Identification of SUMO conjugation sites in the budding yeast proteome

Miguel Esteras¹, I-Chun Liu¹, Ambrosius P. Snijders², Adam Jarmuz¹ and Luis Aragon¹,∗

¹Cell Cycle Group, MRC London Institute of Medical Sciences, Du Cane Road, London W12 0NN, UK.
²Protein Analysis and Proteomics Platform, The Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK.
*Corresponding Author:
Luis Aragon, Cell Cycle Group, MRC London Institute of Medical Sciences, Du Cane Road, London W12 0NN, UK;
E-mail: luis.aragon@csc.mrc.ac.uk

ABSTRACT Post-translational modification by the small ubiquitin-like modifier (SUMO) is an important mechanism regulating protein function. Identification of SUMO conjugation sites on substrates is a challenging task. Here we employed a proteomic method to map SUMO acceptor lysines in budding yeast proteins. We report the identification of 257 lysine residues where SUMO is potentially attached. Amongst the hits, we identified already known SUMO substrates and sites, confirming the success of the approach. In addition, we tested several of the novel substrates using SUMO immunoprecipitation analysis and confirmed that the SUMO acceptor lysines identified in these proteins are indeed bona fide SUMOylation sites. We believe that the collection of SUMO sites presented here is an important resource for future functional studies of SUMOylation in yeast.

INTRODUCTION SUMO (small ubiquitin-related modifier) is a 10-kDa highly conserved protein modifier that reversibly conjugates to specific lysine residues on many target proteins [1]. The functional consequences of SUMO modification include changes in protein stability, localization, DNA-binding, or protein interactions. These SUMO effects can be mediated by providing a new binding interface, masking existing binding sites, or inducing conformational changes on the substrates [2]. SUMO is covalently conjugated to its substrates sequentially by the action of an E1 activating enzyme, an E2 conjugating enzyme, and an E3 ligase [1]. Similar to ubiquitin, SUMO itself can be further SUMOylated via addition of SUMO moieties to lysine residues on SUMO [3]. SUMO can also be released from substrates by SUMO proteases (Ulp1 and Ulp2 in yeast).

A single SUMO gene is found in S. cerevisiae (SMT3), C. elegans (SMO-1) and D. melanogaster (smt3). In contrast, the human genome contains five SUMO variants, SUMO1-SUMO5. SUMO2 and SUMO3 are 97% similar each other but only share 50% of sequence similarity with SUMO1. Accordingly, SUMO1 and SUMO2/3 seem to be functionally distinct with different substrate set [4, 5, 6] role has been yet identified for SUMO4 and even its aptness to be conjugated in vivo remains unclear [7].

A consensus motif for SUMOylation was proposed soon after the mapping of the first SUMO-modified lysine residues. Studies of the first conjugation sites suggested that the acceptor lysines were contained within the consensus ψKxE (where ψ is a large hydrophobic amino acid and x any amino acid) [8]. This motif together with a particular 3D structure in the substrate was proposed to allow the binding of the E2 enzyme, Ubc9, and the consequent transfer of SUMO [9]. In addition to the simple 4 amino acid consensus motif, two more extended versions have also been identified. The first one was the phosphorylation-dependent SUMOylation motif (PDSM), which consists of the core motif succeeded by a phosphorylated serine and a proline (ψKxExxpSP) [10]. The second extended motif is the negatively charged amino-acid dependent SUMOylation motif (NDSM), consisting of the core motif succeeded by a phosphorylated serine and a proline-1 (ψKxExxpSP) [10]. The second extended motif is the negatively charged amino-acid dependent SUMOylation motif (NDSM), consisting of the core motif succeeded by two or more acidic amino acids in the C-terminal tail [11]. Although previously described motifs are found in many substrates, some exceptions have been identified, e.g. the K14 in E2-25K (H. sapiens) and the K164 in PCNA (S. cerevisiae). It is still unclear how Ubc9 recognizes these sites. Whether these unorthodox motifs mimic the 3D structure present at the SUMO consensus motif, or whether they are

doi: 10.15698/mic2017.10.593
Received originally: 19.04.2017;
in revised form: 25.08.2017,
Accepted 27.08.2017,
Published 02.10.2017.

Keywords: SUMO, proteome, budding yeast, mass spectrometry, site-specific SUMOylation.

Abbreviations: aa – amino acid, SUMO – small ubiquitin-related modifier, MS – mass spectrometry, NTA - nitrilotriacetic acid.
no more than rare exceptions remains to be answered. However, they make predictions of whether and where SUMOylation might occur in a given protein challenging.

The identification and quantification of SUMOylation by mass spectrometry (MS) is specially challenging. Most of the in vivo modified proteins have low steady-state SUMOylation and conjugated SUMO is very likely to be lost during the protein extraction and purification. Hence, input protein sample for MS are likely to contain low amounts of SUMOylated peptides. In addition, SUMO-modified lysines keep an amino acid (aa) side chain (5 residues in case of Smt3) after trypsin digestion which belongs to the SUMO modifier. During tandem MS, this aa side chain generates overlapping fragment ions with the ones from the target protein peptide. Standard database matching logarithm find it challenging to assign correct sequences to such a complex ion spectrum. Therefore, our knowledge of site-specific SUMOylation of proteomes is poor particularly when compared to other PTMs, like phosphorylation.

Several studies in mammalian systems have used clever strategies to improve the identification of SUMOylation sites. One of these strategies involves the mutation of all internal lysines in SUMO to arginines to make the mutant SUMO immune to digestion when Lys-C protease is used [12]. This allows digestion of the entire lysate and enrichment of SUMOylated peptides, greatly diminishing the sample complexity. In this study, we have used a similar proteomic approach to identify SUMO-modified proteins and their conjugation sites in the budding yeast proteome. We report over 200 potential SUMO sites. We have chosen a handful of newly identified SUMOylated substrates and demonstrate that mutation of the identified

![FIGURE 1: Proteomic screen to identify SUMO sites in budding yeast proteins.](image-url)

(A) Sequences of wildtype SUMO (SMT3) and the lysine-deficient His6 tagged mutant (SMT3-KallIR I93R) used in the method. (B) Wild-type and SMT3-KallIR I93R strains plated on full media (YPD) and media containing methyl methanesulphonate (MMS). (C) Diagrammatic representation of the purification strategy employed to enrich for SMT3-conjugated peptides. Cell lysates from yeast expressing SMT3-KallIR I93R were digested with the endoprotease LysC, cleaving after lysines. The digested lysates were run on SDS-PAGE. Unconjugated SMT3-KallIR I93R (indicated) and SMT3-KallIR I93R conjugated to target protein fragments were identified. Gel area above the unconjugated SMT3-KallIR I93R band were excised, digested with trypsin, and analyzed by nano LC-MS/MS. And database searched to identify SUMO acceptor lysines in the yeast proteome.
SUMO-conjugation sites prevents their modification in vivo. We present this resource to aid future efforts in the functional characterisation of SUMOylation in these yeast proteins.

RESULTS
Strategy to enrich SUMO-bearing peptides
To identify SUMOylation sites in the yeast proteome we adapted a strategy recently used to discover novel acceptor lysines for SUMO2 in Hela cells [12]. We used a His6-tagged SMT3 (yeast SUMO) allele where all lysines had been replaced by arginines, SMT3-KalIR (Fig. 1A). Similar alleles have been employed previously to identify sites of poly-SUMO chain formation [3]. In addition a mutation at the C-terminus of Smt3 was introduced, isoleucine at position 96 was substituted by arginine (Smt3-KallR-I96R) [13]. Importantly, we confirmed that the SMT3-KalIR allele is conjugated in vivo (Fig. S1) and able to support growth similar to its wild-type Smt3 counterpart (Fig. 1A) [13]. Importantly, we confirmed that the SMT3-KallR-I96R protein is unsensitive to digestion by endoprotease LysC, an enzyme that specifically cleaves after lysine residues. Therefore unconjugated Smt3-KallR and Smt3-KalIR covalently attached to peptides from target proteins can be easily separated by SDS-PAGE from the rest of the proteome fragments after digestion of protein extracts with LysC (Fig. 1C). Excision of the gel area above unconjugated Smt3-KalIR selectively isolates Smt3-modified peptides (Fig. 1C). The excised gel fragments containing the peptides modified were digested with trypsin, which cleaves after arginine and lysine, and therefore removes most of Smt3-KalIR from the substrate peptides. This strategy generates diglycine-modified isopeptides that are more compatible with mass spectrometry identification compared to wildtype conjugates.

