RST domain. Most SROs in group I belonged to structural type A. The members of group II (a and b) without exception belonged to structural type B.

IIa; *Brassica* contained ESTs which can be assigned as orthologous to either SRO2 or 3 (Additional file 1). In contrast, while *Brassica* group IIb orthologs were found for AtSRO5, no sequences similar to AtSRO4 were found. However, it remains unclear if this indicates the absence of an AtSRO4 ortholog from *Brassica*, or if this gene was simply missing from the current EST collections due to low expression levels.

These results demonstrate the presence of group I SROs with a conserved structure and domain architecture in all the genomes studied here and suggests their presence in all extant plant species, while group II SROs are unique to eudicot plants. Intriguingly, both monocot species analyzed possess only group I SROs. The lack of group II in members of the more basal plant groups suggests that the origin of the SROs lies within group I, and that group II represents a later development. It is possible that the group II evolved within eudicots only after the dicot-monocot split, or that at least some monocots, represented in this study by two grasses, have lost these groups after these plant lineages diverged more than 120 million years ago [37]. Resolving this question will require investigation of further genomes especially species from the basal branches of angiosperms and gymnosperms, which are not currently available. Several informative plant species, including loblolly pine (*Pinus taeda*, a coniferous gymnosperm) are currently being sequenced.

The conservation of the RST domain between plant groups

A novel conserved domain in the C-terminus of plant SROs was identified recently [1]. This RST domain is also present in TAF4 (Figure 7A), which is a component of several multimeric protein complexes including primarily the general transcription factor TFIID involved in transcriptional initiation [38,39]. The RST domain is distinct from the conserved TAF4 superfamily-defining domain (PF05236), which is required for the assembly of the TFIID complex (Figure 7A; [1,38]). Here the analysis of the RST domain has been expanded, demonstrating that it is present in all known SRO family members (Figure 5). In the few cases of SROs without an RST domain, the gene annotation was questionable and requires further verification through mRNA support for the gene model (see Methods).

Alignments of the C-terminus of SRO family members from different plant species, representing all groups and subgroups, demonstrated that the RST domain is universally conserved (Figure 7B). The SRO group I was subdivided into three subgroups (Ia, Ib and Ic) based on the sequence of the PARP domain (Figure 6) and analysis of the RST domains resulted in the same grouping (Figure 7B). Members of the groups Ia and Ib have an

approximately 20 AA long extension in the N-terminus of the RST domain compared to the members of the groups Ic and II. Since the group Ib contains SROs from *P. patens* together with SROs from grasses and the eudicots castor bean and poplar, it might represent an ancestral SRO group. A strong conservation of a large number of aliphatic AAs in the N- and C- termini of the RST domain, with a strictly conserved tyrosine in the middle of the domain and two conserved positively charged AAs in the second half of the domain, was striking (Figure 7B). The strong conservation of aliphatic AAs in the C-termini of the SRO proteins points to a conserved alpha-helical structure. This sequence preservation implies strong functional constraints for the RST domain during the diversification of the SRO protein family, possibly ensuring that a critical structure of the SRO C-terminus is retained in spite of sequence divergence.

The functional domains of the *A. thaliana* SRO proteins

The RST domain mediates transcription factor interactions

AtRCD1 interacts with several transcription factors (TFs) in the yeast 2-hybrid system (Y2H) and *in vitro*. The WWE and PARP domains are dispensable for these interactions [1,3]. Analysis of mutants lacking the RST domain of AtRCD1 and AtSRO1 demonstrated the significance of this TF-interacting domain for plant development and stress responses. In contrast to AtRCD1, AtSRO1 only interacts with a subset of these TFs [1]. The C-terminus of AtRCD1 is 18 AAs longer than that of AtSRO1 and thus could account for its broader range of TF interactions. To further characterize the RCD1-TF interactions, we constructed C-terminal truncations of AtRCD1 and tested them for interaction with DREB2A and COL10 (Figure 8), two known AtRCD1 interacting TFs [1]. Deletion of the 18 AA extension did not affect the RCD1-TF interactions and also the next three AAs (Q569-K517) were dispensable. However, deletion of further nine or more AAs (N568-L559), which extend into the conserved RST domain, disrupted interactions supporting the proposed role for RST as a functional protein interaction domain. AA D552 in AtSRO1 is absent from AtRCD1 (and all other SROs), and was thus another candidate for the observed differences in the interactions. However, deletion of this residue did not affect the AtSRO1-TF interactions (data not shown). Thus, the determinants of interaction specificity must lie in the other residues within the RST domain or elsewhere in the protein.

To address if the conserved SRO5-RST domain is also a TF-interaction domain we screened the REGIA (TF) collection [1,40] with AtSRO5, a group IIb SRO, which has been shown to be involved in salt stress responses [18]. AtSRO5 interacted with 13 TFs out of the more

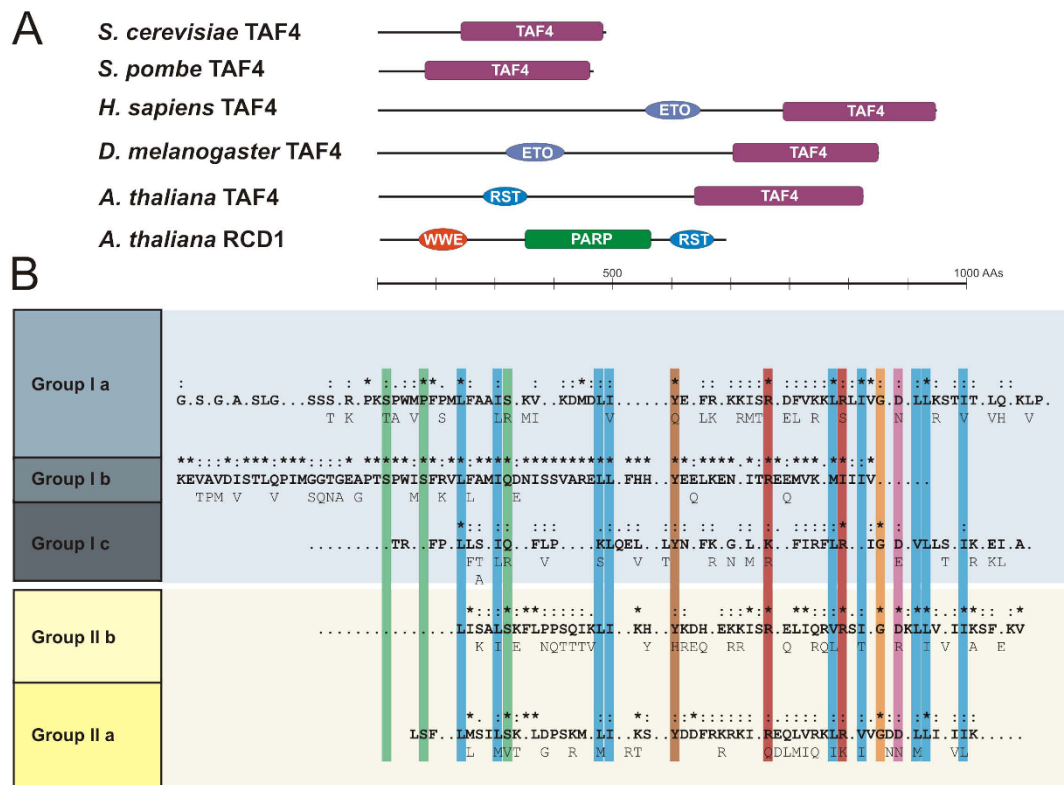


Figure 7 The RST domain of the plant SRO protein family contains a strongly conserved amino acid pattern. (A) Domain structure of AtRCD1 and TAF4s from multiple species (*Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Homo sapiens*, *Drosophila melanogaster*). All TAF4s have the conserved TAF4 superfamily domain (TAF4; PF05236). Yeast TAF4s lack an N-terminal extension while metazoan TAF4s have an extension bearing an ETO domain (ETO/TAFH domain; PF07531), which is a known transcription factor-recruitment domain [43]. Plant TAF4s also have an N-terminal extension that lacks the ETO domain but bears the structurally unrelated plant-specific RST (RCD1-SRO-TAF4; PF12174) domain. TAF4 RST has not been tested for TF interaction, however, the RST domain from AtRCD1 is required for interaction with multiple TFs. AtRCD1 also bears PARP-like (P551059) and WWE (P550918) domains. (B) The C-terminal RST domain of the different groups and subgroups (Ia, Ib, Ic, IIa, IIb) of the plant SRO protein family were aligned using ClustalW and Boxshade. Consensus sequences for each group or subgroup are depicted in bold characters and marked according to similarity: conserved (*), strong similarity (.), weak similarity (.) using Boxshade. Under the sequence, alternatives for AAs are shown. AAs with similar chemical properties are indicated using colored bars. Green indicates polar, non-charged, non-aliphatic residues. Blue indicates the most hydrophobic AAs. Red indicates positively charged AAs. Magenta highlights acidic residues. Orange shows glycine and brown indicates tyrosine.

