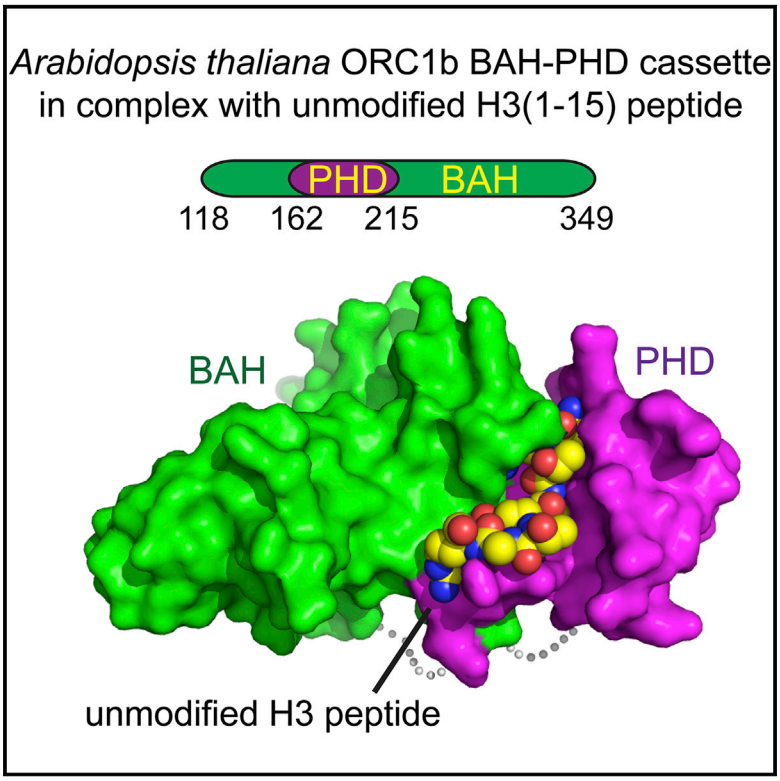
Short Article

Structure

Structural Basis for the Unique Multivalent Readout of Unmodified H3 Tail by *Arabidopsis* ORC1b BAH-PHD Cassette

Graphical Abstract



Authors

Sisi Li, Zhenlin Yang, Xuan Du, ..., Steven E. Jacobsen, Dinshaw J. Patel, Jiamu Du

Correspondence

jmdu@sibs.ac.cn

In Brief

Li et al. report on the structure of *Arabidopsis* ORC1b BAH-PHD cassette in complex with an unmodified H3(1–15) peptide. The structural and biochemical data showed a unique multivalent readout of unmodified H3 tail at H3R2, H3K4, and H3T3, providing a potential link between plant DNA replication and epigenetic regulation.

Highlights

- N-terminal BAH-PHD cassette of plant ORC1 recognizes unmodified H3 tail
- Structure of *Arabidopsis* ORC1b BAH-PHD cassette-unmodified H3 peptide complex
- Structural basis for the recognition of unmodified H3 tail by plant ORC1
- A multivalent histone mark recognition module impacting on epigenetic regulation

Accession Numbers

5HH7

Li et al., 2016, Structure *24*, 486–494
March 1, 2016
http://dx.doi.org/10.1016/j.str.2016.01.004

CellPress

Structure
Short Article

Structural Basis for the Unique Multivalent Readout of Unmodified H3 Tail by *Arabidopsis* ORC1b BAH-PHD Cassette

Sisi Li,[1] Zhenlin Yang,[2,5] Xuan Du,[2,5] Rui Liu,[2] Alex W. Wilkinson,[3] Or Gozani,[3] Steven E. Jacobsen,[4] Dinshaw J. Patel,[1] and Jiamu Du[2,*]

[1]Structural Biology Program, Memorial Sloan-Kettering Cancer Center, New York, NY 10065, USA
[2]Shanghai Center for Plant Stress Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 201602, China
[3]Department of Biology, Stanford University, Stanford, CA 94305, USA
[4]Department of Molecular, Cell, and Developmental Biology, Howard Hughes Medical Institute, University of California at Los Angeles, Los Angeles, CA 90095, USA
[5]University of Chinese Academy of Sciences, Beijing 100049, China
*Correspondence: jmdu@sibs.ac.cn
http://dx.doi.org/10.1016/j.str.2016.01.004

SUMMARY

DNA replication initiation relies on the formation of the origin recognition complex (ORC). The plant ORC subunit 1 (ORC1) protein possesses a conserved N-terminal BAH domain with an embedded plant-specific PHD finger, whose function may be potentially regulated by an epigenetic mechanism. Here, we report structural and biochemical studies on the *Arabidopsis thaliana* ORC1b BAH-PHD cassette which specifically recognizes the unmodified H3 tail. The crystal structure of ORC1b BAH-PHD cassette in complex with an H3(1–15) peptide reveals a strict requirement for the unmodified state of R2, T3, and K4 on the H3 tail and a novel multivalent BAH and PHD readout mode for H3 peptide recognition. Such recognition may contribute to epigenetic regulation of the initiation of DNA replication.

INTRODUCTION

DNA replication is an essential biological event required for faithful inheritance of the genome from parent to offspring. Replication, which initiates at genomic sites named replication origins, is regulated by a number of pathways, many of which are species specific (Aladjem, 2007; DePamphilis et al., 2006). The biochemical machinery of initiation of replication relies on the assembly of the origin recognition complex (ORC), which is conserved from yeast to higher eukaryotes. After ORC assembly, it can subsequently recruit other replication-related factors such as CDT1, CDC6, and hexameric minichromosome maintenance helicase, which together initiate replication (Bell and Stillman, 1992; Leatherwood, 1998). The ORC complex is composed of an assembly of six protein subunits, named ORC1 through ORC6 (Duncker et al., 2009), most of which contain a conserved AAA+ ATPase domain plus a winged helix domain, which function in oligomerization of ORC subunits and the recognition of the origin DNA (Bleichert et al., 2015; Dueber et al., 2007; Gaudier et al., 2007).

A unique feature of ORC1 relative to the other ORC subunits is the presence of an N-terminal BAH domain (Duncker et al., 2009). In yeast, the BAH domain of Orc1p mediates recognition of the silence information regulator 1 protein (Sir1p) to establish epigenetic silencing at target loci (Triolo and Sternglanz, 1996). The structure of yeast Orc1p BAH domain features a classic BAH domain fold with a small insertion of a non-canonical helical domain (H domain), which mediates the interaction with Sir1p (Hou et al., 2005; Hsu et al., 2005; Zhang et al., 2002). In metazoans, the ORC1 N-terminal BAH domain is functionally important, and mutations within this domain in humans are present in patients suffering from Meier-Gorlin syndrome (MGS), a form of primordial dwarfism (Bicknell et al., 2011a, 2011b; Guernsey et al., 2011). Metazoan ORC1 BAH domains, including the human and mouse versions, specifically recognize the histone mark H4K20me2, a mark associated with DNA damage and DNA replication (Kuo et al., 2012). The structure of mouse ORC1 BAH domain in complex with an H4K20me2 peptide highlights a four-residue aromatic cage that specifically recognizes the methyllysine mark (Kuo et al., 2012). Mutations within the BAH domain of ORC1 that abrogate H4K20me2-recognition leads to an MGS-like phenotype in a zebrafish model system. Additional studies also indicated that the human ORC1 BAH domain can directly interact with DNA and that an MGS patient mutation that does not affect H4K20me2-binding is important for DNA binding, particularly in the context of the nucleosome (Zhang et al., 2015). Plant ORC1 proteins possess a BAH domain, but unlike in yeast and metazoans there is a PHD finger embedded within the primary sequence of the BAH domain, thereby generating a plant-specific fused BAH-PHD cassette arrangement (Figures 1A and S1). The PHD finger of *Arabidopsis* ORC1 was reported to bind H3K4me3 peptides by an in vitro pull-down assay (de la Paz Sanchez and Gutierrez, 2009). At the genome level, *Arabidopsis* DNA replication origin sites exhibit correlation with G + C enriched regions, and histone H2A.Z, H3K4me2, H3K4me3, and H4K5ac marks, as well as anti-correlation with H3K4me1 and H3K9me2 marks (Costas

