Title
Mechanism of DNA Methylation-Directed Histone Methylation by KRYPTONITE

Journal
Molecular Cell, 55(3)

ISSN
1097-2765

Authors
Du, Jiamu
Johnson, Lianna M
Groth, Martin
et al.

Publication Date
2014-08-01

DOI
10.1016/j.molcel.2014.06.009

Peer reviewed

Published in final edited form as:
Mol Cell. 2014 August 7; 55(3): 495–504. doi:10.1016/j.molcel.2014.06.009.

Mechanism of DNA methylation-directed histone methylation by KRYPTONITE

Jiamu Du[#1], **Lianna M. Johnson**[#2], **Martin Groth**[2,4], **Suhua Feng**[2,3,4], **Christopher J. Hale**[2], **Sisi Li**[1], **Ajay A. Vashisht**[5], **James A. Wohlschlegel**[5], **Dinshaw J. Patel**[1,#], and **Steven E. Jacobsen**[2,3,4,#]

[1]Structural Biology Program, Memorial Sloan-Kettering Cancer Center, New York, NY 10065, USA

[2]Department of Molecular, Cell and Developmental Biology, University of California at Los Angeles, Los Angeles, CA 90095, USA

[3]Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, University of California at Los Angeles, Los Angeles, CA 90095, USA

[4]Howard Hughes Medical Institute, University of California at Los Angeles, Los Angeles, CA 90095, USA

[5]Department of Biological Chemistry, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, CA 90095, USA

[#] These authors contributed equally to this work.

SUMMARY

In *Arabidopsis*, CHG DNA methylation is controlled by the H3K9 methylation mark through a self-reinforcing loop between DNA methyltransferase CHROMOMETHYLASE3 (CMT3) and H3K9 histone methyltransferase KRYPTONITE/SUVH4 (KYP). We report on the structure of KYP in complex with methylated DNA, substrate H3 peptide and cofactor SAH, thereby defining the spatial positioning of the SRA domain relative to the SET domain. The methylated DNA is bound by the SRA domain with the 5mC flipped out of the DNA, while the H3(1-15) peptide substrate binds between the SET and post-SET domains, with the ε-ammonium of K9 positioned adjacent to bound SAH. These structural insights complemented by *in vivo* functional data on key mutants of residues lining the 5mC and H3K9-binding pockets within KYP, establish how

[#]Correspondence: jacobsen@ucla.edu (S.E.J.) and pateld@mskcc.org (D.J.P.) .

Author Contributions

J.D., L.M.J., D.J.P., and S.E.J. designed the project; J.D., L.M.J., M.G., S.F., C.J.H., S.L., A.A.V., and J.A.W performed experiments; J.D., L.M.J., M.G., S.F., C.J.H., D.J.P., and S.E.J. analyzed the data. J.D., L.M.J., M.G., D.J.P., and S.E.J. wrote the manuscript.

Accession Numbers

The structures have been deposited in the Protein Data Bank under accession ID: 4QEN (KYP-mCHH DNA-SAH complex), 4QEO (KYP-mCHH DNA-SAH-H3 peptide complex), and 4QEP (KYP-mCHG DNA-SAH complex). All genomics data can be accessed at GEO accession GSE57963.

methylated DNA recruits KYP to the histone substrate. Together, the structures of KYP and previously reported CMT3 complexes provide insights into molecular mechanisms linking DNA and histone methylation.

INTRODUCTION

DNA methylation is one of the most important epigenetic marks with functional impact on genomic imprinting, gene silencing, and suppression of repetitive elements (Goll and Bestor, 2005; Law and Jacobsen, 2010). In plants, DNA methylation occurs in three different sequence contexts: CG, CHG (H = C, T or A), and CHH, all of which are highly correlated with the histone lysine methylation modification H3K9me (Cedar and Bergman, 2009; Law and Jacobsen, 2010). CHG DNA methylation is controlled by a plant specific DNA methyltransferase CMT3 and by the histone H3K9 methyltransferase KRYPTONITE (KYP, also known as SUVH4). CMT3 is targeted to H3K9me-containing nucleosomes by a dual recognition mechanism mediated by its BAH and chromo domains (Bartee et al., 2001; Du et al., 2012; Lindroth et al., 2001), and KYP has been shown to be capable of binding to methylated CHH (mCHH) or mCHG DNA through its SRA domain (Jackson et al., 2004; Jackson et al., 2002; Johnson et al., 2007; Malagnac et al., 2002). Therefore, CMT3 can be recruited by H3K9me and further methylate CHG DNA to create binding sites for KYP, as well as its close homologs SUVH5 and SUVH6; in turn, the methylated DNA-recruited KYP can methylate H3K9 to generate the binding sites for CMT3, resulting in a self-reinforcing feedback loop (Law and Jacobsen, 2010). To further investigate the molecular mechanism of the self-reinforcing feedback loop between DNA and histone methylation in plants, we carried out structural and functional studies, which revealed a distinct mechanism by which KYP specifically recognizes mCHH and mCHG DNA, as well as how KYP recognizes its target histone substrates.

RESULTS

Overall Structure of KYP in Complex with mCHH DNA, H3 Peptide and Cofactor SAH

We generated an N-terminal truncated KYP construct (93-624), which includes all its functional domains: the SRA domain, the pre-SET/SET/post-SET domains, and two predicted N-terminal α-helical segments (Figure 1A). The crystal structure of KYP in complex with a 13-bp DNA possessing a central mCHH site and two nucleotide overhangs at both 3′-ends, the cofactor product *S*-adenosylhomocysteine (SAH), and an unmethylated H3(1-15) substrate peptide was solved at 2.0 Å resolution (Figures 1B, and Table 1). Overall, the structure exhibits good electron density except for some loop regions, including the N-terminal disordered region (residues 93-98), the loop linking SRA and pre-SET domains (residues 313-327), and two internal loops within the SET domain (residues 486-490 and residues 500-533), which were not built into the final model. We also solved a 3.1 Å resolution crystal structure of KYP-mCHG DNA-SAH complex (lacking bound H3 peptide), which has an almost identical overall structure and DNA recognition mechanism (Table 1, Figure S1). Despite many attempts, we were unable to grow crystals of mCG DNA-containing KYP complexes, most likely because of the significantly weaker binding of KYP to mCG compared to mCHH and mCHG DNA (Johnson et al., 2007). Because of

