Nucleic Acids Research, 2025, **53**, gkaf259
https://doi.org/10.1093/nar/gkaf259
Gene regulation, Chromatin and Epigenetics

Cooperation of a polymerizing SAM domain and an intrinsically disordered region enables full SAMD1 function on chromatin

Merle Geller[1,†], Yinghua Cao[2,†], Clara Simon[1], Bastian Stielow[1], Jingfei Xu[2], Pengshuai Wei[2], Andrea Nist[3,4], Iris Rohner[1], Lea Marie Jeude[1], Theresa Huber[1], Thorsten Stiewe[3,4], Zhanxin Wang[2,*], Robert Liefke[1,5,*]

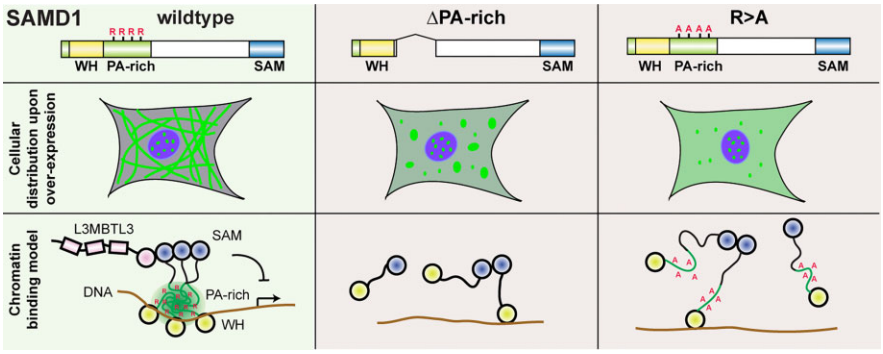
[1]Institute of Molecular Biology and Tumor Research (IMT), Philipps University of Marburg, Marburg 35043, Germany
[2]Key Laboratory of Cell Proliferation and Regulation Biology of Ministry of Education, College of Life Sciences, Beijing Normal University, Beijing 100875, China
[3]Genomics Core Facility, Institute of Molecular Oncology, Member of the German Center for Lung Research (DZL), Philipps University of Marburg, Marburg 35043, Germany
[4]Institute for Lung Health (ILH), Justus Liebig University, Giessen 35392, Germany
[5]Department of Hematology, Oncology, and Immunology, University Hospital Giessen and Marburg, Marburg 35043, Germany

[*]To whom correspondence should be addressed. Email: robert.liefke@imt.uni-marburg.de
Correspondence may also be addressed to Zhanxin Wang. Email: wangz@bnu.edu.cn
[†]These authors contributed equally to this work.

Abstract

Transcription factors orchestrate gene expression through a myriad of complex mechanisms, encompassing collaborations with other transcription factors and the formation of multimeric complexes. The chromatin-binding protein SAMD1 [sterile alpha motif (SAM) domain-containing protein 1] binds to unmethylated CpG-rich DNA utilizing its N-terminal winged-helix (WH) domain. Additionally, its C-terminal SAM domain, which mediates interactions with itself and with L3MBTL3, is crucial for chromatin binding. The precise role of the SAM domain in this process remains unclear. Using structural analyses, we elucidated the distinct homopolymerization modes within the SAM domains of L3MBTL3 and SAMD1, alongside their heterodimerization architecture. Interestingly, SAMD1 necessitates not only the WH and SAM domain but also a proline/alanine-rich intrinsically disordered region (IDR) for efficient chromatin binding. The IDR is essential for the ability of SAMD1 to form large polymers, with its functionality determined by integrity rather than the specific sequence. Mutagenesis studies underscore the critical role of arginines within the IDR for polymerization, chromatin binding, and the biological function of SAMD1. These findings propose a model in which structured and unstructured regions of SAMD1 cooperate in a coordinated fashion to facilitate chromatin binding. This work provides new insights into the diverse mechanisms transcription factors employ to interact with chromatin and regulate gene expression.

Graphical abstract



Introduction

Transcription factors recognize specific DNA sequences and are essential proteins that play a pivotal role in regulating gene expression and chromatin function [1]. These proteins exert their influence through various molecular mechanisms, each contributing to the intricate orchestration of genetic informa-

Received: May 14, 2024. **Revised:** January 30, 2025. **Editorial Decision:** March 18, 2025. **Accepted:** March 21, 2025

tion. The DNA-binding domain is the core of transcription factor functionality [1]. This specialized region enables transcription factors to precisely recognize and bind to specific DNA sequences, thereby dictating their target genes. Transcription factors often work in concert, allowing cooperative binding to DNA and amplifying their influence on gene expression [2]. In line with this, many transcription factors, such as AP-1 [3], bind to DNA as dimers [1]. In certain instances, transcription factors exhibit polymerization, forming larger structures on the chromatin [4]. Some transcription factors establish transcriptional condensates that facilitate the assembly of the transcriptional machinery, enhancing the efficiency and specificity of gene regulation [5].

Understanding the diverse mechanisms employed by transcription factors, from DNA binding cooperativity to dimerization, polymerization and phase separation, is crucial for unraveling the sophisticated control of gene expression that underlies fundamental biological processes.

In our recent investigations, we delineated SAMD1 [sterile alpha motif (SAM) domain-containing protein 1; Uniprot: Q6SPF0] as a transcription factor that exhibits genome-wide association with unmethylated CpG islands (CGIs) [6, 7]. Operating as a transcriptional repressor, SAMD1 interacts with the KDM1A histone demethylase complex and the chromatin regulator L3MBTL3 [6]. Widespread in most tissues [6], SAMD1's abundance is reflective of its significant biological impact. During mouse embryogenesis, its absence results in developmental anomalies, including impaired angiogenesis and embryonic lethality [8]. Strikingly, SAMD1 emerges as a noteworthy player in cancer, exhibiting dysregulation in tumor tissues [6, 7] and playing dual roles—both tumor-suppressive [9] and oncogenic [10], dependent on the specific cancer type. Beyond cancer, SAMD1's influence extends to diverse physiological processes, encompassing muscle adaptation [11], autism [12], antiphospholipid syndrome [13], and arteriosclerosis [14]. These varied associations underscore the versatile and intricate biological functions of SAMD1, positioning it as a multifaceted regulator with implications across a spectrum of physiological and pathophysiological contexts.

At the chromatin, SAMD1 engages directly with unmethylated CpG motifs through its N-terminal winged-helix (WH) domain [6]. This particular domain, classified as a "SAMD1-like WH domain" (InterPro ID: IPR048589), shares structural features with domains found in other chromatin regulatory proteins, such as KAT6A, KAT6B, and ZMYND11 [6, 15]. Notably, in the cases of KAT6A and KAT6B, the analogous WH domain serves a comparable DNA-binding function to that observed for SAMD1 [15, 16]. In addition to the DNA-binding WH domain, SAMD1 harbors a C-terminal SAM domain, renowned for its capacity to interact with other SAM domains, including self-interaction [17]. The SAM domain of SAMD1 directly interacts with the SAM domain of L3MBTL3, important for the recruitment of L3MBTL3 to chromatin [6]. The SAMD1 SAM domain also interacts with itself, forming a closed pentameric ring *in vitro* [6]. This represents alongside the closed octameric ring of SARM1 [18], the only other known closed SAM-domain ring documented to date. Our previous research has established the indispensable roles of both the WH domain and the SAM domain in facilitating SAMD1's efficient chromatin binding [6], yet the molecular mechanisms governing the interplay between these two domains have remained elusive.

This study demonstrates that SAMD1's ability to bind chromatin is closely linked to its polymerization capacity, which arises from the combined contributions of the self-associating SAM domain, and a proline/alanine-rich (PA-rich) intrinsically disordered region (IDR). We find that the presence of arginines within the IDR is particularly essential, while the precise sequence of the IDR appears less critical. This refined understanding of SAMD1's chromatin binding highlights the complex interplay between structured and unstructured regions for chromatin binding.

Materials and methods

Cell culture

HEK293 cells were cultured in Dulbecco's modified Eagle's medium (DMEM)/F-12 + GlutaMAX™ (Gibco, 31331-028) supplemented with 10% fetal bovine serum (FBS; Gibco, 10270-106) and 1% penicillin–streptomycin (P/S; Gibco, 15140-122). PC3 and NCI-H23 cells were cultured in RPMI 1640 + GlutaMAX™ (Gibco, 61870-010) with 10% FBS and 1% P/S.

PaTu8988t cells were cultured in DMEM + GlutaMAX™ (Gibco, 61965-026) with 5% FBS and 1% P/S. PaTu8988t SAMD1 knockout (KO) cells were obtained as described [9]. Rescue with FLAG-HA ER SAMD1 constructs was performed as described [9].

E14 mouse embryonic stem (ES) cells (E14TG2a) were cultured in DMEM + GlutaMAX™ (Gibco, 61965-026), 15% FBS (Gibco, 10270-106), 1× nonessential amino acids (Gibco, 11140-035), 1× sodium pyruvate (Gibco, 11160-039), 1× P/S (Gibco, 15140-122), 0.15% β-mercaptoethanol, and LIF (self-made) on gelatin-coated plates.

SAMD1 mouse ES KO cells were created as described [6]. Mouse ES cells were transfected using the jetPRIME transfection kit (Polyplus, 101000046) according to the manufacturer's instructions.

SAMD1, L3MBTL3, and KDM1A KO mouse ES cells were previously established [6].

Antibodies

The antibodies used for western, immunofluorescence, and chromatin immunoprecipitation (ChIP) are as follows:

Name	Company	Cat. number
Atherin (SAMD1)	Bethyl	A303-578A
SAMD1	Self-made [6]	N/A
L3MBTL3	Self-made [6]	N/A
KDM1A	Abcam	17721
SP1	Self-made [19]	N/A
Beta Actin	Santa Cruz	Sc-47778
Beta Tubulin (KMX-1)	Millipore	MAB3408
FLAG	Sigma–Aldrich	F3165
FLAG	Sigma–Aldrich	F7425
HA	Roche	11087423001
Sheep Anti-Mouse HRP	GE Healthcare	NA931
Donkey Anti-Rabbit HRP	GE Healthcare	NA934
Goat Anti-Rabbit (H + L) Alexa Fluor 488	Thermo Fisher	A-11008
Goat Anti-Rabbit (H + L) Alexa Fluor 546	Thermo Fisher	A-11035
Donkey Anti-Rat (H + L) Alexa Fluor 488	Thermo Fisher	A-21208
Goat Anti-Rat (H + L) Alexa Fluor 546	Thermo Fisher	A-11081

SAMD1 constructs

Wild-type human SAMD1 expression constructs were obtained as described [6]. Zebrafish Samd1a and Samd1b were cloned via polymerase chain reaction (PCR) using complementary DNA (cDNA) from zebrafish. SAMD1 deletion mutants were cloned via extension overlap PCR. The regions within SAMD1 were defined as follows: WH: 27–100; PA: 101–211; central: 212–350; PG: 351–449; and SAM: 450–537. SAMD1 with shuffled or mutated PA-rich sequences were obtained by DNA synthesis. The sequences of all SAMD1 proteins are provided in Supplementary File S1.

Transfection of cells with SAMD1 constructs

The transfection in HEK293 and PaTu8988t cells was performed via polyethylenimine (PEI). Transfection of mouse ES cells was performed using JetPrime (Polyplus; 101000046). For most experiments pDEST lentiviral vectors expressing the respective SAMD1 construct were used for transfection.

Mammalian two-hybrid assay

Mammalian two-hybrid assay was performed using the Mammalian Two-Hybrid Assay Kit from Stratagene/Agilent [catalog no. 211344 (discontinued)]. SAMD1–SAM (452-538), L3MBTL3–SAM (705-780), and SFMBT1–SAM (751-866) were cloned into the pCMV activation domain (AD) or pCMV DNA-binding (BD) vector. For the experiment, 30000 HEK293 cells were plated into one 24-well plate. Two technical replicates were performed for each experiment. The next day, the cells were transfected with 200 ng of pFR-Luc, 0.5 ng of SV-40-RLuc, and 50 ng of the pCMV-AD and pCMV-BD constructs using FuGeneHD (Promega). Two days after transfection, cells were washed one time with phosphate-buffered saline (PBS) and lysed for 20 min with 1× passive lysis buffer (reagents from the DualLuciferase Reporter Assay System Kit, Promega). Firefly luciferase and renilla luciferase activity was determined by using the DualLuciferase Reporter Assay System Kit (Promega). Firefly values were normalized to renilla activity.

Luciferase-based reporter assay

Briefly, 60 000 HEK293 cells were plated into 24-well plates and transfected with 200 ng of pFR-Luc (negative control) or a pFR-Luc plasmid containing a CpG promotor sequence; 0.5 ng of SV-40-RLuc; and 50 ng of a pCMV-AD vector containing either wild-type SAMD1 or different deletion mutants. Forty-two hours after transfection, cells were washed once with PBS and lysed with 1× passive lysis for 20 min (Promega, E1960). Subsequently, firefly and renilla luciferase activity was determined using the Dual Luciferase Reporter Assay (Promega, E1960). Renilla luciferase activity was used for normalization. Two technical replicates were performed for each experiment. Sequences of different CpG promotors were amplified via PCR and cloned into the pFR-Luc reporter plasmid. Their sequences can be found in Supplementary Fig. S3D.

Protein expression and purification for crystallization

Open reading frames of SAMD1 and L3MBTL3 are chemically synthesized with codon optimized for efficient bacterial expression. L3MBTL3–SAM (residues 705–773) was inserted into a hexahistidine-SUMO-tagged pCDFDuet-1 vector. L3MBTL3–SAM (residues 705–775 with L756E/L760E mutations) was inserted into a MBP-tagged pRSFDuet-1 vector. SAMD1–SAM (residues 459–523 with L495R mutation) was inserted into a hexahistidine-SUMO-tagged pCDFDuet-1 vector.

L3MBTL3–SAM was expressed in *Escherichia coli* strain BL21 (DE3) at 37°C till the OD_{600} (optical density at 600 nm) reached around 1.0, then the cells were cooled at 20°C for around 1 h before 0.2 mM isopropyl-β-D-thiogalactopyranosid (IPTG) were added to induce expression overnight. Cells were collected by centrifugation at 5000 × *g* for 10 min. Cell pellets were resuspended with the initial buffer [20 mM Tris (pH 8.0), 500 mM NaCl, and 20 mM imidazole] and then sonicated for around 5 min. The supernatant was collected by centrifugation of the cell lysate at 25 000 × *g* for 1 h. Histidine-SUMO-tagged target protein was isolated through a nickel-charged HiTrap Chelating FF column (GE Healthcare). The histidine-SUMO tag was then cleaved by incubating with histidine-tagged ULP1 protease and dialyzed with the initial buffer at 4°C. The dialyzed solution was then reloaded onto a nickel-charged chelating column to remove both the histidine-tagged SUMO and ULP1. The flow through was diluted to a salt concentration of 100 mM NaCl and was loaded directly onto a Q column (GE Healthcare). The target protein was separated by increasing the salt concentration from the low-salt buffer [20 mM Tris (pH 8.0), 100 mM NaCl, and 2 mM DTT] to the same buffer containing 1 M NaCl through a linear gradient. The target protein was further purified by a Hiload 20016/600 gel-filtration column (GE Healthcare) equilibrated with the low-salt buffer. Purified proteins were concentrated to around 20 mg/ml and stored in a −80°C freezer.

For the L3MBTL3–SAM/SAMD1–SAM complex, two plasmids were co-expressed in E.coli BL21 (DE3). The cell pellets were resuspended with the buffer containing 20 mM Tris (at pH 8.0), 200 mM NaCl, and 20 mM imidazole. After elution of the histidine-SUMO-tagged protein, the MBP tag and the histidine-SUMO tag were cleaved by incubating with TEV proteases and histidine-tagged ULP1 protease overnight. The sample was then diluted to contain 100 mM NaCl and was then loaded directly onto a Q column (GE Healthcare). Target protein was eluted by increasing the salt concentration from 100 mM NaCl to 1 M NaCl. The eluted target protein was then purified through a HiLoad 20016/600 gel filtration column before loading onto a Mono Q 5/50 column (GE Healthcare) for further purification. After these steps, the target protein was concentrated to around 20 mg/ml and stored in a freezer.

