INVESTIGATIONS INTO THE ROLES OF SUMOYLATION IN *ARABIDOPSIS THALIANA*

By

Thérèse C. Rytz

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Genetics)

at the

UNIVERSITY OF WISCONSIN-MADISON

2017

Date of final oral examination: 5/22/2017

The dissertation is approved by the following members of the Final Oral Committee:

Dr. Richard D. Vierstra, Emeritus Professor, Genetics
Dr. Patrick Masson, Professor, Genetics
Dr. Philip Anderson, Professor, Genetics
Dr. Richard M. Amasino, Professor, Biochemistry
Dr. Shigeki Myamoto, Professor, Oncology

ABSTRACT

The covalent attachment of the Small Ubiquitin-like Modifier (SUMO) to other proteins is an essential post-translational modification required in all eukaryotes. This conjugation helps regulate development, cellular homeostasis, and responses to stress through the modification of transcriptional and translation regulators and chromatin modifiers. Particularly influential is the rapid SUMOylation of various nuclear proteins during stress presumably to alter nuclear activities needed for protection. In *Arabidopsis thaliana*, stress-induced SUMOylation has been connected to hundreds of proteins but the underpinning reasons remain unclear.

I investigated three separate aspects of the Arabidopsis SUMO conjugation pathway to further elucidate the roles of SUMOylation in plants. I conducted a phylogenetic study of plant SUMOs and found two types that are universal; a canonical, highly conserved form and a non-canonical form that appears to be lineage specific. These non-canonical SUMO shares little sequence identity among members, implying a role in plant biology that is independent of homology. Plants also have the capacity to assemble concatemers of SUMOs linked internally through lysine-mediated isopeptide linkages. I examined the role(s) of polySUMO chains by studying plants expressing a lysine-null SUMO blocked in chain assembly. Interestingly, the failure to generate SUMO chains has no discernible growth defect under normal and stress conditions, implying that these polymers are not essential. However, the use of lysine-null SUMOs should now facilitate mapping SUMOylation sites by coupling methods to enrich for target peptides bearing a SUMO moiety with identified mass spectrometric (MS) approaches to detect SUMO footprints.

Third, I developed a MS method to connect individual SUMO ligases with specific targets, in which the profile of conjugates affinity-purified from wild-type and ligase mutant

plants are compared by MS. Using this technique, I identified over a hundred proteins SUMOylated by the ligase SIZ1. These SIZ1-dependent substrates include major transcriptional regulators and chromatin modifiers associated with abiotic and biotic stress, thus connecting the role of SIZ1 in stress protection with a suite of affected processes. In summary, the work completed in this thesis provided insight into specific functions of SUMOylation and methods that will ultimately aid in the study of specific SUMO-dependent functions.

ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisor Dr. Richard D. Vierstra. Thank you. I would also like to give a special thanks to Dr. Patrick Masson. Additionally, I would like to thank my thesis committee members: Dr.'s Richard M. Amasino, Phil Anderson and Shigeki Myamoto.

A big thank you goes to all the many past and present members of the Vierstra. I need to especially thank Joseph Walker, Dr. David C. Gemperline, Dr. Richard Marshall, and all previous and current Team SUMO members.

And last, I would like to say thank you to my parents.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

aa	amino acid
ABA	abscisic acid
Arg or R	arginine
At	*Arabidopsis thaliana*
ATG	Autophagy
ATP	adenosine triphosphate
BLASTp	basic local alignment search tool for protein
BME	β-mercaptoethanol
cDNA	complementary deoxyribonucleic acid
CDS	coding sequence
Col-0	Columbia
Cr	crude
cv	column volume
Cys	cysteine
d	day(s)
Da	dalton
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DTT	dithiolthreitol
E1	activating enzyme
E2	conjugating enzyme
E3	ligase
EDTA	ethylenediaminetetraacetic acid
EST	expressed sequence tag
FT	flow through
GM	gamborg's B5 growth medium
His or H	histidine

H_2O_2	hydrogen peroxide
hr	hour
IAA	iodoacetamide
IPTG	isopropyl 1-thio-□-D-galactopyranoside
kDa	kilodalton
LD	long days
Lys or K	lysine
K0	lysine null
min	minute(s)
M	molar
mRNA	messenger ribonucleic acid
MS	mass spectrometry
MUB	membrane-anchored ubiquitin-fold
MW	molecular weight
NaCl	sodium chloride
Ni-NTA	nickel-nitrilotriacetic acid
nm	nanometer
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PMSF	phenylmethylsulfonyl fluoride
PSM	peptide spectral match
PVDF	polyvinylidene fluoride
RING	Really Interesting New Gene
RNA	ribonucleic acid
RT-PCR	reverse transcription-polymerase chain reaction
RUB	Related to Ubiquitin
SA	salicylic acid
SDS	sodium dodecyl sulfate

STUbL	SUMO targeted ubiquitin ligase
SUMO	Small Ub-Like Modifier
SIM	SUMO interacting motif
TAIR	The Arabidopsis Information Resource
TAP	tandem affinity purification
TBS	tris-buffered saline
T-DNA	transfer DNA
TFA	trifluoroacetic acid
Ub	Ubiquitin
Ubl	Ubiquitin-like
UTR	untranslated region
WT	wild type

CHAPTER 1:

INTRODUCTION TO THE SUMO CONJUGATION SYSTEM

Parts from the section on SUMO proteomic studies have been published in Rytz TC, Miller MJ, Vierstra RD (2016). Purification of SUMO Conjugates from Arabidopsis for Mass Spectrometry Analysis. *Methods Mol Biol* 1475:257–81.

Modification of Proteins by SUMO

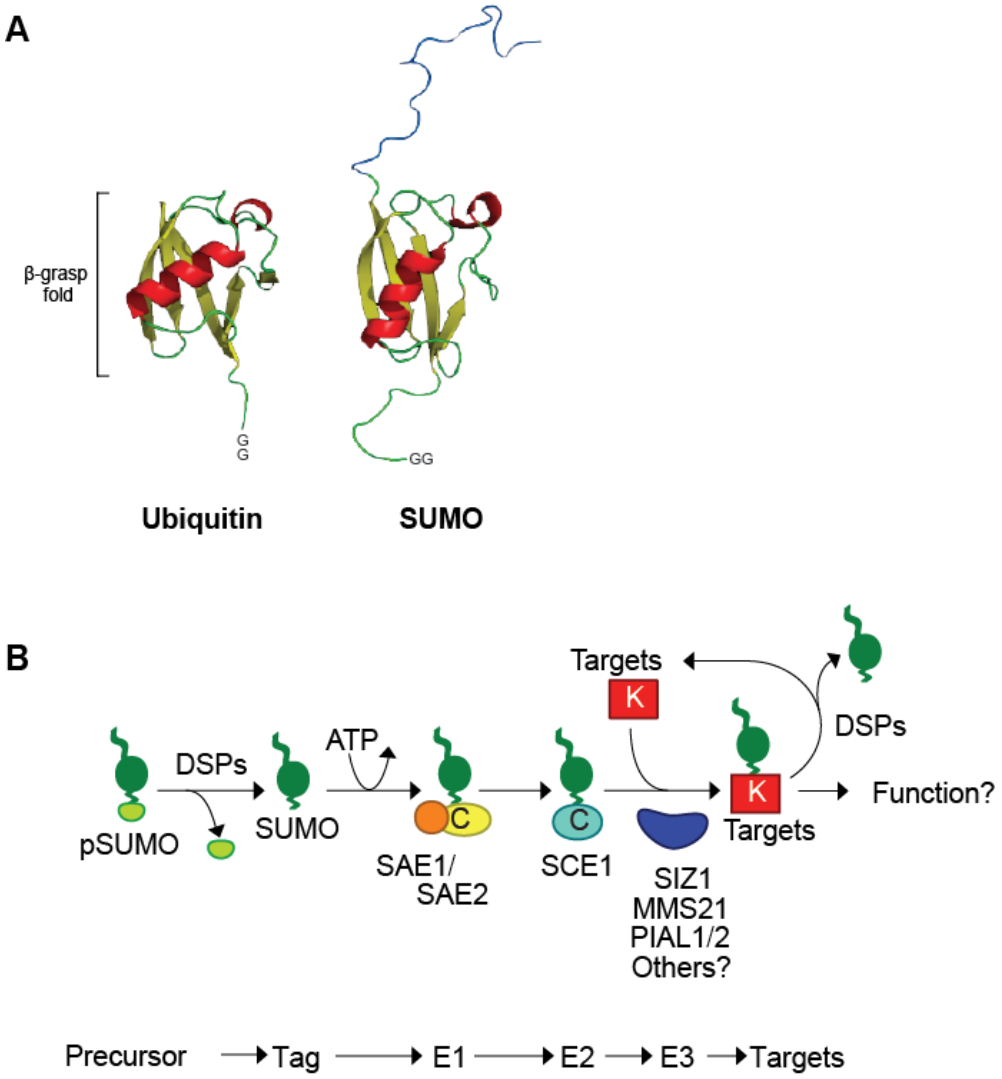
Post-translational modifications of proteins provide an additional level of regulation beyond transcriptional and translational controls by altering the structure, activity, localization and/or half-life of the modified proteins. Hundreds of different types of post-translational modifications have been identified, including the addition of acetyl, methyl and phosphate moieties, sugars, fatty acids, nucleosides as well as other proteins. One major group of proteins that covalently modify other proteins encompasses the ubiquitin-like modifier (Ubl) family. The Ubl proteins share a characteristic beta-grasp fold, consisting of a beta sheet surrounding an alpha-helix. The first member identified was ubiquitin (Ub), but the family has now expanded to include multiple different modifiers such as ATG12, RUB1/Nedd8, MUB1, URM1 and SUMO (for Small Ubiquitin-like Modifier) [1].

SUMO is a conserved ~90 amino acid protein consisting of a 10-20-amino-acid-long flexible N-terminal extension preceding the beta-grasp fold with a di-glycine (di-Gly) motif at the C-terminal end. Although the 3-D structure of SUMO resembles that of Ub (Figure 1-1A), amino acid sequences of SUMO and Ub are only ~20% identical, giving SUMO a differently charged surface. Like Ub, SUMO is conjugated to lysines of other proteins through a thioester bond involving the C-terminal carboxyl group of the modifier and the ϵ-amino group of the target lysine. This modification of proteins with SUMO involves a three-step enzymatic cascade similar to the conjugation pathway of Ub, consisting of a heterodimer SUMO activating enzyme (SAE1 and SAE2) (or E1), a SUMO conjugating enzyme (SCE1) (or E2:) and a SUMO ligase (or E3). SUMO conjugation is a dynamic process, as a collection of SUMO-specific cysteine proteases, or deSUMOylating proteases (DSPs), also exist that will cleave the SUMO moiety from the target lysine to reverse SUMOylation [2–4] (Figure 1-1B).

The SUMO polypeptide is translated as an immature pro-form and requires processing by DSPs to remove the C-terminal propeptide and reveal the di-Gly involved in SUMO conjugation [5–7]. SUMO is then activated by the heterodimeric E1 enzyme through the ATP-dependent formation of a SUMO-adenylate intermediate, which leads to the formation of a thioester bond between the C-terminal carboxyl group of SUMO and the catalytic cysteine residue of the SAE2 subunit and to the release of an AMP moiety [8, 9]. Subsequently, the SUMO moiety is transferred via transesterification to the cysteine residue on the E2 SCE1 [10–12]. Finally, SUMO is transferred from the E2 to target lysines. Both E1 subunits, SAE1 and SAE2, contain the ATP-binding ThiF domain. Besides the active site cysteine, the larger subunit SAE2 also contains an Ubl domain, which is required for E1-E2 interaction [13]. The active cysteine of the E2 is located in a pocket of the enzyme surrounded by residues that facilitate the recognition of target lysines on proteins [14, 15]. In addition to the E1 interacting site, SCE1 also contains a "backside" binding site that allows for the interaction with SUMO moieties [14, 16].

While the E1 and E2 are sufficient for SUMO conjugation *in vivo* [12], the transfer of SUMO from the E2 is often facilitated by SUMO E3 ligases. These ligases do not receive the SUMO moiety through a covalent bond, but provide structural support for the reaction by bringing the SUMO-E2 enzyme complex together with the target protein [2–4]. The main group of SUMO E3 ligases is characterized by an SP-RING (SIZ/PIAS-Really Interesting New Gene) domain, which resembles the RING domain in Ub RING E3 ligase, and interacts with the E2 enzyme. A second type of SUMO ligases exists which do not have the SP-RING domain but still support the SUMOylation of specific targets through interaction with both the E2 and the substrate. These include the nucleoporin RanBP2/Nup358 and the poly- comb protein Pc2 [17, 18].

Figure 1-1. SUMO is covalently attached to lysines on target proteins through a three-step enzymatic cascade. **(A)** The 3D ribbon diagrams of SUMO and ubiquitin. **(B)** Diagram illustrating the Arabidopsis SUMO conjugation pathway members.


A
β-grasp fold
G
G
GG
Ubiquitin
SUMO
B
pSUMO
DSPs
SUMO
ATP
SAE1/
SAE2
C
SCE1
C
Targets
K
SIZ1
MMS21
PIAL1/2
Others?
K
Targets
DSPs
Function?
Precursor
Tag
E1
E2
E3
Targets

The SUMO modification of proteins can be reverse through deconjugation by DSPs. Some proteases are able to both cleave the SUMO propeptide through their peptidase activity and remove SUMO from proteins by isopeptidase cleavage, like *Saccharomyces cerevisiae* Ulp1, while other forms only have isopeptidase activity, such as *S. cervisiae* Ulp2 [7, 19, 20].

The SUMO conjugation pathway is associated with multiple biological processes and is essential for cell viability in animals, yeast and plants [21–23]. SUMOylation is very dynamic and often occurs in response to cellular signals, such as cell-cycle progression, chromosome segregation and DNA damage [24]. Additionally, SUMO is rapidly conjugated to target proteins upon cellular stress, including heat shock, ischemia and osmotic shock [21, 25–27]. In this chapter, I will provide an overview of the SUMO conjugation pathway in *Arabidopsis thaliana*, how targets are selected for conjugation, the subsequent effect of SUMOylation, and the proteomic studies completed to identify SUMO substrates.

The Arabidopsis SUMO Family

In *A. thaliana*, eight genes encode for SUMO [28–30]: *SUMO1* through *SUMO7* and *SUMO9*. Of these eight loci, only *SUMO1*, *SUMO2*, *SUMO3* and *SUMO5* are transcribed [28, 31, 32]. There is no evidence that *SUMO4*, *SUMO6* and *SUMO7* are expressed, nor do they contain the di-Gly motif required for conjugation. *SUMO9* represents a truncated pseudogene [21, 28].

The amino acid sequences of *At*SUMO1 and *At*SUMO2 are 93% identical to each other with only six residues differing between the two isoforms. *At*SUMO1 and *At*SUMO2 are redundant to each other as single null mutants are viable while the *sumo1-1 sumo2-1* double mutant arrests in early embryonic development [21]. Furthermore, *At*SUMO1 and *At*SUMO2

(also referred to as *At*SUMO1/2) are the major SUMO isoforms conjugated to proteins and are rapidly attached to targets upon stress conditions, such as heat-shock or oxidative stress [21, 28, 33].

While *At*SUMO1/2 and their substrates have been a major focus of study, less is known about the roles of *At*SUMO3 and *At*SUMO5. These distinct SUMO isoform share only a 48% and 35% protein identity with *At*SUMO1, respectively, Additionally, *At*SUMO3 and *At*SUMO5 have less than 50% of their amino acids sequences in common with each other. The loss of *At*SUMO3 has little to no effect on plant growth [34]. While artificial over-expression in plants and reconstituted SUMOylation assays have shown that *At*SUMO3 can be conjugated to proteins, no physiologically relevant targets have yet been identified [35–37]. Immunoblot analyses using antibodies to *At*SUMO3 identified low levels of free SUMO3, and failed to detect SUMO3-specific conjugates [28, 34]. No mutant alleles for *At*SUMO5 have been described, and this isoform has not been studies *in vivo*. Interestingly, *in vitro* analysis of Arabidopsis DSPs revealed that SUMO3 and SUMO5 are not readily processed by the plant DSPs [36]. Furthermore, a study of the biochemical properties of the SUMO isoforms found that the SUMO E1 and E2 prefer SUMO1/2 for conjugation, as residues required for interaction with the enzymes have been substituted in SUMO3 and SUMO5 [38]. These observations suggest that *At*SUMO3 and *At*SUMO5 might not be the preferred substrates of the SUMO conjugation pathway; thus, if they do have a role in plant biology it could be quite distinct from those related to SUMO1/2.

While the yeasts *S. cerevisiae* and *Schizosaccharomyces pombe* express only one form of SUMO, mammals express a family of four SUMO isoforms. Mammalian SUMO1, SUMO2 and SUMO3 are ubiquitously expressed whereas SUMO4 is detected in just a few tissues [39]. Like

SUMO1/2 in Arabidopsis, mammalian SUMO2 and SUMO3 (SUMO2/3) are closely related to each other and are essential for viability, while mammalian SUMO1 shares only about 50% protein identity with SUMO2/3 and is not required for cell survival as SUMO2/3 can complete a SUMO1 null mutant [40, 41]. However, mammalian SUMO1 and SUMO2/3 do have distinct cellular functions. While both SUMO1 and SUMO2/3 are conjugated to target proteins, SUMO1 modifies a separate set of proteins [2, 3, 42, 43]. As an example, the RanGTPase-activating protein (RanGAP1) is selectively modified by SUMO1 [40, 44]. Additionally, upon different stresses, such as heat shock and ischemia, SUMO2/3 are rapidly conjugated to target proteins [25, 40, 45]. Therefore, as the mammalian SUMO isoforms have distinct roles, the question remains whether *At*SUMO3 and *At*SUMO5 also serve separate functions from *At*SUMO1/2.

SUMO, like ubiquitin, has the ability to form polySUMO chains through internal linkages. However, the SUMO isoforms vary in their ability to form chains. In Arabidopsis SUMO1/2 readily form chains *in vitro* and polySUMO1 chains have been detected by mass spectrometric analyses of isolated SUMO conjugates [33, 36, 46]. Arabidopsis SUMO3, on the other hand, does not form polySUMO chains *in vivo* [36, 37]. Likewise in mammals, SUMO2/3 are ligated with other SUMO2/3 moieties to form polySUMO chains; however, SUMO1 is unable to be modified by other SUMOs and is not found as chains *in vivo* [47, 48]. PolySUMO chain formation appears to be an integral part of the SUMO conjugation pathway as chains have also been detected in the yeasts [49, 50].

The SUMO Conjugation Pathway in Arabidopsis

The SUMO conjugation pathway is conserved in plants, yeasts and animals [28, 51, 52]. In Arabidopsis, the heterodimeric E1 activating enzyme is encoded by three genes, *SAE1a* and

SAE1b and *SAE2*. SAE1a and SAE1b share an 81% protein sequence identity and are likely redundant to each other [21, 28]. For the E2 conjugating enzyme SCE, two genes are present in the Arabidopsis genome, however only one loci, *SCE1a*, (commonly referred to as *SCE1*) is expressed and *SCE1b* represents a pseudogene due to its lack of expression and truncated coding sequence [28, 31]. To date, four SUMO E3 ligases have been identified in Arabidopsis, SAP and MIZ1 ligase (SIZ)-1 [53, 54], METHYL METHANESULFONATE-SENSITIVE (MMS)-21 (or HIGH PLOIDY (HPY)-2) [55, 56], and PIAS Protein Inhibitors of Activated STATs (PIAL)-1 and PIAL-2 [32]. SIZ1 is the main E3 ligase in plants and is responsible for the majority of protein SUMOylation [21, 53, 57] (Figure 1-2B).

Considering that only one SUMO conjugating enzyme and four SUMO ligases are known to exist in Arabidopsis, the SUMO conjugation system appears to have a strikingly low numbers of E2 and E3 enzymes compared to the ubiquitylation pathway, which has expanded to include over 40 E2 conjugating enzymes and potentially more than 1400 diverse E3 ligases [58–60]. However, the family of SUMO deconjugating enzymes has diversified in Arabidopsis and consists of more than ten distinct SUMO-specific protease, including EARLY IN SHORT DAYS 4 (ESD4), UBIQUITIN-LIKE PROTEASE1a (ULP1b)/ESD4 LIKE SUMO PROTEASE (ELS1), ULP1b, ULP1c/OVERLY TOLERANT TO SALT 2 (OTS2), ULP1d/OTS1, ULP2a, and ULP2b [28, 31, 36, 51, 61–63].

Analysis of SUMO pathway mutants in Arabidopsis revealed that SUMOylation is required for viability. Mutants in *SAE2* (*sae2-1*) and *SCE1* (*sce1-5* and *sce1-6*) and the double mutant *sumo1-1 sumo2-1* are not viable past early embryonic development [21]. However, SUMO E3 ligase mutants are viable, but show acute developmental phenotypes. These

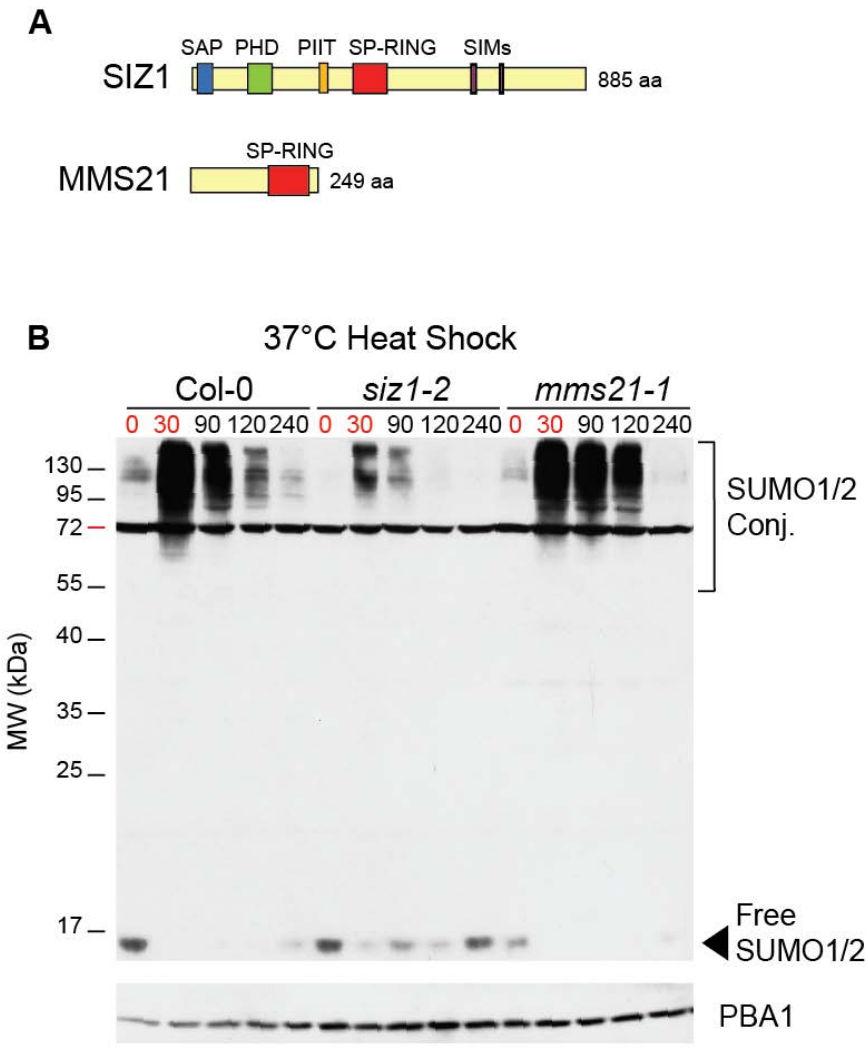
phenotypes differ between the specific ligases, indicating that they modify distinct collections of proteins.

Mutants in SIZ1 (*siz1-2* and *siz1-3*) display a severe dwarf-like phenotype, including curled leaves and decreased cell volumes, and are characterized by highly elevated levels of salicylic acid (SA) [53, 64, 65]. Reduction of SA in *siz1*-2 using an over expression of the bacterial salicylate hydroxylase nahG, rescued some of the E3 ligase mutant phenotypes, such as curled leaves and small cell volume, but these plants were still shortened in stature compared to wild-type [64, 65]. Additionally, SIZ1 mutants are hypersensitive to low phosphate levels, are early flowering, have reduced basal thermotolerance and decreased tolerance to cold [53, 66–68].

The MMS21 mutants (*hpy2-1* and *hpy2-2/mms21-1*), on the other hand, have severely reduced root growth, aberrant root meristems, smaller rosettes with elongated leaves, and display stem fasciation [55, 56]. Most of these phenotypes can be attributed to misregulation of the cell cycle and increased endoreduplication. Additionally, *mms21-1* seedlings are sensitive to DNA damaging agents [69]. The *siz1-2 mms21-1* double mutant is embryonic lethal, suggesting that the combined SUMOylation by SIZ1 and MMS21 is crucial for development. Mutants in PIAL1 and PIAL2 (*pial1-1* and *pial2-1*) do not have apparent phenotypes besides a slightly increased tolerance to salt [32].

Interestingly, many of the phenotypes seen for the E3 ligase mutants have also been observed for SUMO protease mutants. For example, the protease ESD4 is also extremely dwarfed, has increased SA levels, and is early flowering [61]. SUMO mutants containing very low levels of SUMO (*35S promoter:amiR-SUMO2 sumo1-1*) also show an increased level of SA [34]. These results suggests that the balance of free SUMO and SUMO conjugates is tightly

Figure 1-2. The essential Arabidopsis SUMO isoforms (SUMO1 and SUMO2) become rapidly conjugated to proteins in response to heat stress, with SIZ1 directing much of the SUMOylation. (A) Organization of the SIZ1 and MMS21 genes and proteins. The SAP, PHD, PIIT, SP-RING, and SIM domains are highlighted in blue, green, brown, red, and purple, respectively. **(B)** SUMO1/2 are rapidly conjugated to proteins in response to heat stress. This increase in SUMO conjugates upon heat stress is greatly reduced in the SIZ1 mutant *siz1-2*, but not in the MMS21 (HPY2) mutant *mms21-1*. Col-0, wild-type. Plate-grown plants were heat stressed at 37°C for 30 min and returned to room temperature (RT). Tissue samples were collected at the indicated times and subjected to SDS/PAGE and immunoblot analysis with anti-SUMO1 antibodies. An immunoblot analysis with anti-PBA1 antibodies was used to confirm near equal loading.


A
SAP PHD PIIT SP-RING SIMs
SIZ1
885 aa
SP-RING
MMS21
249 aa
B
37°C Heat Shock
Col-0
siz1-2
mms21-1
0 30 90 120 240 0 30 90 120 240 0 30 90 120 240
130
95
72
55
40
35
25
17
MW (kDa)
SUMO1/2
Conj.
Free
SUMO1/2
PBA1

controlled in Arabidopsis and any deviance from this homeostasis induces an increase in SA through unknown mechanisms.

E3 Ligases, Target Selection and Effects of SUMOylation

The four SUMO E3 ligases identified in Arabidopsis are SP-RING type ligases. The main ligase, SIZ1, is part of the SIZ/PIAS family of E3 ligases, which are characterized by the presence of a nuclear localization signal (NLS), a Scaffold Attachment Factor-A/B/Acinus-PIAS (SAP) domain, a PINIT motif, a SP-RING domain, and SUMO-binding sites (SUMO interacting motifs or SIMs). Homologs of the SIZ/PIAS ligases have been identified in animals, yeasts and plants. Besides the main SIZ1/PIAS domains, plants SIZ1s also contain a plant homeodomain (PHD) and the PINIT motif is modified to a PIIT sequence (Figure 2-1A).

The various domains on SIZ1 control the localization of the ligase and its interactions with the E2 and substrates in order to direct the SUMOylation of distinct sets of proteins. The SAP domain has DNA-binding properties and was found to not be required for SUMOylation in plants and yeast; however, this domain is responsible for proper nuclear localization [54, 57, 70, 71]. The PHD domain is a zinc (Zn)-finger motif present in many nuclear-localized proteins and is involved in protein-protein and protein-chromatin interactions, including histones [72, 73]. Mutations in the PHD lead to a decrease in SUMOylation, implying a crucial role for this domain in target identification [54]. For example, the PHD domain is required for SUMOylation of the bromodomain-containing GLOBAL TRANSCRIPTION FACTOR GROUP E3 (GTE3) [74]. Additionally, analysis of rice SIZ1 revealed that the PHD interacts specifically with methylated and trimethylated residues of Histone 3 [75]. The PINIT and SP-RING domains are required for proper interaction with the E2 conjugation enzyme and thus are essential to the

ligase activity of SIZ1 [54, 57, 71, 76]. *At*SIZ1 contains two SIM domains and is thus able to interact with SUMO-modified proteins. While the exact roles of these interacting motifs in the SUMOylation of targets are unclear, loss of a SIM in *At*SIZ1 resulted in increased levels of SA [54, 71, 77]. Interestingly, the mammalian SIZ1/PIAS proteins have repressive activities independent of SUMO ligation by means of the SAP domain [77, 78]; however; no ligase-independent roles have been assigned to SIZ1 in plants.

The SP-RING ligase MMS21 is distinct from SIZ1 and has none of the additional regulatory domains found in the SIZ/PIAS family of E3 ligases [79] (Figure 2-1A). While no specific domains, besides the SP-RING, have been characterized in MMS21, the ligase interacts with the evolutionarily-conserved Structural Maintenance (Smc) complex (also referred to as the Smc5-Smc6 complex) [80–83], which is required for sister chromatid cohesion, mitotic chromosome condensation and recombinational DNA repair [84]. The association of MMS21 with the Smc complex is likely critical for DNA repair, as mutants in the E3 ligase are sensitive to DNA damage [69, 81, 82]. Additionally, in *S. pombe*, SUMOylation is induced by DNA damaging agents in a MMS21-dependent manner [80]. It is unclear if MMS21 acts independently as a SUMO ligase or if the Smc5-Smc6 complex provides assistance in target selection, but initial experiments suggest the complex supports the ligase activity of MMS21 during DNA damage [85, 86].

PIAL1 and PIAL2 are two additional “plant-specific” SP-RING domain containing SUMO E3 ligases with a high protein sequence similarity to each other, but are not related to MMS21 or SIZ1 [32]. Besides the SP-RING, PIAL1 and PIAL2 have two SIMs in the C-terminal region, of which one was found to be involved in SUMO chain formation *in vitro* [32].

In mammals, multiple proteins without the SP-RING domain promote SUMO ligation through various protein interaction domains that bring the target and the conjugation enzyme together, such as the nucleoporin RanBP2/Nup358 and the polycomb protein Pc2 [17, 18]. These ligases are difficult to identify through screens as no consensus E2 interacting domain, like the SP-RING, are present. Additionally, only a few, very specific proteins are targeted by these ligases for SUMOylation. In plants, no SP-RING-less SUMO ligases have been identified; however, this does not rule out their presence. Only 4 ligases have been identified, but hundreds of proteins are SUMOylated, suggesting the presence of additional E3s.

While E3 ligases are thought to direct most of the SUMO addition, the E2 enzyme has been observed to modify protein directly without a ligase. SCE recognizes target lysines through a consensus SUMOylation motif, which consists of a hydrophobic residue preceding the lysine residue followed by an acidic residue two positions downstream of the lysines (ψKxD/E). In mammals, the E2 can directly recognize this sequence and conjugate the target lysine without the assistance of an E3 ligase [87, 88]. The consensus SUMOylation motif can also include stretches of charged residues or phosphorylation sites downstream of the modified lysine, which increase the positive charge of the motif and enhance the non-covalent interaction with the E2 enzyme [15, 88, 89]. In plants, reconstituted E1 and E2 can SUMOylated specific proteins without an E3 ligase, suggesting the E2 is also able recognize SUMOylation motifs [30, 35, 36]. Nevertheless, only about 60% of SUMOylated proteins identified in mammals actually have a consensus SUMOylation motif, indicating that additional components, such as E3 ligases, are also important for the recognition of targets [89].

While the ubiquitylation of a protein generally leads to protein degradation, the outcome of SUMOylation is specific to the target and site of ligation. The addition of SUMO can lead to a

change in a protein's stability, localization, and/or activity. These effects are mainly driven on a molecular level by the masking of existing binding motifs or establishment of new interaction sites [3, 90, 91]. Proteins with SUMO-interacting motifs (SIMs) bind non-covalently to the SUMO moiety on targets, establishing new protein interactions. SIMs are characterized by a core of hydrophobic amino acids flanked by charged residues and interact with the hydrophobic residues in the groove formed by the alpha-helix and beta-sheet of SUMO [92–94]. Variation of the residues of the SIM can lead to binding preferences to specific SUMO isoforms [95]. This selective binding of specific interacting proteins based on sequence differences in the SUMO isoforms is thought to be the main reason for the functional difference observed between the mammalian SUMO isoforms [96, 97]. SIMs have mostly been identified and studied in yeasts and mammals, however a yeast-two hybrid (Y2H) screen in Arabidopsis has identified multiple SUMO interacting proteins, as well as characterized the SIMs in plant proteins [98]. SIMs appear to be conserved between animals, yeast and plants, however an additional type of SIM was found to be present in plants, compromised of a core of hydrophobic residues flanked on both sides with acidic amino acids [98].

Interestingly, SUMO chains present unique binding sites for specific SUMO-interacting proteins. Multiple SUMO-chain binding proteins have been identified in yeasts and animals, including ZIP1, part of the synaptonemal complex involved in meiosis, CENP-E (centromere-associated protein E), required for correct chromosome alignment during mitosis, the ubiquitin E3 ligase heterodimer Slx5–Slx8 *in Saccharomyces cerevisiae* (Rfp1/Rfp2–Slx8 in *Schizosaccharomyces pombe*) and mammalian Ubiquitin E3 ligase RNF4 (RING finger protein 4) which has a much higher binding affinity for SUMO chains than for mono or di-SUMO [99–103]. These proteins are often characterized by containing multiple SIMs in tandem. Although

SUMO-interacting proteins have been identified in Arabidopsis [98], it is unknown if proteins with a preference for SUMO chains exist in plants.

Phosphorylation and acetylation can influence both SUMOylation and the binding affinity of SIMs to SUMO. As these post-translational modifications are transient, they add another level of regulation to target selection and the downstream effects of SUMO addition. In mammals, the phosphorylation of the consensus SUMOylation motif can direct conjugation to that site by selectively enhancing the interaction of the E2 [4, 89, 96]. Interestingly, it has been suggested that the residues on the E2 required for the interaction with this negative charge are not conserved in yeasts, indicating that SUMO substrate selection potentially occurs through differing mechanisms [4, 15]. Additionally, phosphorylation of SIMs can enhance the preference of the binding site to one SUMO isoform over the others [104–106]. While phosphorylation enhances SUMOylation, acetylation is thought to prevent SUMOylation by either modifying the lysine of SUMO consensus motif or the surrounding residues [4, 96, 107]. Acetylation can also prevent SUMO-SIM interactions. For example, the modification of *Hs*SUMO1 with an acetyl molecule prevents it from being bound by SIMs [108]. These modifications are referred to as “acetyl switches” due to their antagonistic role to SUMOylation. It is yet to be determined if acetylation or phosphorylation has an effect on SUMO modification in yeasts or plants.

Ultimately, target selection and the effects of SUMOylation are regulated through various protein-protein interactions, which can be influences by other post-translational modifications. This leads to the dynamic modification of diverse sets of proteins under various cellular conditions; in the end affecting numerous processes.

SUMOylation and the Abiotic Stress Response

One of the more striking features of the SUMO conjugation pathway is the rapid addition of SUMO to proteins in response to abiotic stressors. In plants, the level of SUMO conjugates can be dramatically increased by heat, cold, and drought as well as by treatments with ethanol, hydrogen peroxide and the amino acid analog canavanine [28, 68, 109]. This response to cellular challenges is conserved among eukaryotes. In yeasts and animals, heat, ischemia and other oxidative stresses, osmotic stress, DNA damage and proteosome inhibitors increase global SUMOylation [25, 27, 40, 80, 110]. This universal mechanism suggests that SUMOylation may have a significant role in the response to stress conditions.

In Arabidopsis, heat stress causes SUMO conjugates to accumulate within minutes and remain present for the entirety of the stress before returning to basal levels after the removal of the stimuli [28] (Figure 1-2B). After recovery, a refractory period occurs in which a second heat shock fails to increase SUMO conjugation [21]. This dramatic increase in conjugates upon stress, as well as analyses of SUMO pathway mutants, suggest that SUMOylation is required for stress tolerance.

Additionally, SIZ1 has been implicated in the survival to multiple stress conditions, including thermotolerance to heat and cold, drought tolerance, and nutrient accumulation, often through the modification of specific regulatory factors [53, 66, 68, 109, 111–114]. For example, SIZ1-mediated SUMOylation of ICE1 stabilizes the transcription factor, leading to expression of DREB1A and increased cold tolerance [68]. Furthermore, SUMO addition is involved in the regulation of ABA signaling. Besides regulating seed germination and growth, ABA can mediate plant responses to environmental stresses, such as drought [115–117]. *siz1-2* mutants are hypersensitive to ABA, potentially due to a loss of SUMOylation of ABI5 and MYB30,

whereas seedlings over-expressing SUMO are hyposensitive to ABA [109, 118–120]. Since SIZ1 has been implicated in such diverse processes, further identification of its targets will be necessary for a better understanding of the mechanisms of stress protection provided by SUMOylation.

While the exact functions of stress-induced SUMOylation are unclear, identification of SUMO targets from unstressed and heat-shocked Arabidopsis seedlings revealed that most substrates are nuclear proteins, including transcription factors, chromatin remodelers, RNA splice factors and proteins involved in RNA-directed gene regulation [21, 33, 46]. Similar pools of SUMO conjugates were identified in yeast and animals upon heat shock, highlighting the conserved nature of SUMOylation upon stress [45, 89, 107, 121, 122]. In animals and yeast, SUMOylation occurs at the promoters of active genes during heat stress where it alters the transcription of stress induces gene to promote cell survival [123–126]. Thus, SUMOylation provides stress protection by modifying a large array of key nuclear regulators to alter transcription of stress-induced genes.

Mass Spectrometric Identification of SUMO Conjugates

Proteomic analyses of conjugates by mass spectrometry (MS) have yielded great insight into the cellular processes impacted by SUMOylation. The basic strategy is to enrich for the conjugates based on the SUMO moiety by expression of a tagged SUMO, either stably or transiently, followed by affinity purification of the SUMO and its conjugates based on the tag sequence. Examples include the use of six-histidine (6His), HA, and Flag tags followed by affinity enrichment with nickel chelate or anti-tag antibody columns [33, 37, 42, 45, 46, 122, 127–131]. Extractions often use buffers containing strong denaturants to avoid isolating

proteins non-covalently associated with SUMO, and irreversible cysteine protease inhibitors such as IAA to help block DSPs.

First attempts at purifying SUMO conjugates employed a single affinity purification step using human cell culture lines transiently expressing a tandem 6His-tagged SUMO construction combined with nickel chelate chromatography [42, 127]. The tagged SUMO was expressed in a background of wild-type SUMO, which challenged efficient enrichment. An improved strategy was developed in *S. cerevisiae* by replacing the endogenous SUMO with a 6His-tagged version such that the entire pool of SUMO was tagged [128]. To increase the purity of the preparations and thus provide more confidence of SUMO conjugate identification, tandem affinity methods were subsequently adopted in which two different affinity steps are employed sequentially. Examples include using modified SUMOs bearing two different tag sequences, such as 6His combined with a FLAG epitope, or a tandem affinity purification (TAP) tag consisting of a protein A domain followed by a calmodulin-binding protein [45, 122]. These arrangements allowed for nickel-nitrilotriacetic acid (Ni-NTA) chromatography followed by immunoprecipitation with anti-FLAG antibodies or pulldown with non-specific immunoglobulins and calmodulin, respectively. It is also possible to combine Ni-NTA columns with anti-SUMO antibody columns if a single 6His tag is employed [33]. Vertagaal and colleagues have recently exploited a poly-His tag containing ten tandem histidines that bind tighter to Ni-NTA beads and thus can be washed under stronger denaturing conditions to reduce contaminants [107, 132–134]. These differing purification strategies have yielded large databases of SUMO targets from not only unstressed samples, but also samples subjected to various stresses, including heat shock, and revealed the post-translational modifier's ubiquitous role in conjugating transcription factors, chromatin regulators and RNA-related processes.

Additional studies have employed more targeted approaches to explore subsets of SUMO conjugates. For example, Bruderer et al (2011) made use of tandem SIMs to selectively identify substrates modified by polySUMO chains. Furthermore, a novel purification strategy employing a His-tagged lysine-null SUMO variant resistant to the peptidase activity of LysC allowed for increased detection of SUMO modified lysines [89, 107]. Ascertaining the specific lysine bound by SUMO can assist in studying the role of modification. Ub leaves a di-Gly remainder, or "footprint", on lysine after trypsinization, which can be detected by MS analyses, however, the native remainders of *At*SUMO1/2 or *Hs*SUMOs on lysines after digest results in a peptide too large to be measure by MS. *S. cerevisiae* SUMO does leaves a native di-Gly remainder on the target lysine when cleaved by trypsin due to an arginine upstream of the glycine [135]. Therefore, artificial SUMO footprint mutations have been developed by replacing the residue one to three amino acids upstream of the C-terminal di-Gly motif with an arginine, which leave a 2 to 5 amino acid residue on the modified lysine after tryptic digest [33, 37, 89]. Combining these targeted purifications with quantitative proteomics allows for the study of the dynamics of SUMO modification upon stress treatments in a site-specific manner.

A number of proteomic studies have attempted to provide catalogs of SUMO conjugates from plants, mainly using *A. thaliana* as the model. One of the main difficulties with isolating such conjugates is their low abundance under normal growth conditions with most of the SUMO pool present in a free form. First attempts combined the endogenous expression of a 6His-tagged SUMO1 with a single Ni-NTA chromatography step [37, 136]. To enrich for SUMO conjugates in vivo, Budhiraja et al (2009) also employed a SUMO1 variant with a glutamine-to-alanine mutation at residue 90 that resulted in conjugates that were less readily disassembled by DSPs and thus more stable in extracts. Unfortunately, expression of this SUMO1(Q90-A)

protein negatively affected the phenotype of the plants and resulted in the detection of few potential conjugates, leaving any list developed with this variant open to question. To improve conjugate isolation, Park et al (2011) performed two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) of the Ni-NTA enriched fraction from 6His-SUMO1-expressing plants coupled with matrix-assisted laser-desorption ionization time-of-flight mass spectrometry. This method however is severely limited by the ability of 2D-PAGE to sufficiently separate individual proteins and by its failure to detect low abundance conjugates (which is typically the case) or conjugates that fall outside the optimum range for PAGE analysis (e. g. , too acidic, basic, and/or hydrophobic or very large or small). An alternative approach was to identify SUMO targets by combining two-dimensional liquid chromatography fractionation of the whole plant proteome with an immunoscreen of the resulting fractions for those containing SUMOylated proteins, and then identify the abundant proteins in each fraction by mass spectrometry [137]. Unfortunately, this strategy failed to provide confidence that the proteins in question were directly modified with SUMO as opposed to being more abundant contaminants that co-fractionated with actual targets. Collectively, the low number of targets identified, the lack of overlap among these datasets, and the presence of proteins that probably are not bona fide targets, provided little confidence that the resulting catalogs are not contaminants in these previous studies.

To avoid the above complications, Miller et al (2010) developed a three-step purification strategy for plants that relies on a tagged SUMO that faithfully mimics the wild-type protein, and combines two separate affinity methods to isolate SUMO conjugates based on the SUMO moiety. An initial Ni-NTA column was used to isolate SUMO conjugates, followed by an anti-SUMO immunoprecipitation step to eliminate a majority of contaminants with a subsequent Ni-

NTA to remove the anti-bodies which might have bled from the anti-SUMO affinity column (described in detail by Ref. [138]). By employing strong denaturants during the various purification steps, it was possible to prevent deSUMOylation by proteases, minimize contaminants and avoid isolation of proteins interacting non-covalently with SUMO. Through this stringent purification strategy, over 350 high-confidence SUMO conjugates were identified from unstressed and heat, ethanol or hydrogen peroxide (H2O2) treated seedlings. Additionally, SUMO footprints were identified on 14 proteins, including SUMO1 and SIZ1 [33]. By combining this stringent purification protocol with isobaric tag for relative and absolute quantification (ITRAQ) labeling, the dynamics of SUMO conjugation under a variety of stresses could be measured [46]. This lead to the observation that while most conjugates increase in abundance upon stress, the targets with the most increase in abundance participate in RNA processing and turnover as well as RNA-directed DNA modification, implicating SUMOylation in transcriptome regulation [46]. While these proteomic studies in plants have identified a large set of SUMO conjugates, the list of known targets in animals contains over 1000 proteins [135]. Thus, further proteomic studies using some of the novel methods described above, such as using a 10His-tagged lysine-null SUMO variant could help lead to additional identification of targets and footprints in Arabidopsis.

Conclusions

The post-translational modification of proteins with SUMO is an essential and dynamic process involved in a variety of cellular processes in yeasts, animals and plants. Molecularly, SUMOylation alters the binding properties of targets by providing new binding sites or inhibiting interaction motifs, which ultimately leads to altered protein activity, localization or stability.

However, the exact outcome of SUMOylation is target-specific and can be influenced by phosphorylation, acetylation and ubiquitylation.

One intriguing aspect of SUMOylation is the dramatic increase in conjugates upon various abiotic stresses; an occurrence conserved throughout eukaryotes. The stress-responsiveness of SUMO conjugation, as well as the phenotypic analysis of SUMO pathway mutants, like E3 ligases, imply a crucial role for SUMOylation in providing tolerance to numerous adverse conditions. While the exact role of SUMOylation in stress protection is unknown, proteomic studies of SUMO conjugates highlight an involvement of SUMOylation in transcriptional regulation upon stress through the modification of transcription factors and chromatin modifiers as well as factors involved in RNA processing and RNA-directed DNA modification. Recent anti-SUMO chromatin immunoprecipitations followed by DNA sequencing (ChIP-Seq) experiments and transcriptome analyses indicate that SUMOylation prevents the over-expression of stress-responsive genes [113, 139]. Thus, SUMOylation does not induce the expression of stress-protective genes, but regulates their abundance to prevent detrimental hyper-activation of stress responses.

While large advances have been made to discover SUMO conjugates in plants, more investigation is needed into the regulation of target selection, as well as how SUMO addition alters the activity, interactions, location and/or stability of these proteins, and ultimately how these modifications impact plant growth, development and stress protection. Towards this goal, I investigated three separate aspects of the SUMO conjugation pathway in this thesis to further elucidate the roles of SUMOylation in *Arabidopsis thaliana*.

CHAPTER 2

CONSERVATION OF SUMO HOMOLOGS AND OTHER SUMO FAMILY MEMBERS IN PLANTS

I completed all the analyses in this chapter. Figures were edited by Drs. Robert C. Augustine and Richard D. Vierstra.

This work was published in Augustine RC, York SL, Rytz TC, Vierstra RD (2016) Defining the SUMO System in Maize: SUMOylation Is Up-Regulated During Endosperm Development and Rapidly Induced by Stress. *Plant Physiol.* 171:2191–2210.

ABSTRACT

In *Arabidopsis thaliana*, the SUMO family comprised of eight isoforms, with only SUMO1, SUMO2, SUMO3 and SUMO5 expressed. SUMO1 and SUMO2 are 93% identical to each other and are essential to plant development and growth. Whereas SUMO1 and SUMO2 are the primary stress-responsive isoforms and are conjugated to target proteins, it remains unclear if the less-conserved SUMO3 and SUMO5 form conjugates or even serve a biological function.

To clarify the roles of SUMO3 and SUMO5 in *Arabidopsis*, I conducted a phylogenetic study of the plant SUMO family to determine the evolutionary relationship of SUMO3 and SUMO5 to the other SUMO isoforms in *A. thaliana* and across the plant kingdom. Sequence comparisons revealed that plants express two clades of SUMOs: one highly conserved group of canonical SUMOs, including SUMO1/2, and a more divergent, non-canonical group of SUMOs, including SUMO3 and SUMO5. The non-canonical isoforms have not only low sequence identities to the canonical SUMOs, but also to each other. Non-canonical SUMOs are lineage specific, with SUMO3 and SUMO5 being conserved only within the *Brassicaceae*. Furthermore, two additional SUMO family proteins were identified, a diSUMO-like protein only present in grasses, as well as a novel SUMO variant containing a SUMO-like domain preceded by a long N-terminal extension and lacking the di-glycine motif essential for conjugation.

INTRODUCTION

Post-translational modification of proteins with the Small Ubiquitin-like Modifier (SUMO) is an essential process in eukaryotic organisms and plays a central role in the regulation of development, DNA repair, and biotic and abiotic stresses responses [3, 97, 140]. This 10-kDa modifier is part of the ubiquitin-fold protein family characterized by a central beta-grasp fold of approximately 70 amino acids preceded by a flexible 10 – 20 amino acid N-terminal extension. Typical, SUMOs are translated as propeptides, which are cleaved by SUMO proteases to reveal the C-terminal di-glycine (di-Gly) motif. SUMO conjugation to proteins occurs through the formation of an isopeptide bond between the C-terminal carboxyl group of the di-Gly motif and an ϵ-amino group of specific lysine residues in the target. Most target lysines are located in a consensus SUMOylation motif consisting of a hydrophobic residue preceding the lysine residue followed by an acidic residue two positions downstream of the lysine (ψKxD/E) [89]. The of SUMO to proteins is driven by a three-step enzymatic cascade involving a heterodimeric E1 SUMO activating enzyme (SAE1/2), an E2 SUMO conjugating enzyme (SCE1) and a SUMO E3 ligase. Besides being conjugated to target proteins, SUMO may be ligated to itself to form polySUMO chains.

In *Arabidopsis thaliana,* eight genes encode for SUMO [28, 30]. Of these eight loci, only *SUMO1*, *SUMO2*, *SUMO3* and *SUMO5* are transcriptionally active [28, 31, 32]. SUMO1 and SUMO2 (SUMO1/2) are 93% identical to each other with only six residues differing between the two proteins. They are essential, as the homozygous double mutant *sumo1-1 sumo2-1* arrests in early embryonic development [21]. SUMO1/2 are the primary stress-responsive SUMOs and are rapidly conjugated to target proteins upon challenging conditions, such as heat-shock or oxidative stress [21, 28, 33]. While SUMO1/2 and their targets have been a major focus of

study, less is known about the roles of SUMO3 and SUMO5. The loss of SUMO3 has little to no effect on plant growth [34]. While reconstituted SUMOylation assays have shown that SUMO3 and SUMO5 can be conjugated to other proteins, no physiologically relevant argets have yet been identified [35, 36, 38]. Little is known about SUMO5 as it has not been studied in detail.

