Distinct and Overlapping Sets of SUMO-1 and SUMO-2 Target Proteins Revealed by Quantitative Proteomics

Alfred C.O. Vertegaal1*, Jens S. Andersen2, Stephen C. Ogg3, Ronald T. Hay4, Matthias Mann5 and Angus I. Lamond4

1 Department of Molecular Cell Biology, Leiden University Medical Center, 2300 RC Leiden, the Netherlands
2 Center for Experimental BioInformatics (CEBI), Department of Biochemistry and Molecular Biology, University of Southern Denmark, Campusvej 55, DK-5230, Odense M, Denmark
3 Centre for Molecular Medicine, 61 Biopolis Drive (Proteos), Singapore 138673
4 Wellcome Trust Biocentre, University of Dundee, Dundee DD1 5EH, United Kingdom
5 Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Am Klopferspitz 18, D-82152 Martinsried, Germany

Running title: SUMO-1 and SUMO-2 target protein preferences

* Corresponding author: Alfred C.O. Vertegaal, PhD.
Department of Molecular Cell Biology
Leiden University Medical Center
Postal zone S1-P, P.O. box 9600
2300 RC Leiden, the Netherlands

phone: +31-71-5269212
fax: +31-7-5268290
email: vertegaal@lumc.nl
ABBREVIATIONS

- Arg0, $^{12}$C$_6$$^{14}$N$_4$-Arginine
- Arg6, $^{13}$C$_6$$^{14}$N$_4$-Arginine
- Arg10, $^{13}$C$_6$$^{15}$N$_4$-Arginine
- Bis-TRIS, 2-[bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol
- CHD, Chromodomain Helicase DNA-binding protein
- E1, SUMO activating enzyme
- E2, SUMO protein carrier protein
- E3, SUMO ligase
- LTQ-FT-ICR, linear quadrupole ion trap Fourier transform ion cyclotron resonance
- PML, promyelocytic leukemia protein
- RanGAP1, Ran GTPase-activating protein 1
- SAE, SUMO activating enzyme
- SART, squamous cell carcinoma antigen recognized by T-cells
- SILAC, stable-isotope labeling by amino acids in cell culture
- SUMO, small ubiquitin-like modifier
- Ubc9, ubiquitin-conjugating enzyme 9
- $\psi$KxE/D, consensus sumoylation site where $\psi$ is V, L, I, M or F and x is any amino acid.
- ZF, zinc finger
SUMMARY

The SUMO family in vertebrates includes three different family members that are conjugated as post-translational modifications to target proteins. SUMO-2 and -3 are nearly identical, but differ substantially from SUMO-1. We have used quantitative proteomics to investigate the target protein preferences of SUMO-1 and SUMO-2. HeLa cells were established that stably express His<sub>6</sub>-SUMO-1 or His<sub>6</sub>-SUMO-2. These cell lines and control HeLa cells were labeled with stable arginine isotopes and His<sub>6</sub>-SUMOs were enriched from lysates using immobilized metal affinity chromatography. 53 SUMO conjugated proteins were identified, including 44 novel SUMO targets. 25 proteins were preferentially conjugated to SUMO-1, 19 were preferentially conjugated to SUMO-2 and 9 proteins were conjugated to both SUMO-1 and SUMO-2. SART1 was confirmed by immunoblotting to have both SUMO-1 and SUMO-2 linked forms at similar levels. SUMO-1 and SUMO-2 are thus shown to have distinct and overlapping sets of target proteins, indicating that SUMO-1 and SUMO-2 may have both redundant and non-redundant cellular functions. Interestingly, 14 of the 25 SUMO-1 conjugated proteins contain zinc fingers. While both SUMO family members play roles in many cellular processes, our data show that sumoylation is strongly associated with transcription since nearly one third of the identified target proteins are putative transcriptional regulators.
INTRODUCTION

Conjugation of ubiquitin-like protein modifiers to target proteins regulates a wide variety of cellular processes (1). The ubiquitin family includes Small Ubiquitin-like MOdifiers (SUMOs) that are similar in structure to ubiquitin (2), but in contrast to polyubiquitination, sumoylation does not target proteins for degradation by the proteasome (3-5). Sumoylation can regulate the function of proteins by affecting protein-protein interactions, which in turn can influence subcellular localization. This was first demonstrated for RanGAP1, a protein that is targeted to the nuclear pore component RanBP2 by sumoylation (6-9).

The conjugation pathway of SUMO is similar to the conjugation pathway of ubiquitin and consists of E1, E2 and E3 enzymes (3-5). SUMO is activated by the SUMO activating enzyme 1/2 dimeric E1 enzyme and subsequently SUMO is transferred to target proteins by a single E2 enzyme designated Ubc9. Several E3-like factors have been identified, including RanBP2 and the PIAS family, that enhance SUMO conjugation to proteins (3-5,10,11). Sumoylation is a reversible process; SUMO-specific proteases can remove SUMO from target proteins (12). These SUMO proteases are also essential for SUMO maturation because SUMO precursor proteins require C-terminal cleavage to expose a di-glycine motif essential for conjugation. RNAi and genetic studies of several components of the sumoylation pathway have established that sumoylation is critical for eukaryotic cell viability (13-17).

Many target proteins have been identified for Smt3, the single SUMO in budding yeast (18-23). These include transcription factors, replication factors, RNA binding and
processing proteins, translation factors, transport factors, cytoskeleton components and metabolic enzymes, highlighting the broad impact of SUMO on cellular processes.

In contrast to the single SUMO found in *S. cerevisae, C. elegans* and *D. melanogaster*, higher eukaryotes express multiple different SUMOs. A complex SUMO family has been identified in *A. thaliana* with up to eight members (24,25). Humans express three SUMO family members, SUMO-1, SUMO-2 and SUMO-3. Mature SUMO-2 and SUMO-3 are nearly identical (~95% identity), but differ substantially from SUMO-1 (~50% identity) (26-28). In addition to genes that encode functional SUMOs, extensive sets of SUMO pseudogenes exist (29).

