Supplemental Material can be found at:
http://www.mcponline.org/cgi/content/full/M400154-MCP200
/DC1

A Proteomic Strategy for Gaining Insights into Protein Sumoylation in Yeast*[S]

Carilee Denison‡, Adam D. Rudner‡, Scott A. Gerber‡, Corey E. Bakalarski‡, Danesh Moazed‡, and Steven P. Gygi‡§¶

Sumoylation represents a vital post-translational modification that pervades numerous aspects of cell biology, including protein targeting, transcriptional regulation, signal transduction, and cell division. However, despite its broad reaching effects, most biological outcomes of protein sumoylation remain poorly understood. In an effort to provide further insight into this complex process, a proteomics approach was undertaken to identify the targets of sumoylation *en mass*. Specifically, SUMO-conjugated proteins were isolated by a double-affinity purification procedure from a *Saccharomyces cerevisiae* strain engineered to express tagged SUMO. The components of the isolated protein mixture were then identified by subsequent LC-MS/MS analysis using an LTQ FT mass spectrometer. In this manner, 159 candidate sumoylated proteins were identified by two or more peptides. Furthermore, the high accuracy of the instrument, combined with stringent search criteria, enabled the identification of an additional 92 putative candidates by only one peptide. The validity of this proteomics approach was confirmed by performing subsequent Western blot experiments for numerous proteins and determining the actual sumoylation sites for several other substrates. These data combine with recent works to further our understanding of the breadth and impact of protein sumoylation in a diverse array of biological processes. ***Molecular & Cellular Proteomics 4: 246–254, 2005.***

Post-translational modifications play crucial regulatory roles at the crossroads of almost all cellular processes. It is becoming increasingly clear that some of the most important of these roles are assumed by members of the ubiquitin protein family (1). Next to ubiquitin (Ub),[1] the most intensely studied protein in this family is SUMO (small ubiquitin-like modifer), a protein that shares many mechanistic similarities with the Ub pathway (2–4). One similarity is that, like Ub, SUMO is a relatively small protein (~12 kDa in yeast) that forms an isopeptide bond with the ϵ-amine group of lysine residues in target substrates (2). Frequently, this modification is known to occur at a consensus motif in substrate proteins, defined by the amino acid sequence φKxE/D where φ is a large hydrophobic amino acid (most typically valine, isoleucine, or leucine) and x is any amino acid. However, sumoylation can also occur at lysines that do not reside in this motif, as seen for the modification of K164 in proliferating cell nuclear antigen (5).

A second feature the SUMO pathway shares with ubiquitination is that sumoylation of target proteins requires the sequential action of an E1-activating, E2-conjugating, and, at least in some cases, an E3-ligating enzyme. Specifically, in yeast, SUMO is first activated in an ATP-dependent reaction in which its most C-terminal glycine residue forms a high-energy thioester bond with a cysteine residue in the active site of the E1 heterodimer, Aos1/Uba2 (6). This activated SUMO moiety is subsequently transferred in a transesterification reaction to the E2 enzyme of the SUMO cascade, Ubc9 (7, 8). Ultimately, SUMO is transferred to a substrate lysine residue by the combined activity of Ubc9 and one of various SUMO E3 enzymes such as the yeast proteins Siz1 and Siz2 (9, 10).

A final feature of the SUMO pathway that mirrors ubiquitination is the reversible nature of the process. Multiple SUMO-specific proteases, such as the yeast proteins Ulp1 and Ulp2, are known to reside within cells. These polypeptides can actively cleave SUMO from substrate proteins (11, 12). Therefore, sumoylation is a highly dynamic process in which its steady-state levels are tightly regulated by the complex interplay of the SUMO E1, E2, E3, and proteolytic enzymes.

However, despite these three main mechanistic parallels between the two pathways, it is becoming increasingly apparent from preliminary studies that the functional outcomes signaled by SUMO modifications are very different from those directed by Ub. SUMO modification has not been found in any cases to date to signal protein degradation. In fact, several instances have been described in which SUMO competes with Ub for a given lysine acceptor site, thereby preventing its subsequent poly-ubiquitination and either degradation by the proteasome or other fates induced by ubiquitination (5, 13). A number of additional roles have also been proposed for sumoylation, including affecting protein function, protein subcellular localization, and/or multi-protein complex formation (4, 5, 14–16).

Although these insights have aided our understanding of

From the ‡Department of Cell Biology and §Taplin Biological Mass Spectrometry Facility, Harvard Medical School, Boston, MA 02115

Received, October 8, 2004, and in revised form, November 10, 2004

Published, MCP Papers in Press, November 12, 2004, DOI 10.1074/mcp.M400154-MCP200

[1] The abbreviations used are: Ub, ubiquitin; SUMO, small ubiquitin-like modifier.

© 2005 by The American Society for Biochemistry and Molecular Biology, Inc.
This paper is available on line at http://www.mcponline.org

Molecular & Cellular Proteomics

Downloaded from www.mcponline.org at HARVARD UNIVERSITY on June 2, 2008

sumoylation, details into the biological effects of SUMO are just beginning to be unraveled. In recent years, there has been a surge in interest in the functional outcomes of protein sumoylation. This is largely attributable to the fact that sumoylation is known to affect a broad array of cellular processes. For example, it has been implicated in processes ranging from transcriptional regulation to DNA damage repair to the immune response (2, 5, 13, 17). Interest in protein sumoylation is also largely driven by the knowledge that several components of the SUMO pathways play roles in a number of different diseases, including Alzheimer's and Huntington's diseases (18, 19). Clearly, a key step in gaining insight into a given biological pathway is to first define all of its components. Toward this end, our laboratory recently reported the large-scale identification of over 1,000 substrates of the ubiquitination pathway (20). Here, a similar proteomic undertaking is reported for substrates post-translationally modified by SUMO that should shed light on the crucial process of protein sumoylation.

