The structure of the cysteine-rich region from human histone-lysine N-methyltransferase EHMT2 (G9a)

Keshia M. Kerchner [a], Tung-Chung Mou [b,c], Yizhi Sun [a], Domniţa-Valeria Rusnac [a], Stephen R. Sprang [b,c], Klára Briknarová [a,c,*]

[a] Department of Chemistry and Biochemistry, University of Montana, Missoula, MT 59812, USA
[b] Division of Biological Sciences, University of Montana, Missoula, MT 59812, USA
[c] Center for Biomolecular Structure and Dynamics, University of Montana, Missoula, MT 59812, USA

ARTICLE INFO

ABSTRACT

Euchromatic histone-lysine N-methyltransferase 1 (EHMT1; G9a-like protein; GLP) and euchromatic histone-lysine N-methyltransferase 2 (EHMT2; G9a) are protein lysine methyltransferases that regulate gene expression and are essential for development and the ability of organisms to change and adapt. In addition to ankyrin repeats and the catalytic SET domain, the EHMT proteins contain a unique cysteine-rich region (CRR) that mediates protein–protein interactions and recruitment of the methyltransferases to specific sites in chromatin. We have determined the structure of the CRR from human EHMT2 by X-ray crystallography and show that the CRR adopts an unusual compact fold with four bound zinc atoms. The structure consists of a RING domain preceded by a smaller zinc-binding motif and an N-terminal segment. The smaller zinc-binding motif straddles the N-terminal end of the RING domain, and the N-terminal segment runs in an extended conformation along one side of the structure and interacts with both the smaller zinc-binding motif and the RING domain. The interface between the N-terminal segment and the RING domain includes one of the zinc atoms. The RING domain is partially sequestered within the CRR and unlikely to function as a ubiquitin ligase.

Introduction

Euchromatic histone-lysine N-methyltransferase 2 (EHMT2; also known as G9a) (Brown et al., 2001; Milner and Campbell, 1993) and the closely related euchromatic histone-lysine N-methyltransferase 1 (EHMT1; G9a-like protein; GLP) (Ogawa et al., 2002) are protein lysine methyltransferases that regulate gene expression and are essential for development and the ability of organisms to change and adapt. They play a role in many processes including cellular differentiation, immune response, behavior and cognition (Benevento et al., 2015; Casciello et al., 2015; Kramer, 2016; Scheer and Zaph, 2017; Shankar et al., 2013; Shinkai and Tachibana, 2011). They are primarily responsible for dimethylation of K9 in histone H3 in euchromatic regions, which is associated with transcriptional repression (Ogawa et al., 2002; Tachibana et al., 2001, 2002, 2005); they also methylate other lysine residues in both histone and non-histone proteins and have functions that are independent of their methyltransferase activity (Deimling et al., 2017; Scheer and Zaph, 2017; Shankar et al., 2013; Tachibana et al., 2001).

Human EHMT2 (hEHMT2) protein comprises 1210 residues and can be divided into several distinct regions: the N-terminal one third that is predicted to be mostly disordered, a cysteine-rich region (CRR), eight ankyrin repeats, and a SET domain at the C-terminus, flanked by pre-SET and post-SET regions (Fig. 1A). Human EHMT1 (hEHMT1) is organized in a similar manner. The SET domain together with the pre-SET and post-SET regions is responsible for the lysine methyltransferase activity, and a number of structures with bound products and various inhibitors are available (Chang et al., 2009; Wu et al., 2010). The ankyrin repeats bind to histone-like sequences that are mono- or dimethylated, and structures with and without a bound dimethylated peptide have also been determined (Collins et al., 2008; Dong et al., 2018).

The CRR was reported to mediate interaction of EHMT2 with cyclin D1 (Li et al., 2019), the MED12 subunit of the Mediator complex (Ding et al., 2008), and the neuron restrictive silencer factor (NRSF), also known as RE1-silencing transcription factor (REST) (Roopra et al., 2004). Recruitment of EHMT2 by cyclin D1 is required for association of chromatin with nuclear lamina (Li et al., 2019), and targeting to neuron

restrictive silencer elements (NRSE) by NRSF/REST leads to silencing of neuron-specific genes in non-neuronal cells (Roopra et al., 2004). In addition, the C-terminal portion of the CRR was shown to bind to ZNF200, a transcription factor that contains five zinc fingers (Nishida et al., 2007).

We now present the structure of the CRR, which we determined by X-ray crystallography. We show that the CRR adopts an unusual compact fold that contains an embedded RING domain. We also discuss implications of the structure for the function of the RING domain and analyze the conservation of the CRR.

Results

The CRR binds zinc and adopts a compact globular structure

In our initial studies of the CRR, we focused on residues G380-D552 of hEHMT2, which are relatively well conserved (55% identical) in hEHMT2 and hEHMT1 proteins (Fig. 1B). This stretch of residues is flanked by sequences that do not show any similarity between hEHMT2 and hEHMT1 (Fig. 1B) and are predicted to be disordered (Erdős and Dosztányi, 2020; Kelley et al., 2015).

The CRR contains an array of conserved cysteine and histidine residues (Fig. 1B), suggesting that it binds metal ions. Zinc content analysis confirmed that hEHMT2(380–552) indeed binds zinc, with a stoichiometry of 4 moles of zinc per mole of protein. The two-dimensional (2D) ^{1}H-^{15}N heteronuclear single quantum correlation (HSQC) spectrum of hEHMT2(380–552) showed well-dispersed signals, consistent with a folded protein (Supplementary Fig. S1A). However, the HSQC spectrum also contained a number of strong signals that exhibited negative or only weakly positive ^{1}H-^{15}N nuclear Overhauser effect (NOE) (Supplementary Figs. S1B and S2), which is characteristic of dynamic residues. These strong signals were readily assigned using standard 3-dimensional (3D) triple resonance experiments and were found to be from residues G380-E422 and G551-D552 (Supplementary Fig. S2).

A second construct with a different N-terminus, hEHMT2(419–552), was also prepared and characterized. hEHMT2(419–552) retains the structured portion and all the conserved cysteine and histidine residues, but it lacks most of the residues that are dynamic in hEHMT2(380–552) and is therefore better suited for structural studies. hEHMT2(419–552) is a monomer in solution, as determined by size-exclusion chromatography combined with multi-angle light scattering (SEC-MALS) (results not shown). The signals from the disordered N-terminal residues that were observed in the 2D ^{1}H-^{15}N HSQC spectrum of hEHMT2(380–552) are absent in the 2D ^{1}H-^{15}N HSQC spectrum of hEHMT2(419–552) but signals from other residues are not significantly affected (Supplementary Fig. S1C).

hEHMT2(419–552) was crystallized, and its structure was solved at 1.9 Å resolution using single-wavelength anomalous diffraction (SAD) arising from the protein-bound Zn atoms (Table 1). The asymmetric unit contained four molecules of hEHMT2(419–552), with well-defined electron density for residues G419-R550 in chains A and C and for residues G419-D552 in chains B and D. The conformations of all four molecules are very similar, with average pairwise root-mean-square deviation (RMSD) between backbone atom positions of residues L423-R550 0.52 ± 0.13 Å (Fig. 2A). The CRR adopts a compact fold and, consistent with the zinc content analysis, contains four bound zinc atoms, which we refer to as Zn(a), Zn(b), Zn(c), and Zn(d) (Figs. 1B, 2 and 3).

