Chapter 7

A Novel Method for the Identification of SUMO Acceptor Lysines in a Site-Specific Manner

Ivo A. Hendriks[1], Rochelle C. D'Souza[2],
Matthias Mann[2] and Alfred C.O. Vertegaal[1]

[1]Department of Molecular Cell Biology, Leiden University Medical Centre, Albinusdreef 2, 2333 ZA, Leiden, the Netherlands
[2]Proteomics and Signal Transduction, Max Planck Institute for Biochemistry, 82152 Martinsried, Germany

Chapter 7 has been submitted for publication

ABSTRACT

SUMOylation is a reversible post-translational modification essential for regulation of virtually all nuclear processes. Identification of sites of SUMOylation within proteins by mass spectrometry has been hampered by a bulky tryptic fragment, which thus far necessitated the use of mutated SUMO. Here, we present a SUMO-specific protease-based methodology which circumvents the bulky tryptic fragment altogether, dubbed Protease-Reliant Identification of SUMO Modification (PRISM). PRISM allows for detection of SUMOylated proteins as well as identification of specific sites of SUMOylation while using wildtype SUMO. We successfully confirmed the validity of the methodology by purification of known SUMOylated proteins. Furthermore, PRISM was employed in combination with high-resolution mass spectrometry to identify SUMOylation sites from HeLa cells under standard growth conditions. 389 wildtype SUMOylation sites on endogenous proteins were identified, with half of these sites matching the canonical KxE consensus motif. Thus, we have developed the first methodology capable of studying wildtype SUMO at the site-specific and system-wide level.

INTRODUCTION

Functional proteomes are far more complex when compared to genomes and transcriptomes, primarily due to processing of proteins and extensive post-translational modification of proteins. Hundreds of different modifications exist, ranging from small chemical modifications such as phosphorylation, acetylation and methylation, to modifications by small proteins that belong to the ubiquitin and ubiquitin-like (Ubl) family of proteins. These Ubls are covalently coupled to other proteins, thereby regulating their activity. Virtually all cellular processes are regulated by reversible protein modifications, including replication, transcription, cell-cycle progression and the DNA damage response [1].

Small Ubiquitin-like Modifier (SUMO) is a post-translational modification (PTM) of lysine residues in proteins, and plays a pivotal part in the regulation of many cellular processes ranging from transcription to genome maintenance and cell cycle control to the DNA damage response [2-6]. Precursor SUMO is processed by SUMO-specific proteases in order to generate mature SUMO [7], and subsequently conjugated to target proteins through an enzymatic cascade involving the dimeric E1 activating enzyme SAE1/2, the E2 conjugation enzyme Ubc9, and several catalytic E3 enzymes [8]. SUMOylation is often found to target lysines within the canonical consensus motif [VIL]KxE in proteins, and thus, unlike many other PTMs, shows a remarkable specificity in its conjugation behavior [9, 10]. SUMOylation of proteins is a reversible process, and a handful of SUMO-specific proteases can efficiently remove SUMO from its target proteins [7].

SUMO is essential to the viability of all eukaryotic life, with the exception of some species of yeast and fungi [8]. Ubc9 knockout mice perish at the early post-implantation stage due to chromosome condensation and segregation defects, further stressing the importance of SUMO in maintenance of the genome [11].

In humans, three different SUMOs are expressed; SUMO-1, SUMO-2 and SUMO-3. All SUMOs share the characteristic ubiquitin β-grasp fold, regardless of their limited sequence similarity to ubiquitin. SUMO-2 and SUMO-3 are nearly identical [12], differing by only three N-terminal amino acids, and are typically referred to as SUMO-2/3 as no antibody is able to distinguish between them. SUMO-1 only shares 47% sequence homology with SUMO-2/3, and is thus considered as a separate family member, although all SUMOs are conjugated to their targets by the same enzymatic machinery. SUMO-2/3 are the more abundant forms of the SUMOs [13], and a considerable amount of free SUMO-2/3 exists within the cell, allowing SUMO-2/3 to function efficiently in response to cellular stresses or changes in growth conditions. SUMO-1 is predominantly conjugated to RanGAP1. SUMO-1 and SUMO-2/3 share a significant overlap in conjugation targets, but also retain a differential conjugation specificity [14, 15].

Like other ubiquitin-like modifiers, SUMO is able to form polymeric chains by modifying itself [16, 17], an event that is upregulated under stress conditions

such as heat shock [18]. Furthermore, SUMO can interact non-covalently with other proteins through SUMO Interacting Motifs (SIMs) [8, 19, 20]. A famous examples of this interaction is the SUMO-targeted Ubiquitin Ligase (STUbL) RNF4, which recognizes poly-SUMOylated proteins through its SIMs, and subsequently ubiquitylates these targets [21, 22]. Additional examples of SIM-mediated interactions include the interaction between SUMO-modified RanGAP1 and the nucleoporin RanBP2 [23], and the localization of the transcriptional corepressor Daxx to PML nuclear bodies [24].

There is great interest in SUMO originating from various fields such as chromosome remodeling, DNA damage response, cell cycle control, transcriptional regulation and nuclear organization. SUMOylation has also become increasingly implicated as a viable target in a clinical setting [8, 25-28]. In a screen for Myc-synthetic lethal genes, SAE1 and SAE2 were identified, indicative of Myc-driven tumors being reliant on SUMOylation [26]. Furthermore, SUMOylation is widely involved in carcinogenesis [28]. Nevertheless, the system-wide knowledge of protein SUMOylation is limited to the global protein-level. Specific sites of SUMO modification have mainly been studied at the single protein level using low-throughput methodology. While proteomic approaches have elucidated hundreds of putative target proteins [14, 18, 29, 30], they have failed to elucidate SUMO acceptor lysines.

Increasingly powerful proteomics technologies have facilitated proteome-wide studies of PTMs [31-34]. Various well-studied major PTMs include phosphorylation [35, 36], acetylation [37], methylation [38] and ubiquitylation [39-43], where tens of thousands of modification sites have been identified at the system-wide level. However, site-specific identification of SUMOylation sites significantly trails behind these other PTMs, even though the amount of SUMO-modified proteins is predicted to be within the same order of magnitude as compared to other PTMs.

Besides unfavorable SUMOylation stoichiometry of proteins, the highly dynamic nature of the SUMO modification, and technical difficulties in purifying SUMO from complex samples due to robust and efficient activity of SUMO proteases – the main problem is the cumbersome remnant that is situated on the target peptides after tryptic digestion. In mammalian cells, for all SUMOs, the tryptic remnant exceeds 3 kDa in size, greatly hampering the ability for modified peptides to be resolved from highly complex samples by current MS/MS approaches. The most successful approach in identifying SUMO acceptor lysines to date, has been through use of a mutant SUMO bearing a point mutation. This SUMO-2 contains either the Q87R mutation, which is homologous to the sole yeast SUMO, Smt3, or the T90R mutation, which is homologous to ubiquitin. Additionally, all internal lysines are mutated to arginines to render the mutant SUMO immune to Lys-C. In turn, this allows for pre-digestion of the entire lysate and enrichment of SUMOylated peptides as compared to proteins, greatly diminishing the sample complexity. This approach has allowed for the identification of around 150 SUMOylated lysines

[44, 45].

Regardless of the success of approaches employing mutant SUMO, there are some drawbacks. Firstly, the substitution of all internal lysines to arginines prevents the mutant SUMO itself from being modified, effectively abrogating the ability to form polymeric chains. SUMOylation sites can be mapped while using just the Q87R or T90R mutation, but this prohibits enrichment of modified peptides, a key step in the purification process for all other major PTMs. Secondly, the usage of mutant SUMO necessitates the usage of exogenous SUMO, and thus is incompatible with the identification of lysines modified by endogenous SUMO, for example from clinical samples or animal tissues.

We have successfully developed the PRISM methodology which circumvents the cumbersome tryptic SUMO remnant. PRISM involves chemical blocking of all free lysines in a complex sample, followed by treatment with SUMO-specific proteases, and subsequent identification of the "freed" lysines by high-resolution mass spectrometry. Thus, we provide a key step towards system-wide identification of endogenous protein lysines modified by wild-type SUMO.

RESULTS

SUMO-Specific Proteases Remain Functional under Stringent Buffer Conditions

In order to overcome the cumbersome tryptic fragment left after digestion of wild-type SUMO, a methodology was devised that utilizes SUMO-specific proteases to remove the SUMO, and then employs the "freed" lysine as either a direct identifier or as an intermediate for chemical labeling (i.e. by biotin). The first step in the development of the protocol, was finding buffer conditions stringent enough to lyse cells without loss of SUMOylation due to endogenous proteases. Furthermore, the buffer had to be compatible with the following steps, including chemical labeling of all lysines, function of recombinant SUMO protease, and function of trypsin. Urea was found to be highly efficient, and lysing HeLa cells in 8 M urea in the presence of acetamide swiftly and irreversibly inactivated all endogenous SUMO proteases (**Figure 1**).

Subsequently, the HeLa lysate was diluted to lower concentrations of urea, and the activity of recombinant SENP1 and SENP2 was investigated. Strikingly, we found SENP2 to be stable and able to cleave all SUMO-2/3 from proteins at a concentration of 4 M urea (**Figure 1A**). Under these conditions, SUMO-1 was not affected by SENP2. A reduction to 2 M urea was required for SENP2 to efficiently cleave SUMO-1 off proteins (**Figure 1B**). SENP1 was found to be less effective; only affecting SUMO-2/3 at a concentration of 1 M urea, and SUMO-1 at a concentration of 2 M urea. SENP1 shows a higher affinity for cleaving SUMO-1 as opposed to SUMO-2/3, but overall is less efficient than SENP2. Therefore, SENP2 was chosen as the main protease for identification of SUMO-2/3 sites. As controls, ubiquitin levels were investigated, and found to be completely unaffected by the SUMO proteases

(**Figure 1C**), and equal total protein levels were validated by Ponceau-S (**Figure 1D**).

Sulfosuccinimidyl-acetate Efficiently and Completely Blocks all Free Lysines in Cellular Lysates

Following identification of a suitable SUMO protease, we endeavored to find an efficient and affordable way of blocking all lysines in a complex sample. To this end, sulfosuccinimidyl-acetate (SNHSA) was used to block all lysines. SNHSA irreversibly

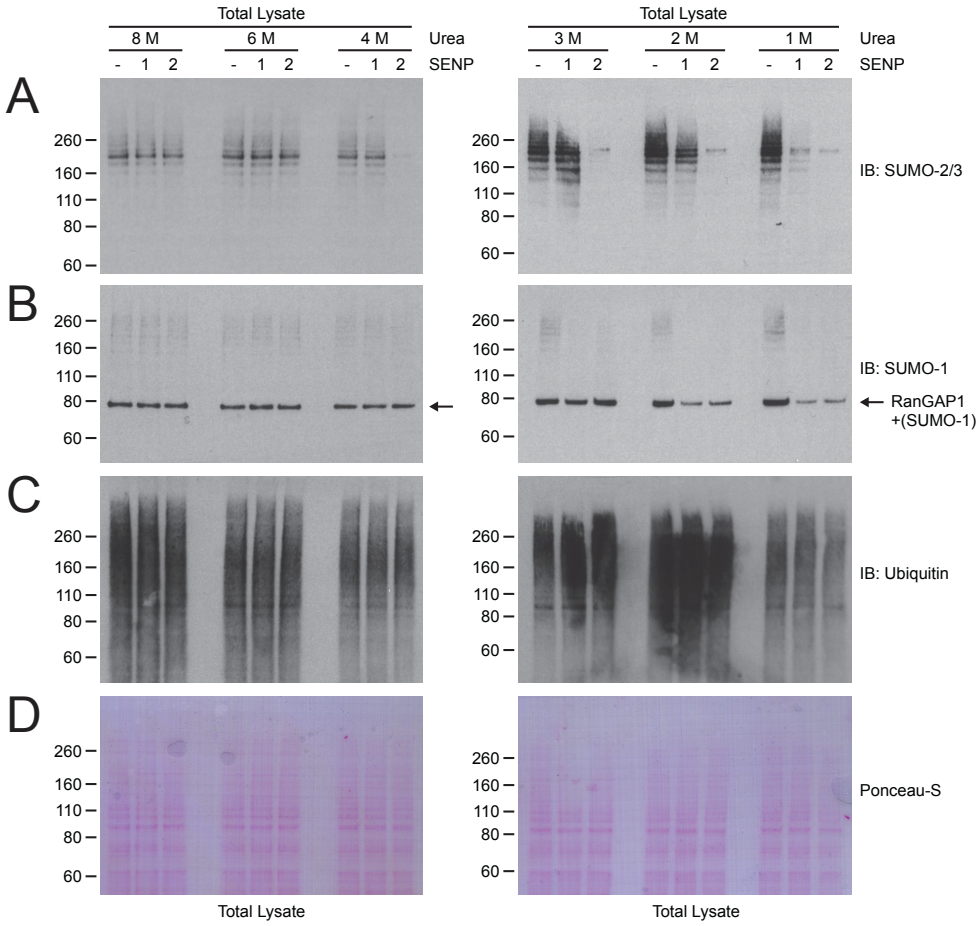


Figure 1. SENP2 is a SUMO-specific protease able to cleave SUMO-2/3 from endogenous proteins under very stringent buffer conditions.