The SAMD1 proteins used for electron microscopy (EM) studies were cloned into a modified pFASTbac vector with an N-terminal 6× His-MBP-TEV tag. Bacmid, virus, and protein production were conducted following a standard Bac-to-Bac baculovirus expression protocol. Sf9 cells infected by p2 viruses were collected by centrifugation (1500 × *g*, 4°C, 20 min), resuspended in lysis buffer [500 mM NaCl, 20 mM Tris (pH 8.0), 20 mM imidazole (pH 8.0), 10% glycerol, and 1 mM phenylmethylsulfonyl fluoride (PMSF)] at 4°C, and then sonicated for around 5 min. The soluble fraction of the cells was collected by centrifugation of the cell lysate at 18 000 rpm for 1 h. His-MBP-TEV-tagged target protein was isolated through a nickel-charged HiTrap Chelating FF column (GE Healthcare). The samples were diluted to half the initial salt concentration (250 mM NaCl) and then were loaded directly

onto a heparin column to remove bound DNA. The His-MBP tag was cleaved by incubating with His-tagged TEV proteases overnight and removed by loading the solution to a nickel column again. Target protein was further purified by a HiLoad 20016/600 gel-filtration column equilibrated with the running buffer [100 mM NaCl, 20 mM HEPES (pH 8.0), and 2 mM DTT]. Peak fractions were pooled and concentrated. All mutations of SAMD1–SAM and L3MBTL3–SAM were generated by PCR method.

Gel-filtration

Gel-filtration experiments with SAM domains were performed using a Superdex 200.

Crystallization and structure resolution

Crystallization was carried out using the hanging-drop, vapor-diffusion method by mixing equal volume of protein and well solution.

The crystals of L3MBTL3–SAM were grown in the solution containing 0.4 M sodium malonate at pH 6.0, 0.1 M MES monohydrate at pH 6.0, and 0.5% (w/v) polyethylene glycol 10 000 at 20°C. Crystallization buffer containing 30% glycerol was used as the cryoprotectant. The crystals of L3MBTL3–SAM/SAMD1–SAM complex were grown in the solution containing 0.01 M magnesium acetate tetrahydrate, 0.05 M MES monohydrate at pH 5.6, and 2.63 M ammonium sulfate at 20°C. Crystallization buffer containing 3 M ammonium sulfate was used as the cryoprotectant.

All the datasets were collected at the Shanghai Synchrotron Radiation Facility beamlines in China at the temperature of −196°C. Datasets for L3MBTL3–SAM and L3MBTL3–SAM/SAMD1–SAM complex were collected at the beamline BL19U1 and BL18U1, respectively. The datasets were processed through the program HKL2000 [20]. The structure of the L3MBTL3–SAM was solved by molecular replace method by PHENIX [21] using SAMD1–SAM structure as the model (PDB ID: 6LUJ). The structure of L3MBTL3–SAM/SAMD1–SAM complex was also solved by molecular replacement, using both L3MBTL3–SAM and SAMD1–SAM as the model. Both structures were manually rebuilt through COOT [22] and further refined by PHENIX.

Sample preparation and data collection for negative-staining EM

For negative-staining sample preparation, proteins were concentrated to 50 nM and dialyzed overnight at 4°C in titration buffer [20 mM Tris (pH 8.0), 50 mM NaCl, and 2 mM DTT]. Then 4 µl sample was added to the glow-discharged grid (200 mesh) and incubated for 1 min. Briefly, 2% uranyl acetate was used to fix and stain the grid and let it dry at room temperature before use. Micrographs were acquired on a 120-kV microscope.

Wound healing

To determine the migration rate of PaTu8988t cells, cells were seeded in culture inserts (Ibidi; 80209). A total of 70 µl of cell suspension at a density of 6×10^5 cells per ml was applied containing either ethanol as a solvent control treatment or 200 nM 4-hydroxytamoxifen (4-OHT). On the next day, PaTu8988t cells were starved with medium containing 0.5% FBS, as well as ethanol or 4-OHT, for 6 h before removing the insert. Photos were taken using an Olympus CKX53 microscope. After 24 h, photos were taken on the same spots and the cell-free area was measured for both timepoints using ImageJ Fiji (version: 2.1.0/1.53r).

Immunofluorescence

HEK293 cells were seeded on coverslips one day prior to transfection. Transfection was performed via PEI using pDEST lentiviral vectors (except for Supplementary Fig. S4H). The medium was changed 18 h after transfection and immunofluorescence staining was started 42 h after transfection. In order to do so, cells were washed three times with PBS. Subsequently, cells were fixed with 4% paraformaldehyde in PBS, permeabilized with 0.5% Triton X-100 in PBS (wash buffer), and blocked with wash buffer + 10% fetal calf serum (FCS) for 30 min each. Cells were incubated with the primary antibody, diluted 1:500 in wash buffer + 10% FCS, for 1 h. Then, cells were washed three times for 10 min with wash buffer before they were incubated with a secondary antibody, diluted 1:1000 in wash buffer + 10% FCS, for 1 h in the dark. Finally, cells were washed three times for 10 min with wash buffer and mounted onto microscopy slides using ProLong™ Antifade Mountant with DAPI (Invitrogen; P36971). Photos were taken using a Leica DM 5500 microscope. For quantification of fiber formation, at least 60 cells per condition were analyzed. Representative images are shown, DAPI and FLAG/HA signals were merged using the software ImageJ2 (version 2.14.0/1.54f). Brightness/ contrast was adjusted for each experiment.

Primary antibodies used for immunofluorescence were as follows: Rabbit polyclonal anti-FLAG antibody (Sigma–Aldrich, F7425); Rat monoclonal anti-HA antibody (Roche, 11867423001); and Rabbit anti-SAMD1 (self-made [6]). Secondary antibodies used for immunofluorescence were as follows: Goat anti-Rabbit IgG (H + L), Alexa Fluor™ 488 (Thermo Fisher Scientific, A-11008); Goat anti-Rabbit IgG (H + L), Alexa Fluor™ 546 (Thermo Fisher Scientific, A-11035); Goat anti-Rat IgG (H + L), Alexa Fluor™ 546 (Thermo Fisher Scientific, A-11081); and Donkey anti-Rat IgG (H + L), Alexa Fluor™ 488 (Thermo Fisher Scientific, A-21208).

Subcellular protein fractionation

Subcellular fractionation was performed using the subcellular protein fractionation kit for cultured cells (Thermo Fisher; 78840). One 10-cm dish corresponding to 20 µl packed cell volume was used as starting material for each sample.

Whole cell extract for western blotting

In order to prepare whole cell extracts, 300 000 mouse ES SAMD1 KO cells or HEK293 were seeded on six-well plates 1 day prior to transfection with either jetPRIME (Polyplus; 101000046) or PEI, respectively. Mouse ES cells were seeded on gelatin-coated six-well plates. Cells were harvested 42 h after transfection by scraping in 60 µl lysis buffer [50 mM Tris/Cl (pH = 7.5), 150 mM NaCl, 1% Triton X-100, 10% glycerol, 1× Protease Inhibitor Cocktail (PIC, Roche; 04 693 116 001), 0.5 mM PMSF, and Benzonase 1:500 (EMD Millipore; 70664-10KUN]. Samples were incubated at 4°C for 1 h while shaking. Subsequently, they were centrifuged at 17 000 × *g* for 10 min and the protein concentration in the super-

natant (whole cell extract) was measured using the DC Protein Assay (Bio-Rad; 5000006).

Western blots

Blotting of sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) gels was performed using the Trans-Blot® Turbo™ Transfer System (Bio-Rad; 1704150). Primary antibodies used for western blots were as follows: Anti-Atherin, Bethyl, 90378 (1:2000); Anti FLAG, Sigma–Aldrich, F3165 (1:2000); Anti Tubulin, Sigma–Aldrich, MAB3408 (1:10 000); Anti beta Actin, Santa Cruz, sc-47778 (1:500); and Anti SP-1, self-made (1:2000). Secondary antibodies used for western blots were as follows: Anti mouse, HRP-linked, NA931 and Anti rabbit, HRP-linked, NA934.

RNA extraction and RT-qPCR

For RNA isolation, PaTu8988t cells were seeded at a density of 5×10^5 cells per well on a six-well plate. After 24 h, the medium was exchanged to medium containing either ethanol as a solvent control treatment or 200 nM 4-OHT. Following 24 h, RNA was prepared according to the manufacturer's manual using the RNeasy Mini Kit (Qiagen; 74004) including an on-column DNA digest.

cDNA synthesis was performed using the Tetro cDNA Synthesis Kit (Meridian Bioscience; BIO-65050) according to the manufacturer's protocol including 10 mM dNTP Mix (NEB; N0447S), RiboSafe RNase Inhibitor (Meridian Bioscience; BIO-65028), and OligodTs. The resulting cDNA was diluted 1:20 before proceeding to analysis via quantitative reverse transcription polymerase chain reaction (RT-qPCR). RT-qPCR was performed using the MyTaq™ Mix (Bioline; BIO-25041) with additional SYBR Green (Thermo Fisher; S7585) and the Stratagene Mx300P Thermocycler (Agilent). Resulting Ct values were normalized to GAPDH.

Following is the list of used RT-qPCR primers:

RT hCDH2 fw	TGGGAATCCGACGAATGGATGAAAG
RT hCDH2 rv	GGTGGAGCTGTGGGGTCATTGTCAG
RT hGAPDH fw	AGCCACATCGCTCAGACAC
RT hGAPDH rv	GCCCAATACGACCAAATC
RT hL3MBTL3 fw	AGGACTACTCGTGAGACGGA
RT hL3MBTL3 rv	AACGTGCCCACTCCATCTTT

Chromatin preparation

Cells were fixed by adding 1% formaldehyde to the cell culture medium. After 10 min, 125 mM glycine was added for 5 min to stop fixation. Then, cells were washed twice with cold PBS and scraped in 1 ml cold buffer [10 mM HEPES/KOH (pH = 6.5), 10 mM ethylenediaminetetraacetic acid (EDTA), 0.5 mM ethyleneglycol-bis(beta-aminoethylether)-tetraacetic acid (EGTA), and 0.25% Triton X-100]. After centrifugation (440 × *g*, 4°C, 5 min), the pellet was resuspended in 1 ml 10 mM HEPES/KOH (pH = 6.5), 10 mM EDTA, 0.5 mM EGTA, and 200 mM NaCl per initial 15-cm dish and incubated on ice for 15 min. After another centrifugation step (440 × *g*, 4°C, 5 min), the pellet was resuspended in ∼250 µl 50 mM Tris/Cl (pH = 8.0), 10 mM EDTA, 1% SDS, and 1× PIC per initial 15-cm dish according to the pellet size. After 20 min incubation on ice, the samples were sonicated using a precooled Bioruptur, Diagenode (2 × 7 min, 30 s on 30 s off). Finally, samples were centrifuged at 17 000 × *g* and 4°C for 10 min and the supernatant containing the sheared chromatin was transferred to fresh 1.5-ml tubes.

Chromatin immunoprecipitation

ChIP was performed according to Diagenode's "One-day ChIP kit" (Diagenode; C0101008) protocol with the following exceptions: First, 10× ChIP buffer was replaced by a self-made buffer consisting of 50 mM Tris/Cl (pH = 7.5), 150 mM NaCl, 5 mM EDTA, 1% Triton X-100, and 0.5% NP-40. Secondly, Dynabeads Protein A (Thermo Fisher Scientific; 10008D) were used instead of the Protein A beads. Thirdly, DNA purifying slurry was replaced by 10% w/v Chelex® 100 Resin (Bio-Rad, #142-1253). For each ChIP, 3 µg of IgG control antibody or a specific antibody (anti SAMD1, self-made, anti L3MBTL3, self-made, or anti FLAG, Sigma–Aldrich, F3165) were used.

When preparing samples for ChIP sequencing (ChIP-seq), the DNA purification steps were altered as follows. First, 230 µl elution buffer (100 mM $NaHCO_3$ and 1% SDS) were added to the beads, then beads were incubated at room temperature while shaking for 30 min for elution. Next, samples were centrifuged at 17 000 × *g* for 1 min and the supernatant was transferred to a new tube containing 8 µl 5M NaCl. Samples were incubated at 65°C overnight for cross-link reversal.

Subsequently, 8 µl of 1M Tris/Cl (pH = 6.5), 4 µl 0.5 M EDTA, and 2 µl of Proteinase K (10 µg/µl) were added to each sample and samples were incubated at 45°C for 1 h while shaking.

Finally, DNA was purified using the QIAquick PCR Purification Kit (Qiagen; 28104). DNA was eluted in 2 mM Tris/Cl and concentrations were measured using the the Quant-iTTM dsDNA Assay Kit (Thermo Fisher Scientific; Q33120) and the NanoDrop™ 3300 (Thermo Fisher Scientific).

Following is the list of used ChIP-quantitative polymerase chain reaction (ChIP-qPCR) primers:

ChIP hCDH2 fw	GAGGAAGCAAAACCACGAGC
ChIP hCDH2 rv	AGAGATCAAGGAGCTGGGGA
ChIP hL3MBTL3 fw	CACCAGCGCAGTTTCCCCTC
ChIP hL3MBTL3 rv	TTCTGTCCCTCCCCTCCCCT
ChIP hNANOS1 fw	GCTCTGGTCTGCAGCCAATG
ChIP hNANOS1 rv	TTATCGGGCCCACTCCTCCC
ChIP hMYC-3000 fw	AACCTCCACTGCCAGAAGTCCTTA
ChIP hMYC-3000 rv	GAAATTTACCTGGCACGTGTCCCT
ChIP mL3mbtl3 fw	TTGACACCGGGTGCTGGGT
ChIP mL3mbtl3 rv	CTCGGCACAAAGGCCACAC
ChIP mCbln1 fw	CTACCTGAAAGCCCACGAAC
ChIP mCbln1 rv	CCGTTGTTATTAGGAGCGGC
ChIP mNanos1 fw	AGTCAGGTCCGCAGCCATTG
ChIP mNanos1 rv	CTTTCTCCTGTGCCCCTCCT
ChIP mIrf2bp2 fw	GTGCTATCTGTGTGACCTGC
ChIP mIrf2bp2 rv	CTGCAGCAGTAGGTCCTTGG
ChIP mPth2 fw	GAGACCTGCCAGATGTCCAG
ChIP mPth2 rv	CTGCATGTAAGAGTCCAGCC
ChIP mPou5f1 fw	GTTTGTGAGGTGTCCGGTGA
ChIP mPou5f1 rv	GCTCACCTAGGGACGGTTTC

Library preparation and next-generation sequencing

Next-generation sequencing was performed at the Genomics Core Facility Marburg (Center for Tumor Biology and Immunology, Hans-Meerwein-Str 3, 35043 Marburg, Germany). For ChIP-seq, the Microplex library preparation kit v2

(Diagenode; C05010012) was used for indexed sequencing library preparation with chromatin immunoprecipitated DNA. Libraries were purified on AMPure magnetic beads (Beckman; A6388). ChIP-seq libraries were quantified on a Bioanalyzer (Agilent Technologies). Next-generation sequencing was performed on an Illumina NextSeq 550.

Bioinformatic analysis

ChIP-seq data were mapped to the mouse genome mm39 using bowtie2 [23], allowing one mismatch. BigWig files were obtained using deepTools/bamCoverage [24, 25]. Heatmaps and profiles were created using Galaxy/deepTools [24, 25]. CGIs annotations were obtained from the UCSC table browser [26]. Following previously published SAMD1 ChIP-seq data were used: mouse ES GSM4287311 [6], HepG2 GSM5731083 [10], and PaTu8988t GSM7664371 [9].

Predictions of binding of SAMD1 WH domains to CpG-rich DNA from the *Cbln1* promoter were performed using the AlphaFold 3 [27] server (https://alphafoldserver.com/) using following DNA sequence: GCTCCGGCCGGGAGTCAGACGAGGGAGCGCGCGCGCGCTCCCGGAC.

The enrichment analysis of amino acids in specific regions of SAMD1 was performed similarly as described in [28]. Specifically, we subtracted the portion of a specific amino acid in a sequence from the portion of the same amino acid in a reference proteome, and divided this to by the portion of the amino acid in the reference proteome.

Analysis of alanine and proline content in proteins of the UniProt databases was performed using a custom R script [29], by which for each protein the 100-amino acids window with the highest proline and alanine content for each protein was identified and returned.