We have previously purified and identified a set of target proteins for human SUMO-2 (30) and other groups have identified target proteins for SUMO-1 and SUMO-3 (31-35). These studies have emphasized the broad impact of SUMO on multiple cellular processes. It is not clear however whether different SUMO family members have unique cellular roles, or whether they act in a redundant manner. Recent developments in quantitative proteomics now enable the systematic investigation of target protein preferences for different SUMO family members. The power of these new techniques has been demonstrated in recent studies. For example, in our laboratories Andersen *et al.* (36) have investigated the flux of 489 endogenous nucleolar proteins in response to metabolic inhibitors, Kratchmarova *et al.* (37) have compared the closely related signaling pathways of epidermal growth factor and platelet-derived growth factor and Trinkle-Mulcahy *et al.* (38) have identified PP1α and -γ binding proteins using the quantitative proteomics technique SILAC (39,40). Here, we have used SILAC to identify and compare target protein sets for SUMO-1 and SUMO-2.
EXPERIMENTAL PROCEDURES

Cell culture. HeLa cells stably expressing His6-SUMO-1 or His6-SUMO-2 were previously described (30,41). HeLa cells were grown in Dulbecco’s modified Eagle’s medium (DMEM), supplemented with 10% FCS and 100 U/ml penicillin and streptomycin (Invitrogen). Stable isotope labeling was carried out essentially as previously described (36,37,42), using $^{12}\text{C}_6^{14}\text{N}_4$-Arginine (referred to as Arg0), $^{13}\text{C}_6^{14}\text{N}_4$-Arginine (referred to as Arg6) or $^{13}\text{C}_6^{15}\text{N}_4$-Arginine (referred to as Arg10) as indicated.

Purification of His$_6$-SUMO conjugated proteins. His$_6$-SUMO conjugates were purified essentially as previously described (30). Briefly, cells were isolated by trypsinization and washed twice with ice-cold PBS. Nuclei were isolated and washed in ice-cold CSK buffer (10 mM PIPES pH 6.8, 300 mM Sucrose, 100 mM NaCl, 3 mM MgCl$_2$, 1 mM EGTA and 0.5% Triton X-100) supplemented with protease inhibitor cocktail 1873580 (Roche Diagnostics GmbH). Subsequently, proteins were solubilized in lysis buffer (8 M Urea, 100 mM Na$_2$HPO$_4$/NaH$_2$PO$_4$, 10 mM Tris/HCl, pH 8.0) and sonicated. His$_6$-SUMO conjugates were enriched on Talon beads (BD Biosciences) and washed extensively with lysis buffer. Conjugates were eluted in lysis buffer containing 200 mM imidazole.

Mass spectrometry and data analysis. Mass spectrometric analysis was performed by nanoscale liquid chromatography-tandem mass spectrometry (LC MS/MS) using a linear ion trap Fourier-transform ion-cyclotron resonance mass spectrometer (LTQ-FT-ICR, Thermo-Finnigan, Bremen). Eluates were analyzed by 1-dimensional gel electrophoresis.
The two gel lanes used were cut in 10 slices and subjected to in-gel digestion with trypsin. The resulting peptides were extracted, concentrated, and then loaded onto a fused silica capillary with a 75 μm ID and an 8 μm tip opening (New Objective, Woburn, MA) filled with Reprosil 3 μm reverse phase material (Dr. Maisch, Ammerbuch, Germany). Peptides were eluted with a 140 min linear gradient of 95% buffer A (0.5% acetic acid in H₂O) to 50% buffer B (80% acetonitrile, 0.5% acetic acid in H₂O). The LTQ-FT-ICR instrument was operated in the data dependent mode to acquire high-resolution precursor ion spectra (from m/z 300-1500, R=25,000, and ion accumulation to a target value of 10,000,000) in the ICR cell. The three most intense ions were sequentially isolated for accurate mass measurements by SIM scans (10 Da mass window, R=50,000, and a target accumulation value of 50,000). The ions were simultaneously fragmented in the linear ion trap with a normalized collision energy setting of 27 % and a target value of 2,000.

Combined peak list were searched in the International Protein Index (IPI) database (http://www.ebi.ac.uk/IPI/IPIhelp.html) using the Mascot program (Matrix Science, London). LTQ-FT-ICR data were searched with a peptide mass tolerance of 5 ppm and a fragment mass tolerance of 0.8 Da. Iterative calibration algorithms on the basis of identified peptides resulted in an average absolute peptide mass accuracy of better than 1 ppm. Stringent criteria were required for protein identification based on the LTQ-FT-ICR data: at least two matching peptides per protein, a mass accuracy within 3 ppm, a Mascot score for individual ions of better than 20, and a delta score of better than 5.

Protein ratios were calculated for each arginine containing peptide as the peak area of Arg6 divided by the peak area of Arg0 and the peak area of Arg10 divided by the
peak area of Arg0 for each single scan mass spectrum. The peptide ratios were averaged for all arginine-containing peptides sequenced for each protein. MSQuant, an in-house developed software program was used to extract information from the Mascot HTML database search files and to manually validate the certainty in peptide identification and in peptide abundance ratio. The program is available as open source at http://msquant.sourceforge.net/.

Proteins, Antibodies, Immunoprecipitations, Protein Electrophoresis and Immunoblotting. SUMO-1 and SUMO-2 proteins were produced in E. coli and purified as described previously (43). The amino acid sequence of the mature protein that we refer to as SUMO-2 is MSEEKPKEGVKTENDHINLKVAGQDGGSVQF-IKRHTPLSKLMKAYCERQGLSMRQIRFRFDGQPINEETDPQAQLEMEDEDTIDVFQQTGG (43). Peptide antibody AV-SM23-0100 against SUMO-2/3 was generated in rabbit using the peptide MEDEDTIDVFQQQQTG (Eurogentec) (30). Peptide antibody 1607 against SART1 was also generated in rabbit by Eurogentec using peptides CSLSIEETNLRAKLGLKPLEV and CNLDEEEQQQDFSASSTT as described previously (44). Monoclonal antibodies 21C7 against SUMO-1 and 19C7 against RanGAP1 were obtained from Zymed, monoclonal antibody His.Tag against His5 was obtained from Novagen, polyclonal antibody AB1380 against Sp100 was obtained from Chemicon, monoclonal antibody 1814460 against GFP was obtained from Roche Diagnostics Corp. and polyclonal antibody SC-551 against RARα was obtained from Santa-Cruz Biotechnology. Secondary antibodies used were anti-rabbit HRP and anti-mouse HRP (1:5000, Pierce Chemical Co.).
HeLa cells were lysed in RIPA buffer (50 mM TRIS-HCl pH7.5, 150 mM NaCl, 1% NP-40, 1% Na-deoxycholate and 5 mM EDTA) supplemented with 1 mM DTT, 1 mM PMSF and 20 mM N-ethylmaleimide and protease inhibitor cocktail 1873580 (Roche Diagnostics GmbH). Lysates were pre-cleared by centrifugation and RanGAP1 and Sp100 were immunoprecipitated using specific antibodies. Species matched control antibodies were directed against GFP and RARα. Lysates were incubated with antibodies at 4 °C for 1 hour, cleared again by centrifugation and immunocomplexes were subsequently purified on protein G Sepharose 4 fast flow beads (Amersham Biosciences) for 3 hours at 4 °C. After extensive washing with lysis buffer, immunoprecipitates were eluted in LDS protein sample buffer (Invitrogen).