A) HeLa cells were lysed in 8 M urea, homogenized, and subsequently diluted to indicated lower concentrations of urea. Lysates were treated with SENP1, SENP2, or mock treated. After protease treatment, lysates were size-separated by SDS-PAGE, transferred to membranes, probed using a SUMO-2/3 antibody, and visualized using chemiluminescence.

B) Same as in section A, but probed using a SUMO-1 antibody. SUMOylated RanGAP1 is indicated with an arrow.

C) Same as in section A, but probed using a ubiquitin antibody.

D) Same as in section A, but with total protein content visualized using Ponceau-S staining.

acetylates all primary amines under alkaline buffer conditions, and is commercially available at relatively low cost. The efficacy of the compound was elucidated by treating HeLa total lysate with SNHSA, and subsequently digesting either mock treated or SNHSA treated lysate with endopeptidase Lys-C and trypsin. Samples were analyzed by Coomassie to visualize total protein content, and SNHSA was observed to remarkably change the banding pattern of the HeLa lysate (**Figure 2A**). Although many size shifts occurred, the banding pattern remained sharp after treatment with SNHSA, indicative of efficient and total labeling. Furthermore, digestion with endopeptidase Lys-C, which specifically cleaves C-terminal of free lysine residues, was found to be completely ineffective on SNHSA treated HeLa lysate, demonstrating a complete protection of free lysines (**Figure 2A**). Trypsin, which additionally cuts after arginines, was still able to fully digest both mock and SNHSA treated lysates. Additionally, the effect of the SNHSA treatment on endogenous SUMO-2/3 was investigated. Similar to total protein levels, SNHSA treated SUMOylated proteins were found to be resilient to digestion by Lys-C (**Figure 2B**). Interestingly, we also observed an increase in SUMO signal after SNHSA treatment, which could be due to the SUMO-2/3 antibody more efficiently recognizing acetylated SUMO, or increased hydrophobicity of proteins altering immunoblotting behavior. The ability of SENP2 to cleave fully acetylated SUMO-2/3 from completely acetylated proteins was investigated, and SENP2 was found to still be able to do so efficiently (**Figure 2C**).

Combining Enrichment of SUMOylated Proteins with Lysine Blocking

In order to further optimize the protocol, the ability to apply the SNHSA labeling "on-beads" was investigated, i.e. during a pulldown or immunoprecipitation. This allowed pre-enrichment of SUMOylated proteins prior to treatment, and furthermore allowed washing away of excess chemical after the blocking process, making subsequent steps easier to perform. To this end, a cell line stably expressing His10-tagged SUMO-2 was utilized. Furthermore, for the purpose of unambiguous monitoring of SUMO levels during optimization of the protocol, lysine-deficient SUMO-2 was employed, which effectively abrogated internal SUMO acetylation and the resulting variability in immunoblot read-out. Pre-enrichment of His-SUMO by nickel-pulldown could be efficiently combined with blocking of all lysines with SNHSA while on-beads (**Figure 2D**). After treatment with SNHSA, acid elution was employed to prevent primary amines from being present in the elution fraction, and thus allowing for a second labeling step. More importantly, the effectiveness of recombinant SENP2 on enriched acetylated SUMOylated proteins was found to remain highly efficient (**Figure 2D**). Ponceau-S staining additionally showed an efficient removal of SUMO from its target proteins, regardless of acetylation status, and a complete immunity of acetylated SUMOylated proteins to Lys-C (**Figure 2E**).

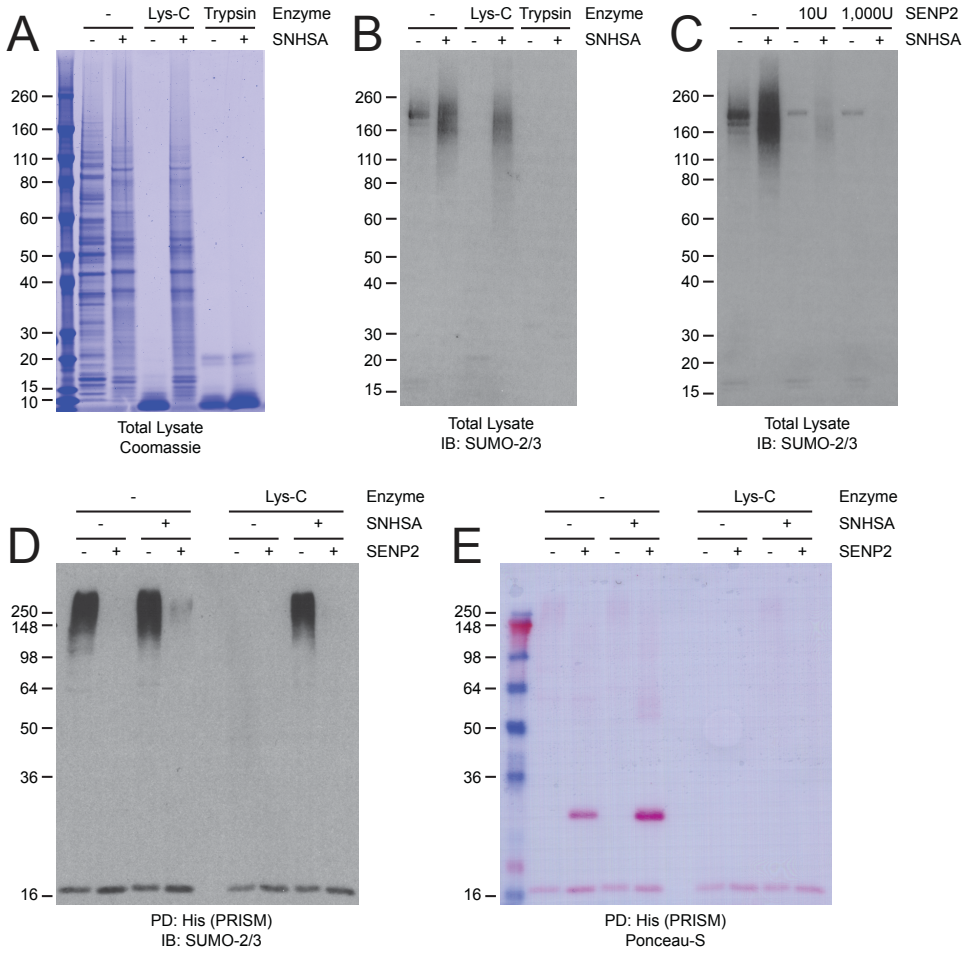


Figure 2. SENP2 is able to remove acetylated SUMO from acetylated proteins after sulfosuccinimidyl-acetate (SNHSA) treatment.

A) HeLa cells were lysed in 8 M urea, homogenized, and subsequently treated with 10 mM SNHSA or mock treated. Next, both lysine-blocked and control samples were treated with either trypsin or Lys-C, or mock treated. All samples were size-separated by SDS-PAGE, and total protein content was visualized using Coomassie.

B) Same as in section A, but with samples transferred to membranes after SDS-PAGE, subsequently probed using a SUMO-2/3 antibody, and visualized using chemiluminescence.

C) HeLa cells were lysed in 8 M urea, homogenized, and subsequently treated with 10 mM SNHSA or mock treated. Next, both lysine-blocked and control samples were treated with either a standard or large amount of SENP2, or mock treated. All samples were size-separated by SDS-PAGE, transferred to membranes, probed using a SUMO-2/3 antibody, and visualized using chemiluminescence.

D) HeLa cells stably expressing lysine-deficient His10-tagged SUMO-2 were lysed and homogenized, and subsequently SUMOylated proteins were enriched by Ni-NTA pulldown. SUMOylated proteins were treated on-beads with SNHSA or mock treated, prior to elution. After elution, proteins were either treated with SENP2 or mock treated. Finally, samples were digested using Lys-C or mock digested. All samples were size-separated using SDS-PAGE, transferred to membranes, probed using a SUMO-2/3 antibody, and visualized using chemiluminescence.

E) Same as in section D, but with total protein content visualized using Ponceau-S staining.

A Second Labeling with Biotin Allows for Repurification of Proteins deSUMOylated by SENP2

We investigated the possibility to benefit from the lysines "freed" by SENP2 in two different ways, the first being by using the free lysine as a target for a second chemical treatment. Here, sulfosuccinimidyl-SS-biotin (SNHSSSB) was employed, which functions in the same way as SNHSA. However, instead of an acetyl, SNHSSSB couples a biotin to the lysine, which is furthermore linked by a disulfide bridge. This then allowed for a second purification of proteins labeled by biotin, where they were previously modified by SUMO-2 (**Figure 3**). The efficacy of this approach was elucidated by monitoring total SUMO-2 throughout the procedure, as well as two known SUMO target proteins, APC4 and TRIM33. The initial step included enrichment of His10-SUMO-2, with acetylation performed on-beads. APC4 was found to be efficiently purified, regardless of acetylation state (**Figure 4A**). TRIM33 seemed to be less efficiently purified after acetylation (**Figure 4C**), but total internal acetylation of the protein may have interfered with the antibody recognizing the protein. Coincidently, total SUMO levels were found to be similar regardless of acetylation, due to the use of lysine-deficient SUMO, and the SUMO antibody used for immunoblot recognizing an epitope that does not contain lysines (**Figure 4E**). Following purification, both the control and acetylated samples were treated with SENP2. Here, both APC4 and TRIM33 were found to be are efficiently deSUMOylated, regardless of acetylation state (**Figure 4A and 4C**).

Subsequently, all samples were treated with SNHSSSB, and an avidin pulldown was performed to enrich biotinylated proteins. The elution was performed in two steps, initially with DTT in order to specifically cleave the disulfide bridges and elute proteins without the biotin remnant, and secondly with LDS in order to achieve total elution. In the case of APC4, a successful and extremely specific enrichment of the protein was found when correctly performing the acetylation and deSUMOylation step (**Figure 4B**). Failure to deSUMOylate resulted in an inability to biotinylate on the freed lysines, and thus no ability to purify APC4. Similar results were observed for TRIM33, but with most of the protein being eluted with LDS as opposed to DTT (**Figure 4D**). This is likely due to the multiple SUMOylation sites in TRIM33, allowing for multiple biotinylation events, which rendered the binding to the avidin beads too strong to be eluted efficiently through reduction. For both APC4 and TRIM33, failure to acetylate the proteins prior to biotinylation results in a massive immunoblot signal, resulting from incomplete and ambiguous biotinylation due to the overabundance of free lysines (**Figure 4B and 4D**). As anticipated, for total SUMO-2/3, immunoblot signal was only observed when neither acetylating nor deSUMOylating (**Figure 4F**).