While *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe, Drosophila melanogaster*, and *Caenorhabditis elegans* encode only a single form of SUMO, four SUMO isoforms exist in mammals: the ubiquitously expressed SUMO1, SUMO2 and SUMO3, and the more distantly related SUMO4 that is detected in just a few tissues [39]. Like SUMO1/2 in *Arabidopsis*, Mammalian SUMO2 and SUMO3 (SUMO2/3) are closely related to each other and are essential [41]. Most proteins are SUMOylated by SUMO2/3 and upon different stress conditions, including heat shock and ischemia, SUMO2/3 are rapidly conjugated to targets; however, SUMO1 is also ligated to proteins [25, 40, 45]. Human SUMO2/3 and SUMO1 serve distinct roles by being conjugated to different sets of proteins [42]. As an example, RanGAP1 is selectively SUMOylated by SUMO1 [44]. Although mammalian SUMO1 only shares a ~50% sequence identity with SUMO2/3 and has set of target proteins separate from SUMO2/3, its functions can be complemented by SUMO2 or SUMO3 [41, 44, 97]. Less is known about the tissue-specific SUMO4, but studies have implicated this isoform in the immune response [39, 42]. The functional difference observed between the mammalian SUMO isoforms has been proposed to be through the selective binding of specific interacting proteins base on sequence differences in the SUMO isoforms [96, 97].

Since *A thaliana* SUMO3 and SUMO5 have not been studied in detail, questions remain whether they are active and have distinct roles from SUMO1/2, as has been observed with the

mammalian SUMO isoforms. In this chapter, I conducted a phylogenetic study of all SUMOs from a range of plant species to determine the conservation of SUMO3 and SUMO5 homologs in other plant species. High conservation might indicate a specific role of these isoforms in plant biology.

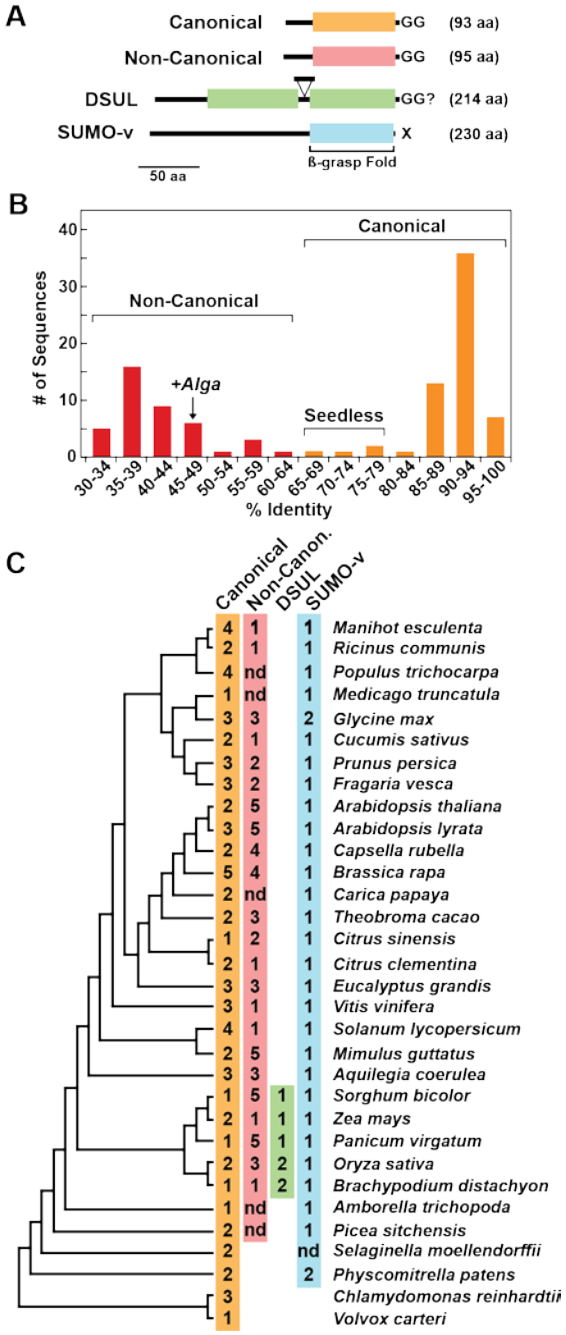
By both sequence comparison and phylogenetic analysis, I found that plants have two distinct clades of SUMO isoforms: a highly conserved group of canonical SUMOs, including Arabidopsis SUMO1/2, and a more divergent, non-canonical group of SUMOs that includes SUMO3 and SUMO5. Non-canonical versions have both low sequence identities to canonical SUMOs as well as to each other. Through the phylogenetic analysis, I discovered that these non-canonical isoforms are lineage specific and found that SUMO3 and SUMO5 are conserved only within the *Brassicaceae*. However, multiple other plant species also contain non-canonical SUMOs implying that these divergent forms might serve a role in plant biology that is independent of sequence homology. Additionally, I identified two additional SUMO family proteins, a diSUMO-like protein only present in grasses, as well as a novel SUMO variant, containing a SUMO domain lacking the C-terminal di-Gly motif preceded by a long N-terminal extension upstream.

RESULTS

Identification of SUMO Homologs in Plants

To identify SUMO homologs in other plants including gymnosperms, dicots, monocots, seedless plants and algae, the predicted proteomes of 32 species available through Phytozome v9.0 (The Plant Genome Resource: www.phytozome.net) were searched by pBLAST using the

Figure 2-1. **Plants express two groups of SUMO isoforms and two SUMO-like proteins.** **(A)** Protein diagram of the canonical, non-canonical, diSUMO (DSUL) and SUMO-v, with the maize protein length in parenthesis. The colored boxes and black represent the beta-grasp fold and N-terminal extension or linker, respectively. The di-glycine motif is noted by the "GG". The variable linker region of DSUL is illustrated by the black triangle. **(B)** Distribution of the percent identity of SUMO orthologs to *At*SUMO1. **(C)** Species tree illustrating the number and distribution of SUMO isoforms, as well as those of the diSUMO and SUMO-V. The description "nd" refers to not detected. The species tree was adapted from Phytozome v9.1.


A
Canonical
GG
(93 aa)
Non-Canonical
GG
(95 aa)
DSUL
GG?
(214 aa)
SUMO-v
X
(230 aa)
ß-grasp Fold
50 aa
B
Canonical
Non-Canonical
+Alga
Seedless
# of Sequences
% Identity
30-34
35-39
40-44
45-49
50-54
55-59
60-64
65-69
70-74
75-79
80-84
85-89
90-94
95-100
C
Canonical
Non-Canon.
DSUL
SUMO-v
Manihot esculenta
Ricinus communis
Populus trichocarpa
Medicago truncatula
Glycine max
Cucumis sativus
Prunus persica
Fragaria vesca
Arabidopsis thaliana
Arabidopsis lyrata
Capsella rubella
Brassica rapa
Carica papaya
Theobroma cacao
Citrus sinensis
Citrus clementina
Eucalyptus grandis
Vitis vinifera
Solanum lycopersicum
Mimulus guttatus
Aquilegia coerulea
Sorghum bicolor
Zea mays
Panicum virgatum
Oryza sativa
Brachypodium distachyon
Amborella trichopoda
Picea sitchensis
Selaginella moellendorffii
Physcomitrella patens
Chlamydomonas reinhardtii
Volvox carteri

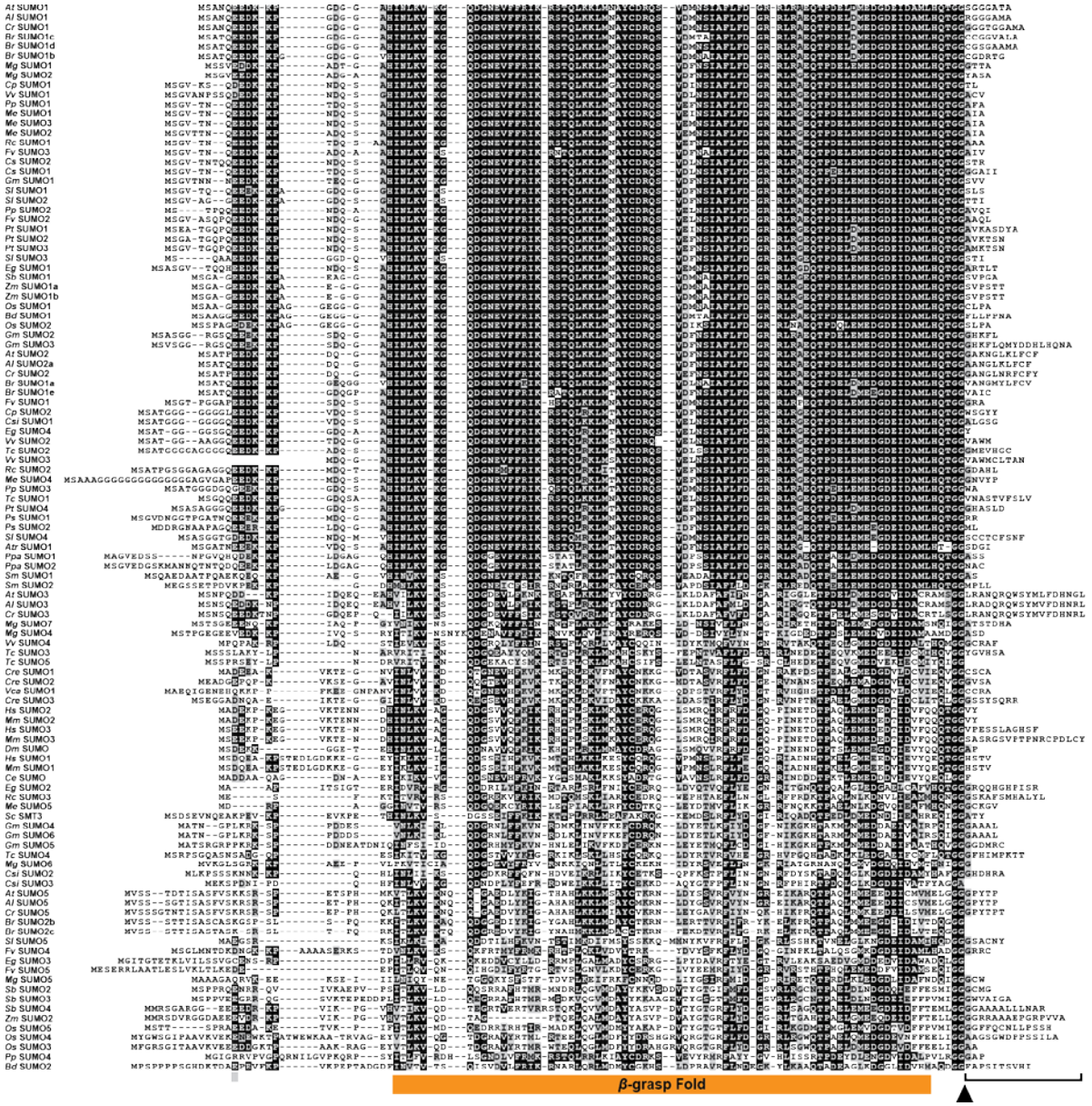
full-length *A. thaliana* SUMO1 (*At*SUMO1) protein sequence as the query. For all species, I identified genes encoding one or more SUMO isoforms. A total of 114 full-length sequences were identified (Figure 1-1C, Appendix 1). Of those, 103 proteins contained a di-Gly motif within 20 amino acids of the C-terminal end indicating that these peptides might be processed by SUMO proteases and subsequently conjugated to proteins. To eliminate errors present in the genome annotation, the genomic sequences were compared to that of *At*SUMO1 and/or to assembled expressed sequence tag (EST) databases available through Phytozome to establish the correct protein sequence. For example, six isolated SUMO genese, *Brassica rappa SUMO1e*, *Capsella rubella SUMO3*, *Prunus persica SUMO2*, *Fragaria vesca SUMO2* and *SUMO5* and *Theobroma cacao SUMO1*, were found to have missannotated start codons.

Sequence Alignments Highlight Two Distinct Groups of SUMO Isoforms

A. thaliana SUMO3 and SUMO5 are divergent SUMO homologs with 54% and 44% sequence identities to *At*SUMO1, respectively, as compared to the 90% protein identity between SUMO1 and SUMO2. To determine the evolutionary relationship of SUMO3 and SUMO5 to SUMOs from other plant species, I performed a sequence comparison and subsequent phylogenetic analysis of the protein sequences of the 103 plant isoforms described above. As the amino acid sequences past the di-Gly motif are highly variable, only the residues up to and including the motif were used in the analysis. The amino acid sequences for human (*Hs*) and mouse (*Mm*) SUMO1, SUMO2 and SUMO3, *C. elegans* SUMO and *S. cervisiae (Sc)* SMT3 were included as outliers.

The MAFT (Multiple Alignment using Fast Fourier Transform) multiple sequence alignment program (http://www.ebi.ac.uk/Tools/msa/mafft/) was used to align the SUMO

Figure 2-2. **Protein alignment of SUMOs illustrates the conservation of canonical SUMOs**. The triangle indicates the location of the di-glycine motif. The horizontal bracket highlights the variable region after the di-glycine motif. The 103 identified full-length plant SUMO protein sequences were aligned with MAFFT multiple sequence alignment program and shaded based on conservation using the Box Shade server. Abbreviations: *Ac, Aquilegia coerulea; Al, Arabidopsis lyrata; At, Arabidopsis thaliana; Atr, Amborella trichopoda; Bd, Brachypodium distachyon; Br, Brassica rapa; Cc, Citrus clementina; Ce, Caenorhabditis elegans; Cp, Carica papaya; Cr, Capsella rubella; Cre, Chlamydomonas reinhardtii; Cs, Citrus sinensis; Dm, Drosophila melanogaster; Eg, Eucalyptus grandis; Fv, Fragaria vesca; Gm, Glycine max; Hs, Homo sapiens; Md, Malus domestica; Me, Manihot esculenta; Mm, Mus musculus; Mt, Medicago truncatula; Os, Oryza sativa; Pp, Prunus persica; Ppa, Physcomitrella patens; Ps, Picea sitchensis; Pt, Populus trichocarpa; Pv, Phaseolus vulgaris; Pvi, Panicum virgatum; Sb, Sorghum bicolor; Sc, Saccharomyces cereviseae; Si, Setaria italica; Sl, Solanum lycopersicum; Sm, Selaginella moellendorffii; Tc, Theobroma cacao; Vc, Volvox carteri; Vv, Vitis vinifera; Zm, Zea mays*


Canonical
Non-Canonical
β-grasp Fold

amino acid sequences to each other (Figure 2-2). The protein alignment brought to light that there are two distinct groups of SUMOs: a group of conserved canonical SUMOs as well as a group of less conserved, or non-canonical, SUMOs. Comparison of their sequence identity to *At*SUMO1 confirmed that the SUMOs separate into two separate sets. The canonical SUMO orthologs are at least a 65% identical to the amino acid sequence of *At*SUMO1, with most having a sequence identity greater than 85%; only the SUMOs from the seedless plants *Selaginella moellendorfii* and *Physcomitrella patens* share less than 85% identity to *At*SUMO (Figure 2-1B). On the other hand, the non-canonical isoforms formed a discrete group that differs considerably from the canonical sequences with only 30-45% residues sharing identities with *At*SUMO1. Besides having a low sequence identity to *At*SUMO1, the non-canonical SUMOs have little sequence homology amongst each other (30-45% identity). Algal SUMOs, having a larger evolutionary distance to land plants, are less well conserved, and thus were grouped with the non-canonical SUMOs.

Among the canonical SUMO, there is very little sequence variation within the beta-grasp fold up to the di-Gly motif, however the length and composition of the N-terminal extensions are highly variable. Interestingly, the one conserved domain in the extension corresponds with a SUMOylation motif (EEDKKP) corresponding to the SUMOylation of *At*SUMO1 at lysine 10 [36]. *At*SUMO1 has two additional mapped SUMOylation sites (lysines 23 and 42) [33], which are also maintained in the canonical clade of SUMOs. Conservation of this motif, as well as K23 and K42, emphasizes a functional role of polySUMO chain formation in plants. Interestingly, the SUMOylation sequence in the N-terminal extension is absent in most non-canonical SUMOs.

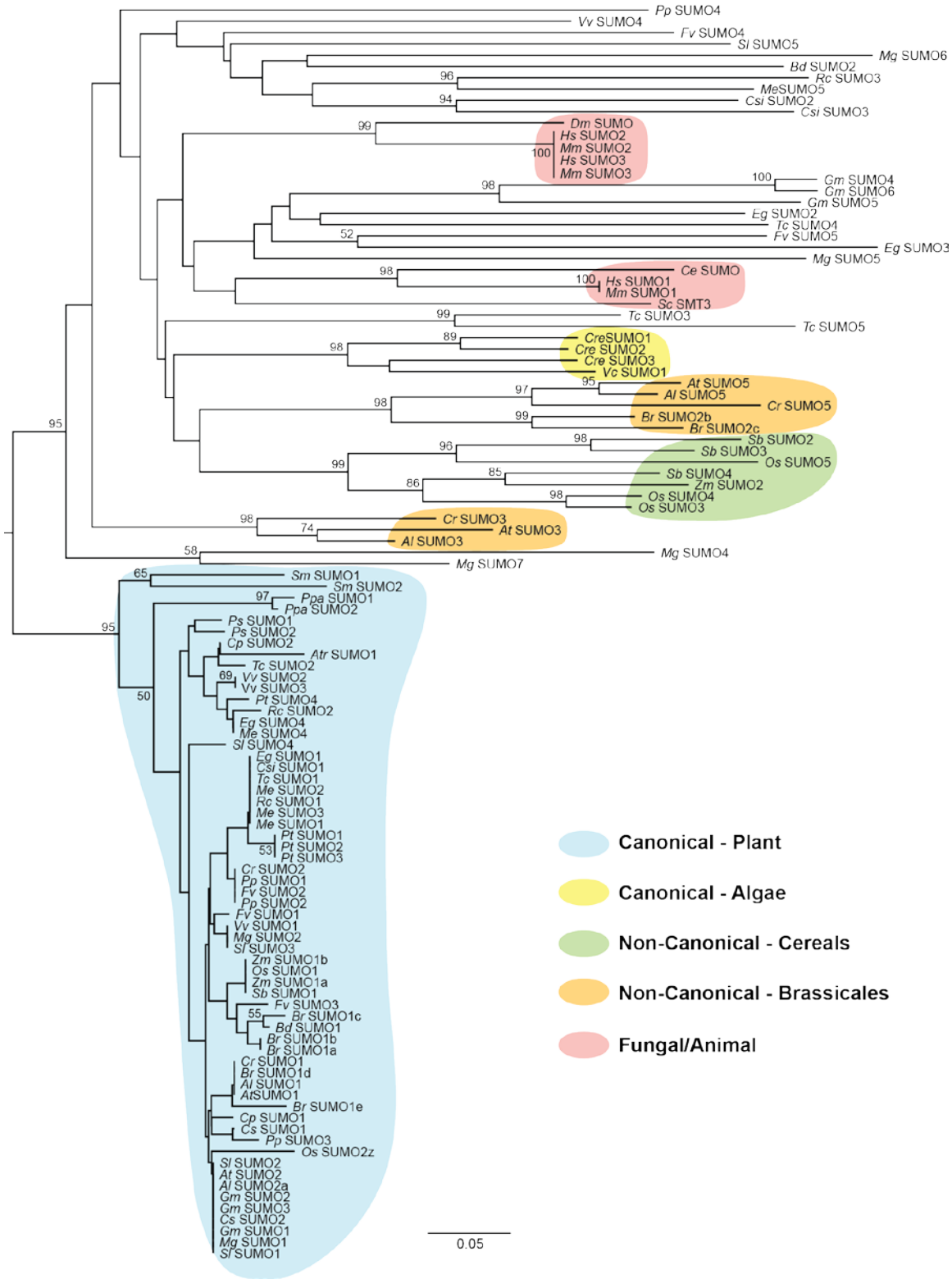
Additionally, analysis of the intron/exon structure of the SUMO isoforms revealed that the canonical SUMOs share identical intron/exon boundaries. However, while some non-

canonical SUMOs have a similar intron/exon organization as canonical SUMOs, numerous isoforms are intron-less, suggesting they have evolved from retrograde integration of mRNA into the genome.

SUMO3 and SUMO5 Are Not Conserved Outside of The *Brassicacae* Family

From the protein alignment, a consensus neighbor-joining tree was generated using Pearson correlations to determine the phylogenetic relationships of the SUMO isoforms. The phylogenetic tree confirms that the SUMO family separates into two distinct evolutionary groups: one clade containing the very closely related canonical SUMOs found in all land plants, including *At*SUMO1/2, and a clade containing more divergent, non-canonical versions, including *At*SUMO3 and 5 (Figure 2-3). The animal and yeast SUMOs, as well as the algal SUMO isoforms, grouped into the more divergent clade, highlighting their evolutionary distance from canonical land plant SUMOs. Interestingly, non-canonical SUMOs do not have a clear evolutionary relationship to each other, but seem to be lineage specific. *At*SUMO3 and *At*SUMO5 were found to be conserved only in the *Brassicaceae* family. Although *At*SUMO3 and *At*SUMO 5 are not present in all plants, most species have at least one non-canonical SUMO, indicating that the existence, but not the sequence, of these isoforms family proteins is conserved (Figure 2-1C).

Figure 2-3. **Phylogenetic tree of plant SUMOs.** The tree was generated from the MAFFT amino acid sequence alignment using the Neighbor-Joining algorithm with Pearson correlation and 1000 iterations in MEGA 6. Protein groups of interest are highlighted. Only percent boot strap values above 50 are illustrate on the consensus tree. Abbreviations: *Ac, Aquilegia coerulea; Al, Arabidopsis lyrata; At, Arabidopsis thaliana; Atr, Amborella trichopoda; Bd, Brachypodium distachyon; Br, Brassica rapa; Cc, Citrus clementina; Ce, Caenorhabditis elegans; Cp, Carica papaya; Cr, Capsella rubella; Cre, Chlamydomonas reinhardtii; Cs, Citrus sinensis; Dm, Drosophila melanogaster; Eg, Eucalyptus grandis; Fv, Fragaria vesca; Gm, Glycine max; Hs, Homo sapiens; Md, Malus domestica; Me, Manihot esculenta; Mm, Mus musculus; Mt, Medicago truncatula; Os, Oryza sativa; Pp, Prunus persica; Ppa, Physcomitrella patens; Ps, Picea sitchensis; Pt, Populus trichocarpa; Pv, Phaseolus vulgaris; Pvi, Panicum virgatum; Sb, Sorghum bicolor; Sc, Saccharomyces cereviseae; Si, Setaria italica; Sl, Solanum lycopersicum; Sm, Selaginella moellendorffii; Tc, Theobroma cacao; Vc, Volvox carteri; Vv, Vitis vinifera; Zm, Zea mays*


Pp SUMO4
Vv SUMO4
Fv SUMO4
Sl SUMO5
Mg SUMO6
Bd SUMO2
Rc SUMO3
MeSUMO5
Csi SUMO2
Csi SUMO3
Dm SUMO
Hs SUMO2
Mm SUMO2
Hs SUMO3
Mm SUMO3
Gm SUMO4
Gm SUMO6
Gm SUMO5
Eg SUMO2
Tc SUMO4
Fv SUMO5
Eg SUMO3
Mg SUMO5
Ce SUMO
Hs SUMO1
Mm SUMO1
Sc SMT3
Tc SUMO3
Tc SUMO5
CreSUMO1
Cre SUMO2
Cre SUMO3
Vc SUMO1
At SUMO5
Al SUMO5
Cr SUMO5
Br SUMO2b
Br SUMO2c
Sb SUMO2
Sb SUMO3
Os SUMO5
Sb SUMO4
Zm SUMO2
Os SUMO4
Os SUMO3
Cr SUMO3
At SUMO3
Al SUMO3
Mg SUMO4
Mg SUMO7
Sm SUMO1
Sm SUMO2
Ppa SUMO1
Ppa SUMO2
Ps SUMO1
Ps SUMO2
Cp SUMO2
Atr SUMO1
Tc SUMO2
Vv SUMO2
Vv SUMO3
Pt SUMO4
Rc SUMO2
Eg SUMO4
Me SUMO4
Sl SUMO4
Eg SUMO1
Csi SUMO1
Tc SUMO1
Me SUMO2
Rc SUMO1
Me SUMO3
Me SUMO1
Pt SUMO1
Pt SUMO2
Pt SUMO3
Cr SUMO2
Pp SUMO1
Fv SUMO2
Pp SUMO2
Fv SUMO1
Vv SUMO1
Mg SUMO2
Sl SUMO3
Zm SUMO1b
Os SUMO1
Zm SUMO1a
Sb SUMO1
Fv SUMO3
Br SUMO1c
Bd SUMO1
Br SUMO1b
Br SUMO1a
Cr SUMO1
Br SUMO1d
Al SUMO1
AtSUMO1
Br SUMO1e
Cp SUMO1
Cs SUMO1
Pp SUMO3
Os SUMO2z
Sl SUMO2
At SUMO2
Al SUMO2a
Gm SUMO2
Gm SUMO3
Cs SUMO2
Gm SUMO1
Mg SUMO1
Sl SUMO1
0.05
Canonical - Plant
Canonical - Algae
Non-Canonical - Cereals
Non-Canonical - Brassicales
Fungal/Animal

Identification of Additional SUMO Family Proteins

Besides identifying canonical and non-canonical SUMOs, the search for homologs also yielded two different SUMO-like proteins: a diSUMO-like (DSUL) protein only identified in grasses, as well as a SUMO variant (SUMO-v) found in all land plants (Figure 2-1A,C). The DSUL consist of two SUMO-like domains joined by a variable linker that may include a di-Gly motif (Figure 2-4). Although the sequences and lengths of the linkers are very variable, the SUMO domains of the DSULs have a high level of conservation to each other. Studies in maize have reported that the DSUL is expressed only in maternal and paternal gametophytic tissues and is required for normal maternal gametophyte development [141, 142]. DSULs do have a di-Gly motif in the C-terminal ends, but it is unknown whether they are processed into an active form and conjugated to other proteins.

The SUMO-v is a ~200 amino acid protein consisting of a SUMO-like domain preceded by a ~100 amino acid extension (Figure 2-5). Surprisingly, SUMO-v does not end in a di-Gly motif, suggesting the protein is not conjugated to other proteins. The N-terminal extension consists of stretches of charged amino acids signifying the region might be unstructured. Additionally, at least one potential SUMO interacting motif (SIM) (residues 54-62 in *At*SUMO-v) is present in the extension. The presence of these SIMs suggests that the N-terminal region of SUMO-v may interact with proteins modified by SUMO. Although SUMO-v cannot be conjugated to proteins as it is lacking the required di-Gly motif, this variant might still interact with the SUMOylation machinery. In support, many of the residues necessary for interaction with the E1 and E2 are conserved in the SUMO-like domain (Figure 2-5) [38]. In addition, a yeast two-hybrid screen using SCE1 identified SUMO-v as an interactor with the E2 SCE1 [35]. Furthermore, SUMO-v might be SUMOylated, as the *A. thaliana* protein contains several

Figure 2-4. Protein alignment of diSUMO-like (DSUL). The DSUL protein sequences were aligned with the T-Coffee multiple sequence alignment program and shaded based on conservation using the Box Shade server. The amino acid sequences of Maize SUMO1 (ZmS1) was added to the sequences to highlight the conservation of the SUMO-like domains. The two SUMO-like beta-grasp folds are illustrated. Abbreviations: *Bd, Brachypodium distachyon; Os, Oryza sativa; Pv, Panicum virgatum; Si, Seteria italic; Sb, Sorghum bicolor; Zm, Zea mays*

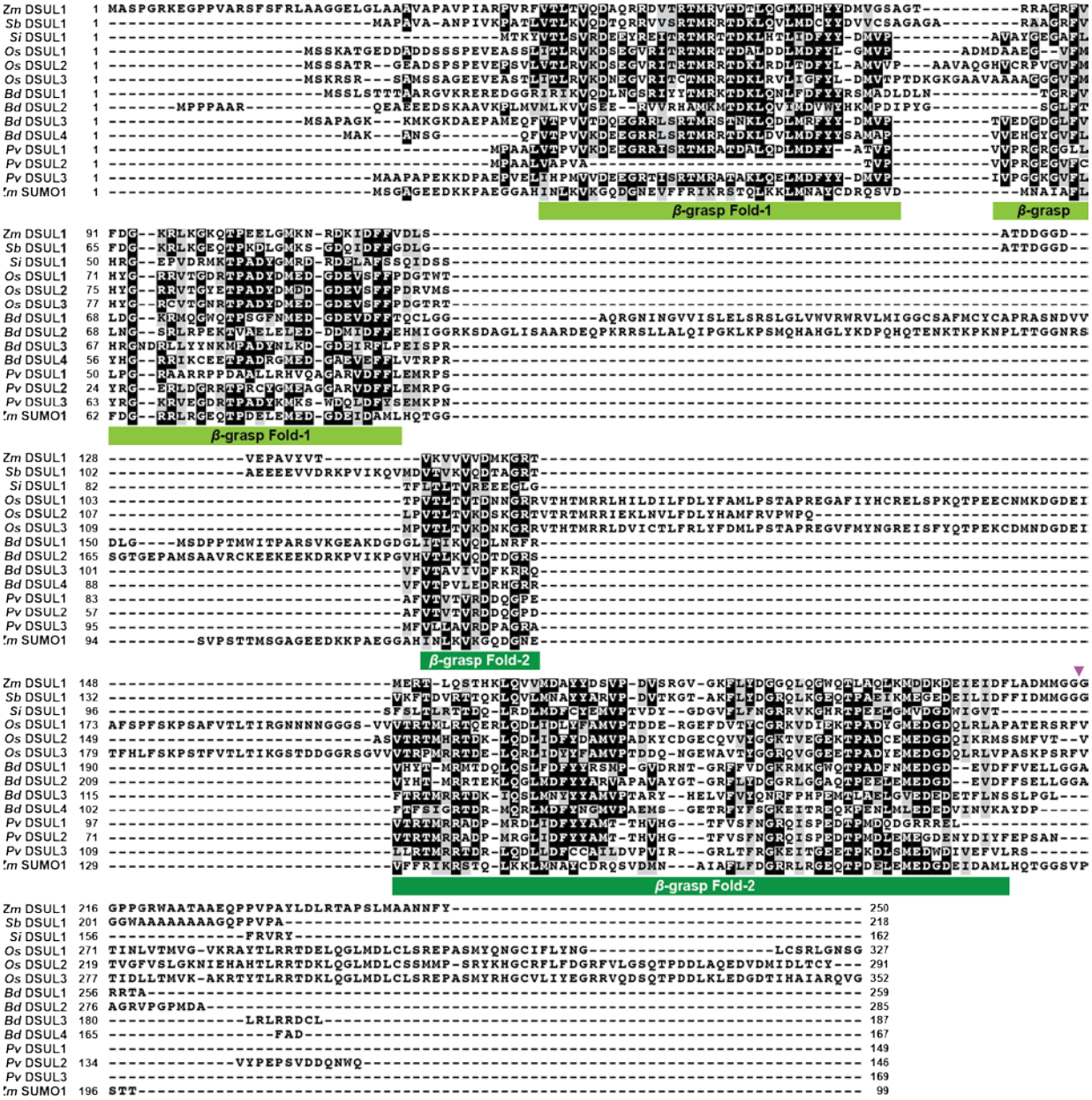
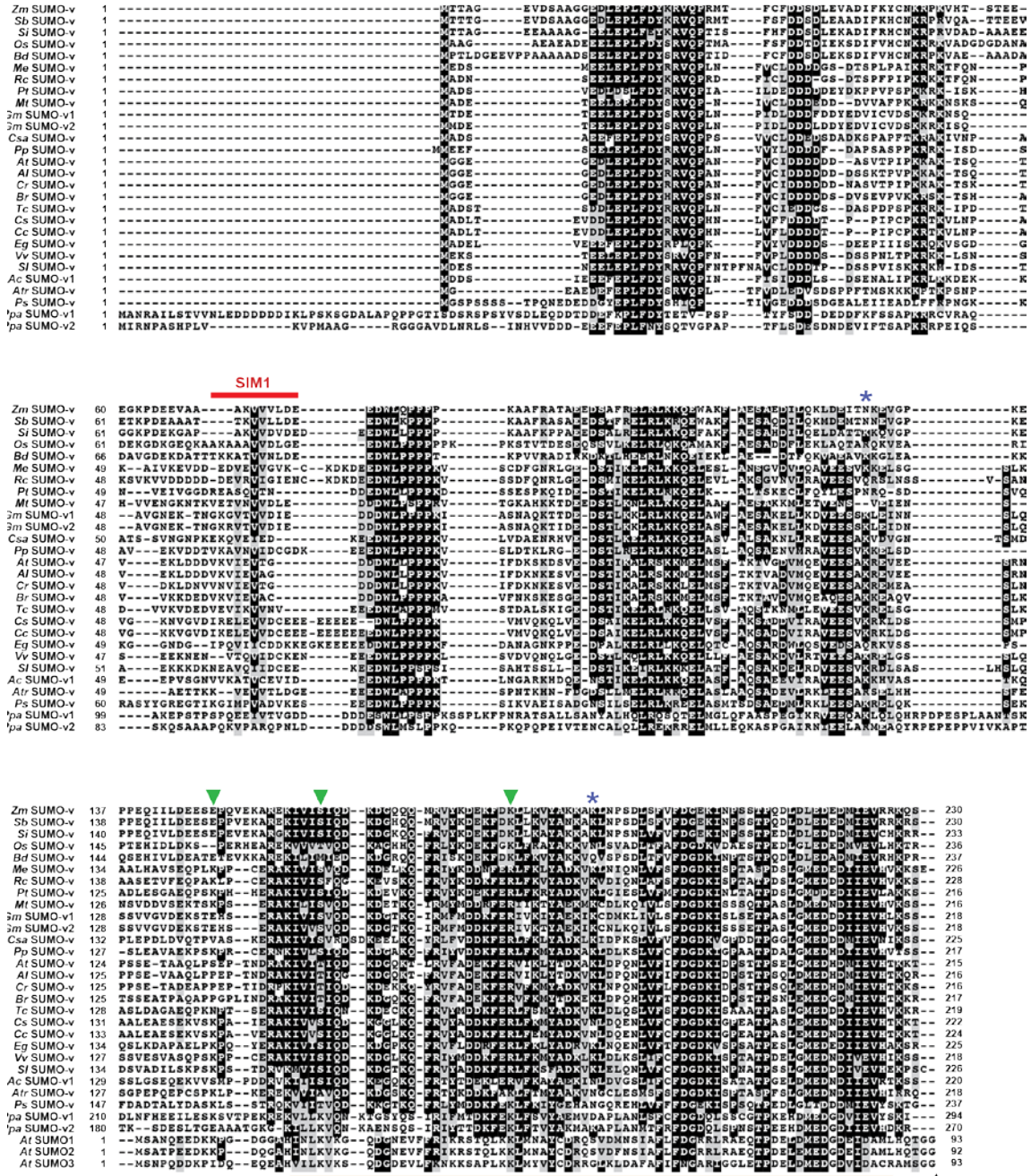

β-grasp Fold-1
β-grasp
β-grasp Fold-1
β-grasp Fold-2
β-grasp Fold-2

Figure 2-5. **Protein alignment of SUMO-v from representative plant species**. The DSUL protein sequences were aligned with the T-Coffee multiple sequence alignment program and shaded based on conservation using the Box Shade server. *At*SUMO1, 2 and 3 were included in the alignment to highlight the conservation of the SUMO-like domain. The putative SUMOylation sites on *At*SUMO-v are highlighted with a star. The black triangle points to the di-glycine motif on the *At*SUMOs and the green arrows highlight the SUMO conjugation sites on *At*SUMO1. Abbreviations: *Ac, Aquilegia coerulea; Al, Arabidopsis lyrata; At, Arabidopsis thaliana; Atr, Amborella trichopoda; Bd, Brachypodium distachyon; Br, Brassica rapa; Cc, Citrus clementina; Ce, Caenorhabditis elegans; Cp, Carica papaya; Cr, Capsella rubella; Cre, Chlamydomonas reinhardtii; Cs, Citrus sinensis; Dm, Drosophila melanogaster; Eg, Eucalyptus grandis; Fv, Fragaria vesca; Gm, Glycine max; Hs, Homo sapiens; Md, Malus domestica; Me, Manihot esculenta; Mm, Mus musculus; Mt, Medicago truncatula; Os, Oryza sativa; Pp, Prunus persica; Ppa, Physcomitrella patens; Ps, Picea sitchensis; Pt, Populus trichocarpa; Pv, Phaseolus vulgaris; Pvi, Panicum virgatum; Sb, Sorghum bicolor; Sc, Saccharomyces cereviseae; Si, Setaria italica; Sl, Solanum lycopersicum; Sm, Selaginella moellendorffii; Tc, Theobroma cacao; Vc, Volvox carteri; Vv, Vitis vinifera; Zm, Zea mays*


SIM1
β-grasp Fold

potential consensus SUMOylation sites including K93, K115 and K176. Although these lysines have a high level of conservation, they are not found in all SUMO-v sequences. Additionally, the residues Lys9, Lys10, Lys 23 and Lys 42 of *At*SUMO1/2, which are modified by SUMO, are not conserved in the SUMO variant. Interestingly, unlike SUMO, all plants have at most two loci encoding SUMO-v.

DISCUSSION AND FUTURE DIRECTIONS

From the phylogenetic analysis completed in this chapter, I found that two distinct clades of SUMO are present in plants. All species contain at least one conserved canonical SUMO. In addition, plants also have at least one non-canonical SUMO isoforms that have reduced sequence conservation to the canonical SUMOs and to each other. These non-canonical forms are lineage specific and most likely arose independently through plant evolution from genome duplications as well as retrograde integration of mRNA into the genome (See [29]). SUMO3 and SUMO5 are non-canonical isoforms that have no evident orthologs outside the *Brassicaceae* family. Although they are not conserved in all plant lineages, SUMO3 and SUMO5 might still serve a functional role in Arabidopsis. Most plant species have at least one non-canonical SUMO, suggesting that the presence of these homologs is being maintained evolutionarily and their divergent sequences may indicate neofunctionalization within the different plant lineages. Thus, SUMO3 and SUMO5 might yet be interesting SUMO homologs to study in *A. thaliana*.

Future work should focus on identifying potential targets as well as interactors of SUMO3 and SUMO5 through pulldowns of tagged forms followed by mass spectrometric analysis. In addition, identification of SUMO5 null mutants is paramount as no mutants of this isoform have yet been characterized. A recent in-depth bioinformatic study of *Brassicaceae*

SUMOs discovered that SUMO5 is under stronger purifying selection than SUMO3, and thus might have a unique role in Arabidopsis [29].

Although seven full-length genes are present in the *A. thaliana* genome, only four are transcribed. Therefore, it is possible that some of the many of the 103 genes analyzed here are not expressed, and thus might be functionally irrelevant in the corresponding organisms. Completing an in-depth expression analysis of SUMO isoforms could shed light on the functional role of non-canonical SUMOs, as little or no expression of the described genes might indicate that these genes have no crucial role in plant biology.

Besides identifying two clades of distinct SUMO isoforms, I found that plants have two additional SUMO-like proteins, a DSUL protein and a SUMO-v. The DSUL is a grass specific protein, and was identified previously to be involved in gametophyte development [141]. SUMO-v is a novel plant protein and its functions are unknown. As it does not contain the di-Gly motif, it is not conjugated to other proteins in the same fashion as canonical SUMOs. SUMO-v might act through its associations to other proteins by the SIMs in the N-terminal extension, its predicted SUMOylation motifs and/or its SUMO beta-grasp fold. Additionally, the N-terminal extension contains multiple stretches of charged residues, which could be unstructured, and thus potentially interact with a variety of substrates.

It has been proposed that plant SUMO-v might have a similar role to *S. cerevisiae* Rad60, *S. pombe* Esc2 and mouse NIP45, which are ~400 kDa proteins containing a 200-kDa unstructured N-terminal region with multiple SIMs followed by two SUMO-like domains in tandem [143]. Rad60, Esc2 and NIP45 participate in recombination, chromatin repair, and chromatin segregation [143–148]. However, no homology is observed between the 200-kDa N-terminal extension of the proteins and the extension of SUMO-v.

As SUMO-v is a unique plant SUMO-like protein with distinctive structural properties, it is an interesting subject for further investigation. Phenotypic analysis of null mutants and identification of interacting partners of SUMO-v will help establish a role of this protein in plant biology.

MATERIALS AND METHODS

Identification of SUMO Homologs and Other SUMO Family Members in Plants

To identify plant SUMO orthologs, the full-length *At*SUMO1 protein sequence was used as a query in pBLAST to search the predicted proteome databases available through Phytozome (Version 9.1; The Plant Genome Rescourse) and the National Center for Biotechnology Information (NCBI). Homologs were confirmed using reciprocal searches. If necessary, the genome annotations were corrected using alignments with *At*SUMO sequences and assembled ESTs databases available on JBrowse on Phytozome. The locus identifier, any annotation changes as well as the abbreviated name are listed in Appendix Table 1. Only full-length SUMO sequences with well established gene annotations and a di-Gly motif in their C-terminal end were considered for subsequent phylogenetic analysis. Protein identity to *At*SUMO1 was determined from the pBLAST results.

DSUL and SUMO-v were first identified by searches using *At*SUMO1. Subsequent searches employed the *Maize* DSUL and *At*SUMO-v full-length protein sequences to search the Phytozome (Version 9.1) and NCBI databases using pBLAST. Amino acid sequences were confirmed using reciprocal searches. The protein sequences for *Vv*SUMOv, *Sb*SUMOv, *Ac*SUMOv and *Atr*SUMOv were re-annotated using alignments with *At*SUMO-v sequences and

assembled ESTs available through JBrowse in Phytozome. *Os*SUMOv was not annotated in the rice genome on Phytozome or NCBI databases, but was found by using the Maize SUMO-v protein sequence to search assembled rice ESTs sequence with tBLAST-N through NCBI.

Protein Alignment and Phylogenetic Analysis

The protein alignment of all SUMOs was completed using the MAFFT sequence alignment program available on the European Molecular Biology Laboratory– European Bioinformatics Institute (EMBL-EBI) website (http://www.ebi.ac.uk/Tools/mesa/miff) using the default settings. Protein alignments for DSUL and SUMO-v were created using the T-Coffee or MUSCLE sequence alignment programs, respectively, available on the EMBL-EBI website (http://www.ebi.ac.uk/Tools/msa/tcoffee and http://www.ebi.ac.uk/Tools/msa/muscle) with the default settings selected. All alignments were visualized and shaded based on conservation using the Box Shade server (http://www.ch.embnet.org/software/BOX_form.html). For the DSUL protein alignment, the amino acid sequences of Maize SUMO1 (ZmS1) was added to the sequences to highlight the conservation of the SUMO-like domains. For the SUMO-v protein alignment, *At*SUMO1, *At*SUMO 2 and *At*SUMO 3 were included in the alignment.

For the phylogenetic tree of SUMO homologs, the MAFFT alignment was imported into MEGA6 [149] and a consensus tree was generated using the Neighbor-Joining algorithm with Pearson correlation and 1000 iterations. The tree was visualized using FigTree v1. 4.

Identification of SIMs and SUMOylation sites

The identification of SIMs on *At*SUMO-v was performed manually based on previously identified SIM motifs in plants and animals, consisting of a hydrophobic core with 3–4 aliphatic

residues, followed by a negatively charged cluster of acidic amino acids [93, 98, 150], and through the program SUMOsp2.0 [151]. SUMOylation motif analysis of *At*SUMO-v was completed using SUMOplot (http://www.abgent.com/sumoplot.html) and SUMOsp2.0. Only high probably canonical SUMOylation sites were considered.

CHAPTER 3

SUMOYLOME PROFILING IN ARABIDOPSIS REVEALS A DIVERSE ARRAY OF NUCLEAR TARGETS MODIFIED BY THE SUMO LIGASE SIZ1 DURING HEAT STRESS

The purification protocol was published in Rytz TC, Miller MJ, Vierstra RD (2016). Purification of SUMO Conjugates from Arabidopsis for Mass Spectrometry Analysis. *Methods Mol Biol* 1475:257–81.

This work will be published in Rytz TC, Miller MJ, Scalf M, McLoughlin F, Augustine RC, Smith LL, Vierstra RD (2017). SUMOylome Profiling in Arabidopsis Reveals a Diverse Array of Nuclear Targets Modified by the SUMO Ligase Siz1 During Heat Stress. *In preparation.*

Drs. Mark Scalf and Fionn McLoughlin performed the mass spectrometric analyses. The RNA-seq analysis was performed by Dr. Robert C. Augustine. I completed the remainder of the work described in this chapter. Figures were edited by Dr. Richard D. Vierstra.

ABSTRACT

The post-translational addition of Small Ubiquitin-like Modifier (SUMO) is an essential protein modification in plants that provides protection against numerous environmental challenges. Ligation is accomplished by a small set of SUMO ligases, with the SAP-MIZ domain-containing (SIZ)-1 and Methyl Methanesulfonate-Sensitive (MMS)-21 ligases having critical roles in stress protection and DNA endoreduplication, respectively. To help identify their cognate targets, we combined Arabidopsis *siz1* and *mms21* mutants with proteomic analyses of SUMOylated proteins enriched using an engineered SUMO suitable for mass spectrometric studies. Through multiple datasets from Arabidopsis seedlings grown at normal temperatures or exposed to heat stress, we identified over 1,000 SUMO targets, most of which are nuclear localized. Whereas no targets could be assigned to MMS21, suggesting that only a few low abundance proteins are substrates, numerous targets could be assigned to SIZ1, including major transcription factors, co-activators/repressors, and chromatin modifiers connected to abiotic and biotic stress responses, some of which associate into multi-subunit regulatory complexes. The list of SIZ1 substrates indicates that SUMOylation by this ligase provides stress protection by modifying a large array of key nuclear regulators.

INTRODUCTION

The covalent attachment of Small Ubiquitin-Like Modifier (SUMO) to other proteins provides an essential mechanism to control the activity, localization, and turnover of many intracellular proteins in eukaryotes [2, 3, 152]. Besides regulating development and cellular homeostasis under normal growth conditions, SUMOylation plays a central role in protection against a variety of abiotic and biotic challenges. As examples, SUMOylation has been connected genetically in plants to basal and acquired thermotolerance, resistance to cold, salt and drought stress, response to phosphate starvation, and innate immunity [26, 153]. Some of these responses are linked to the stress hormones salicylic acid (SA) and abscisic acid (ABA) and their associated signaling pathways [34, 68, 109, 120].

Most notable is the rapid and reversible accumulation of SUMO conjugates during stress, which for heat stress is one of the fastest molecular responses observed, suggesting that the formation of specific SUMO conjugates helps confer stress protection [21, 28]. Indeed, SUMOylation of the transcription factors PHR1, ICE1, HSFA2, ABI5, MYB30, and FLD are associated with tolerance to phosphate starvation, cold and heat tolerance, ABA signaling, and flowering time, respectively [53, 67, 68, 119, 120, 154], and modification of phytochrome B has been connected to light signaling by this photoreceptor [155]. Additionally, SUMOylation of nitrate reductase (NIA1/2) and the DNA chromomethylase (CMT)-3 has been linked to enhanced nitrogen assimilation and the epigenetic regulation of gene expression, respectively [156, 157].

SUMOylation is driven by an E1-E2-E3 reaction cascade by which the SUMO moiety is first activated via ATP hydrolysis by a heterodimeric SUMO-activating enzyme (SAE, or E1) complex comprised of the SAE1 and SAE2 subunits [2, 4]. The charged SUMO is transferred to a SUMO conjugating enzyme (SCE)-1 (or E2) through transesterification, and finally donated to

substrate proteins, often with assistance from a SUMO ligase (or E3). The end result is SUMO covalently coupled by an isopeptide bond that links the C-terminal glycine of SUMO to specific lysines in the target. In many cases, additional SUMOs are then attached, sometimes by using previously bound SUMOs to concatenate poly-SUMO chains [33, 36]. Once bound, the SUMO moieties alter the function(s) of the target proteins, and include changes in their intracellular location, activity, and/or interactions with other cellular factors, including proteins bearing SUMO-interacting motifs (SIMs) [98]. Recent studies showing that SUMO also provides sites for subsequent ubiquitylation [33, 46, 97] raise the possibility that SUMO addition also directs subsequent turnover of the protein by the Ub/26S proteasome proteolytic system. Additionally, SUMO addition can be reversed by a collection of deSUMOylating proteases (DSPs) that specifically cleave the intervening isopeptide bond, thus releasing both the protein and SUMO moiet(ies) intact [61, 62].

Akin to protein modification by ubiquitin (Ub), the specificity of SUMOylation is conferred mainly by the E3s, which help connect specific substrates to the E2-SUMO intermediate and then promote SUMO transfer. Ultimately a myriad of proteins become SUMOylated. In Arabidopsis for example, previous proteomic studies identified over 350 SUMO targets, some of which appear to be dynamically regulated [33, 46]. Most are localized to the nucleus and have functions related to DNA modification, chromatin assembly and structure, transcription, and RNA processing, export and turnover. However, whereas the Ub system employs a large and diverse array of E3s with strong substrate specificities to direct conjugation [158], the SUMO system appears to engage a more limited collection. In Arabidopsis for example, only four E3s have been described thus far, SAP and MIZ1 ligase (SIZ)-1 [53], METHYL METHANESULFONATE-SENSITIVE (MMS)-21 (or HIGH PLOIDY

(HPY)-2) [55, 56], and the PIAS Protein Inhibitors of Activated STATs (PIAL)-1 and PIAL-2 ligases [32]. That such a small set of E3s impacts such a large array of proteins implies that substrate specificity is mostly endowed by additional features beyond the target such as its location and assembly into protein complexes, as well as by the E2 binding directly to the substrate [159].

Important steps toward understanding the molecular ramifications of SUMOylation would be cataloging the proteins modified by each E3, defining how SUMO addition alters the activity, interactions, location and/or half-life of these targets, and ultimately how these modifications impact plant growth, development and stress protection. Toward this goal, we developed here a non-biased proteomic strategy to help assign individual SUMO ligases to specific Arabidopsis targets. It involves combining mutants eliminating specific E3s with a purification background in which the two highly-related SUMO isoforms (SUMO1 and 2) responsible for most SUMOylation were genetically replaced with a variant (6His-SUMO1(H89-R)) engineered to enabled affinity purification of SUMO conjugates and subsequent identification of attachment sites by tandem mass spectrometry (MS) [33, 46]. A stringent three-step purification protocol based on the 6His tag and anti-SUMO antibodies was then employed to isolate SUMO conjugates.

