Screen for multi-SUMO–binding proteins reveals a multi-SIM–binding mechanism for recruitment of the transcriptional regulator ZMYM2 to chromatin

Elisa Aguilar-Martinez, Xi Chen, Aaron Webber, A. Paul Mould, Anne Seifert, Ronald T. Hay, and Andrew D. Sharrocks

Protein SUMOylation has emerged as an important regulatory event, particularly in nuclear processes such as transcriptional control and DNA repair. In this context, small ubiquitin-like modifier (SUMO) often provides a binding platform for the recruitment of proteins via their SUMO-interacting motifs (SIMs). Recent discoveries point to an important role for multivalent SUMO binding through multiple SIMs in the binding partner as exemplified by poly-SUMOylation acting as a binding platform for ubiquitin E3 ligases such as ring finger protein 4. Here, we have investigated whether other types of protein are recruited through multivalent SUMO interactions. We have identified dozens of proteins that bind to multi-SUMO platforms, thereby uncovering a complex potential regulatory network. Multi-SUMO binding is mediated through multi-SIM modules, and the functional importance of these interactions is demonstrated for the transcriptional corepressor ZMYM2/ZNF198 where its multi-SUMO–binding activity is required for its recruitment to chromatin.

SUMO | ZMYM2 | chromatin | SIM | ZNF198

Protein modification with small ubiquitin-like modifier (SUMO) has emerged as a major regulatory event that impacts on the activities of hundreds of proteins associated with a diverse array of biological activities (reviewed in refs. 1–3). In particular, a number of roles have been ascribed to nuclear functions, predominantly in the area of chromatin function, DNA repair, and transcriptional regulation (reviewed in ref. 4). SUMO can act at the molecular level in many different ways but a major mode of action is through providing an additional binding surface for attracting coregulatory proteins to a SUMOylated protein (reviewed in refs. 2 and 5). These interactions are driven by short hydrophobic regions on binding partners known as SUMO-interacting motifs (SIMs) (6, 7). However, their binding affinity for SUMO is generally in the micromolar range (e.g., ref. 6), meaning that additional contacts between the SUMOylated protein and its binding partner are needed to drive interactions. Alternatively, by incorporating multiple SIMs, this can be sufficient to recruit a protein to a SUMOylated binding partner, but multiple SUMO moieties must be present in the form of poly-SUMO chains (reviewed in ref. 8). This is exemplified by PML, which becomes poly-SUMOylated upon treatment of cells with arsenic trioxide, and results in the recruitment of the ubiquitin ligase ring finger protein 4 (RNF4), which contains four closely juxtaposed SIMs (9, 10). More recently, mass spectrometry studies have led to the identification of hundreds of proteins that become poly-SUMOylated upon heat shock, and can then bind to RNF4 (11). The functional importance of these chains has been elucidated in yeast, where numerous roles have been attributed to poly-SUMOylation, including a major defect in higher-order chromatin structure maintenance (12). Thus, the promotion of poly-SUMOylation represents a potential regulatory route for specifically recruiting binding partners that contain multiple SIM motifs.

Many proteins contain multiple sites for modification by SUMO. For example, a bioinformatics approach identified 448 human proteins containing two or more motifs corresponding to the extended negatively charged amino acid-dependent SUMOylation motif (NDSM) (13) Thus, there is a huge potential for widespread multi-SUMOylation of proteins to occur. Indeed, several of these proteins have been shown to be SUMOylated on multiple sites, including megakaryoblastic leukemia (translocation) 1 (MLL1) (14), CREB-binding protein (CBP) (15), and PEA3/ETV4 (16). Furthermore, two recent proteomic studies emphasize the potential for more widespread multi-SUMOylation as they found that a large proportion of all SUMO2-modified proteins contain two or more modified lysine residues, with one study reporting that over one-half of the proteins fell into this category, with proteins like ZNF451 having 40 such sites (17, 18). In principle, multi-SUMOylation could provide a platform for recruiting proteins containing multiple SIMs as has already been observed for poly-SUMOylation in the form of linear chains. Moreover, the aggregation of several SUMOylated proteins into protein complexes, as seen in many transcriptional regulatory complexes, provides yet more potential for presenting a multi-SUMO–binding surface for recruiting coregulatory partners. Indeed, a recent study demonstrated that SUMOylation of many different proteins involved in homologous recombination acts synergistically to promote efficient DNA repair (19).

Significance

Small ubiquitin-like modifier (SUMO) is thought to function by acting as a protein recruitment platform. To date, studies have focused on the role of mono-SUMO and poly-SUMO in the form of linear chains. However, recent findings suggest a role for multi-SUMOylation where several SUMO moieties are spread across numerous proteins found at sites of DNA damage. Here, we used a novel screen to identify dozens of multi-SUMO–binding proteins. We investigated one of these in detail and demonstrate that a multi-SIM–containing SUMO–binding module is required for recruitment of the transcriptional regulator ZMYM2 to chromatin. Because little is known about the function of multi-SUMOylation and multi-SIM–binding proteins, this represents an important conceptual advance in our thinking about how protein SUMOylation might exert its molecular effects.

