Running head: SUMO isoforms and SUMO substrates of Arabidopsis

Corresponding author: Andreas Bachmair, Max F. Perutz Laboratories, Univ. of Vienna, A-1030 Vienna Austria; e-mail andreas.bachmair@univie.ac.at; phone +43 1 4277 74811; fax +43 1 4277 9748.

Research Category: Biochemical Processes and Macromolecular Structures
Substrates Related to Chromatin and to RNA-Dependent Processes Are Modified by Arabidopsis SUMO Isoforms that Differ in a Conserved Residue with Influence on De-Sumoylation

Ruchika Budhiraja, Rebecca Hermkes, Stefan Müller, Jürgen Schmidt, Thomas Colby, Kishore Panigrahi, George Coupland and Andreas Bachmair*

Dept. of Plant Developmental Biology (R. B., R. H., K. P., G. C., A. B.) and Mass Spectrometry Group (J. S., T. C.), Max Planck Institute for Plant Breeding Research, Carl-von-Linné-Weg 10, D-50829 Cologne, Germany; Inst. of Biochemistry, Univ. of Cologne, Joseph-Stelzmann-Str. 52, D-50931 Cologne, Germany (S. M.); Dept. of Biochemistry, Max F. Perutz Laboratories, Univ. of Vienna, Dr. Bohr-Gasse 9, A-1030 Vienna Austria (A. B.)
This work was supported by the Max Planck Society, by the German Research Foundation (SFB 635 to G. C., and grants BA1158/3-1, SPP1365 to A. B.), by the Austrian Science Foundation FWF (grant P 21215-B12 to A. B.), and by pre-doctoral fellowships from the International Max Planck Research School to R. B. and R. H.

These authors contributed equally to the paper.

*Corresponding author; e-mail andreas.bachmair@univie.ac.at; fax +43 1 4277 9748.

The author responsible for distribution of materials in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Andreas Bachmair (andreas.bachmair@univie.ac.at).
(Abstract)
The higher plant Arabidopsis thaliana has eight genes potentially coding for small ubiquitin-related modifier (SUMO) proteins. However, two well-expressed isoforms differ from fungal and animal consensus in a conserved Gln residue situated four residues from the carboxyl terminus. We tested deviations in this position in the background of SUMO1, the isoform with the highest expression level, and found that changes do not prevent conjugation to substrate proteins in vivo. Replacement of this conserved Gln by Ala resulted in a protein that was less readily removed from a substrate by SUMO protease ESD4 in an in vitro reaction, and apparently led to higher levels of SUMO conjugates when expressed in vivo. We used SUMO1 variant with Gln to Ala substitution, as well as SUMO3 and SUMO5 (which carry Met and Leu, respectively, at this position) to enrich in vivo substrates. Identification of the most abundant proteins contained in these fractions indicated that they are involved in DNA-related, or in RNA-dependent processes, such as regulation of chromatin structure, splicing, or translation. The majority of the identified bona fide substrates contain predicted sumoylation sites. A subset of the proteins was expressed in E. coli and could be sumoylated in vitro.
Protein modification allows for adjustment of protein properties following their synthesis, and plays multiple roles in regulation. Among the large and growing number of modification types, attachment of SUMO (small ubiquitin-related modifier) was found to regulate nuclear and cytoplasmic processes. SUMO is covalently linked to substrate proteins by an enzyme cascade of SUMO activating enzyme (SAE) and SUMO conjugating enzyme (SCE), and was shown in several cases to depend on SUMO ligases for substrate selection. SUMO attachment can be reversed by specific cysteine proteases, which recognize the SUMO carboxyl terminus as exclusive substrate, and hydrolyze its linkage to the target protein (Dohmen, 2004; Dye and Schulman, 2007; Geiss-Friedlander and Melchior, 2007; Kerscher et al., 2006; Miura et al., 2007; Novatchkova et al., 2004). While usually only a minor fraction of a substrate protein is linked to SUMO at a given time, the modification can nonetheless be essential, if it constitutes part of an activity cycle (Johnson, 2004). For instance, SUMO has been implicated in assembly and disassembly processes of protein complexes. In some cases, sumoylation antagonizes or promotes other modifications such as ubiquitylation (Kerscher et al., 2006; Tatham et al., 2008).

In plants, SUMO was recently discovered to influence a variety of responses to the environment. SUMO is involved in tolerance to cold, heat, drought and salt stress (Catala et al., 2007; Conti et al., 2008; Kurepa et al., 2003; Miura et al., 2007; Yoo et al., 2006), modulates ABA responses (Lois et al., 2003), has an important role in phosphate homeostasis (Miura et al., 2005), and controls the time of flower initiation (Murtas et al., 2003; Jin et al., 2008). The finding that pathogens manipulate SUMO conjugation (Hotson et al., 2003; Roden et al., 2004), and that sumoylation influences innate immunity (Lee et al., 2007), implies an important role in plant-pathogen relations.

Arabidopsis encodes eight SUMO genes (Kurepa et al., 2003; Novatchkova et al., 2004), at least half of which are expressed to a significant extent. The significance of different isoforms is still unclear, but differences in the activity of SUMO proteases towards different isoforms have been reported (Chosed et al., 2006; Chosed et al., 2007; Colby et al., 2006; Mukhopadhyay and Dasso, 2007; Hay, 2007). Here we show that variations in one conserved residue do not abolish functionality, but nonetheless influence the properties of a SUMO protein that contains the variant residue. Furthermore, we identify putative in vivo sumoylation substrates of Arabidopsis by enrichment of proteins conjugated to tagged SUMO.
isoforms, followed by mass spectrometric identification. Most of the identified proteins are functionally linked to transcription, chromatin modification, RNA metabolism, or translation. SUMO is a predominantly nuclear protein, but sumoylation also occurs in the cytoplasm, and the set of candidate substrates identified in this work encompasses both nuclear and cytoplasmic proteins.

RESULTS

SUMO Isoforms of Arabidopsis Are Differentially Expressed

The Arabidopsis genome encodes eight potential SUMO proteins (Kurepa et al., 2003; Novatchkova et al., 2004). SUMO1, 2, 3 and 5 are highly expressed and therefore probably functional. The relative expression levels are SUMO1 ~ SUMO2 > SUMO3 ~ SUMO5, with SUMO3 and 5 mRNAs being approximately seven-fold less abundant than those of SUMO1 or 2. Putative SUMO proteins 4, 6, 7 and 8 are probably rare, because no cDNAs are found in EST libraries. Microarray data suggest low expression of SUMO4 (about 1% of the SUMO1 level), but are not informative regarding SUMO6, 7 or 8 (http://www.arabidopsis.org/), so that it is unclear under which conditions these open reading frames are expressed.

SUMO Proteins with Amino-Terminal Extension Are Conjugated In Vivo

SUMO isoform 1 (At4g26840; Kurepa et al., 2003; Novatchkova et al., 2004) is highly expressed and well represented in EST data bases. To analyze in vivo sumoylation, a SUMO1 gene with an amino-terminal extension was constructed. One extension (tag1) contains a hexa-histidine tag, followed by a triple HA tag (see Materials and Methods). A second extension (tag3), used for most of the work presented, contains a Strep tag, a triple HA tag, and octa-histidine (extension sequence shown in Fig. 1A). Mature SUMO proteins containing either of these extensions were expressed in Arabidopsis plants. Fig. 2 (lane 3, extract from plants expressing tag1-SUMO1, vs. control lanes 1 and 2, and lane 4, extract from plants expressing tag3-SUMO1) demonstrates that SUMO1 carrying either amino-terminal extension is incorporated into higher molecular weight material.

Because SUMO1 and 2 are highly similar, we did not include SUMO2 in these investigations. Functional equivalence of SUMO1 and 2 was indeed recently demonstrated (Saracco et al., 2007). We did, however, use the distinct isoforms SUMO3 and 5 for in vivo
and in vitro experiments. We thus constructed transgenes for in planta expression of tag3-SUMO3 (At5g55170) and tag3-SUMO5 (At2g32765). Fig. 2 shows that tag3-SUMO3 and tag3-SUMO5 are also incorporated into higher molecular weight material, suggestive of conjugation. There are subtle differences in the pattern of conjugates between SUMO 1, 3 and 5. The high number of bands, however, does not permit conclusions whether the differences pertain to relative abundance of conjugates present for all three isoforms, or whether any of the substrates of SUMO3 and SUMO5 conjugation are distinct from those of SUMO1. For further work, tag3-SUMO1, 3 and 5 were expressed under control of a β-estradiol-inducible promoter (vector pER8; Zuo et al., 2000). Fig. 2, lanes 8 and 9, show the inducible tag3-SUMO5 construct as an example.

