REGULATION OF CHROMATIN-ASSOCIATED PROTEINS BY SUMOYLATION

APPROVED BY SUPERVISORY COMMITTEE

Hongtao Yu, Ph.D.

_____

Joel Goodman, Ph.D.

_____

Elliott Ross, Ph.D.

_____

Michael White, Ph.D.

_____

Acknowledgements

There are many people that I would like to thank for all of their support during my years as a graduate student. First and foremost, I would like to thank my mentor, Hongtao Yu, who has been patient and fair in all situations. Moreover, he has allowed me to function independently and follow my own ideas, which has allowed me to learn more than if I was told what to do. Despite these characterisics which you would find in a large lab, Hongtao has been very accessible, always being available and willing to discuss data. I have learned much as a student in his lab.

I would also like to thank the lab members, past and present, for all their help and guidance. Especially Rajnish Bawardwaj, Josh Bembenek, Jungseog Kang, Zhanyun Tang, Wei Qi, Ryan Potts, Maojun Yang, and of course, Bing Li. They have all helped me significantly in one way or another. I would like to especially acknowledge Jungseog and Maojun, since I worked with them personally on several projects.

Probably the people who need to be acknowledged most are my parents, Terry and Michael Gocke, who made sure that I learned important values such as focus, hard work, and goal setting at a young age. Additionally, they made available education that I otherwise might not have had. Without them I would have accomplished nothing.

By observing their own experiences and taking their words to heart, my three older brothers have taught me what to do and what not to do over the years. I do not think I would have made it through all the changes that I have experienced since highschool without their influence. I especially thank my brother, Tim, and his wife, Fizzah, since they moved here to Dallas from West Virginia. It has made Texas much more home-like! Also, I need to thank my friends from Dallas: Rich, Rose, Pritam, Chad, and Jen, who made Dallas a fun place to be!

Finally, I wish to thank my wife, Annie, who has always been a source of emotional and mental support. She transferred here from the beautiful city of Charleston, SC, and endured two extra years of her thesis just to be with me. I look forward to many more years of such support and hope that I can bring the same to her research.

REGULATION OF CHROMATIN-ASSOCIATED PROTEINS BY SUMOYLATION

By

CHRISTIAN BURRIS GOCKE

DISSERTATION

Presented to the Faculty of the Graduate School of Biomedical Sciences

The University of Texas Southwestern Medical Center at Dallas

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas Southwestern Medical Center at Dallas

Dallas, Texas

June, 2007

REGULATION OF CHROMATIN-ASSOCIATED PROTEINS BY SUMOYLATION

Christian Burris Gocke, Ph.D.

The University of Texas Southwestern Medical Center at Dallas, 2006

Supervising Professor: Hongtao Yu, Ph.D.

Small Ubiquitin-like Modifier (SUMO) regulates diverse cellular processes through its reversible, covalent attachment to target proteins. Many SUMO substrates are involved in transcription and chromatin structure. Sumoylation appears to regulate the functions of target proteins by changing their subcellular localization, increasing their stability, and/or mediating their binding to other proteins. Using an *In Vitro* Expression Cloning (IVEC) approach, we have identified 40 human SUMO1 substrates. We have validated the sumoylation of 24 substrates in living cells. We show that one of these substrates, Mef2C, is coordinately regulated by phosphorylation and sumoylation. The spectrum of human SUMO1 substrates identified in our screen suggests general roles of sumoylation in transcription, chromosome structure, and RNA processing.

Moreover, multiple subunits of a given chromatin-associated complex are targets for SUMO-conjugation. For example, a substrate identified in our screen, lysine-specific demethylase 1 (LSD1), is part of a complex that also contains Histone Deacetylase 1 (HDAC1), that is a SUMO substrate. This prompted me to study the function of this complex and its regulation by SUMO.

Histone methylation regulates diverse chromatin-templated processes, including transcription. Many transcriptional corepressor complexes contain LSD1 and CoREST that collaborate to demethylate mono- and di-methylated histone H3 lysine 4 (H3K4) of nucleosomes. We have determined the crystal structure of the LSD1–CoREST complex. LSD1–CoREST forms an elongated structure with a long stalk connecting the catalytic domain of LSD1 and the CoREST SANT2 domain. LSD1 recognizes a large segment of the H3 tail through a deep, negatively charged pocket at the active site and possibly a shallow groove on its surface. CoREST SANT2 interacts with DNA. Disruption of the SANT2–DNA interaction diminishes CoREST-dependent demethylation of nucleosomes by LSD1. The shape and dimension of LSD1–CoREST suggest its bivalent binding to nucleosomes, allowing efficient H3-K4 demethylation. This spatially separated, multivalent nucleosome-binding mode may apply to other chromatin-modifying enzymes that generally contain multiple nucleosome-binding modules.

The core CoREST corepressor complex, consisting of CoREST, LSD1, and HDAC1/2, represses transcription by coordinately removing histone modifications associated with gene activation. ZNF198 and other MYM-type zinc-finger proteins are also components of this complex. ZNF198, HDAC1, and LSD1 are SUMO substrates, and ZNF198 binds to SUMO non-covalently. We show that ZNF198 and its homologues do not regulate REST-responsive genes. Consistently, binding of REST and ZNF198 to CoREST are mutually exclusive. However, these MYM-domain proteins are required for tethering LSD1 to nuclear compartments and for repression of E-cadherin, a non-REST responsive gene. ZNF198 interacts efficiently only with the intact LSD1-CoREST-HDAC1 ternary complex, but not its individual subunits. ZNF198 also binds specifically to sumoylated, but not unsumoylated HDAC1. These interactions are mediated by tandem zinc-fingers of ZNF198. HDAC1 activity is not stimulated by sumoylation or ZNF198 binding. Sumoylated HDAC1 does not interact with CoREST, and LSD1 sumoylation is inhibited by CoREST binding. Therefore, ZNF198, through its unique and diverse protein-protein interactions, helps to maintain the intact CoREST complex on specific promoters.

Table of Contents

Prior Publications

Yang M, Culhane JC, Szewczuk LM, **Gocke CB**, Brautigam CA, Tomchick DR, Machius M, Cole PA, Yu H. Structural basis of histone demethylation by LSD1 revealed by suicide inactivation. *Nat Struct Mol Biol*. 2007 May 27; [Epub ahead of print]

Yang M, **Gocke CB**, Luo X, Borek D, Tomchick DR, Machius M, Otwinowski Z, Yu H. Structural basis for CoREST-dependent demethylation of nucleosomes by the human LSD1 histone demethylase. *Mol Cell*. 2006 Aug 4;23(3):377-87.

Kang J, **Gocke CB**, Yu H. Phosphorylation-facilitated sumoylation of MEF2C negatively regulates its transcriptional activity. *BMC Biochem*. 2006 Feb 14;7:5.

Gocke CB, Yu H, Kang J. Systematic identification and analysis of mammalian small ubiquitin-like modifier substrates. *J Biol Chem*. 2005 Feb 11;280(6):5004-12. Epub 2004 Nov 23.

Zheng ZM, Quintero J, Reid ES, **Gocke C**, Baker CC. Optimization of a weak 3' splice site counteracts the function of a bovine papillomavirus type 1 exonic splicing suppressor in vitro and in vivo. *J Virol*. 2000 Jul;74(13):5902-10.

Obungu VH, Wang Y, Amyot SM, **Gocke CB**, Beattie DS. Mutations in the tether region of the iron-sulfur protein affect the activity and assembly of the cytochrome bc(1) complex of yeast mitochondria. *Biochim Biophys Acta*. 2000 Feb 24;1457(1-2):36-44.

List of Figures

List of Tables

Abbreviations

μg: micrograms
μl: microliters
μM: micromolar
Ab: antibody
Aos1: activation of Smt3p
ATP: adenosine triphosphate
BSA: bovine serum albumin
cDNA: complementary DNA
ChIP: chromatin immunoprecipitation
DMSO: dimethyl sulfoxide
DNA: deoxyribonucleic acid
ECL: enhanced chemiluminescent reagent
EDTA: ethylenediaminetetra acetic acid
ETOH: ethanol
FBS: fetal bovine serum
FITC: fluorescein isothiocyanate
GFP: green fluorescent protein
HDAC: histone deacetylase
HRP: horseradish peroxidase
Ig: immunoglobulin
IP: immunoprecipitation
IVEC: in vitro expression cloning
kDa: kilodalton
MAP kinase: mitogen activated protein kinase
MEF2: myocyte enhancer factor 2
ml: milliliter
mRNA: messenger RNA
NFkB: nuclear factor kappa B
OD: optical density
PAGE: polyacrylamide gel electrophoresis
PBS: phosphate buffered saline
PCR: polymerase chain reaction
PIAS: protein inhibitor of activated STAT
PML: promyelocytic leukemia protein
PPAR: peroxisome proliferator activated receptor
recombinase activating gene
RNAi: RNA interference
RT-PCR: reverse transcriptase polymerase chain reaction

SDS: sodium dodecyl sulfate
SEM: standard error of the mean
SENP: sentrin-specific protease
siRNA: small interfering RNA
SRC-1: steroid receptor co-activator 1
SUMO: small ubiquitin-like modifier
TE: Tris EDTA
Uba2: ubiquitin-activating enzyme 2
Ubc9: ubiquitin-conjugating enzyme 9
w/w: weight per weight
WT: wildtype

Chapter I: Introduction

Proteins undergo many post-translational modifications, such as proline hydroxylation, serine/threonine or tyrosine phosphorylation, lysine acetylation, and lysine and arginine methylation. Such modifications dramatically change protein function by affecting protein folding and conformation, or creating and destroying interactions with other macromolecules. Therefore, understanding protein modifications that exist in the cell, including the sites and identity of modifications, the mechanism and regulation of attachment and removal, and the downstream consequences of these modifications is essential to understanding cellular biology. We focus on two such modifications: Sumoylation and lysine methylation. We study these modifications in the context of a transcriptional corepressor complex containing lysine specific demethylase 1 (LSD1), REST-corepressor (CoREST), and histone deacetylase 1/2 (HDAC1/2).

Part A: SUMO Background

Small Ubiquitin-Like Modifiers

Ubiquitin is a small protein that can be covalently attached to lysines of other proteins. Attachment of ubiquitin to a substrate (referred to as ubiquitination) in most cases targets proteins for proteasomal degradation (1,2). A large family of ubiquitin-like proteins that share sequence similarity with ubiquitin has been described (3,4). These include ISG15, Nedd8, and small ubiquitin-like modifier (SUMO). SUMO (also known as Sentrin, Smt3, GMP1, PIC1) is conserved from yeast to humans and affects many cellular processes (5,6). Multiple SUMO isoforms are known to exist in humans. SUMO1 is a 101 amino acid protein with 18% identity to ubiquitin. SUMO2 and SUMO3 are nearly identical to each other, but they are only ~45% identical to SUMO1. These isoforms use the same enzymes for attachment and have distinct but overlapping biochemical and biological properties (5).

The Enzymology of SUMO Attachment

The enzymology of sumoylation is similar to ubiquitination (5), requiring a SUMO-activating enzyme (Aos1-Uba2), SUMO-conjugating enzyme (Ubc9), SUMO ligases, and multiple SUMO isopeptidase/proteases (SENPs) (Figure 1). First, cleavage of SUMO precursors by SENP exposes a carboxy-terminal di-

glycine motif. In an ATP-dependent reaction, Aos1-Uba2 heterodimer (E1) initiates thioester bond formation with this di-glycine motif and transfers SUMO to Ubc9 (E2) in a transesterification reaction. Ubc9 in turn transfers SUMO to the ε-amino group of a lysine residue in a substrate, forming an isopeptide bond. Unlike most ubiquitination reactions, efficient *in vitro* sumoylation does not require a SUMO ligase (E3 ligase) because Ubc9 directly binds to the consensus motif ΨKXE (Ψ, a large hydrophobic residue; X, any amino acid) on the substrate (7). However, many SUMO ligases enhance the rate of sumoylation and control substrate specificity (5,8-11).

Many SUMO ligases have been identified in mammals. The SP-RING (Siz1/PIAS-RING) domain family of SUMO ligases including MMS21 (12) and PIASxα/β, PIAS1, PIASy, and PIAS3 (13,14) shares sequence homology with the RING (Really Interesting New Gene) domain family of ubiquitin ligases. SP-RING domains likely directly stimulate SUMO-conjugation activity of Ubc9, similar to RING (15). PIAS proteins and their substrates often have functions in transcription and chromosome biology (10). MMS21 ligase specializes in promoting genome integrity (12,16). RanBP2 and Pc2 are SUMO ligases (17,18) that do not contain SP-RING domains. RanBP2 functions at nuclear pores (19) and kinetochores (20). Pc2 is part of a large transcriptional repressor complex (18). A crystal structure of RanBP2 bound to both SUMO1-RanGAP1 and Ubc9

suggests RanBP2 stimulates sumoylation by interacting non-covalently with both SUMO1 and Ubc9, positioning them for optimal sumoylation of substrates (21).

Poly-Sumoylation and SUMO Isopeptidases

Most SUMO ligases undergo efficient auto-sumoylation *in vitro*, forming conjugates that contain SUMO chains (poly-sumoylation; Figure 1) (14,17). Poly-sumoylation of SUMO1 and SUMO2/3 occurs on many substrates *in vitro* (17,22). However, only SUMO2/3 have been shown to undergo poly-sumoylation *in vivo* (23-25). Chain formation of SUMO2/3 appears to be important for PML body formation *in vivo* (see below) (23). In *S. cerevisiae*, deletion of the SUMO isopeptidase, Ulp2, leads to toxic accumulation of SUMO chains. Mutation of the lysine that is responsible for chain formation on Smt3p (yeast SUMO) partially rescues this phenotype (26).

SUMO conjugation can have dramatic effects on protein function. However, the steady state levels of sumoylation of a given substrate are low in cells. One explanation for this is that sumoylation is dynamically kept in check by SUMO isopeptidases (SENPs 1-7) (27). SENPs have distinct sub-cellular localization patterns. For example, SENP1 is localized to PML nuclear bodies (28), SENP2 to the nuclear pore (29), SENP3 to the nucleolus (30), and SENP6 to the nucleoplasm (23). In addition to their regulation by localization, SUMO isopeptidases show substrate preferences for SUMO isoforms (27). These

differences can be manifested in SUMO isoform specific maturation and/or SUMO isopeptidase activity. Additionally, as mentioned with Ulp2, some isopeptidases can preferentially de-conjugate SUMO-chains (Figure 1) (26,27). SENP6 (also known as SUSP1) was recently shown to prevent the accumulation of SUMO2/3 chains, especially in PML bodies, but has no effect on SUMO1 (23). SENP1 and SENP2 can remove both SUMO1 and SUMO2/3 from substrates, although the K_m and K_{cat} values for different SUMO proteins can vary significantly (27).

Function of SUMO

SUMO-conjugation can have dramatic and diverse effects on substrates (5,8-11). In principle, sumoylation could function through one or a combination of several general mechanisms: 1) competition with other post-translational modifications; 2) direct changes in substrate protein structure; or 3) alterations in affinities for binding partners. For example, conjugation of SUMO to the same lysines that are normally ubiquitinated stabilizes Huntingtin (31) and IκB (32). Alternatively, when SUMO is attached to the nucleotide excision repair enzyme, thymine DNA glycosylase (TDG), SUMO alters its fold, triggering release from the abasic product (33). Sumoylation of RanGAP1 promotes its localization to the nuclear pore by association with RanBP2 (6,34).

Non-covalent SUMO binding is a quality possessed by many proteins including RanBP2, PIAS proteins (13), SENPs (27), Uba2 (35), Ubc9 (36), PML (37), and Daxx (38). Structural and bioinformatic analysis of these and other SUMO-binding proteins yielded a SUMO-interacting consensus motif (SIM) (35,37,39). The basic core of this consensus is [V/I]-X-[V/I]-[V/I] (35). However, an expanded motif K-X_{3-5}-[V/I]-[I/L]-[I/L]-X_3-[D/E/Q/N]-[D/E]-[D/E] has also been defined (39). The SIM forms an extended structure that often extends a beta-sheet found on the conserved surface of SUMO (see below), as is the case for TDG (33) and RanBP2 (21). This binding mode is different from that adopted by ubiquitin-binding motifs (35).

SUMO Substrates and Their Localization

Recently, SUMO-interacting motifs were shown to mediate the formation of PML nuclear bodies (also ND10; PODs) (38,40). PML bodies are detergent-insoluble nuclear sub-domains (41) that host proteins with diverse cellular functions, such as transcription, DNA repair, telomere maintenance, and apoptosis (42). Interestingly, components of the SUMO machinery are also localized to PML bodies (28). Moreover, sumoylation of the transcriptional repressor, PML, is itself required for the formation of PML bodies (43). PML has a SUMO-interaction motif and contains a RING domain that contributes to its efficient auto-sumoylation (40). Current evidence suggests that PML, through these

unique qualities, acts as a scaffold to seed nuclear body formation (40). Sumoylation of many proteins targets them to PML bodies, likely through gained interactions with proteins such as PML (44).

In addition to PML bodies, SENP2, RanBP2, and Ubc9 are concentrated at the nuclear pore (11,45). As discussed above, sumoylation of RanGAP1 is required for its localization to the nuclear pore (6,34). Sumoylation of numerous other factors has been shown to be required for their nuclear localization (46), including NEMO (47), Reptin (48), and CtBP1 (49). Moreover, fusion of many of these proteins with SUMO is sufficient for their nuclear localization (47,48). Conversely, sumoylation of many proteins such as HDAC4 requires their nuclear localization (50). However, SUMO conjugation can also favor nuclear export (51). Mutation of the SUMO sites in the transcription factor Elk-1 increases its speed of nuclear import (52). Despite controlling the localization of some proteins, the localization of many substrates is not affected by SUMO conjugation (22). Thus, SUMO is not a general regulator of protein localization. This supports the notion that SUMO functions through multiple, context-dependent mechanisms.

Sumoylation of Chromatin-Associated Proteins

Many SUMO substrates have chromatin-related functions (5,9,10,22,53-58). Studies in yeast, *Xenopus*, and humans have linked sumoylation of proteins

on chromosome arms or at kinetochores to proper chromosome segregation (59-63). Importantly, SUMO2/3, but not SUMO1, appear to be the preferred SUMO isoforms enriched on mitotic chromosomes (60). SUMO also plays an important role in genome maintenance, such as DNA repair and replication (see MMS21 and TDG above). Given the breadth of SUMO substrates identified so far, sumoylation likely affects most nuclear processes.

One large class of SUMO substrates are those that regulate transcription (10). Basic transcriptional machinery components, such as RNA polymerase II subunits, are sumoylated *in vivo* (54). The list of substrates extends to transcription factors such as nuclear hormone receptors, Jun, Myb, and AP-2, as well as co-factors such as HDAC1, HDAC4, CtBP, GRIP1, and SRC1 (10). Sumoylation of these transcription factors generally leads to transcriptional repression (10). For example, modification of glucocorticoid receptor (GR) represses its ability to synergistically activate transcription at promoters with multiple GR response elements (64). Alanine-scanning mutagenesis on SUMO2 that was fused to GR revealed a conserved surface on SUMO that is required for its repressor function (64,65). This surface of SUMO also mediates its binding to SUMO-interaction motifs. Consistently, Daxx, a PML body localized protein, contains a SIM that mediates recruitment of sumoylated GR into PML bodies for transcriptional repression (38).

Several other transcriptional repressors associate with SUMO non-covalently. These include MTA2, a component of the NuRD (nucleosome remodeling and deacetylase) complex (66,67), and ZNF198 (35,66,68), which binds to a corepressor complex containing CoREST, LSD1, and HDAC1/2 (69-71). The function of SUMO binding in these contexts is unclear, although multiple components of both complexes can be sumoylated (54,72,73). SUMO does not always promote protein-protein interactions. Agonist-dependent sumoylation of PPARγ disrupts an interaction with ubiquitin-proteasome components that is normally needed to clear PPAR promoters of corepressors, such as N-CoR/HDAC3 (74). Sumoylation of chromatin bound MBD1 disrupts its association with the SETDB1 histone H3 lysine 9 methyltransferase (75). However, in this case, SUMO promotes transcriptional activation, underscoring the complexity of SUMO function in transcription. The interplay between sumoylation and non-covalent SUMO binding in complexes that contain multiple SUMO substrates needs to be further explored.

Part B: Chromatin Background

The Structure of Chromatin

The packaging of genomic DNA into chromatin is mediated by multiple proteins, but the central organizers of this process are histones. The nucleosome is the fundamental unit of chromatin. The nucleosome core particle (NCP; see Figure 2, left side) consists of two histone H2A/H2B dimers and one histone H3/H4 tetramer, wrapped approximately 1.7 times by double-stranded genomic DNA (76-78). Protruding from the core of this particle are multiple, highly conserved histone tails that are subject to many types of post-translational modifications (79). Flanking the NCP are often various lengths of linker DNA. These can serve as landing sites for DNA binding proteins (80,81), although nucleosomal DNA is not exempt from DNA binding proteins either (82). Consistently, many promoters and transcriptional response elements are depleted of histone molecules and enriched for general transcriptional machinery (83). Thus, it follows that chromatin must be an important obstruction to DNA binding proteins that allows for complex regulation of chromatin-templated processes, such as transcription. Indeed, in *E. coli*, the existence of a DNA binding motif nearly guarantees its occupancy, whereas in humans, only 1-3% of sequence motifs are actually occupied by their cognate binders (84).

Chromatin is Dynamically Regulated

There are several fundamental means of regulating DNA accessibility in chromatin. The genome encodes *de novo* nucleosomal positioning sequences (80,81), allowing for biased positioning away from enhancers and promoters. Alternatively, linker histones (e.g., histone H1), bind to the junction of linker DNA and the nucleosome core particle, promoting compaction of chromatin (85). Histone-DNA interactions are very stable *in vitro*, but dynamic in cells (85). First, nucleosome remodeling complexes, such as SWI/SNF, contain ATPases that "melt" DNA-histone associations. These complexes catalyze horizontal sliding of nucleosomes (85). Nucleosomes not only restrict the accessibility of DNA, but also impede the movement of DNA-templated enzymes, such as elongating RNA polymerase II (Pol II) (85,86). Consistently, transcriptional elongation *in vitro* is inefficient without the addition of histone chaperones (86). These proteins can catalyze the stepwise eviction or re-deposition of histones onto chromatin *in vitro* and *in vivo* (85). The re-deposition of histones after the passage of Pol II is especially important, since this inhibits initiation from cryptic start sites (86). Other histone chaperones function in histone exchange, in which classical histone molecules are exchanged for histone variants, such as H2A.Z, H2A.X, H3.3, or CENP-A. These variants serve special functions but maintain the core structure of the nucleosome (85).

Histone Modifications

The highly charged and conserved histone tail domains within nucleosomes are subject to many modifications, including lysine acetylation, lysine and arginine methylation, proline isomerization, ADP-ribosylation, phosphorlyation, ubiquitination, and sumoylation (79,87). These modifications either direct local changes in chromatin structure, or more commonly recruit effector proteins that possess specific histone tail recognition modules (79,87). Histone modifications affect multiple chromatin-templated processes, including transcription, DNA repair, and replication (87). I focus my discussion on histone lysine acetylation and methylation in transcription regulation.

Histone Acetylation—Histone acetylation on many residues is strongly associated with transcriptional activation (79). Acetylation functions partially by recruiting effector proteins, but also by neutralizing positive charges of histone tails that contribute to inter-nucleosomal packing (Figure 2; compare top and bottom) (87). Histone acetyltransferases (HATs) show little specificity and fall into three families: GNAT, p300/CBP, or MYST. These enzymes target multiple lysine residues, and are often parts of large transcriptional co-activator complexes (87).

Histone deacetylases (HDACs), similar to HATs, show little specificity towards histone tails, and are generally associated with transcriptional repression (79). There are three major classes of HDACs based on functional and sequence

homology: Class I (HDAC1, HDAC2, HDAC3, HDAC8), Class II (HDAC4-7, 9-11), and class III (Sirtuin family; SirT1-6) (88). The class III sirtuins possess NAD^+-dependent catalytic activity and are structurally distinct (88). Class I/II HDACs share catalytic folds and are subject to inhibition by the same chemical inhibitors (e.g., Trichostatin A). However, class I and II deacetylases have distinct functions. HDAC4, 5, and 7 are normally cytoplasmic and only enter the nucleus after specific stimuli. In contrast, HDAC1-3 constitutively associate with chromatin, and are often part of large corepressor complexes (88).

HDAC1/2 are 85% identical and are partially redundant, as determined by mouse genetic studies (89). HDAC1/2 function largely as corepressors, although they are required for transcriptional activation in some instances (89). There are no other recognizable domains in HDAC1/2 outside their deacetylase domains. The C-terminal regions of HDAC1/2 that are highly charged and subject to many modifications are required for their association within corepressor complexes, such as Sin3, NuRD (Nucleosome remodeling and deacetylase), and CoREST (REST corepressor) (90). Additionally, these complexes often contain other histone modifying activities (e.g., lysine specific demethylase 1 in CoREST; see below) (91) or nucleosome remodeling activities (e.g., Mi2 in NuRD) (67).

Lysine Methylation—Lysine methylation is also dynamically controlled by methyltransferases and demethylases (87). Unlike histone acetylation, the effects

of lysine methylation depend on the specific residue and degree (mono-, di-, trimethylation) of modification (92). Consistently, lysine methyltransferases and demethylases have distinct site and degree specificity (87,92). Additionally, modification at a given site and of the same degrees can be controlled by multiple methyltransferases and demethylases (87). Lysine methylation is catalyzed by SET domain containing methyltransferases, with the exception of Dot1 that is structurally unrelated (87,92). Lysine demethylation is catalyzed by the jumonji-domain type C (JmjC) family of Fe(II)-dependent di-oxygenases or the amine oxidase (AOD) family, including lysine specific demethylase 1 (LSD1) (87).

Histone H3 lysine 9 methylation (H3K9me2/3) is generally associated with transcriptional repression (Figure 2, top) (87,92). The major H3K9 methyltransferases function at different loci in the cell. For example, SUV39H1 is the major H3K9 tri-methyltransferase in heterochromatin, whereas G9a is responsible for most euchromatic H3K9 methyltransferase activity (93). Androgen receptor-mediated activation initiates decreases in H3K9me1/2/3 levels, likely through the combined action of JmjC demethylases (94,95). In this context, H3K9 demethylation also requires LSD1 activity (94-96). However, H3K9 methylation is not always repressive. For example, H3K9me1 is associated with active promoters (97), and H3K9me3, but not H3K9me1/2, has recently been linked to transcriptional elongation (98).

Histone H3K4 methylation, especially di- and trimethylation, is enriched at actively transcribed promoters (83,97,99,100). In yeast, a large complex called COMPASS that contains the methyltransferase SET1 is responsible for H3K4 methylation (92,101). A similar complex exists in humans that contains the MLL family of methyltransferases (101). This complex is responsible for maintaining cellular levels of H3K4me2/3, but not H3K4me1 (Figure 2) (102). Alternatively, SET7/9 transcriptional activator only mediates H3K4 monomethylation (103). H3K4me1 is enriched along with H3K4me2/3 at active promoters (83,97). However, H3K4me1 is specifically enriched at transcriptional enhancers compared to H3K4me2/3 (83). LSD1, the first lysine demethylase discovered, has specific activity towards H3K4me1/2, but not H3K4me3 (91). This limitation is inherent to the chemistry of flavin adenine dinucleotide (FAD)-mediated amine oxidation, which requires a lone pair of electrons for catalysis (Figure 14A) (91). Several H3K4me2/3 jumonji-containing-demethylases have recently been described (104). Some of these demethylases, such as SMCX, associate with sequence-specific repressors that also recruit HDAC1/2 and LSD1 (104-111).

Chromatin Recognition Domains

In vivo, H3K4 methylation status correlates well with transcriptional activity (83,97,99,100). However, this modification has no effect on transcription in an *in vitro* reconstituted model (112). This is explained by the absence of

effectors that bind to this modification *in vivo*. A subunit of the co-activator nucleosome remodeling complex, NURF, contains a PHD (plant homeodomain) zinc-finger, which recognizes H3K4me2/3, and a bromo domain, which binds to acetyled histone tails (Figure 2, bottom) (101,113). Another module, the chromo domain, is found in proteins, such as HP1 (heterochromatin protein 1) . HP1 binds specifically to H3K9me2/3, and has critical functions in pericentromeric heterochromatin maintenance (Figure 2, top) (114,115).

Several other chromatin recognition domains that do not have specificity towards certain modified histone tails have been described. The SWIRM (Swi3p, Rsc8p, and Moira) (116-118) and SANT (SWI-SNF, ADA, N-CoR, and TFIIIB) (119,120) domains are found in many chromatin-associated complexes. These domains can function in histone tail or DNA-binding. For example, the SANT domain of v-Myb binds to DNA, whereas the SANT domain in N-CoR and SMRT contribute to histone tail binding (121-124).

Importantly, chromatin recognition modules, as well as the substrate binding sites of the enzymes themselves, contribute to significant cross-talk between histone modifications (79,87). For example, MLL1 methyltransferase complex shows significantly more activity on acetylated histone tails, as compared to unmodified tails (125). The demethylase activity of LSD1 is inhibited by many modifications on histone H3, including acetylation (69,126-128).

Because many of the domains described above are often found in combination within effector complexes, and because nucleosomes themselves can contain combinations of specific histone modifications, it is likely that the sum of these modules can result in significant specificity and robustness in binding chromatin (101). This is underscored by the amount of cross-talk between histone modifications.

LSD1-Containing Complexes

LSD1 was first discovered as part of a large corepressor complex containing CoREST and HDAC1/2 (Figure 2) (71,129-131). I will refer to this ternary complex as the LCH (LSD1-CoREST-HDAC1/2) core. CoREST was first described as a direct binding partner of the sequence specific repressor, REST (RE1-silencing transcription factor), thus establishing it as a corepressor (132,133). The LCH core also associates with CtBP1/2, incorporating it into an even larger corepressor complex that contains G9a histone H3K9 methyltransferase (130). CtBP1/2 can also target this complex to specific promoters indirectly through its interaction with sequence specific repressors, such as ZEB1/2 (134). This is in contrast to the direct interaction of CoREST with REST (132). MeCP2, a CpG methyl-binding protein, also recruits CoREST complexes to promoters (135). Consistently, methylated CpG islands tend to be devoid of histone acetylation and H3K4 methylation (136). CoREST contains an

ELM2 (Egl-27 and MTA1 homology 2) domain as well as two SANT domains (Figure 14B). It interacts with HDAC1/2 and REST through its ELM2 domain (126)((133) and binds to LSD1 through a linker region that connects its two SANT domains (69,70).

LSD1 consists of an N-terminal SWIRM domain and a large amine oxidase domain (AOD) that is disrupted by a 92 amino acid insert (Figure 14B) (137). Recombinant purified LSD1 can efficiently demethylate histone tail or bulk histone substrates, but demethylation of and efficient binding to nucleosomal substrates requires CoREST (69,70). Stimulation of LSD1 by CoREST requires at least one SANT domain and the linker region of CoREST. Consistent with the importance of their interaction, CoREST is also required for the stability of LSD1 (69). CoREST also coordinates the activities of LSD1 and HDAC1/2 on nucleosomes. For example, HDAC1/2 can stimulate LSD1 activity, and vice versa, but only in the presence of CoREST and on nucleosomal substrates (126). This is consistent with the overriding theme that there is significant cross-talk between histone modifications.

The functions of LSD1 are diverse. In *Drosophila*, the H3K4 demethylase activity of LSD1 is required for subsequent H3K9 methyltransferase activity on histone tails (138). Thus, LSD1 indirectly affects the spread of heterochromatin (138). LSD1 also functions in transcriptional activation. For example, LSD1 is found on the growth hormone promoter in instances of repression as well as

activation (134). Consistently, a fraction of LSD1 associates with MLL methyltransferase complex (139). LSD1 is also directly involved in activation programmes at numerous androgen receptor (AR) and estrogen receptor (ER) responsive promoters (94-96,140). Because of these findings, LSD1 has been proposed to function as a H3K9 demethylase at these AR and ER promoters (96).

Regulation of the LCH Complex by SUMO

I began my graduate studies by performing a systematic screen for SUMO substrates using *in vitro* expression cloning. We used the many substrates we identified and cloned to systematically characterize the specificity of SUMO isopeptidases and the PIAS family of E3 ligases. We also determined the general subcellular localization patterns for SUMO substrates, and tested whether sumoylation regulates protein localization in general. The functional significance of one of the substrates identified in the screen, Mef2C, was also explored. Most SUMO substrates are nuclear proteins with functions in chromatin-templated processes. Moreover, many substrates are clustered into individual chromatin-associated complexes. LSD1 component of the LCH complex was also identified in our SUMO screen. Structural and biochemical studies were used to better understand the mechanism of CoREST stimulation of LSD1 activity *in vitro*. Finally, because HDAC1 and LSD1 are SUMO substrates, and because these

proteins interact with a SUMO interacting protein, ZNF198, we also explored the interplay between sumoylation and complex formation *in vitro*.

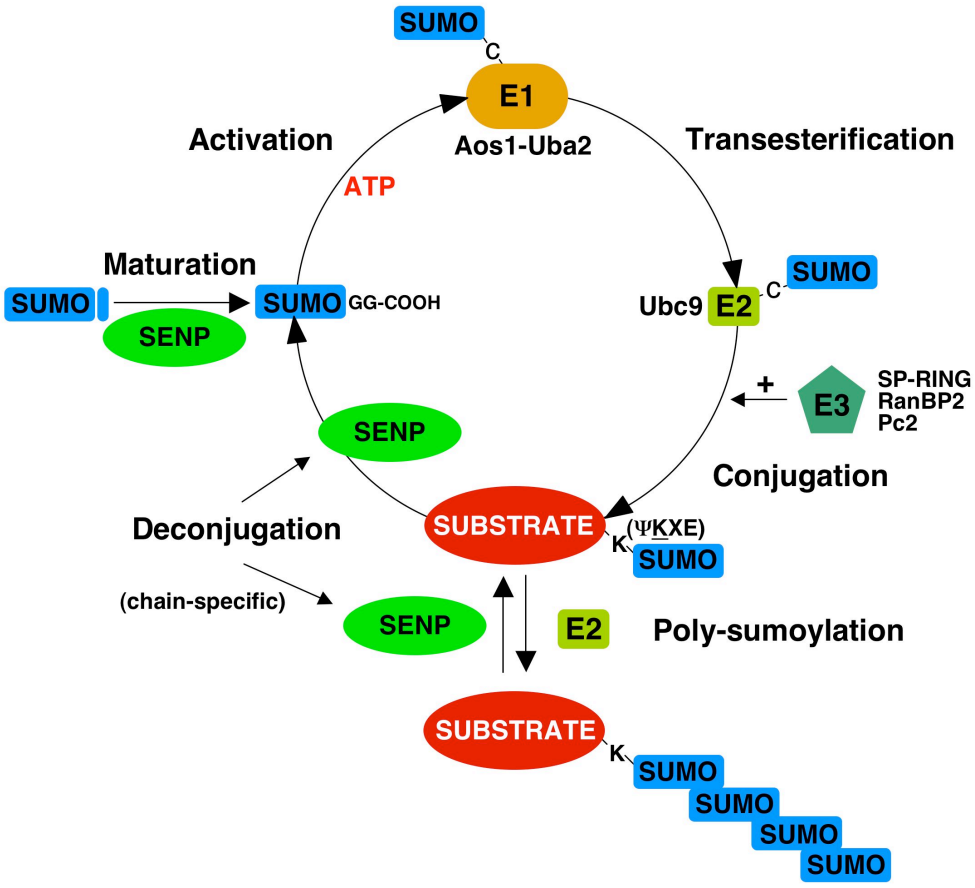

Figure 1. The sumoylation machinery and pathway
The first step of sumoylation, maturation, requires SUMO-proteases/isopeptidases (SENPs) that cleave SUMO to reveal an essential di-glycine motif. SUMO-GG can then be activated in an ATP-dependent process by Aos1-Uba2 heterodimer (E1), resulting in a thioester bond with the carboxy-terminus of SUMO. E1 transfers SUMO to Ubc9 (E2), again as a thioester (transesterification). Finally, E2 conjugates SUMO to a substrate lysine, usually in the context of a consensus ΨKXE (where X is any amino acid and Ψ refers to a large, hydrophobic residue). SUMO E3 ligases can stimulate sumoylation *in vitro* and are likely required *in vivo*. All the SUMO isoforms can form chains *in vitro* (poly-sumoylation), although the importance of this *in vivo* is not clear. Finally, SENPs can deconjugate SUMO from substrates, making sumoylation a reversible modification. Importantly, some SENPs specialize in chain removal.

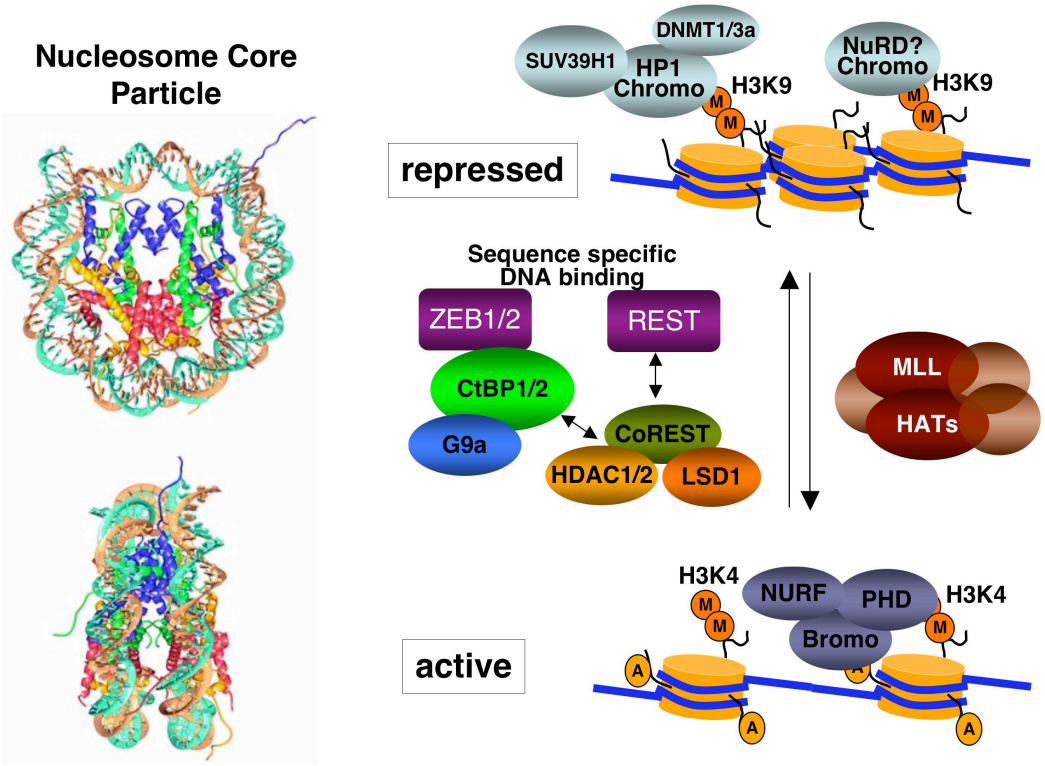

Figure 2: Chromatin Regulation by LSD1-CoREST-HDAC1 complex
Left: Nucleosome core particle consists of a histone octamer wrapped by double-stranded DNA. Obvious from this representation is the overall contribution of DNA to the surface of the core particle, as well as the several histone tails, each of which are subject to a multitude of post-translational modifications. This cartoon is adapted from Tim Richmond's website.
(http://www.biol.ethz.ch/IMB/groups/richmond/projects/nucleosome)
Right: A simplified view of LCH regulation of transcription. CoREST, LSD1, and HDAC1/2 can be recruited to promoters indirectly (via CtBP1/2 or possibly direct chromatin interactions) or directly (via REST). CtBP1 also binds G9a, a H3K9 methyltransferase. The combined function of these complexes is repression, as visualized by the compacted nature of poly-nucleosomes (top). H3K9 methyl-tails (M = methyl) bind to chromo domain containing factors such as HP1 and NuRD, which can repress transcription further. Of course, these processes are reversible by the action of complexes such as MLL, which contain HAT activity as well as H3K4 di/trimethyltransferase activity. Acetylation of tails (A = Acetyl) can inhibit high-order compaction, and the combination of acetylation and H3K4 methylation can recruit activating nucleosome remodeling complexes such as NURF through their Bromo and PHD domains, respectively.

