Structure and methyl-lysine binding selectivity of the HUSH complex subunit MPP8

Nikolaos Nikolopoulos[1,2,‡] Shun-ichiro Oda[1,2,‡], Daniil M. Prigozhin[1,3,‡], and Yorgo Modis[1,2,*]

[1] Molecular Immunity Unit, Department of Medicine, University of Cambridge, MRC Laboratory of Molecular Biology, Cambridge CB2 0QH, UK

[2] Cambridge Institute of Therapeutic Immunology & Infectious Disease (CITIID), Department of Medicine, University of Cambridge, Cambridge CB2 0AW, UK

[3] Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

[‡] These authors contributed equally to this work and are listed alphabetically by last name

* To whom correspondence should be addressed. Tel.: +44 1223 267282; Email: ymodis@mrc-lmb.cam.ac.uk

Key Points

- Structure of MPP8 CTD reveals ankyrin repeats and PINIT-like domain required for HUSH activity
- AlphaFold predicts a large interaction interface between MPP8 CTD and TASOR SPOC and DomI domains
- The MPP8 chromodomain binds trimethyl-lysines in ATF7IP and GLP with higher affinity than H3K9me3

Keywords

Transcriptional repression; epigenetic silencing; histone H3 lysine 9 methylation (H3K9me3); retrotransposon; H3K9 methyltransferase

ABSTRACT

The Human Silencing Hub (HUSH) guards the genome from the pathogenic effects of retroelement expression. Composed of MPP8, TASOR, and Periphilin-1, HUSH recognizes actively transcribed retrotransposed sequences by the presence of long (>1.5-kb) nascent transcripts without introns. HUSH recruits effectors that alter chromatin structure, degrade transcripts, and deposit transcriptionally repressive epigenetic marks. Structural and biochemical data for the three HUSH subunits are sparse. Here, we report the crystal structure of the C-terminal domain (CTD) of MPP8 necessary for HUSH activity. The MPP8 CTD consists of five ankyrin repeats followed by a domain with structural homology to the PINIT domains of Siz/PIAS-family SUMO E3 ligases. AlphaFold-Multimer modeling predicts that the MPP8 CTD forms extended interaction interfaces with a SPOC domain and a domain with a novel fold, DomI, in TASOR. The chromodomain of MPP8, known to bind the repressive mark H3K9me3, binds with similar or higher affinity to H3K9-like sequences in the mammalian H3K9 methyltransferase subunits SETDB1, ATF7IP, G9a, and GLP. Hence, MPP8 promotes heterochromatinization by recruiting H3K9 methyltransferases. Our work identifies novel structural elements in MPP8 that are required for HUSH complex assembly and silencing, thereby fulfilling vital functions in controlling the transcription of retrotransposons of endogenous and viral origin.

INTRODUCTION

Transposable elements (TEs) have been coevolving with the genomes of all organisms since the early stages of life. Some TEs, such as LINEs (long interspersed nuclear elements), are thought to have evolved in early eukaryotes (1). TEs can also be acquired when viral DNA integrates into the genome of a host germline cell. Why TEs are so ubiquitous and abundant, accounting for more than half of the human genome (2), is not fully understood. TEs are known to serve as a genetic reservoir from which new genes, regulatory elements, and transcriptional networks can emerge (2-6). However, inappropriate expression of TEs has been associated with various genetic diseases including autoimmune diseases, neurodegeneration, hemophilia, cystic fibrosis, and cancer (1,7-9). The primary mechanism cells have evolved to control TE expression is epigenetic transcriptional silencing. Transcriptional repression of TEs is particularly important in embryogenesis, chronic infection, and stress responses, when epigenetic modifications are more pro-transcriptional (10). In vertebrates, the Human Silencing

Hub (HUSH) is a key repressor of retrotransposons, the subset of TEs that can replicate by retrotransposition (11). The hundreds of genomic loci repressed by the HUSH complex include recently integrated viral elements, LINE-1 elements, and a subset of endogenous genes generated through retrotransposition of cellular mRNAs (11-14). HUSH target loci have in common that they encode transcripts longer than 1.5 kb, are transcriptionally active, and are products of retrotransposition and therefore lack introns (12).

The HUSH complex consists of three proteins: TASOR, MPP8 (M-phase phosphoprotein 8), and Periphilin-1 (PPLHN1) (11). Periphilin targets HUSH to its genomic loci by binding to nascent transcripts with exons longer than 1.5 kb (12). Periphilin binds nascent RNA with a disordered, arginine/tyrosine-rich N-terminal domain with self-aggregating properties (15). MPP8 was recently reported to contribute to HUSH targeting by binding to termination factor WDR82 that is enriched at sites with high RNA polymerase II occupancy, which include long exons (16). TASOR binds both Periphilin and MPP8 to form the HUSH complex (11,15). TASOR contains three folded domains required for HUSH activity: a pseudo-PARP, a predicted SPOC (Spen Paralogue and Orthologue C-terminal) domain, and a domain of unknown fold, DomI (11,14). The pseudo-PARP binds weakly to ssRNA and is required for HUSH activity, but how this domain contributes to gene silencing remains unclear (14). A structurally homologous pseudo-PARP in TEX15, a piRNA-dependent transposon silencing factor in the male germline, can functionally substitute for the pseudo-PARP of TASOR (17). The SPOC domain is required for binding to MPP8 (11,14). MPP8 recruits ATF7IP, an obligate binding partner and stabilizing factor of the histone H3 lysine 9 (H3K9) methyltransferase SETDB1 (18). HUSH silencing additionally requires MORC2, a DNA-binding ATPase thought to be a chromatin remodeler (19,20). Thus, in the current model of HUSH-dependent repression, Periphilin binds to nascent transcripts of newly retrotransposed sequences, MPP8 recruits SETDB1/ATF7IP to deposit the transcriptionally repressive H3K9me3 mark, and MORC2 is recruited to promote chromatin compaction. Notably, a subset of MPP8 chromatin binding sites lack TASOR and Periphilin, suggesting a HUSH-independent function of MPP8 chromatin binding (16). Consistent with this, MPP8 recruits the NEXT complex to chromatin (via an interaction with ZCCHC8), leading to exosome-dependent decay of non-polyadenylated TE transcripts, whereas HUSH represses polyadenylated transcripts (21). Together, the transcriptional repression activities of MPP8 protect the DNA-hypomethylated pluripotent ground state of stem cells (22). MPP8 depletion causes transcription of RNAs that induce interferon and DNA damage responses, due to recognition by the RNA sensor MDA5 and increased retrotransposition activity, respectively

(23,24). These responses are thought to account for inhibition of acute myeloid leukemia and sensitization of tumor cells to checkpoint blockade upon MPP8 knockout (24,25).

Biochemical and structural data for the protein domains essential for HUSH function are sparse. This study focuses on delineating the structural and biochemical properties of MPP8 and how they contribute to transcriptional repression. MPP8 contains an N-terminal chromodomain (CD, residues 59-118) and a set of five predicted ankyrin repeats (14). The predicted ankyrin repeats (residues 548-728) and the following C-terminal region (residues 729-860) are both absolutely required for binding to TASOR and for HUSH repression of transgenes and LINE-1 elements (14,22). In contrast, HUSH repression of an integrated lentivirus reporter is maintained, although slightly delayed, upon deletion of the CD, or indeed the entire N-terminal region of MPP8 (residues 1-499) (14). The MPP8 CD binds di- or trimethylated H3K9 peptides with micromolar dissociation constants (26-28). A crystal structure of the MPP8 CD bound to the H3K9me3 peptide showed that the trimethyl-lysine is bound in a cage formed by one acidic and three aromatic residues (28). Binding to mono- or unmethylated H3K9 peptides, and to certain methylated or unmethylated H3K4 and H4K20 peptides, is 20- to 600-fold weaker but still measurable (26,28).

The relatively low affinities and methylation specificities of MPP8 CD for lysine-methylated peptides, and its auxiliary role in HUSH-dependent lentivirus repression, raise the question of whether this domain has other binding partners or functions. We note that H3K9 mimic sequences have been identified in all four protein components of two of the most important mammalian H3K9 methyltransferase complexes, G9a/GLP and SETDB1/ATF7IP (29-31). G9a/GLP automethylates the H3K9 mimic sequence in G9a, which can then bind the CD of HP1 (29,31). G9a/GLP also methylates the H3K9 mimic sequence of ATF7IP at Lys16 (30). The resulting ATF7IP K16me3 peptide is specifically recognized by the CD of MPP8 (30). Mutation or deletion of ATF7IP Lys16 abrogated MPP8 chromatin binding and delayed HUSH-dependent repression of a reporter transgene in mouse embryonic stem cells (30), a similar phenotype to deletion of the MPP8 CD or mutation of its H3K9me3 binding motif (14). MPP8 was also reported to interact with H3K9-like sequences in GLP and Dnmt3a, which are both also methylated by G9a/GLP (32,33). Thus, MPP8 may bind to H3K9-like peptides in a broader set of proteins. However, only the interactions of the MPP8 CD with H3K9me3 and Dnmt3aK47me2 have been validated biochemically using purified components (28,32).

Here, we report the crystal structure of the C-terminal domain (CTD) of MPP8 necessary and sufficient for HUSH activity consisting of an ankyrin repeat domain and a β-sandwich with a

PINIT domain fold. We used AlphaFold-Multimer to generate a high-confidence model of the MPP8-TASOR complex, which suggests that the MPP8 CTD binds to a SPOC domain and a domain with a novel fold, DomI, in TASOR. The AlphaFold model of the MPP8-TASOR complex is experimentally supported by HUSH activity assays with truncations of MPP8 and TASOR (14). Isothermal titration calorimetry measurements show that the MPP8 CD binds to H3K9-like peptides from G9a/GLP and SETDB1/ATF7IP with similar or higher affinities to H3K9me3. In the context of the previously reported phenotypes of MPP8 deletions, we propose that these interactions accelerate or amplify transcriptional repression through increased recruitment of H3K9 methyltransferases in a read/write mechanism. Our work identifies novel structural elements in MPP8 required for HUSH assembly and effector recruitment.