Differences Between SUMO Isoforms in a Usually Conserved Residue Do Not Prevent Conjugation, But May Influence Isoform Characteristics

Comparison of the SUMO gene family of Arabidopsis indicated that, interestingly, some of the SUMO isoforms of Arabidopsis do not contain residues conserved in animal or fungal SUMO proteins (Kurepa et al., 2003; Novatchkova et al., 2004). In particular, the usually conserved glutamine residue four amino acids from the carboxyl terminus is replaced by Met and Leu, respectively, in Arabidopsis isoforms SUMO3 and SUMO5 (Fig. 1B). With one exception (vertebrate SUMO4), deviation from the Gln consensus is apparently restricted to flowering plants (Supplementary Fig. 1). Based on structural data, this Gln was predicted to be important for interaction with SUMO-specific proteases (Mossessova and Lima, 2000), and may thus influence the efficiency of de-conjugation. Detection of in vivo conjugates (Fig. 2) indicates that changes at this position, as present in SUMO3 (Met instead of Gln) and SUMO5 (Leu instead of Gln), do not prevent conjugation, suggesting that the conserved Gln residue is not essential for the conjugation reaction. We engineered different residues into this position of SUMO1. Mutation of Gln to a hydrophobic (Leu), a basic (Arg), an acidic (Asp) and a small residue (Ala) were tested (Budhiraja, 2005). Interestingly, high level overexpression of (amino-terminally extended) SUMO1 with the Gln to Leu, or the Gln to Ala change had a deleterious effect on plant growth (shown for the Ala variant in Supplementary Figure 2A; Budhiraja, 2005), whereas similar overexpression of WT SUMO1 (with amino-terminal extension) resulted only in a slight retardation of growth (Supplementary Figure 2B). We concluded that change of Gln to a hydrophobic residue, or to Ala, leads to
changes in biological properties. In order to avoid possible negative effects of high over-expression, we used lower expression levels and an inducible promoter system for the enrichment of \textit{in vivo} substrates described below.

As a first biochemical characterization of SUMO1 Q90A, we capitalize in the following on the consequences of the SUMO1 Gln to Ala (Q90A) mutation on conjugation and de-conjugation reactions. Fig. 2 (lane 7) shows that SUMO1 Q90A expression leads to a similar conjugate pattern as SUMO1 expression. However, we consistently observed a slightly more intense antibody staining (Fig. 2 and data not shown; Coomassie-staining of control gels served to adjust to equal protein loading), suggestive of increased presence of the variant in conjugates.

**The SUMO1 Q90A Mutation Decreases De-sumoylation by SUMO Protease ESD4**

Previous data had indicated that the activity of SUMO proteases towards different SUMO isoforms can differ significantly (Chosed et al., 2006; Chosed et al., 2007; Colby et al., 2006; Mukhopadhyay and Dasso, 2007; Hay, 2007). The availability of sumoylated, recombinant plant proteins as described below allowed an \textit{in vitro} investigation of the differences between SUMO1 and SUMO1 Q90A regarding de-sumoylation. The activity of de-sumoylating enzyme ESD4 (Murtas et al., 2003) on SUMO1 versus SUMO1 Q90A was assessed in a qualitative way. An \textit{in vitro} sumoylation reaction was established using nucleosome assembly factor (NAF, At2g19480; see below and Table 1) as a substrate, and either SUMO1, or SUMO1 Q90A. Sumoylated and non-sumoylated substrate proteins were removed from sumoylation enzymes using affinity beads specific for the Flag tag on NAF. The isolated, bead-associated proteins were incubated with ESD4 protease fragment and analyzed by Western blotting. Figure 3A shows that, whereas SUMO1 Q90A-NAF was gradually de-sumoylated over a 60 minute incubation period, all SUMO1-NAF had been de-conjugated by 15 minutes under the same conditions, suggesting that SUMO1 de-conjugation was at least four times faster. In another experiment with less protease, we observed de-conjugation of 70% of the SUMO1-NAF conjugate after 30 minutes of incubation. However, no de-conjugation of SUMO1 Q90A could be observed under these conditions (Fig. 3B). In the latter experiment, secondary antibodies carried an infrared dye, which allowed quantification by measurement of emitted light after excitation. These experiments were repeated, and were consistent with the notion that SUMO1 Q90A can be de-conjugated by SUMO proteases, but
that its de-conjugation is significantly delayed compared to the un-mutated SUMO1. The exact magnitude of the difference remains to be determined, and may differ for different SUMO proteases.

**Enrichment and Mass Spectrometric Identification of SUMO Conjugates**

Based on tag-SUMO expressing plants, an enrichment procedure was established to facilitate mass-spectrometric identification of *in vivo* sumoylated proteins. As it turned out that de-conjugation was a major problem during enrichment, denaturing conditions were used wherever possible. Fig. 4 documents steps of the enrichment procedure. Both anti-SUMO1, and anti-HA antibodies allowed detection of SUMO conjugates in enriched fractions. However, because the polyclonal anti-SUMO1 serum showed additional bands (Fig. 4A), the monoclonal HA antibody was used preferentially to follow the enrichment. Protein extracts from mature whole plants (see Materials and Methods) were incubated with Ni affinity resin to bind proteins with hexa-histidine tag, which were subsequently eluted. Comparison of the amount of free tag-SUMO contained in crude extracts versus proteins eluted from Ni affinity resin suggested that there was an at least fifty-fold enrichment (Western blot using anti HA antibody, Fig. 4B). Fig. 4C shows a comparison of a Coomassie-stained gel of Ni resin eluate (lane 5) with a parallel Western blot using anti-HA antibody (lane 6). While the Western blot documents the enrichment of sumoylated proteins, these proteins do in general not correspond to visible bands in the Coomassie-stained preparation. However, application of proteins eluted from the Ni column onto anti-HA resin allows a further enrichment. Fig. 4D shows a comparison of Coomassie staining (lane 7) and Western blot (lane 8) with such a fraction. In this case, many major bands of the Western blot coincide with visible bands in the Coomassie-stained gel. We therefore concluded that the protein fraction obtained from the Ni affinity step contains a significant amount of sumoylated proteins, but that these proteins are usually overlaid by protein contaminants. Because the anti-HA affinity step requires semi-native conditions (the sample was diluted to 0.3 M urea in order to allow binding), the procedure results in considerable loss of signal (as judged by Western blotting), and variable yields (see also Discussion). The step, while useful to assess the abundance of sumoylated proteins in enriched fractions, was therefore not used to generate material for mass spectrometry. Fig. 4E shows a typical preparative gel, which was cut into pieces according to molecular weight, and used for mass spectrometry.
The (moderately) increased level of conjugates with SUMO1 Q90A compared to SUMO1 (Fig. 2 and data not shown) motivated the use of this SUMO1 variant to enrich SUMO1 conjugates for identification. SUMO3 and SUMO5 were used in similar experiments. Because we found that high level expression of SUMO1 Q90A, of SUMO3 or SUMO5 with a constitutive promoter delayed growth and could even cause senescence-like symptoms on leaf margins (Supplementary Fig. 2; Budhiraja, 2005), we used estradiol-inducible constructs for these experiments (Fig. 2, lanes 8, 9). Control experiments indicated that the expression levels achieved with the inducible system do not compromise plant growth (data not shown and Supplemental Figure 2). Greenhouse-grown mature transgenic plants were induced by spraying with 5 μM β-estradiol. 16 hours after spraying, the material was harvested and processed as described above and in Materials and Methods. When the procedure was applied to non-transgenic Arabidopsis plants, the amount of protein recovered was slightly smaller than the material recovered from plants induced to express tag3-SUMO1 Q90A (Fig. 4E, lane 10 vs. lane 9). Gel lanes were cut into pieces, to represent defined protein size fractions, and processed for protein identification by LC/MS/MS. Similar protein extracts were prepared from the plants expressing tag3-SUMO3 and tag3-SUMO5.

A number of proteins were identified in extracts from plants that did not contain SUMO transgenes with hexa-histidine tag. Control experiments suggested that the major contaminant, Ribulose bisphosphate carboxylase / oxygenase (RUBISCO; visible as the strongest band at position of the 54 kDa marker in Figs. 4C - E, but present in each size fraction of the gel) is significantly retained on the Sephadex-based matrix irrespective of the Ni-NTA affinity ligand, and constitutes the majority of the non-specific background. Other protein contaminants, such as catalases, may be natural chelators of the Ni affinity ligand. For some proteins, for instance At1g80480 (annotated as PRL1 interacting protein L), the latter property was also supported by poly-histidine tract(s) in the protein sequence. Phospholipase D alpha1 (At3g15730) was another prominent contaminant.