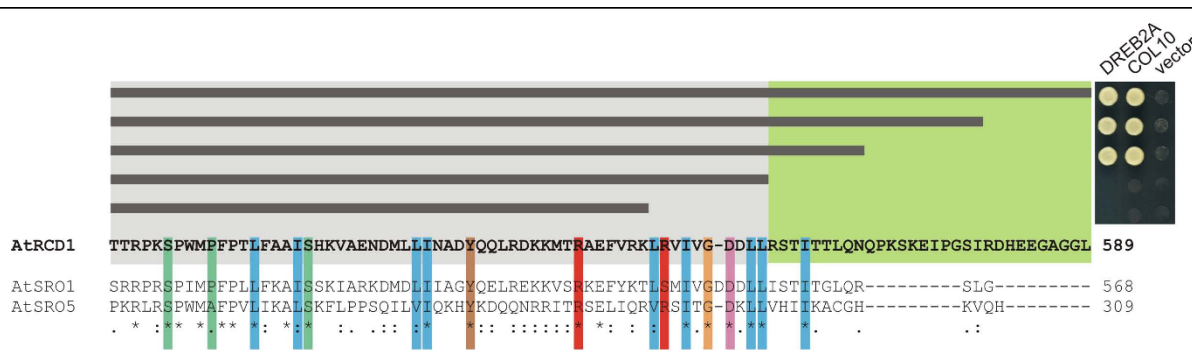


Figure 8 The RST domain of AtRCD1 is required for the TF interactions. The C-terminus of AtRCD1 was truncated to determine the minimum protein length capable of interacting with TFs. The dark gray horizontal bars above the AtRCD1 protein sequence denote the different constructs. Green background indicates interaction and gray background the lack of it. Yeast spots from each interaction test are depicted in the panel on the right. The ClustalW alignment of AtRCD1, AtSR01 and AtSR05 is included for comparison of the RST structure in different proteins. Highlighted AAs in the protein sequences are as in figure 7.

than 1300 present in the collection (Figure 9). Three TFs belong to the AP2/ERF TF family and two to the NAM/NAC and bHLH families each. Five of these TFs interact also with AtRCD1, and DREB2A with both AtRCD1 and AtSRO1 [1]. In addition, AtSRO5 interacted with 3 proteins that were not recovered with full-length AtRCD1 but interacted with a truncated version, which lacks the WWE domain (PCT), thus resembling the AtSRO5 domain structure. Three TFs (AtMYB29, WRKY46 and HSFA1E) were unique interaction partners for AtSRO5, although AtRCD1 interacted with other members of the same TF families [1]. AtSRO5 was previously reported to localize to mitochondria [18]. However, bioinformatic prediction of its subcellular localization rather suggested a different targeting of the protein. This, together with the multiple TF interactions of AtSRO5 prompted us to investigate the cellular distribution of the AtSRO5 protein.

Ectopic expression of AtSRO5-GFP in *A. thaliana* seedlings showed that AtSRO5 localized to several dot-like structures in the nucleus (Figure 10, panels A-C). The results were verified by transient expression of the same construct in *Nicotiana benthamiana* leaves (data not shown). The difference in the observed subcellular localization could be due to the use of different expressions systems. Thus we cannot exclude that AtSRO5 localizes to mitochondria under certain conditions.

These results give possible biological relevance to the interactions between AtSRO5 and TFs. Constant communication between the mitochondria and the nucleus is required for normal cellular function [41]. AtSRO5 might participate in bidirectional interorganellar signaling and play a role in regulating nuclear gene expression through the TF interactions. However, the implications

of AtSRO5 localization to other cellular compartments in addition to the mitochondria require further studies to reveal its significance.

The high number of TF interactions in the Y2H screen demonstrates functional conservation of the RST domain and its importance for protein-protein interaction (this study, [1]). The RST domain is also present in plant TAF4 proteins. Human and *Drosophila* TAF4s have an N-terminal extension carrying the ETO-TAFH domain (Figure 7A). This domain recruits various transcription factors to the TFIID initiation complex and thereby participates in the regulation of transcription [42,43]. The ETO-TAFH domain is missing from plant TAF4 proteins; instead, the TAF4 N-terminus bears the RST domain (Figure 7A). Its presence and position in relation to other domains suggests that the RST domain could be functionally equivalent to other, animal specific, TF-interaction domains. Strong conservation between the RST domains from TAF4 and the SROs could hint towards a common function of TF binding. TF recruitment to TFIID by TAF4 RST is a paradigm for transcriptional regulation. Competition for, or modification of, common TF interaction partners is a model for the modulation of TAF4 dependent processes by the SROs. The future challenge will be to resolve the structure of several highly similar RST domains including AtRCD1, AtSRO1, AtSRO5 and also TAF4s. This together with mutagenesis and deletion studies based on the comparisons (Figures 8 and 9) will help to understand the basis of the specificity of the TF interactions. *In planta* verification of the interactions and competition experiments between SROs, TFs, and TAF4s will be required to determine the significance of the protein-protein interactions.

Name	TF		AtSRO5	AtRCD1		AtSRO1
	AGI	Family		FL	PCT	
DREB2A	At5g05410	AP2/ERF	+	+	+	+
DREB2B	At3g11020	AP2/ERF	+	+	+	-
ATERF14	At1g04370	AP2/ERF	+	-	-	-
ANAC013	At1g32870	NAM/NAC	+	+	+	-
ANAC046	At3g04060	NAM/NAC	+	+	+	-
bZIP	At2g16770	bZIP	+	+	+	-
COL16	At1g25440	zinc finger	+	-	+	-
bHLH	At1g51140	bHLH	+	-	+	-
bHLH	At3g21330	bHLH	+	-	-	-
ARR11	At1g67710	ARR	+	-	+	-
ATMYB29	At5g07690	MYB	+	-	-	-
WRKY46	At2g46400	WRKY	+	-	-	-
HSFA1E	At3g02990	HSF	+	-	-	-

Figure 9 AtSRO5 interacts with transcription factors. Transcription factors interacting with AtSRO5 in a pairwise interaction test against the REGIA TF collection. FL: Full length AtRCD1. PCT: AtRCD1 construct lacking the WWE domain. + interaction observed, - interaction not observed.

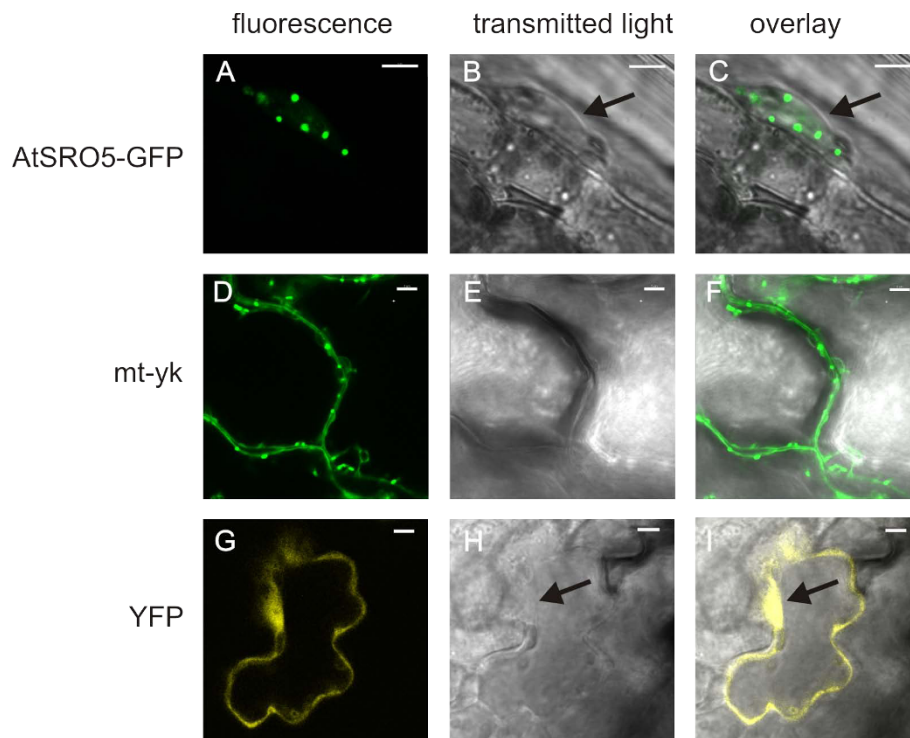


Figure 10 Subcellular localization of AtSRO5. The AtSRO5-GFP fusion protein localized to several dot-like structures in the nucleus in *A. thaliana* seedlings (panels A-C). As comparison, the mitochondrial localization marker line mt-yk (panels D-F) and the nuclear and cytoplasmic localization of YFP protein (panels G-I) are shown. Panels A, D and G display the fluorescent signal, panels B, E and H the light micrograph and panels C, F and I the two overlaid. Scale bar 5 μm ; arrows in B, C, H and I indicate the nucleus.