Overall, we demonstrated the ability to highly specifically purify SUMO target proteins by initially capturing them through the presence of their SUMO, and then re-capturing them through the absence of their SUMO when removed by SENP2.

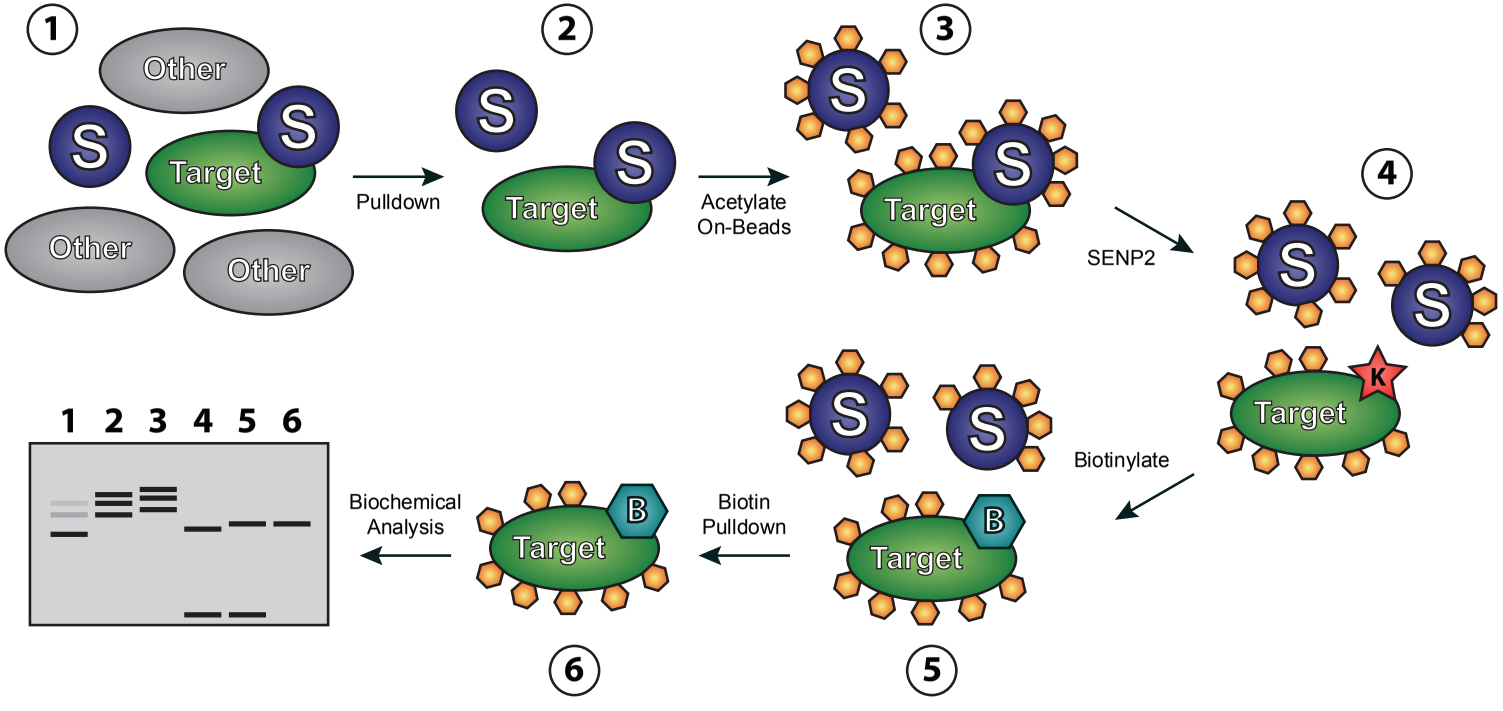


Figure 3. A schematic overview of the PRISM double purification strategy. 1. Cells are lysed under denaturing conditions, inactivating endogenous proteases. 2. SUMOylated proteins are pre-enriched using, in this case, Ni-NTA pulldown to capture the histidine tag. 3. SUMOylated proteins are acetylated on-beads using SNHSA under highly denaturing conditions, enabling efficient blocking of lysines. Excess chemical is washed away, and proteins are eluted using a buffer compatible with a second chemical labeling step. 4. Following elution, lysine-blocked SUMOylated proteins are treated with SENP2, efficiently removing SUMO-2 and freeing up the lysines SUMO-2 was conjugated to. 5. Freed lysines are biotinylated using SNHSSSB. 6. Biotinylated SUMO target proteins are enriched using avidin pulldown, and either specifically eluted using a reducing (DTT) elution buffer, or totally eluted using LDS elution buffer. Finally, SUMO target proteins may be analyzed by various biochemical methods, such as immunoblotting.

Site-Specific Identification of Lysines Modified by Wild-Type SUMO-2/3

In order to extend the PRISM strategy to mapping SUMO-2/3 sites, the methodology was slightly altered, mainly by leaving out the biotinylation and repurification step. Two concentration steps were also included in order to remove free SUMO from the samples (**Figure 5A**). Moreover, it should be mentioned that wild-type SUMO-2 was employed for all proteomics experiments, and Stable Isotope Labeling of Cells (SILAC) was applied in order to "mark" all proteins originating from the cell lysates and rule out contaminants. To this end, both "medium" and "heavy" SILAC labeling was performed, and lysates were mixed in equimolar ratio immediately after cell lysis. After the initial enrichment and acetylation of SUMO-2 target proteins, the samples were concentrated over 100 kDa cut-off filters, specifically removing free unconjugated SUMO-2 (**Figure 5B**). Subsequently, the samples were treated with SENP2 to cleave all SUMO-2 off the target proteins, followed by another 100 kDa concentration step to again remove free SUMO-2 (**Figure 5B**). Additionally, throughout the purification procedure, SUMO target protein APC4 was monitored (**Figure 5C**). Here, APC4 was successfully acetylated and enriched, and efficiently deSUMOylated before appearing in the final concentrated fraction. It should be noted that there is a loss of material throughout the preparation of the samples. This is likely resulting from hydrophobicity changes due to protein acetylation, leading to some protein precipitation during buffer changes and concentration.

Finally, the concentrated, acetylated, and deSUMOylated proteins were digested with trypsin, and analyzed using reversed-phase liquid chromatography followed by high-resolution mass spectrometry. Since all lysines other than the ones freed by SENP2 are blocked, peptides ending in a lysine or peptides which would have been preceded by a lysine can be considered as SUMOylation sites. We identified nearly 6,000 SILAC-labeled peptides resulting from digested acetylated SUMO target proteins (**Figure 5D**). These peptides confidently map to over 700 putative SUMOylated proteins, which is in line with numbers commonly found in the literature [14, 18, 30, 46]. From all peptides, roughly 1/15th contained a C-terminal lysine or were preceded by a peptide containing a C-terminal lysine. A similar number of both N-terminal (ending in a lysine) and C-terminal (preceded by a lysine) reporter peptides were found.

After filtering the data, we found 389 SUMOylation sites, with a near 50% occurrence of the KxE consensus motif (**Figure 5D**). 37 of these sites were identified by both the N-terminal and C-terminal reporter peptides, providing an extremely high identification confidence.

PRISM-identified SUMOylation Sites Adhere to the SUMO Consensus and Predominantly Occur in Nuclear Proteins with DNA-Associated Functions

The KxE frequency of sites identified with both reporter peptides was found to be over 70%, because SUMOylation preferentially occurs on KxE sites, and increased

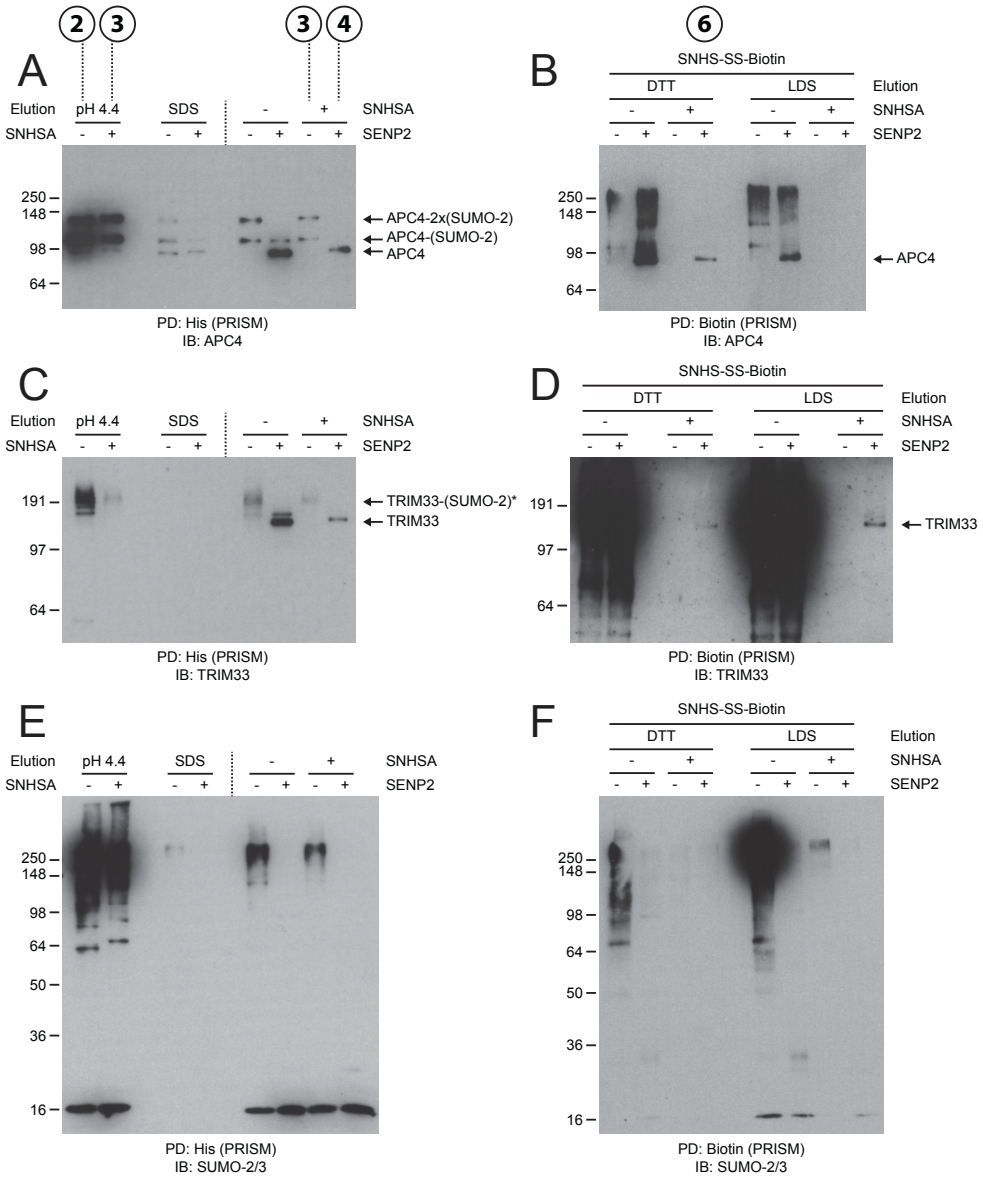


Figure 4. Two known SUMO target proteins, APC4 and TRIM33, are specifically purified using PRISM.