Chapter II. Experimental Methods

Antibodies

The ZNF198 antibody was generated by injecting recombinantly purified ZNF198 residues 923-1377 into rabbits (Zymed). Antibodies were purified from serum with the same fragment. The LSD1 antibody was generated against LSD1ΔN (residues 171-852) using a similar procedure (Yenzym). All other antibodies were commercial. The upstate antibodies with catalogue numbers are α-HDAC1 (05-614), α-CoREST (07-455), α-Histone H4Ac (06-866), α-Histone H3K4Me$_2$, α-Histone H3Ac (06-599), α-REST (07-579). α-Histone H3 antibody (ab1791), α-FLAG M2 (Sigma), and α-Myc (9E10, Roche) were also used.

Plasmids

The coding regions of SUMO1 (1-97), SUMO2 (1-93), PIASy, SENP1, SENP2, SENP3, CoREST, and ZNF198 were amplified from human fetal thymus cDNA library (BD Biosciences) by PCR. Full-length cDNA encoding the SUMO substrates identified in the IVEC screen (see below) were amplified either directly from the original clone (if the clone contained the entire open reading frame) or from human brain or fetal thymus cDNA libraries (BD Biosciences). The PCR products were digested and ligated into pCS2 mammalian expression vectors

containing N-terminal Myc, HA, or GFP tags. Similar methods were used to sub-clone all fragments of ZNF198, CoREST, and LSD1. Ubc9 was also cloned into a pCS2 vector containing a C-terminal Flag tag. The SUMO1 ΔGG, SUMO1 KØ mutant, the dominant-negative Ubc9 mutant, SAP130 K785R, SAP130 K869R, SAP130 K923R, LSD1 K661M, ZNF198 V483A/L484A/V485A, the various MEF2C mutants were constructed with the Quikchange site-directed mutagenesis kit (Qiagen). The pET11c-hAos1, pET28b-hUba2, and pET28b-hUbc9 vectors were gifts from C. Lima and K. Orth. The pGEX-ScUlp1 vector was obtained from M. Hochstrasser. The pGEX-PIASxβ and pGEX-PIAS1 vectors were provided by S. Muller. The Topo IIb vector was a gift from L. Liu. The pSC-B-rat REST/NRSF vector was a gift from Jenny Hsieh. HDAC1-FLAG pCDNA3.1, and pCDNA-MEF2C constructs were obtained from E. Olson. The pCMV5-MKK6-DD plasmids were gifts from M. Cobb.

Protein Expression and Purification

SUMO related proteins—DNA fragments encoding the wild-type or KØ mutant of His_6-SUMO1 and His_6-SUMO2 were subcloned into pET28a. These proteins were expressed in BL21(DE3) and purified using Ni^{2+}-NTA beads per manufacturer's protocols (Qiagen). Ubc9 was expressed and purified similarly. For the expression of Aos1–Uba2, pET11c-hAos1 and pET28b-hUba2 were co-tranformed into BL21(DE3). The resulting Aos1–Uba2 complex was purified by

Ni^{2+}-NTA beads followed by gel filtration chromatography on a Superdex 200 column (Amersham) to remove the excess amount of Aos1. The pGEX-ScUlp1 vector was transformed into BL21. The resulting GST-Ulp1 protein was purified using glutathione-agarose beads (Amersham). All proteins were concentrated to between 1-5 mg/ml in a buffer containing 20 mM Tris (pH 7.7), 100 mM KCl, 1 mM DTT and 10% glycerol, and stored in aliquots at –80˚C.

CoREST complex related proteins—LSD1ΔN (171-852) and the variant containing the K661R mutation were expressed in *E. coli* as glutathione S-transferase (GST) fusion proteins. GST-LSD1ΔN proteins were purified from bacterial lysates by glutathione-sepharose resin. After protease digestion to remove GST, LSD1ΔN proteins were further purified by ion exchange chromatography. CoREST-C (286-482) was expressed with an N-terminal His_6-tag and purified using Ni^{2+} resin followed by ion exchange chromatography. The seleno-methionine derivatives of LSD1ΔN and CoREST-C proteins were purified similarly. LSD1ΔN proteins were then mixed with CoREST-C. The resulting complex was purified by gel filtration chromatography and concentrated to about 10 mg/ml in a buffer containing 25 mM HEPES, pH 7.4, 200 mM NaCl, 1 mM PMSF, and 5 mM DTT. The full-length His_6-LSD1, His_6-CoREST/His_6-LSD1 complex, or His_6-ZNF198 was purified from Sf9 cells using a combination of Ni^{2+}-Sepharose (Amersham) affinity chromatography and ion exchange chromatography (Resource Q, Amersham). HDAC1-FLAG was purified with M2

agarose beads (Sigma) from Sf9 lysates and eluted with the FLAG peptide. GST-CoREST, GST-SUMO, and GST-Mef2C proteins were purified from bacterial lysates using glutathione resin. Proteins were stored in buffer containing 50mM Tris-HCl, pH 8.1, 50 to 200 mM KCl, 10% glycerol, and 1mM DTT.

In Vitro Expression Cloning (IVEC)

Five 96-well plates with 100 cDNAs per well of a human adult brain library (Promega) were used in the IVEC screen for SUMO1 substrates per manufacturer's protocols. Briefly, 1 μl of DNA was *in vitro* transcribed and translated (IVT) in reticulocyte lysate in the presence of ^{35}S-methionine and subjected to *in vitro* sumoylation reactions, which contained 2 μl of IVT product and were performed as described below (see Sumoylation Assays). Positive pools were transformed into DH5a. A total of 96 individual clones per positive pool were picked and cultured overnight in LB/AMP medium on 96-well plates. Aliquots of cultures in each row and each column of these 96-well plates were combined separately. Plasmids were isolated from these cultures and tested in the sumoylation assay as described above. Positive clones were identified and sequenced. The SUMO1 conjugation efficiency of each substrate was quantified using the ImageQuant software (Amersham).

Cell Culture and Transfections

HeLa Tet-on cells (BD Biosciences), 10T1/2, C2C12, and U2OS cells were grown in DMEM (Invitrogen) supplemented with 10% Fetal Bovine Serum, 2 mM L-glutamine, 100 μg/ml penicillin and streptomycin at 37˚C and 5% CO_2. Differentiation of MyoD-transfected 10T1/2 cells was induced by substituting Growth Medium with Differentiation Medium (DMEM supplemented with 2% horse serum, 100 units/ml penicillin, and 0.1 mg/ml streptomycin). Cells were plated and transfected in 12-well plates for promoter-luciferase assays, 6-well plates for RT-PCR or western blotting, 10-cm plates for immunoprecipitations and fractionation experiments, and 4-well chambered slides for indirect immunofluorescence microscopy. Cells were transfected using the Effectene (Qiagen), Oligofectamine (Invitrogen), or Lipofectamine RNAiMax (Invitrogen) reagents according to manufacturer's protocols.

The siRNA sense strands are as follows:

SENP1: 5'-CAGCUGUCCCACAGUGUAUdTdT-3'

SENP2: 5'-GCCCAUGGUAACUUCUGCUdTdT-3'

LSD1-1: 5'-CGGACAAGCUGUUCCUAAAdTdT-3'

LSD1-2: 5'-GGCCUAGACAUUAAACUGAdTdT-3'

ZNF198: 5'-GGGCCAGACAGCUUAUCAAdTdT-3'

ZNF261: 5'-GACCCUGUGUAAGAACUUUdTdT-3'

ZNF262: 5'-CACCACCACUAGUAAAGAUdTdT-3'

Cell Fractionation

HeLa Tet on cells (2 X10 cm plates) were transfected with siRNAs using Lipofectamine RNAi-Max (Invitrogen) for 48 hours, harvested in cold 1X PBS by cell scraping, and fractionated using a protocol adapted from Bruce Stillman's laboratory (141). Cells were resuspended in 1 mL of cold Buffer A (10mM HEPES, pH 7.9, 10mM KCl, 1.5mM $MgCl_2$, 0.34M Sucrose, 10% glycerol, 0.1% Triton X-100, 1mM DTT, 10ug/mL protease inhibitor cocktail, and 0.4mM PMSF) and incubated for 5 minutes on ice. After 5 minutes of centrifugation at 1,300g at 4 degrees, pellets were washed one time in Buffer A and supernatants were further clarified by centrifugation at 14,000g. The nuclear pellet was then extracted for 30 minutes on ice in Buffer C (10mM HEPES, pH 7.9, 10mM KCl, 300mM NaCl, 1.5mM $MgCl_2$, 25% glycerol, 0.1% Triton X-100, 1mM DTT, 10ug/mL protease inhibitor cocktail, and 0.4mM PMSF) or in 2mM EDTA, pH 7.4 (buffered with tetrapropylammonium hydroxide; Sigma) that also contained 1mM DTT and protease inhibitors (but no PMSF). For micrococcal release samples, the nuclear pellet was was treated with 0.5U of micrococcal nuclease (Sigma) in Buffer A plus 1mM $CaCl_2$ at 37 degrees for 2 minutes with intermittent shaking. The nuclease reaction was stopped by the addition of 2mM EGTA and incubated on ice for 5 minutes followed by centrifugation at 1,300g for 5 minutes at 4 degrees. The pellet was then extracted with 2mM EDTA as

above. Equal volumes of samples were normalized, diluted in 2X-SDS sample buffer, boiled, and run on SDS-PAGE for western blotting.

In Situ Extraction and Immunostaining

HeLa Tet-on cells transfected with various plasmids were fixed with 4% paraformaldehyde, permeablized with 0.1% Triton-X100 in PBS, and incubated with 1 μg/ml of anti-Myc (9E10, Roche) or anti-Flag (Sigma). After washing, fluorescent secondary antibodies (Molecular Probes) were added at 1:500 dilutions. The cells were again washed three times with PBS, counter-stained with DAPI, and viewed using a 63X objective on a Zeiss Axiovert 200M microscope. Images were acquired using the Intelligent Imaging software, and pseudo-colored in Adobe Photoshop.

For *in situ* extractions (142), cells were first washed in PBS and then fixed as above (control). For extraction before fixation, cells were washed in CSK buffer (10mM PIPES-KOH, pH 7.0, 100mM NaCl, 300mM Sucrose, 3mM $MgCl_2$) then in CSK buffer + (also contains 0.5 % Triton X-100, 0.5mM PMSF, and 10ug/mL protease inhibitor cocktail) for 5-10 minutes. All extractions were done at room temperature. Samples were then processed normally for immunostaining as above.

Crystallization, Data Collection, and Structure Determination

Crystals were grown at 20°C using the vapor diffusion method in sitting drop mode by mixing 0.8 μl protein with 0.8 μl reservoir solution (0.8 M lithium sulfate, 0.8 M ammonium sulfate, 0.4 M sodium chloride, 0.1 M sodium citrate, pH 5.6, and 10 mM DTT) and equilibrating against 100 μl of reservoir solution. Crystals appeared within 12 hrs and matured in about ten days. The crystals were incubated with reservoir solution supplemented with 23% (v/v) glycerol and 1.0 mM diMeK4H3-21, and then flash-cooled in liquid propane. Crystals exhibit the symmetry of space group I222 with cell dimensions of a = 120 Å, b = 179 Å, c = 235 Å, and contain one complex per asymmetric unit and 82% solvent.

Diffraction data were collected at beamline 19-ID (SBC-CAT) at the Advanced Photon Source (Argonne National Laboratory, Argonne, Illinois, USA) and processed with HKL2000 (143). Both native and selenomethionine-derivatized (SeMet) crystals showed significant anisotropy, with diffraction to a Bragg spacing (d_{min}) of about 2.5 Å along the b and c axes, but only to about 3.1 Å along the a axis, resulting in somewhat lower completeness at the high-resolution limit.

Phases for the SeMet variant were obtained from a single anomalous dispersion (SAD) experiment. Using data to 4.0 Å, fifteen selenium sites were located with a combination of the programs SHELXC, SHELXD, and SHELXE (144). Phases were refined using all data to 2.86 Å with the program MLPHARE

(145), resulting in an overall figure of merit of 0.23. Phases were further improved by density modification with histogram matching in the program DM (146), resulting in a final overall figure of merit of 0.79.

The resulting electron density map was of sufficient quality to automatically construct an initial model using the program ARP/wARP (147). This model was used as a starting model for the refinement of the native complex using the program REFMAC5 (148) from the CCP4 package (149), interspersed with manual rebuilding using the program Coot (150) (Table 2).

NMR Spectroscopy

NMR spectra were acquired with a Varian Inova 600 MHz spectrometer. A 1.0 mM solution of 18 bp duplex DNA oligonucleotide (ATCAATATCCACCTGCAG) was titrated into samples of ^{15}N-labeled the CoREST SANT2^{WT} (residues 373-482), SANT2^{K418E}, SANT2^{N419D}, SANT2^{R426E}, and SANT2$^{R426A/R427A}$. To measure the chemical shift changes (^{15}N and ^{1}H) of CoREST SANT2, a series of ^{15}N/^{1}H heteronuclear single quantum coherence (HSQC) spectra was acquired during the titration. Combined chemical shift changes were calculated using the equation $\Delta d_{combined} = ((\Delta d_{1H})^2 + (0.2 \times \Delta d_{15N})^2)^{1/2}$ and plotted against the molar ratio of DNA/SANT2. The final data were fitted to standard ligand binding curves using SigmaPlot. The dissociation constant (K_d) with standard deviations was calculated using the chemical shift changes from

four SANT2 residues. The histone tail peptides were chemically synthesized and added to samples of 159 μM ^{15}N-labeled CoREST SANT2 at a peptide/SANT2 molar ratio of 1.5. HSQC spectra were acquired before and after the addition of each peptide.

Histone Demethylation and Deacetylation Assays

His_6-LSD1 was incubated with purified CoREST-C^{WT}, CoREST-C^{K418E}, or CoREST-C^{N419D} proteins in a buffer containing 50 mM Tris-HCl (pH 8.0), 50 mM KCl, and 2 mM DTT for 30 min on ice. Stoichiometric amounts of LSD1 and CoREST-C were confirmed by Coomassie blue staining. Nucleosomes were purified essentially as described (151) except that separation of nucleosomes from Histone H1 was performed using a 120 ml sepharose CL-6B column (Amersham) instead of glycerol gradient centrifugation.

Demethylation and deacetylation of bulk histones or mono/di-nucleosomes was performed by incubating varying amounts of LSD1, LSD1–CoREST complexes, or HDAC1-FLAG plus the recombinant proteins indicated in the results to either 10 μg of calf thymus bulk histones (Sigma) or 3 μg of mono/di-nucleosomes prepared from HeLa S3 cells in a buffer containing 50 mM Tris-HCl (pH 8.5), 5% glycerol, 2 mM DTT, and 0.1 mg/ml GST (as a carrier protein) in a total volume of 50 µl for 1 hr at 37°C. Deacetylase assays were performed the same except with 150 mM NaCl in the reaction buffer.

Reactions were stopped with 10 μl of 6X SDS sample buffer, boiled, separated on 16% SDS-PAGE, and blotted with the indicated antibodies (see antibodies)

Promoter Assays

Promoter assays were performed in triplicates with the dual-luciferase reporter assay system (Promega) according to manufacturer's protocols. Luciferase activity was measured with a Turner Designs luminometer and normalized for transfection efficiency using the activity of *Renilla* luciferase. The MEF2-responsive promoter activity assays were performed with a pMEF2×3-Luc construct (provided by E. Olson). The Gal4/LexA-promoter activity assay was performed with a pL8G5-Luc plasmid with or without the transfection of a pLexA-VP16 construct.

Immunoprecipitation and Immunoblotting

SUMO IVEC and Mef2C—Cells were lysed in 400 μl of lysis buffer (50 mM Tris-HCl at pH 7.7, 150 mM KCl, 0.5% NP-40, 5 mM $MgCl_2$, 1 mM DTT, 0.5 μM okadaic acid, 10 mM N-ethylmaleimide, supplemented with protease inhibitors; Sigma) for 30 min on ice. After brief sonication, insoluble materials were pelleted by centrifugation at 15,800×g for 30 min at 4°C. Myc-tagged proteins were immunoprecipitated using 0.4 μg of anti-Myc (9E10) monoclonal antibodies (Roche). After incubation at 4°C for 1 hr, 20 μl of Affi-prep protein A

beads (Bio-Rad) was added to each lysate and incubated for 1 hr. The beads were washed with lysis buffer and eluted by SDS sample buffer. Eluted proteins were resolved by SDS-PAGE and western blotted.

ZNF198 and CoREST complex—For IP and western experiments, cells from a 10-cm dish were washed in cold PBS after transfection for 2 days and then scraped in cold PBS. Cells were lysed in 1mL of Buffer C (see fractionation) supplemented with 0.5 μM okadaic acid and treated similarly as above.

Endogenous IP of ZNF198—Twenty 150cm dishes of HEK293 cells were lysed in 30 mL of 50 mM Tris-HCl at pH 7.7, 500 mM KCl, 0.5% NP-40, 1 mM DTT, 0.5 μM okadaic acid, supplemented with protease inhibitors; Sigma. Lysates were sonicated, cleared by ultracentrifugation and a 0.45 micron filter, and subjected to immunoprecipitation with 100 μl of antibody coupled Affi-prep protein A beads. Following the IP, beads were washed five times with lysis buffer and then eluted three times with 150 μl of 100mM glycine, pH 2.5. Elutions were neutralized with 50 μL of 1M Tris-HCl, pH 7.7, and concentrated to approximately 20 μl. Proteins were subjected to 4-20% gradient SDS-PAGE (Biorad) followed by Colloidal Blue staining (Pierce).

Sumoylation Assays

Either recombinantly purified substrate (ie, HDAC1-FLAG) or plasmids that encode appropriate proteins were *in vitro* transcribed and translated (IVT) in

reticulocyte lysate in the presence of ^{35}S-methionine and subjected to *in vitro* sumoylation reactions, which contained 2 μl of IVT product, 2 μg of AOS1–UBA2, 0.5 μg of UBC9, 1 μg of SUMO1, and 1 μl of Energy Mix (150 mM phosphocreatine, 20 mM ATP, 2 mM EGTA, 20 mM $MgCl_2$, adjust pH to 7.7). Reactions were adjusted to a final volume of 10 μl with the XB buffer (10 mM HEPES, pH 7.7, 1 mM $MgCl_2$, 0.1 mM $CaCl_2$, 100 mM KCl, and 50 mM sucrose). Control reactions contained water and XB buffer. After 1-2 hr at 30°C, reactions were stopped with 10 μl of 2× SDS sample buffer, boiled, and subjected to SDS-PAGE followed by autoradiography. For sumoylation of recombinantly purified proteins, assays were performed similarly, except for the case where HDAC1 activity was compared before and after sumoylation. Here, more SUMO-enzymes were use.

In Vitro FLAG and GST Binding Assays

HDAC1-FLAG (1μg), GST-CoREST (1μg), or GST-SUMO1/2 (10μg) proteins plus other indicated purified proteins were incubated with 5-10 μl of M2 Agarose (Sigma) or glutathione sepharose 4B (Amersham) in 50 μl of binding solution (25 mM Tris, pH 8.0, 150 mM NaCl, 2.5 mM KCl, 0.05% Tween-20, 1mM DTT) for 1 hour, washed and incubated in 50 μl in blocking solution (25 mM Tris, pH 8.0, 150 mM NaCl, 2.5 mM KCl, 0.05% Tween-20, 5% dry milk, 1mM DTT) for 1 hr at RT. Then, indicated recombinant proteins or 5 μl of [^{35}S]*in*

vitro-translated protein were added and incubated for another 1 hr at RT. Beads were washed four times with binding solution, eluted 20 µl 2X SDS sample buffer, and subjected to SDS-PAGE followed by Coomassie staining and autoradiography. For binding reactions containing ZNF198, buffers contained 100 µM $ZnCl_2$ as well.

Reverse Transcription and Quantitative PCR Primer Sets

RNA from U2OS or HeLa cells was extracted from 6-well plates after transfection with siRNA using TriZOL reagent (Invitrogen) followed by RNAeasy RNA purification kit (Qiagen). For microarray experiments, RNA was submitted directly to the microarray core facility without freeze-thawing. Otherwise, RNA was then subjected DNAse digestion and inactivation (Roche) followed by reverse transcription (Invitrogen) using random hexamers as primer templates. 2.5 µl of this cDNA was then used for quantitative PCR using a 2X Sybr Green mix (Biorad). All primers, listed below, were validated using methods described by Bookout A et al.

KRTHB6-F: ggctctgaagaagGATGTGG
KRTHB6-R: ATTGGCCTCCAGGTCTGATT
KRT17-F: ATGCAGGCCTTGGAGATAGA
KRT17-R: agggatgctttCATGCTGAG
CLOCK-F: agcaaccatctcaggctca
CLOCK-R: CCCATGGAGCAACCTAGAAG
IFNGR1-F: tgaacggaagtgagATCCAG
IFNGR1-R: GGCACTGAATCTCGTCACAA
GPCR5A-F: gagacaggggacacgctcta

GPCR5A-R: TGGTTctgcagctgaaaatg
RPIA-F: TAGTCGCTTCATCGTGATCG
RPIA-R: gattcccttgtgccactgat
ROCK2-F: tgacccagcagatgatcaag
ROCK2-R: TTTGtgcctgcatttcattc
PXN-F: GTGTGGAGCCTTCTTTGGTC
PXN-R: tcgaagtagtccttgcgaca
THBS1-F: gccaaagacgggtttcatta
THBS1-R: GGTCCTGAGTCAGCCATGAT
SCN3A-F: atgctgggctttgttatgct
SCN3A-R: TGGCTTGGCTTCAGTTTTCT
hCyclophilin B-F: GGAGATGGCACAGGAGGAA
hCyclophilin B-R: GCCCGTAGTGCTTCAGTTT
E-cadherin-F: GGATGACACAGCGTGAGAGA
E-cadherin-R: acaggatggctgaaggtgac
NCAM2 RT-F: CACGTTCACTGAAGGCGATA
NCAM2 RT-R: gctgccctttgacttcgata
KRT80-F: ACCAGAAGACAGGGGTGTTG
KRT80-R: GCATTGAGAGCCAAGAGGAG

Chromatin Immunoprecipitation

Cells were cross-linked for 8 min at RT in 1% Formaldehyde in complete media. Cross-linking was stopped with the addition of glycine (final 250mM) for 5 minutes, then incubated on ice. Cells were washed two times with ice cold PBS, then lysed in cell lysis buffer (10mM Tris-HCl, pH 7.5, 10mM NaCl, 2mM $MgCl_2$, 0.5% NP40). The nuclear pellet was then resuspended in MNase buffer (10mM Tris-HCl, pH 7.5, 10mM NaCl, 3mM $MgCl_2$, 1mM $CaCl_2$, 4% NP-40) and digested with 1 U of micrococcal nuclease (Sigma) at 37% for 5 minutes. Reaction was stopped with 5mM EGTA on ice, followed by addition of 2mM PMSF, 10μg/ml of protease inhibitors, 200mM NaCl, and 1% SDS. Cells were

sonicated till average fragment sizes were near 1kb, then centrifuged for 20 minutes at 14,000g. Supernatant was diluted 5X with dilution buffer (16.7 mM Tris-HCl, pH 8.1, 167 mM NaCl, 1.2 mM EDTA, 1.1% Triton X-100, and 0.01% SDS), pre-cleared with 100µl of protein-A sepharose, and subjected to IP overnight at 4 degrees with indicated antibodies plus 20 µl of protein A-sepharose. Bound complexes were washed 2 times in wash buffer 1 (2 mM EDTA, 50 mM Tris-HCl, pH 8.0, 0.2% Sarkosyl) followed by 4 times in wash buffer 2 (100 mM Tris-HCl, pH 8.0, 500 mM LiCl, 1% NP-40, 1% Deoxycholate), and then eluted in elution buffer (50 mM $NaHCO_3$, 1% SDS). Protein-DNA complexes were digested for 2 hours with Pronase at 42 degrees, followed by de-crosslinking at 65 degrees over-night. DNA was purified using Qiagen spin columns and eluted into 100 µl of TE. Quantitative PCR was performed with 2.5 µl of this DNA using the following primers:

KRT17 ChIP-F: GGATAGGCTCTCGGTCTCCT
KRT17 ChIP-R: GTCTTTCACCCCACACTGCT
GAPDH ChIP-F: tgtgcccaagacctctttc
GAPDH ChIP-R: tattgagggcagggtgagtc

Chapter III. Systematic Identification and Analysis of Mammalian SUMO Substrates[1]

Introduction

Covalent conjugation of SUMO (sumoylation) is an important post-translational modification that regulates protein functions in eukaryotes (3,5,8-11,153). Three isoforms of SUMO, SUMO1, SUMO2, and SUMO3, exist in mammals (5). SUMO1 consists of 101 amino acids, and shares about 50% sequence identity with SUMO2/3 and 18% sequence identity with ubiquitin (5).

Similar to the ubiquitin system (1,2), conjugation of SUMO to substrate proteins is mediated by a cascade of enzymes, including SUMO isopeptidases (SENPs), SUMO-activating enzyme (a heterodimer of Aos1–Uba2), SUMO-conjugating enzyme (Ubc9) and SUMO ligases (3,5,8-11,153). SUMO precursor proteins are processed by a SUMO protease, exposing di-glycine motifs at their C-termini. In an ATP-dependent reaction, the active site cysteine of Aos1–Uba2 forms a thioester with the C-terminus of SUMO. Aos1–Uba2 transfers SUMO to the Ubc9 SUMO-conjugating enzyme, again as a thioester. Ubc9 then transfers SUMO to the ε-amino group of a lysine residue in the substrate, forming an

[1] This chapter is derived with permission in whole from (22) and partially from (152).

isopeptide bond. Unlike ubiquitination, Ubc9 can catalyze efficient sumoylation of many substrates in the absence of SUMO ligases, largely due to the ability of Ubc9 to directly recognize ΨKXE (Ψ, a hydrophobic residue; X, any residue) sumoylation consensus motifs on substrates (154,155). However, SUMO ligases can increase the rates of sumoylation, especially *in vivo* (3,5,8-11,153).

Several types of SUMO ligases have been identified, including the PIAS family of proteins (14), RanBP2 (17), and Pc2 (18). Interestingly, these SUMO ligases exhibit distinct patterns of subcellular localization (11). Furthermore, SUMO isopeptidases that function both in the maturation of SUMO precursors and in the removal of SUMO from modified substrates also exhibit distinct, defined patterns of subcellular localization. For instance, SENP1 resides in PML nuclear bodies (28). SENP2 localizes to the nucleoplasmic face of the nuclear envelope (29). SENP3 is enriched in the nucleolus (30). The distinct localization patterns of these SUMO ligases and isopeptidases suggest that sumoylation of a given substrate might be regulated by its localization within the cell.

Numerous SUMO substrates have been identified either individually or through proteomic efforts (3,5,8-11,153). The identities of these substrates implicate sumoylation in diverse cellular processes (3,5,8-11,153). Intriguingly, many transcription factors/cofactors and components of the chromatin remodeling complexes have been shown to be sumoylated (5,9,10). Sumoylation of transcription factors and cofactors inhibits transcription in some cases and

activates transcription in others (5,9,10). The fact that sumoylation can affect the activity of transcription factors in seemingly opposite ways highlights the complex effects of sumoylation on protein functions. In contrast to ubiquitination, sumoylation of proteins does not generally target them for degradation. Instead, sumoylation appears to regulate the functions of the target proteins through several distinct, yet not mutually exclusive, mechanisms (3,5,8-11,153). First, sumoylation can affect the subcellular localization and trafficking of target proteins. For example, sumoylation of RanGAP1, a regulator of nucleocytoplasmic transport and the first SUMO-conjugated protein identified, is required for its recruitment to the nuclear pore complex (6,34). Second, by competing with ubiquitin for the same lysine residues as attachment sites, sumoylation can antagonize ubiquitination and stabilize its target proteins, such as IkB (32). Third, attachment of the SUMO moiety to a target protein can create an additional surface for protein–protein interactions and enhance the binding between the SUMO target protein and its binding partner (34,156). Therefore, sumoylation can regulate the function of its substrates in multiple ways.

To gain insights into the general cellular function of sumoylation in mammals, we have used an *In Vitro* Expression Cloning (IVEC) strategy to identify mammalian substrates of the sumoylation pathway (157). Our approach complements the recent proteomic efforts in identifying SUMO target proteins, which is limited by the low steady-state levels of sumoylated forms of proteins,

and is further biased by the relative abundance of substrate proteins in cells (53-58). We have identified 40 human SUMO1 substrates in the IVEC screen. Many of these substrates are involved in transcription, RNA processing, maintenance of genome integrity, and chromatin remodeling. We have confirmed the sumoylation of 24 substrates in living cells. Using this panel of substrates, we have investigated the extent of poly-sumoylation of substrates, the conjugation selectivity of SUMO1 and SUMO2, and the specificities of SUMO isopeptidases and ligases. We have also systematically analyzed the effect of sumoylation on the subcellular localization of target proteins. Our study represents an important first step toward the understanding of the mechanism and function of sumoylation in mammalian cells.

Results

Identification of Human SUMO1 Substrates by IVEC

Since the discovery of SUMO in 1996, many proteins with diverse cellular functions have been shown to be modified by SUMO on a case-by-case basis (3,5,8-11,153). In addition, several recent proteomic studies have utilized the affinity purification of tagged SUMO followed by mass spectrometry to identify new *in vivo* sumoylation substrates in budding yeast and mammalian cells (53-58). However, relatively few mammalian SUMO substrates have been identified in two such efforts, presumably due to the dynamic nature of sumoylation and desumoylation, and the low steady-state levels of SUMO-conjugates (54,55,58). As an alternative, we carried out an *In Vitro* Expression Cloning (IVEC) screen to identify SUMO1 substrates from a human brain cDNA plasmid library (157). We screened about 48,000 independent cDNA clones (five 96-well plates with approximately 100 cDNAs per well). However, due to potential redundancy and poor expression of certain clones in the library, the total number of genes screened is estimated to be around 10,000.

To illustrate the strategy of our screen, the identification of two substrates, STAF65γ and ETV1, is shown in Figure 3. Briefly, the pools of cDNA plasmids were *in vitro* translated in rabbit reticulocyte lysate in the presence of ^{35}S-

methionine. These [^{35}S]proteins were incubated in the absence or presence of purified recombinant Aos1–Uba2, Ubc9, SUMO1, and ATP. The samples were then resolved on SDS-PAGE followed by autoradiography. In the C5 well of plate B and the E5 well of plate E, we observed additional up-shifted bands in the presence of the sumoylation reaction mixture (Figure 3, *A* and *B*). A secondary screen was carried out to identify the SUMO1 substrates within these two wells as STAF65γ and ETV1, respectively (Figure 3, *C* and *D*). The appearance of these up-shifted bands in both cases required the presence of all necessary sumoylation components, such as SUMO1, Aos1–Uba2, and Ubc9 (Figure 5A). Addition of Ulp1 (158), a yeast SUMO isopeptidase, to the reaction mixtures greatly reduced the intensities of these up-shifted bands (Figure 5A). These results confirm that STAF65γ and ETV1 are efficiently sumoylated *in vitro*.

From the screen, we identified 40 human proteins that were sumoylated efficiently *in vitro* (Table 1). Six of the 40 substrates identified in the screen, including hnRNP M, Topoisomerase IIβ, PML, SART1, Similar to MGC25497, and TFII-I, are known SUMO substrates (54,55,58,159-161). A close homolog of SATB1, SATB2, has also been shown to be sumoylated *in vivo* (162). This confirmed the validity of our screen.

Multi- and Poly-Sumoylation of Substrates *In Vitro*

We were amazed by the high efficiency of sumoylation of our substrates *in vitro*. Defined as the percentage of substrates converted to SUMO1-conjugates, the *in vitro* sumoylation efficiency ranged from 12% to 94% (Table 1). In addition, many substrates formed SUMO-conjugates that contained multiple SUMO1 molecules and appeared as ladders on SDS-PAGE (Figure 3 & 4). We tested whether these conjugates contained mono-sumoylation at multiple sites (multi-sumoylation), a SUMO1 chain at a single lysine (poly-sumoylation), or a combination of both. We first counted the number of ΨKXE motifs in our substrates. Many substrates contained multiple such motifs (Table 1). However, there does not appear to be a strict correlation between the number of ΨKXE motifs and the efficiency of sumoylation. Furthermore, we randomly sequenced several cDNAs that were not sumoylated *in vitro* and found that some of them contained ΨKXE motifs (data not shown). Therefore, not surprisingly, the mere presence of ΨKXE motifs is not sufficient to target proteins for sumoylation.

We next constructed a SUMO1 mutant (referred to as SUMO1 KØ) with all 11 lysine residues changed to arginines. SUMO1 KØ is not expected to form SUMO1 chains. Earlier studies have shown that SUMO1 can form chains on a fragment of the SUMO ligase RanBP2 (RanBP2ΔFG) *in vitro* (17). We first examined the auto-sumoylation of the SUMO ligases, PIAS1 (Figure 5B) and RanBP2ΔFG (data not shown). The average molecular mass of SUMO1 KØ

conjugates was much lower than that of SUMO1 WT conjugates, indicating that PIAS1 and RanBP2ΔFG underwent poly-sumoylation. Similarly, Ku80, Mi2, FLASH, Topo IIβ, and hnRNP M also appeared to be poly-sumoylated (Figure 5B and data not shown). In contrast, there was no significant difference between the gel banding patterns of SUMO1 WT and KØ conjugates of ETV1 (Figure 5B). This suggested that ETV1 was only multi-sumoylated, consistent with the fact that ETV1 contained 4 ΨKXE motifs, respectively (Table 1). This also served as an important control for the conjugation efficiency of the SUMO1 KØ mutant. On a cautionary note, ubiquitination can occur at the N-terminus of certain proteins (163). Though unlikely, we cannot rule out the possibility that SUMO1 KØ can still support SUMO chain formation at its N-terminus. Regardless, our data clearly demonstrate that poly-sumoylation can occur on substrates other than SUMO ligases themselves. The remarkably efficient sumoylation of many substrates in the absence of SUMO ligases also indicates that Ubc9 can catalyze efficient SUMO conjugation *in vitro*.

Confirmation of Sumoylation of Substrates *In Vivo*

To determine whether the SUMO1 substrates identified in our *in vitro* screen were sumoylated *in vivo*, we cloned the Myc-tagged full-length cDNAs of 26 substrates into mammalian expression vectors and co-transfected HeLa cells with plasmids encoding GFP-SUMO1. Slower migrating species of the substrates

were observed when the proteins were co-expressed with GFP-SUMO1 (Figure 6A). These slower migrating bands were absent when these substrates were co-expressed with GFP-SUMO1 ΔGG that lacked the C-terminal di-glycine motif and cannot be conjugated to substrates (Figure 6A). Furthermore, overexpression of either the dominant-negative mutant of Ubc9 (DN Ubc9) or the SENP2 SUMO isopeptidase greatly reduced the intensity of these slower migrating bands (Figure 6A). Therefore, these substrates are also sumoylated in living cells. Twenty-four of 26 substrates tested in this *in vivo* assay were shown to be sumoylated (Table 1). In addition, there appeared to be a general positive correlation between the *in vitro* and *in vivo* sumoylation efficiencies. For example, substrates that were efficiently sumoylated *in vitro*, such as TFII-I, ETV1, STAF65γ, and ZNF24, were also efficiently SUMOylated *in vivo* (Figure 6A). This further confirmed the validity of our *in vitro* screen.

We stress that, with a few exceptions, most known SUMO substrates are not sumoylated efficiently in cells. For most substrates, sumoylation cannot be observed in the absence of SUMO overexpression, and less than 5% of a given protein is sumoylated even in the presence of SUMO overexpression. The underlying reason for the low steady-state levels of SUMO-conjugates in cells is unclear at present. However, our data are entirely consistent with published reports for other bona fide SUMO1 substrates. In fact, PML is a well-established SUMO1 substrate and is not efficiently sumoylated under the same conditions

(Figure 6A). Most of the other substrates identified in our IVEC screen are sumoylated more efficiently than PML, and indeed more efficiently than many of the known SUMO substrates in the literature. In addition, several proteomic efforts aimed at identifying SUMO substrates were also performed in the presence of SUMO overexpression or under stress conditions that are known to artificially activate global cellular sumoylation.

Regulation of *In Vivo* Sumoylation by SUMO Isopeptidases and Ligases

Though the *in vivo* sumoylation of our substrates was generally more efficient than those reported for other known SUMO substrates, sumoylation of many of our substrates *in vivo* was nonetheless very inefficient as compared to their *in vitro* sumoylation. In particular, few substrates were visibly sumoylated in the absence of SUMO1 overexpression. We reasoned that the inefficient sumoylation of our substrates in human cells in the absence of SUMO overexpression might be partially due to the actions of SUMO isopeptidases. To test this, we depleted HeLa cells of SENP1 or SENP2 with RNA interference (RNAi). RNAi against SENP1 and SENP2 greatly reduced the protein levels of ectopically expressed HA-SENP1 and HA-SENP2, confirming the efficiency of RNAi (Figure 6B). Interestingly, knock-down of either SENP1 or SENP2 caused a significant increase in the sumoylation of TFII-I and a marginal increase in the sumoylation of SATB1 in the absence of SUMO1 overexpression (Figure 6B).

Therefore, sumoylation is negatively regulated by SUMO isopeptidases *in vivo*. However, RNAi against SENP1 or SENP2 did not increase the sumoylation of 13 other substrates tested (data not shown). This was not surprising given that there are multiple SENPs in mammals.

