

Structural insights into Rhino-mediated germline piRNA cluster formation

Cell Research (2015) 25:525-528. doi:10.1038/cr.2015.10; published online 23 January 2015

Dear Editor,

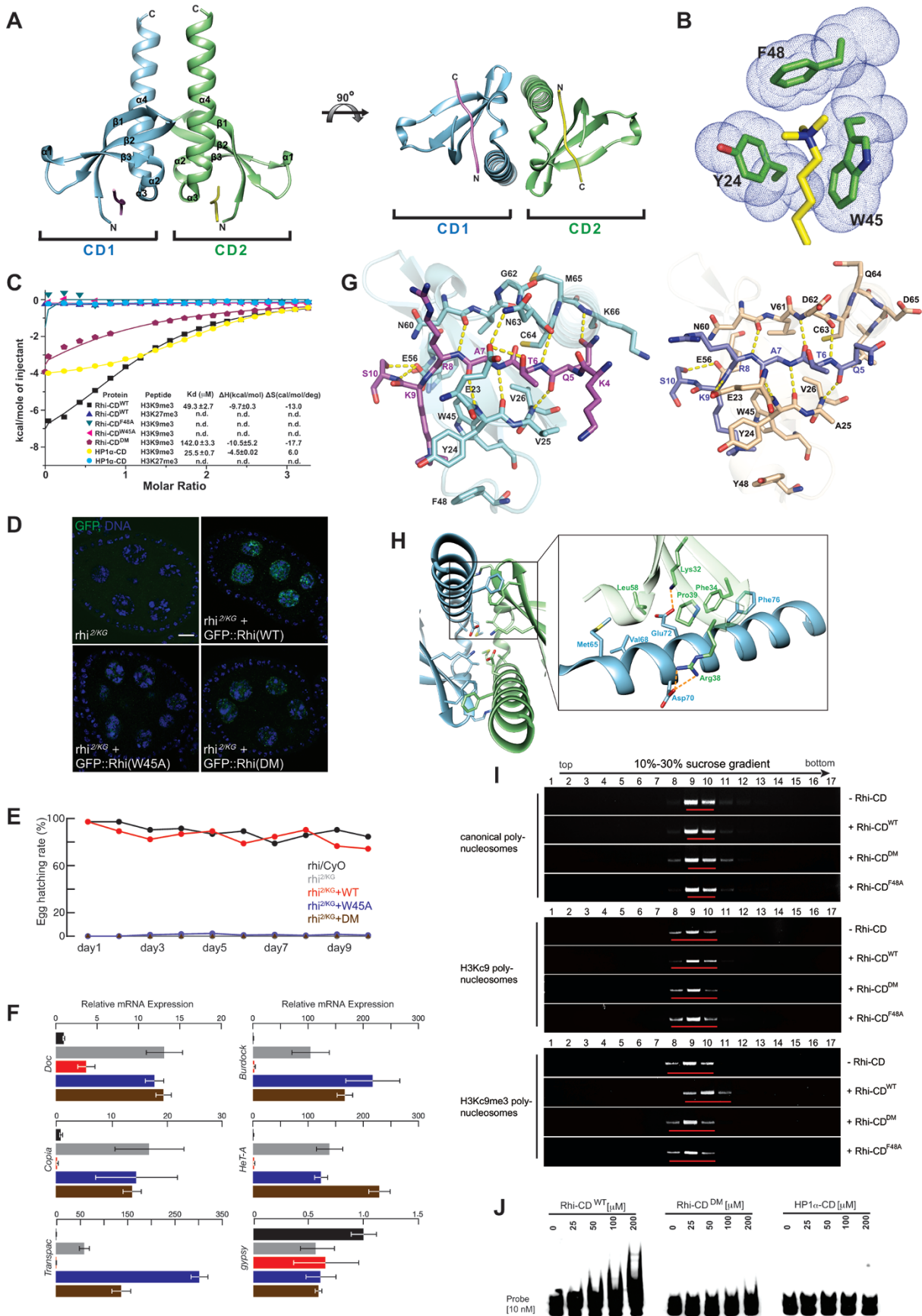
The piRNA pathway silences transposons in animal gonads. Animals lacking piRNA pathway components often show transposon activation in the germline and compromised fertility [1]. In *Drosophila* germ cells, piRNA precursors are transcribed from heterochromatic piRNA cluster regions, and then transported into nuage/mitochondria to be processed into mature piRNAs that are loaded into PIWI clade Argonautes [1]. Guided by piRNAs, PIWI proteins silence transposons at either transcriptional or post-transcriptional level [2]. *Drosophila* ovaries contain two types of piRNA clusters, which are often coated with trimethylated lysine 9 of H3 (H3K9me3). Dual-strand clusters are mainly active in germ cells and produce piRNAs from both genomic strands, while uni-strand clusters are dominant in somatic follicle cells and produce piRNAs from only one genomic strand [2]. However, how piRNA clusters are defined in the genome remains elusive. Recent studies have shown that a Heterochromatin Protein 1 (HP1) homolog Rhino (Rhi), also known as HP1D, is anchored at the dual-strand piRNA clusters to trigger the piRNA production [3, 4]. Rhi is a germline-specific protein that is under rapid and positive selection during evolution [5]. Like other HP1 family proteins, Rhi exists in all *Drosophila* species and shares the similar domain architecture, which contains a chromodomain (CD) at the N-terminus, a chromoshadow domain (CSD) at the C-terminal region, and a hinged region between CD and CSD domains [5].