A) SUMOylated proteins were purified from HeLa cells expressing lysine-deficient His-tagged SUMO-2, using PRISM as described in Figure 3. For diagnostic reasons, the assay was performed with and without the use of SNHSA to block lysines, and eluted proteins were either treated with SENP2 or mock treated. Samples were size-separated using SDS-PAGE, transferred to membranes, probed using an APC4 antibody, and visualized using chemiluminescence. Non-blocked APC4 eluted after Ni-NTA pulldown, Step 2 from Figure 3, is indicated. Lysine-blocked APC4 eluted after Ni-NTA pulldown, Step 3 from Figure 3, is indicated. Lysine-blocked APC4 that was successfully deSUMOylated by SENP2, Step 4 from Figure 3, is indicated.

stoichiometry of modification would facilitate more efficient purification and identification of both reporter peptides. In order to further ascertain the quality of the sites identified by PRISM, an IceLogo was generated, and the identified frequency of amino acids surrounding the SUMOylated lysines was compared to the randomly expected frequency (**Figure 6A**). Here, a strong enrichment for the canonical SUMOylation motif [VI]KxE was observed. Leucine (L) at -1 was neither enriched nor depleted, and no enrichment of aspartic acid (D) at +2 was noted. Contrarily, enrichment of both glutamic and aspartic acid was observed at -2, indicative of the inverted SUMO consensus motif [45]. Furthermore, enrichment of the hydrophobic valine at -3 was found, indicative of the hydrophobic cluster motif [45]. A fill logo directly representing the frequency of all sequence windows was created, demonstrating the clear presence of the [VIL]KxE consensus (**Figure 6B**). A heatmap corresponding to the IceLogo was generated, and displayed a clear enrichment of lysine and glutamic acid in the region surrounding the SUMOylation sites, indicative of solvent exposure (**Figure 6C**). SUMO site sequence windows with an acid at -2 were compared to all other SUMO site sequence windows, and a significant depletion of the glutamic acid at +2 was observed (**Figure 6D**). Thus, the inverted consensus motif is likely to function autonomously, and does not always rely on an acid at +2. Finally, all PRISM-identified SUMOylation sites were mapped to their corresponding proteins, which were subsequently matched to the annotated human proteome.

Term enrichment analysis was performed in order to elucidate the overall functional characteristics and subcellular localization of this group of SUMOylated proteins. For Gene Ontology (GO) Molecular Functions, the heaviest enrichment was found for nucleic acid, DNA and zinc binding categories (**Figure 6E**). Following GO Cellular Compartments, SUMOylated proteins were observed to be primarily located in the nucleus, and further enriched in the nuclear matrix, in nuclear bodies, and at the chromatin (**Figure 6F**). GO Biological Processes revealed involvement of SUMOylated proteins in nucleic acid metabolic processes, transcription regulation, DNA double-strand break processing and RNA splicing (**Figure 6G**). Finally, a general keyword analysis revealed similar terms as the GO analyses, along with an enrichment of SUMOylation occurring on phosphorylated and acetylated proteins, as well as proteins involved in ubiquitin-like (Ubl) protein conjugation (**Figure 6H**).

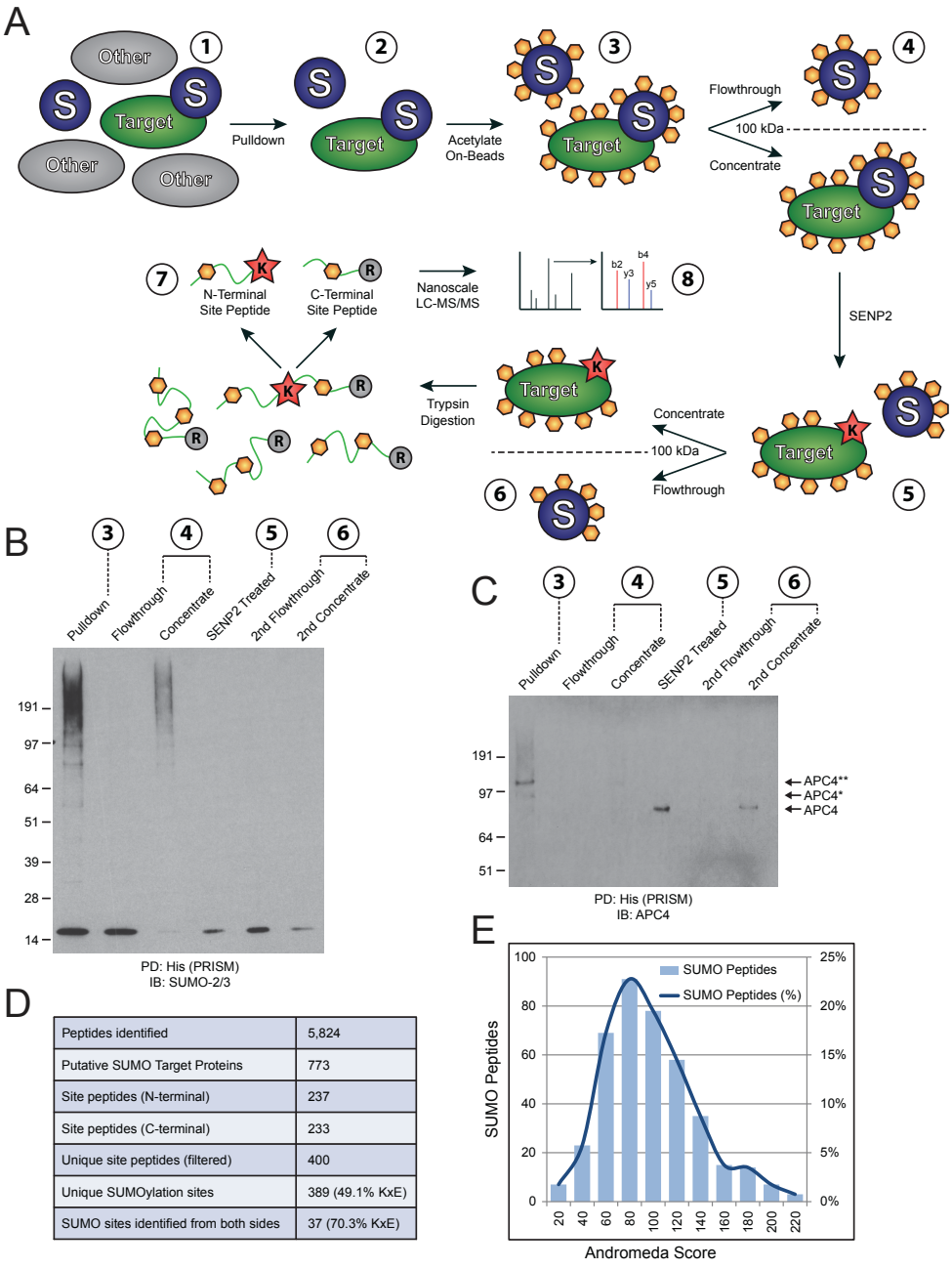
B) All samples described in section A were treated with SNHSSSB, and enriched using avidin pulldown. Elution was initially performed with DTT, and secondly with LDS. Samples were size-separated using SDS-PAGE, transferred to membranes, probed using an APC4 antibody, and visualized using chemiluminescence. APC4 that was lysine-blocked, successfully deSUMOylated by SENP2, and specifically biotinylated and purified, Step 6 from Figure 3, is indicated.

C) Same as in section A, but using a TRIM33 antibody. The asterisk indicates poly-SUMOylated forms of TRIM33.

D) Same as in section B, but using a TRIM33 antibody.

E) Same as in section A, but using a SUMO-2/3 antibody.

F) Same as in section B, but using a SUMO-2/3 antibody.



Peptides identified	5,824
Putative SUMO Target Proteins	773
Site peptides (N-terminal)	237
Site peptides (C-terminal)	233
Unique site peptides (filtered)	400
Unique SUMOylation sites	389 (49.1% KxE)
SUMO sites identified from both sides	37 (70.3% KxE)

Figure 5. PRISM combined with mass spectrometry reveals 389 unique SUMOylation sites, with half adhering to the KxE consensus.

A) A schematic overview of PRISM adapted to system-wide proteomics. 1. Cells are lysed under denaturing conditions, completely inactivating endogenous proteases. 2. SUMOylated proteins are pre-enriched using, in this case, Ni-NTA pulldown to capture the histidine tag. 3. SUMOylated proteins are acetylated on-beads using SNHSA under highly denaturing conditions, enabling

A Comparison of PRISM-identified SUMO Sites and Proteins to Other Studies

To elucidate whether SUMOylated proteins as identified by PRISM are functionally related or interaction partners, an analysis using the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database was performed [47]. Many of the identified SUMOylated proteins were situated in a single large STRING network (**Figure 7A**). The majority of proteins SUMOylated on multiple lysines were also located in this cluster. Overall, at a medium STRING confidence (p>0.4), 78.2% of all identified proteins were tied together into a single cluster, with a ratio enrichment of 14.1 over randomly expected (**Figure 7B**). At high STRING confidence (p>0.7), 53.9% of all identified proteins still resided in the core cluster, and the ratio enrichment over background increased further to 16.7.

Ambiguous identification of putative SUMOylated proteins may often lead to overestimation of a dataset. As PRISM identifies proteins at the site level, interference from background proteins is greatly reduced. However, PRISM does not directly identify a site by modification on a peptide, and thus cannot benefit from the presence of reporter ions. Therefore, to further increase confidence of our dataset, overlap analysis to other SUMOylation studies was performed. PRISM-identified proteins were compared to three major studies [14, 18, 30], and half were found to be previously identified (**Figure 7C**). Comparatively, SUMO targets identified by Bruderer et al. and Becker et al. demonstrated significantly less overlap towards other studies. 21 SUMOylated proteins were identified by all 4 studies. When comparing PRISM-identified proteins to SUMO studies performed in our lab [45,

efficient blocking of lysines. Excess chemical is washed away, and proteins are eluted using a buffer compatible with a second chemical labeling step. 4. Following elution, proteins are concentrated over a 100 kDa filter under denaturing conditions, specifically removing free SUMO but retaining SUMO-modified proteins. 5. Concentrated lysine-blocked SUMOylated proteins are treated with SENP2, efficiently removing SUMO-2 and freeing up the lysines SUMO-2 was conjugated to. 6. DeSUMOylated proteins are concentrated over a 100 kDa filter under denaturing conditions, removing SUMO cleaved off by SENP2 while retaining proteins previously SUMO-modified. 7. Concentrated SUMO target proteins are trypsinized, with trypsin only able to cleave arginines because all lysines in proteins are blocked, with the exception of the lysines that were freed by SENP2. Thus, two reporter peptides are generated per site, either ending in a lysine or preceded by a lysine. 8. Peptides are analyzed using high-resolution mass spectrometry.

B) SUMOylated proteins were purified from medium and heavy SILAC labeled HeLa cells stably expressing His-tagged SUMO-2, using PRISM as described in section A. During various steps of the purification samples were taken for diagnostic purposes. Samples were size-separated using SDS-PAGE, transferred to membranes, probed using a SUMO-2/3 antibody, and visualized using chemiluminescence. Samples are indicated by the corresponding step number from section A.

C) Same as in section B, but using an APC4 antibody.

D) An overview of all identified peptides, putative SUMO target proteins, and all reporter peptides mapping to unique SUMOylation sites. A small number of sites was identified by both reporter peptides.

E) A schematic overview of the Andromeda confidence scores for all peptides mapping to SUMOylation sites.