We next tested the substrate specificity of these SUMO isopeptidases. Despite having lower levels of expression, HA-SENP2 was more efficient in reducing sumoylation of ZNF24 in HeLa cells and in removing conjugation of GFP-SUMO1 to other cellular proteins (Figure 7A). Similar results were observed for MEF2C and ETV1 (data not shown). In fact, overexpression of HA-SENP2 efficiently reduced sumoylation of every single substrate tested *in vivo* (Figure 6A and data not shown). This indicates that, as compared to SENP1 and SENP3, SENP2 is a more efficient SUMO isopeptidase and has little substrate specificity. At present, we do not know whether SENP2 is intrinsically a more efficient enzyme as compared to SENP1 and SENP3 or whether the distinct subcellular localization of these enzymes also contributes to the apparent difference in their efficiency to remove SUMO conjugates *in vivo*. Nevertheless, our findings indicate that overexpression of SENP2 is a reliable way to reduce the global levels of sumoylation in mammalian cells (Figure 7B).

Because we did not include SUMO ligases in our IVEC screen and because the *in vivo* sumoylation of some of these substrates is relatively inefficient, we tested whether sumoylation of our SUMO1 substrates can be

stimulated by PIAS1, PIASxβ, and PIASy *in vivo*. We co-expressed PIASxβ with 22 of our substrates in HeLa cells. Sumoylation of 11 substrates were clearly stimulated by PIASxβ (Figure 8A and Table 1). PIASxβ did not stimulate the sumoylation of all substrates (Table 1). In fact, sumoylation of PML was reduced in the presence of ectopically expressed PIASxβ (Figure 8A). We do not know why overexpression of PIASxβ inhibited sumoylation of certain substrates. It is possible that auto-sumoylation of PIASxβ or sumoylation of its cellular targets might consume/preoccupy components of sumoylation pathway, such as Ubc9 and SUMO. We also tested several substrates with PIAS1 (data not shown). PIAS1 appeared to have similar substrate specificity to PIASxβ. We next tested whether PIASy stimulated sumoylation of our substrates. Among 6 of the 11 PIASxβ substrates, only sumoylation of LOC339287 was enhanced by PIASy, despite the fact that PIASy was expressed to much higher levels than PIASxβ (Figure 8B). This indicates that PIASxβ can stimulate sumoylation of many, but not all, substrates *in vivo* and that PIAS1 and PIASxβ might have broader substrate specificities than PIASy.

Conjugation Selectivity of SUMO1 and SUMO2

SUMO2 and SUMO3 are 96% identical, whereas SUMO1 is about 50% identical to SUMO2/3. While all three SUMO isoforms are conjugated to

substrates by the same enzymes, it has been suggested that SUMO1 and SUMO2/3 might display different substrate specificity (5). Because we performed our IVEC screen with SUMO1, we used our panel of SUMO1 substrates to compare the substrate specificity of SUMO1 and SUMO2 *in vitro* and *in vivo* (Figure 9). All substrates tested were modified equally efficiently by both SUMO1 and SUMO2 *in vitro* (Figure 9A). For ZNF24, modification of SUMO2 even appeared to be more efficient (Figure 9A). Thus, there does not seem to be an inherent difference in substrate specificity between SUMO1 and SUMO2 in the absence of SUMO ligases. We next compared conjugation of SUMO1 and SUMO2 to two substrates, ZNF24 and Ku80, in HeLa cells. Surprisingly, SUMO1 modification was much stronger for both substrates (Figure 9B). The protein levels of GFP-SUMO1 and GFP-SUMO2 were similar (Figure 9B). Global modification of SUMO2 to other cellular proteins was also weaker than SUMO1 conjugation (Figure 9B). These data suggest that conjugation of SUMO2 to target proteins might be more tightly regulated *in vivo*.

Regulation of Subcellular Localization by Sumoylation

We next determined the subcellular localization of our SUMO1 substrates. Consistent with their functions, most of our substrates are enriched in the nucleus (Figure 10A, Table 1). Overexpression of GFP-SUMO1 alone did not significantly alter the localization of any of our substrates (Figure 10A and data

not shown). A significant fraction of ETV1, STAF65γ, or SAP130 was sumoylated when co-transfected with GFP-SUMO1. However, there was no noticeable difference in their localization patterns in the presence or absence of GFP-SUMO1 overexpression (Figure 10A). Therefore, sumoylation does not generally lead to changes in subcellular localization of target proteins.

We noticed that co-expression of Ubc9 and GFP-SUMO1 greatly enhanced the global level of sumoylation (Figure 7) and caused accumulation of GFP-SUMO1 in nuclear foci (Figure 10B). The formation of these nuclear foci was dependent on SUMO-conjugation, as these foci were not observed in cells expressing GFP-SUMO1 ΔGG or the dominant-negative mutant of Ubc9 (Figure 10B). The nature of these nuclear foci was currently unknown. However, they did not perfectly co-localize with PML nuclear bodies (data not shown). Therefore, sumoylation of unknown cellular target proteins is likely responsible for the formation of these nuclear foci. We then tested whether any of our substrates were recruited to these nuclear structures. Interestingly, when both Ubc9 and GFP-SUMO1 were overexpressed, SAP130 (Figure 10D), and to lesser extents, TFII-I and Topo IIβ (data not shown), were enriched in these nuclear foci. To determine whether the recruitment of SAP130 to these foci was also dependent on its own sumoylation, we mutated the lysine residues of three ΨKXE motifs of SAP130 to arginines. The K785R mutation abolished sumoylation of SAP130 *in vivo*, indicating that K785 was the major acceptor site for SUMO1

(Figure 10C). SAP130 K785R was still recruited to the nuclear foci when co-expressed with Ubc9 and GFP-SUMO1 (Figure 10D), indicating that sumoylation of SAP130 itself was not required for its recruitment to these nuclear foci.

MEF2C is Sumoylated at K391 *In Vivo*

To further validate our screening method as a means to identify bona fide, relevant SUMO substrates, we chose to further characterize the functional significance of Mef2C sumoylation. We and Gregoire *et al*. had previously shown that the MEF2 family of transcription factors was efficiently sumoylated (22,164). As shown in Figure 11A, MEF2C was modified in the presence of E1 (AOS1/UBA2), E2 (UBC9), and SUMO1. His_6-SUMO1 resulted in a further shift in gel mobility as compared to untagged SUMO1 (Figure 11A). The yeast SUMO isopeptidase, Ulp1, efficiently reversed this modification (Figure 11A). MEF2C contains a sumoylation motif, ΨKXE (Ψ, hydrophobic residues; X, any residues) (154,155) in the γ-domain of MEF2C (165) that is highly conserved among other members of the MEF2 family (Figure 11B). We mutated the lysine residue in this motif (K391) to arginine and observed the MEF2C-K391R is no longer sumoylated *in vitro* (Figure 11C). We then examined the sumoylation of MEF2C in HeLa and NIH3T3 cells (Figure 11D and data not shown). A significant fraction of Myc-MEF2C, but not Myc-MEF2C-K391R, was converted to a slow-migrating species when it was co-expressed with GFP-SUMO1 (Figure

11D, top panel). We then immunoprecipitated Myc-MEF2C and blotted the immunoprecipitates with anti-GFP. The slower migrating band contained GFP-SUMO1 (Figure 11D, middle panel). Furthermore, we observed that overexpression of SUMO isopeptidase, SENP2, or a dominant-negative mutant of UBC9 (DN-UBC9) greatly reduced the intensity of this slow-migrating band of Myc-MEF2C (Figure 11E). Co-expression of the PIAS family of E3 ligases, PIASxβ, enhanced the sumoylation of Myc-MEF2C-WT (Figure 11F), but not that of Myc-MEF2C-K391R. These data indicate that MEF2C is sumoylated *in vivo* and K391 is the major sumoylation site of MEF2C in living cells.

Sumoylation of MEF2C Reduces its Transcriptional Activity

We next used a luciferase reporter assay to examine whether sumoylation of MEF2C regulates its transcriptional activity. The luciferase reporter construct contained three tandem copies of MEF2-binding sites at the promoter region. The transcriptional activity of MEF2C-K391R (the sumoylation-deficient mutant) was about two-fold higher than that of the wild-type MEF2C (Figure 12A), suggesting that sumoylation of MEF2C inhibits its transcription activity. Overexpression of GFP-SUMO1 downregulates the transcription activity of both MEF2C-WT and K391R (Figure 12A). Because overexpression of GFP-SUMO1 caused a global increase in the sumoylation of many cellular proteins (data not shown), inhibition of MEF2C-K391R by GFP-SUMO1 overexpression was most likely due to the

enhanced sumoylation of other MEF2C regulatory proteins under these conditions. However, we cannot completely rule out the possibility that MEF2C is sumoylated at a second site, the sumoylation of which is below the detection limit of our assay.

We next examined whether sumoylation affected the activation of endogenous MEF2C by the p38MAPK pathway. To do so, we transfected a constitutively active mutant of MKK6 (an upstream activator of p38MAPK) into C2C12 cell lines, together with GFP-SUMO1 or SENP2. Overexpression of GFP-SUMO1 reduced the transcriptional activity of MEF2C whereas overexpression of SENP2 enhanced the activity of MEF2C in the presence of constitutively active MKK6 (Figure 12B). These results are consistent with the notion that sumoylation of MEF2C inhibits the transcription activity of the endogenous MEF2C stimulated by the p38MAPK pathway. However, it is entirely possible that alteration of the sumoylation levels of other MEF2C regulatory proteins is responsible for the observed effects of GFP-SUMO and SENP2.

Embryonic fibroblast cells can be converted into myoblasts upon overexpression of MyoD and MEF2 (166,167). We tested whether MEF2C-K391R was more active than MEF2C-WT in collaborating with MyoD to promote the conversion of 10T1/2 cells into myoblasts. We transfected MyoD- and MEF2C-expressing plasmids into 10T1/2 cells and cultured these cells in

low-serum media to induce the differentiation of the converted myoblasts into myotubes. On the fifth day after the induction of differentiation, the cells were fixed and stained with an antibody against myosin heavy chain (MHC), a well-established myogenic differentiation marker (Figure 12, C and D). Co-expression of MEF2C-WT together with MyoD slightly increased the number of myotubes, as compared to the expression of MyoD alone (Figure 12D). Co-expression of MEF2C-K391R with MyoD increased the myoblast conversion rate by two-fold ($P<0.05$) (Figure 12D). The expression levels of MEF2C-WT and MEF2C-K391R were similar (Figure 12D, left panel). These results suggest that sumoylation of MEF2C down-regulates its transcriptional activity during muscle differentiation.

To further study how sumoylation reduced the transcriptional activity of MEF2C, we measured the transcriptional activity of Gal4 fusion proteins of MEF2C and MEF2C-K391R, using luciferase reporter assays with a reporter construct that contained Gal4-binding sites. As compared to Gal4-MEF2C-WT, Gal4-MEF2C-K391R was much more active in stimulating transcription (Figure 13A). Overexpression of DN-UBC9 or SENP2 greatly increased the transcriptional activity of Gal4-MEF2C-WT, but not Gal4-MEF2C-K391R (Figure 13B). These data indicate that sumoylation also inhibits the transcriptional activity of MEF2C at a promoter that does not contain MEF2C-

binding sites, consistent with the fact that sumoylation does not affect the DNA-binding activity of MEF2C (data not shown).

We next performed luciferase reporter assays with a reporter construct that contained both Gal4- and LexA-binding sites. LexA-VP16 (a fusion protein of the LexA DNA-binding domain and the VP16 transactivation domain) dramatically stimulated the transcription of this reporter gene (Figure 13C). Co-expression of Gal4-MEF2C-WT, but not Gal4-MEF2C-K391R, greatly reduced the transcriptional activity of LexA-VP16 (Figure 13C). This suggests that Gal4-MEF2C might recruit transcriptional repressors to this artificial promoter in a manner that is dependent on its sumoylation.

Discussion

Consistent with several recent studies (53-58), the majority of SUMO1 substrates identified in our screen are involved in transcription, RNA processing, DNA repair, and chromatin remodeling. Additionally, we have shown that sumoylation of MEF2C can significantly modulate its activity in cells, further validating our screen. Other novel, well studied representatives from the above categories are: Symplekin, a factor for polyadenylation of pre-mRNA (168); Ku80, a DNA damage repair factor (169); and Mi2, an ATPase of the NuRD chromatin remodeling complex (67). Further characterization of the role sumoylation plays in regulating these proteins awaits.

Intriguingly, Wohlschlegel et al. showed that, in budding yeast, there is a significant clustering of SUMO substrates in multi-subunit macromolecular complexes (56). Though the number of human substrates identified in our screen is too small for rigorous statistical analysis, we also noticed a tendency of our substrates to be subunits of large protein complexes. For example, TFII-I and BHC110 (*B*RAF-*H*DAC *c*omplex p*110; AOF2; LSD1*) both associate with a novel HDAC1/2-containing complex, called the BRAF-HDAC complex (71). STAF65γ and GCN5 are both subunits of a histone acetyltransferase (HAT) complex, called STAGA (170). Therefore, sumoylation might generally regulate

the activity, stability, and/or biogenesis of large macromolecular complexes with functions in the nucleus.

It remains an open question how sumoylation affects the functions of its target proteins. Unlike ubiquitination that generally targets proteins for degradation (1), sumoylation does not appear to have one defined, general role. Instead, sumoylation has been shown to increase protein stability through antagonizing ubiquitination, to change the localization and/or the kinetics of trafficking of substrates within the cell, and to mediate protein–protein interactions (3,5,8-11,153). Consistent with these findings, we have not yet identified a prevailing mechanism by which sumoylation regulates the functions of the SUMO substrates identified in our screen. For example, sumoylation does not generally affect the steady-state localizations of target proteins. However, it remains possible that sumoylation affects the kinetics of trafficking of these proteins within the cell. Consistent with this notion, sumoylation regulates the kinetics of nucleocytoplasmic shuttling of Elk1 (52). In addition, we have shown that up-regulation of sumoylation may lead to the formation of certain nuclear structures, which then recruits other proteins into these structures. The recruitment of SAP130 into these unknown nuclear foci is very reminiscent of the recruitment of p53, Daxx, and other proteins to the PML nuclear bodies (171-173). Similar to SAP130, mutations of the sumoylation sites within p53 or Daxx did not affect their recruitment to the PML bodies (171,172), though sumoylation

of PML is essential for their recruitment (173). Future studies are needed to address the nature of the nuclear foci formed by the overexpression of both Ubc9 and GFP-SUMO1. Our results presented herein are consistent with the notion that sumoylation might regulate the functions of its substrates with multiple, context-dependent mechanisms.

In conclusion, we have identified 35 novel human SUMO1 target proteins. The identities of these SUMO1 substrates point to important functions of sumoylation in regulating transcription and chromatin structure. Analysis of this panel of SUMO1 substrates has also yielded valuable information about several properties of sumoylation, including the extent of poly-sumoylation *in vitro*, the specificities of SUMO isopeptidases and ligases, the conjugation selectivity of SUMO1 and SUMO2, and the effect of sumoylation on the subcellular localization of its substrates.

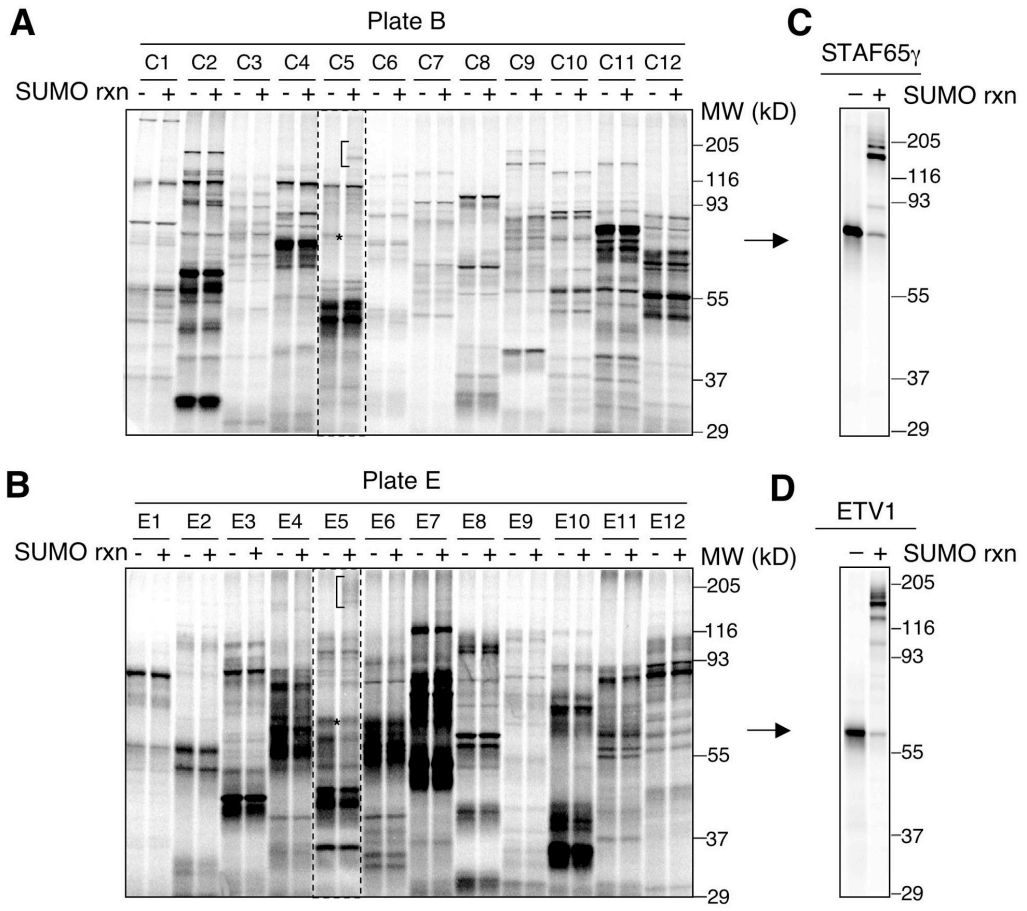

Figure 3. Identification of STAF65γ and ETV1 as SUMO1 Substrates by *In Vitro* Expression Cloning (IVEC).

(A & B) Twelve pools of cDNAs from Row C of Plate B (*A*) or Row E of Plate E (*B*) from a human brain cDNA library were *in vitro* transcribed and translated in the presence of ^{35}S-methionine and subjected to a control reaction (-) or SUMO reaction (+) that contained Aos1–Uba2, Ubc9, SUMO1, and ATP. The reaction mixtures were analyzed by SDS-PAGE followed by autoradiography. Putative substrates in pools B-C5 and E-E5 are boxed with dashed lines. The positions of the un-sumoylated substrates are indicated by asterisks while the putative SUMO-conjugates are marked by brackets.

(C & D) Secondary screen that identifies STAF65γ (*C*) and ETV1 (*D*) as the SUMO1 substrates in pools B-C5 and E-E5, respectively. Reactions were performed as in (*A* & *B*).

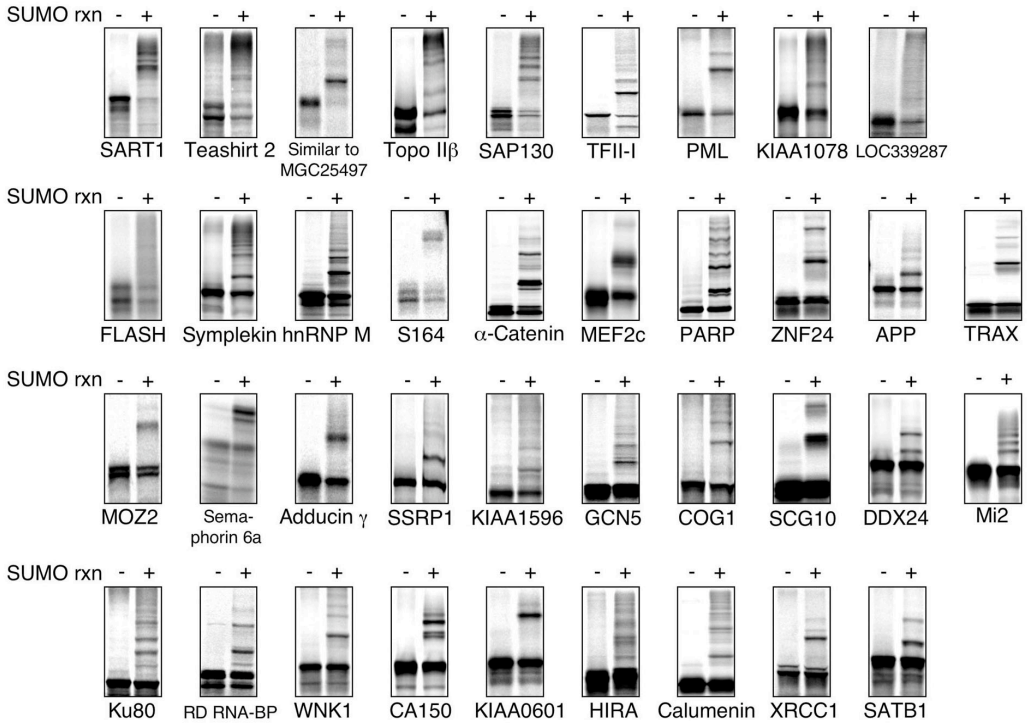

Figure 4. ***In Vitro*** **Sumoylation of the SUMO substrates Identified by IVEC**
Excluding ETV1 and STAF65γ, the other 38 substrates identified in our IVEC screen were translated *in vitro* in the presence of ^{35}S-methionine and incubated with buffer alone (-) or the SUMO reaction mixture (+) containing Aos1–Uba2, Ubc9, SUMO1, and ATP. The reaction mixtures were analyzed by SDS-PAGE followed by autoradiography. The substrates are shown in random order. Quantification of the *in vitro* sumoylation efficiency (defined as the percentage of substrates converted to SUMO1 conjugates) is included in Table 1.

Table 1. Identification and Characterization of Human SUMO-1 Substrates

Gene Name	Accession	Domain(s)	In Vivo Sumoylation	Localization	Stimulated by PIASxβ	[FILV]KxE	Efficiency In Vitro (%)
Transcription Factors/Co-Factors							
ETV1	NP_004947	ETS	+	Nuclear	-	4	94 [a]
Teashirt2 [b]	BAC03610		+	Nuclear	NT [c]	6	83
TFII-I [d]	AAC08315	GTF2I	+	Nuclear	+	4	72
PML [d]	AAB20463	RING/EXOIII	+	NT	-	2	61
BRD8	AAB87858	BROMO	+	Nuclear	-	5	60
FLASH [b,e]	AAD45157		NT	Diffuse [f]	NT	6	55
MEF2C [g]	NP_002388	MADS-Mef2-like	+	Nuclear	+	1	46
ZNF24 [b]	AAB37275	LER(SCAN)/C2H2 ZiF	+	Nuclear	+	1	35
APP [e]	QRHUA4	Amyloid A4/KU	NT	NT	NT	1	34
SSRP1	AAH05116	HMG-box/POB3	+	Nuclear	-	1	29
CA150	AAB80727	WW/FF	+	Nuclear	-	2	21
HIRA	CAA61979	WD40	+	Diffuse	-	1	16
SATB1	AAH01744	CUT/HOX	+	Nuclear	+	1	12
RNA Processing							
SART1 [d]	NP_005137	SART1/Leucine zipper	NT	NT	NT	3	86
Symplekin	NP_004810		+	Nuclear	+	3	54
hnRNP M [d]	NP_005959	RRM	+	Nuclear	NT	4	51
S164 [b]	AAC97961	PWI	+	Nuclear	-	1	49
KIAA1596 [b]	XP_048128	DEAD-box helicase	NT	NT	NT	1	28
DDX24 [b]	AAG02169	DEAD-box helicase	NT	NT	NT	2	23
RD RNA-BP	AAH25235	RRM	+	Nuclear	-	1	22
Genome Integrity							
Topo IIβ [d]	NP_001059	TOPO 2c/4c	NT	Nuclear	NT	2	77
PARP	AAA60137	WGR/PARP	NT	NT	NT	5	36
TRAX	NP_005990	Translin	+	Nuclear	+	1	33
Ku80	A32626	Ku	+	Nuclear	+	2	22
XRCC1	A36353	BRCT	+	Nuclear	-	1	13
Chromatin Modification/Remodeling							
STAF65γ	XP_376044	BTP	+	Nuclear	+	3	80
SAP130	NP_078821		+	Nuclear	NT	3	73
MOZ2 [b]	AAL56647	PHD-ZF/MOZ(SAS)	NT	NT	NT	2	32
GCN5	AAH32743	PCAF/BROMO	NT	NT	NT	0	27
Mi2	NP_001264	Helicase/CHROMO	NT	Nuclear	NT	7	23
BHC110	BAA25527	Amino Oxidase/SWIRM	NT	Nuclear	NT	1	20
Miscellaneous/Unknown							
Similar to MGC25497 [d]	XM_209091 [h]	BTB	+	Nuclear	-	2	82
KIAA1078	NP_982284	Calponin Homology	NT	NT	NT	1	59
LOC339287	XM_290800 [h]		+	Nuclear	+	3	58
α-Catenin	NP_001894	Vinculin	+	Cytosol	-	4	47
Semaphorin 6A	NP_065847	PSI	-	NT	NT	0	31
Adducin γ	AAH62559	Aldolase II	+	Diffuse	+	2	29
COG1	NP_061184		NT	NT	NT	1	26
SCG10	AAB36428	Stathmin	+	Cytosol	+	1	23
WNK1	CAC15059	S/T Kinase	NT	NT	NT	3	22
Calumenin	AAC17216	EF-hand	-	NT	NT	1	13

[a] Percentage of substrates converted to SUMO1-conjugates.

[b] Putative function based on sequence homology only.

[c] Not tested.

[d] Known SUMO substrates.

[e] APP functions in both transcription and Alzheimer's disease; FLASH functions in both transcription and apoptosis.

[f] FLASH C-terminus only.

[g] Mouse MEF2C was used for these studies.

[h] Entry removed from the NCBI database.

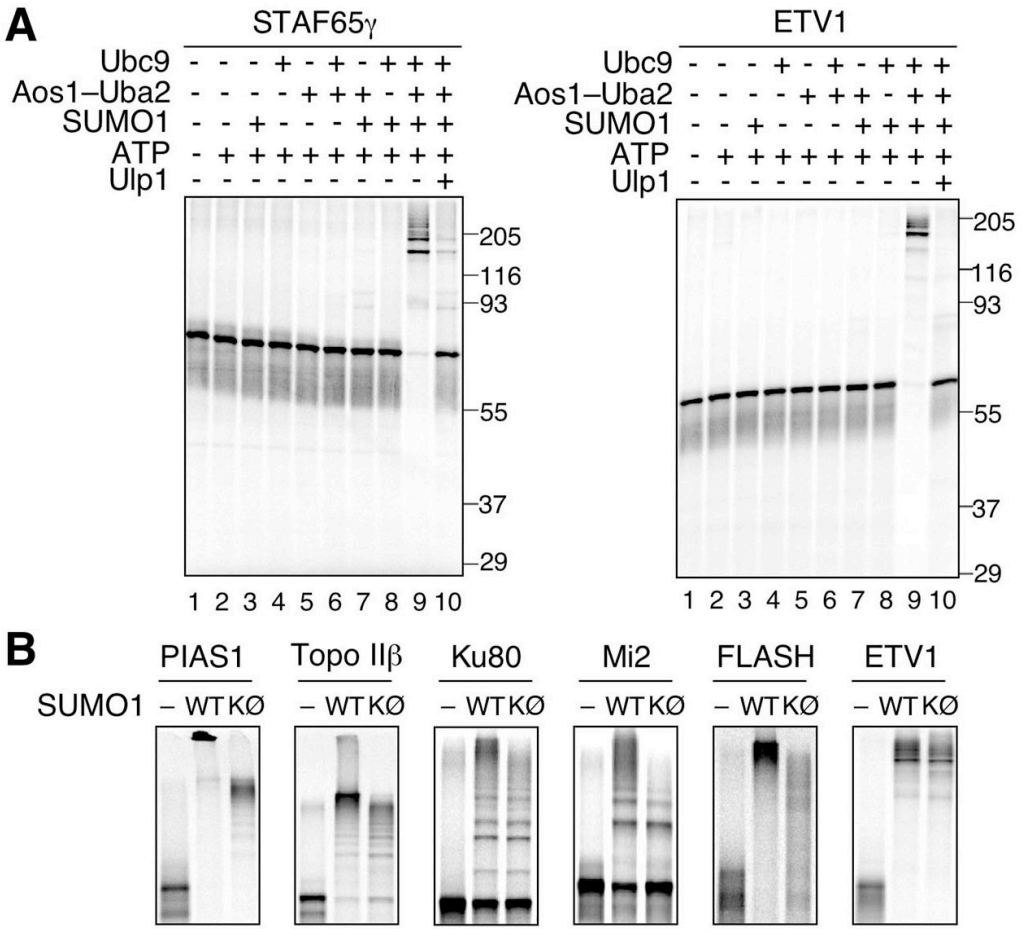

Figure 5. Efficient Multi- and Poly-sumoylation of SUMO1 Substrates *In Vitro*.

(A) Confirmation of STAF65γ and ETV1 as SUMO1 substrates *in vitro*. Plasmids encoding STAF65γ and ETV1 were transcribed and translated in rabbit reticulocyte lysate in the presence of ^{35}S-methionine. The [^{35}S]proteins were incubated with SUMO reaction mixtures containing the indicated components and analyzed by SDS-PAGE followed by autoradiography.

(B) Poly-sumoylation of substrates *in vitro*. The indicated *in vitro* translated [^{35S}]substrates were subjected to either a control reaction (-) or a SUMO reaction containing either His_6-SUMO1 WT (WT) or the His_6-SUMO1 KØ mutant (KØ). The reaction mixtures were analyzed by SDS-PAGE followed by autoradiography. PIAS1 was used as a positive control for poly-sumoylation.

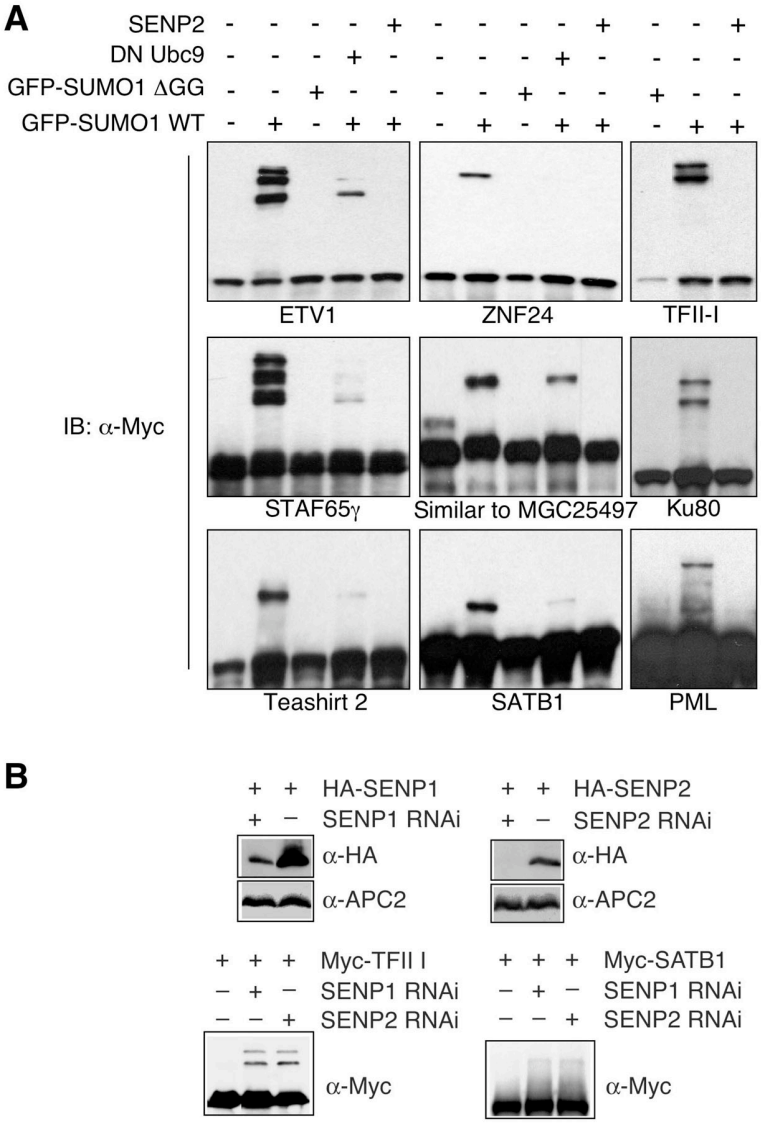

Figure 6. *In Vivo* Sumoylation of SUMO1 Substrates Identified by IVEC.
(A) Myc-tagged SUMO1 substrates identified by IVEC were co-transfected with the indicated plasmids into HeLa cells. Twenty-four to 48 hours after transfection, the total cell lysates were blotted with anti-Myc.
(B) RNAi against either SENP1 or SENP2 enhances sumoylation of TFII-I and SATB1. HeLa cells were transfected with HA-SENP1 or HA-SENP2 plasmids together with siRNAs against SENP1 or SENP2. The total cell lysates were blotted with anti-HA (left two panels). The Myc-TFII-I and Myc-SATB1 plasmids were co-transfected with siRNA against SENP1 or SENP2 into HeLa cells. The total cell lysates were blotted with anti-Myc.

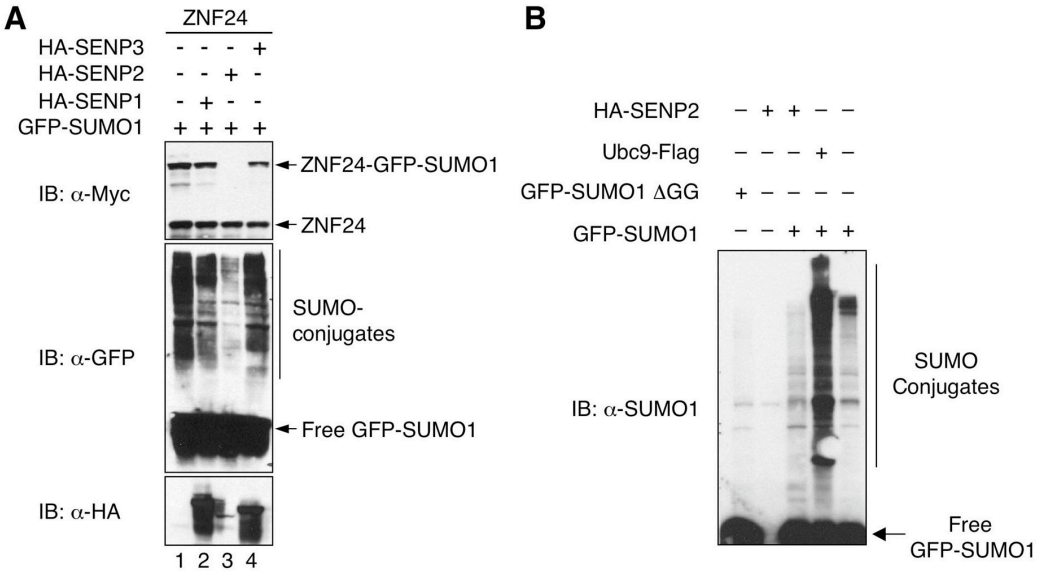

Figure 7. SENP2 is more efficient at de-sumoylation
(A) SENP2 is more efficient in reducing sumoylation in HeLa cells. Myc-tagged ZNF24 was co-expressed with the indicated proteins in HeLa cells for 24 hours. The total cell lysates were blotted with the indicated antibodies. The positions of free ZNF24, the ZNF24-SUMO1 conjugate, free GFP-SUMO1, and GFP-SUMO1 conjugates are labeled.
(B) Overexpression of Ubc9 Enhances Global Level of Sumoylation in HeLa Cells. Lysates of HeLa cells transfected with the indicated plasmids were blotted with anti-SUMO1 antibody (Zymed). The positions of free GFP-SUMO1 and SUMO conjugates are labeled.

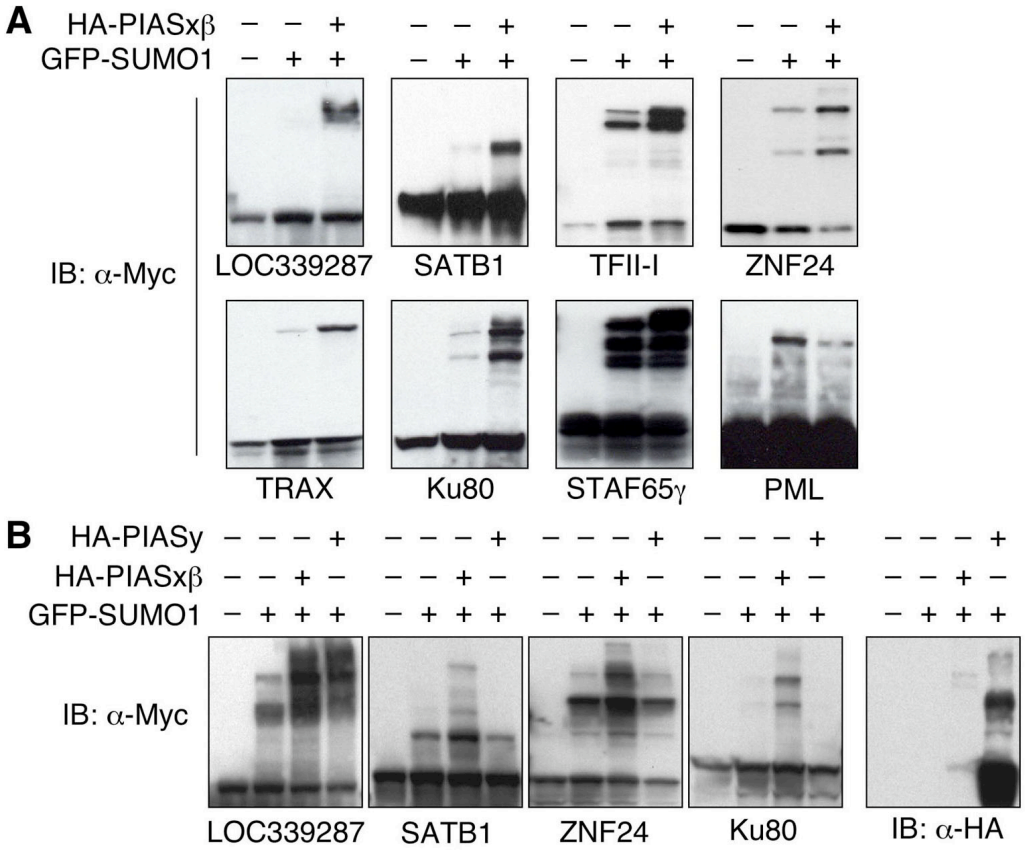

Figure 8. Stimulation of Sumoylation by PIASxβ and PIASy *In Vivo*.
(A & B) Myc-tagged SUMO1 substrates were co-expressed in HeLa cells with the indicated proteins. Twenty-four hours after transfection, the total cell lysates were blotted with anti-Myc or anti-HA. PIASxβ failed to stimulate the sumoylation of PML, which served as a negative control.