As a first test, we studied the SUMOylation patterns before and after a brief heat stress in mutants missing SIZ1 and MMS21, which have been linked to protection against a variety of stresses [53, 68, 109, 156] and DNA endoreduplication and the cell cycle [55, 56], respectively. Both E3s contain the essential SP-RING domain that docks the E2-SUMO intermediate. While MMS21 is devoid of other recognizable features, SIZ1 includes signature Scaffold Attachment Factor-A/B/Acinus-PIAS (SAP), Plant Homeodomain (PHD), a Proline-Isoleucine-Isoleucine-

Threonine (PIIT) sequences, which are followed by a pair of SIM sequences (Figure 1A). The substrate(s) of MMS21 are currently unknown. In contrast, SIZ1 drives much of stress-induced SUMOylation [21], and has been connected to the modification of several Arabidopsis regulators, including PHR1, GTE3, HSFA2, MYB30, CMT3 and NIA1/2 [53, 74, 120, 154, 156, 157].

Whereas prior proteomic studies identified over 350 possible SUMO targets in Arabidopsis) [33, 46], the improved MS instrumentation used here increased this collection to over a thousand. Although no targets could be assigned to MMS21 by label-free quantification, over 100 targets could be assigned with high confidence to SIZ1, especially after heat stress. Most of these SIZ1 substrates reside in the nucleus, and include well-known transcription factors, coactivators/repressors and chromatin modifiers, as well as many proteins involved biotic and abiotic stress responses. Of interest are the co-repressors TOPLESS, its paralog TOPLESS-related 2 (TPR2), and several of their interacting partners ARF2, NAC052 and EMF1 [160]. In addition, a small set of proteins became more SUMOylated in the absence of SIZ1, including BAG7, a heat-induced co-chaperone active in the unfolded protein response, and NSE4 that is part of Smc5-Smc6 DNA repair complex. This deep catalog of SIZ1-dependent SUMO targets described here provides a framework to understand how SUMO and this E3 contribute to plant stress protection.

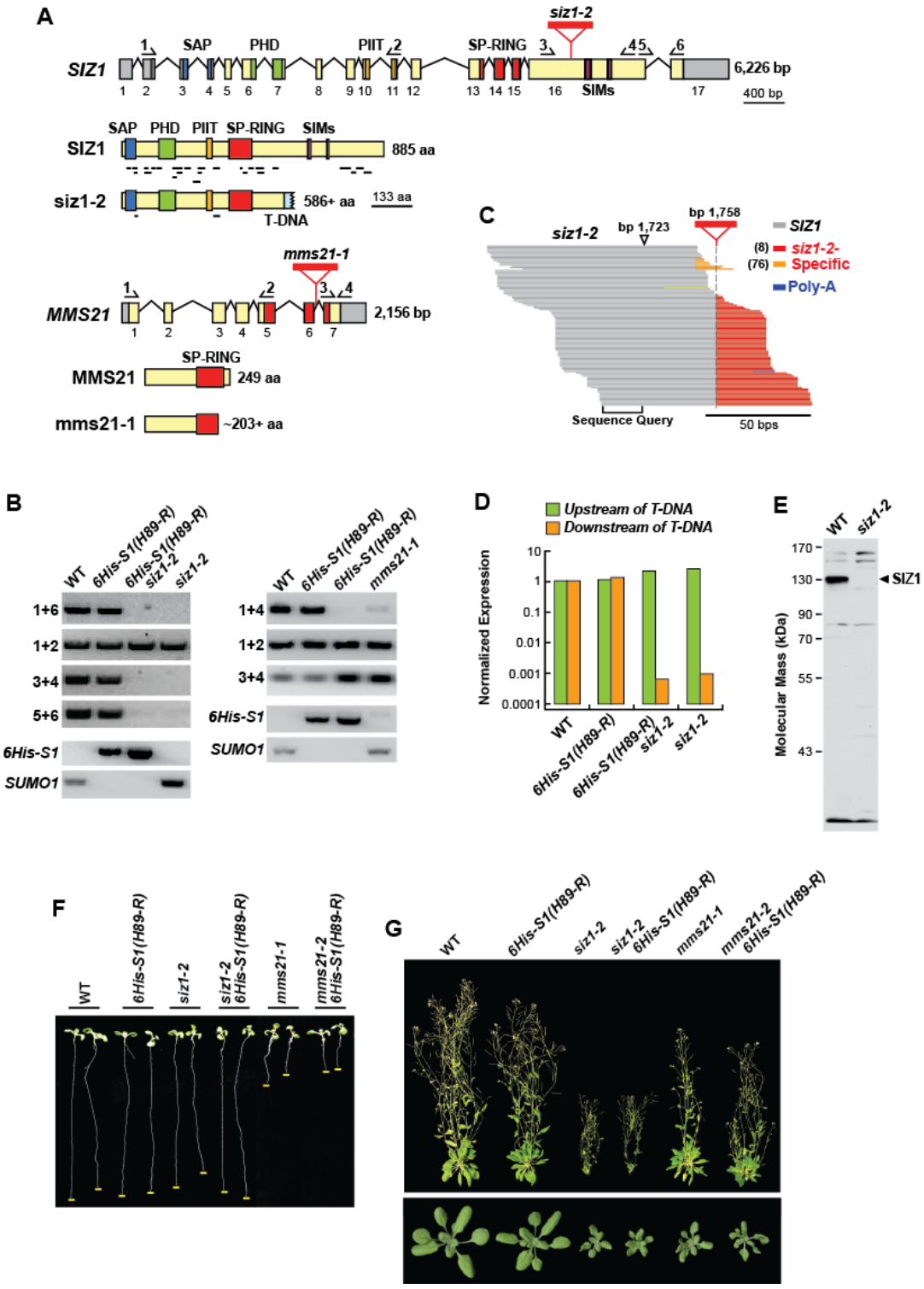
RESULTS

Development of the *siz1-2* and *mms21-1* SUMO Conjugate Purification Lines

As a first step toward defining the SUMOylation substrates of SIZ1 and MMS21, I introgressed the *6His-SUMO1(H89-R) sumo1-1 sumo2-1* purification background into the previously described *siz1-2* and *mms21-1* mutants generated by T-DNA insertional mutagenesis. The T-DNA in *siz1-2* was predicted previously to interrupt the 16th exon at nucleotide 1,753 downstream of the ATG start codon [53]. If transcribed, the resulting mRNA would then encode for a SIZ1 polypeptide that includes the SAP, PHD, PIIT, and SP-RING domains required for full SIZ1 activity *in vitro* and *in vivo* [54, 74], but missing the pair of predicted C-terminal SIM motifs (Figure 3-1A,B). RT-PCR analysis of homozygous *siz1-2* seedling amplified transcripts 5' to the predicted insertion site, but none 3', roughly supporting the position of the T-DNA. However, fine mapping by mRNA sequencing (RNA-seq) revealed that the *siz1-2* insertion site is actually 35-bp downstream at nucleotide 1,758, which would theoretically generate a transcript encoding the first 586 residues of the SIZ1 polypeptide, beyond which continued a stretch of in-frame codons derived from the T-DNA sequence (Figure 3-1C). The most common transcript contained an in-frame T-DNA sequence with at least 12 additional amino-acid codons, suggesting that a sizable non-SIZ1 sequence is potentially appended to the siz1-2 polypeptide during translation. As quantified by real time-PCR, this 5' *siz1-2* transcript accumulates to levels comparable to the full-length *SIZ1* transcript in wild type, suggesting that a partially functional SIZ1 could be synthesized (Figure 3-1D). However, we failed to identify the corresponding siz1-2 polypeptide (Figure 3-1E). Whereas the full-length SIZ1 protein was easily detected by

Figure 3-1. Genetic and phenotypic description of the *siz1* and *mms21* mutants. (A) Organization of the SIZ1 and MMS21 genes and proteins. The SAP, PHD, PIIT, SP-RING, and SIM domains are highlighted in blue, green, brown, red, and purple, respectively. The untranslated regions (UTRs) and introns are shown as gray boxes and lines, respectively. The red triangles show the positions of the T-DNA insertions. The lines underneath the wild-type and mutant SIZ1 proteins locate the peptides identified during our MS analysis of SUMO conjugates. The arrows locate the primers used for RT- PCR in (B). The amino-acid-sequence lengths of the siz1-2 and mms21-1 polypeptides that match their wild-type counterparts are shown. **(B)** RT-PCR analysis of transcripts derived from plants containing wild-type (WT) and mutant versions of *SIZ1* and *MMS21*. The bottom panels reflect genomic PCR analyses demonstrating the presence of the *6His-SUMO1(H89-R)* transgene and absence of an intact *SUMO1* gene in the respective genotypes. **(C)** Alignment of 111 transcripts generated by RNA-seq around the predicted T-DNA insertion site from *siz1-2* plants. *SIZ1* and T-DNA-related sequences are colored in gray and red/orange, respectively. A poly-A tract is indicated in blue. The most common junction between *SIZ1* sequence and predicted T-DNA sequence identifies the T-DNA insertion site at 1,758 bp from the ATG translation start codon. The previously reported insertion site at 1,723 is also shown. **(D)** Quantification of *SIZ1* transcript levels in WT and *siz1-2* plants show in panel (C), using primers that probed the *SIZ1* locus either upstream or downstream of the T-DNA insertion. The values were normalized to those of ACT2, and represented as a ratio to WT. **(E)** Immunoblot detection of SIZ1 protein in 8-d-old unstressed WT and *siz1-2* seedlings. The membrane was probed with anti-SIZ1 antibodies. **(F)** Representative 8-d-old WT, *siz1-2* and *mms21-1* seedlings without or with the SUMO conjugate purification background (*6His-S1(H89-R) sumo1-1 sumo2-1*). Root tips are highlighted by the

yellow line. **(G)** Representative WT, *siz1-2* and *mms21-1* plants without or with the SUMO conjugate purification background (*6His-S1(H89-R) sumo1-1 sumo2-1*) grown for 20 d (bottom) and 40 d (top) in a LD photoperiod.

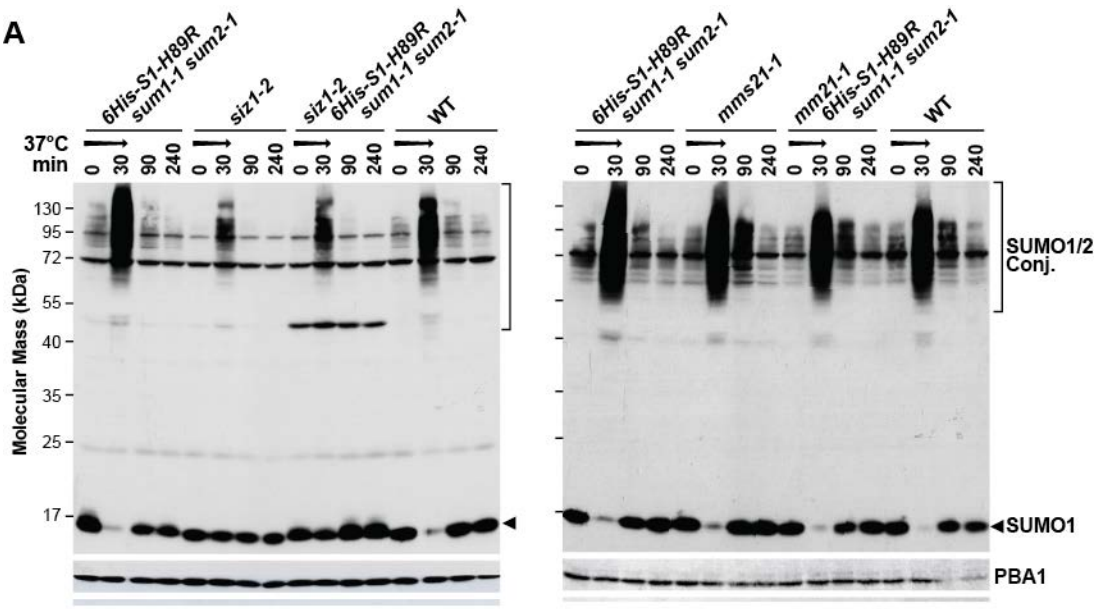

A
siz1-2
SIZ1
SAP
PHD
PIIT
SP-RING
SIMs
6,226 bp
400 bp
885 aa
586+ aa
133 aa
T-DNA
mms21-1
MMS21
2,156 bp
249 aa
~203+ aa
B
WT
6His-S1(H89-R)
siz1-2
mms21-1
1+6
1+2
3+4
5+6
6His-S1
SUMO1
1+4
C
bp 1,723
bp 1,758
(8)
(76)
siz1-2-
Specific
Poly-A
Sequence Query
50 bps
D
Upstream of T-DNA
Downstream of T-DNA
Normalized Expression
E
Molecular Mass (kDa)
170
130
90
70
55
43
F
G

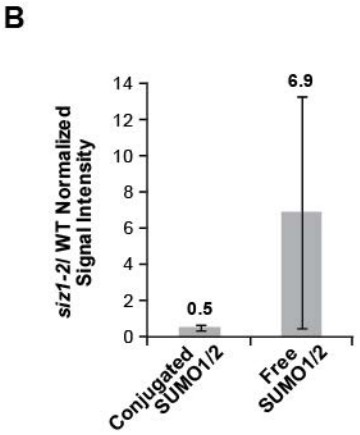
immunoblotting wild-type seedling extracts with anti-SIZ1 antibodies, no smaller species of the expected mass of ~67 kDa could be assigned to the siz1-2 truncation in mutant seedlings. Taken together, the *siz1-2* allele should be considered as an attenuated mutant (and not null) whose protein product might still bind the E2-SUMO donor and possibly its substrates, and/or direct non-specific SUMOylation, but at substantially reduced levels.

The T-DNA sequence in the *mms21-1* allele (also called *hyp2-2* [56]) locates to the 6th intron separating codons for the SP-RING domain [55]. RT-PCR analysis of homozygous plants found transcripts both upstream and downstream of the T-DNA insertion site but failed to detect transcripts encompassing the full *MMS21* coding sequence (Figure3-1B,D). As the resulting polypeptide would be missing part of the SP-RING domain essential to E2-SUMO binding and subsequent transfer, we considered it likely that *mms21-1* is a functional null allele.

Based on the widely separated chromosomal positions of *SUMO1, SUMO2* and *MMS21*, we predicted that creating the homozygous *mms21-1 6His-SUMO1(H89-R) sumo1-1 sumo2-1* lines would be straightforward, and indeed generating the quadruple homozygous mutants fit Mendelian segregation ratios in self crosses. (At present, we do not know the insertion position of the *6His-SUMO1(H89-R)* transgene). However, generating the purification line harboring *siz1-2* was predicted to be more difficult given that the *SIZ1* and *SUMO2* loci are physically linked, being only 1.9-Mbp away from each other on chromosome 5. Here, we screened over 90 offspring from a quadruple heterozygous line to identified 9 individuals that recombined between the *siz1-2* and *sumo2-1* alleles, one of which was homozygous for *sumo1-2* and *sumo2-1*, heterozygous for *siz1-2* and contained the *6His-SUMO1(H89-R)* transgene. Subsequent selfing of this individual generated a line that was homozygous at all four positions (*siz1-2 6His-SUMO1(H89-R) sumo1-1 sumo2-1*), which was confirmed by genomic PCR of its progeny.

Figure 3-2. SUMOylation profile of wild-type (WT), *siz1-2*, and *mms21-1* plants before and after heat stress. (A) Immunoblot analysis of 8-d-old seedlings heat stressed for 30 min at 37°C and collected at the indicated times. The germplasms included the *siz1-2*, and *mms21-1* mutations by themselves or combined with the SUMO-conjugate purification background (*6His-S1(H89-R) sumo1-1 sumo2-1*). The membrane was probed with either anti-SUMO1/2 or anti-PBA1 antibodies (loading control). High molecular mass SUMO conjugates and free SUMO are indicated by the brackets and arrowheads, respectively. **(B)** Quantification of SUMO conjugates in WT versus *siz1-2* seedlings by densitometric scans of the immunoblots of SUMO1/2 in (A). Immunoblot signals generated for SUMO conjugates and free SUMO from heat-shocked (30 min at 37°C) seedlings were visualized using IRDye 800CW or IRDye 680RD goat anti-rabbit secondary antibodies and normalized to those for PBA1. Each bar represents the average of three biological replicates (±SD). The average ratios of WT versus siz1-2 for conjugated and free SUMO ratio are indicated above each bar.


A
6His-S1-H89R sum1-1 sum2-1
siz1-2
siz1-2 6His-S1-H89R sum1-1 sum2-1
WT
6His-S1-H89R sum1-1 sum2-1
mms21-1
mms21-1 6His-S1-H89R sum1-1 sum2-1
WT
37°C
min
0 30 90 240
Molecular Mass (kDa)
130
95
72
55
40
35
25
17
SUMO1/2 Conj.
SUMO1
PBA1

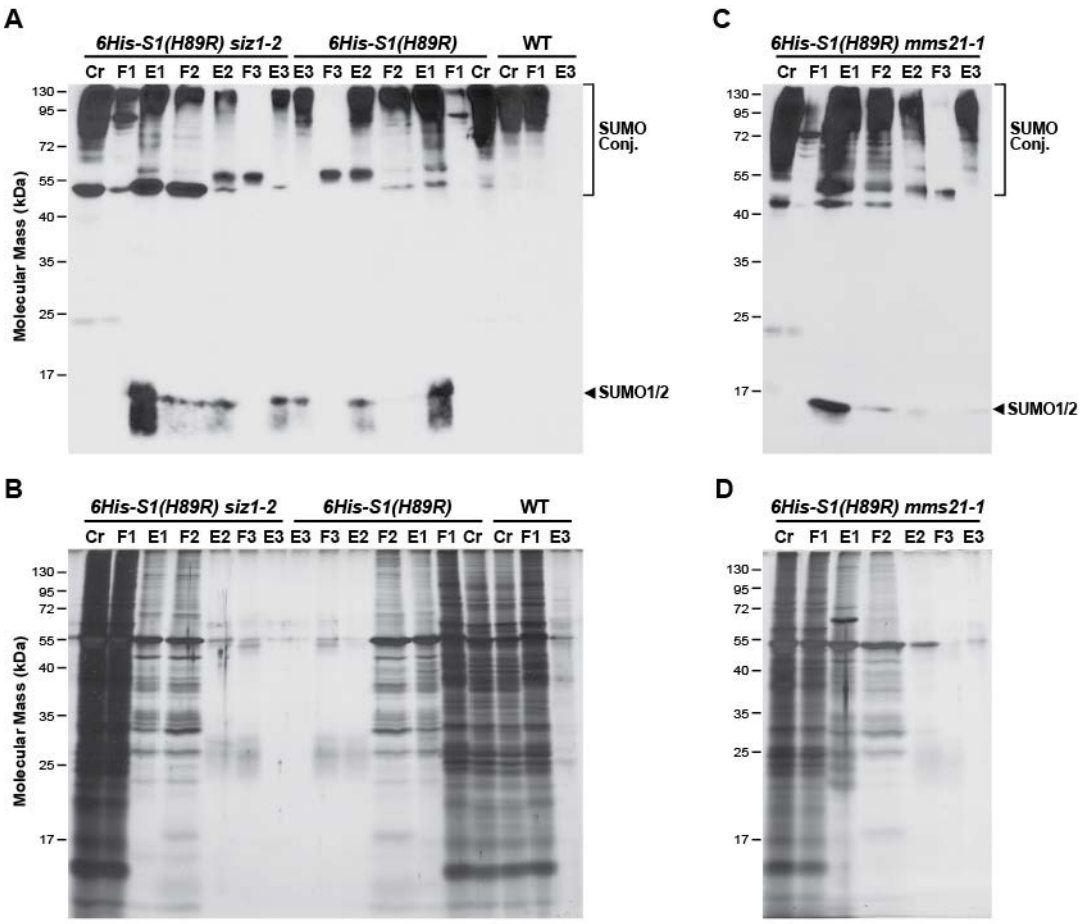

B
siz1-2/ WT Normalized Signal Intensity
14
12
10
8
6
4
2
0
0.5
6.9
Conjugated SUMO1/2
Free SUMO1/2

As shown in Figure 1F,G, the introgressed lines retained the phenotypes of the *siz1-2* and *mms21-1* parents, which included a smaller leaves and dwarfed stature of *siz1-2* plants, and short roots, elongated leaves, ands fascinated stems for *mms21-1* plants [21, 53, 55, 56]. When subjected to heat stress (30 min at 37°C), wild-type Arabidopsis seedlings rapidly accumulate high molecular mass SUMO1/2 conjugates (50 to >170 kDa) with a commensurate depletion of the free SUMO1/2 pool, which is subsequently reversed upon return to 24°C growth temperatures (Figure 3-2A; [21, 28]). This accumulation pattern was preserved in *mms21-1* seedlings but noticeably dampened in the *siz1-2* seedlings (+/- *6His-SUMO1(H89-R) sumo1-1 sumo2-1*) as previously reported [21]. Quantitative immunoblotting revealed a two-fold decrease in SUMO conjugate levels concomitant with a seven-fold retention of the free SUMO pool in *siz1-2* seedlings after the heat stress (Figure 3-2B).

Purification and MS Analysis of SUMO Conjugates in *siz1-2* and *mms21-1* Seedlings.

Using the purification strategy developed by Miller et al. (2010), I generated SUMO conjugate-enriched fractions based on the *6His-SUMO1(H89-R) sumo1-1 sumo2-1* background from 8-d-old wild-type, *siz1-2*, and *mms21-1* seedlings either before or 30 min after a 37°C heat stress. Both the three-step affinity protocol (Nickel- nitrilotriacetic acid (Ni-NTA), anti-SUMO1 antibody, and Ni-NTA) coupled with the inclusion of strong denaturants provided stringent purification with minimal background as can be seen by the absence of SUMO conjugates and protein contaminants when wild-type seedlings without the *6His-SUMO1(H89-R)* transgene were used instead (Figure 3-3). After trypsinization, the protein pools were separated by reverse-phase liquid chromatography and sequences by tandem MS using a LTQ Orbitrap Velos or Q-Exactive mass spectrometers in the electrospray ionization (ESI) mode. The MS2 spectra were

Figure 3-3. Affinity purification of SUMOylated proteins from *6His-S1(H89-R) sumo1-1 sumo2-1* seedlings either wild-type or mutant for the SUMO E3 ligases SIZ1 and MMS21. SUMO conjugates were enriched by the 3-step affinity method based on the 6His-SUMO1(H89-R) variant from 8-day old seedlings heat-stressed for 30 min at 37°C. The three flow-through (FT1, FT2, and FT3) and eluate fractions (E1, E2, and E3) were subjected to SDS-PAGE. Cr, clarified crude extract. The SDS-PAGE loads for each background were proportionally adjusted to allow direct comparison between the purification steps. As a control, wild-type (WT) seedlings were subjected to the same 3-step purification. **(A,C)** Immunoblot analysis of the *siz1-2* (A) and *mms21-1* fractions (B) with anti-SUMO1 antibodies. The higher molecular weight SUMO conjugates and free SUMO band are indicated by the brackets and arrowheads, respectively. **(B,D)** Protein profiles of the *siz1-2* (B) and *mms21-1* fractions (D) samples as detected by silver staining.


A
6His-S1(H89R) siz1-2
6His-S1(H89R)
WT
Cr F1 E1 F2 E2 F3 E3 E3 F3 E2 F2 E1 F1 Cr Cr F1 E3
130
95
72
55
40
35
25
17
Molecular Mass (kDa)
SUMO Conj.
SUMO1/2
C
6His-S1(H89R) mms21-1
Cr F1 E1 F2 E2 F3 E3
130
95
72
55
40
35
25
17
SUMO Conj.
SUMO1/2
B
6His-S1(H89R) siz1-2
6His-S1(H89R)
WT
Cr F1 E1 F2 E2 F3 E3 E3 F3 E2 F2 E1 F1 Cr Cr F1 E3
130
95
72
55
40
35
25
17
Molecular Mass (kDa)
D
6His-S1(H89R) mms21-1
Cr F1 E1 F2 E2 F3 E3
130
95
72
55
40
35
25
17

then searched against the Arabidopsis proteome database available from The Arabidopsis Information Resource (TAIR version 10) to identify possible SUMO conjugates using a ≤1% false discovery rate cut-off. Those proteins routinely identified in wild-plants were then subtracted from this list as likely contaminants.

For the unstressed seedlings (wild type, *6His-SUMO1(H89-R)* and *siz1-2 6His-SUMO1(H89-R)* and *mms21-1 6His-SUMO1(H89-R)*, three biological replicates were generated. For heat-stressed samples, five biological replicates were created with *siz1-2 6His-SUMO1(H89-R)* and *6His-SUMO1(H89-R)* samples, and two with *mms21-1 6His-SUMO1(H89-R)* samples; all but two of which were supported by technical replicates. In general, strong overlaps in protein identifications were seen among replicates, especially for the heat-stressed datasets (Figure 3-4A,B). Subsequent quantification of SUMO conjugate levels using distributed Normalized Spectral Abundance Factor (dNSAF) values generated by Morpheus in combination with Morpheus Spectral Counter [161] also show high correlations among both technical and biological replicates (R^2 values of 0.97-0.99 and 0.5-0.7, respectively; Figure 3-4C,D)

When all the datasets were combined, I identified over 1,400 potential SUMO targets with high statistical support (≤1% FDR). For improved confidence, a collection of 1,060 targets detected by at least two peptide spectral matches (PSMs) were used for subsequent dataset comparisons. This conservative list represents a three-fold increase in SUMO substrates over that previously described by Miller et al. (2010), and is likely derived from the use of more advance Orbitrap mass spectrometers in this study. Remarkably, over 85% of the SUMO targets isolated by Miller et al (2010) were identified here, thus providing confidence in the methodology (Figure 3-5D).

Figure 3-4. Reproducibility between technical and biological replicates for wild-type, *siz1-2* and *mms21-1*. (A & B) Overlap of target proteins between biological replicates for wild-type (WT), *siz1-2* and *mms21-1* identified from heat-shocked (HS) (A) or unstressed (B) seedlings. **(C & D)** Reproducibility of label-free quantification values (dNSAF) between technical (C) and biological (D) replicates for wild-type (WT), *siz1-2* and *mms21-1* after heat shock (+HS). The gray dashed line represents the line of best fit with the slope of the line noted on the graph. The respective R^2 values are shown.

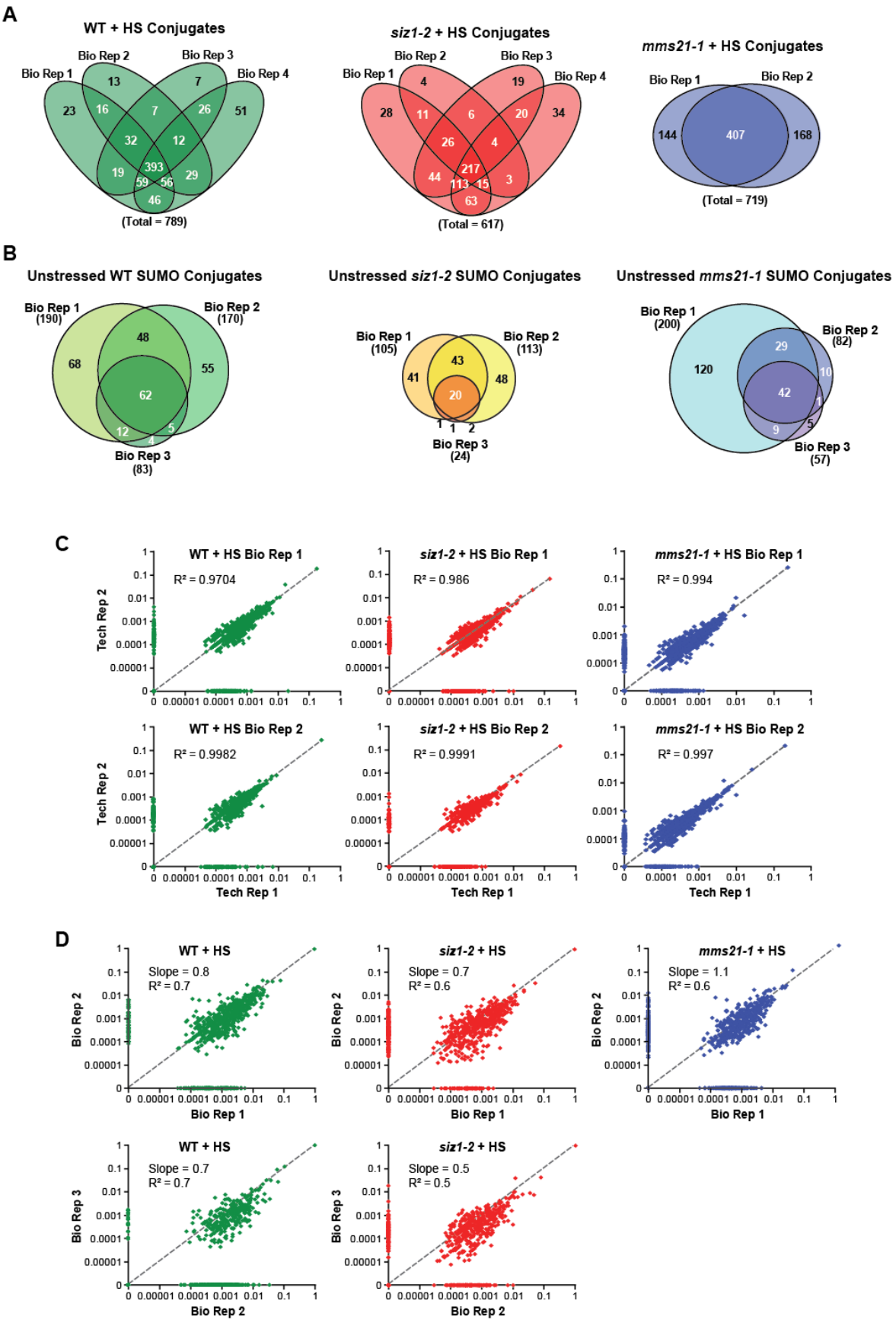
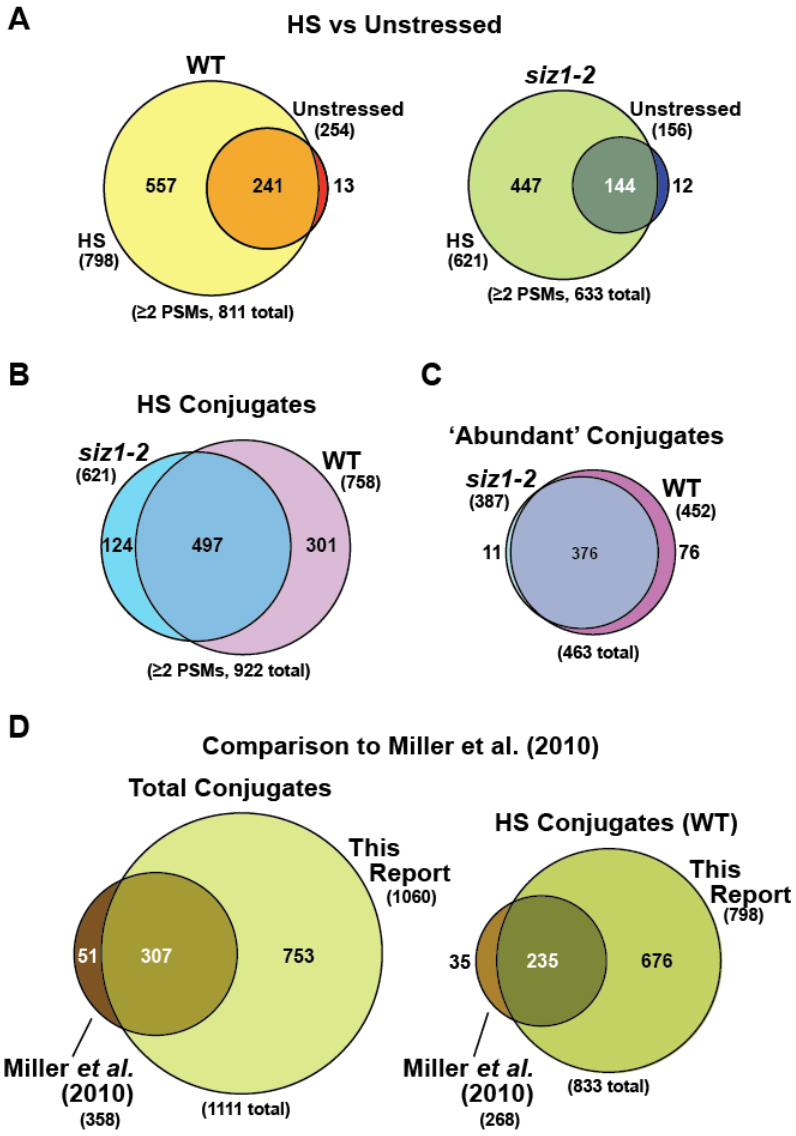

A
WT + HS Conjugates
Bio Rep 1
Bio Rep 2
Bio Rep 3
Bio Rep 4
23
16
13
7
7
26
51
32
12
19
393
29
59
56
46
(Total = 789)
siz1-2 + HS Conjugates
Bio Rep 1
Bio Rep 2
Bio Rep 3
Bio Rep 4
28
11
4
6
19
20
34
26
4
44
217
113
15
3
63
(Total = 617)
mms21-1 + HS Conjugates
Bio Rep 1
Bio Rep 2
144
407
168
(Total = 719)
B
Unstressed WT SUMO Conjugates
Bio Rep 1
(190)
Bio Rep 2
(170)
Bio Rep 3
(83)
68
48
55
62
12
5
4
Unstressed siz1-2 SUMO Conjugates
Bio Rep 1
(105)
Bio Rep 2
(113)
Bio Rep 3
(24)
41
43
48
20
1
1
2
Unstressed mms21-1 SUMO Conjugates
Bio Rep 1
(200)
Bio Rep 2
(82)
Bio Rep 3
(57)
120
29
10
42
1
9
5
C
WT + HS Bio Rep 1
R² = 0.9704
siz1-2 + HS Bio Rep 1
R² = 0.986
mms21-1 + HS Bio Rep 1
R² = 0.994
WT + HS Bio Rep 2
R² = 0.9982
siz1-2 + HS Bio Rep 2
R² = 0.9991
mms21-1 + HS Bio Rep 2
R² = 0.997
Tech Rep 2
Tech Rep 1
D
WT + HS
Slope = 0.8
R² = 0.7
siz1-2 + HS
Slope = 0.7
R² = 0.6
mms21-1 + HS
Slope = 1.1
R² = 0.6
WT + HS
Slope = 0.7
R² = 0.7
siz1-2 + HS
Slope = 0.5
R² = 0.5
Bio Rep 2
Bio Rep 1
Bio Rep 3
Bio Rep 2

Figure 3-5. Venn diagrams showing the distribution of SUMOylated proteins purified from wild-type (WT) and *siz1-2* seedlings before and after heat stress. Eight-d-old green seedlings were either kept at 22°C or heat stressed for 30 min at 37°C before tissue collection and SUMO conjugate purification. **(A)** Overlaps of all SUMO conjugates detected in the wild-type (WT) or *siz1-2* background exposed to heat-stress (HS) versus control conditions (unstressed). **(B)** Comparison of the total collection of SUMO conjugates in WT or *siz1-2* seedlings after heat stress. **(C)** Comparison of the abundant SUMO conjugates in WT or *siz1-2* seedlings after heat stress. Abundant conjugates refer to those detected in 3 or more biological replicates in either WT or *siz1-2* seedlings. **(D)** Comparisons of the SUMO conjugates identified in this study with those identified by Miller et al. (2010). The left diagram includes all SUMO conjugates detected in unstressed samples as well as samples treated with heat shock and hydrogen peroxide. The right panel includes SUMO conjugates that were only detected in heat-stressed samples.


A
HS vs Unstressed
WT
Unstressed
(254)
557
241
13
HS
(798)
(≥2 PSMs, 811 total)
siz1-2
Unstressed
(156)
447
144
12
HS
(621)
(≥2 PSMs, 633 total)
B
HS Conjugates
siz1-2
(621)
WT
(758)
124
497
301
(≥2 PSMs, 922 total)
C
'Abundant' Conjugates
siz1-2
(387)
WT
(452)
11
376
76
(463 total)
D
Comparison to Miller et al. (2010)
Total Conjugates
This
Report
(1060)
51
307
753
Miller et al.
(2010)
(358)
(1111 total)
HS Conjugates (WT)
This
Report
(798)
35
235
676
Miller et al.
(2010)
(268)
(833 total)

Identification of SIZ1 SUMOylation Targets

As expected based on immunoblot detection of SUMO conjugates (see Figure 3-2), unstressed Arabidopsis seedlings contain a small pool of conjugates that rises rapidly during heat stress. For example, whereas only 254 and 156 substrates were detected in wild-type and *siz1-2* seedlings grown at 24°C, respectively, these numbers rose to 798 and 621 targets in seedlings exposed to a 30 min heat stress (Figure 3-2A). Unlike previous reports [46], this rise coincided with a substantial increase in new substrates. One likely reason for this discrepancy is that the deeper collection of proteins found here includes low abundance targets that were previously below detection, especially in unstressed conditions.

Direct comparisons of the conjugate lists from wild-type and *siz1-2* seedlings after heat stress revealed a substantial loss of conjugates in the *siz1-2* background, suggesting that a large collection of these proteins are SIZ1 substrates. When considering all conjugates found in heat-stressed samples (wild type and *siz1-2* with ≥2 PSMs), 301 proteins were absent and only 124 conjugates appeared in the *siz1*-2 mutant from a total list of 922 conjugates (Figure 3-5B). However, when comparing just the 463 conjugates considered to be 'abundant' based on their detection in at least 3 of the 5 biological replicates for either wild type or siz1-2, 76 conjugates were missing from the siz1-2 background with just 11 unique (Figure 3-5C).

To better assess the changes in SUMOylation in *siz1-2* versus wild-type plants, I quantified the abundance of individual substrates using dNSAF values, which are calculated from PSM counts and adjusted based on protein length and shared peptides [161, 162]. To control for variations among the samples, the dNSAF values for all conjugates in each biological replicate were normalized to those obtained for SUMO1, as it was demonstrated previously that

Figure 3-6. Changes in the SUMO conjugate accumulation patterns during heat stress in *siz1-2* versus wild-type seedlings. SUMO conjugates detected with at least two PSM per biological replicate were quantified based on their dNSAF values, which were then normalized based the dNSAF values obtained from SUMO1. **(A)** Average normalized dNSAF values of 922 SUMOylated proteins in *siz1-2* versus wild-type (WT) seedlings (see Figure 3-5B). Each data point represents the average of five biological replicates. Light gray points are conjugates considered to be 'rare' by their detection in less than 3 biological replicates in both backgrounds (*siz1-2* and/or WT). Dark gray points are conjugates considered to be 'abundant' by their detection in 3 or more biological replicates in either background (*siz1-2* and/or WT). Proteins with a significant decrease or increase in SUMOylation in the *siz1-2* mutant as compared to WT ($p \leq 0.05$) are highlighted red and blue, respectively. SUMO targets identified in all WT biological replicates and never or only once in the *siz1-2* mutant (WT > siz1-2) are in orange. Proteins detected in all *siz1-2* biological replicates and never or only once in the WT (*siz1-2* > WT) are in green. The dashed line represents the theoretically situation where conjugate abundance in WT and *siz1-2* is equal. Note that two dNSAF values are assigned to SUMO1 by Morpheus Spectra Counter. **(B)** Volcano plot of the *p*-value for individual SUMO conjugates versus the log2 fold change in WT versus *siz1-2* seedlings. Missing values were imputed for each biological replicate. The colour scheme is the same as in (A). Horizontal dashed lines highlights a *p* -value of 0.05. The vertical dashed lines highlight a 4-fold increase or decrease. **(C)** Expanded view of panel (B) highlighting the proteins with a significant reduction of SUMOylation in the *siz1-2* mutant. Proteins of interest are indicated when possible. Notable proteins are indicated in (A-C).

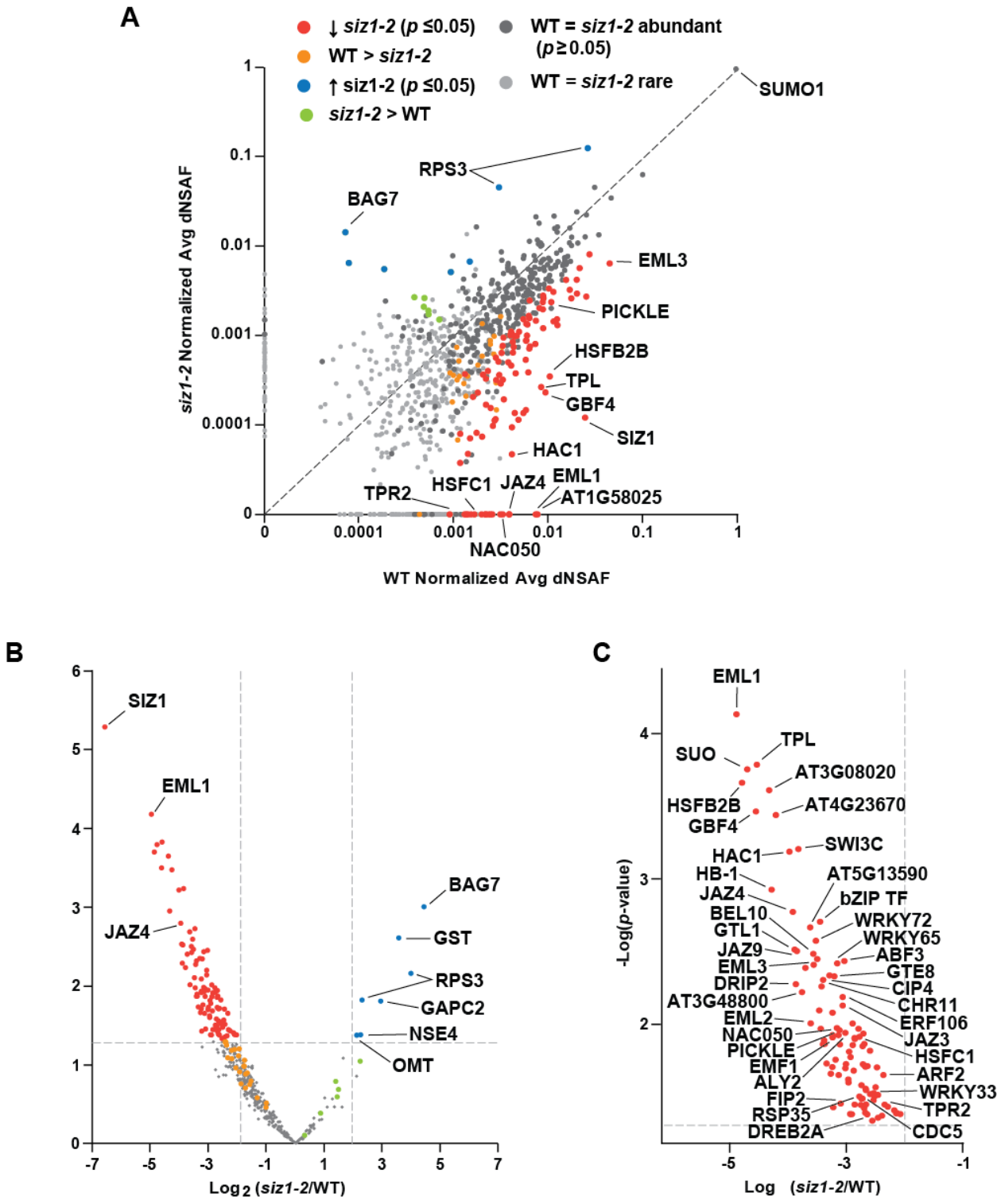
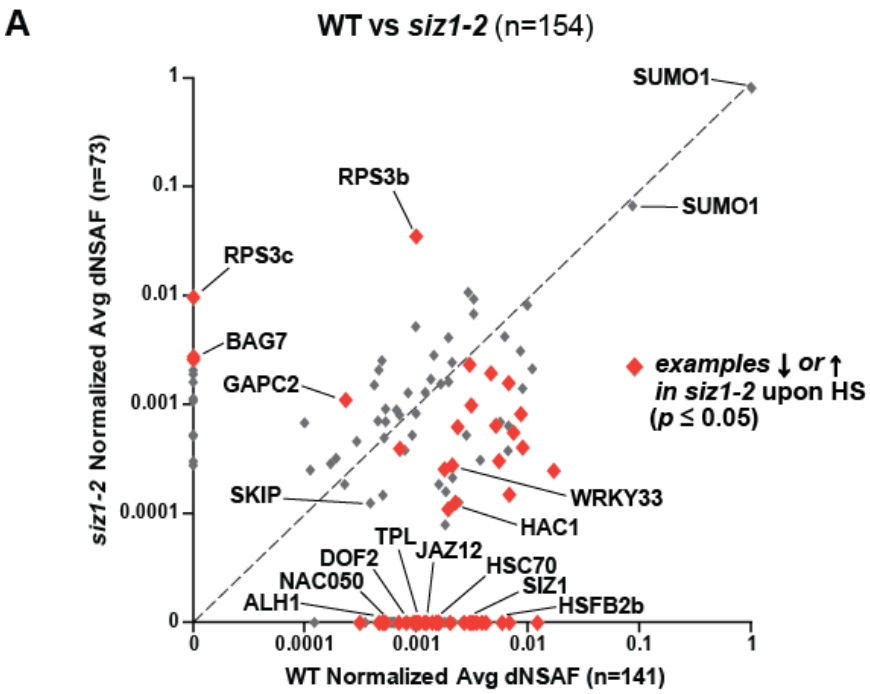
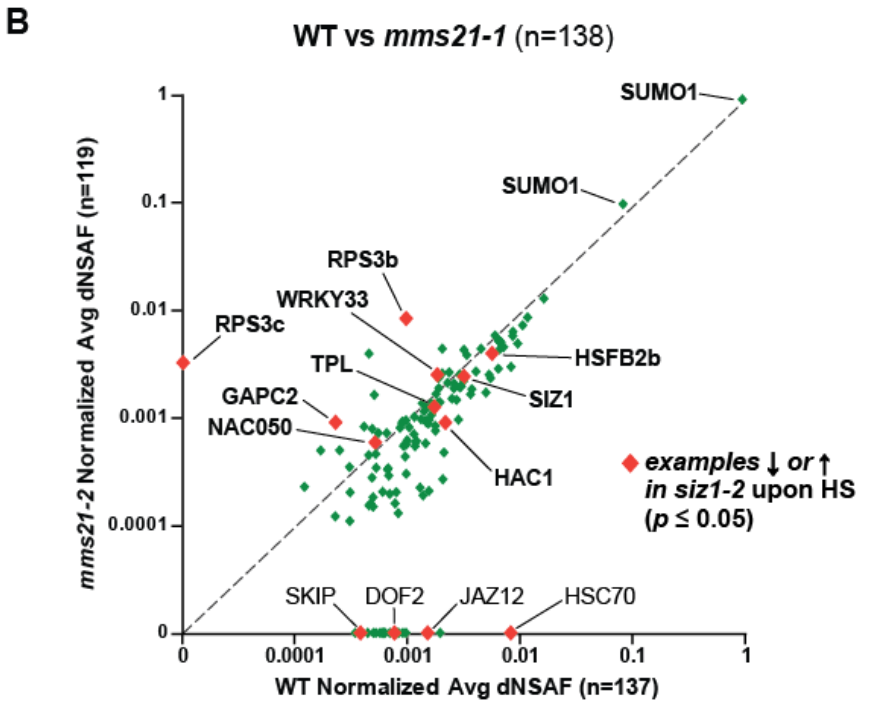

A
↓ siz1-2 (p ≤0.05)
WT > siz1-2
↑ siz1-2 (p ≤0.05)
siz1-2 > WT
WT = siz1-2 abundant (p ≥0.05)
WT = siz1-2 rare
SUMO1
RPS3
BAG7
EML3
PICKLE
HSFB2B
TPL
GBF4
SIZ1
HAC1
EML1
JAZ4
HSFC1
TPR2
AT1G58025
NAC050
siz1-2 Normalized Avg dNSAF
WT Normalized Avg dNSAF
B
SIZ1
EML1
JAZ4
BAG7
GST
RPS3
GAPC2
NSE4
OMT
Log2 (siz1-2/WT)
C
EML1
TPL
SUO
AT3G08020
HSFB2B
AT4G23670
GBF4
SWI3C
HAC1
HB-1
AT5G13590
JAZ4
bZIP TF
BEL10
WRKY72
GTL1
WRKY65
JAZ9
ABF3
EML3
GTE8
DRIP2
CIP4
AT3G48800
CHR11
EML2
ERF106
NAC050
JAZ3
PICKLE
HSFC1
EMF1
ARF2
ALY2
WRKY33
FIP2
TPR2
RSP35
CDC5
DREB2A
-Log(p-value)
Log (siz1-2/WT)

Figure 3-7. Comparison of SUMO conjugate abundance under unstressed conditions in *siz1-2* and *mms21-1* versus wild-type seedlings. (A) Plot of average normalized dNSAF values of SUMOylated proteins from wild-type (WT) and *siz1-2* under unstressed conditions. Notable proteins are identified on the graph. Proteins with a significant decrease in SUMOylation in the *siz1-2* mutant during heat shock (HS) are highlighted in red. The dashed line represents the line of equilibrium. **(B)** Plot of average normalized dNSAF values of SUMOylated proteins from WT and *mms21-1* under unstressed conditions. Notable proteins are identified on the graph. Proteins with a significant decrease in SUMOylation in the *siz1-2* mutant are highlighted in red. The dashed line represents the line of equilibrium.


A
WT vs siz1-2 (n=154)
siz1-2 Normalized Avg dNSAF (n=73)
WT Normalized Avg dNSAF (n=141)
SUMO1
SUMO1
RPS3b
RPS3c
BAG7
GAPC2
SKIP
WRKY33
HAC1
TPL
JAZ12
DOF2
HSC70
NAC050
SIZ1
ALH1
HSFB2b
examples ↓ or ↑ in siz1-2 upon HS (p ≤ 0.05)
1
0.1
0.01
0.001
0.0001
0
0.0001
0.001
0.01
0.1
1


B
WT vs mms21-1 (n=138)
mms21-2 Normalized Avg dNSAF (n=119)
WT Normalized Avg dNSAF (n=137)
SUMO1
SUMO1
RPS3b
WRKY33
RPS3c
HSFB2b
TPL
SIZ1
GAPC2
NAC050
HAC1
SKIP
DOF2
JAZ12
HSC70
examples ↓ or ↑ in siz1-2 upon HS (p ≤ 0.05)
1
0.1
0.01
0.001
0.0001
0
0.0001
0.001
0.01
0.1
1

SUMO remained unchanged in abundance during this short heat stress and equally purified whether it was in its free or conjugated forms [33, 46].

