The Mechanistic Roles for SUMO in Chromatin and Condensates

Andres Reyna

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2025

Reading Committee:

Champak Chatterjee, Chair

Michael H. Gelb

Frantisek Turecek

Program Authorized to Offer Degree:

Chemistry

University of Washington

Abstract

The Mechanistic Roles for SUMO in Chromatin and Condensates

Andres Reyna

Chair of the Supervisory Committee:

Champak Chatterjee

Department of Chemistry

Histone post-translational modifications regulate DNA-templated processes in eukaryotes by shaping chromatin structure and regulating enzyme activity. The unstructured histone termini, or tails, extend outwards from the globular octameric histone core that tightly wraps ~147 bp of DNA to form the nucleosome core particle. The dynamic landscape of histone tail modifications is installed, maintained and remodeled by dedicated nuclear proteins that tune transcriptional outcomes in response to varying cellular needs. One dramatic histone modification that remains poorly studied in cells is conjugation of the histone lysine side-chain with the small ubiquitin like-modifier (SUMO) protein. Prior studies associated the presence of SUMO in chromatin with gene repression, but could not ascribe this effect to specific lysines sites in human histones. Semisynthetic access to site-specifically sumoylated histones, pioneered by our labs, has enabled biochemical investigations of the mechanistic roles for SUMO in chromatin. In this thesis, I examined the mechanistic roles for SUMO in three physiological contexts: its biochemical crosstalk with histone modifications in chromatin, its recruitment of repressive

machinery through SUMO interaction motifs (SIMs) in key protein components of gene repressive complexes, and its ability to regulate the biophysical properties of biomolecular condensates in living cells.

First, to address mechanisms of biochemical crosstalk in chromatin, I focused on the role of sumoylation at H4 Lys12, because it is a validated site of sumoylation in several cell lines and lies in the unstructured H4 tail. By generating site-specifically sumoylated H4K12 (H4K12su) using protein semisynthesis and incorporating it into histone octamers and reconstituted mononucleosomes, we showed that sumoylation inhibits Rad6-Bre1-mediated monoubiquitylation of H2BK120. In complementary cell-based assays, I demonstrated that the in vitro biochemical effect of H4K12su could be recapitulated using SUMO-H4 linear fusions that ensure uniform sumoylation on the intact H4 tail in living cells. By including both SUMO-H4 and a SUMO-H4(□1-11) construct that positions SUMO closer to Lys12, and hence closely mimics H4K12 sumoylation, I demonstrated that H4 sumoylation is inhibitory toward H2B K120 ubiquitylation in cells, a modification required for subsequent H3K4 methylation by the Set1/COMPASS family of H3K4 methyltransferases that are associated with transcription activation and elongation.

Having shown that H4K12su and its isosteres can repress H3K4 methylation by physically hindering Rad6-Bre1 binding to the nucleosome core particle and installing H2BK120ub, I turned to examining the role of H4K12su in recruiting gene repressive enzyme complexes to sumoylated nucleosomes. Toward this goal, I focused on a non-canonical SIM (ncSIM) within the Co-repressor of REST1 protein (CoREST1) and established its direct binding to SUMO3. Two-dimensional 15N-1H HSQC NMR chemical shift perturbations with short ncSIM peptides were employed to map their binding to the canonical SIM-binding groove in SUMO3. Reciprocal pulldowns with either purified SUMO3 or full-length CoREST1 as bait further confirmed their

direct binding, and alanine mutations of key residues in the ncSIM hydrophobic core were found to weaken SUMO3-CoREST binding. Altogether, my results defined the binding interface between the CoREST1 protein and SUMO3 and revealed how SUMO3-CoREST1 binding may position the gene repressive LSD1–CoREST–HDAC1 complex on nucleosomes and influence its demethylase and deacetylase activities on methylated and acetylated nucleosomes, respectively.

In the final part of my thesis, I broadened my studies of the SUMO-SIM interaction to investigate the fundamental biophysical properties of SUMO-dependent biomolecular condensates (BMCs), such as Promyelocytic Leukemia bodies. Toward this, I engineered a synthetic polyvalent platform consisting of poly(SUMO3) and poly(SIM) scaffolds and quantified how the SUMO3-SIM binding affinity changes the internal dynamics of BMCs and the activity of enzymes localized in BMCs. I discovered that the phase boundary (csat) for polyvalent constructs correlates well with monovalent SUMO-SIM affinities (Kd). However, condensate number and size do not correlate with Kd. Condensate stability to chaotoropes and internal mobility, as measured by Fluorescence Recovery After Photobleaching (FRAP), also showed excellent correlation with SUMO-SIM affinities. The strong correlation was also recapitulated in three different human cell lines and demonstrated our ability to generate BMCs with predictable properties in living cells. Finally, I demonstrated that the activity of the enzyme pyroglutamyl peptidase 1 that plays an important role in hormone regulation and the cellular stress response could be modulated by its recruitment to engineered BMCs of different mobilities.

Thus, my mechanistic studies with the small ubiquitin-like modifier protein have led me to identify negative crosstalk between H4K12 sumoylation and H2BK120 monoubiquitylation, to define the specific mode of interaction between SUMO and CoREST1, and to establish an exciting and novel modular framework for condensate formation in vitro and in living cells, which

directly links SUMO-SIM affinities to predictable condensate properties and tunable enzyme activities.

Acknowledgements

Graduate school has been a stretch of long days, late nights, failed experiments that taught me as much as the successful ones, and a lot of growing up in between. Some days felt like they would never end, counted out in spins in the centrifuge, FPLC runs, and half-scribbled ideas in my notebook, but the years still went by faster than I expected. I learned a lot of science along the way, but I also learned how to be more patient with myself, how to sit with uncertainty, and how to keep going when plans fell apart. Being at the end of all this makes it very clear that I did not get here by myself. This section is my attempt to thank the people whose love, guidance, and patience made it possible for me to keep going.

To my parents, Juan Reina and Ofelia Martinez Reyna, there are not enough words to explain what you have done for me. You gave me chances and supported me through experiences you never had yourselves, and you always worked to give the best opportunities to me and my brother. Thank you for every sacrifice that went into making something like this even feel possible. You left so much behind to build a life here and to give me the chance to aim for something bigger and follow my own dreams in a way you never had the chance to. You pushed me to aim higher and to take those chances, while still reminding me where I come from and who I am. When things were not going well in lab, you never started by asking about results. You asked if I was sleeping, if I was eating, if I was okay. Your pride in me never felt tied to a paper, an award, or a title, and that gave me room to try, to fail, to learn, and to keep trying. Whatever this degree ends up meaning on paper, to me it is a reflection of your work, your persistence, and your love. I hope I have made your effort and sacrifices worth it, and I hope this thesis feels like something we earned together.

To my brother, Brian Reyna, thank you for always being on my side. Thank you for paving the way through school and for giving me the support I needed to figure out my own path. Even when my explanations of what I was doing were vague or way too technical, you still listened, asked questions, and celebrated every step forward. Your check-ins and invitations showed up exactly when I needed them, especially on days when I felt stuck or worn down. You reminded me that there is more to life than whatever experiment I was worried about that week. I am proud to be your sibling, and I am really glad we get to share this moment.

I am also very grateful to my extended family for their steady support in the background. Even if the details of my research were unfamiliar, you always asked how I was doing, not just how my experiments were going. Thank you for reminding me that I am more than my work and that my worth does not come from data, deadlines, or degrees. The sense of stability, encouragement, and belonging you provided has been a big part of what kept me moving forward.

To my advisor, Professor Champak Chatterjee, thank you for giving me the chance to join your lab and for helping me find my way through this process. You pushed me to think carefully about the questions I ask and how I design experiments to answer them, and you held my work to a standard that made it better. Thank you for believing in me as a scientist, especially at times when I was not sure of myself. I am especially grateful that you let me chase projects I was genuinely interested in and trusted me with the freedom to try new things. You gave me the support and resources to explore directions that were unfamiliar and encouraged me not to be afraid of stepping outside my comfort zone. That mix of independence and backing has made me a stronger, more confident scientist and has pushed me to do things I never imagined myself doing when I first started graduate school.

Thank you to the members of the Chatterjee lab, past and present. It has been a privilege to work with you. You taught me new techniques, shared protocols, helped me make sense of confusing data, and sat with me through more troubleshooting sessions than I can count. You also made space for laughter, small celebrations, and everyday conversations that made long days in lab feel lighter. I will always remember the coffee runs that broke up the afternoon, the late dinners after a long experiment, and the times we stayed a little longer at the bench just to see one more result come in. From quick tips in passing to long discussions at the whiteboard, you helped turn half-formed ideas into real experiments and made the lab feel like a place I wanted to be, not just a place I had to be. The community you built is a big part of why I made it to the end of this.

I am also grateful to collaborators and colleagues in other groups. Thank you for sharing your expertise, your reagents, and your time. Working with you broadened my training and made my work stronger than it would have been on its own. I appreciate the meetings, the honest feedback, and the new ways of looking at problems that came out of those conversations.

To my friends, especially those back home who stayed close even from a distance, thank you for being such a steady part of my life through all of this. Thank you for the calls that went on longer than we planned, the movie nights we synced over a screen, the gaming sessions that gave me a break from thinking about lab, and the occasional visits that made everything feel normal again for a little while. You listened to me talk about experiments and let me vent, even when the details were so wrapped up in science that they probably sounded like nonsense from the outside. I appreciate how you still tried to follow along and understand what it meant to me. You stayed up with me while I worked up data or finished presentation slides, and you messaged me when I was staying late in lab so it did not feel like I was there by myself. Your

consistency and care, despite the distance and busy schedules, have meant a lot to me and were a big part of what kept me going.

Finally, to everyone who has helped in ways large and small, thank you. This thesis is not just a record of my work in the lab. It is also a record of the support, kindness, and encouragement I have been given over many years. I am very grateful to all of you who helped me reach this point.

TABLE OF CONTENTS

FIGURES AND TABLES

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 1

Introduction

1.1 Chromatin Structure and Regulation

In eukaryotic cells, genomic DNA is packaged in the nucleus as chromatin, a large nucleoprotein complex that serves to compact DNA while regulating access to nucleoproteins.[1,2] The nucleosome is the basic structural unit of chromatin and is composed of ~147 base pairs of DNA wrapped around an octameric complex of histone proteins containing two copies of H2A, H2B, H3, and H4.[3] While nucleosomes serve as a means of offering structural support to package DNA, they are also important regulators of genome access during transcription, replication, repair, and recombination.[2]

Chromatin can be broadly categorized into distinct structures that offer differing levels of accessibility to the underlying DNA. Euchromatin is a less compact structure of chromatin that is transcriptionally active as it allows transcription factors and RNA polymerase to engage DNA.[4] In contrast, heterochromatin is a compact form of chromatin that is generally transcriptionally repressed, and it is often enriched at repetitive elements and structural regions.[5] The two structural states are not static in nature, but, shift dynamically between states during development and in response to environmental signals as primary nucleosome composition is remodeled and chromatin-associated factors are recruited.[4,6]

There are many mechanisms in place that function together to coordinate chromatin accessibility. Histone post-translational modifications (PTMs) alter nucleosome-DNA interactions and recruit

effector proteins which act to shape chromatin accessibility and downstream enzyme activity.[7,8] The observation that combinations of PTMs correlate with distinct functional outcomes led to the "histone code" hypothesis regarding specific patterns of histone marks corresponding to specific DNA-templated processes.[9,10] A key example of this includes histone H3 trimethylation at Lys 9 (H3K9me3) that creates a high-affinity binding site for heterochromatin 1 (HP1) chromodomains, which leads to the repressed heterochromatin structure.[11,12]

Histone variants also add a second layer of complexity. Variants such as H2A.Z and H3.3 can replace canonical histones to further tune nucleosome stability and recognition by readers and remodelers, particularly around gene promoters and unique structural regions such as telomeres and centromeres.[13,14] ATP-dependent remodelers in the SWI-SNF, ISWI, CHD, and INO80 family of proteins can slide, evict, or restructure nucleosomes that work in combination with PTMs and transcription factors to further tune locus-specific accessibility.[15,16]

Taken together, the large number of histone variants and chromatin remodelers, histone modifications and modification-specific readers, writers and erasers make chromatin much more than a static scaffold to wrap genetic DNA and confer upon it the ability to regulate critical cellular processes by responding to cellular and environmental signals and stresses.[17-19]

1.2 Histone Post-Translational Modifications (PTMs)

Histone post-translational modifications can provide a direct way to regulate chromatin states. They can aid in relaxing or tightening DNA-histone contacts thereby altering nucleosome packaging and additionally create new recognition sites for factors that serve to assemble gene transcribing or repair machinery at specific sites.[8,20] These PTMs can act locally at a single nucleosome, as well as cumulatively across promoters, enhancers, and gene bodies. Due to their

association with specific chromatin states, they have also been referred to as histone *marks*. Histone PTM states offer some insight on how a single gene may have different activities across different cell types.[9,10] These PTMs range from smaller chemical modifications of amino acid side-chains such as acetylation, methylation, and phosphorylation to larger protein modifications such as ubiquitylation, sumoylation, and NEDDylation (**Figure 1**).[7,8,21,22] The placement and turnover of a mark may set how quickly a signal is written and how long it is maintained at a specific locus in chromatin.[4,7] These marks additionally influence DNA unwrapping at entry-exit sites and the residence times of chromatin remodelers and readers, which together regulate DNA template access at promoters and enhancers.[23,24]

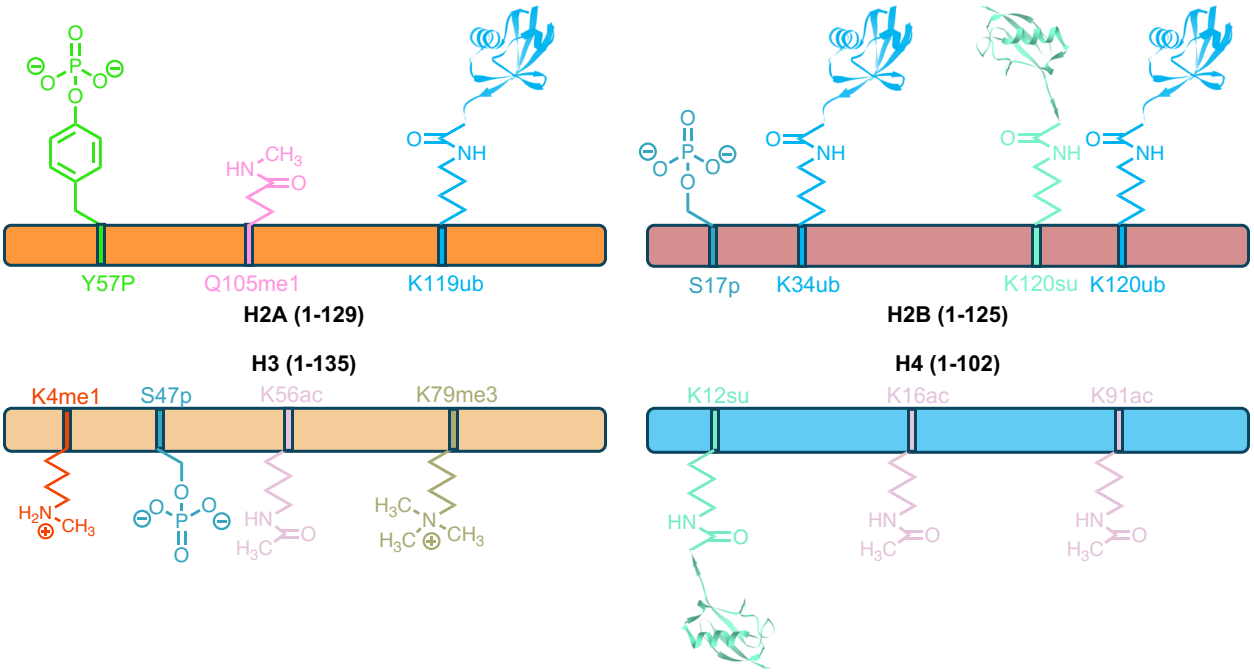


Figure 1. Histone Marks. Schematic representation of the diversity of histone marks that have been discovered on human histones. Chemical groups are indicated as ac= acetyl, m= methyl, p= phosphoryl, su= small ubiquitin-like modifier, ub= ubiquitin. The histones are colored as H2A (orange), H2B (red), H3 (brown) and H4 (blue).

Histone acetylation serves as an example of how a single modification can alter chromatin structure and accessibility. The addition of acetyl marks on lysine side-chains by histone acetyl

transferases (HATs) weakens histone tail-DNA and nucleosome core-DNA interactions, which typically coincides with an open euchromatin structure with greater transcriptional activity.[8,20] A single side-chain modification, such as acetylation at Lys 16 in histone H4 (H4K16ac), can disrupt internucleosomal packing to the extent that it strongly inhibits chromatin compaction.[25] At nucleosomal DNA entry-exit sites, acetylation of histone H3 at Lys 56 (H3K56ac) increases DNA unwrapping and makes nucleosomes easier to disassemble and reassemble during promoter activation and during S-phase chromatin disassembly. These examples illustrate how histone acetylation can dynamically modulate nucleosomes.[26,27] Dynamic cycles of acetylation and deacetylation modulate transcription-initiation without locking promoters into any single structural state.[28,29] Additionally, the same acetylation PTM at different residues in different histones can yield different biochemical outcomes because histone tail geometry and local DNA sequences dictate the degree of DNA unwrapping in nucleosomes.[30,31]

Histone methylation contributes site- and degree-specific information to chromatin that methyllysine reader domains can interpret for distinct biochemical outcomes. For example, trimethylation of histone H3 at Lys 4 (H3K4me3) is concentrated at transcription start sites, while monomethylation of histone H3 at Lys 4 (H3K4me1) is depleted at promoters but is significantly enriched at distal enhancer regions. Thus, the different methylation patterns distinguish genetic elements across a genome.[8,32] Repressive methylation marks at histone H3 at Lys 9 (H3K9me) and at Lys 27 (H3K27me) mark heterochromatin and Polycomb domains, while trimethylation of histone H3 at Lys 36 (H3K36me3) builds up across gene bodies as RNA polymerase II elongates. Additionally, in the budding yeast *Saccharomyces cerevisiae*, Set2-mediated H3K36 methylation recruits the Rpd3S deacetylase to remove acetylation within coding regions, which prevents cryptic transcription initiation. This couples elongation states to local hypoacetylation and stabilizing transcription units.[33-35] Many methyllysine readers containing chromodomains, Tudor domains, and plant homeodomain (PHD) fingers recognize these states and sort loci into stable

recruitment classes without large changes to nucleosome mechanics.[36,37] The specific degree of methylation can also set binding strength and residence time for methyllysine readers, so mono-, di-, and trimethylation states can also encode for the different stabilities and persistence of chromatin-associated complexes.[36]

Histone phosphorylation provides a direct link between intracellular signaling, chromatin function and genome maintenance. Phosphorylation of histone H3 at Ser 10 was observed during mitotic condensation as well as in rapid transcriptional responses, and serves as an example of the same mark participating in distinct functions depending on its genomic context.[38] At DNA double-stranded breaks, H2A.X is phosphorylated on Ser 139, which spreads over megabase-sized domains and serves to scaffold the recruitment of DNA damage response factors.[39,40] Because kinase activity is stimulus dependent, these phosphate modifications are appended and removed on short time scales and create temporary docking sites that organize cell cycle checkpoint and repair proteins.[41-43] For example, during mitosis, the appearance of histone phosphorylation supports chromosome condensation, while during transcriptional activation this mark can accompany promoter-associated acetylation. These provide examples of how context determines the particular downstream output.[43,44] Phosphatases remove the phosphoryl group once the checkpoint is satisfied, thereby releasing assembled factors and restoring the baseline state of chromatin.[45,46]

Histone monoubiquitylation, a well-studied non-degradative form of protein ubiquitylation, leads to both gene activation and repression depending on its precise context. Monoubiquitylation of H2B at Lys 120 (H2BK120ub) facilitates transcriptional elongation and promotes H3K4 and H3K79 methylation by the enzymes Set1/COMPASS and hDot1, respectively. This trans-histone biochemical pathway, or *crosstalk*, has been investigated and structural studies revealed docking of the COMPASS complex on H2BK120ub nucleosomes to stimulate H3K4 methylation.[21,47,48] In

contrast, monoubiquitylation of H2A at Lys 119 (H2AK119ub) by PRC1 is a component of Polycomb-mediated gene repression and heterochromatin formation.[49] Due to its large size relative to acetyl and methyl groups, ubiquitin can also alter access to the nucleosome surface and bias proximal histone mark writers or erasers, with the precise outcomes determined by the ubiquitylated residue and its genomic context.[48,50] Removal of the ubiquitin mark at H2B by deubiquitylating enzymes can limit H3K4 and K3K79 methylation and return elongating regions to a basal inactive state, whereas the maintenance of H2A ubiquitylation supports the compact chromatin structures associated with gene repression.[51,52]

The wide range of histone marks described above rarely act in isolation and often function in a combinatorial fashion. At promoters, H3K4me3 commonly appears along with promoter-proximal acetylation by the histone acetyltransferases CBP/p300 or Gcn5, whereas active enhancer regions typically contain H3K4me1 and gain H3K27ac during transcriptional activation. Poised enhancers marked by only H3K4me1 and lacking H3K27ac are often found within H3K27me3 fields at developmental loci.[32,53-55] Histone mark reader proteins have also evolved to integrate multiple cues toward distinct outputs. The NURF subunit BPTF combines the use of a PHD finger to recognize H3K4me3 along with an adjacent bromodomain to recognize acetylated H4, thereby increasing affinity when both marks are present, which links the recognition of multiple marks to remodeling promoter regions.[56-58] Acting together, marks set quantitative thresholds for factor binding and enzymatic activity across promoters, enhancers, and gene bodies.[7,58] This feature of marks is also seen with the small ubiquitin-like modifier (SUMO) protein that adds an additional protein interacting surface to chromatin and influences chromatin structure and compaction.[59,60]

1.3 Reversible Protein Sumoylation

Like acetylation, methylation, phosphorylation, and ubiquitylation, protein modification by the small ubiquitin-like protein (SUMO), also known as sumoylation, helps to tune nuclear processes. Sumoylation of lysines on target proteins reshapes their interaction surfaces, subnuclear localization, and activity in key biochemical processes such as transcription, DNA repair, chromatin organization, and the cellular stress response.[22,61] Sumoylation functions mainly as a signal that changes how factors assemble and modify chromatin.[62,63] In human cells, SUMO has multiple isoforms– SUMO1, SUMO2, SUMO3, and SUMO4. SUMO2 and SUMO3 share ~98% (precisely 97.83%) sequence identity and can readily form head-to-tail linked polymeric chains. SUMO1 only shares ~47% identity with SUMO2/3 and diverges from these proteins in that it serves to terminate polymeric SUMO2/3 chains. SUMO4, on the other hand, remains less well-characterized and is found to be expressed in only specific human cell types.[22,59]

Protein sumoylation follow a similar enzymatic mechanism as ubiquitylation. SUMO1/2/3 precursors that have additional residues after the C-terminal Gly-Gly sequence are first processed by SUMO-specific proteases to expose the C-terminal diglycine motif. The same proteases also serve to later remove SUMO to render sumoylation dynamic and reversible.[64,65] The heterodimeric E1 enzyme, SAE1-UBA2, then activates SUMO through C-terminal adenylation by ATP and subsequent formation of a thioester intermediate conjugated with UBA2 that is transferred to the only known SUMO E2 enzyme, Ubc9.[66,67] Ubc9 differs from ubiquitin E2 ligases because it recognizes a short linear consensus motif, ψKXE, where ψ is any hydrophobic residue in substrates and can catalyze ubiquitylation on its own whenever this sequence is present. Similar to ubiquitylation, SUMO E3 ligases also confer cellular location and target specificity to sumoylation. For example, the PIAS family of E3 ligases are frequently implicated in nuclear sumoylation, RanBP2/Nup358 undertakes sumoylation at the nuclear pore, and Pc2/CBX4 places SUMO on chromatin-associated targets.[64,68-71]

Acidic residues that flank the *ψKXE* motif, or phosphorylation of nearby serines, can enhance sumoylation and creates phosphorylation-dependent SUMO motifs that tie signaling pathways to site selectivity.[72,73] Recently, proteome-wide screens have catalogued thousands of SUMO acceptor lysine sites in proteins and shown that the SUMO isoforms and the nature of SUMO chains formed can vary under conditions of cellular stress, which is consistent with SUMO participating in numerous cellular processes outside the nucleus.[59]

Histone sumoylation was first suggested to play a role in gene repression and is most often associated with this function in the context of core or linker histones. Early studies revealed that histone sumoylation was often enriched at silent loci, and that mutating sumoylation sites de-repressed transcription. Additionally, these studies suggested that SUMO may recruit corepressors, such as histone deacetylases (HDACs) and heterochromatin associated proteins such as HP1.[74,75] Depending on factors such as the precise lysine residue that is modified, co-occurring modifications, and the recruitment of specific SUMO-binding proteins, SUMO can either stabilize a repressive complex, or it can modulate the transitions between gene activation-repression cycles.[59,63]

Many nuclear proteins also carry SUMO-interacting motifs (SIMs) which are short hydrophobic stretches with flanking acidic residues that bind to a conserved SIM-binding groove in SUMO1/2/3.[76-78] Polymeric SUMO2/3 chains create platforms for multi-SIM-containing proteins, such as the SUMO-targeted ubiquitin ligase (STUbL) protein RNF4 that couples sumoylated proteins in chromatin to protein ubiquitylation during the DNA damage response and in nuclear body turnover.[79-82]

Thus, sumoylation provides a platform that directs protein-protein interactions in and peripheral to chromatin. The precise outcomes of sumoylation depend upon the site of modification, the

SUMO isoform, and mono- or polysumoylation. Collectively, these factors also adjust which complexes occupy specific gene loci and how they modify the function and structure of chromatin.

1.4 The biochemical crosstalk of sumoylated histones

Sumoylation plays an important and diverse role in chromatin regulation on several levels, from directing specific histone modifications to the remodeling of protein-protein interaction networks. At the histone level, sumoylation is most frequently correlated with transcriptional repression by opposing acetylation, a mark typically associated with active transcription, or by restricting access of certain writer enzymes to install repressive modifications.[63,74,83] At the chromatin level, sumoylation often acts by creating, or strengthening, a binding site for SUMO-interacting motifs (SIMs) embedded within repressors, coregulators, and repair factors that reinforce local transcriptional repression.[59,61,76]

Many chromatin-associated protein complexes have components containing SIMs. One such complex is the gene repressive HDAC1-CoREST-LSD1 complex. In this complex, the scaffolding protein CoREST forms a ternary complex with the lysine specific demethylase 1 (LSD1) and class I histone deacetylase 1/2 (HDAC1/2).[84] CoREST also functions as a corepressor for REST at neuronal targets, with early biochemical and genetic work establishing CoREST as essential for the REST-dependent repression of neuronal genes in non-neuronal cells.[85] Biochemical and structural work has shown that CoREST stabilizes LSD1 and enables the efficient demethylation of H3K4me1/2 in the context of a nucleosome, while the ELM2-SANT domain of CoREST recruits HDAC1/2 to remove acetyl marks within the same region.[86-89] In additional kinetics studies, it was suggested that the LSD1-CoREST-HDAC1 complex adopted a flexible architecture and that the LSD1 and HDAC1 enzymatic activities are kinetically coupled when engaged with chromatin, which suggests coordinated deacetylation and demethylation activities during transcriptional repression.[76,84,86-88,90]

SUMO can also function as a specificity determinant for CoREST recruitment to chromatin. CoREST harbors a non-canonical SIM that allows it to bind SUMO2/3. This interaction enhances LSD1 activity on sumoylated and methylated promoters. By disrupting the non-canonical SIM in CoREST, the SUMO-dependent recruitment of CoREST-LSD1 may be disrupted which leads to the weakened repression of some target genes.[91] Consistent with this model, when chemically site-specific sumoylated histone H4 at Lys 12 (H4K12su) was incorporated in H3 K4 methylated nucleosomes, it stimulated intranucleosomal demethylation by the LSD1-CoREST sub-complex, suggesting that sumoylation in chromatin can directly enhance the repressive activity of LSD1-CoREST.[92]

Histone sumoylation can also modulate the installation of gene activating modifications such as histone acetylation. In studies with chemically modified nucleosomes, H4K12su was found to diminish the histone acetyltransferase (HAT) p300-mediated acetylation on the N-terminal H4 tail and to inhibit transcription from chromatin templates. This is consistent with a direct steric constraint on HAT access when SUMO is positioned at K12 in the H4 tail.[83] More broadly, proteomic and genomic experiments clearly indicate that histone sumoylation can not only recruit HDAC-containing complexes, but also serves to limit the installation of gene activating methyl and acetyl marks.[59,63,74]

Two of the three novel areas of research explored in this thesis are directly related to the role of SUMO in stimulating gene repression and preventing the installation of gene activating marks. First, because prior work in the Chatterjee lab had solely defined the CoREST-non-canonical SIM interactions using SIM-containing peptides derived from CoREST, I investigated the direct binding of the full-length SUMO3 and CoREST proteins via *in vitro* pulldowns. Thus, in Chapter 2, I establish that SIM-mediated recognition is sufficient to allow CoREST to bind SUMO3. This study

expands our previous NMR-based binding experiments with a minimal motif and confirms that the interaction between SUMO and CoREST is strongly dependent upon the non-canonical SIM in CoREST. Consistent with this, I further demonstrate that mutations of the SIM in full-length CoREST prevent its binding to SUMO3 in pulldown assays.

In Chapter 3, I explore the biochemical crosstalk between histone sumoylation and histone ubiquitylation in human cells. Ubiquitylation at K120 in H2B (H2BK120ub) is a well-studied mark associated with the transcription elongation-linked methylation of K4 in H3 and is generally associated with active transcription. Chapter 3 probes whether H4K12su and H2BK120ub co-occur on nucleosomes isolated from human chromatin. I demonstrate that a SUMO3-H4 fusion mimic of the gene repressive H4K12su occludes H2BK120ub from nucleosomes in cellular chromatin.

1.5 The role of SUMO in Biomolecular Condensates

While monovalent SUMO-SIM interactions serve to anchor gene repressive enzyme complexes to chromatin, polyvalent SUMO-SIM interactions can scale to mesoscale structures, called Biomolecular Condensates (BMCs). These membrane-less compartments enrich specific proteins and RNAs through weak, multivalent interactions between folded domains, between short linear motifs, and intrinsically disordered regions. BMCs provide cells a dynamic means to control local protein/RNA composition and reaction environments without their encapsulation in membranes (**Figure 2**).[93,94] Cellular conditions such as salt, macromolecular crowding, and PTMs can shift the network of intermolecular interactions necessary for BMC formation and potentially provide a means to govern condensate assembly and function in a stimuli responsive and reversible manner.[93,95]

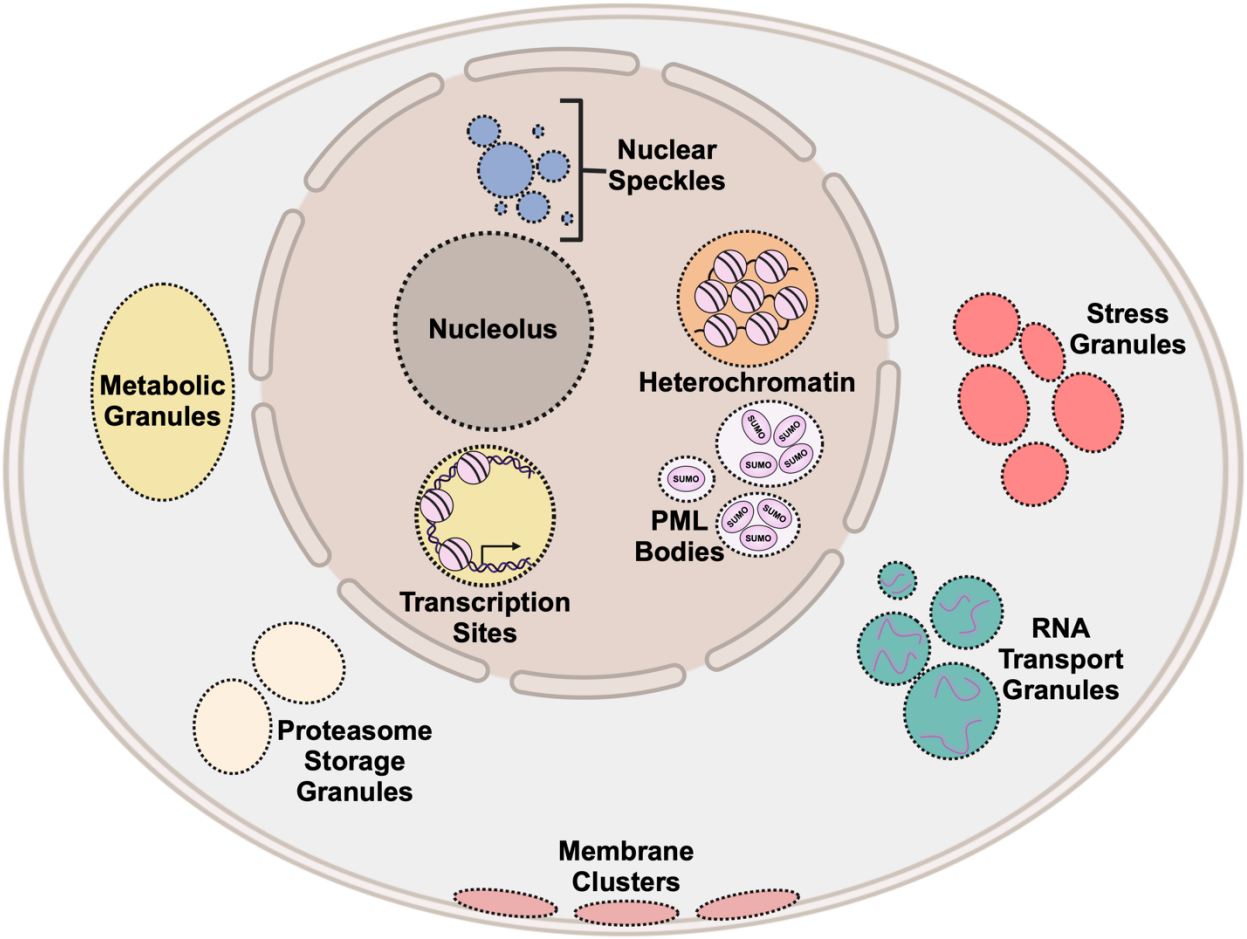


Figure 2. Schematic representation of numerous condensates in the nucleus, cytoplasm and membranes of eukaryotic cells. Some of the represented condensates exist only in specific cell types, such as RNA transport granules (teal) which are specific to neuronal cell types.

SUMO is the scaffolding protein in a special class of BMCs that act as a nuclear regulatory hub involved in tumor suppression, gene expression control, DNA repair, and antiviral defense. These Promyelocytic Leukemia (PML) nuclear bodies are punctate nuclear BMCs built around the principal component, the promyelocytic leukemia protein.[96] PML contains an N-terminal Ring-B Box-Coiled-Coil domain (RBCC) that oligomerizes, three principal SUMO acceptor lysines, and a C-terminal SIM. Sumoylation of PML, in combination with SIM-mediated oligomerization, is required for the formation of PML bodies, which leads to partner protein (client) recruitment in these organelles.[96,97] Under normal conditions, PML bodies are both stable and dynamic, with continuous client exchange. Under arsenic trioxide treatment, which inhibits desumoylating

enzymes, or oxidative and other stresses, PML oligomerization and sumoylation increases, leading to polySUMO chain accumulation, and RNF4 recruitment. RNF4 is a SUMO-targeted ubiquitin ligase that carries multiple SIMs to read the polySUMO signal as well as a RING domain that catalyzes the ubiquitylation of PML, or selected PML body residents, to promote their proteasome-directed turnover and reset PML body composition.[81,82,98,99] Isoform-level analyses show that the PML SIM is dispensable for basal sumoylation and RNF4 recruitment but is required for efficient ubiquitylation and proteasome targeting during arsenic trioxide-induced clearance. This ties SIM content to the efficiency of the stress-response mediated by PML bodies.[100]

Outside of PML bodies, signaling pathways can modulate where SUMO is installed or removed inside condensates. In human cells, activation of the cAMP-Epac1 pathway produces BMCs that enrich components necessary for SUMO conjugation and coincides with increased cellular sumoylation, showing that BMC formation also coincides with increased sumoylation at the cellular scale.[101] Under genotoxic stress, SUMO-rich repair foci assemble, RNF4 is recruited to polysumoylated clients to promote turnover, and SUMO-dependent exchange of chromatin proteins supports repairs.[79,102] Altogether, this illustrates a relationship in which sumoylation helps specify condensate composition, while the condensate environment influences conjugation and deconjugation.[103,104]

SUMO is also found in other nuclear condensates such as nuclear speckles. The sumoylation of spliceosomal factors enables efficient pre-RNA splicing and sumoylation-related enzymes colocalize within these condensates.[105-107] Cajal bodies are another BMC and sumoylation of the coilin protein and its SUMO-dependent interactions influence the number and size of Cajal bodies.[108-110] Thus, SUMO's role in the nucleus extends well beyond being a histone mark for gene repression to also serving as an organizer of nuclear composition and function.[104]

The role of SUMO in forming BMCs is explored in the final Chapter 4 of this thesis. By employing a well-defined polyvalent SUMO3-SIM scaffold to generate BMCs *in vitro* and in living cells, I sought to investigate the direct effect of SUMO-SIM affinity on fundamental properties such as size, stability and internal mobility in BMCs. Further, by recruiting enzymes bearing polyvalent SIMs to BMCs, I demonstrated that enzyme activity in condensates can be up- or downregulated by the mobility of the condensate. The outcomes presented in Chapter 4 serve to highlight the role of SUMO within the broader physical grammar that governs the functions of nuclear constituents.[93-95,101,103,104]

1.6 References

1 Kornberg, R. D. Chromatin structure: a repeating unit of histones and DNA. *Science* **184**, 868–871 (1974). https://doi.org/10.1126/science.184.4139.868

2 Felsenfeld, G. & Groudine, M. Controlling the double helix. *Nature* **421**, 448–453 (2003). https://doi.org/10.1038/nature01411

3 Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**, 251–260 (1997). https://doi.org/10.1038/38444

4 Zentner, G. E. & Henikoff, S. Regulation of nucleosome dynamics by histone modifications. *Nat Struct Mol Biol* **20**, 259–266 (2013). https://doi.org/10.1038/nsmb.2470

5 Grewal, S. I. & Elgin, S. C. Transcription and RNA interference in the formation of heterochromatin. *Nature* **447**, 399–406 (2007). https://doi.org/10.1038/nature05914

6 Bird, A. Perceptions of epigenetics. *Nature* **447**, 396–398 (2007). https://doi.org/10.1038/nature05913

7 Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693–705 (2007). https://doi.org/10.1016/j.cell.2007.02.005

8 Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res* **21**, 381–395 (2011). https://doi.org/10.1038/cr.2011.22

9 Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403**, 41–45 (2000). https://doi.org/10.1038/47412

10 Jenuwein, T. & Allis, C. D. Translating the histone code. *Science* **293**, 1074–1080 (2001). https://doi.org/10.1126/science.1063127

11 Bannister, A. J. *et al.* Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410**, 120–124 (2001). https://doi.org/10.1038/35065138

12 Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. & Jenuwein, T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410**, 116–120 (2001). https://doi.org/10.1038/35065132

13 Talbert, P. B. & Henikoff, S. Histone variants--ancient wrap artists of the epigenome. *Nat Rev Mol Cell Biol* **11**, 264–275 (2010). https://doi.org/10.1038/nrm2861

14 Venkatesh, S. & Workman, J. L. Histone exchange, chromatin structure and the regulation of transcription. *Nat Rev Mol Cell Biol* **16**, 178–189 (2015). https://doi.org/10.1038/nrm3941
15 Clapier, C. R. & Cairns, B. R. The biology of chromatin remodeling complexes. *Annu Rev Biochem* **78**, 273–304 (2009). https://doi.org/10.1146/annurev.biochem.77.062706.153223
16 Clapier, C. R., Iwasa, J., Cairns, B. R. & Peterson, C. L. Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nat Rev Mol Cell Biol* **18**, 407–422 (2017). https://doi.org/10.1038/nrm.2017.26
17 Dekker, J. & Mirny, L. The 3D Genome as Moderator of Chromosomal Communication. *Cell* **164**, 1110–1121 (2016). https://doi.org/10.1016/j.cell.2016.02.007
18 Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012). https://doi.org/10.1038/nature11082
19 Rao, S. S. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014). https://doi.org/10.1016/j.cell.2014.11.021
20 Roth, S. Y., Denu, J. M. & Allis, C. D. Histone acetyltransferases. *Annu Rev Biochem* **70**, 81–120 (2001). https://doi.org/10.1146/annurev.biochem.70.1.81
21 Weake, V. M. & Workman, J. L. Histone ubiquitination: triggering gene activity. *Mol Cell* **29**, 653–663 (2008). https://doi.org/10.1016/j.molcel.2008.02.014
22 Flotho, A. & Melchior, F. Sumoylation: a regulatory protein modification in health and disease. *Annu Rev Biochem* **82**, 357–385 (2013). https://doi.org/10.1146/annurev-biochem-061909-093311
23 Polach, K. J. & Widom, J. Mechanism of protein access to specific DNA sequences in chromatin: a dynamic equilibrium model for gene regulation. *J Mol Biol* **254**, 130–149 (1995). https://doi.org/10.1006/jmbi.1995.0606
24 Narlikar, G. J., Sundaramoorthy, R. & Owen-Hughes, T. Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. *Cell* **154**, 490–503 (2013). https://doi.org/10.1016/j.cell.2013.07.011
25 Shogren-Knaak, M. *et al.* Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* **311**, 844–847 (2006). https://doi.org/10.1126/science.1124000
26 Williams, S. K., Truong, D. & Tyler, J. K. Acetylation in the globular core of histone H3 on lysine-56 promotes chromatin disassembly during transcriptional activation. *Proc Natl Acad Sci U S A* **105**, 9000–9005 (2008). https://doi.org/10.1073/pnas.0800057105
27 Neumann, H. *et al.* A method for genetically installing site-specific acetylation in recombinant histones defines the effects of H3 K56 acetylation. *Mol Cell* **36**, 153–163 (2009). https://doi.org/10.1016/j.molcel.2009.07.027
28 Nicolas, D., Zoller, B., Suter, D. M. & Naef, F. Modulation of transcriptional burst frequency by histone acetylation. *Proc Natl Acad Sci U S A* **115**, 7153–7158 (2018). https://doi.org/10.1073/pnas.1722330115
29 Chen, L. F. *et al.* Enhancer Histone Acetylation Modulates Transcriptional Bursting Dynamics of Neuronal Activity-Inducible Genes. *Cell Rep* **26**, 1174–1188 e1175 (2019). https://doi.org/10.1016/j.celrep.2019.01.032
30 Kono, H. & Ishida, H. Nucleosome unwrapping and unstacking. *Curr Opin Struct Biol* **64**, 119–125 (2020). https://doi.org/10.1016/j.sbi.2020.06.020
31 Ngo, T. T., Zhang, Q., Zhou, R., Yodh, J. G. & Ha, T. Asymmetric unwrapping of nucleosomes under tension directed by DNA local flexibility. *Cell* **160**, 1135–1144 (2015). https://doi.org/10.1016/j.cell.2015.02.001

32 Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39**, 311–318 (2007). https://doi.org/10.1038/ng1966

33 Carrozza, M. J. *et al.* Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* **123**, 581–592 (2005). https://doi.org/10.1016/j.cell.2005.10.023

34 Keogh, M. C. *et al.* Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. *Cell* **123**, 593–605 (2005). https://doi.org/10.1016/j.cell.2005.10.025

35 Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823–837 (2007). https://doi.org/10.1016/j.cell.2007.05.009

36 Musselman, C. A., Lalonde, M. E., Cote, J. & Kutateladze, T. G. Perceiving the epigenetic landscape through histone readers. *Nat Struct Mol Biol* **19**, 1218–1227 (2012). https://doi.org/10.1038/nsmb.2436

37 Lu, R. & Wang, G. G. Tudor: a versatile family of histone methylation 'readers'. *Trends Biochem Sci* **38**, 546–555 (2013). https://doi.org/10.1016/j.tibs.2013.08.002

38 Nowak, S. J. & Corces, V. G. Phosphorylation of histone H3: a balancing act between chromosome condensation and transcriptional activation. *Trends Genet* **20**, 214–220 (2004). https://doi.org/10.1016/j.tig.2004.02.007

39 Rogakou, E. P., Pilch, D. R., Orr, A. H., Ivanova, V. S. & Bonner, W. M. DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J Biol Chem* **273**, 5858–5868 (1998). https://doi.org/10.1074/jbc.273.10.5858

40 Rogakou, E. P., Boon, C., Redon, C. & Bonner, W. M. Megabase chromatin domains involved in DNA double-strand breaks in vivo. *J Cell Biol* **146**, 905–916 (1999). https://doi.org/10.1083/jcb.146.5.905

41 Stucki, M. *et al.* MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks. *Cell* **123**, 1213–1226 (2005). https://doi.org/10.1016/j.cell.2005.09.038

42 Macdonald, N. *et al.* Molecular basis for the recognition of phosphorylated and phosphoacetylated histone h3 by 14-3-3. *Mol Cell* **20**, 199–211 (2005). https://doi.org/10.1016/j.molcel.2005.08.032

43 Drobic, B., Perez-Cadahia, B., Yu, J., Kung, S. K. & Davie, J. R. Promoter chromatin remodeling of immediate-early genes is mediated through H3 phosphorylation at either serine 28 or 10 by the MSK1 multi-protein complex. *Nucleic Acids Res* **38**, 3196–3208 (2010). https://doi.org/10.1093/nar/gkq030

44 Hendzel, M. J. *et al.* Mitosis-specific phosphorylation of histone H3 initiates primarily within pericentromeric heterochromatin during G2 and spreads in an ordered fashion coincident with mitotic chromosome condensation. *Chromosoma* **106**, 348–360 (1997). https://doi.org/10.1007/s004120050256

45 Chowdhury, D. *et al.* gamma-H2AX dephosphorylation by protein phosphatase 2A facilitates DNA double-strand break repair. *Mol Cell* **20**, 801–809 (2005). https://doi.org/10.1016/j.molcel.2005.10.003

46 Cha, H. *et al.* Wip1 directly dephosphorylates gamma-H2AX and attenuates the DNA damage response. *Cancer Res* **70**, 4112–4122 (2010). https://doi.org/10.1158/0008-5472.CAN-09-4244

47 Pavri, R. *et al.* Histone H2B monoubiquitination functions cooperatively with FACT to regulate elongation by RNA polymerase II. *Cell* **125**, 703–717 (2006). https://doi.org/10.1016/j.cell.2006.04.029

48 Worden, E. J., Zhang, X. & Wolberger, C. Structural basis for COMPASS recognition of an H2B-ubiquitinated nucleosome. *Elife* **9** (2020). https://doi.org/10.7554/eLife.53199

49 Wang, H. *et al.* Role of histone H2A ubiquitination in Polycomb silencing. *Nature* **431**, 873–878 (2004). https://doi.org/10.1038/nature02985

50 Hsu, P. L. *et al.* Structural Basis of H2B Ubiquitination-Dependent H3K4 Methylation by COMPASS. *Mol Cell* **76**, 712–723 e714 (2019). https://doi.org/10.1016/j.molcel.2019.10.013

51 Henry, K. W. *et al.* Transcriptional activation via sequential histone H2B ubiquitylation and deubiquitylation, mediated by SAGA-associated Ubp8. *Genes Dev* **17**, 2648–2663 (2003). https://doi.org/10.1101/gad.1144003

52 Blackledge, N. P. *et al.* Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. *Cell* **157**, 1445–1459 (2014). https://doi.org/10.1016/j.cell.2014.05.004

53 Heintzman, N. D. *et al.* Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108–112 (2009). https://doi.org/10.1038/nature07829

54 Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* **107**, 21931–21936 (2010). https://doi.org/10.1073/pnas.1016071107

55 Rada-Iglesias, A. *et al.* A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279–283 (2011). https://doi.org/10.1038/nature09692

56 Wysocka, J. *et al.* A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* **442**, 86–90 (2006). https://doi.org/10.1038/nature04815

57 Li, H. *et al.* Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature* **442**, 91–95 (2006). https://doi.org/10.1038/nature04802

58 Ruthenburg, A. J. *et al.* Recognition of a mononucleosomal histone modification pattern by BPTF via multivalent interactions. *Cell* **145**, 692–706 (2011). https://doi.org/10.1016/j.cell.2011.03.053

59 Hendriks, I. A. & Vertegaal, A. C. A comprehensive compilation of SUMO proteomics. *Nat Rev Mol Cell Biol* **17**, 581–595 (2016). https://doi.org/10.1038/nrm.2016.81

60 Cubenas-Potts, C. & Matunis, M. J. SUMO: a multifaceted modifier of chromatin structure and function. *Dev Cell* **24**, 1–12 (2013). https://doi.org/10.1016/j.devcel.2012.11.020

61 Geiss-Friedlander, R. & Melchior, F. Concepts in sumoylation: a decade on. *Nat Rev Mol Cell Biol* **8**, 947–956 (2007). https://doi.org/10.1038/nrm2293

62 Johnson, E. S. Protein modification by SUMO. *Annu Rev Biochem* **73**, 355–382 (2004). https://doi.org/10.1146/annurev.biochem.73.011303.074118

63 Ryu, H. Y. & Hochstrasser, M. Histone sumoylation and chromatin dynamics. *Nucleic Acids Res* **49**, 6043–6052 (2021). https://doi.org/10.1093/nar/gkab280

64 Gareau, J. R. & Lima, C. D. The SUMO pathway: emerging mechanisms that shape specificity, conjugation and recognition. *Nat Rev Mol Cell Biol* **11**, 861–871 (2010). https://doi.org/10.1038/nrm3011

65 Wilkinson, K. A. & Henley, J. M. Mechanisms, regulation and consequences of protein SUMOylation. *Biochem J* **428**, 133–145 (2010). https://doi.org/10.1042/BJ20100158

66 Lois, L. M. & Lima, C. D. Structures of the SUMO E1 provide mechanistic insights into SUMO activation and E2 recruitment to E1. *EMBO J* **24**, 439–451 (2005). https://doi.org/10.1038/sj.emboj.7600552

67 Olsen, S. K., Capili, A. D., Lu, X., Tan, D. S. & Lima, C. D. Active site remodelling accompanies thioester bond formation in the SUMO E1. *Nature* **463**, 906–912 (2010). https://doi.org/10.1038/nature08765

68 Sampson, D. A., Wang, M. & Matunis, M. J. The small ubiquitin-like modifier-1 (SUMO-1) consensus sequence mediates Ubc9 binding and is essential for SUMO-1

modification. *J Biol Chem* **276**, 21664–21669 (2001). https://doi.org/10.1074/jbc.M100006200
69 Sachdev, S. *et al.* PIASy, a nuclear matrix-associated SUMO E3 ligase, represses LEF1 activity by sequestration into nuclear bodies. *Genes Dev* **15**, 3088–3103 (2001). https://doi.org/10.1101/gad.944801
70 Kagey, M. H., Melhuish, T. A. & Wotton, D. The polycomb protein Pc2 is a SUMO E3. *Cell* **113**, 127–137 (2003). https://doi.org/10.1016/s0092-8674(03)00159-4
71 Kirsh, O. *et al.* The SUMO E3 ligase RanBP2 promotes modification of the HDAC4 deacetylase. *EMBO J* **21**, 2682–2691 (2002). https://doi.org/10.1093/emboj/21.11.2682
72 Yang, S. H., Galanis, A., Witty, J. & Sharrocks, A. D. An extended consensus motif enhances the specificity of substrate modification by SUMO. *EMBO J* **25**, 5083–5093 (2006). https://doi.org/10.1038/sj.emboj.7601383
73 Hietakangas, V. *et al.* PDSM, a motif for phosphorylation-dependent SUMO modification. *Proc Natl Acad Sci U S A* **103**, 45–50 (2006). https://doi.org/10.1073/pnas.0503698102
74 Shiio, Y. & Eisenman, R. N. Histone sumoylation is associated with transcriptional repression. *Proc Natl Acad Sci U S A* **100**, 13225–13230 (2003). https://doi.org/10.1073/pnas.1735528100
75 Nathan, D. *et al.* Histone sumoylation is a negative regulator in Saccharomyces cerevisiae and shows dynamic interplay with positive-acting histone modifications. *Genes Dev* **20**, 966–976 (2006). https://doi.org/10.1101/gad.1404206
76 Song, J., Durrin, L. K., Wilkinson, T. A., Krontiris, T. G. & Chen, Y. Identification of a SUMO-binding motif that recognizes SUMO-modified proteins. *Proc Natl Acad Sci U S A* **101**, 14373–14378 (2004). https://doi.org/10.1073/pnas.0403498101
77 Hecker, C. M., Rabiller, M., Haglund, K., Bayer, P. & Dikic, I. Specification of SUMO1- and SUMO2-interacting motifs. *J Biol Chem* **281**, 16117–16127 (2006). https://doi.org/10.1074/jbc.M512757200
78 Stehmeier, P. & Muller, S. Phospho-regulated SUMO interaction modules connect the SUMO system to CK2 signaling. *Mol Cell* **33**, 400–409 (2009). https://doi.org/10.1016/j.molcel.2009.01.013
79 Galanty, Y., Belotserkovskaya, R., Coates, J. & Jackson, S. P. RNF4, a SUMO-targeted ubiquitin E3 ligase, promotes DNA double-strand break repair. *Genes Dev* **26**, 1179–1195 (2012). https://doi.org/10.1101/gad.188284.112
80 Rojas-Fernandez, A. *et al.* SUMO chain-induced dimerization activates RNF4. *Mol Cell* **53**, 880–892 (2014). https://doi.org/10.1016/j.molcel.2014.02.031
81 Xu, Y. *et al.* Structural insight into SUMO chain recognition and manipulation by the ubiquitin ligase RNF4. *Nat Commun* **5**, 4217 (2014). https://doi.org/10.1038/ncomms5217
82 Tatham, M. H. *et al.* RNF4 is a poly-SUMO-specific E3 ubiquitin ligase required for arsenic-induced PML degradation. *Nat Cell Biol* **10**, 538–546 (2008). https://doi.org/10.1038/ncb1716
83 Leonen, C. J. A. *et al.* Sumoylation of the human histone H4 tail inhibits p300-mediated transcription by RNA polymerase II in cellular extracts. *Elife* **10** (2021). https://doi.org/10.7554/eLife.67952
84 You, A., Tong, J. K., Grozinger, C. M. & Schreiber, S. L. CoREST is an integral component of the CoREST- human histone deacetylase complex. *Proc Natl Acad Sci U S A* **98**, 1454–1458 (2001). https://doi.org/10.1073/pnas.98.4.1454
85 Andres, M. E. *et al.* CoREST: a functional corepressor required for regulation of neural-specific gene expression. *Proc Natl Acad Sci U S A* **96**, 9873–9878 (1999). https://doi.org/10.1073/pnas.96.17.9873

86 Yang, M. *et al.* Structural basis for CoREST-dependent demethylation of nucleosomes by the human LSD1 histone demethylase. *Mol Cell* **23**, 377–387 (2006). https://doi.org/10.1016/j.molcel.2006.07.012
87 Millard, C. J. *et al.* Class I HDACs share a common mechanism of regulation by inositol phosphates. *Mol Cell* **51**, 57–67 (2013). https://doi.org/10.1016/j.molcel.2013.05.020
88 Rivera, C. *et al.* Unveiling RCOR1 as a rheostat at transcriptionally permissive chromatin. *Nat Commun* **13**, 1550 (2022). https://doi.org/10.1038/s41467-022-29261-0
89 Lee, M. G., Wynder, C., Cooch, N. & Shiekhattar, R. An essential role for CoREST in nucleosomal histone 3 lysine 4 demethylation. *Nature* **437**, 432–435 (2005). https://doi.org/10.1038/nature04021
90 Kim, S. A., Chatterjee, N., Jennings, M. J., Bartholomew, B. & Tan, S. Extranucleosomal DNA enhances the activity of the LSD1/CoREST histone demethylase complex. *Nucleic Acids Res* **43**, 4868–4880 (2015). https://doi.org/10.1093/nar/gkv388
91 Ouyang, J., Shi, Y., Valin, A., Xuan, Y. & Gill, G. Direct binding of CoREST1 to SUMO-2/3 contributes to gene-specific repression by the LSD1/CoREST1/HDAC complex. *Mol Cell* **34**, 145–154 (2009). https://doi.org/10.1016/j.molcel.2009.03.013
92 Dhall, A., Weller, C. E., Chu, A., Shelton, P. M. M. & Chatterjee, C. Chemically Sumoylated Histone H4 Stimulates Intranucleosomal Demethylation by the LSD1-CoREST Complex. *ACS Chem Biol* **12**, 2275–2280 (2017). https://doi.org/10.1021/acschembio.7b00716
93 Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Biomolecular condensates: organizers of cellular biochemistry. *Nat Rev Mol Cell Biol* **18**, 285–298 (2017). https://doi.org/10.1038/nrm.2017.7
94 Ditlev, J. A., Case, L. B. & Rosen, M. K. Who's In and Who's Out-Compositional Control of Biomolecular Condensates. *J Mol Biol* **430**, 4666–4684 (2018). https://doi.org/10.1016/j.jmb.2018.08.003
95 Villegas, J. A., Heidenreich, M. & Levy, E. D. Molecular and environmental determinants of biomolecular condensate formation. *Nat Chem Biol* **18**, 1319–1329 (2022). https://doi.org/10.1038/s41589-022-01175-4
96 Shen, T. H., Lin, H. K., Scaglioni, P. P., Yung, T. M. & Pandolfi, P. P. The mechanisms of PML-nuclear body formation. *Mol Cell* **24**, 331–339 (2006). https://doi.org/10.1016/j.molcel.2006.09.013
97 Zhong, S. *et al.* Role of SUMO-1-modified PML in nuclear body formation. *Blood* **95**, 2748–2752 (2000).
98 Lallemand-Breitenbach, V. *et al.* Arsenic degrades PML or PML-RARalpha through a SUMO-triggered RNF4/ubiquitin-mediated pathway. *Nat Cell Biol* **10**, 547–555 (2008). https://doi.org/10.1038/ncb1717
99 Geoffroy, M. C., Jaffray, E. G., Walker, K. J. & Hay, R. T. Arsenic-induced SUMO-dependent recruitment of RNF4 into PML nuclear bodies. *Mol Biol Cell* **21**, 4227–4239 (2010). https://doi.org/10.1091/mbc.E10-05-0449
100 Maroui, M. A. *et al.* Requirement of PML SUMO interacting motif for RNF4- or arsenic trioxide-induced degradation of nuclear PML isoforms. *PLoS One* **7**, e44949 (2012). https://doi.org/10.1371/journal.pone.0044949
101 Yang, W. *et al.* Epac1 activation by cAMP regulates cellular SUMOylation and promotes the formation of biomolecular condensates. *Sci Adv* **8**, eabm2960 (2022). https://doi.org/10.1126/sciadv.abm2960
102 Fukuto, A. *et al.* SUMO modification system facilitates the exchange of histone variant H2A.Z-2 at DNA damage sites. *Nucleus* **9**, 87–94 (2018). https://doi.org/10.1080/19491034.2017.1395543
103 Cheng, X. Protein SUMOylation and phase separation: partners in stress? *Trends Biochem Sci* **48**, 417–419 (2023). https://doi.org/10.1016/j.tibs.2022.12.003