MATERIALS AND METHODS

Protein expression and purification

Synthetic genes codon-optimized for *Escherichia coli* were cloned into the pET15b vector (Novagen) with an N-terminal hexahistidine (His_6)-tag. The MPP8 ankyrin-CTD (residues 546-860, UniProt Q99549) was cloned in frame with the vector's His_6 tag and thrombin cleavage site. The MPP8 CD (residues 55-116) was cloned with a tobacco etch virus (TEV) protease cleavage site (ENLYFQG) replacing the thrombin cleavage site. These constructs were transformed into *E. coli* BL21 (DE3) cells (New England BioLabs).

For expression and purification of MPP8 ankyrin-CTD, overnight *E. coli* cultures were diluted 1:200 into 2 liters of autoinduction media (25 mM Na_2HPO_4, 25 mM KH_2PO_4, 50 mM NH_4Cl, 5 mM Na_2SO_4, 2 mM $MgSO_4$, 0.2 x trace metals mix, 0.5% glycerol, 0.05% glucose, 0.2% α-lactose, 0.1% aspartate, 0.2 mg/ml of 18 amino acids (no C or Y), 1 μM vitamin B12, 0.1 mg/ml ampicillin) and grown at 37˚C for 3 h and then 18˚C for 26 h. Cells were harvested and resuspended in 50 ml Lysis-Wash buffer (20 mM Tris-HCl pH 7.4, 0.15 M NaCl, 1 mM DTT, 20 mM imidazole). Cells were lysed by sonication and centrifuged at 16,000 rpm for 20 min. The supernatant was filtered through a 0.45 μm syringe filter and applied to a 5 ml Ni-NTA agarose column (Qiagen) pre-equilibrated in Lysis-Wash buffer. The column was washed with 50 ml Lysis-Wash buffer. MPP8 ankyrin-CTD was eluted with 25 ml of elution buffer (20 mM Tris-HCl pH 7.4, 0.15 M NaCl, 1 mM DTT, 0.4 M imidazole). To remove the His_6 tag, 0.1 ml of thrombin (Cytiva) was added to the eluate and the mixture was incubated overnight at room temperature with concomitant dialysis against 1 liter of dialysis buffer (20 mM Tris-HCl pH 7.4, 20 mM NaCl, 1 mM DTT, 10% glycerol). The solution was applied to a column of 5 ml Ni-NTA agarose

(Qiagen) equilibrated in Lysis-Wash buffer and the flow-through collected. The sample was then applied to a RESOURCE Q anion exchange column (Cytiva) equilibrated in low-salt buffer (20 mM Tris-HCl pH 8, 20 mM NaCl, 1 mM DTT, 10% glycerol). The protein was eluted with a linear gradient of high-salt buffer (10 mM HEPES pH 7.4, 1 M NaCl, 0.5 mM Tris(2-carboxyethyl)phosphine (TCEP)). Fractions containing MPP8 ankyrin-CTD were pooled, concentrated, and applied to a Superdex 200 10/300 size-exclusion chromatography (SEC) column (Cytiva) equilibrated in SEC buffer (10 mM HEPES pH 7.8, 0.15 M NaCl, 0.5 mM TCEP).

To express and purify MPP8 CD, overnight *E. coli* BL21 (DE3) cultures were diluted 1:200 into 1.6 liters of LB with 0.1 mg ml^{-1} ampicillin, grown at 37˚C to OD_{600} 0.6, induced with 0.1 mM IPTG, and incubated at 37˚C for 4-6 h. Cells were harvested and resuspended in twice the cell pellet volume of lysis buffer (50 mM potassium phosphate pH 8.0, 0.5 M NaCl, 20 mM Imidazole, cOmplete protease inhibitor cocktail (Roche)). Cells were lysed by sonication and centrifuged at 35,000 rpm for 1 h. The supernatant was mixed with 2 ml Ni-NTA agarose (Qiagen) pre-equilibrated in lysis buffer. The mixture was incubated for 1 h, washed with wash buffer (50 mM potassium phosphate pH 8.0, 0.5 M NaCl, 20 mM imidazole), and bound protein was eluted with wash buffer supplemented with 0.3 M imidazole. The buffer was exchanged by serial dilution and concentration to Mono Q buffer (25 mM Tris pH 7.4, 50 mM NaCl, 0.5 mM TCEP). The sample was applied to a Mono Q 5/50 GL column (Cytiva) equilibrated in Mono Q buffer. Bound protein was eluted with a linear gradient of high-salt Q buffer (25 mM Tris pH 7.4, 1 M NaCl, 0.5 mM TCEP). Fractions containing MPP8 CD were pooled, concentrated, and applied to a Superdex 75 10/300 SEC column (Cytiva) equilibrated in SEC75 buffer (50 mM HEPES pH 7.5, 0.15 M NaCl, 0.5 mM TCEP).

Crystallographic structure determination

Crystals were grown at 18˚C by sitting drop vapor diffusion. Purified MPP8 ankyrin-CTD was concentrated to 4.5 mg/ml using a 10-kDa cutoff centrifugal filtration device (Merck Millipore) and mixed with an equal volume (up to 0.5 µl) of reservoir solution. Prism- and needle-shaped crystals were seen in 13 conditions from the Morpheus II screen (Molecular Dimensions), each containing 0.5 mM Oxometalates Mix and either 10% w/v PEG 8000 with 20% v/v 1,5-pentanediol or 12.5% w/v PEG 4000 with 10% w/v 1,2,6-hexanetriol. Crystals suitable for X-ray diffraction experiments were obtained with the following Morpheus II reagents: Precipitant Mix 6 (1:2 dilution), Buffer System 6 pH 8.5 (1:10 dilution), and Oxometalates Mix (1:10 dilution). The final precipitant solution contained 12.5% (w/v) PEG 4000, 20% (v/v) 1,2,6-hexanetriol, 0.5 mM

of each Oxometalate, 0.1 M GlyGly/AMPD pH 8.5. Microseeding was used to obtain larger crystals; the seed stock was made by shearing small crystal by repeated pipetting and performing serial dilutions of the sheared crystal fragments to use as seed stocks. Incubation of crystals in mother liquor supplemented with 10-20% PEG200 for 30 min prior to freezing in liquid nitrogen slightly improved the X-ray diffraction limit to 3.0 Å resolution.

X-ray diffraction data were collected at 100 K at Diamond Light Source (DLS) beamline I04-1 and processed with xia2 (DIALS, AIMLESS) (34). An atomic model predicted by AlphaFold2 as implemented by ColabFold (35) was used as a molecular replacement search model in PHENIX v.1.20 (36). Eight copies of the model were placed in the asymmetric unit (chains A-H). The atomic model was edited with COOT v.0.9.6 (37) and iteratively refined with PHENIX (36). The final model spanned residues 565-860 (residues 546-564 were disordered). See **Table 1** for crystallographic data collection, refinement, and validation statistics.

Isothermal titration calorimetry (ITC)

H3K9 and H3K9-like peptides, unmethylated or lysine-trimethylated, were chemically synthesized (Bio Basic Inc / NBS Biologicals Ltd). Binding of MPP8 CD to H3K9 and H3K9-like peptides was analyzed in 50 mM HEPES pH 7.4, 0.1 M NaCl, 0.5 mM TCEP at 25˚C (298 K), with an AutoiTC200 calorimeter (MicroCal). The sample cell was loaded with 0.37 ml of 50 µM MPP8 CD and the titrant syringe with 750 µM peptide. 20 serial injections of 6 µl peptide solution were performed at 3 min intervals. The stirring speed was 1,000 rpm and the reference power was maintained at 6.06 µcal s^{-1}. The net heat absorption or release associated with each injection was calculated by subtracting the heat associated with the injection of peptide solution to buffer. Thermodynamic parameters were extracted from a curve fit to the data to a single-site model with Origin 7.0 (MicroCal).

Size-exclusion chromatography and multiangle scattering (SEC-MALS)

100 µl samples containing 4 mg ml^{-1} MPP8 CD were analyzed by size-exclusion chromatography (SEC) at 293 K on a Superdex 75 (10/300) column (Cytiva) in 50 mM HEPES pH 7.5, 0.15 M NaCl, 0.5 mM TCEP with a flow rate of 0.5 ml min^{-1} on an Agilent 1200 series liquid chromatography system. The SEC system was coupled to a Wyatt Heleos II 18-angle light scattering instrument coupled to a Wyatt Optilab T-rEX online refractive index detector. Protein concentration was determined from the excess differential refractive index based on 0.185 dRI for 1 g ml^{-1} protein (dn/dc = 0.185 ml g^{-1}). The measured protein concentration and scattering intensity were used to calculate molecular mass from the intercept of a Debye plot using Zimm's

model as implemented in the Wyatt ASTRA software. Bovine serum albumen (100 µl at 2 mg ml^{-1}) was run as a standard for mass determination, and to determine interdetector delay volumes and band broadening parameters.

AlphaFold2 modeling

Atomic models of MPP8 ankyrin-CTD in complex with TASOR SPOC-DomI were predicted with a local installation of AlphaFold2-v2.3.1 (38,39) using multimer mode as implemented by ColabFold v1.5.2 (35) using MMseqs2 for multiple sequence alignment (MSA). The amino acid sequences of human MPP8 residues 535-860 and TASOR residues 352-632 (UniProt Q9UK61-1) were used as the input. Polypeptide geometry in the models was regularized by post-prediction relaxation using the Amber force field. The maximum number of recycles was set to 20. No MSA cutoff was used. The sequence pairing mode was set to pair sequences from the same species and unpaired MSA. The multimer model was set to multimer-v3. Use of available template files from the Protein Data Bank had no significant effect on atomic model prediction. Five atomic models were output from each AlphaFold run. Protein interfaces were analyzed with the Protein interfaces, surfaces, and assemblies (PISA) service at the European Bioinformatics Institute [http://www.ebi.ac.uk/pdbe/prot_int/pistart.html] (40).

Atomic models of MPP8 CD in complex with H3K9 and H3K9-like peptides were generated as described above except that the input sequences consisted of human MPP8 CD (residues 55-116) and one of the following: ATF7IP residues 10-21 (UniProt Q6VMQ6), GLP residues 199-211 (UniProt Q9H9B1), histone 3.1 residues 4-15 (UniProt P68431), G9a residues 179-190 (UniProt Q96KQ7), SETDB1 residues 1165-1175 (UniProt Q155047), and DNMT3a residues 45-51 (UniProt Q9Y6K1). Use of available template files from the Protein Data Bank slightly improved the confidence scores of the models.

