Chapter 8

Discussion

Small Ubiquitin-Like Modifier (SUMO) has been established as a post-translational modification (PTM) of paramount importance to all eukaryotic life [1, 2]. As reviewed in the introduction chapter of this thesis, system-wide proteomics has accelerated understanding of the modified proteome for major post-translational modifications such as ubiquitylation, acetylation, methylation and phosphorylation [3]. However, global and unbiased characterization of SUMOylation has lagged behind, due to unfavorable modification stoichiometry, inefficient purification methods, and a large tryptic remnant on peptides-of-interest.

This work comprises the various strategies we devised in order to circumvent aforementioned obstacles, and a large amount of proteomic data as well as novel functional findings regarding SUMOylation.

SUMOylation Coordinates Transcriptional Repression in Response to DNA damage

SUMO plays a critical role in the deoxyribonucleic acid (DNA) damage response [4, 5]. As reported in Chapter 2, we devised methodology for efficiently purifying SUMOylated proteins by means of a FLAG-tag, and investigated the dynamic nature of SUMOylation in response to the alkylating agent methyl methanesulfonate (MMS). The choice for this compound is based on earlier findings, where cells deficient for SUMOylation were sensitive to MMS [6, 7]. Regardless, as SUMO is involved in virtually all nuclear processes, it would undoubtedly be of interest to investigate SUMOylation dynamics in response to other DNA damage sources such as ionizing radiation, ultraviolet light, hydroxyurea, cisplatin, and Mitomycin-C.

In our mass spectrometry screen, we identified 400 proteins at high confidence through an integrated parental control, and around 800 proteins when accepting proteins at medium confidence. Interestingly, we found roughly 80 proteins to be either upregulated or downregulated in SUMOylation in response to MMS, amounting to 10-20% of all identified SUMOylated proteins. We predominantly found these proteins to be chromatin modifiers, transcription factors, DNA repair factors, and nuclear body components. More strikingly, around half of the identified proteins were clustered in one functional Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) network, giving strength to the theory of SUMO group modification in response to cellular stress [8, 9].

Interestingly, we found many chromatin modifiers to be changed in SUMOylation after DNA damage, and also observed a global transcriptional silencing in response to MMS. The closely related Histone H3 trimethyl Lysine-4 (H3K4me3) demethylases Lysine-Specific Demethylase 5B (JARID1B) and Lysine-Specific Demethylase 5C (JARID1C) were found to be inversely changed in SUMOylation. SUMOylated JARID1B was observed to be rapidly ubiquitylated by RING Finger Protein 4 (RNF4) and degraded, leading to an overall decrease of its SUMOylation. Conversely, JARID1C was found to be increasingly SUMOylated, and moreover JARID1C re-localized from the soluble nucleus to the chromatin in response to MMS. Although we confirmed that SUMOylated JARID1C primarily exists at the chromatin,

the chicken-or-the-egg question remains. Is JARID1C SUMOylated and subsequently transported to the chromatin, or is JARID1C otherwise recruited to the chromatin, and only then SUMOylated because of its proximity to SUMOylation machinery that may target any protein with a SUMO consensus motif ay the chromatin?

Regardless, we demonstrated that DNA damage activates JARID1C, with only the combination of JARID1C overexpression and MMS treatment leading to a rapid global decline of H3K4me3. The function of SUMO on JARID1C is up to speculation, although its observed presence at the chromatin could simply be a matter of timing. We noted that much longer treatment of cells with MMS resulted in the eventual degradation of JARID1C, a process preventable by additional inclusion of proteasome inhibitors (data not shown), hinting that JARID1C may eventually also be degraded through its SUMOylation. It is feasible that JARID1C would have to be cleared from the chromatin once DNA damage has been repaired, to allow the cell to return to normal. SUMOylation could also be directly altering the functional properties of JARID1C, but for such experiments to be performed, a SUMOylation-deficient mutant would have to be generated.

We showed JARID1B to be localized at the chromatin in the absence of DNA damage, and was found to be rapidly and completely degraded in response to MMS treatment. Overexpression of the demethylases results in rapid non-specific removal of H3K4me3. However, at baseline levels, JARID1B has been reported to be quite specific, only removing H3K4me3 from very specific target regions, thereby keeping a set of cell cycle and DNA damage response genes from being activated [10]. Depletion of JARID1B has been reported to lead to cellular senescence [11]. Coincidently, JARID1B is overexpressed in many tumors [12, 13]. Thus, JARID1B may be “guarding” a very specific subset of DNA damage response and apoptosis genes, which are opened up to transcription when JARID1B is rapidly degraded in response to DNA damage. Depletion of JARID1B would in essence force a DNA damage or apoptotic response, whereas overexpression would prevent such responses, and thus granting malignant cells a great survival advantage.

Interestingly, incubation of cells at low oxygen to counter free radical formation leading to continuous damage was found to greatly increase JARID1B levels (data not shown). Under normoxic conditions and without further stress conditions, treatment with the protein synthesis inhibitor cycloheximide resulted in complete degradation of JARID1B in six hours (data not shown). MMS time courses revealed JARID1B to be linearly degraded, a process which could be abruptly halted and stabilized by removing MMS from the cells (data not shown). As such, JARID1B dynamically responds to DNA damage, with increasing damage leading to reduced JARID1B levels. This, in turn, eventually leads to reduced H3K4me3 levels, likely as a secondary effect. Depletion of JARID1B itself by short hairpin ribonucleic acid (shRNA) mediated knockdown resulted in depletion of H3K4me3, and stabilization of JARID1B through knockdown of RNF4 resulted in an increase in H3K4me3 (data not shown). In fact, in most experiments we observed levels of JARID1B and

H3K4me3 to be almost completely synchronized. Interestingly, treatment of cells with a JARID1B/C-specific inhibitor stabilized JARID1B levels in response to cycloheximide and MMS, and similarly stabilized or increased H3K4me3 levels. Thus, the activity of JARID1B might be required for its eventual degradation. All in all, JARID1B is a sophisticated and SUMO-regulated demethylase with a dynamic involvement in the DNA damage response, where it contributes to global transcriptional repression by modulation of the H3K4me3 chromatin mark.

Lastly, in response to MMS, we noted an increase in the transcriptionally repressive chromatin marks Histone H3 dimethyl Lysine-9 and Histone H3 trimethyl Lysine-9 (H3K9me2/3) [14], and found the responsible enzymes Histone-Lysine N-Methyltransferase SET Domain Bifurcated 1 (SETDB1) and Methyl-CpG-Binding Domain Protein 1 (MBD1) to be altered in SUMOylation upon DNA damage [15]. Furthermore, we found CREB-Binding Protein (CBP) and Histone Acetyltransferase p300 (p300) to be increasingly SUMOylated, resulting in decreased histone acetylation and ultimately leading to reduced transcriptional activity [16]. All in all, orchestration of chromatin modifiers by SUMOylation leads to global transcriptional repression. This is a different observation as made in yeast, where SUMOylated proteins involved in homologous recombination were found to be dynamically regulated [8]. Such a difference could in part be explained from the fact that Psakhye et al., used a much higher dose of MMS, and yeast is a completely different organism compared to humans. Furthermore, homologous recombination repair in yeast is far more efficient than in humans, where repair is much slower, and human cells are more inclined to transcriptionally silence themselves in order to gain enough time to repair all damage. Ultimately, we demonstrated the involvement of SUMO in the DNA damage response at a global level, and elucidated a subset of known as well as novel SUMO target proteins that are dynamically regulated by SUMO in response to MMS. Further investigation into these specific targets may uncover new insight into how SUMO contributes to the function of these proteins.

RNF4 and USP11 are Functionally Opposed and Regulate Promyelocytic Leukemia Protein (PML) Nuclear Bodies

The SUMO-targeted ubiquitin (STUbL) E3 ligase RNF4 is the primary human STUbL [17], and is primarily responsible for ubiquitylating SUMOylated proteins and thereby marking them for proteasomal degradation. In Chapter 3, we reported a proteomic screen where we sought to identify direct interaction partners of RNF4, as opposed to the usual SUMO Interaction Motif (SIM) interaction with SUMO by which RNF4 targets are identified. In this screen, we identified the Ubiquitin-Specific Peptidase 11 (USP11) as an interaction partner.

Both RNF4 [18-20] and USP11 [21, 22] have been reported to be functional in the DNA double-strand break repair pathway. The interaction of two proteins with opposed functionality, one being a ubiquitin ligase and the other a ubiquitin protease, especially when both proteins are known to be active in the same biolog-

ical pathway, is striking.