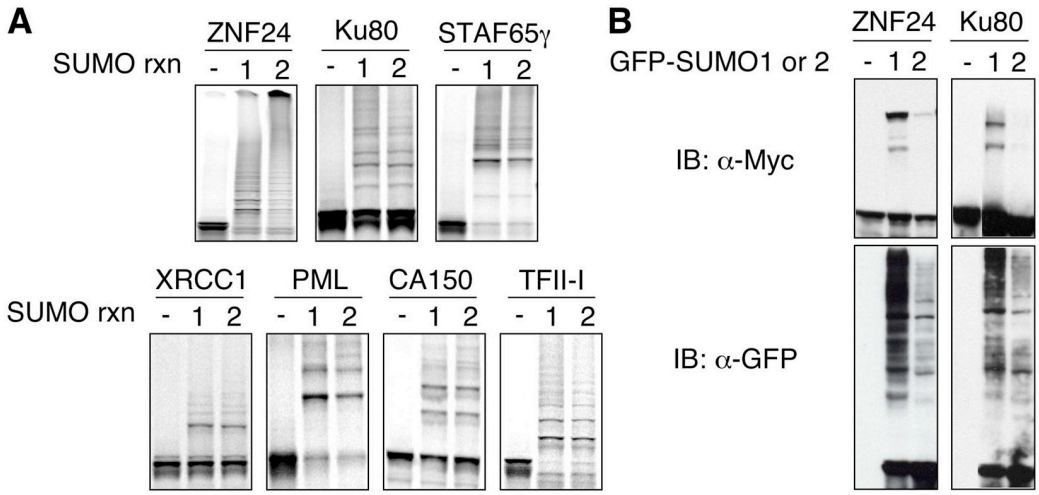

Figure 9. Conjugation Selectivity of SUMO1 and SUMO2.
(A) SUMO1 and SUMO2 are conjugated equally efficiently to a set of SUMO1 substrates identified in our IVEC screen. The indicated substrates were *in vitro* translated in the presence of ^{35}S-methionine and incubated with buffer alone (-) or SUMO reaction mixtures containing His_6-SUMO1 (1) or His_6-SUMO2 (2). The reaction mixtures were analyzed by SDS-PAGE followed by autoradiography.
(B) Myc-tagged ZNF24 or Ku80 were co-expressed in HeLa cells with vector alone (-), GFP-SUMO1 (1), or GFP-SUMO2 (2). Twenty-four hours after transfection, the total cell lysates were blotted with anti-GFP and anti-Myc antibodies.

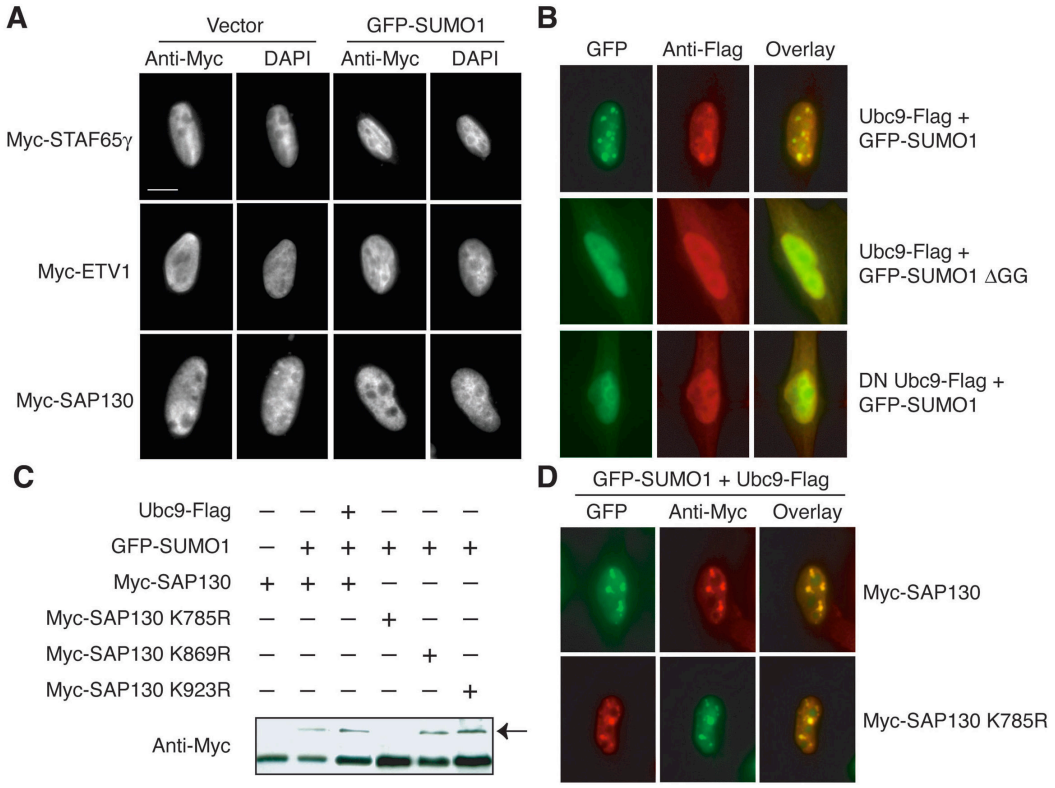

Figure 10. Subcellular Localization of SUMO1 Substrates.
(A) HeLa Tet-on cells were transfected with the indicated plasmids, fixed, and stained with anti-Myc and DAPI (DNA). The scale bar indicates 10 μm.
(B) HeLa Tet-on cells were transfected with the indicated plasmids, fixed, and stained with anti-Flag. GFP is shown in green while anti-Flag staining is shown in red.
(C) Lysates from HeLa cells transfected with the indicated plasmids were blotted with anti-Myc. The position of the SAP130–SUMO conjugate is indicated by an arrow.
(D) HeLa Tet-on cells were transfected with the indicated plasmids, fixed, and stained with anti-Myc. GFP is shown in green while anti-Myc staining is shown in red.

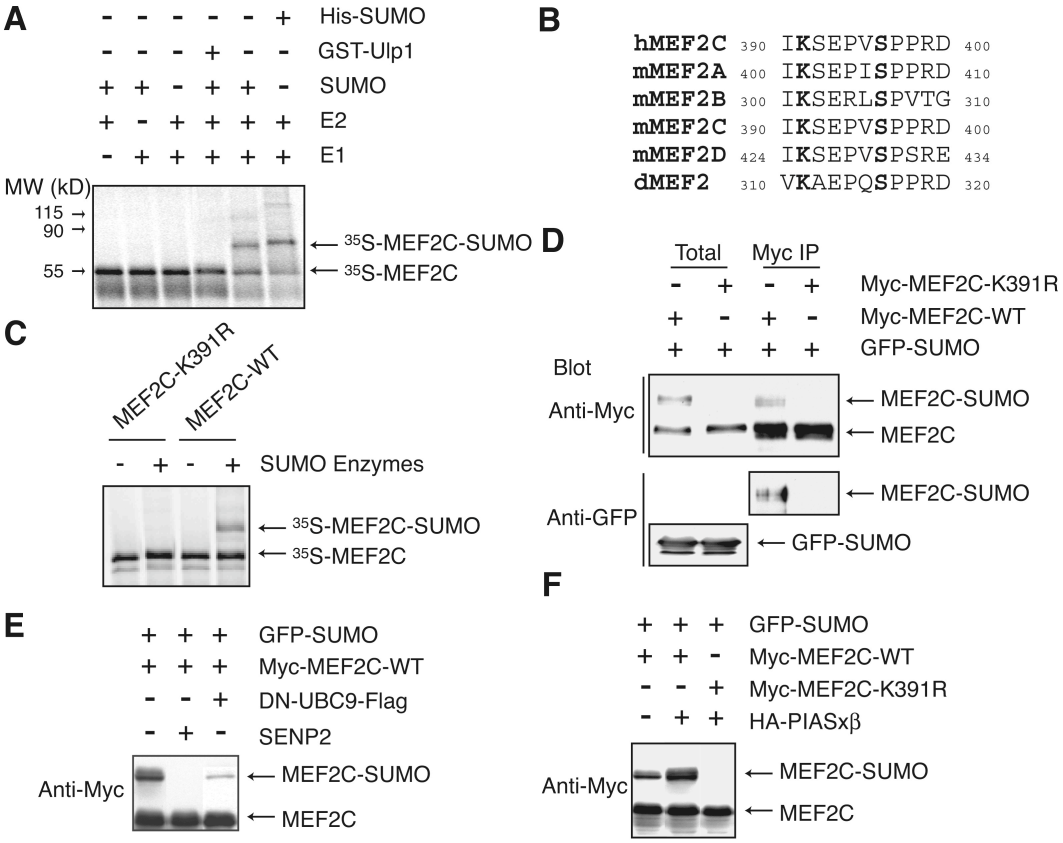

Figure 11. MEF2 proteins are sumoylated.
(A) The [^{35}S]MEF2C protein obtained through in vitro transcription and translation was incubated with SUMO reaction mixtures and analyzed by SDS-PAGE followed by autoradiography.
(B) Sequence alignment of the MEF2 family of proteins. The lysine residue of a sumoylation consensus motif and a serine residue that is phosphorylated are shown in bold.
(C) The [^{35}S]wild-type (WT) and K391R mutant of MEF2C were incubated with SUMO reaction mixtures and analyzed by SDS-PAGE followed by autoradiography.
(D) HeLa cells were transfected with the indicated plasmids. Myc-MEF2C was immunoprecipitated with anti-Myc. Cell lysates and Myc IP were resolved by SDS-PAGE and blotted with anti-Myc or anti-GFP.
(E) HeLa cells were transfected with the indicated plasmids. Cell lysates were resolved by SDS-PAGE and blotted with anti-Myc.
(F) HeLa cells were transfected with the indicated plasmids. Cell lysates were resolved by SDS-PAGE and blotted with anti-Myc.

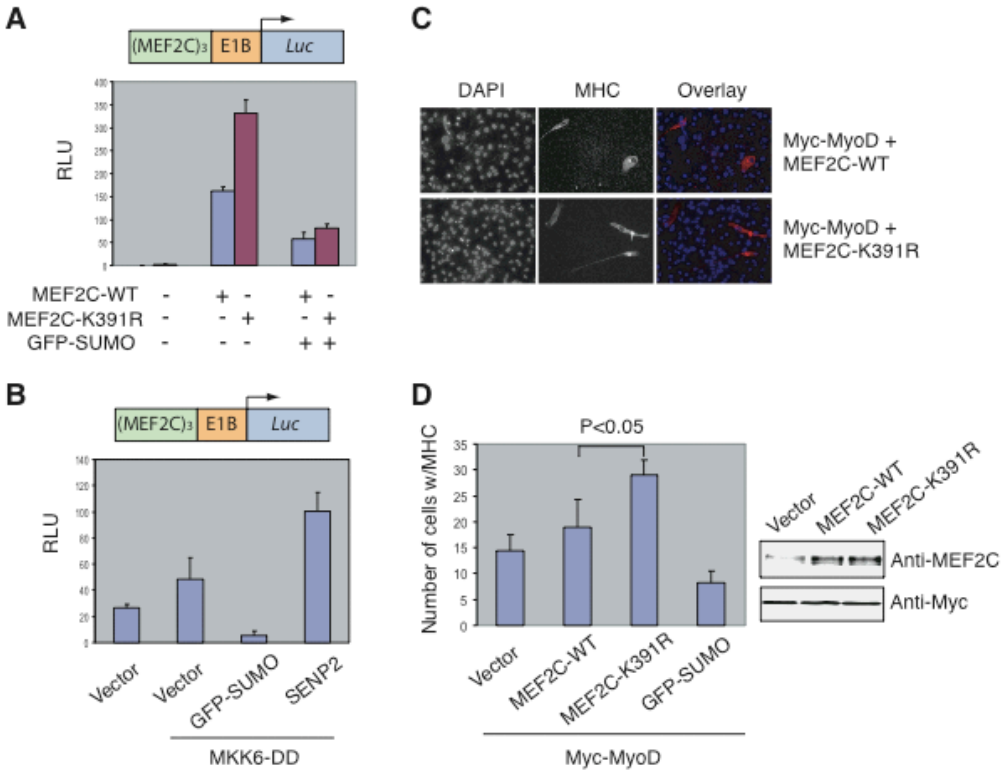

Figure 12. Sumoylation-deficient mutant of MEF2C promotes myogenic conversion more efficiently.

(A) MEF2C-WT- or MEF2C-K391R-expressing plasmids were co-transfected with *MEF2*×3-luciferase reporter, pRL-*tk* reporter, and GFP-SUMO1 or empty vector plasmids into HeLa cells. Firefly luciferase activities were measured and normalized for transfection efficiency by using *Renilla* luciferase activities.

(B) *MEF2*×3-luciferase reporter, pRL-*tk* reporter, and MKK6-DD plasmids were co-transfected with GFP-SUMO1, SENP2, or empty vector plasmids into C2C12 cells. The cells were cultured in differentiation medium for 2 days. Firefly luciferase activities were measured and normalized for transfection efficiency by using *Renilla* luciferase activities.

(C) Vector, MEF2C-WT-, MEF2C-K391R-, or GFP-SUMO1-expressing constructs were co-transfected with Myc-MyoD into 10T1/2 cells. The cells were cultured in differentiation medium for 5 days, fixed, and stained with DAPI (blue) and an anti-myosin heavy chain (MHC) monoclonal antibody (red).

(D) MHC-positive cells were scored by random selection of 20 optical fields of cells in (C). The results of two independent experiments were averaged with the standard deviation indicated. P value was calculated using student t test. Cell lysates were resolved by SDS-PAGE and blotted with anti-MEF2C or Myc.

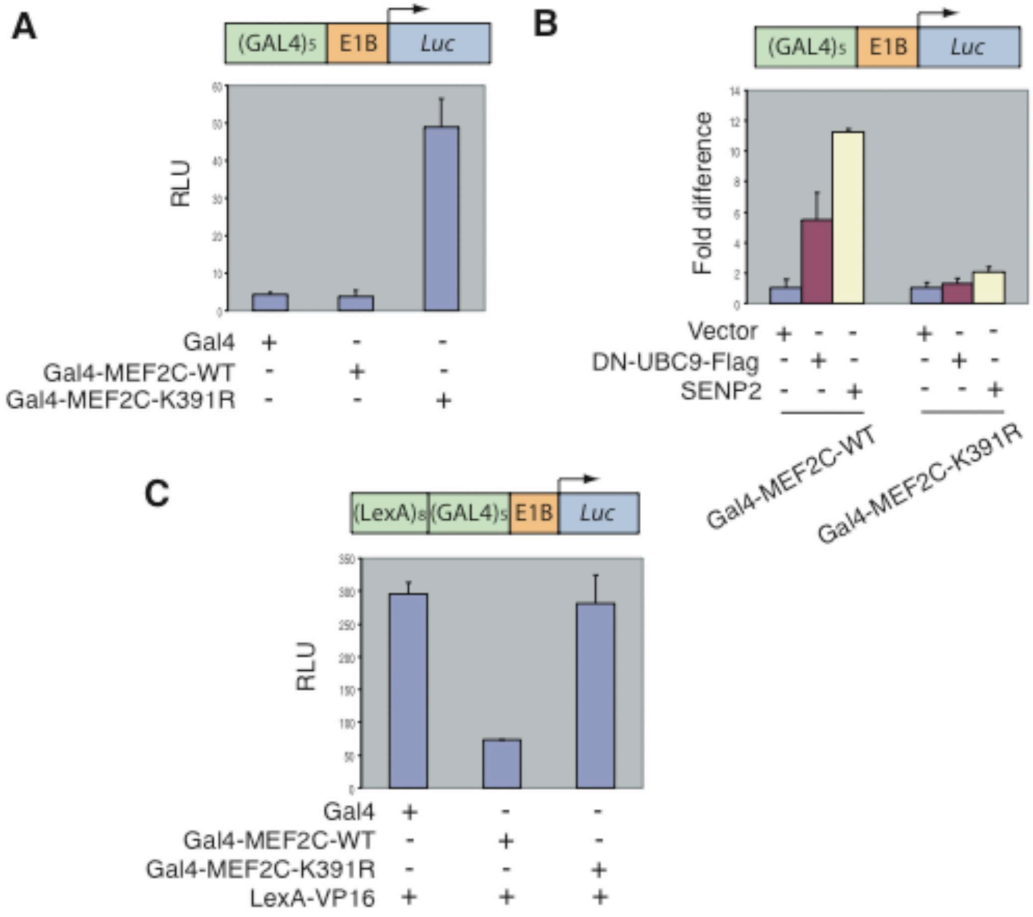

Figure 13. Sumoylation of MEF2C inhibits its transcriptional activity.
(A) Gal4, Gal4-MEF2C-WT, or Gal4-MEF2C-K391R construct was co-transfected with *GAL4*×5-luciferase reporter and pRL-*tk* reporter into HeLa cells. Firefly luciferase activities were measured and normalized for transfection efficiency by using *Renilla* luciferase activities.
(B) Gal4-MEF2C-WT or Gal4-MEF2C-K391R construct was co-transfected with *GAL4*×5-luciferase reporter, pRL-*tk* reporter, and SENP2, DN-UBC9-Flag, or vector construct into HeLa cells. Firefly luciferase activities were measured and normalized for transfection efficiency by using *Renilla* luciferase activities and then divided by the luciferase activities of Gal4-MEF2C-WT or Gal4-MEF2C-K391R, respectively, to show the fold differences.
(C) Gal4, Gal4-MEF2C-WT, or Gal4-MEF2C-K391R construct was co-transfected with *LexA*×8-*GAL4*×5-luciferase reporter, pRL-*tk* reporter, and LexA-VP16 construct into HeLa cells. Firefly luciferase activities were measured and normalized for transfection efficiency by using *Renilla* luciferase activities.

Chapter IV: Structural Basis for CoREST-Dependent Demethylation of Nucleosomes by the Human LSD1 Histone Demethylase[1]

Introduction

A nucleosome core particle––the basic building block of chromatin—consists of 147 base pairs of DNA wrapped around a histone octamer that contains an H3-H4 tetramer and two H2A-H2B dimers (76-78). A linear array of nucleosomes connected by linker DNA is further folded to form dynamic higher-order structures of chromatin (85). Posttranslational modifications of histone tails regulate transcription and other chromatin-templated processes by altering chromatin structure locally and through recruitment of effectors containing protein modules that bind to modified histone tails (79). Lysine acetylation is the best characterized histone modification, which is generally associated with transcriptional activation and is dynamically regulated by histone acetyltransferases (HATs) and deacetylases (HDACs) (79). Histone lysine methylation mediated by multiple classes of methyl transferases has emerged as another important mechanism that regulates chromatin structure and function

[1] This chapter is derived with permission from (174).

(92). Unlike acetylation, histone lysine methylation can either activate or repress transcription, depending on the location and degree (mono-, di- and tri-methylation) of these modifications (92). Two classes of histone demethylases that remove methyl groups from lysines have recently been discovered (91,94,175-177), establishing the dynamic nature of histone methylation.

Lysine-specific demethylase 1 (LSD1; also known as BHC110 and AOF2) is a flavin adenine dinucleotide (FAD)-dependent amine oxidase that removes methyl groups from mono- or di-methylated histone H3 lysine 4 (H3-K4) (91,128) (Figure 14A). LSD1 does not demethylate tri-methylated H3-K4, due to the inherent limitations of the chemistry that it uses to catalyze the demethylation reaction (91). Methylation of H3-K4 is generally associated with active transcription (99,100). Consistently, LSD1 is a component of various transcriptional corepressor complexes that often also contain HDAC1/2 and CoREST (71,129-131). Though LSD1 alone can demethylate H3-K4 in peptides or bulk histones, only the LSD1–CoREST complex is capable of demethylating H3-K4 within nucleosomes (69,70,178). The mechanism by which CoREST stimulates the demethylation of nucleosomes by LSD1 has not been established.

LSD1 consists of an N-terminal SWIRM (Swi3p, Rsc8p and Moira) domain and a C-terminal amine oxidase domain (AOD) that is separated into two halves (AOD_N and AOD_C) by a 92-residue insert (Figure 14B). CoREST consists of an ELM2 (Egl-27 and MTA1 homology 2) domain and two SANT

(SWI-SNF, ADA, N-CoR, and TFIIIB) domains. A truncation mutant of LSD1 lacking its N-terminal 184 residues retained full demethylase activity against methylated H3-K4 peptide substrates (127). A C-terminal fragment of CoREST (CoREST-C, residues 293-482) containing SANT2 and the linker between the two SANT domains efficiently stimulated the demethylase activity of LSD1 toward nucleosomes (69). Here we report the crystal structures of human LSD1 in complex with an LSD1-stimulatory domain of human CoREST. Using nuclear magnetic resonance (NMR) spectroscopy, we also show that the CoREST SANT2 domain binds to DNA. Mutagenesis studies show that DNA-binding by CoREST SANT2 is crucial for the demethylation of H3-K4 within nucleosomes by LSD1–CoREST-C. The shape and dimension of LSD1-CoREST match those of nucleosomes and readily suggest a mechanism by which DNA-binding of CoREST facilitates the histone demethylation of nucleosomes by LSD1.

Structure Determination and Overview of LSD1–CoREST

Human LSD1ΔN (residues 171-852) and CoREST-C (residues 286-482) proteins were expressed and purified from bacteria and mixed to form the LSD1ΔN–CoREST-C complex (hereafter referred to as LSD1–CoREST for simplicity) that was active in demethylating nucleosomes (data not shown). The LSD1–CoREST complex was then crystallized in the presence of the dimethylated K4 N-terminal 21-residue peptide of histone H3 (diMeK4H3-21).

The crystal structure of LSD1–CoREST was determined by single-wavelength anomalous dispersion and molecular replacement (Table 2).

The two halves of LSD1 AOD form one globular domain that consists of two lobes: the substrate-binding lobe and the FAD-binding lobe (Figures 14C and 17A). The active site of AOD is located at the interface of the two lobes. The SWIRM domain packs against the FAD-binding lobe of AOD. The LSD1 insert consists of two long helices (Iα1 and Iα2) that pack against each other in an anti-parallel orientation, forming a long stalk that projects away from AOD. The linker of CoREST folds into two helices (Lα1 and Lα2) that are arranged in a configuration reminiscent of the letter "L". The short Lα1 helix of this L-shaped linker packs against the substrate-binding lobe of AOD whereas the long Lα2 helix forms a parallel coiled-coil with Iα1 of LSD1. The SANT2 domain of CoREST connects to Lα2 through a flexible loop and a 3_{10} helix and lies at the tip of the stalk formed by the LSD1 insert and Lα2 of CoREST. Thus, LSD1–CoREST forms an elongated structure of about 150 Å in length that consists of three parts: the base that contains the SWIRM and AOD domains of LSD1, the stalk that is formed by the LSD1 insert and the CoREST linker, and the head that contains the SANT2 domain of CoREST.

Results and Discussion

The Amine Oxidase Domain of LSD1 and Its Active Site

The catalytic domain of LSD1 is closely related to classical FAD-dependent amine oxidases (91,137). Indeed, LSD1 AOD is structurally highly similar to maize polyamine oxidase (mPAO) (179,180) (Figures 2A and 2B). The root mean square deviation (rmsd) between their Cα traces is 2.8 Å. A nomenclature similar to mPAO is adopted to describe the secondary structural elements of LSD1 (Figures 15 and 16). The substrate-binding lobe of LSD1 comprises a six-stranded β sheet and five α helices (Figure 17A). The insert of LSD1 is located between Sα2 and Sα3. The FAD-binding lobe of LSD1 has an expanded Rossmann fold commonly found in dinucleotide-binding modules. FAD is deeply buried in the core of the protein. The isoalloxazine ring of FAD—and hence the active site of LSD1—is located at the interface between the two lobes of AOD (Figure 17). It shows the same characteristic distortions that have been observed in mPAO (Figure 17C) (179). In mPAO, the N5 atom of FAD that is reduced during the oxidation of the substrate methyl groups forms a hydrogen bond with a water molecule that is in turn positioned by K300 (Figure 17C). K300 has been shown to be critical for the reduction of FAD (181). K661 in LSD1 may play a similar role. F403 and Y439 in mPAO form a so-called

"aromatic sandwich" in mPAO that positions substrates near the isoalloxazine ring of FAD (Figure 17C) (179,180). This "aromatic sandwich" is not conserved in LSD1, with T810 of LSD1 occupying the position of Y439 in mPAO.

Residues that line the rims of the active sites of LSD1 and mPAO are among the least conserved between the two proteins (Figure 15). Consequently, their structural elements in these regions show several important differences, including different orientations of the Sα1 helices and shortening of Fα3 in LSD1 (Figures 17A and 17B). As a result, the active site of LSD1 has one large opening, which accommodates peptide substrates that contain side chains and are larger than polyamines (Figure 17D). In contrast, the active site of mPAO is ideal for binding long, linear polyamines and consists of a long tunnel with two surface openings that are divided by Fα3 (Figures 17B and 17E) (179). In addition, unlike mPAO, the active site in LSD1 does not contain a long tunnel that would allow peptide substrate to thread through its interior. Despite the lack of sequence conservation, the rims of the active sites of mPAO and LSD1 are both lined with multiple, negatively charged residues (Figures 17D and 17E). As the substrates of mPAO and LSD1 are positively charged, the highly negative electrostatic potential of the rims of both enzymes is expected to guide substrates into their active sites.

While this manuscript was under review, the structure of human LSD1 in the absence of CoREST has been reported (182). The structures of the free and

CoREST-bound LSD1 are virtually identical, with only a small difference in the orientations of Iα1 and Iα2 relative to the AOD. The structure of JMJD2, a JmjC-domain-containing histone demethylase, has also recently been determined (183). Though LSD1 cannot demethylate tri-methylated lysines due to the inherent limitations of its chemical mechanism, it is capable of binding to an H3 peptide that contains tri-methylated K4, suggesting that the active site of LSD1 cannot sterically discriminate between mono-/di-methylated and tri-methylated H3 peptides (182). In contrast, JMJD2 selectively demethylates tri-methylated H3-K9 and H3-K36 and possesses a binding pocket specific for the recognition of tri-methylated lysines (183).

Substrate Binding by LSD1

LSD1–CoREST was co-crystallized with a peptide that contains the N-terminal 21 residues of H3 with dimethylated K4. However, there was no interpretable electron density corresponding to the peptide. Because LSD1 specifically demethylates H3-K4 *in vitro*, the N-terminal region of the H3 peptide is expected to bind at the active site. A deep, negatively charged pocket formed by residues N540, W552, D553, D555, D556, P808, and A809 lies in the vicinity of the FAD (Figures 18A and 18B). Mutations of two residues within this pocket, D555 and D556, to alanines abrogate the histone demethylase activity of LSD1 toward the diMeK4H3-21 peptide substrate (data not shown). We propose that

this pocket is the binding site for the positively charged N-terminal amino group of H3-A1 and the side chain of H3-R2. Binding of the N-terminus of the peptide into this pocket would place the side chain of H3-K4 in the vicinity of FAD, in agreement with the substrate specificity of LSD1 toward H3-K4.

Prior biochemical studies have shown that LSD1 recognizes an unusually large, 21-residue segment of the histone H3 tail (127). The active site of LSD1 itself is not large enough to accommodate a 21-residue peptide. However, adjacent to the active site, there is a conspicuous surface groove formed by the interface between the AOD and SWIRM domains of LSD1 (Figure 18B). It is very likely that the C-terminal region of the H3 tail binds at this groove. Indeed, the N-terminal 21 residues of the H3 tail can be easily docked into the active site and the AOD-SWIRM surface groove of LSD1 without creating steric clashes (Figure 18B). In this docking model, the N-terminal 12 residues of diMeK4H3-21 bind at the active site of LSD1 whereas the rest of the peptide is located in the AOD–SWIRM groove (Figure 18B). This model also puts H3-S10 in close proximity to E559 of LSD1, providing a possible explanation for the reported observation that phosphorylation at S10 reduces the affinity of LSD1 toward the H3 peptide (127).

The SWIRM Domain of LSD1

The structure of the SWIRM domain in the presence of LSD1 AOD is almost identical to the recently reported solution structure of the SWIRM domain from human LSD1 in isolation and is highly similar to the structures of the SWIRM domains of Ada2a and Swi3 (116-118). The SWIRM domain of LSD1 consists of six α helices and a 3_{10} helix and packs against the FAD-binding lobe of LSD1 AOD (Figures 18C and 16). The SWIRM domains of Ada2a and Swi3 bind to DNA, with several residues in and around the C-terminal helix (α6 in LSD1 SWIRM) required for DNA binding (116,117). Residues implicated in DNA binding in the SWIRM domains of Ada2a and Swi3 are poorly conserved in LSD1 SWIRM. Furthermore, the C-terminal region of LSD1 AOD partially blocks the putative DNA-binding surface of α6 of LSD1 SWIRM (Figure 18C). Thus, LSD1 SWIRM is unlikely to bind to DNA in a similar manner as the SWIRM domains of Ada2a and Swi3. Indeed, we have failed to detect binding between LSD1 SWIRM and short, synthetic duplex DNA oligonucleotides using NMR spectroscopy (data not shown). Instead, the majority of residues conserved among the SWIRM domains of the LSD1 orthologs are located at the interface between AOD and SWIRM (Figure 18C), suggesting that SWIRM helps to maintain the structural integrity of LSD1 AOD.

As discussed above, LSD1 recognizes an unusually large segment of the histone H3 tail (127). In addition to the expected interactions between the N-

terminal segment of the H3 peptide and the active site of LSD1, the C-terminal portion of the H3 tail may fit into a groove formed between the SWIRM domain and the AOD of LSD1 (Figures 18B). Therefore, the SWIRM domain of LSD1 may also contribute to the binding of the H3 tail by forming one wall of the binding groove for C-terminal segment of the H3 tail (Figures 18B). Consistently, mutations of residues that lie in this groove at the AOD–SWIRM interface abrogated the demethylase activity of LSD1 (182). We emphasize that the H3 tail is likely to bind at the interface between the SWIRM and AOD domains of LSD1. The SWIRM domain alone is unlikely to be sufficient for binding to the H3 tail. Consistently, Tochio *et al*. failed to detect binding between SWIRM and an H3 tail peptide in solution by NMR (118).

Binding between LSD1 and CoREST

CoREST wraps around the stalk formed by the insert of LSD1, creating three major interfaces (Figure 19). Interface I constitutes CoREST Lα1, the loop preceding CoREST Lα1, LSD1 Sα1, and LSD1 Sα2 (Figure 19B). The interactions at this interface are largely hydrophobic in nature. The second interface consists of LSD1 Iα1, LSD1 Iα2, CoREST Lα2, and the loop that follows CoREST Lα2 (Figure 19C). In addition to hydrophobic interactions between CoREST Lα2 and LSD1 Iα1, several ionic interactions exist between Lα2 and Iα2. The third interface is between LSD1 Ia2 and CoREST SANT2

(Figure 19D). Because the linker between the two SANT domains of CoREST is sufficient for LSD1 binding (69), the interactions at interface III might not be essential for binding between LSD1 and CoREST. Instead, they may serve to position CoREST SANT2 for its interactions with nucleosomes or other effectors. In addition to stimulating the demethylation of nucleosomes by LSD1, CoREST is also required for the stability of LSD1 *in vivo* (69). It is apparent from the structure that CoREST binding serves to stabilize the helical conformation of the LSD1 insert.

DNA Binding by the SANT2 Domain of CoREST

A C-terminal CoREST fragment containing the linker and SANT2 is sufficient to stimulate LSD1-dependent demethylation of nucleosomal substrates (69) (see Figure 23 below). Though the CoREST linker alone is capable of binding to LSD1, it is insufficient to stimulate the activity of LSD1 toward nucleosomes (69). Thus, CoREST SANT2 is critical for facilitating LSD1-mediated demethylation of nucleosomes. There are conflicting data with respect to the function of CoREST SANT1 in the stimulation of LSD1. Lee et al. showed that CoREST fragments containing either SANT1 or SANT2 are sufficient for stimulating LSD1-dependent demethylation of nucleosomes (70). In contrast, Shi et al. showed that CoREST fragments containing SANT1 are insufficient to

stimulate LSD1, suggesting a strict requirement of CoREST SANT2 in this process (69).

The SANT domain is present in subunits of many chromatin-remodeling complexes (119,120). The SANT domains of Myb-related proteins interact with DNA whereas the SANT domains in Ada2, SMRT, and c-Myb bind to histone tails (121-124). It has been suggested that CoREST SANT domains might facilitate LSD1-mediated demethylation of nucleosomes by binding to histone tails (69,70). Surprisingly, using isothermal titration calorimetry (ITC) and NMR, we were unable to detect binding between CoREST SANT2 and synthetic peptides corresponding to the N-terminal tails of histone H3 (residues 1-37), H2A (1-20), H2B (1-25), and H4 (1-25) and the C-terminal tail of H2A (110-129) (Figure 20 and data not shown). Apparently, CoREST SANT2 does not bind to isolated, unmodified histone tails.

Unlike the canonical SANT domain that consists of three α helices, CoREST SANT2 consists of four α helices with α1-3 adopting a fold highly similar to the SANT2 domain of v-Myb (Figures 21A and 21B) (121). The rmsd between the Cα traces of the SANT2 domains of CoREST and v-Myb is 1.2 Å. The DNA-binding residues in v-Myb are largely conserved in CoREST SANT2. Furthermore, the molecular surface around the putative DNA-binding α3 helix of CoREST SANT2 is positively charged (Figure 21C). We thus tested whether CoREST SANT2 interacts with DNA by nuclear magnetic resonance (NMR)

spectroscopy. Titration of a double-stranded DNA oligonucleotide into ^{15}N-labeled CoREST SANT2 perturbed the chemical shifts of a subset of residues in the 2D ^{1}H-^{15}N HSQC spectra of CoREST SANT2 (Figure 21D). The dissociation constant (K_d) of SANT2–DNA was determined to be 84 μM (Figure 22A). The lack of sequential assignment prevented unequivocal mapping of the DNA-binding surface of CoREST SANT2. However, the chemical shifts of a tryptophan NεH and the side chain amide group of an asparagine or glutamine were perturbed upon DNA binding (Figure 21D). Consistently, W383 and N419 of CoREST SANT2 contact DNA in the model for the CoREST SANT2–DNA interaction (Figure 21B). Thus, CoREST SANT2 likely binds to DNA in a mode similar to v-Myb, with α3 inserting into the major groove of DNA. To further test this notion, we mutated several residues in and around α3 of CoREST SANT2 and tested the ability of these mutants to interact with DNA by NMR. None of the HSQC peaks of the K418E, N419D, R426E, and R426A/R427A mutants of CoREST SANT2 were shifted upon the addition of DNA, indicating that they all failed to bind to DNA (Figures 21E and 22). Importantly, these mutants were properly folded as revealed by the wide dispersion of their HSQC peaks.

Demethylation of Nucleosomal Substrates by LSD1–CoREST

To directly test whether DNA binding by CoREST SANT2 contributes to the ability of CoREST-C to facilitate nucleosomal demethylation by LSD1, we performed demethylation assays with increasing doses of LSD1 alone or LSD1 in the presence of CoREST-C^{WT} or the DNA-binding mutants, CoREST-C^{K418E} and CoREST-C^{N419D}, using bulk histones or nucleosomes as the substrates (Figure 23). Consistent with previous reports (69,70), LSD1 alone efficiently demethylates bulk histone substrates, but not nucleosomal substrates (Figure 23A). LSD1–CoREST-C^{WT} mediates efficient demethylation of nucleosomes (Figures 23A and 23C). LSD1–CoREST-C^{K418E} and LSD1–CoREST-C^{N419D} are about 5 fold less efficient in mediating the demethylation of nucleosomes as compared to LSD1–CoREST-C^{WT} (Figures 23A and 23C). Importantly, LSD1–CoREST-C^{K418E} and LSD1–CoREST-C^{N419D} demethylate H3-K4 of bulk histones as efficiently as LSD1–CoREST-C^{WT} (Figure 23A). These data indicate that the DNA-binding activity of CoREST SANT2 is critical for nucleosome demethylation by LSD1–CoREST-C. On the other hand, LSD1–CoREST-C^{K418E} and LSD1–CoREST-C^{N419D} still had residual activity toward nucleosomes, suggesting that CoREST might stimulate LSD1 through additional mechanisms. It is conceivable that CoREST-binding might induce a conformational change of LSD1 that enhances its demethylase activity.

CoREST SANT2 is critical for stimulating LSD1-mediated demethylation of H3-K4 in intact nucleosomes, but it is located 100 Å away from the active site of LSD1. LSD1 binds to the unmodified or modified H3 tail with a dissociation constant in the μM range (127). We have shown that CoREST SANT2 binds to DNA with weak affinity, which is critical for the ability of CoREST to stimulate the activity of LSD1 toward nucleosomes. The LSD1–CoREST complex binds to mononucleosomes with higher affinity than either protein alone (70). These findings suggest that LSD1–CoREST binds to nucleosomes through multivalent interactions. Furthermore, CoREST SANT1 shares high sequence similarity with SANT2 and may also bind to DNA. The combination of the LSD1–H3 tail and SANT–DNA interactions with individual dissociation constants between 1-100 μM can yield tight binding between LSD1–CoREST and nucleosomes with a dissociation constant in the nM range. Therefore, one mechanism by which CoREST facilitates the activity of LSD1 toward nucleosomes is to enhance the binding of LSD1 to nucleosomes by providing additional interactions with nucleosomal DNA.

The shape and dimension of LSD1–CoREST readily suggest a model for binding nucleosomes. We docked the structure of LSD1–CoREST to that of a mononucleosome with the structural restraints that one H3 tail binds to the active site of LSD1 and that α3 of CoREST SANT2 inserts into a DNA major groove (Figure 24). A nucleosome consists of about 1.7 turns of DNA wrapped around

the histone octamer in a left-handed superhelical arrangement (77). The H3 tails are nestled in the channels formed by the DNA minor grooves at superhelical locations (SHL) ±6.7 and ±0.7. Reasonable structural models can be obtained if α3 of SANT2 is placed to bind to the major grooves at SHL ±4.5 (Figure 24) or SHL ±1.5 (data not shown). The placement of SANT2 at SHL ±4.5 is more attractive, as this binding mode allows two copies of LSD1–CoREST to bind to both H3 tails concurrently using similar sets of molecular contacts, due to the pseudo-2-fold symmetry of the mononucleosome.

Conclusions

LSD1–CoREST removes methyl groups from mono- and di-methylated H3-K4 in nucleosomes. It forms an elongated structure with a long stalk connecting two nucleosome-binding modules. At the base of the stalk, the histone H3 tail binds to the active site of LSD1 and possibly to a groove at the AOD–SWIRM interface. At the tip of the stalk, the SANT2 domain of CoREST binds to DNA, which is required for the efficient demethylation of nucleosomes by LSD1–CoREST. These findings suggest that LSD1–CoREST is correctly positioned on nucleosomes and possibly chromatin through multiple, weak interactions that are spatially separated. Many chromatin-remodeling complexes contain multiple nucleosome-binding modules. For example, the structure and

domain organization of CoREST are strikingly similar to those of the C-terminal region of the ISWI chromatin remodeling factor (184). Thus, the principle of multivalent binding is most likely applicable to other chromatin-modifying enzymes.