Next to proteins enriched and identified from WT plant extracts, which were obvious contaminants in preparations from transgenic plants, there were many proteins specific for preparations from tag3-SUMO expressing plants. The candidate list contains only proteins for which at least three polypeptides were sequenced from enriched fractions, but no peptide was present in control preparations. The apparent molecular weight of identified proteins, as determined by the position in the gel, was a further criterion to decide whether the protein
was a likely SUMO substrate. The conjugation of transgenic SUMO to a protein is supposed to increase its apparent molecular weight. Thus, only proteins that migrated 20 to 40 kDa above their expected molecular weight were retained on candidate lists. Table 1 shows putative sumoylation targets with the highest scores. The identification method has an obvious bias for more abundant proteins. For instance, peptides from transcription factors were frequently identified (e.g., from IAA18 / At1g51950, from Perianthia / At1g685409, or from FWA / At4g25530), but more than two peptides were rarely identified in one gel piece, so that scores for these candidate substrates stayed relatively low. Likewise, peptide WSVIAR, which occurs in several Myb domain transcription factors, was consistently obtained.

Canonical or high probability sumoylation sites (sumoplot of www.abgent.com) are statistically enriched among candidate proteins (see Discussion). The most noteworthy candidate in this respect is At5g08450, which contains more than ten high probability sumoylation sites. The presence of an Rxt3 domain in this protein suggests participation in histone de-acetylation. An exception to the trend is At1g55300 (TAF7), a subunit of general transcription factor TFIID (Thomas and Chiang, 2006), which has no predicted high probability sumoylation site.

**In Vitro Sumoylation of Candidate Proteins**

The mass spectrometric information regarding sumoylation substrates, as listed in Table 1, was complemented by *in vitro* sumoylation assays with a subset of the identified proteins. Arabidopsis has two forms of SUMO activating enzyme SAE, which share the same large subunit, but differ in the smaller subunit (Kurepa et al., 2003; Novatchkova et al., 2004). Previous knowledge suggested that if produced in *E. coli*, both subunits should be co-expressed for optimal enzyme activity (F. Melchior, personal communication). We therefore constructed a di-cistronic mRNA that expressed the larger SAE subunit, SAE2 (At2g21470; Kurepa et al., 2003; Novatchkova et al., 2004) as the first reading frame, followed by either SAE1a, or by SAE1b with an amino-terminal hexa-histidine tag. The two ORFs were linked as found in a poly-cistronic mRNA of *E. coli* that encodes ribosomal proteins (Yates and Nomura, 1980). We expected that this arrangement results in minimal ribosome disassembly at the end of the first ORF, allowing close to equimolar production of both proteins. SUMO conjugating enzyme SCE (At3g57870), a single copy gene in Arabidopsis, was expressed and
SUMO1 and SUMO1 Q90A were expressed with the amino-terminal extension peptide as used for expression in planta (Fig. 1A). In most experiments, a fragment of SUMO ligase SIZ1 (At5g60410; Miura et al., 2005) was included in the reaction. This SIZ1 clone contained the 478 carboxyl-terminal residues that include the SP-RING. It lacks a putative chromatin-binding domain to increase its solubility in E. coli, but also lacks the PHD finger that was recently shown to contribute to certain sumoylation reactions (Garcia-Dominguez et al., 2008).

SUMO chain formation was used as a test of sumoylation components (Fig. 5). These experiments allowed the following conclusions: Firstly, SAE preparations containing SAE1b fared as well as, and occasionally even better than, those containing SAE1a (lanes SAE1b vs. lanes SAE1a of reaction with SUM1 of Fig. 5). The former isoform was used in substrate-based assays shown in the following. Secondly, consistent with previous reports, SUMO1 formed chains in vitro (Colby et al., 2006; putative sumoylation site indicated in Fig. 1A). It is therefore possible that SUMO1 chain formation is functionally important in plants, similar to the recently reported role in yeast and animals (Uzunova et al., 2007; Tatham et al., 2008). In contrast, experiments with SUMO3 and SUMO5 suggested that these latter two isoforms have a much lower tendency for chain formation in the same assay (data not shown; see also substrate-based assays in Fig. 7). Thirdly, SUMO ligase SIZ1 enhanced in vitro conjugation (lanes +SIZ1 vs. −SIZ in Fig. 5). This effect was also found in most substrate-based assays. Fourthly, SUMO1 variant SUMO1 Q90A behaved similar to SUMO1 in these conjugation reactions (left panel vs. right panel of Fig. 5). This finding could be confirmed in all substrate-based assays (see below).
SUMO3 chains, only one higher MW band was obtained with nucleosome assembly factor (NAF; At2g19480), or with TAF7 (At1g55300). The protein encoded by At5g08450 (labeled RXT3 in Fig. 7) contains 13 predicted high probability sumoylation sites. Its in vitro sumoylation results in multiple bands, consistent with SUMO attachment at several positions. Isolation and mass-spectrometric analysis of SUMO1-NAF and SUMO1-RRM confirmed that sumoylation did not occur in the Flag tag of these proteins (T. Colby, unpublished).

DISCUSSION

In this work, we describe biochemical properties of the Arabidopsis sumoylation system based on in vivo and in vitro experiments, and we compile a short list of potential in vivo sumoylation substrates of Arabidopsis.

Arabidopsis encodes eight SUMO proteins (Kurepa et al., 2003; Novatchkova et al., 2004). Four of them are expressed to a considerable extent. Because they differ in an amino acid residue that is conserved in fungal and most animal SUMO proteins, we tested amino acid changes in the sequence context of SUMO1, the most abundant isoform, which carries the standard residue in this position. SUMO3 and 5 have large hydrophobic residues instead of a conserved Gln four residues from the carboxyl terminus (Novatchkova et al., 2004). Mutations at this position in SUMO1 were tested (this work and Budhiraja, 2005). One of them, SUMO1 Q90A, was used for further characterization. This variant had an in vivo conjugation pattern similar to SUMO1 (Fig. 2), and in vitro sumoylation reactions resulted in an identical pattern of conjugate bands for all proteins tested (Fig. 7). However, the change resulted in decreased de-sumoylation of the variant (Fig. 3). This finding is consistent with structural data (Mossessova and Lima, 2000), and in fact testing variants at this position was in part motivated by the prospects to obtain a “non-cleavable” SUMO1 variant. De-conjugation was assessed by qualitative assays using an in vitro sumoylated substrate, and a fragment of the Arabidopsis SUMO protease ESD4 (Murtas et al., 2003). The data shown in Fig. 3, and additional experiments, indicate that SUMO1 Q90A is cleaved off the model substrate NAF (At2g19480), but at least four times slower than SUMO1. The result has obvious implications for SUMO3 and 5, which have Met and Leu, respectively, in this position. We hypothesize that the previously found differences in in vitro de-conjugation between different Arabidopsis SUMO isoforms (Chosed et al., 2006; Colby et al., 2006) are...
to a significant extent due to the nature of this critical residue. Distinct de-conjugation rates of different isoforms should translate into different lifetimes for conjugated substrate proteins in vivo, and thus imply different roles for different isoforms, even if the spectrum of substrates as such does not differ.

It is likely that changes close to the SUMO carboxyl terminus, as tested in the context of substrate conjugation and de-conjugation, also have an influence on the maturation of SUMO precursors, which is necessary to expose the mature SUMO carboxyl terminus. Assuming that the half life of SUMO isoforms SUMO3 and 5 largely exceeds the average half life of their conjugates, a slow maturation rate would still allow the majority of SUMO protein to be in the mature, i.e. active form, and would thus not have a significant influence on protein properties. However, it is also possible that maturation of SUMO3 and 5 is a regulated step, so that SUMO3/5 precursors can accumulate under certain circumstances. In this latter case, the amount of SUMO3/5 available for conjugation would differ from expectations based on mRNA abundance, but the properties of mature SUMO isoforms remain as suggested by this work. This proposition differs from findings concerning a human SUMO isoform, hSUMO4, which has Pro at position -4. In contrast to Arabidopsis SUMO1 Q90A, which can be slowly de-sumoylated in vitro (Fig. 3), the proteolytic processing of human SUMO1 with Pro at position -4, or of hSUMO4, could not be demonstrated in vitro. It was therefore proposed that hSUMO4 is not conjugated in vivo (Owerbach et al., 2005).