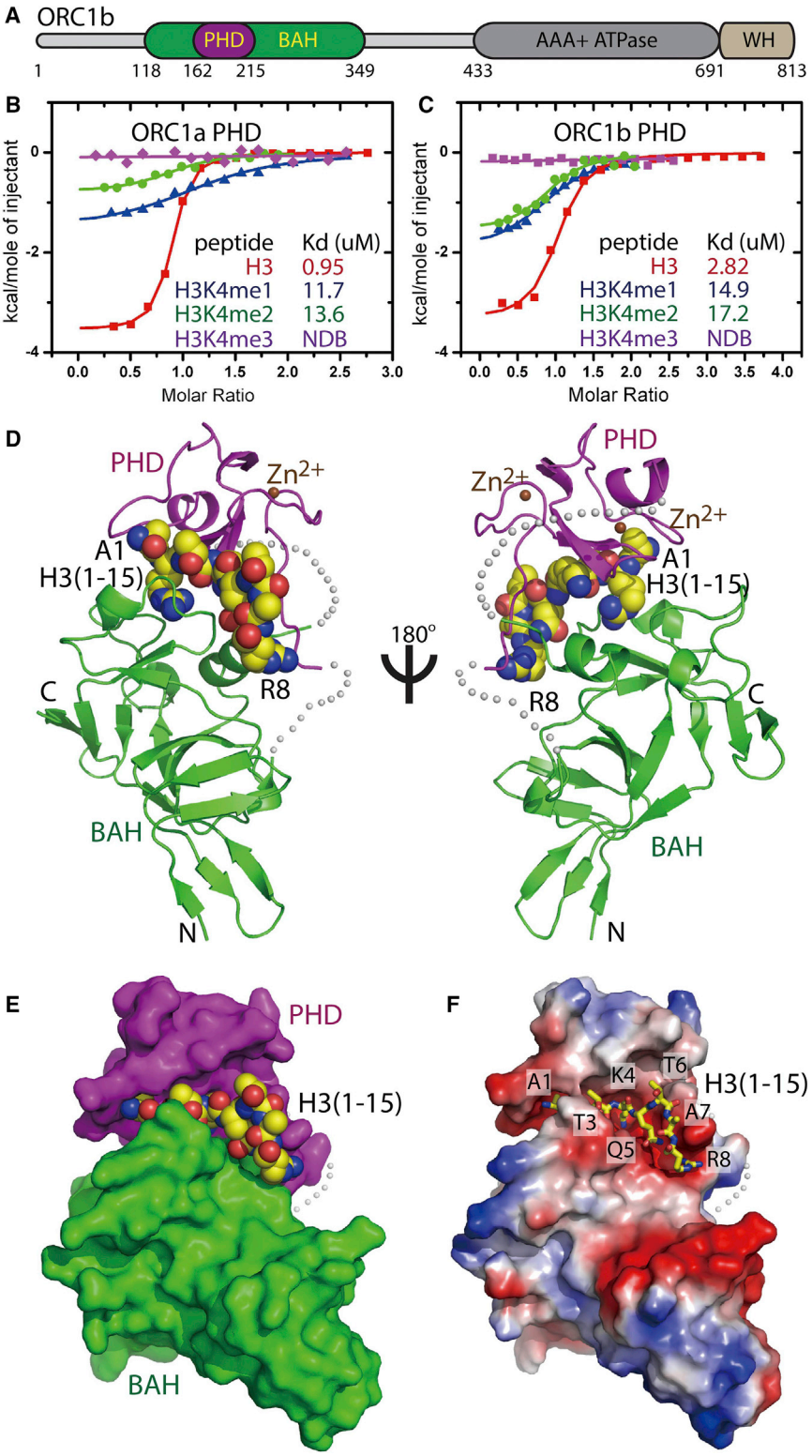


Figure 1. *Arabidopsis* ORC1 Protein Recognizes Unmodified H3 N-Terminal Tail and Overall Structure of ORC1b BAH-PHD Cassette-H3 Peptide Complex

(A) Schematic representation of the domain architecture of *Arabidopsis* ORC1b. WH, winged helix domain.

(B and C) ITC binding between the PHD fingers of *Arabidopsis* ORC1a (B), ORC1b (C), and various H3K4 methylated peptides establishing that the ORC1 PHD fingers prefer to recognize H3K4me0 peptide over methylated H3K4 peptides. NDB, no detectable binding.

(D) Ribbon representation of the ORC1b BAH-PHD cassette in complex with H3(1–15) peptide in two views related by a 180° rotation. The BAH and PHD domains are colored in green and magenta, respectively. The peptide is shown in a space-filling representation. The zinc ions are shown as orange balls. The linkers between BAH and PHD domains are disordered and are shown as silver dashed lines.

(E) Surface view of the BAH and PHD domains in green and magenta, respectively. The PHD finger buds out from one side of the BAH domain. The peptide is clamped in between the two domains.

(F) Electrostatic surface view of the BAH-PHD cassette. The peptide is shown in stick representation. The peptide binds in a negatively charged surface cleft of the BAH-PHD cassette. The N terminus inserts into a negatively charged groove of the protein.

See also Figures S1 and S2.

et al., 2011). *Arabidopsis* has two ORC1 homologs, ORC1a and ORC1b, which share 87% sequence identity throughout the entire sequence and 88% sequence identity within the BAH-PHD region, suggesting a similar function.

Both BAH and PHD domains function as histone mark readers in many other proteins (Musselman et al., 2012; Patel and Wang, 2013; Taverna et al., 2007), raising the possibility that the plant ORC1 BAH-PHD cassette can recognize multiple histone marks. Here, we report on biochemical and structural studies on the *Arabidopsis* ORC1 BAH-PHD cassette. We show that unlike metazoan and yeast ORC1 BAH domains, the plant BAH-PHD cassette specifically recognizes the unmodified H3 tail, while strictly requiring the unmodified states of H3R2, H3T3, and H3K4. The crystal structure of *Arabidopsis* ORC1 BAH-PHD cassette in complex with an unmodified H3(1–15) peptide reveals the structural basis for this specific recognition of the unmodified state of the H3 N-terminal tail. In addition, our structure shows that the BAH domain and PHD finger of ORC1 target the unmodified H3 tail in a multivalent mode involving both domains from opposing sides, rather than the classic linear combinatorial mode using each domain independently, thereby representing a novel multivalent readout mechanism.