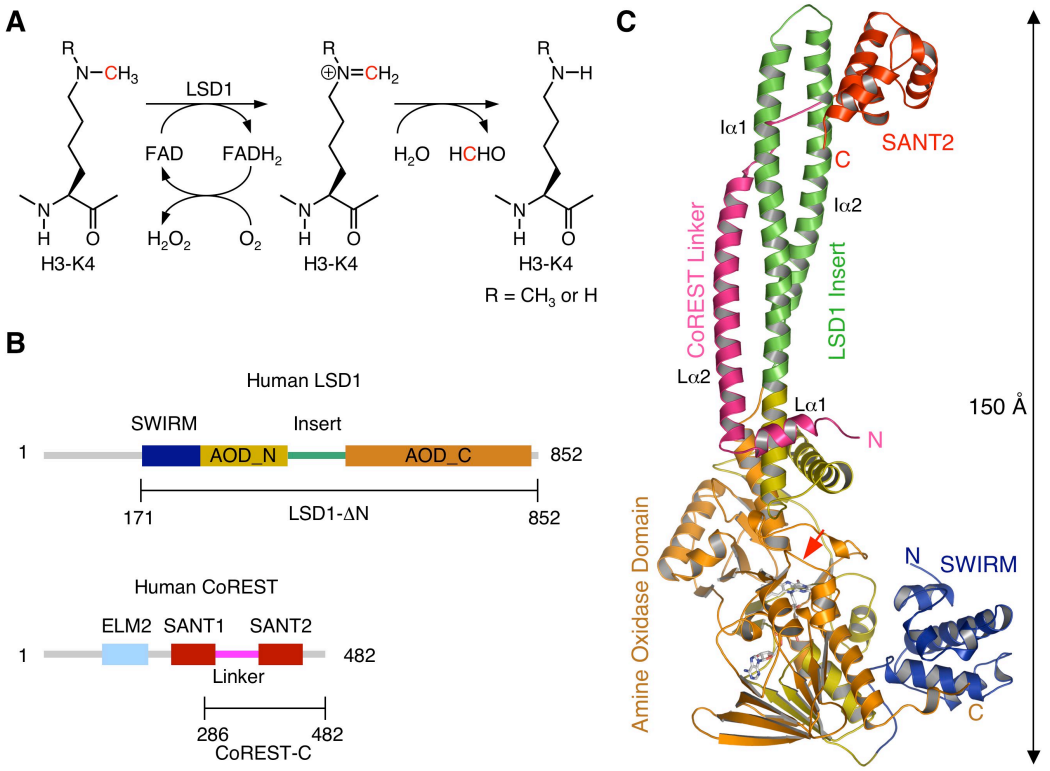

Figure 14. Structure of LSD1–CoREST

(A) Mechanism of LSD1-catalyzed demethylation of H3-K4. The carbon atom that is oxidized to form formaldehyde is shown in red.

(B) Domain structures of human LSD1 (AAH48134) and CoREST. The boundaries of proteins used in crystallization are indicated.

(C) Overall structure of LSD1–CoREST. The color scheme for this and subsequent figures is similar to that used in (B): SWIRM, blue; AOD_N, lime; AOD_C, gold; LSD1 insert, green; CoREST linker, pink; and CoREST SANT2, red. The FAD is shown in stick representation in this and subsequent figures. The red arrow indicates the active site. All structural figures are generated with PyMOL. *Accession Numbers*—Coordinates for the structure reported herein have been deposited in the Protein Data Bank with the ID code 2iw5.

Table 2: Data Collection, Structure Determination and Refinement

	Data collection	
Crystal	SeMet[a]	Native
Cell parameters, a, b, c (Å)	120.8, 178.9, 235.1	120.4, 178.2, 234.9
Resolution range (Å)	49.12-2.86 (2.91-2.86)[b]	49.87-2.57 (2.61-2.57)
Unique reflections	111,780 (4,709)	75,117 (2,423)
Multiplicity	7.1 (6.0)	9.3 (6.2)
Data completeness (%)	97.8 (82.0)	93.6 (60.6)
R_{merge} (%)[c]	9.0 (68.6)	6.1 (63.8)
I/σ(I)	30.9 (2.7)	39.3 (2.0)
Wilson B-value (Å^2)	76.8	70.3
	Phase Determination	
Anomalous scatterer		selenium (15 sites)
Figure of merit (49.1 – 2.86 Å)		0.23 (0.79 after density modification)
	Refinement Statistics	
Crystal		Native
Resolution range (Å)		20.00-2.57
No. of reflections work/free		73,416/1,498
Atoms (non-H protein, FAD)		6,388
Water molecules		50
R_{work}/R_{free} (%)		21.7/22.9
R.m.s.d. bond length (Å)		0.011
R.m.s.d. bond angle (°)		1.44
LSD1 mean B-value (Å^2)		65.1
CoREST mean B-value (Å^2)		72.5
R.m.s.d. B-value (Å^2) backbone/side chain		1.07/2.44
Correlation Coefficient F_o-F_c work/free		0.946/0.942
Residues in most favored regions (%)		91.6
Residues in allowed regions (%)		8.4
Missing residues		LSD1ΔN: 836-852

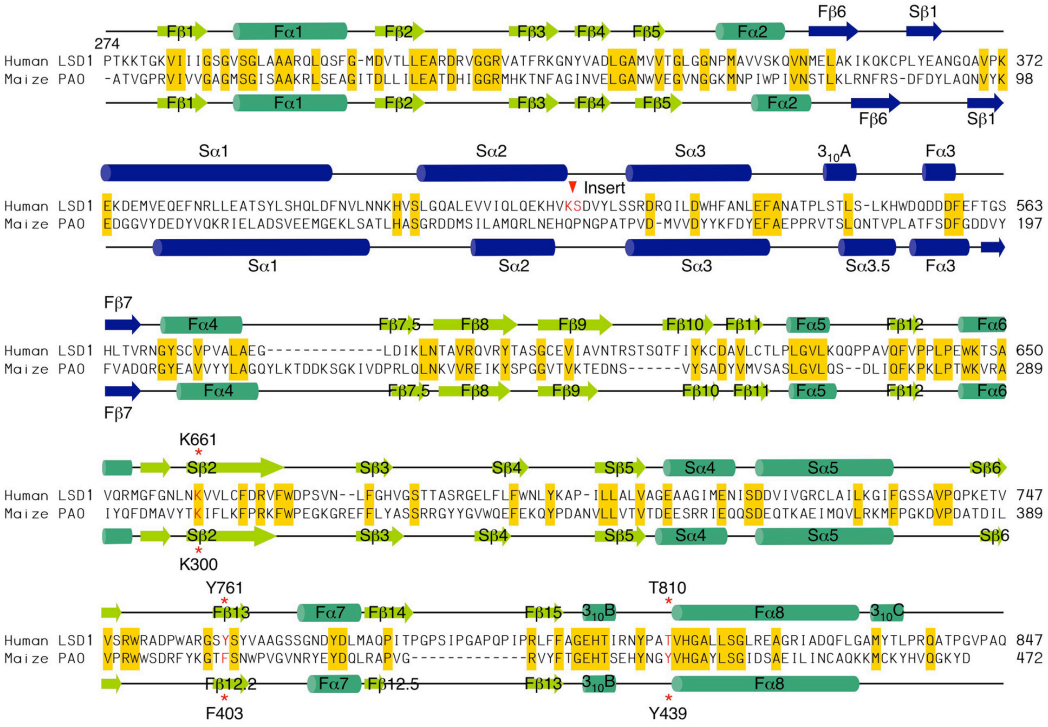

Figure 15. Sequence alignment of human LSD1 and maize PAO.
The secondary structural elements of LSD1 and mPAO are indicated above and below the sequences, respectively. The conserved residues are labeled yellow. Key active site residues are colored red and indicated by asterisks. The position of the LSD1 insert is indicated by a red arrowhead.

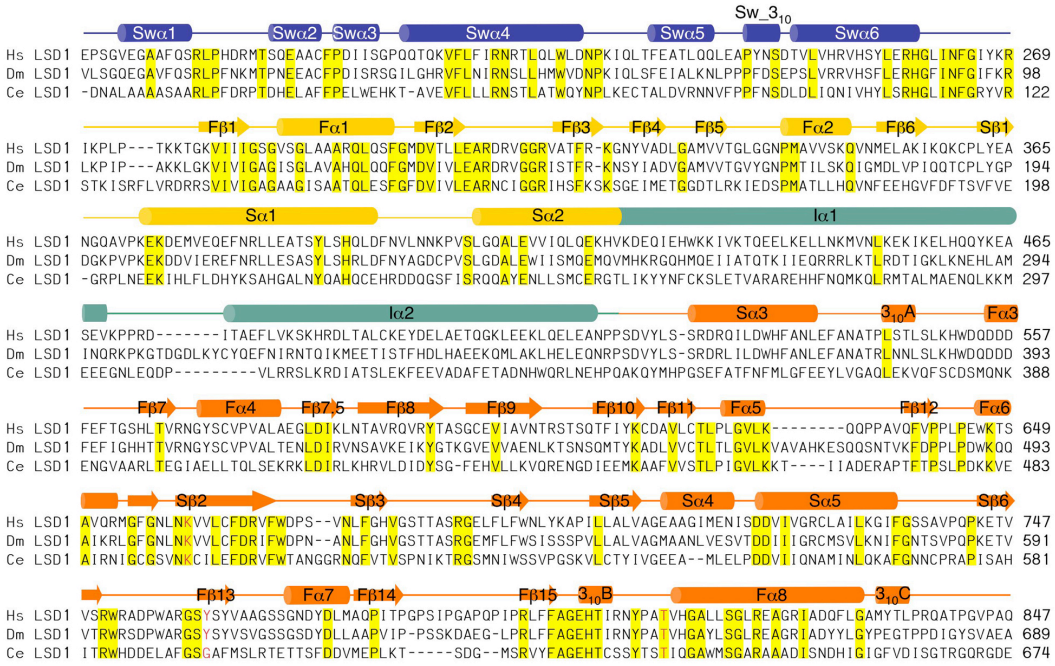

Figure 16. Sequence alignment of LSD1 orthologues

Sequence alignment of LSD1 proteins from human (Hs, *Homo sapiens*), fruit fly (Dm, *Drosophila melanogaster*), and worm (Ce, *Caenorhabditis elegans*). The secondary structural elements of LSD1 are indicated above the sequences. The conserved residues are labeled yellow. Key active site residues are colored red.

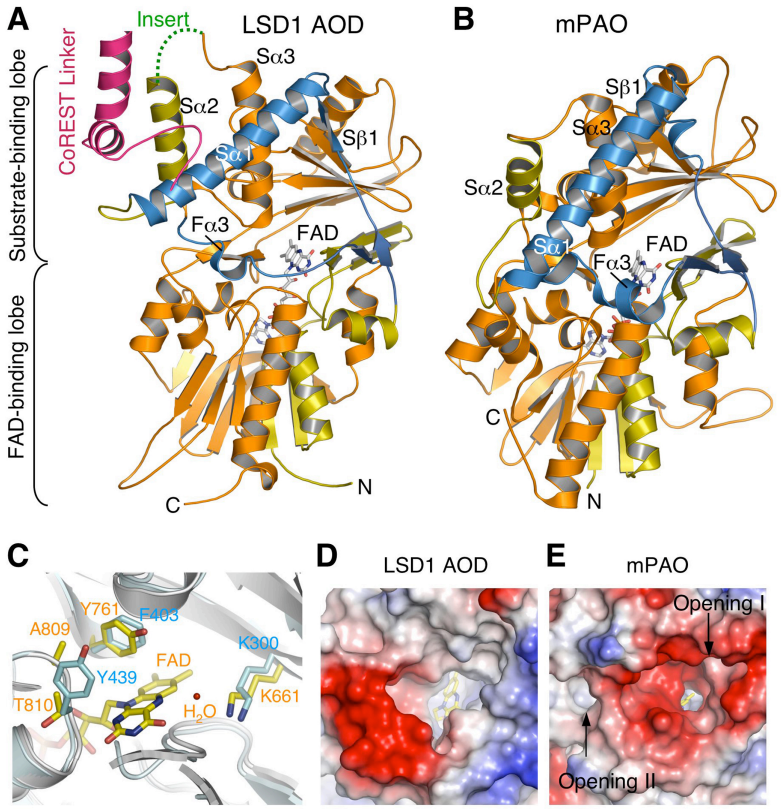

Figure 17. Structure of the Amine Oxidase Domain (AOD) of LSD1

(A) Ribbon drawing of the structure of LSD1 AOD and a portion of the CoREST linker. The structural elements lining the rim of the active site are colored blue. The location of the LSD1 insert is indicated.

(B) Ribbon drawing of the structure of maize PAO. The regions in mPAO that correspond to AOD_N and AOD_C in LSD1 are colored lime and gold, respectively. The structural elements lining the rim of the active site are colored blue.

(C) Overlay of the active site residues of mPAO and LSD1. The ribbons of mPAO and LSD1 are colored cyan and gray, respectively. The active site residues of mPAO and LSD1 are shown as cyan and yellow sticks, respectively. Only FAD in LSD1 is shown for clarity.

(D) Molecular surface of the active site of LSD1 AOD in similar orientation as in (A) with the positive and negative electrostatic potentials colored blue and red, respectively.

(E) Molecular surface of the active site of mPAO in similar orientation as in (B) with superimposed positive and negative electrostatic potentials colored blue and red, respectively. The two openings of the long active site tunnel are indicated.

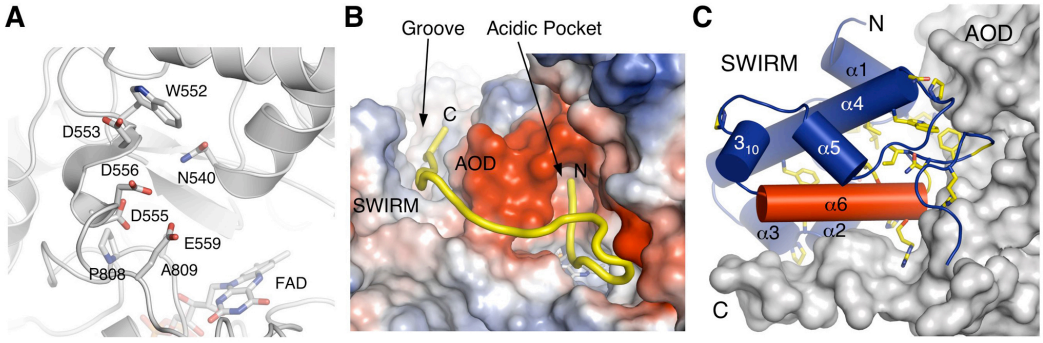

Figure 18. The substrate-binding site of LSD1
(A) A negatively charged pocket in the active site of LSD1. Residues that form this pocket are shown in sticks.
(B) Model of substrate binding to LSD1. The H3 tail (shown as a yellow tube) is docked into a deep acidic pocket in the active site of LSD1 and a surface groove between AOD and SWIRM based on existing biochemical evidence. The molecular surface of LSD1 AOD and SWIRM is colored based on electrostatic potential.
(C) Interactions between SWIRM and AOD of LSD1. The SWIRM domain is shown in cartoon drawing whereas the molecular surface of AOD is shown. Residues conserved among LSD1 orthologs are shown as yellow sticks. The α6 helix (colored red) corresponds to the DNA-binding helix in the SWIRM domains of Ada2a and Swi3.

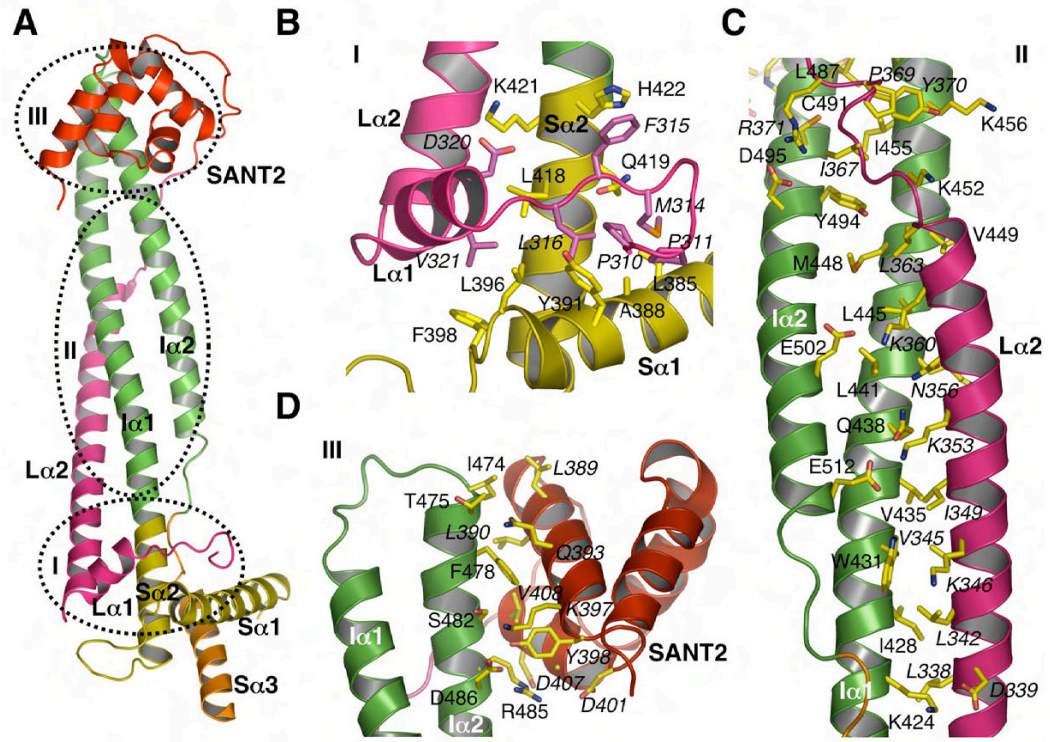

Figure 19. Interactions between LSD1 and CoREST
(A) Overview of the three binding interfaces (I-III) between LSD1 and CoREST. (B-D) Detailed molecular contacts between LSD1 and CoREST at interfaces I-III, respectively. The CoREST residues are labeled in italics.

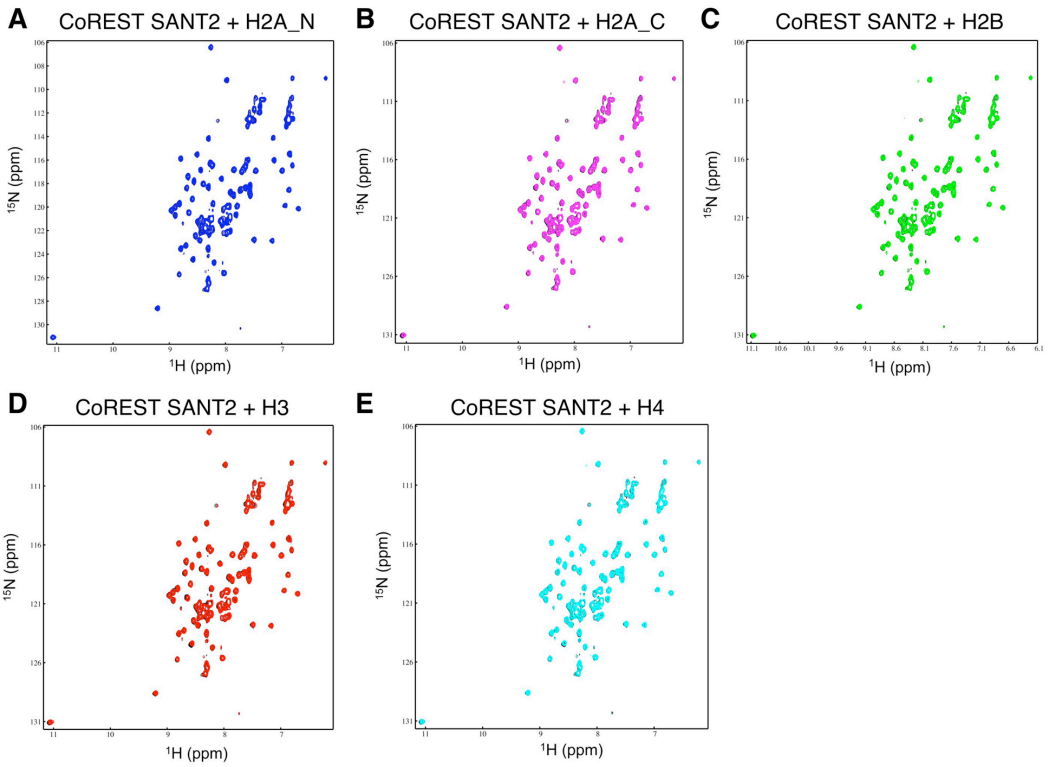

Figure 20. CoREST SANT2 Does Not Bind to Free, Unmodified Histone Tails

Overlay of the $^{15}N/^{1}H$ HSQC spectra of ^{15}N-labeled CoREST SANT2 before (black contours) and after (colored contours) the addition of the histone tail peptides.

(A) The N-terminal H2A peptide (blue contours).
(B) The C-terminal H2A peptide (pink contours).
(C) The N-terminal H2B peptide (green contours).
(D) The N-terminal H3 peptide (red contours).
(E) The N-terminal H4 peptide (cyan contours).

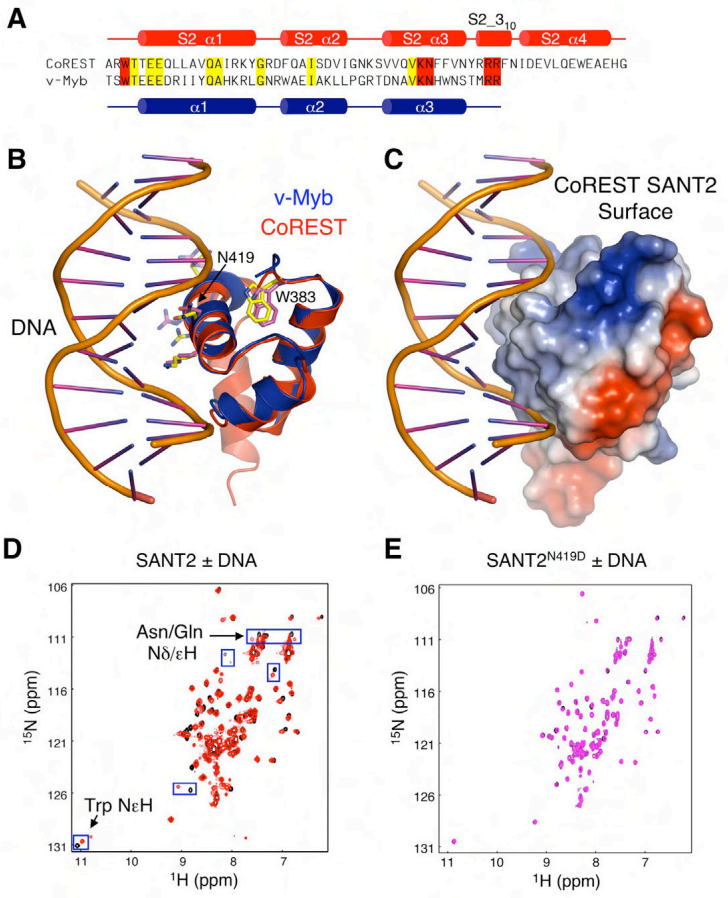

Figure 21. DNA Binding of CoREST SANT2

(A) Sequence alignment of the SANT2 domains of CoREST and v-Myb with their secondary structural elements indicated above and below the sequences, respectively. The conserved DNA-binding residues are labeled red whereas other conserved residues are labeled yellow.

(B) Ribbon drawing of the structure of CoREST SANT2 (red) overlaid on the structure of the v-Myb SANT2 (blue) bound to DNA. The conserved DNA-binding residues are shown as purple (CoREST) and yellow (v-Myb) sticks.

(C) Molecular surface of CoREST SANT2 in the CoREST SANT2–DNA model in the same orientation as in (B).

(D) Overlay of the ^{15}N/^{1}H HSQC spectra of ^{15}N-labeled CoREST SANT2 before (black contours) and after (red contours) the addition of DNA. Peaks that undergo large chemical shift changes are boxed.

(E) Overlay of the ^{15}N/^{1}H HSQC spectra of ^{15}N-labeled CoREST SANT2^{N419D} mutant before (black contours) and after (pink contours) the addition of DNA.

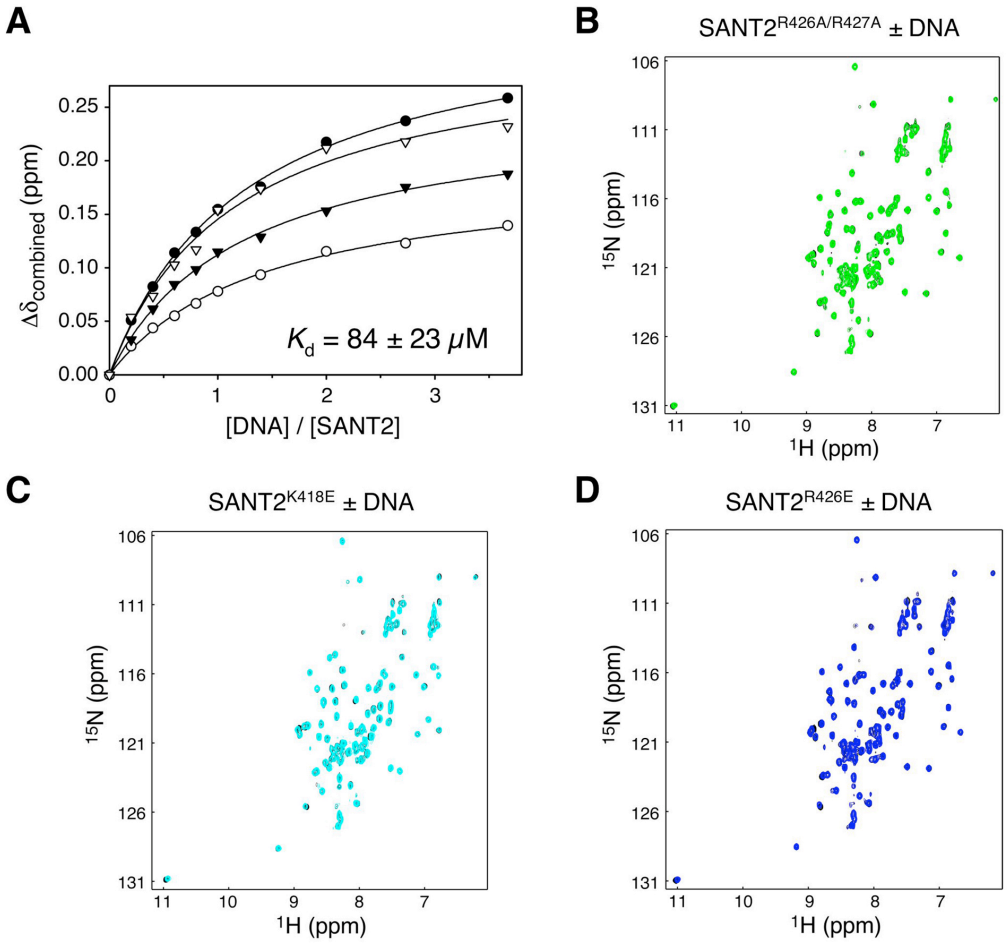

Figure 22. Determination of Dissociation Constant for DNA Binding
(A) Chemical shift changes of four CoREST SANT2 residues are plotted against the molar ratio of DNA/SANT2. K_d of SANT2–DNA and the standard deviation are indicated.
(B) Overlay of the $^{15}N/^{1}H$ HSQC spectra of ^{15}N-labeled CoREST SANT2$^{R426A/R427A}$ before (black contours) and after (green contours) the addition of DNA.
(C) Overlay of the $^{15}N/^{1}H$ HSQC spectra of ^{15}N-labeled CoREST SANT2^{K418E} before (black contours) and after (cyan contours) the addition of DNA.
(D) Overlay of the $^{15}N/^{1}H$ HSQC spectra of ^{15}N-labeled CoREST SANT2^{R426E} before (black contours) and after (blue contours) the addition of DNA.

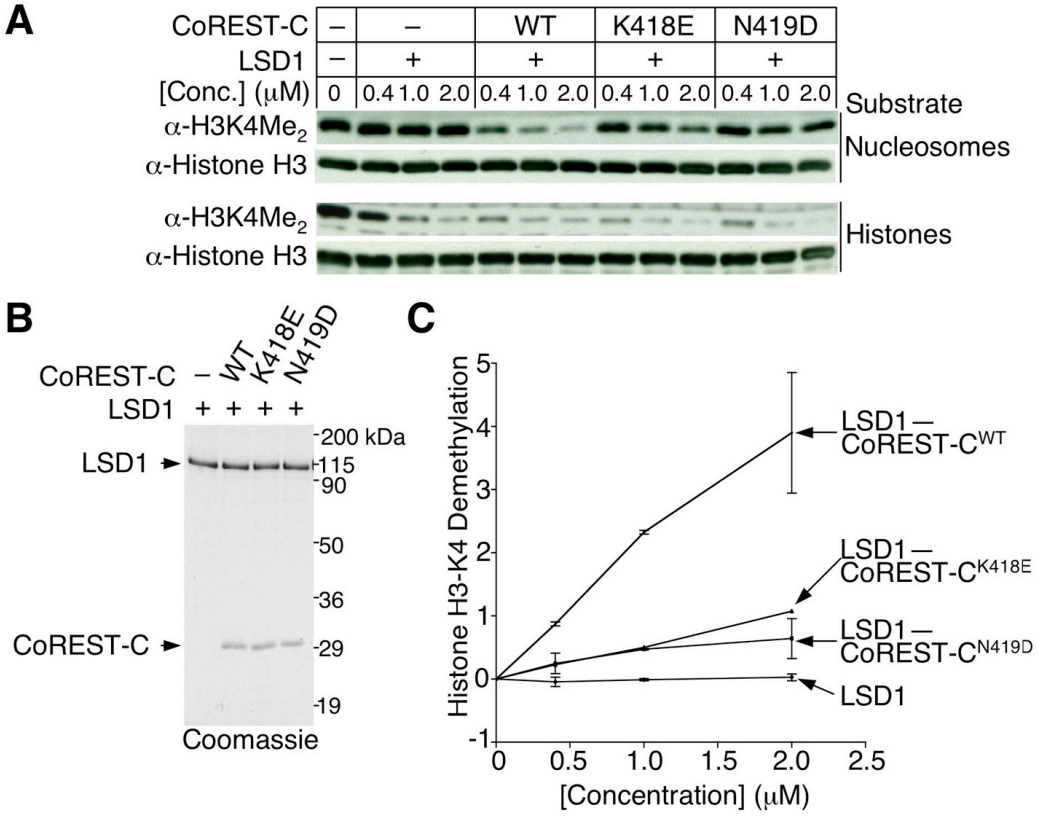

Figure 23. DNA Binding of CoREST SANT2 Is Required for Efficient Demethylation of Nucleosomes by LSD1–CoREST

(A) Bulk histones (bottom two panels) or nucleosomes (top two panels) were incubated with increasing concentrations of LSD1 alone, the His_6-LSD1–CoREST-C^{WT} complex (WT), the His_6-LSD1–CoREST-C^{K418E} complex (K418E), or the His_6-LSD1–CoREST-C^{N419D} complex (N419D). Reactions were separated on SDS-PAGE and blotted with α-$H3K4Me_2$, stripped and re-probed with α-Histone H3 as a loading control.

(B) His_6-LSD1 alone (LSD1) or His_6-LSD1–CoREST-C complexes were separated on 15% SDS-PAGE followed by Coomassie blue staining.

(C) Quantitation of the demethylation of nucleosomal substrates in (A) by densitometry. The intensities of bands in the top panel of (A) were quantified by using ImageQuant 5.2 (Fuji). The extent of nucleosome demethylation (y-axis) was represented as the ratios of the H3K4Me2 signals in the absence or presence of the LSD1-containing enzymes and plotted against enzyme concentrations (x-axis). The average and standard deviation are shown.

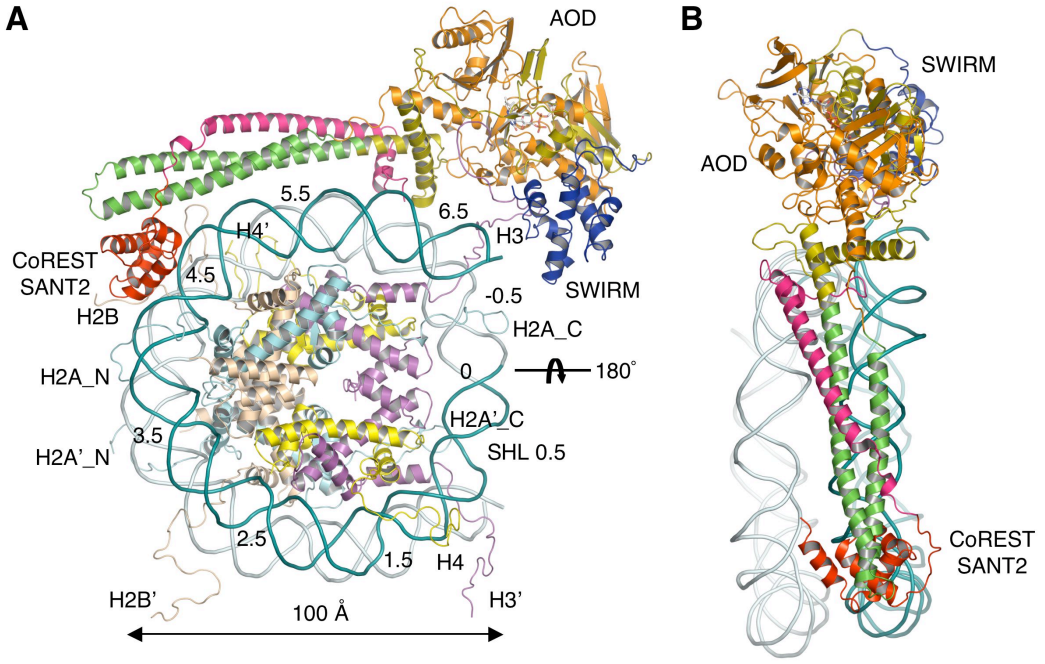

Figure 24. Shape and Dimension of LSD1–CoREST Match Those of the Mononucleosome

(A) LSD1–CoREST is docked onto a nucleosome (PDB ID 1KX5) with one H3 tail inserted into the active site of LSD1 and with CoREST SANT2 binding a DNA major groove. The superhelical locations (SHL) of DNA and the positions of histone tails are labeled. The first turn of the DNA superhelix is shown in cyan whereas the second turn is shown in gray. Histones H2A, H2B, H3, and H4 are colored cyan, wheat, pink, and yellow, respectively. The pseudo-2-fold axis of mononucleosome is indicated.

(B) Model in (A) rotated 90˚ around the vertical axis followed by a 90˚ rotation around the horizontal axis. The histone octamer is not shown for clarity.

Chapter V: Regulation of CoREST Corepressor Complex by Sumoylation

Introduction

The ordered assembly of genomic DNA into a proteinacious substance—chromatin—allows for high-order regulation of DNA-templated processes such as transcription, replication, and DNA repair (87). Chromatin is made up of repeating units of nucleosomes, which consist of one histone H3/H4 tetramer and two H2A/H2B dimers wrapped by double-stranded DNA (76-78). Polymers of nucleosomes flanked by various lengths of linker DNA can fold into compacted high-order structures that are subject to dynamic regulation (85). Histones also possess flexible tails that are susceptible to many types of post-translational modifications, which can directly or indirectly affect chromatin structure (79).

Histone acetylation is strongly associated with transcriptional activation and is dynamically regulated by histone acetyltransferases (HATs) and histone deacetylases (HDACs) (79). The effect of histone lysine methylation, catalyzed by methyltransferases, depends on the specific residue and degree (mono-, di-, or trimethylation) of modification (92). Histone H3 lysine 4 di- and trimethylation

(H3K4me2/3) is associated with active promoters (83,97,99,100), but H3K9me2/3 is usually associated with repression (97). Lysine-specific demethylase 1 (LSD1; also known as BHC110 or AOF2) is a flavin adenine dinucleotide (FAD)-dependent amine oxidase that demethylates histone H3K4me1/2, but not H3K4me3 (91,128). Although LSD1 alone can demethylate bulk histone or peptide substrates, it requires a co-factor, REST corepressor (CoREST), for efficient binding to nucleosomes and demethylation of nucleosomal substrates (69,70,174). LSD1 stability in the cell is dependent on CoREST, highlighting the importance of their association (69). A fraction of the abundant Class I deacetylases, HDAC1 and HDAC2 (85% identical), also associate with LSD1-CoREST, forming a LSD1-CoREST-HDAC1/2 (LCH) core ternary complex (71,129-131). Formation of this complex on chromatin is important because HDAC1/2 and LSD1 can stimulate each others activity through CoREST (126).

Multiple LCH-associated factors have been identified. CtBP1/2 incorporates LCH into a complex containing other repressors such as G9a histone H3K9 methyltransferase (130). CtBP1/2, by binding to sequence specific repressors that have Krüppel-like zinc fingers, such as ZEB1/2, can recruit these complexes to chromatin (134). Alternatively, the Krüppel-like zinc finger protein REST, directly recruits the LCH core to specific promoters (132,133,135). LSD1 is also found at androgen and estrogen responsive promoters, and in this context

acts as a transcriptional activator that can promote the demethylation of H3K9me1/2 (95,96,140). The role of CoREST in this context is unknown.

ZNF198 and ZNF261 bind to LCH complex (69-71). These proteins are members of a protein family characterized by a stretch of unique tandem zinc-fingers called MYM (myeloproliferative and mental retardation) domains (185). The MYM-domains of ZNF198 are frequently fused to FGF receptor kinase in myeloproliferative syndromes (186-188). Disruptions near the ZNF261 gene have been linked to X-linked mental retardation (189). Though there are other human MYM-domain proteins, only ZNF198, ZNF261, and ZNF262 share similar domain architecture with the *Drosophila* ortholog, Without children (dWoc; Figure 25) (190-192). dWoc is essential for viability (190-192), shows a pattern of chromatin binding similar to general transcription factors, and is required to prevent telomeric fusions (190). Recent reports have identified ZNF198 as a non-covalent binding partner for small ubiquitin-like modifier (SUMO; also known as Sentrin, Smt3) (35,66,68). In addition, ZNF198, ZNF262, HDAC1, and LSD1 are known to be covalently conjugated to SUMO (22,54,72,73).

SUMO-conjugation (sumoylation) targets a diverse subset of proteins in the cell (3,5,8-11,153). The enzymology of sumoylation is very similar to ubiquitination, though SUMO is not known to target proteins for degradation (5). Many SUMO substrates are proteins with chromatin-templated functions such as

transcription (5,9,10). Furthermore, in many instances SUMO promotes transcriptional repression (193). Sumoylation of several transcription factors, for example, represses transcription by promoting protein-protein interactions with Daxx (38,194,195), a PML body localized protein (43,173). Interestingly, Daxx, as well as many other proteins, contains a SUMO-interaction motif (SIM) that mediates non-covalent interactions with SUMO (35,37-39). A SIM consensus has been identified, consisting of several hydrophobic residues flanked by a series of acidic residues (35,39). SUMO does not always function by promoting protein-protein interactions. Sumoylation of several factors antagonizes either their acetylation (196-198) or ubiquitination (31,32) due to direct competition for the same lysine. Sumoylation can also disrupt protein-protein interactions, as is the case for the interaction between MBD1 and SETDB1 (75). In this case, SUMO acts as a transcriptional activator, highlighting the diverse functions of this modification.