Comparison of SUMO conjugate abundances by dNSAF in wild-type versus *siz1-2* revealed large deviations from a 1:1 ratio with many proteins underrepresented or absent in the *siz1-2* background both before and after the heat stress (Figure 3-6A; Figure 3-8A). To further analyze the data, statistical differences in protein abundance were assessed by processing the normalized dNSAF values with the Linear Models for Microarray Data (LIMMA) statistical algorithm that calculates moderate *p*-values for each target. To limit the extent of imputations for proteins with null values, we focused only on the 463 SUMO substrates considered to be abundant. Overall, 112 proteins were found to have a significant (moderate p-value ≤ 0.05) change in SUMOylation in the *siz1-2* mutant compared to the WT (Figure 3-6B). Of these, the SUMOylated forms of 18 proteins were not detected and 87 conjugates were significantly decreased in abundance (Figure 3-6 A,C), while 7 proteins were increases in abundance (Figure 3-6 A,B). Of note is another collection of conjugates that were almost always detected in one genotype while often undetected among biological replicates for the other genotype, but were calculated by LIMMA to be above the significance threshold (p-value ≥ 0.05). These could represent a second set of targets whose SUMOylation status might be impacted by the loss of SIZ1, but their scoring was challenged by low MS detection.

Further comparisons of SUMO conjugate abundances by dNSAF highlighted that the conjugate abundances of the 105 proteins found to have reduced conjugation in *siz1-2* after heat shock were already less prevalent in the mutant before heat stress (Figure 3-7A). This observation suggests that the drop in SUMO conjugates seen in the *siz1-2* mutant during

immunoblot analysis (see Figure 3-2A) is due to a general reduction in SUMOylation of targets, and is not solely caused by a decline in heat shock-induced modifications.

SIZ1-Influenced Conjugates are Enriched in Stress Regulators

Further characterization of the 112 conjugates whose SUMOylation was significantly affected by a loss of SIZ1 upon heat-shock highlighted a distinct functional role of these SIZ1-influenced targets compared to the whole SUMOylome. The SUMO substrates with significant changes in SUMOylation in *siz1-2* were predicted to be mainly nuclear-localized or nuclear- and cytoplasmic-localized proteins based on the Panther Gene Ontology (GO) database [163] (Figure 3-8A). Consistent with their localization, GO functional enrichments using the Database for Annotation, Visualization and Integrated Discovery (DAVID) [164] revealed that these targets are substantially enriched in factors involved in transcription, including negative regulation of gene expression, and chromatin remodeling and modification, as well as proteins involved in development and responses to biotic and abiotic stresses (Figure 3-8B).

Most of the functional processes enriched in the SIZ1-influenced conjugates are also enriched for in the whole SUMOylome, but at a lower fold enrichment. Surprisingly however, RNA-related processes, such as regulation of RNA splicing, RNA-directed DNA methylation and gene silencing, and mRNA processing, which have been found to show a significant increase in SUMOylation upon heat-shock in quantitative MS experiment [46], as well as DNA repair and DNA endoreduplication, two process thought to involve MMS21, were only significantly enriched in the whole SUMOylome (Figure 3-9A).

Furthermore, to determine if specific molecular processes are more prevalent in the SIZ1-influenced conjugates compared to all identified targets, the functional enrichment of the high

Figure 3-8. Localization and functional enrichments of SUMOylated proteins from wild-type and *siz1-2*. (A) Pie charts illustrating the known or predicted localization of individual SUMO conjugates. (top) All 921 SUMO conjugates detected in both *siz1-2* and wild-type (WT) plants +/- heat stress (HS) at 37°C for 30 min (see Figure 3-5B). (middle) The 454 abundant SUMO conjugates detection in 3 or more biological replicates in either background (*siz1-2* and/or WT) (see Figure 3-5C). (bottom) The 112 SUMO conjugates that are significantly decreased in siz1-2 versus WT plants after HS**. (B)** GO functional enrichment of all 921 SUMO conjugates (Green) that appear during HS, and the 112 SUMO conjugates that accumulate during HS and appear to be SIZ1-dependent (orange). **(C)** GO functional enrichment of the SIZ1-dependent SUMO targets as compared to the total collection of 921 SUMO targets identified after HS.

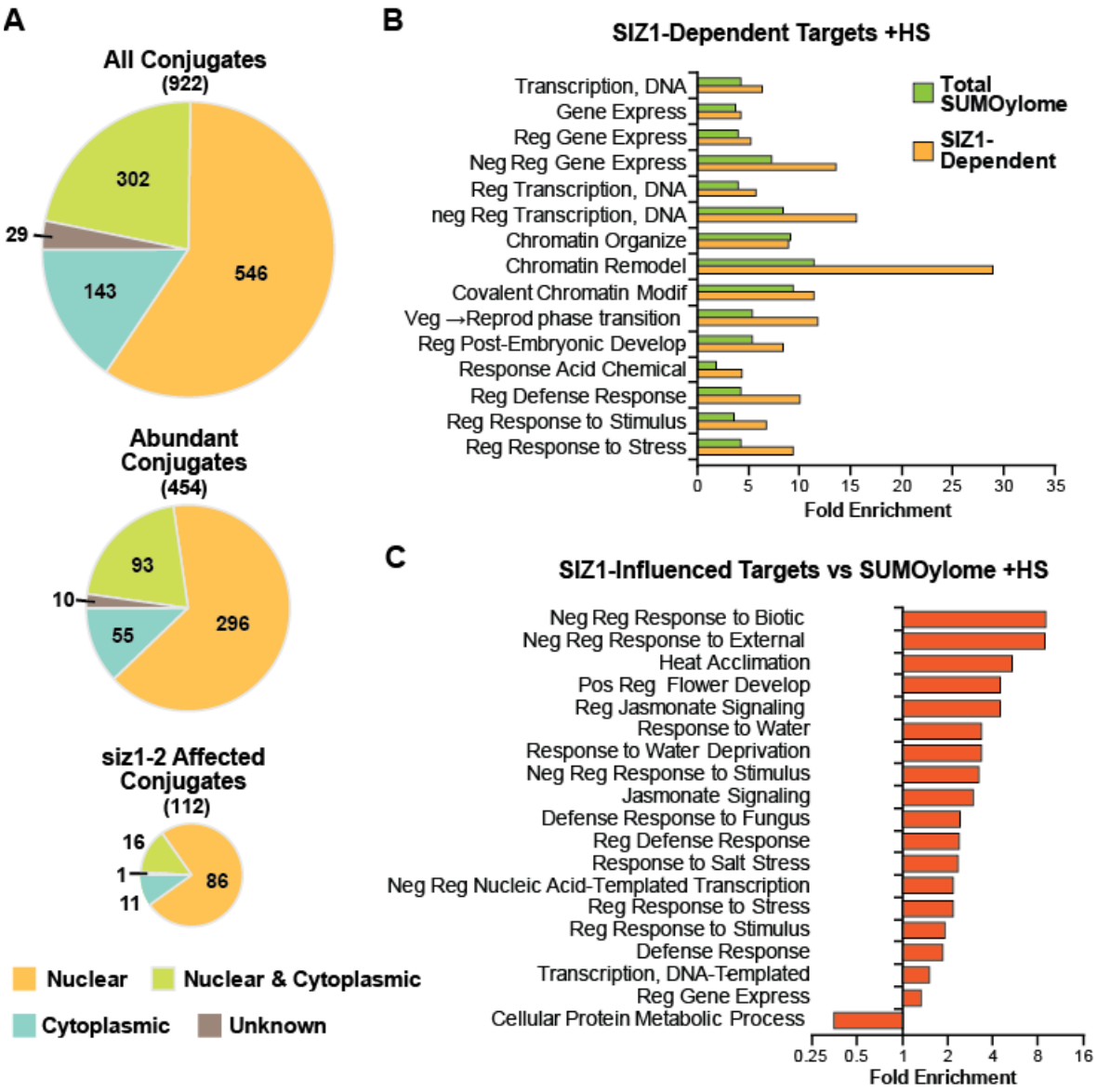
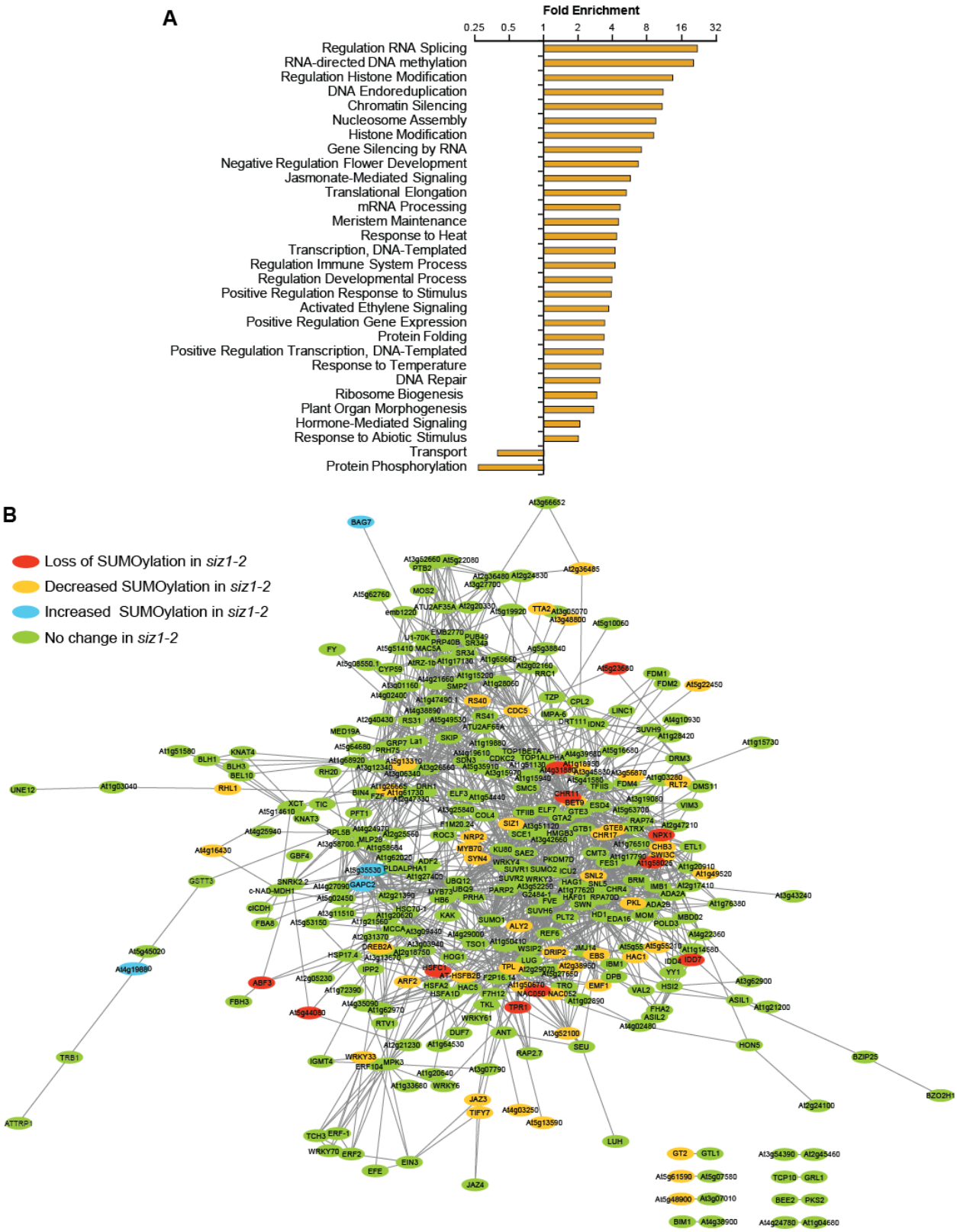

A
All Conjugates
(922)
302
29
143
546
Abundant
Conjugates
(454)
93
10
55
296
siz1-2 Affected
Conjugates
(112)
16
1
11
86
Nuclear
Nuclear & Cytoplasmic
Cytoplasmic
Unknown
B
SIZ1-Dependent Targets +HS
Transcription, DNA
Gene Express
Reg Gene Express
Neg Reg Gene Express
Reg Transcription, DNA
neg Reg Transcription, DNA
Chromatin Organize
Chromatin Remodel
Covalent Chromatin Modif
Veg →Reprod phase transition
Reg Post-Embryonic Develop
Response Acid Chemical
Reg Defense Response
Reg Response to Stimulus
Reg Response to Stress
Total SUMOylome
SIZ1-Dependent
0
5
10
15
20
25
30
35
Fold Enrichment
C
SIZ1-Influenced Targets vs SUMOylome +HS
Neg Reg Response to Biotic
Neg Reg Response to External
Heat Acclimation
Pos Reg Flower Develop
Reg Jasmonate Signaling
Response to Water
Response to Water Deprivation
Neg Reg Response to Stimulus
Jasmonate Signaling
Defense Response to Fungus
Reg Defense Response
Response to Salt Stress
Neg Reg Nucleic Acid-Templated Transcription
Reg Response to Stress
Reg Response to Stimulus
Defense Response
Transcription, DNA-Templated
Reg Gene Express
Cellular Protein Metabolic Process
0.25
0.5
1
2
4
8
16
Fold Enrichment

Figure 3-9. Functional enrichments and interactome of SUMOylated proteins from wild-type and *siz1-2*. (A) Additional GO functional enrichment of all SUMO conjugates identified after HS versus the whole *Arabidopsis* proteome not listed in Figure 3-8B. **(B)** Interactome map of abundant SUMO conjugates. Targets with a significant loss of detection in the *siz1-2* mutant are illustrated with red nodes. Proteins with a significant decrease or increase in SUMOylation in the *siz1-2* mutant are colored yellow and blue respectively. Green nodes represent proteins with no significant change in SUMOylation in *siz1-2*. The interactome was generated using STRING8 database and visualized using Cytoscape. Either the TAIR identifier or the common protein name or abbreviation is shown for each conjugate.


A
Fold Enrichment
0.25
0.5
1
2
4
8
16
32
Regulation RNA Splicing
RNA-directed DNA methylation
Regulation Histone Modification
DNA Endoreduplication
Chromatin Silencing
Nucleosome Assembly
Histone Modification
Gene Silencing by RNA
Negative Regulation Flower Development
Jasmonate-Mediated Signaling
Translational Elongation
mRNA Processing
Meristem Maintenance
Response to Heat
Transcription, DNA-Templated
Regulation Immune System Process
Regulation Developmental Process
Positive Regulation Response to Stimulus
Activated Ethylene Signaling
Positive Regulation Gene Expression
Protein Folding
Positive Regulation Transcription, DNA-Templated
Response to Temperature
DNA Repair
Ribosome Biogenesis
Plant Organ Morphogenesis
Hormone-Mediated Signaling
Response to Abiotic Stimulus
Transport
Protein Phosphorylation
B
Loss of SUMOylation in siz1-2
Decreased SUMOylation in siz1-2
Increased SUMOylation in siz1-2
No change in siz1-2

confidence SIZ1 substrates was calculated using the whole SUMOylome as the background in the analysis. This analysis revealed that the SIZ-influenced conjugates were specifically enriched in proteins involved in abiotic and biotic stress responses, such as heat acclimation, response to drought, hormone signaling and defense responses (Figure 3-8C).

Investigation of the list of SUMOylation targets whose modification is reduced in the *siz1*-2 mutant identified well-known transcription factors, co-activators/repressors and chromatin modifiers, as well as many proteins involved biotic and abiotic stress responses (Table 3-1). These include the heat shock transcription factors HSFC1 and HSFB2B, DREB2A and its ubiquitin E3 ligase DRIP2, ABF3 and JAZ3, -4 and -6, as well as EML1, -2 and-3, IDD7 and WRKY33, which are associated with biotic plant defense. Interestingly, several members of the TOPLESS family, which are part of the 12-member plant Groucho/Tup1 co-repressor family [165], as well as their interacting transcription factors ARF2, NAC052, EMF1, NAC050 and HSFB2B [160], were decreased in SUMOylation in *siz1-2*. However, there was no change in the SUMO conjugation in the *siz1-2* mutant of another plant Groucho/Tup1 co-repressor LEUNIG, its homolog LUH and their interactor SEUSS, suggesting that SIZ1 selectively SUMOylates the TOPLESS complex. Additionally, multiple SIZ1-influenced SUMO conjugates are components the SWI-SNF chromatin remodeling complex (SWI3C, SWI3D, CHR11, CHR17 and PICKLE), implicating SIZ1 in the SUMOylation of multiple subunits of the same protein complex.

Interestingly, none of the 7 conjugates found to have an increase in abundance in the *siz1-2* mutant are transcription factors or chromatin regulators, but include a methyl transferase (OMT), two small ribosomal S3 family proteins (RPS3), a glutathione s-transferase (GST) family protein, and glyceraldehyde-3-phosphate dehydrogenase C2 (GAPC2) (Figure 3-6A,B). Unexpectedly, the co-chaperone BAG7, which is involved in the unfolded protein response

Table 3-1. Arabidopsis SUMOylation Targets Whose Modification is Impacted by SIZ1

Locus	Name	*p*-value[a]	Log2(FC)[b]	Description[c,d]
At3g12140	EML1	***	–	Emsy N Terminus/ plant Tudor-like
At1g48500	JAZ4	***	–	Jasmonate-zim-domain protein 4
At4g34000	ABF3	***	–	ABA responsive element-binding factor 3
At3g10480	NAC50	*	–	NAC domain containing protein 50
At3g24520	HSFC1	*	–	Heat shock transcription factor C1
At1g55110	IDD7	*	–	indeterminate(ID)-domain 7
At3g16830	**TPR2**	*	–	**TOPLESS-related 2**
At5g44180	RINGLET 2	*	-7.89	Homeodomain-like transcriptionregulator
At5g60410	SIZ1	***	-7.79	SUMO E3 Ligase
At3g48050	SUO	***	-7.43	BAH &TFIIS helical bundle-like domain
At1g79000	HAC1	***	-6.80	Histone acetyltransferase/CBP family 1
At1g18800	NRP2	*	-5.61	NAP1-related protein 2
At5g14270	BET9	*	-5.39	Bromodomain&extraterminal domain 9
At1g21700	SWI3C	***	-5.13	SWITCH/sucrose nonfermenting 3C
At1g15750	**TOPLESS**	***	-5.04	**Transcriptional co-repressor**
At4g11660	HSFB2B	***	-4.94	Heat shock transcription factor
At2g04880	WRKY1	*	-4.88	Zinc-dependent activator protein-1
At3g27260	GTE8	***	-4.73	Global transcription factor group E8
At5g15020	SNL2	*	-4.60	SIN3-like 2
At5g15130	WRKY72	***	-4.21	WRKY DNA-binding protein 72
At4g34430	SWI3D	*	-4.08	DNA-binding family
At5g62000	ARF2	*	-4.06	Auxin response factor 2
At1g33240	GTL1	***	-4.02	GT-2-like 1
At3g05380	ALY2	*	-3.78	DIRP, Myb-like DNA-binding domain
At3g17860	JAZ3	***	-3.46	Jasmonate-zim-domain 3
At2g30580	DRIP2	***	-3.20	DREB2A interacting protein
At5g16270	SYN4	*	-3.19	Rad21.3, Sister chromatid cohesion-4
At5g37190	CIP4	***	-3.16	COP1-interacting protein 4
At3g10490	NAC052	*	-3.03	NAC domain containing protein 52
At5g13020	EML3	***	-2.80	Emsy N Terminus/ plant Tudor-like
At2g44440	EML2	***	-2.78	Emsy N Terminus domain
At1g26665	Med10	*	-2.74	Mediator complex, subunit
At1g25540	PFT1	*	-2.57	Phytochrome&flowering time regulator
At2g38470	WRKY33	*	-2.35	WRKY DNA-binding protein 33
At2g25170	PICKLE	*	-2.19	Chromatin remodeling factor CHD3
At1g49480	RTV1	*	-1.98	Related to Vernalization 1
At4g25500	RSP35	*	-1.91	Arginine/serine-rich splicing factor 35
At1g09770	CDC5	*	-1.89	MYBCDC5, cell division cycle 5
At5g05410	DREB2A	*	-1.77	DRE-binding protein 2A
At4g17060	FIP2	*	-1.76	FRIGIDA interacting protein 2
At1g13960	WRKY4	*	-1.60	WRKY DNA-binding protein 4
At1g51130	NSE4	*	2.19	Component Smc5/6 DNA repair complex
At5g35530	Ribosomal S3	***	3.92	Ribosomal S3 family protein
At5g62390	BAG7	***	7.86	BCL-2-associated athanogene 7

[a] ***p-value <0.01; * ≤ 0.05 [b] absent in *siz1-2* [c] Blue, TPL family member
[d] Green, related to abiotic stress

[166, 167], showed over a 200-fold increase in SUMOylation in *siz1-2*, becoming one of the most abundant conjugates in the mutant. Interestingly, Nse4, a component of the Smc5-Smc6 complex which includes MMS21 in yeast and mammals, also showed a significant increase in SUMOylation in the *siz1-2* mutant. I was also able to detect MMS21 in the *siz1-2* samples, where as MMS21 was not present in any wild-type or *mms21-1* samples, suggesting that the SUMOylation of MMS21 is increased in *siz1-2*.

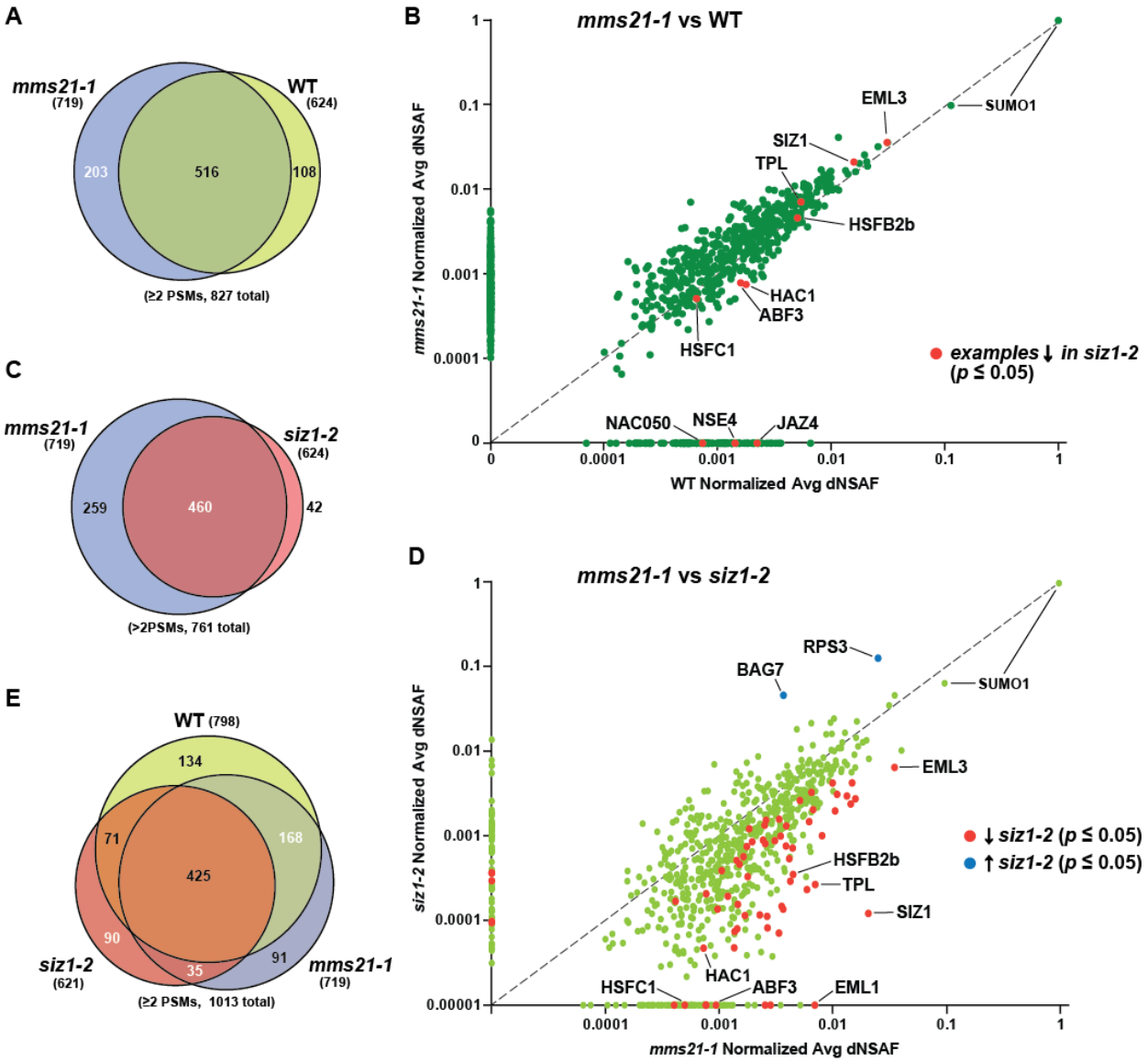
To ensure that the changes in SUMO conjugate abundance observed in *siz1-2* are due to altered SUMO addition and not changes in expression levels, I analyzed RNA sequencing (RNA-seq) data comparing mRNA abundance in the *siz1-2* mutant versus wild-type before or after heat shock [Augustine & Vierstra, unpublished]. None of the 105 proteins whose SUMO modification is influenced by SIZ1 show a significant change in expression before or after stress. Furthermore, only *GAPC2* and *OMT*, which showed elevated SUMOylation in *siz1-2*, have amplified expression in the mutant compared to wild-type. These RNA-seq results provide confidence in that the SUMO conjugates I identified by be modified by SIZ1 are true targets of the E3 ligase.

Surprisingly, I isolated several peptides of SIZ1 from the *siz1-2* mutant (Figure 3-1A) indicating not only that the siz1-2 polypeptide accumulated, but that it was able to interact with the E2 conjugating enzyme in a manner that results in SUMO addition to the E3 ligase. However, the siz1-2 peptides were detected at much reduced levels compared to wild-type (Table 3-1), likely due to an overall decrease in the siz1-2 protein abundance as seen through anti-SIZ1 immunoblot analysis (Figure 3-1E). Only peptides upstream of the TDNA insertion site were identified in *siz1-2* where as SIZ1 peptides covering most regions of the protein were detected in the wild-type. This supports the transcriptional analysis identifying *siz1-2* as an attenuated mutant and not a true null allele.

Attempt to Identify MMS21 SUMOylation Targets

While large differences in protein abundances were observed by comparing normalized dNSAF values of SUMO conjugates isolated from *siz1-2* mutant to wild-type, the SUMO target abundance in the *mms21-1* mutant was similar to wild-type in both unstressed and heat-shocked samples (Figure 3-7B; Figure 3-10B). When considering all conjugates found in heat-stressed samples between wild type and *mms21-1* (with ≥2 PSMs), 108 proteins were absent and 203 conjugates were unique in the *mms21-1* mutant from a total list of 827 conjugates (Figure 3-10A). As was observed by comparisons of SUMO conjugate abundance of *siz1-2* versus wild-type, the target abundances of *mms21-1* versus *siz1-2* after heat-shock revealed large deviations from a 1:1 ratio with many proteins underrepresented or absent in the *siz1-2* background (Figure 3-10D). The abundances of the SIZ1-influenced conjugates in the *mms21-1* mutant reflected those of wild-type before and after stress (Figure 3-7B; Figure 3-10B,D), supporting the conclusion that these targets are selectively modified by SIZ1. As the abundance of SUMO conjugates is similar in *mms21-1* versus wild-type, I was unable to distinguish any MMS21-dependent targets in this analysis, suggesting that MMS21 SUMOylates a small subset of low abundance proteins.

Figure 3-10. Analysis of SUMOylated proteins purified from the *mms21-1* mutant background. (A) Venn diagram of SUMO conjugates overlap from the *mms21-1* mutant and wild-type (WT) after heat shock. **(B)** Plot of average normalized dNSAF values of SUMOylated proteins from WT and *mms21-1*. Notable proteins are identified on the graph. Proteins with a significant decrease in SUMOylation in the siz1-2 mutant are highlighted in red. The dashed line represents the line of equilibrium. **(C)** Venn diagram of SUMO conjugates overlap from the *mms21-1* and *siz1-2* mutant after heat shock. **(D)** Plot of average normalized dNSAF values of SUMOylated proteins from *mms21-1* and *siz1-2*. Notable proteins are identified on the graph. Proteins with a significant decrease or increase in SUMOylation in the siz1-2 mutant are highlighted red and blue, respectively. The dashed line represents the line of equilibrium. Note that two dNSAF values are assigned to SUMO1 by Morpheus Spectra Counter. **(E)** Venn diagram of SUMO conjugates overlap conjugates from WT, *mms21-1* and *siz1-2* after heat shock.


A
mms21-1
(719)
WT
(624)
203
516
108
(≥2 PSMs, 827 total)
B
mms21-1 vs WT
mms21-1 Normalized Avg dNSAF
WT Normalized Avg dNSAF
EML3
SUMO1
SIZ1
TPL
HSFB2b
HAC1
ABF3
HSFC1
NAC050
NSE4
JAZ4
examples ↓ in siz1-2
(p ≤ 0.05)
C
mms21-1
(719)
siz1-2
(624)
259
460
42
(>2PSMs, 761 total)
D
mms21-1 vs siz1-2
siz1-2 Normalized Avg dNSAF
mms21-1 Normalized Avg dNSAF
RPS3
BAG7
SUMO1
EML3
HSFB2b
TPL
SIZ1
HAC1
ABF3
HSFC1
EML1
↓ siz1-2 (p ≤ 0.05)
↑ siz1-2 (p ≤ 0.05)
E
WT (798)
134
71
168
425
90
35
91
siz1-2
(621)
mms21-1
(719)
(≥2 PSMs, 1013 total)

DISCUSSION AND FUTURE DIRECTIONS

As SUMOylation is a crucial post-translational modification with an essential role in the response to cellular challenges, it is important to understand how conjugation of proteins is regulated. In this chapter, I developed an unbiased proteomics strategy to assign targets to the SUMO E3 ligase SIZ1 by comparing the conjugate profiles of the E3 ligase mutants *siz1-2* and *mms21-1* to those of wild-type. Any protein not isolated in the E3 mutant, but present in the wild-type was designated as a potential target of this SUMO ligase.

Immunoblot analyses revealed a 50% reduction of SUMO conjugation in *siz1-2* versus wild-type. MS analysis of the SUMOylome profiles of *siz1-2* and wild-type showed that this decrease is not due to just a loss of specific SUMO substrates. Over 66% of all targets identified in wild-type are still detected in the mutant (Figure 3-4B), however the abundance of many these shared conjugates is considerable lower in *siz1-2* seedlings (Figure 3-6). This reduction in SUMOylation in the *siz1-2* mutant could be due to the presence of a potentially active siz1-2 polypeptide. In my analysis, I discovered that the *siz1-2* locus is still actively transcribed upstream of the T-DNA, but the corresponding protein is poorly translated. Subsequent tandem MS analysis of SUMO conjugates isolated from *siz1-2* seedlings revealed that the siz1-2 protein does accumulate and is SUMOylated but at much reduced levels than in wild-type.

Previous MS analysis of SUMO targets identified SIZ1 as one of the most abundant conjugates upon heat shock, suggesting that auto-SUMOylation of the E3 ligase either through directed modification or by non-specific transfer of SUMO moiety from the E2-SUMO intermediate [33, 46]. Thus, the SUMOylation of the siz1-2 protein suggests the mutant form is able interact with the E2 enzyme possibly through the SP-RING domain retained in siz1-2 and potentially direct some level of SUMO conjugation. Alternatively, the residual SUMOylation of

targets observed in the *siz1-2* mutant may be due to E2-directed modification through the recognition of SUMOylation motifs by the conjugating enzyme [4, 15, 87].

This raises the question if SIZ1 is essential for cell viability. All characterized mutant alleles of *SIZ1* (*siz1-1*, *siz1-2* and *siz1-3*) interrupt the coding region after the SP-RING, and thus might maintain some residual E3 ligase activity if expressed. Therefore, additional *SIZ1* alleles with mutations upstream of the SP-RING need to be identified. Several such lines exist in Col-0, including SALK_058033, SALK_044209 and SAIL_805_A10 (CS876977) available through Arabidopsis Biological Resource Center (ABRC: http://abrc.osu.edu) and GK-217A09-014146 available through ABRC and the Nottingham Arabidopsis Stock Centre (NASC: http://arabidopsis.info).

Overall, I expanded the list of high confidence SUMO conjugates to over 1100 proteins, bringing the number of targets closer to those identified in yeast and mammalian studies (reviewed in [135]). To distinguish targets whose SUMOylation is influenced by SIZ1, I chose to perform label-free quantification by calculating dNSAF values of conjugates and followed this with statistical analyses using LIMMA, which provides a powerful method for stable analysis of smaller sample sizes as compared to an ordinary t-test [168]. I was able to identify over 100 proteins whose modification is impacted by SIZ1. The level of SUMOylation was below detection for 18 proteins and decreased for 87, while the abundance of 7 proteins was increased in *siz1-2* compared to wild-type. These SIZ1-influenced targets were enriched in major transcription factors, translational regulators and chromatin modifiers. Furthermore, comparison of this set of conjugates to the whole SUMOylome revealed that SIZ1 specifically directs the SUMOylation of proteins prevalent in the abiotic and biotic response process, such as heat acclimation, response to drought, hormone signaling and defense responses. These data agree

with previous observation, which have implicated multiple roles of SIZ1 in the response to environmental challenges and pathogens [26, 53, 66, 68, 109, 154].

Further analysis of conjugates whose modification was reduced or not detected in the *siz1-2* mutant revealed the presence of multiple major abiotic transcription factors, including HSFC1, HSFB2B, DREB2A, ABsF3 and JAZ4, as well as factors involved in pathogen defense including EML1, EML2, EML3, IDD7 and WRKY33. Additionally, the plant Groucho/Tup1 co-repressors TPL and TOPLESS-related (TPR) 2 as well as their potential interactors ARF2, NAC052, EMF1, NAC050 were identified, indicating that SIZ1 plays a role in SUMOylating the whole complex. Other TPL family members, TPR1, TPR3 and TPR4, also appeared to have reduced conjugation the *siz1-2* mutant, but not to significant levels by LIMMA. Surprisingly, another SUMOylated transcriptional complex consisting of the plant Groucho/Tup1 co-repressor LEUNIG, its homolog LUH and their interactor SEUSS, showed no loss of SUMOylation in *siz1-2*, suggesting that SIZ1 selectively SUMOylates the TOPLESS complex. As the TOPLESS complex is a central regulator of transcription in Arabidopsis through its interaction with a wide variety of transcription factors, including AUX/IAA and ARFs, AP2/ERF proteins, JAZ proteins, AFB, and TIR-NB-LRR proteins [160, 165], SIZ-dependent SUMOylation could affect the transcriptional level of numerous genes.

I also discovered that the SUMOylation of multiple components of the SWI-SNF chromatin remodeling complex, SWI3C, SWI3D, CHR11, CHR17 and PICKLE, is directed by SIZ1. SWI3C and SWI3D are two of the four SWI3 subunits in Arabidopsis that make up part of the core of the complex while CHR11, CHR17 and PICKLE interact with this core to direct the recruitment of the SWI-SNF complex to genes [169, 170]. This connects SIZ1 to chromatin remodeling in plants and ultimately to transcriptional regulation.

It has been proposed that SUMO E3s are directed to nuclear protein complexes through the interaction with one or more of the subunits and subsequently proceed to SUMOylate the whole complex [159, 171]. This SUMO "spray" mechanism increases the binding of the components to each other through SUMO interacting domains (SIMs), which stabilizes the complex and promotes its activity. Therefore, the identification of the TOPLESS and SWI-SNF complexes as SIZ1-influenced conjugates suggests that their modifications by SIZ1 are required for their stability and activity. In animals and yeast, anti-SUMO chromatin immunoprecipitations followed by DNA sequencing (ChIP-Seq) and RNA-seq experiments have found SUMO to be localized to the promoters of active genes upon heat shock where they reduce the transcription levels of stress-responsive genes [123–126]. This repressive function of SUMO modification on chromatin upon heat shock is proposed to prevent the detrimental hyper-activation of stress responses and thus promote cell survival. [113, 139]. As SIZ1 has been implicated in multiple stress responses, such as thermotolerance to heat and cold, drought survival, and nutrient deprivation [53, 64, 66, 68, 109, 113, 172], the SIZ1-dependent SUMOylation of the co-repressor TOPLESS complex and the chromatin remodeling SWI-SNF complex, as well as various transcription factors, may serve a protective function by regulating the level of stress-responsive genes upon heat shock. Anti-SUMO ChIP-seq combined with transcriptome analysis in *siz1-2* and wild-type plants before and after heat-shock will provide deeper insights in the role of SIZ1-depentent conjugation.

SIZ1 has been connected to drought and heat tolerance in plants [66, 109, 112, 154]. Accordingly, I connected SIZ-dependent SUMOylation to ABA-dependent and ABA-independent signaling pathways, which mediate plant responses to environmental stresses, such as drought and high temperatures [115–117]. I identified DREB2A, and its ubiquitin E3 ligase

DRIP2, as well as ABF3 as targets of SIZ1. These substrates are involved in the same signaling pathway in response to osmotic and heat stress. Under non-stressed conditions DREB2A is degraded by DRIP2, but upon heat stress and drought DREB2A is stabilized and binds to promoters and activates the expression of genes [173]. ABF3 is activated by ABA-dependent phosphorylation, and regulates *DREB2A* expression as well as interacts physically with DREB2A [115]. While it is unknown how the SUMOylation affects the activity of DREB2A, DRIP2 and ABF3, SIZ1-directed modifications could prevent the degradation of DREB2A and additionally control the transcriptional activity of DREB2A and ABF3 to regulate the expression of stress-response factors promoting cell survival.

Interestingly, only seven SUMO targets were found to have an increased abundance in the *siz1-2* mutant and include a methyl transferase (OMT), two small ribosomal S3 family proteins, a glutathione s-transferase family protein, glyceraldehyde-3-phosphate dehydrogenase C2 (GAPC2), as well as Nse4 and BAG7. This enhanced conjugation could be due to the aberrant physiology of *siz1-2* seedlings or could represent targets of other SUMO E3 ligases that are more readily detected in the mutant. Indeed, I found examples of both instances. The elevated levels of GAPC2 and OMT are likely due to an increase in their expression levels in the ligase mutant. On the other hand, Nse4 could be a target of MMS21 as Nse4 and MMS21/Nse2 are components of the same DNA repair complex in animals and yeast [84].

Surprisingly, the SUMOylation of BAG7 shows an over 200-fold increase in *siz1-2* and is one of the most abundant conjugates in the mutant before and after heat shock. BAG7 is a co-chaperon involved in the unfolded protein response (UPR) during heat and cold stress [166, 167]. Interestingly, Arabidopsis BAG7 was recently reported to be SUMOylated upon heat shock [166]. During heat stress, the co-chaperone is translocated from the endoplasmic reticulum to the

nucleus, and subsequent interacts with transcription factors in a SUMOylation-dependent manner to regulate gene expression. This translocation was required for the thermotolerance to high temperatures [166]. In *siz1-2*, BAG7 is SUMOylated under non-stressed conditions, suggesting that the co-chaperone is constitutively activated in the E3 ligase mutant. This increased modification of BAG7 could be caused by an induction of the UPR in *siz1-2*, in part due to the aberrant physiology of the *siz1-2* mutant such as the accumulation of high levels of salicylic acid (SA) and the resulting constitutively active biotic stress response [64].

Previous investigations of specific proteins identified PHR1, GTE3, MYB30, ABI5, and NIA1/2 as SIZ1 targets [53, 68, 74, 120, 156]. Although many of these targets were isolated in my MS analysis of SUMO conjugates, none were found by LIMMA to have reduced SUMOylation in *siz1-2* versus wild-type. In fact, most of these conjugates were not consistently detected in the biological replicates and thus excluded from the analyses. For example, the SIZ1 conjugate PHR1 showed an over 4-fold decrease in conjugation in *siz1-2* versus wild-type, but was found in less than 3 biological replicates in both genotypes and thus excluded. This highlights the limitations of the methodologies used in this study, as I was unable to confidently analyze targets detected at lower abundances in the samples. Therefore, my list of 105 SIZ1 targets is likely a conservative estimate. Improved quantification of protein abundances, through quantitative proteomics such as isobaric tag labeling, could allow for the identification of considerable more E3 ligase targets.

I was unable to identify any MMS21-dependent targets in this analysis as the abundances of most SUMO conjugates are similar in *mms21-1* versus wild-type, suggesting that MMS21 SUMOylates a small subset of proteins. This is supported by the immunoblot analysis of SUMO conjugates in *mms21-1*. Whereas a large portion of SUMO conjugates was lost in *siz1-2*, no

differences in conjugate levels were observed in the *mms21-1* versus wild-type upon heat shock. In support, SIZ1 was implicated in multiple cellular responses to abiotic and biotic signals, and thus is likely to SUMOylate a wide range of conjugates. On the other hand, MMS21 is involved in cell cycle regulation and DNA damage responses [55, 56, 69, 79, 83, 174] and thus may only modify a few proteins involved in cell cycle and DNA repair. Additionally, in yeast, the modification of proteins by MMS21 is specifically involved in the response to DNA damage [80, 82]. Thus, few conjugates of MMS21 might be present under normal conditions or heat shock and could not be detected in my MS analysis. As an alternative approach seedlings could be treated with DNA damaging agents or cell cycle progression inhibitors before purification which might lead to an increase in MMS21-dependent conjugates.

In conclusion, I identified over one hundred SIZ1-influenced targets, including major transcription factors, co-activators/repressors, and chromatin modifiers connected to abiotic and biotic stress responses. This list of substrates indicates that SUMOylation by the E3 ligase SIZ1 provides stress protection by conjugating a large array of key transcriptional regulators to modify the expression of genes under challenging conditions. The identification of specific SIZ1 conjugate targets will support further research into how SUMOylation alters the activity, interactions, location and/or half-life of these targets, and ultimately how their modifications impact plant growth and stress protection.

MATERIALS AND METHODS

Plant Materials and Growth Conditions

The *Arabidopsis thaliana* ecotype Columbia (Col-0) was used as the genetic background for all germplasm unless otherwised noted. The SUMO-conjugate purification line as described by Miller et al. (2010) (*6His-SUMO1(H89R) sumo1-1 sumo2-1*) was introgressed into the *siz1-2* (SALK_065397) [53] or the *mms21-1* (*hyp2*-2, Sail_77_G06) [55, 56] mutants by crossing. Quadruple homozygous lines were identified in the F2 or F3 generations by glufosinate and kanamycin resistance linked to the *sumo1-1* and the *6His-SUMO1(H89R)* loci, respectively, and by genomic PCR for all loci.

Unless otherwise noted, seeds were surface sterilized with bleach and stratified in water at 4°C in the dark for 2 d before sowing. For phenotypic studies, plants were grown at 21°C on soil under long-day photoperiods (LD: 16-hr light, 8-hr dark). For the analysis and purification of SUMO conjugates, seedlings were grown for 8 d at 22°C under continuous light on solid Gamborg's B-5 Basal Medium (GM; Sigma-Aldrich) supplemented with 2% sucrose, and containing a 0.8% agar base that was topped with 0.1% agar in GM. For the heat stress, the plates or cultures were incubated at 37°C for 30 min in a circulating water bath. At the indicated times, the seedlings were harvested and frozen to liquid nitrogen temperatures.

Genomic, RT-PCR, qPCR, and RNA-seq Analyses

Genomic, RT-PCR, and qPCR analysis employed the oligonucleotide primers described in Table 3-2. Genomic analysis of the *SUMO1 and SUMO2* alleles and the *6His-SUMO1(H89-R)* transgene employed the primers developed by Saracco et al (2007) and

Table 3-2. Table of primers used in genomic, RT-PCR and qPCR analyses

	Name	Sequence
qPCR	SIZ1 qPCR 5'- F	TAATCGACCTGGAGGACAGC
	SIZ1 qPCR 5'- R	TCCCCTTTACCCTCTTCTGG
	SIZ1 qPCR 3'- F	CAATGGCATCTGTTCCTGTG
	SIZ1 qPCR 3'- R	CAACTAGGCCATCGTTTGCT
	ACT2-F	GGCATCACACTTTCTACAATGAGC
	ACT2-R	ACCCTCGTAGATTGGCACAG
RT-PCR	SIZ1 P1	TGTCTGGTGTTGAAGACATGG
	SIZ1 P2	ATAATGGGTCCATCGTCGCG
	SIZ1 P3	ACCTGATGGTAGCCTTTGCC
	SIZ1 P4	CAACTAGGCCATCGTTTGCT
	SIZ1 P5	CAGATGCTTCAGCTCAGTCG
	SIZ1 P6	CTTCTGGCGAGGAAATGAAA
	MMS21 P1	AGGATTCAGAACGCTTCTTTGG
	MMS21 P2	ATTCAGTGACAGGCTTCCCG
	MMS21 P3	CCGAGGTAAACTGCAGAATAGC
	MMS21 P4	AGTCTTCAATCACTTCAGCCC
Genomic	MMS21-F	CTCTCCCATGCCTGATGTCT
	MMS21-R	CGAGCGCATCTCCTCTATTT
	SIZ1-F	CTTTGCCCCTCTGCTGTTGA
	SIZ1-R	TGAGTTAATCCCGTTTGTAGTTGC
	LBa1	TGGTTCACGTAGTGGGCCATCG
	LB1	GCCTTTTCAGAAATGGATAAATAGCC

Miller at al (2010), respectively. RNA was extracted from 8-d-old seedlings using the RNeasy Plant Mini Kit (Qiagen) followed by first strand Synthesis with oligo(dT) primers using the SuperScript III First-Strand Synthesis System (Invitrogen-Thermo Fisher Scientific). cDNA and genomic DNA were amplified using EconoTaq Plus Green 2X MasterMix (Lucigen). qPCR was performed with a BioRad CFX Connect Real-Time System together with the LightCycler 480 SYBR Green I Master mix (Roche); transcript abundance was normalized to that generated with *ACT2* based on the comparative threshold method [175].

The RNA-seq datasets for *SIZ1* were generated from total RNA isolated from 7-d-old wild-type and *siz1-2* seedlings grown at 24°C or subjected to 30-min heat stress at 37ºC plus 30 min recovery at 24ºC. TruSeq mRNA libraries were generated by the University of Wisconsin Gene Expression Center with two of the three biological replicates prepared to maintain strand information, and sequenced using the Illumina HiSeq 2000 platform with 2x100 bp paired-end reads. The resulting fastq sequence files were manually searched for reads containing the *SIZ1* query sequence 5'-CCAACGGCATGGAACTTGAT-3' or its reverse complement 5'-ATCAAGTTCCATGCCGTTGG-3', which correspond to the sequence immediately upstream of the T-DNA insertion site reported for *siz1-2* [53].

Immunoblot Analyses

Immunodetection of SUMO1/2 conjugates used frozen tissue pulverized at liquid nitrogen temperatures, mixed with two volumes per mg fresh weight (uL/mg) of twice-strength SDS-PAGE sample buffer, heated to 100°C for 5 min, and clarified at 16,000 X*g*. The clarified extracts were subjected to SDS-PAGE and transferred onto Immobilon-P PVDF membranes (EMD-Millipore). The membranes were blocked with non-fat dry milk in Phosphate Buffered Saline (Fisher Scientific) and probed with rabbit anti-SUMO1 antibodies [28]. Rabbit antibodies

against the proteasome PBA1 subunit were used as the loading control [176]. For the detection of SIZ1, immunoblot analysis was performed as above using anti-SIZ1 antibodies describe in Miller et al (2013).

The relative abundance of SUMO conjugates in *siz1-2* and wild-type plants were quantified by immunoblot analysis of a dilution series of the clarified crude extracts with anti-SUMO1 antibodies followed by IRDye 800CW goat anti-rabbit antibodies (LI-COR), and imaged using the 800 nm channel on the LI-COR Odyssey FC fluorimager. The signal intensities for free SUMO and the smear of SUMO conjugates at the top of immunoblots were quantified by the LICOR imaging software and normalized to the signals obtained with anti-PBA1 antibodies in combination with the IRDye 680RD goat anti-rabbit antibodies (LI-COR) detected at 700 nm.

Affinity Purification of SUMO Conjugates

SUMO conjugates were enriched from *6His-SUMO1(H89R) sumo1-1 sumo2-1* plants using the three-step protocol developed by Miller et al (2010) with slight modifications. Approximately 45 g of frozen tissue was pulverized at liquid nitrogen temperatures and resuspended for 1 hr at 55°C in 90 ml of Extraction Buffer (EXB: 100 mM Na_2HPO_4, 10 mM Tris-HCl (pH 8.0), 300 mM NaCl, and 10 mM iodoacetamide IAA)), containing 7 M guanidine-HCl with 10 mM sodium metabisulfate, and 2 mM phenylmethylsulfonyl fluoride (PMSF) added just before use and the pH readjusted to 8.0. The extract was filtered through two layers of Miracloth (EMD Millipore), clarified by centrifugation at 15,000 X*g*, and incubated overnight at 4°C with Ni-NTA resin (Qiagen) (0.75 mL resin/5 g of tissue) after addition of imidazole to 10 mM. The Ni-NTA beads were washed sequentially with 10 column volumes of EXB containing

6 M guanidine-HCl, 0.25% Triton X-100 and 10 mM imidazole (pH 8.0), 10 column volumes of EXB containing 8 M urea, 0.25% Triton X-100 and 10 mM imidazole (pH 6.8), and fifteen column volumes of EXB containing 8 M urea, 0.25% Triton X-100 and 10 mM imidazole (pH 8.0). SUMO conjugates were eluted with five column volumes of Elution Buffer (ELB: 350 mM imidazole, 100 mM Na_2HPO_4, 10 mM Tris-HCl, and 10 mM IAA (pH 8.0)). The eluant was concentrated by ultrafiltration with a 10-kDa molecular mass cutoff filter (Amicon Ultra-4; EMD Millipore or Vivaspin 6; GE Healthcare Life Sciences).

After two exchanges into ELB without imidazole and reconcentration, the sample was renatured by adding drop-wise to 25 volumes of ice-cold 0.5X RIPA buffer (100 mM $NaHPO_4$ (pH 7.4), 10 mM Tris-HCl, 200 mM NaCl, 2.5% NP40, 1.25% sodium deoxycholate, 0.25% SDS, 10 mM IAA, and 1 mM PMSF). The renatured samples were incubated overnight at 4°C with 0.5 mg of affinity-purified anti-SUMO1 antibodies bound to 500 µL Affi-Gel 10 beads (Bio-Rad). The beads were washed with 10 column volumes of 0.5X RIPA buffer followed by 100 column volumes of 50 mM $NaHPO_4$ (pH 7.4), 100 mM NaCl, 10 mM IAA and 1 mM PMSF. SUMO conjugates were eluted by first incubating the beads for 20 min at 65°C with 1 column volume of 1% SDS and 5% 2-mercaptoethanol, and subsequently washing with 10 column volumes of ELB containing 8 M urea (pH 8.0). The eluates were pooled and mixed with 350 µL of Ni-NTA resin for 4 hr at 22°C, the beads were washed with 70 mL of ELB containing 8 M urea and 10 mM imidazole (pH 8.0), and the bound conjugates were successively eluted with six 500 µL pulses of ELB (without IAA) containing 6 M urea and 300 mM imidazole (pH 8.0). The final elute was concentrated to 100 µL by ultrafiltration as above.