EXPERIMENTAL PROCEDURES

Purification of SUMO Conjugates—A protocol similar to that used by Johnson *et al*. was employed with minor modifications (16). Briefly, a strain expressing $(His)_6$-FLAG-SUMO under the control of a GAL-promoter (the EJY251–11b strain containing p315-P_{GAL}-HFSMT3) was grown in 4 liters of YP media containing 1% galactose and 2% raffinose to an OD_{600} of ~2.00. These cells were harvested by centrifugation, resuspended in lysis buffer (8 M Urea, 50 mM Tris, pH 8.0) containing 10 μg/ml each of aprotinin, leupeptin, and pepstatin A and 1 mM PMSF, and lysed with a French press. The resulting mixture was centrifuged at 4 °C for 30 min at 17,000 rpm in a Sorval SS34 rotor. The supernatant was incubated with a 3-ml slurry of Ni-NTA resin (Qiagen, Valencia, CA) for 2 h at 4 °C and then loaded onto a column. Bound proteins were washed with 50 ml of wash buffer (lysis buffer containing 5 mM imidazole and 300 mM NaCl) and eluted in buffer containing 300 mM imidazole, 1 M urea, and 50 mM Tris, pH 8.0. Fractions containing protein were combined and diluted 10-fold with RIPA buffer (50 mM Tris, pH 7.5, 150 mM NaCl, 5 mM EDTA, and 1% Triton X-100) containing 0.1% SDS, 1 mM PMSF, and 1 mM β-mercaptoethanol. The mixture was then incubated with 0.4 ml of anti-FLAG Sepharose (Sigma, St. Louis, MO) at 4 °C overnight. The next day, the beads were pelleted by centrifugation, transferred to 1.5-ml centrifuge tubes, and washed six times with 1-ml aliquots of RIPA buffer containing 0.1% SDS. Proteins were eluted with buffer containing 100 mM glycine, pH 2.2, 150 mM NaCl, 1% Triton X-100, and 0.1% SDS, and the pH of these elutions was then immediately adjusted with NaOH to pH 7.5. The control preparation for this experiment was prepared in the exact same manner, except the yeast strain SUB61 in which SUMO is not tagged was utilized.

Mass Spectrometry Analysis—Isolated SUMO conjugates were loaded onto one lane of an 8.5% acrylamide gel and separated according to their molecular masses. This same protocol was also followed for the control preparation. The resulting gels were cut into 12 horizontal strips and digested with trypsin as previously described (21). Individual peptide fractions were loaded onto 75-μm inner diameter fused silica columns packed in-house with C18 resin (Michrom Bioresources, Inc., Auburn, CA) and separated using a 35-min gradient from 2.5 to 97.4% ACN. Eluting peptides directly entered an LTQ FT mass spectrometer (Thermo Electron, San Jose, CA) where the instrument alternated between acquiring a full FT-MS scan (*m/z* range of 300 to 1,400) and then 10 subsequent MS/MS spectra of the 10 most abundant precursor ions using the LTQ as the detector. Data were analyzed using the yeast NCBI database in conjunction with the Sequest search algorithm (22). Searches allowed for the following two potential modifications: +15.9949 Da on methionines and 484.2282 Da on lysines (SUMO remnant at sumoylation sites). Furthermore, a dCN value of 0.08 was required for matches and the following XCorr cutoff values for 1+, 2+, and 3+ peptides, respectively, were used: 2, 2.1, and 3. Following the initial filtering of identified peptides, in the case of one and two peptide hits, accurate mass measurements were then calculated with in-house software (VISTA; Bakalarski *et al.*, manuscript in preparation). Figures and tables of the identified proteins were made using information from both the *Saccharomyces cerevisiae* Genome Database (SGD) (www.yeastgenome.org) and Bioknowledge Retriever (proteome.incyte.com). Finally, the sequences of identified proteins were analyzed to locate potential sumoylation consensus motifs.

Western Blotting and Silver Staining Analyses—Aliquots of wild-type lysate (either 5 or 20 μl; prepared from the SUB592 strain) and both the SUMO and control preparations (either 10 or 30 μl) were quenched by the addition of SDS loading buffer to a final concentration of 1× (63 mM Tris, pH 6.8, 2% glycerol, 2.3% SDS, 100 mM DTT, and 0.05% bromphenol blue). These three samples were separated on NuPAGE Novex 4–12% Bis-Tris Gels (Invitrogen, San Diego, CA) using 1× NuPAGE MOPS SDS running buffer and then transferred to nitrocellulose membrane (GE Osmonics Labstore, Minnetonka, MN) at 100 V for 2 h. Subsequent Western blot analysis was then performed with primary antibodies raised against either the FLAG-epitope, Tup1, Rap1, Gin4, Reb1, Rsc2, Spt7, Sir4, Sir3, or Sir2 and either anti-mouse or anti-rabbit horseradish peroxidase-conjugated secondary antibodies (Pierce, Rockford, IL). Results were visualized with the ECL detection system (Amersham Biosciences, Piscataway, NJ). Silver staining was performed with the SilverQuest silver staining kit (Invitrogen).

RESULTS

Purification of SUMO Conjugates from Yeast—To isolate sumoylated substrates, a yeast strain genetically manipulated to express SUMO with both N-terminal $(His)_6$- and FLAG-tags was grown to logarithmic conditions. Due to the dynamic nature of sumoylation, these cells were then lysed in the presence of protease inhibitors and under denaturing conditions. SUMO conjugates were isolated from the resulting mixture by performing Ni-NTA chromatography and a FLAG affinity purification in tandem. To account for background binding of nonmodified histidine-rich and/or highly abundant proteins, a control two-step purification was also performed in parallel with a strain in which SUMO protein was not tagged with either the $(His)_6$- or FLAG-tag sequences.

An aliquot of the SUMO preparation isolated in this manner was subsequently separated by SDS-PAGE and visualized by Western blot analysis performed with an anti-FLAG antibody, as displayed in Fig. 1. As evident from this immunoblot (*A*), a number of sumoylated candidates were successfully isolated in this double-affinity purification procedure, as indicated by their interaction with the FLAG antibody. Further support for the presence of SUMO-modified proteins in this mixture was obtained by separating samples of both the SUMO and control preparations by SDS-PAGE and performing subsequent

