The following resources related to this article are available online at http://stke.sciencemag.org.
This information is current as of August 20, 2014.

| | |
|---|---|
| **Article Tools** | Visit the online version of this article to access the personalization and article tools:<br>http://stke.sciencemag.org/content/7/323/rs2 |
| **Supplemental Materials** | *"Supplementary Materials"*<br>http://stke.sciencemag.org/content/suppl/2014/04/25/7.323.rs2.DC1.html |
| **Related Content** | The editors suggest related resources on *Science*'s sites:<br>http://stke.sciencemag.org/content/sigtrans/2/72/ra24.full.html<br>http://stke.sciencemag.org/content/sigtrans/4/178/rs4.full.html<br>http://stke.sciencemag.org/content/sigtrans/5/253/ra88.full.html<br>http://stke.sciencemag.org/content/sigtrans/5/251/ra84.full.html |
| **References** | This article cites 60 articles, 23 of which you can access for free at:<br>http://stke.sciencemag.org/content/7/323/rs2#BIBL |
| **Glossary** | Look up definitions for abbreviations and terms found in this article:<br>http://stke.sciencemag.org/cgi/glossarylookup |
| **Permissions** | Obtain information about reproducing this article:<br>http://www.sciencemag.org/about/permissions.dtl |

POST-TRANSLATIONAL MODIFICATION

# Proteome-Wide Identification of SUMO2 Modification Sites

Triin Tammsalu,[1]* Ivan Matic,[1]*† Ellis G. Jaffray,[1] Adel F. M. Ibrahim,[2]
Michael H. Tatham,[1] Ronald T. Hay[1]‡

Posttranslational modification with small ubiquitin-like modifiers (SUMOs) alters the function of proteins involved in diverse cellular processes. SUMO-specific enzymes conjugate SUMOs to lysine residues in target proteins. Although proteomic studies have identified hundreds of sumoylated substrates, methods to identify the modified lysines on a proteomic scale are lacking. We developed a method that enabled proteome-wide identification of sumoylated lysines that involves the expression of polyhistidine (6His)–tagged SUMO2 with Thr[90] mutated to Lys. Endoproteinase cleavage with Lys-C of 6His-SUMO2[T90K]–modified proteins from human cell lysates produced a diGly remnant on SUMO2[T90K]-conjugated lysines, enabling immunoprecipitation of SUMO2[T90K]–modified peptides and producing a unique mass-to-charge signature. Mass spectrometry analysis of SUMO-enriched peptides revealed more than 1000 sumoylated lysines in 539 proteins, including many functionally related proteins involved in cell cycle, transcription, and DNA repair. Not only can this strategy be used to study the dynamics of sumoylation and other potentially similar posttranslational modifications, but also, these data provide an unprecedented resource for future research on the role of sumoylation in cellular physiology and disease.

## INTRODUCTION

Posttranslational modification alters the activity, function, and fate of modified proteins. There are many types of posttranslational modifications ranging in size from small chemical groups, such as phosphate, to large protein molecules, like ubiquitin and ubiquitin-like proteins (Ubls). The diversity and number of posttranslational modifications increase the complexity of the proteome by several orders of magnitude. In addition to ubiquitin, the mammalian Ubl family includes at least 11 other proteins that conjugate to lysine residues and share a highly conserved structural fold in spite of low sequence conservation (1). Small ubiquitin-like modifiers (SUMOs) are essential for cell viability (2) and cellular responses to stress conditions, including heat shock (3, 4), proteasome inhibition (5, 6), and DNA damage (7, 8). Mammalian cells express three SUMO paralogs that are conjugated to target proteins. SUMO2 and SUMO3 differ by only three amino acids, and share 46 to 48% amino acid identity with SUMO1 (9). SUMO2 and SUMO3 form chains (10, 11) that can promote ubiquitin-mediated degradation of target proteins (12, 13).

SUMO maturation and conjugation is a multistep process. Translation of mRNAs encoding SUMO1, SUMO2, and SUMO3 produces inactive pro-proteins that are activated by SUMO-specific proteases, which remove the inhibitory C-terminal residues, exposing two glycines known as a diGly motif. SUMO conjugation involves three distinct enzymatic activities, known as E1, E2, and E3. The heterodimeric E1, composed of SUMO-activating enzymes 1 and 2 (SAE1 and SAE2), uses ATP (adenosine 5′-triphosphate) to form a thioester bond between the sulfhydryl group of the Cys in its active site and the carboxyl group of the C-terminal Gly of SUMO. The Cys in the active site of the E2 enzyme Ubc9 accepts SUMO from the E1 enzyme by transthiolation. SUMO E3 ligases catalyze

the transfer of SUMO from Ubc9 to substrate proteins, forming an isopeptide bond between the ε-amino group of the Lys in the substrate protein and the C-terminal carboxyl of Gly in SUMO. SUMO E3 ligases fall into two classes: PIAS [protein inhibitor of activated STAT (signal transducer and activator of transcription)] proteins are similar to RING (really interesting new gene) domain–containing ubiquitin ligases, whereas other SUMO E3 ligases, such as RanBP2 (14), do not contain RING domains (15). The removal of SUMO from substrate proteins is catalyzed by SUMO-specific proteases, including six sentrin-specific proteases (SENPs), which cleave isopeptide bonds between the SUMO C-terminal carboxyl group and substrate proteins (16).

The identity of SUMO-modified proteins is typically determined by either mutational analyses or large-scale proteomic approaches. Both strategies rely on the enrichment of SUMO substrates from complex protein mixtures because the sumoylated form of a protein constitutes only a small proportion of the total protein abundance. Affinity purification can be used to enrich for SUMO substrates from cultured cell lysates (17) and knock-in mice (18) expressing tagged forms of SUMO. Combining stringent enrichment methods with mass spectrometry (MS)–based proteomics enables the identification of hundreds of substrates in a single experiment (4, 6, 19). SUMO-specific antibody-based substrate enrichment methods can be used to identify endogenous SUMO substrates by MS (20, 21). Moreover, SUMO interaction motifs (SIMs) have been exploited to enrich for endogenous SUMO substrates (22). Together, these methods have identified hundreds of putative SUMO substrates, but definitive evidence for sumoylation requires information about the specific lysine that is modified.

Heretofore, there are no MS-based techniques that enable researchers to identify site-specific sumoylation on a scale of hundreds of sites in a single experiment. The primary limitation is the inherent complexity of peptide mixtures derived from protein level purifications of SUMO-modified protein. Peptide level enrichment strategies for other posttranslational modifications, including phosphorylation and acetylation, improve the frequency and quality of modified peptide identification (23, 24). Likewise, immunoaffinity purification of ubiquitylated peptides using antibodies specific to the diGly tryptic remnant on ubiquitin-conjugated lysines (25)

enables the detection of more than 10,000 sites of ubiquitylation in a single study (26, 27).

The identification of sumoylated sites is limited by the fact that "bottom-up" MS-based proteomic studies typically use trypsin for protein digestion, which creates large branched peptides that are not amenable to standard database search algorithms. Branched side-chain remnants resulting from trypsin digestion of mammalian SUMO1, SUMO2, or SUMO3 comprise either 19 or 32 amino acids of the SUMO C terminus covalently bound to Lys on the substrate peptide. Whereas these remnants have a unique mass-to-charge signature, their fragmentation results in complex MS2 spectra that complicate the identification of the substrate sequence. Various approaches used to interpret MS2 spectra of branched sumoylated peptides (28–30) have limited utility in complex mixtures associated with proteomic analyses (31). As an alternative approach, trypsin recognition sites engineered into the C terminus of SUMO can produce shorter branched side chains more amenable to MS-based analyses (32, 33). For example, sumoylation by exogenously expressed Q87R or T90R mutated SUMO2 promotes modified substrates that when digested with trypsin produce remnants of five or two amino acids that can be detected by MS to identify SUMO conjugation sites (29).

To enable the large-scale site-specific analysis of sumoylation, we developed a new peptide-specific enrichment approach. We stably expressed polyhistidine (6His)–tagged SUMO2 with Thr$^{90}$ mutated to Lys (SUMO2$^{T90K}$) in human embryonic kidney (HEK) 293 cells. Cells were heat-shocked to enhance sumoylation, SUMO2$^{T90K}$-conjugated proteins were purified under denaturing conditions using the His tag, and after cleavage with endoproteinase Lys-C, a diGly-Lys–specific antibody was used to enrich SUMO2$^{T90K}$ remnant–containing peptides. Analysis of the resultant peptide mixtures identified more than 1000 sites of modification by SUMO2, providing an unprecedented resource for future studies. In addition, although SUMO2 was used in this study to allow comparison with the existing data, the principles of this strategy could be applied to other SUMOs and Ubls in various cellular systems.