the much higher resolution of the mCHH DNA containing structure of the complex compared with its mCHG DNA containing counterpart, the generally similar recognition mechanism (Figure S1), and KYP forming a self-reinforcing loop with CMT2 and CHH methylation in a manner which parallels the KYP-CMT3 mediated CHG methylation system (Stroud et al., 2013a), we mainly focus on mCHH DNA containing complex in the following presentation. The two α-helices of the N-terminal segment align together in an anti-parallel orientation in the middle of the structure, with the SRA domain positioned on one side and the pre-SET/SET/post-SET domains on the other side (Figures 1B). Several hydrophobic residues of the first α-helix (longer of the two helices) form a continuous hydrophobic surface, which interacts with the hydrophobic surface on the second α-helix, supporting the relative orientation and alignment of the two-helix arrangement (Figure 1C). In addition, the first α-helix is wrapped by the second α-helix, the SRA domain, the pre-SET domain and the SET domain, with these multiple interactions most likely contributing to the stabilization of the overall architecture of the whole protein (Figures 1D-F). The second shorter α-helix mainly interacts with the SRA domain mediated by hydrophobic contacts (Figure 1G). Although the linker between the SRA domain and pre-SET domain (residues 313-327) is disordered, the SRA domain forms both hydrophobic and hydrogen bonding interactions with the pre-SET/SET domains (Figure 1H). Together, the three segments of the protein spanning the two-helix arrangement, the SRA domain and the pre-SET/SET/post-SET domains interact with each other, revealing a rigid overall alignment resulting in a stable scaffold and precise spatial positioning of the SRA domain with respect to the catalytic SET domain.

Recognition of mCHH DNA

The SRA domain of KYP targets methylated DNA using general principles reported previously for recognition of methylated DNA by the SRA domains of UHRF1 (Arita et al., 2008; Avvakumov et al., 2008; Hashimoto et al., 2008) and SUVH5 (Rajakumara et al., 2011) proteins. Nevertheless, unlike the reported structure of the KYP homolog SUVH5 SRA-DNA complex (2:1 ratio of SUVH5 SRA:DNA) (Rajakumara et al., 2011), KYP binds methylated DNA with a 1:1 molar ratio and only the 5-methylcytosine (5mC) base is flipped out of the DNA duplex. The mCHH DNA duplex can be readily traced and is mainly positioned within a basic surface cleft of the SRA domain, with the second α-helix of the two-helix contributing minor intermolecular interactions (Figures 2A, B and S2A-C). The flip-out of 5mC introduces a gap within the DNA duplex, while the remaining bases are undistorted, thereby retaining a slightly twisted B-form conformation (Figure 2B). Leu176 of the thumb-loop inserts into the gap through the minor groove of the DNA duplex, while Leu227 of the NKR finger-loop inserts into the gap through the major groove (Figure 2B). The two inserted leucine residues fill the gap and form hydrophobic stacking interactions with bases of the upstream T6 and downstream A8 bases (Figures 2A and 2B). Moreover, the two-leucine residues are within close enough proximity (3.8-4.2 Å between their Cδ atoms) so as to be within their van der Waals contact distances to form a continuous surface to penetrate the DNA duplex, isolating the 5mC from the duplex and burying it in the SRA domain (Figures S2B, C). The orphaned guanine G7′ base of the unmethylated strand exhibits no base specific recognition with the protein (Figure 2B). By contrast, the UHRF1 SRA domain fills the gap using a valine of the thumb-loop from the minor groove and an

arginine of the NKR finger-loop from the major groove, with the latter arginine pairing with the Hoogsteen edge of the orphaned guanine (Arita et al., 2008; Avvakumov et al., 2008; Hashimoto et al., 2008); similarly, the SUVH5 SRA domain occupies the gap using a glutamine of the thumb-loop through the minor groove, which pairs with the Watson-Crick edge of the orphaned guanine (Rajakumara et al., 2011). Thus, KYP adopts a 5mC flipped-out mechanism, through usage of two leucine residues to occupy the gap introduced by 5mC flipping out, but without pairing with the orphaned guanine.

Conformational Change of DNA in Complex

The insertion of the thumb-loop into the DNA minor groove causes a dramatic conformational change in the DNA. In addition to Leu176, an adjacent residue Trp175, also inserts into the minor groove and forms hydrophobic contacts with the bases of orphaned guanine G7′ base and the next base A8′ (Figures 2A and S2D, E). Such a large magnitude intrusion into the minor groove facilitates the looping-out of the 5mC-containing strand towards the protein (Figure S2E). In essence, the DNA is bent and deviates from the ideal B-form DNA as a consequence of the insertion of the thumb-loop residues Trp175 and Leu176 (Figures S2D, E).

Recognition of the Looped-out 5mC Base

The flipped-out 5mC base is well recognized by the SRA domain using a similar strategy as that reported for the SRA domain of UHRF1 (Arita et al., 2008; Avvakumov et al., 2008; Hashimoto et al., 2008; Rajakumara et al., 2011). In detail, the base is anchored in between Tyr207 and Tyr219 through stacking interactions (stereo view in Figure 2C). The Watson-Crick edge of 5mC forms extensive hydrogen bonding interactions with Ser204, Gln206, Asp210, and Thr220 (Figure 2C). The 5-methyl group inserts into a small hydrophobic pocket formed by Tyr207, Tyr219, Ile179, and the Cα and Cβ of Gln222 (Figure 2C). Several other SRA domain residues of KYP are also involved in the recognition of the DNA backbone within a range spanning 8-bp of DNA (Figure 2A). In addition, the second α-helix of the two-helix uses a positively charged surface to directly interact with the DNA through formation of hydrogen bonds and salt bridges (Figures 2A, B, and S2B, C).

Impact of 5mC-binding Pocket Mutants on *in vitro*-Binding and in vivo-Function

To test the importance of amino acids within KYP associated with binding to methylated DNA, we constructed and purified proteins containing several mutations in the 5mC-binding pocket: L176G, Y207A, D210A, Y219A, and L227G. Binding of each protein with a labeled double-stranded oligonucleotide containing mCHH was visualized using mobility shift assays (Johnson et al., 2007) as shown in Figure 2D. The two tyrosine residues that stabilize the flipped-out 5mC are critical for KYP binding, as is Leu176, which fills the space in the duplex vacated by the flipped-out 5mC. Surprisingly, mutation of Leu227 appears not to be necessary for stable binding, indicating it is not part of the driving force for binding. Mutation of Asp210 had an intermediate effect.