The conserved PARP domain in SRO-proteins: structural vs. functional conservation

Based on the presence of a PARP catalytic domain, it has been presumed that *A. thaliana* RCD1 and SRO proteins could have ADP-ribosyl transferase activity [1,2,6], which seems to be confirmed by the conserved fold structure (Figure 11). The alignment of AtPARPs and AtRCD1 with HsPARP1, for which the 3D structure has been solved, allowed for identification of conserved fold structures as landmarks in *A. thaliana* PARPs (Figure 11A). Generally, the fold structure is well conserved and all of the folds that constitute the active site are present (β sheets 1-6 and α helix 2, Figure 11A). Some additional plant specific folds not present in the HsPARP1 are predicted in AtPARP1 and 2, AtRCD1 and AtSROs (Figure 11A and 11B). These additional predicted features, if present, apparently do not disrupt the activity in AtPARP1, which was shown to exhibit PARP activity (Table 2; [44]). The conserved active site folds also mark the position of the catalytic triad, the three AAs histidine (H), tyrosine (Y) and glutamic acid (E), which is conserved in AtPARP1 and 2 but not AtRCD1 or AtSROs (Figure 11A and 11B; Table 2). The H333 to L and Y365 to H substitutions at the NAD binding sites within the HYE catalytic triad of RCD1

(Table 2) suggest that it has lost the ability to bind NAD.

To test the predictions of activity based on the fold structure of the PARP domain, we expressed the *A. thaliana* full length RCD1 protein and a truncated form containing the PARP and RST domains (PCT; AAs 241-589) as GST-tagged proteins in *Escherichia coli*. The recombinant proteins were partially purified by affinity chromatography with glutathione sepharose and used for testing NAD binding. *Pisum sativum* short-chain alcohol dehydrogenase-like protein A (SAD-A, [45]) was used as positive control.

NAD binding was investigated by covalent cross-linking of bound NAD by ultraviolet irradiation [46,47]. After UV irradiation of sample mixtures containing radioactive NAD and the proteins tested, the proteins were separated by SDS-PAGE and labeling with [α - ^{32}P -NAD] was monitored by autoradiography. To verify the specificity of NAD binding, competition experiments were performed with excess of unlabeled NAD.

The NAD binding of the positive control, SAD-A, was visible as two bands in an autoradiogram (Figure 12A). The major band at 30 kDa corresponds to monomeric form of the enzyme, and the minor band at 60 kDa to the dimer [45]. The presence of 1000-fold excess of

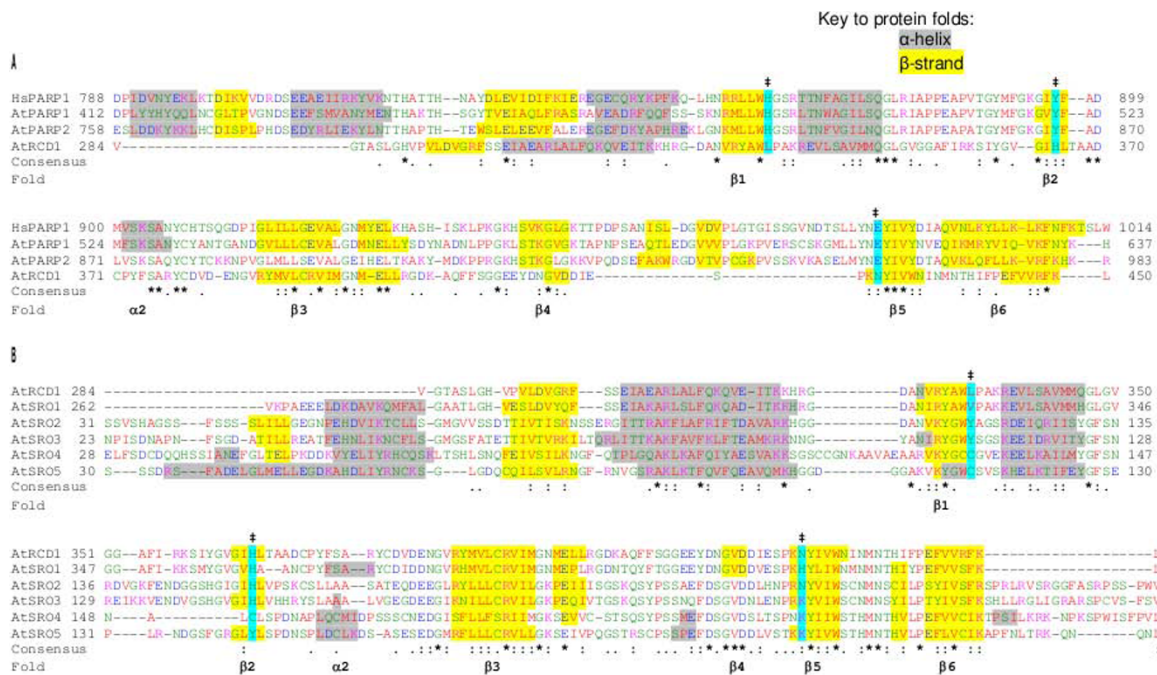
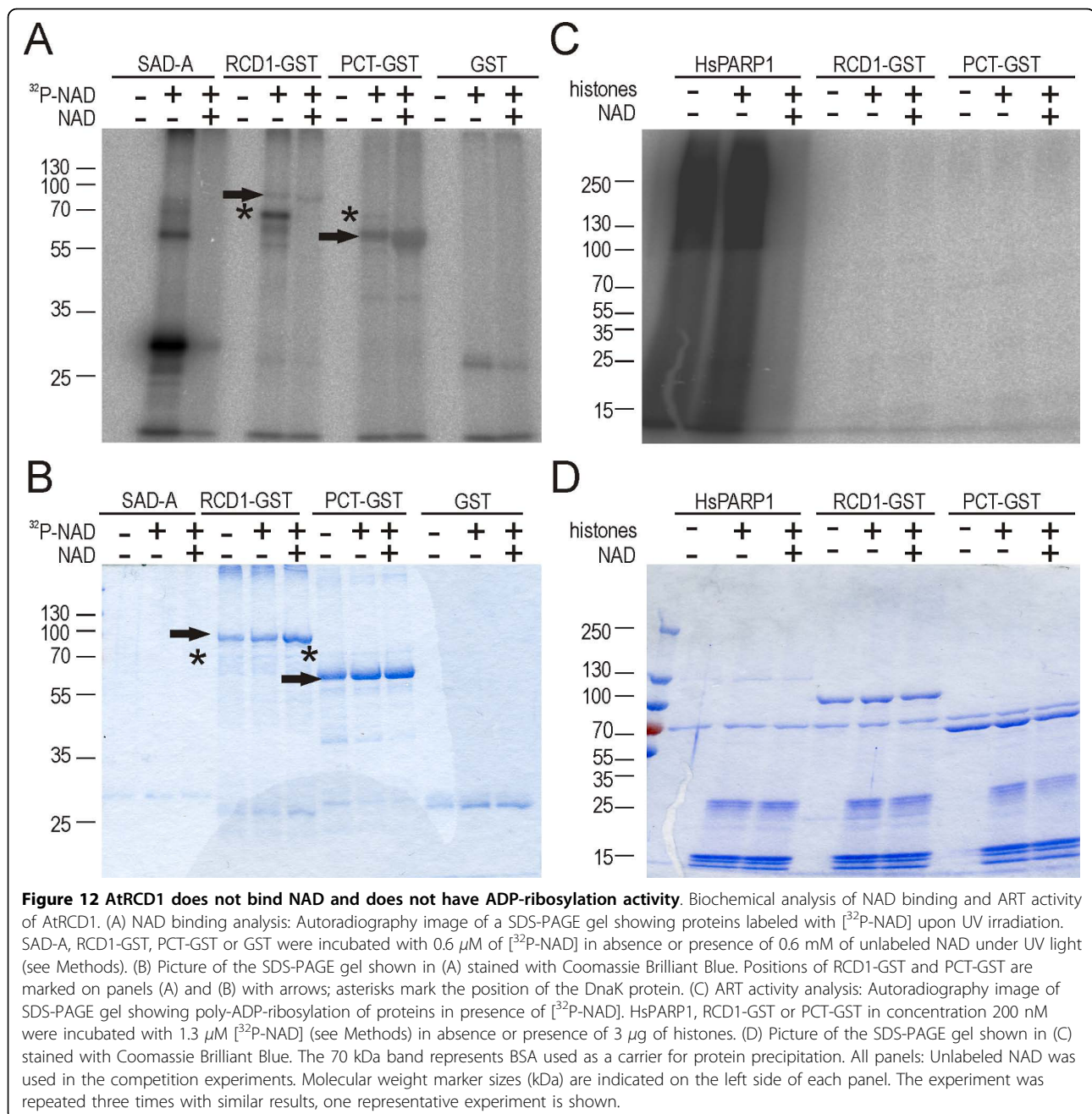


Figure 11 Conserved active site fold structure of the PARP domain. Fold-assisted AA alignments of the PARP catalytic core from (A) human PARP1 (HsPARP1), *A. thaliana* PARP-1 and -2 (AtPARP1, AtPARP2) and RCD1 (AtRCD1) and (B) *A. thaliana* RCD1 and SROs (AtSRO1-5). Consensus of conserved (*) and similar (: and .) AAs and conserved folds (α -helix or β -sheet) are indicated below the alignments. Additionally folds are shaded in the alignment with grey (α -helix) or yellow (β -sheet) backgrounds. Conserved ADP-ribosyl transferase catalytic triad, composed of three AAs at the C-terminus of β -sheet 1, middle of β -sheet 2 and N-terminal end of β -sheet 5, is indicated by turquoise background shading and an (#) above the alignment. Alignments were performed with T-Coffee at EMBL-EBI <http://www.ebi.ac.uk/Tools> and hand-adjusted according to fold predictions performed with Pspired in the Phyre search [53]. AAs were color-coded according to their biochemical properties as in <http://www.ebi.ac.uk/Tools/t-coffee/help.html#color>.