## RESULTS

### Functional analysis of SUMO2$^{T90K}$

Digestion of ubiquitinated proteins with trypsin gives rise to a diGly remnant linked to the acceptor lysine, which can be used to enrich for modified peptides by affinity purification with a selective diGly-Lys–specific monoclonal antibody (anti-KεGG) (25). The Ubl family members, ubiquitin, NEDD8,

and ISG15 each contain an Arg preceding a diGly sequence at the C terminus (Fig. 1A). Thus, trypsin-mediated cleavage of proteins conjugated to these Ubls creates an identical branched side-chain consisting of diGly conjugated to Lys. We mutated Thr$^{90}$ to Lys in SUMO2 (SUMO2$^{T90K}$), creating a Lys N terminal to the diGly motif at the C terminus (Fig. 1A), and generated HEK293 cells stably expressing 6His-SUMO2$^{T90K}$. Similar to parental cells, SUMO2$^{T90K}$-expressing cells had a comparable doubling time and morphology (fig. S1) and responded to heat shock by increasing the abundance of SUMO-modified proteins (Fig. 1B).

To assess whether the SUMO2$^{T90K}$ mutation had any effect on SUMO2 conjugation or cleavage, we performed in vitro enzyme assays. We found
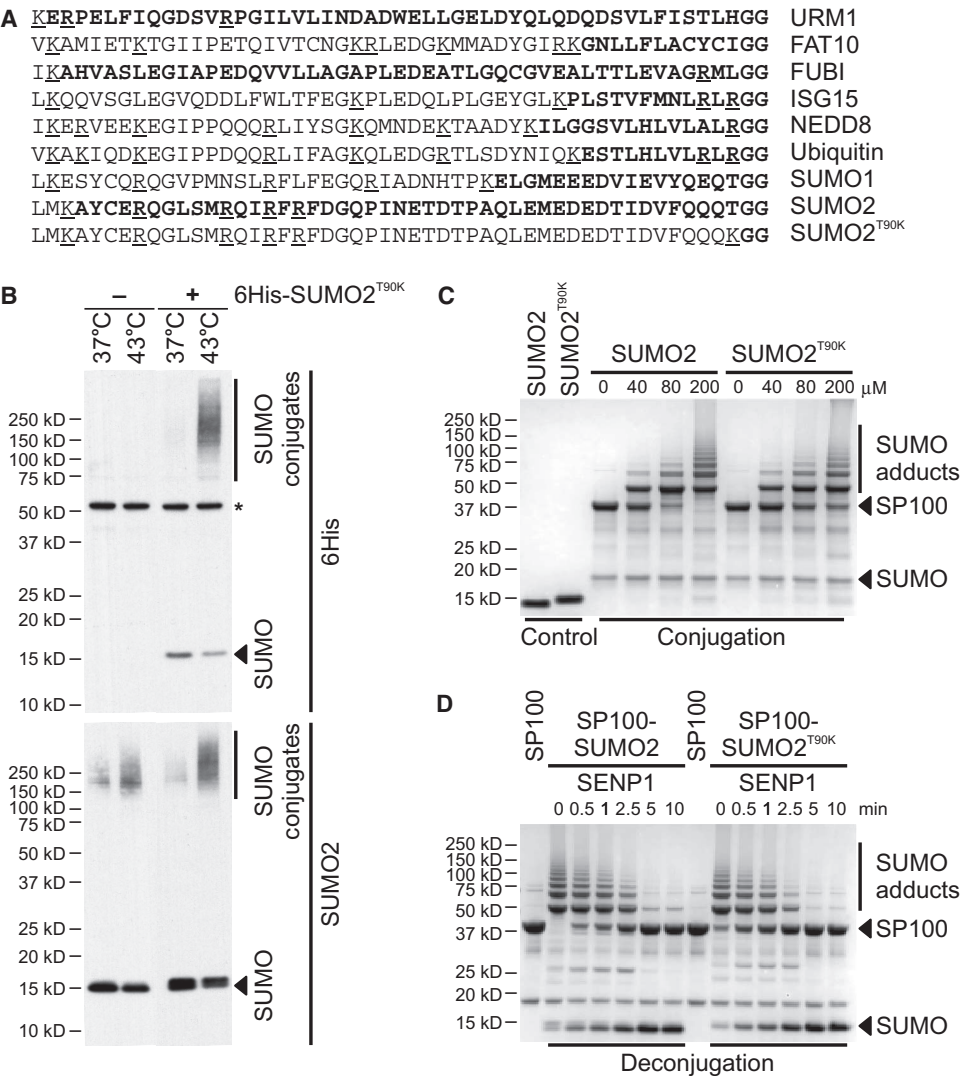


Fig. 1. Alignment of human Ubls and functional analysis of SUMO2$^{T90K}$. (A) Sequence alignment of the C terminus of human Ubl family proteins that terminate in a diGly motif. Sequence of the predicted peptide after Lys-C digestion is highlighted in bold. Lys and Arg are underlined. (B) Western blots for 6His (top) and SUMO2 (bottom) in lysates of parental HEK293 NS3 cells or those stably expressing 6His-SUMO2$^{T90K}$ and either unstressed (37°C) or heat-stressed for 30 min at 43°C. *Nonspecific immunoreactivity. Data are representative of three independent experiments. (C and D) Images of Coomassie-stained protein gels of in vitro sumoylation (C) or desumoylation (D) reactions of recombinant wild-type or T90K mutant SUMO2 using SP100 as a substrate. Data are representative of three independent experiments.

that purified SUMO2 and SUMO2$^{T90K}$ showed little difference in the rate of polymerization and conjugation to the substrate proteins SP100 (Fig. 1C), IRF2, RanGAP1, or PML (fig. S2A). Likewise, SUMO2 and SUMO2$^{T90K}$ had similar effects on the rate of cleavage from these substrates by SENP1 (Fig. 1D and fig. S2B). Finally, SENP1 and SENP2 cleaved the inactive proforms of recombinant SUMO2 and SUMO2$^{T90K}$ at a similar rate (fig. S2C).

## Proteome-wide identification of sumoylation sites in human cells

We established a workflow using 6His-SUMO2$^{T90K}$–expressing cells that enabled us to select for SUMO-modified proteins and enrich for peptides containing diGly-conjugated lysines (Fig. 2A). Cells (~2 × 10$^9$) were heat-stressed at 43°C for 30 min to enhance sumoylation (3) and lysed under denaturing conditions. Proteins conjugated to 6His-SUMO2$^{T90K}$ were isolated by nickel affinity chromatography and subsequently digested with Lys-C, or Lys-C and Glu-C to truncate long substrate peptides. Peptides containing diGly-Lys were enriched with anti-KεGG and analyzed by MS using parameters optimized for the identification of low-abundance peptides. Using this strategy, we identified 2747 peptides, of which 1217 (44%) contained a diGly-Lys motif (table S2). These peptides represented 1002 unique sumoylated sites in 539 proteins (table S1).

To estimate the reproducibility of the SUMO2$^{T90K}$-based enrichment strategy, we repeated the experiment. We performed the SUMO2$^{T90K}$ protein and peptide enrichment workflow in duplicate on a lysate of HEK293 culture. In this analysis, we identified 468 sumoylated sites in common with the 1002 SUMO2$^{T90K}$-modified sites found in the first experiment, of which 323 (69%) were common between replicate purifications (fig. S3), indicating a high degree of reproducibility. In addition, this analysis suggested that relatively small-scale experiments (~6 × 10$^8$ cells) are sufficient to identify large numbers of sumoylated peptides. Moreover, we determined that setting the mass spectrometer to fragment and scan the most abundant peptide from the initial full MS scan ("top 1") with a maximum 1-s fill time yielded almost twice the number of successful identifications of sumoylated peptides as using "top 10" data-dependent acquisition with a 60-ms fill time (Fig. 2B). From these data, we determined that antibody-based enrichment of SUMO2$^{T90K}$-modified peptides increased the percentage of identified sumoylated peptides from 0.05% to greater than 31% compared to nickel affinity purification alone (Fig. 2B).

We also compared our data to published MS-based analyses of posttranslational modifications. SUMO2$^{T90K}$-modified sites overlapped with 55% of SUMO2-modified sites detected in the largest MS-based study of sumoylation (29) and 39% of sites identified in all MS-based studies of sumoylation (34) (Fig. 2C). Slightly less than half of SUMO2$^{T90K}$-modified proteins overlapped with proteins identified in published studies involving protein-level SUMO substrate identification (fig. S4, A and B). However, the fact that these studies used different

cell types, which likely contain different sets of sumoylated proteins, may, at least in part, explain the relatively low overlap. Comparison of functional annotations of sumoylated proteins discovered here and in other large-scale SUMO2 proteomic studies revealed similar enrichment of sumoylated proteins in cellular processes, including gene expression, RNA metabolism, DNA replication, repair and recombination, and cell proliferation, survival, and growth (fig. S4C). Only 146 SUMO2$^{T90K}$-modified sites (14.6%) overlapped with known sites of acetylation (34) (table S1), and 378 SUMO2$^{T90K}$-modified sites (37.8%) overlapped with known sites of ubiquitylation (34) (table S1). Whereas the accuracy of the estimated percentage of overlap among these studies could be compromised by the relative completeness of the data sets, these observations at least suggest that ubiquitination and, less commonly, acetylation can occur on the same residues as sumoylation.