Protein samples were size fractionated on Novex 4-12% Bis-TRIS gradient gels using 4-morpholinepropanesulfonic acid buffer (Invitrogen). Total protein was visualized employing the colloidal blue staining kit according to the instructions of the manufacturer (Invitrogen). For immunoblotting experiments, size fractionated proteins were subsequently transferred onto Hybond-C extra membranes (Amersham Biosciences) using a submarine system (Invitrogen). The membranes were incubated with specific antibodies as indicated. The monoclonal antibody His.Tag against His₅ was used according to the instructions of the manufacturer (Novagen). Bound antibodies were detected via chemiluminescence with ECL Plus (Amersham Biosciences).
RESULTS

SUMO-1 and SUMO2/3 conjugation profiles.

To study the target protein profiles for the nearly identical human SUMO family members SUMO-2 and SUMO-3, a polyclonal antiserum was raised in a rabbit against a peptide from the identical C-terminal regions of both SUMO-2 and SUMO-3. This antiserum specifically recognizes SUMO-2/3, but not SUMO-1 as judged by immunoblotting experiments, whereas the commercially available monoclonal antibody 21C7 specifically recognizes SUMO-1 but not SUMO-2/3 (Fig. 1A and B). Endogenous SUMO target protein profiles in HeLa lysates were studied by immunoblotting, using both this SUMO-2/3 specific antiserum and monoclonal antibody 21C7 (Fig. 1). Using short exposure times, it became apparent that the major 75kDa band recognized by the SUMO-1 antibody (indicated by an asterisk) is not prominent in the SUMO-2/3 target protein profile. Previously, this major SUMO-1 target was shown to be RanGAP1 (8), and these results thus indicate that RanGAP1 is a preferential target for SUMO-1. Other parts of the target protein profiles also differ for SUMO-2/3 and SUMO-1, indicating that substantial numbers of proteins are preferentially conjugated to either SUMO-1 or to SUMO-2/3.

Purification and identification of SUMO-1- and SUMO-2 conjugates by quantitative proteomics.

To investigate the SUMO-1- and SUMO-2 conjugated proteins in more detail and to identify novel SUMO targets, a quantitative proteomics approach was chosen utilizing
the SILAC method (39,40) (Fig. 2). An important technical aspect of the purification of sumoylated proteins is the use of denaturing buffers to inactivate SUMO proteases and to prevent non-sumoylated proteins co-purifying because they are bound to sumoylated proteins in a non-covalent manner. The use of non-denaturing buffers can instead lead to the inclusion of false-positives. We have used His6 to tag SUMO in these experiments because the small hexahistidine tag is less likely to interfere with enzymatic attachment to target proteins and should therefore maximize the range of SUMO targets selected and this tag is compatible with the use of denaturing buffers. Stable HeLa cell lines expressing either His6-SUMO-1 or His6-SUMO-2 were established and labeled with either carbon substituted arginine (Arg6) or carbon plus nitrogen substituted arginine (Arg10), respectively. Control HeLa cells were labeled using unsubstituted arginine (Arg0). Nuclear lysates were prepared in 8M Urea from the three different cell lines because nearly all sumoylated proteins are present in the nucleus (30). 10% of these input fractions were separately affinity purified on Talon beads, size separated by SDS-PAGE and immunoblotted to detect the His-tag (Fig. 3A). Similar expression levels of His6-SUMO-1 and His6-SUMO-2 were found.

The three nuclear lysates were mixed in a 1:1:1 ratio for large-scale His6-SUMO affinity purification on Talon beads. Small aliquots of the eluted fractions were used to detect SUMO-1 and SUMO-2/3 by immunoblotting (Fig. 3B and C). Purified proteins present in the third eluate were size separated by SDS-PAGE in two lanes and stained using colloidal coomassie (Fig. 3D). The top part of the gel lanes that contain the majority of the SUMO conjugates was cut in ten slices and proteins were in-gel digested by trypsin. Peptides were eluted from the gel slices and analyzed by LTQ-FT-ICR. A
complete list of the 1656 arginine-containing peptides matching 324 different proteins identified in this screen is available as supplemental material.

Although the SILAC technique can quantify changes smaller than 10% (45,46), we chose 1.5 as a conservative cutoff ratio. 53 proteins were detected whose identity could be confirmed by at least two arginine-containing peptides and that were enriched at least 1.5 fold in the heavy arginine form (Fig. 3E and table 1). 25 of these proteins were enriched in the Arg6-labelled form, indicating conjugation to SUMO-1, while 19 were enriched in the Arg10-labelled form, indicating conjugation to SUMO-2. 9 proteins were enriched in both heavy arginine labeled forms, as compared with the control fraction, indicating conjugation to both SUMO-1 and SUMO-2.