Tandem Mass Spectrometry

SUMO conjugate preparations were reduced for 1 hr at 22°C with 10 mM dithiothreitol, followed by alkylation with 20 mM IAA for 1 hr [161]. The reaction was quenched with 20 mM dithiotreitol and diluted with 25 mM ammonium bicarbonate. Trypsin (Trypsin Gold, MS grade, Promega) was added at a 1:20 protease to sample ratio by weight and incubated at 37°C for 18 hr. The digests were acidified with 0.5% trifluoroacetic acid (TFA) and desalted with OMIX C18 pipette tips (Agilent), using 75% acetonitrile and 1% formic acid for elution. The samples were vacuum dried, resuspended in 5% acetonitrile and 0.1% formic acid, and subjected to tandem LC-MS using either a LTQ Orbitrap Velos (ThermoFisher) or Q Exactive Plus (ThermoFisher) mass spectrometers operated in the positive ESI mode.

For the LTQ Orbitrap Velos ESI-MS, the tryptic peptides were separated on 50 um x 365 μm fused silica capillary micro-column packed with 20 cm of 1.7-μm-diameter, 130-Å pore size, C18 beads (Waters BEH), with an emitter tip pulled to approximately 1 μm using a laser puller (Sutter Instruments). Peptides were eluted over 120 min at a flow rate of 300 nL/min with a linear gradient of 2% to 30% acetonitrile in 0.1% formic acid. Full-mass scans were performed in the FT Orbitrap at 300–1,500 m/z at a resolution of 60,000, followed by ten MS/MS high-energy collisional dissociation (HCD) scans of the ten highest intensity parent ions at 42% normalized collision energy and 7,500 resolution and a mass range starting at 100 m/z. Dynamic exclusion was set to a repeat count of two over a duration of 30 sec and an exclusion window of 120 sec.

For the Q-Exactive ESI-MS, the tryptic peptides were separated by nano-scale liquid chromatography (LC) using a Dionex Ultimate™ 3000 Rapid Separation LC system (Thermo Scientific) equipped with an 75 μm x 2 cm Acclaim® PepMap 100 guard column followed by a

75 µm x 15 cm analytical Acclaim® PepMap™ RSLC C18 column (2 µm particle size, 100 Å pore size, (Thermo Scientific). The peptides were eluted over 120 min at a flow rate of 250 nL/min with a linear gradient of 1.6% to 32% acetonitrile in 0.1% formic acid. Data-dependent acquisition of full MS scans within a mass range of 380-1500 m/z at a resolution of 70,000 was performed, with the automatic gain control (AGC) target set to 3 x 106, and the maximum fill time set to 200 ms. High energy collision-induced dissociation (HCD) fragmentation of the top 15 most intense peaks was performed with a normalized collision energy of 28, with an AGC target of 2 x 105 counts and an isolation window of 3.0 m/z, excluding precursors that had an unassigned, +1, +7 or +8 charge state. MS2 scans were conducted at a resolution of 17,500, with an AGC target of 2×10^5 and a maximum fill time of 100 msec. Dynamic exclusion was performed with a repeat count of 2 and a duration of 20 sec, while the minimum MS ion count for triggering MS2 was set to 4×10^3 counts.

MS Data Analysis

The MS2 spectra were searched using MORPHEUS version 160 [177] against the Arabidopsis protein database (TAIR10; http://www.arabidopsis.org) along with common contaminants (*e.g.,* trypsin and human keratin). The default search parameters were set to a precursor mass tolerance of 2.100 Da, product mass tolerance of 0.010 Da, maximum FDR of 1%, fixed carbamidomethylation of cytosines and variable methionine oxidation, along with a maximum of 2 missed trypsin cleavages. To provide label-free quantification based on dNSAF values, the datasets were filtered through Morpheus Spectral Counter [161]. Background proteins identified by MS analysis from four biological replicates of wild-type plants were classified as contaminants and removed from the SUMO conjugate datasets. For the heat stress

datasets, only proteins identified by three or more PSMs per biological replicate were included in the final analyses; for the unstressed datasets, all targets were considered due to low protein abundance. All but one heat stress dataset comparing wild type and *siz1-2* were generated from the average of two technical replicates, whereas all unstressed datasets were generated without technical replicates. dNSAF values for SUMOylated targets in each biological replicate were normalized based on the dNSAF value for SUMO1/2.

For the statistical analyses of SUMO conjugates, missing values among biological replicates were imputed with PERCEUS [178] using standard settings and applied to each biological replicate separately. To reduce the imputation frequencies, the dataset were limited to SUMO targets detected in at least three biological replicates in either background. Before imputation, dNSAF values were transformed by taking the $\log_2$ value of the fraction multiplied with 1e10 ($x = \log_2(dNSAF*1e^{10})$). The LIMMA statistical package in R [179] was used to calculate significant differences between SUMO conjugate profiles from wild-type and *siz1-2* plants by adaptation of the source code published by [168] to determine the moderate *p*-value.

Fold enrichments of specific GO functions were obtained using the PANTHER database [163] with either the default *Arabidopsis* proteome or the identified SUMOylome serving as the background for enrichment. GO localizations were predicted with DAVID [164]. The SUMO conjugate interactome was generated by STRING 8 [180] and visualized using Cytoscape version 3.4.0 [181]. Proportional Venn diagrams were generated using Vennerable (https://r-forge.r-project.org/projects/vennerable/).

CHAPTER 4

LYSINE-NULL SUMO: INVESTIGATIONS INTO POLYSUMO CHAINS AND DEVELOPMENT OF A NOVEL MASS SPECTROMETRIC METHOD IN *ARABIDOPSIS THALIANA*

Ryan Nasti assisted with the screening of transgenic plants and Samuel L. York assisted in phenotypic analysis of seedlings. I completed the remainder of the work in this chapter.

ABSTRACT

Conjugation of the Small Ubiquitin-like Modifier (SUMO) to proteins is an essential protein modification in plants and is involved in numerous cellular processes, including translation, transcription, RNA biology, chromatin modification, as well as responses to biotic and abiotic challenges. Targets are not only modified by single SUMOs, but also by chains of internally-linked SUMOs. These SUMO polymers provide a unique binding site that can be recognized by specific SUMO-interacting proteins, such as SUMO targeted ubiquitin ligases (STUbLs). Although polySUMO chain formation has been detected in *Arabidopsis thaliana* [33, 36], their biological function is unknown. In this chapter, I examined the importance of polySUMO chains in Arabidopsis. Surprisingly, I found that SUMO chains are not essential for viability by rescuing the embryonic lethal *sumo1-1 sumo2-1* mutant with a lysine-null *SUMO1*. Plants unable to form polySUMO chains are phenotypically similar to wild-type and are still able to conjugate SUMO to targets upon heat shock. These findings allowed me to adapt a recently developed purification strategy, which employs a lysine-null SUMO and peptidase digest for increased mass spectrometric (MS) identification of conjugation sites, to Arabidopsis. By rescuing the lethal *sumo1-1 sumo2-1* double mutant with a 6His-tagged lysine-null *SUMO1(H89-R)* construction, I was able to enrich for peptides modified with SUMO by combining a Lys-C peptidase digest and nickel affinity purifications. This will allow for identification of the lysines modified by SUMO in Arabidopsis to support further research into how SUMO addition alters the activity, interactions, location and/or stability of targets.

INTRODUCTION

The conserved 10-kDa Small Ubiquitin-like Modifier (SUMO) is part of the ubiquitin-fold family, which is characterized by a shared beta-grasp fold, but otherwise members have little protein sequence homology to each other [1]. SUMO itself consists of a 10-to 20-amino-acid-long N-terminal extension preceding the beta-grasp fold and contains a di-glycine (di-Gly) motif in the C-terminal end of the protein. Canonical SUMOs are processed by SUMO proteases to reveal the di-Gly motif for conjugation to proteins. Like ubiquitin, SUMO is ligated to proteins through the formation of an isopeptide bond with the C-terminal carboxyl group of the glycine residue and ϵ-amino group of a lysine residue in the target. This reaction is driven by a three-step enzymatic cascade involving a heterodimeric E1 SUMO Activating Enzyme (E1: SAE), a SUMO conjugating enzyme (E2: SCE) and SUMO ligases (E3). Most conjugated lysine are located in a consensus SUMOylation motif consisting of a hydrophobic residue preceding the lysine followed by acidic residues two positions downstream of the lysine (ψKxD/E) [4, 89]. Additionally, SUMO is not only conjugated to target proteins, but is also covalently bound to other SUMO moieties to form polySUMO chains.

Chain formation was first discovered for ubiquitin (Ub). PolyUb chains serve diverse roles depending on their architecture as several different lysine residues of Ub are conjugated, which creates distinct structures that present unique binding sites for Ub-interacting proteins. As examples, Lys-48 and Lys-11-linked chains lead to proteosomal degradation of their target while Lys-63-linked chains have mainly nonproteolytic roles in cellular processes such as DNA repair and endocytosis [182].

In animals and yeasts, polySUMO chains are formed through the modification of one or more distinct lysines located in the N-terminal extension of SUMO, creating a flexible chain

[90]. These PolySUMO chains create distinct binding sites for SUMO-interacting proteins. [91, 98]. Multiple SUMO-chain binding proteins have been identified in yeasts and animals, including ZIP1, part of the synaptonemal complex involved in meiosis, CENP-E (centromere-associated protein E), required for correct chromosome alignment during mitosis, the ubiquitin E3 ligase heterodimer Slx5–Slx8 *in Saccharomyces cerevisiae* (Rfp1/Rfp2–Slx8 in *Schizosaccharomyces pombe*) and mammalian Ubiquitin E3 ligase RNF4 (RING finger protein 4) which has a much higher binding affinity for SUMO chains than for mono or di-SUMO [99–103]. These SUMO-targeted ubiquitin ligases (STUbLs) represent a new class of ubiquitin E3 ligases that connect the SUMO pathway with the ubiquitin-proteosome system (UPS). The mammalian STUbL RNF4 has tandem SIMs that allow it to bind and ubiquitylate specifically polySUMO conjugates to regulate protein abundance in processes such as DNA repair [101, 183, 184]. Although SUMO-binding proteins have been identified in Arabidopsis [98], it is unknown if they include interactors with a preference for SUMO chains.

In yeasts, polySUMO chains are not essential for viability, but a loss of SUMO polymers does result in select phenotypes. *Saccharomyces cerevisiae* expressing a version unable to form chains are similar to wild-type, but show a slight defect during meiosis [49, 99]. However, in *Schizosaccharomyces pombe*, the loss of SUMO chains results in nuclear defects and increased sensitivity to genotoxic stresses [50].

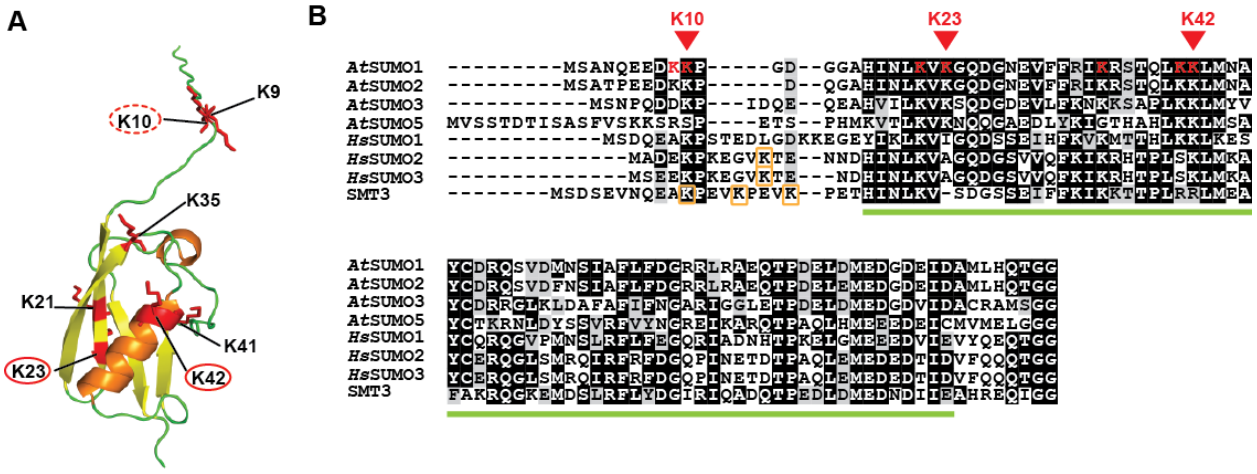
The effects of SUMOylation are varied and can alter the activity, interactions, location and/or stability of targets. Knowing the sites of addition can provide substantial insight into the outcome of SUMOylation. Recent mass spectrometric (MS) methods have made use of a lysine-null (K0) SUMO to identify SUMOylation sites, or “SUMO footprints”, on target proteins [89, 107]. Previous MS analyses to identify SUMO-conjugated lysines were hampered by sample

complexity as only a small subset of peptides contains a modified lysine, and for adequate detection of SUMOylated lysines through MS, the peptides carrying the SUMO footprint need to be enriched. To isolate only SUMOylated peptides, a His-tagged SUMO2(K0) was expressed in mammalian cell lines and combined with a Lys-C peptidase digestion and nickel purifications [89, 107]. The K0 SUMO is resistant to cleavage by Lys-C and remains attached to the modified peptide, which allows for the subsequent purification of only peptides conjugated with SUMO.

In Arabidopsis, we have now identified over 1400 SUMO targets (Miller et al. 2010; Miller et al. 2013; Chapter 3 of this thesis), however the SUMOylation sites on these proteins are unknown. Although a mutation was included in the 6His-tagged SUMO (6His-SUMO1(H89-R)) employed for purification that would allow for identification of modified lysines, only a very small subset of sites was identified [33]. To improve on the detection of conjugated lysines, I decided to adapt the above described purification method for enrichment of SUMOylated peptides in Arabidopsis.

The two main Arabidopsis SUMO isoforms, SUMO1 and SUMO 2, contain seven lysines residues: Lys-9 and Lys-10 located in the flexible N-terminal extension and Lys-21, Lys-23, Lys-35, Lys-41 and Lys-42 of the beta-grasp fold (Figure 4-1A). All lysine residues are solvent exposed, and thus are accessible by the SUMOylation machinery for conjugation, however only Lys-10, Lys-23 and Lys-42 have been identified by mass spectrometric analysis to be SUMOylated [33, 36, 46, Rytz & York et al unpublished data]. Interestingly, Lys-23 was found to be modified in all conditions, while Lys-42 was only conjugated with SUMO upon stress, including heat shock (HS) and hydrogen peroxide treatment [33]. While these lysines are conserved in all expressed Arabidopsis SUMO isoforms (Figure 4-1B), SUMO3 and

Figure 4-1. The 3D structure and amino acid sequence alignment of *Arabidopsis* SUMO1/2 highlighting the lysine residues. (A) 3D ribbon diagram of *At*SUMO1. The seven lysines of SUMO1 are highlighted. SUMO attachment sites K23 and K42 (circled in red) were identified through MS/MS from purified native SUMO conjugates, while K10, dashed circle, was identified as a conjugation site in vitro [33, 36]. SWISS-MODEL was used to generate the 3-D structure of *At*SUMO1 based on the human SUMO3 template (PDB ID 2D07 (2.1Å)). **(B)** Amino acid alignment showing the conservation of lysines across SUMO isoforms in *Arabidopsis thaliana*. Human (*Hs*) SUMO 1, SUMO2 and SUMO3 as well as *Saccharomyces cerevisiae* SUMO (SMT3). All lysines in *Arabidopsis* SUMO1 are highlighted in red. The arrows indicate the known SUMOylation sites in *Arabidopsis*. The SUMOylated lysines in *Hs*SUMO 1, *Hs* SUMO2 and *Hs*SUMO3 SMT3) are highlighted by orange boxes. The sequences were aligned using MUSCLE and shading of the alignment was completed using BOXSHADE. Amino acids after the C-terminal di-glycine motif were omitted from the alignment.


A
K9
K10
K35
K21
K41
K23
K42
B
K10
K23
K42
AtSUMO1 ---------MSANQEEDKKP-----GD--GGAHINLKVKGQDGNEVFFRIKRSTQLKKLMNA
AtSUMO2 ---------MSATPEEDKKP------D--QGAHINLKVKGQDGNEVFFRIKRSTQLKKLMNA
AtSUMO3 ----------MSNPQDDKP---IDQE--QEAHVILKVKSQDGDEVLFKNKKSAPLKKLMYV
AtSUMO5 MVSSTDTISASFVSKKSRSP----ETS--PHMKVTLKVKNQQGAEDLYKIGTHAHLKKLMSA
HsSUMO1 ------------MSDQEAKPSTEDLGDKKEGEYIKLKVIGQDSSEIHFKVKMTTHLKKLKES
HsSUMO2 --------------MADEKPKEGVKTE--NNDHINLKVAGQDGSVVQFKIKRHTPLSKLMKA
HsSUMO3 --------------MSEEKPKEGVKTE---NDHINLKVAGQDGSVVQFKIKRHTPLSKLMKA
SMT3 --------MSDSEVNQEAKPEVKPEVK--PETHINLKV-SDGSSEIFFKIKKTTPLRRLMEA
AtSUMO1 YCDRQSVDMNSIAFLFDGRRLRAEQTPDELDMEDGDEIDAMLHQTGG
AtSUMO2 YCDRQSVDFNSIAFLFDGRRLRAEQTPDELEMEDGDEIDAMLHQTGG
AtSUMO3 YCDRRGLKLDAFAFIFNGARIGGLETPDELDMEDGDVIDACRAMSGG
AtSUMO5 YCTKRNLDYSSVRFVYNGREIKARQTPAQLHMEEEDEICMVMELGGG
HsSUMO1 YCQRQGVPMNSLRFLFEGQRIADNHTPKELGMEEEDVIEVYQEQTGG
HsSUMO2 YCERQGLSMRQIRFRFDGQPINETDTPAQLEMEDEDTIDVFQQQTGG
HsSUMO3 YCERQGLSMRQIRFRFDGQPINETDTPAQLEMEDEDTIDVFQQQTGG
SMT3 FAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGG

SUMO5 are likely not involved in polySUMO chain formation as SUMO3 and SUMO5 were unable to polymerize *in vitro* [36, 38].

While polySUMOs chains do form in Arabidopsis, their functional role in plant biology is unknown. To investigate the importance of polySUMO chains, I studied the loss of SUMO chains by complementing the *sumo1-1 sumo2-1* mutant with lysine-null (K0) SUMO1 construction. Plants unable to form polySUMO chains are not only viable, but are phenotypically similar to wild-type. Like wild-type SUMO, the K0 SUMO variant is rapidly conjugated to proteins upon HS. However, K0 SUMO mutants show a reduction in higher molecular mass conjugates corresponding with a loss of polySUMO chains. Consequently, rescuing the *sumo1-1 sumo2-1* mutant with a lysine-null 6His-tagged SUMO1 bearing the H89-R footprint mutation (*6His-Arg-SUMO1(K0, H89-R)*) allowed for the isolation SUMO-conjugated peptides after LysC digestion. This line is now available for enrichment of conjugated peptides and identification of SUMO footprints.

RESULTS

Plants Unable to Form PolySUMO Chains Are Phenotypically Similar to Wild-Type

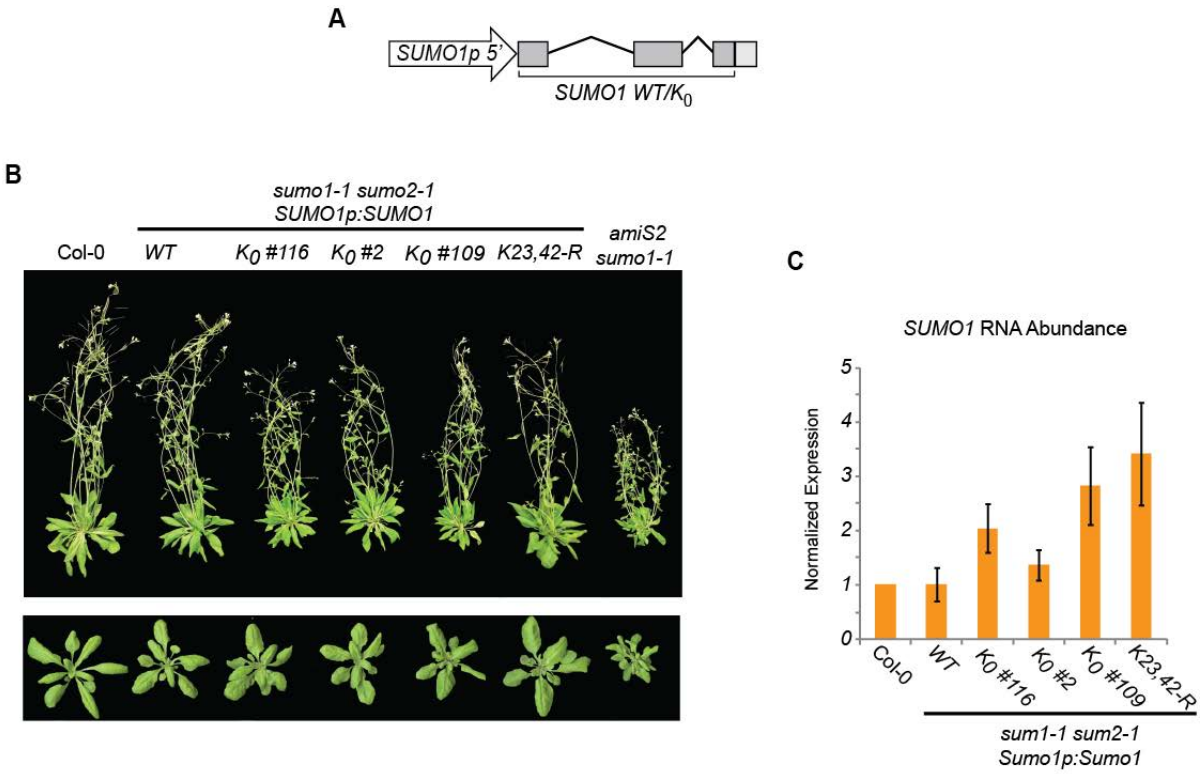
To study the role of polySUMOylation in *Arabidopsis*, a lysine-null genomic *SUMO1* (*SUMO1(K0)*) construction driven by the native promoter was used to rescue the embryonic-lethal *sumo1-1 sumo2-1* mutant (Figure 4-2A). All seven lysine residues of SUMO1 were replaced with arginines through PCR-directed mutagenesis. After transformation of plants heterozygous for *sumo1-1* and homozygous for *sumo2-1*, seedlings were selected for the *sumo1-1* allele and the genomic SUMO transgene and allowed to self cross. From this second

generation, I was able to isolate multiple independent lines expressing K0 SUMO1 (Figure 4-2B,C). Expression of *SUMO1(K0)* was confirmed through sequencing of *SUMO1* cDNA (data not shown). These K0 SUMO1 transformants are comparable to wild-type under normal growth conditions (Figure 4-2B).

Anti-SUMO immunoblots of seedlings confirmed that K0 SUMO1 is conjugated to target proteins upon HS in a manner similar to wild-type (Figure 4-2D). Before HS, few higher molecular weight conjugates were present, but after HS, K0 SUMO was rapidly ligated to target proteins. Comparable to wild-type seedlings, this modification is reversible, as conjugate levels decreased during the recovery from HS and the amount free SUMO returns to basal levels. Interestingly, the amount of SUMO present at higher molecular masses on anti-SUMO immunoblots was reduced in the K0 SUMO mutants, corresponding with a loss of SUMOylation. This was not due to a lack of free SUMO, as a large pool of unconjugated SUMO is available, nor is the binding affinity of the antibody to the lysine-null SUMO altered. Antigenicity tests confirmed the same preference of the antibody to recombinant K0 SUMO1 as wild-type SUMO1 (Figure 4-3). Most likely, this observed decrease is due to a loss of polySUMO conjugates.

As the Lys-23 and Lys-42 were specifically identified to be SUMOylated under stress conditions, I also analyzed a plant line expressing a genomic *SUMO1 K23,42-R* variant from the native *SUMO1* in the *sumo1-1 sumo2-1* background. Like the K0 SUMO lines, the phenotype of the SUMO1(K23,42-R) mutant was comparable to wild-type and SUMOylation of targets upon HS occured in a normal fashion (Figure 4-2D). However, the SUMO(K23,42-R) line is an over expressor of SUMO, but this did not appear to affect the plant phenotype (Figure 4-2B,C).

Figure 4-2. Phenotypic characterization of plants expressing only lysine-null SUMO1. (A) Gene diagram illustrating the genomic constructions used to rescue the *sumo1-1 sumo2-1* embryonic lethal mutant. The 1.5-kbp promoter and 5'-untranslated region (UTR) are represented by the arrow. Exons are shown in dark gray, while the 3' UTR and introns are shown as light gray box and lines, respectively. **(B)** Representative wild-type (Col-0) and *sumo1-1 sumo2-1* plants rescued with *SUMO1, SUMO1(K0)* or *SUMO1(K23,42-R)* genomic constructions grown for 20 d (bottom) and 40 d (top) in a LD photoperiod. **(C)** Quantification of *SUMO1* expression level in wild-type (Col-0) and SUMO1, SUMO1(K0) or SUMO1(K23,42-R) lines. The values were normalized to those of *ACT2*, and represented as a ratio to that of wild-type. **(D)** Immunoblot analysis of 8-d-old seedlings heat stressed for 30 min at 37°C and collected at the indicated times. The membrane was probed with either anti-SUMO1/2 or anti-PBA1 antibodies (loading control). High molecular mass SUMO conjugates and free SUMO are indicated by the brackets and arrowheads, respectively.


A
SUMO1p 5'
SUMO1 WT/K0
B
sumo1-1 sumo2-1
SUMO1p:SUMO1
Col-0
WT
K0 #116
K0 #2
K0 #109
K23,42-R
amiS2
sumo1-1
C
SUMO1 RNA Abundance
Normalized Expression
0
1
2
3
4
5
Col-0
WT
K0 #116
K0 #2
K0 #109
K23,42-R
sum1-1 sum2-1
Sumo1p:Sumo1

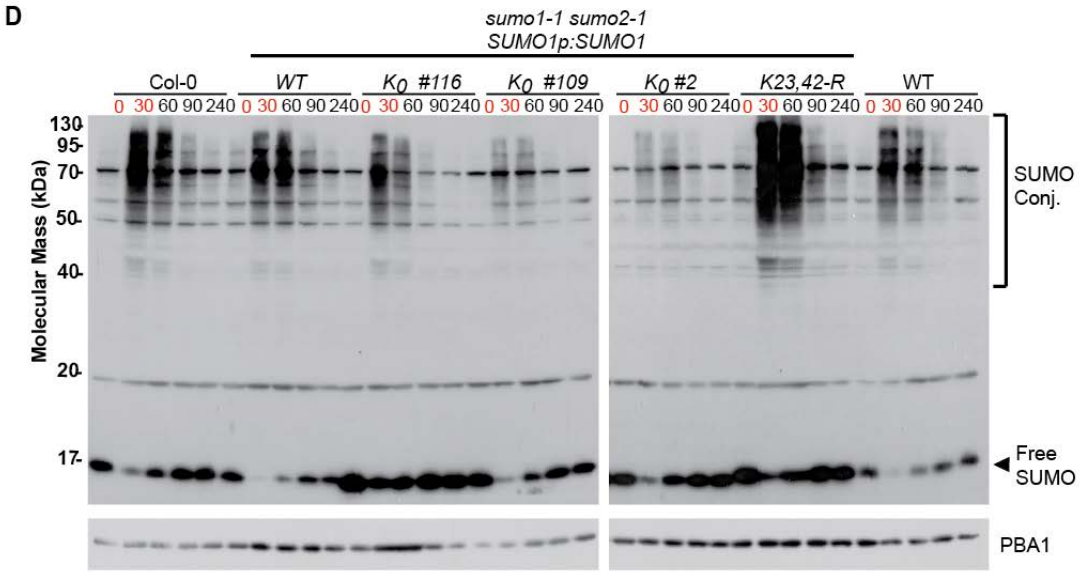
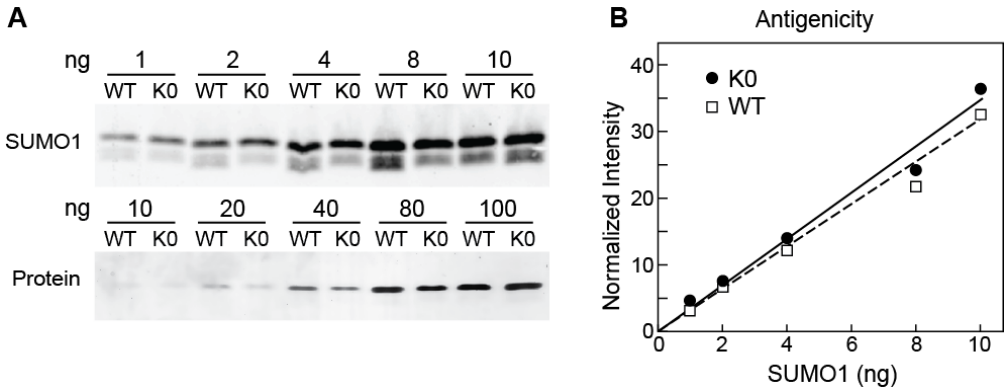

D
sumo1-1 sumo2-1
SUMO1p:SUMO1
Col-0
WT
K0 #116
K0 #109
K0 #2
K23,42-R
WT
0 30 60 90 240
Molecular Mass (kDa)
130
95
70
50
40
20
17
SUMO
Conj.
Free
SUMO
PBA1

Figure 4-3. Wild-type and lysine-null SUMO1 have indistinguishable antigenicity. (A) Comparison of protein antigenicity of anti-SUMO1 antibodies to wild-type (WT) and lysine-null (K0) SUMO. A dilution series of purified recombinant WT and K0 SUMO1 was subjected to anti-SUMO1 immunoblot analysis (upper panel) or staining for protein level (lower panel). The active form of SUMO1 either in the WT form or bearing lysine to arginine mutations were expressed in *E. coli* and separated by SDS-PAGE analysis. Proteins were transferred to PDVF membrane and probed with rabbit anti-SUMO1 antibodies, or stained for total protein. Immunoblot analysis used IRDye 800CW goat anti-rabbit antibodies (LI-COR) to image the blot using the 800 nm channel on the LI-COR Odyssey Classic fluorimager. **(B)** Quantification of the antigenicity shown in the immunoblot analysis from panel (A). The signal intensities for SUMO1 were quantified by the LICOR imaging software and normalized to the background signal. The dashed and the black line represent the best fit for WT or K0 SUMO1 intensity, respectively.


A
ng 1 2 4 8 10
WT K0 WT K0 WT K0 WT K0 WT K0
SUMO1
ng 10 20 40 80 100
WT K0 WT K0 WT K0 WT K0 WT K0
Protein
B
Antigenicity
K0
WT
Normalized Intensity
SUMO1 (ng)
0
10
20
30
40
0
2
4
6
8
10

The Thermotolerance of Seedlings Expressing K0 SUMO Is Comparable to WT

To establish if the loss of polySUMO chains has an effect on the stress response of *Arabidopsis* seedlings, I tested the K0 SUMO lines' thermotolerance to moderately high temperatures (TMHT). Mutants deficient in SUMO (*amiS2 sumo1-1*) or unable to properly SUMOylate a large subset of targets (*siz1-2*) are sensitive to prolonged exposure to moderately high temperatures (35°C), with *amiS2 sumo1-1* seedlings unable to recover from the heat stress (Figure 4-3,B). However, the K0 SUMO1 lines and the SUMO1(K23,42-R) mutant were able to successfully recover from prolonged growth at 35°C. There was no difference in the TMHT of these SUMO mutants versus wild-type plants (Figure 4-3). Thus, SUMO chain formation is not required in the response to moderately high temperature.

The Responses of K0 SUMO seedlings to Various Stress Treatments Are Comparable to wild-type

I subjected the K0 SUMO lines to a gauntlet of stress conditions to associate polySUMO chain formation with specific cellular process. First, I tested three different genotoxic stresses, methyl methanesulfonate (MMS), hydroxy urea (HU) and mitomycin, as lysine-null SUMO mutants in *S. pombe* were sensitive to HU induced DNA damage (Skilton et al. 2009). However, the *Arabidopsis* K0 SUMO mutants do not have an increased sensitivity to HU or the other DNA damaging agents tested and their root growth compared with wild-type (Figure 4-5A,B).

As the *Arabidopsis* SUMO E3 ligase mutants *siz1-2* was found to be hypersensitive to absisic acid (ABA) treatment [119], I tested the response of the K0 SUMO mutants to multiple plant hormones, including ABA, indole-3-acetic acid (IAA) and salicylic acid (SA). No

Figure 4-4. Plants expressing SUMO1 or SUMO1(K0) have similar thermotolerances to moderately high temperature (TMHT). (A) Phenotypic analysis of seedlings treated to moderately high temperature. Seedlings were grown on agar plates at 21°C in long day (LD: 16hr light 8 hr dark) conditions, transferred to 35°C for the indicated number of days, and allowed to recover at 21°C. Plates were image after 20 d of total growth. The following lines were used for the assay wild-type (Col-0), *SUMO1 sumo1-1 sumo2-1* (*WT*), *SUMO1(K0) sumo1-1 sumo2-1 #116* (*K0 #116*), *SUMO1(K0) sumo1-1 sumo2-1 #2* (*K0 #2*), *SUMO1(K0) sumo1-1 sumo2-1 #109* (*K0 #109*), *siz1-2* and *SUMO2 miRNA sumo1-1* line (*amiS2*). **(B-D)** Quantification of the response to the moderately high temperatures based on seedling fresh weight. 20-d old seedlings exposed to 35°C for the indicated number of days were weighed and the results reported as percentage of the total weight of seedlings for each genotype. Each data point represents an average of at least three biological replicates, and error bars measure standard deviation. The same plants lines were used as in panel (A). **(B)** Comparison of Col-0, *WT*, *K0 #116*, the *siz1-2* mutant and the *SUMO2 miRNA sumo1-1* line (*amiS2*). **(C)** Comparison of the multiple independent lines expressing *SUMO1(K0)* to wild-type (Col-0). **(D)** Comparison of the *SUMO1(K23,42-R) sumo1-1 sumo2-1 (K23,42-R)* line to Col-0, *WT sumo1 and K0 #116*.

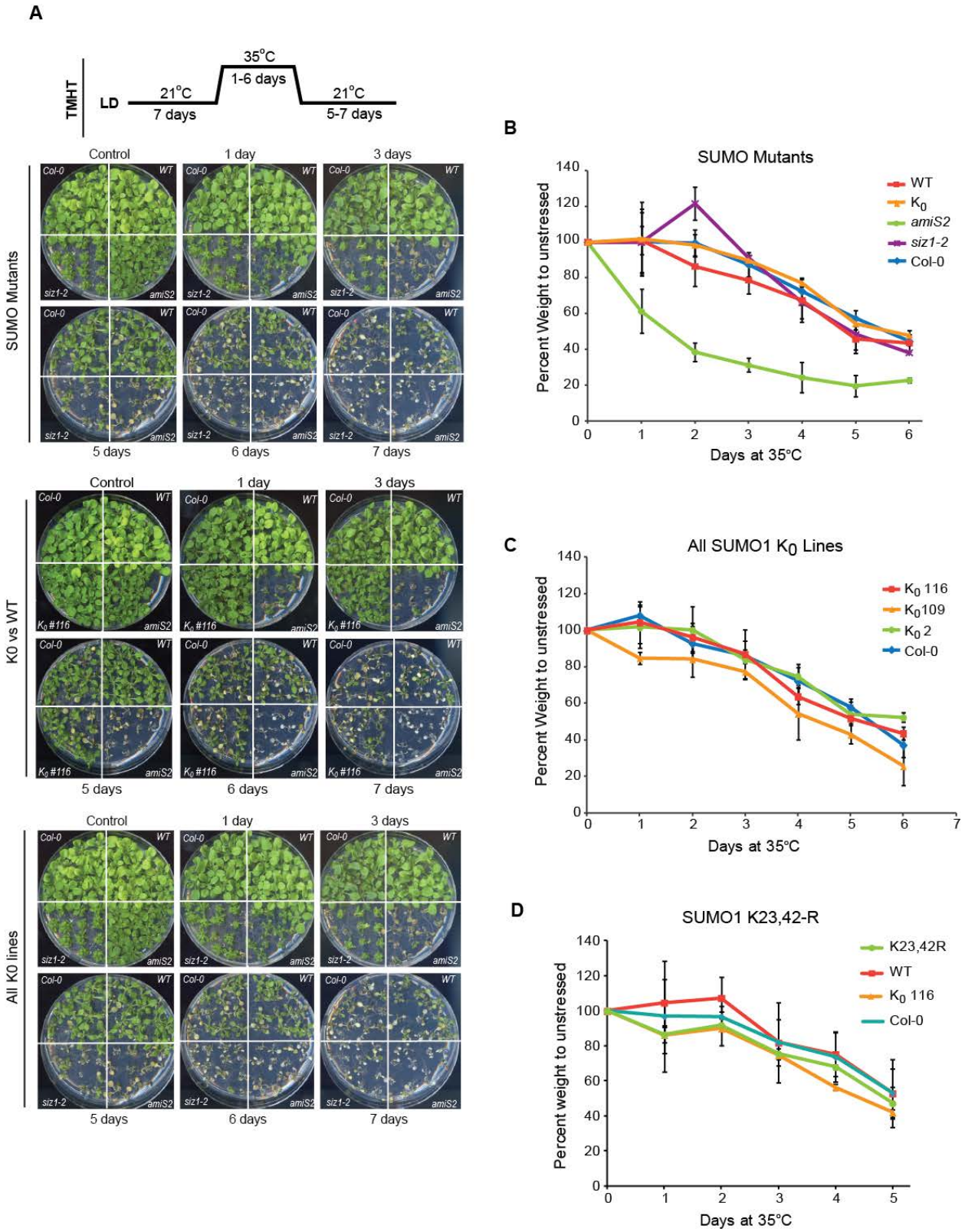

A
TMHT
LD
21°C
7 days
35°C
1-6 days
21°C
5-7 days
Control
1 day
3 days
SUMO Mutants
K0 vs WT
All K0 lines
Col-0
WT
siz1-2
amiS2
K0 #116
5 days
6 days
7 days
B
SUMO Mutants
Percent Weight to unstressed
Days at 35°C
WT
K0
amiS2
siz1-2
Col-0
C
All SUMO1 K0 Lines
Percent Weight to unstressed
Days at 35°C
K0 116
K0109
K0 2
Col-0
D
SUMO1 K23,42-R
Percent weight to unstressed
Days at 35°C
K23,42R
WT
K0 116
Col-0

Figure 4-5. Phenotypic analysis of lysine-null SUMO mutants exposed to DNA damaging agents. (A) Quantification of the root length of seedlings in response to the indicated treatments. Seedlings were grown on agar plates at 21°C in long day (16hr light 8 hr dark) conditions. Root lengths are reported as a fraction compared to the average untreated seedling root length. At least four biological replicates were completed for each condition tested. The lines used are the following: Col-0, *SUMO1 sumo1-1 sumo2-1* (*WT*), *SUMO1(K0) sumo1-1 sumo2-1 #2* (*K0 2*), *SUMO1(K0) sumo1-1 sumo2-1 #109* (*K0 109*), *SUMO1(K0) sumo1-1 sumo2-1 #116* (*K0 116*), and *SUMO1(K23,42-R) sumo1-1 sumo2-1* (*K23,42-R*). **(B)** Dosage response of seedlings treated to the indicated concentration of DNA damage agents. The same plant lines were used and grown as described in (A).

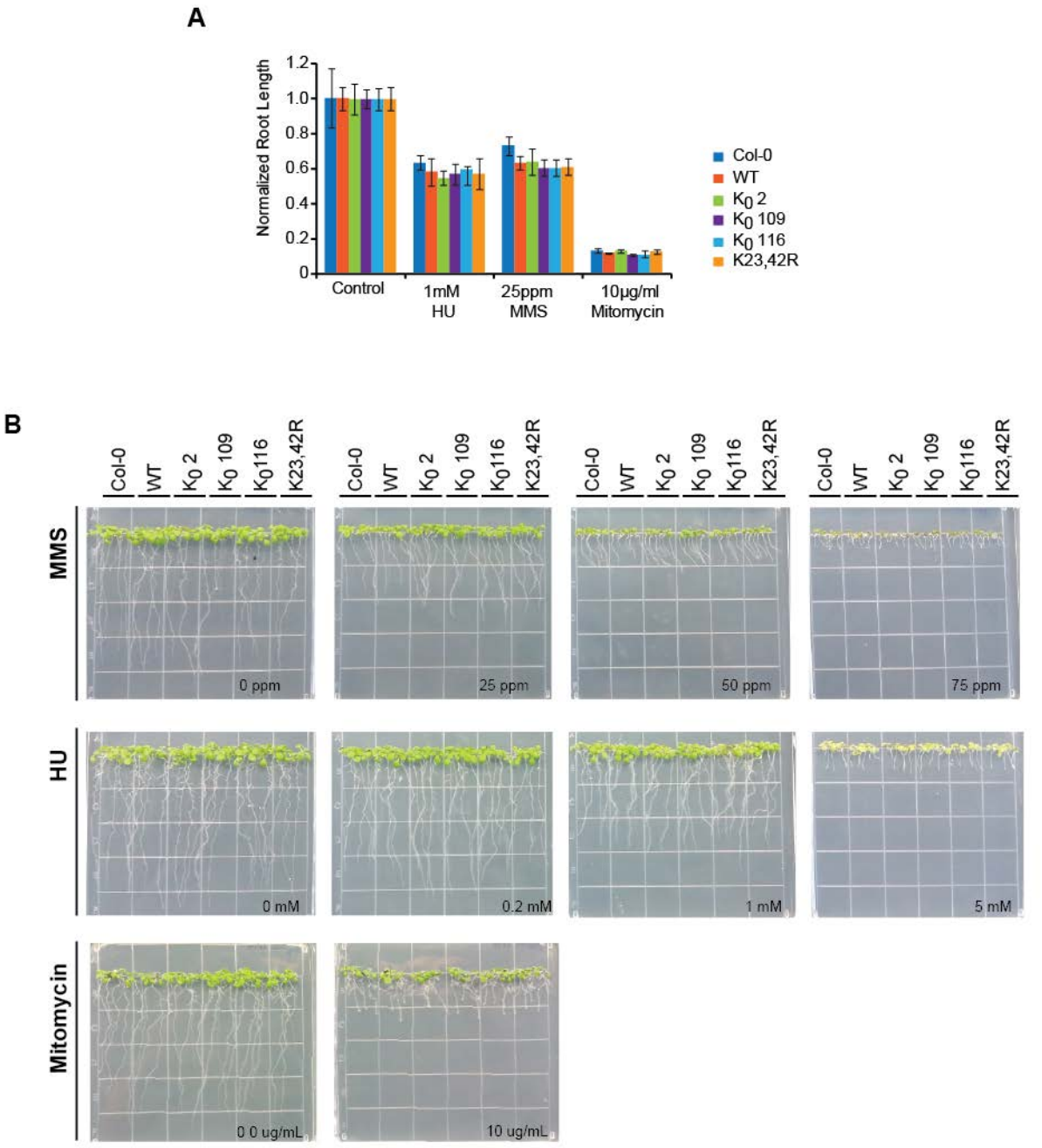
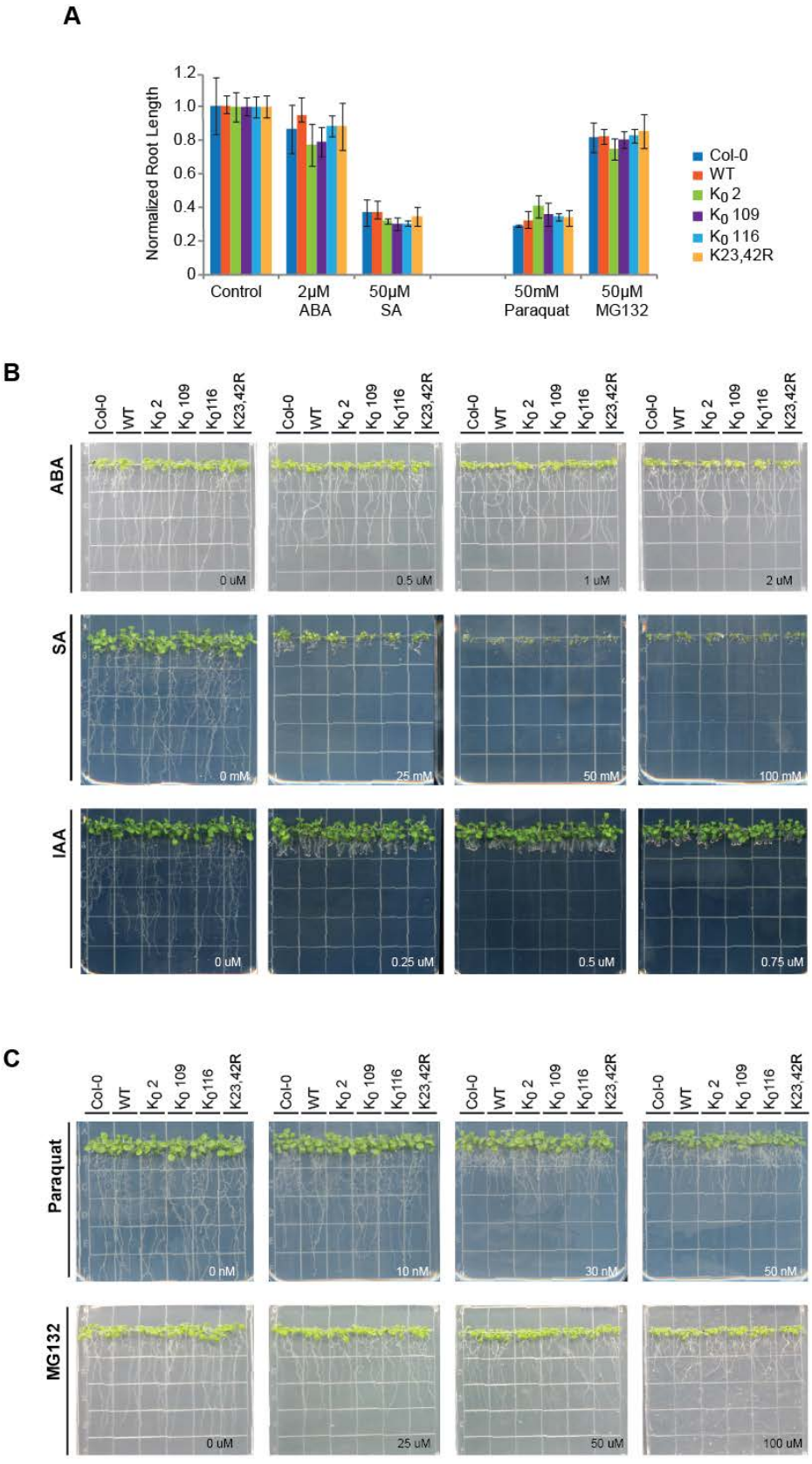

A
Normalized Root Length
1.2
1.0
0.8
0.6
0.4
0.2
0
Control
1mM HU
25ppm MMS
10µg/ml Mitomycin
Col-0
WT
K0 2
K0 109
K0 116
K23,42R
B
Col-0
WT
K0 2
K0 109
K0116
K23,42R
MMS
0 ppm
25 ppm
50 ppm
75 ppm
HU
0 mM
0.2 mM
1 mM
5 mM
Mitomycin
0 0 ug/mL
10 ug/mL

Figure 4-6. Phenotypic analysis of lysine-null SUMO mutants exposed to various hormone treatments, proteosome inhibition, and oxidative stress. (A) Quantification of the root length of seedlings in response to the indicated treatments. Seedlings were grown on agar plates at 21°C in long day (16hr light 8 hr dark) conditions. Root lengths are reported as a fraction compared to the average untreated seedling root length. At least four biological replicates were completed for each condition tested. The lines used are the following Col-0, *SUMO1 sumo1-1 sumo2-1* (*WT*), *SUMO1(K0) sumo1-1 sumo2-1 #2* (*K0 2*), *SUMO1(K0) sumo1-1 sumo2-1 #109* (*K0 109*), *SUMO1(K0) sumo1-1 sumo2-1 #116* (*K0 116*), and *SUMO1(K23,42-R) sumo1-1 sumo2-1* (*K23,42-R*). **(B+C)** Dosage response of seedlings treated to indicated concentration of stressors. The same plant lines were used and grown as described in (A).