Multiple subunits within the same chromatin-associated complex are often subject to sumoylation (22,56), as is the case for the ZNF198-LCH complex (22,54,72,73). Because of the many examples of SUMO affecting protein-protein interactions, it is possible that SUMO may have a general role in regulating the assembly or disassembly of large complexes such as LCH. In this study, we determine the requirements for the interaction between ZNF198 and LCH complex as well as between ZNF198 and sumoylated HDAC1. We also

examined the role of SUMO conjugation and non-covalent SUMO binding in regulating complex formation *in vitro*. Finally, we examined the functional significance for ZNF198 and its family members in regulating LCH complex *in vivo*. These studies overall are an important step in understanding how LCH complexes may be regulated by SUMO.

Results

Domain Analysis of MYM-type Zinc-finger Proteins

The MYM-domain family in humans is made up of five proteins: ZNF198, ZNF261, ZNF262, ZNF237, and ZNF258 (see illustration in Figure 25) (185,199). These proteins are defined by many repeats of zinc-fingers (MYM-domains) that have the unique consensus $CX_2CX_{19-24}[F/Y]CX_3C$ (185). Some of the zinc-fingers in ZNF237 and ZNF258 have cysteine substitutions (Figure 25, asterisk under red bars), whereas the zinc-fingers in ZNF198, ZNF261, and ZNF262 all appear intact. The latter three proteins also possess additional features that differentiate them from ZNF237 and ZNF258. They encode a proline/valine-rich domain of approximately 100 amino acids that is downstream of the MYM-domain region. Downstream from this is a domain that is also conserved in several non-MYM containing proteins (Figure 25, gold colored). Although primary amino acid sequence analysis alone (BLAST) yielded no function for this domain, a search with a Meta-predictor (3D-Jury) returned a high score (> 100) for the DNA breaking-rejoining enzyme fold (Figure 25, bottom) (200-202). Further structural analysis of this domain will be needed to confirm these findings.

ZNF198 Interacts with Only the Intact LSD1-CoREST-HDAC1 Complex

ZNF198 and ZNF261 co-immunoprecipitate (co-IP) with LSD1 substoichiometrically (69-71). To explore the reciprocal interaction, an endogenous IP of ZNF198 was performed from HEK293 and HeLa cells (Figure 26A and data not shown). Interacting proteins were detected by Colloidal blue staining followed by mass spectrometry. LSD1, CoREST (CoREST1), and HDAC1/2 were present at near stoichiometric levels. Co-IP of these proteins decreased after RNAi of ZNF198, ruling out non-specific binding to our anti-ZNF198 antibody (Figure 26E and data not shown). The CoREST homologues, CoREST2 and CoREST3, the MYM-domain protein ZNF262, as well as the abundant proteins tubulin, Hsp70, and dynein were also identified. Overall, these data suggest ZNF198 forms a tight complex with a subpopulation of LCH complex.

We next sought to determine which subunit of the LCH mediates binding to ZNF198. To do this we purified recombinant ZNF198, CoREST, LSD1, and HDAC1 (Figure 26C and 26D). Surprisingly, in GST pull-down assays ZNF198 does not bind efficiently to GST-CoREST alone, or in binary combinations with His-LSD1 or HDAC1-FLAG (Figure 26B, lanes 6-8 and Figure 26D, lanes 2 and 3). However, efficient binding could be achieved in the presence of the ternary LCH complex (Figure 26B, lane 2; Figure 26D, lane 1). Inhibition of HDAC1 or LSD1 activity using Trichostatin A (203,204) (TSA) and/or Tranylcypromine

(205) (TCP), respectively, alone or in combination, had no effect on binding (Figure 26B, lanes 3-5). Consistent with ZNF198 interacting only with LCH ternary complex, depletion of LSD1 by RNAi dramatically reduced the amount of CoREST and HDAC1 that co-IPs with ZNF198 (Figure 26E, top panel, compare lanes 1 and 3). Knock-down of ZNF198, ZNF261, and ZNF262 together (MYM RNAi) had no significant effect on the association of LSD1 with CoREST or HDAC1 (Figure 26E, middle panel, compare lanes 1 and 2). Thus, while ZNF198 favors binding to the LCH complex, it does not appear to stabilize global complex formation.

ZNF198 Interacts with Sumoylated Substrates, but is not an E3 Ligase

Previous reports using yeast two-hybrid and GST pull-downs from cell lysates identified ZNF198 (35,66,68) and LSD1 (66) as non-covalent binding partners for SUMO. To confirm these reports, we used recombinant GST, GST-SUMO1, or GST-SUMO2 in pull-down assays with several different individual [^{35}S]proteins (Figure 27A). Interestingly, ZNF198 bound preferentially to SUMO2 in this assay, though a positive control, PIASxβ (13,37), interacted with both SUMO1 and SUMO2. Binding to LSD1, CoREST, HDAC1, or MMS21 (an efficient SUMO substrate and SUMO E3 ligase (12,16)) was not detected. To determine whether ZNF198 also binds to sumoylated substrates, we performed FLAG IPs with either sumoylated or unsumoylated HDAC1-FLAG as bait

(Figure 27B). Efficient binding of ZNF198 could be detected after only minimal sumoylation of HDAC1, while binding to CoREST was unaffected in a parallel assay.

Many non-covalent binding partners of SUMO are efficiently sumoylated *in vitro* (13,17). To test this notion, we performed an *in vitro* sumoylation assay on [^{35}S]ZNF198 (Figure 28A), a previously identified SUMO-substrate in cells (73). ZNF198 was significantly shifted only after the addition of all the sumoylation machinery, Aos1-Uba2 (E1), Ubc9 (E2), SUMO2, and ATP, but not when SUMO2 was left out. Many efficient SUMO substrates with SUMO binding capacity (such as RanBP2 and PIASxβ) are also SUMO E3 ligases (13,17). To test whether ZNF198 is an E3 ligase, an *in vitro* sumoylation assay on LCH complex with or without recombinant ZNF198 was performed (Figure 28B). Sumoylation of both LSD1 and HDAC1 was detected only after addition of ATP, as determined by western blotting. However, this sumoylation was not stimulated by the addition of ZNF198. Additionally, efficient sumoylation of ZNF198 occurs on fragments that do not bind to SUMO-HDAC1 or SUMO alone (data not shown), suggesting that these two properties of ZNF198 are separable. Therefore, it is unlikely ZNF198 functions as an E3 ligase.

We next wanted to know whether SUMO or ZNF198 affect HDAC1 activity. Previous reports identified the SUMO attachment sites in HDAC as important for its activity, as measured by IP deacetylase assays following

transfection (72,206). To directly test whether sumoylation of HDAC1 is sufficient to stimulate its deacetylase activity, we compared the activity of sumoylated versus un-modified recombinant HDAC1 on bulk histones (Figure 28C). No significant stimulation of activity was observed, as determined by western blotting. Also, ZNF198 did not reproducibly stimulate the activity of either SUMO2-HDAC1, HDAC1 alone, or the LCH complex (Figure 28C). These data suggest SUMO or ZNF198 do not directly stimulate the enzymatic activity of HDAC1.

Antagonism between Sumoylation and LCH Complex Formation *In Vitro*

Because we were unable to assign a clear function to HDAC1 sumoylation, we entertained the possibility that SUMO might be affecting LCH complex formation. Sumoylation in other complexes can affect protein-protein interactions positively (35,37,39) or negatively (74,75). Since CoREST binding to HDAC1 and LSD1 is required for efficient activity on chromatin (126), we tested the impact sumoylation has on binding of these factors to CoREST using GST pull-down assays (Figure 29A and 29B). Surprisingly, SUMO2-HDAC1 could not interact with CoREST, though SUMO2-LSD1 and un-modified HDAC1 binding appeared normal.

Conversely, to determine if LCH complex formation affects sumoylation of individual subunits, we co-expressed GFP-SUMO constructs with HA-LSD1

alone, HDAC1-FLAG alone, or these two proteins together along with HA-CoREST (Figure 29C). The gel mobility of a fraction of LSD1 or HDAC1 was significantly impeded after co-expression with GFP-SUMO1 (Figure 29C) or GFP-SUMO2 (data not shown), consistent with sumoylation. However, if CoREST, HDAC1, and LSD1 were all co-expressed, sumoylation of LSD1, but not HDAC1, was significantly diminished. Similar results were obtained with GFP-SUMO2 co-expression (data not shown). Consistently, purified CoREST inhibited sumoylation of LSD1, but not a non-specific control, N2-Mef2c (152), in an *in vitro* assay (Figure 29D). Thus, CoREST favors binding to the un-sumoylated form of HDAC1 and inhibits LSD1 sumoylation.

MYM-type Zinc-fingers Function as Protein-Protein Interaction Modules

To better understand the interplay between SUMO and complex formation in relation to ZNF198, we mapped the protein-protein interactions of ZNF198 with CoREST complex and SUMO2-HDAC1. Figure 30A illustrates several fragments of ZNF198 that were constructed along with the location of domains and putative SUMO interaction motifs (SIMs) (68). *In vitro* binding assays were used to compare binding of these [^{35}S]fragments to either HDAC1, SUMO2-HDAC1, or LSD1-CoREST-HDAC1 (Figure 30B, see Coomassie and autoradiogram). Though wild-type ZNF198 interacted as expected, further truncating from the N- and C-termini showed that the MYM-domain containing

region was necessary and sufficient for binding (Figure 30B). Interactions were unaffected by the presence of ethidium bromide, suggesting independence from DNA (207). After analysis of many fragments (data not shown) we identified two zinc-fingers, MYM 8-9, as sufficient for binding LCH complex *in vitro* (Figure 30B). Co-IPs with Myc-tagged MYM 8-9 demonstrated an interaction with endogenous LSD1 but not a control, ZNF198 (Figure 30C). A much larger region of the MYM-domains, including the LCH binding region, was required for SUMO2-HDAC1 binding (Figure 30B, compare MYM 1-10 to MYM 1-5 and 6-10). Notably, SUMO2-HDAC1 binding did not require N-terminal SIMs 1/2 (68), and mutation of a third putative SIM within the MYM region also had no effect in this assay (Figure 30B). In summary, MYM-domains 8-9 mediate LCH binding, whereas a much larger region of MYM-domains is required for binding to SUMO-HDAC1.

REST repressor also binds to LCH complex (132,133) and has many tandem C2H2-type zinc-fingers (Krüppel-like) that mediate binding to specific sequences in promoters (RE1-elements) (208,209). Likewise, ZNF198 and its family members also contain many tandem zinc-fingers, albeit of the MYM-type (185,186,188)(Figure 25). The structure of a MYM-domain was recently solved by NMR (210). The fold is strikingly different from that of a Krüppel-like zinc-finger (211) (Figure 30, compare right and left panels). Most importantly, the extended alpha-helix that normally mediates DNA binding in most zinc-finger

proteins (212) is truncated in the MYM-domain structure. Additionally, the location of the coordinated zinc ion is noticeably different between these two structures. This contrast is consistent with a role for MYM-domains in protein-protein interactions.

ZNF198-like Proteins Maintain LSD1 in Insoluble Nuclear Fractions

LSD1 directly associates with chromatin to repress transcription (69,70,91), and the *Drosophila* MYM-domain ortholog, dWoc, is generally localized to chromatin (190). We performed cell fractionation experiments (141) in HeLa cells to determine whether the fractionation profile of ZNF198 and LSD1 were inter-dependent. After LSD1 RNAi, most of ZNF198 still fractionated with the nuclear pellet, even after high salt extraction (Figure 31A and 31B, compare supernatant and pellet). This is consistent with ZNF198 binding to chromatin independently of LCH. After RNAi of ZNF198-like proteins (MYM RNAi), LSD1 levels in the nuclear pellet but not the supernatant were decreased (Figure 31A, compare lanes 4 and 5, 7 and 8). This effect was exacerbated after high-salt extraction of the nuclear pellet (Figure 31B, compare lanes 10 and 11, 13 and 14). Importantly, the unrelated chromatin binding protein MCM7 was unaffected by RNAi (213). HDAC1 levels were also unaffected by LSD1 or MYM-domain RNAi. However, this was expected since only a fraction of cellular HDAC1 associates with LSD1 and ZNF198 (data not shown).

We next wanted to confirm the chromatin-association of ZNF198 and LSD1 by digesting nuclear pellets with mirococcal nuclease followed by EDTA extraction (141) (Figure 31C). Proteins that are not extracted by EDTA after digestion are either bound to nuclease-resistant chromatin or the nuclear matrix (214). As expected, most of MCM7, histone H3, and HDAC1 were extracted after nuclease digestion (Figure 31C, compare lanes 1 and 3). However, much of ZNF198 and LSD1 were resistant to extraction after nuclease digestion. Nevertheless, a significant fraction of LSD1 and ZNF198 was extracted, suggesting that these proteins are chromatin bound. These results are consistent with ZNF198-like proteins stabilizing the LCH complex onto chromatin.

To explore the mechanism by which ZNF198 may tether LSD1 into nuclear compartments, we performed *in situ* extractions of HeLa cells prior to fixation and immunohistochemistry with many different Myc-tagged fragments of ZNF198 (Figure 32B). This is a commonly used method to visualize proteins that are bound to insoluble regions of the nucleus such as chromatin (142). Both endogenous ZNF198 (data not shown) as well as wild-type Myc-ZNF198 (Figure 32A) were diffusely nuclear localized and resistant to extraction. No distinct foci were apparent. However, several fragments of Myc-ZNF198 were not resistant to extraction despite being expressed and nuclear localized (Figure 32A and 32B; see quantitation on right). Importantly, GFP-MCM7, a known chromosome binding protein, was still detected in these samples. Analysis of several fragments

showed that the proline/valine-rich containing region of ZNF198 was necessary for resistance to extraction, consistent with ZNF198 chromatin-association being independent of LCH binding.

Regulation of Gene Expression by ZNF198-like Proteins

Formation of an intact LSD1-CoREST-HDAC1/2 complex on chromatin is functionally important for its corepressor activity (126). Consistently, ZNF198 fused to GAL4 DNA binding domain represses LexA-VP16 *trans*-activation (data not shown). Next, we determined whether ZNF198 or other MYM-domain proteins functionally impact LSD1-responsive promoters. As expected, quantitative RT-PCR analysis demonstrated up-regulation of SCN3A and NCAM2 (REST-dependent repression;(140)), as well as E-Cadherin (CtBP-dependent repression; (130)) after RNAi of LSD1 (Figure 33A). E-Cadherin was de-repressed by MYM domain RNAi, but not SCN3A or NCAM2. Consistently, recombinant full-length ZNF198, as well as MYM-domains 8-10 (LCH binding region), prevented binding of [^{35}S]REST to CoREST *in vitro* (Figure 33B). Thus, ZNF198-like proteins are functionally required for efficient transcriptional repression at some LSD1-responsive promoters.

To identify a broader number of LSD1 target genes for functional assays, we performed microarray analysis on RNA obtained from HeLa Tet-on cells depleted of LSD1 by RNAi. We found 100 up-regulated (Figure 34A and data

not shown) and 48 down-regulated genes, including LSD1 (see AOF2; Figure 34B and data not shown). We chose ten that were up-regulated (Figure 34A, yellow highlight) for further validation. We selected the five genes that were up-regulated most after quantitative RT-PCR confirmation of these hits (Figure 34C, see asterisks). Though all of these mRNAs were further validated with several different LSD1 siRNAs (Figure 35A and data not shown), KRT17 and KRT80 were the only ones that were not also up-regulated by RNAi of an unrelated gene, Mps1 (Figure 35A). Interestingly, KRT17 contains a RE1-element (REST binding consensus) in its promoter (data not shown). REST protein was significantly enriched at this promoter, but not at GAPDH promoter, as determined by chromatin immunoprecipitation (ChIP) analysis with a REST antibody (Figure 35B). As expected, this gene was not subject to regulation by MYM-domain proteins (Figure 35C). The other keratin gene, KRT80, was not REST bound (data not shown), and is up-regulated after RNAi of ZNF198 alone or all three MYM-domain proteins (Figure 35C). Thus, we have identified two novel LSD1-target genes, one of which is regulated by REST (KRT17), the other by ZNF198 (KRT80).

Discussion

LSD1 and HDAC1 are enzymes that often function with CoREST to repress transcription at many promoters (69,70,126). Sumoylation commonly targets chromatin-associated proteins (5,9,10), and can promote (35,37,39) or disrupt (75) protein-protein interactions. Interestingly, ZNF198, a LCH binding protein, also interacts with SUMO non-covalently (35,66,68). We show that ZNF198-like proteins regulate some LSD1-responsive promoters (Figure 33A; Figure 35C). MYM-domain proteins maintain LSD1 on chromatin or within other insoluble nuclear compartments (Figure 31). Additionally, our *in vitro* data suggests ZNF198 may function to mediate SUMO-facilitated effects on complex formation. Further biochemical and functional analysis of ZNF198 and SUMO at a promoter will be important next steps in understanding the full functional relevance of our *in vitro* results.

ZNF198 and Complex Formation

We show that ZNF198 can uniquely bind the ternary LCH complex and not its individual subunits through two zinc-finger motifs (Figure 26; Figure 30). ZNF198 can also interact with SUMO2-HDAC1 (Figure 27B), again through multiple zinc-fingers that overlaps with its LCH binding region (Figure 30B).

One possible explanation for these data is that these interactions are multivalent, occurring only when LSD1 and HDAC1 are positioned properly by CoREST binding or SUMO-conjugation (see model in Figure 36).

Our data suggest ZNF198-like proteins do not regulate LCH complex stability (Figure 26E), though it is possible enough protein was not depleted by RNAi to see an effect. Alternatively, ZNF198 may stabilize only chromatin-bound LCH complex, consistent with our fractionation data (Figure 31B). Indeed, a large amount of LSD1 is present in soluble compartments of the cell (Figure 31A and 31B), and thus is clearly not subject to regulation by ZNF198. Importantly, ZNF198 was partially resistant to EDTA extraction after nuclease digestion. This could be from denaturation during the extraction protocol (see methods). Since dWoc coats pachytene chromosomes, we favor the hypothesis that ZNF198 is also mostly chromatin bound. ZNF198 does not stimulate LCH complex activity or SUMO2-HDAC1 activity on bulk histones (Figure 28C). However, it remains to be seen whether it can stimulate activity on nucleosomal substrates. This is not unlikely since the ternary LCH complex coordinates activity on nucleosomes but not bulk histones (126) (see model in Figure 36).

The Function of HDAC1 Sumoylation

Several studies have been published stating the functional importance of sumoylation in regulating HDAC1 function (72,206). For example, cells stably

expressing wild-type HDAC1 show cell cycle defects. This effect is not seen with a sumoylation deficient mutant (72). The same mutant HDAC1 shows lower deacetylase activity (72,206), though this data is disputed by Colombo et al (215). We show that HDAC1 sumoylation does not significantly stimulate activity on its own, but instead inhibits binding to CoREST. This is consistent with previous findings that the C-terminus of HDAC1 mediates its interactions with co-factors (90). Additionally, mutation of the HDAC1 SUMO-sites does not disrupt co-factor binding (72). It remains to be seen whether sumoylation disrupts binding to other co-factors of HDAC1, or just CoREST.

It is seemingly paradoxical that sumoylation should inhibit binding of HDAC1 to CoREST and be important for its function at the same time. ZNF198 binds efficiently to sumoylated HDAC1 *in vitro* and is important for LCH complex function *in vivo*. We considered several hypotheses based on these data (see model in Figure 36). One intriguing possibility is that the many MYM-domains of ZNF198 serve to stabilize the interaction of sumoylated HDAC1 with chromatin or the LCH complex. Alternatively, turnover of HDAC1 within the LCH complex may be required for efficient activity. Further experiments are needed to determine whether HDAC1 sumoylation is important for its activity *in vivo* in the context of ZNF198 and CoREST.

Conclusion

ZNF198-like proteins regulate some LSD1-responsive promoters, likely by regulating LCH association on chromatin. Moreover, ZNF198 association with chromatin likely occurs through regions outside its MYM-domains. Instead, the MYM-domains are protein-protein interaction motifs that are fundamentally different from Krüppel-like zinc-finger DNA binding domains. SUMO disrupts the interaction of HDAC1 with CoREST. ZNF198 interacts with the ternary LCH complex and with SUMO. ZNF198 could be a mediator for SUMO to regulate CoREST complex on chromatin. Thus, we present multiple mechanisms by which SUMO and ZNF198 could function to regulate LCH complex function *in vivo.*

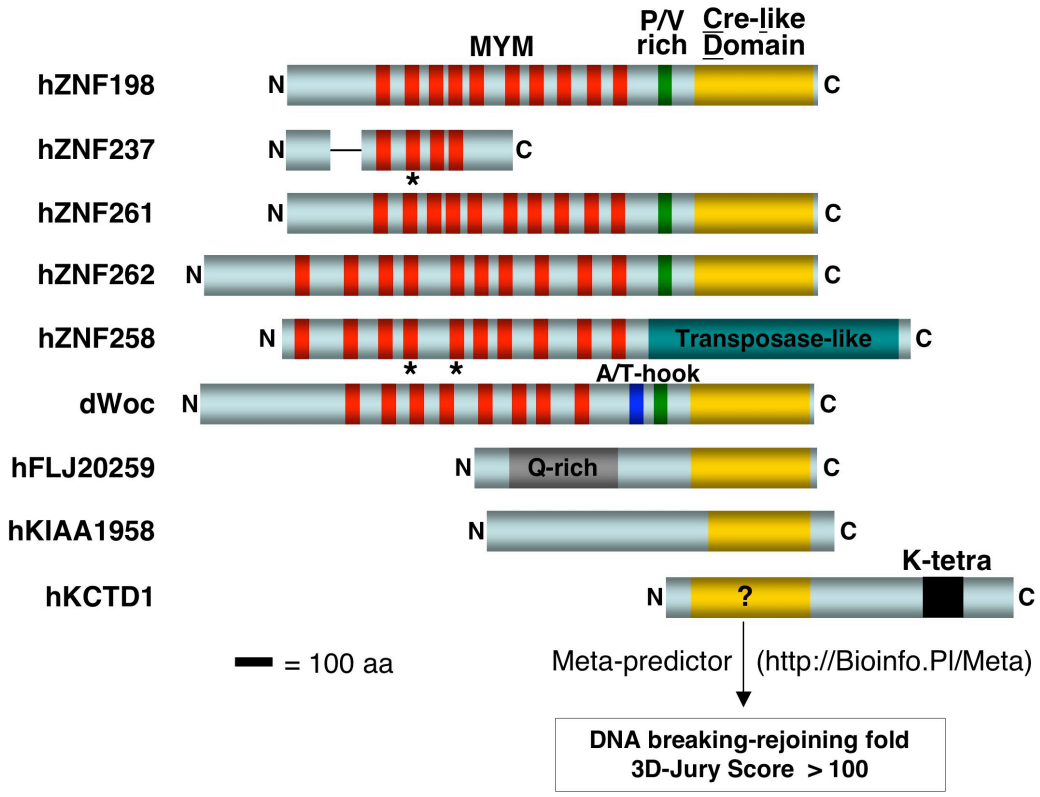

Figure 25. Domain Analysis of ZNF198-like Proteins
The color schemes from this illustration are used throughout the manuscript: MYM-domains (red), naturally occurring cysteine mutations (asterisk), proline/valine-rich (P/V-rich; green), Cre-like domain (CLD; gold), glutamine-rich (Q-rich; gray), potassium-tetramerization domain (K-tetra; black), and a transposase-like domain (teal). BLAST, ClustalW analysis, as well as the SMART database were used to map the domains. Scale bar indicates 100 amino acids. The Cre-like domain of KCTD1 was used for 3D-Jury analysis at the indicated website.

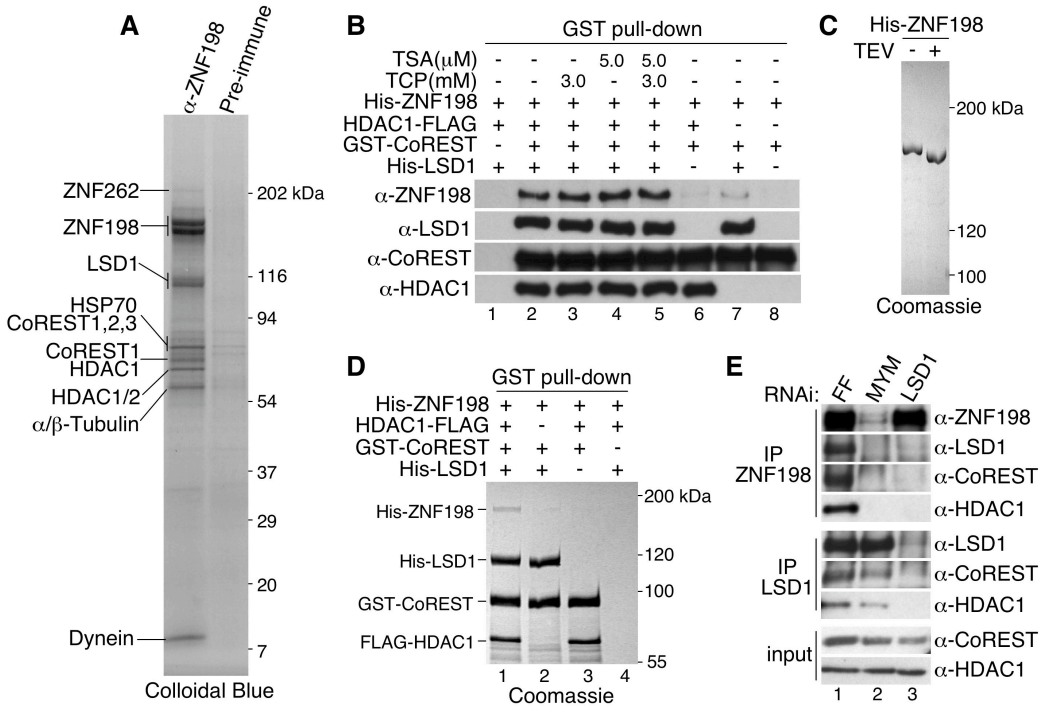

Figure 26. ZNF198 Binds the Intact LCH Complex

(A) IP of ZNF198 from HEK293 whole cell lysates. Bound proteins were stained with Colloidal blue and identified by mass spectrometry, as indicated.

(B) GST pull-down assays were performed by adding His-ZNF198 after pre-binding of the indicated LCH subunits. If TCP or TSA were added, they were present for the entire procedure. After extensive washing, bound proteins were detected by western blotting with the indicated antibodies.

(C) Recombinant His-ZNF198 was purified from Sf9 cells. TEV protease digestion confirmed the identity of the Coomassie stainable band.

(D) GST pull-downs were performed as in (B), except Coomassie staining was used to detect bound proteins, with bands and molecular weights labeled.

(E) IP of either ZNF198 (top panel) or LSD1 (middle panel) from HeLa Tet-on cells transfected with siRNA targeting either FF (FireFly Luciferase), MYM (ZNF198, ZNF261, and ZNF262 together), or LSD1. Bound proteins were detected by western blotting with the indicated antibodies.

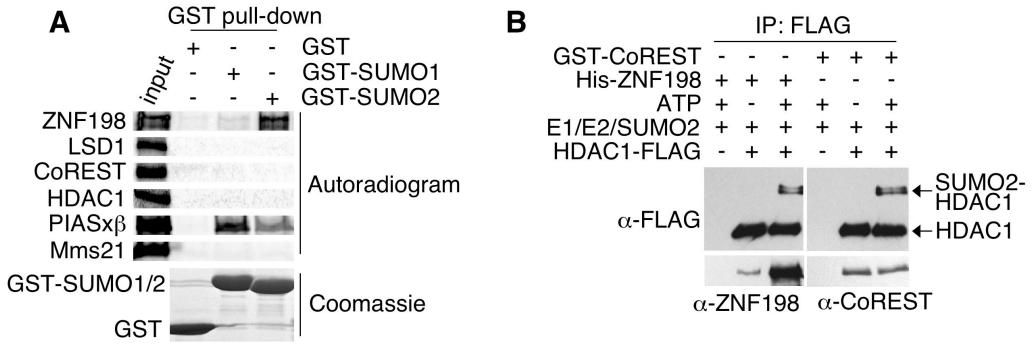

Figure 27. ZNF198 Interacts with SUMO2 and SUMO2-HDAC1
(A) The indicated proteins were *in vitro* transcribed and translated in the presence of ^{35}S-methionine and used as input in GST pull-down assays with the indicated GST-fusions (10 μg each). Bound proteins were detected by autoradiography (top panel) as well as by Coomassie blue staining (bottom panel) as labeled.
(B) IP of recombinant HDAC1-FLAG (1 μg each). SUMO reactions were performed as indicated on anti-FLAG M2 agarose, followed by washing and then addition of either His-ZNF198 or GST-CoREST. Bound proteins were detected by western blotting.

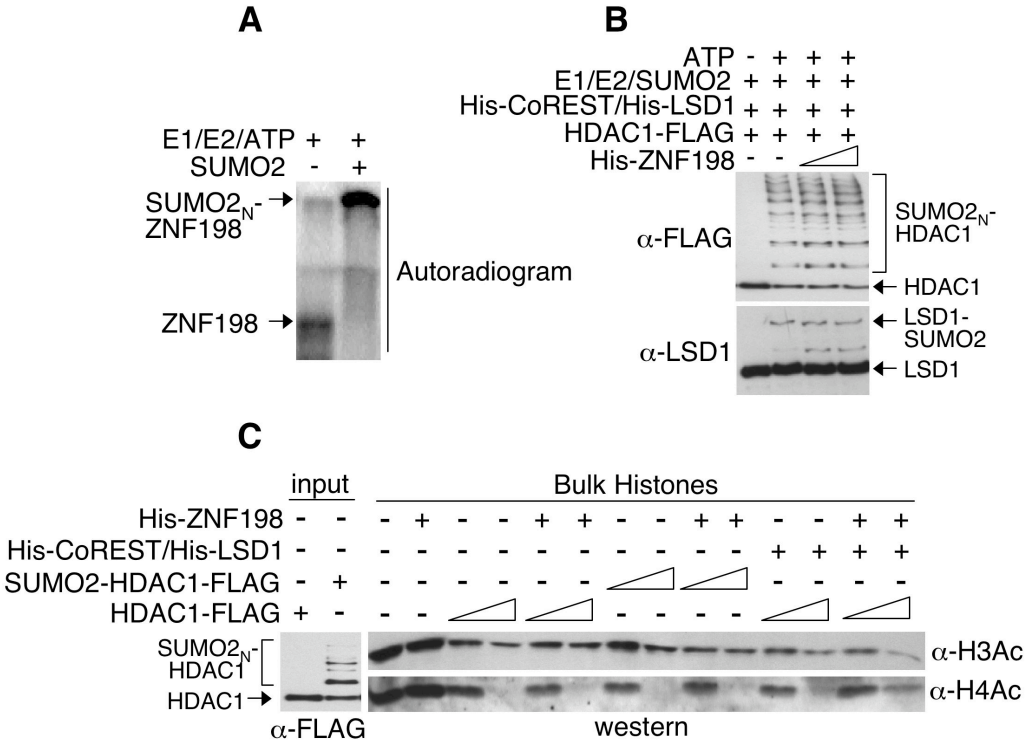

Figure 28. ZNF198 is not an E3 ligase and SUMO or ZNF198 do not Stimulate HDAC1 Activity on Bulk Histones

(A) ZNF198 that was *in vitro*-translated in the presence of [^{35}S]methionine was subjected to reactions containing all sumoylation machinery, +/- SUMO2. Sumoylation was detected by autoradiography after SDS-PAGE. The hypershifted band marked $SUMO2_N$-ZNF198 (top arrow) is in the well of the gel.

(B) LCH components were pre-incubated followed by addition of the sumoylation machinery, with or without ATP as indicated. Following SDS-PAGE, proteins were detected by western blotting. Modified and un-modified LSD1 and HDAC1 are labeled.

(C) *In vitro* deacetylation assays were performed comparing the activity of sumoylated or un-modified HDAC1 (see left panel, α-FLAG) . SUMO reactions were stopped with apyrase prior to the deacetylase reaction. When LSD1-CoREST or ZNF198 were added, they were pre-incubated with HDAC1 prior to substrate addition. Deacetylase activity was measured by western blotting towards acetylated Histone H3 (α-H3Ac) and acetylated Histone H4 (α-H3Ac).

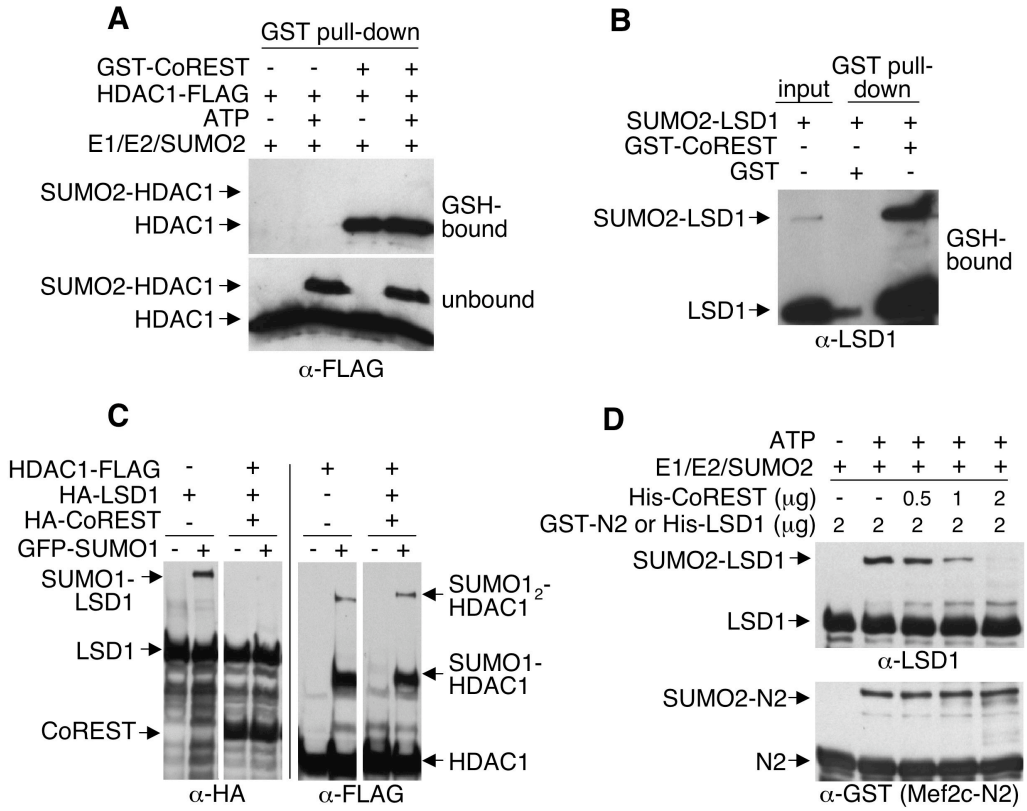

Figure 29. CoREST only Binds Un-Sumoylated HDAC1 and Inhibits LSD1 Sumoylation

(A) HDAC1-FLAG was treated with the indicated sumoylation machinery followed by addition to glutathione beads (GSH) that were pre-bound with GST-CoREST (+ or -). After washing, GSH-bound (top panel) and unbound (bottom panel) HDAC1-FLAG proteins were detected by western blotting. The location of sumoylated and unsumoylated HDAC1 are indicated by arrows.

(B) A GST pull-down assay with sumoylated LSD1 was performed similar to (A).

(C) The indicated tagged constructs were co-transfected for 24 hours in HeLa Tet-on cells. Cells were lysed in SDS sample buffer and subjected to SDS-PAGE followed by western blotting with either α-HA (left two panels) or α-FLAG (right two panels) antibodies. Detected bands are labeled for clarity (see arrows)

(D) LSD1 or GST-Mef2C-N2 (labeled N2 for simplicity) were subjected to sumoylation assays after pre-incubation on ice with increasing concentrations of His-CoREST. Sumoylated bands were detected by western blotting and labeled accordingly.

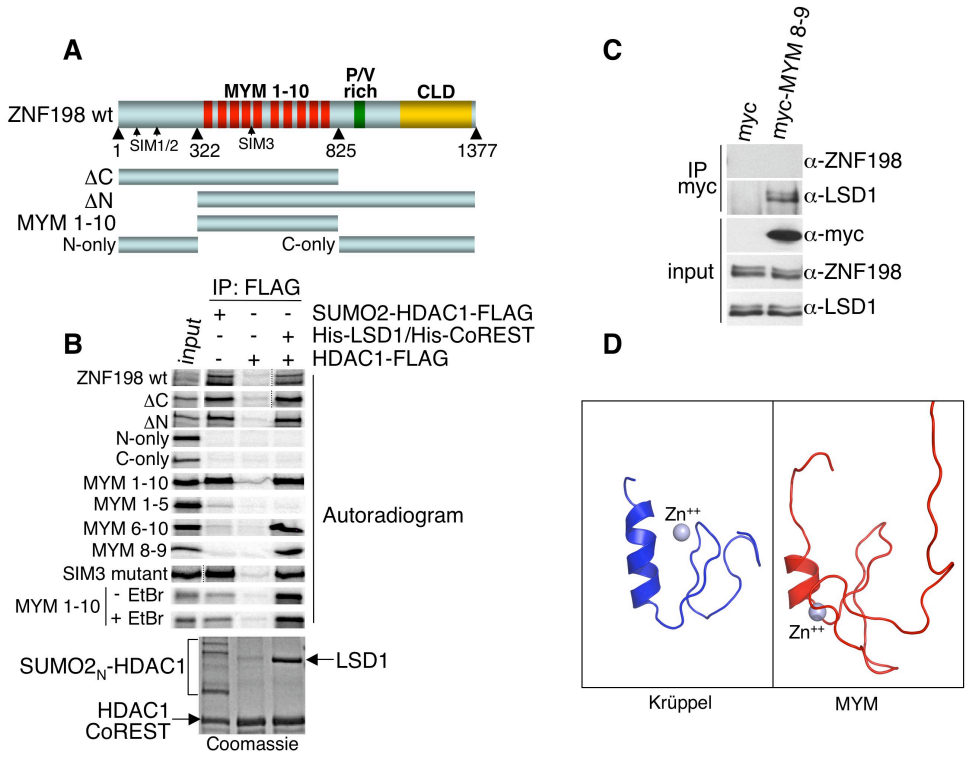

Figure 30. MYM-Domains Mediate Protein-Protein Interactions

(A) Illustration of ZNF198 fragments used in (B). Relevant amino acid numbers (large arrows), domain architecture, and the location of SIMs 1-3 (small arrows) are indicated. wt (wild-type).

(B) FLAG IPs were performed with the indicated proteins similar to Figure 2B, except [^{35}S]fragments of Myc-ZNF198 were used as input. Bound proteins were visualized by autoradiography (top panel) or Coomassie blue staining (bottom panel). MYM-domains are numbered according to their order in the primary amino acid structure in (A). When ethidium bromide was added (100ng/µl; EtBr), it was present for the whole procedure. SIM3 refers to a sumo interacting motif mutant (V483A/L484A/V485A) in a fragment containing amino acids 1-923.