Of the 53 SUMO target proteins selected, 9 have previously been found in SUMO target protein screens. These proteins are RanGAP1 (6,7,8,9), PML (47-50), Delta 1-pyrroline-5-carboxylate synthetase (35,33), 82kDa FMRP interacting protein, Ataxin-2-like protein, PP1 regulator, Hypothetical protein DKFZp434D1319 (33) and SART1 (31-35). Three proteins were previously identified in our screen for SUMO-2 targets, SART1, Ataxin-2-like protein and a homolog of Zinc finger protein 106 (30). The lowest ratio of heavy arginine to light arginine that we observed for peptides from these 9 known SUMO target proteins was 1.53 for the PP1 regulator. This confirms the validity of using 1.5 as the SILAC cutoff ratio to assess specificity. We recognize that using the relatively low cutoff ratio of 1.5 harbors the risk of including false positive SUMO targets in the table. On the other hand, we feel this is justified to avoid rejecting genuine SUMO targets. The relatively low SILAC ratios observed are likely due to the low abundance of sumoylated proteins and the co-purification of contaminating, non-sumoylated proteins.
Many proteins are conjugated to SUMOs via lysines present in the sumoylation consensus motif $\psi$KxE/D where $\psi$ is V, I, L, M or F (3-5). We searched for the presence of this sumoylation consensus motif in the 53 proteins identified. As shown in table 2, a total of 112 consensus sumoylation sites were found in 39 of the 53 proteins (74% of total). Based on a small number of SUMO target proteins, the sumoylation consensus motif was initially defined as $\psi$KxE, where $\psi$ is V, I or L (47). A total of 75 sumoylation sites that match this motif are found in 33 of the 53 proteins (62% of total). These frequencies were compared to the frequencies of sumoylation sites in the 13124 human proteins present in Swissprot release 48.5. 13207 [VILMF]Kx[ED] type sumoylation sites were found in 6849 proteins (52% of total) and 6709 [VIL]KxE type sumoylation sites were found in 4318 proteins (33% of total). Thus, our set of SUMO target proteins is enriched in sumoylation consensus motifs, indicating the validity of our strategy to identify endogenous SUMO target proteins. The 14 proteins that are lacking a consensus site for sumoylation are not necessarily false-positives, however, because several proteins have previously been found to be conjugated to SUMO via lysines that are not situated in sumoylation consensus sites (3,4,19,21).

A striking feature of many SUMO-1 target proteins is the presence of one or more zinc fingers (table 1), in agreement with a previous study (32). 14 of the 25 SUMO-1 conjugated proteins contain zinc fingers, in contrast to 3 zinc finger-containing proteins out of 19 SUMO-2 conjugates and one zinc finger-containing protein out of 9 SUMO-1 and SUMO-2 conjugates. These zinc fingers are mainly C2H2-type, but C3H1-type- and C3HC4-type zinc fingers are also found.
Many zinc finger proteins play a role in transcription, therefore it is not surprising that one third of the identified proteins are transcriptional regulators. This functional group includes 10 preferential SUMO-1 targets, 6 preferential SUMO-2 targets and one protein that is possibly conjugated to both SUMOs, showing that both SUMO family members play a role in transcription. SUMO target proteins are furthermore involved in signaling, metabolism, cell cycle regulation, glycosylation, DNA repair, pre-mRNA splicing, RNA editing and other cellular processes, providing more evidence for the broad impact of SUMOs on cells (table 1).

We compared our data set with data from previous studies on the target proteins that were identified for yeast SUMO, Smt3 (19-23). Eight yeast homologs of the human SUMO target proteins were also identified as targets for Smt3 (table 3). Homologs were found for four preferential SUMO-1 targets, three preferential SUMO-2 targets and three target proteins for SUMO-1 and SUMO-2. A single yeast homolog, CHD1, exists for the human proteins Chromodomain helicase DNA-binding protein 2 and Chromodomain helicase DNA-binding protein HELSNF1 and a single yeast homolog, PBP1, exists for the Ataxin-2 protein and the Ataxin-2-like protein. All the proteins in table 3 contain sumoylation consensus sites, with the exception of Histone H2B and PBP1.

Several examples of peptide spectra corresponding to SUMO-1 and/or SUMO-2 target proteins are given in Fig. 4. Chromodomain Helicase DNA-binding protein 2 (CHD2) is an example of a preferential SUMO-1 target (Fig. 4B) whereas Chromodomain Helicase DNA-binding protein 1 (CHD1) is an example of a non-sumoylated protein (Fig. 4A). Interestingly, CHD1 is missing 5 of the 8 consensus sites for sumoylation that are found in CHD2 (table 2). RanGAP1 is preferentially conjugated
to SUMO-1, but the data also indicate the existence of a SUMO-2 linked form (Fig. 4C). SART1 is a target for both SUMO family members (Fig. 4D), whereas PML (Fig. 4E) and the Ataxin-2 Related Domain Protein (Fig. 4F) are preferential SUMO-2 targets.

**Confirmation of SUMO-1- and SUMO-2 target protein preferences by immunoblotting.**

To confirm that SUMO-1 and SUMO-2 display target protein preferences and also share target proteins, immunoblotting experiments were performed (Fig. 5A). His\textsubscript{6}-SUMO-1 and His\textsubscript{6}-SUMO-2 conjugated proteins were purified separately and control HeLa cells were included in the experiment. Proteins were size-separated by SDS-PAGE, blotted to membranes and SUMO target proteins were detected by specific antibodies. In line with the results obtained for RanGAP1, this protein is detected as a preferential target for SUMO-1, however, a SUMO-2 conjugated form is also detectable (Fig. 5A). Note that SILAC was able to find small amounts of RanGAP1 in the negative control fraction, whereas no RanGAP1 could be found in the negative control by immunoblotting. SILAC is apparently more sensitive in detecting proteins in the negative control fraction than immunoblotting. Sp100 (50) is an example of a specific target for SUMO-2 and SART1 is conjugated to both SUMO family members at similar levels (Fig. 5A).

These results were obtained using endogenous SUMO target proteins, but exogenous His\textsubscript{6}-SUMOs. To investigate also endogenous SUMOs conjugated to endogenous proteins, immunoprecipitation assays were employed (Fig. 5B). A major technical challenge in this type of experiment is the presence of SUMO proteases in extracts that remove SUMOs from target proteins. To inhibit these proteases, N-
ethylmaleimide was added to lysis buffers. Sp100 and RanGAP1 were immunoprecipitated and SUMO attachment was studied by immunoblotting. The antiserum against SART1 failed to function in these immunoprecipitation experiments. The results confirmed that endogenous RanGAP1 is preferentially conjugated to endogenous SUMO-1 and endogenous Sp100 is preferentially conjugated to endogenous SUMO-2/3.
**DISCUSSION**

In contrast with the single SUMO found in lower eukaryotes, vertebrates express three SUMO family members that are attached to substrate proteins (3-5). The recently developed quantitative proteomics tool SILAC (39) enabled us to identify and quantify target proteins conjugated preferentially to SUMO-1 and SUMO-2, respectively. Three sets of proteins were identified: SUMO-1 preferential targets, SUMO-2 preferential targets and also several proteins that are conjugated to both SUMO-1 and to SUMO-2. We confirmed these results by immunoblotting experiments for a subset of SUMO target proteins.