The CRR contains a RING domain

When we compared the structure of the CRR with structures of other proteins using the Dali server (Holm, 2019), similarity was detected between the C-terminal half of the CRR and the really interesting new gene (RING) domain fold. This was not completely unexpected as homology between the C-terminal portion of the CRR and RING domains (sequence identity ∼ 20–30%) was also detectable by the Phyre2 protein fold recognition server (Kelley et al., 2015) (our unpublished results) and had been noted previously by others (Nishida et al., 2007; Roopra et al., 2004). However, we have not identified any RING domains that are part of a larger structure similar to the CRR. With its unique N-terminal half and a total of four bound zinc atoms, the CRR therefore represents a novel zinc-binding motif.

The similarity between the CRR and RING domains is illustrated using the RING domain from human RING1b (hRING1b) protein, which is also known as RING2 (Taherbhoy et al., 2015) (Fig. 4). The backbone RMSD for the 55 residues in the CRR and hRING1b that were identified as structurally equivalent by the Dali server is 2.94 Å; only 10 of these residues (18%) are identical (Fig. 4).

The core of the RING domain consists of a short three-stranded anti-parallel β-sheet (formed by β-strands β1, β2 and β3 in hRING1b, and β3, β4 and β7 in the CRR), a helix immediately following the second β-strand (α-helix labeled α in hRING1b, and a single turn of 3₁₀ helix in the CRR), and two characteristic loops, L1 and L2 (Fig. 4). The RING domain contains two zinc atoms that are bound in an interleaved manner referred to as the cross-brace (Barlow et al., 1994; Budhidarmo et al., 2012): the first and the third pair of zinc-binding residues coordinate one zinc atom (Zn(I) in hRING1b and Zn(c) in the CRR; red in Fig. 4A), and the second and the fourth pair of zinc-binding residues coordinate the other zinc atom (Zn(II) in hRING1b and Zn(d) in the CRR; blue in Fig. 4A). The first zinc-binding site is formed by a CXXC motif in loop L1 and a CXXC motif located at the N-terminal end of helix α in hRING1b
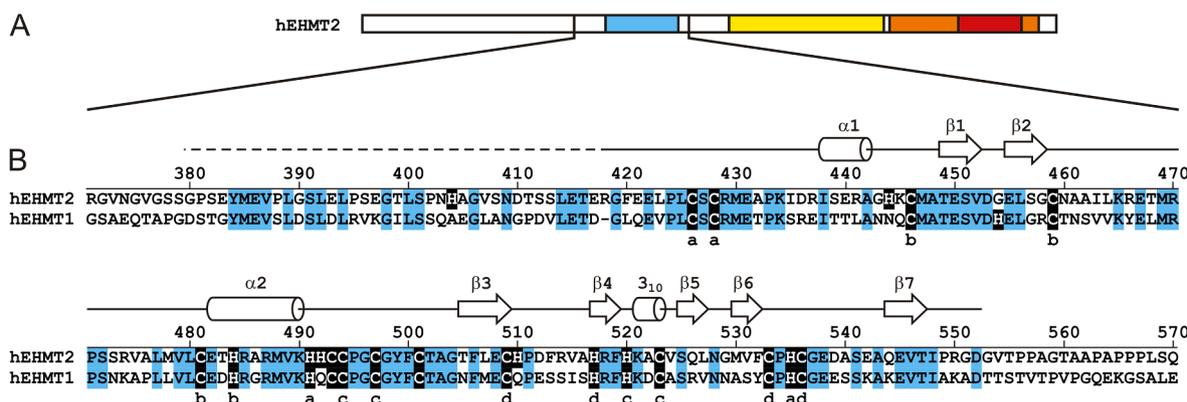
Fig. 1. The cysteine-rich region (CRR) in hEHMT2. (A) A schematic drawing of hEHMT2 with the CRR colored blue, the ankyrin repeats yellow, the SET domain red, and the pre-SET and post-SET regions orange. (B) Alignment of the CRR and adjacent sequences from hEHMT2 (UniProt Q96KQ7) and hEHMT1 (UniProt Q9H9B1). Residues that are identical in both proteins are colored blue, and all cysteine and histidine residues are highlighted in black. Secondary structure that is observed in hEHMT2(419–552) is diagramed above the sequences. The dashed line represents additional residues that were present in the hEHMT2(380–552) construct and were found to be disordered. Letters a, b, c, and d below the sequences denote residues that coordinate Zn(a), Zn(b), Zn(c) and Zn(d), respectively.

Table 1

X-ray crystallography data collection and refinement statistics.

Data set	Set #1	Set #2
Data collection		
Beamline	APS 19-BM	SSRL 12–2
Wavelength (Å)	0.98	1.28
Temperature (K)	100	100
Detector	ADSC Quantum q201r	Pilatus 6M
Rotation range per image (°)	0.5	0.2
Total rotation range (°)	360	360
Exposure time per image (s)	4.0	0.3
Transmission (%)	100	20
Resolution range (Å)*	25.00–1.90 (1.93–1.90)	25.00–2.40 (2.49–2.40)
Space group	P1	P1
Unit cell dimensions		
a, b, c (Å)	48.0, 54.4, 59.3	48.1, 54.3, 59.7
α, β, γ (°)	91.7, 100.9, 96.9	91.5, 101.0, 96.8
Unique reflections*	43,911 (4011)	20,534 (1900)
Redundancy*	2.2 (1.9)	7.2 (6.1)
Completeness (%)*	95.4 (89.9)	88.6 (89.2)
Mean I/σ (I)*	7.6 (1.6)	19.5 (7.5)
Wilson B-factor	16.7	32.7
R_{sym}[†],*	0.12 (0.73)	0.11 (0.36)
$CC_{1/2}$[‡],*	1.0 (0.90)	1.0 (0.95)
Refinement		
R_{work}[§],*	0.1774 (0.3049)	
R_{free}[§],*	0.2173 (0.3373)	
Number of total atoms		
protein	4093	
Zn ions	16	
solvent	579	
Total protein residues	550	
RMS deviations		
Bonds lengths (Å)	0.005	
Bond angles (°)	0.83	
Ramachandran favored (%)[††]	99	
Ramachandran outlier (%)[††]	0	
Average B-factor		
Macromolecules	29.51	
Zn ions	18.95	
Solvent	37.90	
Clashscore[††]	5.23	

* Data for the highest resolution shell are given in brackets. [†] $R_{sym} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the ith observation of the intensity of the reflection hkl. [‡] Correlation coefficients: $CC = \Sigma_i ((x_i - \langle x \rangle)(y_i - \langle y \rangle))/(\Sigma_i ((x_i - \langle x \rangle)^2(y_i - \langle y \rangle)^2)^{1/2}$, where x_i and y_i are the ith of n observations of quantities whose mean values are $\langle x \rangle$ and $\langle y \rangle$; for $CC_{1/2}$, x_i and y_i correspond to intensity measurements derived from each of two randomly selected half-data sets from the set of unmerged data. [§] $R_{work} = \sum_{hkl} ||F_{obs}| - |F_{calc}|| / \sum_{hkl} |F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure-factor amplitudes for each reflection hkl. R_{free} was calculated with 5% of the diffraction data that were selected randomly and excluded from refinement. [††] Calculated using MolProbity (Chen et al., 2010).

(or an HXXC motif encompassing the 3_{10} helical turn in the CRR). The second zinc-binding site is formed by a CXH motif that precedes β2 in hRING1b (or a CX_7H motif that precedes β4 in the CRR) and a CXXC motif in loop L2 (Fig. 4).