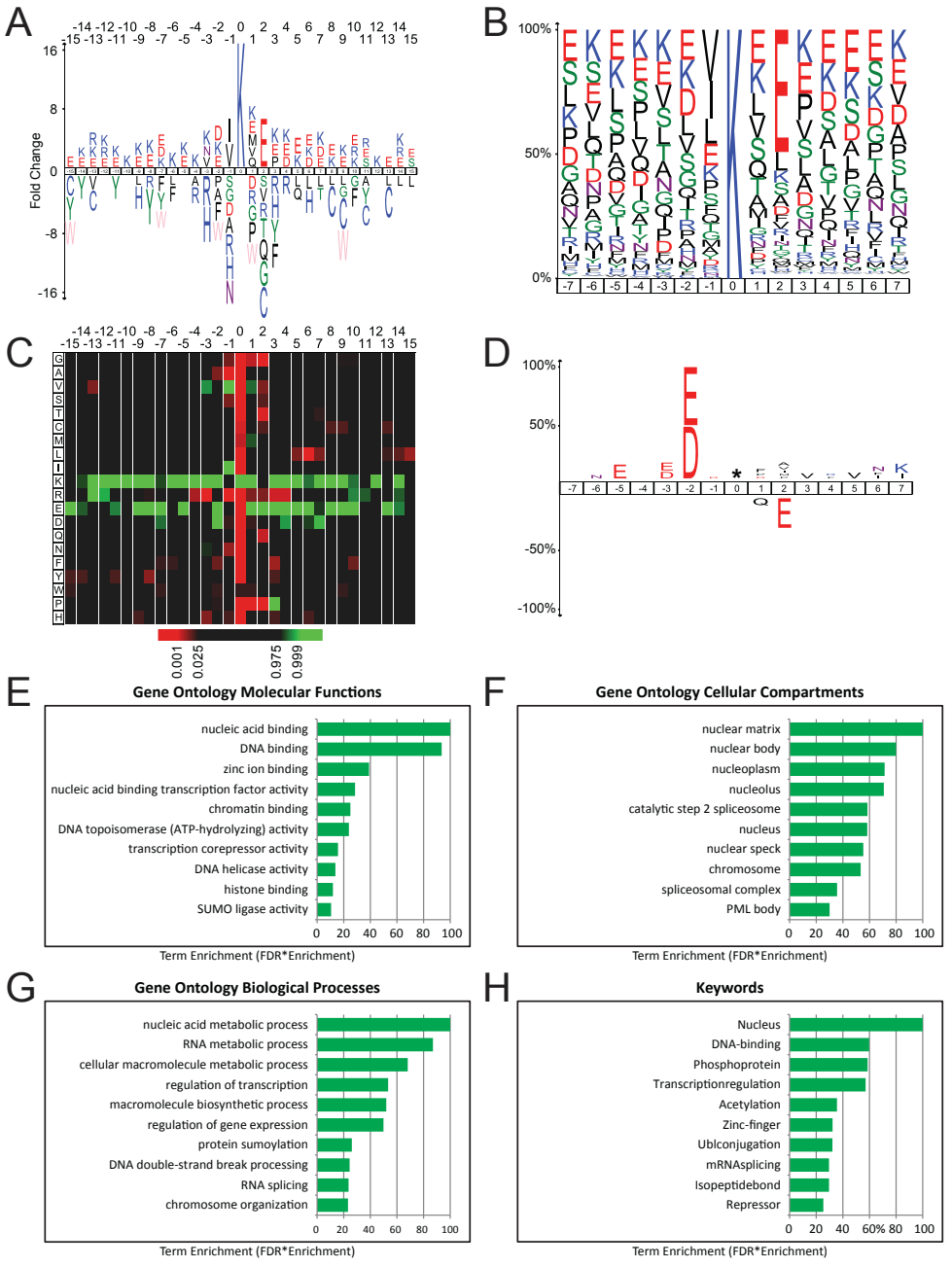


Figure 6. SUMOylation occurs predominantly at the consensus motif [VI]-K-x-E, on nuclear proteins, with functions in transcription, DNA repair, RNA splicing and chromatin remodeling.
A) IceLogo of all PRISM-identified SUMO-2 sites and their surrounding amino acids, ranging from -15 to +15 relative to the modified lysine. Amino acids indicated are contextually enriched or depleted as compared to randomly expected, with the height of the amino acids being representative of fold-change. All changes are significant with $p < 0.05$.

46][Chapter 6], over three quarters were previously detected (**Figure 7D**). Between our studies, overlap was generally highly significant, with the smaller studies being nearly completely enveloped by the larger studies.

At the site level, PRISM was compared against the more commonly applied QQTGG mapping using Q87R SUMO-2 mutants. All large-scale SUMO site studies originate from our lab, and hence the only comparison performed was against our own screens [45, 46][Chapter 6]. A significant overlap between PRISM and the other three studies was observed, with nearly half of all sites being previously identified in other screens (**Figure 7E**). Similarly to comparison at the protein level, the smaller studies were found to be enveloped within the larger studies. Finally, PRISM-identified SUMO sites were compared to known acetylation and ubiquitylation sites, and roughly one-quarter of the SUMOylation sites were found to be targeted by these other major lysine PTMs (**Figure 7F**), indicating extensive competition for certain lysines. Finally, we observed modification of endogenous ubiquitin lysine-63 by wild-type SUMO-2, providing *in vivo* evidence for this novel hybrid Ubl chain.

DISCUSSION

We have optimized the Protease-Reliant Identification of SUMO Modification (PRISM) methodology, which tackles the main problem that persisted in the mass spectrometry field when trying to identify lysines modified by endogenous SUMO. We demonstrated the efficacy of this novel methodology by successfully purifying known SUMO targets from a complex cell lysate. Furthermore, we combined PRISM with high-resolution mass spectrometry, and identified nearly 400 wild-type SUMOylation sites on endogenous protein lysines, purified from HeLa cells. Half of these sites adhered to the [VIL]KxE consensus. SUMOylated proteins were found to be predominantly nuclear, and involved in chromatin remodeling, RNA splicing, transcription, and DNA repair. When compared to other SUMOylation studies, a

B) FillLogo of all PRISM-identified SUMO-2 sites and their surrounding amino acids, ranging from -7 to +7, with the height of the amino acids directly correlating to percentage representation.
C) Heatmap representation of section A, giving a quick overview of enriched (green) and depleted (red) amino acids surrounding SUMOylated lysines. Lysines and glutamic acids are enriched across the entire range surrounding SUMOylation.
D) Sub-IceLogo, comparing sequence windows of inverted SUMO sites (E or D at -2) to sequence windows of non-inverted SUMO sites. Amino acid height corresponds to percentage enrichment or depletion between the datasets representing inverted and non-inverted sites. Displayed amino acids are significantly different between the two datasets, with $p < 0.05$.
E) Term enrichment analysis, comparing all SUMO target proteins identified by PRISM to the human proteome. Gene Ontology Molecular Functions terms were used to find statistical enrichments within the SUMO target protein dataset. Term enrichment score is a composite score based on enrichment over randomly expected and the negative logarithm of the false discovery rate. All listed terms are significant with $p << 0.02$.
F) Same as in section E, using Gene Ontology Cellular Compartments.
G) Same as in section E, using Gene Ontology Biological Processes.
H) Same as in section E, using Keywords.

significant overlap with PRISM-identified SUMO sites and SUMO target proteins was confirmed.

Our dataset provides insight into the SUMO consensus motif and the functional groups of proteins being modified by SUMO, under standard growth conditions. Regardless, the dataset is still fairly modest in size as compared to the other PTMs. The PRISM-identified sites were mapped to just over 200 proteins, which is a somewhat small number for bioinformatics analysis. Therefore, using PRISM to identify sites across multiple cell types, and in response to multiple cellular treatments, will no doubt greatly increase global knowledge about SUMOylation.

Compared to published studies on SUMO, PRISM not only provides the ability to identify wild-type SUMOylation sites, but also identified more sites in one single biological condition. PRISM only falls short to the optimized QQTGG mapping as described in Chapter 6 of this thesis. However, PRISM does not rely on mutant SUMO for its function, and the methodology can likely be further optimized to map completely endogenous SUMO sites.

In total, 171 PRISM-identified sites were previously mapped in our lab with the QQTGG methodology. 114 of these sites, exactly 2-in-3, adhere to the KxE consensus. This is higher than the overall KxE identification rates for the studies separately, which are in both cases around 50%. Sites mapped with both approaches have an extremely low chance to be false positives, and their dual identification is in part a result from higher abundance in the purified samples. This is in agreement with the KxE motif being preferentially targeted by Ubc9 [10, 48], and SUMOylation on KxE motifs likely represents the lion's share of total cellular SUMOylation.

We discovered a quarter of PRISM-identified SUMOylated lysines to overlap with ubiquitylation and acetylation. This finding is nearly synonymous to the findings reported in Chapter 6, during QQTGG mapping. The consistency of this finding across two different methods shows the reliability of both approaches. Of the 171 sites identified by both methodologies, 56 are also used by either ubiquitin, acetylation, or both. Here, the overlap is thus extended to nearly 1-in-3, hinting that the most abundant SUMOylation sites are likely more actively involved in crosstalk and competition with other lysine PTMs.

Similar approaches to PRISM, but then applied to different protein modifications, have been utilized in the last years. Notably, a study on acetylation was published that uses the biotin-switch methodology [49]. While fundamentally highly similar, PRISM solves the tryptic remnant problem that has plagued the identification of endogenous SUMOylation sites. Contrarily, acetylation does not suffer from this limitation, and many thousands of acetylation sites have been published following a direct purification using an acetyl antibody [37]. While investigation of the specific activity of the protease in question remains of interest, the fidelity of these proteases *in vitro* is often not directly comparable to *in vivo* activity of these proteases. Furthermore, PRISM is performed under fully denaturing conditions, ensuring inactivation of all endogenous proteases, and allowing complete blocking

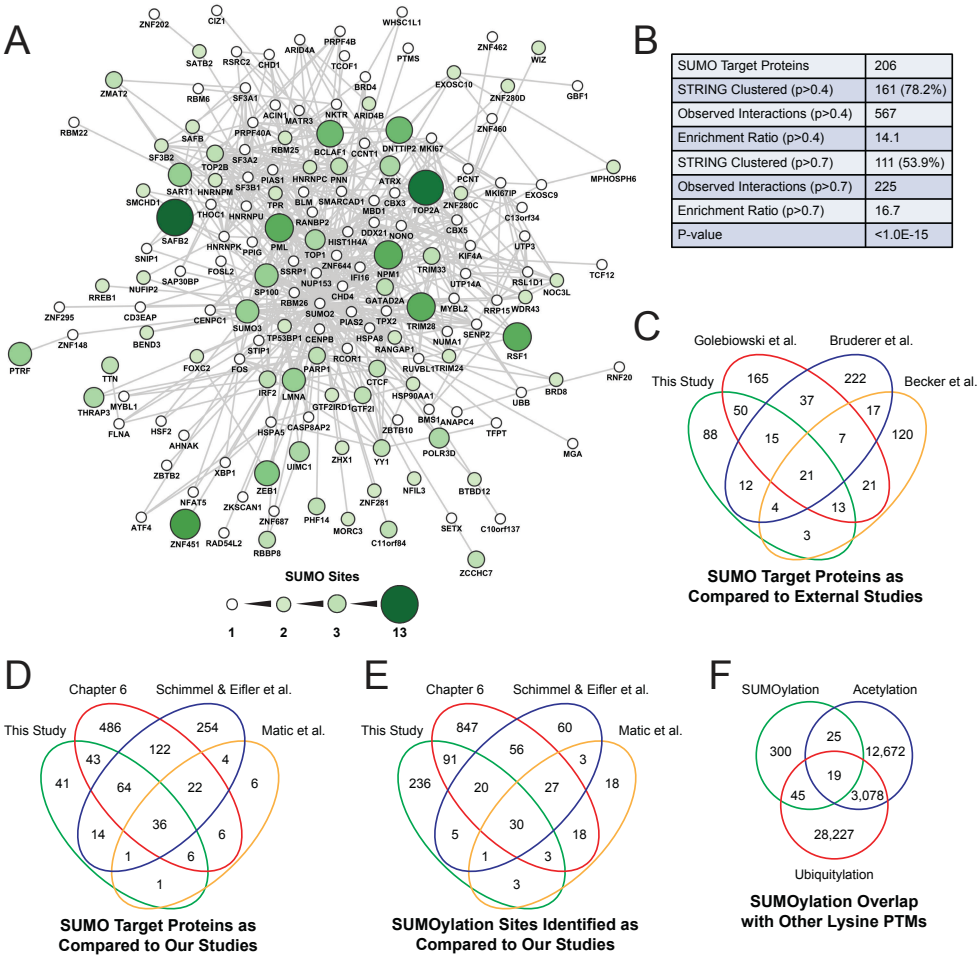


Figure 7. A comparison of PRISM-identified SUMOylation sites and target proteins to other SUMOylation studies.