Downloaded from www.mcponline.org at HARVARD UNIVERSITY on June 2, 2008

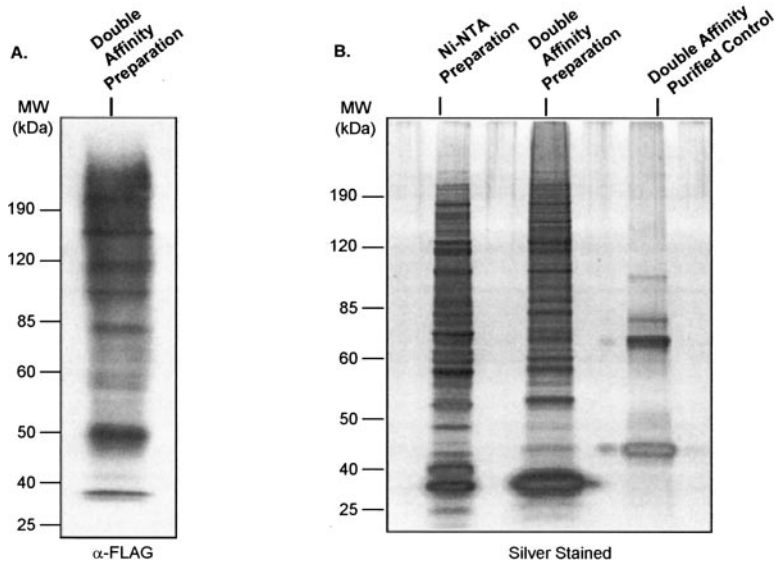

FIG. 1. **Isolation of SUMO conjugates by a double-affinity purification strategy.** *A*, proteins were isolated from a *S. cerevisiae* strain expressing $(His)_6$-FLAG-SUMO by performing a double-affinity purification, involving denaturing Ni-NTA chromatography, followed by an anti-FLAG antibody purification protocol. Isolated polypeptides were then separated by SDS-PAGE and subjected to Western blot analysis with an anti-FLAG antibody. In this manner, a number of bands were observed, suggesting that indeed SUMO conjugates had been successfully obtained by this procedure. *B*, an aliquot of the SUMO conjugates isolated after both the first (*lane 1*) and second (*lane 2*) chromatographic steps, along with a sample isolated from a mock double-affinity purification (*i.e.* from a nontagged yeast strain; *lane 3*), were separated by SDS-PAGE and visualized by subsequent silver-staining analysis. This revealed that indeed a relatively pure preparation of sumoylated substrates had been obtained. Evidence for this is provided by the presence of approximately only six proteins present in the control sample, the two most abundant of which correspond to the light and heavy chains of the antibody used in the purification procedure. Furthermore, a number of nonsumoylated histidine-rich and/or highly abundant proteins present in the preparation after Ni-NTA chromatography appear to be removed in the second purification step of the procedure. This is supported by pattern differences/intensities between *lanes 1* and *2*.

silver staining analysis (*B*). While a multitude of proteins were clearly isolated from the $(His)_6$-FLAG-SUMO strain (*lane 2*), only six bands were present in the control preparation (*lane 3*), and the two most predominant of these bands were found to correspond to antibody light and heavy chains, remnants from the FLAG purification.

A key feature of this procedure was the use of a two-step purification protocol. Yeast are known to contain a number of histidine-rich proteins that bind avidly to Ni-NTA beads, as observed in many proteomic experiments (20, 23). This is supported in the current study by the presence of several silver-stained bands in the SUMO preparation isolated after Ni-NTA chromatography (*lane 1*) that do not bind to the FLAG antibody beads (*lane 2*). Furthermore, it was recently demonstrated that the silver staining pattern of proteins isolated by Ni-NTA chromatography from both a $(His)_8$-SUMO strain and a nontagged strain were virtually indistinguishable (23). Therefore, the double-affinity purification strategy employed provides two distinct advantages. First, substantially less sequencing time in the mass spectrometer is wasted on analyzing nonsumoylated proteins. This time can instead be redirected toward analyzing the desired polypeptides. Second, proteins can be isolated from the double purification with added confidence that they are indeed true targets of the SUMO pathway.

Identification of Sumoylated Candidate Proteins by Liquid Chromatography-Tandem Mass Spectrometry—After concluding that indeed a significantly pure preparation of SUMO conjugates had been obtained, the identity of the various proteins was addressed. To date, a significant hurdle in identifying sumoylated substrates is believed to be their low abundance because in many cases only a small fraction of a substrate is thought to be modified at a given time. Therefore, we reasoned that a protocol centered around an LTQ FT mass spectrometer (Thermo Electron) would be well-suited for the purposes at hand, due to the sub-femtomole sensitivity of the instrument. Furthermore, it was anticipated that the exceptional mass accuracy of the instrument would greatly aid the confirmation process for proteins identified by only one or two peptides.

To implement this approach, SUMO conjugates were first separated in a single gel lane on a polyacrylamide gel. The gel was then divided into 12 different regions and in-gel trypsin digestions were performed. The resulting peptides were extracted, loaded onto a reverse-phase column, and eluted with increasing acetonitrile concentrations directly into the mass spectrometer. In the mass spectrometer, both accurate mass (MS) and sequence (MS/MS) information was acquired for eluting peptides. This information was used by the SEQUEST algorithm to identify the peptides in subsequent database searches performed against the yeast database. It is important to note that peptides identified in this way were subjected to stringent search criteria including high Xcorr values and

Molecular & Cellular Proteomics
Downloaded from www.mcponline.org at HARVARD UNIVERSITY on June 2, 2008

TABLE I
Protein identified as sumolyation substrates by two or more peptides

Protein name	# Peptides	Protein name	# Peptides	Protein name	# Peptides	Protein name	# Peptides
Abf1	20	Gcr1	4	Rap1	6	Spt7	19
Act1	4	Gcr2	3	Rcs1	2	Srp40	3
Adh1	4	Gcy1	4	Reb1	17	Stb3	10
Adh6	4	Gin4	15	Rgr1	2	Sth1	6
Aos1	2	Gnd1	4	Rpa43	3	Sum1	30
Asf2	5	Gsy2	3	Rpb4	6	Swc3	9
Asn2	7	Hap1	5	Rpc37	4	Swc4	6
Bdf1	9	Hmo1	4	Rpc53	5	Swi3	4
Bdp1	10	Hof1	2	Rpl2a	2	Swr1	3
Bir1	5	Hpc2	4	Rpl3	4	Tal1	4
Bmh1	3	Hta1	2	Rpl4a	7	Tdh1	2
Bop3	4	Ies4	2	Rpl7a	3	Tdh2	3
Bud3	9	Iki1	4	Rpl8a	3	Tec1	6
Bud4	13	Ils1	2	Rpl16b	2	Tef1	8
Car2	8	Ipp1	2	Rpl20b	5	Tef4	2
Cbf1	2	Isw1	5	Rpl25	5	Tfg1	6
Cdc3	25	Lap3	3	Rpo21	5	Ths1	2
Cdc11	15	Met17	8	Rpo26	4	Tif1	10
Cdc12	5	Mlp1	3	Rpp2b	4	Toa1	2
Cdc48	12	Mlp2	7	Rps1a	3	Tof2	7
Cet1	4	Mmt2	3	Rps3	4	Top1	5
Cin5	2	Mrp8	9	Rps8a	3	Top2	11
Clf1	2	Ncp1	12	Rsc1	9	Tsa1	2
Cpr1	4	Net1	16	Rsc2	31	Tup1	24
Cti6	3	Ngg1	6	Rsc8	12	Tye7	6
Dcp1	6	Nip1	4	Rsc58	4	Uaf30	2
Dep1	3	Nop12	5	Rtf1	3	Uba2	15
Ebp2	4	Nto1	3	Rvb1	7	Ubc9	6
Eft2	4	Nut1	4	Sah1	3	Ubi4	2
Eno1	4	Orc3	4	Sdc1	3	Utp9	4
Erg20	2	Pgi1	7	Shs1	22	Vid21	7
Fba1	3	Pgk1	9	Sir3	4	Vps72	20
Fpr3	2	Pgm2	4	Sir4	11	Yck2	2
Fpr4	4	Pho8	3	Sko1	8	Ydl025c	3
Gal1	12	Pho23	3	Smc1	5	Yer064c	13
Gal2	2	Pob3	5	Smc3	3	Yil110w	2
Gal7	7	Pcna	15	Smc5	7	Ylr455w	3
Gal11	2	Prp45	3	Sod1	3	Ymr111c	16
Gcn4	3	Psa1	2	Spp41	10	Ypk2	3
Gcn5	8	Pyc1	4	Spt5	6		