104 Keiten-Schmitz, J., Roder, L., Hornstein, E., Muller-McNicoll, M. & Muller, S. SUMO: Glue or Solvent for Phase-Separated Ribonucleoprotein Complexes and Molecular Condensates? *Front Mol Biosci* **8**, 673038 (2021). https://doi.org/10.3389/fmolb.2021.673038
105 Pelisch, F. *et al.* The serine/arginine-rich protein SF2/ASF regulates protein sumoylation. *Proc Natl Acad Sci U S A* **107**, 16119–16124 (2010). https://doi.org/10.1073/pnas.1004653107
106 Pozzi, B. *et al.* SUMO conjugation to spliceosomal proteins is required for efficient pre-mRNA splicing. *Nucleic Acids Res* **45**, 6729–6745 (2017). https://doi.org/10.1093/nar/gkx213
107 Galganski, L., Urbanek, M. O. & Krzyzosiak, W. J. Nuclear speckles: molecular organization, biological function and role in disease. *Nucleic Acids Res* **45**, 10350–10368 (2017). https://doi.org/10.1093/nar/gkx759
108 Navascues, J. *et al.* SUMO-1 transiently localizes to Cajal bodies in mammalian neurons. *J Struct Biol* **163**, 137–146 (2008). https://doi.org/10.1016/j.jsb.2008.04.013
109 Tucker, S. K., McLaurin, D. M. & Hebert, M. D. Cajal body formation is regulated by coilin SUMOylation. *J Cell Sci* **137** (2024). https://doi.org/10.1242/jcs.263447
110 Lett, K. E., McLaurin, D. M., Tucker, S. K. & Hebert, M. D. The Cajal body marker protein coilin is SUMOylated and possesses SUMO E3 ligase-like activity. *Front RNA Res* **1** (2023). https://doi.org/10.3389/frnar.2023.1197990

Chapter 2

CoREST1 reads the SUMO3 code for gene repression

2.1 Introduction

The lysine-specific demethylase 1 (LSD1)–Co-repressor of REST1 (RCOR1 or CoREST1)–histone deacetylase 1 (HDAC1) complex (LCH) is the major subunit of the CoREST1 complex that silences neuronal genes in non-neuronal cells. This complex removes gene-activating acetyl and methyl marks from histones and reinforces a silenced chromatin state. Pioneering biochemical and proteomics approaches established CoREST1 as the corepressor of the REST1 (neuron-restrictive silencer factor, or repressor element 1 silencing transcription factor) protein that scaffolds HDAC1/2 and LSD1 that act on histone substrates.[1,2] X-ray structural studies revealed that CoREST1 binds LSD1 through an elongated helical bundle and the LSD1 Tower

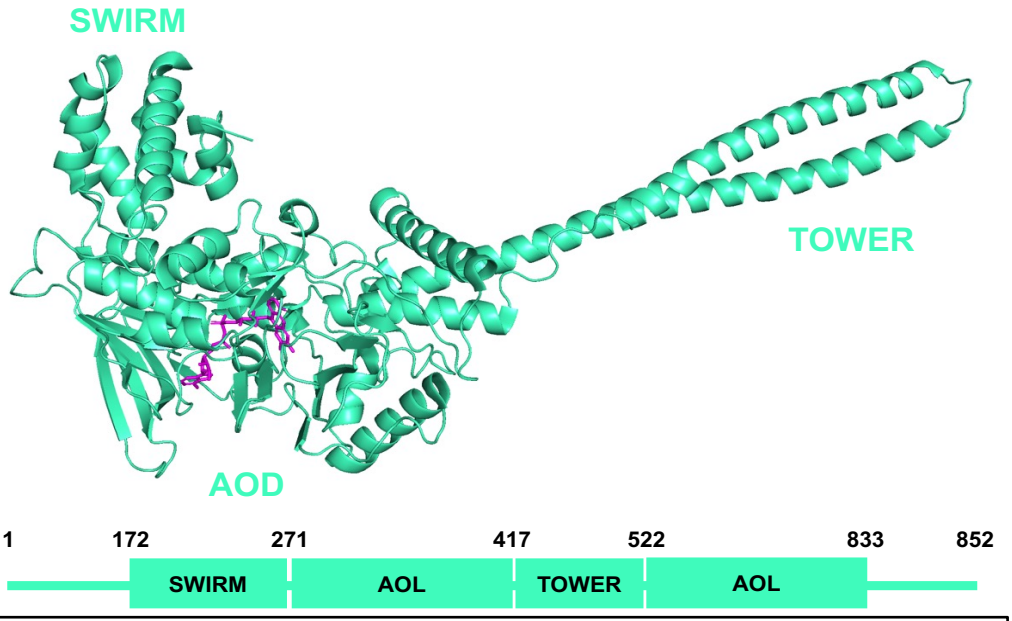


Figure 1. Structure of LSD1. The crystal structure of LSD1(171-836) (PDB: 2V1D), with FAD (magenta) bound in its AOL domain active site. A full-length sequence map showing the relative positions of each of the domains is shown below the structure.

domain helices (**Figure 1**) dock onto the SANT2 module in CoREST1 through a long connecting "stalk," positioning the catalytic amine-oxidase-like (AOL) domain of LSD1 for productive engagement with the H3 tail in nucleosomes (**Figure 2**).[3] LSD1 removes H3K4me1/2 by a flavin adenine dinucleotide (FAD)-dependent oxidative mechanism that generates an imine intermediate that is hydrolyzed to release formaldehyde and the free lysine ε-amine. The chemical mechanism of LSD1 enforces its specificity for mono- and dimethyllysine and the inability to demethylate trimethyllysine.[4] Recent crystallographic studies of LSD1–CoREST1 bound to a mononucleosome revealed how DNA and histone contacts align the LSD1 active site in the amine oxidase domain with the H3 tail in a 1:1 nucleosome complex, and emphasized the important role for CoREST1 in efficient nucleosome demethylation (**Figure 2**).[5,6]

HDAC1 catalyzes Zn^{2+}-dependent hydrolysis of acetyllysine to release acetate and the free lysine ε-amine. Within the LCH sub-complex, HDAC1 activity is linked to CoREST1 through the ELM2–SANT1 cassette (**Figure 3**). Class I HDAC inhibitors are used in the clinic on the strength of its disease relevance; for example vorinostat is FDA-approved for cutaneous T-cell lymphoma.[7] Dual-targeting of the L-C-H complex is now feasible with the discovery of the hybrid inhibitor corin that engages the LCH

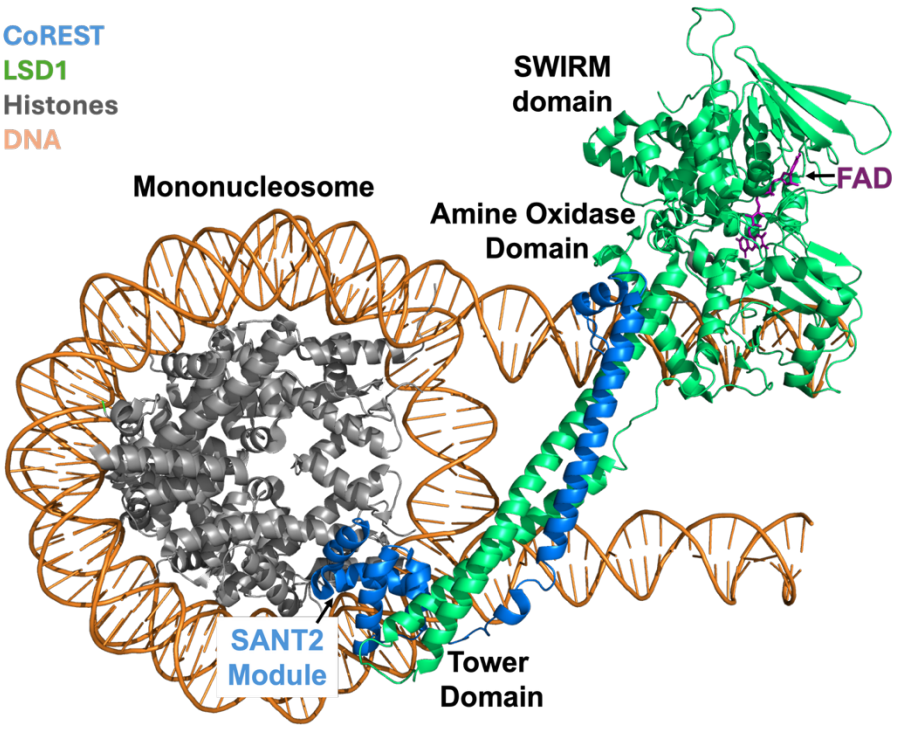


Figure 2. X-ray crystal structure of the LSD1-CoREST1 protein sub-complex bound to its mononucleosome substrate. PDB code 6VYP.

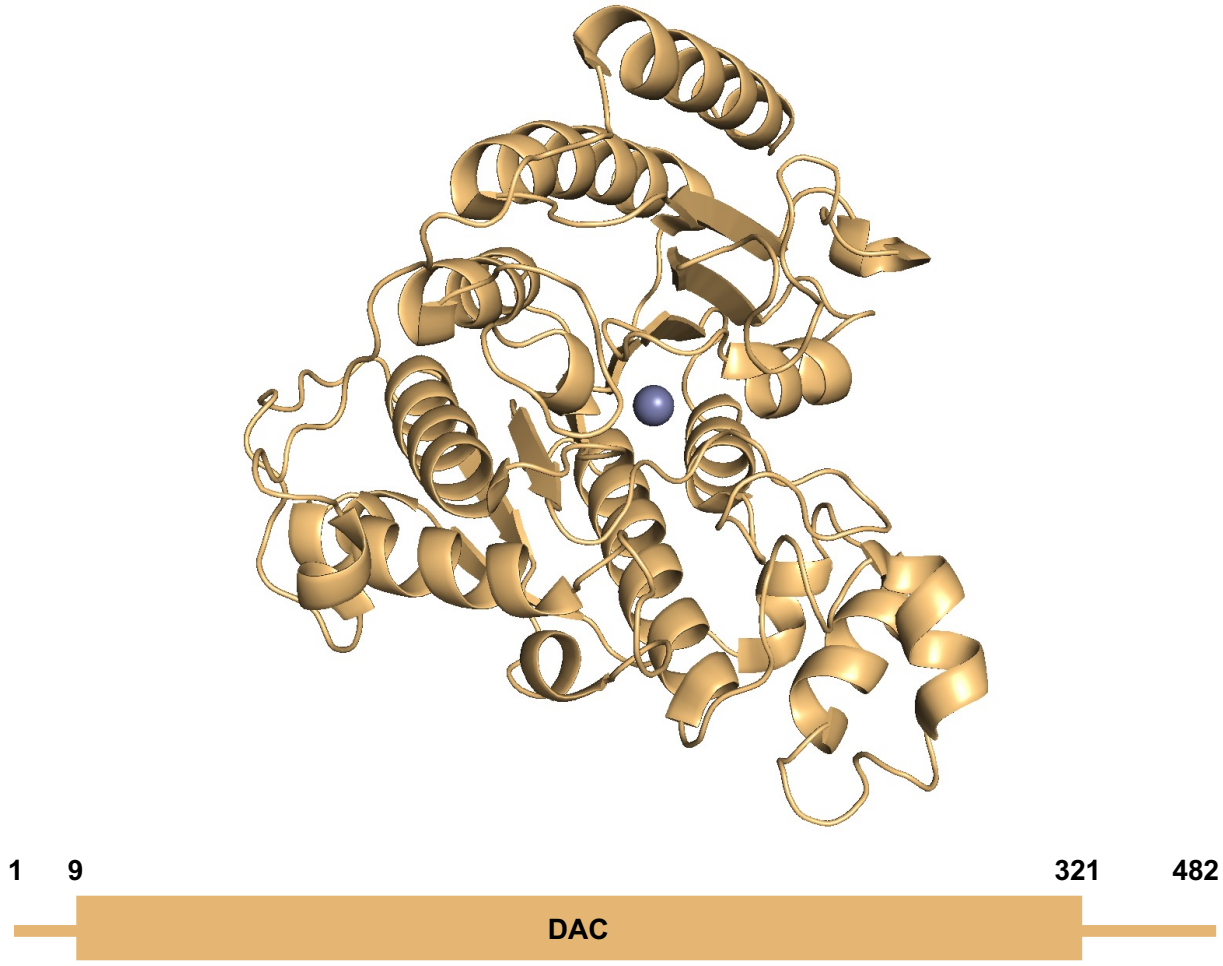


Figure 3. Structure of HDAC1. Crystal structure of HDAC1(8-376) with a Zn^{2+} ion bound at its active site (PDB: 4BKX). A full-length sequence map showing the position of the structured deacetylase domain (DAC) is shown below the structure.

complex and achieves sustained inhibition of complex-embedded HDAC activity while co-targeting LSD1.[8] The application of selective LSD1 inhibitors, including iadademstat, in clinical trials underscores the therapeutic promise of targeting the CoREST1–LSD1 pathway in human hematological and solid cancers.[9]

CoREST1 is a modular scaffold that couples and tunes both demethylase and deacetylase activities. The N-terminal ELM2–SANT1 region (**Figure 4**) recruits and stimulates HDAC1/2, whereas an internal helical linker makes a helical bundle with the LSD1 Tower helices, while the

C-terminal SANT2 domain engages nucleosomal DNA (**Figures 2 and 4**). Together, these contacts enable the LCH complex to efficiently act on acetylated and methylated histone tails.[2,3,6] Additionally, extranucleosomal DNA can also enhance LSD1–CoREST1 activity, which is consistent with the chromatin-tethering role proposed for the CoREST1 SANT2 domain.[6]

The SUMO protein in chromatin may further regulate LHC activity. SUMO3 is installed on histone H4 at Lys12 (H4K12su), and histone sumoylation is broadly associated with transcriptional repression and with the recruitment of proteins that favor heterochromatin formation, including HDACs and heterochromatin protein 1 (HP1).[10] Early biophysical work from the Chatterjee lab revealed that, despite its association with gene repression, H4K12su inhibits chromatin fiber compaction to a degree comparable to a gene activating mark like H4 K16 acetylation. This suggested that any role for H4K12su in gene repression may arise from its biochemical crosstalk with gene activating modifications such as acetylation and methylation.[11] This motivated our lab

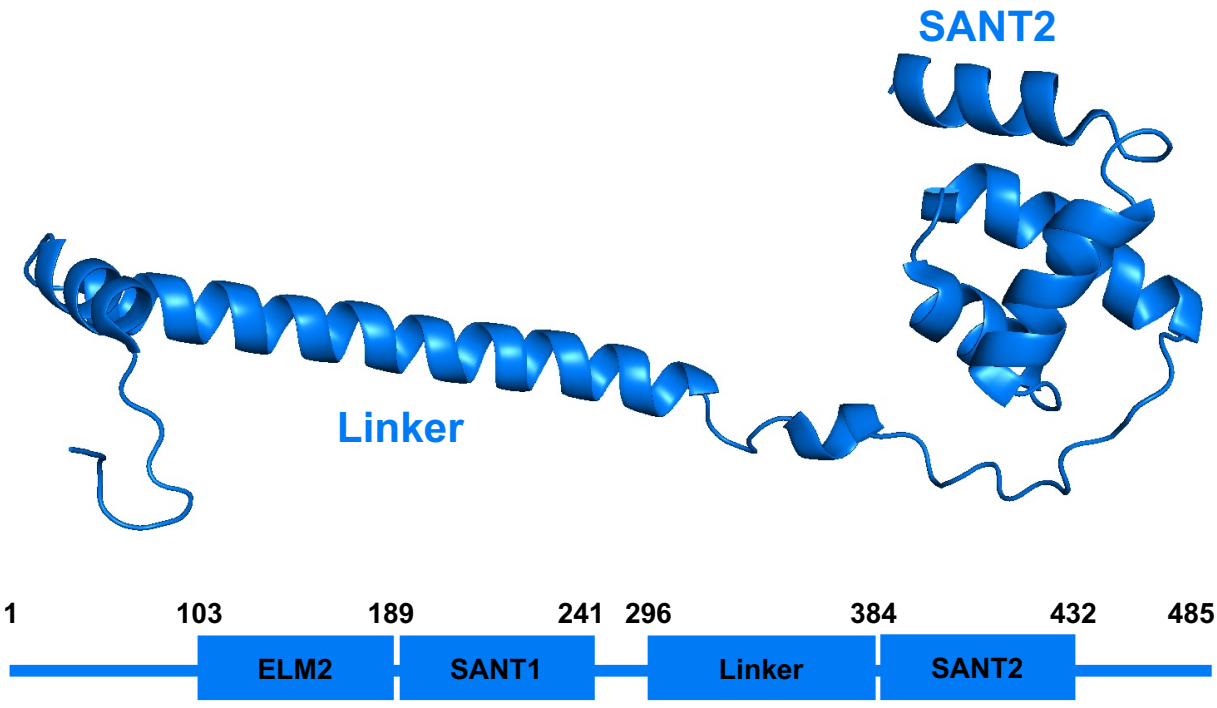


Figure 4. Structure of CoREST1. The crystal structure of CoREST1(308-440) (PDB: 2V1D). A full-length sequence map showing relative positions of domains is shown below the structure.

to investigate a role for SUMO in recruiting and/or stimulating gene-repressive complexes such as the LHC complex that remove gene activating acetyl and methyl marks.

A large majority of proteins that interact with SUMO contain a SUMO-interacting motif (SIM) sequence. Canonical SIMs consist of a hydrophobic core often described by a V/I-V/I-X-V/I-like pattern that is frequently flanked by acidic residues and, in some cases, regulated by proximal phosphoserines that serve to tune SUMO-paralog binding preference by the SIM.[12,13] NMR-based studies with single SIM peptides and high-throughput biochemical assays with peptide-arrays showed that SIMs extend the β-sheet of SUMO in parallel or antiparallel orientations, with key sequence features dictating SUMO-SIM affinity as well as the paralog selectivity of each SIM.[14] Gill and coworkers had previously identified an atypical, or non-canonical, SIM that spans residues ~255–275 in CoREST1 and enables CoREST1 to bind SUMO2/3 but not SUMO1. Mutational scanning of this region in CoREST1 indicated that the acidic patch, hydrophobic core, and a short intervening segment each contribute to SUMO3 binding, consistent with known canonical SIMs.[15]

The Chatterjee lab envisioned that this SIM-mediated binding of CoREST1 to SUMO3 may be one mechanism by which the H4K12su mark stimulates LCH catalysis in chromatin, which leads to gene repression. By incorporating semisynthetic H4K12su and H3K4me2 in nucleosomes and subjecting these to biochemical assays with the LSD1-CoREST1 sub-complex, we showed that site-specifically sumoylated H4 stimulates intranucleosomal demethylation of H3K4me2 by LSD1–CoREST1, and that mutations in the CoREST non-canonical SIM abrogate the stimulation of LSD1 activity. Independent work in the lab also discovered that H4 sumoylation suppresses p300-mediated transcription, reinforcing the repressive role of H4K12su in chromatin.[16,17] Building on our discoveries, we further demonstrated that H4K12su also stimulates nucleosome deacetylation by the HDAC1–CoREST1 sub-complex. The mechanism of stimulation also involves the non-canonical SIM in CoREST1. These findings led us to a simple hypothesis,

namely, that SUMO3 positions CoREST1 on nucleosomes in a manner that stimulates LSD1 and HDAC1 activity. However, there was no direct evidence in our studies that SUMO3 binds the full-length CoREST1 protein. In this chapter, I provide experimental support toward this critical premise of our biochemical model by probing the direct interaction between full-length CoREST1 and full-length SUMO3.

2.2 Results and discussion

2.2.1 CoREST1 binds SUMO3 through a non-canonical SIM (ncSIM)

We set out to test the direct binding between CoREST1 and SUMO3 after observing the SUMO-dependent stimulation of the CoRES1-LSD1 and CoREST1-HDAC1 subcomplexes on methylated and acetylated nucleosomes, respectively. The proposed SUMO3-binding region in CoREST1 spans residues 252–277 and contains an unusual hydrophobic core interspersed by negatively charged residues (D and E) and bracketed by acidic residues, which makes it a non-canonical SIM (ncSIM).[12,15] We employed NMR spectroscopy with isotopically labeled SUMO3 to identify the site of ncSIM binding. By performing HSQC-based titration experiments with ^{15}N-labeled SUMO3 and a 26-mer CoREST1 ncSIM peptide (residues 255-280), we measured an average Kd of ~2 mM for the ncSIM. Furthermore, we observed that SUMO3 residues with the largest chemical-shift perturbation (CSP) surround the known SIM-binding groove formed by α1-β2 in SUMO3 (**Figure 5A-C**).[12,14] This indicated that the ncSIM also binds to the canonical SIM-binding groove in SUMO3. The millimolar K_d we measured by NMR is well above typical cellular concentrations of individual proteins. However, the peptide-based measurement does not capture additional contributions from the full-length CoREST protein or the pre-association of CoREST with chromatin that would increases its local concentration near H4K12su. Therefore, to test the

binding of CoREST1 to SUMO3 at physiologically relevant concentrations we performed pulldown experiments with the purified full-length proteins.

A full-length 2xFLAG–CoREST1(1–485) construct was cloned, overexpressed in *E. coli* BL21(DE3) cells, purified to homogeneity and immobilized on anti-FLAG M2 resin. The resin was washed until no free CoREST1 was detected in the flow-through by SDS-PAGE. Next, 200 μM SUMO3(C47S) was mixed with the CoREST1-bound anti-FLAG M2 resin. Resin-only controls were included in every experiment by incubating the same 200 μM SUMO3(C47S) solution with anti-FLAG M2 resin alone and then performing an identical number of buffer washes prior to protein elution. In this way, the resin-only controls established the maximal amount of SUMO3(C47S) retained on M2 resin as a result of non-specific binding. Gratifyingly, elution of all bound material from the CoREST1-bound M2 resin showed a significant SDS-PAGE gel band at the size of SUMO3, whereas the resin-only control displayed a minimal SUMO3 band under identical experimental conditions (**Figure 5D, top**). Immunoblotting with a SUMO3-specific antibody confirmed that SUMO3 was indeed eluted from the CoREST1-bound M2 resin, but not from the control resin missing CoREST1 (**Figure 5D, bottom**). The enrichment of SUMO3 in CoREST1-pulldowns was consistent across three independent experiments, demonstrating that a stable complex is formed between CoREST1(1–485) and SUMO3 at far less than 2 mM concentrations, and that the predicted ncSIM of CoREST remains functional in the context of the full-length protein.

In order to establish that our observations were not an artefact due to immobilizing CoREST1 on the M2 resin, we also performed the reverse pulldown experiments using SUMO3 as the immobilized bait.

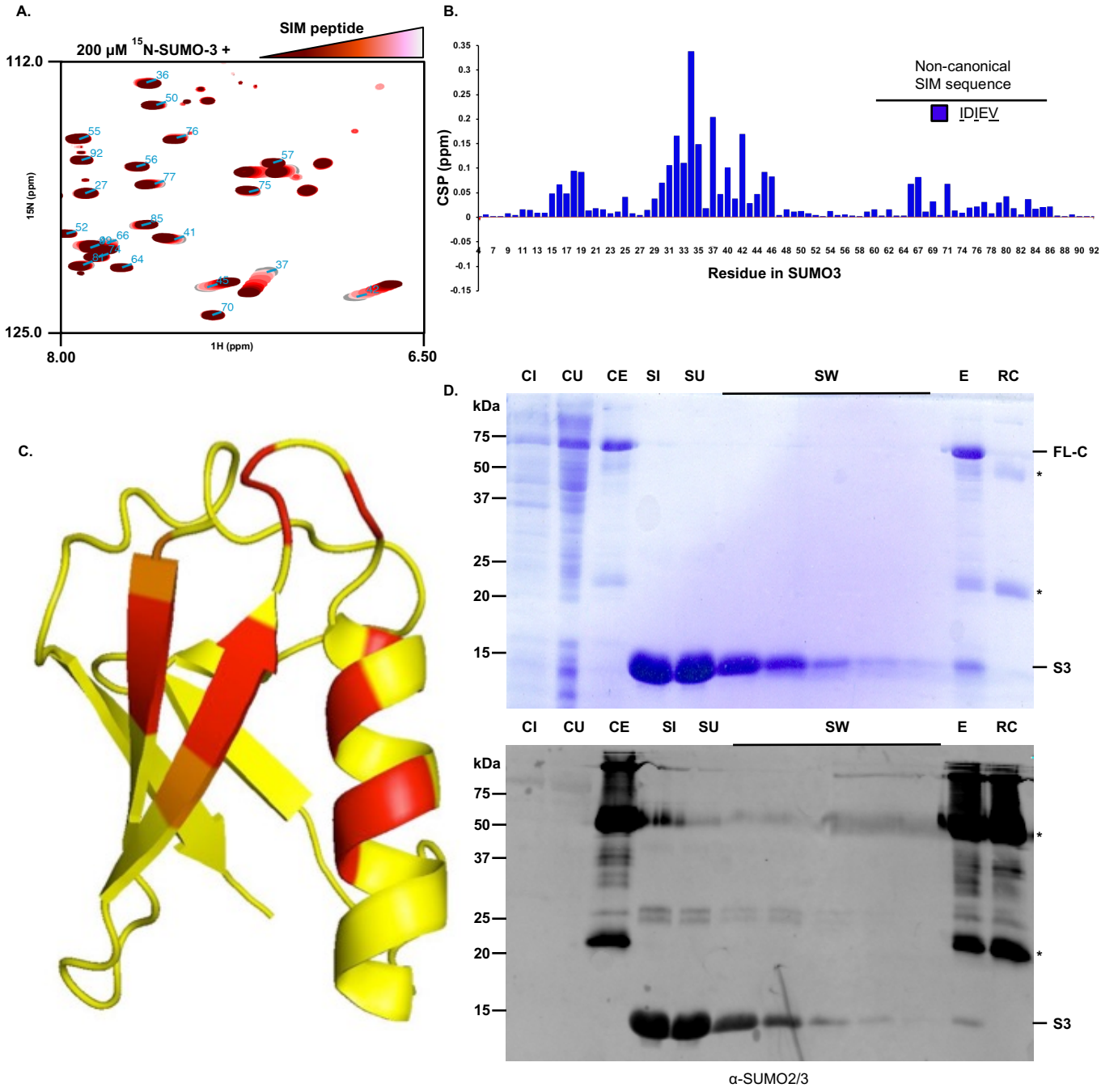


Figure 5. The CoREST1 ncSIM Peptide Binds SUMO3. (**A**) ^{1}H-^{15}N-HSQC NMR spectra of 200 μM SUMO3 titrated with 1 to 24 molar equivalents of CoREST1 (255-280). (**B**) Plot of chemical shift perturbations (CSP) versus amino acid residue within SUMO3 for CoREST1-ncSIM. The largest CSPs from the CoREST1-ncSIM titration correspond to residues K34, T37, and L42 of SUMO3 with next largest CSPs of SUMO3 being L19, F31, and T71. (**C**) Crystal structure with the largest CSPs highlighted. The three residues in red (K34, T37, and L42) indicate the largest chemical shift perturbations. The three residues in orange (L19, F31, and T71) indicate the next three largest CSPs. SUMO PDB: 1U4A. (**D**) Top: Coomassie stained 15% SDS-PAGE showing bead retention for full-length SUMO3 binding to immobilized full-length 2xFLAG-CoREST1 on anti-FLAG resin (E). Bottom: Immunoblot with an anti-SUMO2/3 antibody to confirm the presence of SUMO3 in the elution sample (E). Lanes: CI=2xFLAG-CoREST1 lysate Input, CU=2xFLAG-CoREST1 unbound fraction, CE=Immobilized 2xFLAG-CoREST1, SI=200 μM SUMO3 input, SU=Unbound SUMO3. E=Elution of purified 2xFLAG-CoREST1 bound to SUMO3. RC=Resin control Legend: FL-C=2xFLAG-CoREST1, S3=SUMO3, *= Immunoglobulin G

2.2.2 Reciprocal pulldowns confirm the SUMO3-CoREST1 interaction

To evaluate whether the precise protein immobilized on resin may influence the outcome of pulldown experiments, we repeated the assay using immobilized SUMO3 instead of CoREST1. GST-SUMO3(C47S) was overexpressed in E. coli BL21(DE3) cells, purified to homogeneity and immobilized on Glutathione Sepharose beads. The resin was washed until no free GST-SUMO3(C47S) was detected in the eluate by SDS-PAGE. Then freshly purified 2xFLAG-CoREST1(1-485) in buffered solution was incubated with the SUMO3-bound resin. A resin-control was also undertaken where resin bound to only GST, rather than GST-SUMO3(C47S), was incubated with purified CoREST1(1-485) and subjected to an identical number of buffer washes prior to protein elution. In this way, the resin-only control established the maximal amount of CoREST1(1-485) retained on Glutathione resin as a result of non-specific binding (**Supplemental Figure 1**). Finally, the identity of all proteins bound to immobilized GST-SUMO3(C47S) was confirmed by immunoblotting with a FLAG-specific antibody.

In three independent experiments, a band corresponding to the size of 2xFLAG-CoREST1 was retained on immobilized GST–SUMO3(C47S) in quantities significantly greater than its retention on GST-resin alone (**Figure 6A**). The anti-FLAG immunoblot confirmed the presence of 2xFLAG-CoREST1 when GST-SUMO3(C47S) was present on the resin (**Figure 6B**).

The consistent retention of either protein in pulldowns, when the other protein was immobilized on resin beads, supports a direct binding interaction between SUMO3 and full-length CoREST1 and is consistent with our NMR-based binding measurements with the isolated ncSIM from CoREST1.[12,14]

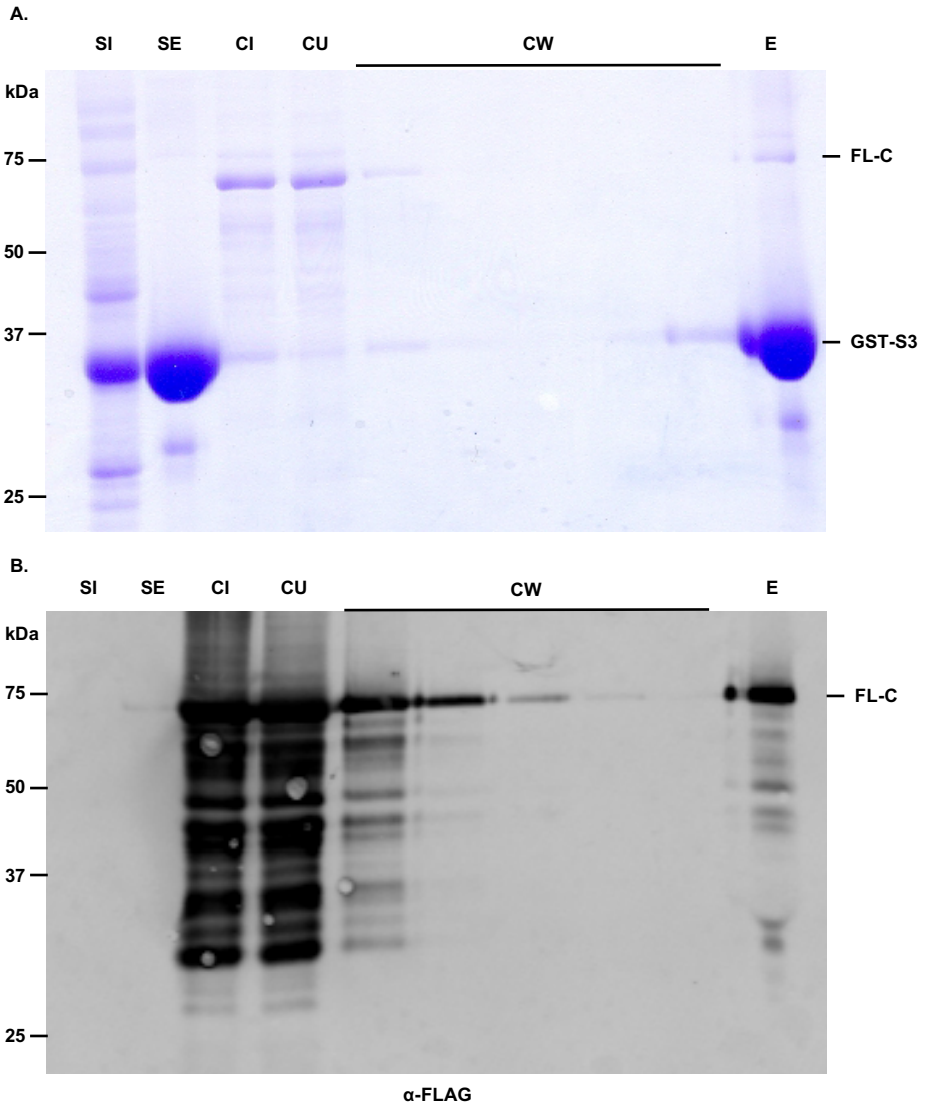


Figure 6. Reciprocal pulldown of full-length GST-SUMO3 with 2xFLAG-CoREST1. (**A**) Coomassie stained 15% SDS-PAGE showing bead retention for full-length 2xFLAG-CoREST1 binding to immobilized full-length GST-SUMO3 on Glutathione Sepharose-4B resin (E). (**B**) Immunoblot with an anti-FLAG antibody to confirm the presence of 2xFLAG-CoREST1 in the elution sample (E). Lanes: SI=GST-SUMO3 lysate Input, SE=Immobilized GST-SUMO3, CI=2xFLAG-CoREST1 lysate input, CU=2xFLAG-CoREST1 unbound fraction. CW=2xFLAG-CoREST1 washes. E=Elution of purified GST-SUMO3 bound to 2xFLAG-CoREST1. FL-C=2xFLAG-CoREST1, GST-S3=GST-SUMO3.

2.2.3 Hydrophobic residues in the CoREST1 nSIM are critical for SUMO3 binding

It is well-established that hydrophobic V/I/L residues in canonical SIM sequences interact with hydrophobic residues Val30, Phe32, Ile34, Leu43 and Ala46 in the SIM-binding groove of SUMO3. However, it remains unknown if the ncSIM hydrophobic core is required for binding to SUMO3. To address this question by 2D-NMR using chemical shift perturbations, we replaced the hydrophobic residues I273, I275 and V277 in the 26-mer CoREST1 ncSIM peptide with Ala (ncSIM-3A) and measured the resultant effect on SUMO3 binding in solution by HSQC experiments with ^{15}N-SUMO3. When compared to the wild-type ncSIM peptide, ncSIM-3A produced markedly attenuated chemical-shift perturbations across the SIM-binding groove made by α1 and β2 in SUMO3. Furthermore, we observed no saturation of CSPs at the highest equivalent of peptide employed (4.8 mM, or 24-fold relative to 200 μM ^{15}N-SUMO3). Due to its poor binding to SUMO3, a dissociation constant could not be calculated for the ncSIM-3A peptide. Nevertheless, our results demonstrated that the hydrophobic residues in the CoREST1 ncSIM peptide are critical for binding SUMO3.

Finally, we asked if the same triple Ala mutation in CoREST(1-485) abolishes binding in pulldown assays. Toward this, GST-SUMO3(C47S) was immobilized on Glutathione Sepharose-4B resin and incubated with either freshly purified 2xFLAG–CoREST1(1-485) or with the triple Ala mutant protein 2xFLAG-CoREST1(1-485)I273A/I275A/V277A. Although wt CoREST showed robust binding to GST-SUMO3(C47S), the mutant CoREST did not show binding above the baseline non-specific binding observed with GST-bound resin alone (**Figure 7B**). We also performed the reciprocal pulldown assays with immobilized 2xFLAG–CoREST1(1-485) or 2xFLAG-CoREST1(1-485)I273A/I275A/V277A incubated with label-free pure SUMO3(C47S) in solution. In these reciprocal pulldowns, the 2xFLAG-CoREST1(1-485)I273A/I275A/V277A mutant failed to retain SUMO3(C47S) on the column (**Figure 7A**), while wt CoREST1 bound SUMO3(C47S) that was visible after elution from the resin. Thus, irrespective of the protein immobilized on resin, 2xFLAG-CoREST1(1-485)I273A/I275A/V277A failed to bind wt SUMO3(C47S) indicating the importance

of the hydrophobic residues in the ncSIM for binding to SUMO3 in the context of full-length CoREST1.

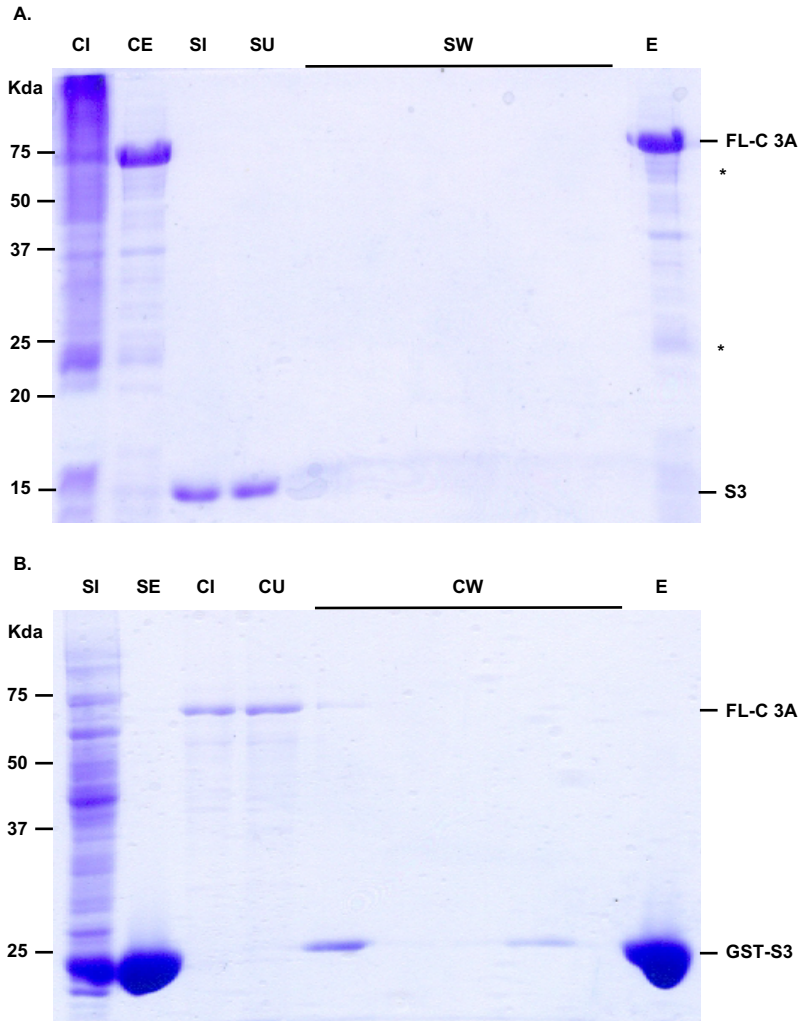


Figure 7. CoREST1-ncSIM-3A does not bind SUMO3 in reciprocal pull-downs. (**A**) Coomassie-stained 15% SDS-PAGE showing no bead retention of full-length SUMO3 after incubation with immobilized full-length 2xFLAG-CoREST1-ncSIM-3A on anti-FLAG resin; eluates contain only the immobilized bait. Lanes: Lanes: CI = 2xFLAG-CoREST1-ncSIM-3A lysate input; CE = 2xFLAG-CoREST1-ncSIM-3A bait elution control from anti-FLAG resin; SI = SUMO3 input; SU = SUMO3 unbound fraction; SW = SUMO3 washes; E = elution from anti-FLAG resin after incubation. (**B**) Reciprocal assay with immobilized full-length GST-SUMO3 on Glutathione Sepharose-4B shows no retention of full-length 2xFLAG-CoREST1-ncSIM-3A; eluates contain only GST-SUMO3. Lanes: SI = GST-SUMO3 lysate input; SE = immobilized GST-SUMO3; CI = 2xFLAG-CoREST1-ncSIM-3A lysate input; CU = 2xFLAG-CoREST1-ncSIM-3A unbound fraction; CW = 2xFLAG-CoREST1-ncSIM-3A washes; E = elution from resin. FL-C 3A = 2xFLAG-CoREST1-ncSIM-3A; GST-S3 = GST-SUMO3.

2.3 Conclusion and outlook

Using both 2-D NMR and pulldown assays, we conclusively established the direct interaction between CoREST1 and SUMO3 that we had hypothesized from nucleosome demethylation and deacetylation assays with sumoylated nucleosomes and the LC and HC sub-complexes. 2-D NMR-based titration experiments with the 26-mer CoREST ncSIM peptide and isotopically labeled ^{15}N-SUMO revealed the low affinity (K_d ~2 mM) but high specificity of the SUMO3-CoREST1 interaction. For the first time, we showed that the CoREST1 ncSIM binds the α1-β2 groove in SUMO3 that is also known to bind canonical SIM sequences. Qualitative binding assays using biochemical pulldown experiments also detected binding when either 2xFLAG–CoREST1(1-485) or GST-SUMO3(C47S) was used as the bait protein. Consistent with previous studies with SIM peptides, Ala mutagenesis of the three hydrophobic residues in the ncSIM core, I273, I275 and V277,significantly attenuated peptide binding in HSQC-based binding experiments and also eliminated binding of the full-length CoREST1 and SUMO3 proteins in pulldowns. These convergent results lend strong support to the hypothesis that H4K12su directly binds CoREST1 and stimulates the activity of LSD1 and HDAC1 enzymes. We propose that SUMO3 in the H4 tail likely recruits and positions the LHC complex in a more productive conformation for demethylation and deacetylation.

Future work in our lab will seek to integrate human genetic variation information for CoREST1 by mining the Catalogue of Somatic Mutations in Cancer (COSMIC) and cBioPortal to identify RCOR1 missense mutations found in the ncSIM. Variants will be mapped onto the CoREST1 sequence and prioritized for investigation by their frequency of occurrence in diseases and by their localization in the acidic or hydrophobic regions of the ncSIM sequence. These mutations

will be introduced into stable human cell lines and their transcriptional outcomes compared with cells containing the wt CoREST1(1–485) gene. Based on the initial work by Gill and coworkers, we expect that a subset, but not all, CoREST1-target genes will be misregulated in the CoREST1 ncSIM mutants due to their inability to read the SUMO code for gene silencing in chromatin.

2.4 Experimental procedures

2.4.1 Molecular cloning

2xFLAG tagging of pET28b–His_6–CoREST1 and pET28b–His_6–CoREST1-ncSIM-3A constructs

The CoREST1 open reading frame refers to the full CoREST1(1–485) coding sequence in pET28b-His_6-CoREST1 or pET28b-His_6-CoREST1-ncSIM-3A. Vectors were linearized by restriction digest with NdeI and BamHI to open the multiple cloning site downstream of the N-terminal His_6-tag. Complementary oligonucleotides encoding a 2xFLAG epitope (Table 2.4.1) were annealed to generate a double-stranded insert with 5′ overhangs compatible with NdeI and BamHI. The digested backbone was treated with calf intestinal phosphatase to reduce vector re-ligation. The dephosphorylated backbone and annealed 2xFLAG insert were ligated with T4 DNA ligase to generate His_6-2xFLAG-CoREST1 and His_6-2xFLAG-CoREST1-ncSIM-3A constructs. Ligation products were transformed into XL10-Gold chemically competent cells and plated on LB-kanamycin agar plates. Plasmid DNA from individual colonies was prepared by miniprep, quantified by absorbance, and submitted for Sanger sequencing to verify the His_6-2xFLAG junction and the CoREST1 N terminus.

Construction of pGEX-2T-GST-SUMO3(C47S)

The pGEX-2T-GST-p53 plasmid provided the vector backbone. The vector was linearized by PCR to remove the p53 insert while retaining the GST open reading frame and the thrombin site using backbone-linearization primers (Table 2.4.1). The SUMO3(C47S) insert was amplified from pTXB1–SUMO3(C47S)-Ava-His_6 plasmid with primers that appended homology to the linearized pGEX-2T ends (Table 2.4.1). PCR products were resolved by agarose electrophoresis and gel purified. The linearized pGEX-2T backbone and SUMO3(C47S) insert were assembled by Gibson Assembly at 50 °C for 60 min using NEB Gibson Assembly Master Mix. Reactions were transformed into XL10-Gold and selected on LB supplemented with ampicillin. Plasmid DNA was verified by Sanger sequencing across both junctions and through the SUMO3(C47S) insert.

Construction of pGEX-2T–GST (GST-only bait).

The pGEX-2T-GST-p53 backbone was used to generate a GST-only construct by introducing a stop codon immediately downstream of the GST coding sequence by site-directed mutagenesis with primers (Table 2.4.1). Post-PCR reactions were treated with KLD mix according to the manufacturer's protocol, transformed into XL10-Gold cells, and selected on LB supplemented with ampicillin. Plasmid DNA was confirmed by Sanger sequencing across the GST terminus to verify the stop codon and reading frame.

Table 2.4.1 List of ssDNA oligos used for molecular cloning.

ssDNA Oligos	**DNA Sequence (5'- to -3')**
2xFLAG_CoREST1_Sense (NdeI)	TAT GGA TTA TAA AGA TGA TGA TGA CAA AGA TTA CAA AGA CGA CGA TGA TAA GCG G
2xFLAG_CoREST1_Antisense (BamHI)	GAT CCC GCT TAT CAT CGT CGT CTT TGT AAT CTT TGT CAT CAT CAT CTT TAT AAT CCA

GST_Stop_FWD	TAA TCG GAT CTG GAA GTT C
GST_Stop_REV	TTT TGG AGG ATG GTC G
SUMO_Gib_FWD	GTG GAT CCA TGT CCG AAG AAA AGC CCA AGG
SUMO_Gib_REV	ATC GTC AGT CCC GTC TGC TGC TGG AAC
GST_Gibson_FWD	GAC GGG ACT GAC GAT CTG CCT CGC
GST_Gibson_REV	TCG GAC ATG GAT CCA CGC GGA ACC AG

2.4.2 Expression, purification, and intein-mediated hydrolysis of SUMO3(C47S)-Ava-His$_6$ to yield tag-free SUMO3(C47S)

Four liters of Luria-Bertani (LB) medium supplemented with ampicillin at 100 µg/mL were inoculated with E. coli BL21(DE3) cells carrying the pTXB1–SUMO3(C47S)-Ava-His$_6$ plasmid and grown at 37 °C with shaking to an OD_{600} of 0.6-0.8. Protein expression was induced with 0.5 mM IPTG and cultures were shifted to 16 °C for 16-18 h. Cells were harvested by centrifugation at 7,000 x g for 30 min at 4 °C. Pellets from each liter were resuspended on ice in approximately 40 mL lysis buffer composed of 50 mM Na_2HPO_4 at pH 8.0, 300 mM NaCl, and 5 mM Imidazole. Cells were lysed by pulse sonication for three rounds of 3 min with 5 min rest intervals on ice. Lysates were clarified by centrifugation at 20,000 x g for 30 min at 4 °C, the supernatant was passed through a 0.45 µm filter, and the filtrate was applied to pre-equilibrated Ni–NTA resin at a loading of 4 mL resin per liter of cell culture. Protein binding was allowed for 30 min at 4 °C with gentle rotation. The flow-through was collected, and the resin was washed with three 50 mL steps of lysis buffer containing 5 mM, then 20 mM, and finally 50 mM Imidazole. Bound protein was eluted in approximately 2 mL fractions in a total volume of 15 mL using lysis buffer containing 250 mM, then 350 mM, then 500 mM Imidazole. Elution fractions containing SUMO3(C47S)-Ava-His$_6$ were pooled and dialyzed at 4 °C against 2 × 2 L of 100 mM Na_2HPO_4 at pH 8.0, 150 mM NaCl,

and 1 mM DTT. After dialysis, the DTT concentration was increased to 50 mM and the sample was incubated at 30 °C for 5 h to induce intein-mediated thiolysis followed by water-mediated hydrolysis of the SUMO3(C47S)-DTT thioester. Reaction progress was monitored by 15% SDS-PAGE with Coomassie staining. After the hydrolysis was complete, the reaction mixture was flash-frozen and lyophilized. The dried material was analyzed by C18 analytical reversed-phase HPLC using a 5–75% B gradient and fractions were analyzed for the presence of hydrolyzed SUMO3(C47S) by ESI mass spectrometry. Preparative C18 reversed-phase HPLC with a 5–75% B gradient was used to purify the SUMO3(C47S) protein. Fractions were first assessed by mass spectrometry, and each clean fraction was subjected to an overnight analytical HPLC run to further confirm purity. Analytically pure fractions were pooled and lyophilized to afford approximately 7 mg of tag-free SUMO3(C47S). Calculated M_{avg} = 10,377.6 Da; found 10,376.5 ± 0.4 Da (**Supplementary Figure 2**).

2.4.3 Reverse-phase High pressure liquid chromatography (HPLC)

SUMO3(C47S) was analyzed (4.6 x 150 mm, 5 µm) and purified (22 x 250 mm, 15 - 20 µm) with C4 and C18 RP-HPLC columns from Vydac (Deerfield, IL) on either a Varian Prostar (Palo Alto, CA) or Agilent (Santa Clara, CA) 1260 Infinity II LC system. The mobile phase consisted of Buffer A (0.1% trifluoroacetic acid in water), and Buffer B (90% acetonitrile in water, 0.1% trifluoroacetic acid). Eluting SUMO3(C47S) was monitored by its absorbance at 214 and 280 nm.

2.4.4 Electrospray ionization mass spectrometry (ESI-MS)

SUMO3 mass spectrometry was performed by direct infusion on a Bruker (Billerica, MA) Esquire ion-trap mass spectrometer operating in positive ion mode.

2.4.5 Biochemical pulldowns with immobilized wt and mutant FLAG-CoREST1.

E. coli BL21(DE3) Rosetta cells carrying pET28b-His_6-2xFLAG-CoREST1 wild type or pET28b-His_6-2xFLAG-CoREST1(I273A/I275A/V277A) mutant were grown in 1 L of 2×YT supplemented with kanamycin at 50 µg/mL and chloramphenicol at 34 µg/mL at 37 °C with shaking to an OD_{600} of 0.6–0.8. Cultures were shifted to 16 °C for 1 h, induced with IPTG at 1 mM, and grown for 16–18 h at 16 °C. Cells were harvested at 7,000 x g for 30 min at 4 °C and resuspended on ice in binding buffer composed of 20 mM Tris pH 7.5, 100 mM NaCl, 0.1 mM EDTA, 0.1% (v/v) IGEPAL, 5% (v/v) glycerol, and 1 mM DTT. PMSF was added to 0.2 mM and cOmplete protease inhibitor cocktail (Roche) was added at 200 µL per liter of culture from a stock made from 1 tablet dissolved in 1 mL water. Cells were lysed by three passes through a French press at about 50 psi followed by brief sonication to reduce viscosity. Lysates were clarified at 20,000 x g for 20 min at 4 °C and the supernatants were passed through 0.45 µm syringe filters.

Anti-FLAG M2 resin was pre-equilibrated on ice by three washes with binding buffer. For each pulldown, 20 µL of settled resin was combined with 1 mL of clarified wt CoREST1 or triple Ala mutant CoREST1- lysate and nutated for 1 h at 4 °C. Resin was pelleted at 6,000 × g in a pre-chilled microcentrifuge, the unbound fraction was removed, and a second 1 mL aliquot of the same lysate was added and incubated for an additional 1 h at 4 °C. The resin was washed five times with 500 µL of binding buffer per wash, and washes continued until wash fractions showed no FLAG-CoREST1 or other impurities by SDS-PAGE. A resin-only tube that received no lysate was treated to the exact same number of washes, in parallel, and served as a control for non-specific retention of SUMO3(C47S) on M2 resin.

For SUMO3 binding, wt CoREST1- or triple Ala mutant CoREST1-bound anti-FLAG M2-resin was incubated with native, label-free SUMO3(C47S) at 200 μM in binding buffer at a total volume of 500 μL for 2 h at 4 °C with gentle nutation. The same batch of SUMO3(C47S) solution was used for both forms of CoREST1. Folowing binding, the resin was pelleted at 6,000 x g, the unbound fraction was saved, and the resin was washed a further five times with 500 μL of binding buffer with each wash collected separately. The resin-only control was incubated with the same 200 μM SUMO3(C47S) solution and washed identically prior to elution, in order to identify the degree of nonspecific retention of SUMO3(C47S) on the anti-FLAG M2 resin.

The washed resins were resuspended in 1x Laemmli sample buffer and heated at 95 °C for 2 min to release all bound proteins. Eluted samples were resolved on 12% SDS–PAGE. All pulldowns were performed in triplicate with resin-only controls processed identically and in parallel for both wt CoREST1 and triple Ala mutant CoREST1 pulldowns.

2.4.6 Biochemical pulldowns with immobilized SUMO3(C47S).

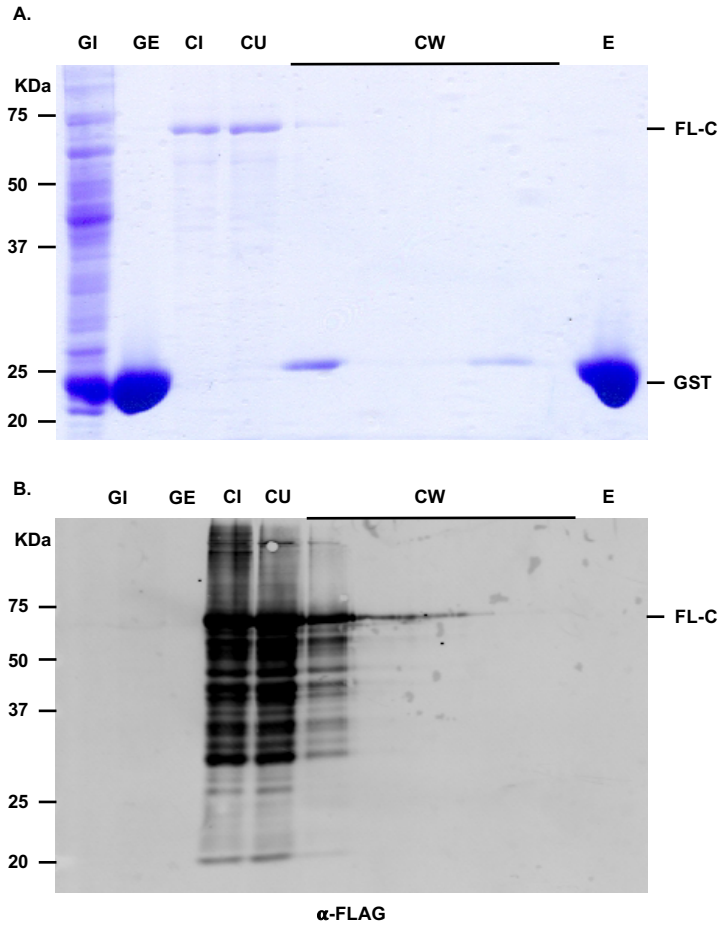
E. coli BL21(DE3) cells carrying the pGEX-2T-GST-SUMO3(C47S) plasmid were grown in 1 L LB containing ampicillin at 100 μg/mL at 37 °C with shaking to an OD_{600} of 0.6-0.8. Protein expression was induced with IPTG at 1 mM and cultures were grown for 4 h at 37 °C. Cells were harvested at 7,000 × g for 30 min at 4 °C and resuspended on ice in binding buffer composed of 20 mM Tris pH 7.5, 100 mM NaCl, 0.1 mM EDTA, 0.1% (v/v) IGEPAL CA-630, 5% (v/v) glycerol, and 1 mM DTT, supplemented with PMSF at 0.2 mM and cOmplete protease inhibitor cocktail at 1×. Cells were lysed by pulsed sonication on ice. Lysates were clarified at 20,000 x g for 30 min at 4 °C and the supernatant was passed through a 0.45 μm syringe filter. In parallel, *E. coli* BL21(DE3) cells carrying pGEX-2T-GST were grown, induced, lysed, clarified, and filtered in the same manner to generate GST bait for resin-control pulldowns.

Glutathione Sepharose 4B resin was pre-equilibrated on ice by three washes with binding buffer. For each pulldown, 20 µL settled resin was combined with 1 mL clarified GST-SUMO3(C47S) lysate and nutated for 1 h at 4 °C. Resin was pelleted at 6,000 x g in a prechilled microcentrifuge, the unbound fraction was removed, and the resin was washed five times with 500 µL binding buffer per wash until wash fractions showed no eluting proteins by SDS–PAGE. A GST-only bait was prepared by binding glutathione resin with clarified pGEX-2T-GST lysate generated as above, and this resin-only control tube was carried through all subsequent steps.