Author contributions: E.A.-M., X.C., R.T.H., and A.D.S. designed research; A.D.S. coordinated the project; E.A.-M., X.C., A.W., A.P.M., and A.S. performed research; E.A.-M., X.C., A.W., A.P.M., A.S., R.T.H., and A.D.S. analyzed data; and E.A.-M., X.C., and A.D.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The data reported in this paper have been deposited in the ArrayExpress database (accession nos. E-MEXP-3880 [microarray data], E-MTAB-2701 [ChIP-seq data], and E-MTAB-3084 [ChIP-seq data]).

1To whom correspondence should be addressed. Email: a.d.sharrocks@manchester.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1509716112/-/DCSupplemental.

Edited by James L. Manley, Columbia University, New York, NY, and approved July 17, 2015 (received for review May 20, 2015)
In this study, we therefore investigated whether multi-SUMOylation could provide a means for driving protein recruitment through multi-SIM–containing protein partners. We used a novel multi-SUMO scaffold to identify over 100 potential multi-SUMO–binding proteins. We verified several of these interactions and demonstrated the importance of a multi-SUMO–binding multi-SIM module for anchoring the transcriptional corepressor protein zinc finger, MYM-type 2 (ZMYM2)/zinc finger protein 198 (ZNF198) to chromatin.

Materials and Methods

Information regarding the plasmids, cell lines, transfections, and reporter gene assays used in this study can be found in SI Materials and Methods.

Microarray Analysis, RT-PCR, and siRNAs. The microarray expression data have been submitted to ArrayExpress (accession no. E-MEXP-3880). Methods of the microarray and quantitative RT-PCR can be found in SI Materials and Methods. SMARTpool siRNA duplexes against ZMYM2,UBE2J1/UBC9, and non-targeting control (NTC) were purchased from Dharmacon.

ChIP-qPCR and ChIP-seq Analysis. ChIP-qPCR was performed at least in duplicate, from at least two independent experiments, and analyzed as described previously (20).

ChIP-seq accession numbers are E-MTAB-2701 (ZMYM2 and FOXO3) and E-MTAB-3084 (SUMO3). Methodology can be found in SI Materials and Methods.

Protein Analysis. All of the proteomic methodology used in this paper, including protein purification, pulldown, SIM peptide competition, nuclear extract preparation, mass spectrometry, Western blotting, immunofluorescence, and surface plasmon resonance can be found in SI Materials and Methods.

SIM Motif Search. A motif string search script was written in Python 3. SIM motifs were searched in either the proteins identified by mass spectrometry or in 1,000 different sets of random proteins with similar length distributions to those identified in this study. Random protein sequences were retrieved from UniProt. To determine the significance of enrichment of this motif in the identified proteins, a one-sample t test was carried out. For annotating our list of proteins identified in the mass spectrometry assay, the presence of SIM motifs was determined using the program GPS-SUMO (21) set to high threshold.

Results

Identification of Multi-SUMO–Binding Proteins. Previous studies established that RNF4 binds to proteins containing multiple SUMO moieties in the form of linear poly-SUMO chains, through its multiple SIM motifs (9). However, it is equally plausible that multi-SIM–containing proteins might be able to recognize multiple single SUMO moieties conjugated to different sites in a protein or protein complex (Fig. L4). To date, RNF4 and the closely related protein Arkadia are the only known multi-SIM–containing proteins with the potential to bind poly- or multi-SUMOylated proteins (9, 22). We therefore sought to investigate how widespread this phenomenon is likely to be, by searching for proteins that could bind to a multi-SUMO platform. To identify multi-SUMO–binding proteins, we created a fusion protein, containing SUMO fused to GST through a linker region corresponding to the coiled-coil pentamization domain of cartilage oligomeric matrix protein (COMP) (23). This allowed us to create a scaffold which presents five copies of SUMO in a form that resembles multi-SUMOylation rather than the single SUMO molecules presented on a standard GST-fusion protein (Fig. 1B). As expected, in comparison with GST-SUMO fusion proteins, GST-COMP-SUMO fusion proteins migrated as higher molecular-weight species in native polyacrylamide gels (Fig. 1C, Bottom; compare lanes 2 and 3 with lanes 5 and 6). To test whether GST-COMP-SUMO acted in the predicted manner, we examined interactions with RNF4. Little binding of RNF4 to GST-COMP or GST-SUMO3 was observed, but binding was readily observable to GST-COMP-SUMO3 (Fig. 1D). Binding appears to be SUMO paralog specific as comparatively little RNF4 binding was observed to GST-COMP-SUMO1 (Fig. S1). Importantly, upon mutation of the SIM2 motif in RNF4, this mutant protein showed reduced binding to GST-COMP-SUMO3 (Fig. 1E), which resembled the reduced binding seen with poly-SUMO chains (9). These findings therefore establish GST-COMP-SUMO3 as a suitable platform for use in the identification of multi-SUMO–binding proteins. To identify novel multi-SUMO–binding proteins, a pulldown assay was carried out using GST-COMP-SUMO3 as bait, and HeLa nuclear extracts as the source of potential interacting proteins. We first removed nonspecific binding proteins by incubating the nuclear extract with glutathione-agarose bead-immobilized GST-COMP (Fig. 1F). Proteins remaining in the supernatant were subsequently incubated with glutathione-agarose bead-immobilized GST-COMP-SUMO3. Following extensive washing, the remaining bound proteins were eluted from the beads by cleaving the GST tag and subjected to SDS/PAGE (Fig. 1G, lane 8). Bound proteins were then eluted from the gel, and their identities were determined by mass spectrometry. We identified a total of 163 proteins, over one-half of which can be considered higher confidence due to not being commonly identified components of coimmunoprecipitation experiments (Dataset S1). This list includes several involved in protein SUMOylation (PIAS1, SENP1, SENP2, and SENP3) and also many associated with transcriptional regulation and DNA repair, two processes that are often associated with SUMOylation (Fig. 1H and Fig. S24; reviewed in ref. 4). Many of these proteins have previously been shown to interact either physically or functionally, thereby suggesting the existence of important underlying SUMO-dependent networks (Fig. S2B). Several of the proteins had previously been shown to bind to monomeric SUMO including BL1 (24) and RCO1 (25).