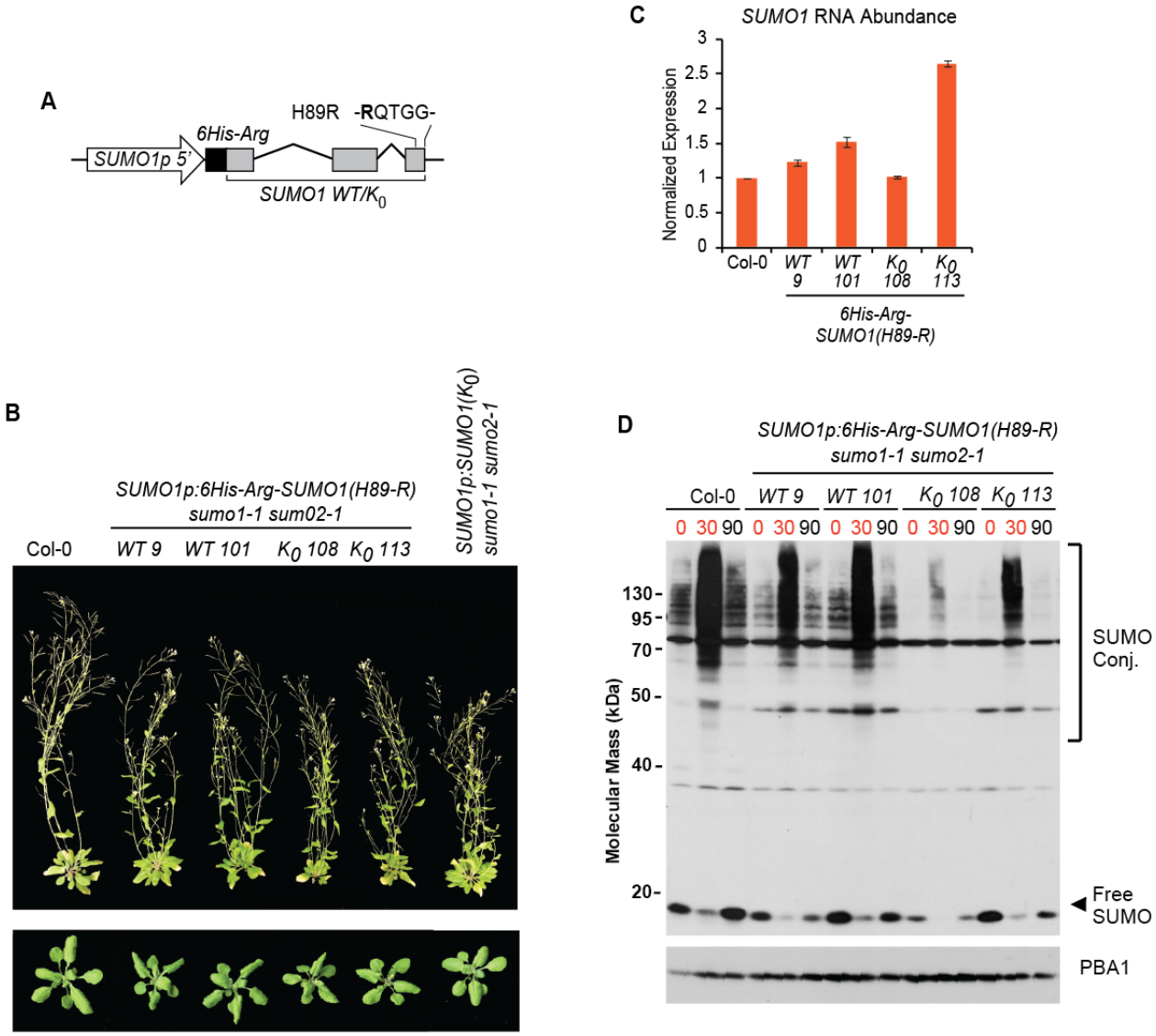

A
Normalized Root Length
1.2
1.0
0.8
0.6
0.4
0.2
0
Control
2μM ABA
50μM SA
50mM Paraquat
50μM MG132
Col-0
WT
K0 2
K0 109
K0 116
K23,42R
B
ABA
SA
IAA
0 uM
0.5 uM
1 uM
2 uM
0 mM
25 mM
50 mM
100 mM
0 uM
0.25 uM
0.5 uM
0.75 uM
C
Paraquat
MG132
0 nM
10 nM
30 nM
50 nM
0 uM
25 uM
50 uM
100 uM

phenotypic differences were observed between the mutant lines and wild-type for the hormone conditions used in this screen. (Figure 4-6A,B). Additionally, I observed the response of the K0 SUMO lines to oxidative and proteosome stress, as both stressors were found to increase SUMO conjugation to proteins in multiple organisms [25, 28, 46, 185, 186]. However, no dissimilar phenotype was observed between the different genotypes upon treatment with paraquat or MG132 (Figure 4-6A,C).

Development of a 6His-Tagged SUMO(K0, H89-R) Line for Footprint Enrichment

As lysine-null SUMO mutants are phenotypically similar to wild-type, I was confident in developing a 6His-tagged SUMO1(K0, H89-R) line that would allow for isolation of SUMOylated peptides for subsequent identification of SUMO footprints through MS. Codons for six His residues followed by an arginine were inserted after the start codon of *SUMO1* in the same genomic construct as was used for the untagged rescue lines (Figure 4-7A). The Arg residue allows for the mass-spectrometric identification of the SUMO N-terminal extension by cleavage of the charged 6His-tag from the peptides during trypsin digest. In previous studies of 6His-SUMO1(H89-R) conjugates, the positively charged histidine residues of the affinity tag prevented the peptide from being analyzed by MS [33]. All seven lysines of SUMO1 were replaced with arginines though PCR-directed mutagenesis. Additionally, the H89-R footprint mutation was added to the construct to allow for detection of SUMO footprints after trypsinization [46]. The *sumo1-1 sumo2-1* mutant was transformed and screened as described for the untagged K0 SUMO lines above. After selection for the transgene and homozygous *sumo1-1* and *sumo2-1* alleles, I isolated three independent insertion lines expressing

Figure 4-7. Phenotypic characterization of plants expressing either wild-type or lysine-null 6His-Arg-SUMO1(H89-R). (A) Gene diagram illustrating the genomic construction used to rescue the *sumo1-1 sumo2-1* embryonic lethal mutant. The 1.5-kbp promoter and 5'-untranslated region (UTR) are represented by the arrow. Exons are shown in dark gray, while the 3' UTR and introns are shown as light gray box and lines, respectively. The black box highlights the 6His-Arg tag. **(B)** Representative plants of wild-type (Col-0) and *sumo1-1 sumo2-1* rescued with either *6His-Arg-SUMO1(H89-R)* or *6His-Arg-SUMO1(K0, H89-R)* construction grown for 20 d (bottom) and 40 d (top) in a LD photoperiod. **(C)** Quantification of *SUMO1* expression level in wild-type and *6His-Arg-SUMO1(H89-R)* and *6His-Arg-SUMO1(K0, H89-R)* lines. The values were normalized to those of *ACT2,* and represented as a ratio to wild-type expression levels. **(D)** Immunoblot analysis of 8-d-old seedlings heat stressed for 30 min at 37°C and collected at the indicated times. The membrane was probed with either anti-SUMO1/2 or anti-PBA1 antibodies (loading control). High molecular mass SUMO conjugates and free SUMO are indicated by the brackets and arrowheads, respectively


A
H89R -RQTGG-
6His-Arg
SUMO1p 5'
SUMO1 WT/K0
C
SUMO1 RNA Abundance
Normalized Expression
3
2.5
2
1.5
1
0.5
0
Col-0
WT 9
WT 101
K0 108
K0 113
6His-Arg-SUMO1(H89-R)
B
SUMO1p:6His-Arg-SUMO1(H89-R)
sumo1-1 sum02-1
SUMO1p:SUMO1(K0)
sumo1-1 sumo2-1
Col-0
WT 9
WT 101
K0 108
K0 113
D
SUMO1p:6His-Arg-SUMO1(H89-R)
sumo1-1 sumo2-1
Col-0
WT 9
WT 101
K0 108
K0 113
0 30 90 0 30 90 0 30 90 0 30 90 0 30 90
Molecular Mass (kDa)
130
95
70
50
40
20
SUMO Conj.
Free SUMO
PBA1

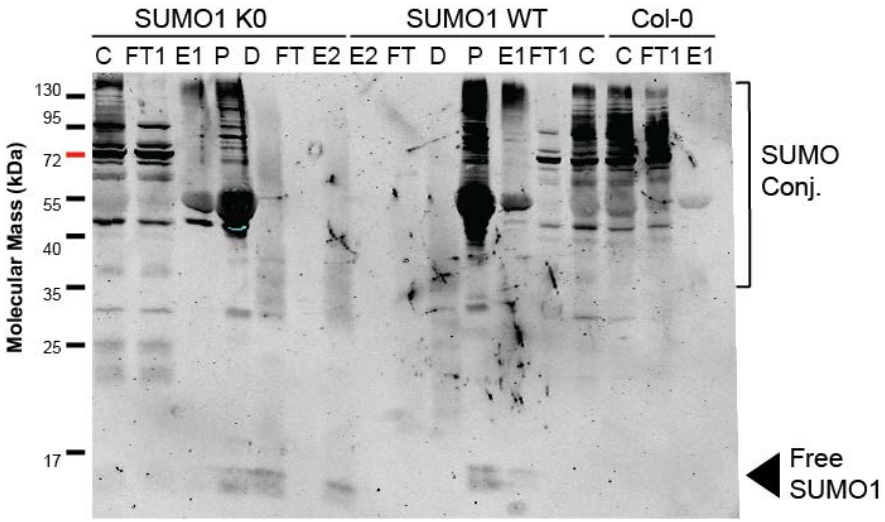
6His-Arg-SUMO1(K0, H89-R) (Figure 4-7B,C). In addition, two transgenic lines expressing 6His-Arg-SUMO(H89-R) were selected as controls for the purification.

The phenotype of the *6His-Arg-SUMO1(K0, H89-R) sumo1-1 sumo2-1* purification lines is comparable to the *6His-Arg-SUMO1(H89-R) sumo1-1 sumo2-1* plants (Figure 7-4B). However, both the 6His-Arg-SUMO(H89-R) and 6His-Arg-SUMO(K0, H89-R) transgenic lines have smaller rosettes and decreased height compared to the wild-type. This is likely to due to a potential negative effect of the 6His-Arg tag on plant growth, and not a result of the lysine mutations as this growth defect is observed in all tagged lines and the rosette size of the untagged K0 SUMO line is comparable to that of the wild-type rosette (Figure 4-7B). It has been observed previously that *Arabidopsis* is sensitive to larger N-terminal tags [33, 37]. While a reduction in growth was observed for the 6His-Arg-tagged lines, both 6His-Arg-SUMO1(H89-R) and 6His-Arg-SUMO1(K0, H89-R) are able to be conjugate to target proteins upon HS, but the amount of SUMO present at higher molecular mass on anti-SUMO immunoblots was reduced in the 6His-Arg-SUMO1(H89-R) K0 seedlings (Figure 4-7D). This decrease was previously observed with the untagged SUMO K0 lines, which do not display the same reduced rosette size, supporting the conclusion that the phenotypes of 6His-Arg-SUMO1 K0 lines are due the interference of the affinity tag and not the lysine mutations.

Enrichment for Peptides Modified by SUMO

To isolate SUMOylated peptides, I employed the same purification strategy as used by Hendricks et al 2014: an initial nickel-affinity chromatography is used to extract SUMO conjugates, followed by a digestion with LysC and a subsequent nickel-affinity step to isolated only peptides modified with SUMO. As seen in Figure 4-8, I was able to successfully enrich for

Figure 4-8. Enrichment of SUMOylated peptides from seedlings expressing *6His-Arg-SUMO1(K0, H89-R)*. Immunoblot analysis of the affinity purification of SUMO conjugated peptides from 8-day old seedlings expressing *6His-Arg-SUMO1(H89-R)* or *6His-Arg-SUMO1(K0, H89-R).* Seedlings were heat-stressed for 30 min at 37°C. The two nickel affinity purification flow-through (FT1 and FT2), the eluate fractions (E1 and E3) and the concentrated first eluate before (P) and after LysC digest (D) were subjected to SDS-PAGE. C, clarified crude extract. As a control, wild-type (Col-0) seedlings were subjected to the same purification protocol. The higher molecular weight SUMO conjugates and free SUMO band are indicated by the brackets and arrowheads, respectively.


SUMO1 K0
SUMO1 WT
Col-0
C FT1 E1 P D FT E2 E2 FT D P E1 FT1 C C FT1 E1
Molecular Mass (kDa)
130
95
72
55
40
35
25
17
SUMO
Conj.
Free
SUMO1

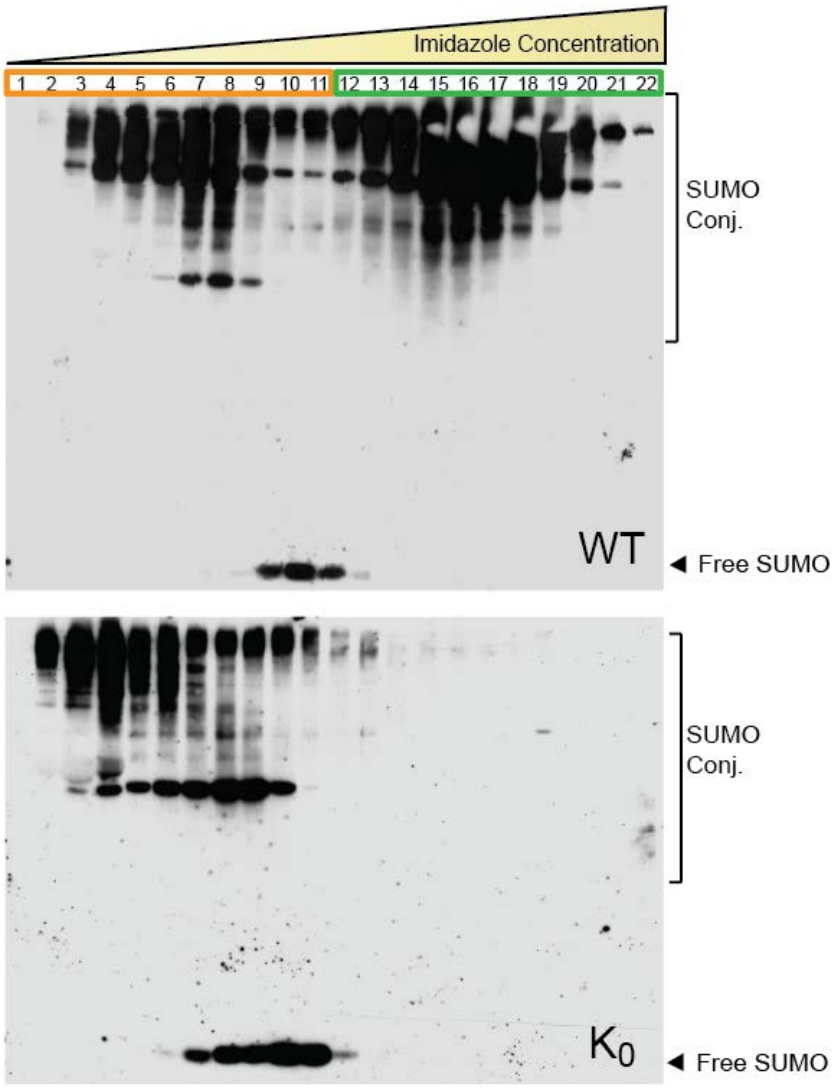
free SUMO (or SUMO bound to a peptide) from seedlings expressing SUMO1(K0, H89-R) with no observable SUMOylated peptides present in the elute of the 6His-SUMO1(H89-R). While the peptidase digestion did not go to completion, as observed through the lower molecular weight smear in the 6His-SUMO1(H89-R) sample, these peptides did not bind to the affinity resin, likely due to the loss of the 6His-tag from a partial digest of the conjugated SUMO moiety.

Identification of Two Distinct Pools of SUMO Conjugates

To observe the loss of polySUMO chain formation in the lysine-null SUMO mutant, I employed an elution strategy to separate proteins based on the number of conjugated SUMO moieties. By using a slow and extended imidazole gradient to elute 6His-tagged SUMO conjugates from nickel-affinity resin, proteins bound by a single 6His-SUMO can be separated from those bound by multiple molecules or polySUMO chains. The more moieties of 6His-SUMO are associated with a protein, the higher its affinity for the nickel resin. Therefore, proteins modified by a single SUMO require a lower imidazole concentration to be released from the resin, while those conjugated with multiple SUMOs or polySUMO chains necessitate a higher imidazole concentration for elution. For this experiment, SUMOylated proteins were first isolated from heat shocked 7-day old seedlings by an initial nickel-affinity chromatography. After removal of the imidazole from the eluant, the conjugates were subjected to a second nickel-affinity purification using an imidazole gradient increase from 30 mM to 250 mM at flow rate of 2.5 mLs/min over 100 minutes.

After anti-SUMO1 immunoblot analysis of the eluant fractions, I identified two distinct pools of SUMO conjugates in wild-type (*6His-Arg-SUMO1(H89-R) sumo1-1 sumo2-1*) (Figure 4-9). An initial pool of SUMOylated proteins elutes in the beginning at low imidazole

Figure 4-9. Fractionation of SUMO modified proteins based on nickel affinity reveals two distinct pools of conjugates. Anti-S1 immunoblot analysis of fractions collected from an extended imidazole gradient elution from nickel-affinity columns. The two distinct pools of SUMO conjugates are illustrated by the orange and green boxes. The higher molecular weight SUMO conjugates and free SUMO band are indicated by the brackets and arrowheads, respectively.


Imidazole Concentration
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
SUMO Conj.
WT
Free SUMO
SUMO Conj.
K0
Free SUMO

concentrations and is followed by free SUMO. A second distinct fraction of SUMO conjugates elutes at higher imidazole concentration. In the K0 SUMO mutant (*6His-Arg-SUMO1(K0,H89-R) sumo1-1 sumo2-1*), only the initial pool of SUMOylated proteins was detected. As the second fraction is not present in the K0 SUMO mutant, it is likely to represent targets modified with polySUMO chains.

DISCUSSION AND FUTURE DIRECTIONS

In mammals and yeasts, polySUMO chains are thought to have significant roles in cellular biology with involvement in many essential processes, such as protein ubiquitylation and degradation, DNA replication, and chromosome segregation [48, 90, 97, 187] by presenting a unique binding site for interacting proteins, including SUMO proteases and SUMO targeted ubiquitin ligases [19, 101, 183, 184, 188].

While SUMO chain formation has been observed in *A. thaliana* [33, 36], its role in plant biology is unknown. In this chapter, I investigated the importance of polySUMO chains in Arabidopsis. I rescued the embryonic lethal *sumo1-1 sumo2-1* mutant with a lysine-null SUMO that prevents the formation of polySUMO chains. This is the first time endogenous SUMO was replaced with a lysine-null version in a multi-cellular organism. Surprisingly, I found that polySUMO chains are not essential for plant viability. The lysine-null SUMO mutants are phenotypically similar to wild-type and accumulate SUMO conjugates upon heat shock. Additionally, their sensitivity to a variety of stresses is comparable to wild-type. In *S. pombe*, mutants unable to form polySUMO chains were sensitive to HU implying a role of SUMO chains in DNA damage and replicative stress [50]. However, Arabidopsis SUMO K0 lines had similar sensitivities to DNA damaging agents as wild-type. This is comparable to observations

in *S. cerevisiae* demonstrating that a lack of polySUMO chains causes no increased sensitivity to stress conditions, including genotoxic treatments [49].

Additionally, the K0 SUMO lines and wild-type respond similarly to stresses that have been found to increase protein SUMOylation, such as oxidative and proteosomal stresses, or to treatments affecting SUMO pathway mutants, such as prolonged exposure to moderately high temperatures and ABA. Although I was unable to distinguish a phenotype associated with a lack of polySUMO chain formation with the conditions tested, there could potentially still be a change in the molecular biology of the plants.

The lysine-null SUMO mutants have reduced amounts of SUMO at higher molecular masses on anti-SUMO immunoblots when exposed to heat shock, likely due to a loss of SUMO chain formation. This is supported by the discover that wild-type seedlings have two distinct pools of SUMO conjugates, but seedlings expressing K0 SUMO have lost the fraction corresponding to proteins potentially modified by polySUMO. Of note is that a loss of SUMO conjugate abundance was also observed in *S. cerevisiae* expressing a lysine-null SUMO [49].

As plants lacking polySUMO chains were phenotypically similar to wild-type and able to accumulate SUMO conjugates in a normal fashion upon HS, it allowed me to replace endogenous SUMO with a 6His-tagged SUMO1(K0, H89-R) for enrichment for SUMOylated peptides and improved MS detection of SUMO footprints. Using a combination of nickel-affinity purifications and LysC digest, I was able to isolate free SUMO (or SUMO bound to a peptide) from seedlings expressing 6His-Arg-SUMO1(K0, H89-R). This purification line is now ready for in-depth analyses of SUMO conjugation sites. Enrichments of peptides modified by SUMO during different cellular challenges, such as HS or treatment with ethanol or hydrogen, could lead to identification of stress-specific conjugation sites. Furthermore, identification of the

modified lysine residues will allow for the determination of plant-specific consensus SUMOylation motifs and lead to better predictions of SUMO targets in plants. Thus, the new purification strategy developed in this chapter will support any future studies of SUMO targets to better define the role of SUMOylation in plants

Although SUMO conjugation to residues other than lysine has not been observed, Ub can be ligated to amino acids other than lysines [189]. Therefore, confirmation is needed that K0 *At*SUMO1 has lost its ability to form poly SUMO chains. Loss of polySUMO chain can be verified through *in vitro* reconstitution of the SUMO conjugation pathway. Additionally, the two distinct pools of SUMO conjugate can be analyzed by MS to determine the presence of polySUMO chains in either fraction. Detection of SUMO modifications on SUMO in the second pool of conjugates will verify the presence of polySUMO chains, and confirm the loss of chain formation in the K0 SUMO mutant.

While I established that polySUMO chains are not required for normal growth of plants, many questions remain about the functional significance of SUMO chain formation in plants, including whether proteins exist which preferentially bind polySUMO chains in plants, as well as their role in ubiquitylation and protein degradation. Continued analysis of the K0 SUMO mutant will provide further insight into polySUMO chain function in plants. Further testing of different cellular challenges, such as UV-induced DNA damage, basal and acquired thermotolerance, additional hormone treatments, could help associate polySUMO chains with specific biological processes, and investigations into the level of ubiquitin conjugates in K0 SUMO plants could connect chain formation and ubiquitin ligation by STUbLs in Arabidopsis. Ultimately, my research will lead to an increased understanding the role of SUMOylation in plants and support the research into specific SUMO targets.

MATERIALS AND METHODS

Plant Materials and Growth Conditions

The *Arabidopsis thaliana* ecotype Columbia (Col-0) was used as the genetic background for all plant lines. The *amiS2 sumo1-1* line was obtained from the Van den Burg lab and is described in [34]. The genomic SUMO1 wild-type and lysine-null (K0) constructs were created by amplification of the *SUMO1* gene with the 3' untranslated region (UTR) and 1.5 kbp of sequence upstream of the start codon, from genomic DNA, by primers flaked by the attB1 and attB2 sequences for insertion into the pDNR221 vector. After ligation into the gateway donor vector, the seven lysines were mutated to arginines through PCR site-directed mutagenesis using the following primers: K9,10-R forward primer, 5'-CAAACCAGGAGGAAGACAGAAGGCCAGGAGAC, K9,10-R reverse primer, 5'-GTCTCCTGGCCTTCTGTCTTCCTCCTGGTTTG, K21,23-R forward primer, 5'-CAATCTCCGAGTCAGGGGACAGGTATCTC, K9,10-R reverse primer, 5'-GTCCC-CTGACTCGGAGATTGATGTGAGCTC, K35-R forward primer, 5'-GGTTTTCTTTAGGA-TCCGGAGAAGCACTCAGCTC, K35-R reverse primer, 5'-GAGCTGAGTGCTTCTC-CGGATCCTAAAGAAAACC, K41,42-R forward primer, 5'-GAAGCACTCAGCTCCGG-CGGCTGATGAATG, and K41,42-R reverse primer, 5'-CATTCATCAGCCGCCGGAGCT-GAGTGCTC. For the SUMO K32,42-R mutant, the following mutagenesis primers were used: K23-R forward primer, 5'- CATCAATCTCAAAGTCCGGGGACAGGTATCTCTC, and K23-R reverse primer, 5'-GAGAGATACCTGTCCCCGGACTTTGAGATTGAT, K42-R forward primer, 5'-GAAGCACTCAGCTCAAGCGGCTGATGAATGCTTAC, and K42-R reverse primer, 5'-TAAGCATTCATCAGCCGCTTGAGCTGAGTGCTTC. A Gateway LR

recombination reaction (Thermo Fisher Scientific) combined the genomic *SUMO1* or *K0 SUMO1* constructs with the pMDC99 plant transformation vector [190]. Plants heterozygous for *sumo1-1* (SAIL_296_C12) and homozygous for *sumo2-1* (Salk_129775), were transformed by the floral dip method [21] and T1 seedlings harboring the *SUMO1* and *K0 SUMO1* transgenes were identified by hygromycin resistance followed by genomic PCR. Selected transformants were allowed to self cross, and F1 seedlings were screened for homozygous *sumo1-1* and *sumo2-1* alleles and the presence of the *SUMO1* transgene by BASTA resistance followed by genomic PCR for all loci.

To construct the 6His-Arg-SUMO1(H89-R) WT and 6His-Arg-SUMO1(K0, H89-R) purification lines, the identical *SUMO1* promoter sequence and *SUMO1* gene, including the 3'-UTR, were amplified separately from genomic DNA. Gateway recombination sites for insertion into pDNR221 and a *XbaI* restriction site were added to the promoter sequence. *XbaI* restriction sites were added to the ends of *SUMO1* and an ATG start codon followed by six histidines and an arginine were inserted before the start codon of *SUMO1*. The ATG-6His-Arg-SUMO1 was inserted into the *XbaI* site of the *pDNR221-SUMO1p*r vector. For the 6His-Arg SUMO1(K0, H89-R) lines, the lysines were mutated to arginines by PCR mutagenesis as described above. The H89-R mutation was added to the *6His-Arg-SUMO1* constructs as described in Miller *et al* 2010. The construct was inserted into the pMDC100 plant transformation vector by Gateway LR recombination reaction [190]. Heterozygous *sumo1-1*, homozygous *sumo2-1* plants were transformed as described above and T1 seedlings harboring the *6His-Arg-SUMO1(H89-R)* and *6His-Arg-SUMO1(K0,H89-R) K0* transgenes were identified by kanamycin resistance followed by genomic PCR. Selected transformants were allowed to self cross and seedlings homozygous

for sumo1-1 and *sumo2-1* and containing the *SUMO1* transgene were screened as described above.

Seeds were surface sterilized with bleach and stratified in water at 4°C in the dark for 2 d before sowing. For plant phenotypic studies, plants were grown at 21°C on soil under long-day photoperiods (LD: 16-hr light, 8-hr dark). For the qPCR analysis of transgenic seedlings, seedlings were grown for 7 d at 21°C under LD photoperiod on solid Gamborg's B-5 Basal Medium (GM; Sigma-Aldrich) supplemented with 2% sucrose, and containing a 0.7% agar base. For the immunoblot analyses and for the purification of SUMO conjugates, seedlings were grown for 7 d in 50 ml liquid cultures containing GM supplemented with 2% sucrose. For the heat stress treatment, the cultures were incubated at 37°C for 30 min in a circulating water bath. At the indicated times, the seedlings were harvested and frozen to liquid nitrogen temperatures.

Plant Stress Conditions

For all conditions tested, seeds were germinated on GM plates supplemented with 2% sucrose and 0.7% agar and transferred to treatment plates after 3 d. For the thermotolerance to moderately high temperatures assay, seedlings were transferred to GM plates supplemented with 2% sucrose and 0.7% agar and grown at 21°C under LD photoperiod for 7 d and transferred to 35°C for the indicated amount of days. After heat treatment, seedlings were allowed to recover at 21°C and plates were imaged and fresh weight was measured when seedlings were 20 d old. For the genotoxic stress treatments, as well as the hormone assays and treatment with other stressors, 3 d-old seedlings were transferred GM plates supplemented with 2% sucrose and to 0.8% agar and the indicated concentration of the specific treatments. For SA and MG132 treatments, the control plates were supplement with ethanol and DMSO, respectively.

Genomic and qPCR Analyses

Genomic analysis of the *sumo1-1 sumo2-1* mutant and the qPCR analysis of *SUMO1* RNA abundance employed the oligonucleotide primers developed by Saracco et al 2007. RNA was extracted from 8-d-old seedlings using the RNeasy Plant Mini Kit (Qiagen) followed by first strand Synthesis with oligo(dT) primers using the SuperScript III First-Strand Synthesis System (Invitrogen-Thermo Fisher Scientific). cDNA and genomic DNA were amplified using EconoTaq Plus Green 2X MasterMix (Lucigen). qPCR was performed with a BioRad CFX Connect Real-Time System together with the LightCycler 480 SYBR Green I Master mix (Roche); transcript abundance was normalized to that generated with ACT2 based on the comparative threshold method [175].

Immunoblot Analyses and Antigenicity Test

Immunodetection of SUMO1/2 conjugates from seedlings used frozen tissue pulverized at liquid nitrogen temperatures, mixed with two volumes per g fresh weight (uL/mg) of twice-strength SDS-PAGE sample buffer, heated to 100°C for 5 min, and clarified at 16,000 Xg. The clarified extracts were subjected to SDS-PAGE and transferred onto Immobilon-P PVDF membranes (EMD-Millipore). The membranes were blocked with non-fat dry milk in PBS, and probed with rabbit anti-SUMO1 antibodies. Rabbit antibodies against the proteasome PBA1 subunit were used as the loading control.

For the antigenicity test of the anti-SUMO antibody, recombinant His-tagged wild-type SUMO1 and lysine-null SUMO1 were isolated from *E. coli* by expression of the active form of SUMO1 (start condon to di-Gly motif) from pDEST17 in Rosetta(DE3)pLysS cells (Novagen – EMDMillipore) followed by purification using nickel-nitrilotriacetic acid agarose (Ni-NTA)

resin (QIAGEN). The antigenicity of the anti-SUMO antibody was quantified by immunoblot analysis of a dilution series of the purified protein with anti-SUMO1 antibodies followed by detection using IRDye 800CW goat anti-rabbit antibodies (LI-COR), and imaged using the 800 nm channel on the LI-COR Odyssey Classic fluorimager. The signal intensities SUMO1 were quantified by the LICOR imaging software and normalized to the background signal.

Purification of SUMOylated Peptides and Imidazole Gradient Elution

For the isolation of SUMOylated peptides from the *6His-Arg-SUMO1(K0, H89-R) sumo1-1 sumo2-1* purification line, a three-step purification strategy was developed based on the SUMO conjugate purification protocol from Miller et al 2010 and the LysC digest procedure from [107]. Approximately 15 g of frozen tissue was pulverized at liquid nitrogen temperatures and resuspended for 1 hr at 55°C in 90 ml of Extraction Buffer (EB: 100 mM Na2HPO4, 10 mM Tris-HCl (pH 8.0), 300 mM NaCl, and 10 mM iodoacetamide IAA)), containing 7 M guanidine-HCl with 10 mM sodium metabisulfate, and 2 mM phenylmethylsulfonyl fluoride (PMSF) added just before use and the pH readjusted to 8.0. The extract was filtered through two layers of Miracloth (EMD Millipore), clarified by centrifugation at 15,000 Xg, and incubated overnight at 4°C with Ni-NTA resin (Qiagen) (0.75 mL resin/5 g of tissue) after addition of imidazole to 10 mM. The Ni-NTA beads were washed sequentially with 10 column volumes of EB (pH 8.0) containing 6 M guanidine-HCl, 0.25% Triton X-100 and 10 mM imidazole, 10 column volumes of EB (pH 6.8) containing 8 M urea, 0.25% Triton X-100 and 10 mM imidazole, and fifteen column volumes of EB (pH 8.0) containing 8 M urea, 0.25% Triton X-100 and 10 mM imidazole. SUMO conjugates were eluted with five column volumes of EB (pH 8.0) containing 8 M urea, and 300 mM imidazole. The eluant was concentrated by ultrafiltration with a 10-kDa

molecular mass cutoff filter (Amicon Ultra-4; EMD Millipore or Vivaspin 6; GE Healthcare Life Sciences). After two exchanges into EB (pH 8.0) containing 8 M urea and reconcentration, mass spectrometric grade LysC (Wako) was added at a factor of 1:100 LysC to protein (by weight). The digest reaction was incubated at room temperature (22°C) for 4 hr in the dark, followed by an the addition of an equal amount LysC and 10mM beta-mercaptoethanol (final concentration) with subsequent incubation over night (approximately 16 hr). The digest was diluted with 6 volumes of EB (pH 8.0) containing 6 M guanidine-HCl, 0.25% Triton X-100 and 10 mM imidazole, and incubated with 500 µL of Ni-NTA resin for 4 hr at 22°C. The Ni-NTA beads were washed sequentially with 10 column volumes of EB (pH 8.0) containing 6 M guanidine-HCl, 0.25% Triton X-100 and 10 mM imidazole, 10 column volumes of EB (pH 6.8) containing 8 M urea, 0.25% Triton X-100 and 10 mM imidazole, and 50 column volumes of EB (pH 8.0) containing 8 M urea and 10 mM imidazole. The bound conjugates were successively eluted with five 500 µL pulses of EB (without IAA) (pH 8.0) containing 6 M urea and 300 mM imidazole.

For the imidazole gradient elution, SUMO conjugates were first isolated from 6His-Arg-SUMO1(H89-R) and 6His-Arg-SUMO1(K0, H89-R) expressing seedlings using a Ni-NTA column as described above. The Ni-NTA elute was concentrated by ultrafiltration with a 10-kDa molecular mass cutoff filter (Amicon Ultra-4; EMD Millipore or Vivaspin 6; GE Healthcare Life Sciences), exchanged twice into EB (pH 8.0) containing 8 M urea and reconcentrated. The final concentrate was resuspended in 30 volumes of EB Buffer (pH 8.0) containing 6 M urea and 10 mM imidazole, and loaded onto a 5mL nickle-loaded Hi-Trap Immobilized Metal ion Affinity Chromatography HP column (GE Life Sciences). After a 10mL wash of EB Buffer (pH 8.0) containing 6 M urea and 10 mM imidazole, the SUMO conjugates were eluted for 100 min using

an 18 to 250 mM imidazole gradient at 2mL/min, and 2.25 mL fractions of were collected. The eluates were subjected to anti-SUMO immunoblot as described above.

CHAPTER 5

CONCLUSIONS AND FUTURE DIRECTIONS

Dr. Robert Augustine assisted in the identification of the *sumov*-3 allele and Joseph Walker expressed and purified the recombinant SUMO-v for antibody production. I completed the remaining experiments described in this chapter.

THESIS CONCLUSION

In preceding chapters, I described my investigations into three distinct aspects of the SUMO conjugation pathway to provide comprehensive insights into the regulation and outcomes of SUMOylation in *Arabidopsis thaliana*. While many questions remain regarding SUMO modifications, I hope the data described in this thesis will support future research into this process and provide tools that can be used to further explore the consequences of SUMO conjugation onto individual targets. With this in mind, I would like to review the different aspects of SUMOylation I investigated to conclude my thesis.

Questions have been raised about the functional roles of the four distinct SUMO isoforms expressed in *Arabidopsis thaliana*. The almost identical proteins SUMO1 and SUMO2 represent the main forms of SUMO used in protein conjugation and are essential for embryogenesis; while the roles of SUMO3 and SUMO5 in plant biology are unknown, and no *in vivo* targets have been identified. To establish conservation of SUMO3 and SUMO5 in plants, which can indicate a functional requirement, I examined the phylogenetic relationships of plant SUMO isoforms in Chapter 2. Through my analysis of over one hundred SUMO sequences from a diverse list of plant species, I discovered that monocots and dicots contain two distinct clades of SUMOs: one group of highly conserved, canonical SUMOs that share 80% or greater protein identities to *At*SUMO1, and a second class of more divergent, non-canonical isoforms, including *At*SUMO3 and *At*SUMO5, that share less than 50% of their protein identity to *At*SUMO1. These non-canonical SUMOs are lineage specific and have little sequence conservation among each other. While I established that *At*SUMO3 and *At*SUMO5 do not have any homologs outside of the *Brassicaceae* family, most land plants contain at least one non-canonical SUMO, suggesting a functional role for these isoforms that is independent of sequence homology. Therefore, further

investigations into the biological role of *At*SUMO3 and *At*SUMO5 are needed and should focus on the isolation of substrates and/or interacting partners of these SUMOs. Through my analysis of SUMO homologs, I also identified two additional distinct SUMO family proteins; a di-SUMO-like (DSUL) protein present only in grasses and a SUMO variant (SUMO-v) found in all land plants. While DSUL has been investigated in maize [141], SUMO-v represents an uncharacterized, plant-specific protein possessing an approximately 100 amino-acid-long N-terminal extension preceding a SUMO-like beta-grasp fold lacking a C-terminal di-glycine motif. These unique characteristics, makes SUMO-v an interesting subject for further investigation.

The identification of proteins modified by each SUMO E3 ligase is a crucial step towards defining how SUMO addition alters the activity, interactions, location and/or half-life of these targets, and ultimately how these modifications impact plant growth, development and stress protection. In Chapter 3, to catalog the targets of the two *Arabidopsis* SUMO ligases SAP and MIZ1 ligase (SIZ)-1 and METHYL METHANESULFONATE-SENSITIVE (MMS)-21 (or HIGH PLOIDY (HPY)-2), I crossed the respective null mutants (*siz1-2* and *mms21-1*) into the 6His-SUMO1(H89-R) purification background, which allowed for the identification of proteins no longer SUMOylated in the mutants. While I was unable to detect potential targets of MMS21, I isolated over one hundred SIZ1-influenced conjugates upon heat shock. These substrates were highly enriched in transcription factors, co-repressors and chromatin regulators connected to abiotic and biotic stresses responses, and included multiple members of the TOPLESS co-repressor complex and the SWI-SNF chromatin remodeling complex. Recent studies in mammals and yeast have associated the presence SUMO at specific genes upon stress with a change in the transcription levels of stress-responsive genes required for cell survival

[139]. Thus, SIZ1-directed SUMOylation could provide stress protection by modifying a large array of key nuclear regulators to alter transcription of stress-induced genes.

Surprisingly, my research in Chapter 3 lead to the discovery that the *siz1-2* mutant allele is not a true null, as assumed in previous publications by multiple labs, but an attenuated mutant, which could have implications on past and future experiments using this allele. The possibility exists that a true SIZ1 null plant is not viable, as no mutant alleles upstream of the SP-RING domain have been characterized to date.

In Chapter 4, I addressed the role of polySUMO chain formation in *Arabidopsis* and developed a plant purification line that will allow for increased detection of modified lysines on SUMO targets through tandem mass spectrometry. Although the modification of SUMO with other SUMO moieties was detected in plants, the function of these chains was unknown. By replacing endogenous SUMO with a lysine-null (K0) version unable to form chains, I found that plants lacking polySUMO chains are comparable to wild-type plants not only in their growth under normal conditions, but also in their response to cellular stressors such as high temperatures and DNA damage, and to hormone treatments. Thus, I was unable to assign an essential role to polySUMO chains in plants with the conditions tested. However, as no aberrant phenotypes were detected in the plants expressing only K0 SUMO1, I was able to develop a purification line expressing a 6His- tagged SUMO1(K0, H89-R) that will allow for the site-specific investigation of SUMO conjugation to proteins. This plant line is now being used in the Vierstra lab to identify the SUMO conjugation sites on proteins under various stress conditions, and will in future permit the targeted study of individual conjugates to expand our understanding of the roles of SUMOylation in plants.

FUTURE DIRECTIONS

In this section, I will describe two studies which represent the next steps in some of my thesis research but are not sufficiently developed to be included in the thesis chapters. First, I will describe the connection between polySUMO chain formation and ubiquitylation. As has been mentioned in this thesis, SUMO-targeted ubiquitin ligases (STUbLs) have been identified in all eukaryotic organisms. The mammalian STUbL, RNF4, specifically binds SUMO chains for ubiquitylation of SUMO modified proteins. It was discovered that Arabidopsis SUMO conjugates are also ubiquitylated during heat stress and SUMO-ubiquitin conjugates have been identified by mass spectrometric analyses (33, Miller et al. *unpublished*). Additionally, it was shown that ubiquitin (Ub) is attached to SUMO and not vice versa (Miller et al. *unpublished*). However, it is unknown how ubiquitylation is targeted to SUMO conjugates. Here, I investigated the role of polySUMO chains in this process through the analysis of the lysine-null (K0) SUMO1 mutant described in Chapter 4.

Second, I will describe the genetic and phenotypic analysis of mutants alleles of Arabidopsis SUMO-v identified in Chapter 2. As SUMO-v is a conserved plant-specific SUMO-like protein with distinctive structural properties, it is an interesting subject for further investigation. The phenotypic analyses of null mutants could establish a role of SUMO-v protein in plant biology.

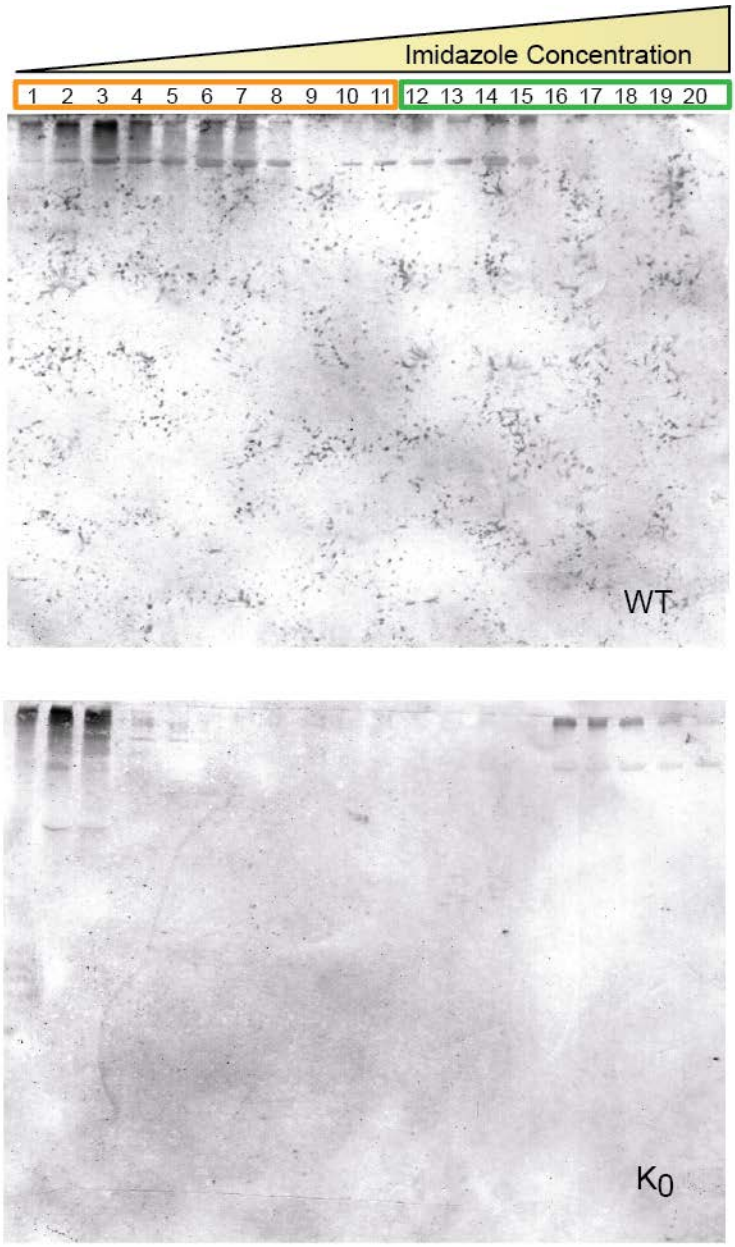
Ubiquitylation of Lysine-Null SUMO Conjugates

In Chapter 4, I identified two distinct SUMO conjugate pools in seedlings expressing a 6His-tagged SUMO1 version by subjecting SUMO conjugates to an extended imidazole gradient elution separating proteins based on the number of bound SUMO moieties (Figure 4-9). An

initial pool of SUMOylated proteins elutes in the beginning at low imidazole concentrations and includes proteins bound by a single SUMO. A second distinct fraction of conjugates, eluting at higher imidazole concentration, likely consists of conjugates bound by multiple moieties or polySUMO chains and is absent in seedlings expressing only a 6His-tagged K0 SUMO1. To ascertain if polySUMO chains are required for the ubiquitylation of SUMO conjugates in *Arabidopsis*, I analyzed the Ub profile of these eluant fractions by immunoblot analysis from seedlings expressing either 6His-tagged SUMO1 or 6His-tagged K0 SUMO1 (Figure 5-1). I expected to find most of the ubiquitylated SUMO substrates in the second pool of SUMO conjugates, however a large fraction of Ub substrates elute in the beginning with the in the first pool of SUMOylated proteins. Additionally, Ub conjugates were still detected in the SUMO mutant and elute in the same first few fractions with a large portion of K0 SUMO conjugates (see Figure 4-9)

These observations suggest that SUMO conjugates can be modified by ubiquitin regardless if they are bound by single SUMOs or polySUMO chains in Arabidopsis. Therefore, plant STUbLs are likely targeted to SUMO conjugates through alternate mechanisms than polySUMO chains. The mammalian STUbL RNF4 selectively binds SUMO chains through a tandem repeat of four SUMO interacting domains (SIMs) [101, 191]. Yeast STUbLs also contain multiple SIMS in their sequence, but these motifs are not as closely spaced as on RNF4 [192, 100, 185]. However, Arabidopsis STUbLs have at most two SIMs that are never located in the same region of the protein [98], and thus, might not require chains for ubiquitylation of SUMO conjugates. Nevertheless, further investigations into the role of SUMO chains and ubiquitylation of SUMO conjugates are essential. First, confirmation is needed that the ubiquitylated proteins observed in Figure 5-1 are SUMO conjugates and not ubiquitylated contaminants that bound

Figure 5-1. Ubiquitin profile of SUMO conjugates fractionated based on nickel-resin affinity. Anti-Ub immunoblot analysis of fractions collected from an extended imidazole gradient elution from nickel-affinity columns. The two distinct pools of SUMO conjugates are illustrated by the orange and green boxes.


Imidazole Concentration
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
WT
K0

non-specifically to the nickel resin. Additionally, a comparison the SUMO-Ub conjugates of the wild-type and the K0 SUMO mutant might detect anomalies in the ubiquitylation of SUMO conjugates due to loss of chains.

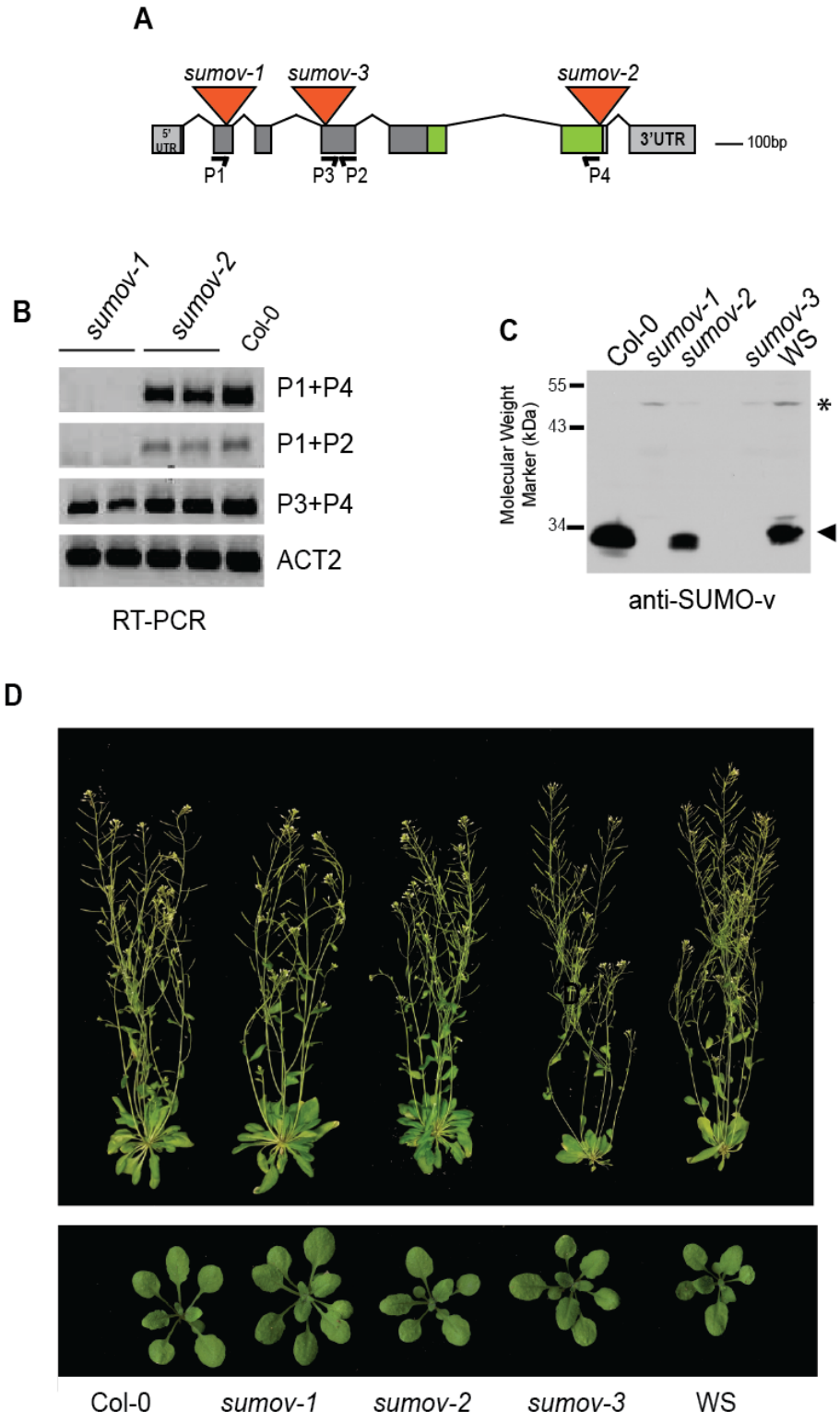
Phenotypic Analysis of SUMO-v Mutants

To investigate the role of SUMO-v in Arabidopsis, I identified and analyzed the phenotypes of two SUMO-v null mutants. Through searches of the Nottingham Arabidopsis Stock Centre (NASC: http://arabidopsis.info) and the Versailles Arabidopsis Stock Center (http://publiclines.versailles.inra.fr/) databases, three separate TDNA insertional mutants were identified in *SUMO-v*. Two alleles, *sumov-1* and *sumov-2,* are in the Col-0 background and the *sumov-3* allele is in the WS background. The TDNA insertions for *sumov-1* and *sumov-3* are located in the 5' end of the gene, which codes for the N-terminal extension (Figure 5-2A). The *sumov-2* insertion occurs at the 3' end of the coding sequence a few codons upstream of the translational stop (Figure 5-2A).

Using RT-PCR and anti-SUMO-v immunoblot analysis, *sumov-1* and *sumov-3* were identified as null mutants. Neither of the mutants produced a full length mRNA, and no SUMO-v protein was detected in the mutants by immunoblot analysis (Figure 5-2 B,C; RT-PCR data not shown for *sumov-3*). The *sumov-2* allele still transcribed a *SUMO-v* mRNA product that is translated into a polypeptide, however the product is slightly smaller than wild-type, indicating that the sumov-3 protein is truncated at the C-terminal end (Figure 5-2C).

SUMO-v is expressed ubiquitously at moderate levels in both vegetative and developmental tissues in Arabidopsis (ThaleMine: https://apps.araport.org/thalemine/portal.do?externalids=AT1G68185) suggesting a prevalent role on plant growth. However,

Figure 5-2. Phenotypic characterization of SUMO-v mutants. (A) Gene diagram illustrating the *SUMO-v* (AT1G68185. 1). Exons are shown in dark gray, while the 5' and 3'-untranslated regions (UTRs) and introns are shown as light gray box and lines, respectively. The TDNA insertion sites are illustred by the orange triangles. The black arrows indicate the primers used for RT-PCR analysis in panel (B). (B) RT-PCR analysis of the expression of *SUMO-v* in the wild-type (Col-0) and *sumov-1* and *sumov-2*. The expression of *ACTIN2* (ACT2) was used as a positive control) (C) Anti-SUMO-v immunoblot analysis of 8-d-old, plate-grown seedlings of the SUMO-v mutants and their respective wild-types (Col-0 and WS). A contaminating band and the band corresponding to unmodified SUMO-v are indicated by the brackets and arrowheads, respectively. (D) Representative wild-type (Col-0 and WS) and *sumov-1, sumov-2* and *sumov-3* plants grown for 20 d (bottom) and 40 d (top) in a LD photoperiod.