Functionally, we observed an opposed effect of RNF4 and USP11 on PML bodies. Depletion of RNF4 led to an increase in the amount of PML bodies, whereas depletion of USP11 led to a decrease in PML bodies. Double depletion of both proteins simultaneously resulted in a phenotype similar to solely RNF4 depletion, hinting that USP11 functions downstream of RNF4. Treatment of cells with MMS normally results in the rapid dissociation of PML bodies [23]. When RNF4 was depleted in cells, PML bodies were found to be resistant to MMS, and took far longer to dissociate, or even were completely stabilized. Similarly, USP11 overexpression in cells was found to stabilize PML bodies in response to MMS treatment, up to a point where under normal conditions in control cells all PML bodies had disintegrated. Likewise, SUMOylated PML was found to be similarly affected by RNF4 and USP11 at the immunoblot level, with RNF4 depletion leading to an accumulation of SUMOylated PML, and USP11 depletion leading to a decrease in SUMOylated PML.

Since USP11 and RNF4 have opposing functions, the question remains why they would be interacting, or otherwise targeted to the same subcellular location. One potential explanation could be protection from auto-ubiquitylation, as is the case with an interaction between the E3 ubiquitin-protein ligase NRDP1 and Ubiquitin Carboxyl-Terminal Hydrolase 8 (USP8) [24]. The E3 ubiquitin-protein ligase ICP0, which is protected from auto-ubiquitylation through an interaction with Ubiquitin Carboxyl-Terminal Hydrolase 7 (USP7) [25]. However, there is no evidence in the literature indicating that RNF4 is ubiquitylated. Alternatively, the E3 ubiquitin-protein ligase Synoviolin (HRD1) is antagonized by interacting Ubiquitin Carboxyl-Terminal Hydrolase 25 (USP25), which removes ubiquitylation from HRD1 substrates [26]. Similar to this, USP11 may serve as an “editor” for RNF4-mediated ubiquitylation, preventing RNF4 from haphazardly ubiquitylating all SUMOylated proteins, thereby dooming them to degradation. We noted a slight decrease in USP11 levels in response to DNA damage, which could be indicative of a release of this counterbalancing mechanism, in turn allowing RNF4 to ubiquitylate SUMOylated PML. We showed that mutating four putative SIMs in USP11 reduced its ability to process hybrid SUMO-ubiquitin chains, although even the mutant USP11 retained a relatively strong interaction with SUMO-ubiquitin chains. Potentially, the ubiquitin binding properties of USP11 are enough to allow for interacting with hybrid chains, but the SIMs could potentiate the ability for USP11 to efficiently process these chains.

It has been reported that USP11 is able to interact with PML [27]. Whereas an interaction between USP11 and PML could also be a reason why USP11 can counteract any ubiquitylation of PML, it would make more sense for USP11 to prevent SUMOylated PML from ubiquitylation by RNF4, and thereby avoid its subsequent degradation. It would also make sense if USP11 could deubiquitylate other proteins with hybrid SUMO-ubiquitin chains, other than just PML. Furthermore, Wu et al. claim an increase in glioma tumor cell malignancy and an increased resistance to

chemotherapeutics upon USP11 depletion. This stands in stark contrast to findings in the literature, where USP11 depletion resulted in hypersensitivity to chemotherapeutics [21, 22], and where a compound designed to inhibit USP11 could antagonize growth of pancreatic cancer cells [28].

The disassociation of PML bodies allows for a proper progression of the DNA damage response by releasing repair proteins that are stored within these nuclear bodies [29], or induction of apoptosis through co-activation of the Cellular Tumor Antigen p53 (p53) and Cyclin-Dependent Kinase Inhibitor 1 (p21) pathways [30, 31]. Therefore, USP11 could be a viable clinical target, with inhibition of USP11 leading to a collapse of PML bodies, and hypersensitivity to DNA damage. Accordingly, some progress has been made on compounds counteracting USP11, displaying the ability to inhibit pancreatic cancer cells [28].

We thus establish a link between the STUbL RNF4, the deubiquitylase USP11, PML nuclear bodies, and the DNA damage response, all mediated by crosstalk between SUMOylation and ubiquitylation. Combination of inhibitors targeting USP11 [28] with other chemotherapeutics, for example Poly ADP Ribose Polymerase (PARP) inhibitors which tumor cells become susceptible to after depletion of USP11 [21], may prove to be a highly viable way of combating certain types of cancer.

SUMO-2/3 is a Highly Dynamic Modifier of Hundreds of Proteins

Recently, more extensive proteomic studies on SUMOylation have been reported, although these studies remained restricted to the protein level [32-35]. The various approaches used for purifying SUMO target proteins are reviewed in Chapter 1. We performed various proteomic screens with the aim of identifying SUMOylation dynamics at the protein level. In Chapter 2, in response to MMS, we optimized purification of a FLAG-tagged SUMO, identifying 400 high confidence SUMOylated proteins, of which 80 dynamically respond. In Chapter 5, the same FLAG methodology was used in the context of the cell cycle, identifying over 500 high confidence SUMOylated proteins, of which over 100 were found to be dynamically regulated throughout cell cycle progression. Other cellular conditions reported so far include the proteasomal inhibitor MG-132 [32], and heat shock [33]. Interestingly, in all cases, dozens of proteins were found to be dynamically regulated by SUMOylation in response to these stresses or at different points in the cell cycle. This strengthens the theory that SUMO-2/3 is primarily conjugated to proteins whenever the cell is undergoing a form of stress, or otherwise requiring a heightened degree of dynamics such as during cell cycle progression. We have observed that slower cycling cells, for example due to serum starvation, exhibit a far lower rate of SUMO conjugation (data not shown). Potentially, cells which are not dividing at all and are kept under conditions where virtually no stress is incurred could be almost completely devoid of SUMOylated proteins, or alternatively have SUMOylation situated on a different set of proteins. The effect of serum starvation on SUMO has not been studied with system-wide proteomics as of yet, and could be an interesting target for furthering

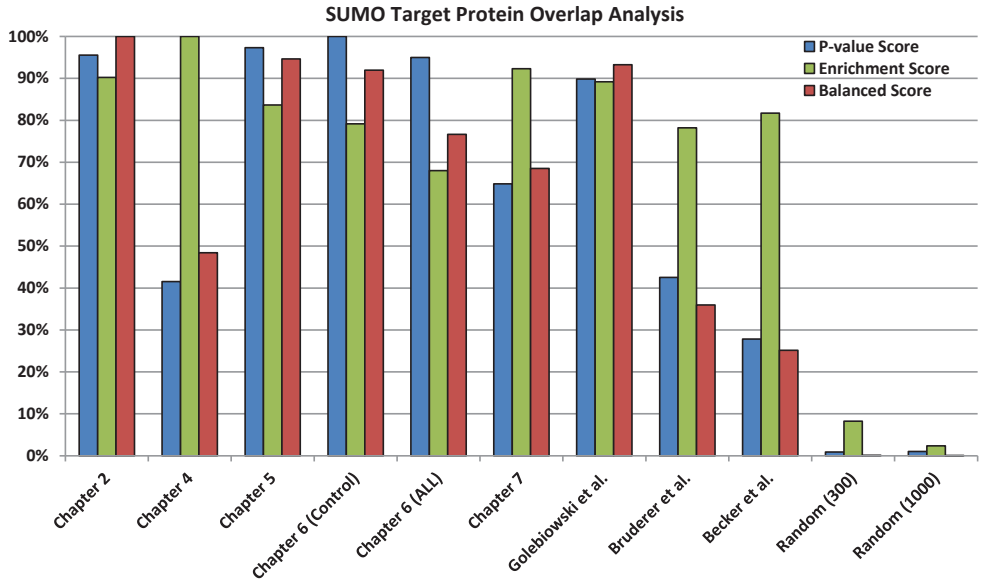

Figure 1. Statistical overlap analysis comparing identified SUMO targets between this thesis and external SUMO studies.
SUMO targets identified at the protein level (Chapters 2 and 5, and all external studies), SUMO targets identified at the site-specific level (Chapters 4, 6 and 7), and two sets containing either 300 or 1000 random proteins were compared against the human proteome and cross-compared to each other for significant overlap by Fisher Exact testing. From Chapter 6, proteins identified exclusively under normal growth conditions and all proteins were considered separately. Overlap between studies was scored based on p-value (biased towards a larger amount of targets identified), scored based on enrichment (biased towards a higher percentage overlap with other studies), and scored with an intermediate model taking both target amount and overlap percentage into account. For each study, scores were cumulated and the highest scoring study was set at 100%. For all separate overlap significance comparisons with non-random sets; $p < 10*E-28$.

insight into the nature of SUMO conjugation.