For analysis of proline and alanine, as well as arginine content in the IDR of SAMD1 in diverse organisms, a custom R script was used. Following NCBI protein entries were used: mammals: NP_612361.1 (*Homo sapiens*), NP_001074884.1 (*Mus musculus*), NP_001075633.1 (*Oryctolagus cuniculus*), XP_024092811.2 (*Pongo abelii*), XP_007507948.2 (*Monodelphis domestica*), XP_038282514.1 (*Canis lupus familiaris*), XP_003760731.3 (*Sarcophilus harrisii*), XP_034802607.2 (*Pan paniscus*), XP_039321970.1 (*Saimiri boliviensis boliviensis*), and XP_023102016.1 (*Felis catus*); birds: XP_041567594.1 (*Taeniopygia guttata*), XP_030330378.1 (*Strigops habroptila*), XP_030826098.1 (*Camarhynchus parvulus*), XP_037982172.1 (*Motacilla alba alba*), and XP_053824310.1 (*Vidua chalybeata*); reptils: XP_026552055.1 (*Pseudonaja textilis*), XP_013916532.1 (*Thamnophis sirtalis*), XP_015675603.1 (*Protobothrops mucrosquamatus*), XP_028576959.1 (*Podarcis muralis*), XP_032992288.1 (*Lacerta agilis*), and XP_034955172.1 (*Zootoca vivipara*); amphibia: XP_029441241.1 (*Rhinatrema bivittatum*), XP_030053005.1 (*Microcaecilia unicolor*), and XP_033780333.1 (*Geotrypetes seraphini*); and fish: NP_001303918.1 (*Danio Rerio*), XP_005163590.1 (*Danio rerio*), SBR05349.1 (*Nothobranchius kuhntae*), SBQ87066.1 (*Nothobranchius kuhntae*), XP_034075592.1 (*Gymnodraco acuticeps*), and XP_034094978.1 (*Gymnodraco acuticeps*).

Statistical analysis

Statistical analysis was performed as described in the figure legends. Error bars indicate standard deviation (SD). The significance of biological experiments was evaluated via ANOVA or unpaired, two-sided Student's *t*-tests. The significance of differences on SAMD1 ChIP-seq levels was evaluated using a two-sided Kolmogorov–Smirnov test. *P*-value (***$P < .001$; **$P < .01$; *$P < .05$).

Results

SAM domains of L3MBTL3 and SAMD1 form distinct homopolymers *in vitro*

SAMD1 and L3MBTL3 are part of a larger cohort of chromatin regulators characterized by a C-terminal SAM domain [30], yet the specific functions of these SAM domains remain elusive. In our previous investigation utilizing a mammalian-2-hybrid system, we discovered that the SAM domain of SAMD1 not only directly interacts with the SAM domain of L3MBTL3 but also exhibits self-association (Fig. 1A) [6]. Interestingly, this interaction is exclusive to SAM domains within SAMD1 and L3MBTL3 (and its close homolog L3MBTL4), as SAMD1's SAM domain does not engage with SAM domains of other nuclear SAM-containing proteins, including SFMBT1 (Fig. 1A) [6]. Similarly, the SAM domain of L3MBTL3 demonstrates self-association and interacts with SAMD1's SAM domain (Fig. 1A), suggesting the propensity for homo- and heteropolymer formation by both SAMD1 and L3MBTL3. Indeed, via structural studies, we have shown that the SAM domain of SAMD1 forms a closed homopentameric ring *in vitro* (Fig. 1B) [6], which is distinct from other known SAM domain polymers [17]. However, the structural characteristics of L3MBTL3's SAM homopolymer and SAMD1/L3MBTL3 SAM heteromer remain unresolved.

To address whether the SAM domain of L3MBTL3 has similar molecular properties as the SAM domain of SAMD1, we crystallized the SAM domain of L3MBTL3 and solved its structure at the resolution of 2.0 Å (Table 1). In the crystal lattice, L3MBTL3 SAM self-associates to form a left-handed spiral, in which every six molecules wind a circle and translate 50 Å (Fig. 1C). The architecture is similar to SAM polymers formed by the ETS translocation variant 6 (ETV6, also TEL) and the polyhomeotic proteins [4, 17], but is quite different from the SAMD1 SAM pentameric polymer (Fig. 1B) [6]. L3MBTL3 SAM domain contains five α helices that fold into a typical globular-shaped SAM domain structure (Fig. 1D). In the polymer, two neighboring molecules contact each other through their conserved mid-loop (ML) and end-helix (EH) surfaces in a head-to-tail fashion (Fig. 1D). The interaction between the ML and EH surfaces is mainly mediated by electrostatic interactions. The negatively charged ML surface of one L3MBTL3 SAM domain made of helices α2, α3, and α4 is attracted by the positively charged EH surface of another L3MBTL3 SAM made of helices α4 and α5 (Fig. 1E). In detail, acidic residues Asp732, Glu733, Asp736, and Asp747 from the ML surface of one molecule form a series of hydrogen bonds with residues Lys755, Leu757, Gly757, Lys761, and Gln745 from the EH surface of a neighboring molecule (Fig. 1D). The interaction between two L3MBTL3 molecules is further stabilized by hydrophobic interactions, as the hydrophobic residues Leu756 and Leu760 that are located at the center of the EH surface of one molecule are fully covered by the hydrophobic patch of the ML surface composed of residues Ile751, Ile735, Met752, Met743, and Leu742 of an

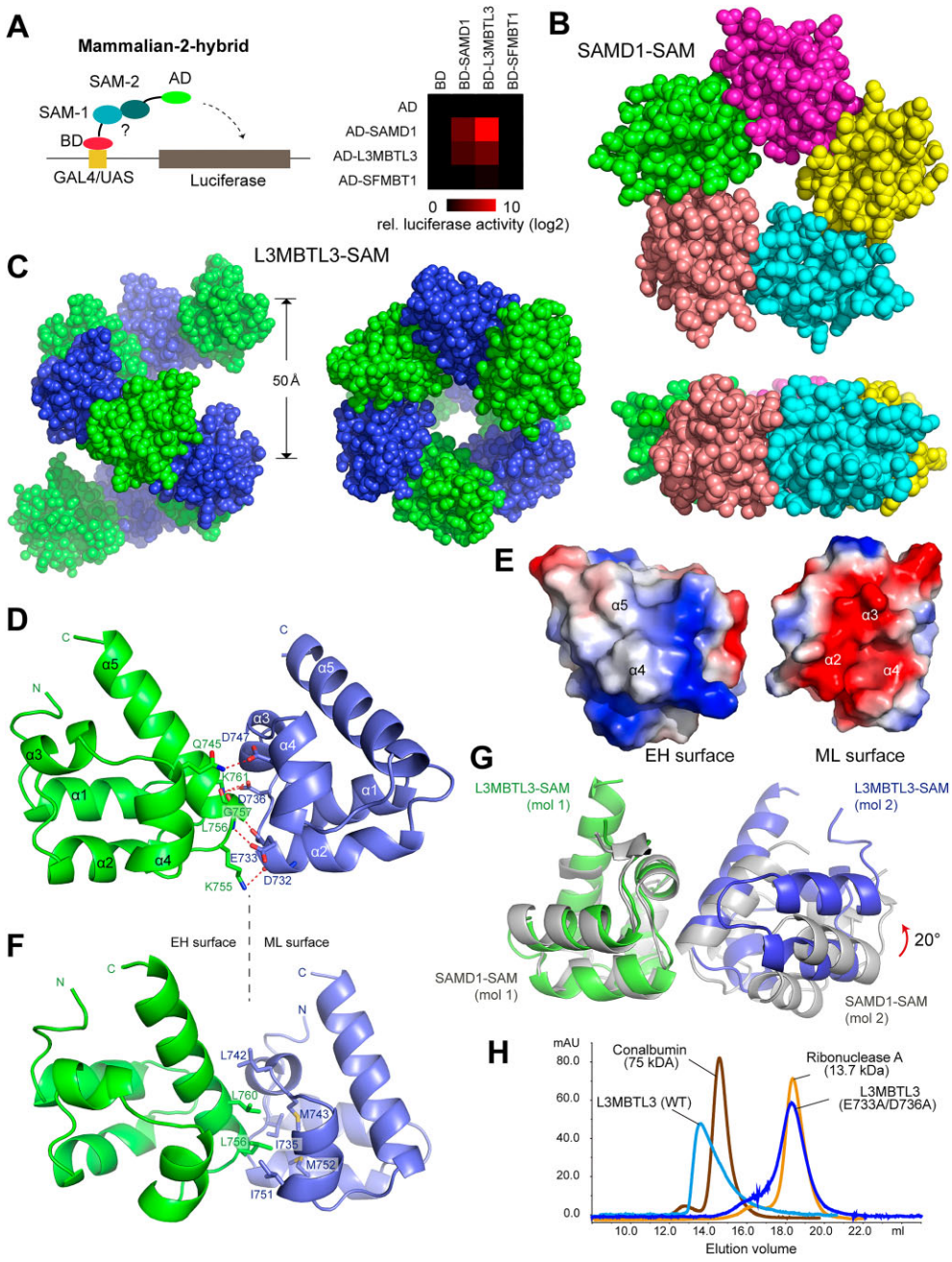


Figure 1. Comparison of L3MBTL3 SAM and SAMD1 SAM domain polymer. (**A**) Mammalian-two-hybrid using SAM domains of SAMD1, L3MBTL3, and SFMBT1 [6]. (**B**) Top and side views of the SAMD1–SAM pentamer (PDB: 6LUJ) [6]. (**C**) Top and side views of the L3MBTL3–SAM polymer (PDB: 8Y76). (**D**) Details of the hydrophilic interactions between two self-associated L3MBTL3 SAM domains. Two molecules are colored in green and blue, respectively. (**E**) Electrostatic surfaces of the EH and ML surfaces of L3MBTL3–SAM. (**F**) Details of hydrophobic contacts between EH and ML surfaces of L3MBTL3–SAM. (**G**) Structural comparison of L3MBTL3 SAM and SAMD1 SAM domain polymers. Left shows the overlapped structures of one molecule each from L3MBTL3 SAM (green) and SAMD1 SAM (gray). Right shows the closely associated molecules of L3MBTL3 SAM (blue) and SAMD1 SAM (gray) from their polymers. (**H**) Overlapped gel-filtration profiles of the wild-type L3MBTL3–SAM and its E733A/D736A mutant. Profiles of standard proteins Conalbumin and Ribonuclease A are shown as references.

adjacent molecule (Fig. 1F). To understand the difference between L3MBTL3 SAM and SAMD1 SAM polymers, we overlapped one molecule each from both polymers. We found that both molecules could be well superimposed, but their neighboring molecules were not well aligned (Fig. 1G). The neighboring molecule of L3MBTL3 SAM rotated by an angle of around 20 degrees relative to the corresponding molecule of SAMD1 SAM. The slight difference in orientation may explain why both polymers extend into different shapes. In solution, L3MBTL3 SAM appeared as a mixture of oligomers as shown from a gel-filtration column (Fig. 1H). When mutations that disrupted the hydrogen-bonding interactions were introduced, the E733A/D736A double mutant of L3MBTL3 SAM changed from an oligomer to a monomer (Fig. 1H), verifying the importance of the hydrogen-bonding interactions for the polymeric state of L3MBTL3 SAM.

Table 1. Crystallization details for L3MBTL3 SAM homopolymer and L3MBTL3–SAM/SAMD1–SAM dimer

	L3MBTL3–SAM (705–773)	**SAMD1–SAM (459–523, L495R)/ L3MBTL3–SAM (705–775, L756E/L760E)**
PDB ID	8Y76	8Y77
Beam line	SSFR-BL19U1	SSFR-BL18U1
Wavelength	0.97 853	0.97 915
Space group	$P6_5$	$C222_1$
Unit cell		
a, b, c (Å)	51.8, 51.8, 49.8	44.0, 95.5, 65.9
α, β, γ (°)	90.0, 90.0, 120.0	90.0, 90.0, 90.0
Resolution (Å)	50–1.99 (2.02–1.99)[a]	50–1.50 (1.53–1.50)
R_{merge}	0.137 (0.719)	0.106 (0.859)
$I/\sigma(I)$	30.6 (4.5)	23.2 (2.0)
Completeness (%)	100.0 (100.0)	99.8 (98.6)
Redundancy	19.6 (17.5)	12.6 (9.7)
No. reflections	5315	22 635
R_{work} / R_{free}	16.9/22.7	18.4/20.8
No. atoms		
Protein	548	1084
Water	55	179
B-factors (Å^2)		
Protein	24.4	16.2
Water	29.3	26.9
R.m.s. deviations		
Bond lengths (Å)	0.009	0.005
Bond angles (°)	0.780	0.749

[a]Values in parentheses are for highest resolution shell.

L3MBTL3–SAM ML surface interacts with the SAMD1–SAM EH surface

Considering the robust interaction between SAMD1 and L3MBTL3 via their SAM domains (Fig. 1A) [6], we also investigated the association of SAMD1–SAM and L3MBTL3–SAM. We found, that although both L3MBTL3–SAM and SAMD1–SAM can self-associate to form homopolymers, when mixed together, they can also associate with one another to form a heteropolymer (Supplementary Fig. S1A and B). As the complex of the wild-type L3MBTL3–SAM and SAMD1–SAM could not be crystallized, we supposed that the mixture of L3MBTL3–SAM and SAMD1–SAM was heterogeneous, as it could contain homopolymers and heteropolymers at the same time. We were then tempted to introduce mutations on either the ML or the EH surfaces of both SAMD1 and L3MBTL3 to disrupt the self-associated binding without disrupting the heterocomplex formation between both proteins. Through extensive analyses, we found that the EH surface mutant L756E/L760E of L3MBTL3–SAM and the ML surface mutant L495R of SAMD1–SAM can disrupt the self-association of both proteins, but would retain their ability to form a heterocomplex (Supplementary Fig. S1C–E). As a contrast, the ML surface mutant L742R of L3MBTL3–SAM could not form a heterocomplex with the wild-type SAMD1–SAM; similarly, the EH surface mutant L509E/L513E of SAMD1–SAM could not form a heterocomplex with the wild-type L3MBTL3–SAM, either (Supplementary Fig. S1F and G). This verified that L3MBTL3–SAM uses its ML surface to interact with the EH surface of SAMD1–SAM (Supplementary Fig. S1H).

Based on these analyses, we crystallized the complex of L3MBTL3–SAM bearing L756E/L760E mutations and SAMD1–SAM bearing L495R mutation and solved the complex structure at the resolution of 1.5 Å (Fig. 2A and Table 1). In the complex structure, one L3MBTL3–SAM molecule interacts with one SAMD1–SAM molecule. The acidic ML surface of L3MBTL3–SAM interacts with the basic EH surface of SAMD1–SAM (Fig. 2B), consistent with our previous analyses. In detail, residues Asp732, Glu733, Asp736, Asp747, and Leu742 from ML surface of L3MBTL3 form multiple hydrogen bonds with the residues Arg508, Leu509, Gly510, Lys514, and Arg498 from the EH surface of SAMD1 SAM (Fig. 2A). The binding between the SAM domains of L3MBTL3 and SAMD1 is also strengthened by hydrophobic interactions. The hydrophobic residues Leu509, Leu502, and Leu513 from the EH surface of SAMD1 stack with the hydrophobic residues of Ile751, Met752, Ile735, Met743, and Leu742 from the ML surface of L3MBTL3 (Fig. 2C).

L3MBTL3–SAM and SAMD1–SAM interact with each other through only one of their binding surfaces, while the other surface of both proteins still retains the self-association capability, indicating that both proteins can still self-associate from the other surface. To verify this hypothesis, we mixed the EH surface mutant of L3MBTL3–SAM (L756E/L760E) with one- and three-fold wild-type SAMD1–SAM, respectively. The L3MBTL3–SAM EH mutant (L756E/L760E) could not self-associate but could hetero-associate with SAMD1–SAM. Via gel-filtration, the heterocomplex peak shifted from a lower molecular weight position (at the molar ratio of 1:1) to a higher molecular weight position (at the molar ratio of 1:3, Supplementary Fig. S1I), indicating self-association of SAMD1–SAM from the other surface of the heterocomplex. Similar results were also seen when the SAMD1–SAM ML mutant (L495R) were mixed with wild-type L3MBTL3–SAM (Supplementary Fig. S1J).

To generate a model of self-association when both proteins form a heterocomplex, we superimposed one molecule each from the SAMD1–SAM pentamer and the spirally polymerized L3MBTL3–SAM with their corresponding SAM domains in the structure of the SAMD1–SAM/L3MBTL3–SAM binary complex, respectively (Fig. 2D). We found that the superimposed SAMD1–SAM pentamer has some clashes with the spirally winded polymer of L3MBTL3–SAM. Molecules 4 and 5 of the pentameric SAMD1–SAM would clash with the spirally extended L3MBTL3–SAM polymer. If only three SAMD1–SAM molecules self-associate, then SAMD1–SAM would not clash with the L3MBTL3–SAM polymer (Fig. 2D). This indicates that although the SAMD1–SAM and L3MBTL3–SAM interaction would partially disrupt the polymerization of both proteins, several copies of both proteins can still self-associate on both sides of the heterocomplex. These results are therefore consistent with the idea that SAMD1 and L3MBTL3 can establish heteropolymers.