## Sequence context of SUMO2 modification sites

Initial findings of a consensus motif for SUMO modification [ψKxE or ψKxD, where ψ represents a hydrophobic residue, and x represents any
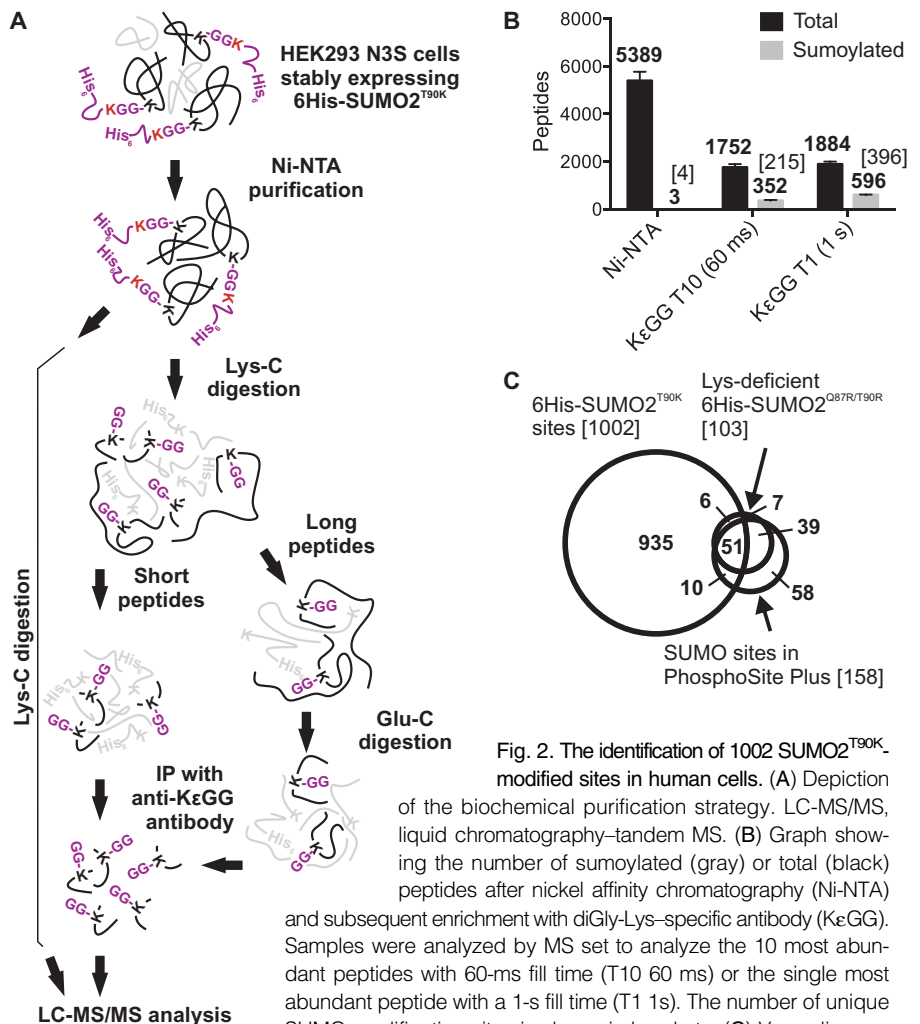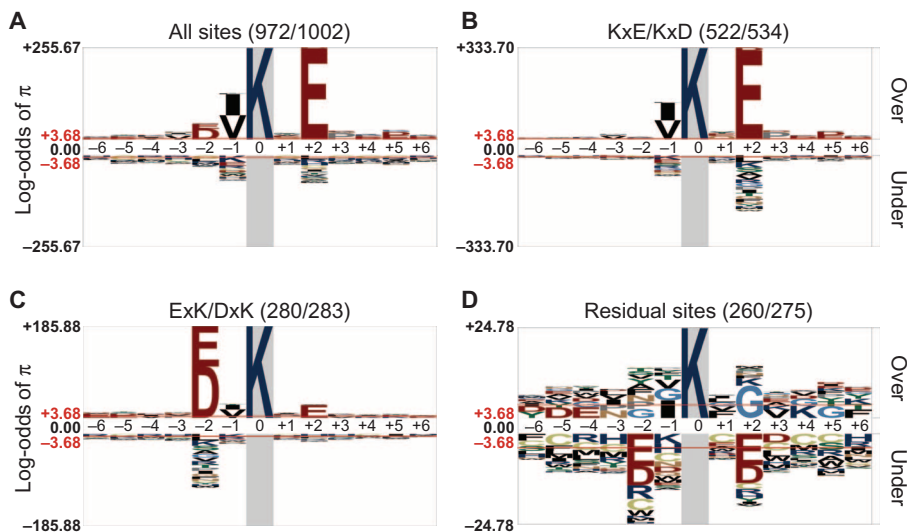


Fig. 2. The identification of 1002 SUMO2$^{T90K}$-modified sites in human cells. (A) Depiction of the biochemical purification strategy. LC-MS/MS, liquid chromatography–tandem MS. (B) Graph showing the number of sumoylated (gray) or total (black) peptides after nickel affinity chromatography (Ni-NTA) and subsequent enrichment with diGly-Lys–specific antibody (KεGG). Samples were analyzed by MS set to analyze the 10 most abundant peptides with 60-ms fill time (T10 60 ms) or the single most abundant peptide with a 1-s fill time (T1 1s). The number of unique SUMO modification sites is shown in brackets. (C) Venn diagram showing comparison among sumoylated sites identified in this study, in an MS-based study using Lys-deficient 6His-SUMO2$^{Q87R/T90R}$–expressing HeLa cells (29), and in all other studies using MS as annotated in the PhosphoSitePlus database (34). The total number of sumoylation sites is shown in brackets.

amino acid (35)] focused experimental investigation toward these sites, which may explain the fact that most SUMO-modified sites discovered to date conform to this consensus. Small-scale proteomic studies confirmed this, but also identified a less common inverted consensus motif DxKψ or ExKψ (29). Our SUMO2[T90K]-modified peptide data set enabled us to investigate the sequences of sumoylation sites using robust statistical analyses. We used pLogo (36) to analyze the sequence of diGly-Lys–containing peptides. This analysis revealed a significant overrepresentation of Glu two amino acids C-terminal to the conjugated Lys (position +2), of hydrophobic residues Ile or Val at position −1, and, to a lesser degree, of the acidic residues Asp or Glu at position −2 (Fig. 3A). This result is consistent with a mixed population of sumoylated peptides containing forward and inverted sites. Unexpectedly, we also found that 90 sumoylated peptides had Glu or Asp at both positions +2 and −2 simultaneously (table S1), suggesting that some sites contain "forward and inverted"–type consensus motifs. We separately analyzed sites containing forward, inverted, or forward and inverted consensus motifs for additional conserved residues. As expected, most sequences with forward consensus motifs contained Glu at position +2 (90%) and a large hydrophobic residue at position −1, typically either Val (31.7%) or Ile (26.5%) (Fig. 3B), although Leu, Phe, and Pro were also overrepresented. For sequences with inverted consensus motifs, there was an approximately equal likelihood of Glu or Asp at position −2, but unexpectedly, no preference for large hydrophobic residues at position +1 (Fig. 3C). Instead, there was an overrepresentation of Val or Ile at position −1 and Glu at position +2, largely in peptides with the forward and inverted motif. Together, sites with forward or inverted consensus motifs represented 72.6% of all SUMO modification sites. The sequences of the remaining sumoylated sites did not show a strong consensus, although there was a small but significant overrepresentation of Gly at position +2 (Fig. 3D).
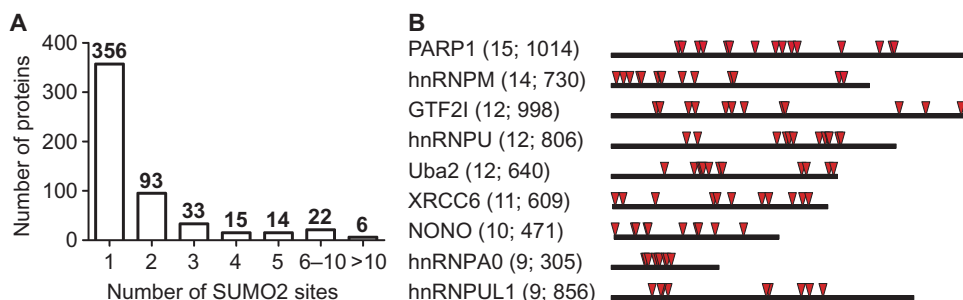
## Potential for multisite modification

Reduced electrophoretic mobility of sumoylated proteins has led researchers to predict that proteins contain multiple SUMO modification sites (22). We found that about one-third of SUMO2[T90K]-modified proteins had more than one sumoylation site (Fig. 4A and table S1). For example, PARP1 had 15 sites, hnRNPM had 14 sites, and hnRNPU, GTF2I, and Uba2 had 12 sites (Fig. 4B and table S1). Because Uba2 is a component of the heterodimeric E1 SUMO-activating enzyme (37, 38), this suggests that it may undergo autosumoylation. Sumoylation sites were clustered in some proteins (Fig. 4B), implying that these regions are more susceptible to modification. We were not able to determine whether multiple sumoylation sites were simultaneously modified for most proteins due to methodological limitations. However, because Lys-C is inhibited by posttranslational modification of Lys, we were able to identify a few peptides with simultaneous sumoylation of adjacent lysines (table S1 and SMfile1.PDF). SUMO1, SUMO2, and SUMO3 have multiple posttranslational modifications, including acetylation at their N terminus and sumoylation on internal lysines (table S1 and SMfile1.PDF). The major site involved in chain formation on SUMO2 and SUMO3 is Lys[11] (10), which we frequently detected as a SUMO2[T90K]-modified site. We also detected doubly modified peptides of SUMO2 and SUMO3, including one modified at Lys[11] and Lys[7], and the other modified at Lys[5] and Lys[7] (table S1 and SMfile1. PDF), suggesting that multiple branching patterns may occur.