As a “royal family” member, the CD has been predominantly characterized as a methylated lysine-binding module [5]. To understand the mechanism by which Rhi specifically recognizes piRNA clusters, we solved a 1.8-Å resolution crystal structure of Rhi-CD in complex with the H3K9me3 peptide by molecular replacement (Supplementary information, Table S1). Rhi-CD shared canonical CD structural features commonly found in HP1 family proteins, including a curved anti-parallel β -sheet comprising three strands (β 1, β 2, and β 3), with a turn α 1 occurring between β 2 and β 3 and turns α 2, α 3, and helix

4 following β 3 (Figure 1A and Supplementary information, Figure S1A). The methylated histone peptide bound to Rhi-CD in a cleft mainly formed by α 2, α 3, and three N-terminal residues. The interaction between Rhi-CD and H3K9me3 was largely composed of main-chain hydrogen bond interactions, involving residues Y24, V26, N60 and N63 of Rhi-CD, and Q5, T6, A7 and R8 of histone H3 peptide. In addition, the main-chain residues R8 and S10 of the peptide formed hydrogen bonds with the side chains of residues N60 and Q56, respectively. We further found that the trimethyllysine of the H3K9me3 peptide is bound in an aromatic cage formed by Y24, W45, and F48 (Figure 1B). As expected, the isothermal titration calorimetry (ITC) assays showed that either W45A or F48A mutant abolished the binding of Rhi-CD to the peptide (Figure 1C).

To examine whether the binding to H3K9me3 is important for Rhi function *in vivo*, we expressed GFP-tagged wild-type Rhi or Rhi^{W45A} protein in the *rhi*^{2/KG} mutant flies. Under laser scanning confocal microscope, wild-type Rhi formed germline-specific nuclear foci, which may represent piRNA cluster regions in the genome. However, the typical Rhi foci were dramatically reduced in the ovaries from Rhi^{W45A} flies (Figure 1D and Supplementary information, Figure S1B). Importantly, transgenic expression of wild-type Rhi, but not Rhi^{W45A} in *rhi*^{2/KG} mutant flies successfully rescued sterility and transposon hyper-expression defects in these animals (Figure 1E and 1F). Collectively, these results suggest that the CD-mediated H3 lysine binding is indispensable for Rhi function *in vivo*.