(C) HeLa Tet-on cells were transfected with the indicated myc-tagged constructs for 24 hours followed by IP with anti-myc Protein A-sepharose. Bound proteins (top panels) as well as input (bottom panels) were detected by western blotting.

(D) Ribbon drawings of a Krüppel-like DNA binding zinc-finger from BKLF (PDB ID: 1P7A; blue, left panel) and a MYM-domain from ZNF237 that is nearly identical to the first MYM-domain of ZNF198 (PDB ID: 2DAS; red, right panel). Overlays of these structures were juxtaposed in the same orientation for clear viewing. Zinc ions are shown as light blue spheres.

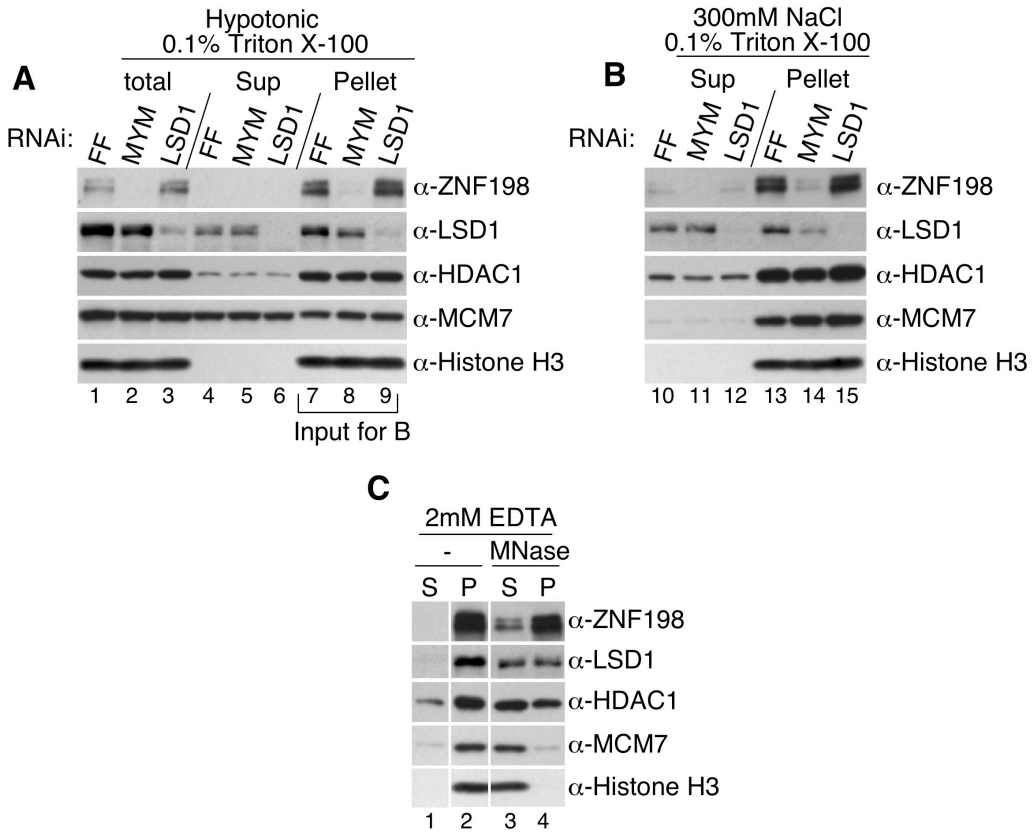

Figure 31. ZNF198-like Proteins Control the Chromatin Association of LSD1

(A & B) After RNAi of the indicated proteins, nuclear pellets were generated by subjecting HeLa Tet-on cells to hypotonic lysis followed by centrifugation (A, see top). Next, nuclear pellets (lanes 7-9) were subjected to extraction with high salt buffer as indicated in (B, see top). Normalized samples from each step were subjected to SDS-PAGE followed by western blotting towards the indicated proteins.

(C) Control samples were subjected to fractionation as in (A), then pellets (ie, lane 7 in (A)) were digested with micrococcal nuclease followed by extraction with 2mM EDTA. Supernatants (S) and pellets (P) were analyzed as in (A) and (B).

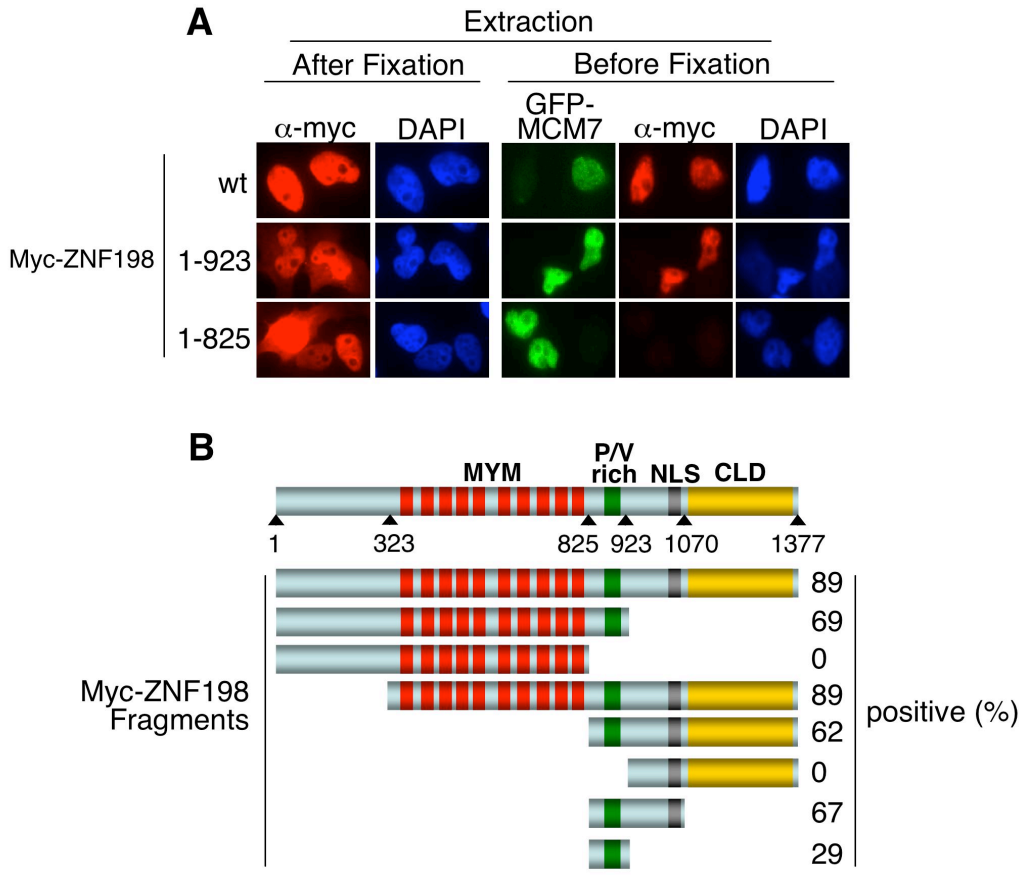

Figure 32. *In Situ* Extraction and Immunohistochemistry of ZNF198
(A) HeLa Tet-on cells were transfected with the indicated Myc-tagged ZNF198 constructs along with GFP-MCM7. Duplicate transfections were either extracted before fixation (right panels) or after fixation (left panels). Samples, which were treated identically after fixation, were stained with anti-myc antibody (red) and DAPI (blue). GFP is shown in green.
(B) Illustrations of the fragments used in (A) as well as the percent (%) of GFP^+ cells that were also myc^+ after pre-extraction (positive). Cells were considered positive if above background staining was observed. Greater than 30 GFP^+ cells from 10 random fields were counted per fragment. Important amino acid residues (arrows) as well as domains (colored bars) are labeled on top.

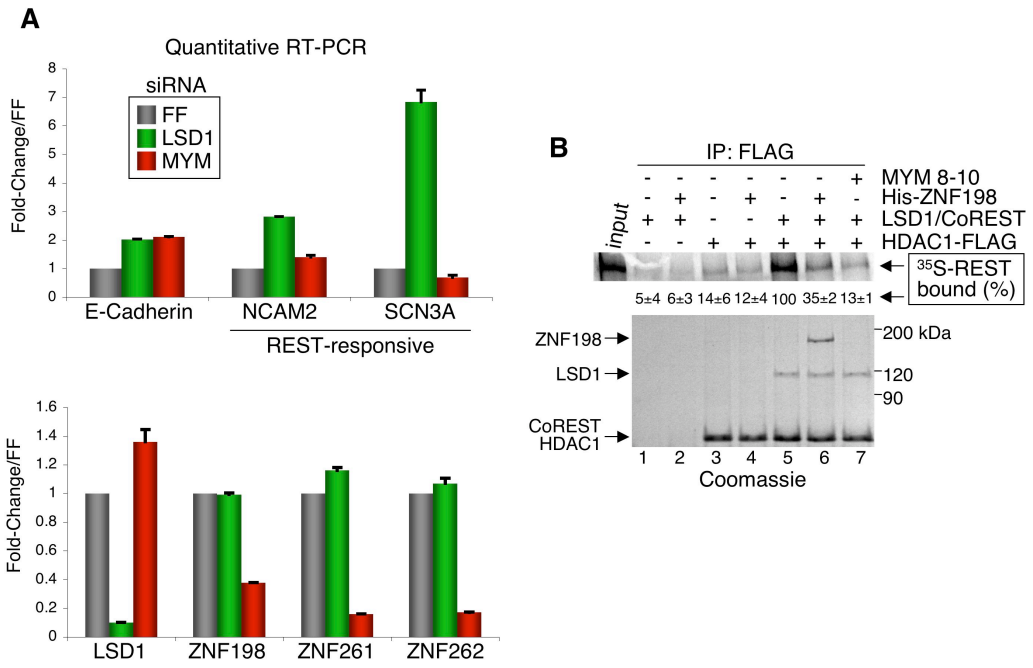

Figure 33. ZNF198-like Proteins Regulate Specific Promoters

(A) U2OS cells were transfected with the indicated siRNA for 3 days, followed by reverse transcription (RT) with random hexamers and quantitative PCR using the indicated primer sets. Cycling-time (Ct) values were normalized to the housekeeping gene cyclophilin B here and in all subsequent experiments. Each PCR reaction was performed in triplicate, and error bars indicate the standard deviation between two separate experiments.

(B) HDAC1-FLAG and His-CoREST/His-LSD1 were pre-bound to anti-FLAG M2 agarose in the combinations indicated. After washing, [^{35}S]REST was added to each binding reaction, +/- the indicated ZNF198 proteins. Bound REST (box and arrows) was detected by autoradiography (top panel). The numbers presented represent the average (± the standard deviation) of two experiments where signals were normalized to lane 5. LCH as well as His-ZNF198 were visualized by Coomassie blue staining and labeled for clarity (arrows, left).

A

Upregulated after LSD1 RNAi = 100 genes

Accession	Symbol	A	B
NM_000422	KRT17	2.02	2.78
NM_182507	KRT80	2.30	3.52
NM_003979	GPCR5A	2.11	3.23
NM_000416	IFNGR1	2.53	2.65
XM_378301		2.43	2.29
NM_144563	RPIA	2.20	2.13
NM_015986	CRLF3	2.40	2.98
NM_014874	MFN2	2.25	2.05
NM_198212	CAV2	2.17	2.89
NM_002284	KRTHB6	2.61	3.00
NM_004406	DMBT1	2.35	2.82
NM_004850	ROCK2	2.12	2.60
NM_002859	PXN	2.23	2.31
NM_024306	FA2H	2.01	4.81
NM_021259	TMEM8	2.42	2.91
NM_020715	PLEKHH1	2.02	2.29
NM_002999	SDC4	2.36	2.24
NM_003930	SCAP2	2.01	2.42
NM_005892	FMNL1	2.43	4.97
NM_004898	CLOCK	2.15	2.26
NM_181552	CUTL1	2.04	2.40
NM_003246	THBS1	5.68	9.12
NM_020925	KIAA1573	2.01	2.04
NM_006033	LIPG	3.14	2.86
NM_021111	RECK	2.75	2.09
NM_001010914	UNQ2430	2.39	4.77
NM_031458	PARP9	2.41	2.03
NM_014786	ARHGEF17	2.17	2.95
NM_020807	ZNF319	2.30	5.63
NM_016205	PDGFC	2.10	2.75

B

Downregulated after LSD1 RNAi = 48 genes

Accession	Symbol	A	B
NM_002647	PIK3C3	2.35	2.64
NM_207368	LOC348262	2.16	3.25
NM_005792	MPHOSPH6	2.05	3.30
AF095719	CPA4	2.23	2.64
NM_013239	PR48	2.02	2.67
NM_005443	PAPSS1	2.06	2.07
NM_015013	† AOF2	6.27	4.65
NM_173548	ZNF584	2.98	2.57
NM_020408	C6orf149	3.28	6.33
AF151100	PHLDA3	2.33	3.34
NM_023007	FLJ12517	2.28	3.35
NM_004490	GRB14	3.13	2.01
NM_206837	NOR1	2.34	2.78
AK000451	FLJ20444	2.00	4.53
NM_024165	PHF1	2.22	2.36
NM_013310	C2orf27	2.28	8.84
AL163303		2.84	2.16
NM_032869	CML66	2.61	4.15
NM_030639	BHLHB9	2.14	2.07
NM_182715	SYPL	3.90	3.40
NM_138393	C19orf32	5.19	2.11
NM_025125	C10orf57	16.56	3.66
NM_019092	KIAA1164	2.26	2.18
NM_152556	FLJ31818	2.09	2.45
AB007875	KIAA0415	3.72	3.77
BC050338		5.88	16.27
BC040618	BCL9L	2.11	497.33
NM_004476	FOLH1	7.41	3.98
XM_378360		3.05	3.74
AL132838		2.02	2.86

† - AOF2 is LSD1

C

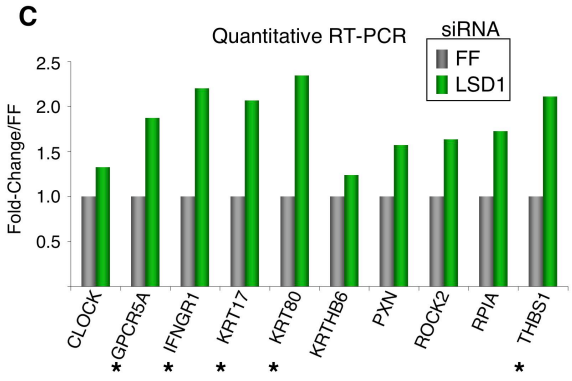

Figure 34. Identification of LSD1 Target Genes by Microarray Analysis
(A and B) Duplicate RNA samples from HeLa Tet-on cells transfected with FF or LSD1 siRNA for 30 hours were analyzed by the UT-Southwestern Microarray Core facility using Illumina BeadChips (human WG-6). A 2-fold change in either direction with p-values < 0.01 (for the numerator) was used as the threshold. 30 genes for each category are shown (ranked by p-value). Tables include gene accession numbers, gene symbols, as well as the fold-change for each duplicate (samples A and B). Yellow highlighted genes were validated in (C). AOF2 (LSD1) is marked by a cross in (B).
(C) RNA from the microarray analysis was subjected to quantitative RT-PCR with the indicated gene specific primers. Values are the average of triplicate PCR experiments normalized to cyclophilin B. Genes marked with an asterisk were selected for further validation.

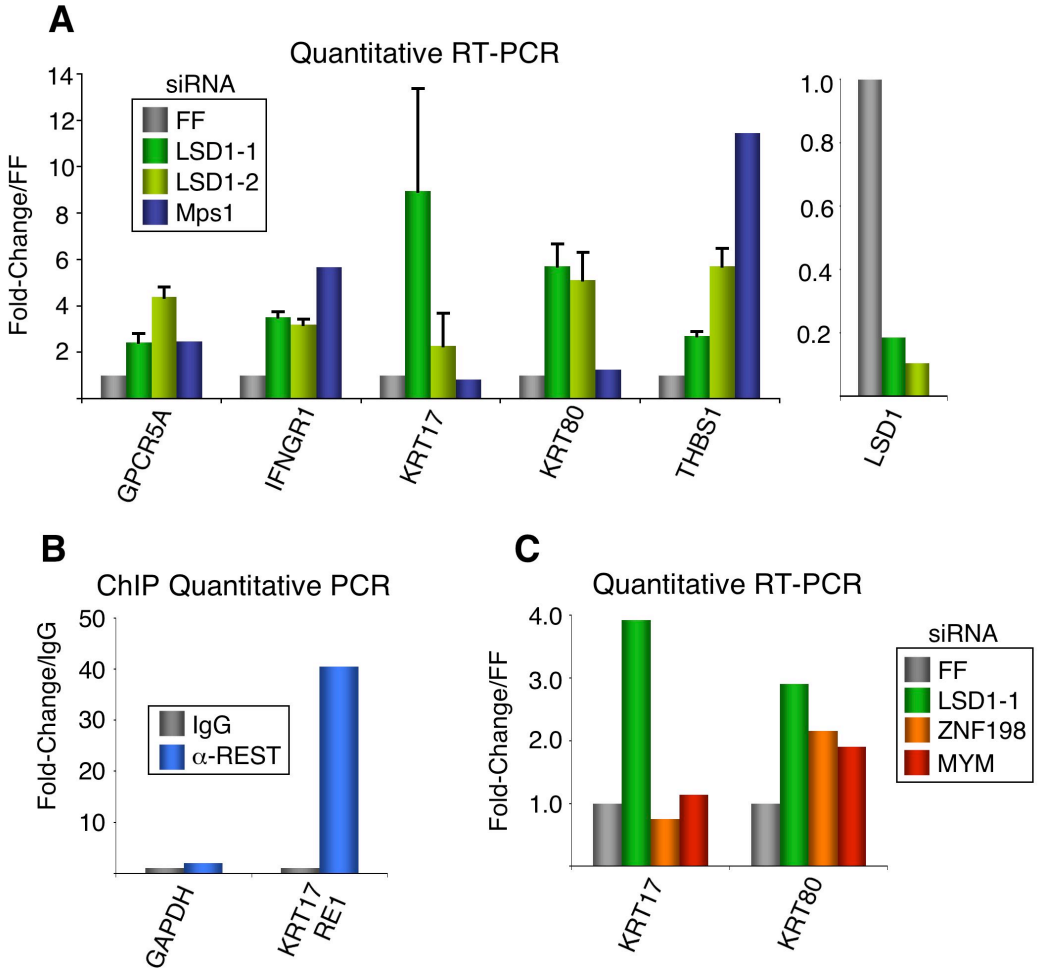

Figure 35. KRT17 is a New RE1-element Containing Gene Regulated by LSD1

(A) HeLa Tet-on cells were subjected to 2 rounds of RNAi with the indicated siRNAs followed by quantitative RT-PCR with the indicated gene specific primers. All PCR reactions were performed in triplicate, and error bars indicate the standard deviation for 2 separate experiments.

(B) Chromatin immunoprecipitation with non-specific rabbit IgG (gray bars) or anti-REST antibody (blue bars) was performed from HeLa Tet-on cells. Bound DNA was subjected to quantitative PCR with primers designed towards the promoter of GAPDH (control) or the KRT17 gene promoter RE1 element. Each PCR was performed in triplicate.

(C) Quantitative RT-PCR was performed as in (A), except only for one round of RNAi.

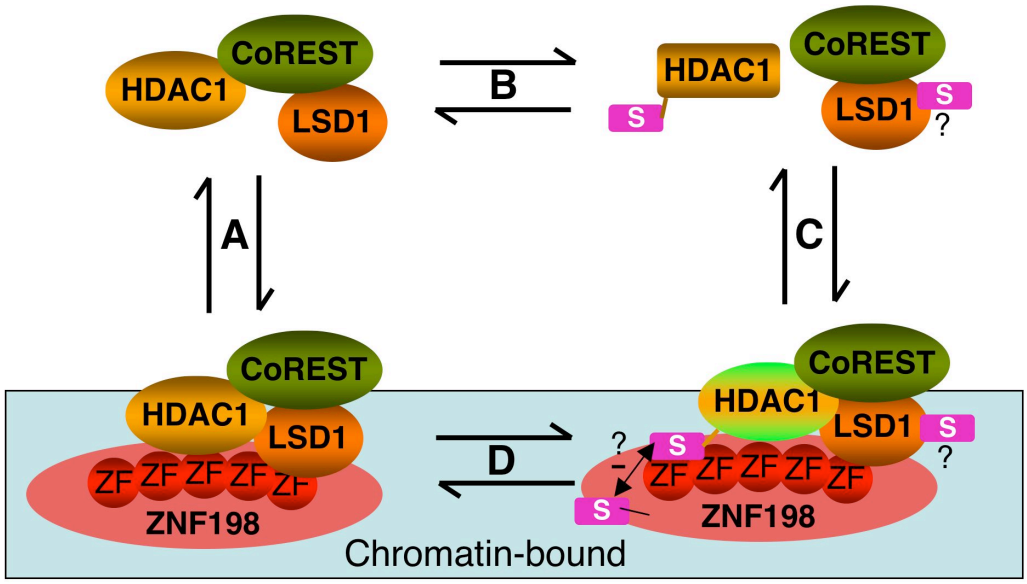

Figure 36: SUMO and ZNF198 Cooperate to Regulate the CoREST Complex

(A) CoREST, LSD1, and HDAC1 form a ternary complex *in vitro*. HDAC1 and LSD1 have been shown to cooperate in their activities towards nucleosomes but not free histones. ZNF198 could stabilize their binding to chromatin.

(B) Sumoylation of HDAC1 disrupts its binding to CoREST. Sumoylation of LSD1 may be inhibited if it is already bound by CoREST.

(C and D) HDAC1 sumoylation is important for its activity, yet it cannot bind to an activator CoREST. The presence of ZNF198 and its many zinc-fingers could create a scaffold where HDAC1 can be stimulated by sumoylation and remain bound in its proper environment for optimal activity (green HDAC1). Alternatively, sumoylation of ZNF198 could act to intra-molecularly inhibit non-covalent binding to SUMO-HDAC1 and destabilize the entire complex on chromatin.

Chapter VI: Discussion and Future Directions

Part A: SUMO Screen and Mef2C Sumoylation

Systematic Identification and Analysis of SUMO Substrates

We have used *in vitro* expression cloning to identify and analyze sumoylated proteins in a systematic manner (22). Before and after publishing the manuscript for our IVEC screen, several proteomic publications identifying SUMO substrates were published (53-58). All of these used tagged-SUMO pull-downs followed by mass spectrometry to identify substrates. Though our approach to identify substrates was different, the conclusions from all these studies was similar: sumoylation targets proteins many cellular processes, especially nuclear processes such as transcription, DNA repair, replication, and RNA biogenesis. Moreover, a tabulation of the current data suggests a clustering of substrates within the same macro-molecular complexes. While SUMO does generally seem to be associated with transcriptional repression (see Mef2C), this is not a rigid rule. Moreover, the mechanism by which SUMO functions in repression is not always the same. SUMO has also been proposed to function generally in protein subcellular localization. We showed that SUMO generally does not promote steady-state changes in the localization of its substrates.

However, since we did not mutate the SUMO sites in all these proteins, we cannot be sure that sumoylation is not required for their localization. Overall, SUMO appears to act through multiple, context-dependent mechanisms.

Regulation of the Transcriptional Activity of MEF2

We also identified Mef2C as a SUMO-substrate in our screen, and extensively studied this modification in cells (22,152). Transcription activity of myogenic bHLH and MEF2 proteins is tightly regulated during muscle differentiation. Multiple pathways exist to ensure the repression of these transcription factors in dividing myoblasts (216-218). For example, Cdk4/Cyclin D represses the activity of MEF2 proteins through blocking their interactions with the GRIP1 co-activator (219), although it is unclear whether Cdk4/Cyclin D phosphorylates MEF2 proteins directly. In addition, another cyclin-dependent kinase, Cdk5, phosphorylates MEF2 proteins and inhibits their transcriptional activity in neurons (220).

Consistent with recent findings by Gregoire and Yang (164), we observed that sumoylation-deficient MEF2C has higher myogenic activity than wild-type (Figure 12*B* and 12*C*), suggesting that sumoylation might be another important mechanism to actively repress the transcriptional activity of MEF2 proteins in dividing myocytes. Surprisingly, only a very small population of MEF2C is modified by SUMO in C2C12 cells (data not shown). It is unclear how

sumoylation of this small population of MEF2C effectively suppresses the activity of cellular MEF2C. On the other hand, this appears to be a recurring theme in the sumoylation of transcriptional factors. One intriguing possibility is that transient sumoylation of these transcriptional factors recruits transcriptional repressors that covalently modify the chromatin at the transcriptional loci, which in turn establishes a relatively long-lived chromatin state that is not permissible for transcription. Alternatively, a "molecular memory" model has been proposed to explain this phenomenon. In this model, a protein molecule that has experienced a sumoylation/desumoylation cycle is proposed to be functionally distinct from one that has never experienced sumoylation (221). Finally, we and others have shown SUMO isopeptidases actively keep steady state levels of SUMO low. Thus, SUMO turnover could be occurring on a large subset of Mef2C, while only a small portion of the protein is sumoylated at any one time.

Part B: CoREST-LSD1 Structure

Nucleosome Recognition

We have solved the structure of LSD1 bound to a fragment of CoREST that is sufficient to stimulate LSD1 activity on nucleosomes. The overall structure of LSD1-CoREST consists of two chromatin binding motifs (LSD1 AOD and CoREST SANT2) that are attached, but also significantly separated, by

an alpha-helical stalk (coiled-coil between LSD1-Iα1 and CoREST-Lα2). The distance between the LSD1 active site and CoREST SANT2 is approximately 100 angstroms, consistent with the dimensions of the mononucleosome (76). We have shown that SANT2 mediated stimulation of LSD1 activity on nucleosomes, but not bulk histones, depends on its DNA binding activity. Others have shown that CoREST or LSD1 only bind to mononucleosomes when they are in complex (70). While we cannot be sure of the exact mode of CoREST-LSD1 binding to nucleosomes, we favor a model similar to that shown in Figure 24.

Many other chromatin-associated complexes contain multiple chromatin recognition motifs such as SANT (87,119). Our structure gives us a glimpse of how multiple such domains might come together to recognize nucleosomes. We propose that chromatin-associated complexes in general recognize chromatin through multiple, weak interactions that can function together from a distance. Such multivalent binding, as opposed to single high affinity interactions, would allow for tight interactions that are highly amenable to dynamic changes, a characteristic needed for such dynamic substrates as nucleosomes.

CoREST-LSD1 Interaction

The α-helices that mediate binding between LSD1 and CoREST are likely to be unstable on their own, explaining the co-dependence of these factors for *in vivo* stability ((69) and data not shown). Whether other co-factors can interact

with LSD1 through its α-helical stalk is unknown (for example, androgen receptor could take the place of CoREST and change the specificity of LSD1 from a K4 to a K9 demethylase). Alternatively, the coiled-coil that is formed between CoREST and LSD1 may create a surface for additional interacting partners (ZNF198, CtBP1/2, etc...)

Specificity of LSD1 and Future Directions

Schüle and colleagues argue that LSD1 is a bona fide H3K9me1/2 demethylase (95,96). This is supported by the fact that LSD1 activity is required for activation of many nuclear receptor responsive genes (95,96,140), and because H3K9 demethylase activity is observed in IP/*in vitro* demethylase assays (96) (where co-IP of TAP-AR and TAP-LSD1 in the presence of androgen agonists was performed). Additionally, in *S. pombe*, LSD1 orthologs show weak activity towards only H3K9me, not H3K4me (222,223). However, several Jumonji-domain H3K9 demethylases are recruited to androgen receptors along with LSD1 (94,95). Additionally, only H3K4 demethylation has been demonstrated with recombinant LSD1. Our lab has recently solved the structure of a propargylamine-derivatized H3 peptide covalently attached to the active site FAD of LSD1 (224). The peptide forms three gamma-turns in order to place itself properly for catalysis, an unusual conformation compared to most H3-tail binding proteins. The energetic restraints of this conformation could explain the relatively

weak activity of LSD1 *in vitro* (K_m ~ 20-200 μM, K_{cat} = ~2 min^{-1}). Importantly, more than three residues N-terminal to the substrate methyllysine cannot be accommodated by the LSD1 active site, explaining its H3K4 specificity. Although unlikely, association of LSD1 with another factor (e.g., AR), could convert LSD1 into a K9 demethylase. However, there is no precedent for this. It remains to be seen whether LSD1 is a bona fide H3K9 demethylase, or whether it is just required for the activity of other demethylases.

Finally, current projects in our lab are trying to understand several aspects of the CoREST complex from a structural view. These projects include solving the structure of full-length CoREST (including SANT1) with LSD1 as well as with the addition of HDAC1. This would help us to better understand how CoREST coordinates HDAC1 and LSD1 activity. The major goal is solving the structure of CoREST-LSD1 with the mononucleosome, which would give us a more complete picture of how chromatin-associated complexes recognize chromatin.

Part C: Regulation of LCH Complex by Sumoylation

Discovery of a DNA Breaking-Rejoining Fold in ZNF198-like Proteins

ZNF198, ZNF261, and ZNF262 contain ten tandem MYM-type zinc-fingers and a proline/valine-rich region. These proteins, and several otherwise

unrelated proteins, contain a conserved domain with the predicted fold of Cre recombinase (Cre-like Domain, CLD; Figure 25). This domain is not required for efficient interactions with SUMO, the CoREST complex, or chromatin in general (Figure 29B; Figure 32A). Recombinantly purified CLD from either ZNF198 or KCTD1 lacked topoisomerase activity *in vitro* (data not shown). Possible explanations include: usage of the wrong substrate, the need for associated factors, or a lack of critical catalytic residues. DNA breaking-rejoining folds are able to recognize bent DNA or secondary structures, such as a holliday junction (225). Interestingly, the binding of dWoc to telomeric regions but not other chromosomal regions can be ablated by a single point mutation in its CLD (190). It is not known whether the LCH complex plays a role in telomere biology in *Drosophila*. It would be interesting to knock-down all three ZNF198-like proteins in mammalians to look for telomeric phenotypes.

Regulation of Gene Expression by ZNF198-like Proteins

Most LSD1 is associated with CoREST, highlighted by the dependence of LSD1 on CoREST for its stability (69). However, it is not known whether all CoREST-bound LSD1 functions in co-repression, or whether CoREST-LSD1 always function together. However, LSD1, but not CoREST, was identified as a component of the H3K4 methyltransferase MLL complex (139). ZNF198 and ZNF261, and likely ZNF262, associate with a large population of LSD1 and

CoREST in the cell ((69,70), data not shown, and Figure 31). However, at REST-responsive promoters we did not see significant changes in LSD1-target gene expression when we removed LSD1 associated MYM-domain proteins (Figure 33A; Figure 35C). This can readily be explained by the competition between ZNF198 and REST for CoREST binding (Figure 33B). It is doubtful that this is an important regulatory function for MYM-domain proteins, since concomitant repression of REST-responsive genes after MYM-domain RNAi was not observed (Figure 33A). Moreover, MYM-domain RNAi did not significantly rescue LSD1 RNAi phenotypes at these promoters (data not shown). However, our data clearly shows that removing ZNF198-like proteins can result in decreases in total levels of LSD1 on chromatin or other insoluble components of the nucleus (Figure 31A). Moreover, GAL4-fused ZNF198 can repress transcription, making it unlikely that ZNF198-bound LCH is inactive or acting in transcriptional activation (data not shown). Finally, since ZNF198 binds uniquely to the three protein complex (Figure 26) and to SUMO (Figure 27), it is likely that this binding could serve important regulatory functions in the context of chromatin and SUMO.

Further open-ended approaches, as well as screening of known LSD1 targets that are not REST-responsive, will be important to identify more targets of ZNF198-like proteins. In addition, it will be critical to develop ChIP assays with both LSD1 and ZNF198 antibodies (which have been unsuccessful for unknown

reasons at this point), to validate direct targets and perform key experiments to determine whether ZNF198-like proteins affect LCH formation on chromatin at specific promoters. This would also allow us to observe sumoylation in the context of the promoter. Especially interesting would be rescue experiments with ZNF198 separation of function mutations or truncations (ie, LCH binding, but no SUMO binding). We have made individual point mutations in the zinc-fingers of ZNF198, but these had no effect on its binding activity. However, there are examples of single cysteine to serine mutations that are not sufficient to ablate the function of other zinc-fingers (211), and it is also possible that other zinc-fingers within ZNF198 are functionally redundant.

Chromatin Binding by ZNF198 and LCH

The LCH binds to chromatin through several chromatin recognition domains. Likewise, with the combination of immunohistochemistry and fractionation experiments we demonstrate that ZNF198 is also likely a chromatin-binding protein (Figure 32; Figure 31C). This association requires the proline/valine-rich region of ZNF198 (Figure 32), a region previously shown to mediate homo-dimerization in co-transfection assays. We have not determined whether the over-expressed protein in our assays is dimerizing with the endogenous protein.

Also, not all of ZNF198 (or LSD1) could be extracted from the nuclear pellet with micrococcal nuclease digestion. ZNF198 may associate with nuclease-resistant regions of the genome, or be denatured during the EDTA extraction. Alternatively, ZNF198 could associate with both chromatin and the nuclear matrix. Some components of the SWI/SNF-chromatin remodeling complex show this characteristic (226). ZNF198 was recently proposed to localize to PML nuclear bodies (73), though we observed no staining consistent with this in HeLa (Figure 32A), U2OS, or HEK293 cells (data not shown). It will be important to sort out exactly which insoluble components LSD1 and ZNF198 associate with (chromatin or matrix) by extracting chromatin with buffers other than EDTA.

SUMOylation and complex formation

We show LSD1-CoREST complex formation antagonizes LSD1 sumoylation (Figure 29). However, we have not been able to map the SUMO attachment sites by mutagenesis or mass spectrometry (with micrograms of SUMO-LSD1). Identifying these sites will be crucial for further understanding LSD1-SUMO function *in vivo*.

As for HDAC1 sumoylation and ZNF198 SUMO binding, another possibility (as opposed to the model in Figure 36) is that SUMO-HDAC1 binding by ZNF198 is not specific to HDAC1, but instead applies to many sumoylated

substrates. In this case, other sumoylated substrates may be able to recruit ZNF198-LCH complex to chromatin (such as sumoylated CtBP), or vice versa. Indeed, previous studies showed that GAL4-SUMO2, which efficiently represses transcription, could recruit histone H3K4 demethylase activity (66). LSD1 loss-of-function did not diminish GAL4-SUMO repressor activity (66). However, SUMO alone binds many repressors besides ZNF198 (35,37,66,68), so this could explain why LSD1 is not necessary for repression in this artificial situation. Finally, the SUMO binding by ZNF198 could allow for auto-sumoylation to compete for LCH, since SUMO binding and LCH binding sites overlap. This would be consistent with our findings that sumoylation and LCH complex binding are generally antagonistic, and would allow for an efficient regulatory mechanism to disrupt the complex when needed. Determining the role of SUMO in regulating the CoREST complex *in vivo* will help answer these questions.

Comparison of ZNF198 and REST

ZNF198 has been compared to REST because it has many zinc-fingers and associates with CoREST(71). Additionally, we show that ZNF198 competes with REST for CoREST binding, suggesting they bind to overlapping interfaces on CoREST. Finally, ZNF198-like proteins likely regulate LCH association with chromatin. However, there are also many differences between ZNF198 and REST. The MYM-domains are fundamentally different from REST zinc-fingers

(Figure 30B and 30D), lacking an extended alpha-helix for binding the major-groove in DNA. Moreover, ZNF198 association with chromatin likely occurs through its proline/valine-rich region (Figure 32). REST is not known to only interact with the three protein complex or with SUMO. Finally, Coomassie stainable amount of ZNF198-like proteins, but not REST, have been identified in IPs of CoREST, HDAC1/2, or LSD1, suggesting MYM-domain proteins may have a more general function in regulating LCH complex function. Still, we cannot be sure that ZNF198 is not functioning similar to REST, targeting the LCH to specific promoters. Either way, ZNF198 could also be a mediator for SUMO to regulate LCH on chromatin. Thus, we present multiple mechanisms by which SUMO and ZNF198 could function to regulate LCH function *in vivo.*

Overall conclusions

After four years of thesis laboratory work I have performed an *in vitro* screen for SUMO substrates, showing that most sumoylated proteins have nuclear functions and are often clustered into single macro-molecular complexes. Identifying LSD1 in our screen led me into the field of chromatin biology, where I contributed significantly to understanding how chromatin-associated complexes function to recognize chromatin. In addition, I have characterized the binding interactions of ZNF198 with LSD1-CoREST-HDAC1 complex and SUMO,

showing that the zinc-fingers in this protein are protein-protein interaction modules that likely mediate multivalent binding interactions. Additionally, I have shown that SUMO and LCH complex formation are generally antagonistic. These *in vitro* studies are an important start to understanding how ZNF198 and SUMO may come together to regulate LCH complexes.