In parallel, we performed *in vitro* histone methyltransferase activity assays to validate the influence of DNA binding to the catalytic activity of the protein. The results indicate that all the mutants preventing the DNA binding, including L176G, Y207A, D210A, Y219A, as

well as L227G which has no effect on DNA binding, have no significant effect on the *in vitro* activity of the enzyme (Figure 2E), reflecting that the binding of DNA by the SRA domain and the catalytic function by the SET domain are biochemically independent. In addition, *in vivo* analysis was also performed by introducing the mutations into a Flag-tagged KYP transgene and transforming modified constructs into a *kyp* loss-of-function line. Initial studies using whole-genome bisulfite sequencing revealed that the wild-type Flag-KYP control construct complements the *kyp* mutant line at only a subset of sites (Figure S3), likely because it is difficult to reinstate the self reinforcing H3K9me/DNA methylation loop at sites where non-CG methylation has been largely lost in the *kyp* mutant. Consistent with this interpretation, the sites that did complement were those retaining relatively high non-CG methylation in the *kyp* mutant background, which likely provide an initial binding site for transformed Flag-KYP. Analysis of SRA mutants L176G, Y219A and D210A at the subset of DMRs that give the highest complementation with wild-type Flag-KYP, revealed complete loss of complementation indicating that SRA function is critical for KYP binding and function (Figure 2F, G). Together, these *in vitro* and *in vivo* functional data are consistent with our structural observation that the SRA domain functions to recruit the pre-SET/SET/post-SET domains to certain loci but without an allosteric regulatory role for its enzymatic activity.

Methylated DNA is Likely Sufficient for Recruitment of KYP to Silent Chromatin

Using an immunoprecipitation-mass spectrometry approach, CMT3 was previously shown to be stably associated with nucleosomes, consistent with its strong dual binding to H3K9 methylation marks through its chromo and BAH domains (Du et al., 2012). Using an identical protocol, we found instead that KYP does not stably associate with histones, or with any other accessory proteins (compare Table S1 with Table 1 from Du et al., 2012), suggesting that KYP's interaction with methylated nucleosomes is most likely transient. The lack of stably associated accessory proteins also suggests that methylated DNA is likely sufficient for recruitment of KYP to silent chromatin. Consistent with this interpretation, and with the specificity of KYP for binding to mCHG and mCHH DNA, we recently showed that a quadruple DNA methyltransferase mutant (*drm1 drm2 cmt2 cmt3*) that eliminates all CHG and CHH methylation also eliminates H3K9 methylation, even though CG methylation is still intact (Stroud et al., 2013a). Thus, our structural and biochemical data on KYP are consistent with available *in vivo* data.

Recognition of Substrate Peptide by pre-SET/SET/post-SET Domains

The pre-SET domain of KYP contains a triangular Zn_3Cys_9 zinc cluster similar to what has been observed in other reported SUV family protein structures (Figure 1B) (Cheng et al., 2005). The cofactor SAH and unmodified H3(1-15) peptide are positioned between the post-SET and SET domains and bound on the opposite side of KYP from that of bound mCHH DNA (Figures 3A, B). The structure suggests that the methylated DNA only serves as a recruitment platform, which targets KYP to mCHH/mCHG containing nucleosomes and directs methylation of the H3K9 tail. The peptide binds within a very narrow channel between the post-SET and SET domains with the H3K9 side chain inserted into a narrow and deep pocket (Figures 3A, B). The post-SET domain adopts a flexible conformation, which is reflected by its extra high B-factor of 123.5 $Å^2$ compared with 40.3 $Å^2$ for the

whole structure, but stabilized in part through coordination of a Zn^{2+} ion by its three cysteine residues together with one additional cysteine residue from the SET domain (Figure 3B).

The ε-ammonium nitrogen of H3K9 forms hydrogen bonds with Tyr475, Glu492, Tyr591, and Tyr593 (Figure 3B). This is consistent with what has been proposed for other reported structures of H3K9 methyltransferaes, in which Tyr residues lining the catalytic pocket are important for activity (Zhang et al., 2003). We generated Y475F, Y591F and Y593F single, as well as Y475/591F and Y475/593F double mutations in order to test the functional effects of the three important tyrosine residues. In our *in vitro* activity assays, the Y475F substitution showed a significant reduction of activity and Y593F, as well as the Y475/593F double mutation led to complete loss of activity, suggesting that these two residues are important for the activity of KYP (Figure 3C, D). In addition, a third tyrosine Tyr591 can form a hydrogen bond with H3K9, which may restrict the orientation of the ε-ammonium nitrogen of H3K9 and thereby restrict KYP to be a mono- and di-methyltransferase. Similar to a previous study showing that a Y591F mutation of KYP converted it from a mono-/di-methyltransferase to a tri-methyltransferase (Ebbs and Bender, 2006), we found that Y591F enhances the di-methyltransferase activity and leads to a gain of tri-methyltransferase activity (Figure 3C, D). Y591F also changes the substrate binding properties of KYP. In contrast to the wild type KYP which is limited to use of unmodified H3 and a low degree of mono-methylated H3 as substrate, Y591F can utilize the unmodified, mono-methylated, as well as di-methylated H3K9 as substrate (Figure 3C). This suggests that Y591 is critical in determining the size of the substrate binding pocket.

Impact of Tyr Residues Lining Catalytic Histone Pocket on *in vivo* KYP Function

The requirement of these residues for *in vivo* KYP function was also measured using whole-genome bisulfite sequencing of plants expressing mutant KYP transgenes (Figures 3E, F) as described above. Mutation of either Y475F or Y593F caused a strong loss of KYP function, consistent with these two residues being important for the activity of KYP. Interestingly, Y591F was able to restore DNA methylation, indicating that the tri-methyltransferase activity does not interfere with the *in vivo* function of KYP (Figure 3E). This observation, along with our previous results that CMT3 binds equally well to the three methylation states of H3K9 (Du et al., 2012; Stroud et al., 2013a), suggests that CMT3 can utilize H3K9me3 as well as H3K9me1 or H3K9me2 as a binding site *in vivo*.

DISCUSSION

Structural Features of KYP Complex that are Distinct from SUVH9 Protein

Our structure of KYP highlights the spatial positioning of the SRA domain with respect to the SET domain, with a pair of N-terminal anti-parallel helices mediating the interaction between the two domains (Figure 1B). This protein architecture has also been observed independently in our recently reported structure of the SUVH9 protein (Johnson et al., 2014). An important distinction is that the structure of SUVH9 was reported in the free state with the protein showing no histone lysine methyltransferase activity due to formation of open peptide and SAH binding pockets incapable of binding these ligands, coupled with our

inability to grow crystals containing bound methylated DNA (Johnson et al., 2014). By contrast, in the present study KYP forms a quartenary complex containing nearly full-length KYP, bound 5mC-containing DNA, H3 peptide and cofactor SAH, thereby opening opportunities for a structure-based understanding of the relationship between DNA and histone methylation.

Although our quartenary KYP structure contains previously identified structural features including the looped out 5mC recognized by the SRA domain (Arita et al., 2008; Avvakumov et al., 2008; Hashimoto et al., 2008; Rajakumara et al., 2011) and histone lysine recognition by the SET domain (Zhang et al., 2003), it is the relative positioning of the SRA and SET domains and their bound ligands within a single protein, and the implications of such alignment for interactions with nucleosomes, that constitutes the significant advance of our structure-function studies.