To verify the mass spectrometry results, we transfected HEK293T cells with expression constructs for epitope-tagged versions of PTRF, BLM, and ZMYM2/ZNF198, and determined whether GST-COMP-SUMO3 could bind to any of these in the context of cell lysates. All three of these proteins bound strongly to GST-COMP-SUMO3 (Fig. 2A, lane 7). In contrast, little binding was seen to GST alone or GST-COMP (Fig. 2B, lanes 2 and 5), and, importantly, little binding was seen to GST-SUMO3 (Fig. 2C, lane 4), demonstrating the importance of presenting SUMO3 in a multimeric form on a pentameric scaffold. Interestingly, there also appears to be SUMO paralog specificity, as comparatively little binding to GST-COMP-SUMO1 was observed (Fig. 2A, lane 6). To further demonstrate the specificity of these interactions, we examined ZBTB33/Kaiso and demonstrated that, although it preferentially binds to GST-COMP-SUMO3 over GST-COMP, the closely related zinc finger and ZBT domain-containing protein ZBTB4 is unable to bind to GST-COMP-SUMO3 (Fig. 2B).

We also identified proteins like SRBC that could not bind specifically to GST-COMP-SUMO3, unless coexpressed with its binding partner PTRF (26) (Fig. S3A and B). This raised the possibility that some of the interactions we have observed might be indirect and mediated by other partners in multiprotein complexes. We therefore further investigated whether ZBTB33 binding to GST-COMP-SUMO3 could be observed when it was expressed in vitro. Again, strong binding was only observed with GST-COMP-SUMO3 among the GST-fusion proteins tested (Fig. S3C). Similarly, we demonstrated direct interactions with multi-SUMO by demonstrating that recombinant ZMYM2(1–200) could bind specifically to recombinant His-tagged COMP-SUMO3 (Fig. S3D).

In summary, we have identified a large number of novel multi-SUMO–binding proteins, and at least some of these are likely to directly interact with SUMO.
Characterization of the SIM–Multi-SUMO Binding Interactions. SUMO2 and 3 have been shown to contain a conserved binding surface that is used for binding to SIMs, which conform to the typical consensus of a short hydrophobic core surrounded by acidic residues such as found in PIAS2 (6, 7). This surface on SUMO2/3 contains a mixture of hydrophobic and basic residues that acts as a reciprocal interface for SIMs (27) (Fig. 3A). We therefore asked whether the proteins that we identified bind through this same surface of SUMO by using a mutant SUMO3 derivative, SUMO3(3A), in which three residues in the SIM binding surface have been replaced with alanines (Fig. 3A). As expected, RNF4 binding to GST-COMP-SUMO3 was reduced upon mutation of the SIM–interaction surface on SUMO3 (Fig. 3B). Similarly, binding of ZMYM2 to SUMO3 was also severely compromised upon introduction of these mutations (Fig. 3B), and three more of the proteins tested, ZBTB33, PTRF, and PIAS1, showed the same binding profile (Fig. 3C). As predicted, SENP3 showed identical binding to wild-type (WT) and mutant versions of SUMO3, in agreement with the observation that SUMO proteases use additional contact regions in their interaction with SUMO (28, 29). Thus, SENP3 uses a different mode of interaction to the other proteins identified in our screen.
Given the observed binding to the SIM interaction surface on SUMO, this indicated that many of the proteins that we identified as binding to multi-SUMO might do so through canonical SIM-like motifs. We therefore searched the sequences of the proteins we identified for the well-characterized motifs conforming to SIM2 in RNF4 [ED][ED][IVL][IVL][DE][IVL][xx][ED] and the SIM in PIAS2 [DE][IVL][IVL][DE][VIL][xx][ED][ED]. These SIM motifs were enriched in the mass spectrometry-identified proteins compared with random background data sets (values of $P < 0.0015$ and $0.00018$, respectively). Moreover, 14 of the proteins contained two or more canonical SIMs, consistent with a potential role in multi-SUMO binding (Dataset S1). To investigate this phenomenon further, we used an alternative approach, involving a peptide competition with a peptide corresponding to the SIM in PIAS2, and incorporated this into binding assays with WT GST-COMP-SUMO3. As expected, increasing amounts of this peptide disrupted interactions between RNF4 and GST-COMP-SUMO3 (Fig. 3D). Similarly, the SIM peptide competed with ZMYM2 for binding to GST-COMP-SUMO3 but was unable to compete with binding of SENP3 (Fig. 3E, top two panels). This is consistent with ZMYM2 and SENP3 each using different interaction surfaces on SUMO3.