Identification of SUMO-acceptor lysines by mass spectrometry
Smt3 purifications for mass spectrometry analysis were performed using 9 liters of cells (harvested at O.D. 0.9) grown in YPD. The Smt3 purifications were performed as described in Material and Methods. Lys-C digested peptides were separated by size in a SDS-PAGE gel. The area that corresponds to peptides between 15 and 25 kDa (containing peptides conjugated to Smt3) was cut from the gel (Fig. 1C), fragmented in smaller horizontal bands and digested with trypsin. These fractions were digested and loaded on the mass spectrometer separately, so that the complexity of the peptide mixture could be reduced. Tryptsin digested peptides were analyzed by nanoscale

TABLE 1. Summary of all sumo-acceptor lysines detected by mass spectrometry.

| Gene   | Modified Sequence                                     | Gene   | Modified Sequence |
|--------|-------------------------------------------------------|--------|------------------|
| ABD1   | NISPIK(gj)IIR                                        | BRX1   | AEAVERK(gj)IK    |
| AFB1   | MDK(gj)IVIVWYEEK                                      | B8P1   | NIK(gj)KEEEDSPEAIK |
| AFB1   | QGQTVK(gj)NDTDDENSK                                   | BUD27  | LEDFK(gj)EYNK    |
| AFB1   | VSNDSK(gj)LDVTDDLEYHLANTHPDNDK                        | BUD3   | FFEEEEKL(gj)EELK |
| ADE1   | SITK(gj)TELDGILPVAR                                   | BUD3   | NIK(gj)QENINSSNLFPEGK |
| AIM44  | LNMK(gj)IOIK                                         | BUD3   | TGNEDVGNPPSNSSIPK(gj)HEKPPAFK |
| ALY2   | FAPDLK(gj)VTLR                                        | BUD4   | AGN(gj)QENNEINIAKEEEIPMTQDGTDLK |
| AOS1   | MDMK(gj)VEKSEDEIAIYDRQJR                              | BUD4   | QENNEINIK(gj)AAEEIEPMTOQDGTIK |
| AOS1   | SIYETVRKDEDEEK(gj)I                               | CB2F   | DNQPIK(gj)KEEINVNEDGPNSTR |
| AOS1   | VKE(gj)LSEDEIALYDR                                   | CBF5   | EDFVK(gj)PEAGASTDSEWLLLK |
| APA1   | ALTFFQDVNLWENPELK(gj)I                               | CBF5   | VNZNTPEQWK(gj)K  |
| APC5   | K(gj)IJDDELSELVEEDK                                   | CDC11  | EAKIK(gj)QEE     |
| ARP7   | LAPK(gj)EENDEMENAEQK                                  | CDC12  | K(gj)YFTDQVK     |
| ARP8   | LTK1(gj)EKDEGYNVAPDK                                 | CDC12  | YK(gj)EEENALK    |
| ASS1   | ILSN1K(gj)TER                                       | CDC19  | IVK(gj)IEIIQGVNIFDoleonLX |
| ATG2   | DEPSQO(gj)IJK                                       | CDC3   | FEAASESDKV(gj)VEPGLMGITSSQSEK |
| BAF1   | QQQGTVK(gj)NDTDDENSK                                 | CDC3   | KLOK(gj)SETELFAR |
| BDP1   | ARQEFK(gj)PHLSLKEEEEQEEEEK                           | CDC3   | SLK(gj)EEQVSK    |
| BDP1   | DK(gj)ILNADIPESRDK                                   | CDC3   | SLKIEEQVSK(gj)QDPEQEER |
| BDP1   | KGAIHTAIQLK                                          | CDC3   | SLKIEEQVSK(gj)QDPEQEER |
| BDP1   | KGTEVGLTDIDLK                                        | CDC48  | EVK(gj)VEGEDVMTDEGAK |
| BDP1   | KGSGGMITMNDLK(gj)IYR                                  | CET1   | KIANGNAVSVK(gj)KKEEANAAVDNIFEEX |
| BDP1   | NTA(gj)IJEEDQTAQQR                                   | CIN5   | MTDATAFPSPVFGFK(gj)JEENK |
| BFR2   | SIADGQISIAIKK(gj)IPVNEK                              | CMD1   | SSNITEEQAEGF(gj)EAAFAIDK |
| BIR1   | EISGK(gj)KETDQDG                                     | CMR1   | IFVTDQSGTK(gj)QEE |
| BIR1   | ILEDVSVK(gj)NETPNEMLLFETGTPASQKEN                    | CMR1   | LSDLUK(gj)DEDESALLEK |
| BIR1   | VKK1(gj)PEEPVPVSVAR                                   | CR12   | IESGVIK(gj)NLDXTSK |
| BMH2   | IVSIEIQK(gj)JEESKEK                                   | CR12   | PK(gj)IESGVIK    |
| BNI1   | KG1(gj)LDENIR                                        | CWC15  | NIK(gj)VEDK      |
| BOP3   | IGASAVALANDNISIKK(gj)IEEDVAR                          | CYC8   | QPTHAIPTQAPATIGITNAEOPQVK(gj)K |
| BRE1   | KIK(gj)LELSDPSEPLTQSVDVIAFQK                         | DAD3   | MEHINSPIQEEVIDK(gj)YK |