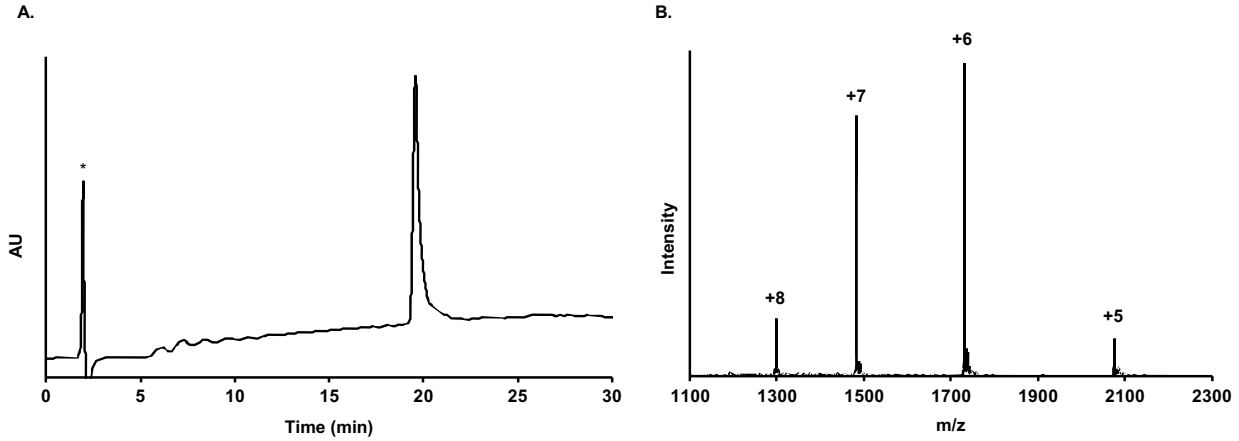
CoREST1 lysates were prepared as described for His_6–2xFLAG–CoREST1 wild type and His_6-2xFLAG-CoREST1(I273A/I275A/V277A). Either GST-SUMO3(C47S)-bound resin or GST-bound resin was incubated with 1 mL of clarified CoREST1 wild type or triple Ala mutant CoREST1 lysate for 1 h at 4 °C with nutation. The resin was pelleted at 6,000 x g, the unbound fraction was saved, and the resin was washed five times with 500 µL binding buffer with each wash collected. The resin-only control was incubated with the same CoREST1 lysates and processed with identical washes to identify nonspecific retention on the resin beads.

Resins were resuspended in 1x Laemmli sample buffer and heated at 95 °C for 2 min. Eluted samples were resolved on 12% SDS–PAGE. All pulldowns were performed in triplicate, with GST-resin-only controls processed in parallel for both wild-type CoREST1 and triple Ala mutant CoREST1 containing lysates.

2.5 Product characterization and supplemental data



Supplementary Figure 1. GST control shows no 2xFLAG-CoREST1 retention. (**A**) Coomassie stained 15% SDS-PAGE showing bead retention for full-length 2xFLAG-CoREST1 incubated with immobilized GST on Glutathione Sepharose-4B resin (E). (**B**) Immunoblot with an anti-FLAG antibody to probe the presence of 2xFLAG-CoREST1 in the elution sample (E). Lanes: GI=GST lysate Input, GE=Immobilized GST, CI=2xFLAG-CoREST1 lysate input, CU=2xFLAG-CoREST1 unbound fraction. CW=2xFLAG-CoREST1 washes. E=Elution of Glutathione Sepharose-4B-GST bound resin following incubation with 2xFLAG-CoREST1. FL-C=2xFLAG-CoREST1, GST=GST



Supplemental Figure 2. Characterization of SUMO3(C47S). (A) C18 analytical HPLC chromatogram of on a gradient of 0–73% B over 30 min. * indicate the buffer peak. (B) ESI-MS of pure SUMO3(C47S). Observed M_{avg}= 10,376.5 ± 0.4 Da, calculated M_{avg} = 10,377.6 Da.

2.6 References

1 Andres, M. E. *et al.* CoREST: a functional corepressor required for regulation of neural-specific gene expression. *Proc Natl Acad Sci U S A* **96**, 9873–9878 (1999). https://doi.org/10.1073/pnas.96.17.9873

2 You, A., Tong, J. K., Grozinger, C. M. & Schreiber, S. L. CoREST is an integral component of the CoREST- human histone deacetylase complex. *Proc Natl Acad Sci U S A* **98**, 1454–1458 (2001). https://doi.org/10.1073/pnas.98.4.1454

3 Yang, M. *et al.* Structural basis for CoREST-dependent demethylation of nucleosomes by the human LSD1 histone demethylase. *Mol Cell* **23**, 377–387 (2006). https://doi.org/10.1016/j.molcel.2006.07.012

4 Forneris, F., Binda, C., Vanoni, M. A., Mattevi, A. & Battaglioli, E. Histone demethylation catalysed by LSD1 is a flavin-dependent oxidative process. *FEBS Lett* **579**, 2203–2207 (2005). https://doi.org/10.1016/j.febslet.2005.03.015

5 Kim, S. A., Zhu, J., Yennawar, N., Eek, P. & Tan, S. Crystal Structure of the LSD1/CoREST Histone Demethylase Bound to Its Nucleosome Substrate. *Mol Cell* **78**, 903–914 e904 (2020). https://doi.org/10.1016/j.molcel.2020.04.019

6 Kim, S. A., Chatterjee, N., Jennings, M. J., Bartholomew, B. & Tan, S. Extranucleosomal DNA enhances the activity of the LSD1/CoREST histone demethylase complex. *Nucleic Acids Res* **43**, 4868–4880 (2015). https://doi.org/10.1093/nar/gkv388

7 Mann, B. S., Johnson, J. R., Cohen, M. H., Justice, R. & Pazdur, R. FDA approval summary: vorinostat for treatment of advanced primary cutaneous T-cell lymphoma. *Oncologist* **12**, 1247–1252 (2007). https://doi.org/10.1634/theoncologist.12-10-1247

8 Kalin, J. H. *et al.* Targeting the CoREST complex with dual histone deacetylase and demethylase inhibitors. *Nat Commun* **9**, 53 (2018). https://doi.org/10.1038/s41467-017-02242-4

9 Salamero, O. *et al.* First-in-Human Phase I Study of Iadademstat (ORY-1001): A First-in-Class Lysine-Specific Histone Demethylase 1A Inhibitor, in Relapsed or Refractory Acute Myeloid Leukemia. *J Clin Oncol* **38**, 4260–4273 (2020). https://doi.org/10.1200/JCO.19.03250

10 Shiio, Y. & Eisenman, R. N. Histone sumoylation is associated with transcriptional repression. *Proc Natl Acad Sci U S A* **100**, 13225–13230 (2003). https://doi.org/10.1073/pnas.1735528100

11 Dhall, A. *et al.* Sumoylated human histone H4 prevents chromatin compaction by inhibiting long-range internucleosomal interactions. *J Biol Chem* **289**, 33827–33837 (2014). https://doi.org/10.1074/jbc.M114.591644

12 Hecker, C. M., Rabiller, M., Haglund, K., Bayer, P. & Dikic, I. Specification of SUMO1- and SUMO2-interacting motifs. *J Biol Chem* **281**, 16117–16127 (2006). https://doi.org/10.1074/jbc.M512757200

13 Chang, C. C. *et al.* Structural and functional roles of Daxx SIM phosphorylation in SUMO paralog-selective binding and apoptosis modulation. *Mol Cell* **42**, 62–74 (2011). https://doi.org/10.1016/j.molcel.2011.02.022

14 Namanja, A. T. *et al.* Insights into high affinity small ubiquitin-like modifier (SUMO) recognition by SUMO-interacting motifs (SIMs) revealed by a combination of NMR and peptide array analysis. *J Biol Chem* **287**, 3231–3240 (2012). https://doi.org/10.1074/jbc.M111.293118

15 Ouyang, J., Shi, Y., Valin, A., Xuan, Y. & Gill, G. Direct binding of CoREST1 to SUMO-2/3 contributes to gene-specific repression by the LSD1/CoREST1/HDAC complex. *Mol Cell* **34**, 145–154 (2009). https://doi.org/10.1016/j.molcel.2009.03.013

16 Dhall, A., Weller, C. E., Chu, A., Shelton, P. M. M. & Chatterjee, C. Chemically Sumoylated Histone H4 Stimulates Intranucleosomal Demethylation by the LSD1-

CoREST Complex. *ACS Chem Biol* **12**, 2275–2280 (2017). https://doi.org/10.1021/acschembio.7b00716

17 Leonen, C. J. A. *et al.* Sumoylation of the human histone H4 tail inhibits p300-mediated transcription by RNA polymerase II in cellular extracts. *Elife* **10** (2021). https://doi.org/10.7554/eLife.67952

Chapter 3

Sumoylated histone H4 inhibits histone H2B ubiquitylation at K120 in human cells

3.1 Introduction

Our cells face the challenging task of packaging approximately six billion base pairs of genomic DNA into the very small nuclear volume of approximately 500 μm^3.[1] To achieve this, about 147 base pairs of double-stranded DNA are wrapped around a cylindrical protein *spool* composed of two H2A-H2B dimers and one H3-H4 tetramer to form the nucleosome core particle; the fundamental repeating unit of chromatin.[3] Although compacting genomic DNA is essential for fitting it into the limited nuclear space, cellular machinery still requires access to large stretches of DNA for transcription, replication and repair. The reversible regulation of compact DNA structures found in chromatin is undertaken by various post-translational modifications (PTMs) found in histones.[4-9] The chemical nature and genomic locations of these modifications, along with their complex interplay and crosstalk, are central to the "histone code" hypothesis[6,7], which posits that histone PTMs, either acting by themselves or in various combinations, encode specific signals that are interpreted, or *read*, by effector proteins to influence chromatin structure and gene expression.[8] Numerous histone modifications, such as methylation, acetylation, phosphorylation, and ubiquitylation, have been extensively studied through biochemical, structural, and cell-based molecular biological approaches. The extensive focus on understanding histone methylation, acetylation and phosphorylation was largely driven by their earlier discovery and high abundance in cells.[8,9] However, the roles for many newer histone PTMs remain to be elucidated due to the lack of chemical and molecular biological tools.

The small ubiquitin-like modifier protein (SUMO), approximately 90-100 amino acids in length, is the largest reversible covalent modification of histones known till date and was first identified >20 years ago. The mechanistic role for sumoylation in chromatin, however, remains poorly understood. SUMO conjugation on lysine side-chain amines proceeds through an E1 (SAE1/SAE2), E2 (Ubc9), and E3 (PIAS family ligases) cascade and protein sumoylation is reversed by SUMO-specific sentrin proteases (SENPs).[10,11] Unlike other well-studied histone PTMs, reliable tools for studying histone sumoylation, such as site-specific antibodies essential for their isolation and identification, are presently absent. Furthemore, sumoylated histones are <5% of total histones founds in cells.[12-16] Highly sensitive proteomics studies have discovered histone sumoylation in budding yeast, plants and humans and SUMO was found to modify all 4 core histones, the linker histone H1, and several histone variants.[16] Based on mutational and knockout or knockdown studies, histone sumoylation has been proposed to play several crucial cellular roles, including heterochromatin formation by the H1 linker histone, the regulation of chromatin structure and transcription, and double-strand break repair.[16] Histone H4 was the first histone reported to be sumoylated in human cells in 2003, with lysine 12 (H4K12su) emerging as a common site for SUMO2/3 modification in multiple proteomic studies.[17] [13,18] Although H4 sumoylation was proposed to participate in gene silencing, how it contributes to this process, especially within human chromatin, remained poorly understood. In particular, its biochemical crosstalk with other key histone PTMs directly involved in gene regulation, such as methylation and ubiquitylation, which are frequently misregulated in human disease, remained unclear. Clarifying the biochemical crosstalk involved in H4K12su is important as it connects a highly dynamic histone mark that is hard to study in native chromatin with important transcriptional outcomes.

To understand the biochemical roles for H4K12su, the Chatterjee lab developed a semisynthetic strategy utilizing disulfide-directed chemical sumoylation to install SUMO site specifically onto

histone H4 at Lys 12. This approach enabled reconstitution of the modified histone into octamers and assembly of defined nucleosomes and chromatin arrays.[19,20] The method for accessing homogenous H4K12su containing a disulfide linkage led to the discovery that sumoylation disrupts chromatin compaction, an unexpected observation for a modification associated with gene repression that is typically associated with a highly compact folded chromatin structure.[19,20] Subsequent biochemical studies with native isopeptide linked H4K12su obtained by chemical synthesis[2,21,22] revealed that H4K12su stimulated the intranucleosomal activity of lysine-specific histone demethylase 1 (LSD1) which demethylates the mono and dimethylated states of histone H3 Lys 4 (H3K4me1/2).[19] These studies revealed that although histone sumoylation biophysically prevents chromatin compaction, it may enable gene silencing by stimulating the activity of enzymes that erase histone modifications associated with gene activation.[16,19] [19,20]

Building on those early discoveries, the Chatterjee lab also examined how H4K12su affects two transcription-related chromatin modifications, acetylation and ubiquitylation. We found that H4K12su perturbs N-terminal histone H4 acetylation by the histone acetyltransferase p300. Additionally, H4K12su also inhibits trimethylation of histone H3 at Lys 4 (H3K4me3) by the Set1/COMPASS complex. These observations were also validated in human cells using a SUMO-H4 fusion model. The combined outcome of inhibiting at least acetylation and methylation, and potentially other factors as well, was reduced *in vitro* transcription by RNA polymerase II.[21]

Our findings raised important questions regarding how H4K12su may interact with other histone modifications that are a pre-requisite for histone methylation. One of these modifications is histone H2B Lys 120 ubiquitylation (H2BK120ub), which is associated with transcriptional elongation and enabling trans-tail signaling to histone H3 methylation at Lys 4 and Lys 79.[23] While we previously showed H4K12su inhibits H3K4me3 by the Set1/COMPASS complex, this was accomplished by using the limited six-subunit extended catalytic module (eCM) of COMPASS that is capable of nucleosome methylation in the absence of H2BK120ub.[24] While this system provided key insights

on the direct inhibition of H3K4 methylation by H4K12su, it did not fully capture how this crosstalk may functions in cells, where nucleosome methylation by the intact 8 protein COMPASS complex is dependent on the presence of H2BK120ub in chromatin.[24,25] This led us to ask if H4K12su may inhibit earlier regulatory steps, including the nucleosome-targeted activity of the Rad6-Bre1 ligase complex that is responsible for depositing H2BK120ub in chromatin.

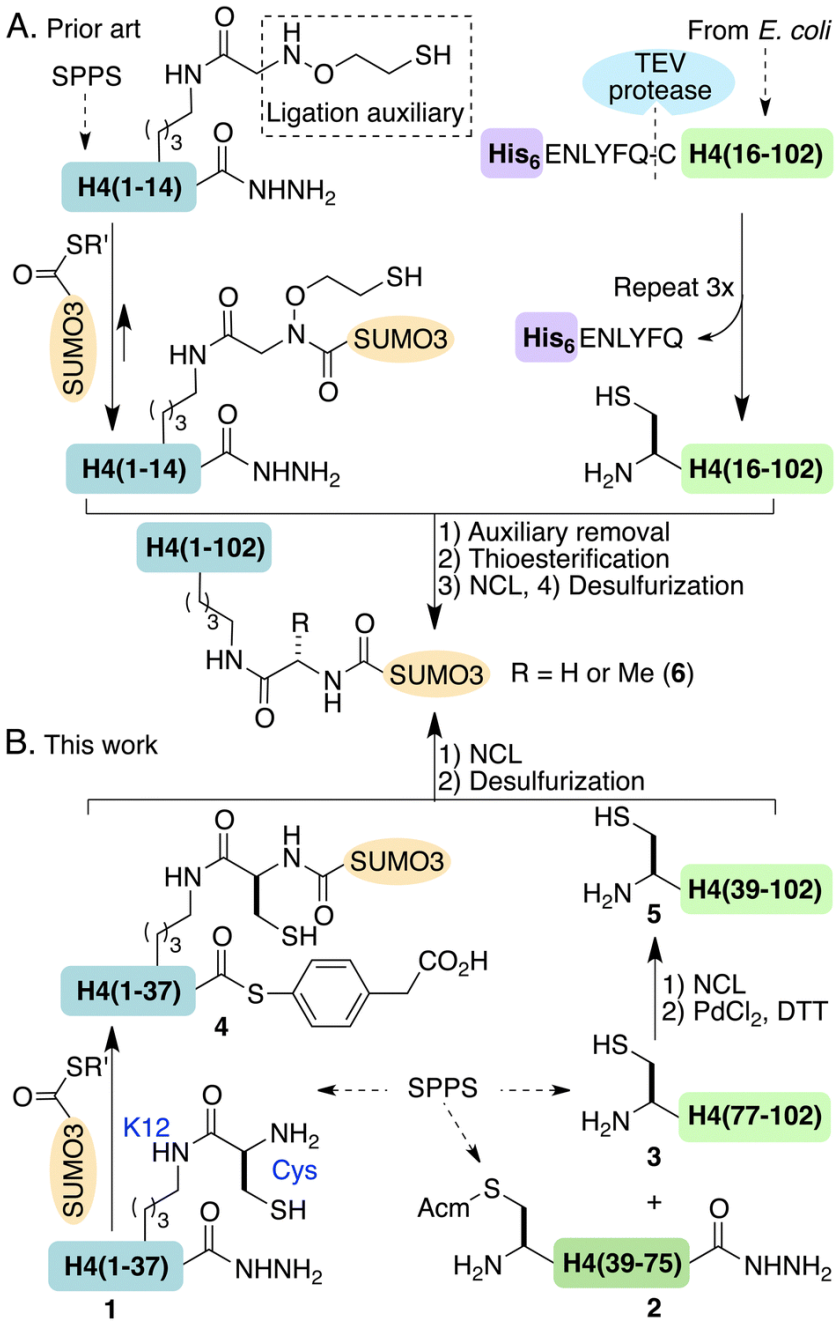
To explore the crosstalk between SUMO and ubiquitin in chromatin, we developed a total chemical synthesis strategy to generate full-length histone H4 in combination with a high-yielding Cys-mediated ligation to conjugate SUMO3 at Lys12 in H4. This novel strategy was designed to overcome key limitations of the earlier semisynthetic method, which restricted access to sufficient quantities of H4K12su for mechanistic studies. Ubiquitylation assays revealed that H4K12su inhibits Rad6-Bre1 activity on nucleosomes, which should, in principle, result in the reduction of H2BK120ub in cellular chromatin. To confirm that the *in vitro* results are true in the cellular context, linear SUMO-H4 fusion constructs were utilized as mimics of H4K12su in human cells. The expression of these constructs led to decreased amounts of H2BK120ub levels in nucleosomes, consistent with our *in vitro* findings. Thus, our results revealed that H4K12su has a negative biochemical relationship with H2BK120ub, suggesting that SUMO contributes to transcriptional repression by disrupting ubiquitylation, methylation and acetylation that are all necessary for efficient transcription elongation.

3.2 Results and discussion

3.2.1 A total chemical synthesis approach to site-specific H4K12 sumoylation and reconstitution into nucleosomes.

The Chatterjee lab had previously reported and utilized a semisynthetic strategy for generating homogenously site-specific sumoylated histone H4 (H4K12su). This was accomplished using an auxiliary-mediated strategy and led to the discovery of negative biochemical crosstalk with histone H3K4 methylation, via stimulation of histone demethylation[19] by LSD1 and by the inhibition of trimethylation by SET1/COMPASS.[21] Our studies also revealed the negative crosstalk of H4K12su with histone acetylation by perturbing acetylation of the N-terminal tail of histone H4 by p300.[21] While our semisynthetic strategy enabled these key discoveries, it remained limited by (1) the use of a ligation auxiliary that is thermodynamically reversible and precludes complete ligation (**Scheme 1A**), and (2) the inefficient cleavage of an N-terminal Ni^{2+}-affinity hexahistidine tag from a truncated form of histone H4, H4(D1-12), due to it's poor solubility, necessitating multiple rounds of tag cleavage and affinity purification that reduced overall yield (**Scheme 1A**).[21,26]

As a means of overcoming these limitations and to enable access to multi-milligram quantities of homogenously sumoylated histones for biochemical studies, we developed a total chemical synthesis strategy to generate H4K12su. To accomplish this, the full-length histone H4 protein was divided into three synthetically tractable fragments—H4(1–37), H4(38–75), and H4(76–102)—assembled via C-to-N-terminal native chemical ligation (**Scheme 1B**). Additionally, a protected Cys residue was coupled to K12 in H4(1–37) and the peptide was synthesized as a C-terminal hydrazide (**Figure S1**). Recombinant pure SUMO3(2–91), containing a Cys47-to-Ser47 substitution (SUMO3(2–91)C47S) and missing the terminal Gly92, was prepared as a C-terminal α-thioester using intein-mediated thiolysis (**Figure S2**). The thioester was ligated to Cys at the K12 site to yield the SUMO3(2–92)C47S,G92C-conjugated N-terminal H4(1–37) fragment. Subsequent hydrazide activation and thioesterification with 2-mercaptophenylacetic acid (MPAA) yielded compound **4** (**Scheme 1B and Figure S3**).



Scheme 1. Comparison of semisynthetic and total synthetic approaches to generate sumoylated H4. SPPS = solid-phase peptide synthesis, NCL = native chemical ligation, DTT = dithiothreitol, Acm = acetamidomethyl, TEV = tobacco etch virus, R′ = –CH_2–CH_2–SO_3Na.

The central fragment H4(38–75)A38C_{Acm} was synthesized as a C-terminal hydrazide with the N-terminal Cys protected with an acetamidomethyl (Acm) group (Figure S4) to enable site-selective

ligation with the C-terminal fragment H4(76–102)A76C, **3** (**Figure S5**). Fragment H4(76–102)A76C was ligated to the central fragment, followed by Pd(II)-mediated removal of the Acm protecting group to yield deprotected H4(38–102)A38C,A76C peptide **5** (**Figure S6 and S7**).[27] Finally, peptide **5** was ligated to the K12 sumoylated H4(1–37) α-thioester, **4**, to generate full-length H4 bearing site-specific SUMO3(G92C) at K12. Radical-mediated desulfurization yielded H4K12su(G92A), hereafter referred to as H4K12su (27% overall yield over two steps), **6** (**Figure 1A and B**).[28]

Each step of the synthesis was monitored by analytical HPLC and confirmed by ESI-MS. Successful conjugation was further validated by SDS-PAGE, which showed a distinct molecular weight shift indicative of SUMO attachment.

To validate the structural and functional integrity of the SUMO(G92A) modification, we assayed its enzymatic desumoylation by the catalytic domain of sentrin-specific protease 2 (SENP2). This demonstrated complete desumoylation of SUMO3(G92A) at a level comparable to wild-type SUMO3 (**Figure S8**). Circular dichroism spectra of SUMO3(2–92)C47S,G92A and H4K12su confirmed their properly folded states in solution (**Figure S9 and S10**). Finally, the synthetic H4K12su was successfully incorporated into histone octamers with histones H2A, H2B, and H3(C110A) via salt dialysis (**Figure 1C**), and reconstituted into mononucleosomes by combining with the high-affinity 147 bp Widom 601 DNA sequence (**Figure 1D**).[19] This demonstrates compatibility of this chemically synthesized H4K12su with chromatin assembly and enables further downstream biochemical assays.

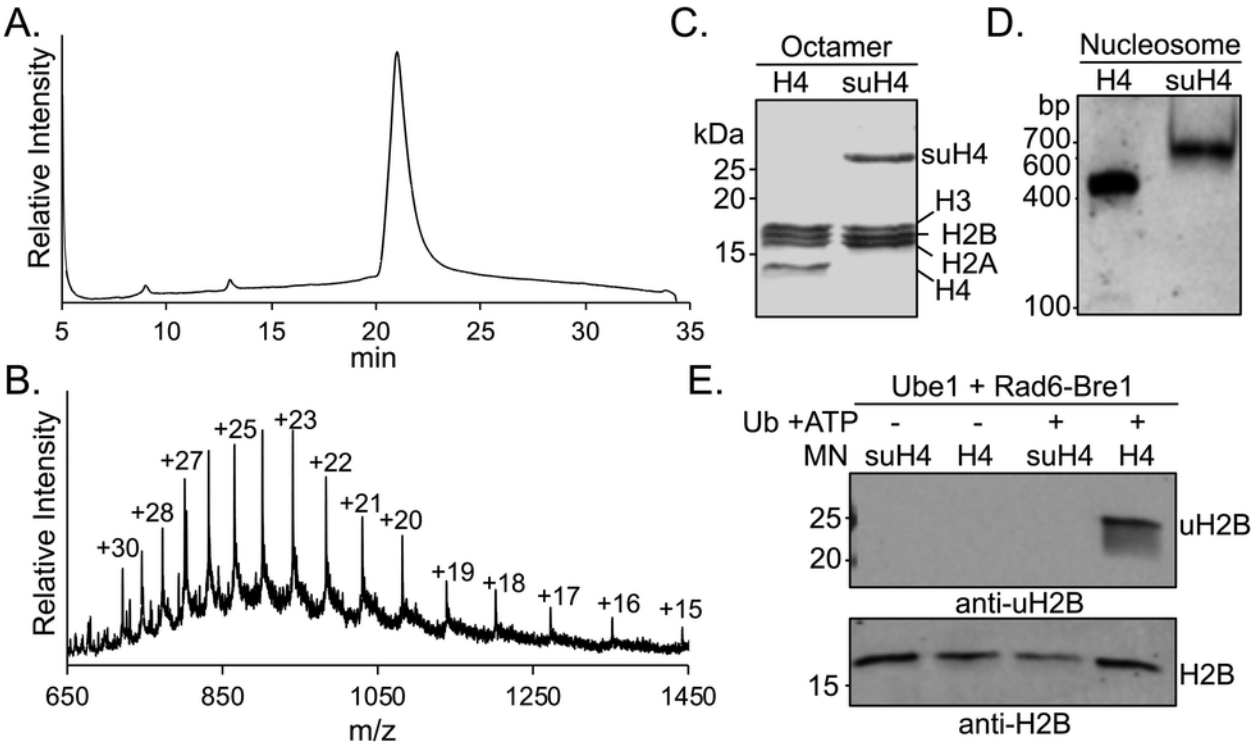


Figure 1. Total chemical synthesis of sumoylated H4 and its biochemical crosstalk with ubiquitylated H2B in vitro. (A) C18 analytical RP-HPLC chromatogram of purified synthetic H4K12su. (B) ESI-MS spectrum of purified synthetic H4K12su. Calculated 21,609.8 Da and observed 21,610.2 ± 1.3 Da. (C) Coomassie-stained 15% SDS-PAGE of octamers assembled with either H4 or H4K12su (suH4). (D) 5% TBE gel of mononucleosomes assembled from wild-type or chemically sumoylated octamers. (E) Rad6–Bre1-mediated nucleosomal ubiquitylation assay. Assay products were probed with an H2BK120ub-specific antibody to detect ubiquitylation and with an H2B-specific antibody as a loading control.

3.2.2 H4K12 sumoylation inhibits Rad6-Bre1-mediated H2B ubiquitylation *in vitro*

The Chatterjee lab has previously investigated the crosstalk between H4K12su, methylation and acetylation of histone tails in nucleosomes. One such investigation was looking at crosstalk between H4K12su and H3K4me3. We found that H4K12su inhibited the installation of H3K4me3 by the 6-protein extended catalytic module of the Set1/COMPASS complex which exhibits Ub-independent methylation activity on modified nucleosomes. This is unlike the eight-subunit intact COMPASS complex that strictly requires the presence of H2BK120ub for activity.[21,25] However,

the biochemical crosstalk of H4K12su with H2BK120ub, the pre-requisite mark for methylation, remained unknown.

In order to address crosstalk with ubiquitylation, we performed *in vitro* ubiquitylation assays using the E2 and E3 ligases, Rad6 and Bre1, respectively, with reconstituted mononucleosomes containing either wild-type (wt) H4 or H4K12su. The Rad6-Bre1 complex was recombinantly expressed in Sf9 insect cells and purified using affinity purification and ion-exchange chromatography by Xiaowen Xie in the Zheng lab (**Figure S11**).[29] Ubiquitylation assays were supplemented with ATP to enable the E1 activation step mediated by Ube1. From these assays, we observed that nucleosomes bearing wt H4 were efficiently ubiquitylated at H2BK120 as seen via western blot analysis with a commercial H2BK120ub-specific antibody. In stark contrast, nucleosomes containing H4K12su showed a dramatic decrease in the level of H2BK120ub, clearly indicating that H4K12su inhibits ubiquitylation.

To probe the mechanism underlying the negative crosstalk, we performed an electrophoretic mobility-shift assay (EMSA) with Rad6-Bre1 and wt or sumoylated mononucleosomes under native conditions (**Figure 2**). Incubation of wt mononucleosomes with Rad6-Bre1 led to a concentration-dependent shift in their native gel migration, consistent with complex formation. However, mononucleosomes containing H4K12su failed to exhibit any gel-shifts across the same concentration range (**Figure 2**). This revealed that the inhibition of H2B ubiquitylation by H4K12su was due to the inability of Rad6-Bre1 to bind sumoylated nucleosomes. Thus, our *in vitro* assay results support a model in which H4K12su acts as a negative regulator of Rad6-Bre1-mediated H2BK120ub by preventing Rad6-Bre1 complex binding to nucleosomes. As H2BK120ub is a necessary modification for stimulating H3K4me3 by the COMPASS complex,[23] the reduction of

ubiquitylation by H4K12su offers an additional mechanism for its ability to inhbit the installation of H3K4me3.

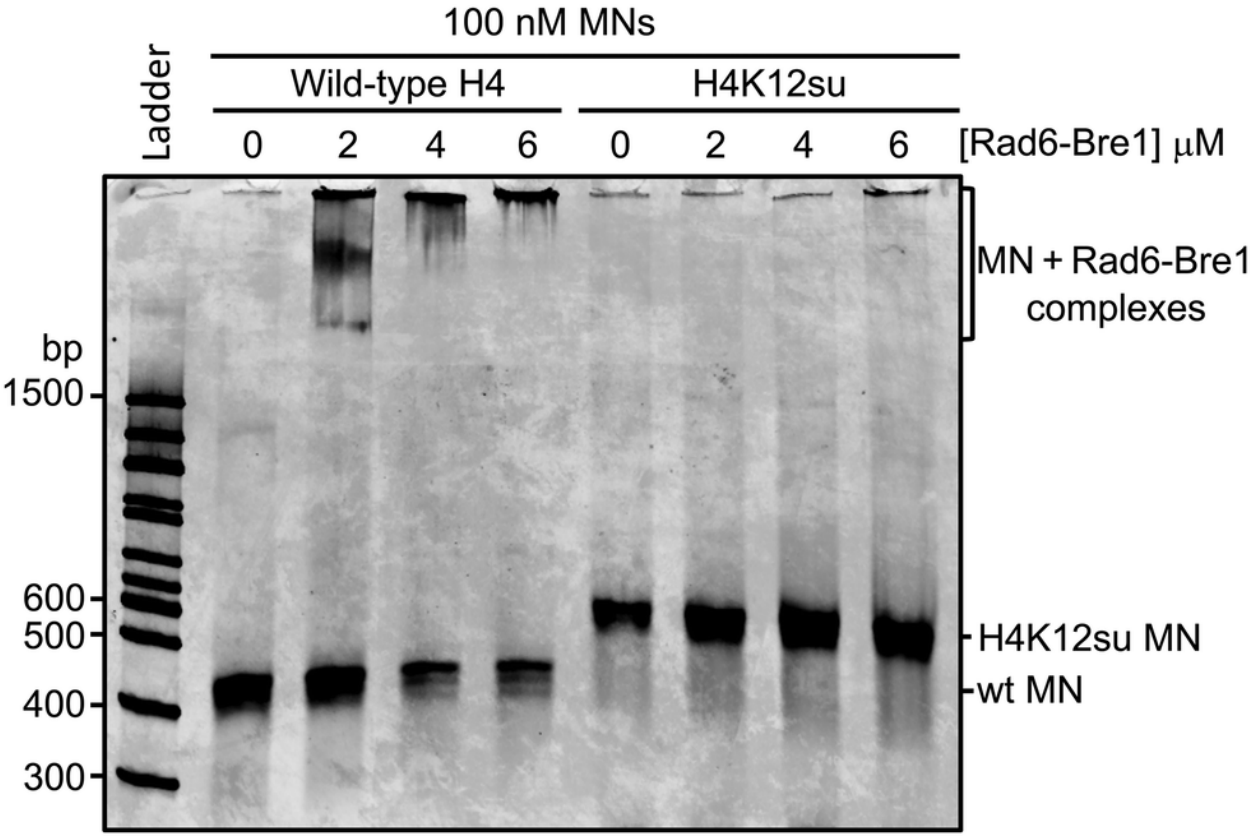


Figure 2. Mononucleosome binding by the Rad6–Bre1 complex. Wild-type and SUMOylated mononucleosomes were incubated with increasing amounts of the Rad6–Bre1 complex and nucleosome-binding was visualized by supershift of the mononucleosome band in 5% TBE gels stained with SYBR safe DNA-stain. MN = mononucleosome, wt = wild-type.

3.2.3 SUMO-H4 mimic expression leads to decreased H2BK120 ubiquitylation *in vivo*

While our *in vitro* assays provided a clear mechanism for the negative crosstalk between H4K12su and H2BK120ub through the physical inability of Rad6-Bre1 to bind to sumoylated nucleosomes, it remained unknown if this negative biochemical crosstalk is also conserved in cells. To address this question, we adopted a genetic fusion strategy to mimic the N-terminal sumoylation of histone H4 in cells. We designed a construct that encoded the protease-stable HA-tagged SUMO3(ΔGG)

protein fused to an N-terminally truncated histone H4 to generate HA-SUMO3(ΔGG)-H4(Δ1-11), referred to here as SU-H4. This genetic fusion places SUMO3, in a non-removable form due to the absence of the di-Gly motif at its C-terminus that is required for SENP activity, in close proximity to the K12 side-chain of H4. This ensured nuclear localization of the fusion and we found it to be compatible with downstream chromatin incorporation and detection with anti-HA antibodies. An immortalized human embryonic kidney cell line, HEK293T, was transiently transfected with the plasmid encoding for SU-H4 under a CMV promoter, and cells were harvested 48 h post-transfection. Expression of the protein was confirmed via western blot analysis using an anti-HA specific antibody.

Following harvesting, cells were lysed and nuclei were isolated by centrifugation. Nuclear chromatin was subjected to limited *Micrococcal nuclease* (MNase) digestion to generate mononucleosomes. Analysis of the digested chromatin by non-denaturing agarose gel electrophoresis showed a low molecular weight ~150 bp (**Figure 3a**). The enrichment of mononucleosomes ensures that H2BK120ub blots report on their occurrence in single nucleosomes that also contain one or two copies of SU-H4; rather than in polynucleosome fragments that may or not contain SU-H4 in every mononucleosome. Mononucleosomes containing SU-H4 were selectively immunoprecipitated using magnetic anti-HA-conjugated affinity beads under native conditions to preserve intact mononucleosomes (Figure 3b). The efficiency of these pulldowns was further validated by additionally immunoblotting for the co-purifying core histone H3, which also verified the successful incorporation of SU-H4 into chromatin.

Next, to assess the effect of SU-H4 on H2BK120 ubiquitylation levels, western blotting analysis was performed on both the total chromatin used as input and the immunoprecipitated SU-H4-containing nucleosomes. A H2BK120ub-specific antibody was employed to evaluate the levels of

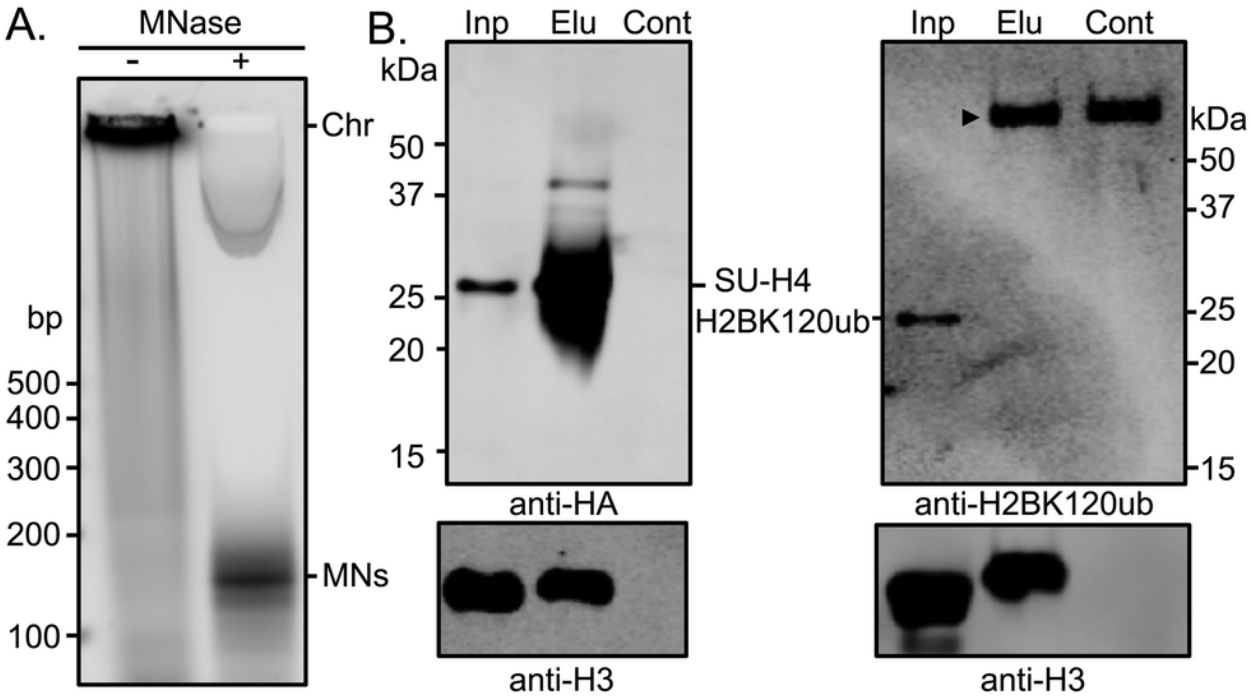


Figure 3. Native chromatin immunoprecipitation of HA-SUMO3(ΔGG)-H4(Δ1–11) from HEK293T cells. (A) Micrococcal nuclease (MNase) digestion of cellular chromatin (Chr) to generate mononucleosomes (MNs). (B) HA-tag-mediated immunoprecipitation of mononucleosomes containing SU-H4. Inp = input nucleosomes, Elu = eluate from anti-HA resin, Cont = resin control, SU-H4 = HA-SUMO3(ΔGG)-H4(Δ1–11). An H3-specific antibody was used as a loading control for the total amount of MNs in the input and elution lanes. A filled triangle indicates antibody-chains released from the resin under denaturing conditions.

ubiquitylation in immunoprecipitated mononucleosomes. Predictably, the total chromatin input displayed robust H2BK120ub signal as it reflects the total amount of ubiquitylated histones in cells. In contrast, SU-H4-containing nucleosomes that were immunoprecipitated from total cellular chromatin exhibited no detectable H2BK120ub across multiple biological replicates (**Figure 3b**). This drastic loss of H2BK120ub within sumoylated nucleosomes suggests that the inhibitory role of H4K12su toward H2BK120ub observed *in vitro* is also conserved in the chromatinized, cellular context.

3.2.4 Control constructs reveal specificity of SUMO-H4-mediated crosstalk

With the exciting confirmatory results in hand, we wanted to further evaluate the precise SUMO-H4 construct employed in our cellular experiments. To exclude the possibility that the observed inhibition of H2BK120ub may be due to the truncation of the H4 N-terminal resiudes 1-11, or indeed other factors that are associated with the first 11-residues in H4, rather than the proximity of SUMO3 to K12 in H4, we performed a series of control experiments.

First, to evaluate the possibility that truncation of the H4 tail itself—independent of SUMO3—may be sufficient to perturb the occurrence of H2BK120ub in mononucleosome, we designed a construct encoding H4(Δ1–11) with an N-terminally fused double Myc-tag to make Myc_2-H4(Δ1–11). This corresponded to the same N-terminal tail truncation of H4 found in SU-H4 but lacked the SUMO3 fusion. HEK293T cells were transfected with this construct followed by chromatin isolation and mononucleosome generation by Mnase digestion. Chromatin immunoprecipitation was performed using magnetic anti-Myc-affinity resin and levels of H2BK120 ubiquitylation were evaluated via western blot, comparing total cellular H2BK120ub content to its presence in immunoprecipitated nucleosomes. Gratifyingly, immunoprecipitated mononucleosomes retained robust H2BK120ub signal (**Figure S12**), which confirmed that N-terminal tail truncation of H4 is not sufficient to completely inhibit H2B monoubiquitylation. Thus, the observed negative crosstalk between SU-H4 and H2BK120ub was not an artifact of the H4(Δ1–11) truncation in the fusion construct.

We next examined whether SUMO3 fused to the N-terminus of the full-length H4 may still inhibit the installation of H2BK120ub in cells. Hence, we generated a plasmid encoding a HA-tagged non-cleavable SUMO3(DGG) fused to the N-terminus of full-length H4 (HA-SUMO3(ΔGG)-H4). This placed SUMO3 11-residues away from K12 and tested if the precise site of sumoylation is important for its function in chromatin. Following transfection and chromatin isolation, mononucleosomes were immunoprecipitated with magnetic anti-HA-affinity resin and the levels of H2BK120ub were evaluated via western blot. Similar to our observations with HA-

SUMO3(ΔGG)-H4(Δ1-11), we also observed the strong reduction of H2BK120ub levels in the HA-SUMO3(ΔGG)-H4 nucleosomes (**Figure S13**).

Collectively, our control experiments demonstrated that the presence of SUMO3 on H4, regardless of whether it is fused to a truncated or full-length H4 protein, was sufficient to disrupt H2BK120ub levels in nucleosomes that contained at least one copy of SUMO. This indicated that the negative crosstalk we observed *in vitro* and in cells arises specifically from the attachment of SUMO3 to the H4 tail, although the precise site of sumoylation is less critical as long as it is near K12. These findings strongly support our model where H4K12su directly interferes with the machinery necessary to install ubiquitin at H2B Lys 120. This negative crosstalk of H4K12su with H2BK120ub, a key modification associated with actively transcribed genes in cells, is consistent with the proposed role of H4 tail sumoylation in gene repression.[17,21]

3.3 Conclusion and outlook

In this study, we developed and streamlined a facile total synthesis strategy to generate histone H4 modified by SUMO3 at lysine 12. This new approach avoided previous limitations such as the necessary TEV-protease-mediated release of H4 from affinity tags used for purification and the use of a thermodynamically reversible 2-aminoxyethanethiol ligation auxiliary with a high-yielding cysteine-mediated native chemical ligation reaction. The novel synthetic H4K12su was validated in biochemical assays by its recognition and cleavage, by the catalytic domain of the SUMO protease SENP2, in addition to its successful incorporation into *in vitro* reconstituted octamers and nucleosomes. These initial successes allowed us to directly probe the functional consequences of H4K12su in a homogenous sumoylated chromatin context.

In both the reconstituted biochemical system and in HEK293T cells, we observed a striking negative crosstalk between H4K12su and ubiquitylation at H2BK120, a modification commonly associated with transcriptional elongation. Our findings reveal a previously unknown mechanism of biochemical crosstalk in which H4K12su suppresses H2BK120ub.

As our initial cellular studies employed a genetic SUMO3-H4 fusion that included a truncated H4 tail, we confirmed that the observed effects were not artifacts of histone tail truncation by demonstrating that truncating H4 alone did not lead to significant changes in total cellular H2BK120ub levels. Additionally, we used a control that fused SUMO3 to full-length H4 to confirm that the inhibition was specifically due to SUMO present in the H4 tail.

Importantly, our findings addressed an open mechanistic question raised by a prior study from our lab (Leonen et al. 2021), which documented the negative crosstalk between H4 sumoylation and H3K4me3 with an enzyme complex that did not require H2BK120ub for its methylation activity. The reagents developed here, including synthetic SUMO3-H4, control constructs, and reconstituted nucleosomes, offer an approachable platform for interrogating SUMO-dependent crosstalk within chromatin.

Looking to the future, the strategies employed herein will enable future investigations of crosstalk involving multiple modifications within H4 that are challenging to address without the total synthetic approach. Future studies in our labs will explore how H4K12su dictates the state of neighboring modifications such as H4K16ac, which is associated with chromatin decompaction and transcriptional activation. The preparation of homogenous synthetic histone H4 bearing both H4K12su and H4K16ac will permit interrogating if these marks function in opposition to each other and if one dominates overall gene function, due to their opposing repressive and activating roles, respectively. Thus, our work has and will continue to contribute to the growing understanding that chromatin modification by SUMO is a versatile modulator of chromatin structure and function. By

establishing a robust route to access site-specifically sumoylated histone and revealing an inhibitory link between H4K12su and H2BK120ub, this study provides a foundation for dissecting SUMO's roles across transcriptional control, DNA repair, and other critical cellular processes.

3.4 Experimental procedures

3.4.1 General Methods

2-Chlorotrityl resin (1.5 mmol/g substitution) was purchased from AnaSpec (Fremont, CA). Fmoc-Gly-Wang resin (0.32 mmol/g substitution) was purchased from EMD Millipore (Gibbstown, NJ). Fmoc-L-amino acids were purchased from Chem-Impex Int'l. (Wood Dale, CA), AGTC Bioproducts (Wilmington, MA), Creosalus (Louisville, KY), and Sigma-Aldrich (St. Louis, MO). Other chemical reagents were purchased from either Fisher Scientific (Pittsburgh, PA) or Sigma-Aldrich Chemical Company (St. Louis, MO). Plasmid mini-prep, PCR purification and gel extraction kits were purchased from QIAGEN (Valencia, CA). Reversed-phase HPLC (RP-HPLC) was performed on either a Varian (Palo Alto, CA) ProStar HPLC or an Agilent (Santa Clara, CA) 1260 Infinity II HPLC using an Agilent Poroshell 120 SB C18 analytical column (2.7 micron, 100 x 4.6 mm) at a flow rate of 1 mL/min, a Grace-Vydac (Deerfield, IL) semi-preparative C18, C4 column (5 micron, 150 x 4.6 mm) at a flow rate of 3.5 mL/min, or a Grace-Vydac preparative C18 column (10 micron, 250 x 22 mm) at a flow rate of 9 mL/min. Mobile phases were Buffer A containing 0.1% trifluoroacetic acid (TFA) in water, and Buffer B containing 90% acetonitrile and 0.1% TFA in water. Solid phase peptide synthesis was performed on a Liberty Blue Automated Microwave Peptide Synthesizer (CEM Corporation, Matthews, NC). Mass spectrometric analysis was conducted on a Bruker (Billerica, MA) Esquire, or a Thermo Scientific (Waltham, MA) LTQ Orbitrap ESI-MS.

3.4.2 Synthesis of H4(1-37)K12$_{Cys}$ (1)

H_2N-SGRGKGGKGLGK(Cys)GGAKRHRKVLRDNIQGITKPAIRRL-C(O)$NHNH_2$

The 2-chlorotrityl hydrazide resin for peptide synthesis was prepared by reacting 2-chlorotrityl chloride resin (1.5 mmol/g, 0.3 mmol) in a 10% (v/v) solution of hydrazine in DMF at 30 °C for 30 min.[30] The reaction was repeated once with fresh 10% (v/v) hydrazine solution. The resin was then treated with 10% (v/v) methanol in DMF for 10 min to cap any unreacted sites on the resin. In order reduce the loading capacity of resin the first amino acid, Leu, was coupled as mixture of Fmoc-Leu-OH and Boc-Leu-OH in a 1:2 molar ratio, respectively. The coupling reaction with Fmoc-Leu-OH (0.4 mmol), Boc-Leu-OH (0.8 mmol) O-(6-Chlorobenzotriazol-1-yl)-N,N,N′,N′-tetramethyluronium hexafluorophosphate (HCTU, 1.1 mmol), and DIEA (2.4 mmol) proceeded for 60 min at 30 °C. Any unreacted hydrazine sites were then capped by treating the resin with acetic anhydride (0.6 mmol), and DIEA (0.6 mmol) for 30 min at 25 °C and the reduced Fmoc-amino acid loading of the resin was confirmed by suspending the Fmoc-protected resin in 20% (v/v) piperidine in DMF to generate the dibenzofulvene-piperidine adduct that was quantified by absorbance at 304 nm (E. Next, each sequential amino acid was coupled in a 5-fold molar excess based on the reduced resin loading using a CEM microwave-assisted peptide synthesizer. Deprotection of the Fmoc- group was achieved by treating the resin with 5% (w/v) piperazine with 0.1 M HOBt in DMF for 3 min at 75 °C. Coupling reactions were undertaken for 10 min at 75 °C with a mixture of Fmoc-amino acid (0.5 mmol), DIC (0.49 mmol), Oxyma (0.49) and DIEA (0.98 mmol) in DMF. For Fmoc-Arg(Pbf)-OH, an additional coupling reaction was performed for 10 min at 75 °C. The Lys at position 12 was orthogonally protected with the 4,4-dimethyl-2,6-dioxocyclohex-1-ylidene (ivDde) protecting group. The N-terminal amino acid, Ser was coupled as Boc-Ser-(O^{t}Bu)-OH. After the completion of linear amino acid couplings, the orthogonal deprotection of ivDde protecting group at K12 was performed by treating the resin-bound peptide with a solution of 5% (v/v) hydrazine in DMF for 5 min. The deprotection was repeated twice to

ensure completion. Finally, the side-chain deprotected peptidyl resin was coupled with Boc-Cys(Trt)-OH (0.5 mmol) with DIC (0.49 mmol), Oxyma (0.49) and DIEA (0.98 mmol) in DMF to attach Cys at K12 as a ligation handle for SUMO3 α-thioester ligation.

Peptide cleavage from the resin. The final resin was successively washed with DMF and DCM, then dried under vacuum. A cleavage cocktail consisting of TFA, triisopropylsilane and water, 95:2.5:2.5 (v/v), was added to the resin and the mixture shaken for 2 h at 25 °C. The resin was filtered, and the combined filtrate was added dropwise to a 10-fold volume of ice-cold diethyl ether and the resulting precipitate was centrifuged at 3000 rpm for 5 min at 25 °C. The clear solution was decanted and the precipitated crude peptide was dissolved in 50% (v/v) AcN in H2O, filtered through a 0.45 µm filter and lyophilized to a dry powder. The dry peptide was dissolved in RP-HPLC Buffer A and then purified by C18 preparative RP-HPLC with a gradient of 0-73% B over 30 min. The isolated yield after lyophilization was 12% based on the initial resin loading. Calculated for 1, M_{avg} 4,055.2 Da and observed 4,054.8 ± 0.6 Da.

3.4.3 Synthesis of H4(38-75)A38C$_{Acm}$ (2)

H_2N-C(Acm)RRGGVKRISGLIYEETRGVLKVFLENVIRDAVTYTEH-C(O)$NHNH_2$

The synthesis of the middle fragment of histone H4 on hydrazide resin followed the same protocol as described for fragment **1**. One key addition was the coupling of Cys(Acm) at the N-terminus using Fmoc-Cys(Acm)-OH (0.5 mmol) with DIC (0.49 mmol), Oxyma (0.49) and DIEA (0.98 mmol) in DMF. The isolated yield of purified peptide was 8% based on the initial resin loading. Calculated for **2**, M_{avg} 4,465.2 Da and observed 4,465.8 ± 0.6 Da.

3.4.4 Synthesis of H4(76-102)A76C (3)

H_2N-CKRKTVTAMDVVYALKRQGRTLYGFGG-CO_2H

Fmoc-Gly-Wang resin (0.32 mmol/gm) was allowed to swell for 10 min in excess DMF. Following resin deprotection with 20% piperidine in DMF (v/v), each amino acid was coupled in 5-fold molar excess based on resin loading using the CEM microwave peptide synthesizer. Deprotection of the Fmoc- group was achieved by treating resin with 5% (w/v) piperazine with 0.1 M HOBt in DMF for 3 min at 75 °C. Coupling reactions were undertaken for 10 min at 75 °C with a mixture of Fmoc-amino acid (0.5 mmol), DIC (0.49 mmol), Oxyma (0.49) and DIEA (0.98 mmol) in DMF. Cleavage and purification protocols were as mentioned for peptide 1. The isolated yield of purified peptide was 14%. Calculated for 3, M_{avg} 3,019.5 Da and observed 3,019.8 ± 0.5 Da.

3.4.5 Overexpression and purification of SUMO3(2-91)C47S-MESNa

E. coli BL21(DE3) cells were transformed with the plasmid pTXB1-SUMO3(2-91)C47S.[2] Cells were grown in 6 L Luria-Bertani medium supplemented with 100 µg/mL of Ampicillin at 37 °C with shaking at 250 rpm until OD_{600} ~0.6-0.8. Overexpression of the desired fusion protein was induced by the addition of 0.3 mM IPTG, and cells were grown for an additional 4 h at 25 ˚C with constant shaking. The cells were then harvested by centrifugation at 7,000xg for 15 min. The cell pellet was resuspended in a lysis buffer consisting of PBS, pH 7.2, containing 1 mM 2-Mercaptoethanesulfonic acid sodium salt (MESNa). Cells were lysed by sonication then centrifuged at 20,000xg for 15 min. The lysate supernatant was passed through a 0.45 µm filter then applied to a 30 mL chitin column pre-equilibrated with lysis buffer. Proteins were bound to the column over a period of 12 h at 4 °C. The column was then washed with 20 column volumes (CV) of lysis buffer followed by 2 CV of PBS at pH 7.75. The SUMO3(2-91)C47S-MESNa α-thioester was generated from its intein-chitin-binding domain (CBD) fusion by incubation with 1.5 CV of PBS, pH 7.75, containing 100 mM MESNa for 72 h at 4 °C. The eluted α-thioester was purified by C18 preparative RP-HPLC employing a gradient of 0-60% B over 60 min. Fractions

containing the desired thioester were identified by ESI-MS. Calculated M_{avg} 10,444.7 Da and observed 10,443.8 ± 1.5 Da. The N-terminal Met of SUMO3 is consistently processed in E. coli by the methionyl aminopeptidase enzyme, leading to the SUMO3(2-91)C47S α-thioester product. Typical yields of the α-thioester were 4-5 mg/L of BL21(DE3) culture.

3.4.6 Synthesis of SUMO3(2-92)C47S,G92A for Circular Dichroism studies

The purified SUMO3(2-91)C47S-MESNa was dissolved in a buffer containing 6 M Gn•HCl, 200 mM NaPi, 100 mM L-Cysteine and 30 mM TCEP at pH 7.0. The reaction mixture was incubated at 25 °C for 4-6 h to allow L-Cys ligation. Reaction progress was followed by analytical C18 RP-HPLC with a gradient of 0-60% B over 30 min. After the reaction was considered to be complete, the reaction mixture was treated with 250 mM TCEP, 30 equivalents of VA-044 (0.08 mmol), and 10% (v/v) ^{t}BuSH and then incubated at 42 °C for 6 h in order to desulfurize the C-terminal Cys92 to Ala. The desulfurized product was purified by C18 analytical RP-HPLC with a gradient 0-60 % B over 30 min. Calculated M_{avg} for SUMO3(2-92)C47S,G92A, 10,373.8 Da and observed 10,374.5 ± 1.4 Da.

3.4.7 Circular Dichroism studies of SUMO3(2-91)C47S,G92A; SUMO3(2-91)C47S-MESNa and H4K12su

Circular Dichroism (CD) spectra were recorded on a JASCO J-720 spectropolarimeter fitted with a Peltier temperature controller. Each pure protein was dissolved at a final concentration of 15 μM in 10 mM NaPi, 100 mM NaF, pH 7.4 at 20 °C in 1 mm quartz cuvettes and measurements were undertaken at a scanning speed of 100 nm/min over a range of 190-270 nm.

3.4.8 Expressed protein ligation of H4(1-37)K12$_{Cys}$ (1) with SUMO3(2-91)C47S-MESNa to generate the sumoylated H4 tail peptide (4)

Purified peptide H4(1-37)K12$_{Cys}$-hydrazide (13 mg, 0.0032 mmol, 2 mM) was dissolved in ligation buffer containing 6 M Gn•HCl, 200 mM NaPi, 100 mM MPAA and 35 mM TCEP at pH 7.2. The dissolved peptide solution was added to SUMO3(2-91)C47S-MESNa (28.2 mg, 0.0027 mmol, 1.6 mM), and the buffer pH was re-adjusted to 7.2 with dilute aqueous NaOH. The reaction mixture was incubated at 37 °C for 6-8 h to allow expressed protein ligation to proceed. Reaction progress was followed by analytical C18 RP-HPLC with a gradient of 0-73% B over 30 min. After the ligation was considered to be complete, the reaction mixture was dialyzed against 500 mL of 6 M Gn•HCl, 200 mM NaPi buffer, pH 7.2, overnight to remove unreacted MPAA and TCEP. The dialyzed solution was treated with 10 equivalents of 20 mM aqueous $NaNO_2$ at -15 °C at pH ~3.0 for 20 min to undertake oxidation of the acyl hydrazide to an acyl azide intermediate. Subsequently, 50 equivalents of MPAA (100 mM) were added to the peptidyl azide in situ and the pH adjusted to 6.0. The displacement of azide by MPAA was followed by C18 RP-HPLC with a gradient of 20-60 %B over 30 min. The same gradient was used to purify the MPAA α-thioester product by C18 semipreparative RP-HPLC with an isolated yield of 40% over three steps. Calculated M_{avg} for MPAA α-thioester form of **4**, 14,359.4 Da and observed 14,359.9 ± 0.7 Da.