Taken together, when comparing all SUMOylated proteins identified thus far by our studies and external studies performed at the protein-level, it becomes apparent that there is quite a significant overlap between certain subsets of studies. Statistical enrichment analysis was performed by means of Fisher Exact testing, with the amount of overlap between studies corrected for their individual size and subsequently compared to each other and to the entire human proteome (**Figure 1**). Individual statistical comparison between studies revealed a very strong statistical overlap between SUMOylated proteins identified in our MMS screen (Chapter 2), SUMOylated proteins identified in the cell cycle screen (Chapter 5), and Golebiowski et al [33]. Granted, all these screens were performed in HeLa cells, although Golebiowski et al. employed a different purification procedure.

Interestingly, SUMOylated proteins identified through SIM-traps [35], as well as endogenous SUMO purified from vertebrate cells and complex animal tissues by

means of SUMO antibodies [34], display a far less significant overlap with the other screens (**Figure 1**). Surprisingly, the most significant overlap between Bruderer et al. and Becker et al. was found as a result of many corresponding background binders. Indeed, reducing the Stable Isotope Labeling of Amino Acids in Culture (SILAC) $\log_2$ ratio thresholds in the MMS or cell cycle screens from 2 to 1.5 or even 1.25 resulted in a statistical significance increase when cross-comparing to Bruderer et al. and Becker et al., hinting that these datasets likely contain a very high amount of false positives. These background binders probably originate directly from interaction with the agarose matrices typically used in purification procedures. The FLAG-tag or protA-TEV-CBP-tags employed in Chapter 2, Chapter 5 and by Golebiowski et al. allowed for a more stringent purification procedure, due to more specialized purification tools and antibodies.

Our MMS screen and cell cycle screens both provide statistically high quality datasets of proteins SUMOylated under standard cellular growth conditions, from their respective integrated parental SILAC controls. Around 500 proteins are confidently identified to be SUMOylated and further improvements in purification protocols and optimized mass spectrometry technology could increase this number up to a 1000. As SUMO is an extremely dynamic modification, this number will indubitably expand further when different cellular stresses, or different cell lines, are considered in the equation. Considering that SUMO is an almost exclusively nuclear modification, the amount of observed proteins to be modified is quite striking, and puts SUMO in the same league as modifications such as ubiquitylation, acetylation and phosphorylation.

Both the MMS and cell cycle screens identify a STRING cluster interactivity enrichment for proteins identified to be SUMOylated, further providing evidence that SUMOylation targets entire interacting groups of proteins. As discussed previously, in response to MMS, SUMOylation was found to target a cluster of predominantly chromatin modifiers and transcription factors. In the cell cycle screen, SUMO targeted clusters of transcription factors with critical functions during cell cycle progression, and displayed cell-cycle-dependent upregulation and downregulation of the SUMOylation of these functionally clustered proteins.

A Strategy to Efficiently Identify SUMOylation Sites

While studying SUMOylation at the target protein level by mass spectrometry is feasible, identification of exact SUMO acceptor lysines using system-wide proteomic approaches has proven difficult, due to technical hurdles as reviewed in Chapter 1.

Virtually all highly successful mass spectrometric approaches that have identified many thousands of ubiquitylation, acetylation, methylation and phosphorylation sites use peptide-specific enrichment. In these approaches, the sample of interest is completely digested by trypsin, and the peptides-of-interest are purified out of the mixture through affinity matrices or antibodies that specifically recognize the modification or modification remnant. Ubiquitylated peptides are enriched

using an antibody directed against a diglycine remnant attached to a lysine residue [36], acetylated peptides are enriched with an antibody directed against an acetylated lysine residue [37], and phosphorylated peptides are enriched using titanium oxide metal affinity chromatography [38]. With SUMO, such peptide enrichment is almost futile, as tryptic digestion of SUMO yields a remnant so large that it cannot be properly analyzed using conventional mass spectrometry approaches.

To circumvent this problem, we generated a His-tagged SUMO-2 with all lysines substituted to arginines. Additionally, these mutant SUMO-2 proteins bear either the Q87R or T90R mutations, which are homologous to arginines situated at the carboxyl-terminus of the sole yeast SUMO (Smt3), and ubiquitin, respectively. These mutations have been reported to not interfere with SUMO conjugation behavior [39, 40], although we noted a slightly less efficient conjugation with the T90R mutant. As reported in Chapter 4, this methodology allowed for the identification of 103 SUMO-2 acceptor lysines in endogenous proteins. The mutant SUMO-2 was transiently overexpressed, and the entire total lysate was digested with Lysyl Endopeptidase (Lys-C) prior to enrichment of the SUMOylated peptides by means of their His-tag. Regardless, our peptide-specific purification resulted in an identification rate which outclassed many other prior published works, where no more than roughly a dozen sites were identified [33, 40-43].

The efficacy of the modified peptide purification becomes apparent when looking at the cell cycle screen (Chapter 5), where outside of hundreds of dynamically regulated SUMOylated proteins, we identified only 202 SUMOylation sites. Here, a FLAG-tagged SUMO was utilized, bearing the Q87R mutation but otherwise remaining wild-type. As such, no Lys-C digestion and site enrichment could be performed. While the number of identified sites has essentially doubled when comparing Chapter 4 to Chapter 5, it should be noted that four years of time passed, and mass spectrometry technology has improved in sensitivity and resolution. Moreover, the cell cycle screen consisted of a far larger amount of samples and technical repeats when compared to the singular control-condition sample from Chapter 4. As all other major PTMs have been successfully characterized using peptide-of-interest enrichment, it goes without saying that for the purpose of identifying SUMOylation sites, the lysine-deficient SUMO remains remarkably effective.

The First System-Wide and Site-Specific Overview of the SUMOylated Proteome

We further optimized the lysine-deficient SUMO purification protocol, through a few notable changes in the purification strategy, as reported in Chapter 6. First of all, we generated a stable cell line with a low expression level of His_{10}-tagged lysine-deficient Q87R SUMO-2. Compared to transient overexpression, where extremely high levels of SUMO present throughout the cell may result in less specific conjugation and result in false positive identification of SUMO sites, a stable cell line with a relatively low level can be expected to behave more like the parental cell line. Regardless, the level of exogenous SUMO-2 still surpasses the endogenous

SUMO-2 by a discernable amount to allow for displacement of endogenous SUMO from target lysines. Secondly, an initial purification for His-tagged SUMO-2 was performed before digestion with Lys-C. As the sample volume could be reduced by more than a thousand-fold in this way, it greatly reduced the amount of enzyme required, and allowed for a more complete digestion of all SUMOylated proteins. The SUMOylated peptides were then re-purified out of the digest, which in combination with the more efficient purification of the extended polyhistidine tag greatly reduced background interference. Thirdly, during concentration of the initially purified SUMOylated into a small volume, unconjugated SUMO could be separated from SUMOylated proteins. In turn, this decreases the "contamination" of the mass spectrometry samples resulting from the presence of large stoichiometric amounts of tryptic SUMO peptides.

We chose to use the Q87R mutation, yielding the glutamine-glutamine-threonine-glycine-glycine (QQTGG) remnant, over the T90R mutation resulting in a diglycine remnant. Our work as reported in Chapter 4, as well as preliminary work performed for Chapter 6 (data not shown), demonstrated that the identified SUMOylation sites are virtually identical between the Q87R and T90R mutants, despite the worry that endogenous ubiquitin could interfere with the T90R diglycine detection. Regardless, a T90R SUMO mutant cannot be used for studying crosstalk between ubiquitin and SUMOylation modifying the same peptide or protein. Secondly, the QQTGG remnant generates unique reporter ions upon high-energy collision fragmentation, which serve to greatly increase the confidence of identification. Thirdly, the QQTGG possesses a unique quality, as after tryptic digest the *de novo* N-terminal glutamine may spontaneously cyclize [44], generating pyroglutamate-QTGG. This "Pyro" remnant has a unique mass signature and charge as compared to QQTGG. In turn, this essentially doubles the chances of detection of the modified peptide due to altered migration behavior in the preceding reversed-phase liquid chromatography, by providing a second opportunity for the modified peptide to be detected from a complex sample. We successfully applied identification of both QQTGG and Pyro in the cell cycle screen (Chapter 5) and the optimized SUMO sites identification screen (Chapter 6), increasing the number of SUMOylation sites identified by up to 50% in individual technical replicates.

All in all, the aforementioned improvements led to the identification of 3,246 SUMOylation sites in endogenous proteins, an increase of 32 times over Chapter 4 and an increase of 16 times over Chapter 5. Over 1,100 sites were detected from cells growing under control conditions, and this number was further expanded by additionally treating cells with the proteasome inhibitor MG-132, the ubiquitin and SUMO protease inhibitor PR-619, and heat shock. The identified SUMO sites mapped to 1,346 proteins in total, which amounts to over 10% of all proteins known to be present in HeLa cells. As such, SUMOylation confidently joins the list of the major post-translational modifications. Further study of SUMOylation sites in response to different stimuli, and moreover in different cell types, will undoubtedly

greatly expand upon the number of SUMOylation sites.