Model of KYP Bound to Methylated Nucleosomal DNA

It has been shown that DNA methylation can be enhanced on Arabidopsis nucleosomal DNA as compared to linker DNA (Chodavarapu et al., 2010). Based on the likelihood that KYP uses a nucleosome as a substrate to methylate H3K9, we speculate that KYP should be capable of binding nucleosomal mCHH/mCHG DNA, given that the bent mCHH DNA observed in our structure of the complex mimics the DNA wrapped around the nucleosome.

Superimposition of the KYP-bound bent DNA with nucleosomal DNA shows that KYP can bind to the nucleosome without steric conflict in a subset of orientations, one of which is shown in Figure 4A. In addition, in some orientations, the distance is sufficient for H3 to extend from the nucleosome core particle to the position of the KYP bounded H3(1-15) peptide. Further, the KYP-bound H3 peptide exhibits directionality with its C-terminus directed towards the nucleosome core, indicating that it is possible for KYP to bind methylated nucleosomal DNA and further methylate the H3 tail on the same nucleosome to maintain the faithful methylation of H3K9me marks (Figures 4A, B). It is worth noting that the methylated DNA sites are arbitrary across the nucleosome and the length and flexibility of the H3 tail may be key factors influencing the ability of KYP to methylate the H3 tail on the same nucleosome. In an alternate circumstance, if KYP were to bind to the nucleosomal DNA where it was too distant to methylate the H3 tail of the same nucleosome, KYP could methylate an H3 tail from an adjacent nucleosome, outlining a plausible methylation spreading mechanism that would act to further maintain histone methylation states at newly synthesized chromatin (Figure 4B). It should be noted that in the model in Figure 4A, a positively-charged region of the pre-SET domain of KYP (highlighted with a green circle in Figure 4A) is positioned adjacent to the negatively-charged sugar-phosphate backbone of nucleosomal DNA. We wish to emphasize that the model of KYP bound to methylated nucleosomal DNA has been proposed so as to stimulate further structural and functional research on this system and like all models will be subject to further testing prior to validation.

DNA Methylation Regulates Histone Methylation

In our previous work, we have elucidated the structural basis for CMT3 association with H3K9me-containing nucleosomes on the basis of CMT3′s BAH and chromo domains using a dual recognition mechanism (Du et al., 2012). In the present study, our structural and *in vivo* functional data establishes the structure-based mechanism by which KYP uses either mCHG or mCHH DNA to guide targeting of H3K9 methylation, showing for the first time how a DNA methylation mark regulates histone methylation at the molecular level. After replication, only half of new nucleosomes retain pre-existing histone methylation marks. The feedback loop between histone and DNA methylation (Figure 4B) likely functions to ensure the maintenance of the silent state of transposons and other methylation-regulated genes.

Chromatin Modification Crosstalk between DNA and Histone Methylation

Our structure-function studies on CMT3 with bound methylated H3K9 peptides (Du et al., 2012) and KYP with bound methylated DNA, H3 peptide and cofactor SAH (this contribution), where each protein contains both writer and reader modules, provides one of the most compelling examples of chromatin modification crosstalk between DNA and histone methylation and how such crosstalk could act epigenetically through cell division.

EXPERIMENTAL PROCEDURES

Protein and DNA Preparation

Recombinant protein expression and purification was described in detail in the Extended Experimental Procedures. DNA oligos were purchased from Keck Oligonucleotide Synthesis Facility at Yale University and Invitrogen Inc.

Crystallization, Structure Determination, and Refinement

Crystallization was undertaken using the hanging drop vapor diffusion method. The diffraction data were collected at beamline X29A at Brookhaven National Laboratory (BNL), New York and beamline 24ID-E Argonne National Laboratory (ANL), Chicago. All the data were processed with the program HKL2000 (Otwinowski and Minor, 1997). The structure of KYP-mCHH DNA-SAH complex was solved using single-wavelength anomalous dispersion method implemented in the program Phenix (Adams et al., 2010). The model building was carried out using the program Coot (Emsley et al., 2010). All other structures were solved using molecular replacement method using the program Phenix (Adams et al., 2010). The statistics of the diffraction data and refinement are summarized in **Table 1**. Additional details are provided in the Extended Experimental Procedures.

Plant Material and KYP Constrcuts

Details of plant material and KYP constructs can be found in the Extended Experimental Procedures.

Western blots, EMSA, and Histone Methyltransferase Assay

The details of western blots, EMSA, and histone methyltransferase assay are described in the Extended Experimental Procedures.

Whole-genome bisulfite sequencing

Whole-genome bisulfite sequencing (BS-Seq) libraries for initial tests of KYP-FLAG complementation were generated as previously reported (Stroud et al., 2013b), while BS-seq libraries for mutant versions of the KYP-FLAG transgene (as well as parallel control libraries) were generated using TruSeq DNA multiplexing kit (Illumina). Additional details are provided in the Extended Experimental Procedures.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful to the staff of beamlines 24ID-C/E at the Argonne National Laboratory and X29A at the Brookhaven National Laboratory for support in diffraction data collection, and Mahnaz Akhavan and the UCLA Broad Stem Cell Research Center BioSequencing facility for high-throughput sequencing. This work was supported by funds from the Abby Rockefeller Mauze Trust and Mloris and STARR Foundations to D.J.P. Work in the Jacobsen lab was supported by NIH grant GM60398. M.G. was an EMBO ALTF 986-2011 Long-term Fellow. S.F. is a Special Fellow of the Leukemia & Lymphoma Society. C.J.H. is a Howard Hughes Medical Institute Fellow of the Damon Runyon Cancer Research Foundation. S.E.J. is an Investigator of the Howard Hughes Medical Institute.

REFERENCES

Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr. 2010; 66:213–221. [PubMed: 20124702]

Arita K, Ariyoshi M, Tochio H, Nakamura Y, Shirakawa M. Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. Nature. 2008; 455:818–821. [PubMed: 18772891]

Avvakumov GV, Walker JR, Xue S, Li Y, Duan S, Bronner C, Arrowsmith CH, Dhe-Paganon S. Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. Nature. 2008; 455:822–825. [PubMed: 18772889]

Bartee L, Malagnac F, Bender J. Arabidopsis cmt3 chromomethylase mutations block non-CG methylation and silencing of an endogenous gene. Genes Dev. 2001; 15:1753–1758. [PubMed: 11459824]

Cedar H, Bergman Y. Linking DNA methylation and histone modification: patterns and paradigms. Nat Rev Genet. 2009; 10:295–304. [PubMed: 19308066]

Cheng X, Collins RE, Zhang X. Structural and sequence motifs of protein (histone) methylation enzymes. Annu Rev Biophys Biomol Struct. 2005; 34:267–294. [PubMed: 15869391]