Microbial Cell | OCTOBER 2017 | Vol. 4 No. 10
| Gene  | Modified Sequence                                 | Gene  | Modified Sequence                                 |
|-------|---------------------------------------------------|-------|---------------------------------------------------|
| DEP1  | LSSLVK(gl)QETLTLSEKL                              | MPP10 | VK(gl)LDFDADEEDEPNAEGVEASDK                        |
| DUN1  | IVFGK(gl)CSFCFK                                   | MRP8  | EFK(gl)DIPDLK                                      |
| EAF7  | EVK(gl)FEDEEEK                                      | MRPL11 | IK(gl)QTVGGKLT                                      |
| EBP2  | SQELK(gl)KEEPTIVTASNLK                             | MRPL22 | SSMK(gl)KATLLL                                      |
| ENO2  | IEEELDGK(gl)AVAYGENFHGDKL                          | MZM1  | K(gl)VDGSSKTGPR                                    |
| ERG10 | AGAK(gl)FGQTQLVDGVR                                 | NCB2  | LHHNSVDVPK(gl)SEDSS                                |
| ESC1  | VNEGEEPEHQAVDPKVEVKG(gl)EEQEEMPSK                  | NET1  | DIDNSK(gl)JDPDR                                    |
| FB1A  | DYIMSPVGNPPEGK(gl)PNK                               | NET1  | EKEDTDK(gl)ILLEK                                   |
| FHL1  | HPQNTTDDNIEVENPVTDDGNL(gl)LELPDLNDNLADFSK          | NET1  | IK(gl)SSIVEEDVSR                                   |
| FLP1  | EMIALK(gl)DETNPIEEWQHIQLK                          | NET1  | ISEIEK(gl)ELKEGPSSPSIALPAK                         |
| GCN2  | LMDISPHLK(gl)K                                    | NET1  | K(gl)KISSIVEEDVSR                                  |
| GCN4  | FIK(gl)TEEDPIK                                      | NET1  | K(gl)SQAEPSVIEPK                                   |
| GCN5  | VK(gl)LENVVEEEQPEAQENKQEGTDK                       | NET1  | NEIDDDDAPVSLYK(gl)SVK                              |
| GPD1  | PFK(gl)TVIGSGNWGTTIAK                               | NET1  | NESAXIDPOQK(gl)ETTSS                               |
| GPM1  | GLVK(gl)HELEGISDAAIK                                | NET1  | SDLFK(gl)MIETGDQTDLPQWFK                            |
| GSH1  | ASGEIPPTAK(gl)FFR                                   | NET1  | SQAEPSVIEPK(gl)JR                                   |
| GZF3  | AISNVK(gl)TETTPHPHFILQSSS                          | NET1  | VADLK(gl)JANSANINGEDELNK                           |
| HAA1  | IGGQSNVK(gl)QENYSK                                 | NFI1  | NENQGTKV(gl)QEQDYDSR                               |
| HAP1  | VK(gl)QESSDELK                                      | NHP10 | KISNIADDODEKEENK(gl)JK                              |
| HDA1  | MDSVMVK(gl)K                                       | NHP10 | VADSK(gl)GGEDGSIVSNN                               |
| HFF1  | K(gl)LRDIQGNTKPAIR                                  | NOP12 | LLNEEAEADDK(gl)PVTQYK                              |
| HHT1  | RFQK(gl)STELLIR                                    | NOP12 | SSSAIDNFNGINEK(gl)JKEISSVSDK                       |
| HHT1  | STGGK(gl)APRK                                       | NOP56 | PTLK(gl)NELAQEGAMELYNK                             |
| HHT1  | YK(gl)PSTVAIR                                      | NOP7  | LDPTEEIDVK(gl)JVESDLASTKL                          |
| HMO1  | DA1IAAPVK(gl)AVR                                   | NSR1  | LSWSDDEWLIK(gl)K                                   |
| HMO1  | TTDPSVKK(gl)I                                       | NTG1  | IK(gl)QEVIPPOVDLWVK                                 |
| HMS1  | DSSLSSAASIVK(gl)KEQLSGLENFLPSK                     | NTG1  | LENDISVK(gl)JVED                                    |
| HPC2  | MQTQTDTNALAVNLTDNSIK(gl)JK                        | NTG1  | RPLVK(gl)TEGPESELLPEK                              |
| HSC82 | KPK(gl)JEEDDEEEEEK                                   | NUM1  | ELSLDK(gl)JELNTNQK                                 |
| HTA1/2| ATK(gl)AQSL                                       | PAA1  | ELIK(gl)JEYDN                                      |
| HTB1/2| AVTK(gl)YSSTTOA                                    | PDC1  | LTQDK(gl)JSFNDSK                                   |
| HTB1/2| KPASK(gl)APAEEK                                     | PDR1  | TSLGTTTVQK(gl)JETDSSSTSFNSPQR                      |
| HTB2  | SSAAEK(gl)KPSK                                     | PG11  | TLSVK(gl)QEFQK                                      |
| IE54  | EPADEDPVEVK(gl)QLEK                                 | PGK1  | VK(gl)ASKEDVQK                                      |
| IE54  | GSEFTASDVK(gl)GSDDK                                 | PGK1  | VLENTIENGDSFDK(gl)JAGAEIVPK                        |
| IE54  | K(gl)KEPAREDPEVK                                    | POB3  | KEESEVNVVVK(gl)KEDEGAEVDQMAVEEK                    |
| IK1   | DIIK(gl)DENR                                       | POL30 | DLSQLDSNNITMTK(gl)IETI                              |
| INO80 | SIAVININKED(gl)DIDSFK                               | POL30 | LMDIDADFLK(gl)JIELQYDSTLSPSFSEF                    |
| IRC20 | K(gl)LEADDK                                        | PRE2  | VK(gl)EEEGFSNVNIG                                  |
| ISW1  | AK(gl)IETDNSVGETQELVAE                              | PRP45 | DVSEK(gl)IIILGAAK                                  |
| ISW1  | DIISPLLNNPTK(gl)JR                                  | PRP45 | K(gl)IQSTVAR                                       |
| ISW1  | LK(gl)EEGSR                                        | PRP45 | LDEAVNK(gl)SEGASGSHGPIQFTK                        |
| KAP123| TSSLQTAFSEPK(gl)JENVR                               | PTA1  | KIK(gl)METEPLAEPEEPPDDRMQK                         |
| KRR1  | DIFAAPEEAEK(gl)IPQNQ                                | PUS1  | K(gl)ADFDDEK(gl)DKK                                 |
| LIF1  | ISQNSVK(gl)MEDDDDDDFQFFGISK                       | RAD16 | NDNDEIEK(gl)JEEER                                  |
| MAG1  | I(g)KREDEIK                                        | RAD52 | K(g)PVFGHSHDQITKLDK                               |
| MCD1  | ELSEEK(g)VEFVDVLK                                  | RAD52 | NLVK(gl)JENTVSR                                    |