Previous studies indicated that the selective binding to H3K9me3 or H3K27me3 is a fundamental feature for CD proteins to regulate the chromatin dynamics and gene expression [6, 7]. In histone H3, the amino acids around K9 site (QTARK⁹S) is similar to K27 site (KAARK²⁷S) with a common motif ARKS. T6 in histone H3 peptide and residues D62 and E23 in HP1 α -CD have been revealed as key determinants to the specific recognition of methylated K9, in which D62 and E23 form a “polar finger” to clasp the side chain of T6 [8]. Interestingly,



the corresponding residues in other chromobox (Cbx) family proteins that bind H3K27me3 are mutated to Leu and Val instead, which may facilitate the hydrophobic interaction with A24 in H3K27me3 peptide [8]. Rhi-CD showed stronger binding selectivity to H3K9me3 than to H3K27me3 peptide with the equilibrium dissociation constant (K_d) of 49 μ M (Figure 1C). We also measured the binding affinity of a mouse HP1 α -CD to H3K9me3, and found a K_d of 25 μ M, indicating that HP1 α -CD has a stronger binding affinity to H3K9me3 than Rhi-CD. Analysis of the surface electrostatic potential map showed that mouse HP1 α -CD had a more negatively charged surface than Rhi-CD in favor of binding to a positively charged histone peptide (Supplementary information, Figure S1C). However, sequence alignment showed that an additional residue G62 was inserted between α 2 and α 4 in Rhi, which is lacking in other CDs of the Cbx family, but highly conserved among *Drosophilidae* Rhi proteins (Supplementary information, Figure S1A). The insertion of G62 led to the formation of turn α 3, which terminates the β -strand hydrogen bonding with histone H3 peptide and alters the side-chain conformation of the neighboring residue N63 (corresponding to D62 in HP1 α -CD; Figure 1G). Moreover, N63 was observed to form hydrogen bond with E23 and stabilize the interaction of E23 with the hydroxyl group of T6, which might be a major contributor to the specific recognition of H3K9 by Rhi-CD.

Intriguingly, we found that Rhi-CD, unlike HP1 α -CD, formed a homodimer through the interaction between helix α 4 of one molecule and residues on β 1, β 2, and α 2 of the other (Figure 1H). F76 on α 4 of CD1 showed a π - π stacking with F34 located on β 1 of the CD2. And L58, M65, and V68 of CD1 and CD2 created a hydrophobic cluster. Additionally, the Rhi-CD dimer was further stabilized by two salt bridges formed by E72 on α 4 of CD1 and K32 on β 1 from CD2, and by D70 of CD1 and R38 of CD2. Sequence alignment showed that residues in-

involved in dimer formation are highly conserved among *Drosophila* Rhi proteins, but less conserved in Cbx family proteins (Supplementary information, Figure S1A). To validate whether the Rhi-CD also forms a homodimer in solution, we generated a F34A/F76A double mutant (Rhi-CD^{DM}) to disrupt the π - π stacking interaction. The Rhi-CD^{DM} representing the monomeric state of Rhi-CD was analyzed by circular dichroism, size exclusive chromatography, and dynamic light scattering assays (Supplementary information, Figure S1D-S1F). Together, these results suggest that Rhi-CD forms a homodimer both in crystal and in solution.

We next examined whether the dimerization is required for H3K9me3 binding. Our ITC assays showed that the binding of Rhi-CD^{DM} to H3K9me3 peptide was about 3-fold and 6-fold weaker than those of wild-type Rhi-CD and HP1 α -CD (Figure 1C), respectively, indicating that dimerization of Rhi proteins is important for its binding to H3K9me3. We further reconstituted H3K9me3 polynucleosomes, an analog of H3K9me3 polynucleosomes and examined the binding activity of Rhi-CD to H3K9me3-containing polynucleosomes by sucrose gradient sedimentation method. When the canonical polynucleosomes without histone modification or H3K9 polynucleosomes was used as a substrate, neither wild-type nor Rhi-CD mutants could result in any shift (Figure 1I). As expected, Rhi-CD^{WT} caused the H3K9me3 polynucleosomes to shift one fraction. However, both Rhi-CD^{F48A} and Rhi-CD^{DM} mutants failed to achieve the polynucleosome migration change (Figure 1I). These results indicate that both H3K9me3 mark on polynucleosomes and dimerization of Rhi-CD are required for Rhi-CD binding to polynucleosomes. To explore the importance of Rhi dimerization *in vivo*, we constructed transgenic flies bearing GFP-tagged Rhi^{DM} (Figure 1D and Supplementary information, Figure S1B). Similar to Rhi^{W45A}, Rhi^{DM} produced much less prominent foci typically observed from wild-type Rhi protein and failed to