Chodavarapu RK, Feng S, Bernatavichute YV, Chen PY, Stroud H, Yu Y, Hetzel JA, Kuo F, Kim J, Cokus SJ, et al. Relationship between nucleosome positioning and DNA methylation. Nature. 2010; 466:388–392. [PubMed: 20512117]

Du J, Zhong X, Bernatavichute YV, Stroud H, Feng S, Caro E, Vashisht AA, Terragni J, Chin HG, Tu A, et al. Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. Cell. 2012; 151:167–180. [PubMed: 23021223]

Ebbs ML, Bender J. Locus-specific control of DNA methylation by the Arabidopsis SUVH5 histone methyltransferase. Plant Cell. 2006; 18:1166–1176. [PubMed: 16582009]

Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr D Biol Crystallogr. 2010; 66:486–501. [PubMed: 20383002]

Goll MG, Bestor TH. Eukaryotic cytosine methyltransferases. Annu Rev Biochem. 2005; 74:481–514. [PubMed: 15952895]

Hashimoto H, Horton JR, Zhang X, Bostick M, Jacobsen SE, Cheng X. The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. Nature. 2008; 455:826–829. [PubMed: 18772888]

Jackson JP, Johnson L, Jasencakova Z, Zhang X, PerezBurgos L, Singh PB, Cheng X, Schubert I, Jenuwein T, Jacobsen SE. Dimethylation of histone H3 lysine 9 is a critical mark for DNA methylation and gene silencing in Arabidopsis thaliana. Chromosoma. 2004; 112:308–315. [PubMed: 15014946]

Jackson JP, Lindroth AM, Cao X, Jacobsen SE. Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. Nature. 2002; 416:556–560. [PubMed: 11898023]

Johnson LM, Bostick M, Zhang X, Kraft E, Henderson I, Callis J, Jacobsen SE. The SRA methyl-cytosine-binding domain links DNA and histone methylation. Curr Biol. 2007; 17:379–384. [PubMed: 17239600]

Johnson LM, Du J, Hale CJ, Bischof S, Feng S, Chodavarapu RK, Zhong X, Marson G, Pellegrini M, Segal DJ, et al. SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation. Nature. 2014; 507:124–128. [PubMed: 24463519]

Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet. 2010; 11:204–220. [PubMed: 20142834]

Lindroth AM, Cao X, Jackson JP, Zilberman D, McCallum CM, Henikoff S, Jacobsen SE. Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation. Science. 2001; 292:2077–2080. [PubMed: 11349138]

Malagnac F, Bartee L, Bender J. An Arabidopsis SET domain protein required for maintenance but not establishment of DNA methylation. EMBO J. 2002; 21:6842–6852. [PubMed: 12486005]

Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. 1997; 276:307–326.

Rajakumara E, Law JA, Simanshu DK, Voigt P, Johnson LM, Reinberg D, Patel DJ, Jacobsen SE. A dual flip-out mechanism for 5mC recognition by the Arabidopsis SUVH5 SRA domain and its impact on DNA methylation and H3K9 dimethylation in vivo. Genes Dev. 2011; 25:137–152. [PubMed: 21245167]

Stroud H, Do T, Du J, Zhong X, Feng S, Johnson L, Patel DJ, Jacobsen SE. The roles of non-CG methylation in Arabidopsis. Nat Struct Mol Biol. 2013a; 21:64–72. [PubMed: 24336224]

Stroud H, Greenberg MV, Feng S, Bernatavichute YV, Jacobsen SE. Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. Cell. 2013b; 152:352–364. [PubMed: 23313553]

Zhang X, Yang Z, Khan SI, Horton JR, Tamaru H, Selker EU, Cheng X. Structural basis for the product specificity of histone lysine methyltransferases. Mol Cell. 2003; 12:177–185. [PubMed: 12887903]

HIGHLIGHTS

- Crystal structure of KYP in complex with mCHH DNA, SAH and H3 peptide
- Two-helix segment of KYP mediates relative alignment of SRA and SET domains
- mCHH DNA and H3 tail are specifically recognized by SRA and SET domains, respectively
- Structural model of KYP and CMT3-controlled DNA and histone methylation feedback loop