The top-ranking atomic coordinates generated by AlphaFold-Multimer for the MPP8-TASOR complex and for the MPP8-peptide complexes shown in this study are available as separate files (**Supplementary Data Sets S1-S6**).

RESULTS

Structure of MPP8 CTD reveals ankyrin repeats and a PINIT-like domain

The region of MPP8 required for transcriptional repression of an integrating lentiviral reporter has been mapped to within residues 500-860 (14). Moreover, residues 528-860 of MPP8 are

sufficient for HUSH complex formation (14). Having established the functional importance of the MPP8 C-terminal domain (CTD), we sought to determine its structure to gain insights on how it contributes to gene silencing as part of the HUSH complex. A 3.0 Å resolution crystal structure of MPP8 CTD (residues 565-860) was determined by molecular replacement using the AlphaFold prediction for the CTD as the search model (**Figure 1A**, **Table 1** and **Supplementary Figure S1**). The crystal structure reveals a bipartite fold consisting of five ankyrin (helix-loop-helix) repeats followed by a β-sandwich domain (**Figure 1B**). The last α-helix of the ankyrin repeats, helix α10 (**Figure 1C**), is twice as long as the other ankyrin-repeat helices and serves as a rigid linker to the β-sandwich domain. The β-sandwich is tightly packed onto the ankyrin repeats. Two strands from the β-sandwich wrap around the C-terminal half of helix α10. The β-sandwich forms additional contacts with the loops of the last two ankyrin repeats (**Figure 1A,B**).

A structural similarity search with DALI (41) revealed structural similarity of the MPP8 β-sandwich domain to the PINIT domains of the Siz/PIAS RING family of SUMO E3 ligases (42). The most similar structures were those of the PINIT domains from *Saccharomyces cerevisiae* Siz1 (43), human PIAS3, and human PIAS2 (**Figure 2A**). The domains have the same topology and overall fold (2.3-2.5 Å RMSD, Z-score 8.4-9.4), although the Siz/PIAS PINIT domains have a flexible N-terminal linker in place of the rigid helical linker (helix α10) in MPP8. The PINIT domain is the substrate recognition domain of Siz/PIAS SUMO E3 ligases and is required for conjugation of SUMO to their substrate proteins (43,44). The PINIT domain of Siz1 is required for SUMOylation of splicing factor Prp45 and the proliferating cell nuclear antigen (PCNA) in yeast (44,45). In humans, PIAS2 and PIAS3 have been reported to SUMOylate various substrates including Aurora-B kinase, Argonaute-2, and the INO80 chromatin remodeling complex (46-49). The PINIT amino acid sequence motif is not conserved in the PINIT-like domain of MPP8 (**Figure 2B**). Indeed, there is no detectable sequence similarity between the PINIT domains of MPP8 and Siz/PIAS. MPP8 also lacks a RING domain and is therefore not an E3 ligase. The presence of a PINIT-like domain in MPP8 remains notable given that HUSH subunits TASOR and Periphilin, and HUSH effectors SETDB1, ATF7IP and MORC2 all contain SUMO acceptor sites (50).

During refinement of the atomic model of MPP8 CTD, we noticed that the eight subunits in the asymmetric unit form four disulfide-linked homodimers. The intermolecular disulfide bond linking the homodimers was formed by cysteine 799, located in the loop between strands β4 and β5 of the PINIT-like domain (**Supplementary Figure S2A**). Moreover, 10-15% of the recombinant MPP8 CTD protein migrated as a dimer in size-exclusion chromatography,

indicating that some dimers formed in solution despite the presence of reducing agent (1 mM DTT or 0.5 mM TCEP) in the buffer (**Supplementary Figure S2B**). However, analysis with the PISA server (39) showed that the dimers have a small interface area (375 Å) and no strong contacts other than the disulfide bond. We therefore conclude that in the context of the reducing intracellular environment the disulfide bond and dimer interface observed in the crystal are unlikely to be physiologically relevant.

Molecular modeling of the MPP8-TASOR complex with AlphaFold-Multimer

We showed previously using purified recombinant proteins that MPP8 CTD (residues 528-860) forms a 1:1 complex with TASOR residues 354-633, which were predicted to contain two folded domains: a SPOC domain and DomI, a domain of unknown structure and function (14). We mapped the minimal region of TASOR required to interact with to within residues 354-633, and showed that deletion of either the SPOC or DomI domains resulted in loss of HUSH-dependent transgene reporter repression (14) (**Figure 3A**). As we were unable to crystallize or obtain a cryoEM reconstruction of an MPP8-TASOR complex, we generated a predicted atomic model with AlphaFold-Multimer (35,38,39) as implemented by ColabFold (35). With only the sequences of human TASOR and MPP8 as the input, AlphaFold generated 1:1 MPP8-TASOR complexes with high confidence scores (average pLDDT = 88-90) and low predicted aligned error (**Figure 3B**). All the generated AlphaFold models had the same overall structure, with or without the use of model templates from the Protein Data Bank. Residues 125-535 of MPP8 were consistently unstructured in AlphaFold predictions, as were amino acids flanking the SPOC and DomI domains of TASOR. Moreover, the pseudo-PARP domain of TASOR did not interact with MPP8 in the predicted models. The boundaries of structural disorder in the AlphaFold-Multimer predictions suggest that the core complex consists of MPP8 CTD residues 536-860 and TASOR SPOC-DomI residues 352-632. These boundaries are experimentally supported by previous coimmunoprecipitation and lentivirus reporter assays with different truncations of MPP8 and TASOR, which mapped the complex forming regions to MPP8 residues 528-860 and TASOR residues 354-633 of TASOR (14).

In the AlphaFold prediction for the MPP8-TASOR complex, the TASOR subunit contains a SPOC domain with very high confidence scores (residues 352-502, pLDDT > 90), and a domain with a novel fold, DomI (residues 518-632, pLDDT > 70-90). The SPOC domain is predicted to bind to the concave surface of the MPP8 ankyrin repeats. DomI consists of four loosely packed α-helices, with an extended 31-amino acid structured loop between the second

and third helices. These structural elements of DomI together occupy an extensive binding footprint on the PINIT-like domain of MPP8 (**Figure 3C**). DomI lacks a stable hydrophobic core, and its fold appears to be dependent on binding to MPP8. In contrast, the SPOC domain appears to form an independent structural unit, although our attempts to express the TASOR SPOC domain by itself in *E. coli* were unsuccessful. The 16-amino acid linker that connects the SPOC and DomI domains does not form any contacts with MPP8 and is predicted to be disordered (pLDDT = 30-50). The MPP8-TASOR interface is extended by residues 536-564 of MPP8, which bind along the surface of the SPOC domain, on the opposite side to the ankyrin-repeat binding surface. We note that MPP8 residues 546-564 were present in the crystallized MPP8 construct but were disordered in the electron density map, presumably due to the absence of TASOR. Together, the large total interface area of 4,374 $Å^2$, the estimated solvation free energy gain (ΔG) of -54 kJ/mol, and the presence of 52 polar contacts in the predicted complex suggest that MPP8 and TASOR form a tight complex (**Figure 3D,E**).

The most similar experimentally determined structures to the predicted TASOR SPOC domain are the SPOC domains from the *Arabidopsis* flowering regulator FPA, and human PHF3, with DALI Z-scores (41) of 12.2 and 11.8, respectively. Notably, the PHF3 SPOC domain binds RNA Pol II CTD di-heptapeptide phosphorylated on Ser2 (Y[pS]PTSPS-Y[pS]PTSPS) with a 1 μM dissociation constant. Binding of PHF3 SPOC to Ser2-phosphorylated Pol II CTD heptapeptide repeats, a hallmark of the elongating form of Pol II, has been proposed to promote Pol II elongation and condensation with mRNA processing factors, thereby contributing to the coordination of transcription elongation with mRNA decay (51). Superposition of the PHF3-Pol II CTD structure onto the predicted MPP8-TASOR structure shows that MPP8 residues 540-554 bind to the equivalent site on the TASOR SPOC domain as the Pol II CTD peptide on the PHF3 structure, such that binding of MPP8 and Pol II CTD heptapeptide to the TASOR SPOC domain would be mutually exclusive (**Supplementary Figure S3**). The arginine and two lysine residues in the PHF3 SPOC domain that coordinate the phospho-serine moieties of the Pol II CTD are not conserved in the TASOR SPOC domain. Moreover, AlphaFold predictions with TASOR SPOC and Pol II CTD di-heptapeptides did not yield a consistent solution and had very low confidence scores (pLDDT = 10-30). Based on these structural analyses, we conclude that it is unlikely that the TASOR SPOC domains binds to phosphorylated Pol II CTD heptapeptides like the PHF3 SPOC domain. Whether the TASOR SPOC domain recognizes another protein or peptide in related to transcription elongation or mRNA processing remains to be determined.

The MPP8 CD binds H3K9me3-like peptides from H3K9 methyltransferases

The chromodomain (CD) of MPP8 binds H3K9me3 with relatively low affinity, with a dissociation constant (K_D) in the micromolar range (26-28), and binding of this mark by MPP8 CD is not absolutely required for HUSH-dependent repression of lentiviruses in cell culture (14). The MPP8 CD binds human Dnmt3aK47me3 and the equivalent mark in mouse with even lower affinity (K_D = 12 μM) (32), but whether the MPP8 CD binds directly to H3K9me3-like marks in other proteins remains unknown. To address this, we measured the binding affinities of MPP8 CD to the H3K9-like sequences in SETDB1, ATF7IP, G9a, and GLP. Our isothermal titration calorimetry (ITC) data show that MPP8 CD binds to the N-terminal tail of ATF7IP trimethylated at lysine 16 with a K_D of 193 ± 31 nM, and to residues 199-211 from GLP with lysine 205 trimethylated with a K_D of 337 ± 64 nM (**Figure 4A**). These binding affinities are fourfold and twofold higher than for H3K9me3, which had a K_D of 769 ± 65 nM in ITC, consistent with previously reported values (26-28). Additionally, MPP8 CD bound to lysine-methylated peptides from G9a (trimethylated at K185) and SETDB1 (trimethylated at K1170) with K_D value of 1.31 and 1.55 μM, respectively. These affinities are slightly lower than H3K9me3 but slightly higher than H3K9me2 (K_D = 1.70 μM). Additionally, H3K27me3 and SETDB1 K1178me3 peptides showed weak binding (K_D > 50 μM). There was no detectable binding of MPP8 CD to any of the peptides listed above in the absence of lysine methylation (**Figure 4B** and **Supplementary Figure S4**). In summary, our ITC data shows that the MPP8 CD binds to lysine-trimethylated peptides in all four components of the SETDB1/ATF7IP and G9a/GLP H3K9 methyltransferase complexes, with binding affinities higher than for H3K9me3 in the case of the ATF7IP and GLP peptides.