Bibliography

1. Hershko, A., and Ciechanover, A. (1998) *Annu. Rev. Biochem.* **67**, 425-479
2. Pickart, C. M. (2001) *Annu. Rev. Biochem.* **70**, 503-533
3. Schwartz, D. C., and Hochstrasser, M. (2003) *Trends Biochem. Sci.* **28**(6), 321-328
4. Hochstrasser, M. (2000) *Science (New York, N.Y* **289**(5479), 563-564
5. Johnson, E. S. (2004) *Annual review of biochemistry* **73**, 355-382
6. Matunis, M. J., Coutavas, E., and Blobel, G. (1996) *J. Cell Biol.* **135**(6 Pt 1), 1457-1470
7. Sampson, D. A., Wang, M., and Matunis, M. J. (2001) *The Journal of biological chemistry* **276**(24), 21664-21669
8. Hilgarth, R. S., Murphy, L. A., Skaggs, H. S., Wilkerson, D. C., Xing, H., and Sarge, K. D. (2004) *J. Biol. Chem.*
9. Muller, S., Ledl, A., and Schmidt, D. (2004) *Oncogene* **23**(11), 1998-2008
10. Seeler, J. S., and Dejean, A. (2003) *Nat. Rev. Mol. Cell Biol.* **4**(9), 690-699
11. Melchior, F., Schergaut, M., and Pichler, A. (2003) *Trends Biochem. Sci.* **28**(11), 612-618
12. Potts, P. R., and Yu, H. (2005) *Molecular and cellular biology* **25**(16), 7021-7032
13. Kotaja, N., Karvonen, U., Janne, O. A., and Palvimo, J. J. (2002) *Mol. Cell Biol.* **22**(14), 5222-5234
14. Johnson, E. S., and Gupta, A. A. (2001) *Cell* **106**(6), 735-744
15. Pickart, C. M., and Eddins, M. J. (2004) *Biochimica et biophysica acta* **1695**(1-3), 55-72
16. Zhao, X., and Blobel, G. (2005) *Proceedings of the National Academy of Sciences of the United States of America* **102**(13), 4777-4782
17. Pichler, A., Gast, A., Seeler, J. S., Dejean, A., and Melchior, F. (2002) *Cell* **108**(1), 109-120
18. Kagey, M. H., Melhuish, T. A., and Wotton, D. (2003) *Cell* **113**(1), 127-137
19. Mahajan, R., Delphin, C., Guan, T., Gerace, L., and Melchior, F. (1997) *Cell* **88**(1), 97-107
20. Joseph, J., Tan, S. H., Karpova, T. S., McNally, J. G., and Dasso, M. (2002) *J. Cell Biol.* **156**(4), 595-602
21. Reverter, D., and Lima, C. D. (2005) *Nature* **435**(7042), 687-692
22. Gocke, C., Yu, H., and Kang, J. (2004) *J. Biol. Chem.* **280**(6), 5004-5012

23. Mukhopadhyay, D., Ayaydin, F., Kolli, N., Tan, S. H., Anan, T., Kametaka, A., Azuma, Y., Wilkinson, K. D., and Dasso, M. (2006) *The Journal of cell biology* **174**(7), 939-949
24. Tatham, M. H., Jaffray, E., Vaughan, O. A., Desterro, J. M., Botting, C. H., Naismith, J. H., and Hay, R. T. (2001) *The Journal of biological chemistry* **276**(38), 35368-35374
25. Li, Y., Wang, H., Wang, S., Quon, D., Liu, Y. W., and Cordell, B. (2003) *Proceedings of the National Academy of Sciences of the United States of America* **100**(1), 259-264
26. Bylebyl, G. R., Belichenko, I., and Johnson, E. S. (2003) *The Journal of biological chemistry* **278**(45), 44113-44120
27. Mukhopadhyay, D., and Dasso, M. (2007) *Trends Biochem Sci*
28. Gong, L., Millas, S., Maul, G. G., and Yeh, E. T. (2000) *J. Biol. Chem.* **275**(5), 3355-3359
29. Hang, J., and Dasso, M. (2002) *J. Biol. Chem.* **277**(22), 19961-19966
30. Nishida, T., Tanaka, H., and Yasuda, H. (2000) *Eur. J. Biochem.* **267**(21), 6423-6427
31. Steffan, J. S., Agrawal, N., Pallos, J., Rockabrand, E., Trotman, L. C., Slepko, N., Illes, K., Lukacsovich, T., Zhu, Y. Z., Cattaneo, E., Pandolfi, P. P., Thompson, L. M., and Marsh, J. L. (2004) *Science (New York, N.Y* **304**(5667), 100-104
32. Desterro, J. M., Rodriguez, M. S., and Hay, R. T. (1998) *Mol. Cell* **2**(2), 233-239
33. Baba, D., Maita, N., Jee, J. G., Uchimura, Y., Saitoh, H., Sugasawa, K., Hanaoka, F., Tochio, H., Hiroaki, H., and Shirakawa, M. (2005) *Nature* **435**(7044), 979-982
34. Matunis, M. J., Wu, J., and Blobel, G. (1998) *J. Cell Biol.* **140**(3), 499-509
35. Song, J., Durrin, L. K., Wilkinson, T. A., Krontiris, T. G., and Chen, Y. (2004) *Proceedings of the National Academy of Sciences of the United States of America* **101**(40), 14373-14378
36. Liu, Q., Jin, C., Liao, X., Shen, Z., Chen, D. J., and Chen, Y. (1999) *The Journal of biological chemistry* **274**(24), 16979-16987
37. Minty, A., Dumont, X., Kaghad, M., and Caput, D. (2000) *The Journal of biological chemistry* **275**(46), 36316-36323
38. Lin, D. Y., Huang, Y. S., Jeng, J. C., Kuo, H. Y., Chang, C. C., Chao, T. T., Ho, C. C., Chen, Y. C., Lin, T. P., Fang, H. I., Hung, C. C., Suen, C. S., Hwang, M. J., Chang, K. S., Maul, G. G., and Shih, H. M. (2006) *Molecular cell* **24**(3), 341-354
39. Hannich, J. T., Lewis, A., Kroetz, M. B., Li, S. J., Heide, H., Emili, A., and Hochstrasser, M. (2005) *The Journal of biological chemistry* **280**(6), 4102-4110

40. Shen, T. H., Lin, H. K., Scaglioni, P. P., Yung, T. M., and Pandolfi, P. P. (2006) *Molecular cell* **24**(3), 331-339
41. Dyck, J. A., Maul, G. G., Miller, W. H., Jr., Chen, J. D., Kakizuka, A., and Evans, R. M. (1994) *Cell* **76**(2), 333-343
42. Matunis, M. J., Zhang, X. D., and Ellis, N. A. (2006) *Developmental cell* **11**(5), 596-597
43. Ishov, A. M., Sotnikov, A. G., Negorev, D., Vladimirova, O. V., Neff, N., Kamitani, T., Yeh, E. T., Strauss, J. F., 3rd, and Maul, G. G. (1999) *The Journal of cell biology* **147**(2), 221-234
44. Seeler, J. S., and Dejean, A. (2001) *Oncogene* **20**(49), 7243-7249
45. Zhang, H., Saitoh, H., and Matunis, M. J. (2002) *Molecular and cellular biology* **22**(18), 6498-6508
46. Pichler, A., and Melchior, F. (2002) *Traffic* **3**(6), 381-387
47. Huang, T. T., Wuerzberger-Davis, S. M., Wu, Z. H., and Miyamoto, S. (2003) *Cell* **115**(5), 565-576
48. Kim, J. H., Choi, H. J., Kim, B., Kim, M. H., Lee, J. M., Kim, I. S., Lee, M. H., Choi, S. J., Kim, K. I., Kim, S. I., Chung, C. H., and Baek, S. H. (2006) *Nature cell biology* **8**(6), 631-639
49. Lin, X., Sun, B., Liang, M., Liang, Y. Y., Gast, A., Hildebrand, J., Brunicardi, F. C., Melchior, F., and Feng, X. H. (2003) *Mol. Cell* **11**(5), 1389-1396
50. Kirsh, O., Seeler, J. S., Pichler, A., Gast, A., Muller, S., Miska, E., Mathieu, M., Harel-Bellan, A., Kouzarides, T., Melchior, F., and Dejean, A. (2002) *EMBO J.* **21**(11), 2682-2691
51. Sobko, A., Ma, H., and Firtel, R. A. (2002) *Dev. Cell* **2**(6), 745-756
52. Salinas, S., Briancon-Marjollet, A., Bossis, G., Lopez, M. A., Piechaczyk, M., Jariel-Encontre, I., Debant, A., and Hipskind, R. A. (2004) *J. Cell Biol.* **165**(6), 767-773
53. Zhou, W., Ryan, J. J., and Zhou, H. (2004) *J. Biol. Chem.* **279**(31), 32262-32268
54. Zhao, Y., Kwon, S. W., Anselmo, A., Kaur, K., and White, M. A. (2004) *J. Biol. Chem.* **279**(20), 20999-21002
55. Vertegaal, A. C., Ogg, S. C., Jaffray, E., Rodriguez, M. S., Hay, R. T., Andersen, J. S., Mann, M., and Lamond, A. I. (2004) *J. Biol. Chem.*
56. Wohlschlegel, J. A., Johnson, E. S., Reed, S. I., and Yates, J. R., 3rd. (2004) *J. Biol. Chem.* **279**(44), 45662-45668
57. Panse, V. G., Hardeland, U., Werner, T., Kuster, B., and Hurt, E. (2004) *J. Biol. Chem.* **279**(40), 41346-41351
58. Li, T., Evdokimov, E., Shen, R. F., Chao, C. C., Tekle, E., Wang, T., Stadtman, E. R., Yang, D. C., and Chock, P. B. (2004) *Proc. Natl. Acad. Sci. U. S. A.* **101**(23), 8551-8556

59. Montpetit, B., Hazbun, T. R., Fields, S., and Hieter, P. (2006) *The Journal of cell biology* **174**(5), 653-663
60. Azuma, Y., Arnaoutov, A., Anan, T., and Dasso, M. (2005) *The EMBO journal* **24**(12), 2172-2182
61. Azuma, Y., Arnaoutov, A., and Dasso, M. (2003) *J. Cell Biol.* **163**(3), 477-487
62. Diaz-Martinez, L. A., Gimenez-Abian, J. F., Azuma, Y., Guacci, V., Gimenez-Martin, G., Lanier, L. M., and Clarke, D. J. (2006) *PLoS ONE* **1**, e53
63. Stead, K., Aguilar, C., Hartman, T., Drexel, M., Meluh, P., and Guacci, V. (2003) *The Journal of cell biology* **163**(4), 729-741
64. Holmstrom, S., Van Antwerp, M. E., and Iniguez-Lluhi, J. A. (2003) *Proceedings of the National Academy of Sciences of the United States of America* **100**(26), 15758-15763
65. Chupreta, S., Holmstrom, S., Subramanian, L., and Iniguez-Lluhi, J. A. (2005) *Molecular and cellular biology* **25**(10), 4272-4282
66. Rosendorff, A., Sakakibara, S., Lu, S., Kieff, E., Xuan, Y., Dibacco, A., Shi, Y., Shi, Y., and Gill, G. (2006) *Proc. Natl. Acad. Sci. U. S. A.*
67. Xue, Y., Wong, J., Moreno, G. T., Young, M. K., Cote, J., and Wang, W. (1998) *Mol. Cell* **2**(6), 851-861
68. Hecker, C. M., Rabiller, M., Haglund, K., Bayer, P., and Dikic, I. (2006) *The Journal of biological chemistry* **281**(23), 16117-16127
69. Shi, Y. J., Matson, C., Lan, F., Iwase, S., Baba, T., and Shi, Y. (2005) *Mol. Cell* **19**(6), 857-864
70. Lee, M. G., Wynder, C., Cooch, N., and Shiekhattar, R. (2005) *Nature* **437**(7057), 432-435
71. Hakimi, M. A., Dong, Y., Lane, W. S., Speicher, D. W., and Shiekhattar, R. (2003) *J. Biol. Chem.* **278**(9), 7234-7239
72. David, G., Neptune, M. A., and DePinho, R. A. (2002) *J. Biol. Chem.* **277**(26), 23658-23663
73. Kunapuli, P., Kasyapa, C. S., Chin, S. F., Caldas, C., and Cowell, J. K. (2006) *Experimental cell research* **312**(19), 3739-3751
74. Pascual, G., Fong, A. L., Ogawa, S., Gamliel, A., Li, A. C., Perissi, V., Rose, D. W., Willson, T. M., Rosenfeld, M. G., and Glass, C. K. (2005) *Nature* **437**(7059), 759-763
75. Lyst, M. J., Nan, X., and Stancheva, I. (2006) *The EMBO journal* **25**(22), 5317-5328
76. Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997) *Nature* **389**(6648), 251-260
77. Davey, C. A., Sargent, D. F., Luger, K., Maeder, A. W., and Richmond, T. J. (2002) *J. Mol. Biol.* **319**(5), 1097-1113

78. Luger, K. (2003) *Curr. Opin. Genet. Dev.* **13**(2), 127-135
79. Jenuwein, T., and Allis, C. D. (2001) *Science (New York, N.Y* **293**(5532), 1074-1080
80. Sekinger, E. A., Moqtaderi, Z., and Struhl, K. (2005) *Molecular cell* **18**(6), 735-748
81. Segal, E., Fondufe-Mittendorf, Y., Chen, L., Thastrom, A., Field, Y., Moore, I. K., Wang, J. P., and Widom, J. (2006) *Nature* **442**(7104), 772-778
82. Li, G., and Widom, J. (2004) *Nat Struct Mol Biol* **11**(8), 763-769
83. Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., Barrera, L. O., Van Calcar, S., Qu, C., Ching, K. A., Wang, W., Weng, Z., Green, R. D., Crawford, G. E., and Ren, B. (2007) *Nature genetics* **39**(3), 311-318
84. Yang, A., Zhu, Z., Kapranov, P., McKeon, F., Church, G. M., Gingeras, T. R., and Struhl, K. (2006) *Molecular cell* **24**(4), 593-602
85. Luger, K., and Hansen, J. C. (2005) *Curr. Opin. Struct. Biol.* **15**(2), 188-196
86. Schwabish, M. A., and Struhl, K. (2006) *Molecular cell* **22**(3), 415-422
87. Kouzarides, T. (2007) *Cell* **128**(4), 693-705
88. Grozinger, C. M., and Schreiber, S. L. (2002) *Chemistry & biology* **9**(1), 3-16
89. Zupkovitz, G., Tischler, J., Posch, M., Sadzak, I., Ramsauer, K., Egger, G., Grausenburger, R., Schweifer, N., Chiocca, S., Decker, T., and Seiser, C. (2006) *Molecular and cellular biology* **26**(21), 7913-7928
90. Pflum, M. K., Tong, J. K., Lane, W. S., and Schreiber, S. L. (2001) *The Journal of biological chemistry* **276**(50), 47733-47741
91. Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J. R., Cole, P. A., Casero, R. A., and Shi, Y. (2004) *Cell* **119**(7), 941-953
92. Martin, C., and Zhang, Y. (2005) *Nat. Rev. Mol. Cell Biol.* **6**(11), 838-849
93. Rice, J. C., Briggs, S. D., Ueberheide, B., Barber, C. M., Shabanowitz, J., Hunt, D. F., Shinkai, Y., and Allis, C. D. (2003) *Molecular cell* **12**(6), 1591-1598
94. Yamane, K., Toumazou, C., Tsukada, Y., Erdjument-Bromage, H., Tempst, P., Wong, J., and Zhang, Y. (2006) *Cell* **125**(3), 483-495
95. Wissmann, M., Yin, N., Muller, J. M., Greschik, H., Fodor, B. D., Jenuwein, T., Vogler, C., Schneider, R., Gunther, T., Buettner, R., Metzger, E., and Schule, R. (2007) *Nature cell biology* **9**(3), 347-353
96. Metzger, E., Wissmann, M., Yin, N., Muller, J. M., Schneider, R., Peters, A. H., Gunther, T., Buettner, R., and Schule, R. (2005) *Nature* **437**(7057), 436-439

97. Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007) *Cell* **129**(4), 823-837
98. Vakoc, C. R., Mandat, S. A., Olenchock, B. A., and Blobel, G. A. (2005) *Molecular cell* **19**(3), 381-391
99. Santos-Rosa, H., Schneider, R., Bannister, A. J., Sherriff, J., Bernstein, B. E., Emre, N. C., Schreiber, S. L., Mellor, J., and Kouzarides, T. (2002) *Nature* **419**(6905), 407-411
100. Bernstein, B. E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D. K., Huebert, D. J., McMahon, S., Karlsson, E. K., Kulbokas, E. J., 3rd, Gingeras, T. R., Schreiber, S. L., and Lander, E. S. (2005) *Cell* **120**(2), 169-181
101. Ruthenburg, A. J., Allis, C. D., and Wysocka, J. (2007) *Molecular cell* **25**(1), 15-30
102. Dou, Y., Milne, T. A., Ruthenburg, A. J., Lee, S., Lee, J. W., Verdine, G. L., Allis, C. D., and Roeder, R. G. (2006) *Nat Struct Mol Biol* **13**(8), 713-719
103. Nishioka, K., Chuikov, S., Sarma, K., Erdjument-Bromage, H., Allis, C. D., Tempst, P., and Reinberg, D. (2002) *Genes & development* **16**(4), 479-489
104. Tahiliani, M., Mei, P., Fang, R., Leonor, T., Rutenberg, M., Shimizu, F., Li, J., Rao, A., and Shi, Y. (2007) *Nature* **447**(7144), 601-605
105. Liang, G., Klose, R. J., Gardner, K. E., and Zhang, Y. (2007) *Nat Struct Mol Biol* **14**(3), 243-245
106. Secombe, J., Li, L., Carlos, L., and Eisenman, R. N. (2007) *Genes & development* **21**(5), 537-551
107. Iwase, S., Lan, F., Bayliss, P., de la Torre-Ubieta, L., Huarte, M., Qi, H. H., Whetstine, J. R., Bonni, A., Roberts, T. M., and Shi, Y. (2007) *Cell* **128**(6), 1077-1088
108. Lee, M. G., Norman, J., Shilatifard, A., and Shiekhattar, R. (2007) *Cell* **128**(5), 877-887
109. Eissenberg, J. C., Lee, M. G., Schneider, J., Ilvarsonn, A., Shiekhattar, R., and Shilatifard, A. (2007) *Nat Struct Mol Biol* **14**(4), 344-346
110. Lee, N., Zhang, J., Klose, R. J., Erdjument-Bromage, H., Tempst, P., Jones, R. S., and Zhang, Y. (2007) *Nat Struct Mol Biol* **14**(4), 341-343
111. Yamane, K., Tateishi, K., Klose, R. J., Fang, J., Fabrizio, L. A., Erdjument-Bromage, H., Taylor-Papadimitriou, J., Tempst, P., and Zhang, Y. (2007) *Molecular cell* **25**(6), 801-812
112. Pavri, R., Zhu, B., Li, G., Trojer, P., Mandal, S., Shilatifard, A., and Reinberg, D. (2006) *Cell* **125**(4), 703-717

113. Wysocka, J., Swigut, T., Xiao, H., Milne, T. A., Kwon, S. Y., Landry, J., Kauer, M., Tackett, A. J., Chait, B. T., Badenhorst, P., Wu, C., and Allis, C. D. (2006) *Nature* **442**(7098), 86-90
114. Bannister, A. J., Zegerman, P., Partridge, J. F., Miska, E. A., Thomas, J. O., Allshire, R. C., and Kouzarides, T. (2001) *Nature* **410**(6824), 120-124
115. Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001) *Nature* **410**(6824), 116-120
116. Da, G., Lenkart, J., Zhao, K., Shiekhattar, R., Cairns, B. R., and Marmorstein, R. (2006) *Proc. Natl. Acad. Sci. U. S. A.* **103**(7), 2057-2062
117. Qian, C., Zhang, Q., Li, S., Zeng, L., Walsh, M. J., and Zhou, M. M. (2005) *Nat. Struct. Mol. Biol.* **12**(12), 1078-1085
118. Tochio, N., Umehara, T., Koshiba, S., Inoue, M., Yabuki, T., Aoki, M., Seki, E., Watanabe, S., Tomo, Y., Hanada, M., Ikari, M., Sato, M., Terada, T., Nagase, T., Ohara, O., Shirouzu, M., Tanaka, A., Kigawa, T., and Yokoyama, S. (2006) *Structure* **14**(3), 457-468
119. Boyer, L. A., Latek, R. R., and Peterson, C. L. (2004) *Nat. Rev. Mol. Cell Biol.* **5**(2), 158-163
120. Aasland, R., Stewart, A. F., and Gibson, T. (1996) *Trends Biochem. Sci.* **21**(3), 87-88
121. Tahirov, T. H., Sato, K., Ichikawa-Iwata, E., Sasaki, M., Inoue-Bungo, T., Shiina, M., Kimura, K., Takata, S., Fujikawa, A., Morii, H., Kumasaka, T., Yamamoto, M., Ishii, S., and Ogata, K. (2002) *Cell* **108**(1), 57-70
122. Boyer, L. A., Langer, M. R., Crowley, K. A., Tan, S., Denu, J. M., and Peterson, C. L. (2002) *Mol. Cell* **10**(4), 935-942
123. Yu, J., Li, Y., Ishizuka, T., Guenther, M. G., and Lazar, M. A. (2003) *EMBO J.* **22**(13), 3403-3410
124. Mo, X., Kowenz-Leutz, E., Laumonnier, Y., Xu, H., and Leutz, A. (2005) *Genes Dev.* **19**(20), 2447-2457
125. Dou, Y., Milne, T. A., Tackett, A. J., Smith, E. R., Fukuda, A., Wysocka, J., Allis, C. D., Chait, B. T., Hess, J. L., and Roeder, R. G. (2005) *Cell* **121**(6), 873-885
126. Lee, M. G., Wynder, C., Bochar, D. A., Hakimi, M. A., Cooch, N., and Shiekhattar, R. (2006) *Molecular and cellular biology* **26**(17), 6395-6402
127. Forneris, F., Binda, C., Vanoni, M. A., Battaglioli, E., and Mattevi, A. (2005) *J. Biol. Chem.* **280**(50), 41360-41365
128. Forneris, F., Binda, C., Vanoni, M. A., Mattevi, A., and Battaglioli, E. (2005) *FEBS Lett.* **579**(10), 2203-2207
129. Humphrey, G. W., Wang, Y., Russanova, V. R., Hirai, T., Qin, J., Nakatani, Y., and Howard, B. H. (2001) *J. Biol. Chem.* **276**(9), 6817-6824

130. Shi, Y., Sawada, J., Sui, G., Affar el, B., Whetstine, J. R., Lan, F., Ogawa, H., Luke, M. P., Nakatani, Y., and Shi, Y. (2003) *Nature* **422**(6933), 735-738
131. You, A., Tong, J. K., Grozinger, C. M., and Schreiber, S. L. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98**(4), 1454-1458
132. Andres, M. E., Burger, C., Peral-Rubio, M. J., Battaglioli, E., Anderson, M. E., Grimes, J., Dallman, J., Ballas, N., and Mandel, G. (1999) *Proceedings of the National Academy of Sciences of the United States of America* **96**(17), 9873-9878
133. Ballas, N., Battaglioli, E., Atouf, F., Andres, M. E., Chenoweth, J., Anderson, M. E., Burger, C., Moniwa, M., Davie, J. R., Bowers, W. J., Federoff, H. J., Rose, D. W., Rosenfeld, M. G., Brehm, P., and Mandel, G. (2001) *Neuron* **31**(3), 353-365
134. Wang, J., Scully, K., Zhu, X., Cai, L., Zhang, J., Prefontaine, G. G., Krones, A., Ohgi, K. A., Zhu, P., Garcia-Bassets, I., Liu, F., Taylor, H., Lozach, J., Jayes, F. L., Korach, K. S., Glass, C. K., Fu, X. D., and Rosenfeld, M. G. (2007) *Nature* **446**(7138), 882-887
135. Lunyak, V. V., Burgess, R., Prefontaine, G. G., Nelson, C., Sze, S. H., Chenoweth, J., Schwartz, P., Pevzner, P. A., Glass, C., Mandel, G., and Rosenfeld, M. G. (2002) *Science (New York, N.Y* **298**(5599), 1747-1752
136. Okitsu, C. Y., and Hsieh, C. L. (2007) *Molecular and cellular biology* **27**(7), 2746-2757
137. Holbert, M. A., and Marmorstein, R. (2005) *Curr. Opin. Struct. Biol.* **15**(6), 673-680
138. Rudolph, T., Yonezawa, M., Lein, S., Heidrich, K., Kubicek, S., Schafer, C., Phalke, S., Walther, M., Schmidt, A., Jenuwein, T., and Reuter, G. (2007) *Molecular cell* **26**(1), 103-115
139. Nakamura, T., Mori, T., Tada, S., Krajewski, W., Rozovskaia, T., Wassell, R., Dubois, G., Mazo, A., Croce, C. M., and Canaani, E. (2002) *Molecular cell* **10**(5), 1119-1128
140. Garcia-Bassets, I., Kwon, Y. S., Telese, F., Prefontaine, G. G., Hutt, K. R., Cheng, C. S., Ju, B. G., Ohgi, K. A., Wang, J., Escoubet-Lozach, L., Rose, D. W., Glass, C. K., Fu, X. D., and Rosenfeld, M. G. (2007) *Cell* **128**(3), 505-518
141. Mendez, J., and Stillman, B. (2000) *Molecular and cellular biology* **20**(22), 8602-8612
142. Todorov, I. T., Attaran, A., and Kearsey, S. E. (1995) *The Journal of cell biology* **129**(6), 1433-1445
143. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276**, 307-326
144. Schneider, T. R., and Sheldrick, G. M. (2002) *Acta Crystallogr. D Biol. Crystallogr.* **58**(Pt 10 Pt 2), 1772-1779

145. Otwinowski, Z. (1991) Isomorphous Replacement and Anomalous Scattering. In: Wolf, W., Evans, P. R., and Leslie, A. G. W. (eds). *Science and Engineering Research Council*
146. Cowtan, K., and Main, P. (1998) *Acta Crystallogr. D Biol. Crystallogr.* **54**(Pt 4), 487-493
147. Perrakis, A., Morris, R., and Lamzin, V. S. (1999) *Nat. Struct. Biol.* **6**(5), 458-463
148. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr. D Biol. Crystallogr.* **53**(Pt 3), 240-255
149. Consortium, T. C. (1994) *Acta Crystallogr. D Biol. Crystallogr.* **50**(Pt 5), 760-763
150. Emsley, P., and Cowtan, K. (2004) *Acta Crystallogr. D Biol. Crystallogr.* **60**(Pt 12 Pt 1), 2126-2132
151. Utley, R. T., Owen-Hughes, T. A., Juan, L.-J., Cote, J., Adams, C. C., and Workman, J. L. (1996) *Methods Enzymol* **274**, 276-291
152. Kang, J., Gocke, C. B., and Yu, H. (2006) *BMC biochemistry* **7**, 5
153. Muller, S., Hoege, C., Pyrowolakis, G., and Jentsch, S. (2001) *Nat. Rev. Mol. Cell Biol.* **2**(3), 202-210
154. Rodriguez, M. S., Dargemont, C., and Hay, R. T. (2001) *J. Biol. Chem.* **276**(16), 12654-12659
155. Bernier-Villamor, V., Sampson, D. A., Matunis, M. J., and Lima, C. D. (2002) *Cell* **108**(3), 345-356
156. Yang, S. H., and Sharrocks, A. D. (2004) *Mol. Cell* **13**(4), 611-617
157. Lustig, K. D., Stukenberg, P. T., McGarry, T. J., King, R. W., Cryns, V. L., Mead, P. E., Zon, L. I., Yuan, J., and Kirschner, M. W. (1997) *Methods Enzymol.* **283**, 83-99
158. Li, S. J., and Hochstrasser, M. (1999) *Nature* **398**(6724), 246-251
159. Vassileva, M. T., and Matunis, M. J. (2004) *Mol. Cell Biol.* **24**(9), 3623-3632
160. Sternsdorf, T., Jensen, K., and Will, H. (1997) *J. Cell Biol.* **139**(7), 1621-1634
161. Mao, Y., Desai, S. D., and Liu, L. F. (2000) *J. Biol. Chem.* **275**(34), 26066-26073
162. Dobreva, G., Dambacher, J., and Grosschedl, R. (2003) *Genes Dev.* **17**(24), 3048-3061
163. Coulombe, P., Rodier, G., Bonneil, E., Thibault, P., and Meloche, S. (2004) *Mol. Cell Biol.* **24**(14), 6140-6150
164. Gregoire, S., and Yang, X. J. (2005) *Mol. Cell Biol.* **25**(6), 2273-2287
165. Zhu, B., and Gulick, T. (2004) *Mol. Cell Biol.* **24**(18), 8264-8275
166. McKinsey, T. A., Zhang, C. L., and Olson, E. N. (2001) *Curr. Opin. Genet. Dev.* **11**(5), 497-504

167. Sartorelli, V., Huang, J., Hamamori, Y., and Kedes, L. (1997) *Mol. Cell Biol.* **17**(2), 1010-1026
168. Takagaki, Y., and Manley, J. L. (2000) *Mol. Cell Biol.* **20**(5), 1515-1525
169. Critchlow, S. E., and Jackson, S. P. (1998) *Trends Biochem. Sci.* **23**(10), 394-398
170. Martinez, E., Palhan, V. B., Tjernberg, A., Lymar, E. S., Gamper, A. M., Kundu, T. K., Chait, B. T., and Roeder, R. G. (2001) *Mol. Cell Biol.* **21**(20), 6782-6795
171. Kwek, S. S., Derry, J., Tyner, A. L., Shen, Z., and Gudkov, A. V. (2001) *Oncogene* **20**(20), 2587-2599
172. Jang, M. S., Ryu, S. W., and Kim, E. (2002) *Biochem. Biophys. Res. Commun.* **295**(2), 495-500
173. Zhong, S., Muller, S., Ronchetti, S., Freemont, P. S., Dejean, A., and Pandolfi, P. P. (2000) *Blood* **95**(9), 2748-2752
174. Yang, M., Gocke, C. B., Luo, X., Borek, D., Tomchick, D. R., Machius, M., Otwinowski, Z., and Yu, H. (2006) *Molecular cell* **23**(3), 377-387
175. Tsukada, Y., Fang, J., Erdjument-Bromage, H., Warren, M. E., Borchers, C. H., Tempst, P., and Zhang, Y. (2006) *Nature* **439**(7078), 811-816
176. Klose, R. J., Yamane, K., Bae, Y., Zhang, D., Erdjument-Bromage, H., Tempst, P., Wong, J., and Zhang, Y. (2006) *Nature*
177. Whetstine, J. R., Nottke, A., Lan, F., Huarte, M., Smolikov, S., Chen, Z., Spooner, E., Li, E., Zhang, G., Colaiacovo, M., and Shi, Y. (2006) *Cell* **125**(3), 467-481
178. Wysocka, J., Milne, T. A., and Allis, C. D. (2005) *Cell* **122**(5), 654-658
179. Binda, C., Coda, A., Angelini, R., Federico, R., Ascenzi, P., and Mattevi, A. (1999) *Structure* **7**(3), 265-276
180. Binda, C., Angelini, R., Federico, R., Ascenzi, P., and Mattevi, A. (2001) *Biochemistry* **40**(9), 2766-2776
181. Polticelli, F., Basran, J., Faso, C., Cona, A., Minervini, G., Angelini, R., Federico, R., Scrutton, N. S., and Tavladoraki, P. (2005) *Biochemistry* **44**(49), 16108-16120
182. Stavropoulos, P., Blobel, G., and Hoelz, A. (2006) *Nat Struct Mol Biol* **13**(7), 626-632
183. Chen, Z., Zang, J., Whetstine, J., Hong, X., Davrazou, F., Kutateladze, T. G., Simpson, M., Mao, Q., Pan, C. H., Dai, S., Hagman, J., Hansen, K., Shi, Y., and Zhang, G. (2006) *Cell* **125**(4), 691-702
184. Grüne, T., Brzeski, J., Eberharter, A., Clapier, C. R., Corona, D. F., Becker, P. B., and Müller, C. W. (2003) *Mol. Cell* **12**(2), 449-460
185. Smedley, D., Hamoudi, R., Lu, Y. J., Cooper, C., and Shipley, J. (1999) *Genomics* **60**(2), 244-247

186. Xiao, S., Nalabolu, S. R., Aster, J. C., Ma, J., Abruzzo, L., Jaffe, E. S., Stone, R., Weissman, S. M., Hudson, T. J., and Fletcher, J. A. (1998) *Nature genetics* **18**(1), 84-87
187. Smedley, D., Hamoudi, R., Clark, J., Warren, W., Abdul-Rauf, M., Somers, G., Venter, D., Fagan, K., Cooper, C., and Shipley, J. (1998) *Human molecular genetics* **7**(4), 637-642
188. Reiter, A., Sohal, J., Kulkarni, S., Chase, A., Macdonald, D. H., Aguiar, R. C., Goncalves, C., Hernandez, J. M., Jennings, B. A., Goldman, J. M., and Cross, N. C. (1998) *Blood* **92**(5), 1735-1742
189. van der Maarel, S. M., Scholten, I. H., Huber, I., Philippe, C., Suijkerbuijk, R. F., Gilgenkrantz, S., Kere, J., Cremers, F. P., and Ropers, H. H. (1996) *Human molecular genetics* **5**(7), 887-897
190. Raffa, G. D., Cenci, G., Siriaco, G., Goldberg, M. L., and Gatti, M. (2005) *Molecular cell* **20**(6), 821-831
191. Warren, J. T., Wismar, J., Subrahmanyam, B., and Gilbert, L. I. (2001) *Molecular and cellular endocrinology* **181**(1-2), 1-14
192. Wismar, J., Habtemichael, N., Warren, J. T., Dai, J. D., Gilbert, L. I., and Gateff, E. (2000) *Developmental biology* **226**(1), 1-17
193. Gill, G. (2005) *Current opinion in genetics & development* **15**(5), 536-541
194. Lin, D. Y., Fang, H. I., Ma, A. H., Huang, Y. S., Pu, Y. S., Jenster, G., Kung, H. J., and Shih, H. M. (2004) *Molecular and cellular biology* **24**(24), 10529-10541
195. Kuo, H. Y., Chang, C. C., Jeng, J. C., Hu, H. M., Lin, D. Y., Maul, G. G., Kwok, R. P., and Shih, H. M. (2005) *Proceedings of the National Academy of Sciences of the United States of America* **102**(47), 16973-16978
196. Stankovic-Valentin, N., Deltour, S., Seeler, J., Pinte, S., Vergoten, G., Guerardel, C., Dejean, A., and Leprince, D. (2007) *Molecular and cellular biology* **27**(7), 2661-2675
197. Nathan, D., Ingvarsdottir, K., Sterner, D. E., Bylebyl, G. R., Dokmanovic, M., Dorsey, J. A., Whelan, K. A., Krsmanovic, M., Lane, W. S., Meluh, P. B., Johnson, E. S., and Berger, S. L. (2006) *Genes & development* **20**(8), 966-976
198. Zhao, X., Sternsdorf, T., Bolger, T. A., Evans, R. M., and Yao, T. P. (2005) *Mol. Cell. Biol.* **25**(19), 8456-8464
199. Sohal, J., Reiter, A., Goldman, J. M., and Cross, N. C. (2000) *Cytogenetics and cell genetics* **89**(1-2), 24-28
200. Ginalski, K., and Rychlewski, L. (2003) *Nucleic acids research* **31**(13), 3291-3292
201. Ginalski, K., Elofsson, A., Fischer, D., and Rychlewski, L. (2003) *Bioinformatics (Oxford, England)* **19**(8), 1015-1018

202. Cheng, C., Kussie, P., Pavletich, N., and Shuman, S. (1998) *Cell* **92**(6), 841-850
203. Laherty, C. D., Yang, W. M., Sun, J. M., Davie, J. R., Seto, E., and Eisenman, R. N. (1997) *Cell* **89**(3), 349-356
204. Yoshida, M., Kijima, M., Akita, M., and Beppu, T. (1990) *The Journal of biological chemistry* **265**(28), 17174-17179
205. Lee, M. G., Wynder, C., Schmidt, D. M., McCafferty, D. G., and Shiekhattar, R. (2006) *Chemistry & biology* **13**(6), 563-567
206. Cheng, J., Wang, D., Wang, Z., and Yeh, E. T. (2004) *Molecular and cellular biology* **24**(13), 6021-6028
207. Lai, J. S., and Herr, W. (1992) *Proceedings of the National Academy of Sciences of the United States of America* **89**(15), 6958-6962
208. Schoenherr, C. J., and Anderson, D. J. (1995) *Science (New York, N.Y* **267**(5202), 1360-1363
209. Chong, J. A., Tapia-Ramirez, J., Kim, S., Toledo-Aral, J. J., Zheng, Y., Boutros, M. C., Altshuller, Y. M., Frohman, M. A., Kraner, S. D., and Mandel, G. (1995) *Cell* **80**(6), 949-957
210. Niraula, T. N., Tomizawa, T., Koshiba, S., Inoue, M., Kigawa, T., Yokoyama, S. . (2005) *to be published*
211. Simpson, R. J., Cram, E. D., Czolij, R., Matthews, J. M., Crossley, M., and Mackay, J. P. (2003) *The Journal of biological chemistry* **278**(30), 28011-28018
212. Lee, M. S., Gippert, G. P., Soman, K. V., Case, D. A., and Wright, P. E. (1989) *Science (New York, N.Y* **245**(4918), 635-637
213. Romanowski, P., Madine, M. A., and Laskey, R. A. (1996) *Proceedings of the National Academy of Sciences of the United States of America* **93**(19), 10189-10194
214. Rose, S. M., and Garrard, W. T. (1984) *The Journal of biological chemistry* **259**(13), 8534-8544
215. Colombo, R., Boggio, R., Seiser, C., Draetta, G. F., and Chiocca, S. (2002) *EMBO reports* **3**(11), 1062-1068
216. Tapscott, S. J., Davis, R. L., Thayer, M. J., Cheng, P. F., Weintraub, H., and Lassar, A. B. (1988) *Science (New York, N.Y* **242**(4877), 405-411
217. Montarras, D., Chelly, J., Bober, E., Arnold, H., Ott, M. O., Gros, F., and Pinset, C. (1991) *New Biol.* **3**(6), 592-600
218. Breitbart, R. E., Liang, C. S., Smoot, L. B., Laheru, D. A., Mahdavi, V., and Nadal-Ginard, B. (1993) *Development* **118**(4), 1095-1106
219. Lazaro, J. B., Bailey, P. J., and Lassar, A. B. (2002) *Genes Dev.* **16**(14), 1792-1805
220. Gong, X., Tang, X., Wiedmann, M., Wang, X., Peng, J., Zheng, D., Blair, L. A., Marshall, J., and Mao, Z. (2003) *Neuron* **38**(1), 33-46

221. Hay, R. T. (2005) *Mol. Cell* **18**(1), 1-12
222. Opel, M., Lando, D., Bonilla, C., Trewick, S. C., Boukaba, A., Walfridsson, J., Cauwood, J., Werler, P. J., Carr, A. M., Kouzarides, T., Murzina, N. V., Allshire, R. C., Ekwall, K., and Laue, E. D. (2007) *PLoS ONE* **2**, e386
223. Lan, F., Zaratiegui, M., Villen, J., Vaughn, M. W., Verdel, A., Huarte, M., Shi, Y., Gygi, S. P., Moazed, D., Martienssen, R. A., and Shi, Y. (2007) *Molecular cell* **26**(1), 89-101
224. Yang, M., Culhane, J. C., Szewczuk, L. M., Gocke, C. B., Brautigam, C. A., Tomchick, D. R., Machius, M., Cole, P. A., and Yu, H. (2007) *Nat Struct Mol Biol*
225. Hede, M. S., Petersen, R. L., Frohlich, R. F., Kruger, D., Andersen, F. F., Andersen, A. H., and Knudsen, B. R. (2007) *Journal of molecular biology* **365**(4), 1076-1092
226. Reyes, J. C., Muchardt, C., and Yaniv, M. (1997) *The Journal of cell biology* **137**(2), 263-274

VITAE

Christian Burris Gocke was born on January 20th, 1979, in Morgantown, West Virginia, to Terry Hill Gocke and Michael Thomas Gocke. He lived in Morgantown till the age of 15, when he moved to Mercersburg, PA, to attend Mercersburg Academy preparatory school. Here, Christian played varsity golf for three years and also won awards for excellence in biology his senior year. He continued his education at Vanderbilt University in Nashville, TN, where he majored in Molecular Biology with Honors Research. Most of his undergraduate research was done in the laboratory of James G. Patton. Following in the footsteps of two of his brothers, his father, and his grandfather, who were either physicians or dentists, Christian enrolled at University of Texas – Southwestern Medical Center Medical Scientist Training Program in 2001. He completed two years of medical school and now is ending his research as a graduate student to continue on his last two years of medical school. He is expected to graduate with his M.D./Ph.D. in 2009.

Permanent Address: 4118 Cove Point Drive
Morgantown, WV 26508