## Evidence for protein group modification

SUMO modification of diverse substrates is transiently increased in response to proteotoxic stress (3, 4, 6). Analysis of the functional annotations of 539 SUMO2[T90K]-modified proteins in our data set derived from heat-shocked cells revealed a statistical overrepresentation of proteins involved in highly interconnected complexes (Fig. 5) that regulate gene expression; DNA replication, recombination, and repair; and cell growth and proliferation (fig. S4C). Zinc finger transcription factors were frequently identified in our analysis with evidence for 85 SUMO2[T90K]-modified sites distributed among 52 different proteins (table S1). Several studies investigating individual zinc finger transcription factors identified sites of SUMO modification in these proteins by mutational analysis (39) and demonstrated that sumoylation is generally associated with transcriptional repression (40). Similarly, 78 sumoylation sites were present in 14 heterogeneous ribonucleoprotein particle proteins (hnRNPs) (Fig. 5), consistent with previous studies (41). Our data extended previous observations by demonstrating extensive SUMO modification of hnRNPs, including 14 sites on hnRNPM, 11 sites on hnRNPU, 9 sites on hnRNPA0, 9 sites on hnRNPUL1, and 8 sites on hnRNPC. In addition, as expected from a previous analysis of endogenous sumoylated proteins (22), we identified multiple SUMO2[T90K]-modified proteins involved in DNA replication, recombination, and repair (Fig. 5). Thus, these results suggest that groups of functionally related proteins are likely subjected to contemporaneous SUMO modification.

## A database of virtual sites to search for modifications with wild-type SUMO

Branched peptides generated from tryptic digestion of proteins modified by wild-type SUMOs

Fig. 3. Sequence analysis of SUMO2[T90K]-modified peptides. (A to D) Sequence logo graphs of amino acid sequence conservation surrounding (A) 1002 SUMO2[T90K]-modified sites, (B) a subset of 534 sites with acidic residue (Glu or Asp) in position +2, (C) a subset of 283 sites with acidic residue (Glu or Asp) in position −2, and (D) 275 residual lysines. The fraction in the parentheses represents the number of sites with a full 13–amino acid sequence divided by the total number of sites that conformed to that motif. The y axis corresponds to the log-odds of the binomial probability (π). Threshold values of 3.68 (P < 0.05) are shown in red and marked with red horizontal lines. Note different scales of y axes.

Fig. 4. Evidence for multisite modification by SUMO2$^{T90K}$. (A) Graph of the number of SUMO2$^{T90K}$-modified sites per protein. (B) Graphical representation of clustering of sumoylation site in highly modified proteins (nine or more sites per protein). The length of the bar corresponds to the number of amino acids in the protein, and the red triangles indicate the location of sumoylation sites. The numbers of sumoylation sites and amino acids in the protein are shown in parentheses.

produce complex MS2 fragmentation spectra that are not recognized by standard annotation algorithms (11, 28). To enable automated searching for these peptides, we used our list of SUMO2$^{T90K}$-modified sites to generate virtual branched peptide databases of artificial peptide sequences. In this database, the C-terminal tryptic peptide of either SUMO1 or SUMO2 was fused to the N terminus of the predicted tryptic peptide corresponding to observed SUMO2$^{T90K}$-modified sites (Fig. 6A and SUMO1.FASTA and SUMO2.FASTA).

Many SUMO2$^{T90K}$-modified peptides were also phosphorylated (table S1). Therefore, we used the virtual branched peptide databases to search raw data files containing high-resolution spectra from an analysis of phosphopeptide-enriched samples (42). We identified a branched peptide using the virtual branched SUMO2 database that contained the long tryptic remnant of wild-type SUMO2 conjugated to Lys$^{648}$ of the transcription factor c-Myb (Fig. 6B). In the SUMO2$^{T90K}$ data, we only detected the Ser$^{653}$ phosphorylated form of the peptide (table S1), consistent with the fact that the sequence surrounding Lys$^{648}$ on c-Myb corresponds to a phosphorylation-dependent sumoylation motif (ψKxExxSP). This observation is consistent with independent studies that show sumoylation of Lys$^{648}$ (43, 44) and phosphorylation of Ser$^{653}$ (45–47).

To identify additional branched peptides modified by wild-type SUMO2, we used MS to reanalyze peptides from a proteomic study using exogenously expressed SUMO2 fused to a tandem affinity purification (TAP) tag (6) and searched the resultant spectra using the virtual branched SUMO2 peptide database. We identified 14 additional sumoylated peptides (SMfile2.PDF), including one containing sumoylation of Lys$^{11}$ of ubiquitin (Fig. 6C). These data validated the SUMO2$^{T90K}$ results and demonstrated that using virtual branched peptide databases based on the SUMO2$^{T90K}$ data for reanalysis of existing MS2 spectra may be useful for identifying sites sumoylated by both endogenous and overexpressed wild-type SUMO2.

## DISCUSSION

Methods to enrich for posttranslationally modified peptides are instrumental in identifying modified sites on a proteome-wide scale by MS. Without enrichment, the relatively low abundance of modified peptides causes most of them to go undetected in complex mixtures. Previous studies have used strategies that employ overexpression of SUMO with mutations in the C terminus that shorten the side-chain remnants on substrate proteins after trypsin digestion (29, 32, 48). These strategies use affinity chromatography to enrich for proteins modified by tagged SUMO but lack efficient peptide-based enrichment steps. To enrich for sumoylated peptides, we created a stable cell line expressing 6His-SUMO2$^{T90K}$ and performed metal affinity purification of SUMO2$^{T90K}$-modified proteins and, after cleavage with Lys-C, antibody-based enrichment of sumoylated peptides with anti-KεGG. Analysis of diGly-Lys–containing peptides by MS with instrument settings optimized for the detection of low-abundance peptides identified 1002 sumoylation sites among 539 proteins. Peptide level enrichment increased the frequency of identification of sumoylated peptides by more than 600-fold.
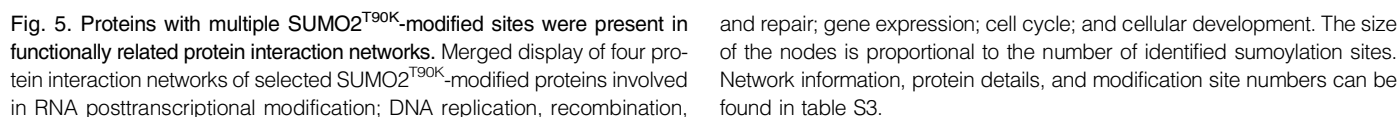
Previous studies suggest that sumoylation primarily occurs on proteins containing the forward consensus motifs ψKxE and ψKxD or the inverted consensus motifs ExKψ and DxKψ (29). However, our study revealed that, for sites with the forward consensus motif, more than 90% contained Glu at the +2 position, suggesting that ψKxE may be a better substrate for sumoylation than ψKxD. Glu and Asp were equally likely at the −2 position in the inverted consensus motifs containing peptides, and a hydrophobic residue at +1 was not overrepresented. Together, the forward and inverted consensus motifs accounted for more than 70% of SUMO2$^{T90K}$-modified peptides. The sequences of the remaining SUMO2$^{T90K}$-modified peptides showed no consensus except for the modest overrepresentation of Gly at position +2. Because Gly does not have a side chain, this may suggest that steric hindrance inhibits sumoylation at nonconsensus motifs.

Previously, we used MS to identify sumoylation sites on proteins that were sumoylated in vitro using wild-type recombinant SUMO2 and digested with trypsin (49). However, this approach is not conducive to the analysis of complex protein mixtures, such as those derived from purification of cell lysates. Here, we used the list of SUMO2$^{T90K}$-modified peptides to create a virtual branched peptide database to search for branched peptides resulting from the trypsin-mediated cleavage of sumoylated proteins. This approach enabled the annotation of previously unassigned MS2 spectra corresponding to proteins sumoylated with endogenous or overexpressed tagged wild-type SUMOs. Thus, this method could be used to search for sumoylation sites in existing proteomic data sets designed to address various biological questions.