Together, these results demonstrate that many of the factors that we have identified as multi-SUMO–binding proteins, share similar binding properties to those previously defined for canonical SIM–SUMO interactions. However, there are likely other proteins typified by SENP3 that use alternative binding modes.

**ZMYM2 Contains Multiple Functional SIMs.** RNF4 contains several SIMs, which enables it to make simultaneous contact with multiple SUMO moieties (9). Thus, it appears likely that, by analogy with RNF4, many of the numerous multi-SUMO–binding proteins we have identified will operate a similar interaction mechanism. One such protein is ZMYM2, which contains two motifs (SIM1 and 2), that closely resemble the canonical SIMs found in RNF4 and PIAS2 with a hydrophobic core and multiple flanking acidic residues and a third motif with only one flanking acidic residue (SIM3) (Fig. 4A). To establish the importance of these motifs in ZMYM2 for multi-SUMO binding, we mutated them by substituting three or four residues of the core hydrophobic motifs for alanines and tested the resulting proteins for binding to GST-COMP-SUMO3. The mutation of individual motifs had differing effects on binding, with SIM1 and SIM3 mutations having moderate and little effect, whereas mutating SIM2 virtually eliminated multi-SUMO binding (Fig. 4B). Such pronounced effects of a single dominant SIM have been observed for RNF4 and Arkadia (9, 22). However, combinatorial mutation of SIM1 and SIM3 caused a loss of binding to GST-COMP-SUMO3 (Fig. 4C), demonstrating their functional importance in multi-SUMO binding.

Next, we asked whether ZMYM2 and several other multi-SUMO proteins we had identified preferred the presentation of multiple SUMO moieties on either a linear poly-SUMO chain or on a multi-SIM scaffold. Although RNF4 preferentially binds multi-SUMO rather than mono-SUMO (Fig. 1), it showed a clear preference for poly-SUMO chains found in GST-SUMO3 × 4 over the multi-SUMO GST-COMP-SUMO3 scaffold (Fig. 4C). In contrast, ZMYM2 bound strongly to both forms of multi-merized SUMO, and ZBTB33 and PTF showed intermediate relative levels of binding to the GST-COMP-SUMO3 scaffold (Fig. 4C). Given the strong multi-SIM–mediated binding characteristics of ZMYM2, we focused on this further. A previous study identified the zinc finger motifs of ZMYM2 as necessary and sufficient for binding to poly-SUMO chains (30). Importantly, full-length ZMYM2 lacking SIM2, which is critically important for binding the multi-SUMO scaffold, still retained poly-SUMO chain binding activity (Fig. 4D). We therefore focused on the N-terminal region of ZMYM2, which contains two functional canonical SIMs, and asked whether this was sufficient for multi-SUMO binding. This region has previously been shown to be required for colocalization with PML and copurification of SUMOylated proteins (30). Although ZMYM2(1–200) was able to efficiently bind directly to His-tagged COMP-SUMO3 in pull-down assays, it was unable to bind to either dimeric or tetrameric SUMO chains (30.E). Furthermore, surface plasmon resonance (SPR) studies demonstrated that ZMYM2(1–200) showed specific binding to both multi-SUMO in COMP-SUMO3 and poly-SUMO in SUMO3 × 4 (Fig. 4E and Fig. S3F). However, only COMP-SUMO3 showed stable binding to ZMYM2 with the majority of COMP-SUMO3 binding having a slow dissociation rate ($k_d = 6.05 \pm 1.11 \times 10^{-2} \text{s}^{-1}$), whereas the poly-SUMO chain had a very rapid dissociation rate ($k_d = 7.51 \pm 1.64 \times 10^{-2} \text{s}^{-1}$).