Arabidopsis SUMO1 variant Q90A, as well as SUMO3 and SUMO5, were expressed in vivo with amino-terminal tags that served for enrichment of potential SUMO conjugates. Experiments to obtain fractions enriched for SUMO conjugates have been described in other organisms (Hannich et al., 2005; Li et al., 2004; Panse et al., 2004; Wohlschlegel et al., 2004; Zhao et al., 2004; Zhuo et al., 2004). Some of these publications, dealing with Saccharomyces cerevisiae, describe a one-step purification scheme employing a hexahistidine tag. However, a similarly straightforward approach in Arabidopsis thaliana appeared more challenging due to the more complex proteome and concomitant decrease in signal to noise ratios. Whereas use of Ni affinity in combination with denaturing protein extract preparation resulted in consistent enrichment of sumoylated proteins as documented in Fig. 4, subsequent use of the HA tag in a two step procedure resulted in variable yield and significant losses. A possible reason was that HA affinity purification with an antibody column does not tolerate completely denaturing conditions, and co-enrichment of proteases
(e.g., At1g09730, a predicted SUMO-specific protease, could be identified as a low scoring component of enriched fractions) might pose a constant danger. For this reason, single step enrichment via hexa-histidine tag under completely denaturing conditions was used for mass spectrometric protein identification. The major contaminant, RUBISCO large subunit, contributes as much as half of the protein mass of these preparations. Peptides with the expected mass of the tryptic RUBISCO digest were therefore automatically discarded and not used for fragmentation. As a conservative estimate, we assume that the enriched fraction contains between two and five per cent sumoylated proteins. This estimate is based on protein gels and gel blots as shown in Fig. 4. The estimate is also consistent with the data collected by mass spectrometry, where more than one in ten peptides were unique to the sample fraction, and not present in fractions prepared from non-transgenic plants.

Potential sumoylation substrates were identified based on the following criteria: Firstly, the proteins co-purified with in vivo expressed tagged versions of SUMO proteins, but were absent from control preparations from non-transgenic plants. Secondly, at least three peptides derived from a putative substrate could be sequenced by LC/MS/MS, and none of these peptides was present in control preparations. Thirdly, the putative substrate proteins enriched in the preparations migrated in SDS PAGE at a molecular weight that was 20 to 40 kDa higher than their predicted position, suggestive of covalent modification by SUMO.

Many high-score candidates contain an RNA binding domain, or are involved in DNA-dependent or chromatin-related processes. A common theme is that candidate substrates participate in assembly and dis-assembly processes. This is true for nucleosome assembly factor, for components of the transcription machinery, for RNA-binding proteins with putative role in splicing, and for components of the translation machinery. Incidentally, the latter proteins are the first candidates for cytoplasmic substrates of sumoylation from the model plant.

A subset of the proteins listed in Table 1 could be expressed in E. coli and was subjected to in vitro sumoylation, further supporting their status as candidate substrates. Only one of the E. coli expressed proteins from Table 1, encoded by At2g43970, tested negative in the in vitro reaction. Many of the putative substrates contained consensus sumoylation sites, which were statistically enriched in the ensemble of candidate substrates. Whereas, for instance, 40 % of the human proteome was reported to contain a sumoylation consensus motif (Zhou et
al., 2005), 70% of the proteins listed in Table 1 contain such a motif, and more than 40% have more than one canonical sumoylation sequence.

A particularly interesting case was At5g08450 (labeled RXT3 in Fig. 7), a protein with similarity to yeast histone de-acetylase component Rxt3. This protein contains more than 10 high probability sumoylation sites, and can be efficiently sumoylated \textit{in vitro}. In contrast, At1g55300, the TAF7 subunit of general transcription factor TFIID (Thomas and Chiang, 2006), does not contain a single site, was identified in enriched preparations, and could nonetheless be sumoylated \textit{in vitro} (TAF7 of Fig. 7). Interestingly, TAF7 contains several sequences that resemble SUMO interaction motifs (SIMs; Hecker et al., 2006), so that association between the SUMO-SCE complex and TAF7 might rely on interaction with the SUMO moiety, not with SCE (which has affinity to exposed sumoylation consensus motifs).

While experiments regarding the Gln residue at position -4 imply a difference between SUMO isoforms 1, 3 and 5 in de-conjugation, the question whether different isoforms differ in substrate specificity, i.e. have distinct conjugation rates for certain substrates, remains open. Generally, \textit{in vitro} sumoylation by SUMO1 was possible with most substrates tested (Table 1), and \textit{in vitro} conjugation by SUMO3 worked less efficiently even with substrates identified as \textit{in vivo} SUMO3 substrates. Therefore, the \textit{in vitro} system did not display strong differences in substrate selection by different isoforms. Similarly, our \textit{in vivo} data had no bearing on this question, because substrate identification from enriched fractions was far below saturation. The absence of a particular substrate from preps made with one isoform versus another did therefore by no means imply that this substrate does not form \textit{in vivo} conjugates with the other isoform.

In summary, the work presented an investigation of Arabidopsis SUMO isoforms and introduced a set of plant substrates of the SUMO conjugation system. Whereas many of these substrates operate in the nucleus, some are cytoplasmic proteins, or occur in the nucleus and in cytoplasm (Table 1). They can be subject of further studies and may serve as standards in future \textit{in vitro} and \textit{in vivo} work. The \textit{in vitro} sumoylation system used in the work is based entirely on Arabidopsis components, and can be used to characterize additional components, such as new substrates, SUMO ligases, or SUMO proteases. Finally, the SUMO conjugate enrichment procedure described can be easily adapted to other ubiquitin-like protein modifiers.
MATERIALS AND METHODS

Plant Growth and Transformation
Plants were grown in the greenhouse under long day conditions (16 hours light, 8 hours darkness). Transgenes were introduced by floral dip transformation. Lines with high expression level were selected after Western blotting of protein extracts. Vector pER8-based SUMO constructs were induced by spraying mature plants (grown on soil for 45 days) with water containing 5 μM β-estradiol. After overnight exposure, plants were harvested on the following day.

Plasmid Construction
Plasmid pSK-TagSUMO-GG contains the sequence of mature SUMO1 with amino-terminal extension tag1 (MAHHHHHHHMGSPYDVPDYAGPYDVPDYAGSPYDVPDYAGY; single letter amino acid code; HA tags are bold and underlined, hexa-histidine tag is in bold). The extension sequence is preceded by an Xho I site, followed by part of the omega translational enhancer of tobacco mosaic virus. The stop codon of the SUMO ORF is followed by a unique Xba I site. The construct was obtained by first amplifying a fragment of SUMO1 from mRNA, using oligonucleotides and standard reverse transcription and PCR methods. The SUMO1 fragment ended with the Bst XI site close to the carboxyl terminus, followed by a newly introduced Xba I site. Thereafter, sequences encoding His6 and triple HA tag were inserted, and the missing portion of the carboxyl end was inserted as an oligonucleotide. For plant expression shown in Fig. 2, an Xho I – Xba I fragment was transferred into Xho I, Xba I digested vector pHi (Schlögelhofer and Bachmair, 2002).

Plasmid pSK-Tag3 contains the extension of all other SUMO constructs used in this work (see text for sequence). The ORF of the extension is preceded by an Xho I site, followed by part of the omega translational enhancer of tobacco mosaic virus. The extension ends with a Kpn I site, followed by an Xba I site. SUMO3 and 5 cDNA sequences were inserted as Kpn I – Xba I fragments. To obtain SUMO1 variants with changes at position 90, oligonucleotides spanning the Cla I – Xba I sites of a pSK based SUMO1 construct were used for cassette exchange, and mutated fragments were transferred thereafter into pSK-Tag3. Xho I – Xba I fragments of Tag3-containing clones were inserted into Xho I – Xba I digested pHi, or into an Xho I – Spe I digested variant of pER8 that had a reduced number of enhancer sequences.
due to a promoter deletion spanning nucleotides 3970 – 4047 (pER8 accession is AF309825). For expression in *E. coli*, Not I Klenow / Nco I fragments containing the coding regions from pSK-Tag3-SUMO1, pSK-Tag3-SUMO1 Q90A, pSK-Tag3-SUMO3 and pSK-Tag3-SUMO5, were inserted into Bam HI Klenow / Nco I digested vector pET9d (Novagen).