Zn(b)-binding motif ("zinc claw")

The RING domain in the CRR is preceded by a zinc-binding fold that we refer to as the Zn(b)-binding motif, or the zinc claw. This motif consists of a short α-helix (α1), a β-hairpin formed by β-strands β1 and β2, an extended loop, and another α-helix (α2) that is shared with the RING domain (Figs. 1B and 2). Zn(b) is coordinated by C446 and C459, which are located in the linkers connecting the β1-β2 hairpin to the rest of the CRR, and C481 and H484 at the N-terminal end of α2 (Figs. 1B, 2 and 3C). The β-hairpin and the extended loop act like two claws that grip the top of the RING domain (Fig. 2B).
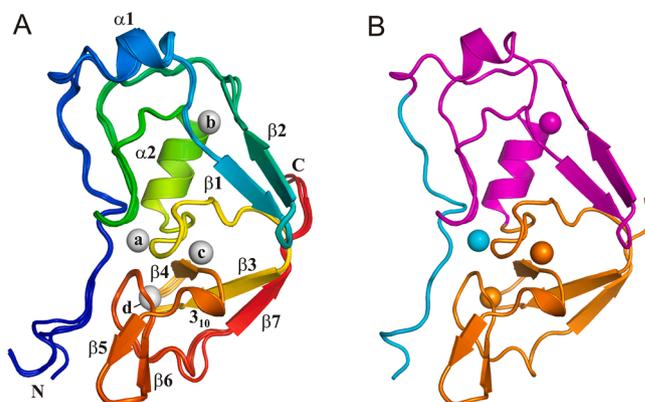
Fig. 2. The structure of the CRR. (A) Superimposed cartoon models of the four hEHMT2(419–552) molecules (chains A-D) in the asymmetric unit. The models are colored using a rainbow spectrum, i.e. the color changes smoothly from blue at the N-terminus through green and yellow to red at the C-terminus. The zinc atoms are shown as grey spheres. Both the zinc atoms and the elements of secondary structure are labeled as in Fig. 1. (B) Organization of the CRR. In a cartoon model of the CRR, the N-terminal segment is colored cyan, the Zn(b)-binding motif (including α2, which is shared with the RING domain) magenta, and the RING domain orange. The RING domain interacts extensively with the Zn(b)-binding motif and the N-terminal segment.

Zinc-binding motifs with similar folds are also part of other protein structures, for example the zinc-binding domain in subunit 2 of the general transcription factor IIH (TFIIH) (PDB ID 5OBZ chain B, F384-D504) (Radu et al., 2017) (not shown), endonucleases from the HNH family (Biertümpfel et al., 2007) (Supplementary Fig. S3A-D), and plant homeodomain (PHD)-like domains (PDB ID 5Z28, 5YUG, and 5YUH), which are found in the VP1/ABI3-LIKE (VAL) family of plant transcription factors (Suzuki et al., 2007) (Supplementary Fig. S3E-H). In the PHD-like domains, the zinc-binding motif and a PHD domain form a compact structure that resembles the structure formed by the Zn(b)-binding motif and the RING domain in the CRR (Supplementary Information and Supplementary Fig. S3E-H).

The N-terminal segment ("zinc strap")

The sequence N-terminal of α1 runs along the concave side of the CRR and is mostly in an extended conformation (Fig. 2). It interacts with the β5-β6 hairpin and the L2 loop in the RING domain, the C-terminal end of α2, and the extended loop in the Zn(b)-binding motif. Its interaction with the RING domain is stabilized by Zn(a), which is coordinated by C426 and C428 from the N-terminal segment, H491, which is positioned between α2 and loop L1 in the RING domain, and H535 in loop L2 of the RING domain (Figs. 1B, 2 and 3A).
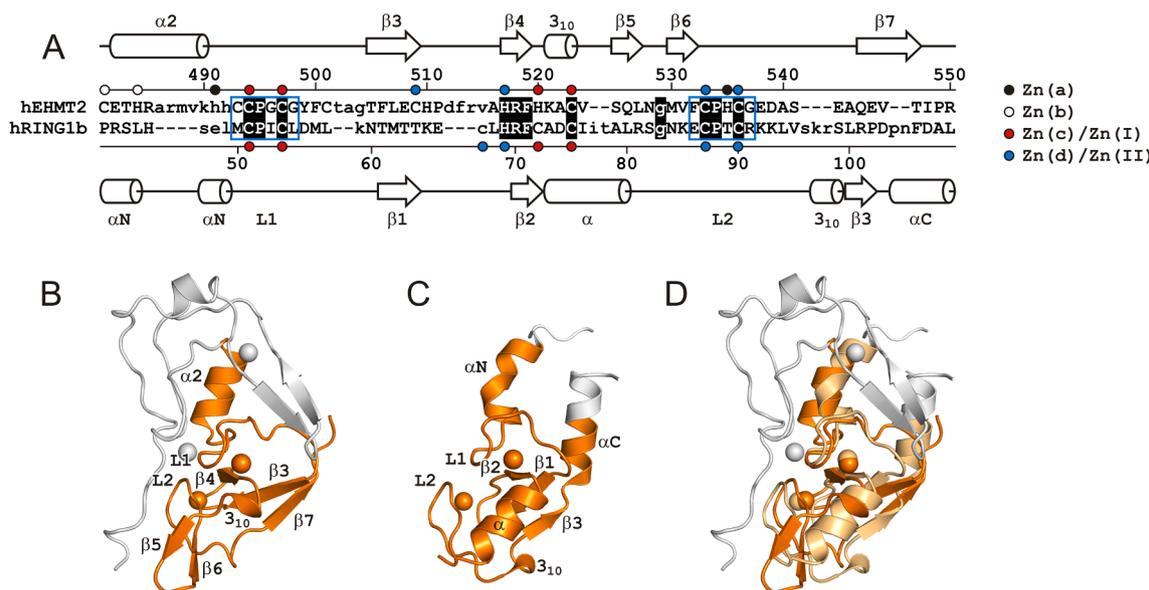
Conservation of the CRR sequence

The degree to which an amino acid position is evolutionarily conserved among related proteins strongly depends on its structural and functional importance. Amino acids in positions that are important for structure or function, e.g. for interaction with other molecules, are typically more conserved than others. We therefore evaluated the degree of evolutionary conservation in the CRR using the ConSurf server (Ashkenazy et al., 2016) and analyzed the conservation in the context of the CRR structure.

The CRR is unique to vertebrate EHMT1 and EHMT2 proteins and their single ortholog in invertebrates. The cysteine and histidine residues that coordinate the four zinc atoms are strictly conserved in EHMT proteins in all species (Fig. 5 and Supplementary Fig. S4), and the zinc-binding pattern and overall structure of the CRR are therefore also likely to be conserved. However, there are few other amino acid positions in the CRR that are invariant across all EHMT proteins.
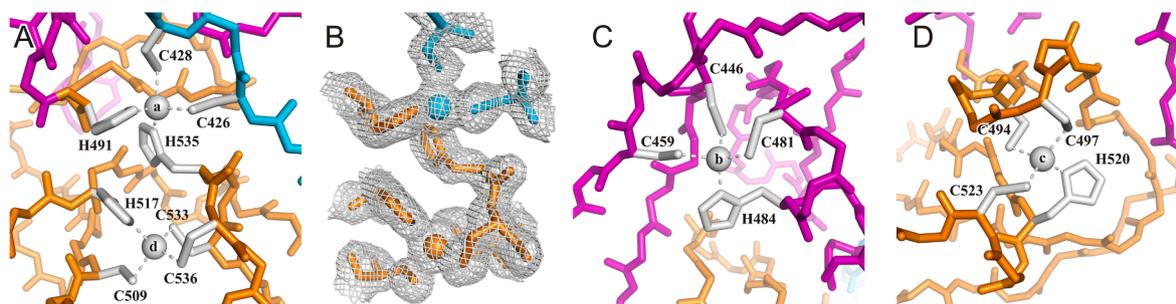
Fig. 3. Close-up view of the four zinc-binding sites in the CRR. (A) Zn(a) is coordinated by C426, C428, H491, and H535 (upper part), and Zn(d) is coordinated by C509, H517, C533, and C536 (lower part). The zinc atoms are depicted as grey spheres. The protein backbone is shown as sticks and is colored as in Fig. 2B, i.e. the N-terminal segment is cyan, the Zn(b)-binding motif magenta, and the RING domain orange. Side chains are displayed only for the zinc-coordinating residues and are colored grey. (B) A 2mFo-DFc map contoured at 1.0 σ (grey mesh) for Zn(a) and Zn(d) atoms and the residues that coordinate them. The view is the same as in panel A. The zinc atoms and the zinc-coordinating residues are colored as in Fig. 2B. (C) Zn(b) is coordinated by C446, C459, C481, and H484. (D) Zn(c) is coordinated by C494, C497, H520, and C523. The coloring scheme in panels C and D is the same as in panel A.


