Small ubiquitin-like modifier (SUMO) is covalently conjugated to its target proteins thereby altering their activity. The mammalian SUMO protein family includes four members (SUMO-1–4) of which SUMO-2 and SUMO-3 are conjugated in a stress-inducible manner. The vast majority of known SUMO substrates are recognized by the single SUMO E2-conjugating enzyme Ubc9 binding to a consensus tetrapeptide (ΨKXE where Ψ stands for a large hydrophobic amino acid) or extended motifs that contain phosphorylated or negatively charged amino acids called PDSM (phosphorylation-dependent sumoylation motif) and NDSM (negatively charged amino acid-dependent sumoylation motif), respectively. We identified 382 SUMO-2 targets using a novel method based on SUMO protease treatment that improves separation of SUMO substrates on SDS-PAGE before LC-ESI-MS/MS. We also implemented a software SUMOFI (SUMO motif finder) to facilitate identification of motifs for SUMO substrates from a user-provided set of proteins and to classify the substrates according to the type of SUMO-targeting consensus site. Surprisingly more than half of the substrates lacked any known consensus site, suggesting that numerous SUMO substrates are recognized by a yet unknown consensus site-independent mechanism. Gene ontology analysis revealed that substrates in distinct functional categories display strikingly different prevalences of NDSM sites. Given that different types of motifs are bound by Ubc9 using alternative mechanisms, our data suggest that the preference of SUMO-2 targeting mechanism depends on the biological function of the substrate. Molecular & Cellular Proteomics 8:1382–1390, 2009.

Small ubiquitin-like modifier (SUMO) is a ubiquitin-like protein covalently conjugated to substrate proteins with a broad spectrum of cellular functions (1). Sumoylation is an essential process, and mice lacking sumoylation die at the early postimplantation stage (2). The biological outcome of sumoylation depends on the substrate and includes alterations in subcellular localization, protein–protein interactions, stability, and transcriptional activity (1, 3–5). Mammalian SUMO-2 and SUMO-3 are almost identical, whereas SUMO-1 shares only 50% sequence identity with the other SUMO proteins. SUMO-1 and SUMO-2/3 also vary functionally; they are associated with distinct but overlapping sets of substrate and effector proteins. Moreover SUMO-2/3 proteins are able to form chains; SUMO-1 cannot (6–8). In contrast to SUMO-1, SUMO-2/3 conjugation increases in response to heat shock and other protein-damaging stress (9). Despite these differences, SUMO-1 functions can be substituted by SUMO-2/3 in vivo (10).

All SUMO proteins are conjugated to the target lysines by a common enzymatic cascade composed of three enzymes: E1, E2, and E3. SUMO conjugation starts by formation of a thioester bond with the activating enzyme E1, a heterodimer of Aos1 and Uba2. Aos1/Uba2 transfers SUMO to the single E2-conjugating enzyme Ubc9, which recognizes and sumoylates the substrate. The Ubc9 substrate recognition is facilitated by specific E3 SUMO ligases (1). Sumoylation is a dynamic and reversible process, and SUMO can be rapidly deconjugated by isopeptidases (11).

The single E2 SUMO-conjugating enzyme Ubc9 binds to different consensus sequences through alternative mechanisms. In most cases, Ubc9 recognizes the substrate by binding to the SUMO consensus tetrapeptide ΨKXE where Ψ stands for a large hydrophobic amino acid and K is the target lysine (12). The ΨKXE tetrapeptide is directly recognized by the active site of Ubc9, and there are examples in which the amino acids immediately preceding or following this motif are not in contact with Ubc9 (13). However, not all consensus sites are sumoylated, and several SUMO substrates have additional specificity determinants that strengthen binding with Ubc9 (14). For example, in the negatively charged amino acid-dependent sumoylation motif (NDSM), acidic amino acid.
Prevalence of Different SUMO-2 Conjugation Sites

ids downstream of the consensus tetrapeptide are essential for SUMO conjugation and bind to a positively charged patch on Ubc9 (15). Another extended consensus motif, the phosphorylation-dependent sumoylation motif (PDSM) consists of the YKXE tetrapeptide followed by a proline-directed phosphorylation site and cannot be sumoylated unless phosphorylated (16). The PDSM and NDSM may represent variations on the same theme; the phosphorylated serine can be substituted for by acidic residues (14). Notably sumoylation has also been reported to occur on non-consensus lysines indicating that the consensus tetrapeptide is not an absolute requirement for sumoylation (17, 18). However, the proteins where SUMO is conjugated on non-consensus sites are not readily found because the presence of either the consensus tetrapeptide or an extended motif is often used as a prediction tool for SUMO substrate identification.

To date, a few proteomics studies of SUMO-2 or SUMO-3 substrates have been reported. Vertegaal et al. (19) used a nuclear body enrichment procedure followed by His purification, in-solution digestion, separation of peptides by HPLC, and MS/MS. Rosas-Acosta et al. (20) used a combination of proteasome inhibition and heat shock treatment followed by tandem affinity purification, enterokinase elution, in-solution digestion, HPLC, and MALDI MS/MS to identify SUMO-1 and SUMO-3 targets. Two studies based on stable isotope labeling by amino acids in cell culture (SILAC) have been published: Vertegaal et al. (7) generated His-tagged cell lines and nuclear lysates followed by SDS-PAGE and LC-MS/MS, whereas Schimmel et al. (21) used HeLa cells overexpressing His-SUMO-2, proteasome inhibition, whole cell lysates, and in-solution digestion followed by LC-MS/MS.