3.4.9 Native chemical ligation of H4(38-75)A38C$_{Acm}$ (2) and H4(76-102)A76C (3) to yield H4(38-102)A38C$_{Acm}$,A76C

The peptide H4(38-75)A38CAcm (2) (14.5 mg, 0.003 mmol, 2 mM) was dissolved in a buffer consisting of 6 M Gn•HCl, and 200 mM NaPi at pH 3.0 and then treated with 10 equivalents of 20 mM aqueous $NaNO_2$ at -15 ºC for 20 min. Subsequently, 50 equivalents of MPAA (100 mM) were added to the peptidyl azide in situ and the pH adjusted to 6.0. The displacement of azide by MPAA

was followed by C4 RP-HPLC with a gradient of 20-80 %B over 30 min. To the reaction mixture of H4(38-75)A38C$_{Acm}$ MPAA α-thioester a solution of H4(76-102)A76C (3) (11.6 mg, 0.0024 mmol) in 6 M Gn•HCl, 200 mM NaPi, and 35 mM TCEP at pH 7.2 was added and the pH of the ligation mixture was re-adjusted to 7.2 with dilute aqueous NaOH. The ligation mixture was incubated at 37 ºC for 6 h and the reaction was followed by C4 analytical RP-HPLC with a gradient 20-80 % B over 30 min. After the ligation was considered complete, the product was purified using C4 preparative RP-HPLC with a gradient of 20-80% B over 60 min. The isolated yield of the ligation product was 35%. Calculated M_{avg} for **5** was 7,451.9 Da and observed 7,450.8 ± 1.1 Da.

3.4.10 Palladium-mediated Acetamidomethyl (Acm) protecting group removal from H4(38-102)A38C$_{Acm}$, A76C (5)

The ligation product H4(38-102)A38CAcm,A76C (8.5 mg, 0.0011 mmol) with N-terminal Acm-protected Cys was dissolved in ligation buffer containing 6 M Gn•HCl, 200 mM NaPi, 100 mM MPAA and 35 mM TCEP at pH 7.2, and incubated with 15 equivalents of $PdCl_2$ for 1 h at 37 °C.[27] The reaction mixture was quenched with 100 mM DTT and centrifuged at 13,000 rpm for 5 min to precipitate the Pd-complex, and the supernatant containing deprotected peptide 5 was purified by C4 preparative RP-HPLC with a gradient of 30-70% B over 60 min. This led to an isolated yield of 85%. Calculated M_{avg} for the Acm-deprotected form of **5**, 7,381.6 Da and observed 7,380.6 ± 1.2 Da.

3.4.11 Native chemical ligation of the sumoylated H4 tail peptide (4) with H4(38-102)A38C,A76C (5) to generate H4(1-102)K12SUMO3(G92C),A38C,A76C and desulfurization to obtain H4K12su (6).

To a solution of the MPAA α-thioester form of peptide **4** (15.6 mg, 0.001 mmol, 2 mM) dissolved in ligation buffer containing 6 M Gn•HCl, 200 mM NaPi, 80 mM MPAA and 40 mM TECP at pH 7.2, the Acm-deprotected H4(38-102)A38C,A76C peptide **5** (7.2 mg, 0.0009 mmol) was added. The pH was re-adjusted to 7.2 and ligation allowed to proceed at 37 °C for 14 h. Reaction progress was followed by C4 analytical RP-HPLC with a gradient of 30-70% B over 30 min. After the ligation was considered complete, the reaction mixture was dialyzed against 500 mL of 6 M Gn•HCl, 200 mM NaPi buffer, pH 7.2, in a 3.5 kDa cut-off dialysis cassette in order to remove excess MPAA. The dialyzed reaction mixture was then treated with 250 mM TCEP, 90 equivalents of VA-044 (0.08 mmol), and 10% (v/v) ^{t}BuSH and then incubated at 42 °C for 6 h in order to desulfurize Cys38 and Cys76 and convert these to the native Ala found in wild-type H4. The reaction was analyzed by C4 analytical RP-HPLC with a gradient of 30-70% B over 30 min and purified by C4 semi-preparative RP-HPLC over a gradient of 30-70%B over 45 min. The isolated yield of pure H4K12SUMO3(C47S,G92A), or H4K12su, was 27%. Calculated M_{avg} for H4K12SUMO3(C47S,G92A), or H4K12su, 21,609.8 Da and observed 21,610.2 ± 1.3 Da. For simplicity, hereon H4K12SUMO3(C47S,G92A) is referred to as H4K12su in the text below.

3.4.12 Octamer assembly

Individual histones H2A, H2B and H3 were dissolved at final concentrations of 4 mg/mL in an unfolding buffer consisting of 7 M Gn•HCl and 20 mM Tris, pH 7.5. Equimolar amounts of the three histones were combined with either wild-type H4 or synthetic H4K12su and the resulting mixture dialyzed into a refolding buffer consisting of 10 mM Tris, 2 M NaCl, 1 mM EDTA at pH 7.5. The self-assembled crude octamers were concentrated with a 10,000 MWCO Vivaspin concentrator and then purified by size-exclusion chromatography using a Superdex S-200 column attached to an AKTA FPLC pump. Fractions containing the desired octameric species were

identified by 18% SDS-PAGE followed by coomassie staining, combined, and concentrated with a Vivaspin 500 concentrator prior to nucleosome assembly.

3.4.13 Mononucleosome (MN) assembly

Octamers and 147 bp of Widom 601 DNA were combined in 10 µL of the high-salt octamer refolding buffer consisting of 10 mM Tris, 2 M NaCl, 1 mM EDTA at pH 7.5 to a final concentration of 2 µM. After incubation at 37 °C for 15 min, 3.3 µL of dilution Buffer 1 (10 mM HEPES, 1 mM EDTA, 0.5 mM PMSF, pH 7.9) was added and the incubation temperature was reduced to 30 °C. Further additions of 6.7, 5.0, 3.6, 4.7, 6.7, 10, 30, and 20 µL of Buffer 1 were undertaken after every 15 min. A final addition of 100 µL of dilution Buffer 2 (10 mM Tris, 1 mM EDTA, 0.5 mM PMSF, 20% (v/v) Glycerol, pH 7.5) was undertaken and after an additional 15 min at 30 °C, the MNs were concentrated by Vivaspin 500 concentrator, and their composition verified by 5% Tris-Borate-EDTA (TBE) non-denaturing polyacrylamide gel electrophoresis followed by staining with SYBR Safe DNA stain.

3.4.14 Desumoylation of synthetic H4K12su by Sentrin-specific protease 2 (SENP2)

The recombinant human His_6-SENP2 (Sentrin-specific protease 2) catalytic domain cloned in pET28a (Addgene Plasmid 16357) was expressed in *E. coli* BL21(DE3) cells and purified using Ni-affinity chromatography using 20 mM Tris and 500 mM Imidazole buffer, pH 7.5, as the eluent.[31] The H4K12su octamer (1 µM) was incubated with His_6-SENP2 (4.5 µM) in desumoylation buffer consisting of 50 mM Tris-HCl, 1 mM DTT, 1 mM PMSF, 0.1 mM EDTA, 10% (v/v) glycerol, pH 7.5 at 4 °C for 6 hrs. The products were resolved by 18% SDS-PAGE followed by coomassie staining and anti-histone H4 immunoblotting.

3.4.15 Overexpression and purification of the full-length Rad6-Bre1 complex.

The full-length *Rad6* and *Bre1* genes were PCR amplified from *Saccharomyces cerevisiae* genomic DNA and sub-cloned into a modified pFastBac vector for co-expression in Sf9 insect cells. Recombinant viruses were produced as per manufacturer instructions, and subsequently amplified in Sf9 monolayer cells to generate high-titer viruses for protein expression in HighFive monolayer cells. Cells were harvested 2-3 days post-infection and lysed by sonication in 40 mM HEPES buffer, pH 7.5, 350 mM NaCl, 5mM b-Mercaptoethanol. The lysate was cleared of insoluble matter by ultracentrifugation at 30,000g for 30 min and the protein complex purified by Glutathione-affinity chromatography (Thermo Fisher) using the GST-tag on Bre1. The protein complex was obtained by overnight on-column cleavage of the GST-tag by TEV protease at 4 °C. The eluted Rad6-Bre1 protein complex was further purified by anion-exchange chromatography on a 1 mL HiTrap Q-HP column (GE Healthcare) followed by size-exclusion chromatography with a Superdex 200 increase column (GE Healthcare) in a final buffer containing 40 mM HEPES, pH 7.5, 100 mM NaCl, and 1mM DTT. The pure proteins were mixed with 20% sterile glycerol and stored in frozen aliquots at -20 °C.

3.4.16 Nucleosome ubiquitylation assays with Rad6-Bre1

For nucleosome ubiquitylation assays, 0.5 µM wild-type or sumoylated mononucleosomes, 100 nM UBE1 (E1 activating enzyme) , 10 µM Rad6 (E2 ligase)-Bre1 (H2BK120-specific E3 ligase) complex, 100 µM Ubiquitin, 3 mM ATP, and 100 µM DTT were mixed in a reaction volume of 10 µL buffered with 50 mM Tris-HCl, pH 8.0, 50 mM NaCl, 50 mM KCl, and 10 mM MgCl2. Assays were undertaken for 90 min at 30 °C and stopped by the addition of 4 µL SDS-containing gel-

loading buffer. Assay products were analyzed by 15% SDS-PAGE followed by immunoblotting with anti-H2B and anti-H2BK120ub-specific antibodies.

3.4.17 Rad6-Bre1 nucleosome binding assays

A fixed concentration of reconstituted wild-type or sumoylated nucleosomes (0.1 μM) was incubated with increasing concentrations (0 to 6 μM) of the Rad6-Bre1 complex in a nucleosome-binding buffer consisting of 50 mM Tris, pH 7.5, 100 mM NaCl, 8 mM MgCl2 at 4 ºC for 45 min. Bound/unbound nucleosomes were resolved by 5% TBE non-denaturing polyacrylamide gel electrophoresis followed by staining with SYBR Safe DNA stain.

3.4.18 Transient Transfection of HEK293T cells with the HA-SUMO3(ΔGG)-H4(Δ1-11) fusion, the HA-SUMO3(ΔGG)-H4 fusion, and the Myc_2-H4(Δ1-11) truncant.

HEK293T human embryonic kidney cells were cultured in T75 flasks containing Dulbecco's modified Eagle medium supplemented with 10% (v/v) fetal bovine serum. Cells were grown at 37 °C in an atmosphere of 5% CO2 until ~50% confluency. Each T75 flask of cells was transiently transfected with 15 μg of pcDNA3.1-HA-SUMO3(ΔGG)-H4(Δ1-11), pcDNA3.1-HASUMO3(ΔGG)-H4, or pcDNA3.1-Myc2-H4(Δ1-11) using Lipofectamine 3000 (Invitrogen). The growth media was changed once 24 h post-transfection and cells grown for an additional 24 h prior to cellular detachment by trypsination. Detached cells were gently collected by centrifugation at 3,000 rpm for 3 min. The resulting cell pellet was washed thrice with Dulbecco's phosphate-buffered saline (PBS) before being subjected to protein purification and subsequent ChIP assays.

3.4.19 Native chromatin immunoprecipitation (ChIP) of sumoylated or truncated H4(Δ1-11) nucleosomes.

HEK293T cells harvested 48 h post-transfection were resuspended in a modified STM-N/D buffer consisting of 250 mM Sucrose, 50 mM Tris-HCl, pH 7.4 at 4°C, 5 mM $MgCl_2$, 10 mM Iodoacetamide (IA), 20 mM N-Ethylmaleimide (NEM), 0.5 % (v/v) IGEPAL CA-630, 1x cOmplete protease inhibitor cocktail, 0.5% (w/v) Sodium deoxycholate in order to lyse cells and nuclei were isolated by centrifugation 2.5 kRCF for 5 min at 4 ºC. The precipitated nuclei were resuspended in Buffer 150 consisting of 20 mM Tris, pH 7.4 at 4°C, 0.15 M NaCl, 5 mM $MgCl_2$, 5 mM $CaCl_2$, 0.1% (v/v) IGEPAL CA-630, 1 mM PMSF, 10 mM IA, 20 mM NEM, 1x cOmplete protease inhibitor cocktail, and 10% (v/v) glycerol followed by digestion with Micrococcal nuclease (MNase) diluted in MNase Buffer consisting of 50 mM Tris-HCl, pH 8.0 at 37 °C, 5 mM $CaCl_2$, 20 mM NEM, 10 mM IA, 0.1% (v/v) Triton X-100, and 1x cOmplete protease inhibitor cocktail. Cellular chromatin (400 µL volume) was digested with MNase for 1 hour at 37 °C with mixing every 30 minutes to ensure solution homogeneity. After 1 h, chromatin digestion was halted by the addition of 54 µL of 0.5 M metal-chelator, EGTA, and the digested chromatin mixture was passed through a 25-gauge needle several times in order to reduce solution viscosity. The digested chromatin was mixed with 450 µL of Buffer 500 consisting of 20 mM Tris-HCl, pH 7.4 at 4 °C, 0.5 M NaCl, 5 mM $MgCl_2$, 5 mM $CaCl_2$, 0.1% (v/v) IGEPAL CA-630, 1 mM PMSF, 10 mM IA, 20 mM NEM, 1x cOmplete protease inhibitor cocktail, and 10% (v/v) glycerol. The mixture of chromatin and Buffer 500 was supplemented with 5 M NaCl to adjust the final salt concentration back to 0.5 M NaCl, then centrifuged at 17.2 kRCF for 20 min at 4 °C to remove any insoluble material. After sample clarification, an aliquot of the soluble fraction was analyzed by non-denaturing agarose gel electrophoresis to ascertain the generation of mononucleosomes through the observation of low molecular weight ~150 bp DNA. The remaining majority of soluble fraction was mixed with anti-HA magnetic resin beads, pre-equilibrated in Buffer 500, for 2 h at 4 °C. The unbound fraction

was subsequently removed by pipetting and the magnetic beads washed several times with Buffer 500. The HA-SUMO3(ΔGG)-H4(Δ1-11), HA-SUMO3(ΔGG)- H4, or Myc_2-H4(Δ1-11) were eluted by boiling the magnetic beads in SDS-containing gel-loading dye for no more than 3 minutes.

3.4.20 Western blotting to detect H2BK120ub and either HA-SUMO3(ΔGG)-H4(Δ1-11), HA-SUMO3(ΔGG)-H4, or Myc_2-H4(Δ1-11).

Immunoprecipitated histones were transferred to a PVDF membrane using 35 V for 2 h at 4 °C in a modified Towbin transfer-buffer consisting of 25 mM Tris-HCl, 192 mM Glycine, 4 mg/L SDS, and 10% (v/v) methanol. After protein transfer, the PVDF membrane was blocked with 5% (w/v) non-fat milk in PBS at 25 °C for 1 h prior to overnight incubation at 4 ºC with specific primary antibodies. Primary antibodies were diluted in PBST (PBS containing 0.05% (v/v) Tween-20), supplemented with 5% (w/v) non-fat milk powder. The next day, membranes were washed in PBST and then incubated for 1 h at 25 °C with the appropriate IR-dye-conjugated secondary antibody diluted in PBST supplemented with 5% (w/v) non-fat milk powder. The membrane was first washed with PBST and then PBS and subsequently scanned on a Li-COR Biosciences Odyssey IR scanner.

3.4.21 Table of Antibodies and Cell Lines used in this study.

No.	Catalog #	Supplier	Protein target	Dilution employed
1.	ab1790	Abcam	Histone H2B	0.1µg/mL (WB 1:10,000)
2.	39623	Active Motif	Histone H2BK120ub	1 µg/mL (WB 1:1,000)

3.	ab10158	Abcam	Histone H4	1 µg/mL (WB 1:1,000)
4.	ab24834	Abcam	Histone H3	1µg/mL (WB 1:1,000)
5.	ab1791	Abcam	Histone H3	1 µg/mL (WB 1:1,000)
6.	3724S	Cell Signaling Technologies	HA-tag	22 ng/mL (WB 1:1,000)
7.	2276S	Cell Signaling Technologies	Myc-tag	80 µg/mL (WB 1:1,000)
8.	926-32210	LI-COR	IR Dye 800CW Goat anti-Mouse	66.67 ng/µL (WB 1:15,000)
9.	926-68071	LI-COR	IR Dye 680RD Goat anti-Rabbit	66.67 ng/µL (WB 1:15,000)
10.	926-32351	LI-COR	IR Dye 800CW Goat anti-Mouse IgG2a	66.67 ng/µL (WB 1:15,000)
11.	CRL-3216	ATCC	HEK293T cell line used for studies of biochemical crosstalk	Lot number 70029111

3.5 Product characterization and supplemental data

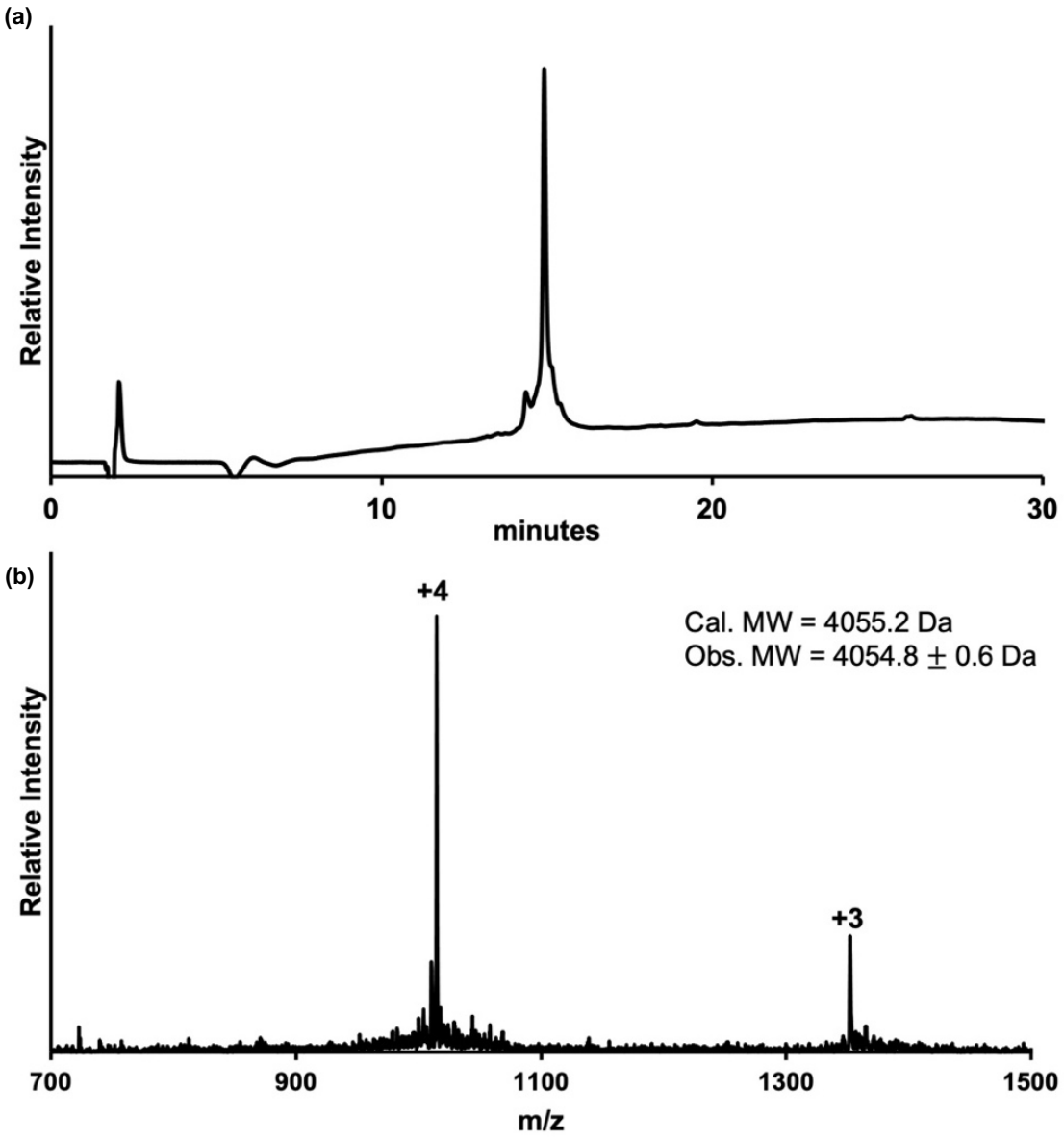


Figure S1. Synthesis of H4(1-37) K12$_{cys}$ (1). (a) C18 analytical RP-HPLC trace of pure **1** injected on a 0-73% B gradient over 30 min. (b) ESI-MS of pure **1**. Calculated M_{avg} 4,055.2 Da and observed 4,054.8 ± 0.6 Da.

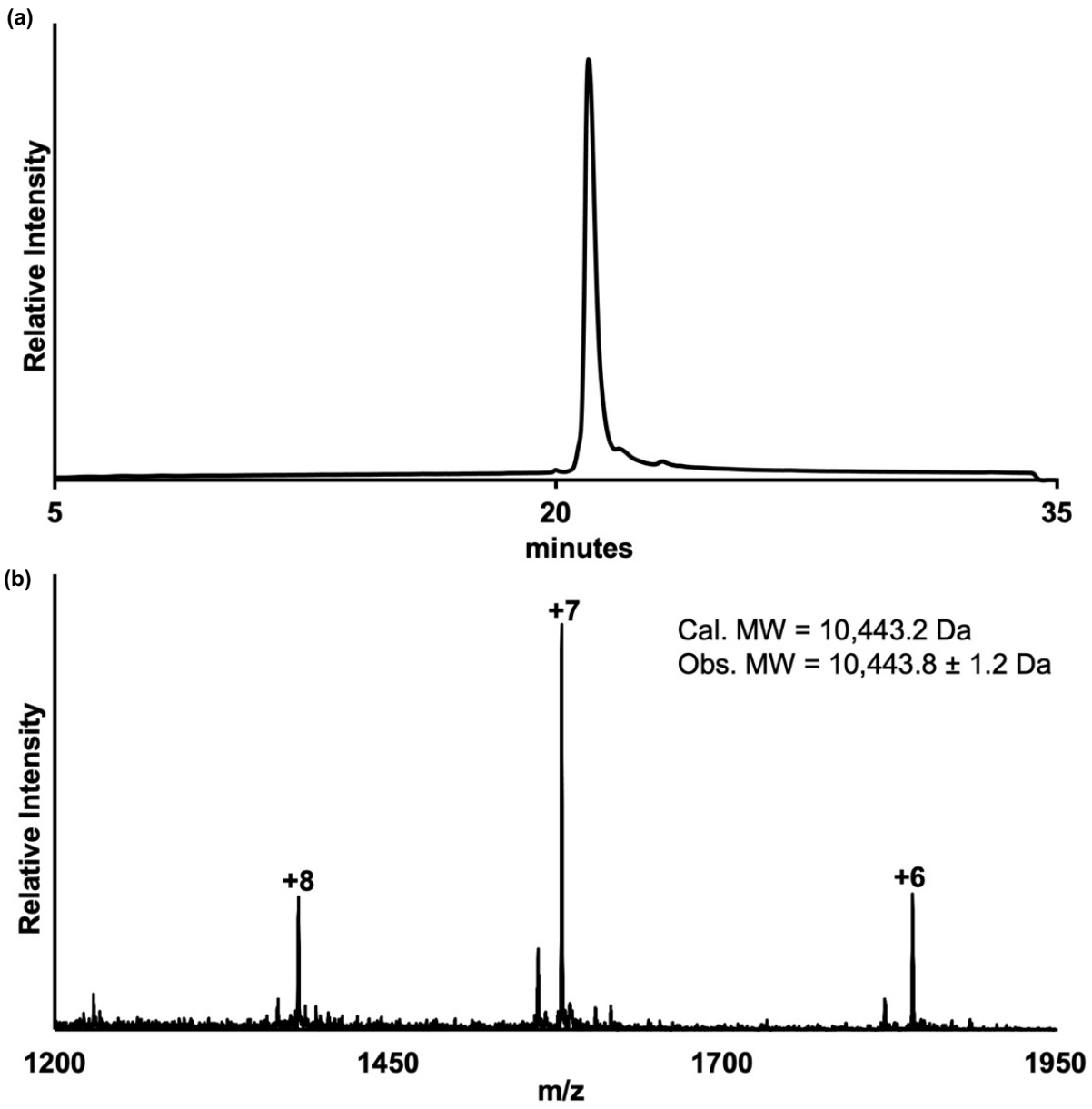


Figure S2. Intein-mediated generation of SUMO-3(2-91)C47S a-thioester. (a) C18 analytical RP-HPLC trace of pure SUMO-3(2-91)C47S a-thioester injected on a 0-60% B gradient over 30 min. (b) ESI-MS of pure SUMO-3(2-91)C47S a-thioester. Calculated Mavg 10,444.7 Da and observed 10,443.8 ± 1.5 Da.

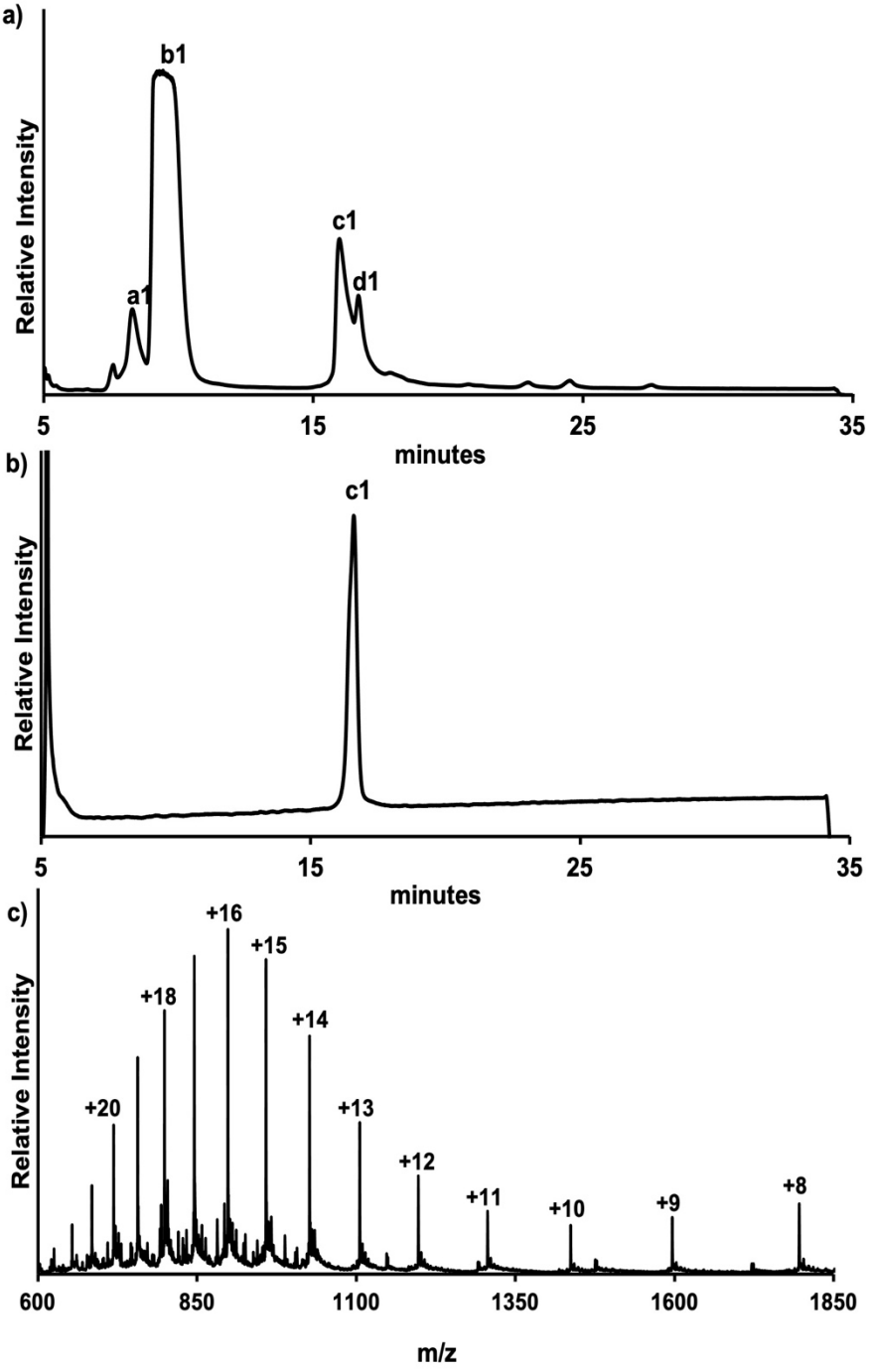


Figure S3. Ligation of H4(1-37)K12$_{cys}$ (1) with SUMO3(2-91)C47S α-thioester. (a) C18 analytical RP-HPLC trace after 8 h of ligation injected on a 0-73% B gradient over 30 min. Peak **a1** corresponds to unreacted H4(1-37)K12$_{cys}$, peak **b1** is the excess MPAA in ligation buffer, peak **d1** is hydrolyzed SUMO3(2-91)-CO_2H, and peak **c1** corresponds to the desired ligation product. (b) C18 RP-HPLC trace of the MPAA α-thioester form of the ligation product (**4**). (c) Calculated M_{avg} for MPAA α-thioester form **4**, 14,359.4 Da and observed 14,359.9 ± 0.7 Da.

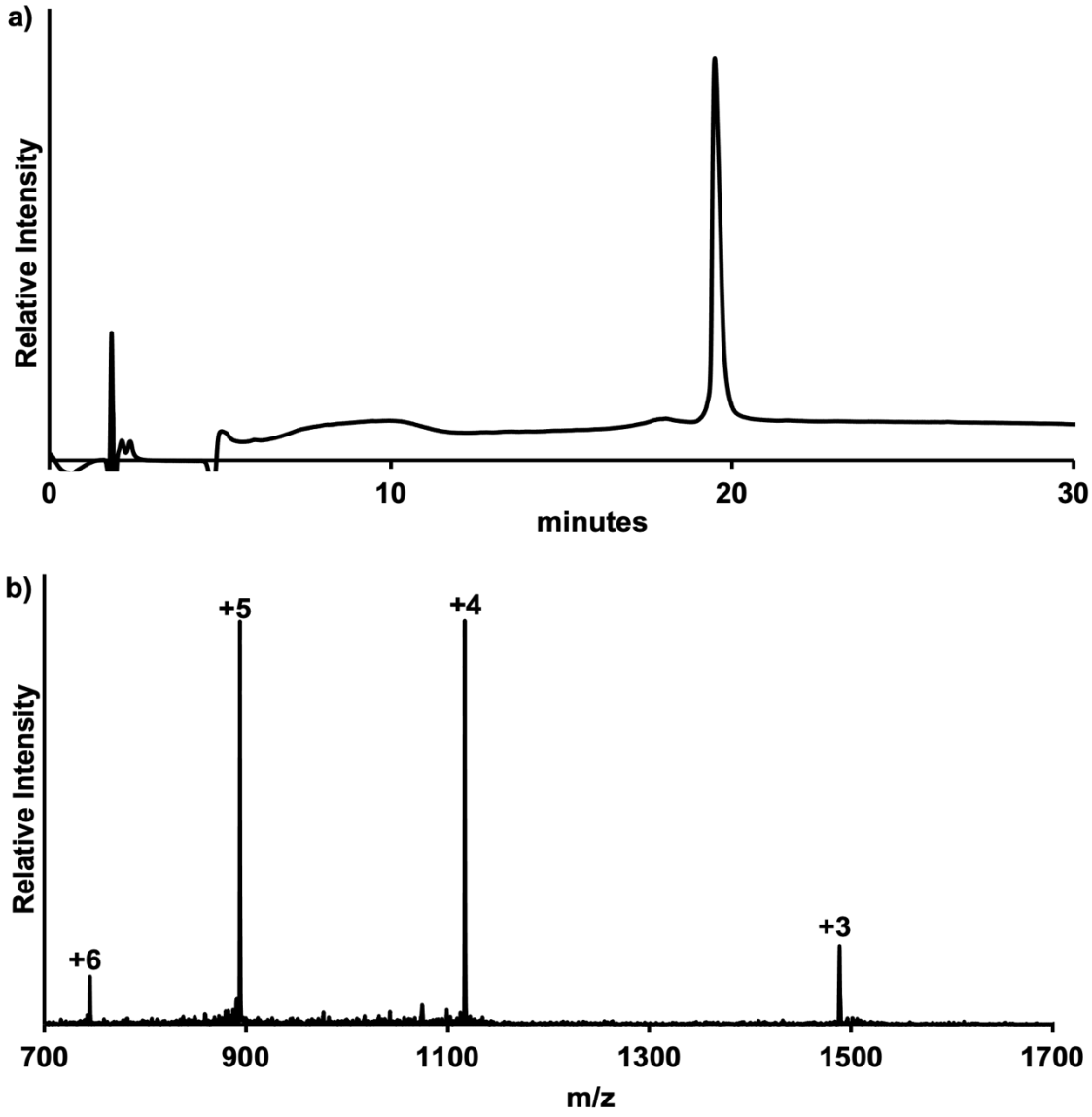


Figure S4. Synthesis of central fragment H4(38-75)A38C$_{Acm}$ (2). (a) C18 analytical RP-HPLC trace of pure **2** injected on a gradient of 20-80% B over 30 min. (b) ESI-MS of pure **2**. Calculated M_{avg} 4,465.2 Da and observed 4,465.8 ± 0.6 Da.

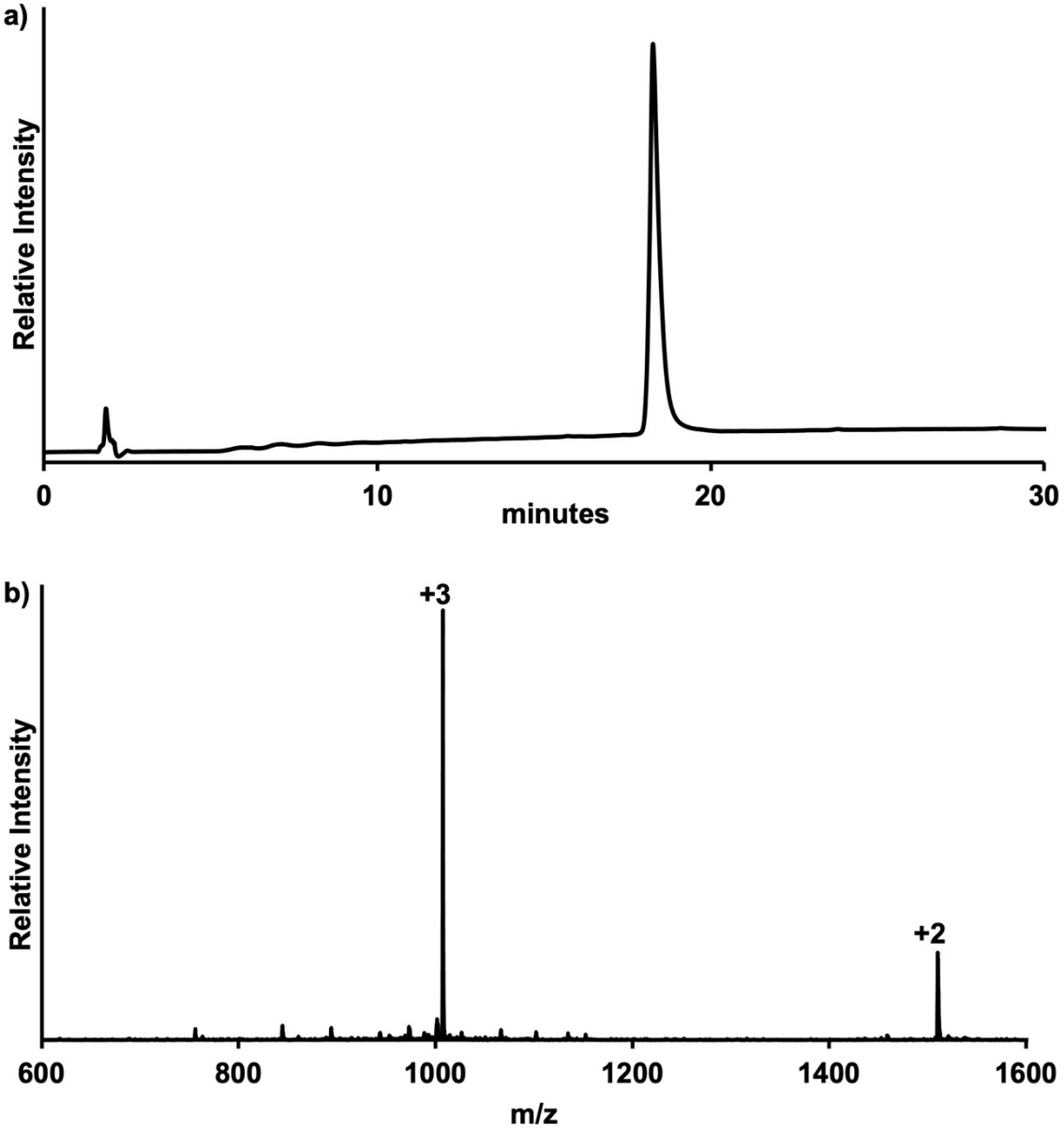


Figure S5. Synthesis of H4(76-102)A76C (3). (a) C18 analytical RP-HPLC trace of pure **3** injected on a gradient of 0-73% B over 30 min. (b) ESI-MS of pure **3**. Calculated for **3**, M_{avg} 3,019.5 Da and observed 3,019.8 ± 0.5 Da.

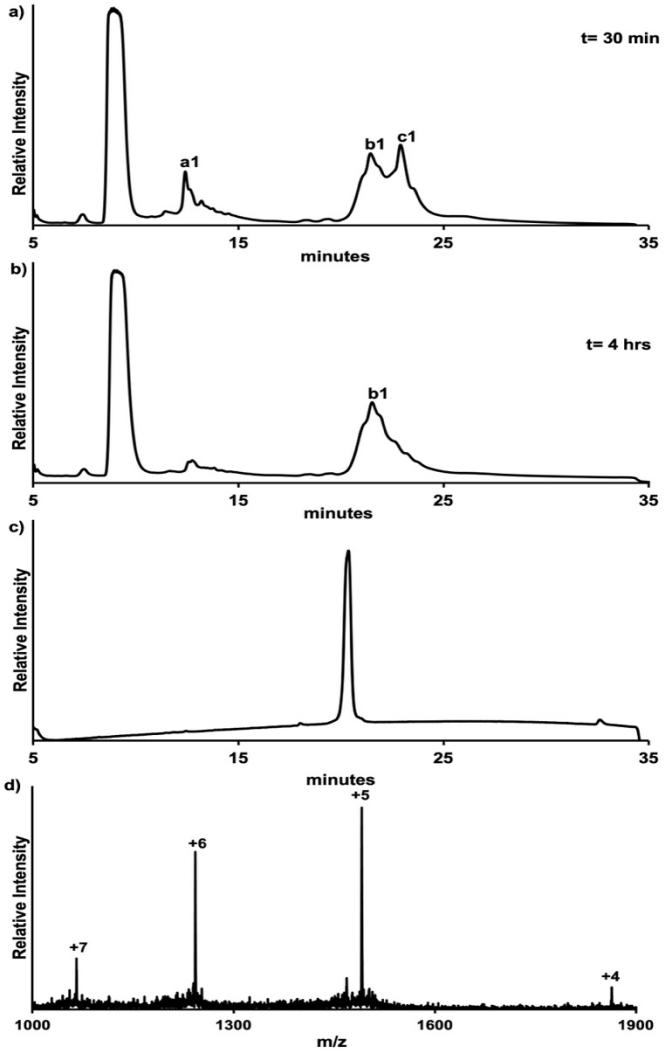


Figure S6. Ligation of H4(38-75)A38CAcm (2) and H4(76-102)A76C (3). C4 analytical RP-HPLC trace after (a) 30 min, and (b) 4 h of ligation. Peak **a1** corresponds to unreacted H4(76-102)A76C, peak **b1** contains the desired product, and peak c1 corresponds to unreacted H4(38-75)A38C$_{Acm}$-MPAA α-thioester. (c) C4 analytical HPLC trace of the purified ligation product on a 20-80% B gradient over 30 min. (d) ESI-MS of purified H4(38-102)A38C$_{Acm}$,A76C. Calculated M_{avg} for H4(38-102)A38C$_{Acm}$,A76C was 7,451.9 Da and observed 7,450.8 ± 1.1 Da.

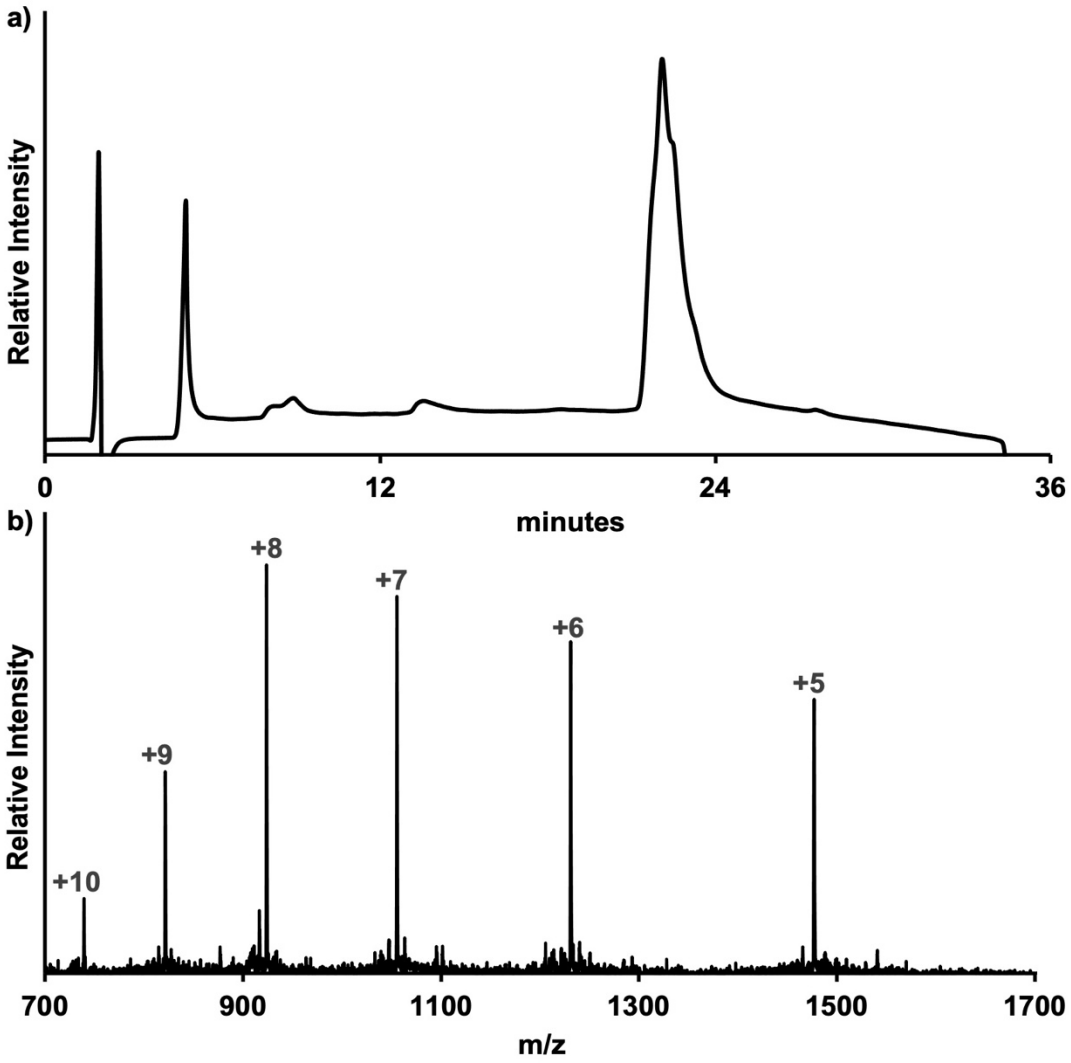


Figure S7. Acm-deprotected H4(38-102)A38C,A76C peptide (5). (a) C4 analytical HPLC trace of pure **5** injected on a gradient of 30-70% B over 30 min. (b) and ESI-MS of purified **5**. Calculated M_{avg} 7,381.6 Da and observed 7,380.6 ± 1.2 Da.

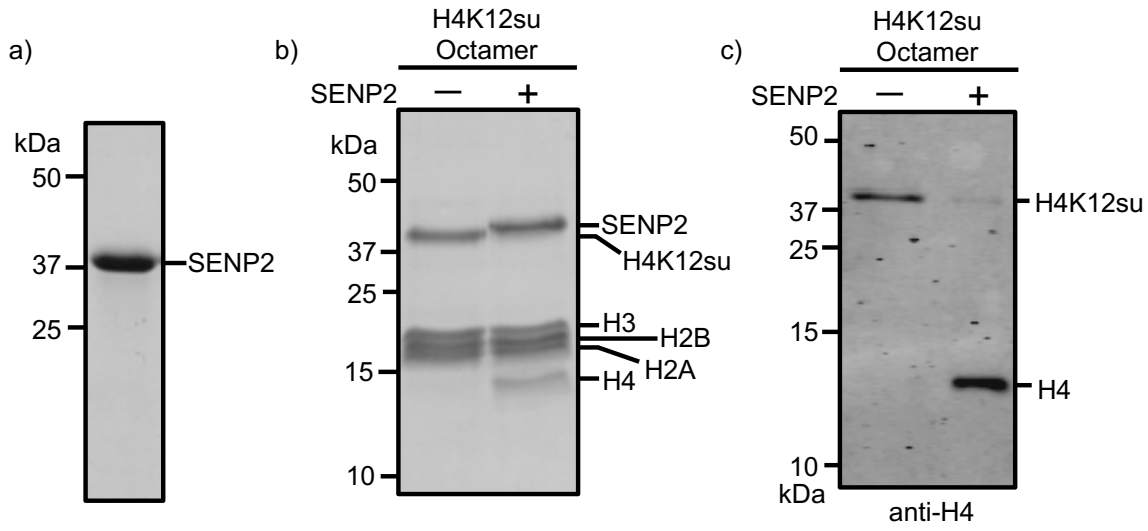


Figure S8. SENP2-mediated deSumoylation of H4K12su. (a) Coomassie-stained 15% SDS-PAGE of pure SENP2 catalytic domain. (b) 18% SDS-PAGE of H4K12su deSumoylation by the SENP2 catalytic domain. (c) Western blot with anti-H4 antibody to detect H4K12su deSumoylation by the SENP2 catalytic domain.

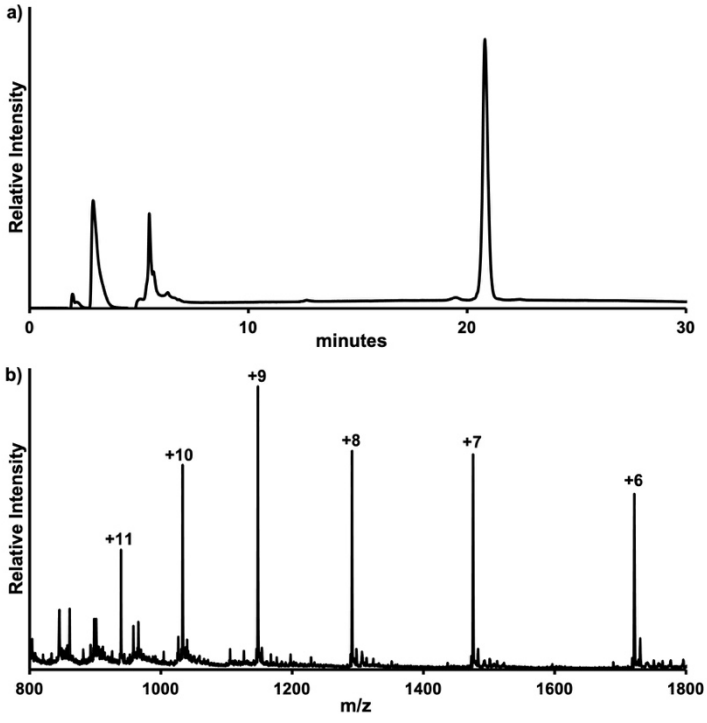


Figure S9. Synthesis of SUMO3(2-92)C47S,G92A. (a) C18 analytical RP-HPLC trace of pure SUMO3(2-92)C47S,G92A injected on a 0-60% B gradient over 30 min. (b) ESI-MS of pure SUMO3(2-92)C47S,G92A. Calculated M_{avg} 10,373.8 Da and observed 10,374.5 ± 1.4 Da.

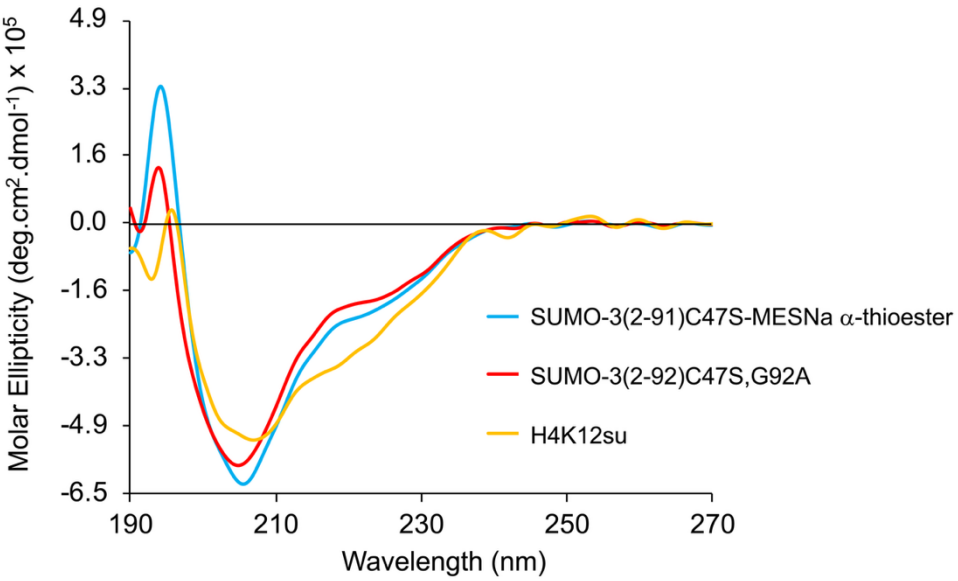


Figure S10. Circular Dichroism spectra of SUMO3(2-92)C47S,G92A; SUMO3(2-91)C47SMESNa α-thioester and H4K12su. The maxima at ~195 nm and minima at ~208 nm are characteristic of the folded SUMO3(2-91)C47S-MESNa α-thioester.[2]

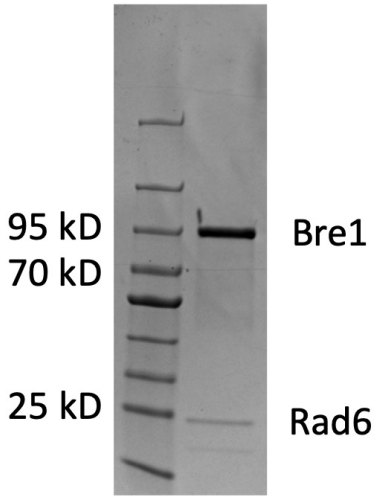


Figure S11. Purification of the Rad6-Bre1 complex. Coomassie-stained 10% SDS-PAGE of the pure Rad6 and Bre1 proteins.

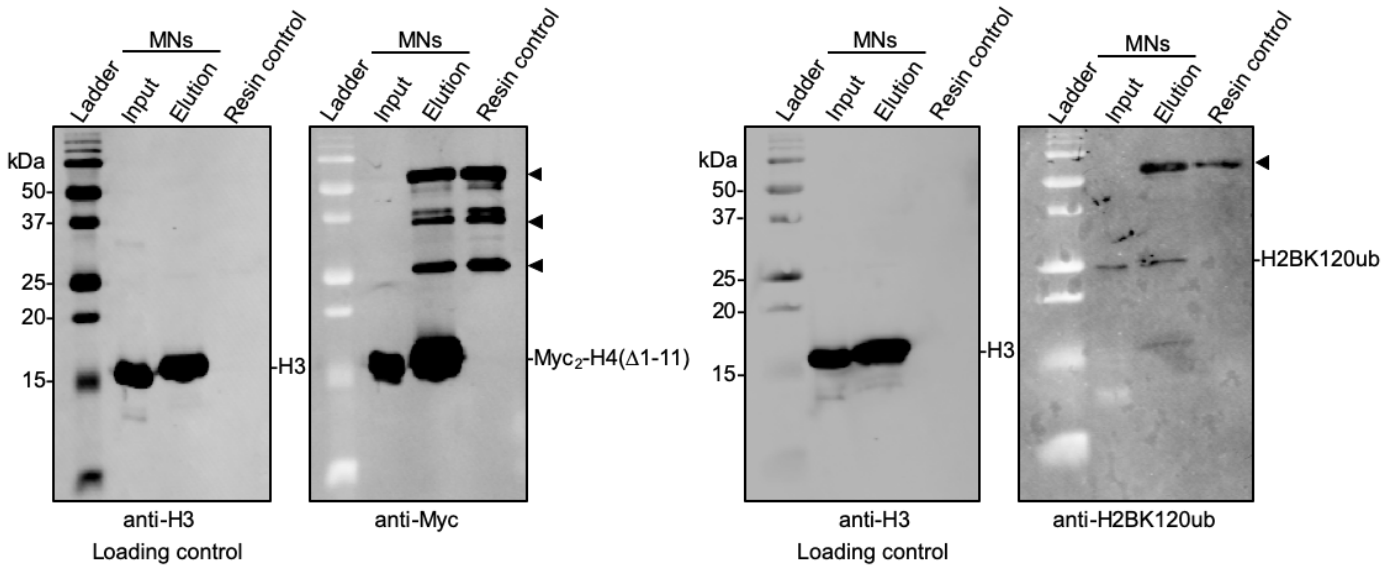


Figure S12. Myc-tag-mediated immunoprecipitation of mononucleosomes containing H4(Δ1-11). Input = input mononucleosomes (MNs) obtained by Micrococcal nuclease digestion of cellular chromatin. Elution = eluate from anti-Myc resin. Resin control = Anti-Myc resin prior to incubation with nucleosomes. Myc_2-H4(Δ1-11) = double Myc-tagged truncated H4(12-102) protein missing tail residues 1-11. An H3-specific antibody was used as a loading control for the total amount of MNs in input and elution lanes. A filled triangle indicates antibody-chains released from the resin under denaturing conditions.

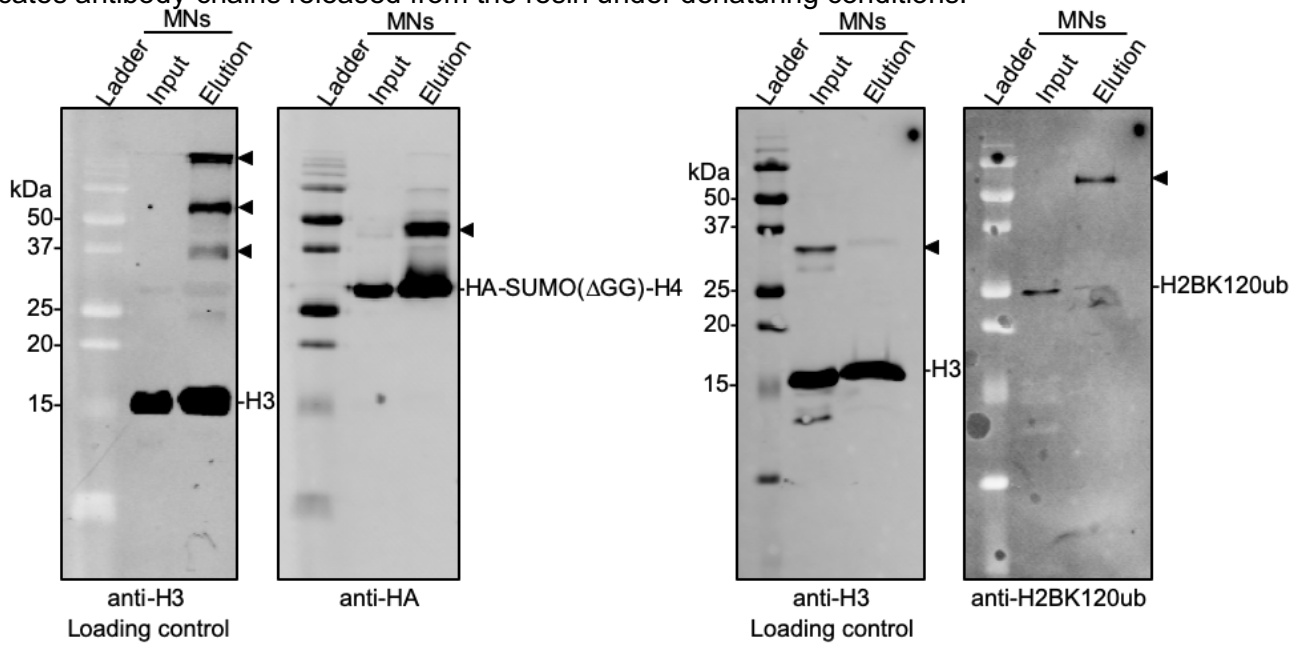


Figure S13. HA-tag-mediated immunoprecipitation of mononucleosomes containing SUMO3(ΔGG)-H4. Input = input mononucleosomes (MNs) obtained by *Micrococcal nuclease* digestion of cellular chromatin. Elution = eluate from anti-HA resin. HA-SUMO3(ΔGG)-H4 = HA-tagged protease-resistant SUMO3 missing its C-terminal Gly91 and Gly92 residues, fused to the N-terminus of full-length histone H4. An H3-specific antibody was used as a loading control for the total amount of MNs in input and elution lanes. A filled triangle indicates non-histone species cross-reactive to the secondary antibody used for imaging western blots.

3.6 References

Portions of this chapter have been published as:

Singh, S. K.; Reyna, A.; Xie, X.; Mao, H.; Ji, M.; Zheng, N.; Hsu, P. L.; Chatterjee, C. Total Chemical Synthesis of Sumoylated Histone H4 Reveals Negative Biochemical Crosstalk with Histone Ubiquitylation. *Chemical Communications* **2023**, *59* (27), 4063–4066. https://doi.org/10.1039/d2cc06683a.