Fig. 4. Similarity between the C-terminal portion of the CRR and RING domains. (A) Structure-based alignment of sequences from the CRR and the RING domain from hRING1b (PDB ID 4S3O chain B) (Taherbhoy et al., 2015). Residues identified by the Dali server (Holm, 2019) as structurally equivalent are shown in upper case. Residues that are identical in both proteins are highlighted with black fill, and cysteine and histidine residues that coordinate zinc are denoted with colored circles. Elements of secondary structure are diagramed next to each protein sequence and labeled; loops L1 and L2 are also labeled and are marked with blue rectangles. (B-D) Cartoon models of the CRR (B), the RING domain from hRING1b (C), and the two models superimposed (D). The orientation of the models in the side-by-side view (B and C) is the same as in the overlay (D). The regions that have similar structures (C481-R550 in the CRR and P42-L109 in hRING1b) are colored orange, and elements of secondary structure within these regions are labeled. The rest of each cartoon model is colored grey. In the overlay (D), the RING domain from hRING1b is colored light orange to make it distinguishable from the CRR.

In vertebrate EHMT2 proteins, the residues in the core of the CRR structure, in particular those in β3, β4 and loops L1 and L2 in the RING domain, are highly conserved (Fig. 5 and Supplementary Fig. S4). These residues are likely to be important for structural integrity. In addition, residues at or near the concave surface are also conserved. This includes residues in the N-terminal segment and the portions of α2 and the extended loop in the Zn(b)-binding motif that interact with it. The convex surface is less conserved, and the residues in α1 and the linker connecting the β1-β2 hairpin with the extended loop in the Zn(b)-binding motif (at the top of the CRR structure in Fig. 5) are the most variable. Some of these trends are also observable in vertebrate EHMT1-like proteins and, to a lesser extent, in invertebrate EHMT proteins; however, especially in the latter, the CRR is highly divergent (Supplementary Fig. S4).

Discussion

The RING domain in the CRR and ubiquitin ligase activity

Most RING domains function as protein ubiquitin ligases (E3), catalyzing transfer of ubiquitin from a ubiquitin-conjugating enzyme (E2) to a lysine side chain in a target protein (Buetow and Huang, 2016; Deshaies and Joazeiro, 2009). To facilitate the transfer, the RING domain binds to the ubiquitin~E2 conjugate (in which the ubiquitin is attached to a cysteine residue in the active site of the E2 via a thioester bond) and stabilizes the conjugate in a reactive conformation (Buetow and Huang, 2016; Dou et al., 2012b; Plechanovová et al., 2012; Pruneda et al., 2012; Zheng and Shabek, 2017). The E2 enzyme interacts with loops L1 and L2 and the α-helix in the RING domain (Budhidarmo et al.,
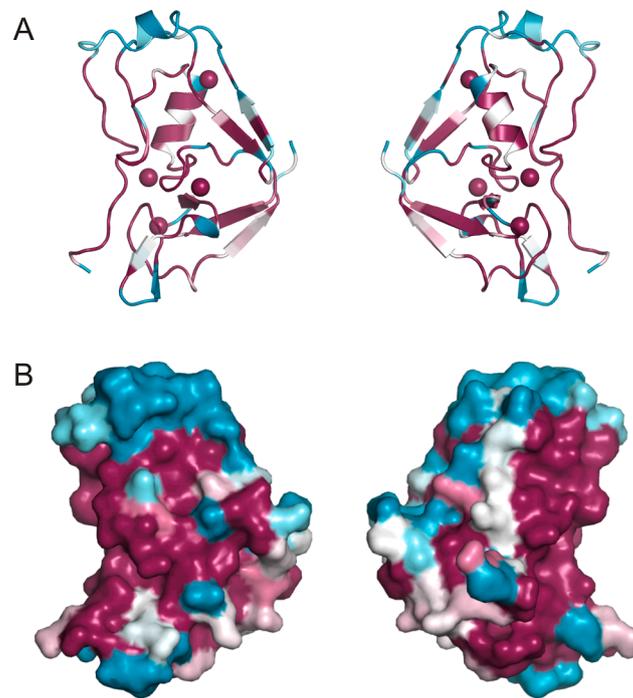
Fig. 5. Conservation of the CRR. Cartoon (A) and contact surface (B) models of the CRR are colored according to conservation (Ashkenazy et al., 2016) in vertebrate EHMT2 proteins. The color changes from magenta in the most conserved positions to cyan in the most variable positions. The two views in each panel are related by a 180° rotation around the vertical axis.

2012; Deshaies and Joazeiro, 2009), as shown in Fig. 6A for a complex containing the RING domain from hRING1b and the UbcH5c E2 enzyme. In the CRR, however, loops L1 and L2 interact with the extended loop in the Zn(b)-binding motif and with the N-terminal segment and are therefore not accessible (Figs. 2B, 4B and 6B).

Some RING domain E3 ubiquitin ligases adopt autoinhibited conformations, in which the E2 binding site is sequestered by interactions with other parts of the protein until a stimulus triggers structural rearrangement (Buetow and Huang, 2016; Dickson et al., 2018; Dou et al., 2012a; Dueber et al., 2011; Fiorentini et al., 2020). The structure of the CRR that we obtained may therefore correspond to such an auto-inhibited state. However, it should be noted that in order for loops L1 and L2 to become exposed, some of the coordination bonds involving Zn (a), which is positioned at the interface between the loops and the N-terminal segment, would need to break (Figs. 2B and 3A).

In addition to being partially sequestered, the RING domain in the CRR lacks several features that are found in many RING domain ubiquitin ligases. In particular, RING domain ubiquitin ligases typically contain a long aliphatic residue (isoleucine, leucine or valine) in the position that immediately precedes the second zinc-coordinating residue in loop L1; this residue forms hydrophobic contacts with a conserved proline in E2 (Deshaies and Joazeiro, 2009). However, the CRR contains a glycine (G496) in this position. RING domain ubiquitin ligases also frequently contain an arginine in the position that follows the last zinc-coordinating residue; this arginine serves as a "linchpin", interacting with both the E2 and ubiquitin moieties of the E2~ubiquitin conjugate (Pruneda et al., 2012). The CRR contains a glycine (G537) in this position (Fig. 4A). Overall, our observations suggest that the RING domain in the CRR is unlikely to function as a ubiquitin ligase.