Consistent with previous studies showing that groups of functionally related proteins are coordinately modified by SUMO (4, 6, 22, 50), we found evidence for group modification. Specifically, 85 sites distributed between 52 different zinc finger proteins, 78 sites on only 14 hnRNPs, and an extensive network of proteins involved in DNA replication, recombination, and repair, and cell cycle bearing many sites of SUMO modification. This supports the idea of protein group modification by SUMO, where the limited number of SUMO E3 ligases and proteases appear to work to direct SUMO to large protein complexes, which may be stabilized by multiple SUMO-SIM interactions (50).

Most proteomic studies to date have focused on SUMO2. However, the method we described here is broadly applicable to other SUMO paralogs and other Ubls. Thus, this method may be useful in addressing key questions about the biology and biochemistry of Ubls. For example, which sumoylation sites are shared among SUMO paralogs and which are unique, is there site specificity in the sumoylation response to stresses, and what are the sites of modification of other Ubls?

Fig. 5. Proteins with multiple SUMO2$^{T90K}$-modified sites were present in functionally related protein interaction networks. Merged display of four protein interaction networks of selected SUMO2$^{T90K}$-modified proteins involved in RNA posttranscriptional modification; DNA replication, recombination, and repair; gene expression; cell cycle; and cellular development. The size of the nodes is proportional to the number of identified sumoylation sites. Network information, protein details, and modification site numbers can be found in table S3.

Fig. 6. Using a virtual branched peptide database to identify wild-type sumoylation sites. (A) Bioinformatic approach for the analysis of MS2 spectra of tryptic branched peptides derived from proteins modified by wild-type SUMO1 or SUMO2. The list of SUMO2[T90K]-modified sites was used as a template to create a database of virtual branched peptides (11). Each virtual peptide consists of a C-terminal tryptic fragment of SUMO1 or SUMO2 joined to the N terminus of the tryptic peptide encompassing the sumoylated Lys. This database was used to annotate spectra in MS-based proteomics data. (B) Annotated MS2 spectra of a peptide of c-Myb modified at Lys[648] by SUMO2[T90K] (top) or by endogenous SUMO (bottom). (C) Annotated MS2 spectra of a peptide of ubiquitin modified at Lys[11] by SUMO2[T90K] (top) or by TAP-SUMO2 (bottom).



## MATERIALS AND METHODS

### Generation of HEK293 N3S cells stably expressing 6His-SUMO2[T90K]

pEFIRESpuro-6His-SUMO2 was created by cloning a polymerase chain reaction–generated 6His-SUMO2 fusion into the Nhe I and Not I sites of the plasmid vector pEFIRES-P-eYFP-C1 (51), replacing the coding sequence of eYFP (enhanced yellow fluorescent protein). The SUMO2[T90K] mutation was introduced into pEFIRESpuro-6His-SUMO2 by site-directed mutagenesis. The SUMO2 coding regions of all plasmids were fully sequenced. HEK293 N3S cells (Sigma-Aldrich, 92052131) grown in suspension culture were transfected using Lipofectamine 2000 (Life Technologies) with pEFIRESpuro-6His-SUMO2[T90K] and selected with puromycin at 2 μg/ml. Thereafter, stable cell populations were maintained in growth medium containing puromycin (1 μg/ml).

### Cell culture and protein extraction

HEK293 N3S cells stably expressing 6His-SUMO2[T90K] were cultured in Dulbecco's modified Eagle's medium (DMEM) (Gibco), supplemented with 10% dialyzed fetal calf serum (FCS), puromycin (1 μg/ml), and penicillin and streptomycin (100 U/ml). Cells were grown on five 175-cm² dishes to about 90% confluency before their transfer into Eagle's minimum essential medium (spinner modification; Sigma-Aldrich) supplemented with 10% FCS (1 μg/ml), 2 mM L-glutamine, and penicillin and streptomycin (100 U/ml). Three liters of cell culture (~2 × 10⁹ cells) was stimulated by heat shock at 43°C, harvested by centrifugation, washed twice with cold 1× DPBS
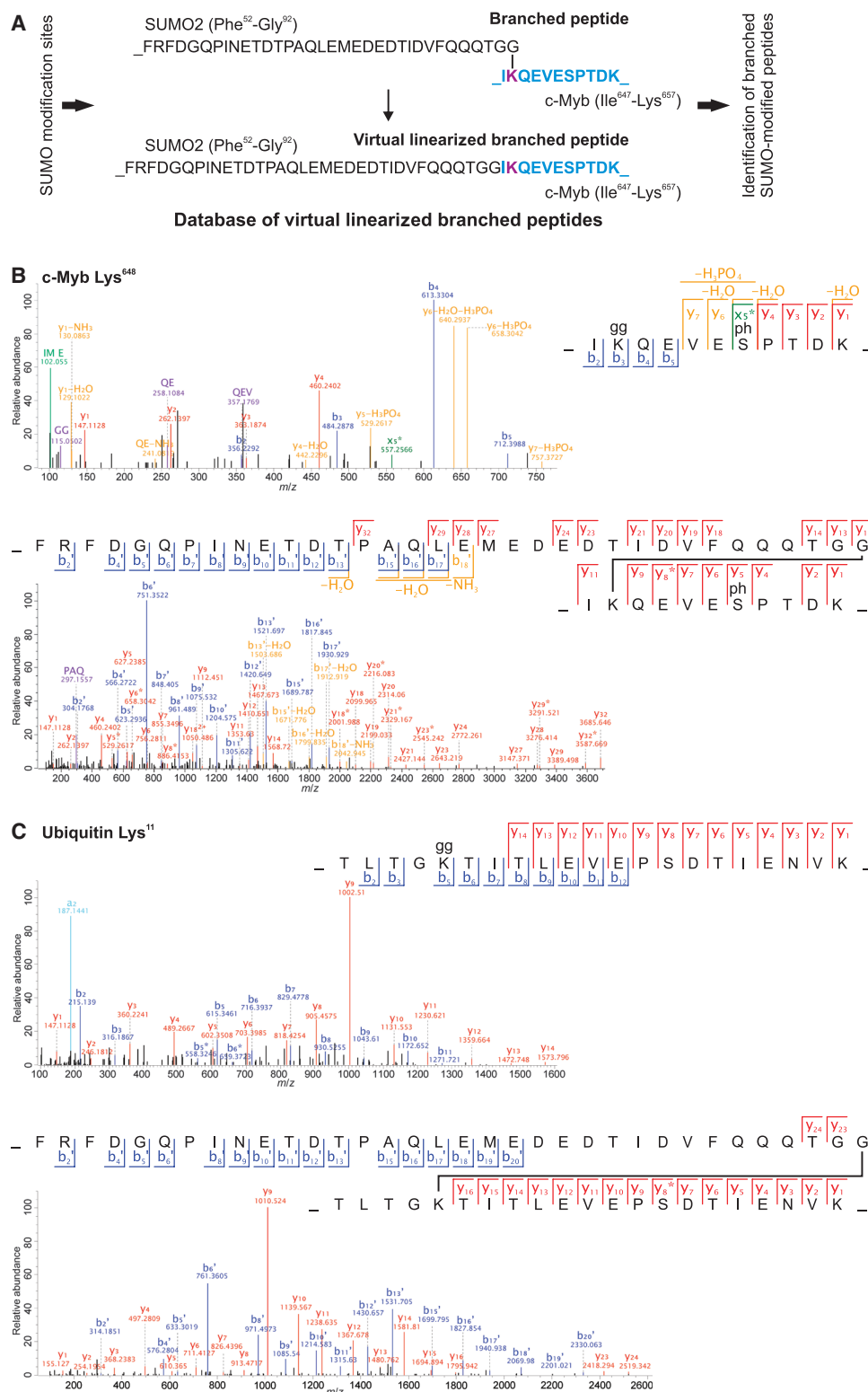
(Dulbecco's phosphate-buffered saline; Gibco), and lysed in cell lysis buffer [6 M guanidinium-HCl, 100 mM sodium phosphate buffer (pH 8.0), 10 mM tris-HCl (pH 8.0), 20 mM imidazole, 5 mM β-mercaptoethanol] (6 ml of lysis buffer per 1 g of cell pellet). DNA was disrupted by short

pulses of sonication, insoluble particles were removed with 0.2-μm sterile filters (Sartorius), and protein concentration was determined by BCA (bicinchoninic acid) assay (Pierce). In the experiment designed to evaluate reproducibility, duplicate 750-ml cultures of HEK293 6His-SUMO2$^{T90K}$ corresponding to $6 \times 10^8$ cells were processed in parallel.

## Nickel affinity chromatography

Nickel affinity purification of 6His-SUMO2$^{T90K}$ conjugates was performed with Ni$^{2+}$-NTA agarose beads (Qiagen) according to a published protocol (52) with minor changes. Ni$^{2+}$-NTA agarose beads (250 μl) were added to the cell lysate (286 mg of total protein for the primary experiment or 48 mg of protein for the experiment designed to evaluate reproducibility) and mixed at 4°C for 16 hours. Beads were washed with cell lysis buffer, wash buffer pH 8.0 [8 M urea, 100 mM sodium phosphate buffer (pH 8.0), 10 mM tris-HCl (pH 8.0), 10 mM imidazole, 5 mM β-mercaptoethanol], wash buffer pH 6.3 [8 M urea, 100 mM sodium phosphate buffer (pH 6.3), 10 mM tris-HCl (pH 8.0), 10 mM imidazole, 5 mM β-mercaptoethanol], and again with wash buffer pH 8.0. Proteins were eluted in three sequential steps with 1.5 column volumes of elution buffer [8 M urea, 100 mM sodium phosphate buffer (pH 8.0), 10 mM tris-HCl (pH 8.0), 200 mM imidazole, 5 mM β-mercaptoethanol].