The purification and identification of sumoylated proteins has been hampered by the low abundance of many SUMO targets, the finding that usually only a small fraction of a protein is sumoylated at any time and the high activity of SUMO proteases (3). We have chosen to deal with these serious technical challenges using the novel approach of combining immobilized metal affinity chromatography with stable isotope labeling. Because it is essential to both enrich the sumoylated target proteins and to block the action of SUMO proteases by using denaturing buffers, this limits in practice the choice of the affinity tag that can be used. The 6His tag is compatible with the use of a denaturing 8M Urea buffer and significant enrichment of tagged proteins can be obtained using immobilized metal affinity chromatography. Nevertheless, the resulting purified fractions, while enriched, are never completely pure and inevitably contain a variety of contaminating, non-specific proteins. Such contaminants are always observed using this and related methodologies and arise for several reasons. For example, they include the
proteins that interact with the immobilized metal Cobalt via internal histidine rich regions and other abundant, “sticky” proteins that bind via lower affinity ionic interactions. SILAC is employed to discriminate between such inevitable contaminants and the *bona fide* SUMO targets by accurately and objectively quantitating the specific enrichment of proteins above background levels. In addition, SILAC enables the quantitation of proteins that are preferentially conjugated to either 6His-SUMO-1 or 6His-SUMO-2. As an example, figure 5A shows that a small amount of SART1 can interact in its non-sumoylated form with immobilized Cobalt. Sumoylated forms can in addition be purified from lysates of 6His-SUMO expressing cells and SILAC is able to detect the larger amounts of SART1 present in the heavy arginine labeled forms, corresponding to the 6His-SUMO-1 and 6His-SUMO-2 conjugated fractions. Importantly, our successful identification of a number of known SUMO target proteins provides a powerful positive control that further underlines the validity of our approach.

Several lines of evidence further support the notion that different SUMO family members display target protein preferences *in vivo*. Preferential conjugation of RanGAP1 to SUMO-1 was previously noted by Saitoh and Hinchey (28). In two proteomics approaches using SUMO-1 and SUMO-3, the sets of identified target proteins were also only partially overlapping, but the interpretation of these results is more complicated due to the use of non-quantitative proteomics approaches (33,35).

Interestingly, proteins are conjugated *in vitro* to SUMO-1 and SUMO-2 by the E2 enzyme Ubc9 with similar efficiency (51). This indicates that, in addition to SUMOs, target proteins and the E1- and E2-enzymes that are employed in sumoylation reactions *in vitro*, additional factors may be present in cells that regulate the preferential usage of
SUMO-1 or SUMO-2. E3-enzymes are likely candidates to fulfill this role \textit{in vivo}. In agreement with this hypothesis, it was shown that adding a fragment of the SUMO E3-ligase RanBP2 to \textit{in vitro} sumoylation assays involving either PML or Sp100 stimulates the preferential usage of SUMO-2 over SUMO-1 (51). This is in line with our results demonstrating the preferential conjugation of Sp100 to SUMO-2. However, it is currently unclear whether RanBP2 also regulates SUMO-2 conjugation of PML and Sp100 in cells. The elucidation of the cellular mechanism underlying target protein preferences for different SUMO family members is therefore an important future objective.

Previously, it has been shown that Sp100 can also be conjugated to SUMO-1 in an interferon-dependent manner (50). This could indicate that SUMO target protein preferences can be stimulus dependent. A more detailed study of conditional sumoylation is required to obtain better insight in the target protein preferences of different SUMO family member upon activation of specific cellular signaling pathways.

In addition to differences in target protein preferences for SUMO-1 and SUMO-2/3, the relative amount of conjugated SUMO compared to free SUMO is also different between these SUMO family members. It has been shown that a large pool of free non-conjugated SUMO-2/3 exists in COS-7 cells compared to SUMO-1 that mainly exists in the protein-conjugated form (28). The free SUMO-2/3 pool is conjugated to target proteins in a stress-dependent manner. In contrast to the situation in COS-7 cells, the pool of free SUMO-2/3 in the HeLa cells used by us appears to be small, although it is probably larger than the pool of free SUMO-1, and many proteins are SUMO-2/3 conjugated in a stress-independent manner (Fig. 1). Thus, cell-type specific differences in conjugation-efficiencies of SUMO-2/3 appear to exist.
The three SUMOs also differ in their ability to form SUMO chains. This occurs via an internal sumoylation site that is present in SUMO-2 and SUMO-3 but is missing in SUMO-1 (43). SUMO-2 chains are formed on PML in vitro and SUMO-2 dimers have been found attached to HDAC4 (43).

Evidence exists that the closely related SUMO-2 and SUMO-3 proteins also display functional differences. Although mature SUMO-2 and SUMO-3 are nearly identical, the precursor proteins differ substantially in their C-termini (43). This could indicate that the processing of the precursor proteins occurs differently, or is mediated by different SUMO proteases. The SUMO protease SENP2 has indeed been shown to catalyze the maturation of pre-SUMO-2 and pre-SUMO-3 with strikingly different efficiencies and this difference can be attributed to the differences in C-termini of these proteins (52). Whether these differences between SUMO-2 and SUMO-3 affect their conjugation to target proteins is currently unclear.