RING domain heterodimerization

Many RING domains form homo- or heterodimers. The RING domain from hRING1b, for example, heterodimerizes with the RING domains from Polycomb group RING finger proteins (PCGF) 1–6 (Buchwald et al., 2006; Li et al., 2006; Taherbhoy et al., 2015). The PCGF RING domains

do not directly interact with an E2 enzyme but are required for full E3 ligase activity of hRING1b (Buchwald et al., 2006). The organization of the complex between the RING1b:PCGF5 heterodimer and the UbcH5c E2 enzyme (Taherbhoy et al., 2015) is illustrated in Fig. 6A. Dimerization between the RING domains is mediated by the core β-sheet and the structural elements flanking the core (α-helices labeled αN and αC in hRING1b) (Budhidarmo et al., 2012), which are on the opposite side of the RING domain than loops L1 and L2 (Fig. 6A). In the CRR, the core β-sheet and the sequences flanking the RING domain core (α2 and the C-terminal end) are exposed and hence available for potential interactions (Fig. 6B). EHMT1 and EHMT2 were previously detected in complexes that contained several other RING domain proteins including RING1a (also known as RING1), RING1b, PCGF6, and the tripartite motif-containing protein 28 (TRIM28; also known as KRAB-associated protein 1, KAP1, and transcription intermediary factor 1β, TIF1β) (Ogawa et al., 2002; Rowbotham et al., 2011); however, it is currently not known whether the EHMT proteins directly interact with these or any other RING domain proteins.

The role of the CRR in chromatin regulation

Experimental evidence available so far suggests that the CRR mediates interactions with several other proteins, namely cyclin D1, NRSF/REST, the Mediator complex, and ZNF200, and that these interactions result in recruitment of EHMT proteins to specific chromatin loci (Li et al., 2019; Roopra et al., 2004; Ding et al., 2008; Nishida et al., 2007). The CRR therefore most likely functions as a novel protein-binding domain. Little information is currently available about how the CRR interacts with its binding partners, but the concave surface on the CRR is, because of its evolutionary conservation, the most likely candidate for a binding site (Fig. 5).

RING domains whose function is unrelated to ubiquitination are uncommon (Deshaies and Joazeiro, 2009), and we are not aware of any RING domains that serve as a building block in a larger protein-binding fold. We believe that the CRR is unique in this respect.

The structures of the ankyrin repeats and the SET domain had been
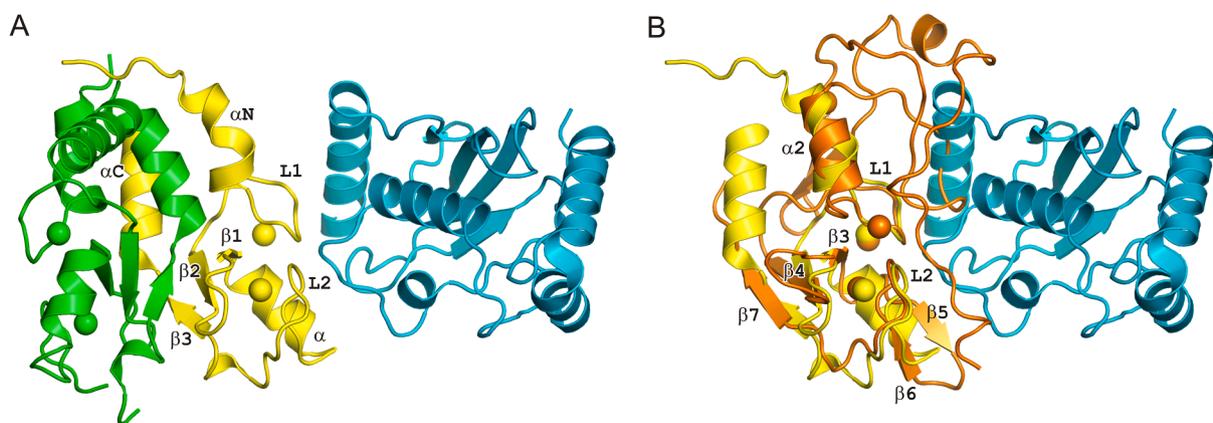
Fig. 6. RING domain interactions. (A) A cartoon model of the RING domain from hRING1b (yellow) bound to the RING domain from human PCGF5 (green) and the UbcH5c ubiquitin-conjugating enzyme (cyan) (PDB ID 4S3O) (Taherbhoy et al., 2015). Residues N-terminal of L36 in hRING1b, which wrap around PCGF5 in an extended conformation, are omitted for clarity. (B) Comparison with the CRR. The RING domain from hRING1b (yellow) bound to UbcH5c (cyan) is displayed and oriented as in panel A, and the CRR (orange) is shown superimposed on hRING1b for comparison. Elements of secondary structure in hRING1b (panel A) and the CRR (panel B) are labeled. The L1 and L2 loops, which form the E2-binding site in RING E3 ligases, are partially occluded in the CRR (right side of the CRR in panel B). The surface that is used by RING domains for dimerization (left side of the CRR in panel B; compare with the dimerization interface in panel A) is accessible.

determined previously (Collins et al., 2008; Wu et al., 2010), and the CRR thus remained the only part of the EHMT proteins that was expected to be ordered but whose structure was not available (Fig. 7). The structure that we present here now fills this gap. We expect that our structure will provide a molecular framework for future studies of the function and interactions of this portion of the EHMT proteins.

Materials and methods

Cloning, expression and purification of hEHMT2(419–552) and hEHMT2 (380–552)

hEHMT2(419–552) was expressed as a fusion protein with an N-terminal glutathione transferase (GST) tag followed by a tobacco etch virus (TEV) protease cleavage site. To generate the vector for expression of the fusion protein, the DNA sequence coding for residues 419–552 of hEHMT2 was amplified by polymerase chain reaction (PCR) from the Mammalian Gene Collection (MGC) hEHMT2 cDNA (Genbank BC018718; Open Biosystems/Dharmacon) and cloned into the pDEST 15 vector using Gateway Recombination Technology (Life Technologies). The TEV site was introduced by primers during the PCR. To improve the yield of soluble protein, expression was carried out in *Escherichia coli* BL21 ArcticExpress (DE3) RIL cells (Agilent Technologies) at 12 °C. The fusion protein was isolated from cell lysates by affinity chromatography on a glutathione-agarose resin. The hEHMT2 (419–552) moiety was then released by on-resin cleavage with a TEV protease and was further purified by anion exchange chromatography on a HiTrap Q HP column (GE Healthcare Life Sciences) and by size-exclusion chromatography (SEC) on Superdex 75 resin (GE Healthcare Life Sciences). To prepare samples for crystallization, the SEC step was performed in HEPES buffer (25 mM HEPES, pH 7.5, 100 mM NaCl, 5% w/v glycerol, 0.025% w/v NaN$_3$, 1 mM TCEP, and 10 μM ZnSO$_4$); when preparing samples for NMR spectroscopy experiments, SEC was performed in phosphate buffer (20 mM potassium phosphate, pH 7.0–7.2, 50 mM KCl, and 2 mM β-mercaptoethanol). hEHMT2 (380–552) was cloned, expressed and purified in a similar manner. After cleavage with TEV protease, the sequences of hEHMT2(380–552) and hEHMT2(419–552) were GSgGPSE…PRGD and gsGFEE…PRGD, respectively, with residues that differ from hEHMT2 sequence (UniProt Q96KQ7) shown in lower case. Protein concentration was calculated from absorbance at 280 nm (Pace et al., 1995).

Zinc content analysis

Total zinc concentration in a sample of hEHMT2(380–552) and in a buffer control was determined using PerkinElmer ELAN DRC-e inductively coupled plasma mass spectrometry (ICP-MS) instrument according to EPA method 200.8. The analysis was performed by the Environmental Biogeochemistry Laboratory at the University of Montana.