## Filter-aided sample preparation and protein digestion

Digestion of 6His-SUMO2$^{T90K}$ proteins was performed on 30-kD cut-off filter units (Sartorius) according to a published protocol (53) with minor changes. Samples were concentrated on filter units, washed twice with UA buffer [200 μl of 8 M urea, 100 mM tris-HCl (pH 7.5)], and treated with 50 mM chloroacetamide in UA buffer for 20 min in the dark. Samples were then washed twice with UA buffer, three times with 200 μl of immunoprecipitation (IP) buffer [50 mM Mops-NaOH (pH 7.2), 10 mM Na$_2$HPO$_4$, 50 mM NaCl], and digested for 16 hours with Lys-C (Wako) in 50 μl of IP buffer at 37°C (enzyme-to-protein ratio, 1:50). Samples were collected and the filters were washed with 50 μl of IP buffer to increase the yield of Lys-C peptides. Peptides retained on the filter units were subsequently digested with endoproteinase Glu-C for 16 hours in 50 μl of IP buffer at 20°C (enzyme-to-protein ratio, 1:100), and after collection of peptides, the filter units were again washed with additional 50 μl of IP buffer. To analyze nickel affinity purifications directly, a small fraction of the nickel column elution (~2 μg) was diluted 10-fold in 8 M urea, 100 mM tris-HCl (pH 7.5), treated with 50 mM chloroacetamide for 1.5 hours at 20°C, then diluted four times with 50 mM ammonium bicarbonate, and mixed with Lys-C (enzyme-to-protein ratio, 1:50), followed by digestion at 20°C for 16 hours.

## Immunopurification of diGly-Lys–containing peptides

Anti-KεGG conjugated to protein A beads (20 μl of beads; PTMScan, Cell Signaling Technology) was washed three times with 1 ml of 50 mM sodium borate (pH 9.0), resuspended in 1 ml of fresh 20 mM dimethyl pimelimidate (DMP), and incubated for 30 min at room temperature while rotating. Beads were washed twice with 1 ml of 200 mM ethanolamine (pH 8.0) for 2 hours at 4°C, three times with 1.5 ml of cold IP buffer, and stored in IP buffer at 4°C. Enrichment of diGly-Lys–containing peptides with anti-KεGG was performed according to a published protocol (54) with minor changes. Anti-KεGG (19 μg) cross-linked to protein A beads (3 μl) was added to peptide mixtures (300 μg for the primary experiment or 360 μg for the experiment designed to evaluate reproducibility) in IP buffer and incubated at 4°C for overnight while rotating. Beads were washed twice with 500 μl of cold 1× DPBS, and peptides were eluted twice with 50 μl of 0.1% trifluoroacetic acid.

## MS analysis and data processing

Before MS analysis, all peptide samples were desalted on self-made reverse-phase C18 (Empore) Stop and Go Extraction Tips (55) and analyzed by liquid chromatography–tandem MS on a Q Exactive mass spectrometer (Thermo Scientific) coupled to an EASY-nLC 1000 Liquid Chromatography system (Thermo Scientific) via an EASY-Spray ion source (Thermo Scientific). Purified peptides were loaded onto 75-μm × 500-mm EASY-Spray column (Thermo Scientific) at a maximum pressure of 800 bars, and various gradient lengths from 90 to 150 min were used with a linear gradient of 5 to 22% of solvent B (100% acetonitrile, 0.1% formic acid) in solvent A (0.1% formic acid), followed by a ramp to 40% of solvent B. Flow rate was set to 250 nl/min, and eluting peptides were injected online into the mass spectrometer. Various Q Exactive settings were tested depending on the complexity of the samples; however, optimal data acquisition for low-complexity diGly-Lys–containing peptides was achieved with the following parameters: Precursor ion full-scan spectra [mass/charge ratio ($m/z$), 300 to 1600] were acquired with a resolution of 70,000 at $m/z$ 400 (target value of 1,000,000 ions, maximum injection time of 20 ms). Up to one data-dependent MS2 spectrum was acquired with a resolution of 35,000 at $m/z$ 400 (target value of 500,000 ions, maximum injection time of 1000 ms). Ions with unassigned charge state, and singly or highly (>8) charged ions were rejected. Intensity threshold was set to $2.1 \times 10^4$ U. Peptide match was set to preferred, and dynamic exclusion option was enabled (exclusion duration, 40 s).

## Data analysis and manual validation of the results

Raw MS data files were processed using MaxQuant software (version 1.3.0.5) (56, 57) and searched against UniProtKB human proteome (canonical and isoform sequences; downloaded in April 2013). Enzyme specificity was set to cleave peptide bonds C-terminally to Lys residues for samples treated with Lys-C only, or C-terminally to Glu, Asp, and Lys residues for samples digested with Lys-C and Glu-C. A maximum number of three or five missed cleavages were allowed for samples cleaved with Lys-C, or Lys-C and Glu-C, respectively. Carbamidomethylation of Cys was set as a fixed modification and oxidation of Met, acetylation of protein N termini, phosphorylation of Ser, Thr, and Tyr, and diGly adduction to Lys (except in peptide C terminus) were set as variable modifications. A minimum peptide length was set to seven amino acids, and a maximum peptide mass was 10,000 daltons. A false discovery rate (FDR) of 1% was set as a threshold at both protein and peptide level, and a mass deviation of 6 parts per million (ppm) was set for main search and 0.5 dalton for MS2 peaks.

MS2 spectra (filtered at 1% FDR) of sumoylated peptides were manually validated and annotated using MaxQuant viewer expert system (58) on the basis of the following criteria: (i) good coverage of y- and b-ion series, (ii) extensive identification rate of intensive fragment ion peaks, (iii) mass error less than 2 ppm after mass recalibration or 4 ppm in case of unsuccessful recalibration, and (iv) preferential fragmentation N-terminally to proline or C-terminally to Glu and Asp residues. All peptides identified from the reverse decoy database were removed, and the probability of site localization was checked to be greater than 75%. Existence of a diagnostic peak corresponding to a fragment ion of diGly residue (GG$^+$, $m/z$ +115.0505 Th) or presence of an ion series with a neutral loss corresponding to single Gly ($m/z$ 57.0215 Th) or an identification of various modified peptides corresponding to the same sumoylation site contributed to the confidence in assigned sumoylation sites. Finally, all accepted modified peptides corresponding to more than one protein were merged into one site identifier indicated in table S1. To investigate experimental reproducibility, a SUMO site was considered "identified" if it matched a peptide from the first list of 1002 modified peptides.

## In vitro SUMO conjugation, deconjugation, and processing assays

Recombinant substrate proteins used in these assays have been described previously (59). All reactions were buffered in 50 mM tris-HCl (pH 7.5). SUMO2 proform processing assays contained 150 mM NaCl, 0.5 mM TCEP, 600 µM SUMO, and 100 nM SENP1 or 200 nM SENP2 recombinant catalytic domains (59), and reactions were incubated at 20°C for 0, 5, 10, 20, 30, 60, 90, 120, 180, 240, or 960 min. Conjugation assays contained 5 mM dithiothreitol, 5 mM MgCl$_2$, 2 mM ATP, 110 nM SAE1 and SAE2, between 0.5 and 2 µM Ubc9, ~10 µM substrate protein, and a range of SUMO2 concentrations (0, 40, 80, or 200 µM) and were incubated at 37°C for 4 hours. Deconjugation assays were prepared as above using 200 µM SUMO followed by addition of SENP1 to 10 nM, and reactions were monitored at 0, 0.5, 1, 2.5, 5, and 10 min at 20°C.

## Immunoblot analysis

HEK293 N3S cells either stably expressing or not 6His-SUMO2[T90K] were grown on 75-cm$^2$ flasks in DMEM supplemented with 10% FCS, puromycin (1 µg/ml) (where required), and penicillin and streptomycin (100 U/ml). Where indicated, heat shock treatment at 43°C for 30 min was applied. Cells were washed twice with 1× DPBS and lysed in 4% SDS, 100 mM tris-HCl (pH 7.5). Immunoblots were prepared using mouse antibody specific to 6His (Clontech no. 631212) or rabbit antibody specific to SUMO2 (Zymed, 91-5100).

## TAP-SUMO2 sample preparation and MS analysis

TAP-SUMO2 protein samples remaining from a previous study (6) were fractionated by protein electrophoresis, the gel was cut into nine slices, and tryptic peptides were extracted, as described previously (60). Peptide mixtures were analyzed by MS as described with the following alterations: High-performance liquid chromatography fractionation occurred over a linear 115-min fractionation containing a gradient from 15 to 32% acetonitrile, with a loop count of 3, an $m/z$ precursor scan range of 800 to 1800, +3 to +7 inclusion charge range, $5 \times 10^5$ AGC (automatic gain control) target or 500-ms maximum injection time, and 35,000 at 400 $m/z$ resolution for MS2 scans.