| MCM1  | QOPOOOQQQQQQOQQQVINAHANSIGNHINOQDVPA-GAIK(gl)DQVYK | RAD59 | NEANTNLYLSATNSKPTFK(gl)JLEDAK                      |
| MET12 | MEMLRNTGKEL(g)I                                   | RAP1  | DSRPK(gl)JTEISTNTNGATEDSTSEK                       |
| MET28 | VAATTAVVV(gl)JEEAPVSTNSEIDK                        | RBA50 | DVHFIK(gl)EEOSENEINIEKINDNDFNDFK                   |
| MET4  | MK(g)QEQQHGSYDSTEFINFGK                            | REB1  | AIIADISITIQHDQFQYINNTAADDNDNEK(gl)JJK              |
| MGA2  | ALK(g)JEEDDEHENK                                    | REB1  | ELVDYFSSNISMK(gl)IEN                                |
| MLP1  | KIK(gl)TEDEEEKETD                                  | REH1  | KGMK(gl)KMQQIEK                                    |
| MLP2  | RVK(gl)EEDIWQSR                                    | REP2  | GAYK(gl)QLNTITEGPK                                 |
|       |                                                    | REP2  | MDDIETAK(gl)NINTVY                                  |
| Gene     | Modified Sequence          | Gene     | Modified Sequence          |
|----------|-----------------------------|----------|-----------------------------|
| REP2     | NLTVK(gl)AR                 | RPS9B    | K(gl)EAESEGAEEAAEEDEE       |
| RHR2     | TYDAIAK(gl)FAPDFADEEVNKLGEIPEK | RPT6    | K(gl)IEFPFPSSVAAR           |
| RLP7     | SSTDQSK(gl)AQTNINSPEIIIRK   | RPT6    | YGEPOK(gl)IVLVK            |
| RNR2     | DSK(gl)INSLNK               | RRG9    | RIIK(gl)SNWKR               |
| RNR2     | STK(gl)QEAGAFTNFED          | RR15    | IFNIAITQVK(gl)TEK           |
| RPA34    | VEGLK(gl)LEHFAFTGYDAEDFVAAEVK | RPP9    | TIDEYNFDAGDLDK(gl)DIIASR    |
| RPB4     | HLK(gl)HENANETTAVEDDDDDLEDDEDVNAADDVFHM-SETREK | RSC2    | TSVK(gl)RESEPUGHTDNEDDEAYTDMIDDNPK |
| RPC37    | SEEVK(gl)AEODTGGEEEDDPVIEEFPKL | RSC4    | UIAKPETVOEVEK(gl)NER        |
| RPC37    | SINDK(gl)LVFTEDDEEDTQDR     | RSC5B   | VK(gl)QEEELNTNEEGINR        |
| RPC53    | EPTPSVK(gl)TEPVGTGLQSYLEER  | RSC8    | LENNGNSVK(gl)K             |
| RPC53    | GFKI(gl)SEGSGGSGVOK         | RSC8    | PFLPENVK(gl)IEVEGDAEGPQVKV |
| RPC53    | LPAFERPAVKEEK(gl)EDMETQADPSK | RTF1    | NAEHVK(gl)KEDSNFDSK         |
| RPC53    | MAKI(gl)YNLTHVSSGLPAGNFVEK | RVB1    | K(gl)EIVNVDNEAK             |
| RPC53    | PAVK(gl)EKEDEMTQADPSK       | SAT4    | DLK(gl)PENLLTHDGVK          |
| RPC53    | VKI(gl)ILIEESK              | SCC2    | K(gl)SIEVSRPEAK            |
| RPC82    | LKI(gl)TEDGFVIPALPAVSK      | SCM4    | TLKI(gl)PESER               |
| RDD3     | DAELYDGVVEEDSAEAK(gl)JTDK    | SDC1    | SVTNQVK(gl)IEEESSTNSVIESEEPK |
| RPL13A/B | GFTLAEVK(gl)AAIATLAAYAR      | SEF1    | DSKI(gl)VSQVQYLSR           |
| RPL18A   | APKI(gl)GQNTLIR             | SGS1    | QLENDIKI(gl)LEVIR           |
| RPL18A   | ALKI(gl)QGGAANK             | HS1     | EIKI(gl)QIENELIR            |
| RPL18A   | VVLK(gl)ALFLK               | HS1     | FLNSPDUPERTI(II)GLR         |
| RPL20A   | DIKI(gl)LPFPHR              | HS1     | SIKI(gl)TESSPK             |
| RPL20A   | TAVDKI(gl)IR                | SI1C    | LTDEEKI(gl)R               |
| RPL25    | APKI(gl)YASK                | SI2R    | IKI(gl)VAQPDSLKR            |
| RPL25    | AVKI(gl)ELYQVDLKV           | SI3R    | KI(gl)IPIESEADDVNNGNIPSQR  |
| RPL25    | LDSYKI(gl)VIEQIPSETAMK      | SI4R    | APFIKI(gl)SESKFPSSDALS     |
| RPL28    | FSVKI(gl)LAEEK              | SI21    | NFLQALNVK(gl)ISDPR          |
| RPL28    | IPNVPVXVKI(gl)AR            | SI21    | STNTDILTKI(gl)GSASSPR      |
| RPL34A/B | AFIKEEQKI(gl)IVK            | SI21    | TLPDKI(gl)SYNIVASEITTPVTNR  |
| RPL3S5B  | EQASQVVDPOI(gl)I            | SI21    | VIPEYGLNSIYYI(gl)QLPNIKG   |
| RPL4A/B  | LNPYAKI(gl)IVFAAEK          | SKO1    | DTVNKI(gl)SENAGYPSVNPRIILDK|
| RPL5     | VAAKI(gl)IIAAGQQ            | SU15    | EVKI(gl)INNYQSPVR          |
| RPL7A/B  | VTKI(gl)JATELLKL            | SU15    | NNVYMTNLKI(gl)IYEK          |
| RPL8B    | NFFIGQAOVQPK(gl)IR          | SMC5    | LDDIVSKI(gl)ISAR           |
| RPO21    | VDLNTDHTDLPSELSSEIGELDLKI(gl)LOVLLDEEYK | SNF2    | DIGAELKI(gl)R            |
| RPP1B    | ALEGKI(gl)DLK               | SODI    | GDAGVSQV(5KI)I#EQASESEPTTVSIEAGNPSNAER |
| RPP2A    | VSVLSALEGKI(gl)JSVDELITEGNEK | SPA2    | TIKI(gl)REEEEDFDRVNHNIQITGAYKT |
| RPS0A/B  | TWEKI(gl)VLVLAAR            | SPC24   | LLKI(gl)TDLGDGER           |
| RPS1     | VSGFKI(gl)DELETV            | SPP41   | GVTTPIKI(gl)I#EODSANVPPVSIASTIEPSODK |
| RPS10A   | HEEDTIKI(gl)JNLVYK          | SPP41   | IPEIKI(gl)NESVGLSNTNITDLSITNILPEITATDVK |
| RPS13    | KI(gl)GLTPSQIEVLKR          | SPP41   | PKI(gl)SEDHEWPILSDSSASONYVDHLK |
| RPS15A/B | GFSFKI(gl)LQEEER            | SPP41   | RPOIKI(gl)PESVINQNLVNYT     |
| RPS15A/B | YPKI(gl)TIDFQTNK            | SPP41   | VKI(gl)QLQDLK              |
| RPS1A    | VTFK(gl)DELETV              | SPT15   | DGTKPATTTQFSEEIDKI(gl)IR   |
| RPS20    | RYDIKELAPQVQKI(gl)IR        | SPT7    | NGFGTVIKI(gl)QEDDOIQFHNDHSINGNEAFK |
| RPS20    | SDFQK(gl)EKVEEQEQQQQQIIK    | SSE1    | GKI(gl)LEEYAPFASDAEK       |
| RPS20    | SDFQKKEKI(gl)VEEQEQQQQQIIK  | SSM4    | LSPKI(gl)DLK               |
| RPS21A/2  | ADDHASVQINAXKI(gl)VIDEER    | STB3    | EVSPPOAIIVKI(gl)SEASSISFSK |
| RPS24A/B | TQFGGGGKI(gl)SVGFGIVYNSVAEAK | STH1    | LIQLEDELXPI(gl)VF          |
| RPS28A/B | MDSKI(gl)TPVTIAK            | STH1    | VFREDIEEHFIVKI(gl)KEDSPEIGR |
| RPS3     | ALPDATVIEPKI(gl)JEEPIILPSVK | STP1    | IKI(gl)SEVNAK              |
| RPS31    | LIFAGKI(gl)IQLEDOR          | SUM1    | IITIKI(gl)SSSenSgNNTNNNTNNTDNVK |
| RPS31    | TLSDYNQK(gl)IESTLHLVLR      | SUM1    | IGI(gl)NEIPINLSLPSSK       |
| RPS56A   | KI(gl)IGEOQTELDTTTPVK       | SUM1    | KI(gl)TPGDEETTTPLVPLNSQPSDTRK |
| RPS58A/B | NVKI(gl)EEETVAK             | SUM1    | LPSGPKI(gl)DVDLTDALTSAONQANSR |
liquid chromatography coupled to high-resolution mass spectrometers (LTQ-Orbitrap Velos) and identified using MaxQuant/Andromeda software [14]. A total of 257 SUMO-acceptor lysines were identified on asynchronous growing yeast cultures (Table 1 & Suppl. Table S2). We were able to detect many SUMO-acceptor lysines previously described in several studies for yeast proteins (Fig. 2A).