1 Kim, D. H. *et al.* Volume regulation and shape bifurcation in the cell nucleus. *J Cell Sci* **128**, 3375–3385 (2015). https://doi.org/10.1242/jcs.166330

2 Weller, C. E. *et al.* Aromatic thiol-mediated cleavage of N-O bonds enables chemical ubiquitylation of folded proteins. *Nat Commun* **7**, 12979 (2016). https://doi.org/10.1038/ncomms12979

3 Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**, 251–260 (1997). https://doi.org/10.1038/38444

4 Dhall, A. & Chatterjee, C. Chemical approaches to understand the language of histone modifications. *ACS Chem Biol* **6**, 987–999 (2011). https://doi.org/10.1021/cb200142c

5 Leonen, C. J. A., Upadhyay, E. & Chatterjee, C. Studies of biochemical crosstalk in chromatin with semisynthetic histones. *Curr Opin Chem Biol* **45**, 27–34 (2018). https://doi.org/10.1016/j.cbpa.2018.02.005

6 Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403**, 41–45 (2000). https://doi.org/10.1038/47412

7 Jenuwein, T. & Allis, C. D. Translating the histone code. *Science* **293**, 1074–1080 (2001). https://doi.org/10.1126/science.1063127

8 Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res* **21**, 381–395 (2011). https://doi.org/10.1038/cr.2011.22

9 Roth, S. Y., Denu, J. M. & Allis, C. D. Histone acetyltransferases. *Annu Rev Biochem* **70**, 81–120 (2001). https://doi.org/10.1146/annurev.biochem.70.1.81

10 Flotho, A. & Melchior, F. Sumoylation: a regulatory protein modification in health and disease. *Annu Rev Biochem* **82**, 357–385 (2013). https://doi.org/10.1146/annurev-biochem-061909-093311

11 Cubenas-Potts, C. & Matunis, M. J. SUMO: a multifaceted modifier of chromatin structure and function. *Dev Cell* **24**, 1–12 (2013). https://doi.org/10.1016/j.devcel.2012.11.020

12 Hendriks, I. A. & Vertegaal, A. C. A comprehensive compilation of SUMO proteomics. *Nat Rev Mol Cell Biol* **17**, 581–595 (2016). https://doi.org/10.1038/nrm.2016.81

13 Tammsalu, T. *et al.* Proteome-wide identification of SUMO modification sites by mass spectrometry. *Nat Protoc* **10**, 1374–1388 (2015). https://doi.org/10.1038/nprot.2015.095

14 Lamoliatte, F. *et al.* Large-scale analysis of lysine SUMOylation by SUMO remnant immunoaffinity profiling. *Nat Commun* **5**, 5409 (2014). https://doi.org/10.1038/ncomms6409

15 Lumpkin, R. J. *et al.* Site-specific identification and quantitation of endogenous SUMO modifications under native conditions. *Nat Commun* **8**, 1171 (2017). https://doi.org/10.1038/s41467-017-01271-3

16 Ryu, H. Y. & Hochstrasser, M. Histone sumoylation and chromatin dynamics. *Nucleic Acids Res* **49**, 6043–6052 (2021). https://doi.org/10.1093/nar/gkab280

17 Shiio, Y. & Eisenman, R. N. Histone sumoylation is associated with transcriptional repression. *Proc Natl Acad Sci U S A* **100**, 13225–13230 (2003). https://doi.org/10.1073/pnas.1735528100
18 Galisson, F. *et al.* A novel proteomics approach to identify SUMOylated proteins and their modification sites in human cells. *Mol Cell Proteomics* **10**, M110 004796 (2011). https://doi.org/10.1074/mcp.M110.004796
19 Dhall, A., Weller, C. E., Chu, A., Shelton, P. M. M. & Chatterjee, C. Chemically Sumoylated Histone H4 Stimulates Intranucleosomal Demethylation by the LSD1-CoREST Complex. *ACS Chem Biol* **12**, 2275–2280 (2017). https://doi.org/10.1021/acschembio.7b00716
20 Dhall, A. *et al.* Sumoylated human histone H4 prevents chromatin compaction by inhibiting long-range internucleosomal interactions. *J Biol Chem* **289**, 33827–33837 (2014). https://doi.org/10.1074/jbc.M114.591644
21 Leonen, C. J. A. *et al.* Sumoylation of the human histone H4 tail inhibits p300-mediated transcription by RNA polymerase II in cellular extracts. *Elife* **10** (2021). https://doi.org/10.7554/eLife.67952
22 Shelton, P. M., Weller, C. E. & Chatterjee, C. A Facile N-Mercaptoethoxyglycinamide (MEGA) Linker Approach to Peptide Thioesterification and Cyclization. *J Am Chem Soc* **139**, 3946–3949 (2017). https://doi.org/10.1021/jacs.6b13271
23 Weake, V. M. & Workman, J. L. Histone ubiquitination: triggering gene activity. *Mol Cell* **29**, 653–663 (2008). https://doi.org/10.1016/j.molcel.2008.02.014
24 Hsu, P. L. *et al.* Structural Basis of H2B Ubiquitination-Dependent H3K4 Methylation by COMPASS. *Mol Cell* **76**, 712–723 e714 (2019). https://doi.org/10.1016/j.molcel.2019.10.013
25 Worden, E. J., Zhang, X. & Wolberger, C. Structural basis for COMPASS recognition of an H2B-ubiquitinated nucleosome. *Elife* **9** (2020). https://doi.org/10.7554/eLife.53199
26 Canne, L. E., Bark, S. J. & Kent, S. B. H. Extending the Applicability of Native Chemical Ligation. *Journal of the American Chemical Society* **118**, 5891–5896 (1996). https://doi.org/10.1021/ja960398s
27 Jbara, M., Maity, S. K., Seenaiah, M. & Brik, A. Palladium Mediated Rapid Deprotection of N-Terminal Cysteine under Native Chemical Ligation Conditions for the Efficient Preparation of Synthetically Challenging Proteins. *J Am Chem Soc* **138**, 5069–5075 (2016). https://doi.org/10.1021/jacs.5b13580
28 Wan, Q. & Danishefsky, S. J. Free-radical-based, specific desulfurization of cysteine: a powerful advance in the synthesis of polypeptides and glycopolypeptides. *Angew Chem Int Ed Engl* **46**, 9248–9252 (2007). https://doi.org/10.1002/anie.200704195
29 Wood, A., Schneider, J., Dover, J., Johnston, M. & Shilatifard, A. The Paf1 complex is essential for histone monoubiquitination by the Rad6-Bre1 complex, which signals for histone methylation by COMPASS and Dot1p. *J Biol Chem* **278**, 34739–34742 (2003). https://doi.org/10.1074/jbc.C300269200
30 Zheng, J. S., Tang, S., Qi, Y. K., Wang, Z. P. & Liu, L. Chemical synthesis of proteins using peptide hydrazides as thioester surrogates. *Nat Protoc* **8**, 2483–2495 (2013). https://doi.org/10.1038/nprot.2013.152
31 Mikolajczyk, J. *et al.* Small ubiquitin-related modifier (SUMO)-specific proteases: profiling the specificities and activities of human SENPs. *J Biol Chem* **282**, 26217–26224 (2007). https://doi.org/10.1074/jbc.M702444200

Chapter 4

Protein Scaffold Affinities Regulate Stress-associated Enzyme Activity in Biomolecular Condensates

4.1 Introduction

Liquid-liquid phase separation (LLPS) is a physical phenomenon in which a soluble mixture of molecules in solution spontaneously divides into two distinct liquid phases upon exceeding threshold conditions of temperature or concentration. Phase separation in cells, also called biomolecular condensation, is driven by favorable interactions between mixtures of proteins, nucleic acids and their ligands. Biomolecular condensates (BMCs) serve to compartmentalize diverse biochemical processes in membraneless organelles. In addition to concentrating various reactants, the core condensate composition is actively set by multivalent *scaffold* molecules that recruit lower valency *clients*, which enables selective enrichment in response to and aligned with cellular needs.[1] BMCs can range from about 50 nm to 5 mm in diameter and can regulate diverse biochemical reactions associated with metabolism, biosynthesis, transcription, stress response and proteolysis.[2] In cells, BMCs often display liquid-like behaviors such as fusion, rounding, and rapid component exchange, and in some contexts mature into less dynamic gel-like states depending on composition and environment.[2,3] Proteins scaffolding BMCs may contain multiple well-structured domains or intrinsically disordered regions that enable their binding to polyvalent targets leading to an extended network of inter-protein interactions in the separated phase. Sequence features within low-complexity segments and short linear motifs create a means to tune condensate formation and material properties.[4] Given its key role in cell

biology, the inability to form functional BMCs underlies human diseases such as leukemia and neurodegeneration. Examples include altered promyelocytic leukemia (PML) nuclear bodies in acute promyelocytic leukemia and liquid-to-solid transitions of the RNA-binding protein FUS in amyotrophic lateral sclerosis.[3,5] Therefore, there is an urgent need to understand the fundamental physical properties that guide BMC formation and function in living organisms.

Although many types of membraneless organelles have been identified in cells, the complex mixtures of proteins and/or nucleic acids, in each organelle pose significant challenges to gaining a fundamental understanding of how the affinities of scaffolding protein-peptide interactions dictate BMC properties such as size, number, stability and dynamics. This comprehensive knowledge is critically needed for engineering separated phases with well-defined and tunable properties that may be harnessed to control enzyme-catalyzed reactions for biotechnological and therapeutic applications. In cells, multivalency is widespread and arises from: (1) multidomain folded proteins with repeating or diverse interaction modules, (2) intrinsically disordered regions (IDRs) containing multiple short linear motifs, and (3) hybrid architectures that combine well-folded domains and IDRs; RNA can function as both a scaffold or as a client depending on specific contexts.[2,6] Naturally occurring scaffolding proteins include the RNA-binding protein FUS, which form liquid droplets that adopt less dynamic states in disease[3], heterochromatin protein HP1α, which forms droplets linked to chromatin compaction[7], and the LAF-1 helicase that phase separates with RNA to build P-granule condensates.[8] Additionally, engineered polyvalent protein scaffolds such as poly-SUMO/poly-SUMO-interacting motif (SIM) serve as excellent scaffolds for testing fundamental rules for BMC composition and client recruitment.[1] Clients can include low-valency variants of these scaffolds and nucleic acids; for example condensates formed by DEAD-box helicase 4 (Ddx4), a germline RNA helicase whose N-terminal intrinsically disordered region is sufficient for phase separation, solubilize a range of nucleic acids in a sequence- and composition-dependent manner.[9]

In mixtures of heterotypic scaffolds, where two different proteins associate to form BMCs, client partitioning depends on the valency of scaffolds and the relative abundance of complementary binding sites. Thus, the composition of BMCs can undergo a sharp transition when the relative concentrations of scaffold molecules cross the phase boundary (c_{sat}) defined by the system's phase diagram.[1,10] Indeed, sumoylation state tunes the composition and number of PML bodies, and increased SUMO2/3 conjugation, due to the loss or inhibition of, the SUMO protease sentrin-specific protease (SENP6), expands PML body number and size while enhancing recruitment of SIM-containing clients.[11,12]

Deciphering the physical determinants of LLPS in intracellular mixtures requires connecting microscopic sequence features and pairwise affinities to macroscopic condensate traits such as density, composition, stability, and internal dynamics across many interacting components. Key unanswered questions regarding LLPS include how valence, affinity, and stoichiometry combine to determine the stability and composition of multicomponent networks, how nucleic acids and small molecules remodel these networks in a concentration-dependent manner, and how compositional diversity relates to function.[13-15] A central challenge toward understanding LLPS lies in relating sequence-level grammar and specific interaction interfaces to mesoscale material properties in cells.[16] These gaps motivate the use of platforms in which valence, sequence context, and binding strength can be varied independently and phase-diagram position measured directly, as demonstrated in modular multivalent systems.[1,17,18]

To achieve the goal of understanding how sequence features and binding affinities regulate BMC formation and function, we were drawn to the small ubiquitin-like modifier isoform 3 (SUMO3)-based LLPS system first reported by Rosen and coworkers. In this elegant binary system, polyvalent interactions between multiple copies of the SUMO3 protein and SIM

peptides drives phase separation.[1] The system closely mimics the membraneless nuclear PML bodies that are formed by the polysumoylated PML protein as a result of its polyvalent binding to SUMO-interacting motifs (SIMs) in PML and other proteins.[5] The multimeric $(SUMO3)_m$-$(SIM)_n$ system was previously applied to study client protein recruitment and to identify molecular features that dictate small-molecule partitioning in condensates.[1,15] At the molecular level, SIMs present a short hydrophobic core, flanked by acidic residues, that docks into a shallow groove formed between the only α-helix and second *β*-strand of SUMO. Precise sequence composition of the SIM leads to varied SUMO2/3 affinities over micromolar ranges.[19-22] Given the wide range of SIMs with previously reported dissociation constants (K_d) for SUMO3 binding, we envisioned the modular SUMO3-SIM platform may be applied toward studying the variance in BMC properties with varying affinities of their scaffolding components.

Early results from our *in vitro* studies revealed a surprising lack of correlation between SUMO3-SIM affinities and the size or number of BMCs formed. No correlation was observed between the length of a SIM, the number of hydrophobic/hydrophilic residues, or its overall charge at pH 7.5 to particle dimensions. This indicates that multiple parameters, beyond the key hydrophobic SUMO-SIM interaction, contribute to the number and size of BMCs formed. Nevertheless, as phase separation is necessarily nucleated by the initial binary interaction between polySUMO3 and polySIM, the minimal concentrations of both polySUMO3 and polySIM required for BMC formation correlated well with their predicted binding affinity. This indicates a conserved mechanism for the initiation of condensate formation among the different SIMs. The stability of BMCs to chaotropic agents, such as urea and guanidine, also correlated well with the affinity of SUMO3-SIM interaction, and tighter binding (lower K_d) SIMs led to greater salt and chaotrope-resistant BMCs. These studies are challenging in cells, where factors such as protein and RNA concentration and adenosine triphosphate (ATP) at physiological millimolar levels may modulate

condensation and material properties, which underscores the importance of our biochemical approach with well-defined components.[23,24]

One highly correlative property across various BMCs was the fluidity or dynamics of the separated phase. Fluorescence recovery after photobleaching (FRAP) experiments performed *in vitro* and in living cells revealed consistently greater dynamics in condensates assembled with weaker binding SIMs (higher K_d). Consistent with this observation, alanine mutants of key hydrophobic residues in tight-binding SIMs, which diminished their affinity for SUMO3, also enhanced BMC dynamics. Finally, to engineer BMCs of varying sizes with varying SUMO3-SIM scaffolds, we found that the addition of increasing amounts of positive charge to the C-terminus of a polySIM construct led to increasing particle sizes and diminished condensate dynamics.

Thus, our studies with the modular polySUMO3-polySIM system identified three unique properties of BMCs that correlate with protein-peptide affinities, namely, (1) the minimal concentration at which phase separation is observed, (2) the stability of the BMC to salts/chaotropic agents, and (3) the dynamics of the separated phase. Importantly, we have discovered that observed trends in BMC dynamics *in vitro* are preserved across three different cell types, which bodes well for future efforts to engineer BMCs in specific cell types with unique dynamic properties. Our surprising discovery that protein-peptide binding affinity alone is not predictive for either condensate size or number, but, predicts BMC dynamics may explain how cells generate membraneless organelles using scaffolds of vastly varying affinities. Thus, we suggest that weakly associating polyvalent scaffolds may be best suited for most biochemical processes that require the free diffusion of enzymes and their substrates.

4.2 Results and discussion

4.2.1 The phase boundaries of polyvalent biomolecular condensates vary with scaffold affinities

Inspired by pioneering work from Rosen and coworkers, we employed the 91-residue protein SUMO3 as a scaffold for BMC generation.[1] While many protein and peptide-based coacervate systems have been reported to date, Rosen's system based on SUMO3 binding to a SUMO3-interacting motif peptide (SIM) benefits from its well-defined 1:1 stoichiometry and a specific region, the SIM-binding groove, of SUMO3 where binding occurs in a groove between *β*2 and *α*1. All known SIMs till date have a hydrophobic core of 3-to-4 residues flanked by acidic or phosphorylatable residues, providing a well-defined region that can be tuned by minor sequence changes.[19,20,25,26] Hence, the SUMO3-SIM system is highly engineerable because varying the SIM changes binding affinity to SUMO3 but retains the mode of binding in the groove. We envisioned that would allow us to directly map scaffold K_d to BMC properties.[1,19,20,22,27]

We began our studies with a SUMO3-SIM system consisting of an N-terminal monomeric enhanced green fluorescent protein (mEGFP) fused with three copies of SUMO3, mEGFP-$(SUMO3)_3$, and another N-terminal mEGFP fused with three copies of the SIM sequence from PIASx, Protein Inhibitor of Activated STAT x, mEGFP-$(PIASx)_3$ (**Table 1**). The repeated SUMO3 and SIM sequences were separated by a conserved 12-mer linker sequence *N*-GGSGGSGGSGGS-*C*. Because higher-valent SUMO:SIM ratios, such as 10:10 or 5:5, exhibit sub-nanomolar affinities,[1] we chose a 3:3 SUMO:SIM ratio to generate BMCs in order to capture the effects of varying SUMO-SIM K_d on BMC properties. We reasoned that biophysical effects of 10-fold changes in SUMO3:SIM affinity (~5 μM to ~50 μM) would be distinguishable in the 3:3 system as its apparent K_d lies in the mid-nanomolar regime.[1]

SIM	Primary sequence	Reported K_d (μM)	Measured K_d (μM)
MCAF1	GSPEFKTIDASVSKKAADSTSQCGKATGSDSSGVIDLTMDDEESGASQD	1.3 ± 0.11[25]	3.8 ± 0.2
PIASx	KVDVIDLTIESSSDEEEDPPAKR	5.07 ± 0.51[20]	8.4 ± 0.8
DAXXn	MATANSIIVLDDDDEDE	38 ± 6[28]	16.6 ± 1.9
SIM2	RETAGDEIVDLTCESLEPRW	40 ± 10[26]	63.8 ± 1.7
DAXXc	KTSVATQCDPEEIIVLSDSD	102.5 ± 13.9[29]	70.3 ± 3.0

Table 1. SIM sequences with literature and standardized ITC affinities. For the SIM modules examined, sequences are shown alongside the published K_d under original conditions and an ITC K_d measured here using matched, standardized conditions.

We observed robust phase separation upon mixing the mEGFP-$(SUMO3)_3$ with mEGFP-$(PIASx)_3$ by using confocal fluorescence microscopy with excitation and emission wavelengths of 488 nm and 507 nm, respectively. The minimal concentration of SUMO3 and peptide at which phase separation was first observed, also known as c_{sat} at the phase boundary,[2,16] correlated well with the previously reported and our own measured K_d of monovalent interactions (Table 1). Thus, tighter SUMO3 binding SIMs such as the MBD1-Containing Chromatin-Associated Factor 1 (MCAF1) and PIASx SIMs formed distinct phases beginning at 60 nM mEGFP-$(SUMO3)_3$ (**Figure 1A**).[1,25] A much weaker binding SIM, such as the C-terminal SIM from Death domain Associated Protein (DAXXc), separated at higher concentrations of 100 nM mEGFP-$(SUMO3)_5$ (**Figure 1A**).[28] SIMs with intermediate affinity such as DAXXn from the N-terminus of DAXX, showed phase boundaries at intermediate concentrations of about 80 nM mEGFP-$(SUMO3)_5$.[26,28,29] Similar correlations in c_{sat} were previously noted with phase-separating systems that employed changing polyvalency to modulate scaffold affinities[1,17] or changing K_d in intrinsically disordered peptide-based systems.[30] The clear correlation between phase

boundaries and SUMO3-SIM K_d indicates that binding is both necessary and sufficient for phase separation in this platform.

Interestingly, however, we observed no correlation between SUMO3-SIM affinities and either condensate size or the number of condensates observed (**Supplementary Figure S1**). We tested if the lack of correlation was a consequence of the *magic number* effect for polyvalent systems, where equivalent numbers of binding sites on both scaffolding proteins may inhibit or entirely preclude phase-separation.[31] Therefore, we generated an asymmetric system of mEGFP-$(SUMO3)_5$ and mEGFP-$(SIM)_3$ which retains the 3:3 ratio of SUMO3-SIM binding but has two additional unfilled valencies in SUMO3 that preclude *magic number* effects. The phase boundary observed for each SUMO3-SIM pair tested in the 5:3 asymmetric system also correlated with reported K_d, and phase boundaries moved to higher concentrations for weaker binders (**Figure 1B**). However, even in the asymmetric 5:3 system, no correlation was observed between SUMO3-SIM affinities and the size or number of condensates formed (**Supplementary Figure S2**). The SIM peptides employed in these assays covered a wide range of molecular weights, isoelectric points (pI), charged residues, and hydrophobicity (**Table 2**) and no single parameter was found to be predictive for the observed particle sizes and numbers of BMCs formed.

SIM	Sequence Length	Molecular Weight (Da)	pI	Grand average of hydropathicity (GRAVY)	Positively Charged Residues	Negatively Charged Residues
MCAF1	49	4,912.16	4.00	-0.706	8%	20%
PIASx	23	2,572.76	4.79	-1.026	13%	26%
DAXXn	17	1,865.94	2.76	-0.412	0%	41%
SIM2	20	2,319.53	4.03	-0.725	10%	30%
DAXXc	20	2,150.34	3.66	-0.240	5%	20%

Table 2. Sequence-derived properties of the SIM peptides. For each SIM peptide, molecular weight, pI, counts of acidic and basic residues, and a hydrophobicity index are reported, highlighting broad variation across sequences.

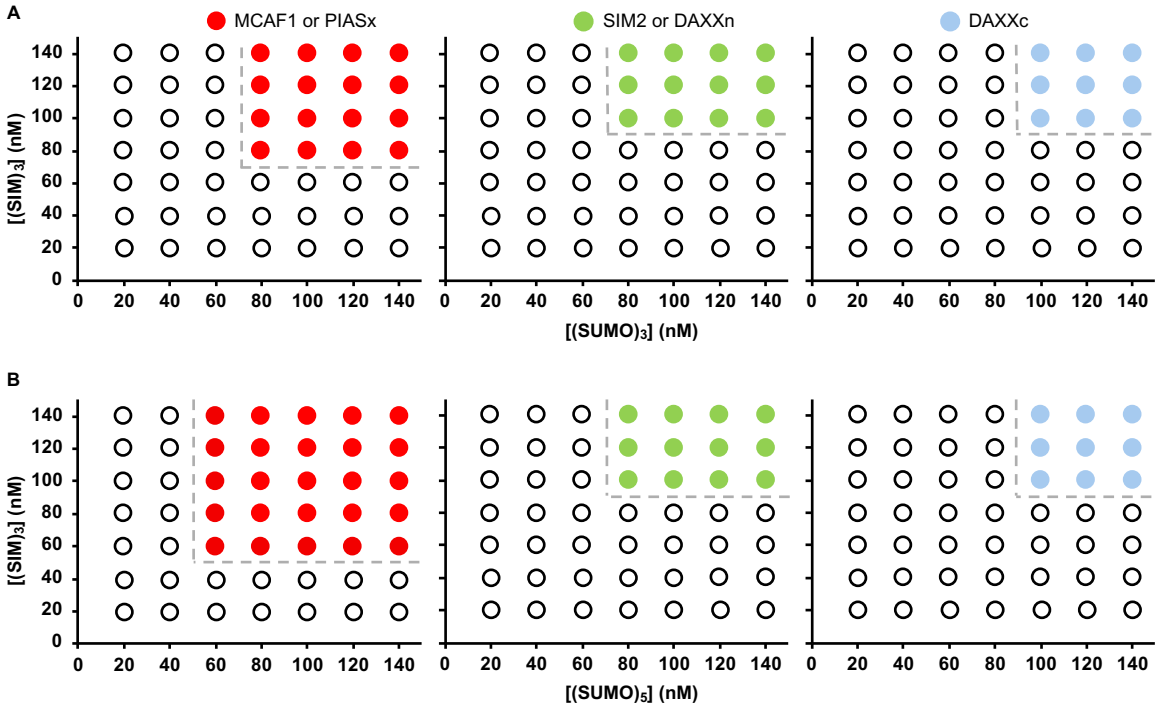


Figure 1. Phase diagram characterization of polymeric SUMO constructs to polymeric SIM constructs. Phase diagram between mEGFP-$(SUMO3)_3$ (**A**) or mEGFP-$(SUMO3)_5$ (**B**) and mEGFP-$(MCAF1)_3$ (red), or mEGFP-$(PIASx)_3$ (red), or mEGFP-$(SIM2)_3$ (green), or mEGFP-$(DAXXn)_3$ (green), or mEGFP-$(DAXXc)_3$ (blue), indicating where the solution consists of a single phase (open dots) or of two phases (colored dots). Axes show concentration of each scaffold. Dotted lines indicate the apparent phase boundary.

4.2.2 Condensate stability varies with scaffold affinities

Having observed a clear correlation between SUMO3-SIM affinities and phase boundaries, we asked if the stability of BMCs to cellular hydrotropes, electrolytes, and chemical denaturants may also correlate with affinities of the scaffolding proteins. The mEGFP-$(SUMO3)_5$ and mEGFP-$(SIM)_3$ system was employed to first generate BMCs at concentrations that promote phase separation for each peptide and were firmly within the phase boundary ($>c_{sat}$). This was followed by the addition of increasing amounts of ATP, KCl, urea or guanidine hydrochloride (GnHCl) (**Figure 2A-D**).

Interestingly, increasing amounts of the hydrotrope ATP, up to 10 mM, did not lead to a significant change in the number of BMCs formed over 24 h (**Figure 2A**). Because ATP's hydrotropic action is expected in the millimolar range, the absence of an effect even at 10 mM indicates that these SUMO-SIM condensates are ATP-resistant within the physiological window, which ranges between ~1-10 mM.[24,32]

In contrast with this observation, even condensates generated by SIMs with low micromolar affinities, such as PIASx and MCAF1 (K_d ~4-8 μM) exhibited a concentration-dependent sensitivity to KCl, urea and guanidine (**Figure 2B-D**). This was evident from the formation of fewer condensates at all concentrations of salt and denaturants tested. Indeed, as little as 50 mM KCl or 250 mM urea/GnHCl reduced the number of BMCs formed by weaker binding peptides such as SIM2 and DAXXc, whereas SIMs such as PIASx and MCAF1 showed less sensitivity to these concentrations of salt and denaturant.

Overall, the SUMO3-SIM binding affinity correlated well with the midpoint of total BMCs observed in the presence of KCl, urea or guanidine. Tighter SUMO3-SIM binding shifted the midpoint to higher perturbant concentrations, indicating that the same specific interactions that lower c_{sat} also confer resistance to destabilization by salts and chaotropes.

Collectively, these results are consistent with BMC stability requiring a well-folded structure of SUMO3 and suggest that in addition to the well-known hydrophobic core interactions there is an electrostatic component to the SUMO3-SIM interaction. This is consistent with the proposed role for negatively charged residues or phosphorylatable residues found near the hydrophobic patch in SIMs.

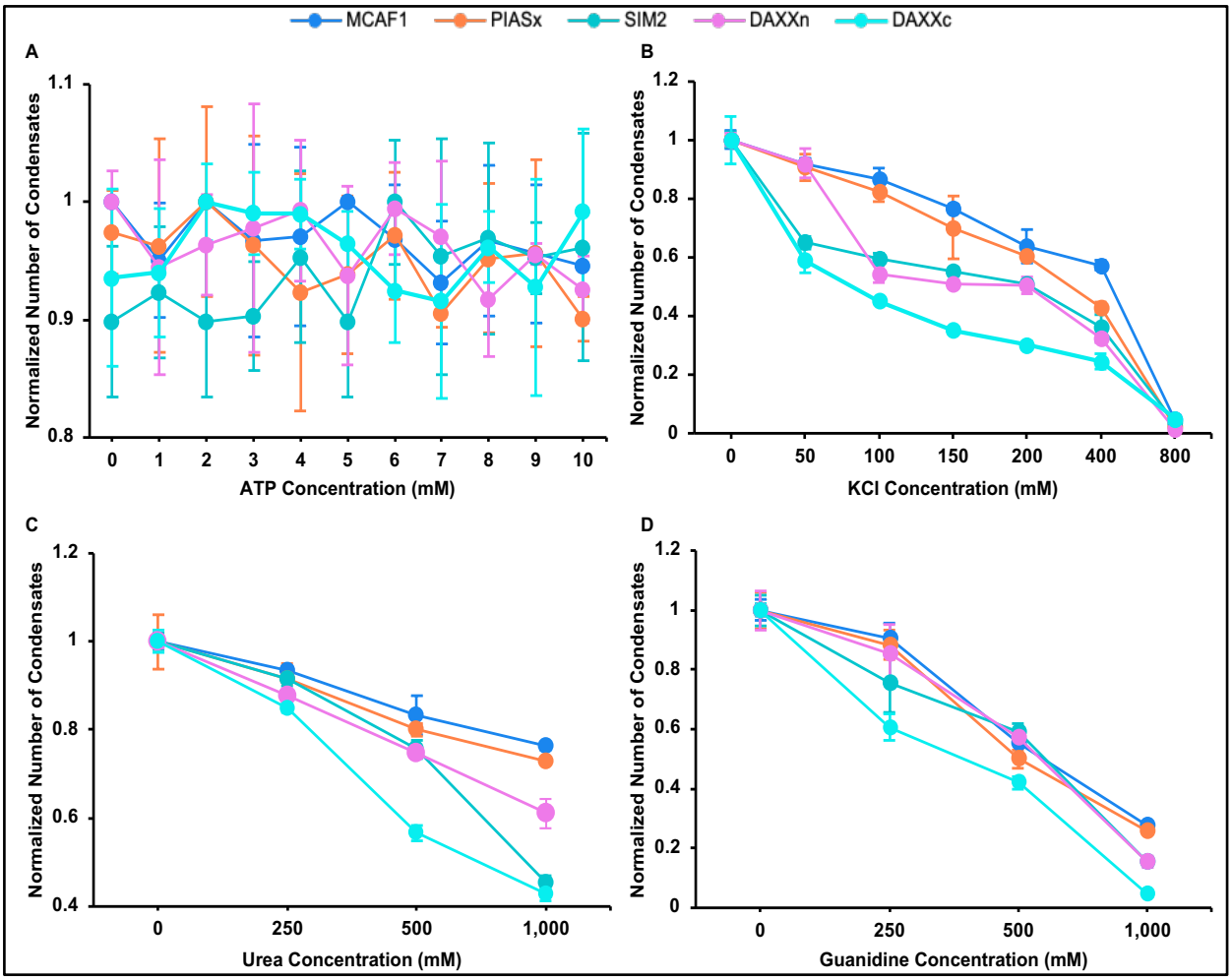


Figure 2. Evaluation of condensates, formed by mEGFP-$(SUMO3)_5$ and mEGFP-$(SIM)_3$, resistance to hydrotrope, salts, and chaotropes. (A) Treating preformed condensates with increasing concentrations of ATP, which disrupts electrostatic interactions in condensates, showed no significant trends in condensate integrity. **(B)** Treating preformed condensates with increasing concentrations of KCl dissolved condensates after 400 mM KCl. Condensates formed via strong SUMO-SIM interactions were more resistant at increased KCl concentrations. **(C,D)** Treating preformed condensates with increasing concentrations of chaotropes dissolved condensates after 1 M of urea (C) or 1M guanidine (D). Condensates formed via strong SUMO-SIM interactions were more resistant toward increasing chaotrope concentrations.

4.2.3 The dynamics of condensates correlate with SUMO3-SIM affinities

Although cellular biomolecular condensates are comprised of many different protein and/or nucleic acid components, it is unclear how the binding affinities of various components, that may

bind the same protein(s), influence the dynamic properties of condensates. This is especially relevant to many enzymatic functions associated with membraneless organelles. Hence, we investigated the dynamics of each unique condensate formed by SUMO3-SIM pairs of varying affinities. We used the standard interpretation of condensate biophysics that fast recovery reflects higher mobility within the particle and/or faster unbinding, whereas slower recovery indicates reduced mobility and longer residence times in the bound form.[33,34]

In this platform, the time required to recover fluorescence in a bleached spot is an indicator of the dynamic nature, or mobility, of components in the BMC. FRAP recovery in condensates can arise from internal rearrangement inside the dense phase, exchange with the surrounding dilute phase, or a combination of the two. Our experimental design was chosen to differentiate these possibilities.[34,35] The constructs mEGFP-$(SUMO3)_5$ and mEGFP-$(SIM)_3$ were mixed at specific concentrations indicated by their phase boundaries, incubated to allow phase separation, and then transferred to glass slides before FRAP imaging. To standardize geometry, we used a circular bleach region corresponding to ~50% of the total visible area, a regime where single-exponential fits typically capture the dominant reaction-diffusion mode in droplets.[34]

We found that condensates formed by tighter SUMO3-SIM interactions, such as with the MCAF1 or PIASx SIMs, recovered fluorescence more slowly when compared to condensates formed by weaker-binding SUMO3-SIM pairs such as SIM2 or DAXXc (**Figure 3A, Supplementary Figure S3**).

The fact that fluorescence recovery was faster in condensates formed by weaker binding pairs suggests that replacement of bleached proteins from the bulk solution is not the major mechanism underlying FRAP, because recruitment from the bulk solution should be favored for tighter binding SUMO3-SIM pairs at equimolar concentrations as weaker binding pairs. Instead,

reorganization within the condensate is a more likely mechanism underlying fluorescence recovery in the $(SUMO3)_5$+$(SIM)_3$ system.

Consistent with the internal reorganization mechanism, when condensates were generated at identical protein concentrations for different SUMO3-SIM pairs, their FRAP half-lives remained largely unchanged, with weaker binders forming more dynamic condensates (**Supplementary Figure S4, Supplemental S5**). Equimolar protein concentrations ensured that any contributions from bulk exchange were only reflective of differences in SUMO3-SIM K_d and highlighted the major contribution of differences in dissociation kinetics between scaffolding pairs.[34] Finally, consistent with our hypothesis, the correlation remained unchanged even when initial concentrations of the weaker-binding SUMO3-SIM pair were lower than for the tighter-binding SUMO3-SIM pair. Thus, the dynamic properties of condensates formed by polyvalent interactions between binding partners with K_d ~1 to 100 μM are dominated by the redistribution of proteins within condensates and not by protein transport from the bulk solution to condensates. The inverse correlation between SUMO3-SIM K_d and FRAP half-lives also indicates that dissociation of polyvalent interactions is the rate-limiting step in the redistribution of proteins in condensates.

Together, our results support a model in which dissociation of polyvalent interactions is the rate-limiting step for redistribution, so that stronger SUMO3-SIM binding yields longer residence times and slower FRAP.[33,34]

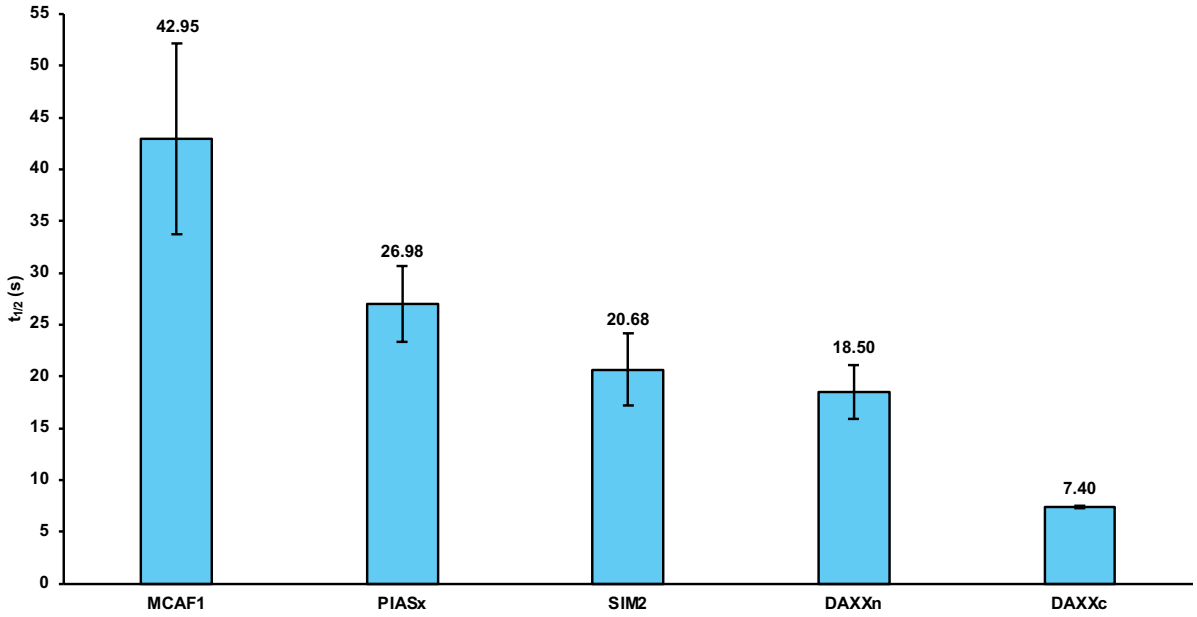


Figure 3. Histogram of measured average half-lives of photobleached condensates. BMCs were formed with SUMO-SIM pairs of different affinities at equimolar concentrations. The mean half-times for each condensate group was calculated from ≥6 similarly sized condensates (**~6 μm^2**) where ~50% of the total visible area was photobleached. Error bars indicate the standard error of the mean.

4.2.4 Hydrophobic core residues in SUMO3 peptides dictate condensate properties

While we discovered that SUMO3-binding SIM peptides with increasing K_d formed increasingly mobile condensates (**Figure 3A**), the different amino acid compositions and lengths of the peptides employed in our studies (**Table 2**) left the possibility that any of several factors besides K_d, such as molecular weight or pI may also underlie observed trends in condensate properties. To eliminate these confounding variables and to test whether affinity alone explains the trends, we implemented a perturbation strategy in which sequence length, net charge, linker spacing, and valence were held constant while affinity was tuned by single-site alanine substitutions in

the SIM hydrophobic core. Alanine preserves backbone geometry and charge while replacing hydrophobic side chains (Val, Ile, Leu) with a smaller methyl group, which reduces hydrophobic packing in the SUMO groove and weakens the extended β-sheet contact with the β2 strand in SUMO3. This substitution also lowers shape complementarity and buried hydrophobic area within the groove, while leaving the flanking acidic residues intact so their contacts with the basic patches of SUMO3 are preserved.[19,20,22,25,36] To confirm that the varying properties we observed primarily arose from changes in K_d we tested the effect of single mutations in two different SIM peptides on condensate stability and dynamics. Prior studies have shown that alanine substitutions at these core hydrophobic positions reduce SUMO binding to varying degrees.[19,20,36] We envisioned that single Ala mutations in the hydrophobic cores of the relatively tight-binding PIASx and MCAF1 peptides would weaken SUMO3 interactions without changing other peptide parameters. Hence, we mutated Val in the conserved hydrophobic core sequence (*N*-VIDL-*C*) of MCAF1 and PIASx. We introduced a Val-to-Ala substitution in all three SIMs, generating mEGFP-(PIASx(V4A))$_3$ and mEGFP-(MCAF1(V34A))$_3$, to preserve the overall scaffold architecture such that changes in c_{sat} or BMC dynamics could be attributed to altered SUMO3-SIM affinity.

Consistent with their expected lower affinity for SUMO3 and with our previous observations with weaker binding peptides, both MCAF1(V34A) and PIASx(V4A) scaffolds exhibited a higher c_{sat} for condensation than their wild-type (wt) counterparts (**Figure 4A**). To confirm their reduced affinity for SUMO3, the wt SIM and Ala mutants were synthesized via solid-phase peptide synthesis (SPPS) and their binding to SUMO3 measured via isothermal titration calorimetry (ITC) by a colleague, graduate student Madyson Briggs, in the Chatterjee lab. These measurements revealed that the K_d increased from 8.4 ± 0.8 μM for wt PIASx to 57.1 ± 3.5 μM for PIASx(V4A) and from 3.8 ± 0.2 μM for wt MCAF1 to 16.5 ± 0.4 μM to MCAF1(V34A),

respectively (**Figure 4B, Table 3**). The change in c_{sat} for the Ala mutants is consistent with previously observed effects of K_d controlling the c_{sat} in IDR peptides.[1,17,27]

In addition to the PIASx(V4A) mutant, another Ala mutants of the PIASx hydrophobic core sequence, Leu-to-Ala (L7A), was also generated. The mEGFP-(PIASx(L7A))$_3$ construct was expressed and purified similar to the wild-type SIM constructs. Surprisingly, however, our attempts to generate condensates with this Ala mutant construct were unsuccessful at all concentrations tested with mEGFP-(SUMO3)$_5$. We surmised that this may be due to their greatly reduced affinity for SUMO3. To test this hypothesis, Madyson Briggs additionally generated this peptide via SPPS to measure their binding affinity to SUMO3 via ITC. Consistent with their inability to form condensates at the concentrations tested *in vitro*, the synthetic peptide for PIASx (L7A) failed to bind SUMO3 at the highest concentration tested in ITC experiments. (**Supplemental Figure S4.4.11.4. Supplementary Figure S6**).

SIM	Primary sequence	Measured K_d (μM)
MCAF1	GSPEFKTIDASVSKKAADSTSQCGKATGSDSSGVIDLTMDDEESGASQD	3.8 ± 0.2
MCAF1 (V34A)	GSPEFKTIDASVSKKAADSTSQCGKATGSDSSGAIDLTMDDEESGASQD	16.5 ± 0.4
PIASx	KVDVIDLTIESSSDEEEDPPAKR	8.4 ± 0.8
PIASx (V4A)	KVDAIDLTIESSSDEEEDPPAKR	57.1 ± 3.5

Table 3. K_d comparison for wild-type SIM sequences for MCAF1 and PIASx with their alanine mutants. For the SIM modules examined, sequences are shown alongside the ITC K_d measured here using matched, standardized conditions.

We next subjected condensates formed by mEGFP-(PIASx(V4A))$_3$ and mEGFP-(MCAF1(V34A))$_3$ at concentrations above their c_{sat} to increasing concentrations of KCl, urea or guanidine. In keeping with other peptides exhibiting weaker affinities for SUMO3 (K_d ~50 μM in **Figure 2**) condensates generated by the Ala mutants were significantly more sensitive to salt and denaturant than the wild-type condensates and closely resembled low-affinity SUMO3 peptides (**Figure 4C**). This pattern is consistent with KCl screening electrostatic contacts between acidic SIM flanks and basic patches on SUMO3, and with urea and guanidine weakening hydrophobic and hydrogen-bonding within the SUMO groove. Guanidine produced stronger effects than urea at matched molarity, as expected for a more chaotropic cosolute.[22,37] Based on the observation that the phase boundary and salt/denaturant sensitivity of Ala mutants resembled other mid-range SUMO3 binding peptides, we asked if they would also show diminished $t_{1/2}$ and greater dynamics in FRAP experiments. FRAP was performed on size-matched condensates with a circular bleach region that covered about half of the droplet area, and fluorescence recoveries were quantified as $t_{1/2}$. Analysis of FRAP half-times for both PIASx(V4A) and MCAF1(V34A) revealed significantly greater dynamics when compared to condensates generated with wt peptides (**Figure 4D, Supplementary Figure S7**). Together, these outcomes indicate that reducing SUMO3–SIM affinity shortens contact lifetimes in the dense phase, at fixed valence, and yields faster internal reorganization.

Collectively, these results strengthen our conclusion that SUMO3–SIM affinity is a primary determinant of condensate behavior. In the native SIM series, weaker binding increased the concentration required for phase separation, heightened sensitivity to KCl, urea, and guanidine, and yielded shorter FRAP half-times. The V4A and V34A substitutions reproduced the same shifts while sequence length, linker spacing, charge, and valence were held constant, indicating

that changes in affinity arising from the hydrophobic core are sufficient to account for these trends.

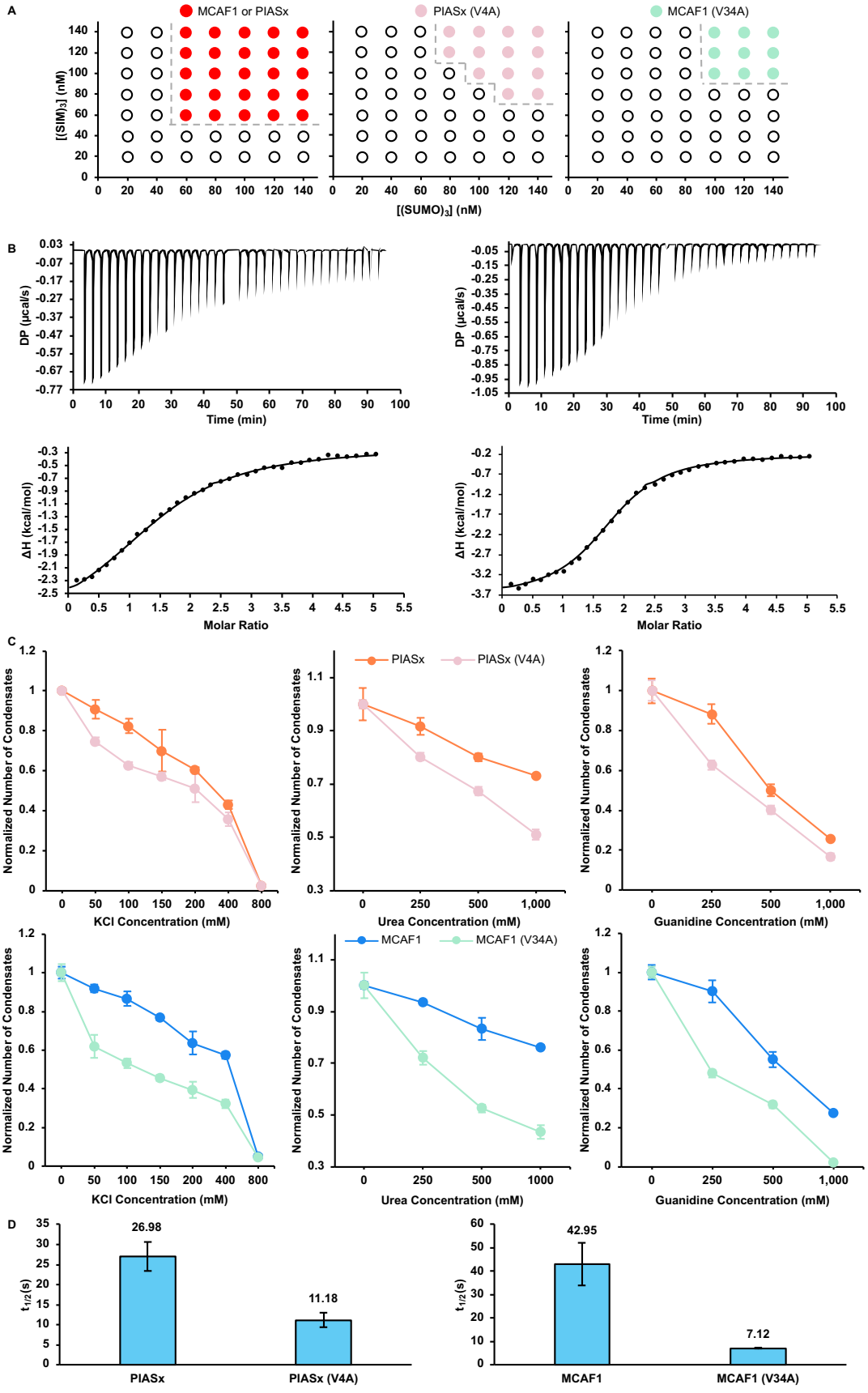

A
MCAF1 or PIASx
PIASx (V4A)
MCAF1 (V34A)
[(SIM)3] (nM)
[(SUMO)3] (nM)
B
DP (µcal/s)
Time (min)
ΔH (kcal/mol)
Molar Ratio
C
Normalized Number of Condensates
PIASx
PIASx (V4A)
KCl Concentration (mM)
Urea Concentration (mM)
Guanidine Concentration (mM)
MCAF1
MCAF1 (V34A)
D
t1/2 (s)
26.98
11.18
42.95
7.12
PIASx
PIASx (V4A)
MCAF1
MCAF1 (V34A)

Figure 4. Alanine mutations in SIM cores reduce SUMO3–SIM affinity, lower dissolution thresholds to salts and chaotropes, and shorten FRAP half-times. (**A**) Two-component phase diagrams for mEGFP-$(SUMO3)_5$ mixed with SIM arrays: mEGFP-$(PIASx)_3$ (left panel), mEGFP-$(PIASx(V4A))_3$ (middle panel), and mEGFP-$(MCAF1(V34A))_3$ (right panel). Open symbols indicate one phase and filled symbols indicate two phases; dashed curves mark apparent phase boundaries. Axes show scaffold concentrations. (**B**) Isothermal titration calorimetry of SUMO3 with PIASx(V4A) (left panel) and MCAF1(V34A) (right panel). Thermograms and integrated heats were fit to a 1:1 heteromeric association model; K_d values are reported in Table 3. (**C**) Stability of preformed condensates at equimolar scaffold levels. Top row, PIASx vs PIASx(V4A); bottom row, MCAF1 vs MCAF1(V34A). KCl titration (left panel), urea titration (middle panel), and guanidine-HCl titration (right panel). Wild-type droplets resist higher concentrations than alanine variants, with complete dissolution by ~400 mM KCl, ~1 M urea, and ~1 M guanidine-HCl. Error bars denote the standard error of the mean. (**D**) FRAP half-times for PIASx vs PIASx(V4A). (**E**) FRAP half-times for MCAF1 vs MCAF1(V34A). Bars or points represent mean ± standard error of the mean, indicating faster internal dynamics for the alanine variants.

4.2.5 Condensate stability is enhanced by electrostatic interactions with SUMO3

Having discovered that condensate stabilities correlate with scaffold-peptide affinities and that weaker binding peptides lead to less stable and more dynamic condensates, we sought additional strategies to control and modulate condensate properties. We were especially intrigued by the challenge of generating larger condensates in a predictable manner, because our results showed no clear correlation between SUMO3 affinities and the size of condensates formed. One reported strategy was to recruit protein cages to condensate interfaces as Pickering agents, where the adsorbed cages lower interfacial energy and slow coalescence, yielding larger apparent droplets without changing the sticker-spacer architecture.[38] Pickering agents are surface-active particles that stabilize liquid-liquid droplets by adsorbing to the interface and creating a steric barrier to fusion; in our context, protein cages would form an interfacial layer that reduces interfacial fluctuations and drainage while leaving scaffold valence unchanged.[38] We envisioned that another solution may lie in accessing additional binding sites on SUMO3 that could lead to greater interactions between condensates. Toward this, we were

attracted by the well-documented acidic patch on SUMO3 (**Figure 5A**) that has been proposed to participate in electrostatic interactions at low pH, when surface histidines are protonated, and by structural maps of acidic surface regions on SUMO3.[39] To preclude any pH dependence, we appended a constitutively charged segment, *N*-$(RK)_5$-*C*, to the C-terminus of mEGFP-$(PIASx)_3$ via a short linker so that the tag remains positively charged at physiological pH and can engage the acidic SUMO3 surface. At pH 7.5, the positive patch may associate with the negative patch of SUMO3, adding a second peptide binding mode to the hydrophobic SUMO3-SIM interaction, potentially leading to the additional nucleation of condensates. Consistent with this hypothesis, we observed greater numbers and sizes of condensates when the mEGFP-$(PIASx)_3$-$(RK)_5$ construct was mixed with mEGFP-$(SUMO3)_3$ (**Figure 5B**). Importantly, the mEGFP-$(PIASx)_3$-$(RK)_5$ construct did not phase separate by itself at the highest concentration tested. This argues that the enhancement requires engagement with SUMO3 rather than non-specific self-association of the basic tag.

To confirm that such dramatic changes were due to the $(RK)_5$ motif, we also tested a truncated half-sequence, mEGFP-$(PIASx)_3$-RKRKR. This shorter sequence generated more and larger condensates than mEGFP-$(PIASx)_3$ alone, but smaller and less than mEGFP-$(PIASx)_3$-$(RK)_5$ (**Figure 5B**). The graded response between RKRKR and $(RK)_5$ is consistent with a simple charge-density dependence that strengthens attraction to the acidic SUMO3 surface.[4,22] The presence of additional electrostatic interactions also rendered both sets of condensates more stable to salts or chaotropes and significantly less dynamic than condensates generated by SIM peptides alone (**Figures 5C-D, Supplementary Figure S8**). Higher concentrations of salt and denaturant were required to reduce condensate number when the basic tag was present, and FRAP half-times increased relative to PIASx-only condensates, in keeping with established salt and chaotrope effects on protein LLPS.[37] These observations are consistent with a bipartite

interaction model where the SIM-groove contact nucleates binding and the basic tag engages the acidic SUMO3 surface to reinforce the dense phase.

Initial attempts to measure binding by ITC with the extended 50-mer PIASx-$(RK)_5$ peptide, which was synthesized by native chemical ligation (NCL) due to its size (**Supplementary Figure S4.4.11.11- S4.4.11.13**), did not yield reliable isotherms in buffers where other SIM K_ds were measured due to insolubility of the peptide. However, at pH 6.5 in MES buffer, the PIASx-$(RK)_5$ peptide exhibited binding to SUMO3 while PIASx did not show any binding. Consistent with this observation, relative to PIASx alone, PIASx-$(RK)_5$ produced more and larger condensates with mEGFP-$(SUMO3)_5$ at lower protein concentrations. These BMCs were also more stable toward KCl, urea, and guanidine, and displayed longer FRAP half-times (**Figures 5B–D**). Taken together, these observations indicate the stronger effective engagement of SUMO3 when the basic tag is present, consistent with the principle that increasing effective interaction strength stabilizes the condensed phase.[2,10] Thus, by adding a secondary electrostatic component that increases overall binding affinity beyond the primary hydrophobic/van der Waals interactions between SUMO3 and its SIMs, it is possible to increase the size, stability, and rigidity of condensates in the polyvalent SUMO3-SIM system. Thus, we have identified electrostatic engagement of the SUMO3 acidic surface as a second interaction mode that enhances effective affinity without changing SUMO3:SIM ratios.[4,22]

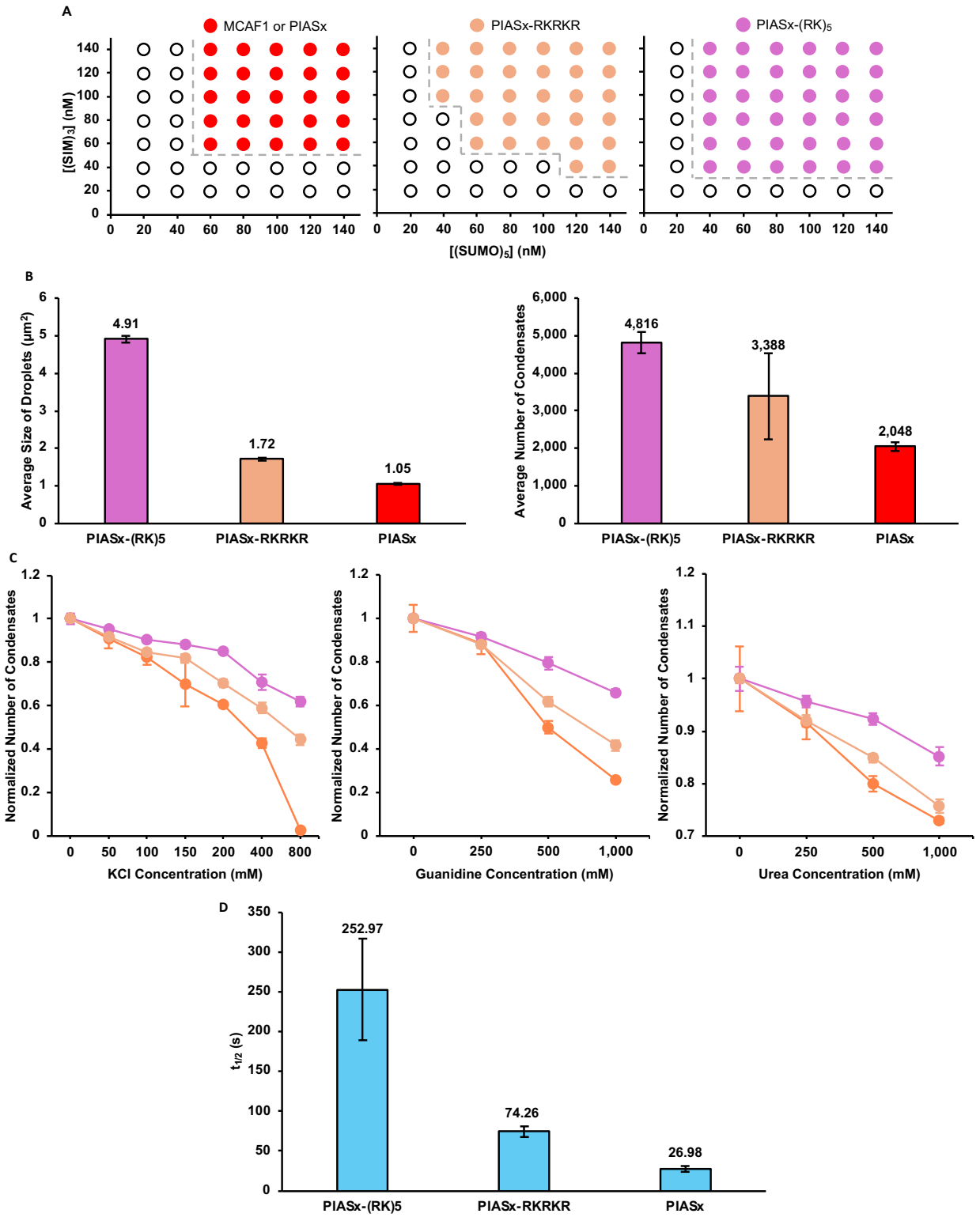

A
MCAF1 or PIASx
PIASx-RKRKR
PIASx-(RK)5
[(SIM)3] (nM)
[(SUMO)5] (nM)
B
Average Size of Droplets (μm2)
4.91
1.72
1.05
PIASx-(RK)5
PIASx-RKRKR
PIASx
Average Number of Condensates
4,816
3,388
2,048
C
Normalized Number of Condensates
KCl Concentration (mM)
Guanidine Concentration (mM)
Urea Concentration (mM)
D
t1/2 (s)
252.97
74.26
26.98

Figure 5. RK extensions on PIASx shift phase boundaries, enlarge condensates, increase stability, and lengthen FRAP half-times. (**A**) Two-component phase diagrams for mEGFP-$(SUMO3)_5$ mixed with SIM arrays: mEGFP-$(PIASx)_3$ (left panel), mEGFP-$(PIASx)_3$-RKRKR (middle panel), and mEGFP-$(PIASx)_3$-$(RK)_5$ (right panel). Open symbols indicate one phase and filled symbols indicate two phases; dashed curves mark apparent phase boundaries. Axes show scaffold concentrations. (**B**) Condensate size (left panel) and number (right panel) for PIASx, PIASx-RKRKR, and PIASx-$(RK)_5$. The RK extension yields larger and more numerous condensates on average compared to PIASx. (**C**) Stability of preformed condensates at equimolar scaffold levels for PIASx, PIASx-RKRKR, and PIASx-$(RK)_5$: KCl titration (left panel), urea titration (middle panel), and guanidine-HCl titration (right panel). RK extension variants resist higher concentrations than wild type, with complete dissolution by ~800 mM KCl, ~2 M urea, and ~2 M guanidine-HCl. Error bars denote the standard error of the mean. (**D**) FRAP half-times for PIASx, PIASx-RKRKR, and PIASx-$(RK)_5$. Points or bars represent mean ± standard error of the mean; RK extension variants exhibit slower internal dynamics.