In summary, we have identified a set of novel potential SUMO target proteins and in addition confirmed several previously identified SUMO conjugates. Some of these proteins are preferentially conjugated to SUMO-1, other proteins are preferentially conjugated to SUMO-2 and a third set of proteins was found to be conjugated to both SUMO-1 and SUMO-2. This indicates that SUMO-1 and SUMO-2 have probably both redundant and non-redundant cellular functions.
ACKNOWLEDGEMENTS

We would like to thank Drs. Peter ten Dijke and Hans Tanke for critically reading the manuscript. AIL is a Wellcome Trust Principal Research Fellow. This work was supported by the Netherlands Organisation for Scientific Research (NWO) to ACOV as part of the Innovational Research Incentives Scheme and by a fellowship from the Dutch Cancer Society to ACOV, by a Wellcome Trust Programme grant to AIL, by a BBSRC grant to RTH and by a generous grant from the Danish National Research Foundation to JSA and MM.
REFERENCES


protein modification system in Arabidopsis. Accumulation of SUMO1 and -2 conjugates is increased by stress. *J. Biol. Chem.* 278, 6862-6872


49. Muller, S., Matunis, M. J., and Dejean, A. (1998) Conjugation with the ubiquitin-related modifier SUMO-1 regulates the partitioning of PML within the nucleus. EMBO J. 17, 61-70


FIGURE LEGENDS

Fig 1. Differential conjugation of endogenous SUMO-1 and endogenous SUMO-2/3 to target proteins. 

(A and B) SUMO-1 and SUMO-2 proteins were produced in *E. coli* and purified. 5 ng protein samples were subjected to SDS-PAGE, transferred to membranes, and probed using monoclonal antibody 21C7 directed against SUMO-1 (A) and polyclonal antibody AV-SM23-0100 directed against SUMO-2/3 (B). (C and D) Whole cell extracts of HeLa cells were separated by SDS-PAGE, transferred to membranes, and probed using antibodies 21C7 or AV-SM23-0100. Unconjugated SUMO is indicated by an arrow and SUMO-conjugated RanGAP1 is indicated by an asterisk. The figure is composed of immunoblotting results representing short exposure times (C) and long exposure times (D).

Fig 2. A quantitative proteomics strategy to identify SUMO-conjugated proteins. 

HeLa cells were labeled with Arg0, HeLa^{His6-SUMO-1} cells were labeled with Arg6 and HeLa^{His6-SUMO-2} cells were labeled with Arg10. Equal amounts of lysates from the three different populations were mixed and proteins conjugated to His<sub>6</sub>-SUMO were affinity purified on Cobalt agarose. The SUMO enriched fraction was separated by SDS-PAGE, the gel lane was cut in slices and the proteins present in these slices were digested by trypsin and identified by mass spectrometry. Peptide mass spectra of proteins were quantified to identify proteins potentially conjugated to SUMO-1 and/or SUMO-2.
Fig 3. Purification of SUMO conjugates. HeLa cells were labeled with Arg0, HeLa \textsuperscript{His6-SUMO-1} cells were labeled with Arg6 and HeLa \textsuperscript{His6-SUMO-2} cells were labeled with Arg10. Nuclei were isolated and lysed in 8M Urea. Equal amounts of protein were used in purification experiments. (A) Three separate purifications were performed on Cobalt columns, using 10% of each lysate, to purify \textsuperscript{His6-SUMO-1} and \textsuperscript{His6-SUMO-2} conjugated proteins separately. Proteins were size-separated by SDS-PAGE, transferred to a membrane and probed using an antibody directed against the His-tag. A non-specific (n.s.) band is indicated by an asterisk. (B & C) Equal amounts of protein from the three different lysates were mixed 1:1:1 and sumoylated proteins were purified on a Cobalt column. Proteins were eluted from the column in fractions. Small aliquots of each fraction were size-separated by SDS-PAGE, transferred to membranes and probed using antibody 21C7 directed against SUMO-1 (B) or antibody AV-SM23-0100 directed against SUMO-2/3 (C). (D) Fraction 3 was size separated by SDS-PAGE and stained for total protein. The top part of the gel-lanes was cut in 10 slices and proteins present in each slice were identified by trypsin digestion and mass spectrometric analysis. (E) Short summary of the identified proteins. 53 potential SUMO target proteins were identified with a SILAC score of at least 1.5, including 25 preferential SUMO-1 targets, 19 preferential SUMO-2 targets and 9 proteins that show little or no preference for SUMO-1 or SUMO-2.

Fig 4. Peptide mass spectra of SUMO target proteins. Peptide mass spectra of proteins were quantified and peptides matching the control fraction (Arg0), the SUMO-1 enriched fraction (Arg6) and the SUMO-2 enriched fraction (Arg10) are indicated. Chromodomain
helicase DNA-binding protein 1 is an example of a non-sumoylated protein (A), Chromodomain helicase DNA-binding protein 2 (B) and RanGAP1 (C) are examples of preferential SUMO-1 targets, PML (E) and Ataxin-2 Related Domain Protein (F) are examples of preferential SUMO-2 targets and SART1 is an example of a protein that is conjugated to both SUMO family members (D).

Fig 5. Preferential conjugation of proteins to SUMO-1 or SUMO-2.
(A) His\textsubscript{6}-SUMO conjugates were purified from HeLa\textsuperscript{His\textsubscript{6}-SUMO-1} nuclei and HeLa\textsuperscript{His\textsubscript{6}-SUMO-2} nuclei. Control purifications from HeLa nuclei were included in the experiment. Proteins were size-separated by SDS-PAGE, transferred to membranes and probed using antibodies directed against endogenous SUMO target proteins. RanGAP1 is preferentially conjugated to SUMO-1, Sp100 is conjugated to SUMO-2 and SART1 is conjugated to both SUMOs at similar levels. (B) Preferential conjugation of endogenous target proteins to endogenous SUMO-1 or endogenous SUMO-2/3. The SUMO-1 target protein RanGAP1 and the SUMO-2 target protein Sp100 were immunoprecipitated from HeLa cell lysates, transferred to a membrane and probed using antibody 21C7 directed against SUMO-1 or antibody AV-SM23-0100 directed against SUMO-2/3.
TABLES

Table 1: SUMO target proteins.
Proteins identified by at least two arginine-containing peptides with a minimum SILAC cutoff ratio of 1.5 are presented here and include 25 SUMO-1 targets (top part), 19 SUMO-2 targets (middle part) and 9 proteins conjugated to both SUMO family members (bottom part). 14 of the 25 SUMO-1 targets contain one or more zinc fingers (ZF). A complete list of identified arginine-containing peptides can be found in the supplemental dataset.

a) SILAC enrichment ratios of Arg6-labelled peptides, indicating conjugation of proteins to SUMO-1.
b) SILAC enrichment ratios of Arg10-labelled peptides, indicating conjugation of proteins to SUMO-2.