Verification of novel SUMO substrates
To confirm that some of the yeast proteins newly identified as potential SUMO substrates are indeed modified by SUMO-conjugation. We tested if the yeast proteins were indeed the product of SUMO-conjugation. (Table 1) are indeed the product of SUMO-conjugation.

| Gene   | Modified Sequence | Gene   | Modified Sequence |
|--------|-------------------|--------|-------------------|
| SUM1   | YVFEPSTK(gl)QESLLSAPSSSR | TUP1   | INTDGSAATTATTETEIK(gl)PK(gl)EEDATPASLHQDHY |
| SUM2   | YLEI1K(gl)SK      | TUP1   | IWHQINANNK(gl)DSK  |
| SUM3   | TTAESTQVDVK(gl)IK | TUP1   | LQOQK(gl)DYDFK    |
| SUM4   | IQKEEPENNVTVEGK(gl)EESQPDENTK | TUP1   | KY(gl)QDEDGDAETATTIPSSAAATSTK |
| SUM5   | AGGEOQDLADI(gl)FR | TUP1   | TYEQ7E K(gl)QQIQWVEAQTEVEFGDSVK(gl)PK |
| SUM6   | LLAQAEEDDVK(gl)AAANLMR | TUP1   | TYES7 E K(gl)QEDPDENWLSDENDGASHINVNK |
| SUM7   | YDHI1K(gl)JEPESEAF1T1K | TUP1   | TYES7 E K(gl)QEDPDENWLSDENDGASHINVNK |
| TAF12  | SAIFK(gl)QETEAPIAISNENTK | TUP1   | TYE7 E K(gl)QEDPDENWLSDENDGASHINVNK |
| TAF14  | TGSASTVK(gl)GSVDLKE | TUP1   | TYES7 E K(gl)QEDPDENWLSDENDGASHINVNK |
| TAP1   | EK(gl)MPDSQANLMORLR | TUP1   | TYES7 E K(gl)QEDPDENWLSDENDGASHINVNK |
| TAP1   | TASSGNPSTSPGAK(gl)ARVAGK | TUP1   | TYES7 E K(gl)QEDPDENWLSDENDGASHINVNK |
| TAP1   | VVDLVEHVK(gl)IA    | TUP1   | TYES7 E K(gl)QEDPDENWLSDENDGASHINVNK |
| TAP1   | KI(gl)EESOSQYIPR  | TUP1   | TYES7 E K(gl)QEDPDENWLSDENDGASHINVNK |
| TAP1   | I1K(gl)QETPAPEGGQK | TUP1   | TYES7 E K(gl)QEDPDENWLSDENDGASHINVNK |
| TAP1   | IEV(gl)PEIETINNATVENDESEK | TUP1   | TYES7 E K(gl)QEDPDENWLSDENDGASHINVNK |
| TAP1   | FKPRT3K(gl)QVQK   | TUP1   | TYES7 E K(gl)QEDPDENWLSDENDGASHINVNK |
| TAP1   | LHOSQGGK(gl)EAFR  | TUP1   | TYES7 E K(gl)QEDPDENWLSDENDGASHINVNK |
| TAP1   | LVEKEFPDK(gl)SLGAASSTSHAK | TUP1   | TYES7 E K(gl)QEDPDENWLSDENDGASHINVNK |
| TAP1   | TKI(gl)QEDGDKV   | TUP1   | TYES7 E K(gl)QEDPDENWLSDENDGASHINVNK |
| TAP1   | KIK(gl)IEDK       | TUP1   | TYES7 E K(gl)QEDPDENWLSDENDGASHINVNK |
| TAP1   | TEEEENAPSSSTTFIK(gl)KEDK | TUP1   | TYES7 E K(gl)QEDPDENWLSDENDGASHINVNK |
| TAP1   | TPSVE5TK(gl)TEEENAPSSSTSFIDKKEK | TUP1   | TYES7 E K(gl)QEDPDENWLSDENDGASHINVNK |
| TAP1   | EIK(gl)LEESNPSLGS | TUP1   | TYES7 E K(gl)QEDPDENWLSDENDGASHINVNK |
| TAP1   | HLFNPDEVK(gl)HIEEEQKQTPEK | TUP1   | TYES7 E K(gl)QEDPDENWLSDENDGASHINVNK |
| TAP1   | VIIPK(gl)NDIISRQESRSIR | TUP1   | TYES7 E K(gl)QEDPDENWLSDENDGASHINVNK |
| TAP1   | VLLSAPLOK(gl)FGLSEELPR | TUP1   | TYES7 E K(gl)QEDPDENWLSDENDGASHINVNK |
| TRX1   | FSEQQPOQFDPK(gl)LDVDELGDVAQK | TUP1   | TYES7 E K(gl)QEDPDENWLSDENDGASHINVNK |
| TUP1   | APESTLK(gl)ETEPPENNTSKE | TUP1   | TYES7 E K(gl)QEDPDENWLSDENDGASHINVNK |
| TUP1   | DAEVEEIEK(gl)TLK  | TUP1   | TYES7 E K(gl)QEDPDENWLSDENDGASHINVNK |
| TUP1   | DYDFFK(gl)MNQQLAEMQGQIR | TUP1   | TYES7 E K(gl)QEDPDENWLSDENDGASHINVNK |
| TUP1   | ETTTLPVK(gl)APESTLK | TUP1   | TYES7 E K(gl)QEDPDENWLSDENDGASHINVNK |

strains carrying Smt3-KallR-16R (Table 1) are indeed the product of SUMO-conjugation. (Table 1) are indeed the product of SUMO-conjugation. (Table 1) are indeed the product of SUMO-conjugation. (Table 1) are indeed the product of SUMO-conjugation. (Table 1) are indeed the product of SUMO-conjugation. (Table 1) are indeed the product of SUMO-conjugation.
(NER) factor Rad16, the replication fork barrier protein Fob1 (Fig. 3A), the RNA Polymerase III subunit Rpc53 (Fig. 3B) and the base excision repair protein Ntg1 (Fig. 3C). Multiple SUMOylated species were observed for Tfg1, Rad16 and Fob1 (Fig. 3A).

**Mutational analysis of SUMO-acceptor sites identified by mass spectrometry**

Following the confirmation that we had identified novel SUMO-substrates during our proteomic analysis, we proceeded to test whether mutation of the lysine residues identified as potential conjugation sites prevented SUMOylation of these substrates. We identified four potential conjugation sites on Rpc53 (K51, 115, 237, 322) and two on Ntg1 (K38, 396). Strains carrying a tagged wild-type copy or a mutant allele where all acceptor lysines identified by mass spectrometry had been substituted for arginine were analyzed for Rpc53 and Ntg1. SUMOylated forms were detected in the immunoprecipitates for all the proteins when Smt3 was tagged (Fig. 3B-C), however SUMO forms were absent when the SUMO-acceptor lysines identified by mass spectrometry were mutated to arginine (Fig. 3B-C), thus these lysines represent bona fide SUMO-acceptor lysines in the target proteins and validate the success of our approach.

**DISCUSSION**

Identifying lysine residues where SUMO is conjugated to

**FIGURE 2:** SUMO-acceptor lysines detected in our mass spectrometry analysis were compared with SUMO-acceptor lysines already described in previous studies done in *S. cerevisiae*. (A) Previous studies were based on either MS or site-directed mutagenesis/immunoblotting, or a combination of both. Most of the SUMO-acceptor lysines previously found were also detected in this study. SUMO substrates (total of 257) identified in the yeast proteome of cells grown asynchronously. Previously published SUMO-substrates were obtained [8, 15-21].

To ensure that diglycine-modified lysines detected by mass spectrometry after Smt3 purification are not due to modification by ubiquitin or Nedd8, a centromeric plasmid 8His-SMT3-KallR-REQIGG-pRS415 expressing the Smt3 variant used previously for the Smt3 purification protocol with the difference that the RGG conjugating terminus was replaced for the native RIEQGG C-terminus was employed. SUMO-acceptor lysines modified by the 8His-Smt3-KallR-REQIGG keep a side chain of 5 aa after trypsin digestion (EQIGG). Therefore, any diglycine-modified lysines detected by mass spectrometry under these conditions can only be due to either false positive hits, or to ubiquitinated or neddylated contaminants. A large culture of 9 l of the strain expressing the 8His-SMT3-KallR-REQIGG variant was grown in YPD and harvested at O.D. 0.9. Smt3 purification and mass spectrometry analysis was performed as described in Material and Methods. We detected 23 diglycine-modified lysines. None of these corresponded with previously detected diglycine-modified lysines in our Smt3-RGG pulldowns. In addition none of these diglycine-modified lysines are within SUMO consensus sequences. This strongly indicates that sumo-acceptor lysines identified after purification of Smt3-RGG pulldowns represent bona fide SUMOylation sites.
substrates is a challenging process that in many cases is resolved by site-directed mutagenesis of potential acceptor lysines. This approach is not only labour intensive but often unable to discriminate between bona fide conjugation sites and lysines that alter the protein biology in a manner that leads to reduced SUMOylation levels (at other domains of the protein). Recent studies in mammalian cells have bypassed the challenges of spectrometric analysis of protein SUMOylation using differential protease cleavage of a modified SUMO2 lacking lysine residues [12]. Here we have employed a similar approach to map acceptor lysines for SUMO in endogenous substrate proteins purified from yeast cells under several conditions. Our analysis revealed a total of 257 SUMO-acceptor lysines. Over 70% of the sites identified adhere to previously described SUMO motifs, including the core consensus motif ([VILP]KxE) [22], the phosphorylation-dependent motif ([VILP]Kx[ED]x[S or xS]) [10], the negatively charged/acidic consensus motif ([VILP]Kx[ED][ED][ED]) [11] and the inverse consensus motif ([ED]xK[VILP]) [12].

SUMOylation substrates were found in all major cellular compartments. However, the presence of SUMOylated proteins was lower in the ER, Golgi’s apparatus, mitochondrion, membrane and cytoplasm when proportionally compared with the whole proteome. On the other hand, the amount of SUMOylated proteins is proportionally higher in ribosomes and much higher in the nucleus. A closer look at nuclear proteins shows that a high proportion of SUMOylated proteins is found in association with chromatin (i.e. transcription factors) (Fig. S2). Functionally, SUMOylated proteins were involved in a wide range of biological processes. There were a large number of SUMOylated proteins involved in nuclear processes such as chromatin organization, transcription and DNA damage. A few SUMO-acceptor lysines were found on proteins from the ubiquitination and protein degradation pathway, including ubiquitin itself. This raises the possibility that SUMOylation of ubiquitin might affect the ubiquitination pathway directly.