Having identified nine lysine-methylated peptides with measurable binding to the MPP8 CD (**Figure 4B**) we generated atomic models of the MPP8-peptide complexes with AlphaFold-Multimer. Although lysine methylation could not be modeled in AlphaFold, we expected the modeling to provide insights on how the amino acid sequence flanking the lysine methylation contributes to MPP8 binding. The MPP8 CD was reported to form homodimers in size-exclusion chromatography (SEC) and in crystals in complex with an H3K9me3 peptide (28). We performed combined SEC and multiangle light scattering (SEC-MALS) and obtained the same SEC profile as previously reported (28), but the MALS signal – which provides a more accurate measurement of the molecular weight in solution than SEC – indicated that the MPP8 CD was a monomer (**Supplementary Figure S2C**). In support of the MPP8 CD being monomeric in solution, analysis with the PISA (39) showed that the crystallographic MPP8 CD dimer interface area (534 Å) and estimated solvation free energy gain (ΔG = -3.6 kcal/mol) were both smaller

than would be expected for the interface to be stable in solution. Moreover, AlphaFold-Multimer predictions with two copies of the MPP8 CD, without H3K9-like peptide or use of Protein Data Bank (PDB) templates, failed to consistently predict the crystallographic MPP8 dimer. We therefore specified a 1:1 MPP8:peptide stoichiometry for AlphaFold modeling. The resulting MPP8-peptide complexes had high confidence scores, with an average pLDDT of 73 for the peptides and 87-91 for the MPP8 CD (**Supplementary Figure S5**). For each complex, the peptide and MPP8 CD in the top five predicted models had the same overall structure as the previously determined crystal structure of the MPP8 CD bound to an H3K9me3 peptide (28) (**Figure 4A,B**). In all the structures, the lysine at the methylation site (position 0) is bound in a cage formed by an aspartate and three aromatic residues from MPP8 (**Figure 4C,D**). All peptides bound by MPP8 contain an alanine a position -2 (**Figure 4C**). The side chain of this alanine packs onto a hydrophobic surface of the CD, forming van der Waals contacts. There is a strong preference of an arginine at position -1, with only the two SETDB1 peptides lacking an arginine at this position. In the AlphaFold models, the side chain of the arginine at position -1 is predicted to form a salt bridge with Glu56 of MPP8, although we note that the Glu56 side chain was disordered in the MPP8-H3K9me3 crystal structure (**Figure 4D**). Position +1 is a serine or threonine in all peptides except the Dnmt3a peptide, which has a valine at this position. The hydroxyl of this serine or threonine contributes to MPP8 binding by forming a hydrogen bond with Glu91 of MPP8. The peptides with highest affinities for MPP8, those from ATF7IP, GLP, and G9a, also contain an aromatic residue (histidine or phenylalanine) at position -4, which is predicted to form hydrophobic contacts with the C-terminal helix of the MPP8 CD. In contrast, the H3K9me3 peptide has a glutamine at position -4. Together, our structural modeling suggests that the key determinants for recognition of a methylated lysine by MPP8 are a strict requirement for an alanine at position -2, strong preferences for arginine at position -1 and serine or threonine at position +1, and a possible preference for an aromatic residue at position -4.

DISCUSSION

The crystal structure of the MPP8 CTD and the AlphaFold models of the MPP8-TASOR complex complete our picture of the of the folded domains essential for HUSH repression of integrated transgenes. This leaves the DomII and PIN domains of TASOR (residues 1212-1466), which are not required for transgene repression, as the only domains in the HUSH complex that are predicted to be folded but structurally unstudied. Our model of the MPP8-

TASOR complex identifies an extensive molecular interaction interface between the MPP8 CTD and the SPOC and DomI domains of TASOR. The model reveals structural homologies between the MPP8 CTD and the substrate recognition domains of Siz/PIAS-family SUMO E3 ligases, and between the SPOC domains of TASOR and PHF3, a protein that contributes to the linkage of transcription elongation with mRNA decay. The TASOR DomI domain is predicted to form a novel fold that is structurally dependent on the MPP8 CTD. A long loop in DomI that binds along a large area of the PINIT-like domain of MPP8 contains a lysine (Lys586) reported to be SUMOylated (50). Lys586 is highly solvent accessible in the predicted structure, suggesting that a function of the MPP8 PINIT-like domain could be to present TASOR K586 as a substrate for a SUMO E3 ligase. This raises the question of whether SUMOylation of TASOR K586 may regulate or license transcriptional repression by the HUSH complex.

Our ITC measurements show that the CD of MPP8 binds to lysine-trimethylated H3K9-like peptides in all four components of two key H3K9 methyltransferases, SETDB1, ATF7IP, G9a, and GLP, with slightly higher affinity than H3K9me3. AlphaFold modeling suggests that the MPP8 CD preferentially recognizes the lysine trimethylated peptides with the sequence ARK[S/T], with a possible preference for an aromatic residue at position -4. A reported phosphorylation site in the MPP8 CD, at threonine 89 (52), maps to the vicinity of the methyl-lysine binding site. The structures of the CD-peptide complexes suggest that a phosphothreonine adduct at this site would be compatible with H3K9-like peptide binding but is likely to alter the binding affinity and potentially also the methylation specificity of MPP8. The interaction of MPP8 with trimethylated peptides from H3K9 methyltransferases suggests that the MPP8 CD can recruit SETDB/ATF7IP to target chromatin loci in addition to sensing (reading) H3K9me3. The H3K9-like peptides from ATF7IP, G9a and GLP are methylated by the G9a/GLP complex, and these methylated peptides can then bind the CD of HP1 (29-33). Thus, recruitment of H3K9 methyltransferases by MPP8 may promote recruitment of HP1 and heterochromatinization through amplification of lysine trimethylation not only on H3K9 but also on H3K9-like peptides in the H3K9methyltransferases themselves. However, binding of the MPP8 CD to H3K9me3 is not absolutely required for targeting HUSH to its chromatin loci, as HUSH can bind chromatin in SETDB1 knockout cells, through binding of Periphilin to nascent transcripts (12).

Our biochemical, structural, and modeling data have major implications for the assembly and oligomeric state of the HUSH complex. We found that MPP8 and TASOR form a heterodimer. We showed previously that in the core Periphilin-TASOR complex two molecules

of Periphilin bind to a single copy of TASOR (15). Consequently, the HUSH complex has a stoichiometry of 1:1:2 MPP8:TASOR:Periphilin, or a multiple thereof. Self-aggregation of the N-terminal domain of Periphilin will draw in additional copies of Periphilin and hence incorporate Periphilin-bound nascent transcripts into higher-order molecular condensates (15). These additional Periphilin molecules could recruit additional copies of MPP8 and TASOR, which would allow simultaneous engagement of HUSH with H3K9me3 and SETDB1/ATF7IP via MPP8, with one copy of MPP8 bound to (reading) an H3K9me3 peptide and the other copy of MPP8 bound to SETDB1/ATF7IP, which could then deposit (write) new H3K9me3 marks. Alternatively, this higher-order assembly may serve to increase the stability and local concentration of HUSH subunits and effectors by providing multivalent interactions between the proteins in the assembly, and with nucleosomes and nascent transcripts at target loci. Stabilization of HUSH-methyltransferase complexes through multivalency, or avidity, could explain why mutation of the lysine in the H3K9-like peptide of ATF7IP caused a loss of MPP8 chromatin binding and delayed HUSH-dependent repression of a reporter transgene in mouse embryonic stem cells (mESCs) (11,30). Similarly, a conditional knockout of SETDB1 led to loss of MPP8 recruitment to HUSH target loci along with a reduction in H3K9me3 (22). Therefore, we propose that multivalent binding of the HUSH complex via MPP8 to SETDB1/ATF7IP contributes to chromatin binding of the complex. Complicating this interpretation, however, there is evidence that MPP8 can bind chromatin outside of the HUSH complex, as a subset of MPP8 chromatin binding sites lack TASOR and Periphilin (16). Consistent with this, MPP8 recruits the NEXT complex to chromatin via an interaction with ZCCHC8, leading to exosome-dependent decay of shorter (<1 kb) non-polyadenylated transcripts, whereas HUSH represses polyadenylated transcripts (21).

The data presented here allows us to propose an updated model for transcriptional repression by the HUSH complex (**Figure 4E**). The N-terminal domain of Periphilin self-aggregates and binds nascent transcripts with intronless, adenine-rich gene bodies longer than 1.5 kb, resulting in large ribonucleoparticles or condensates containing multiple copies of Periphilin bound to each transcript. The pseudo-PARP domain of TASOR may also contribute to transcript binding (14). Transcripts within these ribonucleoparticles would be less accessible to mRNA processing and export machinery, thereby repressing expression. Since ribonucleoparticles will contain many copies of Periphilin, multiple copies of TASOR and then MPP8 are recruited. Some of the MPP8 CDs will bind H3K9me3, if present, and other MPP8 subunits will recruit SETDB1/ATF7IP and possibly G9a/GLP complexes via interactions between the MPP8 CD and methylated H3K9-like sequences in the methyltransferase

components. The recruited SETDB1 deposits new H3K9me3 marks providing a read/write amplification mechanism. MORC2 is subsequently recruited to drive chromatin compaction and remodeling. How MORC2 is recruited remains unclear. Many other questions remain, such as for example what the HUSH-independent function of MPP8 may be, how domains other than the MPP8 CD contribute to recruitment of HUSH to retrotransposons, the roles of the TASOR pseudo-PARP and DomII/PIN domains, and how the requirements for the establishment of transcriptional repression by the HUSH complex may differ from the requirements to maintain repression, once established. Further studies are warranted to address these questions and test the model proposed above.