mass accuracies of less than 10 ppm from expected masses for peptides with two or fewer matching peptides. This same protocol was performed in its entirety for the control sample. All proteins that were unique to the SUMO preparation and that were identified by two or more peptides are found in Table I (159 proteins) and are considered to be true candidates for sumoylation.

In Supplemental Table II of this work, we have also provided a separate list containing 92 putative substrates of sumoylation. This list was compiled from proteins identified by only one peptide where in addition to Xcorr and dCN filtering requirements, this peptide's mass was within 10 ppm of its calculated mass. While it is likely that many of these proteins are true targets of the SUMO pathway (especially in light of their high mass accuracy), it is hard to draw firm conclusions about these data from just one peptide. However, we have provided this list in hope that it may aid researchers interested in these particular proteins in follow-up experiments.

It should be noted that if a particular protein detected in the SUMO preparation was also identified in the control preparation by one or more peptides that met the above-stated search criteria, it was not included in either of these two tables of sumoylation candidates. Instead, these proteins, along with all proteins unique to the mock purification, are tabulated in Supplemental Table III. As predicted from the above-mentioned silver-stained analysis and as evident in this table, very little overlap was observed between the datasets of the SUMO and control purifications. In fact, more proteins than expected appeared to be unique to the control purification. If proteins present in the control sample truly stemmed from nonspecific interactions with the affinity matrix, one might predict that all of these same proteins would have also been

Molecular & Cellular Proteomics

Downloaded from www.mcponline.org at HARVARD UNIVERSITY on June 2, 2008

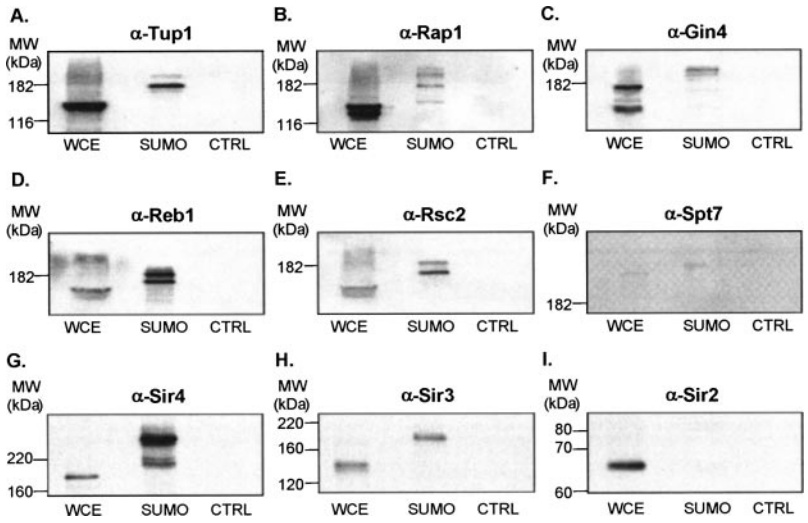

FIG. 2. **Western blot validation of sumoylated substrates.** Aliquots of wild-type cell extract and from both the SUMO and mock purifications were separated by SDS-PAGE and then subjected to Western blot analysis with antibodies raised against one of the following proteins identified by MS: Tup1 (*A*), Rap1 (*B*), Gin4 (*C*), Reb1 (*D*), Rsc2 (*E*), Spt7 (*F*), Sir 4 (*G*), and Sir 3 (*H*). In all of these cases, the protein under study migrated slower in the SUMO preparation (*lane 2*) than the predominant form in the extract (*lane 1*), confirming the sumoylation of these proteins. This is further supported by the absence of these proteins in the control preparation (*lane 3*) and also by the fact that no such shift was apparent when the same analysis was performed for Sir2 (*I*), a protein not identified by MS in this study to be a sumoylated substrate. Interestingly, six of the candidates analyzed in this manner (*A–E* and *G*) appear as multiple bands in the SUMO sample, suggesting they have poly-SUMO chains, are sumoylated on multiple lysine residues, and/or are modified by both SUMO and Ub.

present in the SUMO purification. It seems highly likely that more of these proteins were in fact present in the SUMO preparation but were not identified in the analysis due to the complexity of the sample and the fact that the limited sequencing time in the mass spectrometer was instead spent analyzing more abundant, sumoylated species.

Western Blot Confirmation of Sumoylated Substrates—Importantly, in the above analysis, most proteins known to be sumoylated in yeast were identified. In fact, five of these proteins, Top2, Cdc3, Cdc11, Pol30, and Shs1 were identified by more than 10 peptides each (5, 16, 24). Another protein, Ycs4, was also identified by the high mass accuracy of one peptide, leaving only one known yeast-sumoylated protein (Pds5) not detected in our study (25, 26). These findings lend great support to the value of the dataset obtained.