Table 2 Characteristics of the putative SRO active sites

Name	Identifier	WWE	Catalytic Core Motif	Loop length (between β 4 and β 5)	NAD binding	Predicted Activity	Confirmed Activity
AtPARP1	At4g02390.1	No	H486 Y520 E614	38	yes	PARP	PARP [44]
AtPARP2	At2g31320.1	No	H833 Y867 E960	36	yes	PARP	ND
AtPARP3	At5g22470.1	No	C653 V687 E782	36	yes	PARP	ND
AtRCD1	At1g32230.1	Yes	L333 H365 N428	5	no *	inactive	inactive *
AtSRO1	At2g35510.1	Yes	V329 H361 N422	5	ND	inactive	ND
AtSRO2	At1g23550.1	Yes	Y118 H153 N216	5	ND	inactive	ND
AtSRO3	At1g70440.1	Yes	Y110 H145 K208	5	ND	inactive	ND
AtSRO4	At3g47720.1	Yes	C129 C150 K214	6	ND	inctive	ND
AtSRO5	At5g62520.1	Yes	C113 Y143 K207	5	ND	inactive	ND
HsPARP1	P09874	No	H862 Y896 E988 [24]	37	yes	PARP	PARP
HsPARP7	Q7Z3E1	Yes	H532 Y564 I631 [26]	6	ND	mART	ND
HsPARP10	Q53GL7	No	N886 Y919 I987 [26]	6	yes	mART	mART
HsPARP11	Q9NR21	Yes	H197 Y229 I313 [26]	6	ND	mART	ND
HsPARP12	Q9H0J9	Yes	H564 Y596 I660 [26]	6	ND	mART	ND
HsPARP13	Q7Z2W4	Yes	Y787 Y819 V876 [26]	6	ND	inactive	inactive
HsPARP14	NP_060024	Yes	H1682 Y1714 L1782 [26]	6	yes	mART	mART

Relationship between conserved AA motif of the ADP-ribosyl transferase catalytic triad, the loop length between β -sheets 4 and 5, and predicted or know catalytic activity of selected *A. thaliana* and human PARPs and SRO family members. Presence or absence of the WWE domain (PSS0918) and catalytic activities are indicated. Results of this study are marked with an asterisk (*).



unlabeled NAD resulted in the disappearance of both bands, indicating that the NAD binding was specific (Figure 12A). In contrast, RCD1-GST, PCT-GST and GST did not bind NAD (Figure 12A). The weak bands visible on the autoradiogram at the molecular weights corresponding to RCD1-GST or PCT-GST (indicated by arrows in figures 12A and 12B) or GST alone, respectively, did not disappear in presence of unlabeled NAD, indicating unspecific labeling of the proteins. The 70 kDa band visible in the autoradiogram (Figures 12A and 12B, asterisk) represented a contaminant in the purified

RCD1-GST and PCT-GST samples. It was identified by mass spectrometry as DnaK molecular chaperone from *E. coli*. DnaK protein contains a nucleotide-binding domain explaining its ability to bind NAD.

These results demonstrated that AtRCD1 does not bind NAD and thus should not have ART activity. To verify this, we tested possible poly(ADP-ribosyl) transferase activity of RCD1-GST and PCT-GST directly in a standard ART activity assay using recombinant HsPARP1 as a positive control. HsPARP1 exhibited automodification (Figure 12C, a smear at molecular

mass above 116 kDa) but no auto-poly(ADP-ribosyl) transferase activity was detected for RCD1-GST or PCT-GST (Figure 12C and 12D). Possible substrate modification by RCD1-GST or PCT-GST was analyzed by supplementing the reaction mixture with histones, which are classical PARP targets. Neither RCD1-GST nor PCT-GST exhibited detectable PARP or mART activity (Figure 12C and 12D). Additionally, DREB2A, the most prominent RCD1 interaction partner, could be a possible substrate [1]. However, no PARP or mART activity of RCD1-GST or PCT-GST towards DREB2A-GST was detected (data not shown).

In light of these results, it is remarkable that the SROs structurally resemble PARPs/ARTs so closely. It may be possible that the PARP domain of AtRCD1 and the SROs still has an activity related to ADP-ribosylation. A novel mechanism has been described for HsPARP10, which lacks the catalytic glutamic acid (E), the third conserved AA of the catalytic triad (Table 2). HsPARP10 has still retained mART activity via a novel mechanism in which the active E is provided by the substrate protein [24]. HsPARP10 has a shorter linker sequence between folds $\beta 4$ and $\beta 5$ [24] which facilitates an open active site configuration necessary for the substrate glutamic acid entry into the active site. AtPARPs (AtPARP1, 2, and 3) retain a long $\beta 4$ - $\beta 5$ linker but all AtSROs have the shorter linker (Table 2) suggesting a more open active site fold. The bioinformatic analysis revealed the loss of both conserved NAD contacting H and Y in the *A. thaliana* SRO PARP domains making such a substrate-mediated mART activity unlikely (AtSRO5 is an exception to this, it has lost the H but retained the Y). This is supported by our biochemical analysis which demonstrated that AtRCD1 is not able to bind NAD, and, consequently, does not have mART or PARP activity. Other similar changes in the catalytic triad of the other AtSROs suggest they too may lack the capacity for NAD binding and ART activity (Table 2). Interestingly, this is also true for active sites in SROs from other plant species (Additional file 2), with the notable exception of *P. patens* SROs, which bear more conserved and potentially active catalytic triads.

Conclusions

The SROs are a protein family with a unique domain architecture which is conserved in all land plants. The SRO proteins can be subdivided into two groups representing two different structural types. Different plant groups have experienced expansion of different SRO groups during evolution. Interestingly, the basal plant groups, *P. patens*, a moss, and *S. moellendorffi*, a lycophyte, as well as monocots possess only group I SROs, while eudicots additionally contain group II SROs. Our analysis suggests that the evolutionary origin of the

SROs lies within subgroup Ib, which could be ancestral to all other SROs. Alternatively, monocots and more basal vascular plants might have experienced a secondary loss of group II SROs.

While the N-terminal WWE domain is only present in group I SROs of the structural type A, virtually all SROs analyzed contain a PARP-like domain and a C-terminal RST domain (Figure 7B). The conservation of the C-terminus of the SROs suggests functional constraints and a subsequent requirement for the conservation of a particular structure (Figure 7B). A possible function is the interaction with transcription factors (Figure 8), which has been demonstrated for several *A. thaliana* SROs, including AtSRO5. For a protein localized to mitochondria [18], its ability to interact with several transcription factors in Y2H analysis was unexpected. Our analysis of subcellular localization for AtSRO5 showed that the protein is localized to several dot-like structures in the nucleus (Figure 10) which supports the significance of the TF interactions. Nevertheless, it is possible that AtSRO5 localizes to the mitochondria under certain conditions linking TF interactions to retrograde signaling and mitochondrial metabolism [48].

The PARP-like domain is the most conserved feature of the SROs. However, based on bioinformatic and biochemical evidence (Figures 11 and 12), we suggest that the SROs do not possess PARP or mART activity. Nevertheless, the fold structure of the PARP-like domain is highly conserved (Figure 11). As a comparison, it is estimated, that 10% of the receptor-like protein kinases encoded in the *A. thaliana* genome are inactive but nevertheless expressed and translated and potentially function as co-receptors [49]. What other possible function or activity might those PARP/ART-like domains possess? The structural conservation of an enzymatically inactive domain could facilitate complex formation or stabilization and be an advantage for the organism. Regardless of which activity is eventually discovered in the SROs, they have important functions in plant stress responses and in development.

Methods

Sequence identification and phylogenetic analysis

Protein sequences for SROs of species used in this study were obtained from the respective projects databases (see Table 1 for reference) using HMMER and BLAST searches. Additionally, the genomes of aquatic, photosynthetic, and plant associated microorganisms were queried, including the green algae *Chlamydomonas reinhardtii* and *Ostreococcus tauri*; the yeasts, *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*; the plant pathogenic fungi, *Magnaporthe grisea* and *Botrytis cinerea*; as well as the photosynthetic cyanobacteria *Rhodospira rubra* and *Synechocystis* sp. The genomes

of these microorganisms did not contain genes related to SROs.