Notably, ChIP-qPCR experiments demonstrated a robust dependency of L3MBTL3's chromatin binding on the presence of SAMD1 (Supplementary Fig. S2A) [6], underscoring the crucial nature of the SAM–SAM interaction between SAMD1 and L3MBTL3 for the chromatin association of the latter. The reciprocal scenario, wherein L3MBTL3 was deleted from the cellular milieu, resulted in slightly reduced chromatin binding of SAMD1 (Supplementary Fig. S2A) [6]. KO of KDM1A, another major interaction partner of SAMD1, leads to a stronger reduction of SAMD1 chromatin binding (Supplementary Fig. S2B). This suggests that the chromatin binding of SAMD1 at least partially depends on other chro-

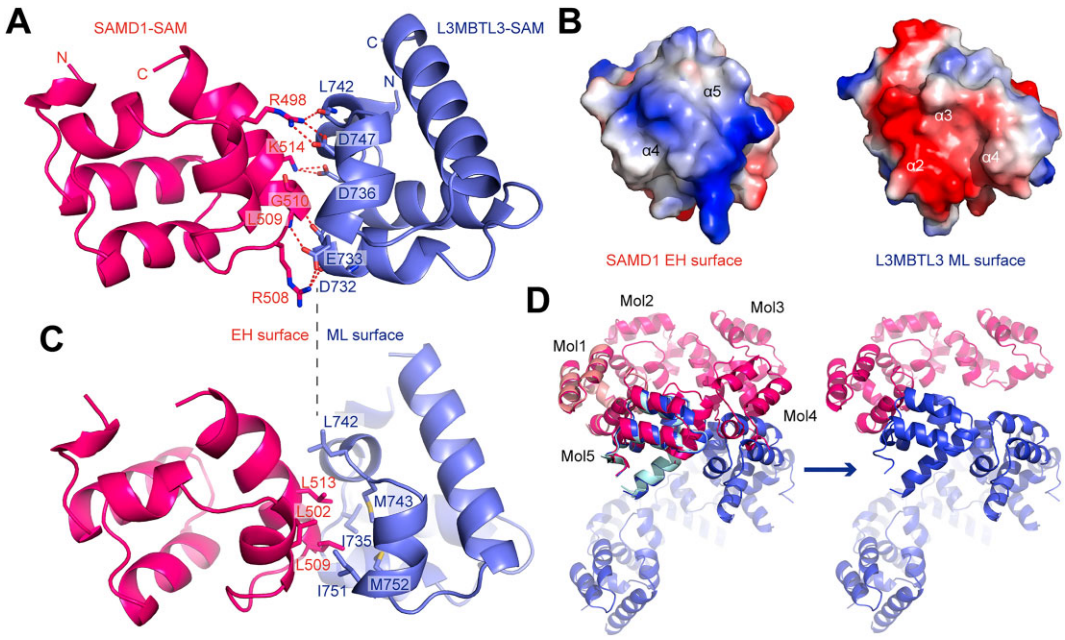


Figure 2. Structural details of the interaction between L3MBTL3–SAM and SAMD1–SAM. (**A**) Detailed hydrogen-bonding interactions between L3MBTL3–SAM and SAMD1–SAM (PDB: 8Y77). (**B**) Electrostatic representation of the binding surfaces of SAMD1–SAM and L3MBTL3–SAM. (**C**) Hydrophobic interactions between SAMD1–SAM and L3MBTL3–SAM. (**D**) A potential model of the wild-type L3MBTL3–SAM/SAMD1–SAM complex that contains the self-associated L3MBTL3–SAM and SAMD1–SAM polymers. To generate the model, one molecule of the SAMD1–SAM pentamer and one molecule of the L3MBTL3–SAM polymer were superimposed with their corresponding molecules in the binary complex of SAMD1–SAM/L3MBTL3–SAM.

matin associated proteins. On the other hand, mutation of SAMD1's WH domain or of the SAM domain strongly abrogated SAMD1's chromatin binding (Supplementary Fig. S2C). This observation supports the hypothesis that the direct interaction with the DNA via the WH domain and self-association ability via the SAM domain are both essential for the chromatin binding of SAMD1.

The WH and SAM domains are required but not sufficient for SAMD1's DNA-binding function

To investigate the requirements of SAMD1 ability to bind to DNA in further detail, we used a luciferase reporter construct driven by promoters containing CGIs (*Cbln1*, *Nanos1*) (Fig. 3A), which contain >20 CpG motifs and are strongly bound by SAMD1 in mouse ES cells [6]. An AlphaFold 3 model [27] suggests that these regions can theoretically be bound by several SAMD1 WH domains (Fig. 3B). Indeed, SAMD1 fused to an AD (AD-SAMD1) can strongly activate the reporter construct, suggesting that SAMD1 can directly bind to CGI-containing promoters (Fig. 3C). In contrast, AD-L3MBTL3 cannot activate the reporters, confirming the notion that L3MBTL3 is not able to directly bind to CGIs (Fig. 3C). Notably, other reporter constructs with lower numbers of CpG motifs, but similar lengths, are less efficiently activated by AD-SAMD1, supporting that SAMD1 requires a certain number or density of CpG motifs for efficient DNA binding (Supplementary Fig. S3A–D). This is also underlined by the observation that SAMD1 preferentially binds to larger CGIs in the genome (Supplementary Fig. S3E).

Importantly, the WH domain alone is not able to activate the *Cbln1* and *Nanos1* reporter (Fig. 3D), suggesting that the isolated DNA-binding module of SAMD1 is not sufficient for DNA binding *in vivo*. Further analyses showed that SAMD1 with mutated SAM or WH domains can also not activate the reporter construct (Fig. 3E), demonstrating that the WH and the SAM domains are both required for DNA binding of SAMD1. This result is consistent with our previous ChIP-qPCR results (Supplementary Fig. S2C) [6] and further supports the idea that beyond the DNA binding also the self-association via the SAM domain is important for efficient DNA binding of SAMD1. However, SAMD1 that lacks the region between the WH and the SAM domain (SAMD1 Δ101–449; SAMD1 ΔIDR) and only consists of the WH and the SAM domain is also not able to activate the reporter (Fig. 3F). This observation suggests that the WH and SAM domain are both required but not sufficient for chromatin binding of SAMD1. Consequently, we speculated that additional regions are essential for the DNA-binding function of SAMD1.

Analysis of SAMD1 utilizing the AlphaFold Database [31, 32] and Metapredict [33] indicates that SAMD1 possesses an extensive IDR between the N-terminal WH domain and the C-terminal SAM domain (Fig. 3G). A closer examination of the amino acid sequence unveils several distinctive features within this region. Specifically, the segment surrounding the WH domain (residues 1–26 and 101–211) displays a pronounced enrichment of proline and alanine residues, defining it as a PA-rich region (Fig. 3H). The region N-terminal to the SAM domain is notably abundant in prolines and glycines, forming a proline/glycine-rich (PG-rich) region (Fig. 3H). Notably, based on the DisProt database [34], prolines, alanines, and glycines are not commonly enriched in IDRs (Fig. 3H).

Low-confidence AlphaFold predictions further hint that the central portion of SAMD1 may harbor alpha helices and beta strands, suggesting the potential for secondary structure formation in this region. This putatively structured region is flanked by unstructured sequences rich in polar and

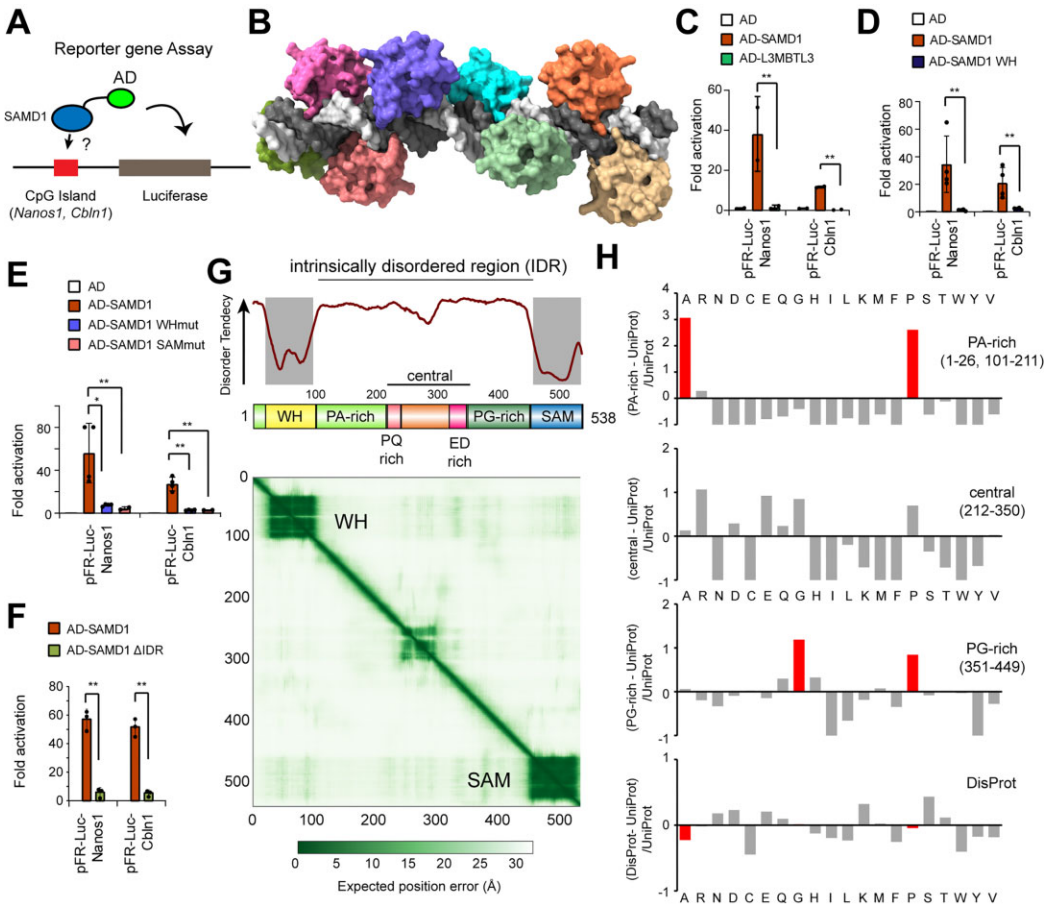


Figure 3. SAMD1 requires multiple features for DNA binding. (**A**) Schematic representation of the reporter gene assay. The luciferase gene is driven by CGIs-containing promoters from Cbln1 or Nanos1. The protein of interest is fused to an AD. This assay investigates the ability of the protein to bind to CGIs. (**B**) AlphaFold 3 model [27] of eight SAMD1 WH domains bound to a CpG-rich DNA sequence from the Cbln1 promoter. (C–F) Reporter gene assay with (**C**) AD-SAMD1 and AD-L3MBTL3, (**D**) the WH domain of SAMD1 compared with SAMD1 full length, (**E**) full-length SAMD1, either wild type or mutated within the WH domain (R45A/K46A) or SAM domain (L509A) [6], and (**F**) full-length SAMD1, or SAMD1 deleted for the region between the WH and the SAM domain (ΔIDR). (**G**) Structural composition (middle panel) of SAMD1 based on metapredict (upper panel) [33] and AlphaFold (lower panel) [31]. (**H**) Relative enrichment of amino acids in the PA-rich, central, and PG-rich regions of SAMD1 as well as of annotated IDRs (DisProt) [34] compared with the human UniProt database. The plots were created using an approach as described in [28]. Alanines, glycines, and prolines are marked red. Data in panel (C)–(F) represent mean ± SD of at least two independent experiments. Statistical analysis was performed using the unpaired two-tailed Student's *t*-test, comparing mutants versus wild-type SAMD1 (**$P < .01$; *$P < .05$).

acidic residues, denoted as proline/glutamine (PQ)-rich and glutamate/aspartate (ED)-rich sequences, respectively (Fig. 3G). These unstructured regions may functionally interact with the structured region interspersed between them. Collectively, we refer to this region as the "central" region. Currently, the relevance of these IDR features for the function of SAMD1 and whether they are involved in chromatin binding is unknown.

Full-length SAMD1 can form large polymers

Given that proteins with IDRs often form specific patterns in the cell [35], we conducted immunofluorescence experiments in HEK293 cells. Unexpectedly, the ectopic overexpression of full-length SAMD1, which increases its levels by ∼400-fold (Supplementary Fig. S4A), resulted not only in its expected localization to the nucleus but also in the formation of fibrous structures in the cytoplasm (Fig. 4A). This observation contrasts with L3MBTL3, which established nuclear speckles (Supplementary Fig. S4B), consistent with previous reports [36]. Notably, the fibrous pattern of SAMD1 was replicated with both untagged and GFP-tagged SAMD1 constructs (Supplementary Fig. S4C), excluding the possibility that the tag influenced this phenomenon. Consistent results were also observed with the mouse ortholog of SAMD1 (Supplementary Fig. S4D) and in alternate cell lines (Supplementary Fig. S4E). Time course experiments suggest that the fibers are fully established ∼48 h after transfection (Supplementary Fig. S4F).

The capacity of SAMD1 to form fiber-like structures was further confirmed by the wild-type full-length SAMD1 under EM (Fig. 4B). These observations suggest that full-length SAMD1 has the intrinsic ability to form large polymers, both *in vivo* and *in vitro*. However, under normal physiological conditions in the cell, endogenous SAMD1 primarily localizes to the nucleus without evident cytoplasmic fibrous patterns (Supplementary Fig. S4G). This suggests that the formation of these visible fibers by SAMD1 may require a certain threshold concentration within the cell, possibly achieved through overexpression. Indeed, weaker overexpression of SAMD1 via an alternative plasmid leads to less fibers (Supplementary Fig. S4H), demonstrating the importance of high SAMD1 protein level in the cells for polymer formation