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

Summary of Diffraction Data	
Crystal	ORC1b BAH-PHD + H3(1–15)
Beamline	SSRF-BL17U1
Wavelength (Å)	1.2827
Space group	$P4_3$
Cell parameters	
$a = b$ (Å)	64.4
c (Å)	79.6
Resolution (Å)	50.0–1.9 (1.97–1.90)[a]
Wilson B factor (Å^2)	27.3
R_{merge} (%)	9.2 (49.1)
Observed reflections	184,019
Unique reflections	25,599
Redundancy	7.2 (7.4)
Average $I/\sigma(I)$	46.3 (5.1)
Completeness (%)	99.8 (99.6)
Refinement and Structure Model	
R/free R (%)	19.4/20.5
No. of atoms	1,874
Protein/peptide	1,614/69
Zn^{2+}	2
Water	189
Average B factor (Å^2)	40.7
Protein/peptide	40.8/39.1
Zn^{2+}	61.2
Water	40.4
RMSDs	
Bond lengths (Å)	0.003
Bond angles (°)	0.858
Ramachandran plot	
Favored	93.8
Allowed	6.2
Generously allowed	0
Disallowed	0
MolProbity clashscore	0.91

[a]Values in parentheses represent highest-resolution shell.

RESULTS

Arabidopsis thaliana ORC1 Recognizes Unmodified H3 Peptide

Our previous genomic data suggested that genome-wide *Arabidopsis* origins of DNA replication are enriched with the histone marks H3K4me3, H3K4me2, and H4K5ac, while being depleted in H3K4me1 and H3K9me2 (Costas et al., 2011). To investigate the molecular basis underlying the linkage of plant DNA replication to histone mark recognition, we focused on the *Arabidopsis* ORC1, since it has a fused BAH-PHD cassette at its N terminus containing two known structural modules that recognize histone marks (Figures 1A and S1). Although it was reported previously that the isolated *Arabidopsis* ORC1b PHD finger alone can recognize the H3K4me3 mark by in vitro pull-down assay (de la Paz Sanchez and Gutierrez, 2009), our isothermal titration calorimetry (ITC) data revealed that both *Arabidopsis* ORC1a and ORC1b PHD fingers exhibit a strong preference for unmodified H3 peptide over its methylated H3K4 peptide counterparts (Figures 1B and 1C).

Overall Structure of ORC1b BAH-PHD Cassette in Complex with H3(1–15) Peptide

To further investigate the molecular mechanism of recognition of the unmodified H3 N-terminal tail by ORC1, we carried out structural and biochemical studies on the complex. Although the PHD fingers of ORC1a and ORC1b did not yield crystal either alone or in complex with unmodified H3(1–15) peptides, we successfully obtained diffraction-quality crystals of *Arabidopsis* ORC1b BAH-PHD cassette in complex with an H3(1–15) peptide. The structure was solved using the single-wavelength anomalous dispersion (SAD) method with zinc anomalous signal and refined to 1.9 Å resolution, yielding an R factor of 19.4% and a free R factor of 20.5% (Table 1 and Figure 1D). In the sequence, the PHD finger is embedded within the BAH domain (Figure 1A), while structurally it buds out from one side of the BAH domain (Figures 1D and 1E). The BAH and PHD domains exhibit well-defined electron density. The two linker regions on both sides of PHD finger that are connected to the BAH domain (residues 156–161 and 216–234) are disordered and were not built into the final model (Figure 1D). Importantly, the peptide is sandwiched in between the BAH and PHD domains and interacts with both these domains (Figures 1D and 1E). The peptide can be fully traced from Ala1 to Arg8, while only the main chain was visible for Lys9, and thus this residue was built as an alanine (Figure S2A). The N-terminal six residues of the peptide adopt an extended conformation and insert into a negatively charged binding groove between the BAH and PHD domains (Figures 1D and 1F). Notably, the peptide makes a nearly 90° sharp turn at Thr6 with the remaining peptide affixed on the protein surface, again in a linear conformation (Figure 1D). The BAH domain has only limited salt-bridge interactions with the PHD domain as observed between BAH domain residues Arg239, Arg288, and Arg306 and PHD finger residues Glu167, Glu182, and Asp163, respectively (Figure S2B), suggestive of a plausible model whereby the H3 peptide mediates the interaction between the two domains and contributes to the stabilization of their relative alignments in the complex.

Interactions between H3(1–15) Peptide and ORC1b BAH-PHD Cassette

The N terminus of the H3(1–15) peptide is deeply buried in a narrow negatively charged binding groove formed between the BAH domain and the PHD finger (Figure 1F). The N-terminal amino protons of the Ala1 residue of the H3 peptide form two hydrogen bonds with the main-chain carbonyl of PHD domain residues Pro203 and Gly205, respectively (Figure 2A). The side-chain methyl group of Ala1 fits into a small hydrophobic pocket formed by Ile181, Pro203, and Trp207 (Figure 2A). The main-chain carbonyl group of Ala1 forms water-mediated hydrogen-bonding interactions with the side chain of BAH domain residues Ser320 and Asn321 (Figure 2A). The Ala1 deeply inserts into the bottom of the binding pocket with optimal structural and chemical shape complementarity.