The assembly scaffold of the *A. lyrata* genome was a kind donation of Prof. Detlef Weigel. *A. lyrata RCD1-SRO* orthologs were identified by genomic blast with the *A. thaliana RCD1-SRO* genomic sequences. The *A. lyrata* sequences were subsequently spliced according to the *A. thaliana* gene models and converted to protein sequences. Some genomes were excluded due to gene models of SRO protein family members with significant dissimilarity to *A. thaliana* gene models and lack of cDNA support for these unique gene models.

The protein domains were identified using SMART [32,33] and ScanProsite [30,31]. cDNA sequences and ESTs were obtained *via* BLAST search through the NCBI webpage <http://www.ncbi.nlm.nih.gov/>. Sequences were, if possible, verified for being full length by comparison to existing ESTs from available collections. Some gene models were included for completeness; however, their dissimilarity to *A. thaliana* SROs and lack of cDNA support made them questionable: the gene models for OsSRO1d and OsSRO1e predicted long C-terminal extensions but ESTs suggested that OsSRO1d ended in the PARP domain and OsSRO1e contained a RST domain of normal length. PpSRO1a and PpSRO1b sequences were likely to be incomplete as the PARP domain extended until the end of the predicted protein. PpSRO1c contained a long C-terminus but ESTs suggested a shorter protein similar to other SROs. The C-terminal part of SmSRO1a from *S. moellendorffi* showed only moderate similarity to the C-terminus of other SROs. The annotation predicted a long C-terminal extension but EST support suggested only a short C-terminal domain. Due to the lack of other SRO sequences from organisms more closely related to *S. moellendorffi*, we were unable to determine if the C-terminus of SmSRO1a represented a unique development or a misannotation. Two additional putative SROs from *S. moellendorffi* were truncated and thus could not be assigned to any group. These sequences from rice, *P. patens*, and *S. moellendorffi* will require future verification.

Sequence alignments were performed using ClustalW2 [50] and colored using the Boxshade programme http://www.ch.embnet.org/software/BOX_form.html. Subsequent phylogenetic analysis was performed using Phylip and MEGA4 [51,52].

Active site alignments were performed with T-Coffee at EMBL-EBI <http://www.ebi.ac.uk/Tools> using only sequences of PARP domains as defined above. Fold predictions utilized Pspred in the Phyre search <http://www.sbg.bio.ic.ac.uk/phyre>[53]. Alignments were then hand adjusted with the guidance of conserved fold structures. Catalytic triad positions were determined as the

positions within conserved folds corresponding to the HYE triad from HsPARP1 and AtPARP1.

Yeast two-hybrid work

Yeast work was conducted as described in [1] using the GAL4-based ProQuest Y2H system (Invitrogen, Carlsbad, CA, USA). 10 mM 3-aminotriazole was used for eliminating autoactivation in all experiments. The primers used for cloning are described in additional file 3.

Gene expression analysis

The *sro5-2* allele was obtained from the GABI-Kat collection at the German Resource Center for Genome Research (line 325B05) [54]. Microarray hybridizations (4 biological repeats) and data analysis were performed as previously described [1]. qPCR experiments for gene expression analysis were done according to Wrzaczek *et al.* [55]. The primers used for qPCR are described in additional file 3.

Affymetrix raw data was downloaded from NASCARays <http://affymetrix.arabidopsis.info/narrays/experimentbrowse.pl> (accession number NASCARRAYS-143, paraquat; NASCARRAYS-353, ZAT12; NASCARRAYS-176, ABA time course experiment 1; NASCARRAYS-192, Ibuprofen), ArrayExpress <http://www.ebi.ac.uk/microarray-as/ae/> (accession numbers E-GEOD-12856, *Blumeria graminis* sp. *hordei*; E-GEOD-5684, *Botrytis cinerea*; E-GEOD-5743, 2,4-Dichlorophenoxyacetic acid (2,4-D); E-ATMX-13, Methyl Jasmonate; E-MEXP-550 polychromatic radiation with decreasing short-wave cut-off in the UV range (UV-B experiment); E-MEXP-739, Syringolin A; E-MEXP-1797, Rotenone), Gene Expression Omnibus <http://www.ncbi.nlm.nih.gov/geo/> (accession numbers GSE5615, Elicitors LPS, HrpZ, Flg22 and NPP1; GSE5685, Virulent and avirulent *Pseudomonas syringae*; GSE9955, BTH experiment 1; GDS417 *E. cichoracearum*; GSE5530, H₂O₂; GSE5621, Cold time course experiment; GSE5622, Osmotic stress time course experiment; GSE5623, Salt time course experiment; GSE5624, Drought time course experiment; GSE5722, O₃; GSE12887, Norflurazon; GSE10732, OPDA and Phytoprostane; GSE7112, ABA experiment 2) and The Integrated Microarray Database System <http://ausubellab.mgh.harvard.edu/imds> (Experiment name: BTH time course, BTH experiment 2).

The raw Affymetrix data was preprocessed with RMA using probe set annotations (custom.cdf files) from <http://brainarray.mbni.med.umich.edu/>, version 11.0.1. Biological repeats of each experiment were combined by computing a mean of the measured gene expression. Gene expression was summarized by computing a log₂ ratio of the treatment and control expressions (differential expression, DE). A visualization of the DE values is shown in figure 2. Variation of differential expression in

an experiment e , $\hat{\sigma}_e^2$, was estimated by summing the variances of (logarithm of) treatment and control gene expressions.

Parametric bootstrapping was implemented by generating 1000 samples for each experiment and each gene from a Gaussian distribution with the estimated DE as the mean and $\hat{\sigma}_e^2$ as the variance.

Bootstrap samples were discretized to down-regulated ($\log_2 \text{DE} < -1$), no regulation ($-1 \geq \log_2 \text{DE} \geq 1$), and up-regulated ($\log_2 \text{DE} > 1$) genes. Bayesian agglomerative hierarchical clustering algorithm was then applied to the discretized bootstrap data [56]. The Bayesian hierarchical clustering algorithm computes the best number of clusters by Bayesian hypothesis testing. For each pair of genes (and experiments, depending on the clustering direction), the number of times they were assigned to the same cluster was computed. These gene (or experiment) similarities were then used as distances for computing the hierarchical clustering (Ward method) shown in figure 2.

Protein localization

The localization of AtSRO5 was predicted using Predotar v. 1.03 <http://urgi.versailles.inra.fr/predotar/predotar.html>, TargetP 1.1 [57], WoLF PSORT [58] and MitoProTII - v1.101 [59]. None of the programs predicted mitochondrial localization. For *in planta* study of the localization, AtSRO5 was cloned into the pB7FWG2.0 [60] binary vector containing eGFP as C-terminal fusion to the protein using the primers described in additional file 3. YFP in pGREENII binary vector was used for nuclear and cytoplasmic localization control [1]. Three-day old *A. thaliana* seedlings were used for transient expression as described in [61]. The fluorescent proteins were visualized using confocal laser scanning microscopy after 36 hours of co-cultivation. The mitochondrial localization control line mt-yk (N16264) was obtained from Nottingham Arabidopsis Stock Centre and imaged at the same age as the transiently transformed plants.

Protein expression and purification

Full-length AtRCD1 and its truncated version, PCT, consisting of PARP and RST domains (AAs 241-589) were cloned into pGEX4T-1 for N-terminal GST fusion using the primers listed in additional file 3. After sequencing, the constructs were transformed into the *E. coli* strain BL21 (DE3) CodonPlus RIL for protein production.

LB medium containing ampicillin ($100 \mu\text{g ml}^{-1}$) and chloramphenicol ($50 \mu\text{g ml}^{-1}$) was inoculated with 1/50 volumes of overnight bacterial culture and grown at 37°C until OD_{600} reached 0.6-0.8. Expression of PCT-GST and DREB2A-GST was induced by adding isopropyl- β -D-galactoside (IPTG) to a final concentration of 0.5 mM, and the culture was transferred to 28°C . After 4 hours,

the cells were harvested by centrifugation at 5000 g and stored at -20°C .

For RCD1-GST expression, benzyl alcohol was added to the cell culture with OD_{600} 0.5-0.6 to a final concentration of 10 mM and the cells were grown for additional 30 min at 22°C [62]. Protein expression was induced by 0.1 mM of IPTG. After 16 hours at 22°C , the cells were harvested by centrifugation at 5000 g at room temperature, resuspended in original volume of fresh LB medium without IPTG and grown for additional 2-3 hours at 22°C . Finally, the cells were harvested by centrifugation at 5000 g and stored at -20°C . The cell pellets were resuspended in a lysis buffer (1/20 of initial culture volume) consisting of 50 mM Tris-HCl, pH7.5, 150 mM NaCl, 5 mM DTT, protease inhibitors cocktail (Complete, Roche Diagnostics GmbH, Mannheim, Germany). The cells were lysed by addition of lysozyme (Roche) to a concentration of 0.2 mg ml^{-1} and incubation for 30 min at 4°C with gentle shaking. Released DNA was then digested by DNase I (Roche) at final concentration of 0.02 mg ml^{-1} in presence of 5 mM MgCl_2 and incubation for another 30 min at 4°C . The cell lysates was clarified by centrifugation at 20000 g for 15 min at 4°C . The GST tagged proteins were purified by affinity chromatography using 1-ml GSTrap columns (GE Healthcare, Chalfont St Giles, UK) according to manufacturer's instructions. SAD-A-His protein was expressed and purified as described [45]. Protein concentration was determined by Bradford method using Protein Assay reagent (Bio-Rad Laboratories Inc., Hercules, CA, USA).