The HUSH complex fulfills a function vital to the cell in protecting the genome and repressing expression of integrated retroelements of endogenous or viral origin. Our work identifies novel structural and biochemical properties of MPP8 that are necessary for HUSH complex assembly and transcriptional repression. In the context of previous studies, the data presented here allows us to update the working model of HUSH-dependent transcriptional repression to include multivalent binding to H3K9me3 and H3K9-like sequences in methyltransferases resulting in spreading of the H3K9me3 mark through a read/write mechanism. HUSH inhibitors could be useful to stimulate autoinflammation in cancer immunotherapy or help bring retroviruses out of latency so they can be treated with antiretrovirals. Targeting the MPP8 CD may be useful in applications that require fractional inhibition of HUSH activity to limit toxicity. A family of peptidomimetic MPP8 CD ligands has been identified, providing proof of principle that MPP8 CD-specific molecules can be developed (53). Our structural data provide a foundation for further design and optimization of HUSH inhibitors targeting MPP8.

DATA AVAILABILITY STATEMENT

The structure factors and atomic coordinates for the MPP8 CTD are available in the Protein Data Bank with code PDB: 8QFB, at https://dx.doi.org/10.2210/pdb8QFB/pdb. The original experimental X-ray diffraction images are available in the SBGrid Data Bank (data.SBGrid.org) with Data ID 1043, at https://dx.doi.org/10.15785/SBGRID/1043. The models are available in ModelArchive (modelarchive.org) with the accession codes ma-t4h1j, ma-awwwj, ma-jz826, ma-zbsvq, ma-mlhvc, and ma-58pyq, at https://dx.doi.org/10.5452/ma-t4h1j, https://dx.doi.org/10.5452/ma-awwwj, https://dx.doi.org/10.5452/ma-jz826,

https://dx.doi.org/10.5452/ma-zbsvq, https://dx.doi.org/10.5452/ma-mlhvc, and https://dx.doi.org/10.5452/ma-58pyq.

SUPPLEMENTARY DATA

See Supplementary Data file for Supplementary Figures S1 to S5.

ACKNOWLEDGEMENTS

We thank Fabrice Gorrec (MRC-LMB Crystallization Facility) for advice on protein crystallization, and staff at the MRC-LMB X-ray Crystallography Facility for advice in crystallographic data collection. We thank Chris Batters (MRC-LMB Biophysics Facility) for advice and training in SEC-MALS. We are grateful to Paul Lehner (University of Cambridge Department of Medicine) for helpful comments on the manuscript. Crystallographic data were collected on beamline I04-1 at Diamond Light Source (DLS).

FUNDING

This work was supported by the Wellcome Trust [101908/Z/13/Z to Y.M.; 217191/Z/19/Z to Y.M.]. Access to Diamond Light Source was supported by the Wellcome Trust; the Medical Research Council; and UKRI [proposal number MX21426]. Open access publication was funded by the University of Cambridge.

AUTHOR CONTRIBUTIONS

Conceptualization, D.M.P., S.O. and Y.M.; Methodology, N.N., D.M.P., S.O. and Y.M.; Investigation, N.N., D.M.P, S.O., and Y.M.; Validation – crystal structure, Y.M.; Writing, Y.M.; Visualization, N.N. and Y.M.; Supervision, Y.M.; Project Administration, Y.M.; Funding Acquisition, Y.M.

CONFLICT OF INTEREST

Y.M. is a consultant for Related Sciences LLC and has profits interests in Danger Bio LLC. None of the other authors have any conflicts of interest.

REFERENCES

1. Goodier, J.L. (2016) Restricting retrotransposons: a review. *Mob DNA*, **7**, 16.

2. Friedli, M. and Trono, D. (2015) The developmental control of transposable elements and the evolution of higher species. *Annu Rev Cell Dev Biol*, **31**, 429-451.
3. Chuong, E.B., Elde, N.C. and Feschotte, C. (2016) Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science*, **351**, 1083-1087.
4. Zhou, L., Mitra, R., Atkinson, P.W., Hickman, A.B., Dyda, F. and Craig, N.L. (2004) Transposition of hAT elements links transposable elements and V(D)J recombination. *Nature*, **432**, 995-1001.
5. Dupressoir, A., Lavialle, C. and Heidmann, T. (2012) From ancestral infectious retroviruses to bona fide cellular genes: role of the captured syncytins in placentation. *Placenta*, **33**, 663-671.
6. Pastuzyn, E.D., Day, C.E., Kearns, R.B., Kyrke-Smith, M., Taibi, A.V., McCormick, J., Yoder, N., Belnap, D.M., Erlendsson, S., Morado, D.R. *et al.* (2018) The Neuronal Gene Arc Encodes a Repurposed Retrotransposon Gag Protein that Mediates Intercellular RNA Transfer. *Cell*, **172**, 275-288.e218.
7. Hung, T., Pratt, G.A., Sundararaman, B., Townsend, M.J., Chaivorapol, C., Bhangale, T., Graham, R.R., Ortmann, W., Criswell, L.A., Yeo, G.W. and Behrens, T.W. (2015) The Ro60 autoantigen binds endogenous retroelements and regulates inflammatory gene expression. *Science*, **350**, 455-459.
8. Hancks, D.C. and Kazazian, H.H., Jr. (2016) Roles for retrotransposon insertions in human disease. *Mob DNA*, **7**, 9.
9. Lamprecht, B., Walter, K., Kreher, S., Kumar, R., Hummel, M., Lenze, D., Kochert, K., Bouhlel, M.A., Richter, J., Soler, E. *et al.* (2010) Derepression of an endogenous long terminal repeat activates the CSF1R proto-oncogene in human lymphoma. *Nat Med*, **16**, 571-579, 571p following 579.
10. Azebi, S., Batsche, E., Michel, F., Kornobis, E. and Muchardt, C. (2019) Expression of endogenous retroviruses reflects increased usage of atypical enhancers in T cells. *EMBO J*, **38**.
11. Tchasovnikarova, I.A., Timms, R.T., Matheson, N.J., Wals, K., Antrobus, R., Gottgens, B., Dougan, G., Dawson, M.A. and Lehner, P.J. (2015) GENE SILENCING. Epigenetic silencing by the HUSH complex mediates position-effect variegation in human cells. *Science*, **348**, 1481-1485.
12. Seczynska, M., Bloor, S., Cuesta, S.M. and Lehner, P.J. (2022) Genome surveillance by HUSH-mediated silencing of intronless mobile elements. *Nature*, **601**, 440-445.

13. Liu, N., Lee, C.H., Swigut, T., Grow, E., Gu, B., Bassik, M.C. and Wysocka, J. (2018) Selective silencing of euchromatic L1s revealed by genome-wide screens for L1 regulators. *Nature*, **553**, 228-232.
14. Douse, C.H., Tchasovnikarova, I.A., Timms, R.T., Protasio, A.V., Seczynska, M., Prigozhin, D.M., Albecka, A., Wagstaff, J., Williamson, J.C., Freund, S.M.V. *et al.* (2020) TASOR is a pseudo-PARP that directs HUSH complex assembly and epigenetic transposon control. *Nat Commun*, **11**, 4940.
15. Prigozhin, D.M., Douse, C.H., Farleigh, L.E., Albecka, A., Tchasovnikarova, I.A., Timms, R.T., Oda, S.I., Adolf, F., Freund, S.M.V., Maslen, S. *et al.* (2020) Periphilin self-association underpins epigenetic silencing by the HUSH complex. *Nucleic Acids Res*, **48**, 10313-10328.
16. Spencley, A.L., Bar, S., Swigut, T., Flynn, R.A., Lee, C.H., Chen, L.F., Bassik, M.C. and Wysocka, J. (2023) Co-transcriptional genome surveillance by HUSH is coupled to termination machinery. *Mol Cell*, **83**, 1623-1639 e1628.
17. Schopp, T., Prigozhin, D.M., Douse, C., Kaji, K., Cook, A.G. and O'Carroll, D. (2023) The DUF3715 domain has a conserved role in RNA-directed transposon silencing. *Rna*, **29**, 1471-1480.
18. Timms, R.T., Tchasovnikarova, I.A., Antrobus, R., Dougan, G. and Lehner, P.J. (2016) ATF7IP-Mediated Stabilization of the Histone Methyltransferase SETDB1 Is Essential for Heterochromatin Formation by the HUSH Complex. *Cell reports*, **17**, 653-659.
19. Tchasovnikarova, I.A., Timms, R.T., Douse, C.H., Roberts, R.C., Dougan, G., Kingston, R.E., Modis, Y. and Lehner, P.J. (2017) Hyperactivation of HUSH complex function by Charcot-Marie-Tooth disease mutation in MORC2. *Nat Genet*, **49**, 1035-1044.
20. Douse, C.H., Bloor, S., Liu, Y., Shamin, M., Tchasovnikarova, I.A., Timms, R.T., Lehner, P.J. and Modis, Y. (2018) Neuropathic MORC2 mutations perturb GHKL ATPase dimerization dynamics and epigenetic silencing by multiple structural mechanisms. *Nat Commun*, **9**, 651.
21. Garland, W., Muller, I., Wu, M., Schmid, M., Imamura, K., Rib, L., Sandelin, A., Helin, K. and Jensen, T.H. (2022) Chromatin modifier HUSH co-operates with RNA decay factor NEXT to restrict transposable element expression. *Mol Cell*, **82**, 1691-1707 e1698.
22. Muller, I., Moroni, A.S., Shlyueva, D., Sahadevan, S., Schoof, E.M., Radzisheuskaya, A., Hojfeldt, J.W., Tatar, T., Koche, R.P., Huang, C. and Helin, K. (2021) MPP8 is essential for sustaining self-renewal of ground-state pluripotent stem cells. *Nat Commun*, **12**, 3034.
23. Tunbak, H., Enriquez-Gasca, R., Tie, C.H.C., Gould, P.A., Mlcochova, P., Gupta, R.K., Fernandes, L., Holt, J., van der Veen, A.G., Giampazolias, E. *et al.* (2020) The HUSH

complex is a gatekeeper of type I interferon through epigenetic regulation of LINE-1s. *Nat Commun*, **11**, 5387.