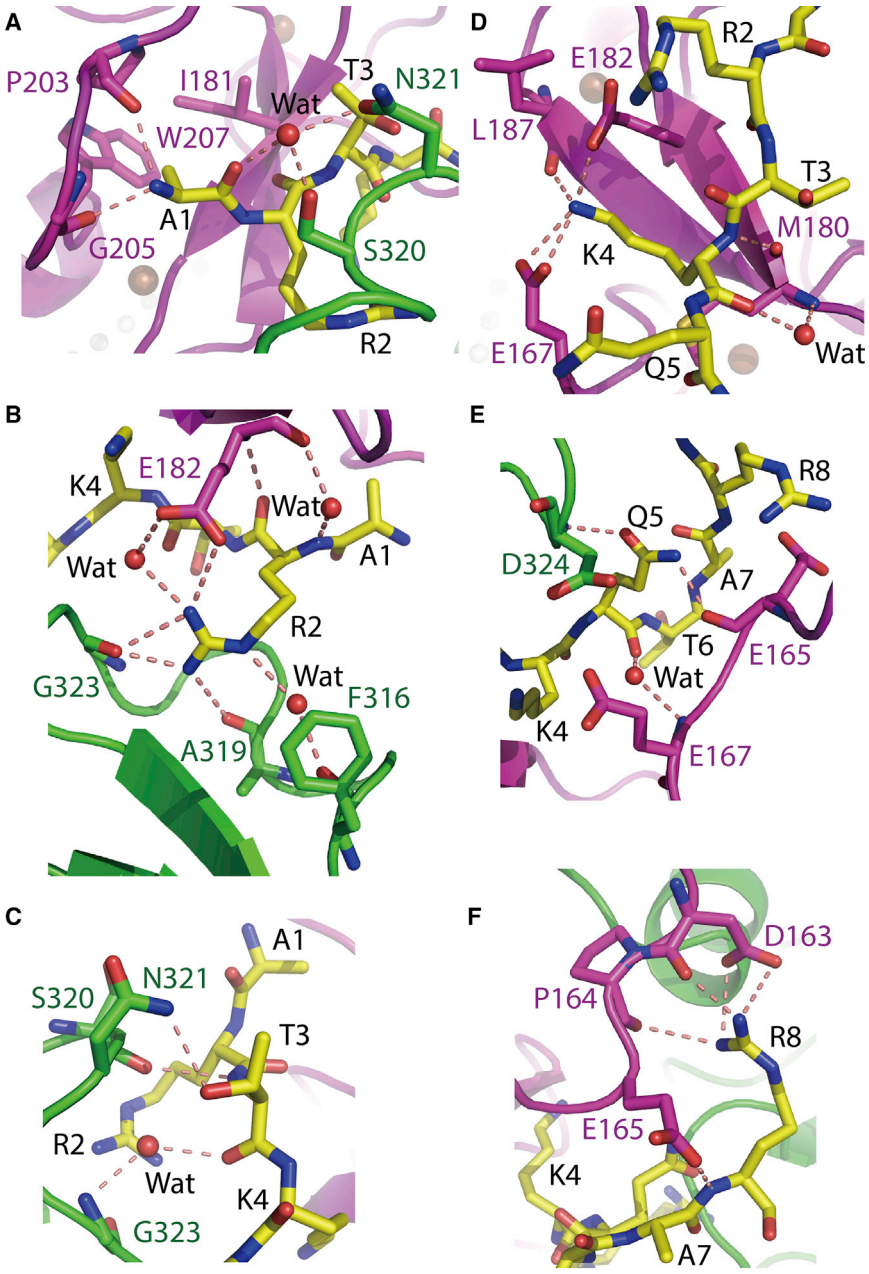


Figure 2. Molecular Basis for the Specific Recognition of the H3 Tail by the ORC1b BAH-PHD Cassette

Specific recognition of H3A1 (A), unmodified H3R2 (B), unmodified H3T3 (C), unmodified H3K4 (D), H3Q5 (E), and unmodified H3R8 (F). The BAH and PHD domains are colored in green and magenta, respectively. The peptide is shown in a stick representation. The hydrogen bonds are shown as dashed red lines. The binding highlights extensive hydrogen-bonding interactions, revealing the structural basis for the specific recognition of the unmodified H3 N-terminal tail.

See also Figure S3.

Arg2 of the H3 peptide is deeply buried between the BAH and PHD domains. The main chain of Arg2 forms a direct hydrogen bond as well as a water-mediated hydrogen bond with Glu182 of the PHD finger (Figure 2B). The side-chain guanidino protons of Arg2 are almost fully hydrogen bonded with surrounding residues, including PHD domain residue Glu182 and BAH domain residues Phe316, Ala319, and Gly323 (Figure 2B). The side chain of the Arg2 is stabilized by an extensive hydrogen-bonding network and almost all of the side-chain guanidino protons form hydrogen-bonding interactions with the protein (Figures 2B and S3), suggesting specific recognition of unmodified H3R2. The side-chain hydroxyl group of Thr3 of the H3 peptide forms a hydrogen bond with the side chain of Asn321 of the BAH domain (Figure 2C). The main-chain amide proton and the carbonyl group of Thr3 also forms a hydrogen bond and a water-mediated hydrogen bond with Ser320 and Gly323 of the BAH domain, respectively (Figure 2C). The whole side chain of Thr3 is clamped between the PHD and the BAH domains, indicating a recognition of unmodified H3T3. The Lys4 of the peptide inserts its side chain into a highly negatively charged pocket of the PHD finger, with its side-chain amino protons forming several hydrogen bonds with the side chains of Glu167 and Glu182 and the main-chain carbonyl of Leu187 (Figure 2D). The main chain of Lys4 also makes direct or water-mediated hydrogen-bonding interactions with Met180 of the PHD finger (Figure 2D). The side chain of Lys4 forms several hydrogen-bonding interactions with the PHD finger pocket, and this pocket is highly negatively charged without any aromatic or hydrophobic residues, indicating a preference for unmodified H3K4.

Gln5 of the H3 peptide forms several hydrogen bonds with the main chain of BAH domain residue Asp324, and PHD domain residues Glu165 and Glu167 (Figure 2E). The Thr6 and Ala7 of the H3 peptide are not involved in the binding with the protein. The guanidino protons of Arg8 form four hydrogen bonds with main chain and side chain of PHD domain residues Asp163 and Pro164 (Figure 2F), the main-chain amide proton also forms a hydrogen bond with the side chain of Glu165 (Figure 2F). Thus, the Arg8 is specifically anchored at this position, which probably marks the N terminus and Arg8 as the second anchoring site in addition to the N-terminal Ala1, to force the peptide to adopt a sharp turn at Thr6.

The Influence of Histone Modification on the Recognition by ORC1b

The N terminus of the H3 tail can be marked with various modifications, such as methylation of H3R2 and H3K4 and

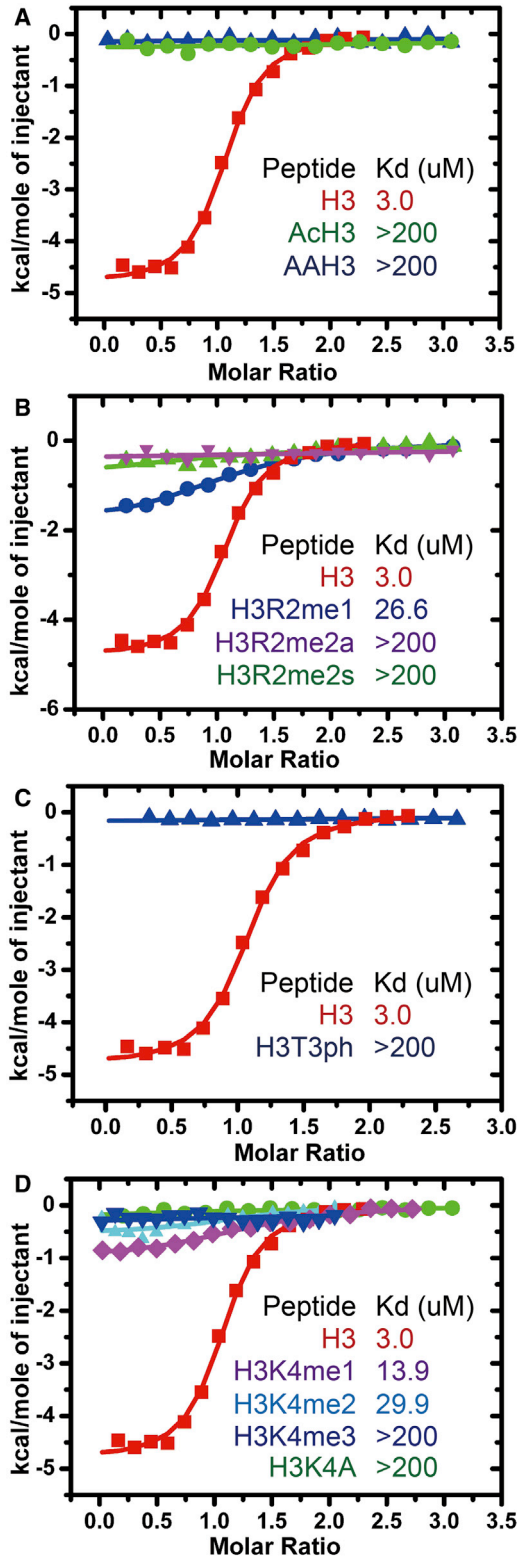


Figure 3. Comparison of Binding Affinity Measured by ITC between the ORC1b BAH-PHD Cassette and Unmodified H3 Tail and Various Modifications of A1, R2, T3, and K4 Positions on H3

(A) Comparison of binding affinities for H3 peptide versus peptide with N-terminal Ala extensions.