Finally, we asked whether we could find evidence for interactions between ZMYM2 and any multi-SUMO–containing proteins in vivo. To do this, we created a doxycycline-inducible

![Figure 2](image-url) Fig. 2. Validation of multi-SUMO–binding interactions. (A) GST pulldown analysis of PTF, BLM, and ZMYM2 binding to the indicated GST-fusion proteins. Epitope-tagged PTF, BLM, or ZMYM2 were overexpressed in HEK293T cells, and total cell lysates were used for the pulldown. Proteins were detected by IB using anti-FLAG (PTF) or anti-Myc (BLM and ZMYM2) antibodies. Arrows indicate the positions of bands corresponding to the full-length proteins. Bait proteins were stained with Ponceau (Bottom). (B) Interaction of ZBTB33 or ZBTB4 with COMP-SUMO3 was analyzed by GST-pulldown. Total cell extracts of transfected HEK293T cells were used for the pulldown. Input and precipitated ZBTB33 and ZBTB4 were detected using an anti-FLAG antibody. Proteins used as bait were stained with Ponceau (Bottom).
stable cell line expressing a FLAG-tagged version of SUMO3 lacking a key lysine residue needed for polychain formation (K11R). Furthermore, to enhance our ability to detect multi-SUMOylated species, we also introduced the Q90P mutation into SUMO3, which renders it noncleavable from substrates by SENPs (31, 32). As predicted, SUMO3 was detected in cells expressing SUMO3(K11R/Q90P) as a series of high-molecular-weight SUMOylated species (Fig. S3G, lane 2). This is suggestive of the generation of a mixture of multi and poly-SUMO species. Importantly, immunoprecipitation of ZMYM2 from these cells yielded coprecipitation of a number of these high-molecular-weight SUMOylated species (Fig. S3G, lane 4). Thus, multi-SUMO–binding proteins show a range of binding propensities for different multi-SUMO scaffolds, with ZMYM2 containing an N-terminal multi-SUMO-binding module and exhibiting the greatest ability to bind to a multi-SUMO scaffold.

ZMYM2 has been shown to form part of the CoREST transcriptional corepressor complex (33, 34), and hence likely plays a role in transcriptional repression. To establish the functional importance of the SIMs, we first analyzed the subcellular distribution of WT ZMYM2, the SIM2 mutant and a form with all three SIMs mutated. However, all of the forms of ZMYM2 remained nuclear (Fig. S4A). Next, we asked whether the nuclear function of ZMYM2 as a co-repressor protein was perturbed upon mutation of its SIMs. Stable U2OS cell lines were created containing an inducible form of FLAG-tagged WT ZMYM2 or mutant forms with either SIM2 or all three SIMs mutated. Equal expression of all three forms of ZMYM2 was observed, and this expression was confirmed by RT-qPCR analysis (Fig. S4B). We then took an unbiased approach to identify the effects of disrupting the multi-SUMO–binding activity of ZMYM2 by using microarray analysis to assess the differences in gene expression in cells ectopically expressing WT or SIM2 mutant version of ZMYM2. We found that 168 genes were up-regulated by more than fivefold and 626 genes were down-regulated by more than fivefold and 626 genes were down-regulated by more than fivefold in cells expressing ZMYM2(SIM2mut) (Dataset S2), indicating a pleiotropic change in the gene expression profiles in these cells. As ZMYM2 is thought to be involved in transcriptional repression (35), we focused on genes that were up-regulated in cells expressing ZMYM2(SIM2mut) to identify effects that were potentially caused as a direct effect of disrupting ZMYM2 function. Three genes in this category, CHGB, EPHX2, and TMEM154, were selected, and their increased expression in cells expressing ZMYM2(SIM2mut) was verified by RT-qPCR analysis (Fig. S4D). In contrast, the control genes ARMC7, CCNF, and RNF7 showed little evidence of up-regulation. Importantly, the
Fig. 4. Mapping the SIMs in ZMYM2. (A) Schematic illustration of ZMYM2 showing the location and sequence of three putative SIMs. Acidic amino acids are colored in red. Regions containing hydrophobic amino acids mutated to alanine are boxed. (B) GST pulldown analysis of the indicated GST-fusion proteins to WT or mutant forms of ZMYM2. Total cell lysates of HEK293T cells overexpressing the ZMYM2 proteins were used for the PD. ZMYM2 was detected by IB using anti-myc antibody. The 5% input is shown. Quantification of ZMYM2 binding to GST-COMP-SUMO3 is shown on the Right as a percentage relative to the input lane and is the average of two experiments. Error bars represent SE. (C) GST pulldown analysis of the interaction of the purified recombinant GST-fusion proteins, COMP, COMP-SUMO3 (multi-SUMO), and SUMO3×4 (poly-SUMO) is presented in the form of a linear fusion protein with the indicated proteins from transiently transfected HEK293T cell lysates. (Top) Schematic illustration of the constructs used for the analysis. (Bottom) Proteins binding to the indicated GST-fusion proteins are shown by immunoblot (IB) with antibodies to the specific tags (Fig. 2) (top panels) and the GST bait proteins are shown on a Ponceau-stained membrane (bottom panel). Quantification of binding of each protein to each of the GST-fusion proteins is shown on the Right as a percentage relative to binding to GST-SUMO3×4 (taken as 100%). Data are the average of three experiments, and error bars represent SD. (D) GST pulldown analysis of GST-COMP-SUMO3 or GST-SUMO3×4 to WT or mutant forms of ZMYM2. Total cell lysates of HEK293T cells overexpressing the ZMYM2 proteins were used for the PD. ZMYM2 was detected by IB using anti-myc antibody. The 5% input is shown. (E) GST pulldown analysis of GST-COMP-SUMO3 or GST-SUMO3×4 to immobilized GST-ZMYM2(1-200) by SPR. At t = 0, either COMP (gray sensorgram), COMP-SUMO3 (blue sensorgram), or SUMO3×4 (orange sensorgram) was injected. Dissociation rates ($k_d$) were $6.05 \pm 1.11 \times 10^{-4}$ s$^{-1}$ (measured between 200 and 300 s) for COMP-SUMO3 and $7.51 \pm 1.64 \times 10^{-2}$ s$^{-1}$ (measured between 62 and 100 s) for SUMO3×4. The $k_d$ values are mean $\pm$ SD from three separate experiments.
same genes were also up-regulated in a second cell line expressing ZMYM2(SIM1,2,3mut) (Fig. S4E). The likely explanation for the effect of SIM mutations is a loss of transcriptional repressive activity of ZMYM2. To confirm that ZMYM2 acted as a repressor of the same target genes, we depleted ZMYM2 and assessed their expression levels. In all cases, increased expression was observed upon ZMYM2 depletion, and comparatively little effect was seen on control genes (Fig. S5A and B). These results suggest that ZMYM2 acts to repress gene expression through its SIM motifs, presumably by interacting with multi-SUMOylated coregulators. If this is the case, reductions in SUMOylation should lead to derepression of the same target genes. We therefore depleted the SUMO E2 conjugating enzyme UBC9/UBE2I (Fig. S5C) and assessed the expression of the same set of genes. In all cases, increased expression was observed upon UBC9/UBE2I depletion (Fig. S5D), demonstrating the importance of SUMOylation in their mediating repression.