NMR spectroscopy

Samples for NMR spectroscopy typically contained ~ 0.3 mM ^{15}N- or ^{13}C- and ^{15}N-labeled proteins in phosphate buffer supplemented with 10% ^{2}H$_2$O. The experiments acquired for hEHMT2(380–552) included 2D ^{1}H-^{15}N HSQC, 2D ^{1}H-^{15}N NOE, and 3D HNCACB, C(CO)NH, HNCO and HN(CA)CO. In addition, a 2D ^{1}H-^{15}N HSQC spectrum was also acquired for hEHMT2(419–552). All data were collected at 25 °C on a Varian 600 MHz NMR System equipped with a triple-resonance probe and were processed and analyzed with Felix 2004 (Felix NMR Inc.). Chemical shift assignments for the flexible region in hEHMT2(380–522) were established using standard procedures.

SEC-MALS

Molar mass of hEHMT2(419–552) in solution was determined by SEC-MALS using a miniDAWN TREOS II MALS detector and an Optilab T-rEX refractometer (Wyatt Technology) connected to a Superdex 75 10/300 column (GE Healthcare Life Sciences). The experiment was performed at room temperature in HEPES buffer without glycerol. The protein was loaded on the column in a volume of 500 μL and at a concentration of 5 mg/mL. The data were analyzed with ASTRA 7.1 software (Wyatt Technology).

Crystallization of hEHMT2(419–552) and structure determination

Crystals were grown in sitting drops that were prepared by mixing 1 μL of 18 mg/mL hEHMT2(419–552) in HEPES buffer with 1 μL of reservoir solution (0.1 M HEPES, pH 8.0, 0.2 M potassium sodium tartrate, and 18% PEG 3350) and equilibrated in a sealed well over the reservoir solution at 20 °C. The crystals were harvested in a cryoprotectant consisting of 30% w/v glycerol and 10 μM ZnSO$_4$ in the reservoir solution and flash-cooled in liquid nitrogen. A 1.9 Å resolution data set
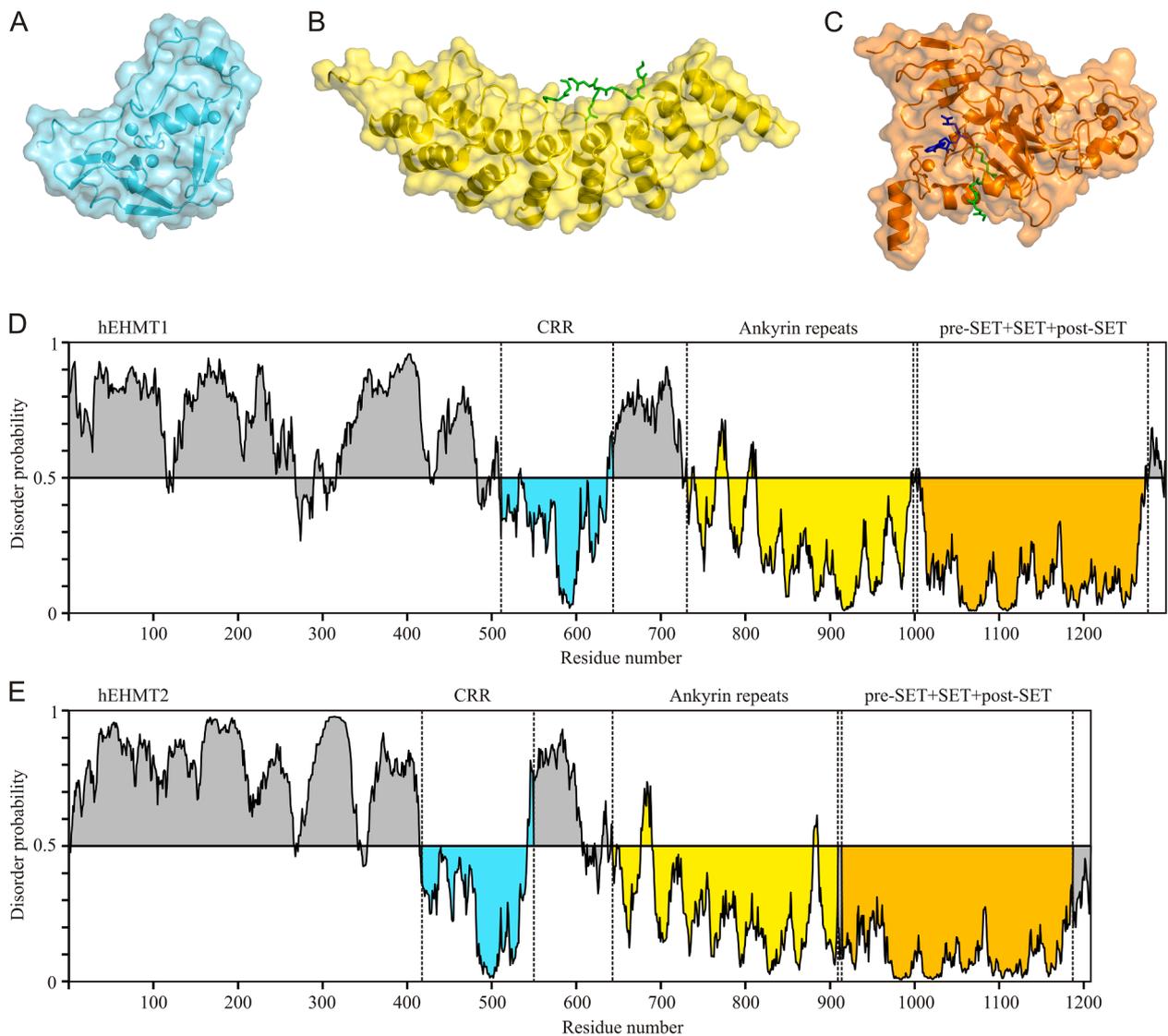
Fig. 7. Structured domains in EHMT proteins. Structures of (A) the CRR (cyan), (B) the ankyrin repeats (yellow; PDB ID 3B95 and 6BY9) (Collins et al., 2008; Dong et al., 2018), and (C) the SET domain with the pre-SET and post-SET regions (orange; PDB ID 3HNA and 5JIN) (Jayaram et al., 2016; Wu et al., 2010). The ankyrin repeats bind to sequences that contain mono- or dimethylated lysine (green sticks; side chain shown only for the dimethyllysine residue). The SET domain catalyzes transfer of a methyl group from S-adenosyl-L-methionine (SAM; blue) to the lysine ε-amino group in a substrate peptide (green). (D and E) Predicted disorder (Erdős and Dosztányi, 2020) in hEHMT1 (D) and hEHMT2 (E). The CRR, the ankyrin repeats, and the SET domain with the pre-SET and post-SET regions are colored cyan, yellow and orange respectively as in panels A-C. The rest of each protein is colored grey and is predicted to be mostly disordered (disorder score > 0.5).

(Set #1) was collected at the Advanced Photon Source (APS) SBC-CAT 19-BM beamline (Argonne, IL), and a 2.4 Å resolution data set (Set #2) with a wavelength of 1.2827 Å, which corresponds to the peak of the Zn absorption edge, was collected at the Stanford Synchrotron Radiation Lightsource SSRL-SMB 12–2 beamline (Melon Park, CA) (Table 1). The diffraction images were indexed, integrated, and scaled with HKL-2000 (Otwinowski and Minor, 1997).

The structure of hEHMT2(419–552) was determined using SAD arising from the protein-bound Zn atoms. The positions of the Zn atoms were identified with the HYSS program; initial phases were calculated with SOLVE/RESOLVE (Terwilliger, 2003) and further improved by density modification in the AutoSol program in the PHENIX software suite (Adams et al., 2010). The resulting electron density map was readily interpretable, and approximately 75% of residues in each protein molecule were built and refined automatically. The higher resolution data set (Set #1) was then used for iterative model rebuilding with COOT (Emsley et al., 2010) followed by refinement of atomic positions and isotropic B-factor parameters in PHENIX. Water molecules were

added progressively, and in later stages of refinement, TLS parameters were employed. The data collection and refinement statistics are listed in Table 1. Figures depicting the protein structures were prepared with the PyMOL Molecular Graphics System version 1.7.6.5 (Schrödinger, LLC). The final atomic coordinates and structure amplitudes have been deposited in the Protein Data Bank (PDB ID: 6MM1).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.yjsbx.2021.100050.