(B) Comparison of binding affinities for H3 peptide versus peptides with mono- and dimethylation at R2.

(C) Comparison of binding affinities for H3 peptide versus peptides with phosphorylation at T3.

(D) Comparison of binding affinities for H3 peptide versus peptides that are methylated at K4 or the K4A modification.

Most of the modifications dramatically decrease the binding affinity between the ORC1b BAH-PHD cassette and modified H3 peptide, revealing a strict requirement of unmodified state of the H3 N-terminal tail and the length of the N terminus of H3. See also Figure S4.

phosphorylation of H3T3. Our structural data indicate a specific recognition of the unmodified H3 tail. This led us to investigate the influence of various histone modifications on the recognition of the H3 tail by the ORC1b BAH-PHD cassette. The N-terminal Ala1 of H3 is anchored in an enclosed pocket between the BAH and PHD domains. A fusion of two extra alanine residues to the N terminus of the H3 peptide (AAH3), or just addition of a small acetyl group by acetylation of the N-terminal amino group (AcH3), significantly decreased the binding affinity between the peptide and ORC1b BAH-PHD cassette from 3.0 μM for the wild-type peptide to a non-detectable binding as shown by our ITC data (Figure 3A). This indicates a strict length requirement of the N terminus of H3 without any other additional residue or group. H3R2 can be methylated in three different states, namely monomethylation, symmetric dimethylation, and asymmetric dimethylation (Di Lorenzo and Bedford, 2011). Our ITC measurements showed that monomethylation of the H3R2 decreases the binding affinity by about 9-fold, while both symmetric and asymmetric dimethylation of the H3R2 dramatically decreases the binding to an undetectable level (Figure 3B). H3T3 can be phosphorylated, and the phosphorylation of Thr3 will result in steric clash with the BAH-PHD cassette, which was confirmed by our ITC data, whereby the H3T3ph peptide lost its binding affinity against the ORC1b BAH-PHD cassette (Figure 3C). H3K4 can be mono-, di-, or trimethylated. The mono- and dimethylation of the Lys4 residue reduces the binding by about 4- and 10-fold, respectively (Figure 3D), most likely due to the disruption of some of the hydrogen-bonding interactions between H3K4 and surrounding residues. The trimethylation of Lys4 or mutation of Lys4 to an alanine, which can eliminate all of the hydrogen bonds between Lys4 side chain and ORC1b BAH-PHD cassette, reduced the binding affinity to an undetectable level (Figure 3D). These data suggested that the *Arabidopsis* ORC1b BAH-PHD cassette strictly requires the unmodified state of the first four residues of the N terminus of the H3 tail.

DISCUSSION

Combinatorial Readout of Unmodified H3 by BAH and PHD Domains

It has been documented that multiple histone marks can be recognized by multiple domain proteins in a combinatorial way (Du and Patel, 2014; Wang and Patel, 2011). In this study, we observed that the PHD finger and BAH domain jointly contribute to the recognition of the unmodified H3 N-terminal peptide. Both the isolated PHD finger and the truncated BAH domain (the insertion PHD finger was replaced by a GSGSGSGS linker) show no significant binding to the unmodified H3(1–15) peptide under the same ITC conditions (150 mM NaCl, 2 mM

β-mercaptoethanol, and 20 mM HEPES [pH 7.5] at 6°C), which is in sharp contrast to the 3 μM binding affinity by the BAH-PHD cassette (Figure S4). (Note: We were able to observe binding between the PHD finger and unmodified H3 peptide by ITC under a low-salt condition of 50 mM NaCl, 2 mM β-mercaptoethanol, and 20 mM HEPES (pH 7.5) at 6°C. However, the BAH-PHD cassette is not stable under this low-salt condition, making it difficult to compare the binding affinities between BAH-PHD cassette and PHD finger.) Thus, the BAH domain and PHD finger are both required for multivalent readout of the H3 peptide.

A Unique Sandwiching Recognition Mode

In most previously reported examples of combinatorial readout of multiple histone marks by multiple reader domain proteins, each reader module reads a specific mark, using a linear readout arrangement (Du and Patel, 2014; Wang and Patel, 2011). In contrast, the ORC1b BAH-PHD cassette reported here exhibits a unique feature in that the PHD finger is fused inside the BAH domain in a sequence context, but buds out from one side of the BAH domain in a spatial context (Figure 1A). The two domains are not linearly positioned along the histone tail as in other cases of combinatorial readout (Du and Patel, 2014; Wang and Patel, 2011). Instead, the two domains clamp the target histone tail from opposing directions, representing a novel multivalent readout mode (Figure S3). The PHD finger is involved in the recognition of H3A1, H3K4, and H3R8, while the BAH domain is involved in the recognition of H3T3. More interestingly, H3R2 and H3Q5 are recognized to the same extent by both the PHD and BAH domains. Therefore, the recognition of residues along the histone H3 peptide represents a mixed rather than a linear distribution of reader domains, and in the current case a single mark needs two domains working together for the recognition, suggesting a novel multivalent readout mode.

Comparison of the Structures of ORC1 BAH Domains from Different Species

Currently the structures of ORC1 BAH domains from three species (*Arabidopsis*, yeast, and mouse) are available for comparison (Hou et al., 2005; Hsu et al., 2005; Kuo et al., 2012; Zhang et al., 2002) (Figure 4). The consensus segment among the three structures is the common BAH domain, which has different extended regions and diversified functions. *Arabidopsis* ORC1 BAH domain has an insertion of a PHD finger, which together with the BAH domain can specifically recognize the unmodified H3 N-terminal tail (Figure 4A). The yeast Orc1p BAH domain has three additional parts, the N-terminal extension, the C-terminal extension, and the inserted H domain, with the H domain mediating the interaction with Sir1p (Hou et al., 2005; Hsu et al., 2005; Zhang et al., 2002) (Figure 4B). The mouse ORC1 BAH domain has no extensions and can use a four-residue aromatic cage to recognize the methyllysine of H4K20me2 (Kuo et al., 2012) (Figure 4C). The superposition of mouse ORC1 BAH domain with *Arabidopsis* ORC1 BAH-PHD cassette shows similar BAH domain topology with a root-mean-square deviation (RMSD) of 2.2 Å for 114 aligned Cα atoms, while the two bound peptides have different orientations (Figure 4D). The four-residue aromatic cage of mouse ORC1 BAH domain is partially conserved in the *Arabidopsis* ORC1 BAH domain. Two of the aromatic residues of Trp248 and Trp270 are positioned in the same region (Figure 4E). However, the putative aromatic cage of *Arabidopsis* ORC1b BAH domain is pre-occupied by the Arg152 of the same molecule (Figure 4E). Arg152 is located within the loop connecting the BAH domain to the N terminus of the PHD finger, which is of a flexible nature. Further studies may be required for identifying additional binding partners, if any, for the BAH domain of *Arabidopsis* ORC1.