SUMO Modifies Clusters of Functionally Related Nuclear Proteins

STRING cluster analysis of all SUMOylated proteins which were filtered for the presence of at least one SUMO site (Chapter 6), provides strong evidence for SUMO group modification. Nearly 60% of all SUMOylated proteins are connected into one core STRING cluster, at high statistical confidence. Furthermore, almost a dozen highly interconnected functional clusters of proteins were discovered within this core cluster, demonstrating SUMO's ability to modify entire functional complexes of proteins. These complexes include the spliceosome, ribosomes, chromatin remodeling complexes, methyltransferase complexes, and deacetylases complexes. Although evidence of SUMO group modification has been provided before [8, 9], this phenomenon has never before been observed at the system-wide level, and not at a scale rivaling modifications such as ubiquitylation and acetylation.

A full term enrichment analysis using protein annotation databases such as Gene Ontology [45] and Comprehensive Resource of Mammalian Protein Complexes (CORUM) [46] generated an exciting summary of processes related to SUMO. The CORUM annotation analysis further reinforced the SUMO group modification observation. Well over a dozen of annotated protein complexes were found to be SUMOylated on the majority of all subunits. For instance, 14 out of 16 subunits of the Histone Deacetylase 2 (HDAC2) complex were SUMOylated, and 8 out of 9 subunits of the Histone Deacetylase 1 (HDAC1) complex. 10 out of 11 subunits of the Lysine-Specific Histone Demethylase 1A (BHC110) complex were SUMOylated, and all 7 out of 7 subunits of the SWI/SNF-Related Matrix-Associated Actin-Dependent Regulator of Chromatin Subfamily E Member 1-Related (BRAF35) and Breast Cancer Type 2 Susceptibility Protein (BRCA2) complex.

The Gene Ontology analysis provides a quick overview of the functions of SUMOylated proteins. At the Biological Processes level, SUMO was among others enriched on proteins involved in nuclear acid metabolic processes, transcription factors, protein-DNA complex organization, nucleosome organization, nuclear messenger RNA splicing, and negative regulation of histone H3 Lysine-4 methylation. At the Molecular Functions level, SUMO was primarily found to be enriched on proteins capable of DNA binding, SUMO ligation, pre-mRNA binding, methyl-CpG binding, zinc ion binding, damaged DNA binding, and demethylase activity. The top 10 hits from a general keywords annotation analysis were; "DNA Binding", "Transcription Regulation", "Isopeptide Bond", "Nucleus", "Ribosome Biogenesis", "Zinc Finger", "Chromatin Regulator", "Repressor", "Citrullination", and "mRNA Splicing". Strikingly, a Kyoto Encyclopedia of Genes and Genomes (KEGG) [47] annotation analysis revealed a modest enrichment in SUMOylation on proteins known to be involved in pathways in cancer, with 37 out of 312 known cancer-involved proteins being SUMOylated. Even more surprisingly, from the 35 of these 37 SUMOylated oncoproteins that could be interpreted by the STRING database, 29 (83%) were

found in one single interaction cluster, further expanded to 34 (97%) by addition of a second interconnected cluster.

Even with so many proteins identified, and a good portion of them predominantly SUMOylated in response to stress conditions, the large majority of SUMO target proteins were found to be localized in the nucleus. Further locational enrichment was observed towards the chromatin, nuclear bodies, and the nucleoli. Although SUMO modification of nuclear proteins comes as no surprise, the ability for SUMO to retain its functionality restricted within the nucleus under stress conditions is notable. Whereas all other major post-translational modifications somewhat ambiguously target proteins all throughout the cell, SUMO almost exclusively modifies multiple important nuclear protein complexes and is involved in crucial nuclear processes such as transcription, RNA splicing, chromosome organization and coordination of the DNA damage response, thus establishing SUMO as a master regulator of virtually all nuclear processes.

Analysis of SUMOylation Sites Reveals Novel Consensus Motifs

For functional studies to be performed on SUMO, mutants have to be generated which are deficient for SUMOylation. Whereas SUMOylation can to some extent be predicted through its conjugation consensus motif, many SUMOylation sites do not adhere to the canonical KxE consensus. Furthermore, trial-and-error site-directed mutagenesis is a highly laborious process, and is unable to distinguish between an actual effective mutation, and a mutation that may alter the localization or structure of the protein, which through secondary effects would lead to the loss of its modification [48]. Thus, the wealth of SUMO sites identified in this work provides an invaluable resource to the SUMO community. Moreover, by studying the amino acid context surrounding all the identified SUMO sites, great insight can be gained in the specificity of SUMO conjugation, as well as about the protein structural context typically surrounding sites of SUMOylation.

Adherence to the core KxE SUMOylation consensus motif was noted to be dependent on stoichiometric abundance, with the more abundant SUMO sites ranging up to 70% adherence to the KxE consensus under standard growth conditions. In the lower percentiles of identification, this abundance dropped down to 50% KxE. Mechanistically, this could be a direct result of Ubiquitin Carrier Protein 9 (Ubc9) preference towards the KxE motif [49, 50], with these sites being more efficiently targeted for SUMOylation. Interestingly, the canonical aspartic acid at +2 of the lysine was not significantly enriched, although it was also not depleted. The efficiency of the glutamic acid was demonstrated in Chapter 6 by mutating the glutamic acid in the Ran GTPase-Activating Protein 1 (RanGAP1) SUMO site motif to an aspartic acid, resulting in reduced SUMOylation efficiency.

Adherence to the SUMOylation consensus motif was observed to be considerably lower in sites exclusively identified in response to heat shock. In case of the protease and proteasome inhibitors, PR-619 and MG-132, a huge drop in faithfulness to

the SUMO consensus motif was observed. In fact, for MG-132 exclusive sites, KxE frequency dropped as low as 10%, only 2-fold over what is observed for all ubiquitin sites. Furthermore, sites identified in response to cellular treatments were found to be more frequently flanked by amino acids normally buried in the hydrophobic core of proteins, such as tyrosines, tryptophans, and phenylalanines, indicative of SUMOylation of misfolded proteins. Interestingly, this provides evidence that SUMO may directly function as a "backup ubiquitin", targeting lysines promiscuously under the opportune conditions. Which SUMO ligases would be involved in such a mechanism, and why proteins SUMOylated at non-consensus sites would be more likely to be degraded, remains unknown.

SUMOylation occurs in regions enriched for glutamic acids and lysines, indicative of solvent-exposed surfaces of proteins. Secondary structure analysis revealed SUMOylated lysines to be solvent-exposed more often than other random lysines contained within SUMOylated proteins, and SUMOylated lysines were found to reside in β-sheet regions more often than random lysines. Furthermore, cysteines were found to be depleted in the immediate vicinity of SUMOylation sites, indicative of the fact that cysteines could interfere with the efficient transfer of SUMO onto the intended lysine. Coincidentally, a similar depletion for cysteines surrounding the lysine was noted for ubiquitylation, but not for methylation and acetylation (Chapter 6).

Most strikingly, as initially reported in Chapter 4, is the discovery of the Inverted SUMO Consensus Motif (ISCM); [ED]xK (**Figure 2**). From the initially identified 103 sites, 24 sites matched this motif. In Chapter 5, from the 202 identified SUMO sites, 42 match the inverted motif. In Chapter 6, from the 1,107 SUMO sites identified under control conditions, 258 match the inverted motif. Thus, the frequency of this occurrence is around 20%, a 2-fold increase over randomly expected. When comparing sites with an acidic residue at -2 to all other sites, depletion in the glutamic acid at +2 is observed, further strengthening the theory that the ISCM is indeed valid (Chapter 6). Mutation of an inverted SUMOylation site in ETS-Related Protein Tel1 (ETV6) also resulted in reduced SUMOylation of the protein. Currently, there is no mechanistic explanation as to how the ISCM would function, and whether there would be any specific E3 ligases responsible for conjugation of SUMO to lysines harbored in such a motif. Interestingly, we noted an equal occurrence of glutamic acid and aspartic acid at -2 in the ISCM, whereas the regular consensus motif contained almost exclusively glutamic acid at +2. This is indicative of a different mechanism involved in SUMO conjugation on these motifs. Potentially, Ubc9 is less efficient in recognition of the forward KxD motif as compared to the KxE motif. This problem is seemingly less prevalent in case of the inverted DxK motif, possibly due to conformational differences in having the aspartic acid placed upstream instead of downstream. Such a difference would have to be elucidated using a structural approach, or at least using binding assays employing peptides bearing the forward or reverse SUMOylation motifs containing either glutamic or