A) STRING network analysis of all PRISM-identified SUMO target proteins. 161 out of 206 proteins formed one core cluster at a medium or higher STRING confidence (p>0.4). The size and color of the proteins is indicative of the amount of identified SUMOylation sites.

B) Statistical data supporting the STRING network analysis. STRING clustering was performed at medium (p>0.4) and high (p>0.7) confidences, separately. Enrichment ratio is the amount of observed interactions divided by the expected interactions, as provided by the online STRING database.

C) Four-way Venn diagram comparing PRISM-identified SUMOylated proteins to SUMO target proteins identified in three external studies.

D) Four-way Venn diagram comparing PRISM-identified SUMOylated proteins to SUMO target proteins identified in three studies from our lab.

E) Four-way Venn diagram comparing PRISM-identified SUMOylated lysines to SUMOylation sites identified in three studies from our lab.

F) Three-way Venn diagram comparing PRISM-identified SUMOylated lysines with all known ubiquitylation and acetylation sites.

of lysines in endogenous proteins. Additionally, for identification of sites by mass spectrometry, PRISM does not utilize biotinylation and subsequent purification. While this could be successful in reducing sample complexity, it does not address any potential background false positive hits resulting from incomplete acetylation of lysines. In order to address this issue, a negative dataset would have to be generated, where the protease step is skipped. Finally, leaving the lysine free after deSUMOylation allows for identification of two reporter peptides, due to trypsin being able to cleave the peptide, with both reporter peptides being shorter and thus easier to resolve. This is especially pivotal in the lysine-blocking context, already resulting in peptides that are on average twice as long as from a non-blocked tryptic digest. Interestingly, because SUMOylation sites are often situated in regions enriched for lysines, PRISM allows for identification of SUMOylated peptides that are lysine-rich up to the point where they would normally be unidentifiable due to being too short.

Conclusively, PRISM could be adapted to be applicable at the completely endogenous level. To this end, a robust and highly efficient SUMO-specific antibody is likely required for pre-enrichment of SUMOylated proteins. In order to achieve total lysine blocking and inactivation of endogenous proteases, the samples would have to be prepared under highly denaturing conditions, and all lysines completely blocked prior to purification. Subsequently, the buffer conditions would have to be mildened to allow for immunoprecipitation of SUMOylated proteins, prior to treatment with SUMO-specific protease. Ultimately, this will allow for system-wide identification of endogenous protein lysines modified by endogenous SUMO in highly complex *in vitro* and *in vivo* samples.

EXPERIMENTAL PROCEDURES

Plasmids

The His10-SUMO-2 we described and used in this manuscript has the following amino acid sequence:
MAHHHHHHHHHHGGSMSEEKPKEGVKTENDHINLKVAGQDGSVVQFKIKRHTPLSKLMKAYCERQGLSMRQIRFRFDGQPINETDTPAQLEMEDEDTIDVFQQQTGG.

The His10-SUMO-2-K0-Q87R we described and used in this manuscript has the following amino acid sequence:
MAHHHHHHHHHHGGSMSEERPREGVRTENDHINLRVAGQDGSVVQFRIRRHTPLSRLMRAYCERQGLSMRQIRFRFDGQPINETDTPAQLEMEDEDTIDVFRQQTGG.

The corresponding nucleotide sequences were cloned in between the PstI and XhoI sites of the plasmid pLV-CMV-IRES-GFP [50].

Cell culture and cell line generation

HeLa cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% Fetal Bovine Serum (FBS) and 100 U/mL penicillin and streptomycin (P/S, Invitrogen). HeLa cell lines stably expressing wild-type His10-SUMO-2 and lysine-deficient His10-SUMO-2 (AllKR-Q87R mutant) were generated. To this end, HeLa cells were infected using a lentivirus encoding CMV-His10-SUMO-2-IRES-GFP at a multiplicity of infection of 2, supplemented by 8 µg/mL of polybrene. One day after infection, the medium was changed to standard DMEM with 10% FBS and P/S. Cells were split and maintained as required. Two weeks after infection, cells were harvested and resuspended at a concentration of 12 million cells per mL. Cells were collected in DMEM lacking phenol red, supplemented

by 2% FBS and P/S. Cells were passed through a 50 µm filter (Partec) to separate single cells from clumped cells and other particulate matter. Following filtration, cells were sorted according to their GFP fluorescence using a FACSAria II (BD Biosciences). Cells were passed through a 100 µm capillary at a pressure of 138 kPa, at a rate of ~5.000 cells per second, and selected for $7.5*10^2$-$3*10^3$ GFP BF 530/30-A intensity, which was 2.5-10 times higher than the background cellular auto-fluorescence of $3*10^2$.

SILAC labeling of HeLa cells stably expressing His10-SUMO-2 for proteomic analysis

For proteomics, HeLa cells stably expressing His10-SUMO-2 were seeded in 4x 15-cm dishes containing either medium SILAC DMEM ([$^2H_4$,$^{12}C_6$,$^{14}N_2$]lysine/[$^{13}C_6$,$^{14}N_4$]arginine) or 4x 15-cm dishes containing heavy SILAC DMEM ([$^{13}C_6$,$^{15}N_2$]lysine/[$^{13}C_6$,$^{15}N_4$]arginine), at a confluence of 25%. After 3 days of growth, all cells were trypsinized, washed twice with PBS, and split to 10x 15-cm dishes for both medium and heavy SILAC DMEM. Following an additional 4 days of growth, cells were harvested and the pellets were vigorously lysed in 25 pellet volumes of 6 M guanidine-HCl, 100 mM sodium phosphate, 10 mM TRIS, buffered at pH 8.0. Immediately following lysis, the lysates were snap frozen and stored at -80 °C until further processing.

Blocking of lysines with Sulfosuccinimidyl Acetate (SNHSA) and removal of SUMO using SUMO proteases in HeLa total lysate

HeLa lysate was generated by scraping HeLa cells off 15-cm dishes after washing twice with ice-cold PBS. Subsequently, cells were pelleted by centrifugation and washed once more with PBS. Cells were vigorously lysed in 8 M urea, 100 mM sodium phosphate, buffered at pH 8.5, freshly supplemented by 10 mM chloroacetamide. Dilution to lower concentration of urea was performed using 100 mM sodium phosphate, buffered at pH 8.5, freshly supplemented by 10 mM chloroacetamide. Blocking of all lysines was performed in 8 M urea buffer, by addition of SNHSA to a concentration of 20 mM. Blocking was performed for 30 minutes at room temperature. Following blocking, remaining SNHSA was quenched by addition of TRIS pH 8.0 to a concentration of 50 mM, and incubated for 30 minutes at room temperature. Removal of SUMO was performed using commercially acquired recombinant catalytic domains of SENP1 (LifeSensors) and SENP2 (LifeSensors). Digestion analysis of acetylated proteins was performed using Sequencing Grade Lys-C (Promega) and Sequencing Grade Trypsin (Promega). Lys-C digestion was performed with 1 µg of enzyme per sample in 8 M urea. Trypsin digestion was performed with 1 µg of enzyme after diluting the urea concentration to 2 M. Digestions were performed overnight at room temperature.

Preparation of His10-SUMO-2 cell lysate

HeLa cells stably expressing wild-type or lysine-deficient His10-SUMO-2 were seeded in 15-cm dishes. After 3 days of growth, cells were harvested and the pellets were vigorously lysed in 25 pellet volumes of 6 M guanidine-HCl, 100 mM sodium phosphate, 10 mM TRIS, buffered at pH 8.0. Immediately following lysis, lysates were snap frozen and stored at -80 °C until further processing. Lysates were thawed at 25 °C and sonicated at power 7.0 (~30 Watt) for 2x 5 seconds per 15 mL of lysate, using a microtip sonicator, and mixing the samples in between sonication steps. Following sonication, lysates were supplemented with β-mercaptoethanol to a concentration of 5 mM, and imidazole (pH 8.0) to a concentration of 50 mM. Subsequently, lysates were tumbled at room temperature for 30 minutes.

Enrichment of His10-SUMO-2

20 µL Ni-NTA beads (Qiagen) were prepared per 1 mL of lysate, and subsequently washed (4x) and equilibrated in 6 M guanidine-HCl, 100 mM sodium phosphate, 10 mM TRIS, buffered at pH 8.0, supplemented by 5 mM β-mercaptoethanol and 50 mM imidazole pH 8.0. The equilibrated Ni-NTA beads were added to the lysate, and tumbled for 5 hours at room temperature. Following incubation of the lysate with Ni-NTA beads, the beads were pelleted by centrifugation at 500 RCF. Next, the beads were washed (resuspended in a buffer, gently mixed for 10 seconds, and then re-pelleted at 500 RCF) in order with the following buffers. Wash Buffer 1: 6 M guanidine-HCl, 100 mM sodium phosphate,

10 mM TRIS, 10 mM imidazole, 5 mM β-mercaptoethanol, 0.2% Triton X-100, at pH 8.0. Washing with Wash Buffer 1 was repeated once. Wash Buffer 2: 8 M urea, 100 mM sodium phosphate, 5 mM β-mercaptoethanol, 0.2% Triton X-100, at pH 8.0. Washing with Wash Buffer 2 was repeated three times.

On-beads lysine acetylation of His10-SUMO-2-conjugated proteins

Following enrichment of proteins modified by His10-SUMO-2, all proteins were acetylated on the beads using Sulfosuccinimidyl Acetate (SNHSA). To this end, after the final wash the beads were resuspended in one bead volume of Acetylation Buffer (AB) supplemented by 20 mM SNHSA. AB is comprised of 8 M urea, 200 mM sodium phosphate, 5 mM β-mercaptoethanol, 0.2% Triton X-100, 50 µg/mL phenol red, buffered at pH 8.0. The SNHSA was weighed and kept in separate tubes, and only added to the AB at the final moment, upon which it was mixed and swiftly added to the Ni-NTA beads. After addition of SNHSA, the beads were rapidly resuspended and tumbled for 10 minutes, upon which 0.5 µL of 6 M NaOH per 100 µL beads was added to raise the pH back up to 8. After an additional 10 minutes of tumbling, a second acetylation round was performed by adding another bead volume of AB supplemented with 20 mM fresh SNHSA to the mixture. Following 10 minutes of tumbling, 1.5 µL 6 M NaOH per 100 µL beads was added to raise the pH over 8 and finalize the reaction. Subsequently, the beads were tumbled for an additional 10 minutes. Following completion of protein acetylation, TRIS pH 8.0 was added to a final concentration of 20 mM to quench any remaining SNHSA. Finally, the beads were pelleted at 500G and subject to the following washes in order. Wash Buffer 3: 8 M urea, 100 mM sodium phosphate, 5 mM β-mercaptoethanol, 0.2% Triton X-100, buffered at pH 8.0. Note: For proteomic samples, Triton X-100 was left out. Washing with Wash Buffer 3 was repeated once. Wash Buffer 4: 8 M urea, 100 mM sodium phosphate, 0.1% Triton X-100, buffered at pH 6.3. Note: For proteomic samples, Triton X-100 was left out. Washing with Wash Buffer 4 was repeated once.

Acid elution of acetylated His10-SUMO-2-conjugated proteins, for biotinylation

Following the final wash step, the proteins were eluted off the Ni-NTA beads using one buffer volume of 8 M urea, 56 mM citric acid, 44 mM sodium citrate, self-buffered at pH 4.4 (Elution Buffer). After addition of the Elution Buffer, the beads were tumbled for 15 minutes to ensure proper suspension of the beads and elution of the proteins. Next, the beads were pelleted at 500G and the eluted proteins were transferred to a separate tube. The beads were resuspended in one additional buffer volume of Elution Buffer, for an additional 15 minutes, to yield a second fraction of eluted proteins. Both fractions of eluted proteins were pooled and passed through a twice-washed 0.45 µM filter column (MilliPore) to separate any remaining beads from the proteins. Following clearance of the eluted proteins, buffer conditions were neutralized to pH 8.0 by addition of 1/40th elution volume of 6 M sodium hydroxide.