4.2.6 The highest affinity scaffolding component dictates dynamics of complex biomolecular condensates

Advancing from binary SUMO3-SIM scaffold systems, we asked how ternary or quaternary complexes of scaffolding proteins may modulate condensate dynamics. This is especially relevant toward understanding the dynamic properties of complex coacervates formed in cells, where many interacting proteins are simultaneously present, such as in PML bodies. PML bodies are enriched for SUMO conjugates and SIM-bearing partners, and several SUMO–SIM interactions contribute to their assembly and turnover.[5,40] DAXX contains a SIM required for PML nuclear body localization, and MCAF1 carries a SIM that preferentially engages SUMO2/3, placing both within SUMO-rich nuclear assemblies.[41,42] Building upon trend observed in binary mixtures, namely, that higher SUMO3-SIM affinity slowed FRAP recovery, we asked whether BMC dynamics in mixtures would appear as an average of all the components, or, would be dominated by the scaffold with the strongest SUMO3 engagement.

We first generated condensates from the lowest-affinity mEGFP-$(SUMO3)_5$ and mEGFP-$(DAXXc)_3$ pair for 12 h, then added the tighter binding mEGFP-$(MCAF1)_3$ protein and incubated the condensates for an additional 12 h. This produced well-defined puncta on glass slides (**Figure 6A**). The order of addition allowed droplets to mature with the weak pair before introducing a tighter SUMO3 binder. Control condensates for comparison were generated with SUMO3 and either weak or strong SIM as a binary mixture. The ternary mixtures formed after 24 h exhibited dynamic properties in FRAP experiments that lay between the highly dynamic DAXXc and less dynamic MCAF1-mediated condensates (**Figure 6B**). This intermediate mobility in FRAP supports the idea of co-assembly of both scaffolds within a single network, rather than selective partitioning of one or other component.

Separation of these condensates from the bulk medium by centrifugation followed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis revealed the presence of both low and high affinity peptides in the ternary condensate (**Supplementary Figure S9**). This supports co-assembly of scaffolds within a single network rather than selective partitioning and supports the view that slower recovery reflects longer-lived SUMO3-SIM contacts within a shared network rather than a change in composition. A similar shift in FRAP to a less dynamic state was also observed when condensates initially generated with DAXXc were mixed with free mEGFP-$(PIASx)_3$-$(RK)_5$ (**Supplemental Figure S10**). This directional change matches the ternary result and was observed while puncta remained well-defined.

Consistent with the observed trends in ternary mixtures, condensates formed by complex quaternary mixtures containing mEGFP-$(SUMO3)_5$, mEGFP-$(DAXXc)_3$, mEGFP-$(MCAF1)_3$, and mEGFP-$(PIASx)_3$-$(RK)_5$ exhibited dynamics that most closely resembled that of the tightest SUMO3-binding scaffold, mEGFP-$(PIASx)_3$-$(RK)_5$ (**Figure 6C, Supplementary Figure S11**). FRAP half-lives in these quaternary mixtures approached those of mEGFP-$(PIASx)_3$-$(RK)_5$

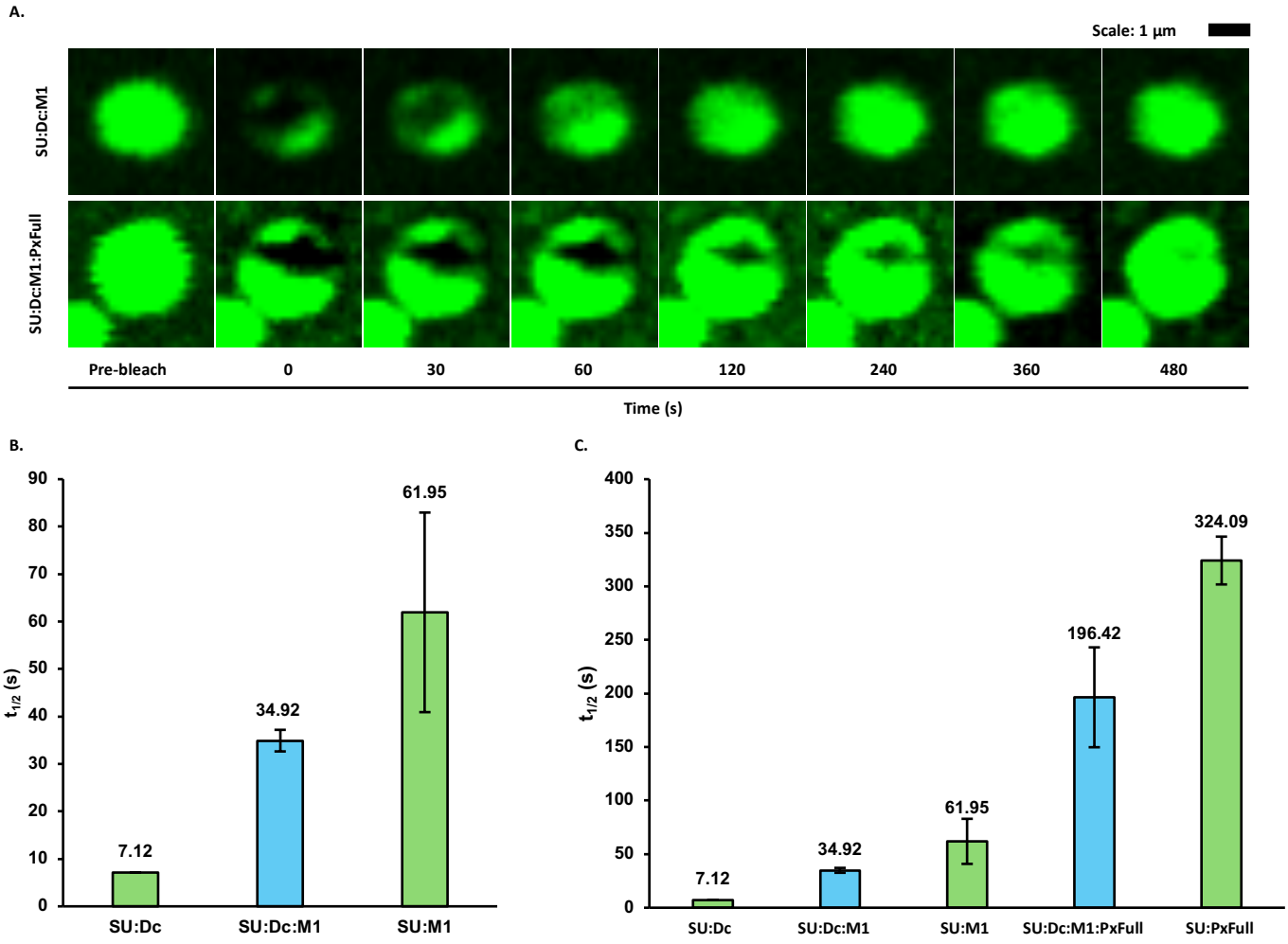


Figure 6. FRAP reveals composition- and affinity-dependent dynamics in binary, ternary, and quaternary SUMO assemblies. (**A**) Representative image of preformed SUMO:DAXXc condensates (left) and after addition of mEGFP-$(MCAF1)_3$ to generate the SUMO:DAXXc:MCAF1 trimeric mixture (right). Scale bars (black boxes), 10 µm. (**B**) FRAP half-times comparing binary mixtures SU:Dc and SU:M1 (green) with the trimeric SU:Dc:M1 assembly (blue). Points or bars represent mean ± standard error of the mean. (**C**) FRAP half-times comparing binary SU:Dc, SU:M1, and SU:PxFull mixtures (green) with trimeric SU:Dc:M1 and quaternary SU:Dc:M1:PxFull assemblies (blue). Points or bars represent mean ± standard error of the mean.

Abbreviations: SU, mEGFP-$(SUMO3)_5$; Dc, mEGFP-$(DAXXc)_3$; M1, mEGFP-$(MCAF1)_3$; PxFull, mEGFP-$(PIASx)_3$-$(RK)_5$.

rather than the arithmetic average of the individual components. Our results suggest that when multiple proteins in a condensate have similar binding modes, the tightest pair of interacting proteins contribute the most to the overall condensate dynamics. Thus, when considering SUMO-rich nuclear condensates, our data provides a mechanistic basis for understanding client exchange and dynamics in these multi-component assemblies.

4.2.7 Enzyme activity in condensates is modulated by scaffold affinities

Previous studies have demonstrated that condensates can enhance or limit enzymatic activity by modulating kinetic parameters such as K_m and k_{cat} for specific enzyme-substrate pairs. We wondered whether our observed differences in condensate mobility would lead to measurable differences in the activity of enzymes recruited to the dense phase. A systematic evaluation of enzyme function in closely related condensates that vary only in scaffold affinity, but have the same mode of binding and identical valency, has not been previously reported. Therefore, we used the mEGFP-$(SUMO3)_5$ and mEGFP-$(SIM)_3$ platform to create droplets of different mobilities and investigated how mobility influences enzyme function under identical conditions of temperature, buffer and substrate. We employed the SUMO:DAXXc and SUMO:PIASx condensates, due to their higher and lower mobility, respectively, at identical concentrations of SUMO and SIM proteins.

For the enzyme, we chose the cysteine protease pyroglutamyl peptidase 1, or PGP1. Pyroglutamyl peptidase is as an w-peptidase that hydrolyzes the amino terminal pyroglutamate (pGlu) residue from various pyroglutamylated substrate proteins.[43] Three forms of PGP have been identified in mammalian tissues. Type 1 is a monomeric, cytosolic, cysteine peptidase displaying broad substrate specificity and a low molecular mass of ~24 kDa, while Types 2 and 3 are membrane-associated high molecular weight enzymes that specifically act on Thyrotropin-releasing hormone. As a fluorogenic substrate for the reaction, we selected and synthesized a previously characterized probe (**Supplemental Figure 12**) that incorporates L-Pyroglutamic acid C-terminally conjugated with a hemicyanine fluorophore (pGlu-HCA).[44] Hydrolysis of pGlu-

HCA by PGP1 releases the fluorescent product HCA (**Supplemental Figure S12**), which we followed by its characteristic near-IR fluorescence emission at 700 nm during confocal microscopy.

To ensure recruitment of PGP1 into different condensates, we expressed and purified two trivalent SIM fusions, PGP1-$(PIASx)_3$ and PGP1-$(DAXXc)_3$, and then labeled them with Alexa Fluor 568 (AF568). AF568 was chosen to label the two enzyme constructs in order to measure their recruitment into condensates while preventing spectral overlap with mEGFP and HCA. To establish baseline kinetic parameters of the labeled enzymes, Michaelis-Menten steady-state kinetics were measured on a plate reader for each enzyme in solution using varying concentrations of the pGlu-HCA substrate. Reactions were monitored by HCA fluorescence at λ_{ex} 670 nm and λ_{em} 700 nm, and initial velocities were fit to the Michaelis–Menten equation to obtain K_m and k_{cat}. Under these conditions, AF568-PGP1-$(PIASx)_3$ exhibited K_m= 5.42 ± 2.81 μM and k_{cat}= 48.29 ± 19.82 s^{-1}, whereas AF568-PGP1-$(DAXXc)_3$ had K_m= 2.55 ± 1.64 μM and k_{cat} = 11.17 ± 4.01 s^{-1}. Thus, addition of the C-terminal SIM extensions and non-specific amine-labeling by AF568 had small measurable effects on the overall activities of the two PGP1 forms, however, their catalytic efficiencies were very similar. For AF568-PGP1-$(PIASx)_3$ k_{cat}/K_m was 8.91 ± 5.89 $\mu M^{-1}s^{-1}$ and for AF568-PGP1-$(DAXXc)_3$ k_{cat}/K_m was 4.38 ± 3.23 $\mu M^{-1}s^{-1}$.

With the active and well-characterized enzymes in hand, we next measured their catalytic activities when recruited to condensates formed under identical conditions. Toward this, confocal microscopy was used to follow pGlu-HCA hydrolysis in SUMO+PIASx and SUMO+DAXXc droplets. These droplets were assembled at 2 μM SUMO3 and 2 μM of the respective SIM construct. Condensates formed after 16 h of incubation were treated with 50 nM AF568-PGP1-$(PIASx)_3$ or 50 nM AF568-PGP1-$(DAXXc)_3$ and incubated for 1 h to allow for enzyme recruitment to condensates. The presence of enzymes in condensates was then

confirmed by AF568 fluorescence at 603 nm. Finally, pGlu-HCA was added to 2 μM and HCA fluorescence was recorded at λ_{ex} 670 nm with a detector window of 710-799 nm. The encapsulation efficiency for pGlu-HCA was 89.55 ± 0.72% in SUMO+PIASx condensates and 88.38 ± 0.54% in SUMO+DAXXc condensates, indicating that initial concentrations of the probe were near identical in both types of condensates. Droplet and dilute-phase regions in the same wide-field image were segmented, background fluorescence from the dilute-phase was subtracted, and the early linear portion of each trace was used to measure initial rates of hydrolysis. Substrate-only and condensate-only controls showed no increase in fluorescence signal at 700 nm over a period of 2 h. Under the identical 2 mM substrate and 50 nM enzyme concentrations, the product HCA fluorescence was consistently visible in SUMO+DAXXc droplets well before its appearance in SUMO+PIASx droplets (**Figure 7B-C**). Thus, confocal microscopy with condensates of differing mobility clearly indicated that more mobile SUMO+DAXXc condensates led to faster enzyme activity than the less mobile SUMO+PIASx condensates.

As recruitment of additional proteins may inadvertently change the properties of a condensate, we also undertook FRAP experiments with condensates after PGP1 recruitment. A circular region accounting for ~50% of the visible droplet area was bleached in each condensate, and the recovery of fluorescence was visualized by confocal microscopy. The percentage of fluorescence recovered as a function of time was fit to a single-exponential model to obtain the half-time ($t_{1/2}$) for FRAP. Interestingly, enzyme recruitment slowed the internal dynamics in both sets of condensates by about 3.3-fold. However, the relative order of mobility remained unchanged with SUMO+PIASx+PGP1 condensates recovering significantly slower than SUMO+DAXXc+PGP1 condensates (**Supplementary Figure S13**). These control FRAP experiments confirmed that the initial mobility differences in condesnates persisted under the conditions used for imaging PGP1 activity.

To quantitate the differences in enzyme activities observed by confocal microscopy, we measured initial rates in bulk solutions of SUMO+PIASx or SUMO+DAXXc condensates containing PGP1. Assays were conducted in 96-well plates by adding pre-formed condensates to each well followed by the addition of 50 nM of AF568-PGP1-$(PIASx)_3$ or AF568-PGP1-$(DAXXc)_3$. Reactions were started by the addition of 4 μM pGlu-HCA to each well and

hydrolyzed HCA fluorescence was observed at 700 nm. Initial rates were calculated from the first five minutes of activity and averaged across three wells. As expected, substrate-only and condensate-only controls showed no detectable fluorescence over 2 h (**Figure 7A**). Relative to the initial rate of free AF568-PGP1-$(DAXXc)_3$, SUMO+DAXXc+AF568-PGP1-$(DAXXc)_3$

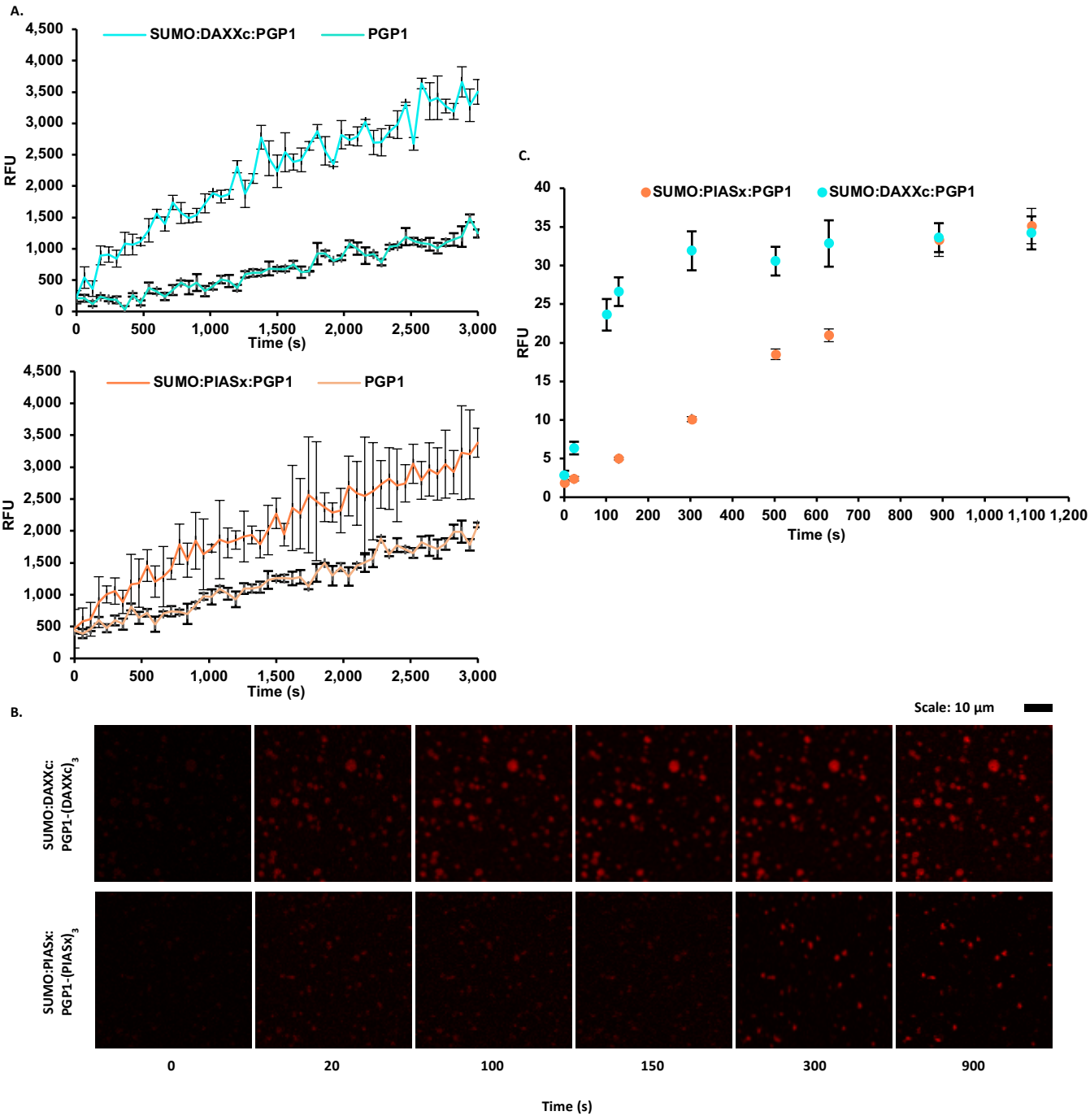


Figure 7. Scaffold affinity sets condensate mobility and tunes PGP1 catalysis in SUMO3:SIM droplets. (**A**) Bulk kinetics by plate reader at identical conditions. Top panel, AF568-PGP1-(DAXXc)3 in solution vesus the matching SUMO-DAXXc condensates. Bottom panel, AF568-PGP1-(PIASx)3 in solution versus the matching SUMO-PIASx condensates. Initial rates were extracted over the first five minutes and averaged across three wells. Rates are mean ± SEM and show ~5.1-fold enhancement in SUMO-DAXXc and ~2.7-fold in SUMO-PIASx relative to enzyme in solution. (**B**) Confocal images of preformed SUMO-PIASx and SUMO-DAXXc droplets after recruitment of the matching AF568-PGP1 fusion and addition of pGlu-HCA. Product HCA fluorescence was monitored at λ_{ex} 670 nm with a 710-799 nm detector window. Scale bars, 10 μm. (**C**) Time courses of HCA signal within segmented droplets (20 droplets) under identical conditions. Product signal appears sooner in SUMO-DAXXc droplets than in SUMO-PIASx droplets.

condensates exhibited an apparent rate increase of ~5.1-fold. In comparison, SUMO+PIASx+ AF568-PGP1-$(PIASx)_3$ condensates gave a smaller enhancement of ~2.7-fold over free AF568-PGP1-$(PIASx)_3$ in solution. Collectively our measurements in single condensates and in the bulk solution revealed that recruitment to condensates elevates PGP1 activity. Furthermore, the increase in PGP1 activity correlates well with increased mobility of the condensates, which suggests that condensate mobility may be used to control enzyme activity in cells.

4.2.8 Condensate dynamics in cells reflect SUMO-SIM affinities

Having established the strong correlation between binding affinity and condensate dynamics *in vitro*, we sought to test if this correlation is recapitulated in living cells. The ability to control condensate properties, such as composition and dynamics, in living cells is highly desirable for efforts to engineer metabolic pathways, to isolate disease-associated enzymes, and to regulate gene transcription. In the eukaryotic cell nucleus, SUMO conjugates and SIM-bearing client proteins concentrate in PML nuclear bodies, providing an important cellular context in which SUMO-SIM contacts contribute to condensate assembly and turnover.[40] Rosen and coworkers previously demonstrated that linear fusions of poly-SUMO3-poly-SIM repeats may phase separate in cells by overcoming limitations imposed by the mismatched expression levels of separate polySUMO3 and polySIM genes.[1] Therefore, we generated a series of linear fusion

constructs containing the monomeric Cherry fluorescent protein (mCherry), five SUMO3 and three SIM repeats, namely mCherry-$(SUMO)_5$-$(SIM)_3$. In this linear construct, the SIM component could be readily changed to cover a broad range of K_d values. The mCherry fluorescent protein enabled condensate visualization at a longer excitation wavelength of 587 nm that is less damaging to cells than the lower wavelengths used for mEGFP visualization. Three SIMs with distinct affinities for SUMO3, PIASx ~8 μM, SIM2 ~60 μM, and DAXXc ~70 μM were chosen to generate the mCherry-$(SUMO3)_5$-$(SIM)_3$ fusions. HeLa cells were separately transfected with each fusion construct and imaged for condensate formation after 48 h. Consistent with their behavior *in vitro*, all constructs formed distinct puncta in cells (**Figure 8A**). The puncta were similar in size, with most measurements clustered at ~1–6 μm^2 when measured as the maximal in-plane cross-sectional area for each droplet. Consistent with our observations *in vitro*, there was no correlation between K_d and size. Surprisingly, in contrast with sumoylated PML bodies that are formed in the nucleus, our engineered puncta were localized to the cytoplasm,.[40] The fluidity of these cytoplasmic puncta formed in HeLa cells was measured by FRAP in an area corresponding to ~50% of the droplet area. In keeping with the trends observed *in vitro*, condensates formed by mCherry-$(SUMO3)_5$-$(PIASx)_3$ recovered slower than mCherry-$(SUMO3)_5$-$(SIM2)_3$ and the latter recovered slower than mCherry-$(SUMO3)_5$-$(DAXXc)_3$ (**Figure 8B**). The fact that relative dynamics observed *in vitro* within condensates made from mEGFP labeled scaffolds present *in trans* are also recapitulated in cells with mCherry labeled scaffolds fused *in cis,* indicates that K_d clearly drives condensate dynamics for polyvalent protein-protein interactions. We further confirmed that the observed trends in mobilities were not limited to HeLa cells alone by repeating the experiments in HEK293T and H1299 cells (**Supplementary Figure S14**).

We next asked if addition of the basic $(RK)_5$ appendage to the linear fusion construct would have a similar effect in cells as we observed *in vitro*. When HeLa cells were transfected with the

mCherry-$(SUMO3)_5$-$(PIASx)_3$-$(RK)_5$ construct, large puncta were observed within the cell nucleus with much slower dynamics than mCherry-$(SUMO3)_5$-$(PIASx)_3$ alone (**Figure 8B and 8C**). The nuclear localization of mCherry-$(SUMO3)_5$-$(PIASx)_3$-$(RK)_5$ is not surprising because polyarginine sequences can promote nuclear import via classical basic nuclear localization motifs recognized by importin-α/β.[45,46] Nevertheless, the localization of the $(RK)_5$ tag in the nucleus led us to ask if reducing the number of arginines may potentially lead to cytoplasmic localization. Truncating the RK repeats in mCherry-$(SUMO3)_5$-$(PIASx)_3$-RKRKR, was indeed sufficient to generate condensates in the cytoplasm that exhibited slower dynamics than mCherry-$(SUMO3)_5$-$(PIASx)_3$ alone (**Figure 8B and 8D**). Thus, our cellular measurements revealed that modulating scaffold affinities by changing SIM sequences and by including additional sites of interaction in SUMO enable regulating condensate mobility in both the cytoplasm and nucleus.

In summary, the genetically fused mCherry-$(SUMO3)_5$–$(SIM)_3$ platform fully recapitulates the affinity-mobility relationship observed *in vitro*. The same hierarchy of FRAP half-lives were conserved across three different cell lines, HeLa, HEK293T, and H1299 cells, and in both the cytoplasmic and nuclear spaces. The localization of condensates could be readily tuned by introducing the short basic $(RK)_5$ segment rather than by any changes to the SUMO3–SIM interface. As these FRAP measurements were performed in living cells, the recovery half-lives report intra-condensate dynamics within the crowded cellular milieu and further solidify the affinity–mobility relationship discovered in our *in vitro* platform.[47]

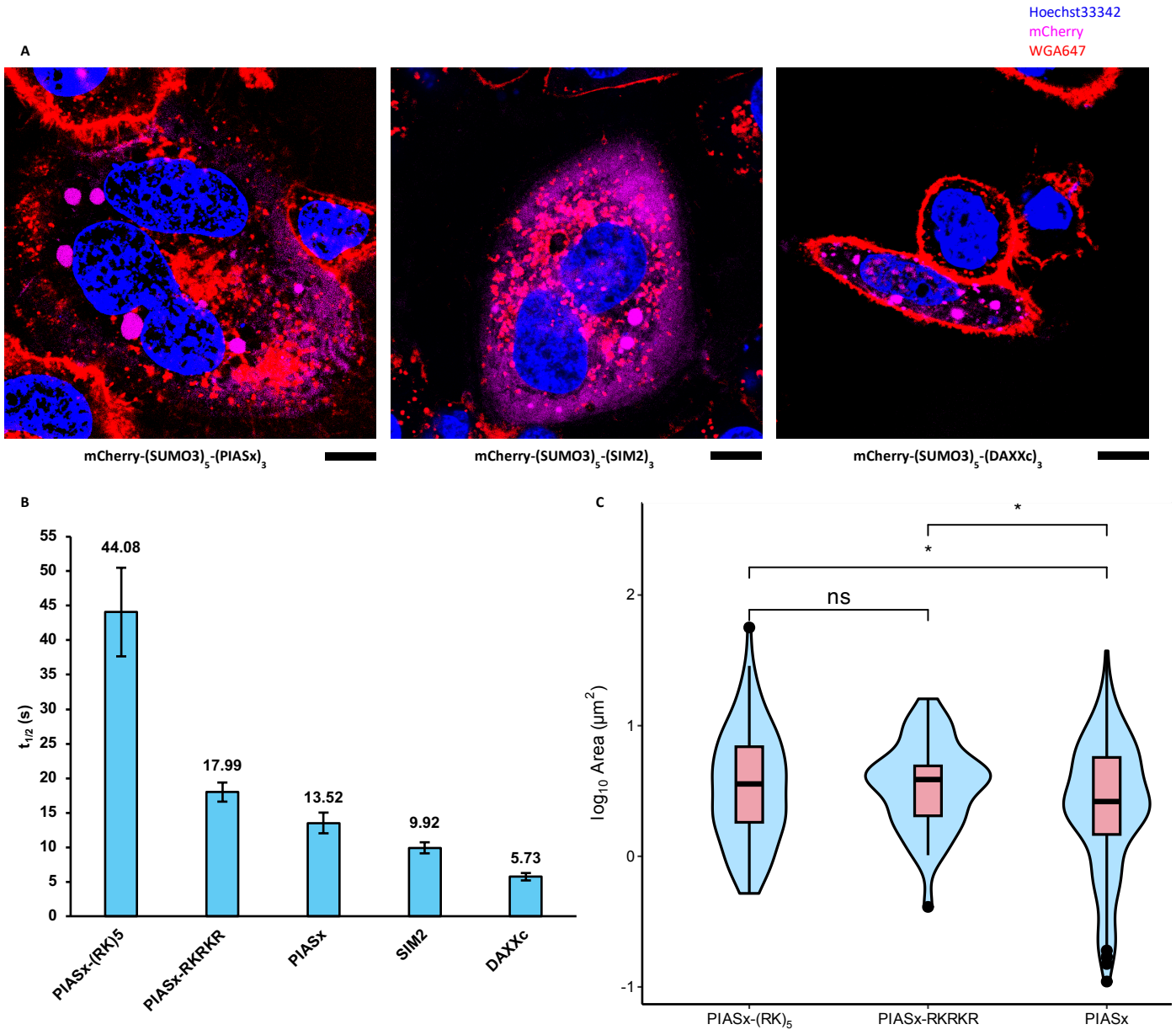


Figure 8. Cellular SUMO3–SIM condensates show cytoplasmic localization and affinity-dependent FRAP. (**A**) Representative fluorescence overlays of HeLa cells expressing mCherry-$(SUMO3)_5$-$(PIASx)_3$ (left panel), mCherry-$(SUMO3)_5$-$(SIM2)_3$ (middle panel), or mCherry-$(SUMO3)_5$-$(DAXXc)_3$ (right panel). Blue corresponds to the nuclear stain Hoechst 33342, magenta corresponds to mCherry, and red corresponds to the membrane stain WGA647. Condensates are predominantly cytoplasmic. Scale bars (black boxes), 10 μm. (**B**) FRAP half-times in cells for the indicated constructs. As observed in vitro, stronger SUMO:SIM interactions correspond to slower recovery. Each point represents one condensate; bars denote mean ± standard error of the mean. (**C**) Condensate size in cells for PIASx and its RK-extension variants (RKRKR and $(RK)_5$). RK extensions yield larger condensates on average than PIASx. Points or bars denote mean ± standard error of the mean. Significance notation: ns, not significant; *, $p < 0.05$

4.3 Conclusions and outlook

By employing a polyvalent SUMO3-SIM-based system that phase separates at physiologically relevant concentrations of SUMO3, we sought to learn the properties of coacervates best predicted by the binding affinities, or K_d, of scaffold proteins. By varying SIM affinity for SUMO, while holding scaffold valence and SUMO architecture constant, we isolated the single parameter, K_d, and measured its direct effect on condensate properties. Our results clearly revealed that scaffold K_d alone is a poor predictor of condensate size or number. However, condensate stability and dynamics are both closely related to the K_d of scaffold proteins. In both salt- and chaotrope-induced dissolution of condensates and FRAP experiments *in vitro* and in living cells, tighter SUMO3-SIM interactions lowered the c_{sat} and condensate mobility. This strongly indicated that scaffold interaction lifetime sets the pace of network relaxation under our conditions, a relationship anticipated by polymer-network models that link condensate relaxation to sticker lifetimes and connectivity.[48] While the entire sequence of amino acids in the SUMO-interacting motif (SIM) peptide scaffold poorly correlates with condensate properties, 3-to-4 amino acids in the SIM hydrophobic core that bind a groove in SUMO3 formed by its only α-helix and second β2-strand strongly dictate condensate properties. Hence, single Ala mutations of Val residues in the hydrophobic cores of PIASx and MCAF1 had significant detrimental effects on both condensate stability and dynamics. The conservative V-to-A mutations recapitulated the effects of K_d-based trends seen with different SIM sequences without changing SIM length, charge, linker spacing, or valence, which points to binding free energy at the SUMO groove as the dominant contributor to condensate properties.[22] Furthermore, by identifying a secondary electrostatic interaction between the positively charged peptide $(RK)_n$ peptide and

the acidic surface of SUMO3, we demonstrated that condensate size, stability, and dynamics may be further modulated in a predictable manner both *in vitro* and in living cells. Short basic extensions also provided a practical handle for subcellular localization while preserving the affinity-driven hierarchy of condensate mobilities. To understand how condensate mobility may directly influence important cellular functions such as the stress response, we measured activity of the pyroglutamyl peptidase type 1 enzyme (PGP1) recruited to condensates using condensate-matching SIM repeats. By confocal microscopy we observed that condensates that recovered fluorescence faster in FRAP experiments also exhibited higher initial rates of product-formation rates than droplets with slower fluorescence recovery in FRAP. This indicates that the scaffold-dependent exchange timescale can also influence reaction flux in puncta of similar composition and size.

Membraneless organelles in cells, such as proteasome storage bodies and stress granules, contain many proteins that either act as scaffolds to favor phase separation, or as clients recruited to the separated phase. Our studies with mixtures of scaffolding SIM peptides with varying affinities for SUMO revealed that the dynamics of increasingly complex condensates containing ternary and quaternary mixtures of scaffolds are dictated by the dynamics of the tightest SUMO-SIM interaction. Thus, condensates become increasingly less mobile as tighter binding SIMs are present in scaffold mixtures**.** This suggests that the overall properties of any given membraneless organelles is likely reflective of the tightest binding scaffolding proteins, which warrants further studies in cells. Finally, toward engineering well-defined and predictable condensates in living cells, we demonstrated that the polySUMO and polySIM scaffolds may be fused *in cis* to ensure the expression of equivalent ratios of the two, that trends in mobility of condensates formed by in cis fusions also correlate with SUMO-SIM K_d. Altogether, the results of our combined experiments in vtiro and in living cells help us define a practical design rule for

condensate formation that is based on a single readily measurable and changeable parameter, the affinity of scaffold interactions.

In the immediate future, we are working toward creating a truly orthogonal SUMO-SIM scaffolding pair that shows minimal binding to endogenous SUMO isoforms and SIM-containing proteins in cells. This will enable engineering well-defined condensates that do not recruit SUMO or SIM binding proteins and are compositionally homogeneous. One reasonable path toward our goals is to re-sculpt the conserved SIM-binding groove by mutating wild-type SUMO3 (mutSUMO) and to then identify a mutant SIM (mutSIM) peptide that binds the engineered SUMO3.[22] This biologically orthogonal mutSUMO-mutSIM pair would permit identical cell-based studies of condensate mobility in a more homogeneous background. By incorporating the inducible control of local concentration or valence, for example using light-activated optoDroplet recruitment in living cells,[49] the same mutSUMO-mutSIM framework could be employed to test how *in vitro* measured interaction lifetimes set recovery on defined timescales in cytoplasmic or nuclear environments. Additional experiments include estimating on- and off-rates for individual scaffolds in cells from calibrated FRAP or tracking analyses, comparing SUMO1 versus SUMO2 or SUMO3 selectivity within a shared scaffold architecture, and measuring how client load or enzymatic activity influences condensate properties across different crowding regimes. We anticipate that affinity-tuned and orthogonal scaffolds combined with localization tags will yield programmable condensates whose turnover, composition, and cellular placement can be adjusted independently, opening a route to controlled catalysis and signaling in living systems.

4.4 Experimental procedures

4.4.1 General methods

DNA synthesis was performed by Integrated DNA Technologies (Coralville, IA), and Sanger sequencing was performed by Eurofins Genomics (Louisville, KY). Plasmid miniprep, PCR purification, and gel extraction kits were from Qiagen (Germantown, MD). HisPur Ni-NTA resin was from Thermo Scientific (Waltham, MA). HeLa, HEK293T, and H1299 cells were cultured in T75 flasks or 8-well μ-Slides (Ibidi, Gräfelfing, Germany) in Dulbecco's Modified Eagle Medium (DMEM; HeLa and HEK293T) or Roswell Park Memorial Institute Medium (RPMI 1640; H1299) supplemented with 10% fetal bovine serum at 37 °C in a humidified incubator with 5% CO_2. Routine confocal microscopy was performed on a Leica SP8 X scanning confocal microscope using Leica LAS X. PCRs used Phusion High-Fidelity DNA Polymerase in HF buffer (Thermo Scientific, Waltham, MA) unless otherwise noted. All buffers were prepared with 18.2 MΩ·cm water. Protein concentrations were determined either by SDS-PAGE with Coomassie staining relative to bovine serum albumin standards or, for constructs bearing mEGFP, by absorbance at 488 nm using ε_{488} = 56,000 M^{-1} cm^{-1}. Peptides were synthesized by Fmoc solid-phase peptide synthesis on CEM Liberty Blue 1.0 and CEM Liberty PRIME 2.0 microwave synthesizers from CEM Corporation (Matthews, NC), using the resins and protocols specified in each subsection. HPLC: Reverse-phase chromatography used Buffer A (0.1% TFA in H_2O) and Buffer B (90% acetonitrile, 9.9% H_2O, 0.1% TFA) on Agilent 1260 Infinity II systems from Agilent Technologies (Santa Clara, CA), with C18 or C4 columns as specified; analytical runs were typically 0–73% B over 30 min at 1.0 mL min^{-1}, and semi-preparative or preparative runs used the indicated C18 or C4 columns at 4–20 mL min^{-1} with UV detection at 214 and 280 nm. Where Waters columns were used, they were sourced from Waters Corporation (Milford, MA). Mass spectrometry: Routine peptide and protein masses were measured by direct-infusion electrospray ionization on a Bruker Esquire ion-trap from Bruker (Billerica, MA) or a Finnigan LTQ ion-trap from Thermo Fisher Scientific (Waltham, MA) operated in positive-ion mode.

4.4.2 Molecular cloning of DNA and sequencing

All PCR amplifications and site-directed mutagenesis used Phusion polymerase (Phusion High-Fidelity PCR Master Mix with HF Buffer, Thermo Scientific).

The plasmid pTEV19-His_{10}-mEGFP-$(SUMO3)_3$ containing three repeats of the human SUMO3 gene separated by Gly- and Ser-rich flexible linkers was purchased from Addgene (plasmid #127093). A stop codon was introduced by PCR-based site-directed mutagenesis using primers listed in (Table 4.4.2) to terminate translation immediately after the final SUMO3 repeat, preventing expression of the downstream TEV site. The modified pTEV19 plasmid containing five repeats of SUMO3 with intervening Gly- and Ser-rich linkers, pTEV19-His10-mEGFP-$(SUMO3)_5$, was generated by restriction-digest cloning on the pTEV19-His_{10}-mEGFP-$(SUMO3)_3$ plasmid as the vector source, using AatII and XhoI restriction sites (New England Biolabs (NEB), Ipswich, MA). The cut vector was treated with calf intestinal phosphatase to prevent self-ligation (NEB, Ipswich, MA) and purified by gel excision from agarose using the QIAquick PCR & Gel Cleanup Kit (Qiagen, Germantown, MD). The same restriction enzymes were used to cut a synthetic gene fragment encoding two additional SUMO3 repeats with Gly- and Ser-rich linkers and flanked by 5' AatII and 3' XhoI restriction sites (Table 4.4.1). The digested insert and dephosphorylated vector were ligated with T4 DNA ligase (NEB, Ipswich, MA).

All pTEV19-His_{10}-mEGFP-$(SIM)_3$ constructs used *in vitro* were generated by restriction-digest cloning of the plasmid GFP-$(SIM)_3$ purchased from Addgene (plasmid #126955). The vector was digested with BsrGI and XhoI and treated with calf intestinal phosphatase to prevent self-ligation. The same restriction enzymes were used to cut a synthetic gene fragment corresponding to each SIM peptide, separated by Gly- and Ser-rich linkers and flanked by 5' BsrGI and 3' XhoI restriction sites (Table 4.4.1). The digested insert and dephosphorylated vector were ligated with T4 DNA ligase.

Two PIASx variants were prepared by PCR-based site-directed mutagenesis. For mEGFP-$(PIASx)_3$, a stop codon was introduced using primers listed in (Table 4.4.2) to terminate translation immediately after the final PIASx repeat. For mEGFP-$(PIASx)_3$-RKRKR, a second stop codon was introduced further downstream using primers listed in (Table 4.4.2). The mEGFP-$(MCAF1)_3$ (V34A) construct was synthesized and purified by Genscript.

To generate PGP1 fusion constructs in pTEV19, the host vector pTEV19-His_{10}-mEGFP-$(SUMO3)_5$, synthesized and purified by GenScript, was linearized by PCR amplification using Phusion polymerase with primers listed in (Table 4.4.2) to remove the mEGFP-$(SUMO3)_5$ cassette and to introduce 20–40 bp overlaps compatible with HiFi assembly for insertion of either the $(PIASx)_3$ or $(DAXXc)_3$ fragments. The $(PIASx)_3$ and $(DAXXc)_3$ inserts were generated by PCR amplification from pTEV19-His_{10}-mEGFP-$(PIASx)_3$ and pTEV19-His_{10}-mEGFP-$(DAXXc)_3$, respectively, using Phusion polymerase and primers listed in (Table 4.4.2) designed to provide sequence overlap with the linearized host vector. Assembly reactions were performed using NEBuilder HiFi DNA Assembly Master Mix (NEB, Ipswich, MA) at an ~2:1 insert:vector molar ratio and incubated at 50 °C for 60 min according to the manufacturer's guidelines.

For mammalian expression, the plasmid RFP-$(SUMO)_6$-$(SIM)_{10}$ (pC1-mCherry-$(SUMO3)_6$-$(PIASx)_{10}$) was purchased from Addgene (plasmid #122027) and served as a positive control for condensate formation in cells. pC1-mCherry-$(SUMO3)_5$-$(PIASx)_3$, pC1-mCherry-$(SUMO3)_5$-$(SIM2)_3$, pC1-mCherry-$(SUMO3)_5$-$(DAXXc)_3$, and pC1-mCherry-$(SUMO3)_5$-$(PIASx)_3$-$(RK)_5$ in-cis fusion constructs genes utilized for transient transfection in human cells were custom synthesized, purified, and cloned into the host pC1-mCherry-$(SUMO3)_6$-$(PIASx)_{10}$ plasmid between restriction sites BsrGI and SalI by GenScript (Piscataway, NJ). The plasmid pC1-mCherry-$(SUMO3)_5$-$(PIASx)_3$-$(RK)_5$ was used to generate the plasmid pC1-mCherry-$(SUMO3)_5$-

$(PIASx)_3$-RKRKR. Using the primers below (Table 4.4.2), the portion of the gene encoding for the $(RK)_5$ extension was truncated to RKRKR through exclusion of them via PCR.

Assembled or ligated products were transformed into XL10-Gold ultracompetent E. coli (Agilent Technologies, Santa Clara, CA). Constructs were screened by agarose gel electrophoresis and verified by DNA sequencing.

Table 4.4.1 List of synthetic gene fragments used for molecular cloning.

Synthetic Gene Fragments	DNA Sequence (5'- to -3')
3xSIM2_BsrGI_XhoI	CTC GGC ATG GAC GAG CTG TAC AAG TCC GGA GGT AGT GGC GGA AGC GGT GGC TCT TGG GGT GGG AGC CGC GAA ACC GCC GGG GAT GAG ATC GTG GAT TTA ACT TGC GAG TCA TTG GAG CCT CGC TGG GGC GGT AGC GGT GGC TCT GGC GGG AGC GGC GGC TCG CGT GAA ACA GCA GGT GAT GAA ATT GTC GAT CTG ACG TGC GAA TCC CTG GAG CCA CGT TGG GGC GGC TCG GGA GGT TCG GGC GGC AGC GGC GGA AGC CGC GAG ACT GCA GGG GAT GAA ATT GTG GAT CTT ACA TGT GAG AGT TTG GAA CCA CGT TGG TAA TAA GGG GGC TCA TGC GGG GGT TCT CTC GAG AAC CTG TAT TTT CAG GGC
3xDAXXn_BsrGI_XhoI	CTC GGC ATG GAC GAG CTG TAC AAG TCC GGA GGT AGT GGC GGA AGC GGT GGC TCT TGG GGT GGG AGC ATG GCT ACT GCC AAT TCT ATC ATC GTA TTA GAC GAT GAC GAT GAA GAC GAA GGC GGT AGC GGT GGC TCT GGC GGG AGC GGC GGC TCG ATG GCG ACG GCT AAC TCG ATT ATC GTC TTG GAT GAT GAC GAC GAA GAT GAA GGC GGC TCG GGA GGT TCG GGC GGC AGC GGC GGA AGC ATG GCA ACC GCT AAT TCC ATT ATT GTT TTG GAC GAC GAT GAC GAG GAT GAA TAA TAA GGG GGC TCA TGC GGG GGT TCT CTC GAG AAC CTG TAT TTT CAG GGC
3xDAXXc_BsrGI_XhoI	CTC GGC ATG GAC GAG CTG TAC AAG TCC GGA GGT AGT GGC GGA AGC GGT GGC TCT TGG GGT GGG AGC AAA ACG AGT GTT GCG ACC CAA TGC GAT CCG GAA GAG ATC ATC GTT CTG AGC GAT AGT GAT GGC GGT AGC GGT GGC TCT GGC GGG AGC GGC GGC TCG AAG ACC AGC GTG GCC ACC

	CAG TGC GAT CCA GAA GAA ATT ATT GTG TTA AGT GAT AGC GAC GGC GGC TCG GGA GGT TCG GGC GGC AGC GGC GGA AGC AAA ACC AGT GTT GCG ACT CAG TGT GAC CCA GAG GAG ATT ATC GTC CTG TCG GAT AGC GAC TAA TAA GGG GGC TCA TGC GGG GGT TCT CTC GAG AAC CTG TAT TTT CAG GGC
3xMCAF1_BsrGI_XhoI	CTC GGC ATG GAC GAG CTG TAC AAG TCC GGA GGT AGT GGC GGA AGC GGT GGC TCT TGG GGT GGG AGC GGC TCC CCT GAG TTT AAG ACG ATT GAT GCC AGT GTG AGT AAG AAA GCC GCT GAT TCC ACT TCT CAA TGT GGC AAA GCC ACG GGC TCG GAC AGC TCT GGG GTG ATT GAT TTG ACT ATG GAT GAC GAA GAA TCT GGG GCA TCA CAA GAT GGC GGT AGC GGT GGC TCT GGC GGG AGC GGC GGC TCG GGC AGC CCG GAG TTC AAA ACC ATT GAT GCA TCC GTT TCG AAG AAG GCG GCT GAT TCA ACG TCC CAA TGC GGC AAG GCT ACG GGT TCC GAT AGT AGC GGT GTG ATT GAC CTT ACA ATG GAT GAT GAA GAA AGC GGT GCG TCA CAG GAC GGC GGC TCG GGA GGT TCG GGC GGC AGC GGC GGA AGC GGT AGC CCG GAA TTC AAA ACG ATC GAT GCC TCC GTA AGT AAA AAA GCT GCT GAC AGC ACA AGC CAG TGT GGC AAA GCT ACA GGA AGC GAC AGT TCA GGG GTA ATT GAC TTA ACA ATG GAC GAC GAG GAG AGT GGT GCG TCT CAA GAT TAA TAA GGG GGC TCA TGC GGG GGT TCT CTC GAG AAC CTG TAT TTT CAG GGC
3xPIASx(V4A)_BsrGI_XhoI	CTC GGC ATG GAC GAG CTG TAC AAG TCC GGA GGT AGT GGC GGA AGC GGT GGC TCT TGG GGT GGG AGC AAA GTA GAT GCG ATC GAC TTG ACC ATC GAA AGC AGC AGC GAC GAA GAA GAA GAT CCG CCT GCC AAA CGC GGC GGT AGC GGT GGC TCT GGC GGG AGC GGC GGC TCG AAA GTT GAT GCC ATT GAC CTT ACA ATT GAA TCT TCG AGC GAT GAA GAG GAA GAT CCT CCA GCG AAA CGT GGC GGC TCG GGA GGT TCG GGC GGC AGC GGC GGA AGC AAA GTC GAT GCA ATT GAC TTA ACG ATC GAA TCT AGC AGC GAT GAA GAA GAA GAT CCA CCG GCT AAA CGT TAA TAA GGG GGC TCA TGC GGG GGT TCT CTC GAG AAC CTG TAT TTT CAG GGC
3xPIASx(I5A)_BsrGI_XhoI	CTC GGC ATG GAC GAG CTG TAC AAG TCC GGA GGT AGT GGC GGA AGC GGT GGC TCT TGG GGT GGG AGC AAA GTA GAT GTG GCG GAC TTG ACC ATC GAA AGC AGC AGC GAC GAA GAA GAA GAT CCG CCT GCC AAA CGC GGC GGT AGC GGT GGC TCT GGC GGG AGC GGC GGC TCG AAA GTT GAT GTC GCC GAC CTT ACA ATT GAA TCT TCG AGC

	GAT GAA GAG GAA GAT CCT CCA GCG AAA CGT GGC GGC TCG GGA GGT TCG GGC GGC AGC GGC GGA AGC AAA GTC GAT GTA GCA GAC TTA ACG ATC GAA TCT AGC AGC GAT GAA GAA GAA GAT CCA CCG GCT AAA CGT TAA TAA GGG GGC TCA TGC GGG GGT TCT CTC GAG AAC CTG TAT TTT CAG GGC
3xPIASx(L7A)_BsrGI_XhoI	CTC GGC ATG GAC GAG CTG TAC AAG TCC GGA GGT AGT GGC GGA AGC GGT GGC TCT TGG GGT GGG AGC AAA GTA GAT GTG ATC GAC GCG ACC ATC GAA AGC AGC AGC GAC GAA GAA GAA GAT CCG CCT GCC AAA CGC GGC GGT AGC GGT GGC TCT GGC GGG AGC GGC GGC TCG AAA GTT GAT GTC ATT GAC GCC ACA ATT GAA TCT TCG AGC GAT GAA GAG GAA GAT CCT CCA GCG AAA CGT GGC GGC TCG GGA GGT TCG GGC GGC AGC GGC GGA AGC AAA GTC GAT GTA ATT GAC GCG ACG ATC GAA TCT AGC AGC GAT GAA GAA GAA GAT CCA CCG GCT AAA CGT TAA TAA GGG GGC TCA TGC GGG GGT TCT CTC GAG AAC CTG TAT TTT CAG GGC
2xSUMO3_AatII_XhoI	GAC GAG GAC ACC ATT GAC GTC TTC CAG CAA CAA ACT GTA GTG GGG GGC TCC GGC GGA AGC GGG GGG TCG GGA GGT TCC ATG TCT GAA GAG AAA CCG AAG GAA GGG GTG AAA ACA GAG AAC GAT CAC ATT AAT CTG AAG GTT GCG GGG CAG GAT GGT TCC GTC GTG CAG TTC AAA ATT AAG CGT CAT ACT CCA CTG TCG AAA TTG ATG AAG GCA TAC TGT GAA CGT CAA GGC CTG TCA ATG CGC CAA ATC CGC TTC CGT TTT GAT GGC CAA CCA ATT AAT GAA ACT GAC ACC CCA GCA CAA TTA GAG ATG GAA GAC GAA GAT ACG ATC GAT GTG TTT CAA CAA CAG ACT GTC GTC GGG GGT AGC GGG GGA AGT GGT GGT TCG GGA GGG TCG ATG TCA GAG GAG AAA CCT AAA GAG GGA GTC AAA ACT GAG AAT GAT CAT ATC AAT CTT AAG GTA GCA GGA CAG GAC GGT AGC GTT GTA CAG TTC AAG ATT AAA CGC CAC ACT CCC TTG AGC AAG CTT ATG AAA GCT TAT TGC GAG CGC CAA GGA TTA TCT ATG CGC CAG ATC CGC TTT CGT TTT GAC GGC CAA CCC ATT AAC GAA ACT GAC ACT CCC GCG CAA CTG GAA ATG GAA GAT GAA GATA CGA TTG ACG TTT TTC AGC AGC AGA CCG TAG TTG GTG GCT CCT GCG GTG GCT CTC TCG AGA ACC TGT ATT

Table 4.4.2 List of ssDNA primers used for molecular cloning.

Primer	DNA Sequence (5'- to -3')
SUMO3_Stop_FWD	TAA GGT GGC TCC TGC GGT G
SUMO3_Stop_REV	AAC TAC GGT CTG CTG CTG AAA AAC
PIASx_Stop_FWD	TAA GGG GGC TCA TGC GGG
PIASx_Stop_REV	ACG TTT AGC CGG TGG ATC TTC
PIASx_RKRKR_FWD	ACG TAA ACG TTA ACG TAA ACG TAA AG
PIASx_RKRKR_REV	TTA CGG CCC TGA AAA TAC
PGP1_vec_PIASx_HiFi_FWD	TAA ACG TTA AGG TGG CTC CTG CGG TGG C
PGP1_vec_PIASx_HiFi_REV	CAT CTA CTT TGG ATC CGC CAG ACC CGC C
PIASx_ins_PGP1_HiFi_FWD	TGG CGG ATC CAA AGT AGA TGT GAT CGA CTT GAC
PIASx_ins_PGP1_HiFi_REV	AGG AGC CAC CTT AAC GTT TAG CCG GTG G
PGP1_vec_DAXXc_HiFi_FWD	TAG CGA CTA AGG TGG CTC CTG CGG TGG C
PGP1_vec_DAXXc_HiFi_REV	CAC TCG TTT TGG ATC CGC CAG ACC CGC C
DAXXc_ins_PGP1_HiFi_FWD	TGG CGG ATC CAA AAC GAG TGT TGC GAC C
DAXXc_ins_PGP1_HiFi_REV	AGG AGC CAC CTT AGT CGC TAT CCG ACA GG

4.4.3 Recombinant protein expression and purification

mEGFP fusions (pTEV19-His_{10}-mEGFP-$(SUMO3)_3$, pTEV19-His_{10}-mEGFP-$(SUMO3)_5$, pTEV19-His_{10}-mEGFP-$(SIM)_3$).

Two liters of Luria-Bertani medium, supplemented with 100 µg/mL ampicillin and 34 µg/mL chloramphenicol, were inoculated with E. coli BL21(DE3)pLysS cells transformed with pTEV19-His_{10}-mEGFP-$(SUMO3)_3$, pTEV19-His_{10}-mEGFP-$(SUMO3)_5$, or pTEV19-His_{10}-mEGFP-$(SIM)_3$, where the SIM is interchanged with the respective SIM peptide. The cells were grown at 37 °C, shaking at 250 rpm to OD_{600} 0.6–0.8. Protein expression was induced with 1 mM IPTG and cultures were shifted to 16 °C for ~18 h. Cells were harvested by centrifugation at 7,000 × g at 4

°C, and the pellet was either flash frozen in liquid nitrogen and stored at -80 °C or lysed in buffer containing 50 mM sodium phosphate, pH 7.5, 300 mM NaCl, and 1 mM DTT, supplemented with 1 mM iodoacetamide, 2 mM N-ethylmaleimide, and 1x cOmplete protease inhibitor cocktail (Roche). Cells were lysed by French press followed by brief sonication for 10 pulses. Lysed cells were clarified by centrifugation at 20,000 × g for 30 min at 4 °C. The supernatant was passed through a 0.45 µm PES filter and loaded onto 5 mL Ni-NTA resin. Protein was incubated for 1 h at 4 °C with gentle nutation. The flow-through was collected and the resin was washed with lysis buffer containing 20 mM imidazole (5 column volumes, CV) followed by 50 mM imidazole (5 CV), and proteins were eluted with 500 mM imidazole (2.5 CV). Fractions were analyzed by SDS-PAGE, and fractions containing the target protein were pooled and dialyzed at 4 °C twice against 2 x 4 L of 20 mM Tris, pH 7.5, and 1 mM DTT, first for 3 h then overnight. The dialyzed protein was purified by anion-exchange chromatography on a SOURCE 15Q 4.6/100 PE column (GE Healthcare, Chicago, IL). A linear 0–100% gradient over 40 CV was used to increase NaCl concentration. Buffer A was 20 mM Tris, pH 7.5, and 1 mM DTT. Buffer B was 20 mM Tris, pH 7.5, 1 M NaCl, and 1 mM DTT. Fractions were analyzed by SDS-PAGE, and the most concentrated and highest-purity fractions were pooled. The pooled elution was dialyzed against 2 × 4 L of microscopy buffer composed of 20 mM HEPES, pH 7.5, 150 mM KCl, 1 mM $MgCl_2$, 1 mM EGTA, and 1 mM DTT. Protein concentration was determined by absorbance at 488 nm using the reported mEGFP extinction coefficient (56.0 x 10^{-3} M-1 cm^{-1}; Kremers et al., 2011). Proteins were aliquoted into single-use volumes, flash frozen in liquid nitrogen, and stored at −80 °C until use.