To improve the sensitivity of SUMO substrate identification, we developed a novel strategy for SUMO-2 substrate purification and identification. The novelty lies within the desumoylation of substrates prior to SDS-PAGE using a recombinant SUMO protease, which considerably reduces the complexity of the SDS-PAGE-separated sample and increases the number of identified substrates. Bioinformatics analysis of the 382 identified SUMO-2 substrates revealed that the SUMO consensus site is significantly enriched within the SUMO-2 proteome. However, 52% of the identified proteins lacked any known consensus sequence, suggesting that SUMO conjugation to non-consensus sites is more common than previously anticipated. In addition, many substrates containing NDSM, but not PDSM, were present in our data set, indicating that these extended consensus motifs are differentially targeted by stress-induced SUMO-2 conjugation. To further characterize biological functions of the identified proteins, we did Gene Ontology (GO) (22) term enrichment analysis. Interestingly GO categories of DNA- and nucleotide-related functions were selectively enriched in NDSM-containing SUMO-2 substrates. Given that different types of consensus sites are bound by Ubc9 using alternative mechanisms (13, 14, 16), our results reveal an unexpected connection between the SUMO-2 targeting mechanism and the biological function of the substrate.

EXPERIMENTAL PROCEDURES

Plasmid Constructs—The Myc-PARP-1 and pSG5-His-HA-SUMO-2 plasmids have been described earlier (16, 23).

Cell Culture and Transfection—Human K562 erythroleukemia cells were maintained and transfected as described earlier (24). To generate a stable cell line overexpressing His-hemagglutinin (HA)-tagged SUMO-2 (K562HA-SUMO-2), K562 cells were cotransfected with 30 μg of pSG5-His-HA-SUMO-2 and 3 μg of empty pcDNA1.1 by electroporation (975 microfarads, 220 V) using a Bio-Rad Gene Pulser electroporator. Transfected cells were allowed to recover for 2 days, and neomycin-resistant cells were selected in medium containing G418 (500 μg/ml; Invitrogen) for 2 weeks. Drug-resistant cells were diluted and selected for single cell clones. The K562HA-SUMO-2 cells were routinely maintained in medium containing G418 (500 μg/ml). Human HeLa cervical cancer cells were grown in Dulbecco’s modified Eagle’s medium supplemented with 10% fetal calf serum and antibiotics (penicillin and streptomycin) in a humidified 5% CO2 atmosphere at 37 °C. HeLa cells were transfected by electroporation as described earlier (24).

Heat Shock Treatments—Cells were heat-shocked for the indicated times in water baths so that the temperature of the medium was 43 °C. HeLa cells were heat-shocked in plastic bags sealed with Heat Sealer (Wallac). K562 cells were heat-shocked in cell culture bottles (Falcon).

Western Blotting—For Western blotting, cell pellets were first suspended in 1 volume of PBS. The suspension of cells was lysed by boiling in 2 volumes of Laemmli sample buffer. The lysates were separated by 8% SDS-PAGE, transferred to a nitrocellulose membrane, and blotted with α-HA antibody (Covance).

Immunoprecipitation—For large scale immunoprecipitation, cells were suspended in PBS and lysed by boiling in 1 volume of 1% SDS in PBS. Lysates were mixed with 10 volumes of 1% Triton X-100 in PBS, sonicated, and cleared by centrifugation at 18,000 rpm for 10 min. Cleared lysates were mixed with 9 volumes of 1% BSA and 1% Triton X-100 in PBS. N-Ethylmaleimide was added to a final concentration of 20 mm. Lysates were precleared with IgG-Sepharose (Amersham Biosciences) and incubated with α-HA-agarose (Sigma) for 2 h at room temperature. Beads were washed intensively, suspended in high salt protease buffer (Invitrogen), and resuspended in low salt protease buffer (Invitrogen) before treating with recombinant Ulp-1 SUMO protease (Invitrogen) at 37 °C for 1.5 h and boiling in Laemmli sample buffer. A similar protocol was used for small scale immunoprecipitations but without SUMO protease treatment unless otherwise indicated. α-Myc antibodies (Sigma) were used together with protein G beads (GE Healthcare) to immunoprecipitate Myc-tagged poly(ADP-ribose) polymerase 1 (PARP-1). Immunoprecipitated proteins were separated by SDS-PAGE and either silver-stained or transferred to a nitrocellulose membrane and blotted with α-topoisomerase I (Santa Cruz Biotechnology), α-DDX21 (25), α-PARP-1 (Sigma), α-Myc (Sigma), or α-SUMO-2/3 (Zymed Laboratories Inc.) antibodies.