To further validate our proteomics strategy, two additional means of confirming the sumoylation of identified proteins were sought. The first approach was to validate the sumoylation of numerous candidate proteins by immunoblotting. Specifically, samples containing aliquots of either wild-type cell extract, the double-affinity purified SUMO preparation, or the control preparation were separated by SDS-PAGE. Western blot analyses were then performed for sumoylated protein candidates that specific antibodies could readily be obtained for. This included the following eight candidate substrates: Tup1, Rap1, Gin4, Reb1, Sir4, Rsc2, Sir3, and Spt7. As indicated in Fig. 2, in all eight of these cases examined, the proteins were found to migrate significantly slower in the SUMO preparation (*lane 2*) than their predominant species in the yeast extract (*lane 1*), suggesting that indeed these proteins are post-translationally modified by SUMO. In fact, for six of the proteins studied, several such bands were observed, suggesting that these proteins are either modified by chains of SUMO, on various different acceptor sites and/or by both SUMO and Ub, a protein also identified in the MS analysis. The nature of these modifications is supported by the absence of these proteins in the control purification (*lane 3*). Furthermore, an additional control experiment was performed in which this same Western blot analysis was performed for Sir2, a protein not identified in the MS analysis. As expected, a strong signal was seen for this protein in the wild-type cell extract. However, this protein was not detected in either the SUMO or control preparations. The proteins examined by immunoblotting are believed to be a reasonable test for the dataset obtained because the proteins examined vary significantly in the number of peptides by which they were identified, ranging from 0 (Sir2) to 4 (Sir3) to 31 (Rsc2).

Identification of Sumoylation Sites by MS/MS—A second line of work pursued to engender confidence in the dataset obtained with our proteomics approach was to try to identify actual sites of modification in the sumoylation candidates. As demonstrated in Fig. 3*A*, trypsinolysis of sumoylated substrates produces three characteristic peptides: SUMO peptides, substrate peptides, and a characteristic sumoylation site peptide. This latter peptide has the following two unique features: a missed tryptic cleavage at the site of modification and also an EQIGG remnant, the five most C-terminal amino acids of SUMO, on this modified lysine. Therefore, in an effort to characterize such sites, the above SEQUEST searches were performed allowing for potentially both a missed tryptic

Molecular & Cellular Proteomics
Downloaded from www.mcponline.org at HARVARD UNIVERSITY on June 2, 2008

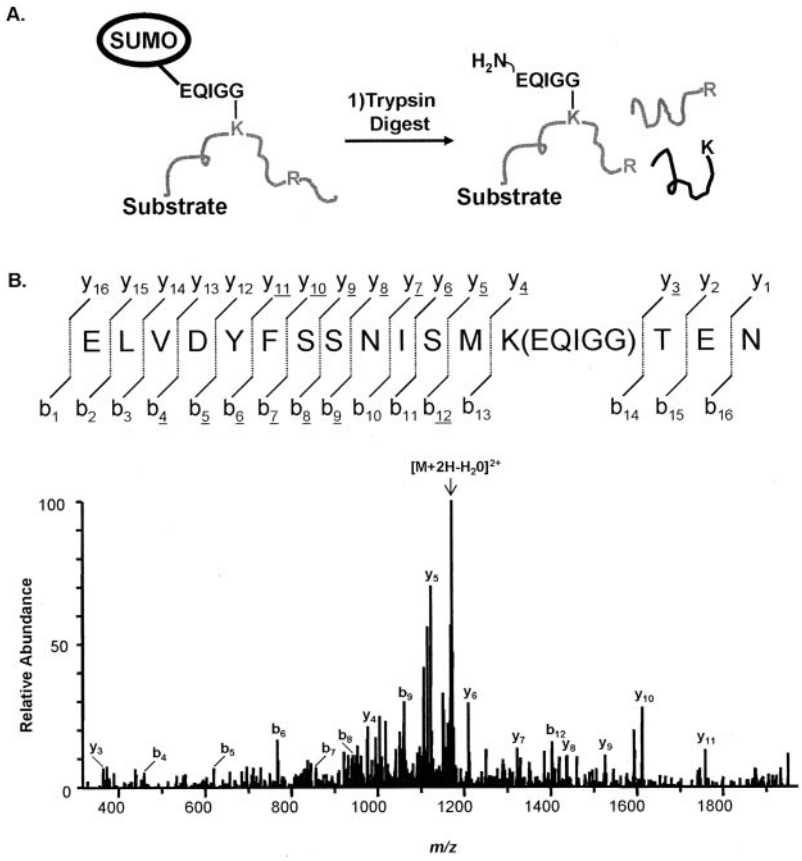

FIG. 3. **Identification of sumoylation sites by MS/MS.** *A*, as illustrated, three distinct types of peptides are created upon trypsinolysis of a sumoylated protein. These include tryptic peptides of both the substrate and SUMO, along with a characteristic sumoylation site peptide. This last type of peptide is characterized by the following two distinct features: 1) a missed tryptic cleavage at the modified lysine and 2) a 484.2-amu mass addition to this same lysine that corresponds to the five most C-terminal amino acids of SUMO (EQIGG). *B*, both of these characteristics of a sumoylation site peptide can be utilized to identify sumoylation sites by MS. This is done by entering them in as possible modifications in database searches of MS/MS data. Knowledge of this information was successfully used in the current study to identify six such sites. This is illustrated for a novel Reb1 site in this figure.

cleavage and a 484.2 amu mass addition to lysine residues, the monoisotopic mass of the EQIGG remnant. Fig. 3*B* displays a novel sumoylation site for the Reb1 protein identified in this manner. In addition to this site, five other sites were identified in this work, three more of which are also novel. The sequences of all identified sumoylation site peptides are listed in Table II.

It is believed that more sumoylation sites were not identified in this work due to the low abundance of these peptides in the sea of peptides created upon trypsinolysis. Future works will be required to develop methods of enriching for EQIGG-modified peptides. However, in addition to this approach in the current study, we also attempted to identify potential sumoylation sites in the identified proteins by analyzing their sequences with the GPMAW computer program for SUMO consensus motifs. Motifs identified in this manner are tabulated in Supplemental Table I. Approximately 86% of the proteins identified by two or more peptides and 71% of the one-peptide hits contained at least one SUMO consensus motif. While SUMO conjugation of substrates is known to not occur at such a motif in all cases studied to date, localization of these motifs suggests strong candidate lysines in the proteins that might possibly be modified and, hence, a starting point for subsequent site-directed mutagenesis studies aimed at identifying the biological effect of sumoylation in specific instances.

TABLE II
Sumoylation sites identified by MS/MS

Asterisks denotes modified sumoylated lysine. Residues before and after the identified peptides are also shown.