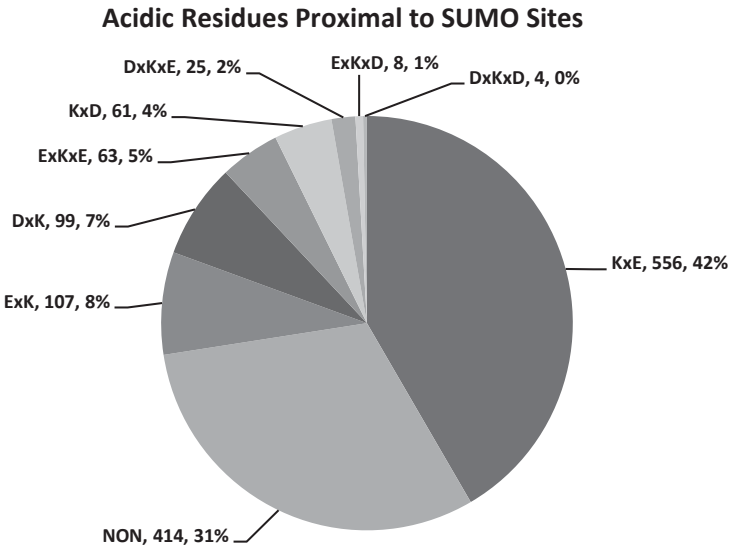

Figure 2: An overview of acidic residues located 2 positions upstream or downstream of SUMOylated lysines.
1,337 lysines were found to be SUMOylated under control conditions in this thesis, and were divided into categories depending on the presence of acidic residues proximal to the SUMOylated lysine. KxE is the dominant motif, with the canonical KxD motif utilized at less than its randomly expected frequency (=5.2%). Inverted SUMOylation sites without an acidic residue at +2 represent 15% of all lysines, and 31% of SUMOylated sites are entirely non-consensus.

aspartic acids.

An enrichment of bulky hydrophobic amino acids was also noted at -2 and -3 of the SUMOylation sites in Chapter 4, a motif dubbed the Hydrophobic Cluster SUMOylation Motif (HCSM). The SUMO target proteins RanGAP1 and Zinc Finger and BTB Domain-Containing Protein 1 (ZBTB1) both contain a HCSM, and mutation of the bulky hydrophobic amino acids resulted in abolishment of their SUMOylation. With the vastly increased amount of sites mapped in Chapter 6, statistical evidence now also exists that mainly valine is enriched at the -3 position (**Figure 3**). Interestingly, this is at a two amino acid jump away from the bulky hydrophobic amino acid normally preceding SUMOylation at -1, which would make sense from a structural point of view, with amino acids often alternating facing inward and outward in a protein chain. The HCSM could enhance binding of Ubc9, thereby increasing efficiency of SUMOylation [51].

A Negatively charged amino acid-Dependent SUMOylation Motif (NDSM) has been reported [52], which is enriched for acidic residues downstream of the SUMOylated lysine. These residues can aid in recognition of the substrate by Ubc9, and data-mining for proteins containing a NDSM has predicted novel SUMO target proteins [52]. The NDSM partially overlaps with our findings, where the region directly surrounding SUMOylated lysines is often enriched for charged residues. In

addition, we noted a larger enrichment of both glutamic and aspartic acids situated from +4 until +8 of the SUMOylated lysine. Thus, studying the amino acid context around a potential SUMO target lysine may greatly increase the accuracy of *in silico* prediction of SUMOylation sites.

Core SUMO Consensus Motif	**[IV]-K-x-E**
Enriched SUMO Consensus Motif	**[IVM]-K-[QEM]-E-P**
Inverted SUMO Consensus Motif	**[ED]-x-K**
Hydrophobic Cluster SUMO Motif	**V-x-[IVM]-K**
Negative-charge Dependent SUMO Motif	**K-x-E-x-x-E-E-E-E**
Phosphorylation Dependent SUMO Motif	**K-x-E-x-x-[ST]{phos}**
Acetylation Dependent SUMO Motif	**K-x-x-x-(x)-K{acetyl}**

Figure 3. SUMOylation consensus motifs.
The core consensus of SUMO can be summarized as [IV]KxE, with these amino acids enriched by at least 3-fold over randomly expected. Expanding the consensus with amino acids that are contextually at least significantly enriched amounted to the enriched SUMO consensus motif. The inverted SUMO motif has an acidic residue at -2 instead of +2, and the hydrophobic cluster motif is an extension of the core motif with another hydrophobic amino acid, predominantly a valine, at -3. SUMOylation may also be dependent on other nearby modifications, such as phosphorylation, which predominantly occurs at +5. Acetylation-dependent SUMOylation of histones was observed with acetylation at +4 or +5.

Crosstalk between SUMOylation and Other Post-Translational Modifications

More insight into crosstalk between SUMOylation and other PTMs was gained through refinement of the Phosphorylation Dependent SUMOylation Motif (PDSM) (**Figure 3**). Although this dependence on phosphorylation had been reported before [53], we found far more occurrences of the PDSM through our proteomics approach. In Chapter 4, PDSM-dependent SUMOylation of Nucleolar Protein 58 (NOP58) is reported. In Chapter 6, we discovered 46 instances of phosphorylation and SUMOylation simultaneously on the same peptide (Chapter 6). About half of these peptides were exclusively detected with both modifications occurring simultaneously, indicative of a full dependency. Nonetheless, the non-uniquely phosphorylated SUMO peptides could still be more efficiently SUMOylated when the phosphorylation is present, or the phosphorylation may have been removed after SUMOylation had already occurred. Furthermore, there is a heavy enrichment for phosphorylation at +5 of the SUMOylated lysine, for both non-uniquely and exclusively phosphorylated SUMO peptides (**Figure 3**).

Interestingly, two phosphorylation-dependent SUMOylation events were observed with the phosphorylation at +2, where the acid normally resides in the SUMO consensus motif. Phosphorylation could in this case directly serve as the acid required for proper recognition of the site by Ubc9. Additionally, a single phosphorylation-dependent SUMOylation was observed in conjunction with a phosphorylated

threonine, located at the dominant +5 spacing from the SUMOylated lysine.

We found evidence for an Acetylation Dependent SUMOylation Motif (ADSM), with two SUMOylation sites mapped in the histones H3 and H4 exclusively in combination with nearby acetylation (**Figure 3**). In Histone H3, lysine-18 was found to be SUMOylated with lysine-23 acetylated. Interestingly, Histone H3 acetyl Lysine-23 (H3K23ac) has been reported to serve as a docking site in tandem with non-methylated H3K4 for the chromatin regulator Transcription Intermediary Factor 1-alpha (TRIM24) [54], in which we identified 6 SUMOylation sites. In Histone H4, lysine-12 was found to be SUMOylated with lysine-16 acetylated. Histone H4 acetyl Lysine-16 (H4K16ac) is a hallmark for active gene transcription, which correlates with the methylation of H3K4 [55].

SUMOylation has previously been reported to alter the behavior of kinases; in essence SUMOylation-dependent phosphorylation [56]. We also observed many kinases to be SUMOylated, as well as phosphatases, methyltransferases, demethylases, acetyltransferases, deacetylases, and ubiquitin ligases. In total, 141 of such enzymes were found to be SUMOylated, demonstrating the immense complexity of the PTM crosstalk landscape.

Our findings provide novel insight into crosstalk between SUMOylation and ubiquitylation, with the discovery of five lysines in ubiquitin being modified by SUMO, indicative of formation of heterogeneous SUMO-ubiquitin chains. Modification of ubiquitin lysine-11 and lysine-63 by SUMO was observed under control conditions, whereas after stressing the cells, SUMOylation additionally occurred on lysine-6, lysine-27, and lysine-48. Lysine-11-linked ubiquitin-ubiquitin chains are known to be important in cell cycle control [57], whereas lysine-63-linked ubiquitin-ubiquitin chains are important in the DNA damage response [58]. Moreover, after MG-132 treatment, we found endogenous modification of the Neural precursor cell expressed developmentally down-regulated protein 8 (NEDD8) by SUMO on lysine-48. In ubiquitin-ubiquitin chains, lysine-48 is a signal that targets proteins for proteasomal degradation [59]. The Nedd8-SUMO-2 linkage on lysine-48 could have a similar function considering this modification was only found in response to MG-132. Naturally, the function of SUMO-2 conjugated to these lysines in ubiquitin and Nedd8 may not be synonymous to the respective ubiquitin-ubiquitin linkages, and thus further investigation is needed in order to elucidate the functionality and biological relevance of these novel hybrid chains. Finally, we also reconfirmed extensive modification of endogenous SUMO-1, SUMO-2 and SUMO-3, by SUMO-2 [60, 61].