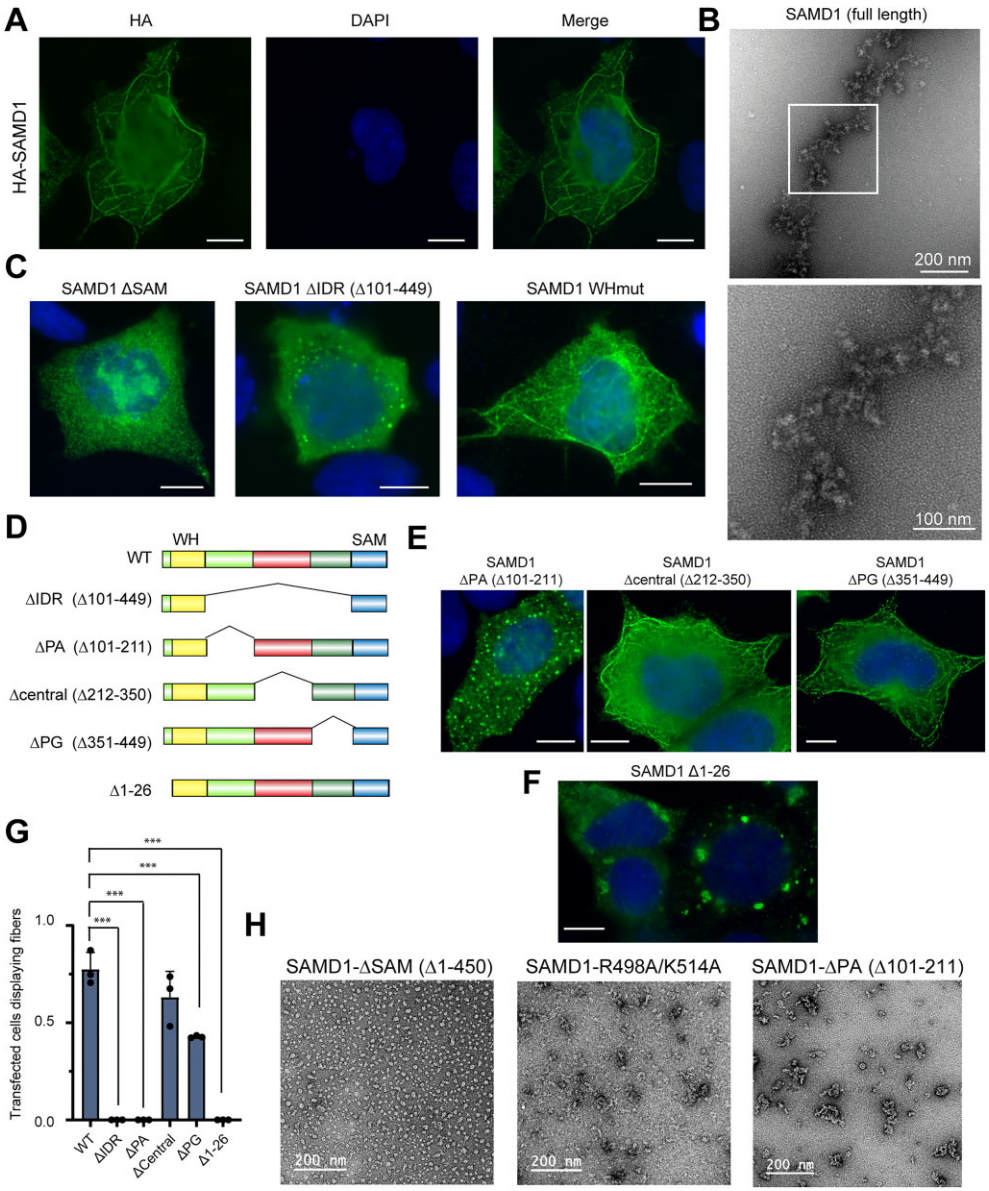


Figure 4. SAMD1 can polymerize *in vitro* and *in vivo*. (**A**) Representative immunofluorescence image of overexpressed SAMD1 in HEK293 cells. (**B**) Negative-staining EM image of full-length SAMD1. (**C**) Representative immunofluorescence image of overexpressed SAMD1 ΔSAM, SAMD1 ΔIDR, and SAMD1 WHmut (R45A/K46A) in HEK293 cells. (**D**) Schematic representation of deletion constructs used for experiments. FL = full length. (**E**, **F**) Representative immunofluorescence images of overexpressed SAMD1 IDR deletion constructs in HEK293 cells. (**G**) Quantification of fiber formation of panels (E) and (F). Shown is the relative number of transfected cells displaying fibers. Data represent mean ± SD of three biological replicates. Significance was evaluated using a two-way ANOVA with Dunnett's multiple comparisons test, comparing SAMD1 mutant constructs to wild-type (WT) SAMD1. (***$P < .001$; **$P < .01$; *$P < .05$). (**H**) Negative-staining EM of SAMD1 ΔSAM, SAMD1-R498A/K514A, and SAMD1 ΔPA. Scale bar in panels (A), (C), (E), and (F) 10 μm. Cells in panels (A), (C), (E), and (F) were stained with an anti-HA antibody (green) and DAPI (blue).

in the cytoplasm. Notably, a high proportion of ectopically overexpressed SAMD1 remains nuclear and associated with chromatin (Supplementary Fig. S4I). This suggests that only a fraction of overexpressed SAMD1 contributes to the formation of cytoplasmic fibers, which however becomes prominently visible. Importantly, when using SAMD1 without the SAM domain, we observed a diffuse pattern, while deletion of the entire IDR led to the formation of cytoplasmic foci (Fig. 4C). This observation suggests that SAMD1's ability to form cytoplasmic fibers depends on both the IDR and the SAM domain. In contrast, mutation of the WH domain of SAMD1, which abolishes SAMD1's DNA binding capacity [6], did not hinder fiber formation (Fig. 4C), suggesting that the DNA binding ability of SAMD1 is not a prerequisite for this phenomenon.

Considering the robust interaction between SAMD1 and L3MBTL3 via their SAM domains (Fig. 2) [6], we investigated whether their co-overexpression would influence each other's cellular localization. Remarkably, a co-localization of both proteins in cytoplasmic fibers can be observed in many cells where they were co-expressed (Supplementary Fig. S5A), implying the formation of heteropolymeric structures encompassing both SAMD1 and L3MBTL3. This co-localization required the presence of the SAM domains of both proteins (Supplementary Fig. S5B and C), emphasizing the pivotal role of these domains.

Together we concluded that although the observed fiber formation in the cytoplasm likely reflects a nonphysiological behavior, it represents an easy-to-use tool to assess the ability of SAMD1 to polymerize. Thus, we decided to make use of this feature to investigate which specific region of the IDR is required for polymerization and whether it would be important for the chromatin binding and biological function of SAMD1.

SAMD1 polymerization requires a PA-rich IDR

To probe the significance of distinct regions for SAMD1's ability to form polymeric structures, we generated multiple deletion mutants (Fig. 4D). Strikingly, deletion of the PA-rich region yielded a pattern resembling that of the entire IDR deletion (Fig. 4D, E, and G), a finding reproducible across different cell lines (Supplementary Fig. S6A). Similarly, deletion of the region N-terminal to the WH domain (Δ1–26) also abolished fiber formation (Fig. 4F and G). Conversely, deletion of both the central region and PG-rich region still permitted cytoplasmic fiber formation, albeit to a lesser extent (Fig. 4D, E, and G). Notably, deletion of the PG-rich region occasionally resulted in the appearance of interrupted fibers in some cells, suggesting its partial contribution to fiber stability. Deletions of smaller segments (∼25 amino acids) within the PA-rich region also reduces fiber formation (Supplementary Fig. S6B–D). These results underscore the essential need for the structural integrity of the entire PA-rich region to sustain SAMD1's polymerization capability.

The formation of long fibers was also not observed in the absence of the PA-rich region *in vitro* (Fig. 4H). However, SAMD1 lacking the PA-rich region still exhibited some degree of structure formation in this experiment (Fig. 4H), suggesting that its absence allows for partial polymerization. This contrasts to SAMD1 without SAM domain, which does not appear to form any multimeric structures (Fig. 4H). Taken together, our findings highlight the pivotal role of the PA-rich region, in addition to the SAM domain for SAMD1's optimal polymerization ability.

SAMD1's PA-rich region is unique and has gradually evolved

Proline-rich IDRs have been implicated in facilitating liquid–liquid phase separation [35, 37], suggesting a potential similar role for the PA-rich region of SAMD1. However, it is noteworthy that the PA-rich region of SAMD1 contains an exceptionally high proportion of proline (33.6%) and alanine (38.7%) residues, totaling 72.3% combined. When looking at a 100-amino acid window C-terminal to the WH domain (106–206), it is even 78.2%. Analysis of the UniProt database [38] demonstrated that this constitutes the highest percentage of prolines and alanines within such a window for all human proteins (Supplementary Fig. S7A and B). A similar high rank was found in other mammalian organisms (Supplementary Fig. S7C). Also, when looking at alanine and proline content separately, the PA-rich region ranks highly within the human proteome (rank 8 and 46 for alanine and proline, respectively). This high proline and alanine content is also uncommon for IDRs, as indicated by data from the DisProt database [34], where annotated IDRs exhibit an average of 8.9% prolines and 7.4% alanines. Intriguingly, the PA-rich region of SAMD1 appears to be evolutionarily recent, exclusively present in mammalian SAMD1 proteins (Supplementary Fig. S7D). An analysis across species revealed a gradual increase in the percentage of prolines and alanines within the disorder region of SAMD1, rising from ∼13% in fish to 40% in mammals (Supplementary Fig. S7E). At the same time, the overall content of prolines and alanines in the proteomes remains rather constant. This evolutionary trend suggests a potential advantage conferred by a higher proline and alanine content in SAMD1, hinting at its functional significance in the biological context.

In zebrafish, two variants of SAMD1, Samd1a and Samd1b, are present. While these variants exhibit highly similar WH and SAM domains compared with their human counterpart, they lack an evident PA-rich region (Supplementary Fig. S7D). Consistent with the notion that the PA-rich region plays a crucial role in polymerization, these zebrafish proteins fail to form fiber-like structures upon overexpression in HEK293 cells (Supplementary Fig. S8A). Samd1a predominantly localizes to the nucleus without displaying any discernible pattern, whereas Samd1b exhibits a cytoplasmic distribution and forms few conspicuous structures aside from larger foci (Supplementary Fig. S8A). When analyzing the chromatin-binding capacity of these proteins in mouse ES cells, we found that the zebrafish Samd1 proteins very inefficiently bind to mammalian chromatin, compared with human and mouse SAMD1 (Supplementary Fig. S8B). This may in part be due to weaker expression (Supplementary Fig. S8C) but could also be due to distinct functionality of the proteins. The WH domains of the zebrafish SAMD1 proteins are >75% identical to human SAMD1 (Supplementary Fig. S7D), thus likely having a similar DNA-binding function. This suggests that other differences may prevent efficient chromatin binding of zebrafish SAMD1 in mammalian cells. We speculated that the N-terminus of human SAMD1, containing the PA-rich region and the WH domain, may rescue the ability of the zebrafish SAMD1 to polymerize and to bind to chromatin. Thus, we created a hybrid protein that consists of the N-terminus of human SAMD1 and the C-terminus of zebrafish Samd1b (Supplementary Fig. S8D). In mouse ES cells, this hybrid protein is expressed at a similar level to zSamd1b (Supplementary Fig. S8E). Interestingly, the hybrid protein can establish fibers in the cytoplasm of HEK293 cells (Supplementary Fig. S8F) and can bind to chromatin in mouse ES cells (Supplementary Fig. S8G).

Overall, these findings support the hypothesis that the specific characteristics of the PA-rich region, surrounding the DNA binding WH domain, might have evolved in mammals to facilitate more efficient polymerization and chromatin binding.

SAMD1 chromatin binding involves the PA-rich region

Next, we sought to further investigate how the IDR region of SAMD1 might play a role in chromatin binding. Using the luciferase assay, described above, we observed that the deletion of both the PA-rich and PG-rich regions led to an impairment in the ability of the SAMD1 fusion protein to activate the reporter construct (Fig. 5A), suggesting a diminished efficiency of SAMD1 in binding to the CGIs. Conversely, deletion of the central domain exhibited minimal impact on reporter construct activation (Fig. 5A).

To investigate the impact of these deletions on chromatin binding, we expressed various SAMD1 constructs in mouse

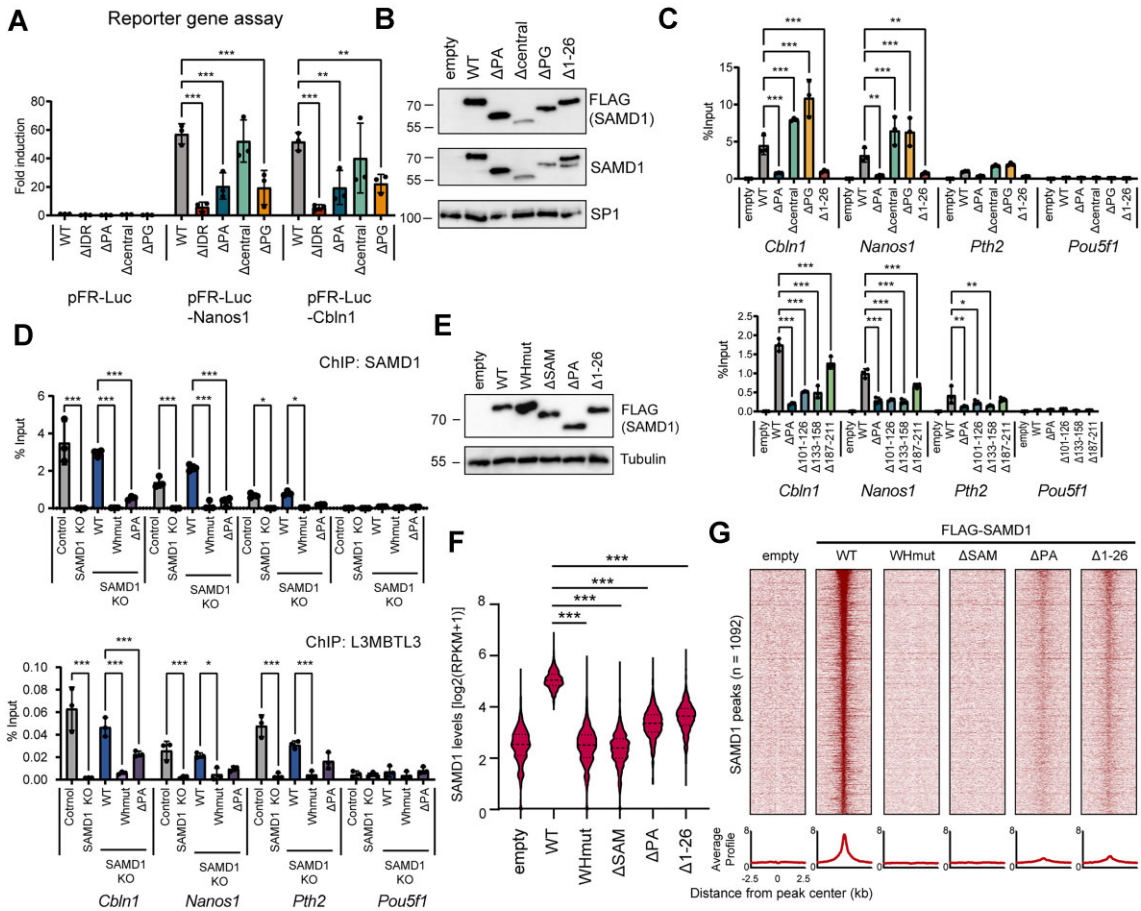


Figure 5. Role of PA-rich region for chromatin binding. (**A**) Reporter gene assay at CGI-driven luciferase genes (see Fig. 3A). (**B**) Western blot and (**C**) SAMD1 ChIP-qPCR of SAMD1 KO mouse ES cells [6] rescued with various SAMD1 constructs. (**D**) SAMD1 and L3MBTL3 ChIP-qPCR after SAMD1 rescue in SAMD1 KO mouse ES cells. (**E**) Western blot of SAMD1 KO mouse ES cells rescued with various SAMD1 constructs. (**F**) Quantification and (**G**) heatmap and profiles of SAMD1 ChIP-seq results at SAMD1-bound regions. See also Supplementary Fig. S9B. Statistical analysis in panel (F) was performed using a Kolmogorov–Smirnov test. Data in panels (A), (C), and (D) are represented as the mean of three biological replicates ± SD. The *P*-values were calculated using a two-way ANOVA with Dunnett's test for multiple comparison (***$P < .001$; **$P < .01$; *$P < .05$).

ES cells (Fig. 5B). Deletion of the PA-rich region led to an enhanced level of SAMD1 in the nucleoplasm, as assessed by fractionation experiments (Supplementary Fig. S9A), suggesting a reduced ability to bind to chromatin. To investigate this aspect on specific genomic loci, we performed extensive ChIP-qPCR experiments across several SAMD1 target genes. Consistent with the reporter assay and the fractionation experiment, we observed a substantial impact on SAMD1's chromatin binding when the PA-rich region is deleted (Fig. 5C). A reduced chromatin binding was also evident when removing the PA-rich region N-terminal to the WH domain or deleting smaller segments of the PA-rich region (Fig. 5C). In contrast to these observations, the deletion of the central or PG-rich region resulted in an increase in SAMD1's chromatin binding in this experiment (Fig. 5C). One possible explanation could be that negative regulators of SAMD1 bind to these regions. Their absence may then allow more efficient chromatin binding in the cellular context.

In our prior investigations, we underscored the significance of SAMD1 as a pivotal factor in recruiting L3MBTL3 to chromatin (Supplementary Fig. S2A) [6]. To discern the specific role of the PA-rich region in facilitating L3MBTL3's chromatin binding, we restored SAMD1 KO mouse ES cells with either wild-type SAMD1, a WH-mutant SAMD1, or SAMD1 lacking the PA-rich region (Fig. 5D). Both mutant proteins retained the SAM domain, crucial for interacting with L3MBTL3 (Fig. 1A and 2) [6]. We observed that while the wild-type SAMD1 effectively restored L3MBTL3 chromatin binding (Fig. 5D), neither the WH mutant SAMD1 nor the SAMD1 lacking the PA-rich region could achieve this rescue (Fig. 5D). This trend closely paralleled SAMD1's own chromatin-binding capability (Fig. 5D). These findings suggest that the PA-rich region not only directly contributes to SAMD1's chromatin binding but also indirectly influences the recruitment of SAMD1's interaction partners, including L3MBTL3.

Together these data further support the involvement of the PA-rich region for efficient chromatin binding of SAMD1.