Analysis of SUMO Protease Activity—Lysates from heat-shocked K562HA-SUMO-2 cells were prepared and immunoprecipitated with α-HA-agarose (Sigma) as described above. Equal amounts of immunoprecipitate-containing beads were treated with or without recombinant Ulp-1 SUMO protease (Invitrogen) in low salt SUMO protease buffer (Invitrogen) at 37 °C for 1.5 h. The beads were repeatedly washed with PBS and boiled in Laemmli sample buffer. Non-deconjugated SUMO substrates were analyzed by Western blotting with α-HA antibodies (Covance).

Nano-LC/ESI-MS/MS Analysis—Immunoprecipitated and deSUMOylated samples from 1 × 108 K562HA-SUMO-2 or parental K562
cells were separated by SDS-PAGE, and the gel was silver-stained. The lane from heat-treated K562HA-SUMO-2 cells was cut in 1-mm slices, and each slice was washed twice with 0.2 M NH4HCO3, 10% ACN and dehydrated with ACN. Proteins were reduced and alkylated by treating them with 20 mM DTT and 55 mM iodoacetamide, respectively. After washing with 50 mM NH4HCO3 and ACN, proteins were digested in gel with trypsin (Promega, Madison, WI) and incubated overnight at 37 °C. Tryptic peptides were extracted from the gel pieces with 50% ACN, 5% HCOOH. The peptide extracts were vacuum centrifuged to dryness and stored at −20 °C until analyzed by mass spectrometry.

Dried peptides were dissolved to 10 μl of 0.1% formic acid (FA) and analyzed by automated nanoscale capillary LC-MS/MS using an Ultimate capillary LC system, Famos autosampler, and Switchos II column switching unit (LC Packings). The LC system was coupled to a quadrupole TOF mass spectrometer (Q-Star Pulsar, Applied Biosystems/MDS Sciex, Toronto, Canada). Samples were desalted and preconcentrated on line with a 0.3 × 5-mm PepMap C18 μ-precolumn (LC Packings). Peptide separation was achieved by using a 75-μm × 150-mm PepMap C18 (3 μm, 100 Å) analytical column (LC Packings) and a two-stage gradient. First mobile phase B concentration was raised from 5% to 20% in 5 min and then from 20 to 60% in 20 min using a flow rate of 200 nl/min. Mobile phase A was a mixture of 0.1% FA and 5% ACN. Mobile phase B consisted of 0.1% FA and 95% ACN. The mass spectrometer was programmed to accumulate signal for 1 s for the TOF-MS scan after which the two most intense peaks were selected for 3-s product ion scans. Analyst QS (Applied Biosystems) was used for instrument control.

The data from different gel slices were combined and analyzed using Analyst QS 1.1 (Applied Biosystems) and Mascot in-house server and Daemon (2.2.0; Matrix Science). Raw data files from the Q-Star Pulsar were converted to peak lists for MASCOT searches using the following data import filter options in Daemon: (a) default precursor charge stages 2+ and 3+; (b) MS-MS averaging: reject spectra if less than five peaks or precursor charge <10 or precursor >100,000; precursor mass tolerance for grouping, 0.2; maximum number of cycles between groups, 10; and minimum number of cycles per group, 1; (c) information-dependent acquisition survey scan centroid parameters: redetermine precursor charge; percent height, 50%; and merge distance, 0.01 amu; and (d) MS-MS data centroid/threshold: remove peaks within intensity less than 0.01% of highest; centroid all MS/MS data; percent height, 50%; and merge distance, 0.1 amu.

The data were searched against the Swiss-Prot database release 54.0, which has 276,256 sequences in total and 16,891 sequences after taxonomy filter. The following search criteria were used: taxonomy, Homo sapiens; mass error tolerance for parent ion and fragment ions, 0.3 and 0.3 Da, respectively; fixed modification, cysteine carbamidomethylation; variable modification, methionine oxidation; enzyme, trypsin; and number of missed cleavages, one. The same settings were used when the MASCOT 2.2.0 in-build feature was used to search the data set against the decoy data base (26). This setting gave a false discovery rate of 2.6%. In the MASCOT peptide summary report the settings were as follows: significance threshold, p < 0.05; require bold red; and expect cutoff, 0.05. All the proteins that were identified only with one peptide were manually removed from the data set. Keratins were removed as common contaminants.

**GO Enrichment Analysis**—GO is a widely used controlled vocabulary to describe gene and gene product attributes that comprises over 26,000 terms in three major categories: biological process, molecular function, and cellular component (22, 27). Identification of statistically enriched GO terms for a given group of SUMO substrates was done by comparing GO term frequencies in the substrate group against the background. GO annotations for each protein were fetched from Ensembl through the BioMart system (28). GO terms are structured as a directed acyclic graph so that the nodes deeper in the tree are more specific than the nodes close to the root node. This hierarchy was taken into account when counting the number of annotated proteins as follows. A protein was associated with a certain GO term if it was annotated either with a term itself or a child of the term. Fisher’s exact test was performed to derive a p value for each GO term in a given group of substrates (e.g. NDSM) using the other group (e.g. non-consensus) or the complete proteome as background. Fisher’s exact test uses a 2 × 2 contingency table to evaluate the significance of the association between two variables (29). To correct for multiple hypothesis comparisons we used Holm’s method (30) to adjust p values when using the complete proteome as the background distribution. To show only the most relevant GO terms, parent terms having a child term with a lower p value were excluded from the final result list.