Gene	Protein	Peptide
YLR314C	CDC3	R.FEAAESDVK*VEPGLGMGITSSQSEK.G
YOL127W	RPL25	R.LDSYK*VIEQPITSETAMK.K
YBR088C	PCNA	R.DLSQLSDSINIMITK*ETIK.F
YMR192W	YMR192W	R.DHLLK*MGFDDMLELLKSGLLDAYIK.Q
YDR227W	SIR4	K.KPLMVK*NVKPSSPPDVK.S
YBR049C	REB1	K.ELVDYFSSNISMK*TEN.-

Analysis and Categorization of Identified Proteins—In an attempt to gain insights into the 159 potential SUMO substrates identified, the *S. cerevisiae* Genome (SGD) and the Bioknowledge Retriever databases were used to group these proteins according to their molecular environments (*A*) and functional categories (*B*); this information is graphically displayed in Fig. 4. Furthermore, this information, along with the molecular functions of these polypeptides, is tabulated in Supplemental Table I. From these analyses, several conclusions can be drawn, many of which are also supported by other recent proteomic studies of sumoylation (23, 27, 28). First, sumoylation has been traditionally viewed as a predominantly nuclear process. Indeed, ~43% of the proteins identified in this work are reported to be nuclear proteins, found

Molecular & Cellular Proteomics

Downloaded from www.mcponline.org at HARVARD UNIVERSITY on June 2, 2008

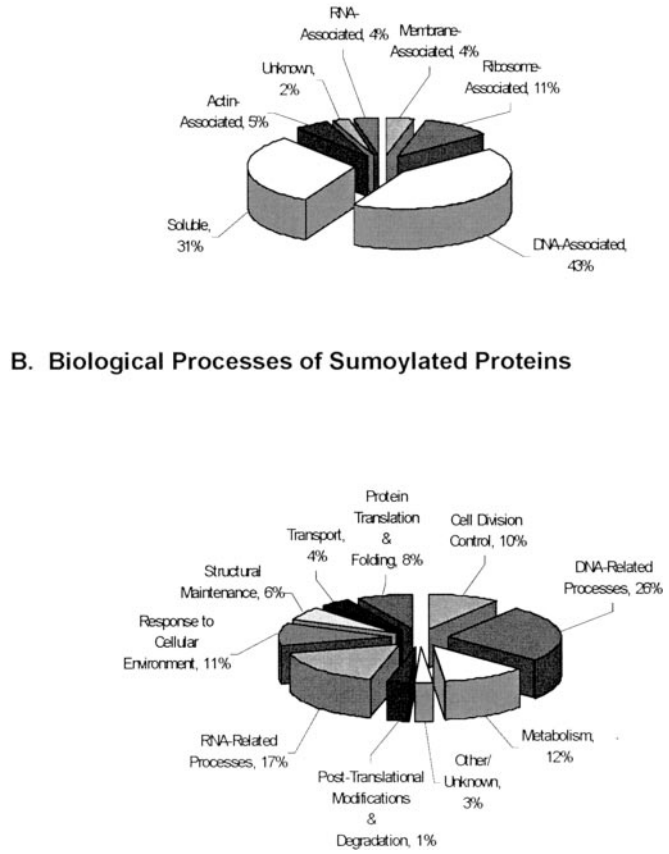

FIG. 4. **Graphical representation of both the molecular environments and functional categories of SUMO candidate substrates.** Both the *S. cerevisiae* Genome Database (SGD) and the Bioknowledge Retriever Database were used to graphically display the molecular environments (*A*) of sumoylated candidates identified (see Supplemental Table I). As expected, the majority of these proteins were found to be nuclear. However, a surprisingly high number of these proteins were also found to be cytosolic and/or localized to various other subcellular compartments. These same database sources were also utilized to graphically represent the diverse functional categories in which the identified sumoylated substrates are known to be involved in (*B*). The largest such category was found to be comprised of proteins involved in DNA-related processes. However, sumoylation clearly plays a role in other biological phenomena, including RNA-related processes, cell division control, metabolism, and protein translation and folding.

either directly or indirectly associated with DNA. However, it is interesting to note that a large percent of cytosolic, membrane, and/or other subcellular organelle proteins were also detected. For example, 31% of the proteins localized in soluble fractions, while 11 and 5% are, respectively, ribosome- and actin-associated. These findings seem to suggest that sumoylation is a less centralized process than previously thought. An important caveat to these conclusions is that it is possible that sumoylation of these substrates normally found in these various subcellular compartments targets them for the nucleus. Distinguishing between these two different scenarios will require further study.

Second, sumoylation substrate proteins belong to an even broader range of biological processes than originally predicted for sumoylation. As seen in similar works, the largest group apparently targeted by the SUMO pathway is composed of proteins involved in almost all DNA-related processes (26%) (23, 27, 28). For example, a number of both transcriptional activators and represssors like Abf1, Gcr1, and Sko1, along with subunits of all three RNA polymerases, were detected. Furthermore, numerous factors involved in chromatin remodeling, chromatin silencing, transcriptional elongation, DNA replication, and/or DNA repair were identified. In addition to these broad array of DNA-related processes, a substantial number of proteins in additional pathways are also clearly regulated by sumoylation. For instance, a significant fraction of the identified candidates are involved in various metabolic pathways (12%), RNA-related processes such as RNA processing (17%), responses to changes in cellular environment (11%), protein translation and folding (8%), and cell division control (10%). It is interesting to note that even members of the SUMO conjugation pathway itself can be modified by SUMO, as seen in a number of other recent works (23, 27, 28). For example, Aos1, Uba2, and Ubc9 were all identified in the current study. Clearly, the role of sumoylation in cellular processes is even more diverse than previously thought.

A third trend apparent from these data is that multiple members of the same protein complexes appear to be sumoylated, a trend also recently noted by Wohlschlegel *et al.* (23). Due to the highly denaturing conditions (8 M urea) employed in the purification protocol, it is not believed that a significant fraction of nonsumoylated members of a protein complex were isolated via their association with a sumoylated protein. As previously reported in the literature, a number of septins known to be sumoylated in the bud neck of dividing yeast were found to be sumoylated in the current work, such as Shs1, Cdc3, and Cdc11 (16). Further examples of this trend are provided by numerous members of the Rsc chromatin remodeling complex, such as Rsc1, Rsc2, Rsc8, Rsc58, and Sth1, and also by protein component of complexes present at silent-mating type loci, such as Sir3, Sir4, Rap1, and Abf1. The biological significance of this redundancy in the regulation of a given complex is currently unknown and will be an interesting phenomenon to address in future works.