A
sumov-1
sumov-3
sumov-2
5' UTR
3'UTR
100bp
P1
P3
P2
P4
B
sumov-1
sumov-2
Col-0
P1+P4
P1+P2
P3+P4
ACT2
RT-PCR
C
Col-0
sumov-1
sumov-2
sumov-3 WS
Molecular Weight Marker (kDa)
55
43
34
*
anti-SUMO-v
D
Col-0
sumov-1
sumov-2
sumov-3
WS

SUMO-v null mutants in two different ecotypes are phenotypically similar to the corresponding wild-types (Figure 5-2D). Additionally, I subjected the SUMO-v null mutants to a variety of cellular stresses, including moderately high temperatures, DNA damaging agents, oxidative stress, and proteosome inhibition, as well as plant hormone treatments. The responses of the mutant seedlings were comparable to the wild-type for all conditions tested. (Figure 5-3, Figure 5-4 and Figure 5-5).

Furthermore, I observed no differences in the conjugation of SUMO to proteins upon heat shock between *sumov-1*, *sumov*-2 and wild-type (Figure 5-6A) suggesting SUMO-v is not required for proper SUMOylation of targets during this stress response. Additionally, the SUMO-v protein levels did not change in abundance during heat shock, however a slight increase in SUMO-v was observed during the recovery period after stress (Figure 5-6B). Although a control is needed to ensure equal loading for the immunoblot, the increase in protein abundance indicates the expression of *SUMO-v* is elevated during stress recovery and should be confirmed through quantitative PCR.

While I was unable to associate a phenotype to the loss of SUMO-v, further testing of the mutants with additional stressors, such as UV-induced DNA damage, basal and acquired thermotolerance and additional hormone treatments, could help elucidate the role of SUMO-v in Arabidopsis.

SUMO-v has been suggested to be the plant equivalent of the SUMO domain-containing proteins Rad60 in *Schizosaccharomyces pombe*, Esc2 in *Saccharomyces cerevisiae* and Nip45 in mammals (or RENi family) [143]. The RENi family proteins are characterized by a 200 amino acid flexible region containing SIMs followed by two distinct SUMO-like domains [143, 147, 193, 184]. These proteins are involved in the response to DNA damage, as their loss results in

Figure 5-3 Seedlings expressing lacking SUMO-v have similar responses to various hormone treatments, proteosome inhibition, and oxidative stress as wild-type. (A) Dosage response of seedlings treated to indicated concentration of stressors. Seedlings were grown on agar plates at 21°C in long day (16hr light 8 hr dark) conditions *sumov-1 #2* and *sumov-1 #18* represent siblings.

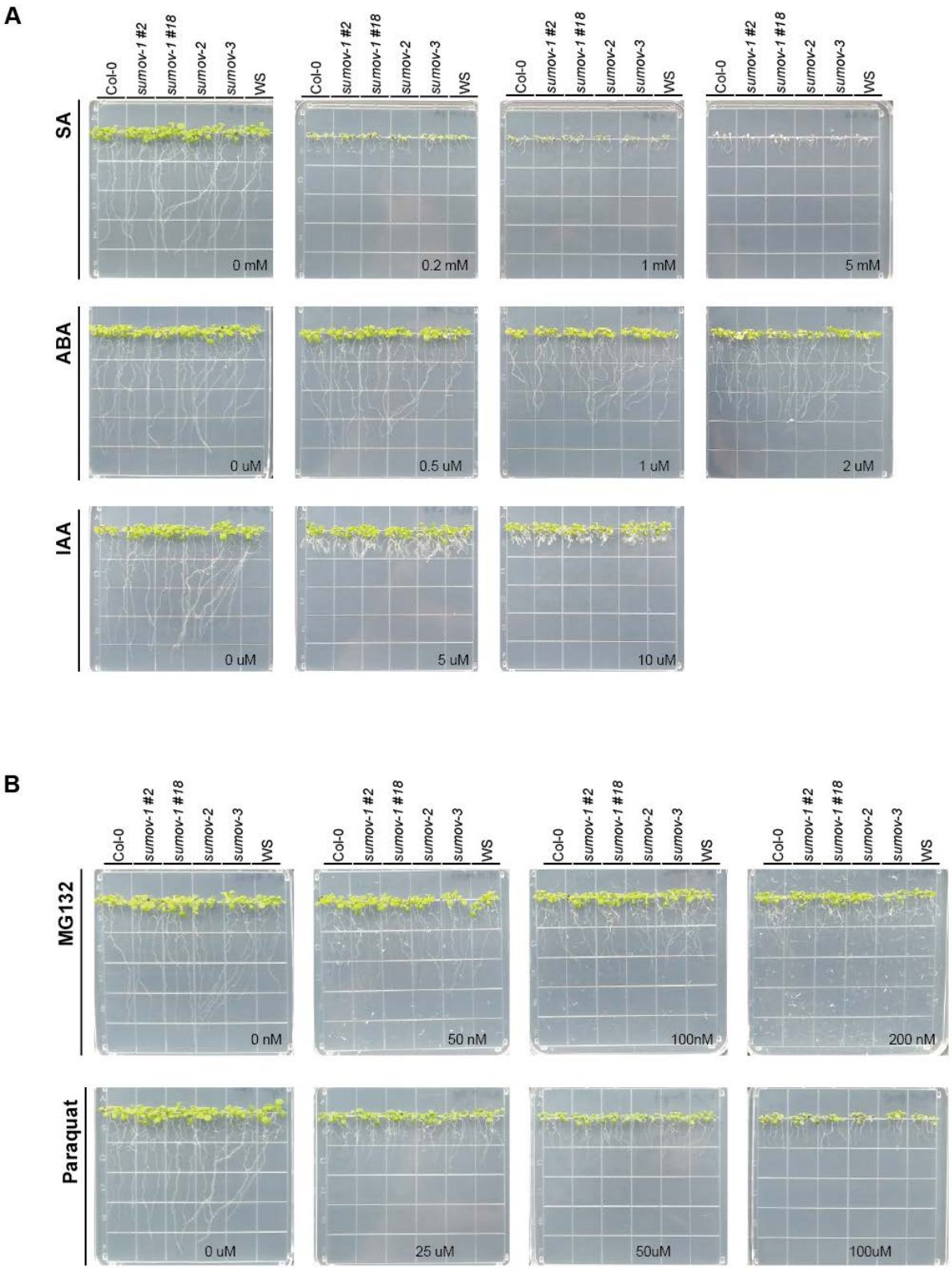

A
SA
ABA
IAA
Col-0
sumov-1 #2
sumov-1 #18
sumov-2
sumov-3
WS
0 mM
0.2 mM
1 mM
5 mM
0 uM
0.5 uM
1 uM
2 uM
0 uM
5 uM
10 uM
B
MG132
Paraquat
0 nM
50 nM
100nM
200 nM
0 uM
25 uM
50uM
100uM

Figure 5-4. Seedlings expressing lacking SUMO-v have similar sensitivities to DNA damage as wild-type. (A) Dosage response of seedlings treated to indicated concentration of DNA damaging agents. Seedlings were grown on agar plates at 21°C in long day (16hr light 8 hr dark) conditions *sumov-1 #2* and *sumov-1 #18* represent siblings.

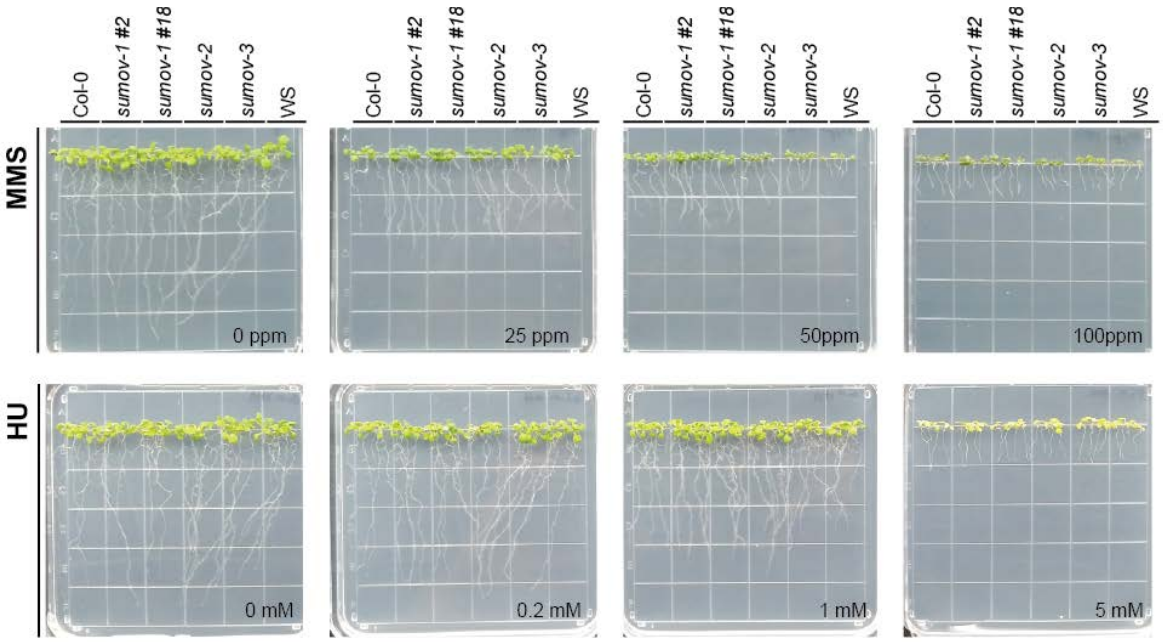

Col-0
sumov-1 #2
sumov-1 #18
sumov-2
sumov-3
WS
MMS
HU
0 ppm
25 ppm
50ppm
100ppm
0 mM
0.2 mM
1 mM
5 mM

Figure 5-5. SUMO-v mutants are comparable to wild-type in their response to moderately high temperatures. (A) Phenotypic analysis of seedlings treated to moderately high temperature. Seedlings were grown on agar plates at 21°C in long day (LD: 16hr light 8 hr dark) conditions, transferred to 35°C for the indicated number of days, and allowed to recover at 21°C. Plates were image after 20 d of total growth. (B) Quantification of the response to the moderately high temperatures based on seedling fresh weight. 20-d old seedlings exposed to 35°C for the indicated number of days were weighed and the results reported as percentage of the total weight of seedlings for each genotype. Each data point represents an average of at least three biological replicates, and error bars measure standard deviation. The same plants lines were used as in panel (A).

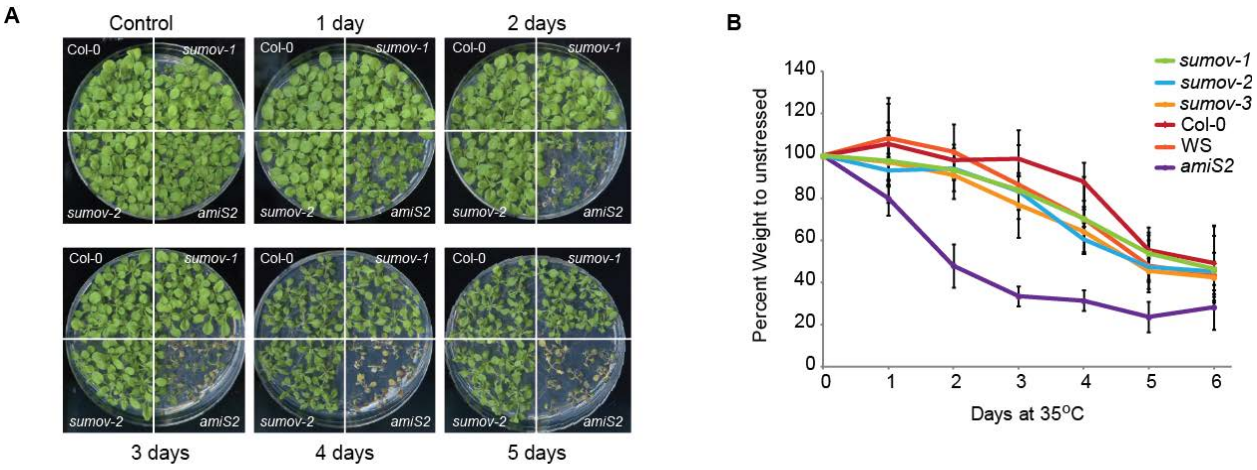
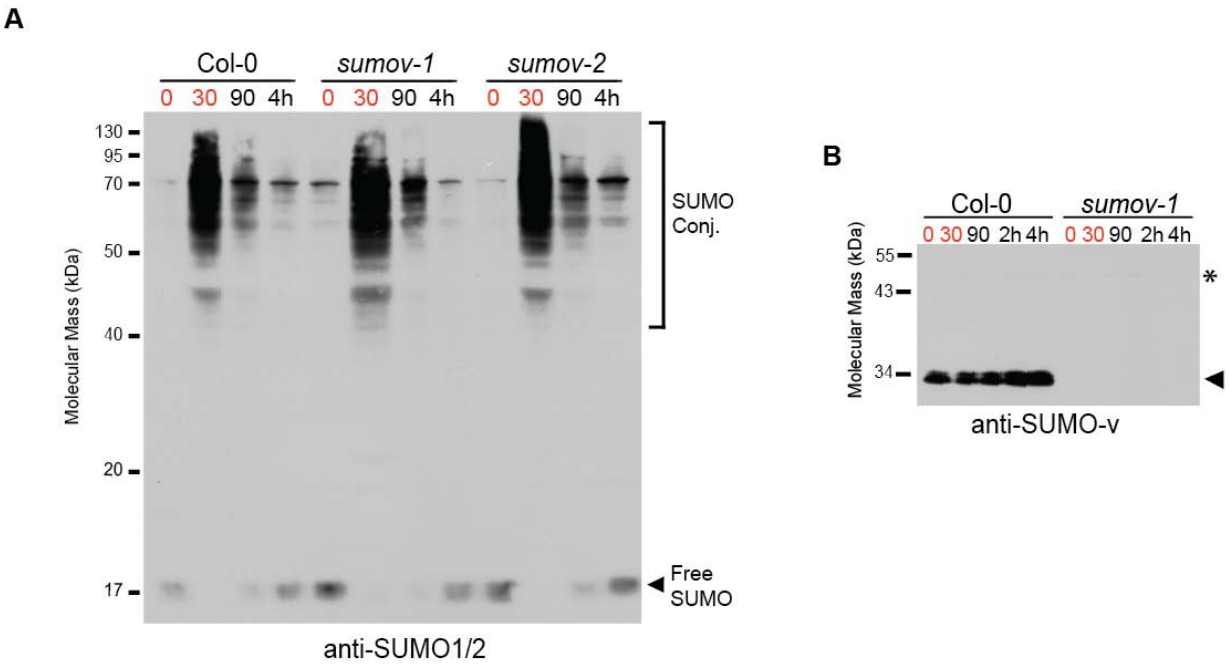

A
Control
1 day
2 days
Col-0
sumov-1
sumov-2
amiS2
3 days
4 days
5 days
B
Percent Weight to unstressed
Days at 35°C
sumov-1
sumov-2
sumov-3
Col-0
WS
amiS2

Figure 5-6. Immunoblot analysis of seedlings exposed to heat shock. (A) Anti-SUMO1 immunoblot analysis of 8-day-old, plate-plate grown seedlings exposed to a 30 min heat shock (HS). Samples were collected at the indicated times and subjected to SDS-PAGE. The membrane was probed with anti-SUMO1/2 antibodies.). High molecular mass SUMO conjugates and free SUMO are indicated by the brackets and arrowheads, respectively. (B) Anti-SUMO-v immunoblot analysis from HS time course from panel (A). A contaminating band and the band corresponding to unmodified SUMO-v are indicated by the brackets and arrowheads, respectively.


A
Col-0
sumov-1
sumov-2
0 30 90 4h 0 30 90 4h 0 30 90 4h
130
95
70
50
40
20
17
Molecular Mass (kDa)
SUMO Conj.
Free SUMO
anti-SUMO1/2
B
Col-0
sumov-1
0 30 90 2h 4h 0 30 90 2h 4h
55
43
34
Molecular Mass (kDa)
*
anti-SUMO-v

increased sensitivities to DNA damage [193, 145, 144, 148, 194]. However, the Arabidopsis SUMO-v null mutants did not have an increased sensitivity to DNA damaging agents compared to WT, suggesting SUMO-v is not involved in DNA repair like the RENi proteins.

Interestingly, a yeast two-hybrid screen identified SUMO-v as an interactor of the SUMO conjugating enzyme SCE1. Rad60 and Nip45 were found to interact non-covalently with the E2 enzyme through one of their SUMO-like domains to inhibit SUMO chain formation [184, 195]. Thus, *Arabidopsis* SUMO-v could serve a similar role in plants regulating polySUMO chain formation. Further studies of SUMO-v are clearly warranted. For example, over-expression of the gene or identification of SUMO-v interacting partners through immunoprecipitation could provide further insight into the processes involving this SUMO-domain containing protein.

MATERIALS AND METHODS

Anti-Ub Immunoblot Analysis of K0 SUMO Mutant

The eluant fractions of the extended imidazole gradient elution of SUMO conjugates from plants expressing either 6His-Arg-SUMO1(H89-R) or 6His-Arg-SUMO1(K0, H89-R) (see Chapter 4 for methods) were subjected to SDS-PAGE and subsequent anti-Ub immunoblot analysis as described in [196].

Plant Materials and Growth Conditions for SUMO-v Mutants

The *A. thaliana* ecotype Col-0 T-DNA insertion mutants *sumov-1* (GK-161H01) and *sumov-2* (GK-110E01) were identified in the Nottingham Arabidopsis Stock Centre (NASC:

http://arabidopsis.info) and the *A. thaliana* ecotype WS T-DNA insertion mutant *sumov-3* (DY30) from the Versailles Arabidopsis Stock Center (http://publiclines.versailles.inra.fr/).

Seeds were surface sterilized with bleach and stratified in water at 4°C in the dark for 2 d before sowing. For plant phenotypic studies, plants were grown at 21°C on soil under long-day photoperiods (LD: 16-hr light, 8-hr dark). For the RT-PCR analysis and heat shock treatment, seedlings were grown for 8 d at 21°C under continuous light on solid Gamborg's B-5 Basal Medium (GM; Sigma-Aldrich) supplemented with 2% sucrose and containing a 0.7% agar base. For the heat shock treatment, the plates were incubated at 37°C for 30 min in a circulating water bath and allowed to recover at 21°C. At the indicated times, the seedlings were harvested and frozen to liquid nitrogen temperatures.

Genomic and qPCR Analyses

The genotypes of the mutants were analyzed by PCR of genomic DNA with 5'-and 3'-gene-specific primers together or in combination with T-DNA-specific left-border primer o8409 (5'-ATATTGACCATCATACTCATTGC) for *sumov-1* or *sumov-2* lines. Gene-specific primers are the following SUMO-v1 (5'-CATGTAGAGCTTCCCATGGAC) and SUMO-v1 (5'-CAAGCCCACAATACCAAAAG) used for *sumov-1* and SUMO-v3 (5' TGCTTGCTCTCTGTCATACAATG) and SUMO-v4 (5'- AGAGAGATCCCTTCCCATGC) used for *sumov-2*.

RNA was extracted from 8-d-old plate grown seedlings using the RNeasy Plant Mini Kit (Qiagen) followed by first strand Synthesis with oligo(dT) primers using the SuperScript III First-Strand Synthesis System (Invitrogen-Thermo Fisher Scientific). cDNA and genomic DNA were amplified using EconoTaq Plus Green 2X MasterMix (Lucigen). The following primers

were used for RT-PCR: P1 (5’- GGAGAAGATCTAGAGCCACTTTTTG), P2 (5’- CTCCTCAACTTCCTGCATCAC), P3 (5’- GTGATGCAGGAAGTTGAGGAG) and P4 (5’- TCATATCATGATCCTCCATGC).

Plant Stress Conditions

For all conditions tested, seeds were germinated on 0.7% agar GM plates supplemented with 2% sucrose and transferred to treatment plates after 3 d. For the thermotolerance to moderately high temperatures assay, seedlings were transferred to GM plates supplemented with 2% sucrose and 0.7% agar grown at 21°C under LD photoperiod for 7 d and transferred to 35°C for the indicated amount of days. After heat treatment, seedlings were allowed to recover at 21°C and plates were imaged and fresh weight was measured when seedlings were 20 d old. For the genotoxic stress treatments, as well as the hormone assays and treatment with other stressors, 3 d-old seedlings were transferred to GM plates supplemented with 2% sucrose and 0.8% agar the indicated concentration of the specific treatments, and grown vertically at 21°C under LD photoperiod for 7 d. For SA and MG132 treatments, the control plates were supplement with ethanol and DMSO, respectively.

Immunoblot Analyses

Immunodetection of SUMO1/2 conjugates and SUMO-v from seedlings used frozen tissue pulverized at liquid nitrogen temperatures, mixed with two volumes per g fresh weight (uL/mg) of twice-strength SDS-PAGE sample buffer, heated to 100°C for 5 min, and clarified at 16,000 Xg. The clarified extracts were subjected to SDS-PAGE and transferred onto Immobilon-P PVDF membranes (EMD-Millipore). The membranes were blocked with non-fat

dry milk in PBS, and probed with rabbit anti-SUMO1 or anti-SUMO-v antibodies. The polyclonal anti-SUMO-v antibodies were raised against 6His-tagged full-length recombinant *Arabidopsis* SUMO-v. The coding sequence of *SUMO-v* was amplified from cDNA isolated from A. thaliana Col-0 and cloned into the pET23b vector in frame with the N-terminal 6His-tag. Recombinant protein was expressed in Rosetta(DE3)pLysS *Escherichia coli* cells (Novagen – EMDMillipore) and subsequently purified using nickel-nitrilotriacetic acid agarose (Ni-NTA) resin (QIAGEN).

LITERATURE CITED

1. Vierstra RD (2012) The Expanding Universe of Ubiquitin and Ubiquitin-Like Modifiers. Plant Physiol 160:2–14. doi: 10.1104/pp.112.200667

2. Hay RT (2005) SUMO: a history of modification. Mol Cell 18:1–12. doi: 10.1016/j.molcel.2005.03.012

3. Geiss-Friedlander R, Melchior F (2007) Concepts in sumoylation: a decade on. Nat Rev Mol Cell Biol 8:947–56. doi: 10.1038/nrm2293

4. Anckar J, Sistonen L (2007) SUMO: getting it on. Biochem Soc Trans 35:1409–13. doi: 10.1042/BST0351409

5. Melchior F, Schergaut M, Pichler A (2003) SUMO: Ligases, isopeptidases and nuclear pores. Trends Biochem Sci 28:612–618. doi: 10.1016/j.tibs.2003.09.002

6. Yates G, Srivastava AK, Sadanandom A (2016) SUMO proteases: Uncovering the roles of deSUMOylation in plants. J Exp Bot 67:2541–2548. doi: 10.1093/jxb/erw092

7. Li SJ, Hochstrasser M (1999) A new protease required for cell-cycle progression in yeast. Nature 398:246–251. doi: 10.1038/18457

8. Desterro JMP, Rodriguez MS, Kemp GD, Ronald T H (1999) Identification of the enzyme required for activation of the small ubiquitin-like protein SUMO-1. J Biol Chem 274:10618–10624. doi: 10.1074/jbc.274.15.10618

9. Johnson ES, Schwienhorst I, Dohmen RJ, Blobel G (1997) The ubiquitin-like protein Smt3p is activated for conjugation to other proteins by an Aos1p/Uba2p heterodimer. EMBO J 16:5509–5519. doi: 10.1093/emboj/16.18.5509

10. Johnson ES, Blobel G (1997) Ubc9p is the conjugating enzyme for the ubiquitin-like protein Smt3p. J Biol Chem 272:26799–26802. doi: 10.1074/jbc.272.43.26799

11. Desterro JMP, Thomson J, Hay RT (1997) Ubch9 conjugates SUMO but not ubiquitin. FEBS Lett 417:297–300. doi: 10.1016/S0014-5793(97)01305-7

12. Okuma T, Honda R, Ichikawa G, et al. (1999) In VitroSUMO-1 Modification Requires Two Enzymatic Steps, E1 and E2. Biochem Biophys Res Commun 254:693–698. doi: 10.1006/bbrc.1998.9995

13. Lois LM, Lima CD (2005) Structures of the SUMO E1 provide mechanistic insights into SUMO activation and E2 recruitment to E1. EMBO J 24:439–51. doi: 10.1038/sj.emboj.7600552

14. Stewart MD, Ritterhoff T, Klevit RE, Brzovic PS (2016) E2 enzymes: more than just middle men. Nat Publ Gr 26:423–440. doi: 10.1038/cr.2016.35

15. Bernier-Villamor V, Sampson DA, Matunis MJ, Lima CD (2002) Structural basis for E2-mediated SUMO conjugation revealed by a complex between ubiquitin-conjugating enzyme Ubc9 and RanGAP1. Cell 108:345–356. doi: 10.1016/S0092-8674(02)00630-X

16. Liu Q, Jin C, Liao X, et al. (1999) The binding interface between an E2 (UBC9) and a ubiquitin homologue (UBL1). J Biol Chem 274:16979–16987. doi: 10.1074/jbc.274.24.16979

17. Kagey MH, Melhuish TA, Powers SE, Wotton D (2005) Multiple activities contribute to Pc2 E3 function. EMBO J 24:108–119. doi: 10.1038/sj.emboj.7600506

18. Saitoh H, Pizzi MD, Wang J (2002) Perturbation of SUMOlation Enzyme Ubc9 by Distinct Domain within Nucleoporin RanBP2/Nup358. J Biol Chem 277:4755–4763. doi: 10.1074/jbc.M104453200

19. Mukhopadhyay D, Dasso M (2007) Modification in reverse: the SUMO proteases. Trends Biochem Sci 32:286–295. doi: 10.1016/j.tibs.2007.05.002

20. Li SJ, Hochstrasser M (2000) The yeast ULP2 (SMT4) gene encodes a novel protease specific for the ubiquitin-like Smt3 protein. Mol Cell Biol 20:2367–2377. doi: 10.1128/MCB.20.7.2367-2377.2000

21. Saracco S a, Miller MJ, Kurepa J, Vierstra RD (2007) Genetic analysis of SUMOylation in Arabidopsis: conjugation of SUMO1 and SUMO2 to nuclear proteins is essential. Plant Physiol 145:119–34. doi: 10.1104/pp.107.102285

22. Tian L, Wang J, Fong MP, et al. (2003) Genetic control of developmental changes induced by disruption of Arabidopsis histone deacetylase 1 (AtHD1) expression. Genetics 165:399–409.

23. Takahashi Y, Iwase M, Konishi M, et al. (1999) Smt3, a SUMO-1 homolog, is conjugated to Cdc3, a component of septin rings at the mother-bud neck in budding yeast. Biochem Biophys Res Commun 259:582–7. doi: 10.1006/bbrc.1999.0821

24. Eifler K, Vertegaal ACO (2015) SUMOylation-Mediated Regulation of Cell Cycle Progression and Cancer. Trends Biochem Sci 40:779–793. doi: 10.1016/j.tibs.2015.09.006

25. Lee YJ, Hallenbeck JM (2013) SUMO and ischemic tolerance. NeuroMolecular Med 15:771–781. doi: 10.1007/s12017-013-8239-9

26. Castro PH, Tavares RM, Bejarano ER, Azevedo H (2012) SUMO,a heavyweight player in plant abiotic stress responses. Cell Mol Life Sci 69:3269–3283. doi: 10.1007/s00018-012-1094-2

27. Tempé D, Piechaczyk M, Bossis G (2008) SUMO under stress. Biochem Soc Trans 36:874–8. doi: 10.1042/BST0360874

28. Kurepa J, Walker JM, Smalle J, et al. (2003) The small ubiquitin-like modifier (SUMO) protein modification system in Arabidopsis. Accumulation of SUMO1 and -2 conjugates is increased by stress. J Biol Chem 278:6862–72. doi: 10.1074/jbc.M209694200

29. Hammoudi V, Vlachakis G, Schranz ME, van den Burg HA (2016) Whole-genome duplications followed by tandem duplications drive diversification of the protein modifier SUMO in Angiosperms. New Phytol 211:172–85. doi: 10.1111/nph.13911

30. Augustine RC, York SL, Rytz TC, Vierstra RD (2016) Defining the SUMO System in Maize: SUMOylation Is Up-Regulated during Endosperm Development and Rapidly Induced by Stress. Plant Physiol 171:2191–2210. doi: 10.1104/pp.16.00353

31. Novatchkova M, Tomanov K, Hofmann K, et al. (2012) Update on sumoylation: defining core components of the plant SUMO conjugation system by phylogenetic comparison. New Phytol 195:23–31.

32. Tomanov K, Zeschmann A, Hermkes R, et al. (2014) Arabidopsis PIAL1 and 2 promote SUMO chain formation as E4-type SUMO ligases and are involved in stress responses and sulfur metabolism. Plant Cell 26:4547–60. doi: 10.1105/tpc.114.131300

33. Miller MJ, Barrett-Wilt GA, Hua Z, Vierstra RD (2010) Proteomic analyses identify a diverse array of nuclear processes affected by small ubiquitin-like modifier conjugation in Arabidopsis. Proc Natl Acad Sci U S A 107:16512–7. doi: 10.1073/pnas.1004181107

34. van den Burg HA, Kini RK, Schuurink RC, Takken FLW (2010) Arabidopsis Small Ubiquitin-Like Modifier Paralogs Have Distinct Functions in Development and Defense. Plant Cell 22:1998–2016. doi: 10.1105/tpc.109.070961

35. Elrouby N, Coupland G (2010) Proteome-wide screens for small ubiquitin-like modifier (SUMO) substrates identify Arabidopsis proteins implicated in diverse biological processes. Proc Natl Acad Sci U S A 107:1–6. doi: 10.1073/pnas.1005452107

36. Colby T, Matthäi A, Boeckelmann A, Stuible H-P (2006) SUMO-conjugating and SUMO-deconjugating enzymes from Arabidopsis. Plant Physiol 142:318–32. doi: 10.1104/pp.106.085415

37. Budhiraja R, Hermkes R, Müller S, et al. (2009) Substrates related to chromatin and to RNA-dependent processes are modified by Arabidopsis SUMO isoforms that differ in a conserved residue with influence on desumoylation. Plant Physiol 149:1529–40. doi: 10.1104/pp.108.135053

38. Castaño-Miquel L, Seguí J, Lois LM (2011) Distinctive properties of Arabidopsis SUMO paralogues support the in vivo predominant role of AtSUMO1/2 isoforms. Biochem J 436:581–90. doi: 10.1042/BJ20101446

39. Citro S, Chiocca S (2013) Sumo paralogs: redundancy and divergencies. Front Biosci (Schol Ed) 5:544–53.

40. Saitoh H, Hinchey J (2000) Functional heterogeneity of small ubiquitin-related protein modifiers SUMO-1 versus SUMO-2/3. J Biol Chem 275:6252–6258. doi: 10.1074/jbc.275.9.6252

41. Wang L, Wansleeben C, Zhao S, et al. (2014) SUMO2 is essential while SUMO3 is dispensable for mouse embryonic development. EMBO Rep 15:878–85. doi: 10.15252/embr.201438534

42. Vertegaal ACO, Andersen JS, Ogg SC, et al. (2006) Distinct and overlapping sets of SUMO-1 and SUMO-2 target proteins revealed by quantitative proteomics. Mol Cell Proteomics 5:2298–310. doi: 10.1074/mcp.M600212-MCP200

43. Flotho A, Melchior F (2013) Sumoylation: A Regulatory Protein Modification in Health and Disease. Http://DxDoiOrg/101146/Annurev-Biochem-061909-093311. doi: 10.1146/annurev-biochem-061909-093311

44. Evdokimov E, Sharma P, Lockett SJ, et al. (2008) Loss of SUMO1 in mice affects RanGAP1 localization and formation of PML nuclear bodies, but is not lethal as it can be compensated by SUMO2 or SUMO3. J Cell Sci 121:4106–4113. doi: 10.1242/jcs.038570

45. Golebiowski F, Matic I, Tatham MH, et al. (2009) System-wide changes to SUMO modifications in response to heat shock. Sci Signal 2:ra24. doi: 10.1126/scisignal.2000282

46. Miller MJ, Scalf M, Rytz TC, et al. (2013) Quantitative proteomics reveals factors regulating RNA biology as dynamic targets of stress-induced SUMOylation in Arabidopsis. Mol Cell Proteomics 12:449–63. doi: 10.1074/mcp.M112.025056

47. Tatham MH, Jaffray E, Vaughan OA, et al. (2001) Polymeric Chains of SUMO-2 and SUMO-3 are Conjugated to Protein Substrates by SAE1/SAE2 and Ubc9. J Biol Chem 276:35368–35374. doi: 10.1074/jbc.M104214200

48. Bruderer R, Tatham MH, Plechanovova A, et al. (2011) Purification and identification of endogenous polySUMO conjugates. EMBO Rep 12:142–148. doi: 10.1038/embor.2010.206

49. Bylebyl GR, Belichenko I, Johnson ES (2003) The SUMO isopeptidase Ulp2 prevents accumulation of SUMO chains in yeast. J Biol Chem 278:44113–20. doi: 10.1074/jbc.M308357200

50. Skilton A, Ho JCY, Mercer B, et al. (2009) SUMO chain formation is required for response to replication arrest in S. pombe. PLoS One 4:e6750. doi: 10.1371/journal.pone.0006750

51. Novatchkova M, Budhiraja R, Coupland G, et al. (2004) SUMO conjugation in plants. Planta 220:1–8. doi: 10.1007/s00425-004-1370-y

52. Lin XL, Niu D, Hu ZL, et al. (2016) An Arabidopsis SUMO E3 Ligase, SIZ1, Negatively Regulates Photomorphogenesis by Promoting COP1 Activity. PLoS Genet 12:1–21. doi: 10.1371/journal.pgen.1006016

53. Miura K, Rus A, Sharkhuu A, et al. (2005) The Arabidopsis SUMO E3 ligase SIZ1 controls phosphate deficiency responses. Proc Natl Acad Sci U S A 102:7760–5. doi: 10.1073/pnas.0500778102

54. Cheong MS, Park HC, Hong MJ, et al. (2009) Specific domain structures control abscisic acid-, salicylic acid-, and stress-mediated SIZ1 phenotypes. Plant Physiol 151:1930–42. doi: 10.1104/pp.109.143719

55. Huang L, Yang S, Zhang S, et al. (2009) The Arabidopsis SUMO E3 ligase AtMMS21, a homologue of NSE2/MMS21, regulates cell proliferation in the root. Plant J 60:666–78. doi: 10.1111/j.1365-313X.2009.03992.x

56. Ishida T, Fujiwara S, Miura K, et al. (2009) SUMO E3 ligase HIGH PLOIDY2 regulates endocycle onset and meristem maintenance in Arabidopsis. Plant Cell 21:2284–97. doi: 10.1105/tpc.109.068072

57. Reindle A, Belichenko I, Bylebyl GR, et al. (2006) Multiple domains in Siz SUMO ligases contribute to substrate selectivity. J Cell Sci 119:4749–57. doi: 10.1242/jcs.03243

58. Callis J (2014) The ubiquitination machinery of the ubiquitin system. Arabidopsis Book 12:e0174. doi: 10.1199/tab.0174

59. Kraft E, Stone SL, Ma L, et al. (2005) Genome analysis and functional characterization of the E2 and RING-type E3 ligase ubiquitination enzymes of Arabidopsis. Plant Physiol 139:1597–611. doi: 10.1104/pp.105.067983

60. Hua Z, Zou C, Shiu S-H, Vierstra RD (2011) Phylogenetic Comparison of F-Box (FBX) Gene Superfamily within the Plant Kingdom Reveals Divergent Evolutionary Histories Indicative of Genomic Drift. PLoS One 6:e16219. doi: 10.1371/journal.pone.0016219

61. Murtas G, Reeves PH, Fu Y, et al. (2003) A nuclear protease required for flowering-time regulation in Arabidopsis reduces the abundance of SMALL UBIQUITIN-RELATED MODIFIER conjugates. Plant Cell 15:2308–19. doi: 10.1105/tpc.015487

62. Conti L, Price G, O'Donnell E, et al. (2008) Small ubiquitin-like modifier proteases OVERLY TOLERANT TO SALT1 and -2 regulate salt stress responses in Arabidopsis. Plant Cell 20:2894–908. doi: 10.1105/tpc.108.058669

63. Kong X, Luo X, Qu G-P, et al. (2017) Arabidopsis SUMO protease ASP1 positively regulates flowering time partially through regulating FLC stability. J Integr Plant Biol 59:15–29. doi: 10.1111/jipb.12509

64. Lee J, Nam J, Park HC, et al. (2007) Salicylic acid-mediated innate immunity in Arabidopsis is regulated by SIZ1 SUMO E3 ligase. Plant J 49:79–90. doi: 10.1111/j.1365-313X.2006.02947.x

65. Miura K, Lee J, Miura T, Hasegawa PM (2010) SIZ1 controls cell growth and plant development in Arabidopsis through salicylic acid. Plant Cell Physiol 51:103–13. doi: 10.1093/pcp/pcp171

66. Yoo CY, Miura K, Jin JB, et al. (2006) SIZ1 small ubiquitin-like modifier E3 ligase facilitates basal thermotolerance in Arabidopsis independent of salicylic acid. Plant Physiol 142:1548–58. doi: 10.1104/pp.106.088831

67. Jin JB, Jin YH, Lee J, et al. (2008) The SUMO E3 ligase, AtSIZ1, regulates flowering by controlling a salicylic acid-mediated floral promotion pathway and through affects on FLC chromatin structure. Plant J 53:530–40. doi: 10.1111/j.1365-313X.2007.03359.x

68. Miura K, Jin JB, Lee J, et al. (2007) SIZ1-mediated sumoylation of ICE1 controls CBF3/DREB1A expression and freezing tolerance in Arabidopsis. Plant Cell 19:1403–14. doi: 10.1105/tpc.106.048397

69. Yuan D, Lai J, Xu P, et al. (2014) AtMMS21 regulates DNA damage response and homologous recombination repair in Arabidopsis. DNA Repair (Amst) 21:140–7. doi: 10.1016/j.dnarep.2014.04.006

70. Aravind L (2000) SAP – a putative DNA-binding motif involved in chromosomal organization. Trends Biochem Sci 25:112–114. doi: 10.1016/S0968-0004(99)01537-6

71. Takahashi Y, Kikuchi Y (2005) Yeast PIAS-type Ull1/Siz1 is composed of SUMO ligase and regulatory domains. J Biol Chem 280:35822–8. doi: 10.1074/jbc.M506794200

72. Bienz M (2006) The PHD finger, a nuclear protein-interaction domain. Trends Biochem Sci 31:35–40. doi: 10.1016/j.tibs.2005.11.001

73. Mellor J (2006) It Takes a PHD to Read the Histone Code. Cell 126:22–24. doi: 10.1016/j.cell.2006.06.028

74. Garcia-Dominguez M, March-Diaz R, Reyes JC (2008) The PHD domain of plant PIAS proteins mediates sumoylation of bromodomain GTE proteins. J Biol Chem 283:21469–77. doi: 10.1074/jbc.M708176200

75. Shindo H, Suzuki R, Tsuchiya W, et al. (2012) PHD finger of the SUMO ligase Siz/PIAS family in rice reveals specific binding for methylated histone H3 at lysine 4 and arginine 2. FEBS Lett 586:1783–1789. doi: 10.1016/j.febslet.2012.04.063

76. Yunus A a, Lima CD (2009) Structure of the Siz/PIAS SUMO E3 ligase Siz1 and determinants required for SUMO modification of PCNA. Mol Cell 35:669–82. doi: 10.1016/j.molcel.2009.07.013

77. Rytinki MM, Kaikkonen S, Pehkonen P, et al. (2009) PIAS proteins: Pleiotropic interactors associated with SUMO. Cell Mol Life Sci 66:3029–3041. doi: 10.1007/s00018-009-0061-z

78. Sharrocks AD (2006) PIAS proteins and transcriptional regulation--more than just SUMO E3 ligases? Genes Dev 20:754–8. doi: 10.1101/gad.1421006

79. Ishida T, Yoshimura M, Miura K, Sugimoto K (2012) MMS21/HPY2 and SIZ1, Two Arabidopsis SUMO E3 Ligases, Have Distinct Functions in Development. PLoS One 7:1–10. doi: 10.1371/journal.pone.0046897

80. Andrews EA, Palecek J, Sergeant J, et al. (2005) Nse2, a component of the Smc5-6 complex, is a SUMO ligase required for the response to DNA damage. Mol Cell Biol 25:185–96. doi: 10.1128/MCB.25.1.185-196.2005

81. Potts PR, Yu H (2005) Human MMS21/NSE2 is a SUMO ligase required for DNA repair. Mol Cell Biol 25:7021–32. doi: 10.1128/MCB.25.16.7021-7032.2005

82. Zhao X, Blobel G (2005) A SUMO ligase is part of a nuclear multiprotein complex that affects DNA repair and chromosomal organization. Proc Natl Acad Sci U S A 102:4777–82. doi: 10.1073/pnas.0500537102

83. Xu P, Yuan D, Liu M, et al. (2013) AtMMS21, an SMC5/6 Complex Subunit, Is Involved in Stem Cell Niche Maintenance and DNA Damage Responses in Arabidopsis Roots. Plant Physiol 161:1755–1768. doi: 10.1104/pp.112.208942

84. Stephan AK, Kliszczak M, Morrison CG (2011) The Nse2/Mms21 SUMO ligase of the Smc5/6 complex in the maintenance of genome stability. FEBS Lett 585:2907–2913. doi: 10.1016/j.febslet.2011.04.067

85. Kliszczak M, Stephan AK, Flanagan AM, Morrison CG (2012) SUMO ligase activity of vertebrate Mms21/Nse2 is required for efficient DNA repair but not for Smc5/6 complex stability. DNA Repair (Amst) 11:799–810. doi: 10.1016/j.dnarep.2012.06.010

86. Bermúdez-López M, Pociño-Merino I, Sánchez H, et al. (2015) ATPase-Dependent Control of the Mms21 SUMO Ligase during DNA Repair. PLOS Biol 13:e1002089. doi: 10.1371/journal.pbio.1002089

87. Rodriguez MS, Dargemont C, Hay RT (2001) SUMO-1 conjugation in vivo requires both a consensus modification motif and nuclear targeting. J Biol Chem 276:12654–9. doi: 10.1074/jbc.M009476200

88. Yang S-H, Galanis A, Witty J, Sharrocks AD (2006) An extended consensus motif enhances the specificity of substrate modification by SUMO. EMBO J 25:5083–93. doi: 10.1038/sj.emboj.7601383

89. Matic I, Schimmel J, Hendriks IA, et al. (2010) Site-Specific Identification of SUMO-2 Targets in Cells Reveals an Inverted SUMOylation Motif and a Hydrophobic Cluster SUMOylation Motif. Mol Cell 39:641–652. doi: 10.1016/j.molcel.2010.07.026

90. Ulrich HD (2008) The fast-growing business of SUMO chains. Mol Cell 32:301–5. doi: 10.1016/j.molcel.2008.10.010

91. Kerscher O (2007) SUMO junction-what's your function? New insights through SUMO-interacting motifs. EMBO Rep 8:550–5. doi: 10.1038/sj.embor.7400980

92. Hannich JT, Lewis A, Kroetz MB, et al. (2005) Defining the SUMO-modified proteome by multiple approaches in Saccharomyces cerevisiae. J Biol Chem 280:4102–4110. doi: 10.1074/jbc.M413209200

93. Hecker C, Rabiller M, Haglund K, et al. (2006) Specification of SUMO1- and SUMO2-interacting motifs. J Biol Chem 281:16117–27. doi: 10.1074/jbc.M512757200

94. Song J, Durrin LK, Wilkinson TA, et al. (2004) Identification of a SUMO-binding motif that recognizes SUMO-modified proteins. Proc Natl Acad Sci U S A 101:14373–8. doi: 10.1073/pnas.0403498101

95. Chupreta S, Holmstrom S, Subramanian L, Iñiguez-Lluhí JA (2005) A small conserved surface in SUMO is the critical structural determinant of its transcriptional inhibitory properties. Mol Cell Biol 25:4272–82. doi: 10.1128/MCB.25.10.4272-4282.2005

96. Wilkinson K a, Henley JM (2010) Mechanisms, regulation and consequences of protein SUMOylation. Biochem J 428:133–45. doi: 10.1042/BJ20100158

97. Hay RT (2013) Decoding the SUMO signal. Biochem Soc Trans 41:463–73. doi: 10.1042/BST20130015

98. Elrouby N, Bonequi MV, Porri A, Coupland G (2013) Identification of Arabidopsis SUMO-interacting proteins that regulate chromatin activity and developmental transitions. Proc Natl Acad Sci U S A 110:19956–61. doi: 10.1073/pnas.1319985110

99. Cheng C-H, Lo Y-H, Liang S-S, et al. (2006) SUMO modifications control assembly of synaptonemal complex and polycomplex in meiosis of Saccharomyces cerevisiae. Genes Dev 20:2067–81. doi: 10.1101/gad.1430406

100. Sun H, Leverson JD, Hunter T (2007) Conserved function of RNF4 family proteins in eukaryotes: targeting a ubiquitin ligase to SUMOylated proteins. EMBO J 26:4102–12. doi: 10.1038/sj.emboj.7601839

101. Rojas-Fernandez A, Plechanovová A, Hattersley N, et al. (2014) SUMO chain-induced dimerization activates RNF4. Mol Cell 53:880–892. doi: 10.1016/j.molcel.2014.02.031

102. Zhang XD, Goeres J, Zhang H, et al. (2008) SUMO-2/3 Modification and Binding Regulate the??Association of CENP-E with Kinetochores and??Progression through Mitosis. Mol Cell 29:729–741. doi: 10.1016/j.molcel.2008.01.013

103. Westerbeck JW, Pasupala N, Guillotte M, et al. (2014) A SUMO-targeted ubiquitin ligase is involved in the degradation of the nuclear pool of the SUMO E3 ligase Siz1. Mol Biol Cell 25:1–16. doi: 10.1091/mbc.E13-05-0291

104. Chang CC, Naik MT, Huang YS, et al. (2011) Structural and Functional Roles of Daxx SIM Phosphorylation in SUMO Paralog-Selective Binding and Apoptosis Modulation. Mol Cell 42:62–74. doi: 10.1016/j.molcel.2011.02.022

105. Stehmeier P, Muller S (2009) Phospho-Regulated SUMO Interaction Modules Connect the SUMO System to CK2 Signaling. Mol Cell 33:400–409. doi: 10.1016/j.molcel.2009.01.013

106. Cappadocia L, Mascle XH, Bourdeau V, et al. (2015) Structural and Functional Characterization of the Phosphorylation-Dependent Interaction between PML and SUMO1. Structure 23:126–138. doi: 10.1016/j.str.2014.10.015

107. Hendriks I a, D'Souza RCJ, Yang B, et al. (2014) Uncovering global SUMOylation signaling networks in a site-specific manner. Nat Struct Mol Biol 21:927–36. doi: 10.1038/nsmb.2890

108. Ullmann R, Chien CD, Avantaggiati ML, Muller S (2012) An Acetylation Switch Regulates SUMO-Dependent Protein Interaction Networks. Mol Cell 46:1–12. doi: 10.1016/j.molcel.2012.04.006

109. Catala R, Ouyang J, Abreu I a, et al. (2007) The Arabidopsis E3 SUMO ligase SIZ1 regulates plant growth and drought responses. Plant Cell 19:2952–66. doi: 10.1105/tpc.106.049981

110. Zhou W, Ryan JJ, Zhou H (2004) Global analyses of sumoylated proteins in Saccharomyces cerevisiae. Induction of protein sumoylation by cellular stresses. J Biol Chem 279:32262–32268. doi: 10.1074/jbc.M404173200

111. Zhang S, Qi Y, Liu M, Yang C (2013) SUMO E3 Ligase AtMMS21 Regulates Drought Tolerance in Arabidopsis thaliana. J Integr Plant Biol 55:83–95. doi: 10.1111/jipb.12024

112. Miura K, Okamoto H, Okuma E, et al. (2013) SIZ1 deficiency causes reduced stomatal aperture and enhanced drought tolerance via controlling salicylic acid-induced accumulation of reactive oxygen species in Arabidopsis. Plant J 73:91–104. doi: 10.1111/tpj.12014

113. Crozet P, Margalha L, Butowt R, et al. (2016) SUMOylation represses SnRK1 signaling in Arabidopsis. Plant J 85:120–133. doi: 10.1111/tpj.13096

114. Castro PH, Verde N, Lourenço T, et al. (2015) SIZ1-Dependent Post-Translational Modification by SUMO Modulates Sugar Signaling and Metabolism in Arabidopsis thaliana. Plant Cell Physiol 56:2297–2311. doi: 10.1093/pcp/pcv149

115. Yoshida T, Mogami J, Yamaguchi-Shinozaki K (2014) ABA-dependent and ABA-independent signaling in response to osmotic stress in plants. Curr Opin Plant Biol 21:133–139. doi: 10.1016/j.pbi.2014.07.009

116. Nakashima K, Yamaguchi-Shinozaki K, Shinozaki K (2014) The transcriptional regulatory network in the drought response and its crosstalk in abiotic stress responses including drought, cold, and heat. Front Plant Sci 5:170. doi: 10.3389/fpls.2014.00170

117. Finkelstein RR, Gampala SSL, Rock CD (2002) Abscisic acid signaling in seeds and seedlings. Plant Cell 14 Suppl:S15–S45. doi: 10.1105/tpc.010441.would

118. Lois LM, Lima CD, Chua N-H (2003) Small ubiquitin-like modifier modulates abscisic acid signaling in Arabidopsis. Plant Cell 15:1347–59. doi: 10.1105/tpc.009902

119. Miura K, Lee J, Jin JB, et al. (2009) Sumoylation of ABI5 by the Arabidopsis SUMO E3 ligase SIZ1 negatively regulates abscisic acid signaling. Proc Natl Acad Sci U S A 106:5418–23. doi: 10.1073/pnas.0811088106

120. Zheng Y, Schumaker KS, Guo Y (2012) Sumoylation of transcription factor MYB30 by the small ubiquitin-like modifier E3 ligase SIZ1 mediates abscisic acid response in Arabidopsis thaliana. Proc Natl Acad Sci 109:12822–12827. doi: 10.1073/pnas.1202630109