24. Gu, Z., Liu, Y., Zhang, Y., Cao, H., Lyu, J., Wang, X., Wylie, A., Newkirk, S.J., Jones, A.E., Lee, M., Jr. *et al.* (2021) Silencing of LINE-1 retrotransposons is a selective dependency of myeloid leukemia. *Nat Genet*, **53**, 672-682.
25. Griffin, G.K., Wu, J., Iracheta-Vellve, A., Patti, J.C., Hsu, J., Davis, T., Dele-Oni, D., Du, P.P., Halawi, A.G., Ishizuka, J.J. *et al.* (2021) Epigenetic silencing by SETDB1 suppresses tumour intrinsic immunogenicity. *Nature*, **595**, 309-314.
26. Quinn, A.M., Bedford, M.T., Espejo, A., Spannhoff, A., Austin, C.P., Oppermann, U. and Simeonov, A. (2010) A homogeneous method for investigation of methylation-dependent protein-protein interactions in epigenetics. *Nucleic Acids Res*, **38**, e11.
27. Bua, D.J., Kuo, A.J., Cheung, P., Liu, C.L., Migliori, V., Espejo, A., Casadio, F., Bassi, C., Amati, B., Bedford, M.T. *et al.* (2009) Epigenome microarray platform for proteome-wide dissection of chromatin-signaling networks. *PLoS One*, **4**, e6789.
28. Chang, Y., Horton, J.R., Bedford, M.T., Zhang, X. and Cheng, X. (2011) Structural insights for MPP8 chromodomain interaction with histone H3 lysine 9: potential effect of phosphorylation on methyl-lysine binding. *J Mol Biol*, **408**, 807-814.
29. Sampath, S.C., Marazzi, I., Yap, K.L., Sampath, S.C., Krutchinsky, A.N., Mecklenbrauker, I., Viale, A., Rudensky, E., Zhou, M.M., Chait, B.T. and Tarakhovsky, A. (2007) Methylation of a histone mimic within the histone methyltransferase G9a regulates protein complex assembly. *Mol Cell*, **27**, 596-608.
30. Tsusaka, T., Kikuchi, M., Shimazu, T., Suzuki, T., Sohtome, Y., Akakabe, M., Sodeoka, M., Dohmae, N., Umehara, T. and Shinkai, Y. (2018) Tri-methylation of ATF7IP by G9a/GLP recruits the chromodomain protein MPP8. *Epigenetics Chromatin*, **11**, 56.
31. Chin, H.G., Esteve, P.O., Pradhan, M., Benner, J., Patnaik, D., Carey, M.F. and Pradhan, S. (2007) Automethylation of G9a and its implication in wider substrate specificity and HP1 binding. *Nucleic Acids Res*, **35**, 7313-7323.
32. Chang, Y., Sun, L., Kokura, K., Horton, J.R., Fukuda, M., Espejo, A., Izumi, V., Koomen, J.M., Bedford, M.T., Zhang, X. *et al.* (2011) MPP8 mediates the interactions between DNA methyltransferase Dnmt3a and H3K9 methyltransferase GLP/G9a. *Nat Commun*, **2**, 533.
33. Kokura, K., Sun, L., Bedford, M.T. and Fang, J. (2010) Methyl-H3K9-binding protein MPP8 mediates E-cadherin gene silencing and promotes tumour cell motility and invasion. *EMBO J*, **29**, 3673-3687.

34. Winter, G. (2010) xia2: an expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography*, **43**, 186-190.
35. Mirdita, M., Schutze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S. and Steinegger, M. (2022) ColabFold: making protein folding accessible to all. *Nat Methods*, **19**, 679-682.
36. Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr*, **66**, 213-221.
37. Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*, **60**, 2126-2132.
38. Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Žídek, A., Bates, R., Blackwell, S., Yim, J. *et al.* (2021) Protein complex prediction with AlphaFold-Multimer. *bioRxiv*, 2021.2010.2004.463034.
39. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A. *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**, 583-589.
40. Krissinel, E. and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol*, **372**, 774-797.
41. Holm, L. and Rosenstrom, P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res*, **38**, W545-549.
42. Duval, D., Duval, G., Kedinger, C., Poch, O. and Boeuf, H. (2003) The 'PINIT' motif, of a newly identified conserved domain of the PIAS protein family, is essential for nuclear retention of PIAS3L. *FEBS Lett*, **554**, 111-118.
43. Yunus, A.A. and Lima, C.D. (2009) Structure of the Siz/PIAS SUMO E3 ligase Siz1 and determinants required for SUMO modification of PCNA. *Mol Cell*, **35**, 669-682.
44. Streich, F.C., Jr. and Lima, C.D. (2016) Capturing a substrate in an activated RING E3/E2-SUMO complex. *Nature*, **536**, 304-308.
45. Pfander, B., Moldovan, G.L., Sacher, M., Hoege, C. and Jentsch, S. (2005) SUMO-modified PCNA recruits Srs2 to prevent recombination during S phase. *Nature*, **436**, 428-433.
46. Ban, R., Nishida, T. and Urano, T. (2011) Mitotic kinase Aurora-B is regulated by SUMO-2/3 conjugation/deconjugation during mitosis. *Genes Cells*, **16**, 652-669.
47. Josa-Prado, F., Henley, J.M. and Wilkinson, K.A. (2015) SUMOylation of Argonaute-2 regulates RNA interference activity. *Biochem Biophys Res Commun*, **464**, 1066-1071.

48. Cox, E., Hwang, W., Uzoma, I., Hu, J., Guzzo, C.M., Jeong, J., Matunis, M.J., Qian, J., Zhu, H. and Blackshaw, S. (2017) Global Analysis of SUMO-Binding Proteins Identifies SUMOylation as a Key Regulator of the INO80 Chromatin Remodeling Complex. *Molecular & cellular proteomics : MCP*, **16**, 812-823.
49. Salas-Lloret, D., Jansen, N.S., Nagamalleswari, E., van der Meulen, C., Gracheva, E., de Ru, A.H., Otte, H.A.M., van Veelen, P.A., Pichler, A., Goedhart, J. *et al.* (2023) SUMO-activated target traps (SATTs) enable the identification of a comprehensive E3-specific SUMO proteome. *Sci Adv*, **9**, eadh2073.
50. Hendriks, I.A., Lyon, D., Young, C., Jensen, L.J., Vertegaal, A.C. and Nielsen, M.L. (2017) Site-specific mapping of the human SUMO proteome reveals co-modification with phosphorylation. *Nat Struct Mol Biol*, **24**, 325-336.
51. Appel, L.M., Franke, V., Bruno, M., Grishkovskaya, I., Kasiliauskaite, A., Kaufmann, T., Schoeberl, U.E., Puchinger, M.G., Kostrhon, S., Ebenwaldner, C. *et al.* (2021) PHF3 regulates neuronal gene expression through the Pol II CTD reader domain SPOC. *Nat Commun*, **12**, 6078.
52. Ochoa, D., Jarnuczak, A.F., Vieitez, C., Gehre, M., Soucheray, M., Mateus, A., Kleefeldt, A.A., Hill, A., Garcia-Alonso, L., Stein, F. *et al.* (2020) The functional landscape of the human phosphoproteome. *Nat Biotechnol*, **38**, 365-373.
53. Waybright, J.M., Clinkscales, S.E., Barnash, K.D., Budziszewski, G.R., Rectenwald, J.M., Chiarella, A.M., Norris-Drouin, J.L., Cholensky, S.H., Pearce, K.H., Herring, L.E. *et al.* (2021) A Peptidomimetic Ligand Targeting the Chromodomain of MPP8 Reveals HRP2's Association with the HUSH Complex. *ACS Chem Biol*, **16**, 1721-1736.
54. Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H. and Ferrin, T.E. (2021) UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci*, **30**, 70-82.
55. Bond, C.S. (2003) TopDraw: a sketchpad for protein structure topology cartoons. *Bioinformatics*, **19**, 311-312.
56. Crooks, G.E., Hon, G., Chandonia, J.M. and Brenner, S.E. (2004) WebLogo: a sequence logo generator. *Genome Res*, **14**, 1188-1190.