Together, these results therefore demonstrate the functional importance of the SIMs found in ZMYM2 and indicate that a multi-SUMO-binding mode is used in mediating its transcriptional regulatory activities.

**Multi-SUMO-Binding Activity Promotes ZMYM2 Recruitment to Chromatin.** Having demonstrated that the SIM motifs of ZMYM2 play an important role in its gene-regulatory activities, we next investigated the molecular basis to this phenomenon. First, we tested whether mutation of the SIMs in ZMYM2 affected its transcriptional repression activity. We used a reporter assay using a luciferase reporter, driven by a LexA-VP16 fusion protein, and asked whether a Gal4-ZMYM2 fusion protein could repress transcription in this context (Fig. 5A). Here, we fused full-length ZMYM2 to the Gal4 DNA binding. Gal4-ZMYM2 was a highly active transcriptional repressor (Fig. 5A). However, this transcriptional repressive activity was retained in the Gal4-ZMYM2(SIM1,2,3mut) protein, which lacked all three SIMs we have identified (Fig. 5A). Thus, the SIM motifs are not required for the intrinsic transcriptional repressive properties of ZMYM2.

A second possibility is that the SIMs could be required for recruitment of ZMYM2 to chromatin. To address this possibility, we used ChIP-seq to analyze the binding of exogenous ZMYM2 to chromatin in our U2OS cell lines containing inducible forms of FLAG-tagged WT ZMYM2 or ZMYM2 with the critical SIM for multi-SUMO binding (SIM2) mutated. We initially identified 4,361 binding regions in the ZMYM2(WT) cell line but only 3,168 binding regions in the ZMYM2(SIM2mut) cell line. Taking the union of ZMYM2 binding sites for the WT and SIM2mut proteins gave 6,693 distinct regions, of which 3,069 loci were identified where significantly stronger binding (more than four-fold greater tag density; Poisson P value < 1 × 10^{-4}) of WT ZMYM2 was observed. The remainder showed similar or lower levels of WT ZMYM2 compared with the ZMYM2(SIM2mut) mutant version (see examples in Fig. 5B). We validated three sites that showed stronger binding by the WT protein and five sites that showed equivalent levels of binding by ChIP-qPCR (Fig. 5C).