Imidazole elution of acetylated His10-SUMO-2-conjugated proteins, for proteomics

Following the final wash step, the proteins were eluted off the Ni-NTA beads using one buffer volume of 8 M urea, 100 mM sodium phosphate, 10 mM TRIS and 500 mM imidazole, buffered at pH 7.0. After addition of the elution buffer, the beads were tumbled for 30 minutes to ensure proper suspension of the beads and elution of the proteins. Next, the beads were pelleted at 500 RCF and the eluted proteins were transferred to a separate tube. The beads were then resuspended in one additional buffer volume of elution buffer, and incubated while tumbling for an additional 30 minutes, to yield a second fraction of eluted proteins. Both fractions of eluted proteins were pooled and passed through a twice-washed 0.45 µM filter column (MilliPore) to separate any remaining beads from the proteins.

Concentration of acetylated His10-SUMO-2-conjugated proteins, for proteomics

Purified acetylated His10-SUMO-2 conjugated proteins were concentrated on a twice-washed 100 kDa cut-off filter (Vivacon 500, Sartorius Stedim, pre-washed with Elution Buffer). To facilitate concentration, only 400 µL of sample was loaded on a filter per cycle, and then concentrated at 8.000 RCF for 10 minutes at a controlled temperature of 20 °C. After each cycle, another 400 µL of sample was loaded on top of the concentrate and concentrated in a similar fashion. Samples were concentrated to a volume equal to approximately 1/40th to 1/100th of the starting volume. The final concentrated

proteins were removed from the filter by placing it upside-down into a clean 1.5 mL microcentrifuge tube and centrifuging it in an open table-top centrifuge at 500G for 30 seconds.

Specific removal of His10-SUMO-2 from proteins by SENP2

Sample were slowly diluted to 3 M urea by addition of 4 volumes of 1.75 M urea, 100 mM sodium phosphate, buffered at pH 8.5, while gently mixing the samples during the dilution process to prevent abrupt changes in buffer condition. After dilution, DTT was added to a final concentration of 2 mM. Subsequently, per 15-cm plate of cells used for the sample, 2.5 µg (250 U) of dialyzed (primary-amine depleted) recombinant His10-SENP2 catalytic domain was added to the samples. Samples were gently mixed and left for 24 hours at room temperature, in the dark, and undisturbed. For proteomic analysis only, the concentration of urea was raised back up to 8 M by addition of two volumes of 10.5 M urea. The sample were then concentrated over a pre-washed 100 kDa cut-off filter, as described previously, to approximately 1/25th to 1/50th of the starting volume.

Labeling of SENP2-cleared lysines with Sulfo-NHS-SS-Biotin

After removal of all SUMO-2 from the proteins, Triton X-100 was added to a concentration of 0.5%, and samples were heated to 37 °C for 30 minutes to precipitate His10-SENP2. Following incubation, samples were centrifuged at 10.000 RCF to clear precipitated SENP2. Samples were then transferred to a clean tube and treated with 1 mg (0.83 mM) of Sulfosuccinimidyl-SS-Biotin (SNHSSSB). SNHSSSB was immediately added to the samples after being dissolved in 8 M urea, 100 mM sodium phosphate, buffered at pH 8.5. Following 2 hours of incubation at room temperature, TRIS pH 8.0 was added to a final concentration of 50 mM, and samples were incubated for another 30 minutes at 30 °C to quench any remaining SNHSSSB. One volume of 8 M urea was added to the samples to raise the overall concentration of urea up to 5 M.

Enrichment of biotinylated proteins

200 µL of Neutravidin beads (Thermo) were washed (4x) and equilibrated in 8 M urea, 100 mM sodium phosphate, 10 mM TRIS, 0.2% Triton X-100, buffered at pH 8.5 (Neutravidin Wash Buffer). The equilibrated neutravidin beads were added to the samples and tumbled for 3 hours at room temperature. Following incubation, the beads were pelleted by centrifugation at 500 RCF, and washed 6x with Neutravidin Wash Buffer. Following washing, the Neutravidin beads were eluted for 10 minutes at 30 °C with one bead volume of Neutravidin Wash Buffer supplemented with 100 mM of dithiothreitol (DTT). A secondary elution was performed using one bead volume 1x LDS Sample Buffer (NuPAGE) supplemented with 100 mM DTT, for 15 minutes at 50 °C.

Electrophoresis and immunoblot analysis

Protein samples were size-fractionated on Novex 4-12% Bis-Tris gradient gels using MOPS running buffer (Invitrogen), or on home-made 10% polyacrylamide gels using TRIS-Glycine buffer. Size-separated proteins were transferred to Hybond-C membranes (Amersham Biosciences) using a submarine system (Invitrogen). Gels were Coomassie stained according to manufacturer's instructions (Invitrogen). Membranes were stained for total protein loading using 0.1% Ponceau-S in 5% acetic acid (Sigma). Membranes were blocked using PBS containing 0.1% Tween-20 (PBST) and 5% milk powder for one hour. Subsequently, membranes were incubated with primary antibodies as indicated, in blocking solution. Incubation with primary antibody was performed overnight at 4°C. Subsequently, membranes were washed three times with PBST and briefly blocked again with blocking solution. Next, membranes were incubated with secondary antibodies (donkey-anti-mouse-HRP or rabbit-anti-goat-HRP, Pierce) for one hour, before washing three times with PBST and two times with PBS. Membranes were then treated with ECL2 (Pierce) as per manufacturer's instructions, and chemiluminescence was captured using Biomax XAR film (Kodak).

Primary antibodies

Primary antibodies used in this study were Mouse α SUMO-2/3 (8A2, Abcam), Mouse α SUMO-1

(33-2400, Zymed), Mouse α Ubiquitin (P4D1, sc-8017, Santa Cruz), Rabbit α APC4 (A301-176A, Bethyl), Rabbit α TRIM33 (kind gift from AG Jochemsen).

In-solution digestion and desalting of the peptides

Acetylated deSUMOylated proteins in 8 M urea were supplemented with ammonium bicarbonate (ABC) to 50 mM. Subsequently, dithiothreitol (DTT) was added to a concentration of 1 mM, and samples were left to incubate at room temperature for 30 minutes. Next, chloroacetamide was added to a concentration of 5 mM, and samples were incubated at room temperature for 30 minutes. At this point, samples were gently diluted 4-fold using 50 mM ABC. Subsequently, 1 µg of Sequencing Grade Modified Trypsin (Promega) was added to the samples. Digestion with trypsin was performed overnight, at room temperature, still and in the dark. In-solution digested peptides were cleaned, desalted and concentrated on triple-disc C18 reverse phase StageTips [51], before being eluted twice with 25 µL 80% acetonitrile in 0.1% formic acid. Desalted peptides were vacuum centrifuged at room temperature until 10% of the original volume remained, prior to online nanoflow liquid chromatography-tandem mass spectrometry.

LC-MS/MS analysis

25% of the digest was analyzed per run by means of nanoscale LC-MS/MS using an EASY-nLC system (Proxeon) connected to a Q-Exactive (Thermo) using Higher-Collisional Dissociation (HCD) fragmentation. Samples were eluted off a reversed phase C18 column packed in-house, using a 4 hour gradient ranging from 0.1% formic acid to 80% acetonitrile/0.1% formic acid, at a flow rate of 250 nL per minute. The mass spectrometer was operated in data-dependent acquisition mode using a top 10 method. The resolution of full MS acquisition was 70,000, with an AGC target of 3e6 and a maximum injection time of 20 ms. Scan range was 300 to 1750 m/z. For tandem MS/MS, the resolution was 17,500 with an AGC target of 1e5 and a maximum injection time of 120 ms. An isolation window of 2.2 m/z was used, with a fixed first mass of 100 m/z. Normalized collision energy was set at 25%. An underfill ratio of 0.1% was set, leading to a minimum intensity threshold of 8.3e2. Singly charged objects were rejected, and peptide matching was preferred. A 45-second dynamic exclusion was used.

Data processing

Analysis of the raw data was performed using MaxQuant version 1.4.0.8 [52, 53]. MS/MS spectra were filtered and deisotoped, and the 15 most abundant fragments for each 100 m/z were retained. MS/MS spectra were filtered for a mass tolerance of 6 ppm for precursor masses, and a mass tolerance of 20 ppm was used for fragment ions. Peptide and protein identification was performed through matching the identified MS/MS spectra versus a target/decoy version of the complete human Uniprot database, in addition to a database of commonly observed mass spectrometry contaminants. Up to 5 missed tryptic cleavages were allowed, to compensate for extensive internal acetylation within peptides due to the PRISM methodology. Cysteine carbamidomethylation was set as a fixed peptide modification. Peptide pairs were searched with a multiplicity of 2, allowing medium labeled and heavy labeled SILAC peptides. Medium peptides were set to be labeled with Arginine-6 (monoisotopic mass of 6.020129) and Lysine-4-Acetyl (monoisotopic mass of 46.035672). Heavy peptides were set to be labeled with Arginine-10 (monoisotopic mass of 10.008269) and Lysine-8-Acetyl (monoisotopic mass of 50.024763). Protein N-terminal acetylation and methionine oxidation were set as variable peptide modifications. Moreover, to allow identification of peptides ending in a “free” lysine, a “negative” weight acetyl (monoisotopic mass of -42.010565) was set as a variable peptide C-terminal lysine modification. Up to 5 peptide modifications were allowed. Peptides were accepted with a minimum length of 6 amino acids, a maximum size of 4.6 kDa, and a maximum charge of 7. The processed data was filtered by posterior error probability (PEP) to achieve a protein false discovery rate (FDR) of below 1% and a peptide-spectrum match FDR of below 1%. Peptides ending with a lysine or being preceded by a lysine were assumed to be corresponding to previously SUMOylated lysines. Peptides were additionally filtered to have an Andromeda score of at least 20, and detected as either a medium or a heavy labeled SILAC peptide.

IceLogo and heatmap generation
For SUMOylation site analysis of all identified sites, amino acid sequence windows of 15 amino acids downstream as well as 15 amino acids upstream of the modified lysine were extracted from the corresponding proteins. IceLogo software version 1.2 [54] was used to overlay sequence windows in order to generate a consensus sequence, and compensated against the expected random occurrence frequencies of amino acids across all human proteins (IceLogo). Alternatively, subsets of modification sites were compared directly to other subsets of modification sites, generating consensus sequences showing differential occurrence of amino acids between the subsets (Sub-IceLogo). Heatmaps were generated in a similar fashion to IceLogos. For IceLogos, Sub-IceLogos and heatmaps, all amino acids shown as enriched or depleted are significant with $p < 0.05$.

Term enrichment analysis
Statistical enrichment analysis for protein and gene properties was performed using Perseus software [55]. The human proteome was annotated with Gene Ontology terms [56], including Biological Processes (GOBP), Molecular Functions (GOMF), and Cellular Compartments (GOCC). Additional annotation was performed with Keywords. SUMOylated proteins were compared by annotation terms to the entire human proteome, using Fisher Exact Testing. Benjamini and Hochberg FDR was applied to p-values to correct for multiple hypotheses testing, and final corrected p-values were filtered to be less than 2%. Final scoring of terms was performed by multiplying the 2-Log of the enrichment ratio by the negative 10-Log of the FDR, which allowed ranking of terms by both their enrichment and confidence.

STRING network analysis
STRING network analysis was performed using the online STRING database [47], using all SUMOylated proteins as input. Protein interaction enrichment was performed based on the amount of interactions in the networks, as compared to the randomly expected amount of interactions, with both variables directly derived from the STRING database output. Enrichment analysis was performed allowing network interactions at medium or greater confidence ($p>0.4$), and separately by allowing network interactions at high or greater confidence ($p>0.7$). Visualization of the interaction network was performed using Cytoscape version 3.0.2 [57].