TABLE and FIGURE LEGENDS

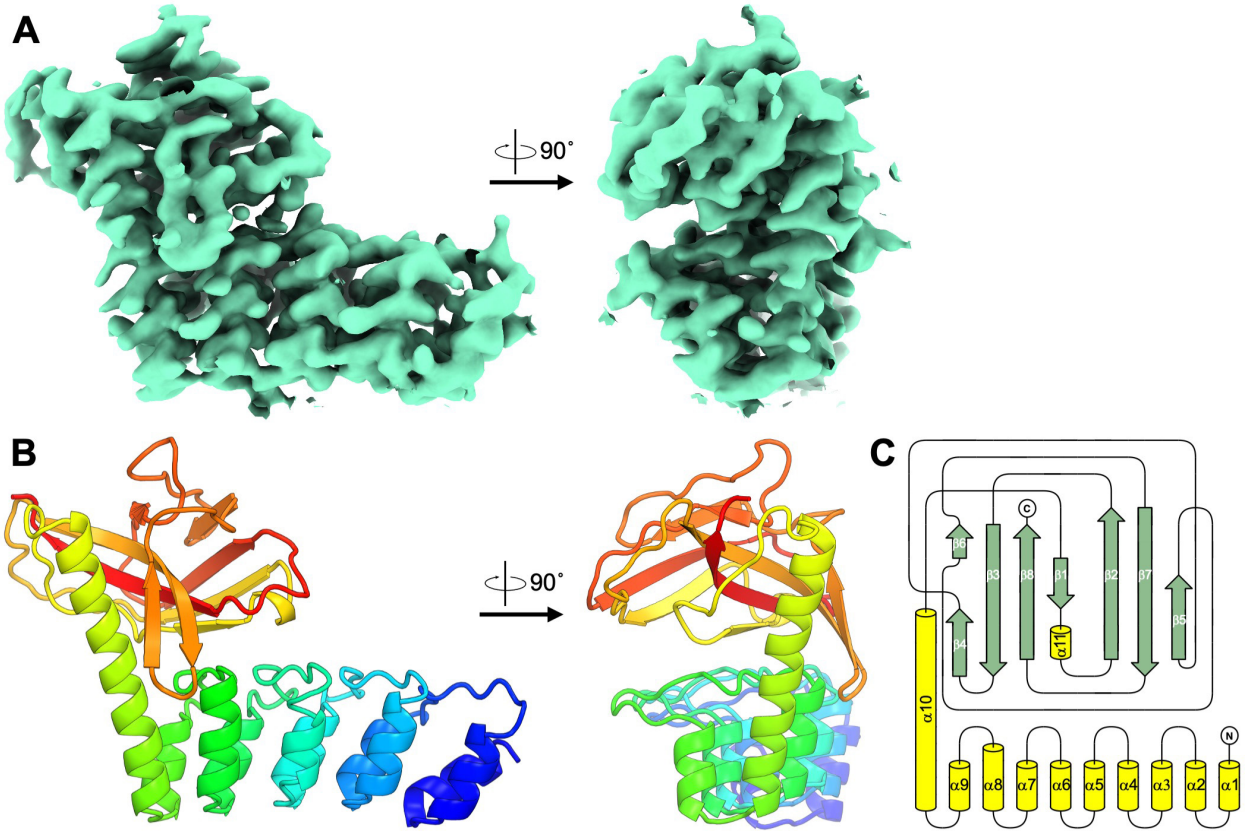


Figure 1. Crystal structure of the MPP8 C-terminal domain (CTD).

(**A**) Weighted $2F_o - F_c$ electron density map for one subunit of the MPP8 CTD. Images created with ChimeraX (54) using contour level 0.104. See **Table 1** for crystallographic data collection, refinement, and validation statistics. Representative electron density snapshots are shown in **Supplementary Figure S1**.

(**B**) Crystal structure of the MPP8 CTD in cartoon representation with rainbow coloring (blue, N-terminus; red, C-terminus).

(**C**) Secondary structure topology of the MPP8 CTD. Diagram created with TopDraw (55).

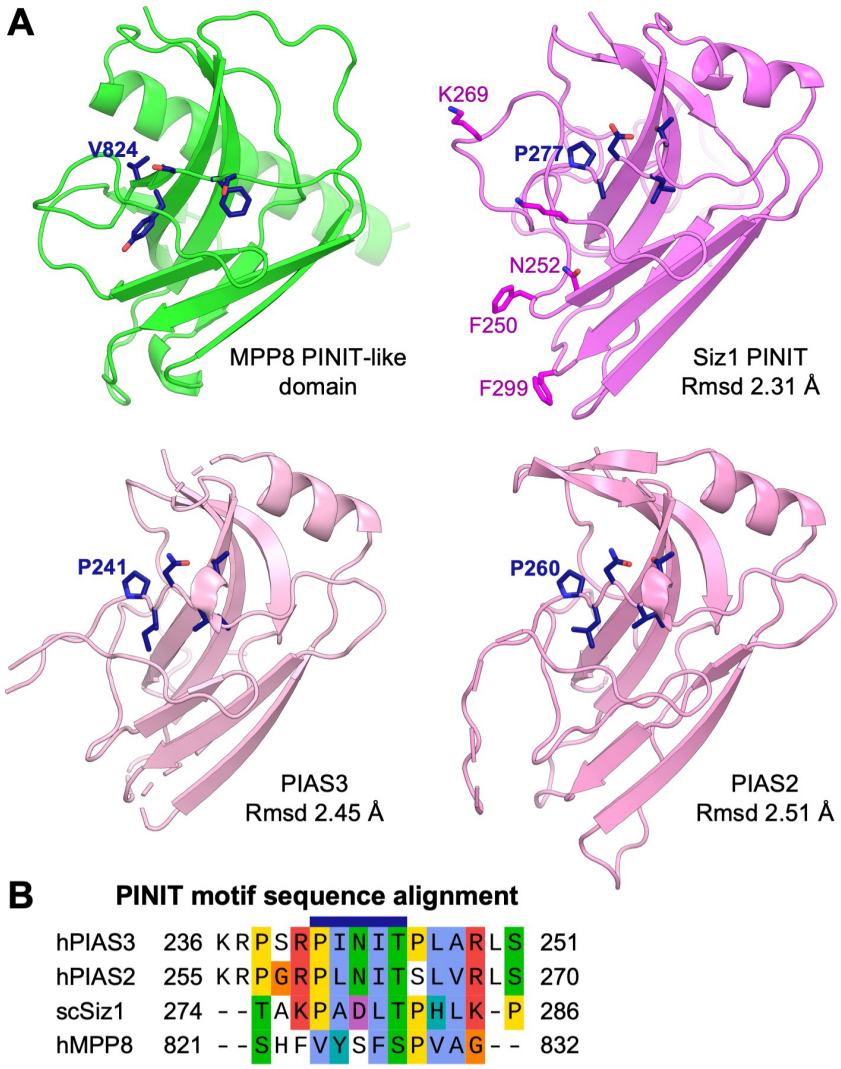


Figure 2. Structural homology of the MPP8 CTD β-sandwich to the PINIT domains from Siz1, PIAS2 and PIAS3.

(**A**) Crystal structures of the MPP8 PINIT-like domain and the PINIT domains from Siz/PIAS-family E3 SUMO ligases *Saccharomyces cerevisiae* Siz1 (43), human PIAS2 (PDB 4FO9), and human PIAS3 (PDB 4MVT). Rmsd, root mean square deviation between atoms in a superposition with MPP8.

(**B**) Multiple sequence alignment of the PINIT motif (highlighted with a blue bar) and flanking sequences of PIAS3, PIAS2, Siz1 with the MPP8 sequence aligned based on structural superpositions.

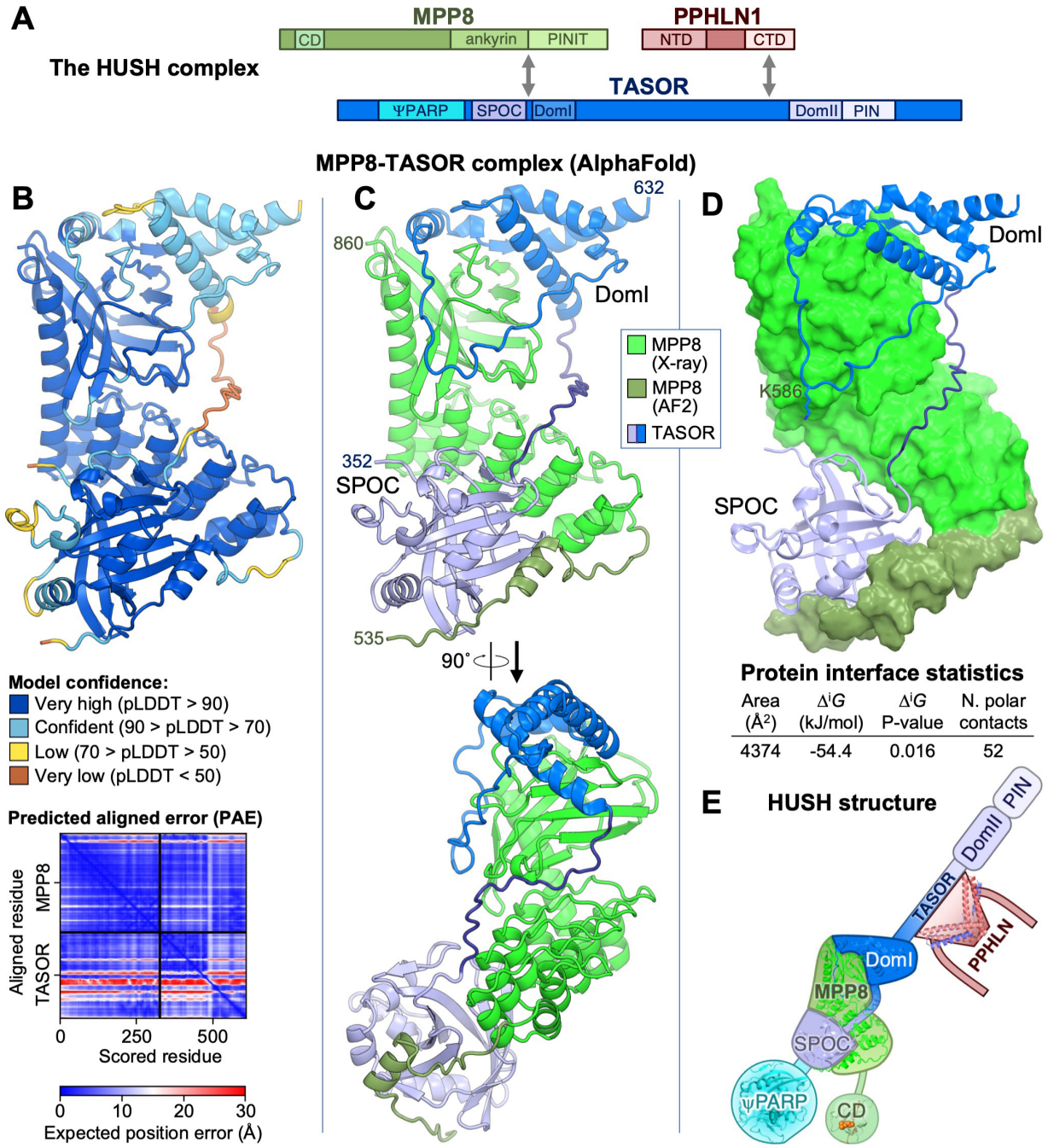


Figure 3. Structural modeling of the MPP8-TASOR complex with AlphaFold-Multimer.