**RESULTS**

**A Novel Proteomics Strategy to Identify SUMO-2 Substrates**—We designed a proteomics-based approach for identification of SUMO-2 targets (Fig. 1A). Our purification strategy was based on denaturing cell lysis and immunopurification followed by SUMO-2 deconjugation using a SUMO protease prior to separation by SDS-PAGE and mass spectrometric identification. Sumoylation was induced by heat shock to maximize the number of identified SUMO-2 substrates (9). Sufficient quantities of cell lysate for substrate identification were produced by human K562 erythroleukemia cells stably expressing His-HA-tagged SUMO-2 (K562HA-SUMO-2). The ectopically expressed SUMO-2 behaved similarly to the endogenous SUMO-2/3 proteins because it was largely unmodified in untreated cells and efficiently conjugated to substrates upon heat shock (Fig. 1B). Denaturing cell lysis is an important step for restoring the SUMO-2 conjugates because it results in rapid inactivation of endogenous SUMO proteases. Lysates need to be renatured before immunoprecipitation of SUMO-2 substrates.

The novelty of our method lies within the *in vitro* SUMO deconjugation prior to SDS-PAGE, which allows concentration of substrates containing several sumoylation sites or SUMO chains to single bands on the SDS-PAGE gel (Fig. 1C). The SUMO deconjugation step reduced complexity of the sample and increased the number of identified proteins (data not shown). We chose a gel-based procedure because it allowed us to efficiently separate large quantities of purified proteins, which could subsequently be easily visualized and compared with mock and control samples. For the desumoylation reaction, we used recombinant yeast Ulp1, which can efficiently deconjugate SUMO-2 from HA beads as demonstrated by disappearance of the high molecular weight SUMO-2 conjugates (Fig. 1D).

**Identification of 382 SUMO-2 Substrates**—Immunoprecipitated proteins from the K562HA-SUMO-2 cells and parental K562 cells were prepared for mass spectrometric analysis. After desumoylation, the proteins were separated by SDS-PAGE and visualized by silver staining. The heat-treated K562HA-SUMO-2 sample showed strong staining throughout the molecular
weight range. In contrast, samples from parental K562 cells and untreated K562HA-SUMO-2 cells showed, except for the strong IgG band, only a weak staining indicative of specific immunoprecipitation (Fig. 2A). The heat-shocked K562HA-SUMO-2 lane was cut in slices that were treated with trypsin. After digestion tryptic peptides were extracted from the gel pieces and analyzed by LC-MS/MS. Data collected from individual gel bands were merged and searched against a database using stringent criteria; i.e. at least two identified peptides with a significance threshold of $p < 0.05$ were required for a positive protein hit. We used a decoy database method that gave a false discovery rate of 2.6%. The mass spectrometric analysis of the heat-treated K562HA-SUMO-2 lane led to the identification of 382 SUMO-2 substrates (supplemental Table 1).

Based on a literature survey in PubMed, 115 of the identified substrates were novel as they were not previously reported to be conjugated by any of the four mammalian SUMO paralogues (7, 19–21, 31–45). We also analyzed several gel slices from heat-shocked parental K562 and untreated K562HA-SUMO-2 cells. These samples contained no hits exceeding our threshold, likely reflecting the low stoichiometry of SUMO-2 conjugation in unstressed cells (Fig. 1B).

The purification was validated by Western blotting of the immunoprecipitated and desumoylated samples. DNA topoisomerase I, nucleolar RNA helicase 2 (DDX21), and PARP-1 were identified as positive hits by mass spectrometry, and they
were readily detected in the heat-shocked K562HA-SUMO-2 sample (Fig. 2B). In contrast, the control K562HA-SUMO-2 showed only a weak signal in long exposures, and no signal could be detected in the parental K562 cells. Next we analyzed the sumoylation of PARP-1. HeLa cells were transfected with Myc-PARP-1 and SUMO-2 and exposed to heat shock. PARP-1 was immunoprecipitated using antibodies against the Myc epitope. The immunoprecipitates were blotted against SUMO-2/3 and the Myc tag to detect PARP-1. A ladder of sumoylated forms after SDS-PAGE indicated that PARP-1 is multisumoylated in a heat shock-inducible manner (Fig. 2C).

**Differential Prevalence of NDSM and PDSM Sites in SUMO-2 Substrates**—As sumoylation is often coupled to transcriptional repression (1), we were surprised not to find significant enrichment of transcriptionally related GO categories (Fig. 3A and Table I). Instead strongly enriched GO categories belonging to the molecular function ontology included “RNA processing” (8.2-fold) and “RNA binding” (7.3-fold). The most frequent GO term “heterogeneous nuclear ribonucleoprotein complex” belonged to the cellular component ontology and was 64 times more abundant in our data set than in the genome (supplemental Table 2).