References

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C., Zwart, P.H., 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. 66 (2), 213–221. https://doi.org/10.1107/S0907444909052925.

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., Ben-Tal, N., 2016. ConSurf2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res. 44 (W1), W344–W350. https://doi.org/10.1093/nar/gkw408.

Barlow, P.N., Luisi, B., Milner, A., Elliott, M., Everett, R., 1994. Structure of the C_3HC_4 domain by ^1H-nuclear magnetic resonance spectroscopy. A new structural class of zinc-finger. J. Mol. Biol. 237 (2), 201–211. https://doi.org/10.1006/jmbi.1994.1222.

Benevento, M., van de Molengraft, M., van Westen, R., van Bokhoven, H., Nadif Kasri, N., 2015. The role of chromatin repressive marks in cognition and disease: A focus on the repressive complex GLP/G9a. Neurobiol. Learn. Mem. 124, 88–96. https://doi.org/10.1016/j.nlm.2015.06.013.

Biertümpfel, C., Yang, W., Suck, D., 2007. Crystal structure of T4 endonuclease VII resolving a Holliday junction. Nature 449 (7162), 616–620. https://doi.org/10.1038/nature06152.

Brown, S.E., Campbell, R.D., Sanderson, C.M., 2001. Novel NG36/G9a gene products encoded within the human and mouse MHC class III regions. Mamm. Genome 12 (12), 916–924. https://doi.org/10.1007/s00335-001-3029-3.

Buchwald, G., van der Stoop, P., Weichenrieder, O., Perrakis, A., van Lohuizen, M., Sixma, T.K., 2006. Structure and E3-ligase activity of the Ring-Ring complex of Polycomb proteins Bmi1 and Ring1b. EMBO J. 25 (11), 2465–2474. https://doi.org/10.1038/sj.emboj.7601144.

Budhidarmo, R., Nakatani, Y., Day, C.L., 2012. RINGs hold the key to ubiquitin transfer. Trends Biochem. Sci. 37 (2), 58–65. https://doi.org/10.1016/j.tibs.2011.11.001.

Buetow, L., Huang, D.T., 2016. Structural insights into the catalysis and regulation of E3 ubiquitin ligases. Nat. Rev. Mol. Cell Biol. 17 (10), 626–642. https://doi.org/10.1038/nrm.2016.91.

Casciello, F., Windloch, K., Gannon, F., Lee, J.S., 2015. Functional role of G9a histone methyltransferase in cancer. Front. Immunol. 6, 487. https://doi.org/10.3389/fimmu.2015.00487.

Chang, Y., Zhang, X., Horton, J.R., Upadhyay, A.K., Spannhoff, A., Liu, J., Snyder, J.P., Bedford, M.T., Cheng, X., 2009. Structural basis for G9a-like protein lysine methyltransferase inhibition by BIX-01294. Nat. Struct. Mol. Biol. 16 (3), 312–317. https://doi.org/10.1038/nsmb.1560.

Chen, V.B., Arendall III, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., Richardson, D.C., 2010. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. D Biol. Crystallogr. 66 (1), 12–21. https://doi.org/10.1107/S0907444909042073.

Collins, R.E., Northrop, J.P., Horton, J.R., Lee, D.Y., Zhang, X., Stallcup, M.R., Cheng, X., 2008. The ankyrin repeats of G9a and GLP histone methyltransferases are mono- and dimethyllysine binding modules. Nat. Struct. Mol. Biol. 15 (3), 245–250. https://doi.org/10.1038/nsmb.1384.

Deimling, S.J., Olsen, J.B., Tropepe, V., 2017. The expanding role of the Ehmt2/G9a complex in neurodevelopment. Neurogenesis 4 (1), e1316888. https://doi.org/10.1080/23262133.2017.1316888.

Deshaies, R.J., Joazeiro, C.A.P., 2009. RING domain E3 ubiquitin ligases. Annu. Rev. Biochem. 78, 399–434. https://doi.org/10.1146/annurev.biochem.78.101807.093809.

Dickson, C., Fletcher, A.J., Vaysburd, M., Yang, J.-C., Mallery, D.L., Zeng, J., Johnson, C.M., McLaughlin, S.H., Skehel, M., Maslen, S., Cruickshank, J., Huguenin-Dezot, N., Chin, J.W., Neuhaus, D., James, L.C., 2018. Intracellular antibody signalling is regulated by phosphorylation of the Fc receptor TRIM21. eLife 7, e32660. https://doi.org/10.7554/eLife.32660.

Ding, N., Zhou, H., Esteve, P.-O., Chin, H.G., Kim, S., Xu, X., Joseph, S.M., Friez, M.J., Schwartz, C.E., Pradhan, S., Boyer, T.G., 2008. Mediator links epigenetic silencing of neuronal gene expression with X-linked mental retardation. Mol. Cell 31 (3), 347–359. https://doi.org/10.1016/j.molcel.2008.05.023.

Dong, A., Wei, Y., Li, A., Tempel, W., Han, S., Sunnerhagen, M., Penn, L., Bountra, C., Arrowsmith, C.H., Edwards, A.M., Tong, Y., 2018. Crystal structure of EHMT1. PDB ID: 6BY9. https://doi.org/10.2210/pdb6BY9/pdb.

Dou, H., Buetow, L., Hock, A., Sibbet, G.J., Vousden, K.H., Huang, D.T., 2012a. Structural basis for autoinhibition and phosphorylation-dependent activation of c-Cbl. Nat. Struct. Mol. Biol. 19 (2), 184–192. https://doi.org/10.1038/nsmb.2231.

Dou, H., Buetow, L., Sibbet, G.J., Cameron, K., Huang, D.T., 2012b. BIRC7-E2 ubiquitin conjugate structure reveals the mechanism of ubiquitin transfer by a RING dimer. Nat. Struct. Mol. Biol. 19 (9), 876–883. https://doi.org/10.1038/nsmb.2379.

Dueber, E.C., Schoeffler, A.J., Lingel, A., Elliott, J.M., Fedorova, A.V., Giannetti, A.M., Zobel, K., Maurer, B., Varfolomeev, E., Wu, P., Wallweber, H.J.A., Hymowitz, S.G., Deshayes, K., Vucic, D., Fairbrother, W.J., 2011. Antagonists induce a conformational change in cIAP1 that promotes autoubiquitination. Science 334 (6054), 376–380. https://doi.org/10.1126/science.1207862.

Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K., 2010. Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. 66 (4), 486–501. https://doi.org/10.1107/S0907444910007493.

Erdős, G., Dosztányi, Z., 2020. Analyzing protein disorder with IUPred2A. Curr. Protoc. Bioinformatics 70 (1), e99. https://doi.org/10.1002/cpbi.99.

Fiorentini, F., Esposito, D., Rittinger, K., 2020. Does it take two to tango? RING domain self-association and activity in TRIM E3 ubiquitin ligases. Biochem. Soc. Trans. 48 (6), 2615–2624. https://doi.org/10.1042/BST20200383.

Holm, L., 2019. Benchmarking fold detection by DaliLite v.5. Bioinformatics 35 (24), 5326–5327. https://doi.org/10.1093/bioinformatics/btz536.