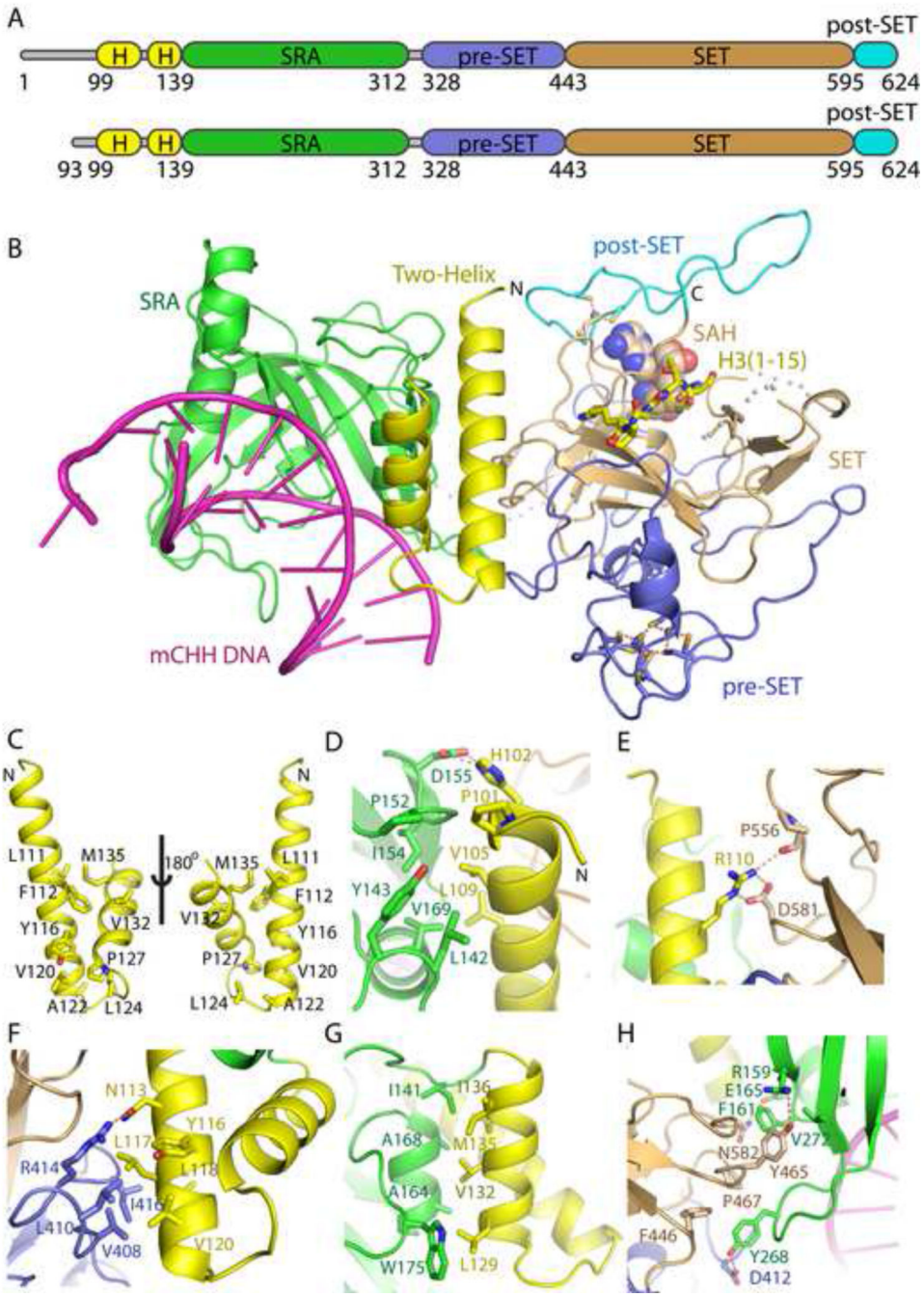

Figure 1. Overall Structure of KYP in Complex with mCHH DNA, SAH and H3(1-15) Peptide

(A) Color-coded domain architecture of full length KYP and KYP(93-624) construct used in this study.

(B) Ribbon representation of the overall structure of KYP in complex with mCHH DNA, SAH and H3(1-15) peptide. The N-terminal anti-parallel two-helix alignment, SRA, pre-SET, SET and post-SET domains of KYP are color-coded in yellow, green, blue, orange, and cyan, respectively. The mCHH DNA, the SAH cofactor, and the H3 peptide are shown in magenta ribbon, space filling, and stick model, respectively. Some disordered loops,

which were not built in the final model, are shown as dashed lines. The Zn_3Cys_9 triangular zinc cluster in the pre-SET domain is highlighted with ball and stick model.
(C) The hydrophobic interactions within the anti-parallel two-helix alignment shown in two views by a 180° rotation. The interacting residues are highlighted in a stick representation.
(D) The N-terminal part of the first helix forms extensive hydrophobic and hydrogen bonding interactions with the SRA domain. The interacting residues are shown in stick representation and hydrogen bonds are shown by dashed red line.
(E) The middle part of the first helix forms hydrogen bonding as well as salt bridge interactions with the SET domain.
(F) The C-terminal part of the first helix forms hydrophobic interactions and hydrogen bonding interaction with the pre-SET domain.
(G) One side of the short helix has several hydrophobic residues that form extensive hydrophobic interactions with the SRA domain.
(H) The SRA domain forms both hydrophobic and hydrogen bonding interactions with the pre-SET and SET domains.
See also Figure S1.

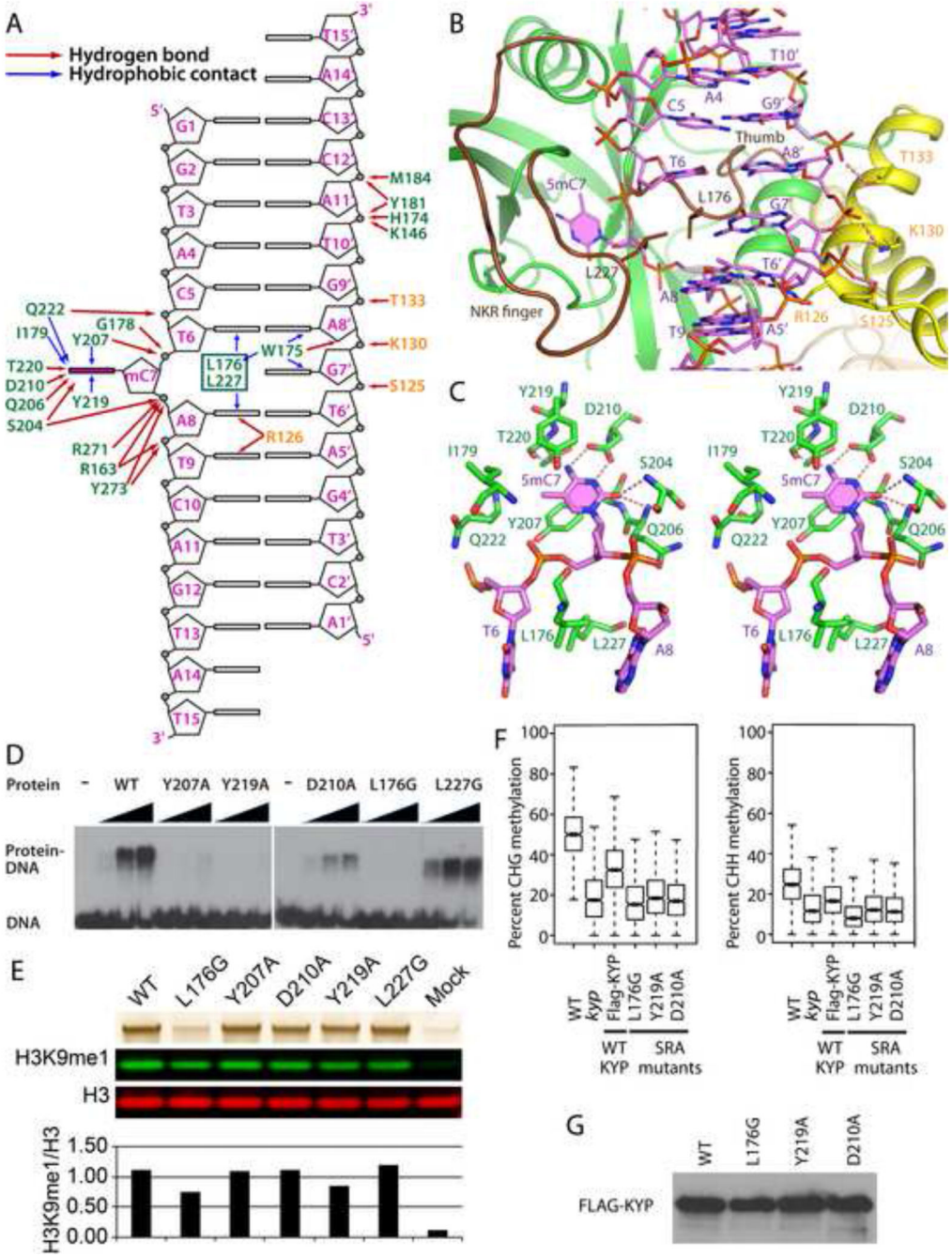

Figure 2. Recognition of mCHH DNA by KYP

(A) Schematic representation of the interactions between KYP and DNA. Hydrogen bonds are shown by red arrows and hydrophobic contacts by blue arrows. The SRA domain residues and the two-helix alignment residues are colored in green and yellow, respectively. (B) The SRA pocket accommodating the flipped-out 5mC and the two-helix alignment residues interacting with the DNA. The 5mC base is highlighted by a solid magenta hexagon. The thumb- and NKR finger-loops are colored in brown with the Leu176 and

Leu227 highlighted in a stick representation. The hydrogen bonds between the two-helix alignment residues and mCHH DNA are shown in dashed red lines.