SUMOylation is known to play an important role in the metabolism of the chromatin and in gene expression. We identified a large number of SUMO-acceptor lysines on nuclear proteins. Most of these proteins are related to transcription and chromatin organization. In accordance with previous studies we found SUMO-acceptor lysines on components of the general transcription machinery as well as on components of gene-specific transcription pathways. SUMO-acceptor lysines were identified in six subunits of RNA polymerases and on components of the TFIIID, TFIIF and TFIIB, including the TATA-binding protein (TBP). Other components of the general transcription found SUMOylated included the transcription repressors Ncb2, Tup1 and Cyc8. SUMOylated proteins related to gene-specific transcription include the transcription factors Crz1 and Sko1 (stress response genes), Gzf3 (of nitrogen catabolic genes), and Cin5 (drug response genes).

Control of chromatin function is likely to involve not only SUMOylation of core histones but also proteins involved in the modification and exchange of histones. We found SUMO-acceptor lysines on histone acetylases (i.e. histone acetyltransferases).
proteins like Cbf2, and on components of other centromeric complexes (CPC complex, Ndc80 or Dam1 complexes). We mapped SUMO-acceptor function is Top2, which controls local chromatin structures. One of the key SUMO-substrate in this regulatory function is Top2, which controls local chromatin structures. The detailed functional significance of Top2 SUMOylation and of SUMOylation of other components with centromeric functions and in chromosome segregation pathways remains unclear. We hope that the SUMO-acceptor dataset presented here will be useful for future functional studies of SUMOylation in yeast.

MATERIALS AND METHODS

Yeast strains
Strains used in this study are isogenic to BY4741 (MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0), W303 (Mata can1-100 leu2-3 his3-11 trpl-1 ura3-1 ade2-1) or D55a (MATa his3Δ200 leu2-3 lys2-801 trpl-1 ura3-52). Strains used are listed in Supplemental Table S1.

Yeast media and cell cycle synchronization
Cells were grown on complete media YPED in broth or solid form (3% yeast extract, 1% peptone, 1% glucose/dextrose, 2% agar for solid media). For plates containing MMS (in DMSO), the genotoxin was added to warm YPD.

Sample preparation for western blot analysis
Samples were prepared by TCA extraction. Extracts were prepared as follows, cells were collected by centrifugation (4000 rpm, 2 minutes) and washed with 20% TCA. The TCA was added to the lysate, and mixed by pipetting. The precipitated proteins were collected by centrifugation (14000 rpm, 10 minutes 4°C) and then pellets were washed with 750 µl 100% ethanol. Proteins were solubilized in 50 µl 1 M Tris pH 8 and 100 µl x2 SDS-PAGE loading buffer (60 mM Tris pH 6.8, 2% SDS, 10% glycerol, 0.2% bromophenol blue) and boiled for 5 minutes at 95°C. Insoluble material was removed by centrifugation (14000 rpm, 5 minutes, room temperature) and the supernatant either stored at -20°C or loaded immediately onto a SDS-PAGE mini-gel. Samples were either run on an 8% acrylamide gel in Tris-Glycine SDS running buffer using the Bio Rad Mini PROTEAN 3 system or on Pre-Cast 4-12% Bis-Tris gels in NuPAGE MOPS (all Invitrogen). SDS-PAGE gels were transferred to polyvinylidene fluoride transfer membrane (Hybond-P, Amersham Biosciences) in either the Bio-Rad Mini Trans-Blot Electrophoretic Transfer Cell or by using the XCell SureLock Mini Cell Transfer module (Invitrogen). The Bio-Rad system was used in conjunction with Tris-Glycine blotting buffer (National Diagnostics) containing 20% methanol, and run for 1h at 200V or overnight at 30V. The NuPAGE system was used with NuPAGE transfer buffer containing 20% methanol and run for 1h at 30V.

Purification of His-tagged proteins
For His-tag purifications 100 OD of cells were harvested (4000 rpm, 2 minutes), washed once with water, and the cell pellet frozen at -80°C. The cell pellet was resuspended in 500 µl Buffer A (8 M Urea, 100 mM NaH₂PO₄, 10 mM Tris HCl, 0.05% Tween pH 8.1), an equal volume of glass beads were added, and the cells lysed by 1 45s cycle, power 6 in a FastPrep FP120 (BIO 101) machine. Tubes were pierced with a hot needle and placed onto fresh eppendorfs and spun (1000 rpm, 2 minutes) to collect lysate minus glass beads. The glass beads were washed with 1 ml 5% TCA and this was added to the lysate, and mixed by pipetting. The precipitated proteins were collected by centrifugation (14000 rpm, 10 minutes 4°C) and then pellets were washed with 750 µl 100% ethanol. Proteins were solubilized in 50 µl 1 M Tris pH 8 and 100 µl x2 SDS-PAGE loading buffer (60 mM Tris pH 6.8, 2% SDS, 10% glycerol, 0.2% bromophenol blue) and boiled for 5 minutes at 95°C. Insoluble material was removed by centrifugation (14000 rpm, 5 minutes, room temperature) and the supernatant either stored at -20°C or loaded immediately onto a SDS-PAGE mini-gel. Samples were either run on an 8% acrylamide gel in Tris-Glycine SDS running buffer using the Bio Rad Mini PROTEAN 3 system or on Pre-Cast 4-12% Bis-Tris gels in NuPAGE MOPS (all Invitrogen). SDS-PAGE gels were transferred to polyvinylidene fluoride transfer membrane (Hybond-P, Amersham Biosciences) in either the Bio-Rad Mini Trans-Blot Electrophoretic Transfer Cell or by using the XCell SureLock Mini Cell Transfer module (Invitrogen). The Bio-Rad system was used in conjunction with Tris-Glycine blotting buffer (National Diagnostics) containing 20% methanol, and run for 1h at 200V or overnight at 30V. The NuPAGE system was used with NuPAGE transfer buffer containing 20% methanol and run for 1h at 30V.

Immunological detection
Membranes were blocked in 5% skimmed milk powder in PBS with 0.1% Tween 20 (PBS-T) for 1h or overnight at 4°C, then incubated with either mouse monoclonal anti-c-myc IgG1x antibody 9E10 (Roche) or anti-HA IgG1 antibody 12CA5 (Roche) in blocking solution for between 1 hour at room temperature to overnight at 4°C. Following several washes in PBS-T, membranes were incubated with the sheep anti-mouse IgG Horse-Redish Peroxidase-linked antibody (GE Healthcare) at a 1/10000 dilution in blocking solution. After several further washes in PBS-T, membranes were incubated with the ECL Plus Western Blotting Detection System (GE Healthcare) followed by exposure to ECL Hyperfilm (GE Healthcare), to detect the secondary antibody.

Purification of SUMOylated proteins for mass spectrometry
Nine liters of culture harvested at O.D.0.9 were used in each experiment. Cells were pelleted and washed in chilled PBS with 20 mM NEM. Cells were resuspended in 250 ml of lysis buffer (1.85 N NaOH, 1.85% β-mercaptoethanol) and incubated. Proteins were collected by centrifugation (2000 rpm, 2 minutes, 4°C) and then pellets were washed with 750 µl 100% ethanol. Proteins were solubilized in 50 µl 1 M Tris pH 8 and 100 µl x2 SDS-PAGE loading buffer (60 mM Tris pH 6.8, 2% SDS, 10% glycerol, 0.2% bromophenol blue) and boiled for 5 minutes at 95°C. Insoluble material was removed by centrifugation (14000 rpm, 5 minutes, room temperature) and the supernatant either stored at -20°C or loaded immediately onto a SDS-PAGE mini-gel. Samples were either run on an 8% acrylamide gel in Tris-Glycine SDS running buffer using the Bio Rad Mini PROTEAN 3 system or on Pre-Cast 4-12% Bis-Tris gels in NuPAGE MOPS (all Invitrogen). SDS-PAGE gels were transferred to polyvinylidene fluoride transfer membrane (Hybond-P, Amersham Biosciences) in either the Bio-Rad Mini Trans-Blot Electrophoretic Transfer Cell or by using the XCell SureLock Mini Cell Transfer module (Invitrogen). The Bio-Rad system was used in conjunction with Tris-Glycine blotting buffer (National Diagnostics) containing 20% methanol, and run for 1h at 200V or overnight at 30V. The NuPAGE system was used with NuPAGE transfer buffer containing 20% methanol and run for 1h at 30V.