To evaluate the use of various consensus sites in SUMO-2 modification, we classified the substrates according to the type of consensus sites present. To achieve this we developed novel software called SUMOFI, which searches for motifs that are used for targeting of SUMO substrates from a selected set of proteins using ScanProsite (46). SUMOFI defines Ψ as Met, Leu, Val, Ile, or Phe and NDSM according to Yang et al. (15). With the help of SUMOFI we divided our substrates into the following groups. Classical refers to SUMO-2 substrates containing the consensus tetrapeptide alone, NDSM or PDSM refers to SUMO-2 substrates containing NDSM or PDSM, non-NDSM refers to SUMO-2 substrates lacking NDSM, and non-consensus refers to SUMO-2 substrates lacking the consensus tetrapeptide. Despite the significant enrichment of classical and NDSM, a majority of SUMO-2 substrates (52%) were non-consensus (Fig. 3B and Table II). To evaluate whether this was due to a heat shock-specific effect, we analyzed the proportion of non-consensus from a recent SUMO-2 proteomics study where no heat shock was used to induce SUMO-2 conjugation (21). The results were strikingly similar, i.e. 51% non-consensus, indicating that a large proportion of the SUMO-2 proteome is indeed modified on sites lacking any known SUMO consensus sequence. The proportion of NDSM was remarkably different as NDSM were significantly enriched compared with the genome, but only one PDSM was included in our data set (Fig. 3B and Table II). This implies that, unlike PDSM sites, NDSM sites specifically are targeted by heat shock-induced SUMO-2 modification, although Ubc9 likely recognizes both extended motifs in a similar manner (15).

**The Occurrence of NDSM Sites Depends on the Biological Function of the Substrate**—To understand how the different motifs are utilized in SUMO-2 modification, we performed GO classification for substrate groups according to the type of consensus sites. Certain GO terms, but not all SUMO-2-enriched GO terms, showed prominent differences between NDSM and non-NDSM (Fig. 3C and Table III). GO categories associated with DNA- and nucleotide-related functions such as “DNA metabolic process” and “helicase activity” were more abundant in NDSM than non-NDSM. The same GO terms were enriched in NDSM also when compared with classical. Importantly when examining all NDSM site-containing proteins in the
There is no significant enrichment of either DNA metabolic process or helicase activity (15). In contrast, some SUMO-2-enriched terms, such as RNA binding and RNA processing, were equally abundant in NDSM, classical, and non-NDSM (Fig. 3C and Table III).

**DISCUSSION**

We present a novel strategy for SUMO-2 substrate purification and identification. In technical terms the innovation lies within the desumoylation of substrates prior to SDS-PAGE using a recombinant SUMO protease, which considerably reduces the complexity of the SDS-PAGE-separated sample and increases the number of identified substrates. This approach of protease treatments should be widely applicable for example when identifying substrates of ubiquitin and other ubiquitin-like modifiers, which are conjugated as chains of variable length. There are several commercially available re-

---

**Table I**

| GO term               | GO term ID | p value (adjusted) |
|-----------------------|------------|--------------------|
| Purine nucleotide binding | GO:0017076 | 2.13e-02           |
| RNA binding           | GO:0003723 | 5.83e-13           |
| RNA processing        | GO:0006396 | 3.74e-08           |
| Translation           | GO:0006412 | 3.00e-03           |
| Cation binding        | GO:0043169 | 1.28e-11           |
| Intrinsic to membrane | GO:0031224 | 1.33e-12           |

*p values adjusted for multiple comparisons when using the complete proteome as background distribution.

---

**Table II**

| Consensus type | p value         |
|----------------|-----------------|
| Classical      | 0.023           |
| NDSM           | 3.12e-06        |

---

**Fig. 3. Consensus types and GO categories in SUMO-2 substrates.**

A. GO categories associated with RNA-related processes are enriched in heat shock-induced SUMO-2 substrates. Heat shock-induced SUMO-2 substrates were compared with genome background for identification of positively and negatively enriched GO terms belonging to the molecular function ontology. B. Many SUMO-2 substrates lack the consensus tetrapeptide. Prevalences of classical, NDSM, PDSM, and non-consensus were compared between our list of heat shock-inducible SUMO-2 substrates and genome. C. Several GO categories are selectively associated with NDSM. Proportions of GO categories associated with genome, non-NDSM, classical, and NDSM were analyzed. Classical, SUMO-2 substrates containing the consensus tetrapeptide alone; NDSM, SUMO-2 substrates containing NDSM; PDSM, SUMO-2 substrates containing PDSM; non-consensus, SUMO-2 substrates lacking consensus tetrapeptide; non-NDSM, SUMO-2 substrates lacking NDSM.
combinant enzymes with deubiquitinating activity, such as His₆-UbP7CD (Boston Biochemicals), that could be used to facilitate identification of ubiquitin substrates.

Ubc9 uses alternative mechanisms to bind dissimilar SUMO motifs (13–16). Different SUMO motifs are thus targeted by alternative mechanisms. To understand how these different motifs are utilized in SUMO-2 modification, we made GO classifications of substrate groups according to the type of consensus sites present. Similar analyses can now be easily performed by the SUMO investigators using our SUMOFI Web server. The striking finding that GO categories associated with DNA- and nucleotide-related functions were enriched in associated with DNA- and nucleotide-related functions were.