Efficient binding of SAMD1 to CGIs requires the WH domain, the SAM domain, and the PA-rich region

To scrutinize the impact of specific domains on chromatin binding on a genome-wide level, we conducted ChIP-seq experiments in mouse ES cells, where SAMD1 binds to unmethylated CGIs [6]. We introduced various FLAG-tagged SAMD1

variants in SAMD1 KO mouse ES cells, to eliminate interference with the endogenous SAMD1 protein (Fig. 5E). In alignment with our previous findings [6], SAMD1 proteins lacking the SAM domain or harboring a mutated WH domain exhibited complete abolishment of chromatin binding (Fig. 5F and G and Supplementary Fig. S9B). This reaffirms the indispensability of the SAM domain and the WH domain for SAMD1's chromatin binding. Notably, upon deletion of the PA-rich region, either N-terminal (1–26) or C-terminal (101–210) to the WH domain, chromatin binding was significantly reduced, though not entirely eliminated (Fig. 5F and G and Supplementary Fig. S9B). This underscores the importance of the PA-rich regions for optimal chromatin binding of SAMD1, while highlighting their lower significance compared with the WH domain and the SAM domain.

The specific amino acid sequence of the PA-rich region is not essential for polymerization and chromatin binding

To better understand the role of the PA-rich region, we conducted an in-depth exploration of this region. Alongside its abundance in hydrophobic amino acids [46 prolines (P), 53 alanines (A), 6 glycines (G), and 2 leucines (L)], this region also harbors a subset of charged or polar amino acids, including 4 serines (S), 6 threonines (T), 2 glutamines (Q), 2 glutamates (E), and 12 arginines (R) (Fig. 6A). These amino acids could potentially establish interactions between SAMD1 molecules via charged interactions or hydrogen bonds.

Initially, we aimed to discern whether the specific sequence of the PA-rich region holds significance, given that short motifs in the sequence may be relevant for protein–protein interactions [39]. We examined the consequences of shuffling the amino acid sequences, before and after the WH domain (Fig. 6A), creating three different SAMD1 versions (shuffle #1–#3). Surprisingly, SAMD1 with a shuffled PA-rich regions still led to fiber-like structures upon overexpression (Fig. 6D and E), suggesting that the exact amino acid sequence is not crucial for SAMD1's polymerization ability. Instead, we hypothesized that an interplay of hydrophobic and charged amino acids within the PA-rich region, independent of the specific sequence, could be pivotal for stabilizing SAMD1 polymers.

To further investigate the relevance of the prolines and alanines within this region, we replaced them with serines and glycines (A > GS and P > GS, respectively). Glycines and serines tend to interact minimally with other amino acids [40], making them suitable replacements to assess the importance of specific amino acids. Interestingly, these mutations retained the ability of SAMD1 to form the cytoplasmic fibers (Fig. 6D and E). This observation suggests that the prolines and alanines per se are not essential for the fiber formation of SAMD1. In ChIP-qPCR experiments, we found that SAMD1 with A > GS and P > GS mutations still allow chromatin binding (Fig. 6G). However, in this experiment, we observed slightly different chromatin-binding capacities of the three shuffled mutants (Fig. 6G). This observation suggests that although the specific amino acid sequence appears not to be essential for the polymerization ability, the correct positioning of certain amino acids within the PA-rich region could be necessary for optimal chromatin binding.

Arginines within the PA-rich region are involved in polymerization and chromatin binding of SAMD1

To assess the importance of the non-hydrophobic amino acids in the PA-rich region, we mutated these amino acids into alanines (STEQR > A) (Fig. 7A). Strikingly, this alteration abolished fiber formation (Fig. 7B and C), indicating the indispensability of at least some of these amino acids for SAMD1 polymerization. We then individually mutated these amino acids into alanines. Mutations of threonines and serines to alanines (ST > A) and glutamines and glutamic acids to alanines (EQ > A) had minimal effects on fiber formation (Fig. 7B and C). However, mutating nine arginines to alanines (9R > A) resulted in the abrogation of fiber formation (Fig. 7B and C), mirroring the phenotype observed with the STEQR > A mutant version. Analysis of the chromatin-binding capacity of these mutants showed that both the STEQR > A and the 9R > A mutant SAMD1 have reduced chromatin-binding capacity, while both the EQ > A and ST > A mutant still bind robustly to chromatin (Fig. 7D). This observation suggests that the arginines are the most relevant amino acids in the PA-rich region for both polymerization and chromatin binding, while other non-polar amino acids have a lesser impact. Arginines are implicated in charged–charged interactions and are crucial for forming hydrogen bonds and cation–pi interactions [41–43], potentially explaining their significant impact. Notably, both the 9R > A and STEQR > A mutants displayed reduced migration in SDS–PAGE (Fig. 7E), suggesting alterations in the protein's properties.

To investigate the roles of the arginines in more depth, we created two additional mutants with only six and three mutated arginines (6R > A, 3R > A) (Fig. 7A). We also mutated the arginines to lysines, which retains the positive charge, but allows less hydrogen bonds (Fig. 7A). These mutants showed a weaker shift in the migration behavior in SDS–PAGE, compared with the 9R > A mutant (Fig. 7G). In all three cases, we found strongly impaired cytoplasmic fiber formation (Fig. 7F and H), further supporting that the arginines are important for the function of the PA-rich region for SAMD1 polymerization. Surprisingly, however, in ChIP-qPCR experiments, the 3R > A mutant as well as the 9R > K mutant still show robust chromatin binding, supporting that a certain level of positive charges within the IDR are important for the chromatin binding of SAMD1 (Fig. 7I).

For most PA-region mutants, the ability to form fibers and bind chromatin is well-aligned (Fig. 7J). There are only two notable exceptions: (i) Shuffle mutants #2 and #3 can form cytoplasmic polymers but exhibit reduced chromatin-binding capacity, and (ii) the 3R > A and 9R > K mutants fail to form fibers in the cytoplasm yet retain the ability to bind chromatin (Fig. 7J). These exceptions suggest that the role of the PA-rich region in polymerization and chromatin binding may involve partially distinct mechanisms. For most mutants, both functions are either preserved or disrupted, resulting in similar effects on polymerization and chromatin binding. The exceptions highlight cases where one function is possibly more strongly impaired than the other, leading to these discrepant outcomes.

Overall, these results support the hypothesis that the arginines within the PA-rich region exert a pronounced influence on SAMD1's functionality. Notably, in contrast to the prolines and alanines, the amount of arginines remained relatively constant upon evolution (Supplementary Fig. S7E and F).

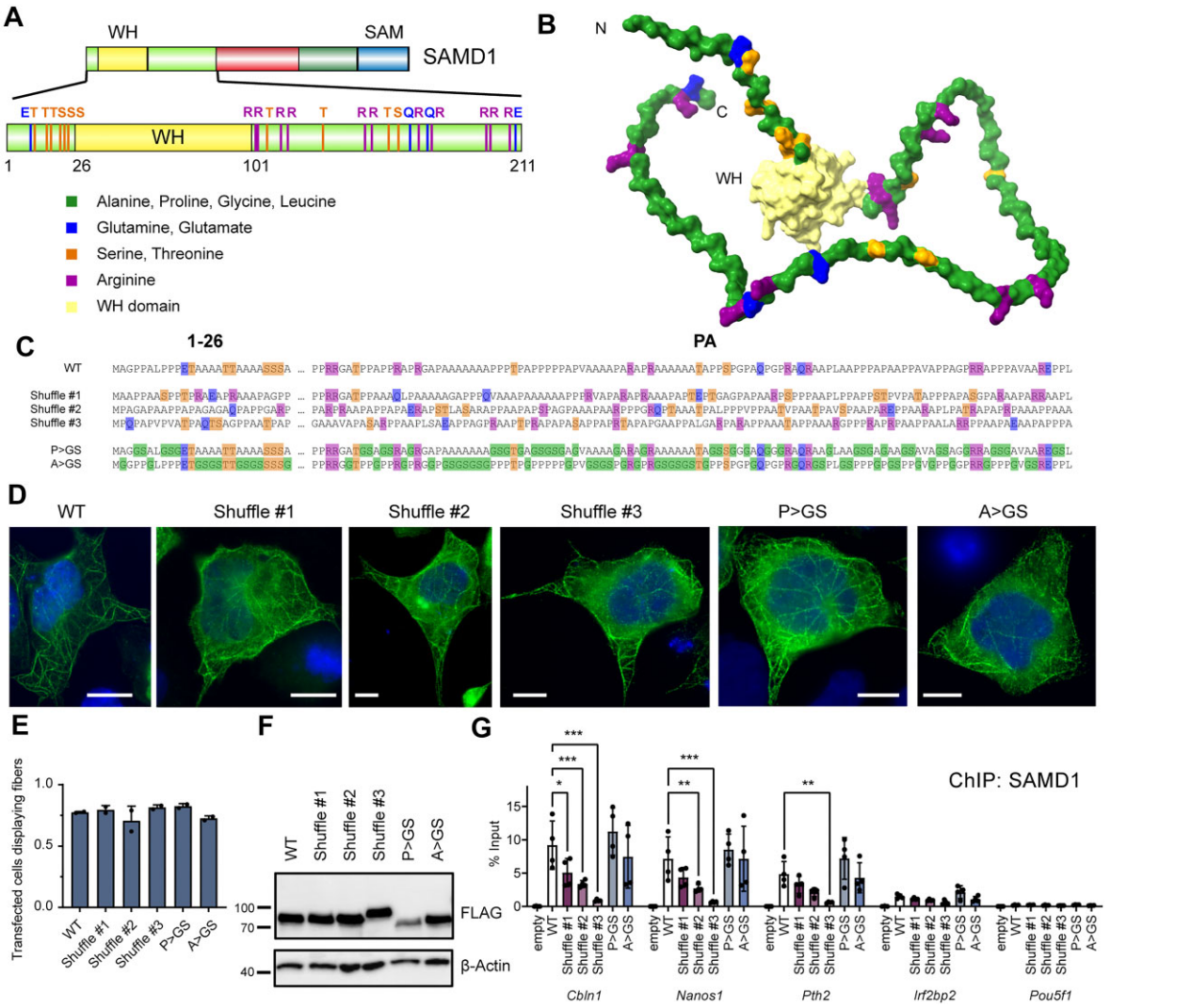


Figure 6. The specific sequence of the PA-rich region appears to be of low relevance. (**A**) Schematic overview of the PA-rich region surrounding the WH domain. Charged and polar amino acids (15%) are marked in blue (Q/E), orange (S/T), and red (R), respectively. All other amino acids in this region are hydrophobic and are marked green (P/A/G/L; 85%). (**B**) AlphaFold-predicted structure [31] of the PA-rich region surrounding the WH domain (yellow). The amino acids are colored as in panel (A). (**C**) Sequences of wild-type PA-rich region and its modifications, used for the experiments. (**D**) Representative immunofluorescence images of FLAG-SAMD1 with a shuffled PA-rich region of SAMD1 or SAMD1 in which all charged and polar amino acids in the PA-rich region are mutated to alanine (STEQR > A). Cells were stained with an anti-FLAG antibody (green) and DAPI (blue). Scale bar 10 μm. (**E**) Quantification of immunofluorescence experiments in panel (D), compared with wild-type SAMD1. Data represent mean ± SD of two biological replicates. (**F**) Western blot of SAMD1 constructs upon overexpression in HEK293 cells. (**G**) ChIP-qPCR experiments of overexpressed SAMD1 in mouse ES cells with indicated modifications. Data represent mean ± SD of four biological replicates. Statistical analysis was performed using a two-way ANOVA with Dunnett's multiple comparison test, comparing mutant constructs with wild-type (WT) SAMD1 (***$P < .001$; **$P < .01$; *$P < .05$).

Mutations of arginines within the PA-rich region impair the biological function of SAMD1

To ascertain the significance of arginines for SAMD1's biological function, we leveraged pancreatic cancer cell lines, specifically PaTu8988t cells, where SAMD1 KO led to markedly increased migration (Fig. 8A and B) [9]. This migratory phenotype can be rescued by inducing nuclear localization of an ER-SAMD1 fusion protein via tamoxifen (4-OHT) [9]. Crucially, the 9R > A mutant, though comparably expressed to the wild-type (Fig. 8C), failed to rescue this phenotype (Fig. 8D and E).

To delve into the underlying molecular mechanisms, we examined SAMD1's direct target genes, *CDH2* and *L3MBTL3*, which exhibit upregulation upon SAMD1 deletion in PaTu8988t cells [9]. *CDH2* emerged as a pivotal downstream target of SAMD1, driving the observed enhanced migration [9]. The L3MBTL3 protein directly interacts with SAMD1 (Figs 1A and 2) [6], but its gene is also a common downstream target in various cell lines [6, 9, 10]. We observed a significant reduction in chromatin binding of SAMD1 and recruitment of L3MBTL3 at these genes with the 9R > A mutant version of SAMD1 (Fig. 8F). Similarly, transcription levels of these genes were less effectively rescued with the 9R > A mutant (Fig. 8G). These findings underscore the critical role of arginines within the PA-rich region for chromatin binding and the biological function of SAMD1.

In summary, SAMD1's capacity to bind chromatin and to fulfill its biological function hinges not only on the DNA binding WH domain, and the SAM domain, but also necessitates its PA-rich region.

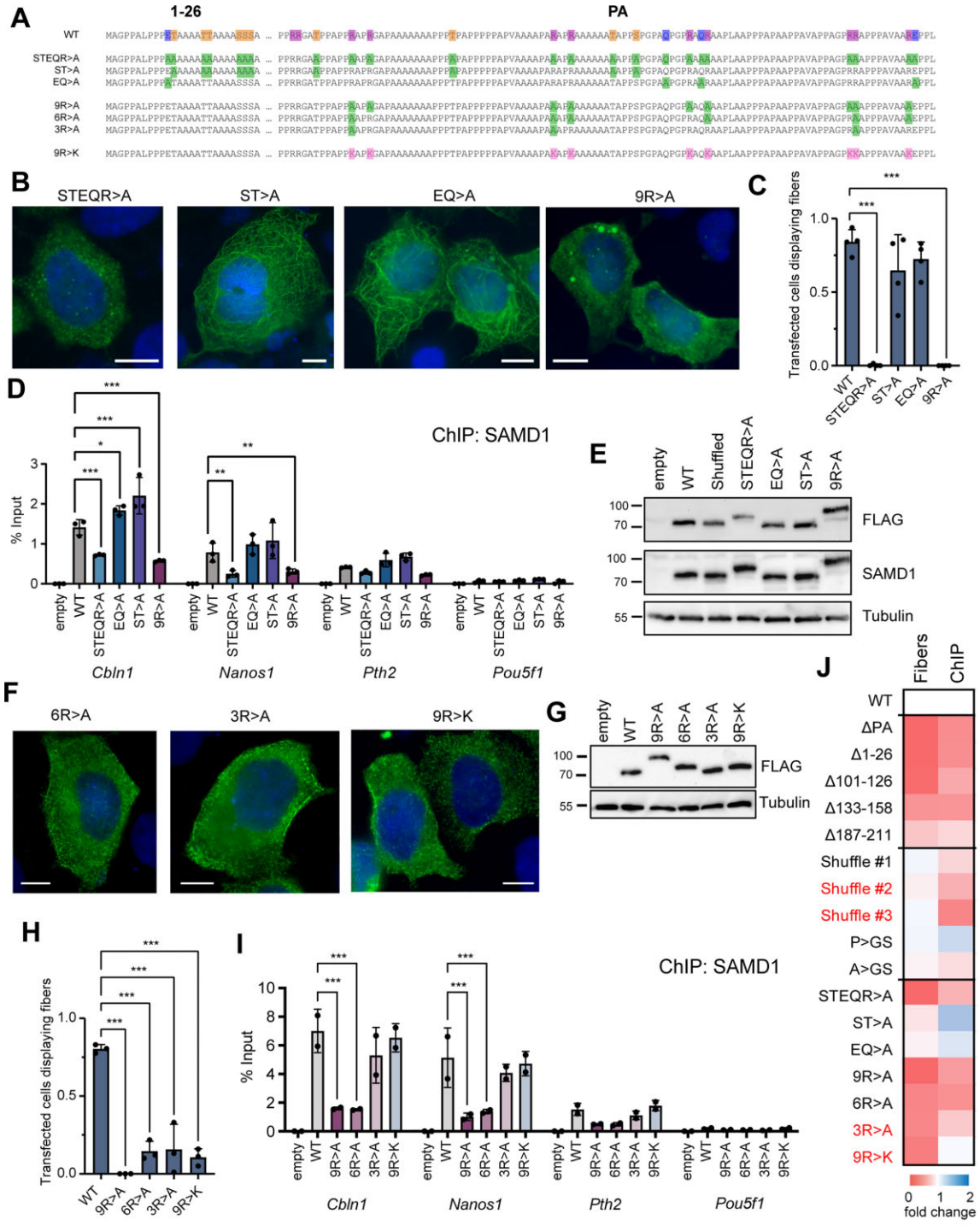


Figure 7. Arginines within the PA-rich region of SAMD1 are involved in SAMD1 polymerization and efficient chromatin binding. (**A**) Sequences of modified PA-rich regions of SAMD1 used for experiments. (**B**) Representative immunofluorescence images of in HEK293 cells ectopically expressing FLAG-tagged SAMD1 in which serine/threonines (ST > A), glutamines/glutamatic acids (EQ > A), or nine arginines (9R > A) within the PA-rich region were mutated to alanines. (**C**) Quantification of panel (B). (**D**) SAMD1 ChIP-qPCR and (**E**) western blot of SAMD1 KO mouse ES cells rescued with indicated SAMD1 mutants. (**F**) Representative immunofluorescence images of HEK293 cells transiently overexpressing SAMD1 mutants in which varying amount of arginines within the PA region have been mutated to alanines or lysines. (**G**) Western blot of SAMD1 KO mouse ES cells rescued with indicated SAMD1 arginine mutant constructs. (**H**) Quantification of panel (F). (**I**) SAMD1 ChIP-qPCR of mouse ES cells rescued with indicated SAMD1 arginine mutant constructs. (**J**) Overview of influence of PA-region mutants on fiber formation ("Fibers") and chromatin binding ("ChIP"), based on results from Figs 4–7 and Supplementary Fig. S6. "Fibers" is the mean relative number of transfected cells displaying fibers upon transient overexpression of the given construct in HEK293 cells with the value for full-length SAMD1 set as 1. "ChIP" refers to SAMD1 ChIP experiments upon overexpression in mouse ES cells. The values reflect the mean relative ChIP-qPCR signal of mutant SAMD1 versus wild-type SAMD1 at the Cbln1, Nanos1, and Pth2 promoter. Data in panels (C), (D), and (H) represent mean ± SD of at least three biological replicates. Data in panel (I) represent mean ± SD of two replicates. Statistical analysis was performed using a two-way ANOVA with Dunnett's multiple comparison test, comparing mutant constructs with wild-type (WT) SAMD1. (***$P < .001$; **$P < .01$; *$P < .05$). Scale bar in 10 μm in panels (B) and (F). Cells in panels (B) and (F) were stained with an anti-FLAG antibody (green) and DAPI (blue).