For expression of SAE in *E. coli*, cDNAs encoding SAE2, SAE1a and SAE1b were obtained from RIKEN depository (pda 10760, pda 07771 and pda 08247, respectively; http://www.brc.riken.jp/lab/epd/Eng/; Seki et al., 2002). Using oligos CCA TGG TGT ACA GGC CAG ATC TGA GCC TGC TTC TAA GAA GAG AAG ACT and GAG CTC ATC TCC GTC CAT GGC ACC ATG GTG ATG ATG GTG ATG GGT CAT TAT TCA ACT CTT ATC TTC TT, a carboxyl-terminal fragment of SAE2 was amplified, which connected to SAE1a with sequence TAATG (TAA as stop codon of SAE2, ATG as start codon of SAE1a), and simultaneously extended the SAE1a reading frame by His6-codons. The PCR fragment was inserted into Ecl 136II digested pSK. A Sca I – Sac I fragment from cDNA pda 08247 was inserted into the Spe I (recessive ends filled with Klenow DNA polymerase) – Sac I digested pSK-based vector. The latter vector was digested with Bsr GI and Bgl II, and a Bsr GI – Bam HI fragment from SAE2 cDNA was inserted. Vector pET9d was digested with Nco I and Bam HI, and an Nco I – Bam HI fragment from SAE2 cDNA was inserted. Thereafter, the pSK based plasmid was digested with Bam HI and Bsr GI, and the insert was combined with the Bam HI – Bsr GI-digested pET-based vector. Finally, a spurious frameshift mutation in the SAE2 part was removed by inserting an Afl II – Bsr GI fragment amplified from cellular mRNA by reverse transcription and PCR. The clone was called pETS1AE1b2. To obtain the analogous clone with isoform SAE1a, (called pETS1AE1a2), a Pvu II – Sal I fragment from pETS1AE1b2 was inserted into Ecl 136II – Sal I digested vector pSK. The ensuing clone was digested with Nco I and Sal I, and oligonucleotides CAT GCG AGC ATG GAC GGA GAA GAG CCC GGG ATC C and TCG AGG ATC CCG GGC TCT TCT C were annealed and inserted. The resulting plasmid was digested with Sap I and Sma I, and a Sap I – Pme I fragment from the SAE1a cDNA was inserted. Finally, a Bsr GI – Bam HI fragment from this construct was used to replace the Bsr GI – Bam HI fragment of pETS1AE1b2, to result in pETS1AE1a2. For expression of SCE1 in *E. coli*, a cDNA fragment was amplified by reverse transcription and PCR from cellular mRNA. The fragment was flanked by Nco I (at the amino terminus) and Sma I (after the stop codon) and cloned between Nco I and filled-in Bam HI sites of vector pET9d. A similar construct with a Cys to
Ser (TGT to AGT) change (mutated clone kindly provided by Dr. C. Hardtke, Univ. of Lausanne) was also prepared and called pETSCE C94S. A fragment of SUMO ligase SIZ1, spanning amino acid residues 370 to 873 of the protein, was inserted into vector pDEST17 for expression in *E. coli*.

Protein substrates for *in vitro* sumoylation assays were obtained in the following way: GST-S was expressed by induction of pET42c (Novagen). For UBC27Flag expression, UBC27 was amplified from cDNA pda06125 (RIKEN), using oligos CCG CCC GTC ATA TGA TAG ATT TCA GTC GAA TC and GGC CTC GAG CTT GTC ATC GTC GTC CTT GTA GTC GGT ACC AGC AGA ACA GAG CTT TTC C. The ensuing fragment was digested with Nde I and Xho I, and inserted into Nde I Xho I digested vector pET42c (Novagen) to give pETUBC27FlagHis. pETGST-S-Flag, the vector for expression of GST-S-Flag, was constructed by inserting an Nde I Kpn I fragment from pET42c into Nde I Kpn I digested vector pETUBC27FlagHis, and replacing a Swa I Spe I fragment of the ensuing construct by a Swa I Spe I fragment that was generated using oligos CCA GAA CCA CTA GTT GAA CCA TCC AGG GGA GGA T and GCT GAA AAT GTT CGA AGA TCT GTC GGT ACC AGC AGA ACA GAG CTT TTC C. The ensuing fragment was digested with Nde I and Xho I, and inserted into Nde I Xho I digested vector pET42c (Novagen) to give pETUBC27FlagHis. pETGST-S-Flag, the vector for expression of GST-S-Flag, was constructed by inserting an Nde I Kpn I fragment from pET42c into Nde I Kpn I digested vector pETUBC27FlagHis, and replacing a Swa I Spe I fragment of the ensuing construct by a Swa I Spe I fragment that was generated using oligos CCA GAA CCA CTA GTT GAA CCA TCC AGG GGA GGA T and GCT GAA AAT GTT CGA AGA TCT GTC GGT ACC AGC AGA ACA GAG CTT TTC C. The ensuing fragment was digested with Nde I and Xho I, and inserted into Nde I Xho I digested vector pET42c (Novagen) to give pETUBC27FlagHis. pETGST-S-Flag, the vector for expression of GST-S-Flag, was constructed by inserting an Nde I Kpn I fragment from pET42c into Nde I Kpn I digested vector pETUBC27FlagHis, and replacing a Swa I Spe I fragment of the ensuing construct by a Swa I Spe I fragment that was generated using oligos CCA GAA CCA CTA GTT GAA CCA TCC AGG GGA GGA T and GCT GAA AAT GTT CGA AGA TCT GTC GGT ACC AGC AGA ACA GAG CTT TTC C. The ensuing fragment was digested with Nde I and Xho I, and inserted into Nde I Xho I digested vector pET42c (Novagen) to give pETUBC27FlagHis. pETGST-S-Flag, the vector for expression of GST-S-Flag, was constructed by inserting an Nde I Kpn I fragment from pET42c into Nde I Kpn I digested vector pETUBC27FlagHis, and replacing a Swa I Spe I fragment of the ensuing construct by a Swa I Spe I fragment that was generated using oligos CCA GAA CCA CTA GTT GAA CCA TCC AGG GGA GGA T and GCT GAA AAT GTT CGA AGA TCT GTC GGT ACC AGC AGA ACA GAG CTT TTC C. The ensuing fragment was digested with Nde I and Xho I, and inserted into Nde I Xho I digested vector pET42c (Novagen) to give pETUBC27FlagHis. pETGST-S-Flag, the vector for expression of GST-S-Flag, was constructed by inserting an Nde I Kpn I fragment from pET42c into Nde I Kpn I digested vector pETUBC27FlagHis, and replacing a Swa I Spe I fragment of the ensuing construct by a Swa I Spe I fragment that was generated using oligos CCA GAA CCA CTA GTT GAA CCA TCC AGG GGA GGA T and GCT GAA AAT GTT CGA AGA TCT GTC GGT ACC AGC AGA ACA GAG CTT TTC C. The ensuing fragment was digested with Nde I and Xho I, and inserted into Nde I Xho I digested vector pET42c (Novagen) to give pETUBC27FlagHis. pETGST-S-Flag, the vector for expression of GST-S-Flag, was constructed by inserting an Nde I Kpn I fragment from pET42c into Nde I Kpn I digested vector pETUBC27FlagHis, and replacing a Swa I Spe I fragment of the ensuing construct by a Swa I Spe I fragment that was generated using oligos CCA GAA CCA CTA GTT GAA CCA TCC AGG GGA GGA T and GCT GAA AAT GTT CGA AGA TCT GTC GGT ACC AGC AGA ACA GAG CTT TTC C. The ensuing fragment was digested with Nde I and Xho I, and inserted into Nde I Xho I digested vector pET42c (Novagen) to give pETUBC27FlagHis.
amplify a 3’ portion, and restriction enzymes Aat II and Sex AI to insert the central fragment from pda08474. To express the ORF of At5g08450 (RXT3) with amino-terminal GST tag (without S peptide), cDNA pda08750 was used as template together with oligos CCG CGC CAA TTG GTC ATA TGA GTG GTG TTC CAA AGA GAT and CCG CGC GGT ACC TTA ACC GTT CTA GAT T, the ensuing PCR fragment was digested with Mfe I and Kpn I and inserted into Mfe I Kpn I digested vector pET42c. After digestion of the latter vector with Bam HI and partial digestion with Xba I, a Bam HI Xba I fragment from cDNA pda08750 was inserted to give vector pET RXT3. For expression of TAF7 as a GST fusion, an Neo I – Not I fragment from pUNI clone U63389 (Yamada et al., 2003) was inserted between Nco I and Not I sites of vector pET42c. Thereafter, the ensuing vector was digested with Bgl II and Mfe I, treated with Klenow fragment and re-ligated to obtain pET-TAFΔS.