SUMO target protein overlap analysis
For SUMO target protein analysis, all proteins identified in this work with at least one SUMO site were selected. For comparative analysis, identified SUMO target proteins were compared to other studies. SUMO-2 target proteins from Becker et al. [14] were selected by a SUMO-2 / Control ratio of greater than 2, and additionally filtered for a SUMO-2 intensity of greater than 10% as compared to SUMO-1 and control. SUMO-2 target proteins from Golebiowski et al. [18] were selected for a SUMO-2 / Control SILAC ratio of greater than 1.5. SUMO-modified proteins from Bruderer et al. [30] were selected as identified proteins filtered for an observed molecular weight shift of 2 times the molecular weight of SUMO or greater, as compared to the expected protein molecular weight. SUMO-2 target proteins from Matic et al. [45] were considered to be all proteins in which at least one site of SUMOylation was identified. SUMO-2 target proteins from Chapter 6 in this thesis, were considered to be all proteins in which at least one SUMOylation site was identified under standard growth conditions. SUMO-2 target proteins from Schimmel et al. [46], were selected for a SUMO-2 / Control SILAC ratio of greater than 2. Where required, gene IDs were mapped to the corresponding Uniprot IDs. Additionally, where multiple Uniprot IDs were listed for a singular protein identification, a major Uniprot ID was selected by selecting the first Uniprot ID in the list starting with a P, or otherwise the first Uniprot ID starting with a Q, or otherwise the first Uniprot ID in the list. Perseus software was used to generate a complete gene list for all known human proteins, and all identified SUMO target proteins from our study as well as the above-mentioned studies were aligned based on matching Uniprot IDs. Results were visualized using Venn diagrams.

SUMOylation and PTM site overlap analysis
For comparative analysis, all SUMOylation sites identified by Matic et al. [45], sites from Schimmel et al. [46], and sites identified in Chapter 6 under standard growth conditions, were assigned to matching Uniprot IDs and sequence windows were parsed. Furthermore, 31,369 MS/MS-identified ubiquitylation sites and 15,794 acetylation sites were extracted from PhosphoSitePlus (PSP; PhosphoSitePlus®, www.phosphosite.org, [58]), and sequence windows were assigned. For each dataset, duplicate sequence windows were removed. Perseus software was used to generate a matrix where all sequence windows from all PTMs were cross-referenced to each other. Corresponding parental proteins were assigned, and multiple modifications targeting the same lysines were visualized using Venn diagrams.

ACKNOWLEDGMENTS

This work is supported by the Netherlands Organization for Scientific Research (NWO) (A.C.O.V.), the European Research Council (A.C.O.V.). and the Max Planck Society (M.M). The authors declare no conflict of interest. We would like to thank Jürgen Cox for including functionality into MaxQuant which greatly aided our ability to filter peptides-of-interest.

Reference List

1. Vertegaal, A. C. (2011) Uncovering ubiquitin and ubiquitin-like signaling networks. *Chem. Rev.* 111, 7923-7940
2. Hay, R. T. (2005) SUMO: a history of modification. *Mol. Cell* 18, 1-12
3. Geiss-Friedlander, R., and Melchior, F. (2007) Concepts in sumoylation: a decade on. *Nat. Rev. Mol. Cell Biol.* 8, 947-956
4. Jackson, S. P., and Durocher, D. (2013) Regulation of DNA damage responses by ubiquitin and SUMO. *Mol. Cell* 49, 795-807
5. Ulrich, H. D., and Walden, H. (2010) Ubiquitin signalling in DNA replication and repair. *Nat. Rev. Mol. Cell Biol.* 11, 479-489
6. Gill, G. (2005) Something about SUMO inhibits transcription. *Curr. Opin. Genet. Dev.* 15, 536-541
7. Mukhopadhyay, D., and Dasso, M. (2007) Modification in reverse: the SUMO proteases. *Trends Biochem. Sci.* 32, 286-295
8. Flotho, A., and Melchior, F. (2013) Sumoylation: a regulatory protein modification in health and disease. *Annu. Rev. Biochem.* 82, 357-385
9. Rodriguez, M. S., Dargemont, C., and Hay, R. T. (2001) SUMO-1 conjugation in vivo requires both a consensus modification motif and nuclear targeting. *J. Biol. Chem.* 276, 12654-12659
10. Sampson, D. A., Wang, M., and Matunis, M. J. (2001) The small ubiquitin-like modifier-1 (SUMO-1) consensus sequence mediates Ubc9 binding and is essential for SUMO-1 modification. *J. Biol. Chem.* 276, 21664-21669
11. Nacerddine, K. et al. (2005) The SUMO pathway is essential for nuclear integrity and chromosome segregation in mice. *Dev. Cell* 9, 769-779
12. Wang, Y., and Dasso, M. (2009) SUMOylation and deSUMOylation at a glance. *J. Cell Sci.* 122, 4249-4252
13. Saitoh, H., and Hinchey, J. (2000) Functional heterogeneity of small ubiqui-

tin-related protein modifiers SUMO-1 versus SUMO-2/3. *J. Biol. Chem.* 275, 6252-6258
14. Becker, J. et al. (2013) Detecting endogenous SUMO targets in mammalian cells and tissues. *Nat. Struct. Mol. Biol.* 20, 525-531
15. Vertegaal, A. C. et al. (2006) Distinct and overlapping sets of SUMO-1 and SUMO-2 target proteins revealed by quantitative proteomics. *Mol. Cell Proteomics* 5, 2298-2310
16. Vertegaal, A. C. (2010) SUMO chains: polymeric signals. *Biochem. Soc. Trans.* 38, 46-49
17. Tatham, M. H. 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
18. Golebiowski, F. et al. (2009) System-wide changes to SUMO modifications in response to heat shock. *Sci. Signal.* 2, ra24
19. Burgess, R. C., Rahman, S., Lisby, M., Rothstein, R., and Zhao, X. (2007) The Slx5-Slx8 complex affects sumoylation of DNA repair proteins and negatively regulates recombination. *Mol. Cell Biol.* 27, 6153-6162
20. Perry, J. J., Tainer, J. A., and Boddy, M. N. (2008) A SIM-ultaneous role for SUMO and ubiquitin. *Trends Biochem. Sci.* 33, 201-208
21. Sun, H., Leverson, J. D., and Hunter, T. (2007) Conserved function of RNF4 family proteins in eukaryotes: targeting a ubiquitin ligase to SUMOylated proteins. *EMBO J.* 26, 4102-4112
22. Tatham, M. H. et al. (2008) RNF4 is a poly-SUMO-specific E3 ubiquitin ligase required for arsenic-induced PML degradation. *Nat. Cell Biol.* 10, 538-546
23. Mahajan, R., Delphin, C., Guan, T., Gerace, L., and Melchior, F. (1997) A small ubiquitin-related polypeptide involved in targeting RanGAP1 to nuclear pore complex protein RanBP2. *Cell* 88, 97-107
24. Lin, D. Y. et al. (2006) Role of SUMO-interacting motif in Daxx SUMO modification, subnuclear localization, and repression of sumoylated transcription factors. *Mol. Cell* 24, 341-354
25. Mei, D. et al. (2013) Up-regulation of SUMO1 pseudogene 3 (SUMO1P3) in gastric cancer and its clinical association. *Med. Oncol.* 30, 709
26. Kessler, J. D. et al. (2012) A SUMOylation-dependent transcriptional subprogram is required for Myc-driven tumorigenesis. *Science* 335, 348-353
27. Wang, Q. et al. (2013) SUMO-specific protease 1 promotes prostate cancer progression and metastasis. *Oncogene* 32, 2493-2498
28. Bettermann, K., Benesch, M., Weis, S., and Haybaeck, J. (2012) SUMOylation in carcinogenesis. *Cancer Lett.* 316, 113-125
29. Schimmel, J. et al. (2008) The ubiquitin-proteasome system is a key component of the SUMO-2/3 cycle. *Mol. Cell Proteomics* 7, 2107-2122
30. Bruderer, R. et al. (2011) Purification and identification of endogenous polySUMO conjugates. *EMBO Rep.* 12, 142-148
31. Witze, E. S., Old, W. M., Resing, K. A., and Ahn, N. G. (2007) Mapping protein post-translational modifications with mass spectrometry. *Nat. Methods* 4, 798-806
32. Pandey, A., and Mann, M. (2000) Proteomics to study genes and genomes. *Nature* 405, 837-846
33. Mann, M., and Jensen, O. N. (2003) Proteomic analysis of post-translational modifications. *Nat. Biotechnol.* 21, 255-261
34. Olsen, J. V., and Mann, M. (2013) Status of large-scale analysis of post-translational modifications by mass spectrometry. *Mol. Cell Proteomics.*
35. Huttlin, E. L. et al. (2010) A tissue-specific

atlas of mouse protein phosphorylation and expression. *Cell* 143, 1174-1189
36. Olsen, J. V. et al. (2006) Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* 127, 635-648
37. Choudhary, C. et al. (2009) Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* 325, 834-840
38. Guo, A. et al. (2013) Immunoaffinity Enrichment and Mass Spectrometry Analysis of Protein Methylation. *Mol. Cell Proteomics.*
39. Kim, D. Y., Scalf, M., Smith, L. M., and Vierstra, R. D. (2013) Advanced proteomic analyses yield a deep catalog of ubiquitylation targets in Arabidopsis. *Plant Cell* 25, 1523-1540
40. Emanuele, M. J. et al. (2011) Global identification of modular cullin-RING ligase substrates. *Cell* 147, 459-474
41. Povlsen, L. K. et al. (2012) Systems-wide analysis of ubiquitylation dynamics reveals a key role for PAF15 ubiquitylation in DNA-damage bypass. *Nat. Cell Biol.* 14, 1089-1098
42. Wagner, S. A. et al. (2011) A proteome-wide, quantitative survey of in vivo ubiquitylation sites reveals widespread regulatory roles. *Mol. Cell Proteomics* 10, M111
43. Kim, W. et al. (2011) Systematic and quantitative assessment of the ubiquitin-modified proteome. *Mol. Cell* 44, 325-340
44. Galisson, F. et al. (2011) A novel proteomics approach to identify SUMOylated proteins and their modification sites in human cells. *Mol. Cell Proteomics* 10, M110
45. Matic, I. 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
46. Schimmel, J. et al. (2014) Uncovering SUMOylation Dynamics during Cell-Cycle Progression Reveals FoxM1 as a Key Mitotic SUMO Target Protein. *Mol. Cell*
47. Franceschini, A. et al. (2013) STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* 41, D808-D815
48. Yunus, A. A., and Lima, C. D. (2005) Purification and activity assays for Ubc9, the ubiquitin-conjugating enzyme for the small ubiquitin-like modifier SUMO. *Methods Enzymol.* 398, 74-87
49. Andersen, J. L. et al. (2011) A biotin switch-based proteomics approach identifies 14-3-3zeta as a target of Sirt1 in the metabolic regulation of caspase-2. *Mol. Cell* 43, 834-842
50. Vellinga, J. et al. (2006) A system for efficient generation of adenovirus protein IX-producing helper cell lines. *J. Gene Med.* 8, 147-154
51. Rappsilber, J., Mann, M., and Ishihama, Y. (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* 2, 1896-1906
52. Cox, J. et al. (2011) Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome. Res.* 10, 1794-1805
53. Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* 26, 1367-1372
54. Colaert, N., Helsens, K., Martens, L., Vandekerckhove, J., and Gevaert, K. (2009) Improved visualization of protein consensus sequences by iceLogo. *Nat. Methods* 6, 786-787
55. Cox, J., and Mann, M. (2012) 1D and 2D annotation enrichment: a statistical method

integrating quantitative proteomics with complementary high-throughput data. *BMC. Bioinformatics.* 13 Suppl 16, S12

56. Ashburner, M. et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 25, 25-29
57. Shannon, P. et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498-2504
58. Hornbeck, P. V. et al. (2012) PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic Acids Res.* 40, D261-D270