UV photoaffinity labeling

Samples of total volume $30 \mu\text{l}$ containing 30 pmol of protein in 50 mM Tris-HCl, pH7.5, 100 mM NaCl, 5 mM MgCl_2 , 1 mM DTT and $0.6 \mu\text{M}$ of [α - ^{32}P NAD] ($0.8 \text{ mCi mmol}^{-1}$) (NEN, PerkinElmer, Inc. Boston, MA, USA) were incubated in a 96-well plate on an ice bath. Unlabeled NAD in concentration 0.6 mM was added to the mixtures in competition experiment. The UV irradiation was performed for 15 min as described in [46]. The proteins were then precipitated by addition of equal volume of ice-cold 22% trichloroacetic acid and incubation on ice for at least 30 min. After centrifugation for 10 min at 16000 g the protein pellet was washed once with cold acetone, air-dried, and resuspended in $10 \mu\text{l}$ of SDS-PAGE sample buffer [63].

In vitro PARP activity assay

Samples corresponding to 200 nM of proteins were incubated for 20 min at 22°C in assay buffer ($50 \mu\text{l}$) consisting of 50 mM Tris-HCl, pH 7.5, 100 mM NaCl, 5 mM MgCl_2 , 1 mM DTT, $10 \mu\text{g ml}^{-1}$ activated DNA (calf thymus nicked DNA, Sigma Aldrich, St. Louis, MO, USA) and $1.3 \mu\text{M}$ [α - ^{32}P NAD] ($0.8 \text{ mCi mmol}^{-1}$).

Recombinant HsPARP1 (Sigma) was used as a positive control. 3 µg of total histones (calf thymus histones, Roche) or DREB2A-GST were added as acceptor proteins. 1 mM unlabeled NAD was added in competition experiment. The reaction was stopped by addition of ice-cold trichloroacetic acid as described above. 5 µg BSA were added to the reaction mixture just before protein precipitation as a carrier.

SDS-PAGE and autoradiography

The proteins were separated on SDS-PAGE (12% or 4-15%) according to the protocol of [63]. After protein visualization with Coomassie Brilliant Blue, the gels were dried and subjected to autoradiography. The autoradiography images were analysed with Fuji BAS-1500 phosphorimager.

In-gel digestion and mass spectrometry

In-gel digestion and sample preparation for mass spectrometry was performed as described [64]. MALDI TOF (matrix-assisted laser desorption-ionisation time-of-flight) analysis was performed on reflector mode on a Voyager DE-PRO mass spectrometer (Applied Biosystems, Foster City, CA, USA).

Additional file 1: Neighbour-joining phylogenetic tree of

Arabidopsis thaliana and Brassica SROs. The gene duplication leading to the *AtRCD1/AtSRO1*, *AtSRO2/AtSRO3* and *AtSRO4/AtSRO5* gene pairs in *Arabidopsis* is also present in *Brassica*. Individual protein-coding ESTs from *Brassica* can be assigned to *AtRCD1* or *AtSRO1*, *AtSRO2* or *AtSRO3* or *AtSRO5*. No EST was identified for *AtSRO4*. This indicates that the gene duplication event leading to the gene pairs occurred early during the evolution of the Brassicaceae family before the split between the *Arabidopsis* and *Brassica* genera. Representative ESTs from *Brassica napus*, *Brassica rapa* and *Brassica oleracea* coding for SRO proteins were extracted via NCBI blast and the PARP domain was identified using Prosite. The PARP domains of the *Brassica* SROs were aligned with the PARP domains of the members of the *A. thaliana* and *A. lyrata* SRO protein families an unrooted Neighbour-joining tree was constructed using MEGA4.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-11-170-S1.PNG>]

Additional file 2: Active site alignments for all plant SROs.

Alignments of the region around the active site catalytic triad of all SROs analyzed here. Alignments were hand-adjusted according to the positions of conserved folds in *AtRCD1*, *AtPARP1* and *AtPARP2* as in figure 10, only the regions immediately surrounding the predicted catalytic amino acids (highlighted in red) are presented.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-11-170-S2.PDF>]

Additional file 3: Primers used for analysis. All primer sequences used for qPCR analysis in the manuscript are listed in this file.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-11-170-S3.PDF>]

Acknowledgements

Tuomas Puukko and Marjukka Uuskallio are acknowledged for superb technical assistance. We thank Dr. Mikael Brosché for his insight and

comments on the manuscript and Prof. Åke Strid for providing the SAD-A-His expression vector. The funding sources acknowledged for this work are: Finnish Academy Centre of Excellence program (2006-2011) and research grant to JK (# 121576), Postdoctoral grant to KO (# 115034) and Helsinki University Postdoctoral Grant to MW. PJ and TB were supported by the Viikki Graduate School in Biosciences. The REGIA TF ORF Library was generated as part of the EU-funded project REGIA (<http://cordis.europa.eu/QLG-CT11999-00876>) and we acknowledge Dr. Franziska Turck for providing the collection in the Y2H strains. Prof. Gerco Angenent and Dr. Richard Immink are gratefully acknowledged for their collaboration in the Y2H work. We thank Dr. Saijalliisa Kangasjärvi (University of Turku, Finland) for excess light treated plant material and Dr. Sophia Mersmann (Max-Planck Institute for Plant Breeding Research) and Dr. Silke Robatzek (The Sainsbury Laboratories, UK) for plant material. Dr. Jorma Vahala is acknowledged for help with the *P. trichocarpa* genome and Mr. Korbinian Schneeberger and Prof. Detlef Weigel for the *A. lyrata* genome. We thank Dr. Natalia Battchikova (University of Turku, Finland) for mass spectrometric analysis of *AtRCD1*-copurifying protein and acknowledge the national Finnish DNA Microarray Centre (Turku Centre for Biotechnology) for the manufacturing of microarray slides.

Authors' contributions

PJ, KO, MW, JPV and JK designed research. PJ, KO, MW, JPV, TB, RAR and JS carried out research. PJ, KO, MW, JPV, TB, JS and JK analyzed the data. PJ, KO, MW and JK wrote the paper. All authors have read and approved the final manuscript.

Received: 6 November 2009 Accepted: 12 March 2010

Published: 12 March 2010

References

1. Jaspers P, Blomster T, Brosché M, Salojärvi J, Ahlfors R, Vainonen JP, Reddy RA, Immink R, Angenent G, Turck F, Overmyer K, Kangasjärvi J: **Unequally redundant RCD1 and SRO1 mediate stress and developmental responses and interact with transcription factors.** *Plant J* 2009, **60**(2):268-79.
2. Teotia S, Lamb RS: **The paralogous genes *RADICAL-INDUCED CELL DEATH 1* and *SIMILAR TO RCD ONE 1* have partially redundant functions during *Arabidopsis thaliana* development.** *Plant Physiol* 2009, **151**(1):180-98.
3. Belles-Boix E, Babychuk E, Van Montagu M, Inzé D, Kushnir S: **CEO1, a new protein from *Arabidopsis thaliana*, protects yeast against oxidative damage.** *FEBS Lett* 2000, **482**(1-2):19-24.
4. Overmyer K, Tuominen H, Kettunen R, Betz C, Langebartels C, Sandermann H, Kangasjärvi J: **Ozone-sensitive *Arabidopsis rcd1* mutant reveals opposite roles for ethylene and jasmonate signaling pathways in regulating superoxide-dependent cell death.** *Plant Cell* 2000, **12**(10):1849-62.
5. Ahlfors R, Lång S, Overmyer K, Jaspers P, Brosché M, Tauriainen A, Kollist H, Tuominen H, Belles-Boix E, Piippo M, Inzé D, Palva ET, Kangasjärvi J: ***Arabidopsis* RADICAL-INDUCED CELL DEATH1 belongs to the WWE protein-protein interaction domain protein family and modulates abscisic acid, ethylene, and methyl jasmonate responses.** *Plant Cell* 2004, **16**(7):1925-37.
6. Ahlfors R, Macioszek V, Rudd J, Brosché M, Schlichting R, Scheel D, Kangasjärvi J: **Stress hormone-independent activation and nuclear translocation of mitogen-activated protein kinases in *Arabidopsis thaliana* during ozone exposure.** *Plant J* 2004, **40**(4):512-22.
7. Fujibe T, Saji H, Arakawa K, Yabe N, Takeuchi Y, Yamamoto KT: **A methyl viologen-resistant mutant of *Arabidopsis*, which is allelic to ozone-sensitive *rcd1*, is tolerant to supplemental ultraviolet-B irradiation.** *Plant Physiol* 2004, **134**(1):275-85.
8. Katiyar-Agarwal S, Zhu J, Kim K, Agarwal M, Fu X, Huang A, Zhu JK: **The plasma membrane Na⁺/H⁺ antiporter SOS1 interacts with RCD1 and functions in oxidative stress tolerance in *Arabidopsis*.** *Proc Natl Acad Sci USA* 2006, **103**(49):18816-21.
9. Ahlfors R, Brosché M, Kollist H, Kangasjärvi J: **Nitric oxide modulates ozone-induced cell death, hormone biosynthesis and gene expression in *Arabidopsis*.** *Plant J* 2008, **58**(1):1-12.
10. Briggs GC, Osmont KS, Shindo C, Sibout R, Hardtke CS: **Unequal genetic redundancies in *Arabidopsis*—a neglected phenomenon?** *Trends Plant Sci* 2006, **11**(10):492-8.