**Plant Extraction**

SUMO substrates for mass spectrometric analysis were obtained from induced plants by immersion of typically 200 g plant material in 5 volumes of buffer A (6 M Guanidine hydrochloride, 0.1 M Tris Cl and 0.1 M Na phosphate buffer pH 8 supplemented with 20 mM β-mercaptoethanol, 10 mM Na metabisulfite, 3% polyvinyl-pyrrolidone, 5% (w/v) sucrose and 5 mM imidazole). The material was homogenized with a blender (POLYTRON PT 2100, Kinematica). The slurry was stirred for 30 minutes and then centrifuged at 3000 x g for 30 minutes. The supernatant was filtered through MiraCloth, centrifuged again (30 000 x g, 60 min), and incubated overnight at room temperature (constant agitation) with 4 ml of Ni-NTA resin (Qiagen). The resin was filled into a disposable column (Poly-Prep; BioRad) and washed with 5 - 10 volumes of buffer B (8 M urea, 0.1 M Na phosphate, 0.1 M Tris Cl pH 8). Protein was eluted with 1 – 2 column volumes of buffer C (8 M urea, 0.2 M acetic acid). The eluate was neutralized with 1 M Tris Cl pH8 and concentrated by ultrafiltration (Centriprep and Centricon tubes, 10 kDa cutoff; Millipore). To remove background, the eluate of Ni-NTA resin was in some cases re-applied onto a fresh batch of Ni-NTA resin, followed by elution as described above. For the experiment shown in Fig. 4D, the Ni-NTA eluate was diluted to 0.3 M urea. Triton X 100 (1% final concentration), SDS (0.1 % final concentration), protease inhibitor cocktail (Roche) and 2-mercaptoethanol (1 mM) were added, and the material was applied to anti-HA resin (Roche). Bound proteins were eluted using 8M urea in glycine buffer (100 mM, pH 2.2), followed by neutralization of the eluate.
Mass Spectrometric Analysis

*In-gel digestion:* Proteins were separated on standard SDS-polyacrylamide gels. Coomassie-stained bands were excised from the gel, and treated as described (Shevchenko et al., 1996), except that no CaCl₂ was added during digestion with trypsin (sequencing grade, Promega), and final extraction of peptides was carried out with 100 µl 1 % Trifluoroacetic acid (TFA) for 30 min at 37 °C. The extraction was repeated once with 100 µl 0.1 % TFA – acetonitrile (1:2). The volume of the combined supernatants was reduced to 5 µl in a vacuum centrifuge and 20 µl 0.1 % TFA were added to each sample.

*LC-MS/MS of in-gel digested proteins:* Liquid chromatography (LC)-MS data were acquired on a quadrupole-TOF mass spectrometer (Q-Tof II, Micromass, Manchester, UK) equipped with a Z spray source. Samples were introduced by an Ultimate nano-LC system (LC Packings) equipped with the Famos autosampler and a Switchos column switching module. The column setup comprises a 0.3 mm by 1 mm trapping column and a 0.075 by 150 mm analytical column, both packed with 3 µm Atlantis dC18 (Waters). A total of 10 µl was injected onto the trap column and desalted for 1 min with 0.1% trifluoroacetic acid at a flow rate of 10 µl/min. Peptides were eluted onto the analytical column by a gradient of 2% acetonitrile in 0.1% formic acid to 40% acetonitrile in 0.1% formic acid over 55 min at a column flow rate of ca. 200 nl/min, resulting from a 1:1,000 split of the 200 µl/min flow delivered by the pump. The electrospray ionization interface comprised an uncoated 10 µm i.d. PicoTip spray emitter (New Objective) linked to the HPLC flow path using a 7 µl dead volume stainless mounted onto the PicoTip holder assembly (New Objective). Stable nanospray was established by the application of 1.7 to 2.4 kV to the stainless steel union. The data-dependent acquisition of MS and tandem MS (MS/MS) spectra was controlled by the Masslynx 4.0. Survey scans of 1 s covered the range from m/z 360 to 1,200. Doubly and triply charged ions rising above a given threshold were selected for MS/MS experiments. In MS/MS mode the mass range from m/z 50 to 1,200 was scanned in 1 s, and 3 scans were added up for each experiment. Micromass-formatted peak-lists were generated from the raw data by using the Proteinlynx software module. Proteins were identified by searching the NCBI nr public database (National Center for Biotechnology Information) using a local
installation of MASCOT 1.9. (Matrix Science). A mass deviation of 0.5 Da was allowed for peptide and fragment ions.

**Protein Expression**
Proteins were expressed in *E. coli* strain ROSETTA (DE3) pLysS (Novagen), or in strain BL21 (AtSIZ1 fragment). Purification via Ni-NTA resin (Qiagen) or glutathione-agarose (Novagen) followed the resin manufacturer’s recommendations. If necessary, buffer changes and concentration of protein samples were carried out by centrifugation (Vivaspin 500, Sartorius). SCE1 was purified from pelleted, induced *E. coli* cells. After addition of 1/30 vol lysis buffer (50 mM Na phosphate pH 6.5, 50 mM NaCl, 1 mM DTT, protease inhibitor mix Complete (Roche)), the cells were subjected to freeze-thaw cycles and centrifuged (1 hr 100 000 x g, 4°C). The supernatant was applied to SP Sepharose (GE Healthcare) pre-equilibrated with lysis buffer. After washing (3 bed volumes of lysis buffer), batch elution was carried out with lysis buffer plus 300 mM NaCl. Active fractions were subjected to buffer change by ultrafiltration (20 mM Hepes/KOH pH 7.3, 10 mM K-Acetate, 2 mM Mg-Acetate, 0.5 mM EGTA, 1 mM DTT and protease inhibitors Aprotinin, Leupeptin and Pepstatin (1 μg/ml each)).

**In Vitro Sumoylation**
Typical sumoylation reactions were carried out for four hours, or overnight at 30°C in 50 μl reaction volume, containing 5 mM ATP, 5 mM Mg++, 20 mM Tris pH 7.5, 100 μg tag-SUMO, 4 μg SAE, 0.6 μg SCE and 0.15 μg SIZ1 protein fragment. Between 1 and 10 μg substrate was added. Aliquots of 10 μl were separated by SDS-PAGE, and processed by Western blotting to detect substrate-specific bands using either Flag antibody-alkaline phosphate conjugate (Sigma), or anti-GST antibody followed by alkaline phosphatase-coupled secondary antibody.

**In Vitro De-Sumoylation**
An *in vitro* sumoylation reaction with substrate NAF was incubated with anti Flag resin (Sigma) in buffer (50mM Tris pH7.5, 150 mM NaCl, 1 mM EDTA and 1% Triton-X 100) and washed twice with TBS, and finally with SUMOylation assay buffer. *E. coli*-expressed ESD4 fragment was purified using Ni++ affinity, and the eluate in buffer containing 50 mM
NaH₂PO₄, 300 mM NaCl and 250 mM Imidazole (pH 8.0) was added to the resin-bound substrate. Aliquots were withdrawn, and the reaction was terminated by addition of N-ethyl maleimide to a final concentration of 2 mM at the indicated times. Sample buffer was added and aliquots were applied for gel electrophoresis and Western blotting.

**Protein Detection**

Proteins were separated on lab-made minigels except for Fig. 2, lanes 8 and 9, where a pre-cast gradient gel (3 to 12 %; Invitrogen) was used, and detected by Western blotting essentially as described (Stary et al., 2003), using Immobilon (Millipore) membrane for protein support, followed by detection with alkaline phosphatase-coupled secondary antibody (Sigma). Primary antibodies were directed against protein tags HA (Roche), Flag (Sigma), or GST (Sigma), or against SUMO (Murtas et al., 2003; Abcam). For experiments shown in Fig. 3, proteins were transferred to nitrocellulose membrane and detected using anti Flag primary antibody, followed by either HRP-coupled secondary antibody (Amersham ECL, GE Healthcare; Fig. 3A), or IR Dye 800-coupled secondary antibody (Rockland; detection by the Odyssey Infrared Imager, LiCor; Fig. 3B) as described (Garzón et al., 2007). Protein concentration in plant extracts was assessed by gel electrophoresis of aliquots, followed by Coomassie staining. Adjusted volumes with equal protein content were applied to the blotting gels.