(C) The detailed recognition of the flipped out 5mC base by residues lining the binding pocket within the SRA domain.

(D) Electrophoretic mobility-shift assays using a mCHH double-stranded DNA and increasing levels (50, 100, and 200 ng) of the indicated protein. Similar results were observed using mCHG substrate (data not shown).

(E) *In vitro* methylation of H3 by KYP SRA domain mutants. KYP protein (upper panel, silver-stained) was incubated with S-adenosyl methionine (SAM) and recombinant H3. Histone methyltransferase activity was tested by quantitative western blot using primary antibodies against H3K9me1 and H3 and infrared secondary antibodies (green: 800 nm, red: 680 nm).

(F) Boxplots of CHG and CHH context DNA methylation at a subset of *kyp* CHG hypomethylated DMRs that show complementation upon transformation with a wild-type KYP construct. All the whiskers on the box plots represent plus/minus 1.5x iqr (inter quartile range).

(G) Western blot analysis of lines used in the complementation studies showing expression levels comparable to the Flag-KYP/*kyp* (WT).

See also Figures S2, S3 and Table S1.

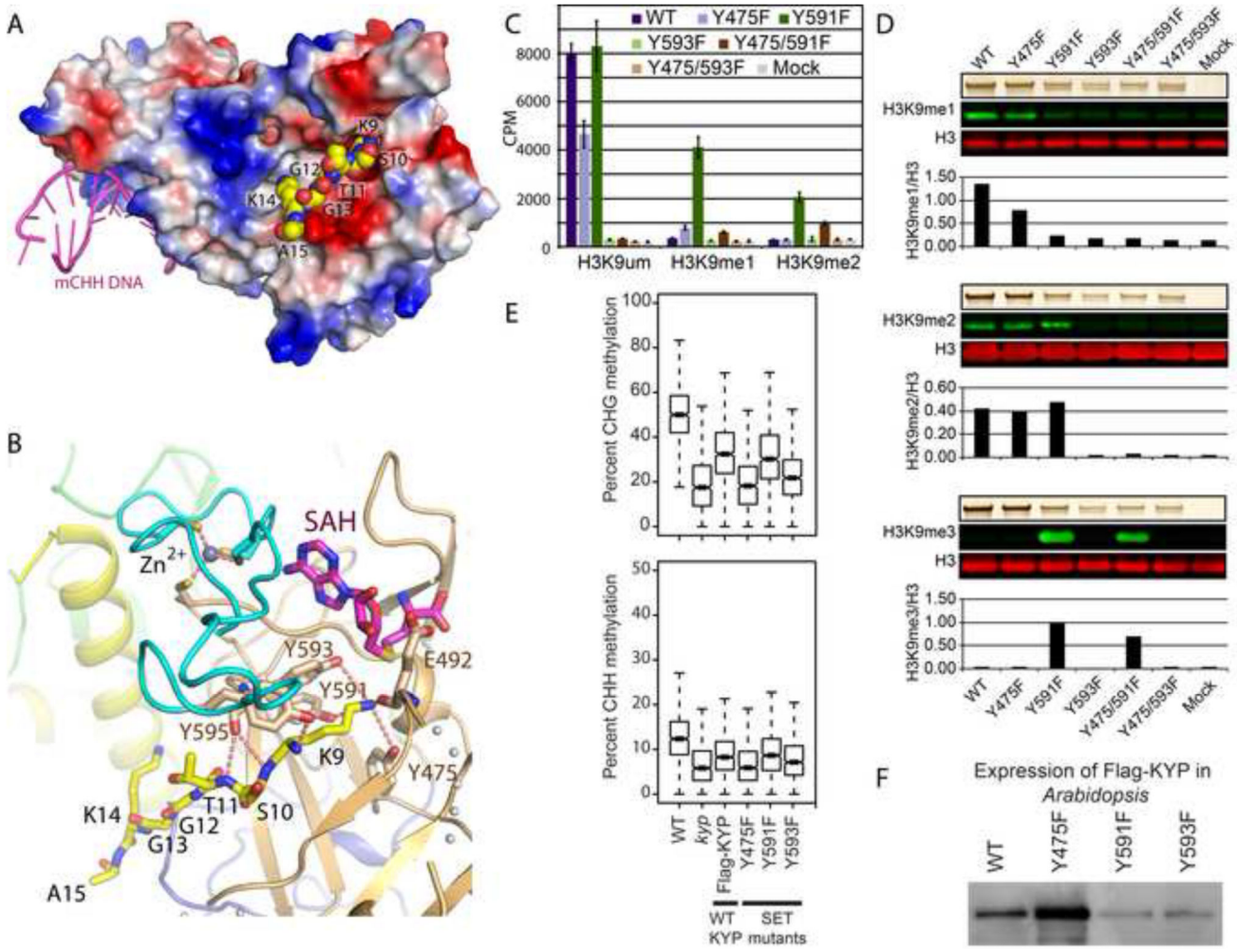

Figure 3. Recognition of the Substrate H3 Peptide by the SET Domains of KYP

(A) An electrostatics surface representation of KYP. The bound H3 peptide, in a space filling representation, is inserted into a negatively-charged binding channel. The mCHH DNA in magenta cartoon representation aligns along the opposite side of KYP.

(B) The intermolecular interactions amongst SAH, peptide and KYP in the complex. The peptide fits into a narrow channel between the SET (orange) and post-SET (cyan) domains, with the H3K9 inserting into a narrow deep pocket, where it is stabilized through formation of extensive intermolecular hydrogen bonding interactions. The zinc-binding motif, which stabilizes the fold of the post-SET domain, is highlighted with ball and stick representation.

(C) Methyltransferase activity of KYP WT and SET domain mutants. Radioactivity (CPM) of H3 peptides was measured after incubation of unmethylated (H3K9um), monomethylated (H3K9me1) or dimethylated (H3K9me2) substrate with KYP protein and tritiated SAM (n = 3, ± S.D.).

(D) Differential *in vitro* methylation of H3 by KYP SET domain mutants. KYP protein (upper panels, silver-stained) was incubated with SAM and recombinant H3. Histone methyltransferase activity was tested by quantitative western blots using primary antibodies

against H3 and H3K9me1, H3K9me2 or H3K9me3 and infrared secondary antibodies (green: 800 nm, red: 680 nm).

(E) Boxplots of CHG and CHH context DNA methylation at a subset of *kyp* CHG hypomethylated DMRs that show complementation upon transformation with a wild-type KYP construct (see Supplemental Methods).