We identified 4,361 binding regions in the ZMYM2(WT) cell line but only 3,168 binding regions in the ZMYM2(SIM2mut) cell line. This suggests that the SIMs are required for recruitment of ZMYM2 to a subset of genomic loci. One prediction of our results is that the genomic loci that show differential requirement for the ZMYM2 SIMs for its recruitment to chromatin, should harbor either a multi-SUMOylated protein or a cluster of SUMOylated proteins. To address this issue, we first compared our ZMYM2

---

**Fig. 5.** ZMYM2 SIMs are required for its recruitment to chromatin. (A) Reporter gene assay with constant LexA-VP16 but increasing amounts of the indicated Gal4-ZMYM2 derivatives. A schematic of the reporter gene is shown above, and the expression of the different Gal4 fusion proteins is shown in the Western blot as an Inset. Lamin B is a loading control, and the vertical line indicates that an intervening irrelevant lane has been removed. Data are shown relative to the internal β-galactosidase control (normalized to Gal4 DBD alone, taken as 1) and are the mean of two independent experiments, each performed in triplicate. (B) University of California, Santa Cruz genome browser view of three example loci and the binding peaks of the indicated ZMYM2 derivatives. (C) ChIP analysis of ZMYM2(WT) or ZMYM2(SIM2mut) binding to the indicated loci in the respective stable U2OS cell lines. Data are presented relative to input for each locus (n = 3), and statistically significant differences are indicated (P value < 0.005). Control ChIPs with IgG are shown at each locus for each cell line.
ChIP-seq data with ChIP-seq data for SUMO3 from the same cell line. SUMO binding occurs around many sites where we observed binding of WT ZMYM2 (Fig. 6A). To establish whether this association is more relevant to ZMYM2 or is typical of other transcriptional regulators, we compared SUMO occupancy around the summits of the binding regions for the transcriptional activator FOXO3 and found that SUMO3 is substantially more enriched around ZMYM2 binding summits (Fig. 6B). Given this association between ZMYM2 and SUMO, we next examined whether SUMO occupancy was higher around sites where ZMYM2 binding was affected by mutation of its SIMs as might be expected if SUMO is involved in ZMYM2 recruitment. ZMYM2-binding regions were partitioned into deciles, depending on the enrichment of ChIP-seq signal for WT versus SIM mutant form of ZMYM2. Importantly, the top decile, where binding of WT ZMYM2 was favored, showed the largest SUMO tag density, and this enrichment for SUMO co-occupancy decreased in the bottom decile, where the dependency on SIMs for ZMYM2 binding was lowest (Fig. 6C). Consistent with these results, stronger binding of SUMO was demonstrated at several of the sites that we identified as preferentially bound by WT ZMYM2 (e.g., TNXL4 and EMX2; Figs. 5C and 6D). Finally, we asked whether the coassociation with SUMO on chromatin is functionally linked to ZMYM2 recruitment. To do this, we took advantage of a ChIP-seq dataset for SUMO done under heat shock conditions (36). Although heat shock is generally thought to promote SUMOylation, decreases in SUMOylation occurred at many loci. Importantly, this occurred at all three loci where we saw differential binding between WT and SIM mutant forms of ZMYM2 and was accompanied by reduced binding of endogenous ZMYM2 following heat shock (e.g., EMX2; Fig. 6D). In contrast, heat shock did not affect ZMYM2 binding at loci such as PADI4 and ZC3H3 where basal levels of SUMO are low and there is little difference in binding of WT and SIM2 mutant forms of ZMYM2 (Fig. 6D).

In summary, these results are consistent with a model whereby the multiple SIM motifs in ZMYM2 are required for recruitment to a subset of chromatin regions in a SUMO-dependent manner (Fig. 6E).

**Discussion**

Protein SUMOylation has been implicated in numerous nuclear processes, and in particular chromatin regulation, transcriptional control, and DNA repair have emerged as major areas of influence (reviewed in ref. 4). Although it is now appreciated that a major role for SUMO is to provide a binding platform for other proteins, we still know relatively little about how specificity is generated or how these interactions are controlled. Recently, poly-SUMOylation has been shown to act as a binding platform...
for SUMO-targeted ubiquitin ligases (STUbLs), thereby coupling protein SUMOylation and ubiquitination (9, 22). We have extended these findings and demonstrate that STUbLs might also be targeted to substrates by interacting with alternative multi-SUMO interfaces, although poly-SUMO chain binding was clearly preferred. Nevertheless, these findings on STUbLs raised the possibility that multi-SUMO might act as a binding platform for other multi-SIM–containing ligases. Here, we identified dozens of potential multi-SUMO–binding proteins, validated our findings on several of these proteins, and, through detailed analysis of ZMYM2, identified a multi-SIM module and demonstrated its functional importance.