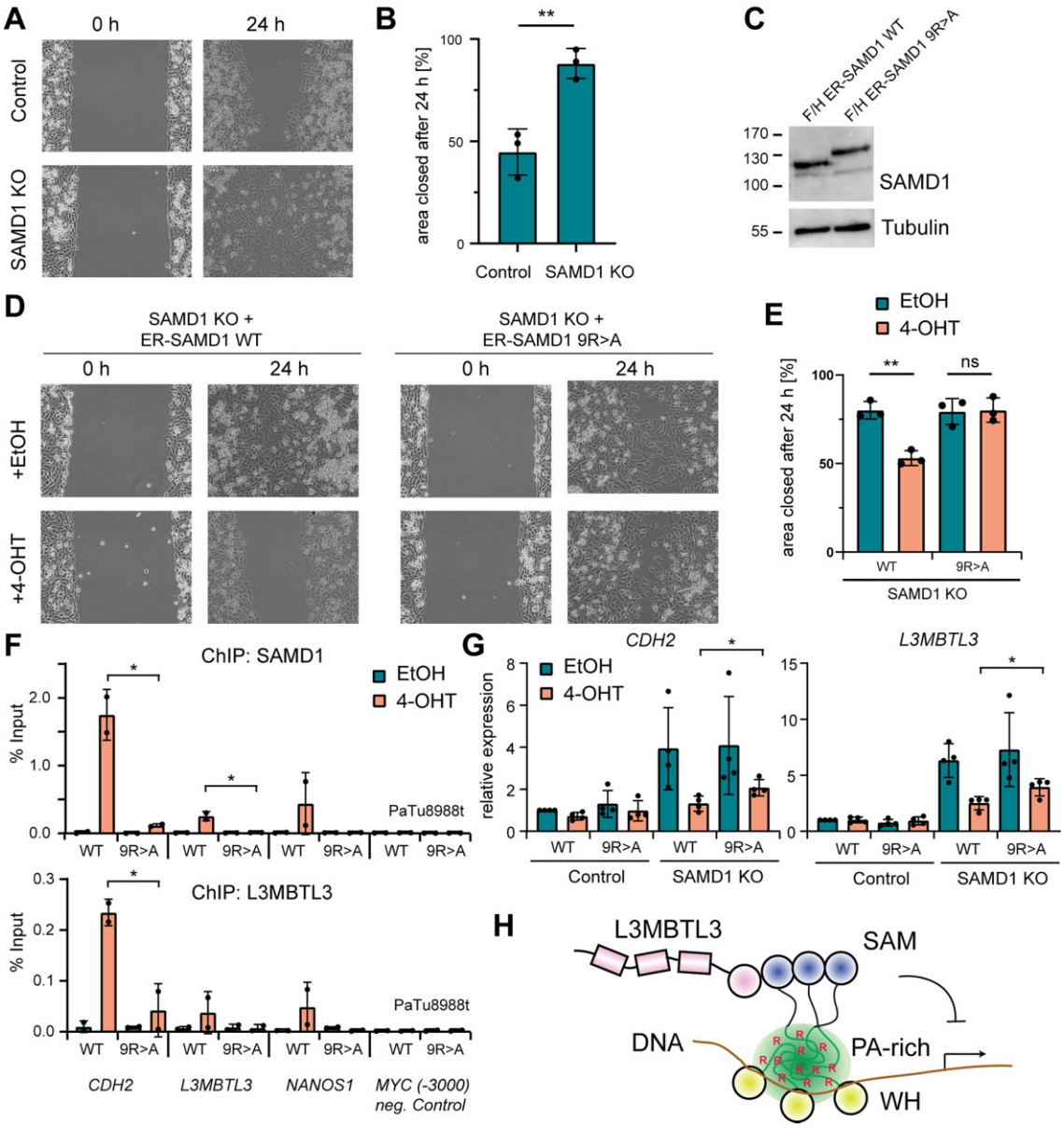


Figure 8. Mutation of arginines within the PA-rich region impairs SAMD1's biological function. (**A**) Wound healing assay in PaTu8988t cells in wild type and SAMD1 KO cells [9]. (**B**) Quantification of the wound healing assay from panel (A). (**C**) Western blot of rescue of the PaTu8998t SAMD1 KO cells with FLAG-HA (F/H) ER SAMD1 wild type or R > A mutant. (**D**) Wound healing assay using rescue cells from panel (C). The nuclear localization of FLAG-HA-ER-SAMD1 is triggered by the addition of tamoxifen (4-OHT). (**E**) Quantification of experiment from panel (D). (**F**) ChIP-qPCR using antibodies for SAMD1 and L3MBTL3 at SAMD1 target genes. (**G**) RT-qPCR of two SAMD1 target genes [9] in CRISPR control and SAMD1 KO PaTu8988t cells, expressing FLAG-HA-ER-SAMD1 wild type or R > A mutant. (**H**) Model of SAMD1 chromatin binding. Aside the DNA binding via the WH domain, SAMD1 requires its PA-rich region, including its arginines, for efficient chromatin binding. This region may be important to establish direct contact with DNA but may also be involved to enhance polymerization of SAMD11. Data in panels (B), (E), (F), and (G) represent mean ± SD of (panels B and E) three, (panel F) two, and (panel G) three biological replicates. Statistical analysis was performed using the two-sided unpaired *t*-test (***$P < .001$; **$P < .01$; *$P < .05$).

Discussion

Transcription factors play a pivotal role in deciphering genomic information and orchestrating gene transcription [1]. Typically, these proteins leverage specific DNA-binding domains to facilitate their interaction with genomic targets. Collaboration with other transcription factors and chromatin regulators is a common strategy employed by these molecules in their quest to locate and regulate genes [2]. While some transcription factors, such as ETV6, have been suggested to enhance their chromatin binding through homopolymerization [4, 44], the intricacies of how polymerization contributes to their chromatin-related functions remain elusive. This article sheds light on SAMD1, an atypical transcription factor that relies on the combined synergy of a DNA-binding domain, a polymerizing SAM domain, and an IDR for efficient chromatin binding. This novel perspective on SAMD1 expands our understanding of the diverse mechanisms employed by transcription factors in gene regulatory processes.

The SAM domain, a ubiquitous protein domain found in diverse cellular proteins within both the cytoplasm and nucleus [45], exhibits a remarkable capacity for interactions, both with other SAM domains and itself [30]. The versatility

of SAM domains is underscored by the formation of distinct polymer structures, ranging from dimers to open spiral-like and closed polymers [4, 6, 17, 18, 46]. SAM domain polymers have been postulated to play a pivotal role in subnuclear clustering and the recruitment of the polycomb repressive complex 1 [46–48]. Moreover, SAM polymerization has been associated with the chromatin dynamics of ETV6, contributing to its spreading on chromatin and facilitating transcriptional repression [4, 44]. Similarly, in plants, the SAM domain polymerization of the transcription factor LEAFY is required for chromatin binding and flower development [49].

To delve deeper into the intricate roles of SAM domains in chromatin regulation, our investigation initially focused on deciphering the structural nuances of the SAM domains within SAMD1 and L3MBTL3, two proteins that exhibited robust mutual association [6]. Surprisingly, our findings unveiled striking disparities in their *in vitro* polymerization behavior. The SAMD1 SAM domain demonstrated the formation of a closed pentameric ring [6], in stark contrast to the L3MBTL3 SAM domain, which adopted an open spiral structure reminiscent of ETV6, the polyhomeotic proteins and LEAFY (Fig. 1) [4, 17, 49]. This structural divergence hinted at distinct functional roles for these two domains. This notion gained further support from our observations of disparate cellular outcomes upon ectopic expression. While L3MBTL3 expression led to the formation of nuclear foci (Supplementary Fig. S4B), SAMD1 manifested as cytoplasmic fibers (Fig. 4A). These fibers likely reflect a nonphysiological assembly of SAMD1 in the cells but offer an attractive tool to investigate the polymerization ability of SAMD1. Notably, Cryo-EM experiments with purified SAMD1 corroborated the intrinsic ability of SAMD1 to form larger polymers (Fig. 4B).

A pivotal revelation emerged as we probed the molecular requirements for SAMD1's polymerization ability. Beyond the polymerizing SAM domain, SAMD1 relies on an extensive PA-rich IDR for this unusual polymerization capacity (Fig. 4E and F). Prolines have been implicated to play an important role in IDRs as disorder-promoting residue [50]. Unexpectedly, the observed fiber formation in cytoplasm does not appear to require a specific sequence, nor the high content of prolines and alanines, as shuffled sequences and sequences where the prolines and alanines are replace by more neutral glycines and serines still allow polymerization (Fig. 6). Instead, we found that small deletions (Fig. 4 and Supplementary Fig. S6), as well as mutations of arginines (Fig. 7) have the most disruptive effect on SAMD1's ability to establish cytoplasmic fibers. This observation suggests that the length of the IDR as well as the arginines are the most critical features. The length of the IDR could be important to establish a certain distance between the WH domain and the putatively structured middle region. Given that the exact function of the middle region is currently unclear, it remains to be clarified how exactly the spacing between the WH domain and the middle domain could contribute to efficient chromatin binding. Arginine has a positive charge that is delocalized over a planar guanidinium group, allowing the formation of simultaneous hydrogen bonds with multiple acceptors and or is often critical for the function of IDRs [41]. This feature may allow intermolecular interactions and may therefore support the formation of SAMD1 polymers. We found that mutation of the arginines to lysines prevents cytoplasmic fiber formation, but still allows chromatin binding, while mutation of arginines to alanines prevents both (Fig. 7F, H, and I). This observation suggests that the the arginines may not only contribute to SAMD1's self-association, but their positive charge may also be involved to directly contact DNA, as proposed for other IDRs [51].

In absence of the PA-rich region, we often observe cytoplasmic foci, similar to liquid–liquid phase separated membraneless organelles (Fig. 4E) [52]. Currently, the nature of this foci is unclear. However, given that SAMD1 without the PA-rich region still allows the formation of smaller complex in EM (Fig. 4H), it suggests that SAMD1 can still form some kind of polymers via its SAM-domain that may cluster together in these foci. SAMD1 without the SAM domain is not forming larger cytoplasmic foci (Fig. 4C), supporting the notion that the SAM domain is important for this observation. In the future, it will be of interest to better understand whether these foci colocalize with other features, and why they are only visible in the cytoplasm.

Using various SAMD1 mutants, we found that the polymerization ability in the cytoplasm of SAMD1 correlates in most cases with its chromatin-binding ability (Fig. 5). This observation prompts the hypothesis that SAMD1 necessitates the formation of multimers for chromatin binding and gene repression (Fig. 8H). This is perhaps not unexpected, given that the DNA binding affinity of the WH domain is relatively low [6]. Polymerization would allow SAMD1 to synergistically engage with multiple CpG motifs at CGIs, thereby strongly amplifying its binding affinity. Supporting this hypothesis, SAMD1 exhibits a preference for binding to larger CGIs, characterized by a higher density of CpG motifs (Supplementary Fig. S3E). Importantly, SAMD1 without the PA-rich region and the region N-terminal to the WH domain, still shows some residual chromatin binding in ChIP-seq experiments (Fig. 5G). This contrasts to SAMD1 without the SAM domain, which is absent from chromatin (Fig. 5G). We speculate that SAMD1 with impaired PA-rich region can, in contrast to SAMD1 without SAM domain, still form some smaller multimers, thus allowing weak chromatin binding (Fig. 8H). Indeed, EM results indicate that SAMD1 can still multimerize even without the PA-rich region, as shown by larger clusters (Fig. 4H).

Notably, in some cases, we observe robust chromatin binding but no cytoplasmic polymerization, and vice versa (Fig. 7J), implying that the function of the IDR for polymerization and chromatin binding is not entirely overlapping. Further work will be required to elucidate the specifics of these potentially independent functions.

The potential multimeric binding of SAMD1 to chromatin also raises the question of whether the chromatin association of SAMD1 can be finely regulated by modulating its polymerization capacity. The SAM domain stands out as a potential regulatory target. Specifically, lysine residues within this domain may undergo ubiquitination, as indicated by the PhosphositePlus database [53], introducing a disruptive element to the SAM–SAM interactions. Furthermore, the PA-rich region presents another avenue for regulation. Arginines within this region may undergo methylation by protein arginine methyltransferases [53, 54], a modification known to disrupt hydrogen-bond interactions [52, 55, 56]. Phosphorylation events on serines and threonines within the PA-rich region also emerge as plausible regulatory checkpoints [56]. The high amount of prolines within the PA-rich region may additionally be subject to regulation by peptidyl-*cis*/*trans* isomerases [57, 58]. These regulatory events likely modulate SAMD1's polymerization ability, further fine-tuning its chromatin-binding dynamics.

SAMD1's polymerization ability likely not only plays a pivotal role in chromatin interaction but also could have broader implications, potentially facilitating the interconnection of spatially distant promoters containing CGIs. This mechanism bears significance akin to the proposed concept of transcription factories [59], suggesting a role in the coordinated regulation of genes. Conversely, aberrations in SAMD1 polymerization could potentially contribute to pathological conditions, such as neurological diseases [12]. Therefore, it becomes imperative for future investigations to delve into the regulatory pathways governing SAMD1 polymerization and its influence on spatial genome organization. Unraveling these intricacies is crucial for a comprehensive understanding of the biological significance of SAMD1 in cellular processes and disease contexts.

IDRs can be found in the majority of proteins, but their specific functions are often unclear [35]. In the context of transcription factors, IDRs have been proposed to establish phase-separated condensates for gene activation [60] or to direct the chromatin binding via multiple weak interactions with the DNA [51]. For SAMD1, a similar model emerges: the IDR may be important for intermolecular interactions important to establish multi- or polymeric complexes on the chromatin. This in turn, provides the ground for multivalent association of the DNA-binding domain with DNA (Fig. 8H). In addition, the positive charge of the arginines within the PA-rich region may further increase DNA interactions. Such multivalent chromatin-binding mechanism may particularly be relevant for transcription factors for which the DNA-binding domain alone is not sufficient for chromatin interaction [61, 62]. Our findings underscore the inadequacy of solely relying on DNA-binding domains for chromatin interaction.