Figure 1 Structure of the *Drosophila* Rhi chromodomain dimer in complex with H3K9me3. **(A)** Overall structure of Rhi-CD dimer and H3K9me3 complex in perpendicular views. **(B)** Aromatic residues that coordinate the methylated lysine group. **(C)** ITC measurements of binding affinities of wild-type and mutant Rhi-CD as well as HP1 α -CD to trimethylated histone peptides, as indicated. **(D)** GFP staining indicating the localization of wild-type and mutant Rhi. Scale bar, 10 μ m (applies to all panels). **(E)** Summary of egg hatching rate from *rhi* heterozygotes, mutants and mutants with wild-type or mutant rescue construct. **(F)** Steady-state transposon mRNA levels from the indicated genotypes. Color scheme is the same as **E**. *gypsy* is a transposon family mainly active in somatic follicle cells, where Rhi is not produced (the bottom right panel). Data were collected from three biological replicates, and error bars stand for standard deviations. **(G)** Interaction networks of Rhi-CD (cyan) with H3K9me3 peptide (magenta) (left) and HP1 α -CD (light brown) with H3K9me3 peptide (slate blue) (right). Yellow dashed lines indicate the hydrogen bonds. **(H)** A schematic diagram showing the interactions between CD1 and CD2 of the Rhi-CD dimer. **(I)** Binding of wild-type and mutant Rhi-CD to polynucleosomes *in vitro* analyzed by sucrose gradient sedimentation analysis. Each sample was fractionated and DNA collected from individual fractions was sized on 1% agarose gel. **(J)** EMSA showing that only Rhi-CD dimer binds a dsDNA probe (17 bp). Concentrations of wild-type and mutant Rhi-CD are indicated, as well as HP1 α -CD.

rescue *rhi* mutation defects: sterility and transposon activation (Figure 1E and 1F). Therefore, the dimerization of Rhi is obligatory for its function *in vivo*.

Our crystal structure showed that two histone H3K9me3 peptides simultaneously bound to the Rhi-CD dimer in an anti-parallel way. Given that it is sterically unfavorable for Rhi-CD to bind both H3 tails from the same nucleosome, a special recognition of H3K9me3 heterochromatin by Rhi-CD dimer should be required. Due to the lack of information of heterochromatin structure, we proposed a model based on a recently published cryo-EM structure of the chromatin fiber, which shows a double helix twisted by tetranucleosomal units [9]. According to the structure, histone tails from the polynucleosomes on both strands of the chromatin double helix may adopt an anti-parallel conformation in favor of a Rhi-CD dimer binding (Supplementary information, Figure S1G). Upon binding, Rhi-CD is probably located in the groove of the double helix formed by the linker DNAs connecting neighboring nucleosomes, which leads to a hypothesis that Rhi-CD may interact with DNA directly. To test this model, we performed electrophoretic mobility shift assay (EMSA), fluorescence polarization and bio-layer interferometry using dsDNA as probes and the results showed that Rhi-CD^{WT} bound to dsDNA with a relatively higher affinity compared to Rhi-CD^{DM} and HP1 α -CD with a K_d of $\sim 51.0 \mu\text{M}$ (Supplementary information S1H and S1I). Rhi-CD^{WT} produced a dosage-dependent migration of the probe, while Rhi-CD^{DM}, as well as HP1 α -CD, failed to shift the same probe (Figure 1J). Surface electrostatic potential analysis indicated a potential DNA-binding surface on the positively charged C-terminal helix $\alpha 4$ of Rhi-CD (Supplementary information, Figure S1J).