Competition between SUMOylation and Other Lysine PTMs

Comparison of tandem mass spectrometry (MS/MS)-identified SUMO sites to all known ubiquitylation, acetylation and lysine-methylation sites revealed subsets of lysines which were utilized by two, three or even all four of these modifications (Chapter 6). Nearly 25% of identified SUMOylated lysines were reported to

be ubiquitylated, indicative of extensive competition between these two major post-translational modifications. Additionally, around 10% of SUMOylated lysines were found to overlap with acetylation, and 1% of SUMOylated lysines overlapped with methylation. Granted, the amount of known acetylation and methylation sites are considerably smaller when compared to ubiquitin. Thus, from the perspective of the ubiquitin, acetylation and methylation datasets, 2% of these lysines were found to be modifiable by SUMO. No particular enrichment between SUMO and any of the three lysine-PTMs was observed when comparing the different overlaps, suggesting an unbiased degree of crosstalk or competition between SUMO and the other lysine modifications. It should be noted that overlap between these modifications does not imply direct competition per say, as the modification of the same lysine in the same protein by different modifiers could occur at alternative locations in the cell, in response to varying cellular stresses, or be otherwise differentially regulated. Regardless, from a statistical point of view, the overlap observed in the same lysines being targeted by different PTMs is highly significant.

Interestingly, 10 lysines were identified to be modified by SUMOylation, and known to be modifiable by ubiquitylation, acetylation and methylation. 7 out of these 10 lysines were situated in the H2B, H3 en H4 histones, which could be explained by the exceptionally dynamic nature of histone modifications. The other 3 shared sites were located in p53, Chromobox Protein Homolog 3 (CBX3), and Eukaryotic Translation Elongation Factor 1 alpha 1 (EEF1A1). Furthermore, 200 SUMO sites were found to overlap with two other PTMs simultaneously; either ubiquitylation and acetylation, or ubiquitylation and methylation. No sites were found which were used by SUMO, acetylation and methylation, but not used by ubiquitylation. This demonstrates both the promiscuous nature of ubiquitin towards available lysines, as well as the meticulous nature of the reported proteome-wide studies on ubiquitin.

STRING analysis was performed on proteins containing lysines which were found to be competitively modified by SUMOylation, acetylation and ubiquitylation. Strikingly, we found over 80% of these proteins to be situated in a single functional cluster, for SUMO overlap with either of the other lysine PTMs, as well as with both. We also noted an increase in interaction connectivity between the proteins in the cluster, in addition to the aforementioned high participation rate of all proteins in the core cluster. Taken together, this is a clear indication of the importance of PTMs in the mechanistic and functional interaction between proteins. Furthermore, investigation of these clusters of proteins revealed them to be especially highly SUMOylated, averaging 4.8 SUMOylation sites per protein, as compared to the all-around average of 2.4 SUMO sites per protein. Ultimately, SUMOylation functions in concert with the other major PTMs, either through crosstalk or competitively, co-regulating a dense functional cluster of heavily and dynamically modified proteins.

A First Step towards Identification of Endogenous SUMOylation Sites

Finally, in order to take a step towards the identification of truly endogenous SUMOylation sites, we developed a methodology for mapping sites of SUMOylation using wild-type SUMO-2 (Chapter 7). This methodology, Protease-Reliant Identification of SUMO Modification (PRISM), functions by chemically blocking unmodified lysines in a sample. Subsequently, the sample is treated with SUMO-specific proteases, which remove the SUMO and leave free lysines where SUMO used to be. The validity of this method was confirmed by means of a second chemical labeling step, which targeted the freed lysines using a biotin compound. In combination with pre-enrichment of SUMOylation, this allowed for two known SUMO targets, Anaphase-Promoting Complex Subunit 4 (APC4) and E3 Ubiquitin-Protein Ligase TRIM33 (TRIM33), to be successfully purified using PRISM. In essence, proteins first have to be purified by their SUMOylation, and subsequently re-purified by their protease-dependent loss of SUMOylation, ensuring exceptionally specific purification.

Using PRISM in combination with mass spectrometry, where we utilized trypsin to cleave only at lysines which had been previously modified by SUMO, 389 SUMO acceptor lysines were identified. It should be noted that we used a stable cell line expressing His10-tagged, but otherwise wild-type, SUMO-2. This allowed for pre-enrichment of SUMO prior to chemical labeling, which made the protocol less laborious, and reduced background interference. We also included SILAC labeling as an internal control, so we could ascertain that the peptides-of-interest originated from cell culture, and were not introduced into the assay at a later point. Even though we equally used both heavy and medium SILAC labeling without any differential treatment, and thus expected ratios near 1 to be representative of reliable hits, we found the SILAC ratios to not necessarily correlate with discovery confidence. Thus, all peptides identifying SUMO sites were accepted as long as they were detected in either the medium or heavy SILAC channel, and filtering of the data was instead performed by Andromeda scoring.

Regardless, identification of nearly 400 sites from a singular sample of cells under standard growth conditions is still considerable when compared to other SUMO site identification projects (**Figure 4**). PRISM utilizes wild-type SUMO, and therefore does not suffer from a lack of chain formation ability. PRISM could also be modified to be applicable at the endogenous level, by using a SUMO antibody to pre-enrich endogenously SUMOylated proteins. Furthermore, since peptides generated in the PRISM protocol are vastly different from standard tryptic peptides, being highly hydrophobic and on average twice as large due to acetylation of all lysines, PRISM is able to produce and identify peptides which would be unresolvable by conventional means.

PRISM's main weak point is that it does not identify a modification site directly. There is no tryptic remnant or no actual modification present on the previously modified lysine. Whereas a second-step biotinylation could be performed, this would not actually reduce false positive hits resulting from an incomplete first

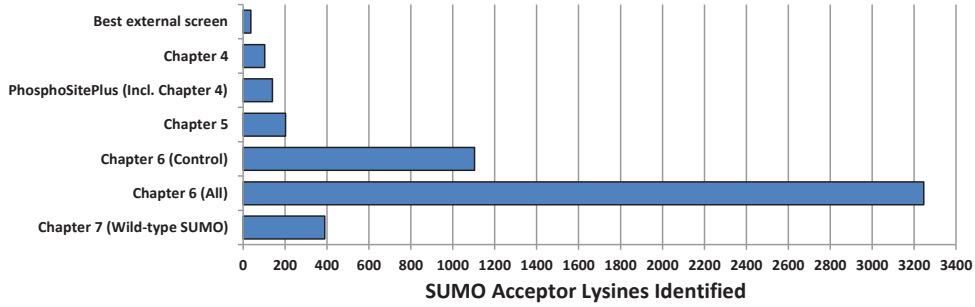

Figure 4. Overview of the amount of SUMO sites mapped in this thesis.
The best external screen published to date identified 37 SUMO sites [62], still falling short of the QQTGG mapping performed in Chapter 4 [63]. 140 SUMO MS/MS-identified SUMO sites are known on PhosphoSitePlus, including the sites in Chapter 4. Increased sample size and improved mass spectrometry technology allowed for the identification of 202 sites in Chapter 5 [64]. Improved QQTGG methodology in Chapter 6 allowed for a sharp incline in known SUMOylation sites, even when only considering cells cultured under standard growth conditions (control). Chapter 7 falls short in sheer number of SUMO sites identified, but regardless takes second place and moreover does so using a wild-type instead of a mutant SUMO.

chemical blocking of all lysines. To counteract this weakness, a control dataset would have to be created, where no protease is used. Hits from this control dataset would then qualify as false positives, and be subtracted from the main dataset. Regardless, the chemical acetylation of lysines is highly efficient under denaturing conditions, and thus false positive generation is limited.

The iteration of PRISM that uses “freed” lysines to identify sites of modification suffers from an additional weakness, in that no peptide-specific enrichment is performed. Out of 6,000 peptides identified, only 400 corresponded to sites, just over 6%. Granted, every SUMOylation site can be identified by two separate reporter peptides, as the progenitor peptide would be cut into two halves by trypsin. Regardless, this problem could be overcome by the inclusion of the biotinylation step, and subsequent purification of biotinylated peptides.

Nonetheless, the quality of the 389 wild-type SUMO-2 sites identified from a complex sample is high. Around 50% of the sites match the KxE consensus, which is a solid indication of a high-quality dataset. Furthermore, all 389 PRISM-identified sites and their corresponding 206 proteins were compared to all other SUMO datasets, and a highly significant overlap was noted (Chapter 7). 236 sites were uniquely identified by PRISM, and the other 163 sites were previously detected in our QQTGG screens. When considering the protein level, only 41 proteins were uniquely identified by PRISM, with the other 165 proteins having been identified as SUMOylated by other studies. Thus, 60% of sites detected by PRISM are novel, versus 20% novel proteins. The ratio between these two is indicative that PRISM is detecting different sites in the same SUMOylated proteins, which the QQTGG mapping is unable to detect. Interestingly, 144 sites were detected with both PRISM

and QQTGG mapping under standard growth conditions. These sites represent an extremely confident resource, with the chances of erroneous detection being virtually zero. 99 (69%) of these sites match the KxE consensus, and 76 (53%) match [VIL]KxE.