Functional Insight into the Epigenetic Regulation of Plant ORC1

Arabidopsis has two ORC1 homologs, ORC1a and ORC1b, which share a very similar sequence (Figure S5), indicating that ORC1a most likely has the same binding specificity as ORC1b. Although the ORC1a BAH-PHD cassette behaves poorly in ITC, our ITC data of ORC1a PHD finger revealed its preference for unmodified H3 peptide over H3K4me peptides (Figure 1B). The specific recognition of unmodified H3 N-terminal tail by *Arabidopsis* ORC1 proteins is in contrast with the observation that replication origin sites are enriched in H3K4me3 and H3K4me2 sites (Costas et al., 2011). A plausible explanation for this disagreement between biochemical and genomic data may be that the ORC1 BAH-PHD cassette here only serves as an anchoring site to stabilize the complex on chromatin. The targeting of ORC to certain chromatin loci could depend on other unknown factors. Indeed, the targeting of ORC has multiple mechanisms across different species. In the single-cell budding yeast, replication origin sites are determined strictly in a DNA sequence-specific manner (Bell and Stillman, 1992; Wyrick et al., 2001), while in fission yeast the replication origin sites have diversified and lack a consensus sequence (Antequera, 2004; Hayashi et al., 2007). In higher multicellular organisms, many additional factors are involved and the targeting of ORC is likely complex. In plants, in addition to H3K4me2 and H3K4me3, other factors such as G + C sequence enrichment, histone H2A.Z, and H4K5ac were also associated with *Arabidopsis* replication origin sites (Costas et al., 2011), indicating that the targeting of ORC is complex and involves multiple factors. This mode of regulation may have multiple epigenetic regulators such as an unknown H3K4me3 reader, H4K5ac reader, and so forth, and like other biological processes may not rely on a single factor. Thus, it seems likely that the ORC complex may be captured by unmodified H3 through its BAH-PHD cassette, while other factors may target the complex to H3K4me3-enriched or H4K5ac-enriched regions. The tethering of the ORC complex to the unmodified H3 tail may also function to stabilize ORC chromatin associations. Finally, the *Arabidopsis* ORC1-H3 interaction may not function in origin recognition but rather in a later step during replication regulation.

EXPERIMENTAL PROCEDURES

Protein Preparation

Constructs of ORC1 were cloned into pGEX-6p-1 vector (GE Healthcare) and expressed in *Escherichia coli*. The proteins were purified using a GSTrap FF column, a Q FF column, and a Superdex G200 column (GE Healthcare). The peptides were ordered from the Tufts University peptide synthesis facility

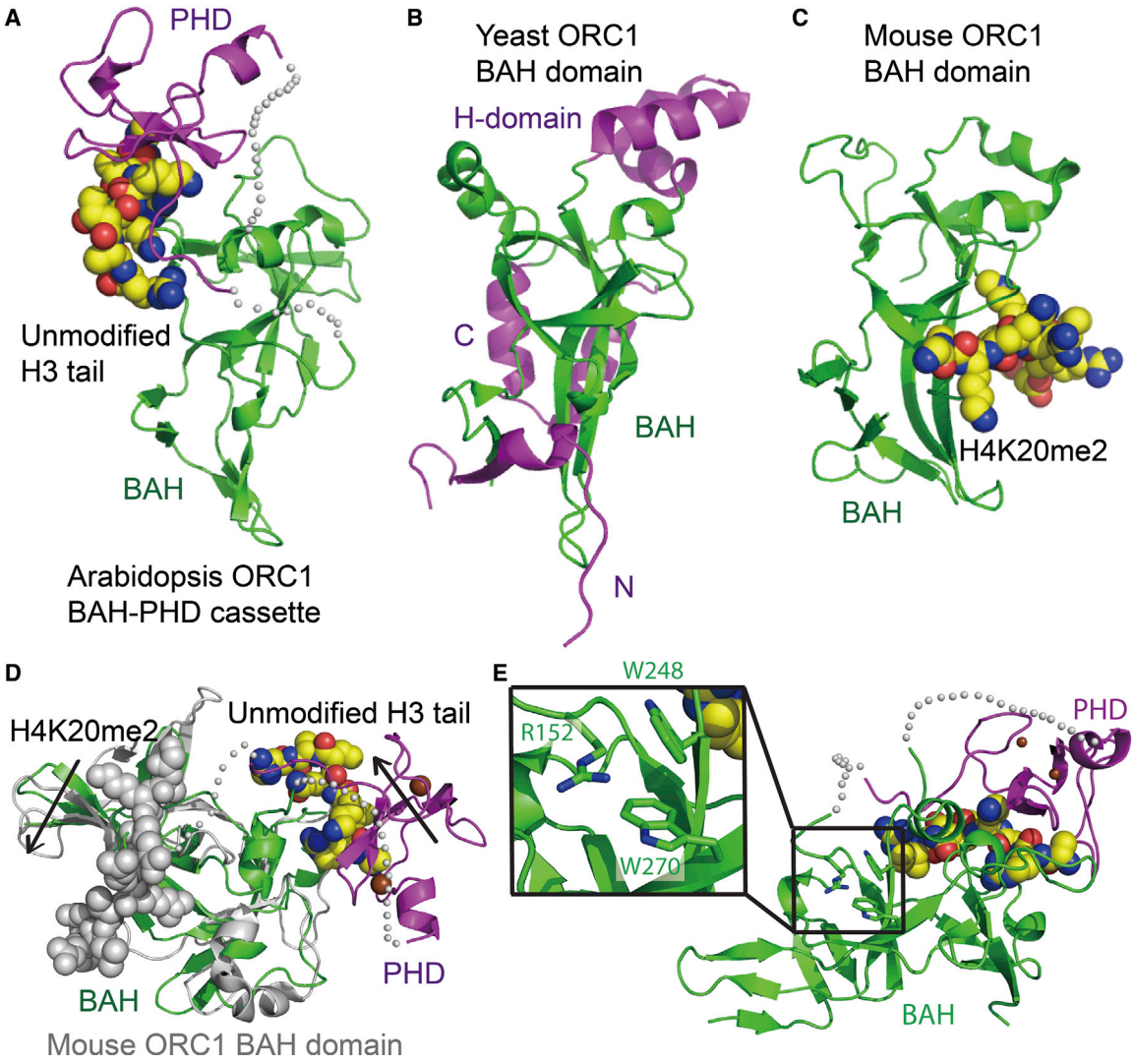


Figure 4. Comparison of Structures of Known ORC1 BAH Domains from Different Species

(A) Ribbon diagram of the crystal structure *Arabidopsis* ORC1b BAH-PHD cassette in complex with an unmodified H3 peptide. The structure is colored the same as Figure 1D. The PHD domain buds out from one side of the BAH domain, and the peptide is clamped in between the BAH and PHD domains.

(B) The crystal structure of yeast ORC1 BAH domain (PDB: 1M4Z). It lacks the PHD domain but has an additional N-terminal, C-terminal, and internal inserted H domain extensions, colored in magenta; the core BAH domain is colored in green. The H domain can mediate the interaction with Sir1p and regulates ORC1 function.

(C) The crystal structure of mouse ORC1 BAH domain in complex with an H4K20me2 peptide (PDB: 4DOW). Mouse BAH domain is a BAH-only domain without other extended segments. The bound peptide is shown in space-filling representation. Note that the BAH domain in (A), (B), and (C) is aligned in the same orientation.