11. Caruso A, Chedford F, Carpin S, Depierreux C, Delmotte FM, Kahlem G, Morabito D: **Physiological characterization and identification of genes differentially expressed in response to drought induced by PEG 6000 in *Populus canadensis* leaves.** *J Plant Physiol* 2008, **165**(9):932-41.
12. da Cruz Gallo de Carvalho MC, Caldas DGG, Carneiro RT, Moon DH, Salvatierra GR, Franceschini LM, de Andrade A, Celedon PAF, Oda S, Labate CA: **SAGE transcript profiling of the juvenile cambial region of *Eucalyptus grandis*.** *Tree Physiol* 2008, **28**(6):905-19.
13. Quaggiotti S, Barcaccia G, Schiavon M, Nicolé S, Galla G, Rossignolo V, Soattin M, Malagoli M: **Phytoremediation of chromium using *Salix* species: cloning ESTs and candidate genes involved in the Cr response.** *Gene* 2007, **402**(1-2):68-80.
14. Walter S, Brennan JM, Arunachalam C, Ansari KI, Hu X, Khan MR, Trognitz F, Trognitz B, Leonard G, Egan D, Doohan FM: **Components of the gene network associated with genotype-dependent response of wheat to the *Fusarium* mycotoxin deoxynivalenol.** *Funct Integr Genomics* 2008, **8**(4):421-7.
15. Sanchez-Ballesta MT, Lluch Y, Gosalbes MJ, Zacarias L, Granell A, Lafuente MT: **A survey of genes differentially expressed during long-term heat-induced chilling tolerance in citrus fruit.** *Planta* 2003, **218**(1):65-70.
16. Bechtold U, Richard O, Zamboni A, Gapper C, Geisler M, Pogson B, Karpinski S, Mullineaux PM: **Impact of chloroplastic- and extracellular-sourced ROS on high light-responsive gene expression in *Arabidopsis*.** *J Exp Bot* 2008, **59**(2):121-33.
17. Taylor NL, Heazlewood JL, Day DA, Millar AH: **Differential impact of environmental stresses on the pea mitochondrial proteome.** *Mol Cell Proteomics* 2005, **4**(8):1122-33.
18. Borsani O, Zhu J, Verslues PE, Sunkar R, Zhu JK: **Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in *Arabidopsis*.** *Cell* 2005, **123**(7):1279-91.
19. Babajani G, Effendya J, Plant AL: ***SI-SRO1* increases salt tolerance and is a member of the radical-induced cell death 1-similar to *RCD1* gene family of tomato.** *Plant Science* 2009, **176**(2):214-222.
20. Aravind L: **The WWE domain: a common interaction module in protein ubiquitination and ADP-ribosylation.** *Trends Biochem Sci* 2001, **26**(5):273-5.
21. Zweifel ME, Leahy DJ, Barrick D: **Structure and Notch receptor binding of the tandem WWE domain of Deltex.** *Structure* 2005, **13**(11):1599-611.
22. Schreiber V, Dantzer F, Ame JC, de Murcia G: **Poly(ADP-ribose): novel functions for an old molecule.** *Nat Rev Mol Cell Biol* 2006, **7**(7):517-28.
23. Hakmé A, Huber A, Dollé P, Schreiber V: **The macroPARP genes *Parp-9* and *Parp-14* are developmentally and differentially regulated in mouse tissues.** *Dev Dyn* 2008, **237**(1):209-15.
24. Kleine H, Poreba E, Lesniewicz K, Hassa PO, Hottiger MO, Litchfield DW, Shilton BH, Lüscher B: **Substrate-assisted catalysis by PARP10 limits its activity to mono-ADP-ribosylation.** *Mol Cell* 2008, **32**(1):57-69.
25. Hassa PO, Haenni SS, Elser M, Hottiger MO: **Nuclear ADP-ribosylation reactions in mammalian cells: where are we today and where are we going?** *Microbiol Mol Biol Rev* 2006, **70**(3):789-829.
26. Otto H, Reche PA, Bazan F, Dittmar K, Haag F, Koch-Nolte F: **In silico characterization of the family of PARP-like poly(ADP-ribosyl)transferases (pARTs).** *BMC Genomics* 2005, **6**:139.
27. Corda D, Di Girolamo M: **Functional aspects of protein mono-ADP-ribosylation.** *EMBO J* 2003, **22**(9):1953-8.
28. Ma S, Bohnert HJ: **Integration of *Arabidopsis thaliana* stress-related transcript profiles, promoter structures, and cell-specific expression.** *Genome Biol* 2007, **8**(4):R49.
29. Li L, Foster CM, Gan Q, Nettleton D, James MG, Myers AM, Wurtele ES: **Identification of the novel protein QQS as a component of the starch metabolic network in *Arabidopsis* leaves.** *Plant J* 2009, **58**(3):485-98.
30. de Castro E, Sigrist CJA, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, Bairoch A, Hulo N: **ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins.** *Nucleic Acids Res* 2006, **34**(Web Server):W362-5.
31. Hulo N, Bairoch A, Bulliard V, Cerutti L, De Castro E, Langendijk-Genevaux PS, Pagni M, Sigrist CJA: **The PROSITE database.** *Nucleic Acids Res* 2006, **34**(Database):D227-30.
32. Letunic I, Goodstadt L, Dickens NJ, Doerks T, Schultz J, Mott R, Ciccarelli F, Copley RR, Ponting CP, Bork P: **Recent improvements to the SMART domain-based sequence annotation resource.** *Nucleic Acids Res* 2002, **30**(1):242-4.
33. Letunic I, Doerks T, Bork P: **SMART 6: recent updates and new developments.** *Nucleic Acids Res* 2009, **37**(Database):D229-32.
34. Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH: **Syntenic and collinearity in plant genomes.** *Science* 2008, **320**(5875):486-8.
35. Romanel EAC, Schrago CG, Couñago RM, Russo CAM, Alves-Ferreira M: **Evolution of the B3 DNA binding superfamily: new insights into REM family gene diversification.** *PLoS One* 2009, **4**(6):e5791.
36. Sonnhammer ELL, Koonin EV: **Orthology, paralogy and proposed classification for paralog subtypes.** *Trends Genet* 2002, **18**(12):619-20.
37. Sanderson MJ, Thorne JL, Wikström N, Bremer K: **Molecular evidence on plant divergence times.** *American Journal of Botany* 2004, **91**:1656-1665.
38. Wright KJ, Marr MT, Tjian R: **TAF4 nucleates a core subcomplex of TFIID and mediates activated transcription from a TATA-less promoter.** *Proc Natl Acad Sci USA* 2006, **103**(33):12347-52.
39. Lago C, Clerici E, Mizzi L, Colombo L, Kater MM: **TBP-associated factors in *Arabidopsis*.** *Gene* 2004, **342**(2):231-41.
40. Paz-Ares J, The Regia Consortium: **REGIA, an EU project on functional genomics of transcription factors from *Arabidopsis thaliana*.** *Comp Funct Genomics* 2002, **3**(2):102-8.
41. Cannino G, Di Liegro CM, Rinaldi AM: **Nuclear-mitochondrial interaction.** *Mitochondrion* 2007, **7**(6):359-66.
42. Wei Y, Liu S, Lausen J, Woodrell C, Cho S, Biris N, Kobayashi N, Wei Y, Yokoyama S, Werner MH: **A TAF4-homology domain from the corepressor ETO is a docking platform for positive and negative regulators of transcription.** *Nat Struct Mol Biol* 2007, **14**(7):653-61.
43. Marr MT: **TAF4 takes flight.** *Proc Natl Acad Sci USA* 2009, **106**(5):1295-6.
44. Doucet-Chabeaud G, Godon C, Brutusco C, de Murcia G, Kazmaier M: **Ionising radiation induces the expression of *PARP-1* and *PARP-2* genes in *Arabidopsis*.** *Mol Genet Genomics* 2001, **265**(6):954-63.
45. Scherbak N, Brosché M, Ala-Häivälä A, Strid H, Ohrfelt A, Nilsson F, Strid A: **Expression of *Pisum sativum* SAD polypeptides in production hosts and in planta: tetrameric organization of the protein.** *Protein Expr Purif* 2009, **63**(1):18-25.
46. Carroll SF, Lory S, Collier RJ: **Ligand interactions of diphtheria toxin. III. Direct photochemical cross-linking of ATP and NAD to toxin.** *J Biol Chem* 1980, **255**(24):12020-4.
47. Ma Y, Ludden PW: **Role of the dinitrogenase reductase arginine 101 residue in dinitrogenase reductase ADP-ribosyltransferase binding, NAD binding, and cleavage.** *J Bacteriol* 2001, **183**(1):250-6.
48. McBride HM, Neuspiel M, Wasiaik S: **Mitochondria: more than just a powerhouse.** *Curr Biol* 2006, **16**(14):R551-60.
49. Castells E, Casacuberta JM: **Signalling through kinase-defective domains: the prevalence of atypical receptor-like kinases in plants.** *J Exp Bot* 2007, **58**(13):3503-11.
50. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG: **Clustal W and Clustal X version 2.0.** *Bioinformatics* 2007, **23**(21):2947-8.
51. Felsenstein J: **PHYLP - Phylogeny Inference Package (Version 3.2).** *Cladistics* 1989, **5**:164-166.
52. Tamura K, Dudley J, Nei M, Kumar S: **MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0.** *Mol Biol Evol* 2007, **24**(8):1596-9.
53. Kelley LA, Sternberg MJE: **Protein structure prediction on the Web: a case study using the Phyre server.** *Nat Protoc* 2009, **4**(3):363-71.
54. Rosso MG, Li Y, Strizhov N, Reiss B, Dekker K, Weisshaar B: **An *Arabidopsis thaliana* T-DNA mutagenized population (GABI-Kat) for flanking sequence tag-based reverse genetics.** *Plant Mol Biol* 2003, **53**(1-2):247-59.
55. Wrzaczek M, Brosché M, Kollist H, Kangasjärvi J: ***Arabidopsis* GRI is involved in the regulation of cell death induced by extracellular ROS.** *Proc Natl Acad Sci USA* 2009, **106**(13):5412-7.
56. Savage RS, Heller K, Xu Y, Ghahramani Z, Truman WM, Grant M, Denby KJ, Wild DL: **R/BHC: fast Bayesian hierarchical clustering for microarray data.** *BMC Bioinformatics* 2009, **10**:242.
57. Emanuelsson O, Nielsen H, Brunak S, von Heijne G: **Predicting subcellular localization of proteins based on their N-terminal amino acid sequence.** *J Mol Biol* 2000, **300**:1005-16.
58. Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Collier CA, Nakai K: **WoLF PSORT: protein localization predictor.** *Nucleic Acids Res* 2007, **35**:W585-W587.