## Virtual branched peptide database

Virtual branched peptide databases were created as a modification of the procedure described by Matic *et al.* (11). A python script was created to extract protein sequence information for each SUMO-modified protein to determine the tryptic peptide encompassing each SUMO target lysine. Cleavages at KP (lysine-proline) and RP (arginine-proline) were ignored, and protein N-terminal methionines were omitted. Each tryptic peptide was appended at the N terminus with either the tryptic fragment of SUMO1 C terminus (IADNHTPKELGMEEEDVIEVYQEQTGG) or that of SUMO2 C terminus (FRFDGQPINETDTPAQLEMEDEDTIDVF-QQQTGG), which include a single missed cleavage in this region of SUMO. Raw data files were searched against the databases (SUMO1. FASTA and SUMO2.FASTA) using MaxQuant version 1.3.0.5 (56, 57). Maximum peptide size was set to 10,000 daltons. A whole human proteome database was used for first search, and the branched peptide database(s) was used for main search. Missed cleavages were set to at least 3, and no FDR filtering was used at the protein or peptide level. All spectra were computationally (58) and manually validated.

## Bioinformatics analysis

Sequence analysis was performed with pLogo (36). Because not all identified peptides could be assigned to a single protein, multiple 13–amino acid sequences were used for input in these cases. N- or C-terminal sequences that did not cover the 13 residue window were omitted from the output. Residues were scaled relative to their Bonferroni-corrected statistical significance using human proteome as a background data set (645,531 lysines). Protein functional annotation and network analyses were created using Ingenuity Pathway Analysis (Ingenuity Systems, Qiagen).

## SUPPLEMENTARY MATERIALS

www.sciencesignaling.org/cgi/content/full/7/323/rs2/DC1
Fig. S1. Comparison of parental and 6His-SUMO2[T90K]–expressing HEK293 cells.
Fig. S2. Comparison of SUMO2 and SUMO2[T90K] in conjugation and deconjugation reactions with various substrates.
Fig. S3. Overlap of SUMO2[T90K] sites in replicate experiments.
Fig. S4. Overlap of sumoylated proteins with published protein level proteomic studies.
Table S1. Sumoylation sites identified in SUMO2[T90K]-expressing cells.
Table S2. diGly-modified peptides identified in SUMO2[T90K]-expressing cells.
Table S3. Functional annotation and network analysis of SUMO2[T90K]-expressing cells.
SMfile1.PDF. MS2 spectra of multiply sumoylated peptides identified from SUMO2[T90K]-expressing cells.
SMfile2.PDF. MS2 spectra of branched sumoylated peptides identified from SUMO2-expressing cells.
SUMO1.FASTA. SUMO1 virtual branched peptide database.
SUMO2.FASTA. SUMO2 virtual branched peptide database.

## REFERENCES AND NOTES

1. R. Geiss-Friedlander, F. Melchior, Concepts in sumoylation: A decade on. *Nat. Rev. Mol. Cell Biol.* **8**, 947–956 (2007).
2. K. Nacerddine, F. Lehembre, M. Bhaumik, J. Artus, M. Cohen-Tannoudji, C. Babinet, P. P. Pandolfi, A. Dejean, The SUMO pathway is essential for nuclear integrity and chromosome segregation in mice. *Dev. Cell* **9**, 769–779 (2005).
3. H. Saitoh, J. Hinchey, Functional heterogeneity of small ubiquitin-related protein modifiers SUMO-1 versus SUMO-2/3. *J. Biol. Chem.* **275**, 6252–6258 (2000).
4. F. Golebiowski, I. Matic, M. H. Tatham, C. Cole, Y. Yin, A. Nakamura, J. Cox, G. J. Barton, M. Mann, R. T. Hay, System-wide changes to SUMO modifications in response to heat shock. *Sci. Signal.* **2**, ra24 (2009).
5. J. Schimmel, K. M. Larsen, I. Matic, M. van Hagen, J. Cox, M. Mann, J. S. Andersen, A. C. Vertegaal, The ubiquitin-proteasome system is a key component of the SUMO-2/3 cycle. *Mol. Cell. Proteomics* **7**, 2107–2122 (2008).
6. M. H. Tatham, I. Matic, M. Mann, R. T. Hay, Comparative proteomic analysis identifies a role for SUMO in protein quality control. *Sci. Signal.* **4**, rs4 (2011).
7. Y. Galanty, R. Belotserkovskaya, J. Coates, S. Polo, K. M. Miller, S. P. Jackson, Mammalian SUMO E3-ligases PIAS1 and PIAS4 promote responses to DNA double-strand breaks. *Nature* **462**, 935–939 (2009).
8. J. R. Morris, C. Boutell, M. Keppler, R. Densham, D. Weekes, A. Alamshah, L. Butler, Y. Galanty, L. Pangon, T. Kiuchi, T. Ng, E. Solomon, The SUMO modification pathway is involved in the BRCA1 response to genotoxic stress. *Nature* **462**, 886–890 (2009).
9. R. T. Hay, SUMO: A history of modification. *Mol. Cell* **18**, 1–12 (2005).
10. M. H. Tatham, E. Jaffray, O. A. Vaughan, J. M. Desterro, C. H. Botting, J. H. Naismith, R. T. Hay, Polymeric chains of SUMO-2 and SUMO-3 are conjugated to protein substrates by SAE1/SAE2 and Ubc9. *J. Biol. Chem.* **276**, 35368–35374 (2001).
11. I. Matic, M. van Hagen, J. Schimmel, B. Macek, S. C. Ogg, M. H. Tatham, R. T. Hay, A. I. Lamond, M. Mann, A. C. Vertegaal, In vivo identification of human small ubiquitin-like modifier polymerization sites by high accuracy mass spectrometry and an in vitro to in vivo strategy. *Mol. Cell. Proteomics* **7**, 132–144 (2008).
12. M. H. Tatham, M. C. Geoffroy, L. Shen, A. Plechanovova, N. Hattersley, E. G. Jaffray, J. J. Palvimo, R. T. Hay, RNF4 is a poly-SUMO-specific E3 ubiquitin ligase required for arsenic-induced PML degradation. *Nat. Cell Biol.* **10**, 538–546 (2008).
13. V. Lallemand-Breitenbach, M. Jeanne, S. Benhenda, R. Nasr, M. Lei, L. Peres, J. Zhou, J. Zhu, B. Raught, H. de Thé, Arsenic degrades PML or PML–RARα through a SUMO-triggered RNF4/ubiquitin-mediated pathway. *Nat. Cell Biol.* **10**, 547–555 (2008).
14. A. Pichler, A. Gast, J. S. Seeler, A. Dejean, F. Melchior, The nucleoporin RanBP2 has SUMO1 E3 ligase activity. *Cell* **108**, 109–120 (2002).
15. D. Reverter, C. D. Lima, Insights into E3 ligase activity revealed by a SUMO–RanGAP1–Ubc9–Nup358 complex. *Nature* **435**, 687–692 (2005).
16. R. T. Hay, SUMO-specific proteases: A twist in the tail. *Trends Cell Biol.* **17**, 370–376 (2007).
17. G. Filosa, S. M. Barabino, A. Bachi, Proteomics strategies to identify SUMO targets and acceptor sites: A survey of RNA-binding proteins SUMOylation. *Neuromolecular Med.* **15**, 661–676 (2013).
18. M. Tirard, H. H. Hsiao, M. Nikolov, H. Urlaub, F. Melchior, N. Brose, In vivo localization and identification of SUMOylated proteins in the brain of His$_6$-HA-SUMO1 knock-in mice. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 21122–21127 (2012).

19. A. C. Vertegaal, J. S. Andersen, S. C. Ogg, R. T. Hay, M. Mann, A. I. Lamond, Distinct and overlapping sets of SUMO-1 and SUMO-2 target proteins revealed by quantitative proteomics. *Mol. Cell. Proteomics* 5, 2298–2310 (2006).

20. V. Matafora, A. D'Amato, S. Mori, F. Blasi, A. Bachi, Proteomics analysis of nucleolar SUMO-1 target proteins upon proteasome inhibition. *Mol. Cell. Proteomics* 8, 2243–2255 (2009).

21. J. Becker, S. V. Barysch, S. Karaca, C. Dittner, H. H. Hsiao, M. Berriel Diaz, S. Herzig, H. Urlaub, F. Melchior, Detecting endogenous SUMO targets in mammalian cells and tissues. *Nat. Struct. Mol. Biol.* 20, 525–531 (2013).

22. R. Bruderer, M. H. Tatham, A. Plechanovova, I. Matic, A. K. Garg, R. T. Hay, Purification and identification of endogenous polySUMO conjugates. *EMBO Rep.* 12, 142–148 (2011).