SUMO3(C47S) for ITC. Three liters of Luria-Bertani medium supplemented with 100 µg/mL ampicillin were inoculated with E. coli BL21(DE3) transformed with the plasmid pTXB1-SUMO3(1-91)C47S-SUMO3(2-91)C47S-His_6. The cells were grown at 37 °C with shaking at 250 rpm to OD_{600} 0.6–0.8. Protein expression was induced with 0.8 mM IPTG and cultures were

incubated for 3 h at 37 °C with shaking at 250 rpm. Cells were harvested by centrifugation at 7,000 × g for 20 min at 4 °C. The cell pellet was lysed in buffer containing 50 mM sodium phosphate (pH 7.7), 300 mM NaCl, and 5 mM imidazole supplemented with 0.2 mM phenylmethylsulfonyl fluoride (PMSF) and 1x cOmplete protease inhibitor cocktail (Roche). Cells were lysed by sonication, and lysates were clarified by centrifugation at 20,000 × g for 30 min at 4 °C. The clarified lysate was passed through a 0.45 µm filter and loaded onto 5 mL Ni-NTA resin pre-equilibrated with lysis buffer. Protein was incubated for 45 min at 4 °C with gentle nutation. The column was washed with 10 column volumes (CV) of lysis buffer, followed by lysis buffer containing 20 mM imidazole (10 CV) and 50 mM imidazole (10 CV), and proteins were eluted with 500 mM imidazole (5 x 3 CV). Fractions were analyzed by SDS-PAGE and fractions containing the diSUMO3 fusion were pooled and dialyzed at 4 °C against 2 × 4 L of 50 mM sodium phosphate (pH 7.7), 150 mM NaCl, and 1 mM DTT. To liberate untagged SUMO3 from the fusion construct, the catalytic domain of SENP2 was added to a final concentration of 0.5 µM and the reaction was incubated at 30 °C for 2 h. After incubation, the reaction was diluted 1:1 with 50 mM sodium phosphate (pH 7.7), 150 mM NaCl to halve the DTT concentration. The diluted reaction mixture was loaded onto 5 mL Ni-NTA resin pre-equilibrated with reaction buffer lacking DTT to bind the tagged diSUMO3 fusion and the catalytic domain of SENP2. The flow-through was collected as untagged SUMO3. The column was washed 4 times with 2 CV of reaction buffer lacking DTT containing 5 mM imidazole. Untagged SUMO3 collected in the flow-through and washes were dialyzed against water 2 x 4 L. The dialysate was then collected, flash frozen, and lyophilized to dryness. The crude protein was purified via C4 preparative RP-HPLC (5–75% B over 45 min) to yield the cleaved monomeric SUMO3. Fractions were characterized by ESI-MS. The calculated average mass for SUMO3 was 10,377.6 Da, and the observed mass was 10,377.2 ± 1.6 Da.

PGP1 fusions (pTEV19-His$_{10}$-PGP1-(PIASx)$_3$ and pTEV19-His$_{10}$-PGP1-(DAXXc)$_3$). Two liters of Luria-Bertani medium, supplemented with 100 µg/mL ampicillin and 34 µg/mL chloramphenicol, were inoculated with E. coli BL21(DE3)pLysS cells (Novagen, Darmstadt, Germany) transformed with pTEV19-His10-PGP1-(PIASx)3 or pTEV19-His10-PGP1-(DAXXc)3. The cells were grown at 37 °C, shaking at 250 rpm to OD_{600} 0.6–0.8. Protein expression was induced with 1 mM IPTG and cultures were shifted to 16 °C for ~18 h. Cells were harvested by centrifugation at 7,000 × g at 4 °C, and the pellet was either flash frozen in liquid nitrogen and stored at −80 °C or lysed in buffer containing 50 mM sodium phosphate, pH 7.5, 300 mM NaCl, and 1 mM DTT. Cells were lysed by French press followed by brief sonication for 10 pulses. Lysed cells were clarified by centrifugation at 20,000 × g for 30 min at 4 °C. The supernatant was passed through a 0.45 µm PES filter and loaded onto 5 mL Ni-NTA resin. Protein was incubated for 1 h at 4 °C with gentle nutation. The flow-through was collected and the resin was washed with lysis buffer containing 20 mM imidazole (5 column volumes, CV) followed by 50 mM imidazole (5 CV), and proteins were eluted with 500 mM imidazole (2.5 CV). Fractions were analyzed by SDS-PAGE, and fractions containing the target protein were pooled and dialyzed at 4 °C twice against 2 x 4 L of 20 mM Tris, pH 7.5, and 1 mM DTT, first for 3 h then overnight. The dialyzed protein was purified by anion-exchange chromatography on a SOURCE 15Q 4.6/100 PE column (GE Healthcare, Chicago, IL). A linear 0–100% gradient over 40 CV was used to increase NaCl concentration. Buffer A was 20 mM Tris, pH 7.5, and 1 mM DTT. Buffer B was 20 mM Tris, pH 7.5, 1 M NaCl, and 1 mM DTT. Fractions were analyzed by SDS-PAGE, and the most concentrated and highest-purity fractions were pooled. The pooled elution was then dialyzed against 2 x 4 L 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 137 mM NaCl, 2.7 mM KCl, and 1 mM DTT, pH 7.5 (PBS). Protein concentration was determined by SDS-PAGE and Coomassie staining relative to bovine serum albumin standards of known concentration. Glycerol was added to 10% v/v, and samples were flash frozen and stored at −80 °C until use.

4.4.4 General protocol for confocal microscopy imaging of condensates

All confocal microscopy was performed on a Leica SP8 X scanning confocal microscope from Leica (Gräfelfing, Germany).

***In vitro* samples.** Imaging was performed with a 40x oil-immersion objective for *in vitro* samples. A tunable white-light laser (WLL) was set to 488 nm to excite mEGFP, and emission was collected on PMT detectors as follows: 497-577 nm when used alone, with Alexa Fluor 568 (AF568), and when used with pGlu-HCA to minimize crosstalk. For AF568-labeled PGP1-$(PIASx)_3$ and PGP1-$(DAXXc)_3$, the WLL was set to 577 nm, and emission was collected at 582–677 nm when imaged with mEGFP, and 582-650 nm when imaged with pGlu-HCA. For pGlu-HCA, the WLL was set to 670 nm, and emission was collected at 710-799 nm. Line-sequential scanning was used for multi-color acquisitions to reduce crosstalk. The z-focus was set at the glass surface, and the x and y coordinates were centered in each well.

In-cellulo imaging. Cellular imaging was performed using a 60x oil-immersion objective. A fixed-wavelength 405 nm laser was used to excite Hoechst 33342, and emission was collected on PMT detectors at 410-587 nm. For condensates containing mCherry fusions, the WLL was set to 587 nm to excite mCherry, and emission was collected at 587-592 nm. For WGA647, the WLL was set to 650 nm to excite AF647, and emission was collected at 592-653 nm. Line-sequential scanning was used for multicolor acquisitions to minimize crosstalk. A z-stack was acquired over the condensates, and the middle optical section was used for imaging.

4.4.5 Phase diagram mapping

mEGFP-$(SUMO3)_3$ and mEGFP-$(SIM)_3$ or mEGFP-$(SUMO3)_5$ and mEGFP-$(SIM)_3$ proteins were mixed in experimental buffer composed of 20 mM HEPES (pH 7.4), 150 mM KCl, 1 mM $MgCl_2$, 1 mM EGTA, and 1 mM DTT, at the indicated concentrations in a final volume of 80 µL. Mixtures were prepared in 96-well plates (Cellvis) surface-treated with bovine serum albumin (BSA) and incubated for 20–24 h at room temperature. Wells were imaged at 40x magnification using a Leica SP8X scanning confocal microscope. Condensate formation was scored as present or absent to construct the phase boundary for each SUMO3–SIM pair.

4.4.6 Measuring the stability of condensates to changing salt, ATP, and chaotropes (Decomposition)

For all decomposition experiments, mixtures were prepared in 96-well plates (Cellvis) surface-treated with bovine serum albumin (BSA) and incubated for 20–24 h at room temperature prior to imaging.

To test the effects of KCl concentration on the stability of condensates, mEGFP-$(SUMO3)_5$ and mEGFP-$(SIM)_3$ proteins were mixed in the experimental buffer, composed of 20 mM HEPES (pH 7.5), 1 mM $MgCl_2$, 1 mM EGTA, and 1 mM DTT. Experimental buffer containing concentrated KCl was added to the preformed condensates to reach final KCl concentrations of 0–3 M in a total volume of 80 µL.

To test the effects of ATP on condensate stability, mEGFP-$(SUMO3)_5$ and mEGFP-$(SIM)_3$ proteins were mixed in the experimental buffer, composed of 20 mM HEPES (pH 7.5), 150 mM KCl, 1 mM $MgCl_2$, 1 mM EGTA, and 1 mM DTT. Experimental buffer containing concentrated ATP was added to the preformed condensates to reach final ATP concentrations of 0–50 mM in a total volume of 80 µL.

To test the effects of chaotropes, such as guanidine-HCl and urea, on condensate stability, mEGFP-$(SUMO3)_5$ and mEGFP-$(SIM)_3$ proteins were mixed in the experimental buffer, composed of 20 mM HEPES (pH 7.5), 150 mM KCl, 1 mM $MgCl_2$, 1 mM EGTA, and 1 mM DTT. Experimental buffer containing concentrated guanidine-HCl or urea was added to the preformed condensates to reach final concentrations of 0–3 M in a total volume of 80 µL.

4.4.7 Fluorescence recovery after photobleaching (FRAP)

In vitro phase-separated condensates were subjected to the same preparation as described above. A Leica SP8X scanning confocal microscope with a 40x oil-immersion objective was used with the FRAP module in Leica LAS X software. Pre-bleach images were acquired for baseline normalization. A zoom factor was applied to center an individual condensate, and the region of interest (ROI) was defined as approximately 50% of the visible droplet area. The ROI was photobleached with a 488 nm argon laser at 100% power. Recovery was monitored by time-lapse acquisition until fluorescence reached a plateau, and intensities were background-subtracted and normalized to the pre-bleach mean.

For cellular experiments, cells were transfected and counterstained as described in 4.4.25 Counterstaining of human cells for confocal microscopy. A Leica SP8X scanning confocal microscope with a 63x oil-immersion objective was used with the FRAP module in Leica LAS X software. FRAP acquisition settings matched the *in vitro* assays. The ROI was photobleached with a 587 nm WLL at 100% power. Recovery was monitored by time-lapse acquisition until fluorescence reached a plateau, and intensities were background-subtracted and normalized to the pre-bleach mean.

4.4.8 Statistical analysis of FRAP data

Data were analyzed in ImageJ/Fiji. Fluorescence intensities were background-subtracted and normalized to the pre-bleach mean of each droplet ROI. To account for acquisition-induced photobleaching during the post-bleach recovery phase, intensities were corrected using an unbleached reference ROI. The resulting normalized intensities, $I(t)$, were fit by nonlinear least-squares to a single-exponential model $I(t) = I_\infty + (I_0 - I_\infty)e^{-kt}$, where I_0 is the intensity in the bleached ROI immediately after bleaching, I_∞ is the plateau intensity after recovery, and k is the apparent first-order rate constant. The characteristic time constant τ was defined as $1/k$. Half-times, $t_{1/2}$, were calculated as $\ln(2)/k$.

4.4.9 Reverse Phase High Performance Liquid Chromatography (RP-HPLC)

RP-HPLC used Buffer A (0.1% TFA in H_2O) and Buffer B (90% acetonitrile, 9.9% H_2O, 0.1% TFA) as the mobile phases. Analytical RP-HPLC was performed on an Agilent 1260 Infinity II quaternary system with either a Grace-Vydac C18 column (5 µm, 4.6 × 150 mm) or a Waters C4 column (5 µm, 4.6 × 150 mm) at 1.0 mL/min. A typical analytical gradient was 0–73% B over 30 min. Semi-preparative and preparative RP-HPLC were performed on an Agilent 1260 Infinity II preparative HPLC with either a Zorbax StableBond preparative C18 column (21.2 × 250 mm), a Hypersil Gold preparative C4 column (25 × 250 mm), a Zorbax StableBond semi-preparative C18 column (9.4 × 250 mm), or a Waters semi-preparative C4 column (10 × 250 mm). Preparative runs were performed at 20 mL/min and semi-preparative runs at 4 mL/min. Elution was monitored by UV–vis at 214 and 280 nm.

4.4.10 Electrospray Ionization Mass Spectrometry (ESI-MS)

Routine peptide and protein mass measurements were acquired by direct infusion electrospray ionization on a Bruker Esquire ion trap mass spectrometer (Billerica, MA) or on a Finnigan LTQ ion trap mass spectrometer (Thermo, Waltham, MA), operated in positive ion mode.

4.4.11 Synthesis, purification, and characterization of SIM peptides used for ITC

General. Peptides were prepared by Fmoc solid-phase peptide synthesis (SPPS) on the indicated resins with microwave assistance (CEM Liberty Blue 1.0 or CEM Liberty PRIME 2.0). Where noted, initial resin loading was intentionally reduced to limit aggregation and deletion products. Crude peptides were precipitated in ice-cold diethyl ether, collected by centrifugation, dissolved, and purified by C18 preparative RP-HPLC using the gradients specified below. Collected fractions were pooled and analyzed by ESI-MS; purity was confirmed by analytical RP-HPLC. Where applicable, immediate acetyl capping was performed after specific Asp couplings to avoid the formation of aspartic-acid deletion peptides. HPLC solvents: Buffer A (99.9% H_2O, 0.1% trifluoroacetic acid (TFA)); Buffer B (90% acetonitrile, 9.9% H_2O, 0.1% TFA). All gradients are reported as %B.

(1) PIASx wild-type:

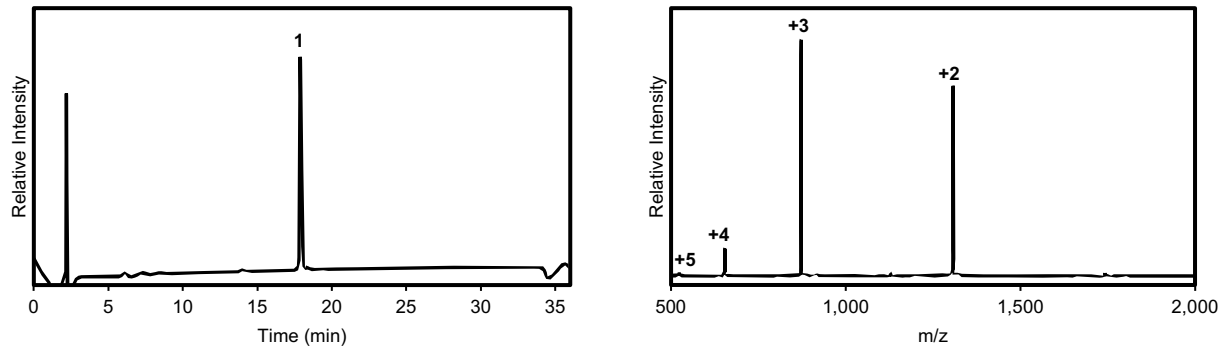
Ac-KVDVIDLTIESSSDEEEDPPAKR-$CONH_2$

Rink amide (MBHA) resin (0.667 meq g^{-1}) was swollen in 1:1 DMF/DCM and deprotected with 20% (v/v) piperidine in DMF for 30 min. To reduce loading, the third residue (Ala) was coupled as a 1:1 mixture of Fmoc-Ala-OH and Boc-Ala-OH (0.8 mmol each) with Oxyma (0.8 mmol) and DIC (0.8 mmol) for 60 min at 25 °C. Reduced loading was verified by measuring the dibenzofulvene–piperidine adduct at 301 nm after treatment with 20% (v/v) piperidine in DMF. Chain elongation proceeded on the Liberty Blue using 5 equivalents of each Fmoc amino acid

relative to the reduced loading. Fmoc deprotection used 5% (w/v) piperazine with 0.05 M HOBt in DMF for 3 min at 75 °C. Standard couplings were 10 min at 75 °C with Fmoc-AA (0.5 mmol), DIC (0.49 mmol), Oxyma (0.49 mmol), and DIEA (0.98 mmol) in DMF; additional double couplings were used for Fmoc-Arg(Pbf)-OH, Fmoc-Thr(OtBu)-OH, Fmoc-Val-OH, Fmoc-Lys(Boc)-OH, Fmoc-Asp(OtBu)-OH, and Fmoc-Leu-OH as needed. The N terminus was capped with acetic anhydride (0.5 mmol) and DIEA (0.5 mmol) for 10–15 min at 25 °C.

Cleavage used TFA/triisopropylsilane/H_2O, 90:5:5 (v/v), for 2 h at 25 °C. The ether-precipitated pellet was dissolved, filtered (0.45 µm), lyophilized, and purified by C18 preparative RP-HPLC (17–48%B over 40 min). Pooled fractions were analyzed by ESI-MS. Final QC used RP-HPLC (0–73% over 30 min) and ESI-MS. Calculated M_{avg} = 2,613.3 Da; found 2,613.3 ± 0.4 Da.

Supplementary Figure S4.4.11.1



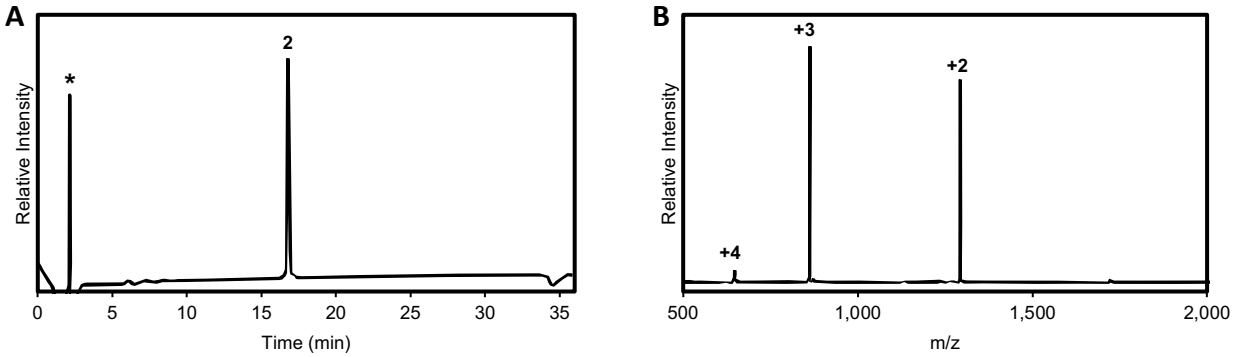
Supplemental Figure S4.4.11.1. Characterization of PIASx wt (1). (A) C18 analytical HPLC chromatogram of **1** on a gradient of 0–73% B over 30 min. * indicate the buffer peak. **(B)** ESI-MS of pure **1**. Observed $[M]^+$ = 2,613.3 ± 0.4 Da, calculated $[M]^+$ = 2,613.3 Da.

(2) PIASx (V4A):

Ac-KVDAIDLTIESSSDEEEDPPAKR-$CONH_2$

Synthesized as in **(1)**. Calculated M_{avg} = 2,585.7 Da; found 2,585.4 ± 0.4 Da.

Supplementary Figure S4.4.11.2



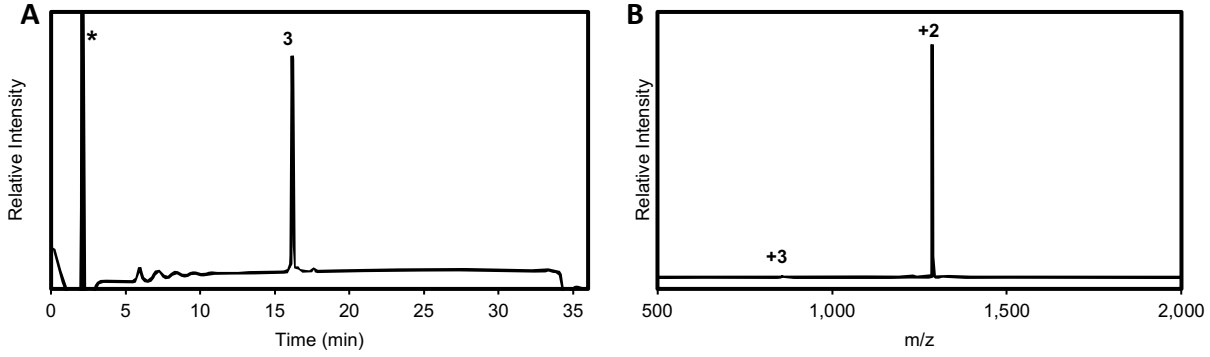
Supplementary Figure S4.4.11.2. Characterization of PIASx V4A (**2**). **(A)** C18 analytical HPLC chromatogram of **2** on a gradient of 0–73% B over 30 min. * indicate the buffer peak. **(B)** ESI-MS of pure **2**. Observed $[M]^+$ = 2,585.4 ± 0.4 Da, calculated $[M]^+$ = 2,585.7 Da.

(3) PIASx (I5A):

Ac-KVDVADLTIESSSDEEEDPPAKR-$CONH_2$

Synthesized as in **(1)**. To avoid the formation of aspartic-acid deletion peptides, Asp3 and Asp6 were double-coupled and immediately acetyl-capped with acetic anhydride (0.5 mmol) and DIEA (0.5 mmol) after these couplings. Calculated M_{avg} = 2,571.7 Da; found 2,571.5 ± 0.2 Da.

Supplementary Figure S4.4.11.3



Supplemental Figure S4.4.11.3. Characterization of PIASx (I5A) (3). (A) C18 analytical HPLC chromatogram of **3** on a gradient of 0–73% B over 30 min. * indicate the buffer peak. **(B)** ESI-MS of pure **3**. Observed $[M]^+$ = 2,571.5 ± 0.2 Da, calculated $[M]^+$ = 2,571.7 Da.

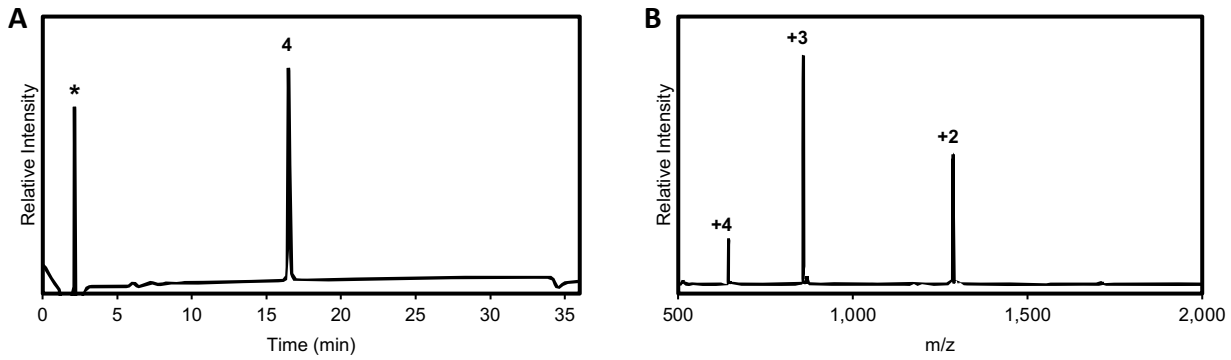
(4) PIASx (L7A):

Ac-KVDVIDATIESSSDEEEDPPAKR-$CONH_2$

Synthesized as in **(1)**. Calculated M_{avg} = 2,571.7 Da; found 2,571.2 ± 0.3 Da.

Supplementary Figure S4.4.11.4

.



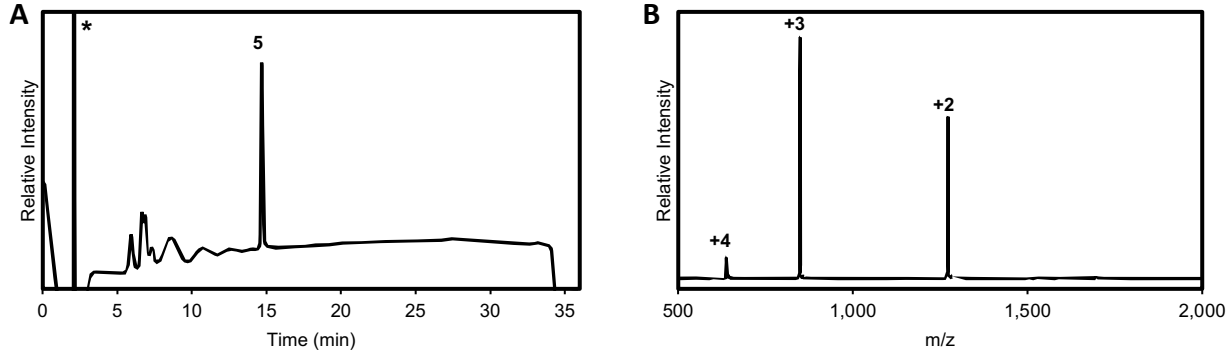
Supplemental Figure S4.4.11.4. Characterization of PIASx L7A (4). (A) C18 analytical HPLC chromatogram of **4** on a gradient of 0–73% B over 30 min. * indicate the buffer peak. **(B)** ESI-MS of pure **4**. Observed $[M]^+$ = 2,571.2 ± 0.3 Da, calculated $[M]^+$ = 2,571.7 Da.

(**5**) PIASx (V4A, L7A):

Ac-KVDAIDATIESSSDEEEDPPAKR-$CONH_2$

Synthesized as in **(1)**. Calculated M_{avg} = 2,543.6 Da; found 2,543.6 ± 0.6 Da.

Supplementary Figure S4.4.11.5



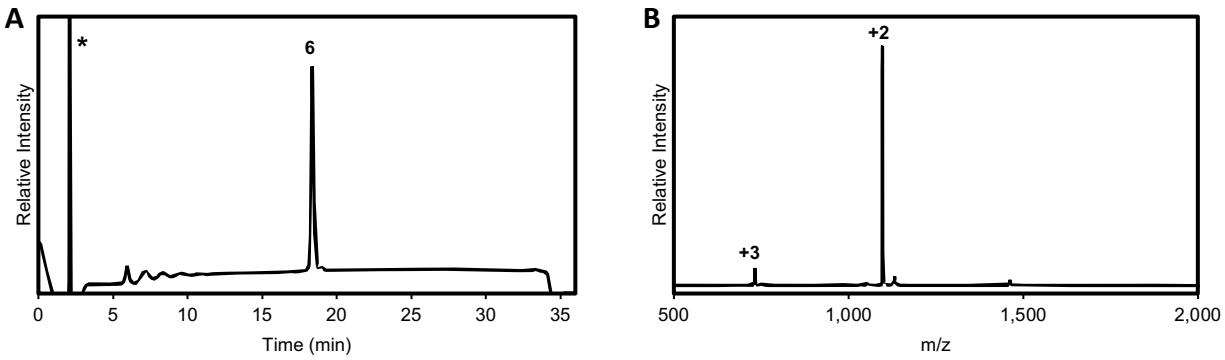
Supplemental Figure S4.4.11.5. Characterization of PIASx (V4A,L7A) (5). (A) C18 analytical HPLC chromatogram of **5** on a gradient of 0–73% B over 30 min. * indicate the buffer peak. **(B)** ESI-MS of pure **5**. Observed $[M]^+$ = 2,543.6 ± 0.6 Da, calculated $[M]^+$ = 2,543.6 Da.

(**6**) DAXXc:

Ac-KTSVATQCDPEEIIVLSDSD-$CONH_2$

Fmoc-Asp-Wang resin (0.57 meq g^{-1}) was swollen in DMF for 20–30 min and deprotected with 20% (v/v) piperidine in DMF. Loading was reduced as in **(1)** using Fmoc-Ser(OtBu)-OH (0.4 mmol). Elongation used a CEM Liberty PRIME 2.0 with the "no-wash" method[50] and 5 equivalents per coupling. After elongation, capping and cleavage followed **(1)**. Purification used C18 preparative RP-HPLC (20–44% B over 40 min). Final QC used RP-HPLC (0–73% B over 30 min) and ESI-MS. Calculated M_{avg} = 2,192.3 Da; found 2,192.2 ± 0.1 Da.

Supplementary Figure S4.4.11.6



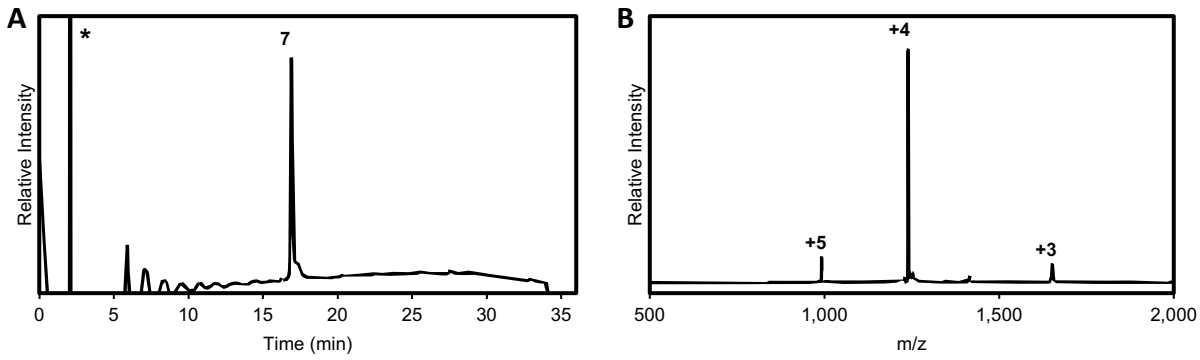
Supplemental Figure S4.4.11.6. Characterization of DAXXc (6). (A) C18 analytical HPLC chromatogram of **6** on a gradient of 0–73% B over 30 min. * indicate the buffer peak. **(B)** ESI-MS of pure **6**. Observed $[M]^+$ = 2,192.2 ± 0.1 Da, calculated $[M]^+$ = 2,192.3 Da.

(7) MCAF1 wild type:

Ac-GSPEFKTIDASVSKKAADSTSQCGKATGSDSSGVIDLTMDDEESGASQD-$CONH_2$

Rink amide (MBHA) resin (0.667 meq g^{-1}) was swollen in 1:1 DMF/DCM and deprotected with 20% piperidine in DMF (v/v). Loading was reduced as in **(1)** using Fmoc-Asp(OtBu)-OH (0.24 mmol). Elongation used a CEM Liberty PRIME 2.0 with the "no-wash" method.[50] Additional double couplings were applied for Fmoc-Ile-OH, Fmoc-Thr(OtBu)-OH, Fmoc-Val-OH, Fmoc-Lys(Boc)-OH, Fmoc-Ala-OH, Fmoc-Leu-OH, Fmoc-Gly-OH, Fmoc-Cys(Trt)-OH, Fmoc-Ser(OtBu)-OH, and Fmoc-Gln(Trt)-OH. Cleavage used TFA/triisopropylsilane/H_2O/anisole, 85/5/5/5 (v/v). Purification used C18 preparative RP-HPLC (19–43%B over 45 min). Final QC used RP-HPLC (0–73% over 30 min) and ESI-MS. Calculated M_{avg} = 4,953.2 Da; found 4,953.1 ± 0.1 Da.

Supplementary Figure S4.4.11.7



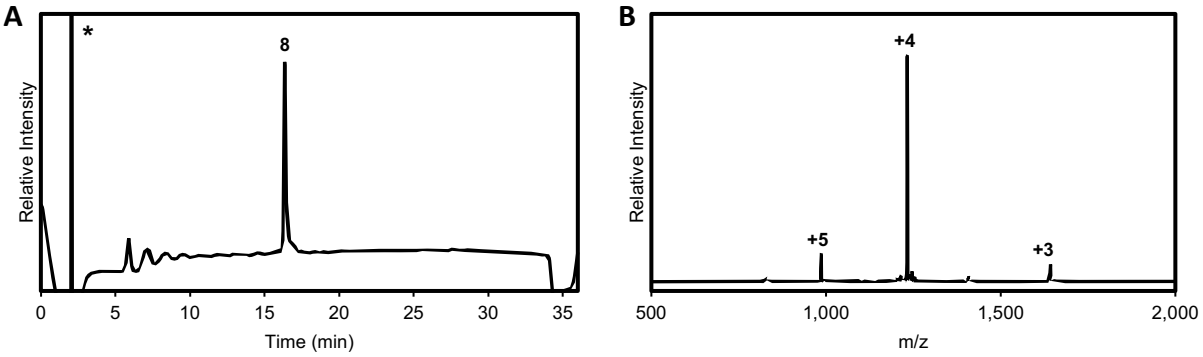
Supplemental Figure S4.4.11.7. Characterization of MCAF1 (7). (A) C18 analytical HPLC chromatogram of **7** on a gradient of 0–73% B over 30 min. * indicate the buffer peak. **(B)** ESI-MS of pure **7**. Observed $[M]^+$ = 4,953.1 ± 0.1 Da, calculated $[M]^+$ = 4,953.2 Da.

(8) MCAF1 (V34A):

Ac-GSPEFKTIDASVSKKAADSTSQCGKATGSDSSGAIDLTMDDEESGASQD-$CONH_2$

Synthesized as in **(7)**. Calculated M_{avg} = 4,925.1 Da; found 4,925.8 ± 0.9 Da.

Supplementary Figure S4.4.11.8



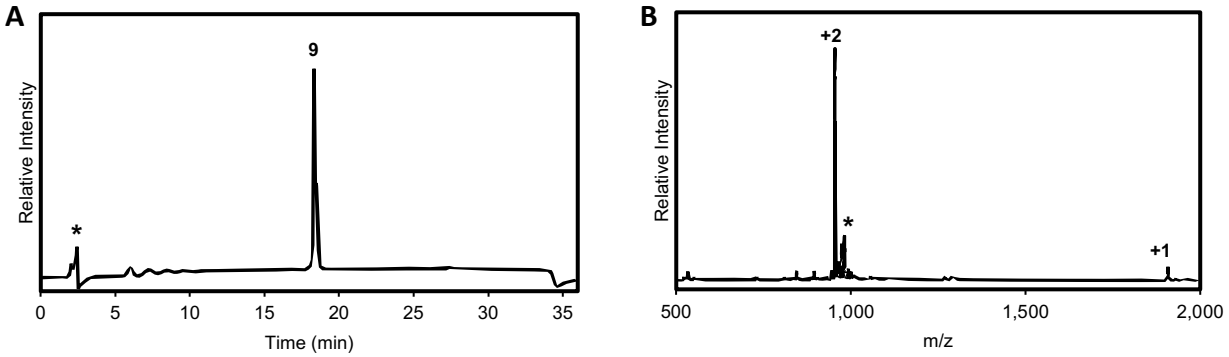
Supplemental Figure S4.4.11.8. Characterization of MCAF1 (V34A) (8). (A) C18 analytical HPLC chromatogram of **8** on a gradient of 0–73% B over 30 min. * indicate the buffer peak. **(B)** ESI-MS of pure **8**. Observed $[M]^+$ = 4,925.8 ± 0.9 Da, calculated $[M]^+$ = 4,925.1 Da.

(9) DAXXn:

Ac-MATANSIIVLDDDDEDE-$CONH_2$

Synthesized as in **(7)**. Cleavage used TFA/triisopropylsilane/H_2O/anisole/thioanisole, 85/5/5/2.5/2.5 (v/v). Purification used C18 preparative RP-HPLC (17–48% B over 40 min). Final QC used RP-HPLC (0–73% over 30 min) and ESI-MS. Calculated M_{avg} = 1,906.9 Da; found 1,907.2 ± 0.3 Da.

Supplementary Figure S4.4.11.9



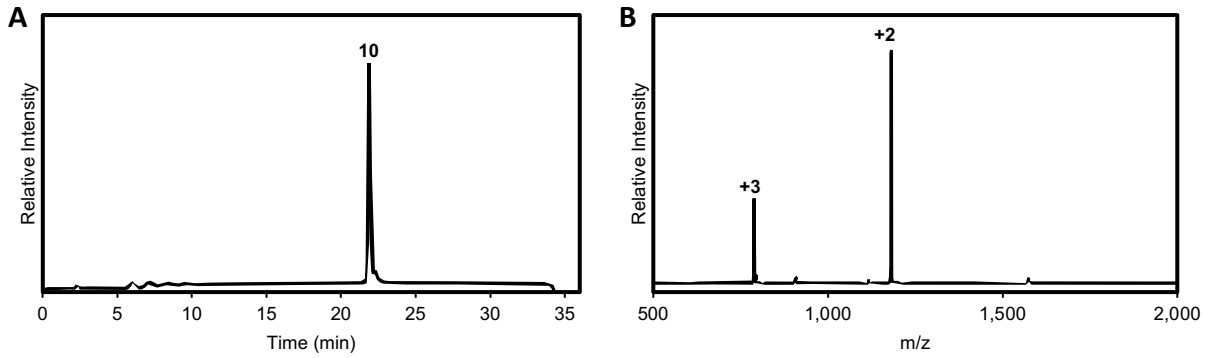
Supplemental Figure S4.4.11.9. Characterization of DAXXn (9). (A) C18 analytical HPLC chromatogram of **9** on a gradient of 0–73% B over 30 min. * indicate the buffer peak. **(B)** ESI-MS of pure **9**. Observed $[M]^+$ = 1,907.2 ± 0.3, calculated $[M]^+$ = 1,906.9 Da. * indicate salt adducts.

(10) SIM2:

Ac-RETAGDEIVDLTCESLEPRW-CONH$_2$

Synthesized as in (7), with additional double couplings for Fmoc-Glu(OtBu)-OH and Fmoc-Arg(Pbf)-OH. Cleavage used TFA/triisopropylsilane/H_2O/anisole/thioanisole, 85/5/5/2.5/2.5 (v/v). Purification used C18 preparative RP-HPLC (25–55% B over 40 min). Final QC used RP-HPLC (0–73% B over 30 min) and ESI-MS. Calculated M_{avg} = 2,360.5 Da; found 2,361.1 ± 0.4 Da.

Supplementary Figure S4.4.11.10



Supplemental Figure S4.4.11.10. Characterization of SIM2 (10). (A) C18 analytical HPLC chromatogram of **10** on a gradient of 0–73% B over 30 min. **(B)** ESI-MS of pure **10**. Observed $[M]^+$ = 2,361.1 ± 0.4 Da, calculated $[M]^+$ = 2,360.5 Da.

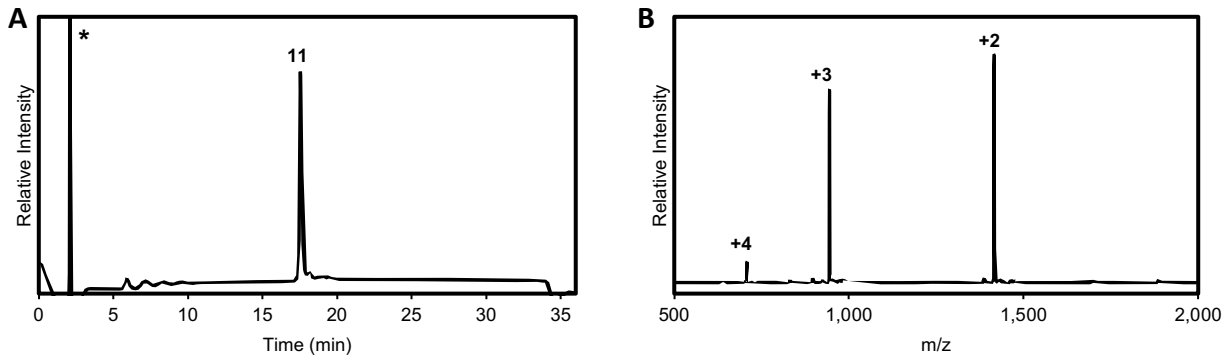
Synthesis of PIASX-(RK)$_5$ peptide by native chemical ligation (NCL)

(A) PIASx hydrazide:

Ac-KVDVIDLTIESSSDEEEDPPAKRGGS-C(O)NHNH$_2$

2-Chlorotrityl hydrazide resin was prepared by reacting 2-chlorotrityl chloride resin (1.5–2.0 meq g^{-1}) with 10% (v/v) hydrazine in DMF at 25 °C for 30 min, repeated once with fresh hydrazine. Residual sites were capped with 10% (v/v) methanol in DMF for 10 min. SPPS followed **(1)**, with reduced loading using Fmoc-Gly-OH (0.2 mmol). Calculated M_{avg} = 2,829.9 Da; found 2,829.6 ± 0.4Da.

Supplementary Figure S4.4.11.11



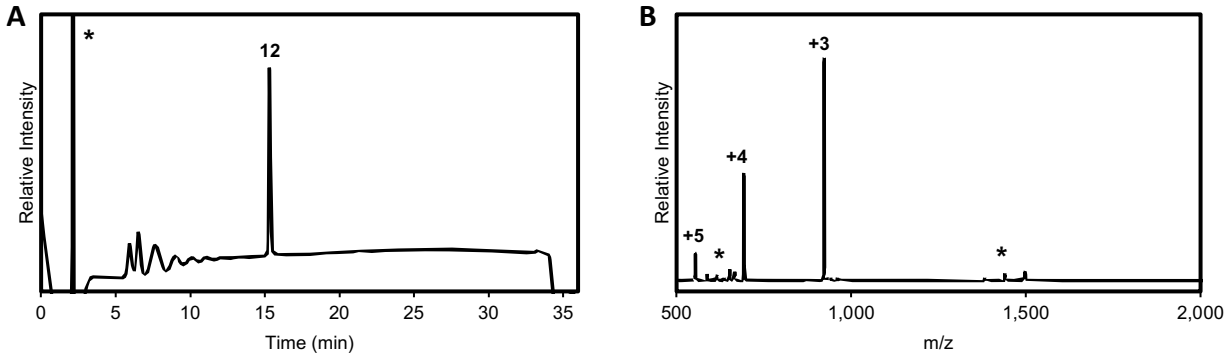
Supplemental Figure S4.4.11.11. Characterization of hydrazide peptide (11). (A) C18 analytical HPLC chromatogram of **11** on a gradient of 0–73% B over 30 min. * indicate the buffer peak. **(B)** ESI-MS of pure **11**. Observed $[M]^+$ = 2,829.6 ± 0.4 Da, calculated $[M]^+$ = 2,829.9 Da.

(B) PIASx cysteine peptide:

H_2N-CGGSLENLYFQGRKRKRKRKRKG-$CONH_2$

Synthesized as in (1) with additional double couplings for Fmoc-Glu(OtBu)-OH, Fmoc-Asn(Trt)-OH, and Fmoc-Cys(Trt)-OH. The N terminus was left free for NCL. Calculated M_{avg} = 2,765.3 Da;found 2,765.0 ± 0.1 Da.

Supplementary Figure S4.4.11.12

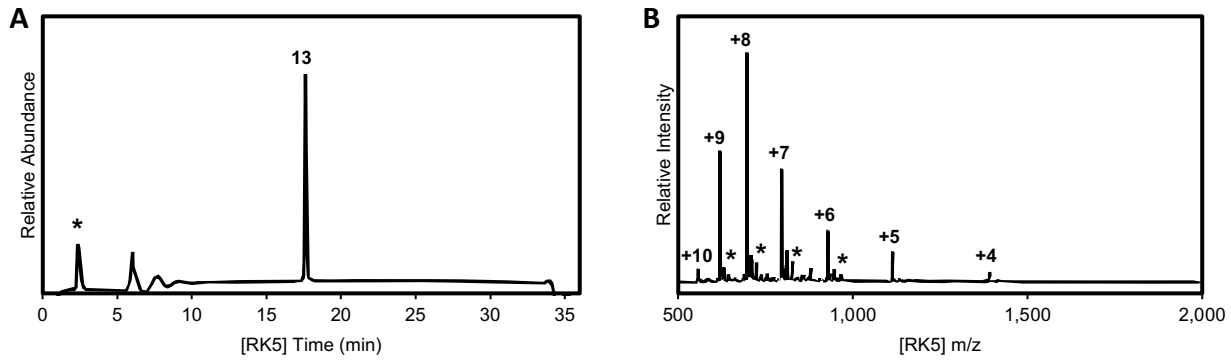


Supplemental Figure S4.4.11.12. Characterization of cysteine peptide (12). (**A)** C18 analytical HPLC chromatogram of **12** on a gradient of 0–73% B over 30 min. * indicate the buffer peak. **(B)** ESI-MS of pure **12**. Observed $[M]^+$ = 2,765.0 ± 0.1 Da, calculated $[M]^+$ = 2,765.3 Da. * indicate fragments.

(C) NCL to generate PIASx-$(RK)_5$

Peptide **(A)** was dissolved in 0.2 M sodium phosphate, pH 3.0, with 6 M guanidine HCl. Sodium nitrite (10 equivalents) was added at −20 °C for 20 min to form the acyl azide. Peptide **(B)** (1.5 equivalents) was dissolved in the same buffer and mixed with MPAA (100 equivalents). After conversion of **(A)** to the acyl azide, the solution of **(B)**+MPAA was added to initiate NCL at 25 °C. Aliquots were removed hourly, quenched with phosphate buffer and ~80 equivalents TCEP, and analyzed by RP-HPLC and ESI-MS. Upon completion, the reaction was quenched with ~80 equivalents TCEP and purified by C18 RP-HPLC (20–60%B over time). Product identity was confirmed by ESI-MS. Calculated M_{avg} = 5,563.2 Da; found 5,564.3 ± 1.0 Da.

Supplementary Figure S4.4.11.13



Supplemental Figure S4.4.11.13. Characterization of PIASx-$(RK)_5$ (13). (A) C18 analytical HPLC chromatogram of **13** on a gradient of 0–73% B over 30 min. * indicate the buffer peak. **(B)** ESI-MS of pure **13**. Observed $[M]^+$ = 5564.3 ± 1.0 Da, calculated $[M]^+$ = 5,563.2 Da. * indicate TFA adducts.

4.4.12 Isothermal titration calorimetry (ITC) of SUMO3 with SIM peptides

ITC experiments were performed in a buffer composed of either 20 mM Tris (pH 7.6) or 20 mM MES (pH 6.5), 150 mM KCl, on a MicroCal PEAQ-ITC (Malvern Panalytical, Westborough, MA) or an Affinity ITC (TA Instruments, New Castle, DE). The titration program consisted of one 0.4 µL priming injection followed by 18 injections of 2 µL, spaced 150 s, at 25 °C. SUMO3 was dialyzed into the chosen buffer. The same dialysis buffer was used to dissolve each SIM

peptide, and peptide solutions were adjusted to match the protein pH. Following instrument protocols, the sample cell and syringe were washed with 20% Contrad70, rinsed with H2O, and dried with methanol. The sample cell was then rinsed manually with H2O four times, followed by the working buffer three times. Three hundred microliters of protein solution were loaded into the cell and 40 µL of peptide solution into the syringe. Heats of injection were integrated to generate binding isotherms, and ΔH values were fit with the instrument binding model to obtain binding affinities.

For SIM2, PIASx (V4A), PIASx (L7A), MCAF1 (V34A), Malvern Panalytical's concatenation workflow was used to extend the isotherm to higher ligand-to-protein ratios without increasing peptide concentration. Two sequential titrations were performed per replicate: after the first titration, the syringe was refilled with peptide while the same protein sample remained in the cell. The two titrations were concatenated in the software and analyzed as a single isotherm.

4.4.13 Synthesis of amino hemicyanine (HCA)

HCA was synthesized as described[51] with minor modifications. Briefly, 3-nitrophenol (461 mg, 3.75 mmol) and K2CO3 (458 mg, 3.75 mmol) were suspended in 20 mL acetonitrile under an argon atmosphere. IR-780 (1.0 g, 1.33 mmol) was added, and the reaction mixture was stirred for 4 h. The solvent was then evaporated, and the residue was dissolved in DCM, washed with H2O (3x), and dried over Na_2SO4. The organics were evaporated, and the residue was dissolved in 40 mL MeOH and maintained under argon.

$SnCl_2$ (6.0 g, 30 mmol) in concentrated HCl (5 mL) was added via syringe, and the reaction mixture was gently refluxed overnight. The next morning, the reaction was quenched by the addition of saturated Na_2CO_3 and filtered. The filtrate was diluted with DCM, washed with H2O

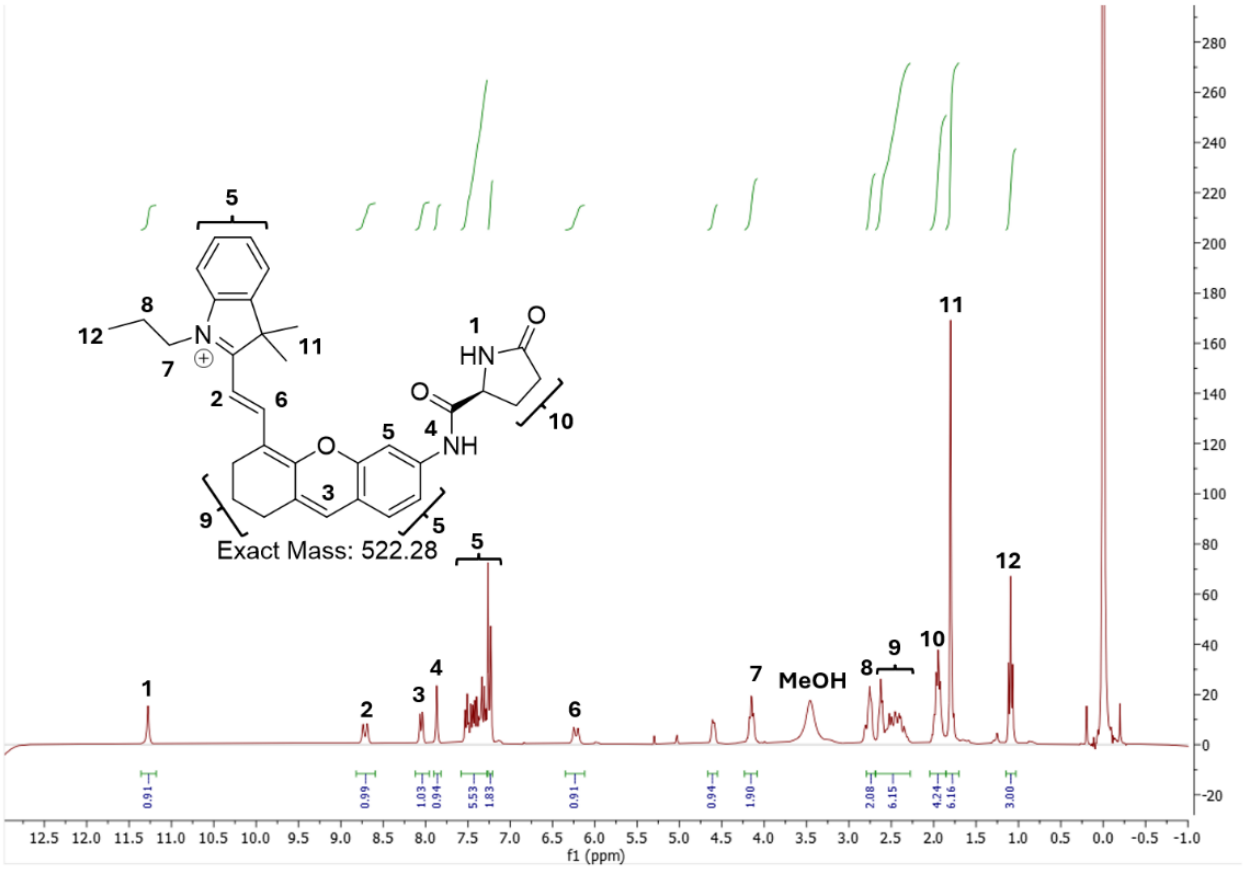
(3x), and dried over Na_2SO_4. After solvent evaporation, the crude residue was purified by silica gel chromatography (2–20% MeOH/DCM) to afford HCA as a green solid (162 mg, 26%).

4.4.14 Synthesis of L-pyroglutamic acid hemicyanine (pGlu-HCA)

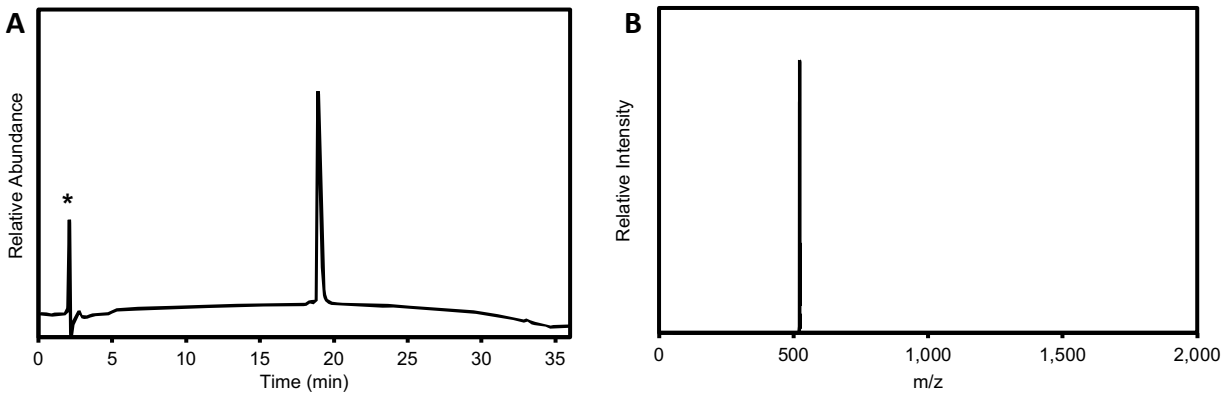
Boc-L-pyroglutamic acid (71 mg, 0.3 mmol) and HATU (230 mg, 0.6 mmol) were suspended in dry DCM (15 mL) at 0 °C under an argon atmosphere. DIPEA (196 μL, 0.9 mmol) was added *via* syringe, and the reaction mixture was stirred for 30 min on ice. HCA (162 mg, 0.3 mmol) dissolved in dry DCM (2 mL) was then added *via* syringe, and the reaction was brought to room temperature and stirred vigorously for 36 h under an argon atmosphere. The reaction mixture was then diluted with DCM, washed with H_2O (3x), and dried over Na_2SO_4 before solvent evaporation.

The residue was then suspended in DCM (10 mL) at 0 °C, and trifluoroacetic acid (5 mL) was added dropwise. The reaction mixture was then brought to room temperature and stirred for 3 h. After solvent evaporation, the crude residue was resuspended in MeOH and purified by C18 preparative RP-HPLC (30-90% B over 40 min). Fractions deemed pure by C18 analytical RP-HPLC and ESI-MS were combined and lyophilized to yield pGlu-HCA (12.4 mg, 6%) as a violet solid.

Supplementary Figure S4.4.14.1 and Supplementary Figure S4.4.14.2



Supplemental Figure S4.4.14.1. *H-NMR of pGlu-HCA ($CDCl_3$, 300 MHz).*



Supplemental Figure S4.4.14.2. Characterization of pGlu-HCA. (A) C18 analytical HPLC chromatogram on a gradient of 20–100% B over 30 min. * indicate the buffer peak. (B) ESI-MS of pure pGlu-HCA. Observed $[M]^+$ = 522.36 Da, calculated $[M]^+$ = 522.28 Da.

4.4.15 Labeling of PGP1-$(PIASx)_3$ and PGP1-$(DAXXc)_3$

PGP1-$(PIASx)_3$ and PGP1-$(DAXXc)_3$ were labeled with Alexa Fluor 568 following the manufacturer's protocol (Invitrogen). Labeled proteins were purified on the gel filtration column supplied with the kit. The protein concentrations were determined by absorbance at 577 nm and 280 nm using a NanoDrop 2000c UV–Vis spectrophotometer (ThermoFisher Scientific, Waltham, MA). Concentrations were calculated using the equation below (ε280 = 25,900 M^{-1} cm^{-1} for both enzymes):

$$C_{protein}\ (M) = \frac{[A_{280} - 0.46\ x\ A_{577}]\ x\ Dilution\ factor}{25{,}900\ M^{-1}cm^{-1}}$$

Where A_{280} and A_{577} denote absorbance at 280 nm and 577 nm, respectively.

4.4.16 Michaelis-Menten kinetics analysis of AF568-labeled PGP1-$(PIASx)_3$ and AF568-labeled PGP1-$(DAXXc)_3$

Flat-bottom 96-well plates were passivated with 0.1 mg/mL bovine serum albumin (BSA) in 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 137 mM NaCl, 2.7 mM KCl (pH 7.5) (PBS) for 1 h at room temperature. Wells were washed with H2O (5x) followed by PBS (3x) to remove residual BSA. Reactions (200 µL) were prepared in PBS in these BSA-passivated wells at 37 °C. pGlu-HCA was titrated at 1, 2.66, 7, 13, and 19 µM. Reactions were initiated by adding either AF568-labeled PGP1-$(PIASx)_3$ or AF568-labeled PGP1-$(DAXXc)_3$ to a final concentration of 50 nM and monitored in a BioTek Synergy 2 microplate reader (Winooski, VT) at $\lambda_{ex}/\lambda_{em}$ = 670/700 nm, recording every 30 s for 2 h. For each enzyme (AF568-labeled PGP1-$(PIASx)_3$ and AF568-labeled PGP1-$(DAXXc)_3$), initial velocities (v_0) were computed from the linear portion of the first 5 min of each progress curve using linear regression. Background (no-enzyme and enzyme-only) slopes were subtracted.

v_0 values were plotted versus substrate concentration and fit by nonlinear least squares to the Michaelis–Menten equation, $v_0 = (V_{max}[S])(K_M + [S])$, in R (RStudio; packages: ggplot2 and broom) to obtain K_m and V_{max}. k_{cat} was calculated as $V_{max}/[E]_T$ using the active enzyme concentration. All conditions were identical for AF568-labeled PGP1-$(PIASx)_3$ and AF568-labeled PGP1-$(DAXXc)_3$.

4.4.17 Spectroscopic analysis of reaction kinetics

Bulk fluorescence measurements of PGP1 activity were carried out in flat-bottom 96-well plates. Plates were first passivated with 0.1 mg/mL bovine serum albumin (BSA) in phosphate-buffered saline (PBS; 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 137 mM NaCl, 2.7 mM KCl, pH 7.5) for 1 h at room temperature. Wells were washed five times with H_2O followed by three washes with PBS to remove residual BSA.