Jayaram, H., Hoelper, D., Jain, S.U., Cantone, N., Lundgren, S.M., Poy, F., Allis, C.D., Cummings, R., Bellon, S., Lewis, P.W., 2016. S-adenosyl methionine is necessary for inhibition of the methyltransferase G9a by the lysine 9 to methionine mutation on histone H3. Proc. Natl. Acad. Sci. U.S.A. 113 (22), 6182–6187. https://doi.org/10.1073/pnas.1605523113.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., Sternberg, M.J.E., 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protoc. 10 (6), 845–858. https://doi.org/10.1038/nprot.2015.053.

Kramer, J.M., 2016. Regulation of cell differentiation and function by the euchromatin histone methyltransferases G9a and GLP. Biochem. Cell Biol. 94 (1), 26–32. https://doi.org/10.1139/bcb-2015-0017.

Li, Z., Cao, R., Wang, M., Myers, M.P., Zhang, Y., Xu, R.-M., 2006. Structure of a Bmi-1-Ring1B polycomb group ubiquitin ligase complex. J. Biol. Chem. 281 (29), 20643–20649. https://doi.org/10.1074/jbc.M602461200.

Li, Z., Jiao, X., Di Sante, G., Ertel, A., Casimiro, M.C., Wang, M., Katiyar, S., Ju, X., Klopfenstein, D.V., Tozeren, A., Dampier, W., Chepelev, I., Jeltsch, A., Pestell, R.G., 2019. Cyclin D1 integrates G9a-mediated histone methylation. Oncogene 38 (22), 4232–4249. https://doi.org/10.1038/s41388-019-0723-8.

Milner, C.M., Campbell, R.D., 1993. The G9a gene in the human major histocompatibility complex encodes a novel protein containing ankyrin-like repeats. Biochem. J. 290 (3), 811–818. https://doi.org/10.1042/bj2900811.

Nishida, M., Kato, M., Kato, Y., Sasai, N., Ueda, J., Tachibana, M., Shinkai, Y., Yamaguchi, M., 2007. Identification of ZNF200 as a novel binding partner of histone H3 methyltransferase G9a. Genes Cells 12 (7), 877–888. https://doi.org/10.1111/j.1365-2443.2007.01098.x.

Ogawa, H., Ishiguro, K., Gaubatz, S., Livingston, D.M., Nakatani, Y., 2002. A complex with chromatin modifiers that occupies E2F- and Myc-responsive genes in G_0 cells. Science 296 (5570), 1132–1136. https://doi.org/10.1126/science.1069861.

Otwinowski, Z., Minor, W., 1997. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. 276, 307–326. https://doi.org/10.1016/S0076-6879(97)76066-X.

Pace, C.N., Vajdos, F., Fee, L., Grimsley, G., Gray, T., 1995. How to measure and predict the molar absorption coefficient of a protein. Protein Sci. 4 (11), 2411–2423. https://doi.org/10.1002/pro.5560041120.

Plechanovová, A., Jaffray, E.G., Tatham, M.H., Naismith, J.H., Hay, R.T., 2012. Structure of a RING E3 ligase and ubiquitin-loaded E2 primed for catalysis. Nature 489 (7414), 115–120. https://doi.org/10.1038/nature11376.

Pruneda, J.N., Littlefield, P.J., Soss, S.E., Nordquist, K.A., Chazin, W.J., Brzovic, P.S., Klevit, R.E., 2012. Structure of an E3:E2~Ub complex reveals an allosteric mechanism shared among RING/U-box ligases. Mol. Cell 47 (6), 933–942. https://doi.org/10.1016/j.molcel.2012.07.001.

Radu, L., Schoenwetter, E., Braun, C., Marcoux, J., Koelmel, W., Schmitt, D.R., Kuper, J., Cianférani, S., Egly, J.M., Poterszman, A., Kisker, C., 2017. The intricate network between the p34 and p44 subunits is central to the activity of the transcription/DNA repair factor TFIIH. Nucleic Acids Res. 45 (18), 10872–10883. https://doi.org/10.1093/nar/gkx743.

Roopra, A., Qazi, R., Schoenike, B., Daley, T.J., Morrison, J.F., 2004. Localized domains of G9a-mediated histone methylation are required for silencing of neuronal genes. Mol. Cell 14 (6), 727–738. https://doi.org/10.1016/j.molcel.2004.05.026.

Rowbotham, S.P., Barki, L., Neves-Costa, A., Santos, F., Dean, W., Hawkes, N., Choudhary, P., Will, W.R., Webster, J., Oxley, D., Green, C.M., Varga-Weisz, P., Mermoud, J.E., 2011. Maintenance of silent chromatin through replication requires SWI/SNF-like chromatin remodeler SMARCAD1. Mol. Cell 42 (3), 285–296. https://doi.org/10.1016/j.molcel.2011.02.036.

Scheer, S., Zaph, C., 2017. The lysine methyltransferase G9a in immune cell differentiation and function. Front. Immunol. 8, 429. https://doi.org/10.3389/fimmu.2017.00429.

Shankar, S.R., Bahirvani, A.G., Rao, V.K., Bharathy, N., Ow, J.R., Taneja, R., 2013. G9a, a multipotent regulator of gene expression. Epigenetics 8 (1), 16–22. https://doi.org/10.4161/epi.23331.

Shinkai, Y., Tachibana, M., 2011. H3K9 methyltransferase G9a and the related molecule GLP. Genes Dev. 25 (8), 781–788. https://doi.org/10.1101/gad.2027411.

Suzuki, M., Wang, H.H.-Y., McCarty, D.R., 2007. Repression of the LEAFY COTYLEDON 1/B3 regulatory network in plant embryo development by VP1/ABSCISIC ACID INSENSITIVE 3-LIKE B3 genes. Plant Physiol 143 (2), 902–911. https://doi.org/10.1104/pp.106.092320.

Tachibana, M., Sugimoto, K., Fukushima, T., Shinkai, Y., 2001. SET domain-containing protein, G9a, is a novel lysine-preferring mammalian histone methyltransferase with hyperactivity and specific selectivity to lysines 9 and 27 of histone H3. J. Biol. Chem. 276 (27), 25309–25317. https://doi.org/10.1074/jbc.M101914200.

Tachibana, M., Sugimoto, K., Nozaki, M., Ueda, J., Ohta, T., Ohki, M., Fukuda, M., Takeda, N., Niida, H., Kato, H., Shinkai, Y., 2002. G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. Genes Dev. 16 (14), 1779–1791. https://doi.org/10.1101/gad.989402.

Tachibana, M., Ueda, J., Fukuda, M., Takeda, N., Ohta, T., Iwanari, H., Sakihama, T., Kodama, T., Hamakubo, T., Shinkai, Y., 2005. Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3–K9. Genes Dev. 19 (7), 815–826. https://doi.org/10.1101/gad.1284005.

Taherbhoy, A.M., Huang, O.W., Cochran, A.G., 2015. BMI1-RING1B is an autoinhibited RING E3 ubiquitin ligase. Nat. Commun. 6, 7621. https://doi.org/10.1038/ncomms8621.

Terwilliger, T.C., 2003. SOLVE and RESOLVE: automated structure solution and density modification. Methods Enzymol. 374, 22–37. https://doi.org/10.1016/S0076-6879(03)74002-6.

Wu, H., Min, J., Lunin, V.V., Antonshenko, T., Dombrovski, L., Zeng, H., Allali-Hassani, A., Campagna-Slater, V., Vedadi, M., Arrowsmith, C.H., Plotnikov, A.N., Schapira, M., 2010. Structural biology of human H3K9 methyltransferases. PLoS ONE 5 (1), e8570. https://doi.org/10.1371/journal.pone.0008570.

Zheng, N., Shabek, N., 2017. Ubiquitin ligases: structure, function, and regulation. Annu. Rev. Biochem. 86, 129–157. https://doi.org/10.1146/annurev-biochem-060815-014922.