23. C. Choudhary, C. Kumar, F. Gnad, M. L. Nielsen, M. Rehman, T. C. Walther, J. V. Olsen, M. Mann, Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* 325, 834–840 (2009).

24. J. V. Olsen, B. Blagoev, F. Gnad, B. Macek, C. Kumar, P. Mortensen, M. Mann, Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* 127, 635–648 (2006).

25. G. Xu, J. S. Paige, S. R. Jaffrey, Global analysis of lysine ubiquitination by ubiquitin remnant immunoaffinity profiling. *Nat. Biotechnol.* 28, 868–873 (2010).

26. W. Kim, E. J. Bennett, E. L. Huttlin, A. Guo, J. Li, A. Possemato, M. E. Sowa, R. Rad, J. Rush, M. J. Comb, J. W. Harper, S. P. Gygi, Systematic and quantitative assessment of the ubiquitin-modified proteome. *Mol. Cell* 44, 325–340 (2011).

27. S. A. Wagner, P. Beli, B. T. Weinert, M. L. Nielsen, J. Cox, M. Mann, C. Choudhary, A proteome-wide, quantitative survey of in vivo ubiquitylation sites reveals widespread regulatory roles. *Mol. Cell. Proteomics* 10, M111 013284 (2011).

28. P. G. Pedrioli, B. Raught, X. D. Zhang, R. Rogers, J. Aitchison, M. Matunis, R. Aebersold, Automated identification of SUMOylation sites using mass spectrometry and SUMmOn pattern recognition software. *Nat. Methods* 3, 533–539 (2006).

29. I. Matic, J. Schimmel, I. A. Hendriks, M. A. van Santen, F. van de Rijke, H. van Dam, F. Gnad, M. Mann, A. C. Vertegaal, 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 (2010).

30. H. H. Hsiao, E. Meulmeester, B. T. Frank, F. Melchior, H. Urlaub, "ChopNSpice" a mass spectrometric approach that allows identification of endogenous small ubiquitin-like modifier-conjugated peptides. *Mol. Cell. Proteomics* 8, 2664–2675 (2009).

31. T. Makhnevych, Y. Sydorskyy, X. Xin, T. Srikumar, F. J. Vizeacoumar, S. M. Jeram, Z. Li, S. Bahr, B. J. Andrews, C. Boone, B. Raught, Global map of SUMO function revealed by protein-protein interaction and genetic networks. *Mol. Cell* 33, 124–135 (2009).

32. M. Knuesel, H. T. Cheung, M. Hamady, K. K. Barthel, X. Liu, 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 (2005).

33. H. A. Blomster, S. Y. Imanishi, J. Siimes, J. Kastu, N. A. Morrice, J. E. Eriksson, L. Sistonen, In vivo identification of sumoylation sites by a signature tag and cysteine-targeted affinity purification. *J. Biol. Chem.* 285, 19324–19329 (2010).

34. P. V. Hornbeck, J. M. Kornhauser, S. Tkachev, B. Zhang, E. Skrzypek, B. Murray, V. Latham, M. Sullivan, 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 (2012).

35. M. S. Rodriguez, C. Dargemont, R. T. Hay, SUMO-1 conjugation in vivo requires both a consensus modification motif and nuclear targeting. *J. Biol. Chem.* 276, 12654–12659 (2001).

36. J. P. O'Shea, M. F. Chou, S. A. Quader, J. K. Ryan, G. M. Church, D. Schwartz, pLogo: A probabilistic approach to visualizing sequence motifs. *Nat. Methods* 10, 1211–1212 (2013).

37. J. M. Desterro, M. S. Rodriguez, G. D. Kemp, R. T. Hay, Identification of the enzyme required for activation of the small ubiquitin-like protein SUMO-1. *J. Biol. Chem.* 274, 10618–10624 (1999).

38. E. S. Johnson, G. Blobel, Ubc9p is the conjugating enzyme for the ubiquitin-like protein Smt3p. *J. Biol. Chem.* 272, 26799–26802 (1997).

39. U. Karvonen, T. Jääskeläinen, M. Rytinki, S. Kaikkonen, J. J. Palvimo, ZNF451 is a novel PML body- and SUMO-associated transcriptional coregulator. *J. Mol. Biol.* 382, 585–600 (2008).

40. S. Chupreta, H. Brevig, L. Bai, J. L. Merchant, J. A. Iñiguez-Lluhí, Sumoylation-dependent control of homotypic and heterotypic synergy by the Kruppel-type zinc finger protein ZBP-89. *J. Biol. Chem.* 282, 36155–36166 (2007).

41. M. T. Vassileva, M. J. Matunis, SUMO modification of heterogeneous nuclear ribonucleoproteins. *Mol. Cell. Biol.* 24, 3623–3632 (2004).

42. P. Mertins, J. W. Qiao, J. Patel, N. D. Udeshi, K. R. Clauser, D. R. Mani, M. W. Burgess, M. A. Gillette, J. D. Jaffe, S. A. Carr, Integrated proteomic analysis of post-translational modifications by serial enrichment. *Nat. Methods* 10, 634–637 (2013).

43. Ø. Dahle, T. Ø. Andersen, O. Nordgård, V. Matre, G. Del Sal, O. S. Gabrielsen, Transactivation properties of c-Myb are critically dependent on two SUMO-1 acceptor sites that are conjugated in a PIASy enhanced manner. *Eur. J. Biochem.* 270, 1338–1348 (2003).

44. J. Bies, J. Markus, L. Wolff, Covalent attachment of the SUMO-1 protein to the negative regulatory domain of the c-Myb transcription factor modifies its stability and transactivation capacity. *J. Biol. Chem.* 277, 8999–9009 (2002).

45. C. Weber, T. B. Schreiber, H. Daub, Dual phosphoproteomics and chemical proteomics analysis of erlotinib and gefitinib interference in acute myeloid leukemia cells. *J. Proteomics* 75, 1343–1356 (2012).

46. C. Choudhary, M. Mann, Decoding signalling networks by mass spectrometry-based proteomics. *Nat. Rev. Mol. Cell Biol.* 11, 427–439 (2010).

47. E. Pani, M. Menigatti, S. Schubert, D. Hess, B. Gerrits, K. H. Klempnauer, S. Ferrari, Pin1 interacts with c-Myb in a phosphorylation-dependent manner and regulates its transactivation activity. *Biochim. Biophys. Acta* 1783, 1121–1128 (2008).

48. H. A. Blomster, V. Hietakangas, J. Wu, P. Kouvonen, S. Hautaniemi, L. Sistonen, Novel proteomics strategy brings insight into the prevalence of SUMO-2 target sites. *Mol. Cell. Proteomics* 8, 1382–1390 (2009).

49. H. J. Cooper, M. H. Tatham, E. Jaffray, J. K. Heath, T. T. Lam, A. G. Marshall, R. T. Hay, Fourier transform ion cyclotron resonance mass spectrometry for the analysis of small ubiquitin-like modifier (SUMO) modification: Identification of lysines in RanBP2 and SUMO targeted for modification during the E3 autoSUMOylation reaction. *Anal. Chem.* 77, 6310–6319 (2005).

50. I. Psakhye, S. Jentsch, Protein group modification and synergy in the SUMO pathway as exemplified in DNA repair. *Cell* 151, 807–820 (2012).

51. S. Hobbs, S. Jitrapakdee, J. C. Wallace, Development of a bicistronic vector driven by the human polypeptide chain elongation factor 1α promoter for creation of stable mammalian cell lines that express very high levels of recombinant proteins. *Biochem. Biophys. Res. Commun.* 252, 368–372 (1998).

52. M. H. Tatham, M. S. Rodriguez, D. P. Xirodimas, R. T. Hay, Detection of protein SUMOylation in vivo. *Nat. Protoc.* 4, 1363–1371 (2009).

53. J. R. Wiśniewski, A. Zougman, N. Nagaraj, M. Mann, Universal sample preparation method for proteome analysis. *Nat. Methods* 6, 359–362 (2009).

54. N. D. Udeshi, P. Mertins, T. Svinkina, S. A. Carr, Large-scale identification of ubiquitination sites by mass spectrometry. *Nat. Protoc.* 8, 1950–1960 (2013).

55. J. Rappsilber, M. Mann, Y. Ishihama, Protocol for micro-purification, enrichment, prefractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* 2, 1896–1906 (2007).

56. J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* 26, 1367–1372 (2008).

57. J. Cox, N. Neuhauser, A. Michalski, R. A. Scheltema, J. V. Olsen, M. Mann, Andromeda: A peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* 10, 1794–1805 (2011).

58. N. Neuhauser, A. Michalski, J. Cox, M. Mann, Expert system for computer-assisted annotation of MS/MS spectra. *Mol. Cell. Proteomics* 11, 1500–1509 (2012).

59. L. Shen, M. H. Tatham, C. Dong, A. Zagórska, J. H. Naismith, R. T. Hay, SUMO protease SENP1 induces isomerization of the scissile peptide bond. *Nat. Struct. Mol. Biol.* 13, 1069–1077 (2006).

60. A. Shevchenko, H. Tomas, J. Havlis, J. V. Olsen, M. Mann, In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* 1, 2856–2860 (2006).