Table 2: Consensus sumoylation sites found in the identified SUMO target proteins.
A total of 112 [VILMF]Kx[ED] consensus sites for sumoylation were identified in 39 of the 53 SUMO targets. Indicated are the positions of the sumoylated lysines in the respective proteins and the tetramer sequences involved.

Table 3: A comparison between previously published Smt3 target proteins and the SUMO target proteins identified in this study.
Ten proteins identified in our screen for SUMO-1 and SUMO-2 targets have yeast homologs that were previously identified as targets for Smt3. Consensus sumoylation sites are found in all these proteins, except for Histone H2B and PBP1.
Supplemental Table 1: Proteins Identified by LTQ-FT-ICR.
Figure 1

A

recombinant protein

SUMO-1  SUMO-2

anti SUMO-1

B

recombinant protein

SUMO-1  SUMO-2

anti SUMO-2/3

C

HeLa  HeLa

anti SUMO-1  anti SUMO-2/3

short exposure

* RanGAP1

D

HeLa  HeLa

anti SUMO-1  anti SUMO-2/3

long exposure

free SUMO
Figure 2

HeLa

HeLa

HeLa

Arg0

Arg6

Arg10

Mix lysates 1:1:1

Affinity purification of Hi6-SUMO

Size separation of proteins by SDS-PAGE

Cutting of gel slices and digestion of proteins with trypsin

Protein identification and quantitation by LTQ-FT-ICR
Figure 3

A. Western blots probed with anti-His5 of His6-SUMO-1, His6-SUMO-2, and HeLa cell lysate show bands at 191, 97, 64, 51, 39, 28, 19, and 14 kDa. The asterisk indicates a non-significant (n.s.) band.

B. Western blots probed with anti-SUMO-1 of eluted fractions 1 to 7 show bands at 191, 97, 64, 51, 39, 28, 19, and 14 kDa. Free SUMO is marked by an arrow.

C. Western blots probed with anti-SUMO-2/3 of eluted fractions 1 to 7 show bands at 191, 97, 64, 51, 39, 28, 19, and 14 kDa. Free SUMO is marked by an arrow.

D. Coomassie blue staining of a gel showing molecular weight markers and fraction 3.

E. Venn diagram showing the overlap of SUMO-1 and SUMO-2 with 25, 9, and 19 elements in common.
Figure 4
Figure 5

A

HeLa  His6-SUMO-1  His6-SUMO-2

anti RanGAP1

anti Sp100

anti SART1

64  64  97

B

Immunoprecipitations

αRanGAP1  αSp100  control 1  control 2

anti SUMO-1

anti SUMO-2/3

64  64
<table>
<thead>
<tr>
<th>Accession number</th>
<th>Protein name</th>
<th>Zinc Fingers</th>
<th># Arg peptides</th>
<th>SUMO-1, 2 control (a) StDev</th>
</tr>
</thead>
<tbody>
<tr>
<td>SWISS-PROT:Q92466</td>
<td>pre-mRNA splicing transcriptional regulator, helicase</td>
<td>3</td>
<td>2</td>
<td>4.57</td>
</tr>
<tr>
<td>SWISS-PROT:Q01082</td>
<td>uncharacterized novel protein</td>
<td>3.65</td>
<td>0.59</td>
<td>0.73</td>
</tr>
<tr>
<td>SWISS-PROT:Q8NFD5</td>
<td>zinc finger HIT domain containing protein 4</td>
<td>2.39</td>
<td>1.48</td>
<td>2.79</td>
</tr>
<tr>
<td>SWISS-PROT:P05187</td>
<td>metabolism</td>
<td>1.10</td>
<td>1.52</td>
<td>2.49</td>
</tr>
<tr>
<td>SWISS-PROT:Q8IVS2</td>
<td>zinc finger protein 96</td>
<td>1.58</td>
<td>2.08</td>
<td>2.50</td>
</tr>
<tr>
<td>SWISS-PROT:O15294</td>
<td>protein phosphatase 1 regulator</td>
<td>4.00</td>
<td>0.25</td>
<td>0.00</td>
</tr>
</tbody>
</table>