DISCUSSION

In the current study, we report the identification of 159 candidate SUMO substrates by two or more peptides, and an additional 92 putative substrates by one peptide. The success of our proteomics approach is believed to be largely attributable to both the double-affinity purification as well as the use of new mass spectrometry instrumentation. The double-affin-

Downloaded from www.mcponline.org at HARVARD UNIVERSITY on June 2, 2008

ity purification procedure employed ensured that a significantly pure preparation of SUMO conjugates was obtained. This helped maximize the amount of time spent acquiring MS information for the sumoylated substrates. The purity of the sample also provided added confidence that the proteins identified were not false positives and, instead, true targets of SUMO. Furthermore, it helped reduce a potentially large source of false negatives present in many similar proteomic-type experiments. This source is illustrated by the recent work of Wohlschlegel *et al.* in which SUMO conjugates were isolated by denaturing Ni-NTA chromatography from a yeast strain expressing $(His)_8$-SUMO (23). When comparing the isolated sample with that obtained in a mock purification from a nontagged strain, it was found that many of the same proteins were present in both samples. Therefore, the authors eliminated these proteins as potential sumoylation candidates. However, it is highly likely that some of these proteins are true sumoylated candidates that happen to also bind nonspecifically to the Ni-NTA beads in the nontagged conditions. In fact, the authors demonstrate by subsequent Western blot analysis that this was the case for Cdc48. Having a significantly cleaner control preparation greatly reduces this source of false negatives.

A second important feature of our proteomic strategy was the use of the LTQ mass spectrometer for the collection of MS/MS spectra. This was critical because, in many cases, only a small fraction of a protein is believed to be sumoylated at a given time. The high sensitivity of the instrument presumably enabled the identification of significantly more SUMO substrates than found in recent studies utilizing less-sensitive instrumentation (27, 28). Furthermore, the excellent mass accuracy was exploited by the detection of all full MS spectra in the FT-ICR cell, greatly engendering confidence in the proteins detected, in particular those identified by only one or two peptides.

The value of the dataset obtained in this manner is underscored by several different lines of evidence: i) a number of known SUMO targets were identified in this work; ii) the sumoylation of multiple candidates was confirmed by Western blotting procedures; and iii) for several proteins, precise sites of sumoylation were identified by MS. However, despite these facts, it is important to note that, as seen in any proteomic study, presumably some fraction of these candidate proteins are not true SUMO substrates. Despite the careful controls, it is still possible that a small percent of proteins were carried nonspecifically through the double-affinity purification. In particular, it seems highly likely that some or all of the GAL proteins identified are artifacts because galactose media was used to induce the GAL promoter regulating SUMO expression. We also suspect that additional proteins identified that are known to be highly abundant species, such as some of the ribosomal and heat shock factors, might also be false positives. Furthermore, despite the highly denaturing purification conditions employed, it is still possible that some nonsumoylated proteins present in very stable complexes were carried through the purification via their association with true sumoylated substrates. Addressing these issues in such instances will require further experiments, such as more immunoblotting studies and/or sumoylation site identifications, performed on a case-by-case basis.

As mentioned previously, due to its functional significance, protein sumoylation is an area of intense research. Therefore, a number of proteomic studies have been performed recently to identify the substrates of sumoylation both in yeast and higher eukaryotes (23, 27–31). In the current work, a significant fraction of the proteins identified are novel potential substrates of the SUMO pathway. However, equally important, a large number of the proteins detected have also been identified in other recent works. Therefore, in addition to providing new insights, this study also provides confirmation in many instances to these previous reports. To aid the complementary nature of these experiments, we have tabulated the proteins identified, along with the number of peptides with which these identifications were made, in each of the four such works performed to date in *S. cerevisiae* (23, 27, 28). This tabulation can be found in Supplemental Table IV. This comparative list should help compile the wealth of information obtained in recent works and help direct more focused studies in the future that will ultimately result in a better understanding of protein sumoylation.

Sumoylation is known to be a highly dynamic process and the profile of sumoylated substrates is known to vary dramatically throughout the cell cycle and in response to various stimuli (16, 27). Therefore, a particular line of future work that should be aided by this information involves using quantitative MS tools, such as isotope-coded affinity tags (ICAT), stable isotope labeling with amino acids in cell culture (SILAC), and/or absolute quantification (AQUA) technologies, to identify how the modifications of these SUMO substrates change in various cellular conditions and genetic backgrounds (32–34). Furthermore, the use of proteomic approaches involving double-affinity purification procedures and LTQ FT mass spectrometry as described here should be applicable to future studies aimed at increasing the number of sumoylated substrates identified to date in higher eukaryotes. Finally, this approach should prove to be a valuable strategy for studying the substrates of various other Ub-like proteins, including Ned8, Atg12, and ISG15.

Acknowledgments—Our thanks to Erica Johnson (Thomas Jefferson University) for yeast strain EJY251-11b and Daniel Finley (Harvard Medical School) for the control strain. We also thank Brad Cairns (University of Utah) for α-RSC2, Fred Winston (Harvard Medical School) for α-Spt7, Doug Kellogg (University of California at Santa Cruz) for α-Gin4, David Shore (Unige) for α-Rap1, and Jonathan Warner (Einstein College of Medicine) for α-REB1.

* This work was supported by National Institutes of Health Grants HG00041 and GM67945 (to S. P. G.) and the Jane Coffin Childs Memorial Fund for Medical Research (to A. D. R.). The costs of pub-

Downloaded from www.mcponline.org at HARVARD UNIVERSITY on June 2, 2008

lication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[S] The on-line version of this article (available at http://www.mcponline.org) contains supplemental material.

¶ To whom correspondence should be addressed: Department of Cell Biology and Taplin Biological Mass Spectrometry Facility, 240 Longwood Avenue, Harvard Medical School, Boston, MA 02115. Tel.: 617-432-3155; Fax: 617-432-1144; E-mail: steven_gygi@hms.harvard.edu.