**ACKNOWLEDGMENTS**

We wish to thank K. Luxa, M. Lehnen, A. Krull and A. Harzen for expert technical help, Dr. Y.-F. Fu for SAE2 cDNA and the SIZ1 expression clone, Dr. C. Hardtke for the SCE1 C94S clone, Dr. F. Melchior for advice on SCE and SAE purification, and to Dr. N. Elrouby for reading of the manuscript.

**LITERATURE CITED**


Owerbach D, McKay EM, Yeh ETH, Gabbay KH, Bohren KM (2005) A proline-90 residue unique to SUMO-4 prevents maturation and sumoylation. Biochem Biophys Res Commun 337: 517-520


FIGURE LEGENDS

Figure 1. A, Amino-terminal extension of AtSUM1 (capital letters) by tag3 to allow detection and enrichment of sumoylated protein substrates with minimal disruption of functionality. tag3 (single letter code in small letters) consists of a Strep tag (bold), three HA tags (bold, underlined), and an octa-histidine sequence (bold). K residue in position 10 of the SUM1 sequence (bold) was previously identified as a site of SUMO attachment in SUMO chains. B, Alignment of carboxy termini of mature SUMO proteins (single letter code). HsSUM1, human SUMO1; ScSMT3, SUMO of baker’s yeast; AtSUM1, 2, 3, 5, the four highly expressed SUMO isoforms of Arabidopsis. The glutamine at position four from the carboxyl terminus (bold, pos. 90 in AtSUM1) is conserved in animal and fungal SUMO proteins, but replaced by a hydrophobic amino acid in AtSUM3 and AtSUM5. In AtSUM1 Q90A, this residue was changed to Ala in the sequence of Arabidopsis SUMO1.

Figure 2. SUMO isoforms with amino-terminal extension are conjugated to protein substrates in vivo. Extracts from plants expressing different SUMO constructs were used for protein blotting and detection with antibody directed against the HA tag of the extension. SUMO1 (extended by tag1 at the amino terminus as indicated in the text; lane 3) was compared to extract from plants expressing SUMO1 without the two carboxyl-terminal
residues (SUMO1 ΔGG; lane 2), indicating dependence of the conjugation reaction on an intact carboxyl terminus. Lane 1 shows an extract from non-transgenic plants. SUMO isoforms SUMO3 and 5 (lanes 5 and 6, respectively), as well as a variant of SUMO1 with change Q90A (lane 7), amino-terminally extended by tag3 (cf. Fig. 1), can also form conjugates in vivo. The conjugate patterns of SUMO1 and SUMO1 Q90A are similar. Lanes 8 and 9, example of SUMO expression by an inducible promoter. Extract from cells transformed with an inducible tag3-SUMO5 construct contain conjugates to this protein only after induction (lane 9 vs. 8). Bars to the right indicate position of unconjugated tag-SUMO, dotted line indicates position of SUMO conjugates. Molecular weight marker sizes in kDa are indicated to the left.

**Figure 3.** Qualitative assessment of in vitro de-sumoylation of a substrate conjugated to SUMO1 versus SUMO1 Q90A. NAF protein with Flag tag was purified from *E. coli* and subjected to in vitro sumoylation, using either SUMO1 or SUMO1 Q90A. After re-isolation with anti-Flag resin, the material was incubated with SUMO protease ESD4 and harvested at the indicated times. Reaction products were detected on Western blots, using anti Flag antibody. SUMO1 Q90A-NAF (A, bottom panel) was gradually de-sumoylated, while SUMO1-NAF (A, top panel) was de-conjugated quantitatively within 15 minutes. At lower ESD4 concentration (panel B), 70% of SUMO1-NAF were de-sumoylated in a 30 minute incubation, while SUMO1 Q90A-NAF was almost unchanged. Band intensities in panel B were quantified with secondary antibody coupled to infrared dye, and detection of light emission. Unmodified NAF protein was used as an internal standard.

**Figure 4.** Enrichment of sumoylated proteins. A, An enriched protein fraction from tag-SUMO expressing plants was used for antibody detection using either rat monoclonal anti-HA antibody (lane 1), or polyclonal anti-SUMO antiserum (lane 2). The polyclonal serum detects additional bands. B, Crude plant extract (lane 3) was compared to an enriched fraction (lane 4) using gel blot and anti-HA detection. The intensity difference of the tag-SUMO band (bar to the right) suggests an approximately fifty-fold enrichment of SUMO and sumoylated proteins. C, Side-by-side comparison of a Coomassie-stained preparative gel of an enriched protein fraction (lane 5) with a Western blot of the same material (lane 6). The blot
demonstrates enrichment of sumoylated proteins, but the bands detected by the anti-HA antibody do not co-migrate with prominent bands of the Coomassie-stained gel. D, Material as shown in panel C was further purified using anti-HA affinity chromatography. The Coomassie-stained gel (lane 7) has prominent bands at the position of the most intense bands of the Western blot (lane 8), indicating that preparations as shown in panel C do contain significant amounts of SUMO conjugates. E, Coomassie-stained gel loaded with protein extract enriched by one step (Ni⁺⁺ affinity, cf. panel C), as used for mass spectrometric protein identification. Lane 9, extract from tag-SUMO transgenic plants; lane 10, material from non-transgenic plants (background control). In panels A and D, prominent bands visualized by two distinct detection methods are linked by horizontal lines. Dot, major contaminant RUBISCO large subunit; bar to the right, position of tag-SUMO. Molecular weight marker positions are shown to the left.

**Figure 5.** Test of sumoylation enzymes by formation of SUMO chains. SUMO activating enzymes SAE1a / SAE2 or SAE1b / SAE2 (labeling on top of gel pictures) were incubated with SCE (+ label), or without SCE (- label) together with tag3-SUMO1 (left panels), or with tag3-SUMO1 Q90A (right panel), in presence (+) or absence (-) of SUMO ligase SIZ1. Note that SUMO1 can form SUMO-SUMO linkages, in a pattern that is identical to that of SUMO1 Q90A. Bars to the right indicate position of di- tri- and tetra-SUMO (several bands, presumably due to conformation differences, or to different SUMO attachment sites). Bands were detected with antibody against the HA epitope present on SUMO.

**Figure 6.** *In vitro* sumoylation of protein tags. UBC27-Flag, a ubiquitin conjugating enzyme with a Flag tag extension, or glutathione S transferase encoded by vector pET42c (GST-S) can serve as a negative control. However, extension of GST-S by a Flag tag converts GST into a substrate (GST-S-Flag; dot to the right indicates position of the sumoylated form). Sumoylation is abolished by deletion of the S-peptide from the linker region between GST core and Flag peptide (panel GST-Flag). Control lanes labeled (-) contain an inactive (Cys 94 to Ser change) version of SCE, (+) indicates presence of WT SCE. Proteins were detected by Western blotting, using antibodies against Flag, of GST tags.
Figure 7. In vitro sumoylation of potential in vivo substrate proteins. Proteins identified as potential in vivo substrates were expressed in E. coli, purified and incubated with sumoylation enzymes. NAF, At2g19480; RRM, At3g56860; RXT3, At5g08450; TAF7, At1g55300. Control lanes labeled SCE (-) contain an inactive (Cys 94 to Ser change) version of SCE. SUMO1 and SUMO1 Q90A give identical patterns, while SUMO3 results in a single, considerably weaker sumoylation band of NAF or TAF7. Proteins NAF and RRM have a carboxyl-terminal Flag peptide which was used for antibody-based detection, whereas RXT3 and TAF7 have an amino-terminal GST tag (without S peptide) and were detected by anti-GST antibody.
Table 1. Putative *in vivo* sumoylation substrates of Arabidopsis\(^1\).