Purification of His-tagged proteins
For His-tag purifications 100 OD of cells were harvested (4000 rpm, 2 minutes), washed once with water, and the cell pellet frozen at -80°C. The cell pellet was resuspended in 500 µl Buffer A (8 M Urea, 100 mM NaH₂PO₄, 10 mM Tris HCl, 0.05% Tween pH 8.1), an equal volume of glass beads were added, and the cells lysed by 1 45s cycle, power 6 in a FastPrep FP120 (BIO 101) machine. Tubes were pierced with a hot needle and placed onto fresh eppendorfs and spun (1000 rpm, 2 minutes) to collect lysate minus glass beads. The glass beads were washed with 1 ml 5% TCA and this was added to the lysate, and mixed by pipetting. The precipitated proteins were collected by centrifugation (14000 rpm, 10 minutes 4°C) and then pellets were washed with 750 µl 100% ethanol. Proteins were solubilized in 50 µl 1 M Tris pH 8 and 100 µl x2 SDS-PAGE loading buffer (60 mM Tris pH 6.8, 2% SDS, 10% glycerol, 0.2% bromophenol blue) and boiled for 5 minutes at 95°C. Insoluble material was removed by centrifugation (14000 rpm, 5 minutes, room temperature) and the supernatant either stored at -20°C or loaded immediately onto a SDS-PAGE mini-gel. Samples were either run on an 8% acrylamide gel in Tris-Glycine SDS running buffer using the Bio Rad Mini PROTEAN 3 system or on Pre-Cast 4-12% Bis-Tris gels in NuPAGE MOPS (all Invitrogen). SDS-PAGE gels were transferred to polyvinylidene fluoride transfer membrane (Hybond-P, Amersham Biosciences) in either the Bio-Rad Mini Trans-Blot Electrophoretic Transfer Cell or by using the XCell SureLock Mini Cell Transfer module (Invitrogen). The Bio-Rad system was used in conjunction with Tris-Glycine blotting buffer (National Diagnostics) containing 20% methanol, and run for 1h at 200V or overnight at 30V. The NuPAGE system was used with NuPAGE transfer buffer containing 20% methanol and run for 1h at 30V.

Immunological detection
Membranes were blocked in 5% skimmed milk powder in PBS with 0.1% Tween 20 (PBS-T) for 1h or overnight at 4°C, then incubated with either mouse monoclonal anti-c-myc IgG1x antibody 9E10 (Roche) or anti-HA IgG1 antibody 12CA5 (Roche) in blocking solution for between 1 hour at room temperature to overnight at 4°C. Following several washes in PBS-T, membranes were incubated with the sheep anti-mouse IgG Horse-Redish Peroxidase-linked antibody (GE Healthcare) at a 1/10000 dilution in blocking solution. After several further washes in PBS-T, membranes were incubated with the ECL Plus Western Blotting Detection System (GE Healthcare) followed by exposure to ECL Hyperfilm (GE Healthcare), to detect the secondary antibody.

Purification of SUMOylated proteins for mass spectrometry
Nine liters of culture harvested at O.D.0.9 were used in each experiment. Cells were pelleted and washed in chilled PBS with 20 mM NEM. Cells were resuspended in 250 ml of lysis buffer (1.85 N NaOH, 1.85% β-mercaptoethanol) and incubat-
ed in ice for 30 minutes. TCA (trichloroacetic acid) was added to a final concentration of 25% and kept in ice for a further 30 minutes. Sample were spun at 15,000 rpm for 10 minutes in acetone-resistant tubes. The pellet washed with 250 ml acetone and spun again. The pellet was air-dried for 10 minutes and resuspended in 100 ml of binding buffer (6 M guanidine hydrochloride, 100 mM NaH2PO4, 10 mM Tris-HCl, 20 mM NEM, pH 8) by vortexing for 10 minutes at RT. Resuspended proteins were recovered after a spin at 15,000 rpm for 10 minutes. All spins were performed at 4°C. Protein extraction after elution from agarose beads (see Purification of His-tagged proteins) was adjusted to pH 7 and filtered through a 50 kDa MWCO protein filter (Vivaspin 20, GE Healthcare) up to 1 ml (one filter was used for every 12 ml of elution sample). Buffer A (w/o Tween®-20) was added in a proportion 10:1 and the sample filter again up to 1 ml. Sample was placed in a 10 kDa MWCO protein filter (Vivaspin 2, GE Healthcare), and filtered up to a volume 1/100000 of the original volume of the culture (i.e. 90 µl from a 9 l culture). Protein sample was then digested with Lys-C endonuclease (Roche) at final concentration of 0.02 µg/µl. The reaction was incubated at 37°C for 12 hours. A standard 4% β-mercaptoethanol/2% SDS loading buffer was added to the sample, left on the bench for 15 minutes and loaded into a 12% Tris-Bis NuPAGE® gel (Invitrogen). Gel was run using an Invitrogen system with NuPAGE® MOPS SDS running buffer (Invitrogen) for 2 h at 140 V. After the running, the gel was stained using SimplyBlue™ SafeStain (Invitrogen) following manufacturer’s protocol. The band of interest was cut from the gel. Peptides were in-gel digested with Lys-C endonuclease (Roche) following manufacturer’s protocol. The band of interest was cut from the gel. Peptides were in-gel digested using trypsin. First, gel bands were destained using 40% acetonitrile, 200 mM ammonium bicarbonate. Gel pieces were shrunk with acetonitrile and enough trypsin solution (10 ng/µl trypsin in 10% acetonitrile in 50 mM ammonium bicarbonate) was added to cover the gel pieces. Gel pieces were incubated overnight at 37°C. Peptides were extracted with 5% formic acid in acetonitrile. The extracts were dried to completeness in a vacuum concentrator and resuspended in 10 µl of 0.1% trifluor acetic acid (TFA). Peptide mixtures were loaded onto a 75 µm i.d. c18 column (pepmap, ThermoFisher) and eluted over a 60 minute gradient stretching from 3% acetonitrile to 50% acetonitrile (both in 0.1% formic acid) at 300 nl/minute using an RSLC Ultimate 3000 onto a LTQ-Orbitrap-Velos (both ThermoFisher). Profile spectra were acquired in the Orbitrap (60,000 resolution) and the 6 most intense ions excluding +1 charge state were selected for fragmentation in the linear ion trap (LTQ). Raw data was processed using MaxQuant and searched using the embedded Andromeda routine [14]. Data was searched against a *Saccharomyces cerevisiae* database (to which the recombinant Smt3 was manually added) using default settings. Mass tolerances for Ms and MS/MS data were 10 part per million (ppm) and 0.5 Da respectively. Methionine oxidation and glygly modification of lysine were allowed as variable modification. The false discovery rates at the peptide, protein and site level were 0.005, 0.01 and 0.01 respectively. Statistical analysis and charts were created using the SPSS software environment. Graphic representation of the local context of the SUMO-acceptor lysines was created using ice-Logo [23]. SUMO motifs were assigned to sumo-acceptor lysines using the 3of5 web application for pattern matching [24]. Subcellular distribution and function of SUMOylated proteins was obtained from the Gene Ontology (GO) slim mapper part of the *Saccharomyces* Genome Database (SGD, http://www.yeastgenome.org).

**ACKNOWLEDGMENTS**

We would like to thank members of the Aragon laboratory for discussions and critical reading of the manuscript. This work was funded by the Intramural research programme of the Medical Research Council UK and ERC grant (Chrm Stability-202337).