**Table 1**

Protein name | Accession number | Functional classification | Zinc Fingers | # Arg peptides | SUMO-1/2 control (a) StDev
---|---|---|---|---|---
...
<table>
<thead>
<tr>
<th>Accession number</th>
<th>Sumoylation sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>SWISS-PROT:Q96QC0</td>
<td>179VKAE</td>
</tr>
<tr>
<td>SWISS-PROT:Q01082</td>
<td>1108IKNE, 1957IKAE</td>
</tr>
<tr>
<td>SWISS-PROT:P54886-1</td>
<td>258MKTD</td>
</tr>
<tr>
<td>SWISS-PROT:O14647</td>
<td>125IKEE, 623LKND, 702VKKD, 842IKGE, 1285IKTD, 1347VKKE, 1356LKEE, 1379VKDD</td>
</tr>
<tr>
<td>SWISS-PROT:Q9C0J8</td>
<td>321LKEE, 530LKLE, 560LKIE</td>
</tr>
<tr>
<td>SWISS-PROT:Q9H0J9</td>
<td>not found</td>
</tr>
<tr>
<td>TREMBL:Q6BEP6</td>
<td>615IKQE, 885IKLE, 913VKQE</td>
</tr>
<tr>
<td>TREMBL:Q9P2A7</td>
<td>243LKQE, 306LKEE, 1004MKKE</td>
</tr>
<tr>
<td>SWISS-PROT:O15294</td>
<td>119LKPD, 337IKRE , 459LKPD</td>
</tr>
<tr>
<td>TREMBL:Q569L7</td>
<td>not found</td>
</tr>
<tr>
<td>SWISS-PROT:Q14865</td>
<td>105VKLE, 130VKTE, 284VKCE, 410IKGE, 445IKHE, 629VKVD, 774FKHE, 803LKQE</td>
</tr>
<tr>
<td>TREMBL:Q5BJG6</td>
<td>123FKAD, 389MKSD</td>
</tr>
<tr>
<td>TREMBL:Q9HA80</td>
<td>48LKDE, 174VKME, 475IKTE, 516LKVE</td>
</tr>
<tr>
<td>SWISS-PROT:Q92610</td>
<td>not found</td>
</tr>
<tr>
<td>TREMBL:Q9NYW8</td>
<td>59IKLE</td>
</tr>
<tr>
<td>SWISS-PROT:O43309</td>
<td>20VKIE, 196LKQE</td>
</tr>
<tr>
<td>SWISS-PROT:Q15697</td>
<td>not found</td>
</tr>
<tr>
<td>SWISS-PROT:Q9NS39</td>
<td>not found</td>
</tr>
<tr>
<td>TREMBL:Q96G95</td>
<td>not found</td>
</tr>
<tr>
<td>SWISS-PROT:Q96PQ6-2</td>
<td>not found</td>
</tr>
<tr>
<td>TREMBL:Q5T0B9</td>
<td>198IKAE</td>
</tr>
<tr>
<td>TREMBL:Q7Z7K6</td>
<td>not found</td>
</tr>
<tr>
<td>SWISS-PROT:Q9NX05-1</td>
<td>747MKSD, 809LKIE, 1042IKEE</td>
</tr>
<tr>
<td>SWISS-PROT:Q9C086</td>
<td>155LKKE, 202LKRE</td>
</tr>
<tr>
<td>SWISS-PROT:Q9HCK8</td>
<td>2MKGE, 233LKEE, 330IKPE, 375VKKE, 485LKED, 734LKTE, 755LKED, 1133LKDD, 1487MKIE, 1746LKLE, 1783VLKE, 1977IKDE</td>
</tr>
<tr>
<td>SWISS-PROT:P05187</td>
<td>297MKYE</td>
</tr>
<tr>
<td>SWISS-PROT:Q9H2Y7</td>
<td>578LKIE, 911LKQE, 1200IKQE</td>
</tr>
<tr>
<td>TREMBL:Q53GB5</td>
<td>94VKRE, 141IKKE, 709VKIE, 742MKTE</td>
</tr>
<tr>
<td>SWISS-PROT:P27105</td>
<td>92IKVD, 187IKVE</td>
</tr>
<tr>
<td>SWISS-PROT:Q16778</td>
<td>not found</td>
</tr>
<tr>
<td>TREMBL:Q9HCL9</td>
<td>354VKVE</td>
</tr>
<tr>
<td>SWISS-PROT:P46060</td>
<td>306IKRD, 481FKDE, 524LKSE</td>
</tr>
<tr>
<td>TREMBL:Q7L2E3</td>
<td>not found</td>
</tr>
<tr>
<td>SWISS-PROT:Q92466</td>
<td>not found</td>
</tr>
<tr>
<td>SWISS-PROT:Q7Z417</td>
<td>74LKHE, 262FKPD</td>
</tr>
<tr>
<td>SWISS-PROT:Q8WWM7</td>
<td>573LKEE, 641IKGE</td>
</tr>
<tr>
<td>SWISS-PROT:Q9Y618</td>
<td>672LKME, 863VKSE, 892LKAE, 1168VKGE, 1379LKRE, 1458LKYD</td>
</tr>
<tr>
<td>SWISS-PROT:Q8NFD5</td>
<td>1016LKAD</td>
</tr>
<tr>
<td>TREMBL:Q8WUY5</td>
<td>301MKAD, 746LKSE</td>
</tr>
<tr>
<td>SWISS-PROT:P30533</td>
<td>97LKLD</td>
</tr>
<tr>
<td>SWISS-PROT:Q8NCA5</td>
<td>not found</td>
</tr>
<tr>
<td>TREMBL:Q9H8X6</td>
<td>not found</td>
</tr>
<tr>
<td>SWISS-PROT:O00470</td>
<td>72LKRD</td>
</tr>
<tr>
<td>TREMBL:Q9NSV0</td>
<td>237LKIE</td>
</tr>
<tr>
<td>SWISS-PROT:P42694</td>
<td>261VKVE, 1845LKSE</td>
</tr>
<tr>
<td>TREMBL:Q59FP9</td>
<td>188LKHE, 254LKCD, 470IKME, 596LKID</td>
</tr>
<tr>
<td>SWISS-PROT:Q13129</td>
<td>693LKAE, 839VKLE, 1224LKGD, 1561VKLE, 1599IKEE, 1611IKKE</td>
</tr>
<tr>
<td>SWISS-PROT:Q8IVS2</td>
<td>325VKWE</td>
</tr>
<tr>
<td>SWISS-PROT:Q02218</td>
<td>not found</td>
</tr>
<tr>
<td>SWISS-PROT:Q99700</td>
<td>435LKRE, 800FKND</td>
</tr>
<tr>
<td>SWISS-PROT:P49761-3</td>
<td>285LKPE</td>
</tr>
<tr>
<td>SWISS-PROT:Q8N3C0-1</td>
<td>1709IKKD, 1837LKPE</td>
</tr>
<tr>
<td><strong>Table 2</strong></td>
<td></td>
</tr>
<tr>
<td>Protein name</td>
<td>Accession number</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Spectrin beta chain, brain 1</td>
<td>SWISS-PROT:Q01082</td>
</tr>
<tr>
<td>UDP-N-acetylglucosamine--peptide N</td>
<td>SWISS-PROT:O15294</td>
</tr>
<tr>
<td>ZMYND11 protein</td>
<td>TREMBL:Q5BJG6</td>
</tr>
<tr>
<td>Nuclear receptor co-repressor 2</td>
<td>SWISS-PROT:Q9Y618</td>
</tr>
<tr>
<td>Ataxin-2-like protein</td>
<td>SWISS-PROT:Q8WWM7</td>
</tr>
<tr>
<td>Ataxin-2</td>
<td>SWISS-PROT:Q99700</td>
</tr>
<tr>
<td>Histone H2B.q</td>
<td>SWISS-PROT:Q16778</td>
</tr>
<tr>
<td>Alkaline phosphatase</td>
<td>SWISS-PROT:P05187</td>
</tr>
</tbody>
</table>