REFERENCES

1. Schwartz, D. C., and Hochstrasser, M. (2003) A superfamily of protein tags: Ubiquitin, SUMO and related modifiers. *Trends Biochem. Sci.* **28,** 321–328
2. Johnson, E. S. (2004) Protein modification by SUMO. *Annu. Rev. Biochem.* **73,** 355–382
3. Melchior, F. (2000) SUMO-nonclassical ubiquitin. *Annu. Rev. Cell Dev. Biol.* **16,** 591–626
4. Seeler, J. S., and Dejean, A. (2003) Nuclear and unclear functions of SUMO. *Nat. Rev. Mol. Cell. Biol.* **4,** 690–699
5. Hoege, C., Pfander, B., Moldovan, G. L., Pyrowolakis, G., and Jentsch, S. (2002) RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. *Nature* **419,** 135–141
6. Johnson, E. S., Schwienhorst, I., Dohmen, R. J., and Blobel, G. (1997) The ubiquitin-like protein Smt3p is activated for conjugation to other proteins by an Aos1p/Uba2p heterodimer. *EMBO J.* **16,** 5509–5519
7. Johnson, E. S., and Blobel, G. (1997) Ubc9p is the conjugating enzyme for the ubiquitin-like protein Smt3p. *J. Biol. Chem.* **272,** 26799–26802
8. Schwarz, S. E., Matuschewski, K., Liakopoulos, D., Scheffner, M., and Jentsch, S. (1998) The ubiquitin-like proteins SMT3 and SUMO-1 are conjugated by the UBC9 E2 enzyme. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 560–564
9. Johnson, E. S., and Gupta, A. A. (2001) An E3-like factor that promotes SUMO conjugation to the yeast septins. *Cell* **106,** 735–744
10. Takahashi, Y., Kahyo, T., Toh-E, A., Yasuda, H., and Kikuchi, Y. (2001) Yeast Ull1/Siz1 is a novel SUMO1/Smt3 ligase for septin components and functions as an adaptor between conjugating enzyme and substrates. *J. Biol. Chem.* **276,** 48973–48977
11. Li, S. J., and Hochstrasser, M. (1999) A new protease required for cell-cycle progression in yeast. *Nature* **398,** 246–251
12. Li, S. J., and Hochstrasser, M. (2000) The yeast ULP2 (SMT4) gene encodes a novel protease specific for the ubiquitin-like Smt3 protein. *Mol. Cell. Biol.* **20,** 2367–2377
13. Desterro, J. M., Rodriguez, M. S., and Hay, R. T. (1998) SUMO-1 modification of IκBα inhibits NF-κB activation. *Mol. Cell* **2,** 233–239
14. Matunis, M. J., Coutavas, E., and Blobel, G. (1996) A novel ubiquitin-like modification modulates the partitioning of the Ran-GTPase activating protein RanGAP1 between the cytosol and the nuclear pore complex. *J. Cell Biol.* **135,** 1457–1470
15. Mahajan, R., Delphin, C., Guan, T., Gerace, L., and Melchior, F. (1997) A small ubiquitin-related polypeptide involved in targeting RanGAP1 to nuclear pore complex protein RanBP2. *Cell* **88,** 97–107
16. Johnson, E. S., and Blobel, G. (1999) Cell cycle-regulated attachment of the ubiquitin-related protein SUMO to the yeast septins. *J. Cell Biol.* **147,** 981–994
17. Verger, A., Perdomo, J., and Crossley, M. (2003) Modification with SUMO. A role in transcriptional regulation. *EMBO Rep.* **4,** 137–142
18. Li, Y., Wang, H., Wang, S., Quon, D., Liu, Y. W., and Cordell, B. (2003) Positive and negative regulation of APP amyloidogenesis by sumoylation. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 259–264
19. Steffan, J. S., Agrawal, N., Pallos, J., Rockabrand, E., Trotman, L. C., Slepko, N., Illes, K., Lukacsovich, T., Zhu, Y. Z., Cattaneo, E., Pandolfi, P. P., Thompson, L. M., and Marsh, J. L. (2004) SUMO modification of Huntingtin and Huntington's disease pathology. *Science* **304,** 100–104
20. Peng, J., Schwartz, D., Elias, J. E., Thoreen, C. C., Cheng, D., Marsischky, G., Roelofs, J., Finley, D., and Gygi, S. P. (2003) A proteomics approach to understanding protein ubiquitination. *Nat. Biotechnol.* **21,** 921–926
21. Gygi, M.P., Licklider, L. J., Peng, J., and Gygi, S. P. (2004) *Proteins & Proteomics: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 514–519
22. Eng, J. K., McCormack, A. L., and Yates, J. R. (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5,** 976–989
23. Wohlschlegel, J. A., Johnson, E. S., Reed, S. I., and Yates, J. R. (2004) Global analysis of protein sumoylation in *Saccharomyces cerevisiae. J. Biol. Chem.* **279,** 45662–45668
24. Bachant, J., Alcasabas, A., Blat, Y., Kleckner, N., and Elledge, S. J. (2002) The SUMO-1 isopeptidase Smt4 is linked to centromeric cohesion through SUMO-1 modification of DNA topoisomerase II. *Mol. Cell* **9,** 1169–1182
25. D'Amours, D., Stegmeier, F., and Amon, A. (2004) Cdc14 and condensin control the dissolution of cohesin-independent chromosome linkages at repeated DNA. *Cell* **117,** 455–469
26. Stead, K., Aguilar, C., Hartman, T., Drexel, M., Meluh, P., and Guacci, V. (2003) Pds5p regulates the maintenance of sister chromatid cohesion and is sumoylated to promote the dissolution of cohesion. *J. Cell Biol.* **163,** 729–741
27. Zhou, W., Ryan, J. J., and Zhou, H. (2004) Global analyses of sumoylated proteins in *Saccharomyces cerevisiae*: Induction of protein sumoylation by cellular stresses. *J. Biol. Chem.* **279,** 32262–32268
28. Panse, V. G., Hardeland, U., Werner, T., Kuster, B., and Hurt, E. (2004) A proteome-wide approach identifies sumoylated substrate proteins in yeast. *J. Biol. Chem.* **279,** 41346–41351
29. Li, T., Evdokimov, E., Shen, R. F., Chao, C. C., Tekle, E., Wang, T., Stadtman, 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
30. Vertegaal, A. C., Ogg, S. C., Jaffray, E., Rodriguez, 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
31. 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
32. Gygi, S.P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **17,** 994–999
33. Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., and Mann, M. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **1,** 376–386
34. Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W., and Gygi, S. P. (2004) Absolute quantification (AQUA) of proteins and phosphoproteins from cell lysates by tandem MS. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 6940–6945

Downloaded from www.mcponline.org at HARVARD UNIVERSITY on June 2, 2008