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Characteristics</th>
<th>ΨKxD/E(^2)</th>
<th><em>in vivo</em> conjugate to sumoylation</th>
<th><em>in vitro</em> sumoylation (^3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>RNA-dependent processes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At1g29400</td>
<td>AML5; has 1 RRM2, 2 RRM1 domains</td>
<td>1 (1) SUM3</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>At3g56860</td>
<td>contains 2 RRM1 domains</td>
<td>1 (0) SUM1</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>At2g43970</td>
<td>contains LA domain</td>
<td>0 (1) SUM1</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>At2g47020</td>
<td>translation release factor</td>
<td>1 (2) SUM3</td>
<td>n. t.</td>
<td></td>
</tr>
<tr>
<td>At3g60240</td>
<td>eIF4G</td>
<td>4 (2) SUM3</td>
<td>n. t.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Chromatin-related processes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At1g55300</td>
<td>TAF7 (TFIID subunit)</td>
<td>0 (0) SUM3, 5</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>At2g19480</td>
<td>nucleosome assembly factor</td>
<td>2 (0) SUM1</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>At5g08450</td>
<td>contains Rxt3 domain</td>
<td>3 (10) SUM3</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>At5g39660</td>
<td>zfDOF transcription factor</td>
<td>1 (1) SUM1</td>
<td>n. t.</td>
<td></td>
</tr>
<tr>
<td>At5g43130</td>
<td>TFIID subunit</td>
<td>2 (4) SUM1</td>
<td>n. t.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Others</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At1g27430</td>
<td>contains GYF domain</td>
<td>1 (4) SUM1, 3</td>
<td>n. t.</td>
<td></td>
</tr>
<tr>
<td>At1g64330</td>
<td>contains KIP domain, coiled-coil region</td>
<td>2 (1) SUM1, 3</td>
<td>n. t.</td>
<td></td>
</tr>
<tr>
<td>At5g13480</td>
<td>FY; WD40 repeat protein</td>
<td>0 (0) SUM1, 3</td>
<td>n. t.</td>
<td></td>
</tr>
<tr>
<td>At5g52300</td>
<td>RD29B desiccation induced</td>
<td>3 (2) SUM1</td>
<td>n. t.</td>
<td></td>
</tr>
</tbody>
</table>

\(^1\)Criteria for inclusion of a protein in Table 1 were robust detection in preparations, absence from control preparations and apparent molecular weight, as detailed in the text.

\(^2\)In addition to canonical SUMO attachment sites (Ψ is an aliphatic hydrophobic residue, x means any residue, D/E is either aspartic, or glutamic acid), many proteins contain near-canonical sequences as further potential SUMO attachment sites (number of such sites given in parentheses; www.abgent.com/doc/sumoplot).

\(^3\)The *in vitro* sumoylation reactions referred to in this table used SUMO1 and its variant SUMO1 Q90A, with identical results.
**Figure 1.** A, amino-terminal extension of AtSUM1 (capital letters) by tag3 to allow detection and enrichment of sumoylated protein substrates with minimal disruption of functionality. tag3 (single letter code in small letters) consists of a Strep tag (bold), three HA tags (bold, underlined), and an octa-Histidine sequence (bold). K residue (bold) in position 10 of the SUM1 sequence was previously identified as a site of SUMO attachment in SUMO chains. B, alignment of carboxyl termini of mature SUMO proteins (single letter code). HsSUM1, human SUMO1; ScSMT3, SUMO of baker’s yeast; AtSUM1, 2, 3, 5, the four highly expressed SUMO isoforms of Arabidopsis. The glutamine at position four from the carboxyl terminus (bold, pos. 90 in AtSUM1) is conserved in animal and fungal SUMO proteins, but replaced by a hydrophobic amino acid in AtSUM3 and AtSUM5. In AtSUM1 Q90A, this residue was changed to Ala in the sequence of Arabidopsis SUM1.
Figure 2. SUMO isoforms with amino-terminal extension are conjugated to protein substrates \textit{in vivo}. Extracts from plants expressing different SUMO constructs were used for protein blotting and detection with antibody directed against the HA tag of the extension. SUMO1 (extended by tag1 at the amino terminus as indicated in the text; lane 3) was compared to extract from plants expressing SUMO1 without the two carboxyl-terminal residues (SUMO1 ΔGG; lane 2), indicating dependence of the conjugation reaction on an intact carboxyl terminus. Lane 1 shows an extract from non-transgenic plants. SUMO isoforms SUMO3 and 5 (lanes 5 and 6, respectively), as well as a variant of SUMO1 with change Q90A (lane 7), amino-terminally extended by tag3 (cf. Fig. 1), can also form conjugates \textit{in vivo}. The conjugate patterns of SUMO1 and SUMO1 Q90A are similar. Lanes 8 and 9, example of SUMO expression by an inducible promoter. Extract from cells transformed with a tag3-SUMO5 construct contain conjugates to this protein only after induction (lane 9 vs. 8). Bars to the right indicate position of unconjugated tag-SUMO, dotted line indicates position of SUMO conjugates. Molecular weight marker sizes in kDa are indicated to the left.
Figure 3. Qualitative assessment of \textit{in vitro} de-sumoylation of a substrate conjugated to SUMO1 versus SUMO1 Q90A. NAF protein with Flang tag was purified from \textit{E. coli} and subjected to \textit{in vitro} sumoylation, using either SUMO1 or SUMO1 Q90A. After re-isolation with anti-Flag resin, the material was incubated with SUMO protease ESD4 and harvested at the indicated times. Reaction products were detected on Western blots, using anti Flag antibody. SUMO1 Q90A-NAF (A, bottom panel) was gradually de-sumoylated, while SUMO1-NAF (A, top panel) was de-conjugated quantitatively within 15 minutes. At lower ESD4 concentration (panel B), 70\% of SUMO1-NAF was de-conjugated in a 30 minute incubation, while SUMO1 Q90A-NAF was almost unchanged. Band intensities in panel B were quantified with secondary antibody coupled to infrared dye, and detection of light emission. Unmodified NAF protein was used as an internal standard.
**Figure 4.** Enrichment of sumoylated proteins. A, An enriched protein fraction from tag-SUMO expressing plants was used for antibody detection using either rat monoclonal anti-HA antibody (lane 1), or polyclonal anti-SUMO antibody (lane 2). The polyclonal serum detects additional bands. B, Crude plant extract (lane 3) was compared to an enriched fraction (lane 4) using gel blot and anti-HA detection. The intensity difference of the tag-SUMO band (bar to the right) suggests an approximately fifty-fold enrichment of SUMO and sumoylated proteins. C, Side-by-side comparison of a Coomassie-stained preparative gel of an enriched protein fraction with a Western blot of the same material. The blot demonstrates enrichment of sumoylated proteins, but the bands detected by the anti-HA antibody do not co-migrate with prominent bands of the Coomassie-stained gel. D, Material as shown in panel C was further purified using anti-HA affinity chromatography. The Coomassie-stained gel (lane 7) has prominent bands at the position of the most intense bands of the Western blot (lane 8), indicating that preparations as shown in panel C do contain significant amounts of SUMO conjugates. E, Coomassie-stained gel loaded with protein extract enriched by one step (His affinity, cf. panel C), as used for mass spectrometric protein identification. Lane 9, extract from tag-SUMO transgenic plants; lane 10, material from non-transgenic plants (background control). In panels A and D, prominent bands visualized by two distinct detection methods are linked by horizontal lines. Dot, major contaminant RUBISCO large subunit; bar to the right, position of tag-SUMO. Molecular weight marker positions are shown to the left.
Figure 5. Test of sumoylation enzymes by formation of SUMO-SUMO chains. SUMO activating enzymes SAE1a / SAE2 or SAE1b / SAE2 (labeling on top of gel pictures) were incubated with SCE (+ label), or without SCE (- label) together with tag3-SUMO1 (left panels), or with tag3-SUMO1 Q90A (right panel), in presence (+) or absence (-) of SUMO ligase SIZ1. Note that SUMO1 can form SUMO-SUMO linkages, in a pattern that is identical to that of SUMO1 Q90A. Bars to the right indicate position of di- tri- and tetra-SUMO (several bands, presumably due to conformation differences, or to different SUMO attachment sites).
Figure 6. In vitro sumoylation of protein tags. UBC27-Flag, a ubiquitin conjugating enzyme with a Flag tag extension, or glutathione S transferase encoded by vector pET42c (GST-S) can serve as a negative control. However, extension of GST-S by a Flag tag converts GST into a substrate (GST-S-Flag; dot to the right indicates position of the sumoylated form). Sumoylation is abolished by deletion of the S-peptide from the linker region between GST core and Flag peptide (panel GST-Flag). Control lanes labeled (-) contain an inactive (Cys 94 to Ser change) version of SCE, (+) indicates presence of WT SCE.
Figure 7. *In vitro* sumoylation of potential *in vivo* substrate proteins. Proteins identified as potential *in vivo* substrates were expressed in *E. coli*, purified and incubated with sumoylation enzymes. NAF, At2g19480; RRM, At3g56860; RXT3, At5g08450; TAF7, At1g55300. Control lanes labeled SCE (-) contain an inactive (Cys 94 to Ser change) version of SCE. SUMO1 and SUMO1 Q90A give identical patterns, while SUMO3 results in a single, considerably weaker sumoylation band of NAF or TAF7. Proteins NAF and RRM have a carboxyl-terminal Flag peptide for antibody-based detection, whereas RXT3 and TAF7 have an amino-terminal GST tag (without S peptide) and were detected by anti-GST antibody.