Further refinement of PRISM, along with further repeats under different cellular stress conditions, could no doubt boost the amount of PRISM-identified sites to well over 1,000. Expansion of PRISM to incorporate a monoclonal antibody, and thus gain the ability to target endogenous SUMO, would ultimately allow for the system-wide identification of endogenous SUMOylation sites from complex mammalian tissues and patient samples.

The Future of SUMOylation Proteomics

In conclusion, the studies contained within this thesis have optimized and innovated the strategy for the system-wide study of SUMO by proteomics. Moreover, the comprehensive list of validated SUMO target proteins and the large quantity of SUMO acceptor lysines will form an invaluable resource to the SUMO research field.

Regardless, much work remains to be done on SUMO proteomics, especially for the study of SUMO acceptor lysines. The methodology described within this thesis can be applied to chart the SUMO proteome in a site-specific manner and in response to a wide range of different cellular stresses. Furthermore, SUMO dynamics could be studied at a more detailed level by performing time course experiments and thus investigating the behavior of SUMO over a wider range time, as opposed to taking a single snapshot. Currently, most identifications have been performed in HeLa cells. Expanding the characterization of the SUMO proteome to cell lines originating from different tissues, i.e. Michigan Cancer Foundation-7 (MCF-7, breast cancer) or U-2 Osteosarcoma (U2-OS, osteosarcoma), will undoubtedly yield acceptor lysines unique to these cell lines. Detailed investigation of differences between such cell lines could yield insight into the relationship between SUMO and different types of cancer.

Other than cellular stresses, another interesting venue of research would be investigation of the SUMO landscape after depletion of any of the SUMO E3 ligases or SUMO-specific proteases. Unlike ubiquitin, the SUMO system only has a limited number of E3 ligases and proteases, and functional redundancy may thus be limited. As such, system-wide studies of SUMO target proteins and SUMO acceptor lysines in the absence of any of these SUMO-related enzymes could help unraveling the function of these core regulatory components of the SUMO system. In addition to ligases and proteases, SUMO-targeted ubiquitin ligases such as RNF4 and RING Finger protein 111 (RNF111) could be depleted, or even SUMO-targeted ubiquitin proteases such as USP11. The resulting build-up or depletion in SUMO target proteins could then be studied at a global level through mass spectrometry, at the SUMOylation and ubiquitin level.

The site-specific analyses described in this thesis are predominantly of a

qualitative nature, and not of a quantitative one. Whereas the SILAC methodology has been successfully applied to find differentially regulated SUMO target proteins in response to cellular stresses or throughout the cell cycle (Chapter 2 and 5) [32, 33], or otherwise identified SUMO target proteins as compared to a parental control (Chapter 2 and 5)[33], SILAC has to date not been successfully combined with site-specific study of SUMOylation. During preliminary studies for Chapter 6, we did investigate a combination of SILAC with the QQTGG approach, and successfully identified 76 SUMO sites in a quantitative manner (data not shown). However, technical difficulties resulting from the SILAC approach, such as decreased SUMO conjugation and increased fragmentation complexity, culminated in a dataset of rather underwhelming size. Thus, further optimization of the purification strategy, and scaling up of experiments in general, may be required in order to quantitatively investigate hundreds of SUMOylation sites.

Comparison of SUMOylated lysines to other lysine modifications in Chapter 6 yielded a considerable overlap between these sites. Granted, since this overlap does not necessarily mean competition or co-occurrence, further study is required to conclusively comment on this phenomenon. One possible approach would be a double purification strategy, where purified SUMOylated proteins are re-purified for the presence of a second modification such as ubiquitin or acetylation, and vice versa. Tryptic digests of these proteins could then elucidate whether the modifications co-occur or compete, and in general shed more light on the modification of proteins by multiple modifiers simultaneously. Granted, for such an approach to be successful in combination with SUMO, a Q87R SUMO mutant, or the PRISM method, would have to be applied in order to allow identification of SUMO sites.

In Chapter 6, hybrid ubiquitin-SUMO chains were identified. The exact structure and biological relevance of these novel chains remains to be investigated. One avenue of study would be through triple-MS, also known as top-down proteomics, which enables study of complete proteins or even entire protein complexes [65]. Such an approach would grant insight into the exact architecture and branching of these chain structures, as well as the amount of ubiquitin-likes that are chained together. The effect of individual ubiquitin proteases and SUMO proteases on these hybrid chains could also be studied in a similar fashion, either *in vitro* or by purification from cells after depletion of these proteases.

Finally, the study of SUMOylation should ultimately be extended to the endogenous level, so that the SUMOylated proteome may truly be charted on equal footing with modifications such as ubiquitin and acetylation. As SUMO has become increasingly implicated in contributing to the regulatory pathways involved in cancer [2, 66, 67], the ability to efficiently study SUMO in patient material or animal tissues is of paramount importance. The PRISM strategy described in this thesis takes a first step towards endogenous identification of SUMOylation sites. Refinement of the method to include a SUMO-targeted monoclonal antibody would be the second step, and would eliminate the need for exogenous expression of

mutant or otherwise epitope-tagged SUMO. In combination with high-resolution mass spectrometry, investigation of SUMO in patient material and animal tissues would become viable, and the role of SUMO in these cellular systems, as well as its contribution to cancer and other diseases, may finally be elucidated.

Reference List

1. Geiss-Friedlander, R., and Melchior, F. (2007) Concepts in sumoylation: a decade on. *Nat. Rev. Mol. Cell Biol.* 8, 947-956
2. Flotho, A., and Melchior, F. (2013) Sumoylation: a regulatory protein modification in health and disease. *Annu. Rev. Biochem.* 82, 357-385
3. Vertegaal, A. C. (2011) Uncovering ubiquitin and ubiquitin-like signaling networks. *Chem. Rev.* 111, 7923-7940
4. Ulrich, H. D., and Walden, H. (2010) Ubiquitin signalling in DNA replication and repair. *Nat. Rev. Mol. Cell Biol.* 11, 479-489
5. Jackson, S. P., and Durocher, D. (2013) Regulation of DNA damage responses by ubiquitin and SUMO. *Mol. Cell* 49, 795-807
6. Maeda, D. et al. (2004) Ubc9 is required for damage-tolerance and damage-induced interchromosomal homologous recombination in S. cerevisiae. *DNA Repair (Amst)* 3, 335-341
7. Hoege, C., Pfander, B., Moldovan, G. L., Pyrowolakis, G., and Jentsch, S. (2002) RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. *Nature* 419, 135-141
8. Psakhye, I., and Jentsch, S. (2012) Protein group modification and synergy in the SUMO pathway as exemplified in DNA repair. *Cell* 151, 807-820
9. Johnson, E. S. (2004) Protein modification by SUMO. *Annu. Rev. Biochem.* 73, 355-382
10. Bueno, M. T., and Richard, S. (2013) SUMOylation negatively modulates target gene occupancy of the KDM5B, a histone lysine demethylase. *Epigenetics.* 8
11. Ohta, K. et al. (2013) Depletion of JARID1B induces cellular senescence in human colorectal cancer. *Int. J. Oncol.* 42, 1212-1218
12. Kuzbicki, L., Lange, D., Straczynska-Niemiec, A., and Chwirot, B. W. (2013) JARID1B expression in human melanoma and benign melanocytic skin lesions. *Melanoma Res.* 23, 8-12
13. Radberger, P., Radberger, A., Bykov, V. J., Seregard, S., and Economou, M. A. (2012) JARID1B protein expression and prognostic implications in uveal melanoma. *Invest Ophthalmol. Vis. Sci.* 53, 4442-4449
14. Chi, P., Allis, C. D., and Wang, G. G. (2010) Covalent histone modifications--miswritten, misinterpreted and mis-erased in human cancers. *Nat. Rev. Cancer* 10, 457-469
15. Lyst, M. J., Nan, X., and Stancheva, I. (2006) Regulation of MBD1-mediated transcriptional repression by SUMO and PIAS proteins. *EMBO J.* 25, 5317-5328
16. Girdwood, D. et al. (2003) P300 transcriptional repression is mediated by SUMO modification. *Mol. Cell* 11, 1043-1054
17. 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
18. Yin, Y. et al. (2012) SUMO-targeted ubiquitin E3 ligase RNF4 is required for the response of human cells to DNA damage. *Genes Dev.* 26, 1196-1208
19. Galanty, Y., Belotserkovskaya, R., Coates, J., and Jackson, S. P. (2012) RNF4, a SUMO-tar-

geted ubiquitin E3 ligase, promotes DNA double-strand break repair. *Genes Dev.* 26, 1179-1195