59. Claros MG, Vincens P: **Computational method to predict mitochondrially imported proteins and their targeting sequences.** *Eur J Biochem* 1996, **241**(3):779-86.
60. Karimi M, Inzé D, Depicker A: **Gateway vectors for *Agrobacterium*-mediated plant transformation.** *Trends Plant Sci* 2002, **5**(5):193-5.
61. Li JF, Park E, von Arnim AG, Nebenführ A: **The FAST technique: a simplified *Agrobacterium*-based transformation method for transient gene expression analysis in seedlings of *Arabidopsis* and other plant species.** *Plant Methods* 2009, **5**:6.
62. de Marco A, Vigh L, Diamant S, Goloubinoff P: **Native folding of aggregation-prone recombinant proteins in *Escherichia coli* by osmolytes, plasmid- or benzyl alcohol-overexpressed molecular chaperones.** *Cell Stress Chaperones* 2005, **10**(4):329-39.
63. Laemmli UK: **Cleavage of structural proteins during the assembly of the head of bacteriophage T4.** *Nature* 1970, **227**(5259):680-5.
64. Shevchenko A, Wilm M, Vorm O, Mann M: **Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels.** *Anal Chem* 1996, **68**(5):850-8.
65. The Arabidopsis Genome Initiative: **Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*.** *Nature* 2000, **408**(6814):796-815.
66. The International Brachypodium Initiative: **Genome sequencing and analysis of the model grass *Brachypodium distachyon*.** *Nature* 2010, **463**(7282):763-8.
67. Yu J, Hu S, Wang J, Wong GKS, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, Cao M, Liu J, Sun J, Tang J, Chen Y, Huang X, Lin W, Ye C, Tong W, Cong L, Geng J, Han Y, Li L, Li W, Hu G, Huang X, Li W, Li J, Liu Z, Li L, Liu J, Qi Q, Liu J, Li L, Li T, Wang X, Lu H, Wu T, Zhu M, Ni P, Han H, Dong W, Ren X, Feng X, Cui P, Li X, Wang H, Xu X, Zhai W, Xu Z, Zhang J, He S, Zhang J, Xu J, Zhang K, Zheng X, Dong J, Zeng W, Tao L, Ye J, Tan J, Ren X, Chen X, He J, Liu D, Tian W, Tian C, Xia H, Bao Q, Li G, Gao H, Cao T, Wang J, Zhao W, Li P, Chen W, Wang X, Zhang Y, Hu J, Wang J, Liu S, Yang J, Zhang G, Xiong Y, Li Z, Mao L, Zhou C, Zhu Z, Chen R, Hao B, Zheng W, Chen S, Guo W, Li G, Liu S, Tao M, Wang J, Zhu L, Yuan L, Yang H: **A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*).** *Science* 2002, **296**(5565):79-92.
68. Rensing SA, Lang D, Zimmer AD, Terry A, Salamov A, Shapiro H, Nishiyama T, Perroud PF, Lindquist EA, Kamisugi Y, Tanahashi T, Sakakibara K, Fujita T, Oishi K, Shin-I T, Kuroki Y, Toyoda A, Suzuki Y, Hashimoto SI, Yamaguchi K, Sugano S, Kohara Y, Fujiyama A, Anterola A, Aoki S, Ashton N, Barbazuk WB, Barker E, Bennetzen JL, Blankenship R, Cho SH, Dutcher SK, Estelle M, Fawcett JA, Gundlach H, Hanada K, Heyl A, Hicks KA, Hughes J, Lohr M, Mayer K, Melkozernov A, Murata T, Nelson DR, Pils B, Prigge M, Reiss B, Renner T, Rombauts S, Rushton PJ, Sanderfoot A, Schween G, Shiu SH, Stueber K, Theodoulou FL, Tu H, Peer Van de Y, Verrier PJ, Waters E, Wood A, Yang L, Cove D, Cuming AC, Hasebe M, Lucas S, Mishler BD, Reski R, Grigoriev IV, Quatrano RS, Boore JL: **The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants.** *Science* 2008, **319**(5859):64-9.
69. Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen GL, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J, Degroove S, Déjardin A, Depamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehlting J, Ellis B, Gendler K, Goodstein D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R, Huang W, Islam-Faridi N, Jones S, Jones-Rhoades M, Jorgensen R, Joshi C, Kangasjärvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A, Kalluri U, Larimer F, Leebens-Mack J, Lepié JC, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson DR, Nelson C, Nieminen K, Nilsson O, Pereda V, Peter G, Philippe R, Pilate G, Poliakov A, Razumovskaya J, Richardson P, Rinaldi C, Ritland K, Rouzé P, Ryaboy D, Schmutz J, Schrader J, Segerman B, Shin H, Siddiqui A, Sterky F, Terry A, Tsai CJ, Uberbacher E, Unneberg P, Vahala J, Wall K, Wessler S, Yang G, Yin T, Douglas C, Marra M, Sandberg G, Peer Van de Y, Rokhsar D: **The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray).** *Science* 2006, **313**(5793):1596-604.
70. Velasco R, Zharkikh A, Troggio M, Cartwright DA, Cestaro A, Pruss D, Pindo M, Fitzgerald LM, Vezzulli S, Reid J, Malacarne G, Iliev D, Coppola G, Wardell B, Micheletti D, Macalma T, Facci M, Mitchell JT, Perazzolli M, Eldredge G, Gatto P, Oyzerski R, Moretto M, Gutin N, Stefanini M, Chen Y, Segala C, Davenport C, Demattè L, Mraz A, Battilana J, Stormo K, Costa F, Tao Q, Si-Ammour A, Harkins T, Lackey A, Perbost C, Taillon B, Stella A, Solovyev V, Fawcett JA, Sterck L, Vandepoele K, Grandio SM, Toppo S, Moser C, Lanchbury J, Bogden R, Skolnick M, Sgaramea V, Bhatnagar SK, Fontana P, Gutin A, Peer Van de Y, Salamini F, Viola R: **A high quality draft consensus sequence of the genome of a heterozygous grapevine variety.** *PLoS One* 2007, **2**(12):e1326.

doi:10.1186/1471-2164-11-170

Cite this article as: Jaspers et al.: The RST and PARP-like domain containing SRO protein family: analysis of protein structure, function and conservation in land plants. *BMC Genomics* 2010 **11**:170.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