(D) The superposition of the structures of BAH domains of *Arabidopsis* ORC BAH-PHD cassette in complex with H3(1–15) peptide (in colored scheme) and mouse ORC1 BAH domain in complex with an H4(14–25)K20me2 peptide (in silver). Both peptides are shown in space-filling representation. Although the BAH domains have similar conformation, the two peptides have different binding sites and orientations.

(E) The BAH domain of *Arabidopsis* ORC1 has a partially conserved aromatic cage at the same position as mouse BAH domain aromatic cage for accommodating the methyllysine of H4K20me2. An enlarged view of the potential aromatic cage of *Arabidopsis* ORC1b BAH domain is highlighted in the box and reveals an arginine residue occupying the aromatic cage in an autoinhibitory mode.

See also Figure S5.

and GL Biochem. Detailed information can be found in Supplemental Experimental Procedures.

Crystallization

Crystallization was conducted using the hanging-drop vapor diffusion method. A SAD dataset was collected at the beamline BL17U1 at Shanghai Synchrotron Radiation Facility (SSRF), and processed with the program HKL2000 (Otwinowski and Minor, 1997). The statistics of the diffraction data are summarized in Table 1. Detailed information can be found in Supplemental Experimental Procedures.

Structure Determination and Refinement

The structure was solved using the SAD method with the program PHENIX (Adams et al., 2010). The model building was carried out using the program

Coot (Emsley et al., 2010). The statistics of the refinement and structure models are shown in Table 1. Detailed information can be found in Supplemental Experimental Procedures.

Isothermal Titration Calorimetry

ITC binding experiments were carried out on a Microcal calorimeter ITC 200 instrument. Detailed information can be found in Supplemental Experimental Procedures.

ACCESSION NUMBERS

Coordinates and structure factors for *Arabidopsis* ORC1b BAH-PHD cassette in complex with unmodified H3(1–15) peptide have been deposited in the PDB with the accession code PDB: 5HH7.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures and five figures and can be found with this article online at http://dx.doi.org/10.1016/j.str.2016.01.004.

AUTHOR CONTRIBUTIONS

S.L., Z.Y., X.D., R.L., A.W.W., and J.D. carried out the experiments and data analysis. O.G., S.E.J., D.J.P., and J.D. conceived and initiated the project and wrote the paper.

ACKNOWLEDGMENTS

We are grateful to the staff members at beamline BL17U1 at Shanghai Synchrotron Radiation Facility for their support in diffraction data collection, and Dr. Peng Zhang for ITC experiments. This work was supported by the Thousand Young Talent Program of China and the Chinese Academy of Sciences to J.D., the LLSCOR Program Project and STARR Foundation grants to D.J.P, and NIH grants GM060398 to S.E.J. and GM079641 to O.G. S.E.J. is an Investigator of the Howard Hughes Medical Institute. D.J.P. acknowledges support by the Memorial Sloan-Kettering Cancer Center Support Grant/Core Grant (P30 CA008748). O.G. is a co-founder of EpiCypher, Inc.

Received: October 28, 2015
Revised: December 28, 2015
Accepted: January 9, 2016
Published: February 11, 2016

REFERENCES

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. *66*, 213–221.

Aladjem, M.I. (2007). Replication in context: dynamic regulation of DNA replication patterns in metazoans. Nat. Rev. Genet. *8*, 588–600.

Antequera, F. (2004). Genomic specification and epigenetic regulation of eukaryotic DNA replication origins. EMBO J. *23*, 4365–4370.

Bell, S.P., and Stillman, B. (1992). ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. Nature *357*, 128–134.

Bicknell, L.S., Bongers, E.M., Leitch, A., Brown, S., Schoots, J., Harley, M.E., Aftimos, S., Al-Aama, J.Y., Bober, M., Brown, P.A., et al. (2011a). Mutations in the pre-replication complex cause Meier-Gorlin syndrome. Nat. Genet. *43*, 356–359.

Bicknell, L.S., Walker, S., Klingseisen, A., Stiff, T., Leitch, A., Kerzendorfer, C., Martin, C.A., Yeyati, P., Al Sanna, N., Bober, M., et al. (2011b). Mutations in ORC1, encoding the largest subunit of the origin recognition complex, cause microcephalic primordial dwarfism resembling Meier-Gorlin syndrome. Nat. Genet. *43*, 350–355.

Bleichert, F., Botchan, M.R., and Berger, J.M. (2015). Crystal structure of the eukaryotic origin recognition complex. Nature *519*, 321–326.

Costas, C., de la Paz Sanchez, M., Stroud, H., Yu, Y., Oliveros, J.C., Feng, S., Benguria, A., Lopez-Vidriero, I., Zhang, X., Solano, R., et al. (2011). Genome-wide mapping of *Arabidopsis thaliana* origins of DNA replication and their associated epigenetic marks. Nat. Struct. Mol. Biol. *18*, 395–400.

de la Paz Sanchez, M., and Gutierrez, C. (2009). *Arabidopsis* ORC1 is a PHD-containing H3K4me3 effector that regulates transcription. Proc. Natl. Acad. Sci. USA *106*, 2065–2070.

DePamphilis, M.L., Blow, J.J., Ghosh, S., Saha, T., Noguchi, K., and Vassilev, A. (2006). Regulating the licensing of DNA replication origins in metazoa. Curr. Opin. Cell Biol. *18*, 231–239.

Di Lorenzo, A., and Bedford, M.T. (2011). Histone arginine methylation. FEBS Lett. *585*, 2024–2031.

Du, J., and Patel, D.J. (2014). Structural biology-based insights into combinatorial readout and crosstalk among epigenetic marks. Biochim. Biophys. Acta *1839*, 719–727.

Dueber, E.L., Corn, J.E., Bell, S.D., and Berger, J.M. (2007). Replication origin recognition and deformation by a heterodimeric archaeal Orc1 complex. Science *317*, 1210–1213.

Duncker, B.P., Chesnokov, I.N., and McConkey, B.J. (2009). The origin recognition complex protein family. Genome Biol. *10*, 214.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. *66*, 486–501.

Gaudier, M., Schuwirth, B.S., Westcott, S.L., and Wigley, D.B. (2007). Structural basis of DNA replication origin recognition by an ORC protein. Science *317*, 1213–1216.

Guernsey, D.L., Matsuoka, M., Jiang, H., Evans, S., Macgillivray, C., Nightingale, M., Perry, S., Ferguson, M., LeBlanc, M., Paquette, J., et al. (2011). Mutations in origin recognition complex gene ORC4 cause Meier-Gorlin syndrome. Nat. Genet. *43*, 360–364.

Hayashi, M., Katou, Y., Itoh, T., Tazumi, A., Yamada, Y., Takahashi, T., Nakagawa, T., Shirahige, K., and Masukata, H. (2007). Genome-wide localization of pre-RC sites and identification of replication origins in fission yeast. EMBO J. *26*, 1327–1339.

Hou, Z., Bernstein, D.A., Fox, C.A., and Keck, J.L. (2005). Structural basis of the Sir1-origin recognition complex interaction in transcriptional silencing. Proc. Natl. Acad. Sci. USA *102*, 8489–8494.

Hsu, H.C., Stillman, B., and Xu, R.M. (2005). Structural basis for origin recognition complex 1 protein-silence information regulator 1 protein interaction in epigenetic silencing. Proc. Natl. Acad. Sci. USA *102*, 8519–8524.