SUMOFI Web server. The striking finding that GO categories be easily performed by the SUMO investigators using our the type of consensus sites present. Similar analyses can now be easily performed by the SUMO investigators using our SUMOFI Web server. The striking finding that GO categories associated with DNA- and nucleotide-related functions were enriched in associated with DNA- and nucleotide-related functions were.

SUMOFI Web server. The striking finding that GO categories associated with DNA- and nucleotide-related functions were enriched in associated with DNA- and nucleotide-related functions were. Similar analyses can now be easily performed by the SUMO investigators using our SUMOFI Web server. The striking finding that GO categories associated with DNA- and nucleotide-related functions were enriched in associated with DNA- and nucleotide-related functions were.

**TABLE III**

| GO term                  | GO term ID   | Classical⁴ | Non-NDM⁵  | Genome  |
|--------------------------|--------------|------------|-----------|---------|
| ATPase activity          | GO:0016887   | 0.002      | 0.0004    | 0.001   |
| DNA metabolic process    | GO:0006259   | 0.013      | 0.016     | 0.002   |
| Helicase activity        | GO:0004386   | 0.006      | 0.001     | 0.005   |
| Purine nucleotide binding| GO:0017076   | 0.009      | 0.0001    | 6.28e−05|
| RNA binding              | GO:0003723   | 0.787      | 0.876     | 0.001   |
| RNA processing           | GO:0006396   | 0.5        | 0.228     | 0.0002  |
| Translation              | GO:0006412   | 0.83       | 0.81      | 0.064   |

⁴ SUMO-2 substrates containing the consensus tetrapeptide alone.
⁵ SUMO-2 substrates lacking NDSM.

**Acknowledgments**—We thank Michael Hottiger and Paul Hassa for the PARP-1 construct and Anna Aakula for skillful technical assistance. We also thank Garry Corthals and the Proteomics Core Facility of the Turku Centre for Biotechnology. We are grateful to Minna Poukkula, Thomas Sandmann, and John Eriksson for support and scientific insight, and the members of our laboratory and especially Susumu Imanishi from the Eriksson laboratory for critical comments on the manuscript.
REFERENCES