20. Vyas, R. et al. (2013) RNF4 is required for DNA double-strand break repair in vivo. *Cell Death Differ.* 20, 490-502
21. Wiltshire, T. D. et al. (2010) Sensitivity to poly(ADP-ribose) polymerase (PARP) inhibition identifies ubiquitin-specific peptidase 11 (USP11) as a regulator of DNA double-strand break repair. *J. Biol. Chem.* 285, 14565-14571
22. Schoenfeld, A. R., Apgar, S., Dolios, G., Wang, R., and Aaronson, S. A. (2004) BRCA2 is ubiquitinated in vivo and interacts with USP11, a deubiquitinating enzyme that exhibits prosurvival function in the cellular response to DNA damage. *Mol. Cell Biol.* 24, 7444-7455
23. Conlan, L. A., McNees, C. J., and Heierhorst, J. (2004) Proteasome-dependent dispersal of PML nuclear bodies in response to alkylating DNA damage. *Oncogene* 23, 307-310
24. Wu, X., Yen, L., Irwin, L., Sweeney, C., and Carraway, K. L., III. (2004) Stabilization of the E3 ubiquitin ligase Nrdp1 by the deubiquitinating enzyme USP8. *Mol. Cell Biol.* 24, 7748-7757
25. Canning, M., Boutell, C., Parkinson, J., and Everett, R. D. (2004) A RING finger ubiquitin ligase is protected from autocatalyzed ubiquitination and degradation by binding to ubiquitin-specific protease USP7. *J. Biol. Chem.* 279, 38160-38168
26. Blount, J. R., Burr, A. A., Denuc, A., Marfany, G., and Todi, S. V. (2012) Ubiquitin-specific protease 25 functions in Endoplasmic Reticulum-associated degradation. *PLoS. One.* 7, e36542
27. Wu, H. C. et al. (2014) USP11 regulates PML stability to control Notch-induced malignancy in brain tumours. *Nat. Commun.* 5, 3214
28. Burkhart, R. A. et al. (2013) Mitoxantrone Targets Human Ubiquitin-Specific Peptidase 11 (USP11) and Is a Potent Inhibitor of Pancreatic Cancer Cell Survival. *Mol. Cancer Res.* 11, 901-911
29. Dellaire, G., and Bazett-Jones, D. P. (2004) PML nuclear bodies: dynamic sensors of DNA damage and cellular stress. *Bioessays* 26, 963-977
30. Guo, A. et al. (2000) The function of PML in p53-dependent apoptosis. *Nat. Cell Biol.* 2, 730-736
31. Wang, Z. G. et al. (1998) PML is essential for multiple apoptotic pathways. *Nat. Genet.* 20, 266-272
32. 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
33. Golebiowski, F. et al. (2009) System-wide changes to SUMO modifications in response to heat shock. *Sci. Signal.* 2, ra24
34. Becker, J. et al. (2013) Detecting endogenous SUMO targets in mammalian cells and tissues. *Nat. Struct. Mol. Biol.* 20, 525-531
35. Bruderer, R. et al. (2011) Purification and identification of endogenous polySUMO conjugates. *EMBO Rep.* 12, 142-148
36. Xu, G., Paige, J. S., and Jaffrey, S. R. (2010) Global analysis of lysine ubiquitination by ubiquitin remnant immunoaffinity profiling. *Nat. Biotechnol.* 28, 868-873
37. Choudhary, C. et al. (2009) Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* 325, 834-840
38. Larsen, M. R., Thingholm, T. E., Jensen, O. N., Roepstorff, P., and Jorgensen, T. J. (2005) Highly selective enrichment of phosphorylated peptides from peptide mixtures using

titanium dioxide microcolumns. *Mol. Cell Proteomics.* 4, 873-886

39. Knuesel, M., Cheung, H. T., Hamady, M., Barthel, K. K., and Liu, X. (2005) A method of mapping protein sumoylation sites by mass spectrometry using a modified small ubiquitin-like modifier 1 (SUMO-1) and a computational program. *Mol. Cell Proteomics* 4, 1626-1636
40. Wohlschlegel, J. A., Johnson, E. S., Reed, S. I., and Yates, J. R., III. (2006) Improved identification of SUMO attachment sites using C-terminal SUMO mutants and tailored protease digestion strategies. *J. Proteome. Res.* 5, 761-770
41. Denison, C. et al. (2005) A proteomic strategy for gaining insights into protein sumoylation in yeast. *Mol. Cell Proteomics* 4, 246-254
42. Blomster, H. A. et al. (2010) In vivo identification of sumoylation sites by a signature tag and cysteine-targeted affinity purification. *J. Biol. Chem.* 285, 19324-19329
43. Hsiao, H. H., Meulmeester, E., Frank, B. T., Melchior, F., and Urlaub, H. (2009) "ChopNSpice," a mass spectrometric approach that allows identification of endogenous small ubiquitin-like modifier-conjugated peptides. *Mol. Cell Proteomics* 8, 2664-2675
44. Twardzik, D. R., and Peterkofsky, A. (1972) Glutamic acid as a precursor to N-terminal pyroglutamic acid in mouse plasmacytoma protein (protein synthesis-initiation-immunoglobulins-pyrrolidone carboxylic acid). *Proc. Natl. Acad. Sci. U. S. A* 69, 274-277
45. Ashburner, M. et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 25, 25-29
46. Ruepp, A. et al. (2008) CORUM: the comprehensive resource of mammalian protein complexes. *Nucleic Acids Res.* 36, D646-D650
47. Kanehisa, M., and Goto, S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27-30
48. Wilson, V. G., and Heaton, P. R. (2008) Ubiquitin proteolytic system: focus on SUMO. *Expert. Rev. Proteomics* 5, 121-135
49. 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
50. 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
51. Bernier-Villamor, V., Sampson, D. A., Matunis, M. J., and Lima, C. D. (2002) Structural basis for E2-mediated SUMO conjugation revealed by a complex between ubiquitin-conjugating enzyme Ubc9 and RanGAP1. *Cell* 108, 345-356
52. Yang, S. H., Galanis, A., Witty, J., and Sharrocks, A. D. (2006) An extended consensus motif enhances the specificity of substrate modification by SUMO. *EMBO J.* 25, 5083-5093
53. Hietakangas, V. et al. (2006) PDSM, a motif for phosphorylation-dependent SUMO modification. *Proc. Natl. Acad. Sci. U. S. A* 103, 45-50
54. Tsai, W. W. et al. (2010) TRIM24 links a non-canonical histone signature to breast cancer. *Nature* 468, 927-932
55. Dou, Y. et al. (2005) Physical association and coordinate function of the H3 K4 methyltransferase MLL1 and the H4 K16 acetyltransferase MOF. *Cell* 121, 873-885
56. Yao, Q., Li, H., Liu, B. Q., Huang, X. Y., and Guo, L. (2011) SUMOylation-regulated

protein phosphorylation, evidence from quantitative phosphoproteomics analyses. *J. Biol. Chem.* 286, 27342-27349

57. Bremm, A., and Komander, D. (2011) Emerging roles for Lys11-linked polyubiquitin in cellular regulation. *Trends Biochem. Sci.* 36, 355-363
58. Spence, J., Sadis, S., Haas, A. L., and Finley, D. (1995) A ubiquitin mutant with specific defects in DNA repair and multiubiquitination. *Mol. Cell Biol.* 15, 1265-1273
59. Ikeda, F., and Dikic, I. (2008) Atypical ubiquitin chains: new molecular signals. 'Protein Modifications: Beyond the Usual Suspects' review series. *EMBO Rep.* 9, 536-542
60. Vertegaal, A. C. (2010) SUMO chains: polymeric signals. *Biochem. Soc. Trans.* 38, 46-49
61. Vertegaal, A. C. (2007) Small ubiquitin-related modifiers in chains. *Biochem. Soc. Trans.* 35, 1422-1423
62. Lamoliatte, F. et al. (2013) Targeted Identification of SUMOylation Sites in Human Proteins Using Affinity Enrichment and Paralog-specific Reporter Ions. *Mol. Cell Proteomics* 12, 2536-2550
63. 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
64. Schimmel, J. et al. (2014) Uncovering SUMOylation Dynamics during Cell-Cycle Progression Reveals FoxM1 as a Key Mitotic SUMO Target Protein. *Mol. Cell*
65. Catherman, A. D., Skinner, O. S., and Kelleher, N. L. (2014) Top Down Proteomics: Facts and Perspectives. *Biochem. Biophys. Res. Commun.*
66. Lee, L., Sakurai, M., Matsuzaki, S., Arancio, O., and Fraser, P. (2013) SUMO and Alzheimer's Disease. *Neuromolecular. Med.*
67. Lee, Y. J., and Hallenbeck, J. M. (2013) SUMO and Ischemic Tolerance. *Neuromolecular. Med.*