Reactions (200 µL) were assembled in BSA-passivated wells in PBS at 37 °C. For condensate-containing samples, preformed SUMO:PIASx or SUMO:DAXXc condensates were prepared as described above and diluted into PBS immediately before use. Aliquots of these condensate suspensions were added to the wells, followed by AF568-labeled PGP1-$(PIASx)_3$ or AF568-labeled PGP1-$(DAXXc)_3$ to a final enzyme concentration of 50 nM. For matched solution controls, the same concentration of AF568-labeled PGP1 construct was added to wells containing PBS lacking condensates.

Reactions were initiated by addition of the fluorogenic substrate pGlu-HCA to a final concentration of 4 µM. Formation of the fluorescent HCA product was monitored in a microplate reader (BioTek Synergy 2, Winooski, VT) using excitation at 670 nm and emission at 700 nm. Fluorescence intensity was recorded as a function of time for each well. Substrate-only and

condensate-only controls, lacking AF568-labeled PGP1, were included to confirm that changes in signal arose from enzyme-catalyzed hydrolysis.

For each condition, fluorescence time courses from three technical replicate wells were collected. Traces were background-subtracted using the corresponding substrate-only control and, where indicated, normalized to the initial fluorescence at the start of the reaction. For each replicate, an apparent initial rate was obtained from a linear fit to the early, approximately linear portion of the fluorescence versus time trace. These slopes were averaged across technical replicates and used to compare the relative extent of PGP1-catalyzed product formation between solution and condensate conditions.

4.4.18 Confocal microscopy analysis of reaction kinetics

Condensates were prepared as previously described. AF568-labeled PGP1-$(PIASx)_3$ was added to solutions of preformed SUMO:PIASx condensates and AF568-labeled PGP1-$(DAXXc)_3$ was added to solutions of preformed SUMO:DAXXc condensates at a final concentration of 50 nM. After 1 h of incubation with the enzymes, pGlu-HCA was added to the solution of condensates at a final concentration of 2 μM. For control solutions, preformed condensates were not treated with enzyme. The reaction was monitored over time for ~30 min using a Leica SP8X scanning confocal microscope, with λ_{ex} = 670 nm and λ_{em} = 710–799 nm. For each condition, 20 condensates were analyzed in ImageJ FIJI to quantify mean fluorescence over time.

4.4.19 Encapsulation Efficiency of pGlu-HCA

Condensates for SUMO:PIASx and SUMO:DAXXc were prepared as described above. In the absence of AF568-labeled PGP1-$(PIASx)_3$ or AF568-labeled PGP1-$(DAXXc)_3$, 2 μM pGlu-HCA

was added to each mixture. The fluorescence signal was acquired on a Leica SP8X scanning confocal microscope, with λ_{ex} = 600 nm and λ_{em} = 610–699 nm. Fluorescence intensity was measured inside condensates and in the surrounding dilute phase on n = 10 condensates and n = 10 dilute-phase ROIs per image, 30 s after pGlu-HCA addition. The encapsulation efficiency (EE) was calculated using:

$$\%EE = \left(\frac{Intensity_{droplets}}{Intensity_{droplets} + Intensity_{background}}\right) \text{x } 100$$

4.4.20 Preparation of pC1-mCherry plasmids for transfection

DNA for transient transfection was prepared by Miraprep of E. coli cells as previously described using a Qiagen (Germantown, MD) DNA miniprep kit. Briefly, transformed E. coli XL10-Gold cells (Agilent, Santa Clara, CA) were grown overnight at 37 °C in 50 mL Luria-Bertani medium supplemented with ampicillin (100 µg/mL). Cells were collected by centrifugation and resuspended in Buffer P1 supplemented with fresh RNase. After alkaline lysis and neutralization, the supernatant was cleared by centrifugation. The supernatant was diluted with an equal volume of 96% (v/v) ethanol prior to loading onto four Qiagen miniprep spin columns. DNA was washed and eluted according to the Qiagen protocol. Purity of the eluted DNA was checked by measuring the A_{260}/A_{280} ratio on a NanoDrop 2000c spectrophotometer and by agarose gel electrophoresis. The correct gene sequences for all constructs were confirmed by DNA sequencing prior to transfection in human cells.

4.4.21 Transient transfection of human cells

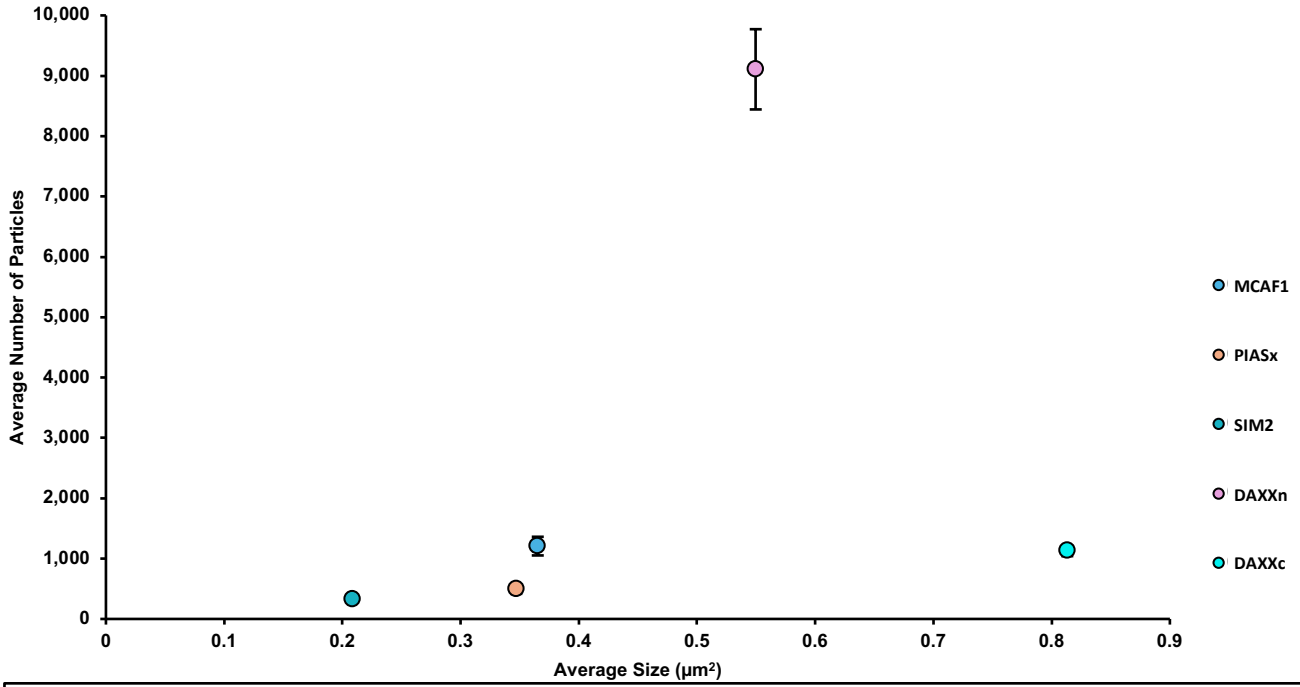
HeLa, HEK293T, and H1299 cells were maintained in DMEM (HeLa, HEK293T) or RPMI 1640 (H1299) supplemented with 10% fetal bovine serum at 37 °C in a humidified incubator with 5% CO_2. HeLa, HEK293T, and H1299 cells were cultured to ~60% confluence before transient transfection with pC1-mCherry plasmids. The cell growth medium was replaced at least 1 h

before transfection. Cells were transfected using Lipofectamine 3000 (Invitrogen) at 0.26 µg DNA per well in 8-well µ-Slides (Ibidi), and the medium was replaced 24 h later. Cells were incubated in transfection medium for an additional 24 h at 37 °C in a humidified incubator with 5% CO_2 prior to counterstaining as described in 4.4.21 and confocal microscopy.

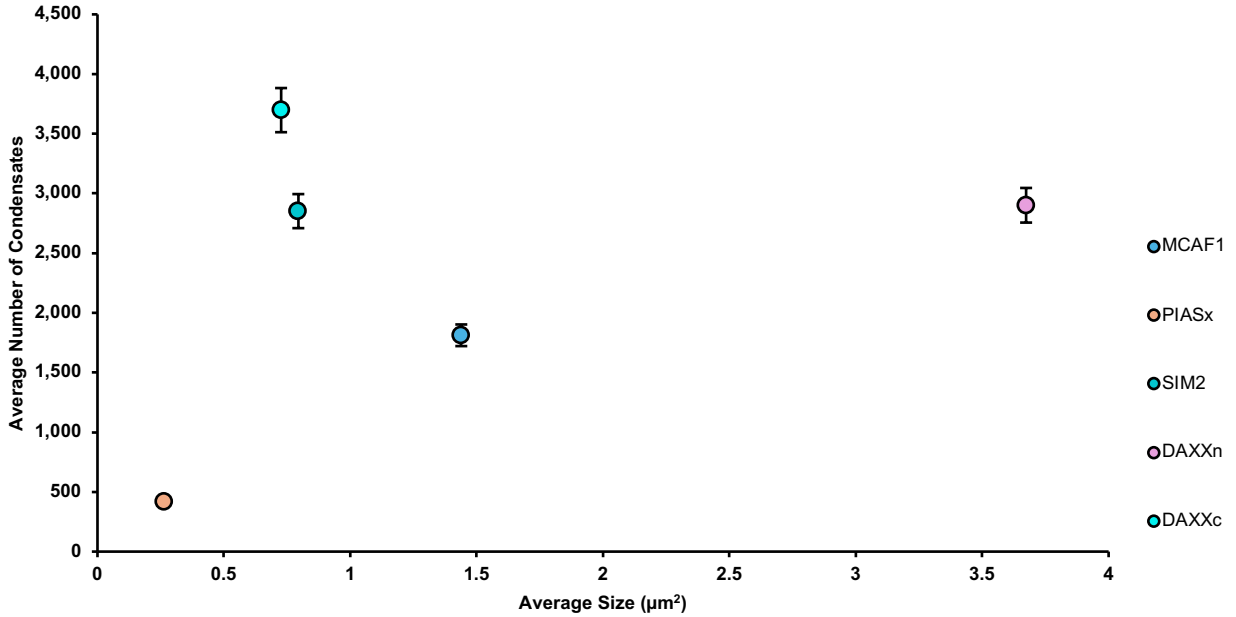
4.4.22 Counterstaining of human cells for confocal microscopy

HeLa, HEK293T, and H1299 cells were washed 3x with warm DPBS prior to counterstaining with Hoechst 33342 (Thermo Scientific) and wheat germ agglutinin conjugated to Alexa Fluor 647 (WGA647, Invitrogen). A mixture of Hoechst 33342 at 1 µg/mL and WGA647 at 5 µg/mL was prepared in warm DPBS. This mixture was added to cells and incubated for 10 min at room temperature, protected from light. Following this, cells were washed 3x with warm DPBS to remove residual stain and returned to warm growth medium. Confocal microscopy was performed immediately after counterstaining.

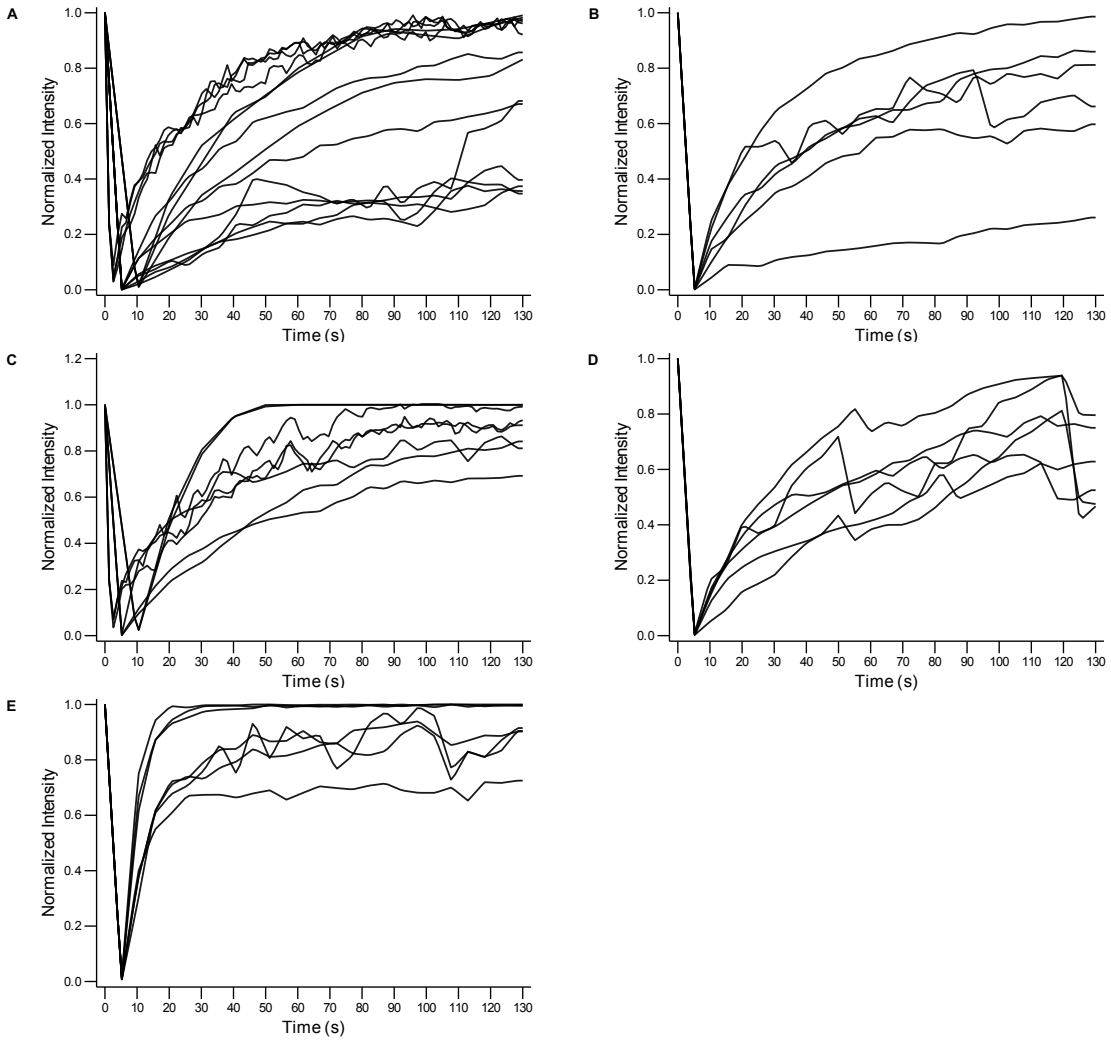
4.5 Product characterization and supplemental data



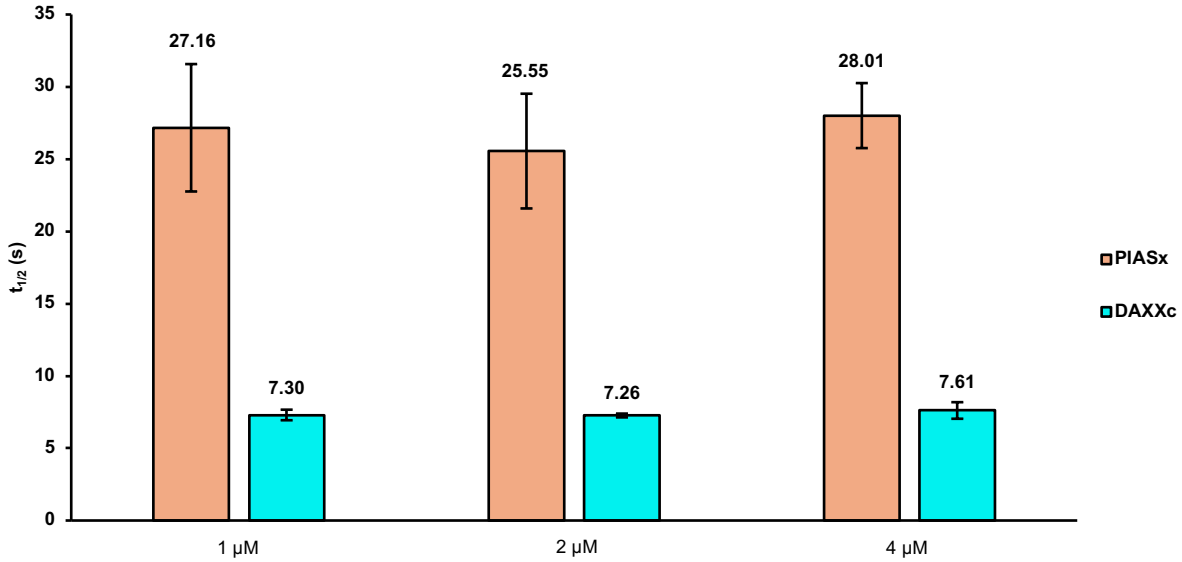
Supplementary Figure S1. SUMO3–SIM K_d is not predictive of condensate size or number in the 3:3 scaffold system. Each point represents a SUMO–SIM pair measured under matched conditions in the fixed-valence 3:3 system. Error bars show standard deviation. No clear relationship between size and number was observed across the panel.



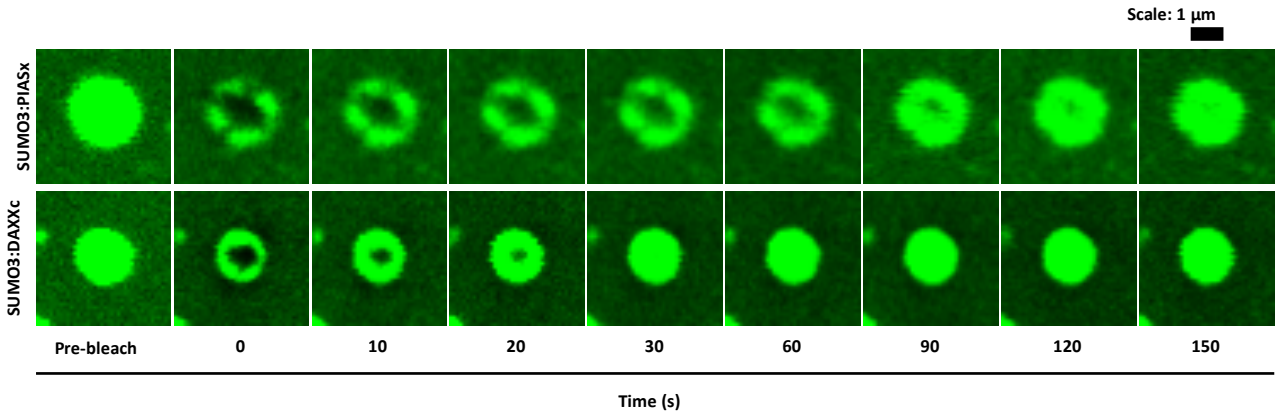
Supplementary Figure S2. SUMO3–SIM Kd is not predictive of condensate size or number in the 5:3 scaffold system. Each point represents a SUMO–SIM pair measured under matched conditions in the fixed-valence 5:3 system. Error bars show standard deviation. No clear relationship between size and number was observed across the panel.



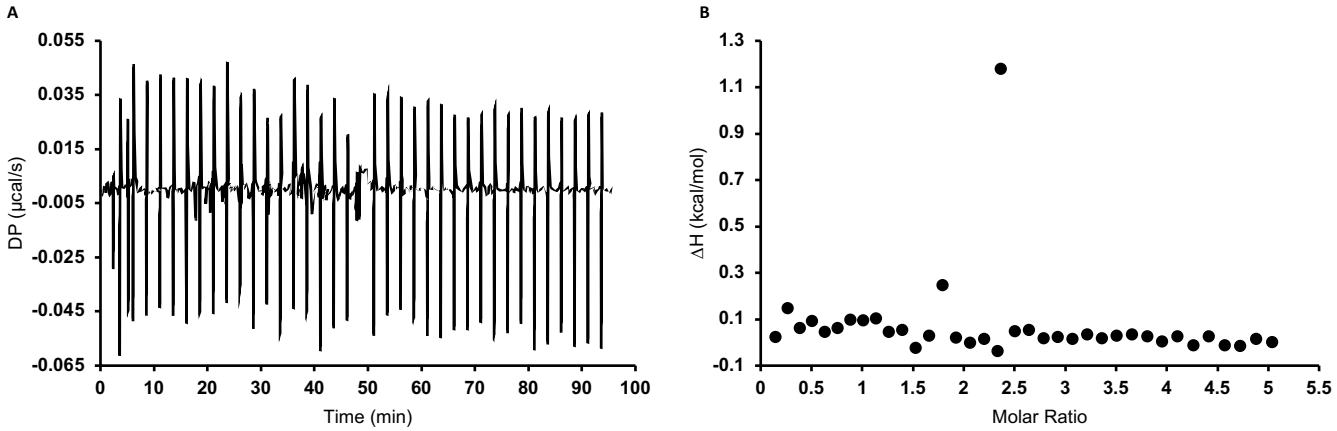
Supplementary Figure S3. Individual FRAP recovery traces for SUMO3–SIM condensates used for quantitation. Panels show single-droplet recovery curves used to extract half-times for the indicated assemblies: (**A**) SUMO:MCAF1, (**B**) SUMO:PIASx, (**C**) SUMO:SIM2, (**D**) SUMO:DAXXn, (**E**) SUMO:DAXXc. A circular region covering ~50% of the visible droplet area was photobleached with a 488 nm laser. Traces are normalized to the pre-bleach frame, and half-times were obtained from single-exponential fits to these primary data.



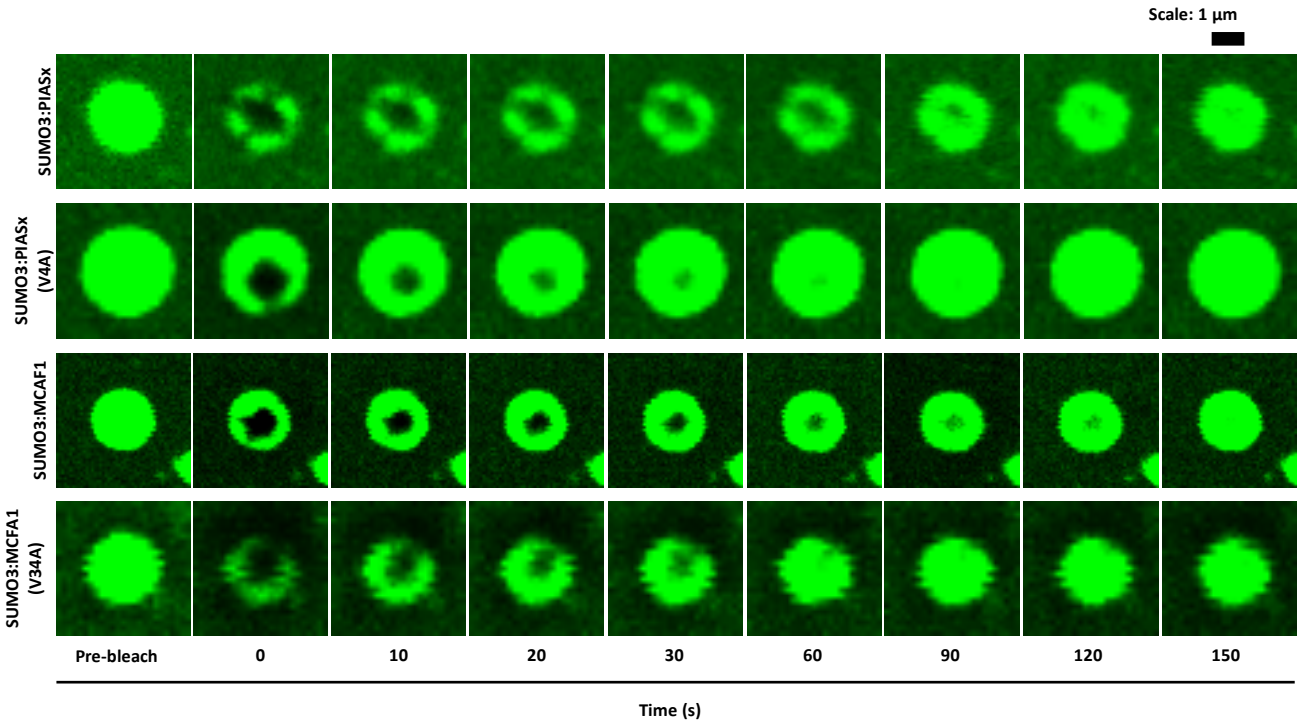
Supplementary Figure S4. FRAP half-times are consistent across all tested concentrations, with DAXXc recovering faster than PIASx. Histograms show FRAP half-times for PIASx and DAXXc SUMO3–SIM droplets at 1, 2, and 4 µM under equimolar scaffold concentrations. PIASx is the tighter binder and DAXXc the weaker binder in the 5:3 system. BMCs were formed at matched input composition. For each condition, the mean half-time was calculated from ≥6 similarly sized condensates (~6 µm²), with ~50% of the visible droplet area photobleached. Error bars denote the standard error of the mean. Half-times were largely unchanged across this concentration range.



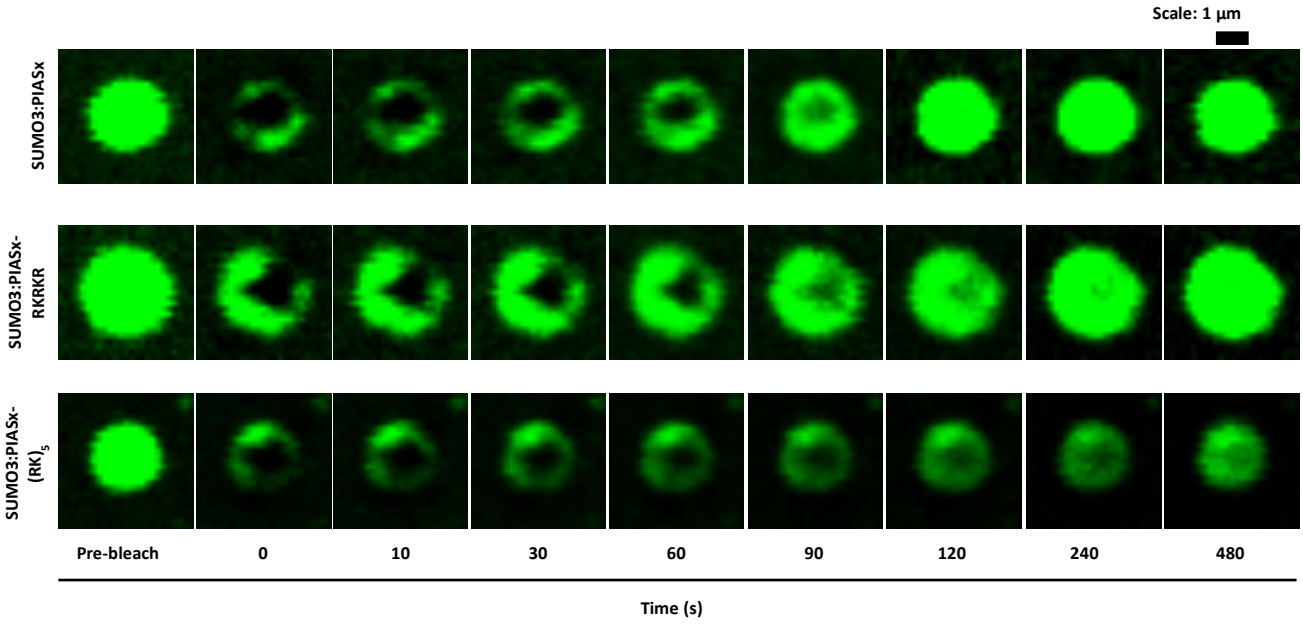
Supplemental Figure S5. Time-lapse FRAP at matched concentration: SUMO:PIASx versus SUMO:DAXXc. Pre-bleach, post-bleach, and recovery frames show FRAP for SUMO:PIASx (top) and SUMO:DAXXc (bottom) condensates at 2 µM per scaffold under matched input composition and buffer. A circular region covering ~50% of the visible droplet area was bleached using a 488 nm laser. Images are displayed on the same intensity scale. Scale bars, 10 µm. Recovery is faster for SUMO:DAXXc than for SUMO:PIASx.



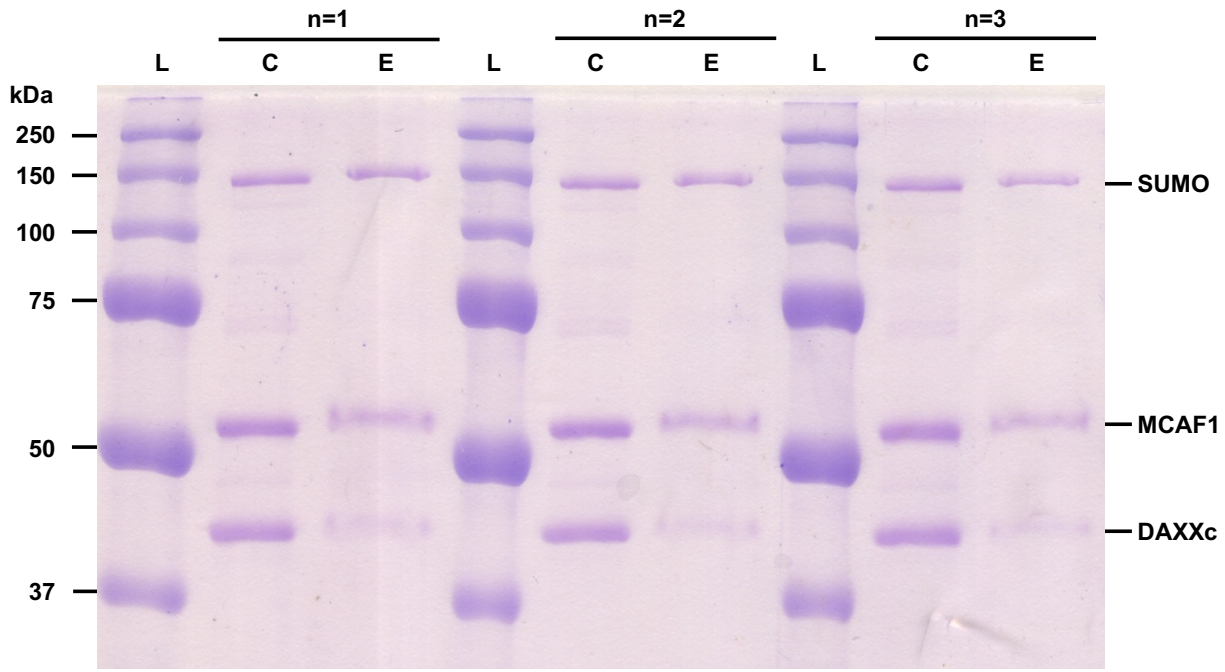
Supplemental Figure S6. PIASx(L7A) does not measurably bind SUMO3 by ITC. Thermogram (**A**) and integrated heats (**B**) were analyzed with a 1:1 heteromeric association model; no fit yielded a finite K_d at the highest peptide concentration tested, consistent with no detectable binding.



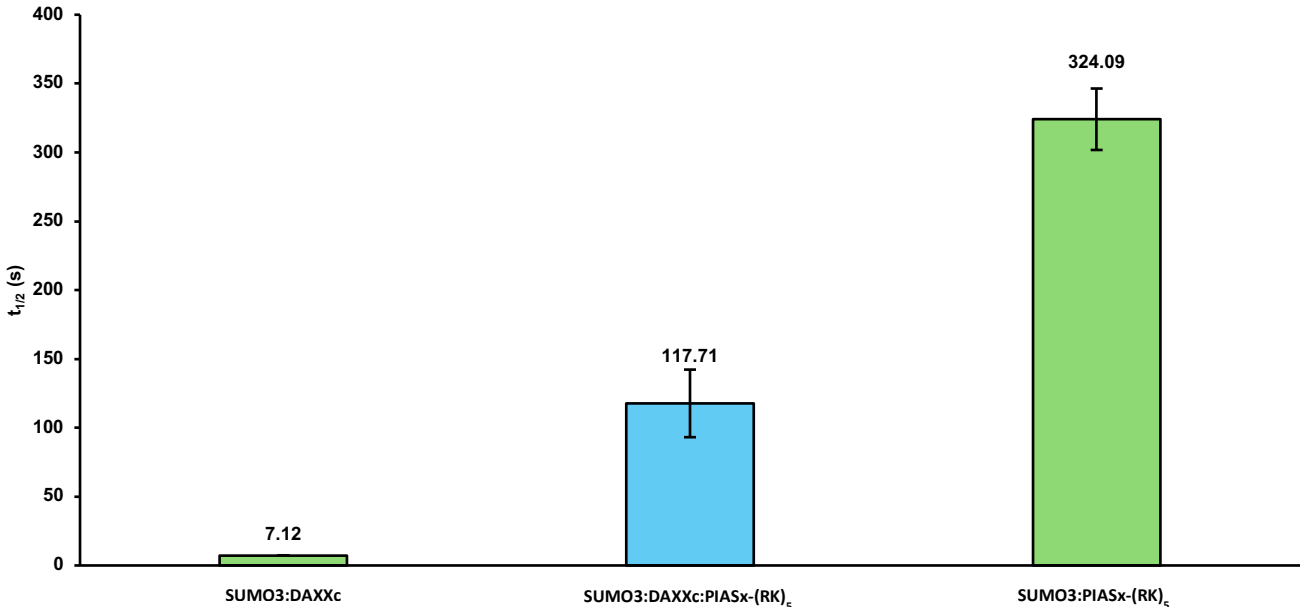
Supplemental Figure S7. Time-lapse FRAP: wild type versus alanine variants for PIASx and MCAF1. Pre-bleach, post-bleach, and recovery frames show FRAP for four SUMO3–SIM condensate types arranged by row: SUMO:PIASx (row 1), SUMO:PIASx(V4A) (row 2), SUMO:MCAF1 (row 3), and SUMO:MCAF1(V34A) (row 4). A circular region covering ~50% of the visible droplet area was bleached using a 488 nm laser. Images are displayed on the same intensity scale. Scale bars, 10 µm. Recovery is faster for the alanine variants than for their wild-type counterparts.



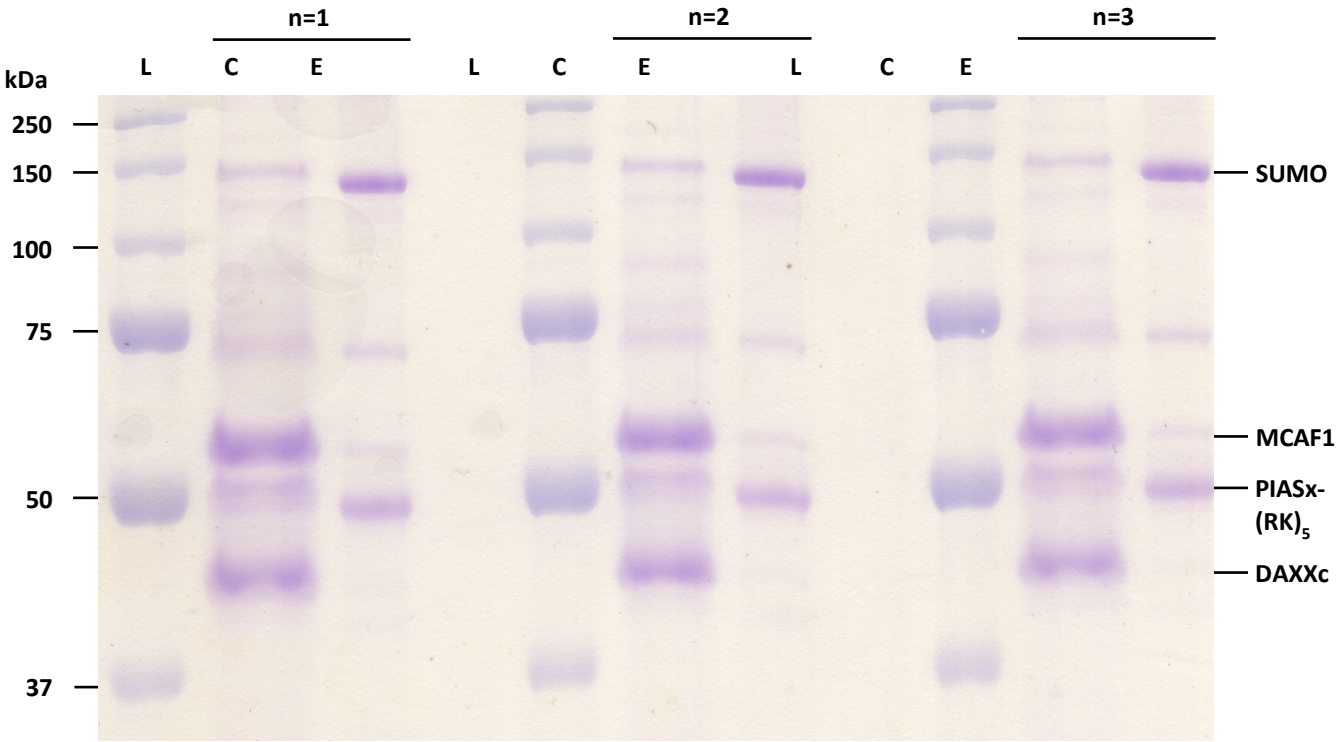
Supplemental Figure S8. Time-lapse FRAP: PIASx versus RK-extended variants. Pre-bleach, post-bleach, and recovery frames show FRAP for three condensate types arranged by row: SUMO:PIASx (top), SUMO:PIASx-RKRKR (middle), and SUMO:PIASx-$(RK)_5$ (bottom). A circular region covering ~50% of the visible droplet area was bleached using a 488 nm laser. Images are displayed on the same intensity scale. Scale bars, 10 µm. Recovery is slower for the RK-extended variants than for PIASx.



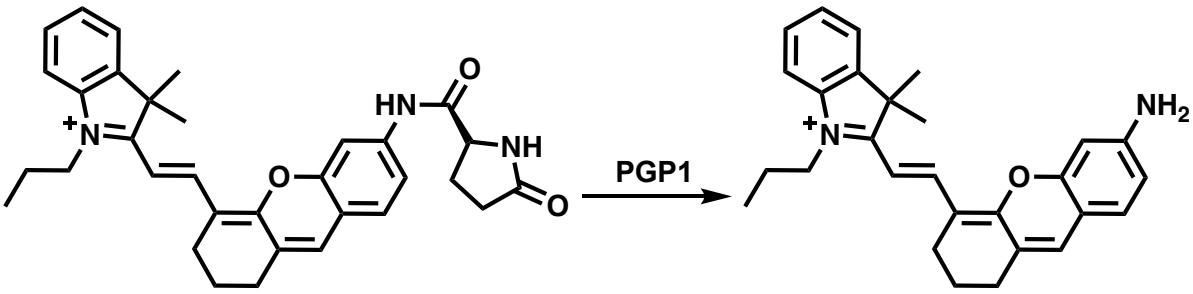
Supplemental Figure S9. SDS-PAGE of sedimented SUMO:DAXXc:MCAF1 condensates shows co-assembly with MCAF1 enrichment. Preformed ternary condensates of mEGFP-$(SUMO3)_5$, mEGFP-$(DAXXc)_3$, and mEGFP-$(MCAF1)_3$, across three replicates, were centrifuged to separate droplets from bulk. SDS-PAGE of experimental lanes shows bands for both MCAF1 and DAXXc, consistent with co-assembly within a single condensed phase rather than selective partitioning. Analyzing protein bands indicates MCAF1 is 2.719 ± 0.284-fold higher than DAXXc under these conditions. L, ladder. C, input control with proteins at equal concentration. E, experimental sample after sedimentation.



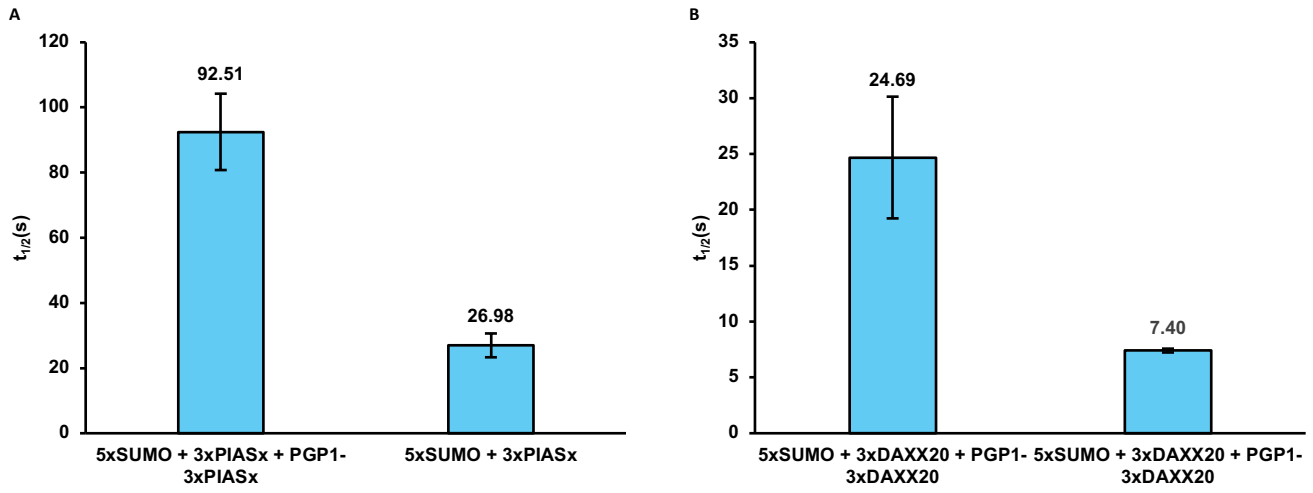
Supplemental Figure S10. Adding PIASx-(RK)$_5$ to preformed SUMO-DAXXc condensates slows FRAP recovery. Histograms show FRAP half-times for SUMO-DAXXc condensates before and after addition of mEGFP-(PIASx)$_3$-(RK)$_5$ as well as SUMO-PIASx-(RK)$_5$ condensates. For each condition, the mean half-time was calculated from ≥6 similarly sized condensates (~6 μm²), with ~50% of the visible droplet area photobleached using a 488 nm laser. Error bars represent standard error of the mean.



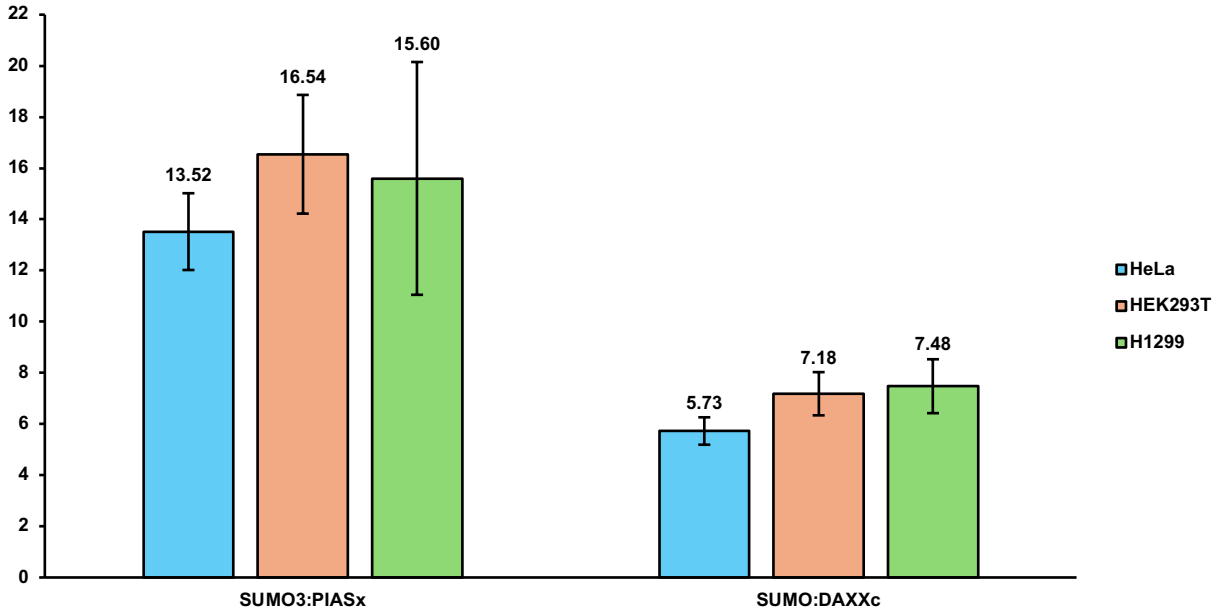
Supplemental Figure S11. SDS-PAGE of sedimented SUMO:DAXXc:MCAF1:PIASx-$(RK)_5$ condensates shows co-assembly with PIASx-$(RK)_5$ enrichment. Preformed quaternary condensates of mEGFP-$(SUMO3)_5$, mEGFP-$(DAXXc)_3$, mEGFP-$(MCAF1)_3$, and mEGFP-$(PIASx)_3$-$(RK)_5$ across three replicates, were centrifuged to separate droplets from bulk. SDS-PAGE of experimental lanes shows bands for both MCAF1 and PIASx-$(RK)_5$, indicating enrichment of both components and competing out the lowest affinity partner, DAXXc. L, ladder. C, input control with proteins at equal concentration. E, experimental sample after sedimentation.



Supplemental Figure S12. pGlu-HCA substrate and released HCA product by PGP1 catalysis. Left, pGlu-HCA containing an N-terminal L-pyroglutamate. Right, HCA generated after PGP1-catalyzed hydrolysis of the pyroglutamate. Product fluorescence was monitored at λ_{ex} 670 nm and λ_{em} 700 nm.



Supplemental Figure S13. PGP1 recruitment slows FRAP relative to enzyme-free condensates, with SUMO:PIASx remaining slower than SUMO:DAXXc. (**A**) FRAP half-times for SUMO:PIASx condensates with and without PGP1. (**B**) FRAP half-times for SUMO:DAXXc condensates with and without PGP1. For each condition, ~50% of the visible droplet area was photobleached using a 488 nm laser, and recovery was fit to a single-exponential to obtain $t_{1/2}$. Bars show mean ± standard error of the mean. PGP1 recruitment slows recovery in both systems, and the relative order is preserved with SUMO:PIASx recovering more slowly than SUMO:DAXXc.



Supplemental Figure S14. FRAP recoveries for PIASx and DAXXc condensates are consistent across HeLa, HEK293T, and H1299 cells. Histograms show FRAP half-times for mCherry-$(SUMO3)_5$-$(PIASx)_3$ and mCherry-$(SUMO3)_5$-$(DAXXc)_3$ condensates in HeLa, HEK293T, and H1299 cells. Bars indicate mean ± standard error of the mean. For each condition, n=6 similarly sized condensates were analyzed with ~50% of the visible droplet area photobleached. Within each construct, half-times were similar across cell lines.

4.6 References

1 Banani, S. F. *et al.* Compositional Control of Phase-Separated Cellular Bodies. *Cell* **166**, 651–663 (2016). https://doi.org/10.1016/j.cell.2016.06.010

2 Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Biomolecular condensates: organizers of cellular biochemistry. *Nat Rev Mol Cell Biol* **18**, 285–298 (2017). https://doi.org/10.1038/nrm.2017.7

3 Patel, A. *et al.* A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* **162**, 1066–1077 (2015). https://doi.org/10.1016/j.cell.2015.07.047

4 Wang, J. *et al.* A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* **174**, 688–699 e616 (2018). https://doi.org/10.1016/j.cell.2018.06.006

5 Lallemand-Breitenbach, V. & de The, H. PML nuclear bodies: from architecture to function. *Curr Opin Cell Biol* **52**, 154–161 (2018). https://doi.org/10.1016/j.ceb.2018.03.011

6 Hyman, A. A., Weber, C. A. & Julicher, F. Liquid-liquid phase separation in biology. *Annu Rev Cell Dev Biol* **30**, 39–58 (2014). https://doi.org/10.1146/annurev-cellbio-100913-013325

7 Larson, A. G. *et al.* Liquid droplet formation by HP1alpha suggests a role for phase separation in heterochromatin. *Nature* **547**, 236–240 (2017). https://doi.org/10.1038/nature22822

8 Elbaum-Garfinkle, S. *et al.* The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. *Proc Natl Acad Sci U S A* **112**, 7189–7194 (2015). https://doi.org/10.1073/pnas.1504822112

9 Nott, T. J. *et al.* Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Mol Cell* **57**, 936–947 (2015). https://doi.org/10.1016/j.molcel.2015.01.013

10 Ditlev, J. A., Case, L. B. & Rosen, M. K. Who's In and Who's Out-Compositional Control of Biomolecular Condensates. *J Mol Biol* **430**, 4666–4684 (2018). https://doi.org/10.1016/j.jmb.2018.08.003

11 Hattersley, N., Shen, L., Jaffray, E. G. & Hay, R. T. The SUMO protease SENP6 is a direct regulator of PML nuclear bodies. *Mol Biol Cell* **22**, 78–90 (2011). https://doi.org/10.1091/mbc.E10-06-0504

12 Keiten-Schmitz, J., Roder, L., Hornstein, E., Muller-McNicoll, M. & Muller, S. SUMO: Glue or Solvent for Phase-Separated Ribonucleoprotein Complexes and Molecular Condensates? *Front Mol Biosci* **8**, 673038 (2021). https://doi.org/10.3389/fmolb.2021.673038

13 Martin, E. W. *et al.* Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science* **367**, 694–699 (2020). https://doi.org/10.1126/science.aaw8653

14 Banerjee, P. R., Milin, A. N., Moosa, M. M., Onuchic, P. L. & Deniz, A. A. Reentrant Phase Transition Drives Dynamic Substructure Formation in Ribonucleoprotein Droplets. *Angew Chem Int Ed Engl* **56**, 11354–11359 (2017). https://doi.org/10.1002/anie.201703191

15 Ambadi Thody, S. *et al.* Small-molecule properties define partitioning into biomolecular condensates. *Nat Chem* **16**, 1794–1802 (2024). https://doi.org/10.1038/s41557-024-01630-w

16 Alberti, S., Gladfelter, A. & Mittag, T. Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* **176**, 419–434 (2019). https://doi.org/10.1016/j.cell.2018.12.035

17 Li, P. *et al.* Phase transitions in the assembly of multivalent signalling proteins. *Nature* **483**, 336–340 (2012). https://doi.org/10.1038/nature10879
18 Banjade, S. & Rosen, M. K. Phase transitions of multivalent proteins can promote clustering of membrane receptors. *Elife* **3** (2014). https://doi.org/10.7554/eLife.04123
19 Hecker, C. M., Rabiller, M., Haglund, K., Bayer, P. & Dikic, I. Specification of SUMO1- and SUMO2-interacting motifs. *J Biol Chem* **281**, 16117–16127 (2006). https://doi.org/10.1074/jbc.M512757200
20 Namanja, A. T. *et al.* Insights into high affinity small ubiquitin-like modifier (SUMO) recognition by SUMO-interacting motifs (SIMs) revealed by a combination of NMR and peptide array analysis. *J Biol Chem* **287**, 3231–3240 (2012). https://doi.org/10.1074/jbc.M111.293118
21 Xu, Y. *et al.* Structural insight into SUMO chain recognition and manipulation by the ubiquitin ligase RNF4. *Nat Commun* **5**, 4217 (2014). https://doi.org/10.1038/ncomms5217
22 Yau, T. Y., Sander, W., Eidson, C. & Courey, A. J. SUMO Interacting Motifs: Structure and Function. *Cells* **10** (2021). https://doi.org/10.3390/cells10112825
23 Maharana, S. *et al.* RNA buffers the phase separation behavior of prion-like RNA binding proteins. *Science* **360**, 918–921 (2018). https://doi.org/10.1126/science.aar7366
24 Patel, A. *et al.* ATP as a biological hydrotrope. *Science* **356**, 753–756 (2017). https://doi.org/10.1126/science.aaf6846
25 Sekiyama, N. *et al.* Structure of the small ubiquitin-like modifier (SUMO)-interacting motif of MBD1-containing chromatin-associated factor 1 bound to SUMO-3. *J Biol Chem* **283**, 35966–35975 (2008). https://doi.org/10.1074/jbc.M802528200
26 Kung, C. C. *et al.* Structural analysis of poly-SUMO chain recognition by the RNF4-SIMs domain. *Biochem J* **462**, 53–65 (2014). https://doi.org/10.1042/BJ20140521
27 Choi, J. M., Holehouse, A. S. & Pappu, R. V. Physical Principles Underlying the Complex Biology of Intracellular Phase Transitions. *Annu Rev Biophys* **49**, 107–133 (2020). https://doi.org/10.1146/annurev-biophys-121219-081629
28 Escobar-Cabrera, E. *et al.* Characterizing the N- and C-terminal Small ubiquitin-like modifier (SUMO)-interacting motifs of the scaffold protein DAXX. *J Biol Chem* **286**, 19816–19829 (2011). https://doi.org/10.1074/jbc.M111.231647
29 Chang, C. C. *et al.* Structural and functional roles of Daxx SIM phosphorylation in SUMO paralog-selective binding and apoptosis modulation. *Mol Cell* **42**, 62–74 (2011). https://doi.org/10.1016/j.molcel.2011.02.022
30 Sun, Y. *et al.* Phase-separating peptide coacervates with programmable material properties for universal intracellular delivery of macromolecules. *Nat Commun* **15**, 10094 (2024). https://doi.org/10.1038/s41467-024-54463-z
31 Xu, B. *et al.* Rigidity enhances a magic-number effect in polymer phase separation. *Nat Commun* **11**, 1561 (2020). https://doi.org/10.1038/s41467-020-15395-6
32 Greiner, J. V. & Glonek, T. Intracellular ATP Concentration and Implication for Cellular Evolution. *Biology (Basel)* **10** (2021). https://doi.org/10.3390/biology10111166
33 Sprague, B. L. & McNally, J. G. FRAP analysis of binding: proper and fitting. *Trends Cell Biol* **15**, 84–91 (2005). https://doi.org/10.1016/j.tcb.2004.12.001
34 Taylor, N. O., Wei, M. T., Stone, H. A. & Brangwynne, C. P. Quantifying Dynamics in Phase-Separated Condensates Using Fluorescence Recovery after Photobleaching. *Biophys J* **117**, 1285–1300 (2019). https://doi.org/10.1016/j.bpj.2019.08.030
35 Zhang, Y. *et al.* The exchange dynamics of biomolecular condensates. *Elife* **12** (2024). https://doi.org/10.7554/eLife.91680
36 Song, J., Durrin, L. K., Wilkinson, T. A., Krontiris, T. G. & Chen, Y. Identification of a SUMO-binding motif that recognizes SUMO-modified proteins. *Proc Natl Acad Sci U S A* **101**, 14373–14378 (2004). https://doi.org/10.1073/pnas.0403498101

37 Vazquez, D. S., Toledo, P. L., Gianotti, A. R. & Ermacora, M. R. Protein conformation and biomolecular condensates. *Curr Res Struct Biol* **4**, 285–307 (2022). https://doi.org/10.1016/j.crstbi.2022.09.004
38 Folkmann, A. W., Putnam, A., Lee, C. F. & Seydoux, G. Regulation of biomolecular condensates by interfacial protein clusters. *Science* **373**, 1218–1224 (2021). https://doi.org/10.1126/science.abg7071
39 Ding, H. *et al.* Solution structure of human SUMO-3 C47S and its binding surface for Ubc9. *Biochemistry* **44**, 2790–2799 (2005). https://doi.org/10.1021/bi0477586
40 Corpet, A. *et al.* PML nuclear bodies and chromatin dynamics: catch me if you can! *Nucleic Acids Res* **48**, 11890–11912 (2020). https://doi.org/10.1093/nar/gkaa828
41 Lin, D. Y. *et al.* Role of SUMO-interacting motif in Daxx SUMO modification, subnuclear localization, and repression of sumoylated transcription factors. *Mol Cell* **24**, 341–354 (2006). https://doi.org/10.1016/j.molcel.2006.10.019
42 Sasai, N., Saitoh, N., Saitoh, H. & Nakao, M. The transcriptional cofactor MCAF1/ATF7IP is involved in histone gene expression and cellular senescence. *PLoS One* **8**, e68478 (2013). https://doi.org/10.1371/journal.pone.0068478
43 Cummins, P. M. & O'Connor, B. Pyroglutamyl peptidase: an overview of the three known enzymatic forms. *Biochim Biophys Acta* **1429**, 1–17 (1998). https://doi.org/10.1016/s0167-4838(98)00248-9
44 Gong, Q. *et al.* A Ultrasensitive Near-Infrared Fluorescent Probe Reveals Pyroglutamate Aminopeptidase 1 Can Be a New Inflammatory Cytokine. *Adv Sci (Weinh)* **5**, 1700664 (2018). https://doi.org/10.1002/advs.201700664
45 Lange, A. *et al.* Classical nuclear localization signals: definition, function, and interaction with importin alpha. *J Biol Chem* **282**, 5101–5105 (2007). https://doi.org/10.1074/jbc.R600026200
46 Kosugi, S. *et al.* Six classes of nuclear localization signals specific to different binding grooves of importin alpha. *J Biol Chem* **284**, 478–485 (2009). https://doi.org/10.1074/jbc.M807017200
47 Rivas, G. & Minton, A. P. Macromolecular Crowding *In vitro*, *In vivo*, and In Between. *Trends Biochem Sci* **41**, 970–981 (2016). https://doi.org/10.1016/j.tibs.2016.08.013
48 Mittag, T. & Pappu, R. V. A conceptual framework for understanding phase separation and addressing open questions and challenges. *Mol Cell* **82**, 2201–2214 (2022). https://doi.org/10.1016/j.molcel.2022.05.018
49 Shin, Y. *et al.* Spatiotemporal Control of Intracellular Phase Transitions Using Light-Activated optoDroplets. *Cell* **168**, 159–171 e114 (2017). https://doi.org/10.1016/j.cell.2016.11.054
50 Collins, J. M. *et al.* Total wash elimination for solid phase peptide synthesis. *Nat Commun* **14**, 8168 (2023). https://doi.org/10.1038/s41467-023-44074-5
51 He, X. *et al. In vivo* imaging of leucine aminopeptidase activity in drug-induced liver injury and liver cancer via a near-infrared fluorescent probe. *Chem Sci* **8**, 3479–3483 (2017). https://doi.org/10.1039/c6sc05712h