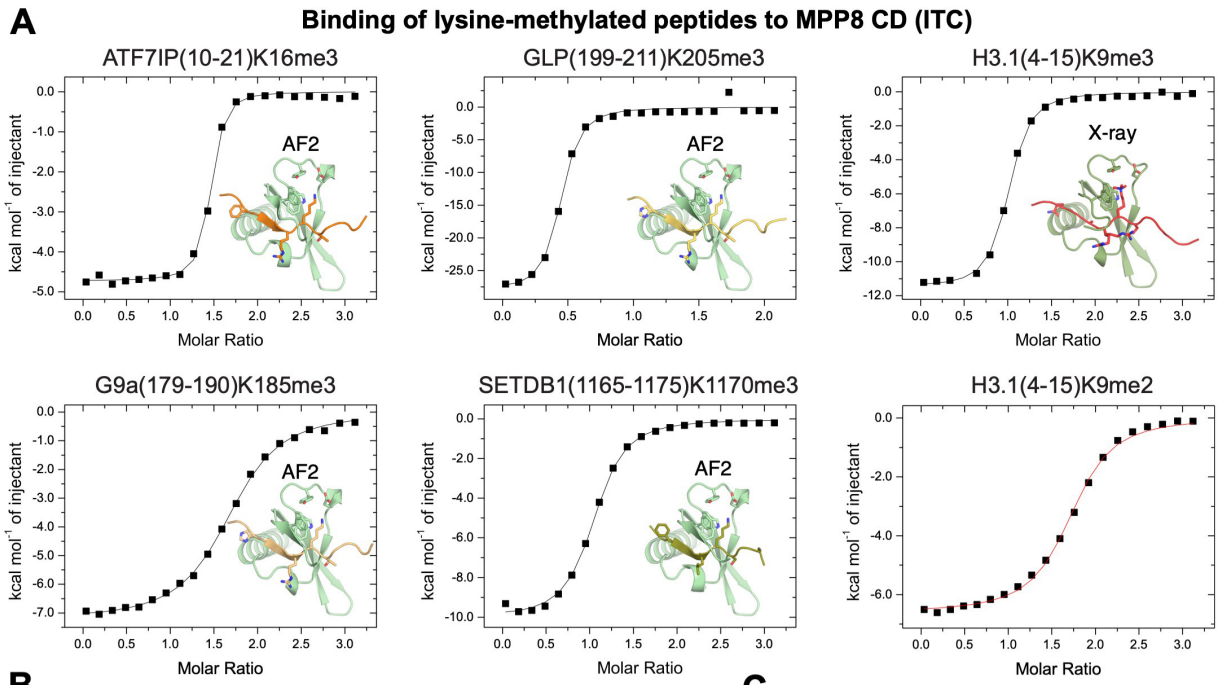
(**A**) Domain diagram of the HUSH complex (drawn to scale).

(**B**) AlphaFold model of the MPP8-TASOR complex colored by model confidence (pLDDT, according to the color code shown). Lower panel: predicted aligned error plot.

(**C**) AlphaFold model of the MPP8-TASOR complex colored by domain.

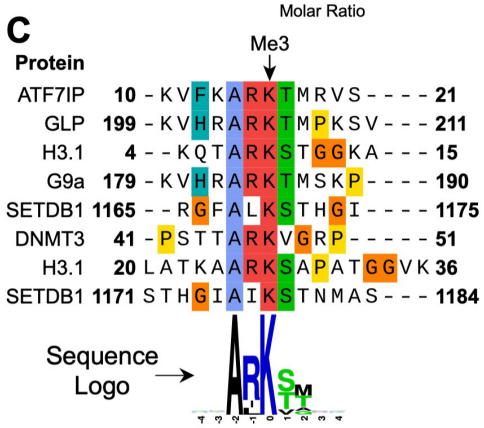
(**D**) AlphaFold model of the MPP8-TASOR complex with MPP8 in surface representation and with PISA protein interface parameters. See **Supplementary Figure S3** for comparison

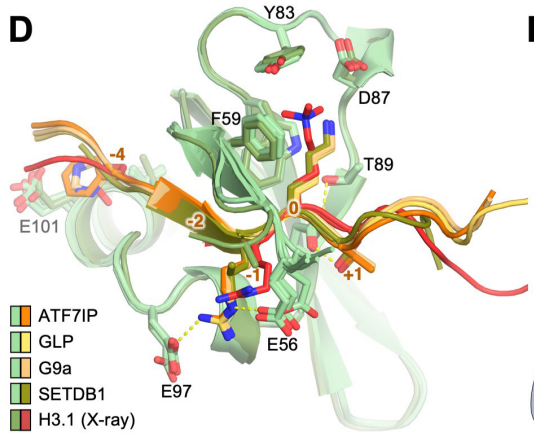
between the AlphaFold model of the TASOR SPOC domain and the crystal structure of the PHF3 SPOC domain bound to RNA Pol II CTD di-heptapeptide (Ser2-phosphorylated).
(**E**) Schematic of the HUSH complex with available structures overlaid.



B

Protein	Res. range	Methylation	K_D	N	ΔH (kJ/mol)	$T\Delta S$ (kJ/mol)	Interface area (Å^2)
ATF7IP	10-21	K16me3	193 ± 31 nM	1.4	-19.8 ± 0.1	18.5	-
GLP	199-211	K205me3	337 ± 64 nM	0.4	-118 ± 2	-80.5	-
H3.1	4-15	K9me3	769 ± 65 nM	1.0	-48.1 ± 0.4	-13.1	721#
G9a	179-190	K185me3	1.31 ± 0.07 µM	1.7	-30.2 ± 0.2	3.34	-
SETDB1	1165-1175	K1170me3	1.55 ± 0.12 µM	1.0	-42.0 ± 0.4	-8.94	-
H3.1	4-15	K9me2	1.70 ± 0.14 µM	1.7	-27.6 ± 0.2	4.85	-
DNMT3	41-51	K47me2	12 µM*	-	-	-	643#
H3.1	20-36	K27me3	>50 µM	-	-	-	-
SETDB1	1171-1184	K1178me3	>80 µM	-	-	-	-
ATF7IP	10-21	None	No binding	-	-	-	720‡
GLP	199-211	None	No binding	-	-	-	717‡
H3.1	4-15	None	No binding	-	-	-	758‡
G9a	179-190	None	No binding	-	-	-	726‡
SETDB1	1165-1175	None	No binding	-	-	-	661‡
DNMT3	45-51	None	No binding	-	-	-	738‡
SETDB1	1171-1184	None	No binding	-	-	-	734‡





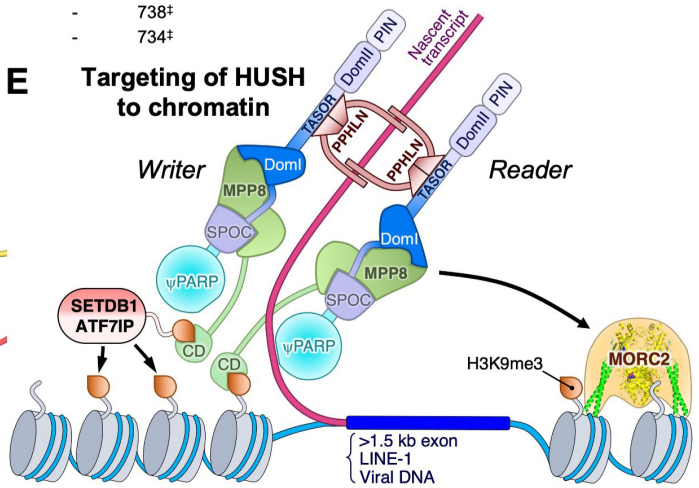


Figure 4. Binding of the MPP8 CD to lysine-methylated H3K9 and H3K9-like peptides. (**A**) Isothermal titration calorimetry (ITC) thermograms from titration of six different lysine methylated peptides into a solution of MPP8 CD. Insets, AlphaFold models ("AF2") and crystal

structure ("X-ray") (28) of the MPP8 CD bound to the indicated peptide. See **Supplementary Figure S4** for the complete set of ITC isotherms and thermograms.

(**B**) Table of dissociation constants (K_D), binding stoichiometry (*N*) and thermodynamic parameters derived from ITC data. The listed protein interface areas are for crystal structures ("#") of MPP8 CD bound to H3K9me3 (PDB 3QO2) (28) and hDnmt3aK47me2 (PDB 3SVM) (32); and AlphaFold models ("‡") of MPP8 CD bound to the indicated H3K9-like sequences.

(**C**) Multiple sequence alignment of H3K9 and the H3K9-like sequences from the ITC measurements. UniProt accession numbers: ATF7IP, Q6VMQ6; GLP, Q9H9B1; H3.1, P68431; G9a, Q96KQ7; SETDB1, Q15047; DNMT3a, Q9Y6K1. Lower panel: sequence logo generated with WebLogo (56).

(**D**) Superposition of the AlphaFold models and crystal structure shown in (A). Numbers in orange show the residue position relative to the methyl-lysine.

(**E**) Model for targeting of the HUSH complex with multivalent engagement to actively transcribed retrotransposons.

Table 1. Crystallographic data collection, refinement, and validation statistics for MPP8 CTD.

Space group	$P\ 2_1\ 2_1\ 2_1$
Cell dimensions	
a, b, c (Å)	102.56 179.53 228.08
α, β, γ (°)	90, 90, 90
Resolution (Å)	93.5 – 3.0 (3.107 – 3.0)
Total reflections	1,097,353 (94,635)
Unique reflections	84,956 (8,368)
Multiplicity	12.9 (11.3)
Completeness	98.6 (94.2)
$\langle I \rangle / \sigma(I)$	14.36 (1.46)
Wilson *B*-factor	99.7
R_{merge}	0.167 (3.20)
R_{pim}	0.0482 (0.979)
CC1/2	0.999 (0.266)
Refinement and Validation	
Resolution (Å)	43 - 1.72 (1.79 - 1.72)
No. reflections, working set	83,763 (7,905)
No. reflections, test set	4,291 (420)
R_{work}	0.2696 (0.5247)
R_{free}	0.2911 (0.5203)
No. non-hydrogen atoms	18,150

No. ligand or solvent atoms	0
No. MPP8 chains per a.s.u.	8
Average *B*-factor	116.85
RMSD	
Bond lengths (Å)	0.014
Bond angles (°)	1.68
Ramachandran plot	
Favored (%)	91.23
Allowed (%)	7.01
Outliers (%)	1.75
Clashscore (Phenix 1.20.1)	35.2
PDB accession code	8QFB

Numbers in parentheses are for the highest resolution shell