**SUPPLEMENTAL MATERIAL**

All supplemental data for this article are available online at www.microbialcell.com.

**CONFLICT OF INTEREST**

The authors declared that there is no conflict of interest arising from this work.

**COPYRIGHT**

© 2017 Esteras et al. This is an open-access article released under the terms of the Creative Commons Attribution (CC BY) license, which allows the unrestricted use, distribution, and reproduction in any medium, provided the original author and source are acknowledged.

Please cite this article as: Miguel Esteras, I-Chun Liu, Ambrosius P. Snijders, Adam Jarmuz & Luis Aragon (2017). Identification of SUMO conjugation sites in the budding yeast proteome. *Microbial Cell* 4(10): 331-341. doi: 10.15698/mic2017.10.593

**REFERENCES**

1. Melchior F (2000). SUMO - nonclassical ubiquitin. *Annu Rev Cell Dev Biol* 16:591-626.

2. Geiss-Friedlander R, Melchior F (2007). Concepts in sumoylation: a decade on. *Nat Rev Mol Cell Biol* 8(12): 947-956.

3. Bylebyl GR, Belichenko I, Johnson ES (2003). The SUMO isopeptidase Ulp2 prevents accumulation of SUMO chains in yeast. *J Biol Chem* 278(45): 44113-44120.

4. Saitoh H, Hinchey J (2000). Functional heterogeneity of small ubiquitin-related protein modifiers SUMO-1 versus SUMO-2/3. *J Biol Chem* 275(9): 6252-6258.

5. Vertegaal AC, Andersen JS, Ogg SC, Hay RT, Mann M, Lamond Al (2006). Distinct and overlapping sets of SUMO-1 and SUMO-2 target proteins revealed by quantitative proteomics. *Mol Cell Proteomics* 5(12): 2298-2310.

6. Liang YC, Lee CC, Yao YL, Lai CC, Schmitz ML, Yang WM (2016). SUMOS, a Novel Poly-SUMO Isoform, Regulates PML Nuclear Bodies. *Sci Rep* 6:26509.

7. Owerbach D, McKay EM, Yeh ET, Gabbay KH, Bohren KM (2005). A proline-90 residue unique to SUMO-4 prevents maturation and sumoylation. *Biochem Biophys Res Commun* 337(2): 517-520.

**CONFLICT OF INTEREST**

The authors declared that there is no conflict of interest arising from this work.

**COPYRIGHT**

© 2017 Esteras et al. This is an open-access article released under the terms of the Creative Commons Attribution (CC BY) license, which allows the unrestricted use, distribution, and reproduction in any medium, provided the original author and source are acknowledged.

Please cite this article as: Miguel Esteras, I-Chun Liu, Ambrosius P. Snijders, Adam Jarmuz & Luis Aragon (2017). Identification of SUMO conjugation sites in the budding yeast proteome. *Microbial Cell* 4(10): 331-341. doi: 10.15698/mic2017.10.593

**REFERENCES**

1. Melchior F (2000). SUMO - nonclassical ubiquitin. *Annu Rev Cell Dev Biol* 16:591-626.

2. Geiss-Friedlander R, Melchior F (2007). Concepts in sumoylation: a decade on. *Nat Rev Mol Cell Biol* 8(12): 947-956.

3. Bylebyl GR, Belichenko I, Johnson ES (2003). The SUMO isopeptidase Ulp2 prevents accumulation of SUMO chains in yeast. *J Biol Chem* 278(45): 44113-44120.

4. Saitoh H, Hinchey J (2000). Functional heterogeneity of small ubiquitin-related protein modifiers SUMO-1 versus SUMO-2/3. *J Biol Chem* 275(9): 6252-6258.

5. Vertegaal AC, Andersen JS, Ogg SC, Hay RT, Mann M, Lamond Al (2006). Distinct and overlapping sets of SUMO-1 and SUMO-2 target proteins revealed by quantitative proteomics. *Mol Cell Proteomics* 5(12): 2298-2310.

6. Liang YC, Lee CC, Yao YL, Lai CC, Schmitz ML, Yang WM (2016). SUMOS, a Novel Poly-SUMO Isoform, Regulates PML Nuclear Bodies. *Sci Rep* 6:26509.

7. Owerbach D, McKay EM, Yeh ET, Gabbay KH, Bohren KM (2005). A proline-90 residue unique to SUMO-4 prevents maturation and sumoylation. *Biochem Biophys Res Commun* 337(2): 517-520.
8. Johnson ES, Blobel G (1999). Cell cycle-regulated attachment of the ubiquitin-related protein SUMO to the yeast septins. J Cell Biol 147(5): 981-994.

9. Bernier-Villamor V, Sampson DA, Matunis MJ, Lima CD (2002). Structural basis for E2-mediated SUMO conjugation revealed by a complex between ubiquitin-conjugating enzyme Ubc9 and RanGAP1. Cell 108(3): 345-356.

10. Yang XI, Gregoire S (2006). A recurrent phospho-sumoyl switch in transcriptional repression and beyond. Mol Cell 23(6): 779-786.

11. Wohlschlegel JA, Johnson ES, Reed SI, Yates JR, 3rd (2006). Improved identification of SUMO attachment sites using C-terminal SUMO mutants and tailored protease digestion strategies. J Proteome Res 5(4): 761-770.

12. Cox I, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M (2011). Andromeda: a peptide search engine integrated into the MaxQuant environment. J Proteome Res 10(4): 1794-1805.

13. Albuquerque CP, Yeung E, Ma S, Fu T, Corbett KD, Zhou H (2015). A Chemical and Enzymatic Approach to Study Site-Specific Sumoylation. PLoS One 10(12): e0143810.

14. Denison C, Rudner AD, Gerber SA, Bakalarski CE, Moazed D, Gygi SP (2005). A proteomic strategy for gaining insights into protein sumoylation in yeast. Mol Cell Proteomics 4(3): 246-254.

15. Zhou W, Ryan JJ, Zhou H (2004). Global analyses of sumoylated proteins in Saccharomyces cerevisiae. Induction of protein sumoylation by cellular stresses. J Biol Chem 279(31): 32262-32268.

16. Zhou W, Ryan JJ, Zhou H (2004). Global analyses of sumoylated proteins in Saccharomyces cerevisiae. Induction of protein sumoylation by cellular stresses. J Biol Chem 279(31): 32262-32268.

17. Silver HR, Nissley JA, Reed SH, Hou YM, Johnson ES (2011). A role for SUMO in nucleotide excision repair. DNA Repair (Amst) 10(12): 1243-1251.

18. Chen XL, Silver HR, Xiong L, Belichenk I, Adegite C, Johnson ES (2007). Topoisomerase I-dependent viability loss in Saccharomyces cerevisiae mutants defective in both SUMO conjugation and DNA repair. Genetics 177(1): 17-30.

19. Nathan D, Ingvarsosottir K, Sterner DE, Bylebly GR, Dokmanovic M, Dorsey JA, Whelan KA, Krsmanovic M, Lane WS, Meluh PB, Johnson ES, Berger SL (2006). Histone sumoylation is a negative regulator in Saccharomyces cerevisiae and shows dynamic interplay with positive-acting histone modifications. Genes Dev 20(8): 966-976.

20. Sacher M, Pfander B, Hoege C, Jentsch S (2006). Control of Rad52 recombination activity by double-strand break-induced SUMO modification. Nat Cell Biol 8(11): 1284-1290.

21. Rodriguez MS, Dargemont C, Hay RT (2001). SUMO-1 conjugation in vivo requires both a consensus modification motif and nuclear targeting. J Biol Chem 276(16): 12654-12659.

22. Colaert N, Helsens K, Martens L, Vandekerckhove J, Gevaert K (2009). Improved visualization of protein consensus sequences by iceLogo. Nat Methods 6(11): 786-787.

23. Mehrle A, Rosenfelder H, Schupp I, del Val C, Arlt D, Hahnke F, Bechtel S, Simpson J, Hofmann O, Hide W, Glatting KH, Huber W, Pepperkok R, Pousta A, Wiemann S (2006). The LIFEdb database in 2006. Nucleic Acids Res 34(90001): D415-418.