This study has several limitations. The structural studies utilized isolated SAM domains for both SAMD1 and L3MBTL3. It remains unclear whether the spiral-like structure of L3MBTL3–SAM and the pentameric ring of SAMD1–SAM are also formed by the full-length L3MBTL3 and SAMD1 proteins, respectively. In the context of the full-length proteins and/or *in vivo*, the formation of polymers may be influenced by additional factors, potentially leading to alternative compositions not fully represented by the crystal structures. Additionally, the ability of SAMD1 to form cytoplasmic fibers likely occurs only under conditions of overexpression in cells. It would be interesting to determine whether such SAMD1 fibers can also exist under physiological or pathological conditions. Nevertheless, the markedly reduced chromatin binding in the absence of the SAM domain and the PA-rich region supports the importance of SAMD1 multimerization for chromatin binding, beyond the DNA-binding WH domain.

In summary, our study has unveiled a molecular mechanism that integrates the functionalities of a DNA-binding domain, a polymerizing SAM domain, and an IDR, that are collectively involved to facilitate chromatin binding (Fig. 8H). This discovery expands the repertoire of potential mechanisms through which transcription factors enhance their binding to specific target loci, offering a novel perspective on the multifaceted orchestration of molecular events involved in chromatin regulation.

Acknowledgements

We thank Arne Skerra (Technical University of Munich, Germany) for the discussion. We thank Matthias Hammerschmidt (University of Cologne, Germany) for providing zebrafish cDNA. We thank the Microscopy and Genomics core facilities of the University of Marburg. We thank the staff from the BL18U1 and BL19U1 beamlines at the Shanghai Synchrotron Radiation Facility (SSRF) in China for assistance during data collection.

Author contributions: Conceptualization: M.G., Z.W. and R.L.; Data curation and formal analysis: M.G., Y.C., C.S.; B.S.; J.X., P.W., A.N., I.R.; L.J., T.H. and R.L.; Resources: A.N. and T.S. Funding acquisition, project administration, and supervision: Z.W. and R.L.; Visualization, writing and editing: M.G., Z.W. and R.L.

Supplementary data

Supplementary data is available at NAR online.

Conflict of interest

None declared.

Funding

This work was supported by the German Research Foundation [109546710, 416910386, and 516068166], the Deutsche José Carreras Leukämie-Stiftung [DJCLS 06 R/2022] to R.L., the National Natural Science Foundation of China [32071204 and 32125008], and the National Key R&D Program of China [2023YFA1801900] to Z.W. Supported by a Research Grant of the University Medical Center Giessen and Marburg (UKGM). Funding to pay the Open Access publication charges for this article was provided by the Open Access Publishing Fund of Philipps-Universität Marburg.

Data availability

ChIP-seq data were uploaded to the Gene Expression Omnibus (GEO) database, with the accession numbers GSE254666. Coordinates of crystal structures were deposited in the protein data bank (http://rcsb.org) with the entries 8Y76 and 8Y77. All unique reagents generated in this study are available upon reasonable request from the corresponding authors.

References

1. Lambert SA, Jolma A, Campitelli LF *et al.* The human transcription factors. *Cell* 2018;**172**:650–65. https://doi.org/10.1016/j.cell.2018.01.029
2. Morgunova E, Taipale J. Structural perspective of cooperative transcription factor binding. *Curr Opin Struct Biol* 2017;**47**:1–8. https://doi.org/10.1016/j.sbi.2017.03.006
3. Shaulian E, Karin M. AP-1 as a regulator of cell life and death. *Nat Cell Biol* 2002;**4**:E131–6. https://doi.org/10.1038/ncb0502-e131
4. Kim CA, Phillips ML, Kim W *et al.* Polymerization of the SAM domain of TEL in leukemogenesis and transcriptional repression. *EMBO J* 2001;**20**:4173–82. https://doi.org/10.1093/emboj/20.15.4173
5. Wagh K, Garcia DA, Upadhyaya A. Phase separation in transcription factor dynamics and chromatin organization. *Curr Opin Struct Biol* 2021;**71**:148–55. https://doi.org/10.1016/j.sbi.2021.06.009
6. Stielow B, Zhou Y, Cao Y *et al.* The SAM domain-containing protein 1 (SAMD1) acts as a repressive chromatin regulator at

unmethylated CpG islands. *Sci Adv* 2021;**7**:eabf2229. https://doi.org/10.1126/sciadv.abf2229
7. Stielow B, Simon C, Liefke R. Making fundamental scientific discoveries by combining information from literature, databases, and computational tools—an example. *Comput Struct Biotechnol J* 2021;**19**:3027–33. https://doi.org/10.1016/j.csbj.2021.04.052
8. Campbell B, Weber LM, Engle SJ *et al.* Investigation of SAMD1 ablation in mice. *Sci Rep* 2023;**13**:3000. https://doi.org/10.1038/s41598-023-29779-3
9. Simon C, Brunke ID, Stielow B *et al.* SAMD1 suppresses epithelial-mesenchymal transition pathways in pancreatic ductal adenocarcinoma. *PLoS Biol* 2024;**22**:e3002739. https://doi.org/10.1371/journal.pbio.3002739
10. Simon C, Stielow B, Nist A *et al.* The CpG island-binding protein SAMD1 contributes to an unfavorable gene signature in HepG2 hepatocellular carcinoma cells. *Biology (Basel)* 2022;**11**:557.
11. Dungan CM, Brightwell CR, Wen Y *et al.* Muscle-specific cellular and molecular adaptations to late-life voluntary concurrent exercise. *Function (Oxf)* 2022;**3**:zqac027. https://doi.org/10.1093/function/zqac027
12. Annear DJ, Vandeweyer G, Sanchis-Juan A *et al.* Non-Mendelian inheritance patterns and extreme deviation rates of CGG repeats in autism. *Genome Res* 2022;**32**:1967–80. https://doi.org/10.1101/gr.277011.122
13. An R, Yang Y, Liu L *et al.* SAMD1 attenuates antiphospholipid syndrome-induced pregnancy complications. *Immun Inflam Dis* 2023;**11**:e1006. https://doi.org/10.1002/iid3.1006
14. Lees AM, Deconinck AE, Campbell BD *et al.* Atherin: a newly identified, lesion-specific, LDL-binding protein in human atherosclerosis. *Atherosclerosis* 2005;**182**:219–30. https://doi.org/10.1016/j.atherosclerosis.2005.01.041
15. Weber LM, Jia Y, Stielow B *et al.* The histone acetyltransferase KAT6A is recruited to unmethylated CpG islands via a DNA binding winged helix domain. *Nucleic Acids Res* 2023;**51**:574–94. https://doi.org/10.1093/nar/gkac1188
16. Becht DC, Klein BJ, Kanai A *et al.* MORF and MOZ acetyltransferases target unmethylated CpG islands through the winged helix domain. *Nat Commun* 2023;**14**:697. https://doi.org/10.1038/s41467-023-36368-5
17. Kim CA, Gingery M, Pilpa RM *et al.* The SAM domain of polyhomeotic forms a helical polymer. *Nat Struct Biol* 2002;**9**:453–7.
18. Jiang Y, Liu T, Lee CH *et al.* The NAD (+)-mediated self-inhibition mechanism of pro-neurodegenerative SARM1. *Nature* 2020;**588**:658–63. https://doi.org/10.1038/s41586-020-2862-z
19. Volkel S, Stielow B, Finkernagel F *et al.* Zinc finger independent genome-wide binding of Sp2 potentiates recruitment of histone-fold protein Nf-y distinguishing it from Sp1 and Sp3. *PLoS Genet* 2015;**11**:e1005102. https://doi.org/10.1371/journal.pgen.1005102
20. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 1997;**276**:307–26. https://doi.org/10.1016/S0076-6879(97)76066-X
21. Liebschner D, Afonine PV, Baker ML *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol* 2019;**75**:861–77. https://doi.org/10.1107/S2059798319011471
22. Emsley P, Lohkamp B, Scott WG *et al.* Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 2010;**66**:486–501. https://doi.org/10.1107/S0907444910007493
23. Langmead B, Trapnell C, Pop M *et al.* Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 2009;**10**:R25. https://doi.org/10.1186/gb-2009-10-3-r25
24. Ramirez F, Dundar F, Diehl S *et al.* deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* 2014;**42**:W187–91. https://doi.org/10.1093/nar/gku365
25. Galaxy Community. The Galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update. *Nucleic Acids Res* 2024;**52**:W83–94. https://doi.org/10.1093/nar/gkae410
26. Nassar LR, Barber GP, Benet-Pages A *et al.* The UCSC Genome Browser database: 2023 update. *Nucleic Acids Res* 2023;**51**:D1188–95. https://doi.org/10.1093/nar/gkac1072
27. Abramson J, Adler J, Dunger J *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 2024;**636**:E4. https://doi.org/10.1038/s41586-024-07487-w
28. Vacic V, Uversky VN, Dunker AK *et al.* Composition Profiler: a tool for discovery and visualization of amino acid composition differences. *BMC Bioinformatics* 2007;**8**:493–500. https://doi.org/10.1186/1471-2105-8-211
29. Gentleman RC, Carey VJ, Bates DM *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 2004;**5**:R80. https://doi.org/10.1186/gb-2004-5-10-r80
30. Kim CA, Bowie JU. SAM domains: uniform structure, diversity of function. *Trends Biochem Sci* 2003;**28**:625–8. https://doi.org/10.1016/j.tibs.2003.11.001
31. Varadi M, Anyango S, Deshpande M *et al.* AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res* 2022;**50**:D439–44. https://doi.org/10.1093/nar/gkab1061
32. Jumper J, Evans R, Pritzel A *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* 2021;**596**:583–9. https://doi.org/10.1038/s41586-021-03819-2
33. Emenecker RJ, Griffith D, Holehouse AS. Metapredict: a fast, accurate, and easy-to-use predictor of consensus disorder and structure. *Biophys J* 2021;**120**:4312–9. https://doi.org/10.1016/j.bpj.2021.08.039
34. Quaglia F, Meszaros B, Salladini E *et al.* DisProt in 2022: improved quality and accessibility of protein intrinsic disorder annotation. *Nucleic Acids Res* 2022;**50**:D480–7. https://doi.org/10.1093/nar/gkab1082
35. Holehouse AS, Kragelund BB. The molecular basis for cellular function of intrinsically disordered protein regions. *Nat Rev Mol Cell Biol* 2024;**25**:187–211. https://doi.org/10.1038/s41580-023-00673-0
36. James LI, Barsyte-Lovejoy D, Zhong N *et al.* Discovery of a chemical probe for the L3MBTL3 methyllysine reader domain. *Nat Chem Biol* 2013;**9**:184–91. https://doi.org/10.1038/nchembio.1157
37. Zhang X, Vigers M, McCarty J *et al.* The proline-rich domain promotes Tau liquid-liquid phase separation in cells. *J Cell Biol* 2020;**219**:e202006054. https://doi.org/10.1083/jcb.202006054
38. UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2025. *Nucleic Acids Res* 2025;**53**:D609–17. https://doi.org/10.1093/nar/gkae1010
39. Weatheritt RJ, Gibson TJ. Linear motifs: lost in (pre)translation. *Trends Biochem Sci* 2012;**37**:333–41. https://doi.org/10.1016/j.tibs.2012.05.001
40. Reddy Chichili VP, Kumar V, Sivaraman J. Linkers in the structural biology of protein-protein interactions. *Protein Sci* 2013;**22**:153–67. https://doi.org/10.1002/pro.2206
41. Martin EW, Holehouse AS. Intrinsically disordered protein regions and phase separation: sequence determinants of assembly or lack thereof. *Emerg Top Life Sci* 2020;**4**:307–29. https://doi.org/10.1042/etls20190164
42. Vernon RM, Chong PA, Tsang B *et al.* Pi–Pi contacts are an overlooked protein feature relevant to phase separation. *eLife* 2018;**7**:e31486. https://doi.org/10.7554/eLife.31486
43. Gupta MN, Uversky VN. Biological importance of arginine: a comprehensive review of the roles in structure, disorder, and functionality of peptides and proteins. *Int J Biol Macromol* 2024;**257**:128646. https://doi.org/10.1016/j.ijbiomac.2023.128646

44. Zhang J, Graham TG, Vivekanand P *et al.* Sterile alpha motif domain-mediated self-association plays an essential role in modulating the activity of the *Drosophila* ETS family transcriptional repressor Yan. *Mol Cell Biol* 2010;**30**:1158–70. https://doi.org/10.1128/MCB.01225-09
45. Ray S, Hewitt K. Sticky, Adaptable, and Many-sided: SAM protein versatility in normal and pathological hematopoietic states. *Bioessays* 2023;**45**:e2300022. https://doi.org/10.1002/bies.202300022
46. Frey F, Sheahan T, Finkl K *et al.* Molecular basis of PRC1 targeting to polycomb response elements by PhoRC. *Genes Dev* 2016;**30**:1116–27. https://doi.org/10.1101/gad.279141.116
47. Isono K, Endo TA, Ku M *et al.* SAM domain polymerization links subnuclear clustering of PRC1 to gene silencing. *Dev Cell* 2013;**26**:565–77. https://doi.org/10.1016/j.devcel.2013.08.016
48. Seif E, Kang JJ, Sasseville C *et al.* Phase separation by the polyhomeotic sterile alpha motif compartmentalizes polycomb group proteins and enhances their activity. *Nat Commun* 2020;**11**:5609. https://doi.org/10.1038/s41467-020-19435-z
49. Sayou C, Nanao MH, Jamin M *et al.* A SAM oligomerization domain shapes the genomic binding landscape of the LEAFY transcription factor. *Nat Commun* 2016;**7**:11222. https://doi.org/10.1038/ncomms11222
50. Theillet FX, Kalmar L, Tompa P *et al.* The alphabet of intrinsic disorder: I. Act like a Pro: on the abundance and roles of proline residues in intrinsically disordered proteins. *Intrinsically Disord Proteins* 2013;**1**:e24360. https://doi.org/10.4161/idp.24360
51. Brodsky S, Jana T, Mittelman K *et al.* Intrinsically disordered regions direct transcription factor in vivo binding specificity. *Mol Cell* 2020;**79**:459–71. https://doi.org/10.1016/j.molcel.2020.05.032
52. Nott TJ, Petsalaki E, Farber P *et al.* Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Mol Cell* 2015;**57**:936–47. https://doi.org/10.1016/j.molcel.2015.01.013
53. Hornbeck PV, Kornhauser JM, Tkachev S *et al.* PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic Acids Res* 2012;**40**:D261–70. https://doi.org/10.1093/nar/gkr1122
54. Owen I, Shewmaker F. The role of post-translational modifications in the phase transitions of intrinsically disordered proteins. *Int J Mol Sci* 2019;**20**:5501. https://doi.org/10.3390/ijms20215501
55. Hofweber M, Dormann D. Friend or foe-Post-translational modifications as regulators of phase separation and RNP granule dynamics. *J Biol Chem* 2019;**294**:7137–50. https://doi.org/10.1074/jbc.TM118.001189
56. Borcherds W, Bremer A, Borgia MB *et al.* How do intrinsically disordered protein regions encode a driving force for liquid-liquid phase separation?, *Curr Opin Struct Biol* 2021;**67**:41–50. https://doi.org/10.1016/j.sbi.2020.09.004
57. Hanes SD. Prolyl isomerases in gene transcription. *Biochim Biophys Acta* 2015;**1850**:2017–34.
58. Babu M, Favretto F, Rankovic M *et al.* Peptidyl prolyl isomerase a modulates the liquid–liquid phase separation of proline-rich IDPs. *J Am Chem Soc* 2022;**144**:16157–63. https://doi.org/10.1021/jacs.2c07149
59. Sutherland H, Bickmore WA. Transcription factories: gene expression in unions?, *Nat Rev Genet* 2009;**10**:457–66. https://doi.org/10.1038/nrg2592
60. Boija A, Klein IA, Sabari BR *et al.* Transcription factors activate genes through the phase-separation capacity of their activation domains. *Cell* 2018;**175**:1842–55. https://doi.org/10.1016/j.cell.2018.10.042
61. Kribelbauer JF, Rastogi C, Bussemaker HJ *et al.* Low-affinity binding sites and the transcription factor specificity paradox in eukaryotes. *Annu Rev Cell Dev Biol* 2019;**35**:357–79. https://doi.org/10.1146/annurev-cellbio-100617-062719
62. Már M, Nitsenko K, Heidarsson PO. Multifunctional intrinsically disordered regions in transcription factors. *Chem* 2023;**29**:e202203369. https://doi.org/10.1002/chem.202203369

Received: May 14, 2024. **Revised:** January 30, 2025. **Editorial Decision:** March 18, 2025. **Accepted:** March 21, 2025