1. Geiss-Friedlander, R., and Melchior, F. (2007) Concepts in sumoylation: a decade on. Nat. Rev. Mol. Cell Biol. 8, 947–956
2. Nacerdine, K., Lehembre, F., Bhaumik, M., Artus, J., Cohen-Tannoudji, M., Babinet, C., Pandolfi, P. P., and Dejean, A. (2005) The SUMO pathway is essential for nuclear integrity and chromosone segregation in mice. Dev. Cell 9, 769–779
3. Song, J., Durrin, L. K., Wilkinson, T. A., Krontiris, T. G., and Chen, Y. (2004) Identification of a SUMO-binding motif that recognizes SUMO-modified proteins. Proc. Natl. Acad. Sci. U.S.A. 101, 14373–14378
4. Lallemand-Breitenbach, V., Jeanne, M., Benhenda, S., Naar, R., Lei, M., Peres, L., Zhou, J., Zhu, J., Raught, B., and de The, H. (2008) Arsenic degrades PML or PML-RARα through a SUMO-triggered RNF4/ubiquitin-mediated pathway. Nat. Cell Biol. 10, 547–555
5. Tatham, M. H., Geoffroy, M. C., Shen, L., Plechanovova, A., Hattersley, N., Jaffray, E. G., Palvimo, J. J., and Hay, R. T. (2008) RNF4 is a poly-SUMO-specific E3 ubiquitin ligase required for arsenic-induced PML degradation. Nat. Cell Biol. 10, 538–546
6. Tatham, M. H., Jaffray, E., Vaughan, O. A., Desterro, J. M., Botting, C. H., Naismith, J. M., and Hay, R. T. (2001) Polymorphic chains of SUMO-2 and SUMO-3 are conjugated to protein substrates by SAE1/SAE2 and Ubch9. J. Biol. Chem. 276, 35368–35374
7. Verteaga, A. C., Andersen, J. S., Ogg, S. C., Hay, R. T., Mann, M., and Lamond, A. I. (2006) Distinct and overlapping sets of SUMO-1 and SUMO-2 target proteins revealed by quantitative proteomics. Mol. Cell Proteomics 5, 2298–2310
8. Hecker, C. M., Rabiller, M., Haglund, K., Bayer, P., and Dikic, I. (2006) Specification of SUMO1- and SUMO2-interacting motifs. J. Biol. Chem. 281, 16117–16127
9. Saitoh, H., and Hinchey, J. (2000) Functional heterogeneity of small ubiquitin-related protein modifiers SUMO-1 versus SUMO-2/3. Mol. Cell. Biol. 20, 5083–5093
10. Zhang, F. P., Mikkonen, L., Toppari, J., Palvimo, J. J., and Lamond, A. I. (2004) A proteomic study of SUMOylation-dependent SUMO modification. Proc. Natl. Acad. Sci. U.S.A. 101, 45–50
11. Chung, T. L., Hsiao, H. H., Yeh, Y. Y., Shia, H. L., Chen, Y. L., Liang, P. H., Wang, A. H., Khoo, K. H., and Shoelung-Li, S. (2004) In vitro modification of human centromere protein CENP-C fragments by small ubiquitin-like modifier (SUMO) protein: definitive identification of the modification sites by tandem mass spectrometry analysis of the isopeptides. J. Biol. Chem. 279, 39653–39662
12. Pichler, A., Knipscheer, P., Oberhofer, E., van Dijk, W. J., Korner, R., Olsen, J. V., Jentusch, S., Melchior, F., and Sixma, T. K. (2005) SUMO modification of the ubiquitin-conjugating enzyme E2–25K. Nat. Struct. Mol. Biol. 12, 264–269
13. Verteaga, A. C., Ogg, S. C., Jaffray, E., Rodríguez, M. S., Hay, R. T., Andersen, J. S., Mann, M., and Lamond, A. I. (2004) A proteomic study of SUMO-2 target proteins. J. Biol. Chem. 279, 33791–33798
14. Rosas-Acosta, G., Russell, W. K., Deyrieux, A., Russell, D. H., and Wilson, V. G. (2003) A universal strategy for proteomic studies of SUMO and other ubiquitin-like modifiers. Mol. Cell Proteomics 2, 66–72
15. Schimmel, J., Larsen, K. M., Matic, I., van Hagen, M., Cox, J., Mann, M., Andersen, J. S., and Verteaga, A. C. (2008) The ubiquitin-proteasome system is a key component of the SUMO-2/3 cycle. Mol. Cell. Proteomics 7, 2107–2122
16. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, A. M., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. 25, 25–29
17. Hassa, P. O., Covic, M., Hasan, S., Imhof, R., and Hottinger, M. O. (2001) The enzymatic and DNA binding activity of PARP-1 are not required for NF-κB coactivator function. J. Biol. Chem. 276, 45588–45597
18. Hietakangas, V., Ahlskog, J. K., Jakobsson, A. M., Hellesuo, M., Sahlgren, N. M., Holmenberg, C. I., Mikhalov, A., Palvimo, J. J., Pirkkalä, L., and Sistonen, L. (2003) Phosphorylation of serine 303 is a prerequisite for the stress-inducible SUMO modification of heat shock factor 1. Mol. Cell. Biol. 23, 2953–2968
19. Valdez, B. C., Henning, D., Busch, R. K., Woods, K., Flores-Rozas, H., Hurwitz, J., Perlaky, L., and Busch, H. (1996) A nuclear RNA helicase recognized by autoimmune antibodies from a patient with watermelon starch disease. Nucleic Acids Res. 24, 43–48
20. Elias, J. A., Haas, W., Faherty, B. K., and Gygi, S. P. (2005) Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. Nat. Methods 2, 667–675
21. Rhee, S. Y., Wood, V., Dolinski, K., and Draghici, S. (2008) Use and misuse of the gene ontology annotations. Nat. Rev. Genet. 9, 509–515
22. Flice, P., Aken, B. L., Beal, K., Ballester, C., Caccamo, M., Chen, Y., Clarke, L., Coates, G., Cunningham, F., Cutts, T., Down, T., Dyer, S. C., Eyer, T., Fritzgerald, S., Fernandez-Banet, J., Graf, S., Haider, S., Hammond, M., Holland, R., Howe, K. L., Howe, K., Johnson, N., Jenkinson, A., Kahari, A., Keefe, D., Kokocinski, F., Kulesha, E., Lawson, D., Longden, I., Megy, K., Meidi, P., Overduin, B., Parker, A., Pritchard, B., Prlic, A., Rice, S., Rios, D., Schuster, M., Sealy, I., Slater, G., Smedley, D., Spudich, G., Trevanian, S., Villa, A. J., Vogel, J., White, S., Wood, M., Brown, E., Cox, T., Cutts, T., Durbin, R., Eyer, T., Herrero, J., Hubbard, T. J., Kasprzyk, A., Proctor, G., Smith, J., Ureta-Vidal, A., and Searle, S. (2008) Ensemble 2008. Nucleic Acids Res. 36, D707–D714
23. Forthofer, R. N., Lee, E. S., and Hernandez, M. (2007) Biostatistics: a guide to design, analysis, and discovery, 2nd ed., pp. 279–280, Academic Press, Burlington, MA
24. Holm, S. (1979) A simple sequentially rejective multiple test procedure. Scand. J. Stat. 6, 65–70
25. Guo, D., Han, J., Adam, B. L., Colburn, N. H., Wang, M. H., Dong, Z., Eizirik, D. L., She, J. X., and Wang, C. Y. (2005) Proteomic analysis of SUMO4 substrates in HEK293 cells under serum starvation-induced stress. Biochem. Biophys. Res. Commun. 337, 1308–1318
26. Klein, U. R., Haendler, M., Nigg, E. A., and Muller, S. (2009) RanBP2 and RanBP3 function in a mitotic SUMO2/3 conjugation-deconjugation cycle on borealin. Mol. Biol. Cell 20, 410–418
27. Jacobs, A. M., Nicol, S. M., Hislop, R. G., Jaffray, E. G., Hay, R. T., and Fuller-Pace, F. V. (2007) SUMO modification of the DEAD box protein p68 modulates its transcriptional activity and promotes its interaction with HDAC1. Oncogene 26, 5866–5876
28. Arguisinska, J., Zhou, K., Donnelly, R. J., Hay, R. T., and Lee, C. G. (2004) A functional interaction between RHA and Ubch9, an E2-like enzyme.
Prevalence of Different SUMO-2 Conjugation Sites