In *Drosophila*, piRNA clusters often located near the boundary zone between euchromatin and heterochromatin at the pericentromeric and subtelomeric regions, and are featured by the heterochromatic marker, H3K9me3 [11]. Our recent study, along with a report by Mohn *et al.* [4], suggests that Rhi uniquely marks the regions for piRNA production [3]. In the present study, we provide the structural evidence that Rhi binds with histone H3K9me3 through its CD in a dimerization-dependent manner. Our data also suggest that both H3K9me3 binding and dimerization are critical for its function *in vivo*.

Noticeably, besides piRNA clusters, H3K9me3 also covers a broad spectrum of genome; however, it remains largely unknown why Rhi is only recruited to piRNA cluster regions. Our findings may provide critical molecular insights into the recognition of piRNA clusters by Rhi. One intriguing model is that the density of H3K9me3 on clusters may favor the recruitment of the Rhi dimer and the requirement that only anti-parallel H3 tails can be simultaneously bound by the Rhi dimer may

keep the polynucleosome in piRNA cluster in an “open” conformation, which thus promotes the accessibility of Pol II to H3K9me3 heterochromatin and initiates the biogenesis of piRNA. Further structural studies on Rhi binding with H3K9me3 polynucleosome will elucidate how Rhi marks the dual-strand piRNA clusters in the genome.

The atomic coordinate has been deposited in the Protein Data Bank with the accession code 4U68.

Note

Le Thomas *et al.* [10] also reported Rhi-CD structure with H3K9me3 peptide when this manuscript was under review.

Acknowledgments

We thank Prof William E Theurkauf at University of Massachusetts Medical School for providing the Rhino plasmids. This work was supported by grants from the Ministry of Science and Technology of China (2011CB966304, 2012CB910502), the National Natural Science Foundation of China (31270774), and the Strategic Priority Research Program (XDB08010202). We thank the staff at BL17U of Shanghai Synchrotron Radiation Facility (SSRF) for their help with data collection.

Bowen Yu^{1,5}, Madeline Cassani², Min Wang^{3,5}, Mofang Liu⁶, Jinbiao Ma⁴, Guohong Li³, Zhao Zhang², Ying Huang¹

¹State Key Laboratory of Molecular Biology, National Center for Protein Science Shanghai, ⁶Collaborative Innovation Center of Genetics and Development, Shanghai Key Laboratory of Molecular Andrology, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China; ²Department of Embryology, Carnegie Institution for Science, Baltimore, MD 21218, USA; ³National Laboratory of Biomacromolecules, Chinese Academy of Sciences, Beijing 100101, China; ⁴State Key Laboratory of Genetic Engineering, Collaborative Innovation Center of Genetics and Development, Department of Biochemistry, School of Life Sciences, Fudan University, Shanghai 200438, China; ⁵University of Chinese Academy of Sciences, Beijing 100049, China

Correspondence: Ying Huang^a, Zhao Zhang^b

^aE-mail: huangy@sicb.ac.cn

^bE-mail: zhang@ciwemb.edu

References

- Siomi MC, Sato K, Pezic D, *et al.* *Nat Rev Mol Cell Biol* 2011; **12**:246-258.
- Guzzardo PM, Muerdter F, Hannon GJ. *Curr Opin Genet Dev* 2013; **23**:44-52.
- Zhang Z, Wang J, Schultz N, *et al.* *Cell* 2014; **157**:1353-1363.
- Mohn F, Sienski G, Handler D, *et al.* *Cell* 2014; **157**:1364-1379.
- Vermaak D, Henikoff S, Malik HS. *PLoS Genet* 2005; **1**:96-108.
- Martin C, Zhang Y. *Nat Rev Mol Cell Biol* 2005; **6**:838-849.
- Hublitz P, Albert M, Peters AH. *Int J Dev Biol* 2009; **53**:335-354.
- Kaustov L, Ouyang H, Amaya M, *et al.* *J Biol Chem* 2011; **286**:521-529.
- Song F, Chen P, Sun D, *et al.* *Science* 2014; **344**:376-380.
- Le Thomas A, Stuwe E, Li S, *et al.* *Genes Dev* 2014; **28**:1667-1680.
- Yamanaka S, Siomi MC, Siomi H. *Mob DNA* 2014; **5**:22.

(Supplementary information is linked to the online version of the paper on the *Cell Research* website.)