Both RNF4 and ZMYM2 can bind to multi-SUMO either as a multimerized SUMO array or as a poly-SUMO chain, and in keeping with this observation, each protein contains multiple SIMs, presumably so that each SIM can bind simultaneously to a different SUMO moiety (Fig. 1A). In principle, the binding modes for poly-SUMO chains and multi-SUMO platforms might be expected to be similar, but although poly-SUMO chains are likely to be uniform linear arrays, the structure of multi-SUMO platforms is dependent on the relative stereospecific presentation of the SUMO conjugates on the protein surface. Although ZMYM2 has the capacity to recognize both forms of multi-SUMO scaffold, RNF4 clearly has a preference for multi-SUMO in the context of poly-SUMO chains. Other proteins show intermediate preferences for the two forms of multimerized SUMO, presumably due to differences in how their SIMs are presented. Specific multi-SUMO platforms could be assembled either in cis on a single protein, or in trans on two or more proteins in a complex (Fig. 6E). Hints at both types of mechanisms are provided by the recent demonstration that SUMOylation of many different proteins involved in homologous recombination acts synergistically to promote efficient DNA repair (19). Conversely, juxtaposed SIMs can be presented in the form of a flexible chain as observed in RNF4 or be presented according to their locations on the surface of a folded protein, and such a scenario appears likely for ZMYM2. It is also possible that individual SIMs could be contributed by several different members of a protein complex. Thus, there are a large number of possible assembly mechanisms involving multiple SIMs and multi-SUMO–containing surfaces.

One emerging concept is that one or a subset of SIMs is “dominant” and defined by the large effects on binding seen upon their mutation (22). In the case of ZMYM2, one of the SIMs (SIM2) is “dominant,” whereas the roles of the other SIMs are only revealed through combinatorial mutation. It is not clear what causes this dominant effect, but a likely cause is that these SIMs might orchestrate the clustering of the other SIMs into a suitable conformation that enables multi- or poly-SUMO binding. In this context, it is notable that, unlike RNF4, ZMYM2 can bind equally well to a multi-SUMO platform and poly-SUMO chains, thereby further suggesting that the overall conformation of the SIM clusters and the SUMO binding platform is important for directing binding events.

It is also important to consider the functional significance of multi-SUMO binding and why this might be important. One attractive hypothesis is that a multi-SUMO–binding surface will only be created when several SUMOylated proteins aggregate together or high levels of SUMOylation are initiated on one or more proteins. This would then provide a molecular rheostat whereby recruitment of multi-SIM–containing partner proteins would only be triggered once high-level SUMOylation is achieved. Similar mechanisms involving a different posttranslational modification operate in the context of cell cycle control. Here, multisite phosphorylation, as exemplified by CDK-mediated Sic1 phosphorylation, triggers “switch-like” behavior through promoting SCF recruitment and Sic1 destruction (37).

Another interesting finding of our results was that, when tested for SUMO paralog binding, all of the multi-SUMO proteins that we identified showed a strong preference for SUMO2/3 binding over SUMO1 (Fig. 2). This might be expected given the fact that we used a GST-COMP-SUMO3 scaffold to isolate these proteins in the first place, but this does emphasize that SUMO1 and SUMO3 binding through SIMs is likely to be a highly selective process. Indeed, others have documented selective binding to SUMO3 over SUMO1 such as was observed for USP25 binding to SUMO-thioester charged UBC9/UBE2I (38). Furthermore, we also find that RNF4 preferentially binds multi-SUMO3 over multi-SUMO1. Conversely, several proteins have been shown to be able to bind both SUMO1 and SUMO3 (7), with acidic residues around SIMs apparently favoring SUMO1 binding. However, given that the SIMs in ZMYM2 resemble canonical SIMs with surrounding acidic residues and that ZMYM2 preferentially binds to SUMO3, this does not fit with this simplified view and suggests that there is much to learn about the specificity mechanisms used in SUMO paralog recognition. Moreover, such specificity mechanisms would open up additional regulatory possibilities. For example, it is possible that, in some cases, a mixture of SUMO1 and SUMO3 could constitute a multi-SUMO–binding surface, and this would result in the recruitment of proteins with a distinct mixture of SIMs recognizing these two different paralogs.

ZMYM2 has previously been identified in a proteomic screen for GST-SUMO3–binding proteins (39), and in a yeast two-hybrid screen using GAL-SUMO1 and GAL-SUMO3 as baits (7), whereas the roles of the other SIMs are

**Acknowledgments.** We thank Karen Palmer for excellent technical assistance; Catherine Millar, Alan Whitmarsh, Shen-Hsi Yang, and other members of our laboratory for comments on the manuscript and stimulating discussions; Richard Kammerer, Frances Fuller-Pace, Hongtao Yu, Ian Hickson, Pierre-Antoine Defossez, Egor Prokhortchouk, Richard Anderson, and Ingrid Grummt for reagents; Martin Love for help with the tissue culture; Stacey Warwood in the Core Biological Mass Spectrometry Facility; Peter March in
the Bioimaging Facility, Michael Smiga in the Genomics Technology Facility, and Ian Donaldson and Leo Zeef in the Bioinformatics Facility. This work was supported by the Welcome Trust and a Royal Society-Wolfson award (to A.D.S.) and the Wellcome Trust (098391/Z/12/Z) (to R.T.H.).