specific for sumo-1. J. Mol. Biol. 341, 15–25
35. Jakobs, A., Himstedt, F., Funk, M., Korn, B., Gaestel, M., and Niedenthal, R. (2007) Ubc9 fusion-directed SUMOylation identifies constitutive and inducible SUMOylation. Nucleic Acids Res. 35, e109
36. Shiio, Y., and Eisenman, R. N. (2003) Histone sumoylation is associated with transcriptional repression. Proc. Natl. Acad. Sci. U. S. A. 100, 13225–13230
37. Li, T., Evdokimov, E., Shen, R. F., Chao, C. C., Tekle, E., Wang, T., Stadtmann, E. R., Yang, D. C., and Chock, P. B. (2004) Sumoylation of heterogeneous nuclear ribonucleoproteins, zinc finger proteins, and nuclear pore complex proteins: a proteomic analysis. Proc. Natl. Acad. Sci. U. S. A. 101, 8551–8556
38. Ganesan, A. K., Kho, Y., Kim, S. C., Chen, Y., Zhao, Y., and White, M. A. (2007) Broad spectrum identification of SUMO substrates in melanoma cells. Proteomics 7, 2216–2221
39. Vassileva, M. T., and Matunis, M. J. (2004) SUMO modification of heterogeneous nuclear ribonucleoproteins. Mol. Cell. Biol. 24, 3623–3632
40. Zhao, Y., Kwon, S. W., Anselmo, A., Kaur, K., and White, M. A. (2004) Broad spectrum identification of cellular small ubiquitin-related modifier (SUMO) substrate proteins. J. Biol. Chem. 279, 20999–21002
41. Zhang, Y. Q., and Sarge, K. D. (2008) Sumoylation regulates lamin A function and is lost in lamin A mutants associated with familial cardiomyopathies. J. Cell Biol. 182, 35–39
42. Liu, X., Liu, Z., Jiang, S. W., Ma, Z., Shirmura, K., Kang, S., Dong, S., Chen, J., Fukasawa, K., and Ye, K. (2007) Sumoylation of nucleophosmin/B23 regulates its subcellular localization, mediating cell proliferation and survival. Proc. Natl. Acad. Sci. U. S. A. 104, 9679–9684
43. Mao, Y., Sun, M., Desai, S. D., and Liu, L. F. (2000) SUMO-1 conjugation to topoisomerase I: a possible repair response to topoisomerase-mediated DNA damage. Proc. Natl. Acad. Sci. U. S. A. 97, 4046–4051
44. Agostinho, M., Santos, V., Ferreira, F., Costa, R., Cardoso, J., Pinheiro, I., Rino, J., Jaffray, E., Hay, R. T., and Ferreira, J. (2008) Conjugation of human topoisomerase 2α with small ubiquitin-like modifiers 2/3 in response to topoisomerase inhibitors: cell cycle stage and chromosome domain specificity. Cancer Res. 68, 2409–2418
45. Wang, Q. E., Zhu, Q., Wani, G., El-Mahdy, M. A., Li, J., and Wani, A. A. (2005) DNA repair factor XPC is modified by SUMO-1 and ubiquitin following UV irradiation. Nucleic Acids Res. 33, 4023–4034
46. de Castro, E., Sigrist, C. J., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P. S., Gasteiger, E., Bairocch, A., and Hulo, N. (2006) ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. Nucleic Acids Res. 34, W362–W365
47. Hay, R. T. (2005) SUMO: a history of modification. Mol. Cell 18, 1–12
48. Zhu, J., Zhu, S., Guzzo, C. M., Ellis, N. A., Sung, K. S., Choi, C. Y., and Matunis, M. J. (2006) Small ubiquitin-related modifier (SUMO) binding determines substrate recognition and paralog-selective SUMO modification. J. Biol. Chem. 283, 29405–29415
49. Ravid, T., and Hochstrasser, M. (2008) Diversity of degradation signals in the ubiquitin-proteasome system. Nat. Rev. Mol. Cell Biol. 9, 679–690
50. Hou, D., Cenciarelli, C., Jensen, J. P., Nguyen, H. B., and Weissman, A. M. (1994) Activation-dependent ubiquitination of a T cell antigen receptor subunit on multiple intracellular lysines. J. Biol. Chem. 269, 14244–14247
51. Schwartz, D., and Gygi, S. P. (2005) An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. Nat. Biotechnol. 23, 1391–1398