121. Panse VG, Hardeland U, Werner T, et al. (2004) A proteome-wide approach identifies sumoylated substrate proteins in yeast. J Biol Chem 279:41346–51. doi: 10.1074/jbc.M407950200

122. Denison C, Rudner AD, Gerber SA, et al. (2005) A proteomic strategy for gaining insights into protein sumoylation in yeast. Mol Cell Proteomics 4:246–54. doi: 10.1074/mcp.M400154-MCP200

123. Neyret-Kahn H, Benhamed M, Ye T, et al. (2013) Sumoylation at chromatin governs coordinated repression of a transcriptional program essential for cell growth and proliferation. Genome Res 23:1563–1579. doi: 10.1101/gr.154872.113

124. Niskanen E a., Malinen M, Sutinen P, et al. (2015) Global SUMOylation on active chromatin is an acute heat stress response restricting transcription. Genome Biol 16:153. doi: 10.1186/s13059-015-0717-y

125. Ng CH, Akhter A, Yurko N, et al. (2015) Sumoylation controls the timing of Tup1-mediated transcriptional deactivation. Nat Commun 6:6610. doi: 10.1038/ncomms7610

126. Seifert A, Schofield P, Barton GJ, Hay RT (2015) Proteotoxic stress reprograms the chromatin landscape of SUMO modification. Sci Signal 8:rs7-rs7. doi: 10.1126/scisignal.aaa2213

127. Vertegaal ACO, Ogg SC, Jaffray E, et al. (2004) A proteomic study of SUMO-2 target proteins. J Biol Chem 279:33791–8. doi: 10.1074/jbc.M404201200

128. Wohlschlegel JA, Johnson ES, Reed SI, Yates 3rd JR (2004) Global analysis of protein sumoylation in Saccharomyces cerevisiae. J Biol Chem 279:45662–45668. doi: 10.1074/jbc.M409203200\rM409203200 [pii]

129. Wohlschlegel JA, Johnson ES, Reed SI, Yates JR (2006) Improved identification of SUMO attachment sites using C-terminal SUMO mutants and tailored protease digestion strategies. J Proteome Res 5:761–70. doi: 10.1021/pr050451o

130. Schimmel J, Larsen KM, Matic I, et al. (2008) The ubiquitin-proteasome system is a key component of the SUMO-2/3 cycle. Mol Cell Proteomics 7:2107–2122. doi: 10.1074/mcp.M800025-MCP200

131. Blomster HA, Hietakangas V, Wu J, et al. (2009) Novel proteomics strategy brings insight into the prevalence of SUMO-2 target sites. Mol Cell Proteomics 8:1382–90. doi: 10.1074/mcp.M800551-MCP200

132. Hendriks IA, D'Souza RC, Chang J-G, et al. (2015) System-wide identification of wild-type SUMO-2 conjugation sites. Nat Commun 6:7289. doi: 10.1038/ncomms8289

133. Xiao Z, Chang J-G, Hendriks IA, et al. (2015) System-wide Analysis of SUMOylation Dynamics in Response to Replication Stress Reveals Novel Small Ubiquitin-like Modified Target Proteins and Acceptor Lysines Relevant for Genome Stability. Mol Cell Proteomics 14:1419–34. doi: 10.1074/mcp.O114.044792

134. Hendriks IA, Treffers LW, Verlaan-de Vries M, et al. (2015) SUMO-2 Orchestrates Chromatin Modifiers in Response to DNA Damage. Cell Rep 10:1778–1791. doi: 10.1016/j.celrep.2015.02.033

135. Hendriks IA, Vertegaal ACO (2016) A comprehensive compilation of SUMO proteomics. Nat Rev Mol Cell Biol 17:581–95. doi: 10.1038/nrm.2016.81

136. Park HC, Choi W, Park HJ, et al. (2011) Identification and molecular properties of SUMO-binding proteins in arabidopsis. Mol Cells 32:143–151. doi: 10.1007/s10059-011-2297-3

137. López-Torrejón G, Guerra D, Catalá R, et al. (2013) Identification of SUMO targets by a novel proteomic approach in plants(F). J Integr Plant Biol 55:96–107. doi: 10.1111/jipb.12012

138. Rytz TC, Miller MJ, Vierstra RD (2016) Purification of SUMO Conjugates from Arabidopsis for Mass Spectrometry Analysis. Methods Mol Biol 1475:257–81. doi: 10.1007/978-1-4939-6358-4_18

139. Wilson NR, Hochstrasser M (2016) The Regulation of Chromatin by Dynamic SUMO Modifications. pp 23–38

140. Park HJ, Kim WY, Park HC, et al. (2011) SUMO and SUMOylation in plants. Mol Cells 32:305–316. doi: 10.1007/s10059-011-0122-7

141. Srilunchang K, Krohn NG, Dresselhaus T (2010) DiSUMO-like DSUL is required for nuclei positioning, cell specification and viability during female gametophyte maturation in maize. Development 137:333–45. doi: 10.1242/dev.035964

142. Studier FW (2005) Protein production by auto-induction in high-density shaking cultures. Protein Expr Purif 41:207–234. doi: 10.1016/j.pep.2005.01.016

143. Novatchkova M, Bachmair A, Eisenhaber B, Eisenhaber F (2005) Proteins with two SUMO-like domains in chromatin-associated complexes: the RENi (Rad60-Esc2-NIP45) family. BMC Bioinformatics 6:22. doi: 10.1186/1471-2105-6-22

144. Morishita T, Tsutsui Y, Iwasaki H, Shinagawa H (2002) The Schizosaccharomyces pombe rad60 Gene Is Essential for Repairing Double-Strand DNA Breaks Spontaneously Occurring during Replication and Induced by DNA-Damaging Agents The Schizosaccharomyces pombe rad60 Gene Is Essential for Repairing Double-Strand DN. doi: 10.1128/MCB.22.10.3537

145. Boddy MN, Shanahan P, Mcdonald WH, et al. (2003) Replication Checkpoint Kinase Cds1 Regulates Recombinational Repair Protein Rad60. 23:5939–5946. doi: 10.1128/MCB.23.16.5939

146. Heideker J, Prudden J, Perry JJP, et al. (2011) SUMO-targeted ubiquitin ligase, Rad60, and Nse2 SUMO ligase suppress spontaneous Top1-mediated DNA damage and genome instability. PLoS Genet 7:e1001320. doi: 10.1371/journal.pgen.1001320

147. Raffa GD, Wohlschlegel J, Yates JR, Boddy MN (2006) SUMO-binding motifs mediate the Rad60-dependent response to replicative stress and self-association. J Biol Chem 281:27973–81. doi: 10.1074/jbc.M601943200

148. Prudden J, Perry JJP, Arvai AS, et al. (2009) Molecular mimicry of SUMO promotes DNA repair. Nat Struct Mol Biol 16:509–16. doi: 10.1038/nsmb.1582

149. Tamura K, Stecher G, Peterson D, et al. (2013) MEGA6 : Molecular Evolutionary Genetics Analysis Version 6 . 0. 30:2725–2729. doi: 10.1093/molbev/mst197

150. Da Silva-Ferrada E, Xolalpa W, Lang V, et al. (2013) Analysis of SUMOylated proteins using SUMO-traps. Sci Rep 3:1690. doi: 10.1038/srep01690

151. Zhao Q, Xie Y, Zheng Y, et al. (2014) GPS-SUMO : a tool for the prediction of sumoylation sites and SUMO-interaction motifs. 1–6. doi: 10.1093/nar/gku383

152. Miura K, Hasegawa PM (2010) Sumoylation and other ubiquitin-like post-translational modifications in plants. Trends Cell Biol 20:223–32. doi: 10.1016/j.tcb.2010.01.007

153. Park HJ, Yun DJ (2013) New Insights into the Role of the Small Ubiquitin-like Modifier (SUMO) in Plants. Int Rev Cell Mol Biol. doi: 10.1016/B978-0-12-405210-9.00005-9

154. Cohen-Peer R, Schuster S, Meiri D, et al. (2010) Sumoylation of Arabidopsis heat shock factor A2 (HsfA2) modifies its activity during acquired thermotholerance. Plant Mol Biol 74:33–45. doi: 10.1007/s11103-010-9652-1

155. Sadanandom A, Ádám É, Orosa B, et al. (2015) SUMOylation of phytochrome-B negatively regulates light-induced signaling in Arabidopsis thaliana. Proc Natl Acad Sci U S A 112:11108–13. doi: 10.1073/pnas.1415260112

156. Park BS, Song JT, Seo HS (2011) Arabidopsis nitrate reductase activity is stimulated by the E3 SUMO ligase AtSIZ1. Nat Commun 2:400. doi: 10.1038/ncomms1408

157. Kim DY, Han YJ, Kim S Il, et al. (2015) Arabidopsis CMT3 activity is positively regulated by AtSIZ1-mediated sumoylation. Plant Sci 239:209–215. doi: 10.1016/j.plantsci.2015.08.003

158. Hua Z, Vierstra RD (2011) The cullin-RING ubiquitin-protein ligases. Annu Rev Plant Biol 62:299–334. doi: 10.1146/annurev-arplant-042809-112256

159. Jentsch S, Psakhye I (2013) Control of nuclear activities by substrate-selective and protein-group SUMOylation. Annu Rev Genet 47:167–86. doi: 10.1146/annurev-genet-111212-133453

160. Causier B, Ashworth M, Guo W, Davies B (2012) The TOPLESS Interactome: A Framework for Gene Repression in Arabidopsis. Plant Physiol 158:423–438. doi: 10.1104/pp.111.186999

161. Gemperline DC, Scalf M, Smith LM, Vierstra RD (2016) Morpheus Spectral Counter: A computational tool for label-free quantitative mass spectrometry using the Morpheus search engine. Proteomics 16:920–4. doi: 10.1002/pmic.201500420

162. Zhang Y, Wen Z, Washburn MP, Florens L (2010) Refinements to label free proteome quantitation: How to deal with peptides shared by multiple proteins. Anal Chem 82:2272–2281. doi: 10.1021/ac9023999

163. Thomas PD, Campbell MJ, Kejariwal A, et al. (2003) PANTHER: a library of protein families and subfamilies indexed by function. Genome Res 13:2129–41. doi: 10.1101/gr.772403

164. Huang DW, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4:44–57. doi: 10.1038/nprot.2008.211

165. Liu Z, Karmarkar V (2008) Groucho/Tup1 family co-repressors in plant development. Trends Plant Sci 13:137–44. doi: 10.1016/j.tplants.2007.12.005

166. Li Y, Williams B, Dickman M (2016) Arabidopsis B-cell lymphoma2 (Bcl-2)-associated athanogene 7 (BAG7)-mediated heat tolerance requires translocation, sumoylation and binding to WRKY29. New Phytol. doi: 10.1111/nph.14388

167. Williams B, Kabbage M, Britt R, Dickman MB (2010) AtBAG7, an Arabidopsis Bcl-2-associated athanogene, resides in the endoplasmic reticulum and is involved in the unfolded protein response. Proc Natl Acad Sci U S A 107:6088–6093. doi: 10.1073/pnas.0912670107

168. Kammers K, Cole RN, Tiengwe C, Ruczinski I (2015) Detecting significant changes in protein abundance. EuPA Open Proteomics 7:11–19. doi: 10.1016/j.euprot.2015.02.002

169. Gentry M, Hennig L (2014) Remodelling chromatin to shape development of plants. Exp Cell Res 321:40–46. doi: 10.1016/j.yexcr.2013.11.010

170. Kwon CS, Wagner D (2007) Unwinding chromatin for development and growth: a few genes at a time. Trends Genet 23:403–412. doi: 10.1016/j.tig.2007.05.010

171. Psakhye I, Jentsch S (2012) Protein group modification and synergy in the SUMO pathway as exemplified in DNA repair. Cell 151:807–20. doi: 10.1016/j.cell.2012.10.021

172. Castro PH, Verde N, Lourenço T, et al. (2015) SIZ1-Dependent Post-Translational Modification by SUMO Modulates Sugar Signaling and Metabolism in Arabidopsis thaliana. Plant Cell Physiol 56:2297–311. doi: 10.1093/pcp/pcv149

173. Nakashima K, Ito Y, Yamaguchi-Shinozaki K (2009) Transcriptional regulatory networks in response to abiotic stresses in Arabidopsis and grasses. Plant Physiol 149:88–95. doi: 10.1104/pp.108.129791

174. Liu M, Shi S, Zhang S, et al. (2014) SUMO E3 ligase AtMMS21 is required for normal meiosis and gametophyte development in Arabidopsis. BMC Plant Biol 14:153. doi: 1471-2229-14-153 [pii] 10.1186/1471-2229-14-153

175. Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res 29:e45. doi: 10.1093/nar/29.9.e45

176. Yang P, Fu H, Walker J, et al. (2004) Purification of the Arabidopsis 26 S proteasome: Biochemical and molecular analyses revealed the presence of multiple isoforms. J Biol Chem 279:6401–6413. doi: 10.1074/jbc.M311977200

177. Wenger CD, Coon JJ (2013) A Proteomics Search Algorithm Specifically Designed for High-Resolution Tandem Mass Spectra. J Proteome Res 12:1377–1386. doi: 10.1021/pr301024c

178. Tyanova S, Temu T, Sinitcyn P, et al. (2016) The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat Methods 13:731–40. doi: 10.1038/nmeth.3901

179. Ritchie ME, Phipson B, Wu D, et al. (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 43:e47–e47. doi: 10.1093/nar/gkv007

180. Szklarczyk D, Franceschini A, Wyder S, et al. (2015) STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res 43:D447-52. doi: 10.1093/nar/gku1003

181. Shannon P, Markiel A, Ozier O, et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13:2498–504. doi: 10.1101/gr.1239303

182. Komander D, Rape M (2012) The Ubiquitin Code. Annu Rev Biochem 81:203–229. doi: 10.1146/annurev-biochem-060310-170328

183. Yin Y, Seifert A, Chua JS, et al. (2012) SUMO-targeted ubiquitin E3 ligase RNF4 is required for the response of human cells to DNA damage. Genes Dev 26:1196–208. doi: 10.1101/gad.189274.112

184. Sekiyama N, Arita K, Ikeda Y, et al. (2010) Structural basis for regulation of poly-SUMO chain by a SUMO-like domain of Nip45. Proteins 78:1491–502. doi: 10.1002/prot.22667

185. Uzunova K, Gottsche K, Miteva M, et al. (2007) Ubiquitin-dependent Proteolytic Control of SUMO Conjugates. J Biol Chem 282:34167–34175. doi: 10.1074/jbc.M706505200

186. Tatham MH, Matic I, Mann M, Hay RT (2011) Comparative proteomic analysis identifies a role for SUMO in protein quality control. Sci Signal 4:rs4. doi: 10.1126/scisignal.2001484

187. Ulrich HD (2014) Two-way communications between ubiquitin-like modifiers and DNA. Nat Struct {&} Mol Biol 21:317–324. doi: 10.1038/nsmb.2805

188. Perry JJP, Tainer JA, Boddy MN (2008) A SIM-ultaneous role for SUMO and ubiquitin. Trends Biochem Sci 33:201–8. doi: 10.1016/j.tibs.2008.02.001

189. McDowell GS, Philpott A (2013) Non-canonical ubiquitylation: mechanisms and consequences. Int J Biochem Cell Biol 45:1833–42. doi: 10.1016/j.biocel.2013.05.026

190. Curtis MD, Grossniklaus U (2003) A gateway cloning vector set for high-throughput functional analysis of genes in planta. Plant Physiol 133:462–9. doi: 10.1104/pp.103.027979

191. Tatham MH, Geoffroy M-C, Shen L, et al. (2008) RNF4 is a poly-SUMO-specific E3 ubiquitin ligase required for arsenic-induced PML degradation. Nat Cell Biol 10:538–546. doi: 10.1038/ncb1716

192. Mullen JR, Brill SJ (2008) Activation of the Slx5-Slx8 ubiquitin ligase by poly-small ubiquitin-like modifier conjugates. J Biol Chem 283:19912–19921. doi: 10.1074/jbc.M802690200

193. Yu Q, Kuzmiak H, Olsen L, et al. (2010) Saccharomyces cerevisiae Esc2p interacts with Sir2p through a small ubiquitin-like modifier (SUMO)-binding motif and regulates transcriptionally silent chromatin in a locus-dependent manner. J Biol Chem 285:7525–36. doi: 10.1074/jbc.M109.016360

194. Ohya T, Arai H, Kubota Y, et al. (2008) A SUMO-like domain protein, Esc2, is required for genome integrity and sister Chromatid Cohesion in Saccharomyces cerevisiae. Genetics 180:41–50. doi: 10.1534/genetics.107.086249

195. Prudden J, Perry JJP, Nie M, et al. (2011) DNA repair and global sumoylation are regulated by distinct Ubc9 noncovalent complexes. Mol Cell Biol 31:2299–310. doi: 10.1128/MCB.05188-11

196. Marshall RS, McLoughlin F, Vierstra RD (2016) Autophagic Turnover of Inactive 26S Proteasomes in Yeast Is Directed by the Ubiquitin Receptor Cue5 and the Hsp42 Chaperone. Cell Rep 16:1717–1732. doi: 10.1016/j.celrep.2016.07.015

APPENDIX 1

List of SUMO protein sequences used in the phylogenetic analysis of SUMO isoforms. The italicized gray sequences in parenthesis illustrate the residues deleted during re-annotation of the gene. Non-italicized sequences in parenthesis in *Eg*SUMO4 highlight the new annotation of the C-terminal sequences of the gene. Abbreviations: *Ac, Aquilegia coerulea; Al, Arabidopsis lyrata; At, Arabidopsis thaliana; Atr, Amborella trichopoda; Bd, Brachypodium distachyon; Br, Brassica rapa; Cc, Citrus clementina; Ce, Caenorhabditis elegans; Cp, Carica papaya; Cr, Capsella rubella; Cre, Chlamydomonas reinhardtii; Cs, Citrus sinensis; Dm, Drosophila melanogaster; Eg, Eucalyptus grandis; Fv, Fragaria vesca; Gm, Glycine max; Hs, Homo sapiens; Md, Malus domestica; Me, Manihot esculenta; Mm, Mus musculus; Mt, Medicago truncatula; Os, Oryza sativa; Pp, Prunus persica; Ppa, Physcomitrella patens; Ps, Picea sitchensis; Pt, Populus trichocarpa; Pv, Phaseolus vulgaris; Pvi, Panicum virgatum; Sb, Sorghum bicolor; Sc, Saccharomyces cereviseae; Si, Setaria italica; Sl, Solanum lycopersicum; Sm, Selaginella moellendorffii; Tc, Theobroma cacao; Vc, Volvox carteri; Vv, Vitis vinifera; Zm, Zea mays*

Phytozyme annotation	Abrev. name	Sequence	% id to *At*S1	Introns
>Athaliana|AT4G26840|SUMO1	AtSUMO1	MSANQEEDKKPGDGGAHINLKVKGQDGNEVFFRIKRSTQLKKLMNAYCDRQSVDMNSIAFLFDGRRLRAEQTPDELDMEDGDEIDAMLHQTGGSGGGATA*	100	2
>Athaliana|AT5G55160|SUMO2	AtSUMO2	MSATPEEDKKPDQGAHINLKVKGQDGNEVFFRIKRSTQLKKLMNAYCDRQSVDFNSIAFLFDGRRLRAEQTPDELEMEDGDEIDAMLHQTGGGAKNGLKLFCF*	93.5	2
>Athaliana|AT5G55170|SUMO3	AtSUMO3	MSNPQDDKPIDQEQEAHVILKVKSQDGDEVLFKNKKSAPLKKLMYVYCDRRGLKLDAFAFIFNGARIGGLETPDELDMEDGDVIDACRAMSGGLRANQRQWSYMLFDHNGL*	53.9	2
>Athaliana|AT2G32765|SUMO5	AtSUMO5	MVSSTDTISASFVSKKSRSPETSPHMKVTLKVKNQQGAEDLYKIGTHAHLKKLMSAYCTKRNLDYSSVRFVYNGREIKARQTPAQLHMEEEDEICMVMELGGGGPYTP*	43.4	1
>Alyrata|492116|(SUMO1)	AlSUMO1	MSANQEEDKKPGDGGAHINLKVKGQDGNEVFFRIKRSTQLKKLMNAYCDRQSVDMNSIAFLFDGRRLRAEQTPDELDMEDGDEIDAMLHQTGGRGGGAMA*	100	2
>Alyrata|950171|(SUMO2)	AlSUMO2a	MSATQEEDKKPDQGAHINLKVKGQDGNEVFFRIKRSTQLKKLMNAYCDRQSVDFNSIAFLFDGRRLRAEQTPDELEMEDGDEIDAMLHQTGGAANGLKLFCF*	94.6	2
>Alyrata|917212|(SUMO3)	AlSUMO3	MSNSQEDDKNPIDQEQEAHVILKVKSQDGDEVLFKIKKSTPLRKLMYAYCDRRGLKLDAFAFMLDGARIRGTQTPDELDMEDGDEIDACRAMSGGLRADQRQWSYMVFDHNRL*	60	2
>Alyrata|321011|(SUMO5)	AlSUMO5	MVSSSGTISASFVSKRSRSPETPHQKITLKVKNQQGAEDLYKIGAHAHLKKLMSAYCMKRNLDYGSVRFVYNGREIKARQTPAQLKMEEEDEICSVMELGGGGPYTP*	44.7	2
>Crubella|Carubv10006104m.g	CrSUMO1	MSANQEEDKKPGDGGAHINLKVKGQDGNEVFFRIKRSTQLKKLMNAYCDRQSVDMNSIAFLFDGRRLRAEQTPDELDMEDGDEIDAMLHQTGGGGGTGGAMA*	100	2
>Crubella|Carubv10027422m.g	CSUMO2	MSATPEEDKKPDQGAHINLKVKGQDGNEVFFRIKRSTQLKKLMNAYCDRQSVEFNSIAFLFDGRRLRAEQTPDELEMEDGDEIDAMLHQTGGGANGLNRFCFY*	92.5	2
>Crubella|Carubv10027334m.g	CrSUMO3	*(TPYNNPKRPRSDRTFLLFTSPDRESDKRK)*MSNSQEEDKTNPGDQEPQIILKVKSQDGDEVFFSIKKSTQVKRLMYAYCDRRGLKLDAFAFVFDGARIRGQETPFELKMESGDVIDACRTLSGGLRANQRQWSYMVFDHNRL*	58.5	2
>Crubella|Carubv10024362m.g	CrSUMO5	MVSSSGTNTISASFVSKRSRSPEPPHQKLTLKVKNQQGAEDVYKIGAHAHLKKLMIAYCVKRNLEYGAVRFIYNQKHIKPRQTPAQLRMKEEDEILSVMELGGGGPYTPT*	31.8	2
>Brapa|Bra002928	BrSUMO1a	MSATQEEDKKPGEQGGVHINLKVKGQDGNEVFFKIKRSTQLKKLMNAYCDRQSVDLNAIAFLFDGRRLRAEQTPDELDMEDGDEIDAMLHQTGGVANGMYLFCV*	92.6	2
>Brapa|Bra026399	BrSUMO1b	MSATQEEDKKPGGDGGVHINLKVKGQDGNEVFFRIKRSTQLKKLMNAYCDRQSVDMNAIAFLFDGRRLRAEQTPDELDMEDGDEIDAMLHQTGGCGDRTG*	95.7	2
>Brapa|Bra019083	BrSUMO1c	MSATQEEDKKPGDGGAHINLKVKGQDGNEVFFRIKRSTQLKKLMNAYCDRQSVDMTAIAFLFDGRRLRAEQTPDELDMEDGDEIDAMLHQTGGCCGGVALA*	96.8	2
>Brapa|Bra010425	BrSUMO1d	MSATQEEDKKPGDGGAHINLKVKGQDGNEVFFRIKRSTQLKKLMNAYCDRQSVDMNSIAFLFDGRRLRAEQTPDELDMEDGDEIDAMLHQTGGCGSGAAMA*	98.9	2
>Brapa|Bra035560	BrSUMO1e	*(MRPRSLFQSFYYSPSLDILQSQNFFLHRSEFLFSTPTGESGKSK)*MSATQEEDKKPGDQGPAHINLKVKGQDGNEVFFRIKRATQLKKLMTAYCDRQSVDFNSIAFLFDGRRLRAEQTPDELDMEEGDEIDAMLHQTGGVAIC*	92.6	2
>Brapa|Bra005558	BrSUMO2b	MVSSSTTISASCASKGSPSLSPQKKITLKVKAQQDGGEDIYKIGYGAHLKKLMDAYCTKRNLERTTVRFIFRYKELKPRQTPAQLMMEEGDIIDIVTDQGGG*	44.2	2
>Brapa|Bra021812	BrSUMO2c	MVSSSTTISASTASKSRSLTPQRKITLKVKTQQDGREDVYKIGYNAHMKKLMDACCTKRNFEKDTVRFIFGRKELKPRQTPAQLMMEEGDIIDLVTEQGGG*	41.6	2

>Cpapaya|evm.TU.supercontig_132.20	CpSUMO1	MSGVKSQDEDKKPNDQSAHINLKVKGQDGNEVFFRIKRSTQLKKLMGAYCDRQSVDINSIAFLFDGRRLRAEQTPDELEMEDGDEIDAMLHQTGGTL*	91.1	2
>Cpapaya|evm.TU.supercontig_8.137	CpSUMO2	MSATGGGGGGGLEEDKKPVDQSAHINLKVKGQDGNEVFFRIKRSTQLRKLMTAYCDRQSVDFNSIAFLFDGRRLRGEQTPDELEMEDGDEIDAMLHQTGGWSGYY*	90.9	2
>Csinensis|orange1.1g034043m.g	CsSUMO1	MSATGGGGGGGQEEDKKPVDQSAHINLKVKGQDGNEVFFRIKRSTQLKKLMNAYCDRQSVELNSIAFLFDGRRLRGEQTPDELEMEDGDEIDAMLHQTGGALGSG*	92.1	2
>Csinensis|orange1.1g045424m.g	CsSUMO2	MLKPSSSKNNKKPQHLINLIIKSQDGDKRFFQFNHDVEIKRLLIKYCETKSQPFKSTPFLINGNRFDYSKTADQLGLKDGDEIDAMYHAFGGGHDHRA*	43.4	2
>Csinensis|orange1.1g048514m.g	CsSUMO3	MEKSPDNIPDQHFINLVVKGQDNDPLYFEFRRDWEIKKLLITYCEKKDAQYGTFPFLINGNRFPHIRTPDQLGLKDGDEIVATFYAGGA*	39.5	2
>Egrandis|Eucgr.K00756	EgSUMO1	MSASGVTQQHEEDKKPNDQSAHINLKVKGQDGNEVFFRIKRSTQLKKLMNAYCDRQSVEMNSIAFLFDGRRLRGDQTPDELEMEDGDEIDAMLHQTGGARTLT*	87.1	2
>Egrandis|Eucgr.H00049	EgSUMO2	MAAPITSIGTERIDVRVRGQDDRILYFKINRTARLSRLFNIYCERRQLDVQTVQFLYEGNRITGNQTPQALGLEDGAELCAFVHQTGGGRQQHGHPISR*	42.5	2
>Egrandis|Eucgr.L03437	EgSUMO3	MGITGTETKLVILSSVGCENSRPDPILLRVQKQKEDDVCYLLDRRMPLGALMADYCSRRGLPYDAVRFTYEGTRVLEAKSAEDVGMDDEDVIDAWADQLGG*	31.6	0
>Egrandis|Eucgr.H00789	EgSUMO4	MSATGGGGSGQEEDKKPGDQAAHINLKVKGQDGNEVFFRIKRSTQLRKLMTAYCDRQSVELNSIAFLFDGRRLRGEQTPDE(LEMEDGDEIDAMLHQTGGY*)	86.5	2
>Vvinifera|GSVIVG01030502001	VvSUMO1	MSGVANPSSQDEDKKPNDQSGHINLKVKGQDGNEVFFRIKRSTQLKKLMNAYCDRQSVDLNSIAFLFDGRRLRGEQTPDELEMEDGDEIDAMLHQTGGACV*	89	2
>Vvinifera|GSVIVG01003307001	VvSUMO2	MSATGGAAGGQEEDKKPTDQGAHINLKVKGQDGNEVFFRIKRSTQLRKLMSAYCDRQSVELNSIAFLFDGRRLRGEQTPDELEMEDGDEIDAMLHQTGGVAWM*	88	2
>Vvinifera|GSVIVG01003301001	VvSUMO3	MDQGAHINLKVKGQDGNEVFFRIKRSTQLRKLMSAYCDRQSVELNSIAFLFDGRRLRGEQTPDELEMEDGDEIDAMLHQTGGVAWMCLTAN*	91.4	2
>Vvinifera|GSVIVG01021058001	VvSUMO4	MPQPAKRPLDQSTIEVKVKSQDGRQLYFRINRSTPLQRLLVAYCQQINIDYKTMQFVYNGNRVTAKQTPEQLGMEDGDEIDALTHQMGGGCRAF*	50.6	2
>Ppersica|ppa012976m.g	PpSUMO2	*(MTFCTLAQTHDFTNERRGERQENPNFIRFTKTIERPRETETETHSQKTK)*MSTPQQEEDKKPNDQAAHINLKVKGQDGNEVFFRIKRSTQLKKLMNAYCDRQSVELNSIAFLFDGRRLRAEQTPDELEMEDGDEIDAMLHQTGGAVQI*	90.2	2
>Ppersica|ppa013880m.g	PpSUMO1	MSGVTNQEEDKKPTDQSAHINLKVKGQDGNEVFFRIKRSTQLKKLMNAYCDRQSVEFNSIAFLFDGRRLRAEQTPDELEMEDGDEIDAMLHQTGGAFA*	92.3	2
>Ppersica|ppa013826m.g|ppa	PpSUMO3	MSATGGGDGQGEEKKPGDQSAHINLKVKGQDGNEVFFRIKQSTQLKKLMTAYCDRQSVDMNSIAFLFDGRRLRAEQTPEELEMEDGDEIDAMLHQTGGWA*	91	2
>Ppersica|ppa013753m.g|ppa013753m	PpSUMO4	MGIGRRVPVGPQRNILGVPKQRPSYITLFVRDHLSGNDLVFRMKRSTQLRRLKIAYCDRKSVEVYRMRFAYYGVHLISSRTPDEYDLENGDVIDALPVLRGGGAP*	45.2	2
>Gmax|Glyma08g43290	GmSUMO1	MSGVTNNNEEDKKPTEQGAHINLKVKGQDGNEVFFRIKRSTQLKKLMNAYCDRQSVDFNSIAFLFDGRRLRAEQTPDELEMEDGDEIDAMLHQTGGSVV*	90.3	2
>Gmax|Glyma08g46500	GmSUMO2	MSASGGRGSQEEEKKPSDQGAHINLKVKGQDGNEVFFRIKRSTQLKKLMNAYCDRQSVDFNSIAFLFDGRRLRAEQTPDELEMEDGDEIDAMLHQTGGGHKFL*	92.3	2
>Gmax|Glyma18g35450|Glyma18g35450.1	GmSUMO3	MSVSGGRGSQEEEKKPSDQGAHINLKVKGQDGNEVFFRIKRSTQLKKLMNAYCDRQSVDFNSIAFLFDGRRLRAEQTPDELEMEDGDEIDAMLHQTGGGHKFLQMYDDHLHQNA*	92.3	2
>Gmax|Glyma18g35450	GmSUMO4	MATNGPLKRKSPPDDESVNLKIKLQDGRNLFFKVNRDMKLINVFKEFCDRQKLDYETLKFIYDGFNIKGKHTAKMLNMEDDAEIVAIRPQIGGGAAAL*	35.5	2

>Gmax|Glyma08g11770	GmSUMO5	MATSRGRPPKRKSPDDNEATDNIQINFSIIDQDGRHMYFKVNHNLELIKVFKDFCERKNLEYETMQFLCDGIHIKGKHTPKMLNMEDDAEIFAATHQVGGGGDMRC*	34.8	2
>Gmax|Glyma08g11781|Glyma08g11781.1	GmSUMO6	MATNGPLKRKSPPDDDSVNLKIKLQDGRNLFFKVNRDLKLINVFKEFCDRQNLDYETLKFIYDGFNIKGKHTARMLNMEDDAEIVAIRSQIGGGAAAL*	35.5	2
>Ptrichocarpa|Potri.002G224700|Potri.002G224700.1	PtSUMO1	MSEATGQPQEEDKKPNDQSAHINLKVKGQDGNEVFFRIKRSTQLKKLMNAYCDRQSVEINSIAFLFDGRRLRGEQTPDELDMEDGDEIDAMLHQTGGAVKASDYA*	93.3	2
>Ptrichocarpa|Potri.002G224800|Potri.002G224800.1	PtSUMO2	MSGATGQPQEEDKKPNDQSAHINLKVKGQDGNEVFFRIKRSTQLKKLMNAYCDRQSVEFNSIAFLFDGRRLRGEQTPDELDMEDGDEIDAMLHQTGGAVKTSN*	93.3	2
>Ptrichocarpa|Potri.014G158300|Potri.014G158300.1	PtSUMO3	MSGVTGQPQEEDKKPNDQSAHINLKVKGQDGNEVFFRIKRSTQLKKLMNAYCDRQSVEFNSIAFLFDGRRLRGEQTPDELDMEDGDEIDAMLHQTGGAMKTSN*	93.3	2
>Ptrichocarpa|Potri.014G190300|Potri.014G190300.1	PtSUMO4	MSASAGGGQEEDKKPGGDQSAHINLKVKGQDGNEVFFRIKRSTQLRKLMTAYCDRQSVEFNSIAFLFDGRRLRGEQTPDELDMEDGDEIDAMLHQTGGGHASLD*	91.1	2
>Rcommunis|29762.t000027	RcSUMO1	MSGVTNQEEDKKPTDQSAAHINLKVKGQDGNEVFFRIKRSTQLKKLMNAYCDRQSVEFNSIAFLFDGRRLRGEQTPDELEMEDGDEIDAMLHQTGGAAA	90.2	2
>Rcommunis|30204.t000013	RcSUMO2	MSATPGSGGAGAGGQEEDKKPMDQTAHINLKVKGQDGNEMFFRIKRSTQLRKLITAYCDRQSVEFNSIAFLFDGRRLRGEQTPDELEMEDGDEIDAMLHQTGGGDAHL	87.6	2
>Rcommunis|29904.t000041	RcSUMO3	MESFKTITVRVRSQDGREKVFRIKMDTQMSKLIARYCEDRQWEPHTAEFLLNGLRFPRDKTPAQLNLKDNVLIEAMMHQNGGGSKAFSMHALYL	40.8	2
>Mesculenta|cassava4.1_020028m.g	MeSUMO1	MSGVTNQEEDKKPNDQSAHINLKVKGQDGNEVFFRIKRSTQLKKLMNAYCDRQSVEINSIAFLFDGRRLRGEQTPDELEMEDGDEIDAMLHQTGGAIA*	91.2	2
>Mesculenta|cassava4.1_019995m.g	MeSUMO2	MSGVTTNQEEDKKPADQSAHINLKVKGQDGNEVFFRIKRSTQLKKLMNAYCDRQSVEMNSIAFLFDGRRLRGEQTPDELEMEDGDEIDAMLHQTGGAIA*	90.3	2
>Mesculenta|cassava4.1_020007m.g	MeSUMO3	MSGVTNQEEDKKPNDQSAHINLKVKGQDGNEVFFRIKRSTQLKKLMNAYCDRQSVEMNSIAFLFDGRRLRGEQTPDELEMEDGDEIDAMLHQTGGAIA*	92.3	2
>Mesculenta|cassava4.1_019468m.g	MeSUMO4	MSAAAGGGGGGGGGGGGGAGVGAPEEDKKPMDQSAHINLKVKGQDGNEVFFRIKRSTQLRKLMTAYCDRQSVEFNSIAFLFDGRRLRGEQTPDELEMEDGDEIDAMLHQTGGGNVYP*	89.8	2
>Mesculenta|cassava4.1_031997m.g	MeSUMO5	MDRPAGGINVTVRSQDGQEKCYRIKLETPIAKLLRFYCDTKQLEYDTMVFLIKGRRFNQKKTPAELNLKDGVQIEAFMHQNGGGCKGV*	41.7	2
>Mguttatus|mgv1a017022m.g	MgSUMO1	MSSVEDDKKPADTGAHINLKVKGQDGNEVFFRIKRSTQLKKLMNAYCDRQSVDFNSIAFLFDGRRLRAEQTPDELEMEDGDEIDAMLHQTGGGTTA*	94.3	2
>Mguttatus|mgv1a017044m.g	MgSUMO2	MSGVEEDKKPADGAAHINLKVKGQDGNEVFFRIKRSTQLKKLMNAYCDRQSVDFNSIAFLFDGRRLRGEQTPDELEMEDGDEIDAMLHQTGGYASA*	94.3	2
>Mguttatus|mgv1a016909m.g	MgSUMO4	MSTPGEGEEVEDKKPIVQSRYITIKVNSNYKQDENAVFFKIKRNVKLKVLIRAYRERQSVDDSIVYLYNGTKIGDEDTPDSLEMEDVDEIDAMAAMDGGASD*	52.2	2
>Mguttatus|mgv1a020750m.g	MgSUMO5	MAAAGAQRVKEEKSEIIILNIQYNETGGQKYSFSTFTDVPLKEIFRKFCNNQDLIYGSIRFMIDGDRIRETQTPRDLKLEDGDLIDAFNDQIGGGCW*	36.8	1
>Mguttatus|mgv1a017058m.g	MgSUMO6	MVKGLSGRKKPAEEPVLPKVTICIAQDGDEVYFRYVRNKKIQNLLTLYCKEKNIDYRSVEFLFNGKRIATGRNANQLGMVDGDQIDVMTHNIGGG*	42	2
>Mguttatus|mgv1a022663m.g	MgSUMO7	MSTSGEEENQKPIAQPGYVNIKVNSQDGKQVFFRINRNTPLKKLMCAYRAKESLDNSIVFLFNGGRIRETHTPDKLEMKDGDEIDAMSNQIGGATSTDHA*	55.4	2

>Sbicolor|Sb03g043870	SbSUMO1	MSGAGEEDKKPAEAGGAHINLKVKGQDGNEVFFRIKRSTQLKKLMNAYCDRQSVDMNAIAFLFDGRRLRGEQTPDELEMEDGDEIDAMLHQTGGSVPGA*	90.4	2
>Sbicolor|Sb02g037195	SbSUMO2	MSPPRQENRRQVIVKAEPVPSITLKVLDQQSRRAFHTMRMNDRLQGVMDAYYKKVSDDVTYGTGIFMFDGSVRLRGCNTPAELDLNDGDQIEFFESMIGGGCMG*	33	0
>Sbicolor|Sb02g037200	SbSUMO3	MSPPVEEGRRQGSVKTEPEDDPLITLKVLDQEGRRAFHTMRMSDKVQGVMDAYYKKAAGEVTYGSGTFMFDGSVRLRGCNTPAELDLNDGDEIEFFPVMIGGGWVAIGA*	40.5	0
>Sbicolor|Sb02g037220	SbSUMO4	MMRSGARGGEEEEDRKPVIKPGVHVTIKVQDTEGRTVERTVRRSTQKLQVVMDAYYASVPDVTYGTGRFLYDGGRLSAGQTPAELEMEEGDEIDFFTEMLGGGGAAAALLLNAR*	46.4	0
>Zmays|GRMZM2G053898	ZmSUMO1a	MSGAGEEDKKPAEGGAHINLKVKGQDGNEVFFRIKRSTQLKKLMNAYCDRQSVDMNAIAFLFDGRRLRGEQTPDELEMEDGDEIDAMLHQTGGSVPSTT*	91.4	2
>Zmays|GRMZM2G082390	ZmSUMO1b	MSGAGEEDKKPAEGGAHINLKVKGQDGNEVFFRIKRSTQLKKLMNAYCDRQSVDMNAIAFLFDGRRLRGEQTPDELEMEDGDEIDAMLHQTGGSVPSTT*	91.4	2
>Zmays|GRMZM2G305196	ZmSUMO2	MMRSDVRGGDAEEEVDRKPVIKPGVHVTLKVQDTAGPTQELQALMDAYYASVPDVAYGTGRFLYDGGRLTGAHTPAELGMEEQDEIDFFTELLGGGGRRAAAEPGRPVVA*	37.6	0
>Osativa|LOC_Os01g68940	OsSUMO1	MSSPAGEDEKKPAGGEGGGAHINLKVKGQDGNEVFFRIKRSTQLKKLMNAYCDRQSVDIKSIAFLFDGRRLNAEQTPDQLEMEDGDEIDAMLHQTGGSLPA*	84.2	2
>Osativa|LOC_Os01g68950	OsSUMO2	MSAAGEEDKKPAGGEGGGAHINLKVKGQDGNEVFFRIKRSTQLKKLMNAYCDRQSVDMNAIAFLFDGRRLRGEQTPDELEMEDGDEIDAMLHQTGGCLPA*	89.6	2
>Osativa|LOC_Os07g38660	OsSUMO3	MYGWSGIPAAVKVEKENEWKTPATWEWKAPATRVAGEYVTLKVQGTDGRAVYRTMLRTEELQGLMDFYYDRSHGRVQRGTGRFLFDGRRLRGWQTPAELQMEDGDEVNFFEELIGGAAGSGWDPPSSILA*	45.1	0
>Osativa|LOC_Os07g38690	OsSUMO4	MFGRSGITAAVKVEEEDDGKTPAAKRAGEYVTLKVQDTDGRAVYRTMRWTEQLQGLMDFYYDRAHGRVQRGTGRFLYDGRRLSGWQTPAELDMEDGDEVDFFEELIGGAA*	44.2	0
>Osativa|LOC_Os07g38650	OsSUMO5	MSTTSPRAEEDAKETVKPIFITLKVMDQEDRRIRHTIRMADKLQVVMDMYYAKAPDVTYGTGTFLFDGIRLKGDMTPMGLEMVDGDTVDFFPVMIGGGGFFQCNLLPSSH*	34.4	0
>Bdistachyon|Bradi2g58830	BdSUMO1	MSAAGGEEDKKPAGGEGGGAHINLKVKGQDGNEVFFRIKRSTQLKKLMNAYCDRQSVDMTAIAFLFDGRRLRAEQTPDELEMEDGDEIDAMLHQTGGFLLPPNA*	88.4	2
>Bdistachyon|Bradi2g55140	BdSUMO2	MPSPPPPSGHDKTDAEPEVFKPVKPEPTADGDFINVTVTSQISVDVLFRIKRNARLQRLMDMYCGKHSLDPRAVRFLNDEGKYLKAAQTADEAGLKDGGLIDVHMAQDGGFAPSITSVHI*	41.9	0
>Smoellindorffii|171605	SmSUMO1	MSQAEDAATPQAEKQEQKPAEGVHINVKVKSQDGNEVFFRIKKNTQFRKLMTAYCQRQSVEADAIAFLFDGRRLRADQTPEELEMEDGDEIDAMLHQTGGAS*	73	2
>Smoellindorffii|79293	SmSUMO2	MEGSSETPDVKPEKKPGDHMNLKVKSQDGNEICFSIRRNTRLAKLMKAYCERMSVAPDSIAFLLDGKRLREDQTPEELEMEDGDEIDAMLHQTGGMPLL*	70.9	2
>Ppatens|Pp1s61_57V6	PpaSUMO1	MAGVEDSSNPGVQHQDEKKPLDGAGQHINLKVKGQDGGEVFFRIKSTATLRKLMNAYCDRQSVDPSSIAFLFDGRRLRAEQTPAELDMEDGDEIDAMLHQTGGASS*	82.2	2
>Ppatens|Pp1s317_15V6	PpaSUMO2	MSGVEDGSKMANNQTNTQDQEEKKPLDGAGQHINLKVKGQDGGEVFFRIKSTATLRKLMNAYCDRQSVDPSSIAFLFDGRRLRADQTPAELEMEDGDEIDAMLHQTGGNAC*	82	2
>Creinhardtii|g16733|g16733.t1	CreSUMO2	MEADGEPQPKVKSEGAVINLVVKDQQGTEVHFKVKTKTRLEKVFNAYCNKKGMDTASVRFLFDGERVNANSTPEQLEMADGDVIDCVIEQVGGGVSA*	48.4	4
>Creinhardtii|g16734|g16734.t2	CreSUMO3	MSEGGADNQAEIKTEGGIINLVVKDQEGSEVHFKVKMKTKLEKVIDAYCKKKALDASTIRFLYDGNRVNPTNTPAELGMEDGDTIDCLITQLGGGSSYSQRR*	47.8	4
>Vcarteri|Vocar20000106m.g	VcSUMO1	MAEQIGENEHQKKPPFKEEGNPANVINLVVKDQTGNEVHFKVKMKTKLDKVFTAYCNKKGQDPSTVRFLYDGTRVHGHSTPDELGMEDGDVLDCVI	46.2	4

		EQLGGCCRA*		
gene07612-v1.0-hybrid	FvSUMO5	*(MERYGKTNISSVERFRKISISS)*MESERRLAATLESLVKLTLKEDEEPITLQVQNQIHGDIFYRTGRTVSLGNVLKDYCERKGLVYEEMRFIYDGRRVRSTHTPHQLEMEDDFVIDAMSEQIGG*	36.9	0
gene31679-v1.0-hybrid	FvSUMO4	MSGLMNTDKDNGKKPAAAASERKSTDVNLKVKSQKFRTMYFRMKRHTPLQKLVDVYTRKYDVYSFKFLYDGQGINPKLTALQSGMKDGDEIDAMLHADGGGRRC*	39.4	2
gene03443-v1.0-hybrid	FvSUMO3	MSGVTNQEEDKKPADQAAHINLKVKSQDGNEVFFRIKRNTQLKKLMNAYCDRQSVDFNAIAFLFDGRRLRAEQTPDELEMEDGDEIDAMLHQTGGAIV*	89.9	2
gene01213-v1.0-hybrid	FvSUMO2	*(MEPQGPWAEYPKRKIKTRRRGVEREAQNPNFHNLERKRGGEEEEEEAK)*MSGVASQPQEEDKKPNDQGAHINLKVKGQDGNEVFFRIKRSTQLKKLMNAYCDRQSVELNSIAFLFDGRRLRAEQTPDELEMEDGDEIDAMLHQTGGAAQL*	92.1	2
gene17600-v1.0-hybrid	FvSUMO1	MSGTPGGAPEEDKKPSDQSAHINLKVKGQDGNEVFFRIKHSTQLKKLMNAYCDRQSVDMNSIAFLFDGRRLRPEQTPDELEMEDGDEIDAMLHQTGGGRA*	92.9	2
>Tcacao|Thecc1EG002167|Thecc1EG002167t1	TcSUMO1	*(KKKKTLILQKKKLNLRRIGSRKKSLSRER)*MSGQQEEDKKPGDQSAAHINLKVKGQDGNEVFFRIKRSTQLKKLMNAYCDRQSVELNSIAFLFDGRRLRGEQTPDELEMEDGDEIDAMLHQTGGVNASTVFSLV*	90.2	2
>Tcacao|Thecc1EG002904|Thecc1EG002904t1	TcSUMO2	MSATGGGGAGGGGQEEDKKPADQSAHINLKVKGQDGNEVFFRIKRSTQLRKLMTAYCDRQSVDLSSIAFLFDGRRLRGEQTPDELEMEDGDEIDAMLHQTGGGMEVHGC*	89.5	2
>Tcacao|Thecc1EG020389|Thecc1EG020389t1	TcSUMO3	MSSSLAKYLPNARVRITIKNQDGQEAYYQMKRTTPLRKLMNAHCSKYSFEPNTVAFLFDGRRLNEDETPEQVKMEDEEEIDCMIHQVGGYGVHSA*	50	2
>Tcacao|Thecc1EG020307|Thecc1EG020307t1	TcSUMO4	MSRPSGQASNSADGQPESIKITVKGQDGSTVVYKIGRKIKLSKLLHSYCQRKQLDYRTVRFVHEGRHVPGQHTADKLKLEDGAEIFCMFLQTGGGFHIMPKTT*	41.9	2
>Tcacao|Thecc1EG020390|Thecc1EG020390t1	TcSUMO5	MSSPRSEYLPNDRVRITVKNQDGEKACYSMKRTSPLCKLMKAHCSIFSLELNTASFLFGSRCLHEDETPEQVGMEDVEKIECMIYQIGG*	38.4	2
Solyc07g064880.2	SlSUMO1	MSGVTQQEEEKKPAGDQGGHINLKVKSQDGNEVFFRIKRSTQLKKLMNAYCDRQSVDFNSIAFLFDGRRLRAEQTPDELEMEDGDEIDAMLHQTGGSLS*	89.9	2
Solyc12g006010.1	SlSUMO2	MSGVAGGEEDKKPAGDQSGHINLKVKSQDGNEVFFRIKRSTQLKKLMNAYCDRQSVDFNSIAFLFDGRRLRAEQTPDELEMEDGDEIDAMLHQTGGTTI*	89.9	2
Solyc07g049360.2	SlSUMO3	MSQAAEEDKKPGGDQVHINLKVKSQDGNEVFFRIKRSTQLKKLMNAYCDRQSVDFNSIAFLFDGRRLRGEQTPDELEMEDGDEIDAMLHQTGGSTI*	86.7	2
Solyc09g059970.2	SlSUMO4	MSASGGTGDEDKKPNDQMVHINLKVKGQDGNEVFFRIKRSTQMRKLMNAYCDRQSVDMNSIAFLFDGRRLRAEQTPDELEMEEGDEIDAMLHQTGGSCCTCFSNF*	89.5	2
Solyc09g091890.2	SlSUMO5	MAEGSRKSIKLKIKAQDDTILHFKVNTSTIMKDIFMSYSSKKQMMNYKVFRFFLDGKRLSSHKTVNELGLKNGDEIDAMIHQDGGGSACNY*	39.2	2
AMTR_s00228p00023500 [Amborella trichopoda]	AtrSUMO1	MSGATNEEEKKPVDQSAHINLKVKGQDGNEVFFRIKRSTQLRKLMTAYCDRQSVDFNSIAFLFDGRRLRGQTPDELEMEGDEIDAMLHTGGSDGI*	85.1	3
ABK22096.1| unknown [Picea sitchensis]	PsSUMO1	MSGVDNGGTPGATNQEEEKKPMDQGAHINLKVKGQDGNEVFFRIKRSTQLRKLMNAYCDRQSVDFNSIAFLFDGRRLRGEQTPEELEMEDGDEIDAMLHQTGGRR	88.9	N/A
ABK23000.1| unknown [Picea sitchensis]	PsSUMO2	MDDRGNAAPAGQEEERKPLDQGAHINLKVKGQDGNEVFFRIKRSTQLRKLMNAYCDRQSIDFNSIAFLFDGRRLRGEQTPDELEMEEGDEIDAMLHQTGGML	87.6	N/A