(F) Western blot analysis of lines used in the complementation studies showing expression levels comparable to the Flag-KYP/*kyp* (WT).

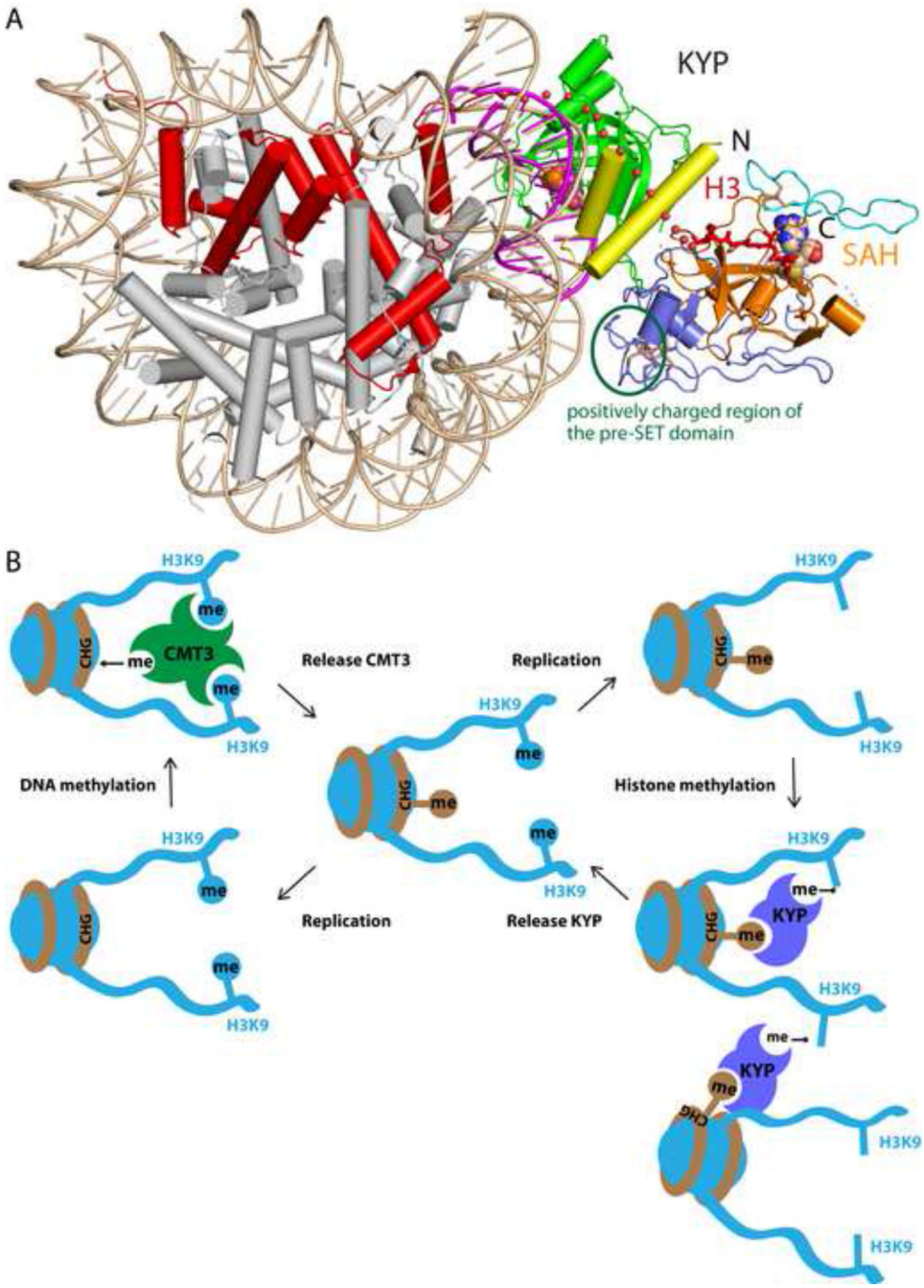

Figure 4. A Working Model for the Epigenetics Mechanism Controlling H3K9me by KYP

(A) Modeling of KYP on the nucleosomal DNA indicates that KYP can be bound to methylated nucleosomal DNA and further methylate the H3 tail of the same nucleosome. The KYP is colored as Figure 1B. The H3 is highlighted in red. The nucleosomal DNA and other histone proteins are colored in wheat and silver, respectively. The flipped out 5mC is highlighted in space filling model to indicate its position. A green circle marks a positively charged region within the pre-SET domain that is adjacent the nucleosomal DNA, indicating plausible interaction between them.

(B) A schematic model of CMT3 and KYP controlled self-reinforcing feedback loop between mCHG and H3K9me.

Table 1
Summary of X-ray Diffraction Data and Structure Refinement Statistics

Summary of diffraction data			
Crystal	KYP-mCHH -SAH	KYP-mCHH -SAH-H3(1-15)	KYP-mCHG -SAH
Beamline	BNL-X29A	BNL-X29A	ANL-24ID-E
PDB code	4QEN	4QEO	4QEP
Wavelength (Å)	1.2830	1.0750	0.9793
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell parameters			
a, *b*,*c* (Å)	54.4, 98.0, 122.6	55.5, 96.8, 122.4	54.3, 95.6, 121.7
Resolution (Å)	50.0-2.0 (2.07-2.00)[a]	50.0-2.0 (2.07-2.00)	50.0-3.1 (3.21-3.10)
R_{merge} (%)	6.7 (61.0)	7.4 (76.9)	15.5 (61.5)
Observed reflections	609,995	319,270	38,784
Unique reflections	44,950	45,280	10,851
Average I/σ(I)	48.1 (3.7)	36.1 (2.6)	9.5 (2.0)
Completeness (%)	99.9 (100.0)	99.9 (100.0)	89.6 (92.8)
Redundancy	13.6 (12.7)	7.1 (7.0)	3.6 (3.5)

Refinement and structure model			
R / Free *R* factor (%)	17.6 / 21.1	18.8 / 22.6	21.2 / 25.1
non-H atoms	4624	4570	4210
Protein / peptide	3,721 / -	3721 / 44	4690 / -
DNA / SAH	490 / 26	490 / 26	490 / 26
Zn^{2+} / Water	4 / 383	4 / 285	4 / -
B factor (Å^2)	40.3	51.9	43.1
Protein / peptide	40.0 / -	50.0 / 95.4	40.4 / -
DNA / SAH	56.0 / 38.7	64.1 / 77.3	61.6 / 78.3
Zn^{2+} / Water	42.0 / 43.2	44.9 / 46.8	45.1 / -
RMS deviations			
Bond lengths (Å)	0.008	0.011	0.016
Bond angles (°)	1.410	1.367	1.601

[a] Values in parentheses are for highest