Kuo, A.J., Song, J., Cheung, P., Ishibe-Murakami, S., Yamazoe, S., Chen, J.K., Patel, D.J., and Gozani, O. (2012). The BAH domain of ORC1 links H4K20me2 to DNA replication licensing and Meier-Gorlin syndrome. Nature *484*, 115–119.

Leatherwood, J. (1998). Emerging mechanisms of eukaryotic DNA replication initiation. Curr. Opin. Cell Biol. *10*, 742–748.

Musselman, C.A., Lalonde, M.E., Cote, J., and Kutateladze, T.G. (2012). Perceiving the epigenetic landscape through histone readers. Nat. Struct. Mol. Biol. *19*, 1218–1227.

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

Patel, D.J., and Wang, Z. (2013). Readout of epigenetic modifications. Annu. Rev. Biochem. *82*, 81–118.

Taverna, S.D., Li, H., Ruthenburg, A.J., Allis, C.D., and Patel, D.J. (2007). How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. Nat. Struct. Mol. Biol. *14*, 1025–1040.

Triolo, T., and Sternglanz, R. (1996). Role of interactions between the origin recognition complex and SIR1 in transcriptional silencing. Nature *381*, 251–253.

Wang, Z., and Patel, D.J. (2011). Combinatorial readout of dual histone modifications by paired chromatin-associated modules. J. Biol. Chem. *286*, 18363–18368.

Wyrick, J.J., Aparicio, J.G., Chen, T., Barnett, J.D., Jennings, E.G., Young, R.A., Bell, S.P., and Aparicio, O.M. (2001). Genome-wide distribution of ORC and MCM proteins in *S. cerevisiae*: high-resolution mapping of replication origins. Science *294*, 2357–2360.

Zhang, Z., Hayashi, M.K., Merkel, O., Stillman, B., and Xu, R.M. (2002). Structure and function of the BAH-containing domain of Orc1p in epigenetic silencing. EMBO J. *21*, 4600–4611.

Zhang, W., Sankaran, S., Gozani, O., and Song, J. (2015). A Meier-Gorlin syndrome mutation impairs the ORC1-nucleosome association. ACS Chem. Biol. *10*, 1176–1180.

Structure, Volume *24*

Supplemental Information

Structural Basis for the Unique Multivalent Readout of Unmodified H3 Tail by *Arabidopsis* ORC1b BAH-PHD Cassette

Sisi Li, Zhenlin Yang, Xuan Du, Rui Liu, Alex W. Wilkinson, Or Gozani, Steven E. Jacobsen, Dinshaw J. Patel, and Jiamu Du

Supplemental Information for:

Structural basis for the unique multivalent readout of unmodified H3 tail by Arabidopsis ORC1b BAH-PHD cassette

Sisi Li[1], Zhenlin Yang[2,5], Xuan Du[2,5], Rui Liu[2], Alex W. Wilkinson[3], Or Gozani[3], Steven E. Jacobsen[4], Dinshaw J. Patel[1] and Jiamu Du[2,*]

[1]Structural Biology Program, Memorial Sloan-Kettering Cancer Center, New York, NY 10065, USA; [2]Shanghai Center for Plant Stress Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 201602, China; [3]Department of Biology, Stanford University, Stanford, CA 94305, USA; [4]Howard Hughes Medical Institute and Department of Molecular, Cell, and Developmental Biology, University of California at Los Angeles, Los Angeles, CA 90095, USA; [5]University of Chinese Academy of Sciences, Beijing 100049, China.

[*] Correspondence: jmdu@sibs.ac.cn (Jiamu Du)

Inventory for Supplemental Items:

Supplemental Experimental Procedures

Protein preparation

Various constructs of *Arabidopsis thaliana* ORC1a and ORC1b were cloned into pGEX-6p-1 vector (GE Healthcare), which fuses a GST tag and a PreScission cleavage site to the N-terminus of the target protein. The plasmid was transformed into *E. coli* strain BL21 (DE3) RIL (Stratagene). The cells were cultured at 37 °C to OD_{600} reached 0.7, then the media was cooled to 16 °C and 0.2 mM IPTG was added to induce the protein expression overnight. The recombinant expressed protein was firstly purified using a GSTrap FF column (GE Healthcare). The GST tag was cleavage by PreScission protease. The protein was further purified by a Q FF column (GE Healthcare) and a Hiload Superdex G200 16/60 column (GE Healthcare). The isolated BAH domain of ORC1b was constructed by linking replacing residues 156-234 by a $(GS)_4$ linker. The protein was expressed, and purified with same protocol. The peptides were ordered from the Tufts University peptide synthesis facility and GL Biochem (Shanghai) Ltd..

Crystallization

Before crystallization, purified ORC1b BAH-PHD cassette (residues 118-349) protein was concentrated to 15 mg/ml and mixed with unmodified H3(1-15) peptide with a molar ratio of 1 : 3 at 4 °C for 30 minutes. Crystallization was conducted at 20 °C using the hanging drop vapor diffusion method in a condition of 0.2 M Na-formate and 20% PEG3350. The crystals were soaked into the reservoir solution supplement with 15% glycerol, and then mounted on a nylon loop followed by flash cooling into liquid

nitrogen. A SAD data set was collected at the zinc peak wavelength (1.2827 Å) at the beamline BL17U1 at Shanghai Synchrotron Radiation Facility (SSRF), and processed with the program HKL2000(Otwinowski and Minor, 1997). The statistics of the diffraction data are summarized in **Table 1**.

Structure determination and refinement

The structure of ORC1b BAH-PHD cassette in complex with an unmodified H3(1-15) peptide was solved using single-wavelength anomalous dispersion method as implemented in the program Phenix(Adams et al., 2010). The model building was carried out using the program Coot (Emsley et al., 2010) and structural refinement using the program Phenix(Adams et al., 2010). Throughout the refinement, a free *R* factor was calculated using 5% random chosen reflections. The stereochemistry of the structural models was analyzed using the program Procheck(Laskowski et al., 1993). The statistics of the refinement and structure models are shown in **Table 1**. All the molecular graphics were generated with the program Pymol (DeLano Scientific LLC).

Isothermal Titration Calorimetry

Isothermal titration calorimetry (ITC) binding experiments were carried out on a Microcal calorimeter ITC 200 instrument at 6 °C. For the ORC1a and ORC1b PHD fingers alone, the protein samples were dialyzed against a buffer of 50 mM NaCl, 2 mM β-mercaptoethanol, and 20 mM HEPES, pH 7.5 at 4 °C. Then, the protein samples were diluted and the lyophilized peptides were dissolved with the same buffer. Because the

BAH-PHD cassette is not stable in low salt condition, the ITC experiments of the BAH-PHD cassette, the BAH domain and the PHD domain as reference were carried out under a buffer condition of 150 mM NaCl, 2 mM β-mercaptoethanol, and 20 mM HEPES, pH 7.5. The titration was according to standard protocol and the data were fit using the program Origin 7.0 with a 1:1 binding model.

Supplemental References

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., *et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta crystallographica. Section D, Biological crystallography *66,* 213-221.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta crystallographica. Section D, Biological crystallography *66,* 486-501.

Laskowski, R.A., Macarthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr *26,* 283-291.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. *276,* 307-326.

Supplemental Figures

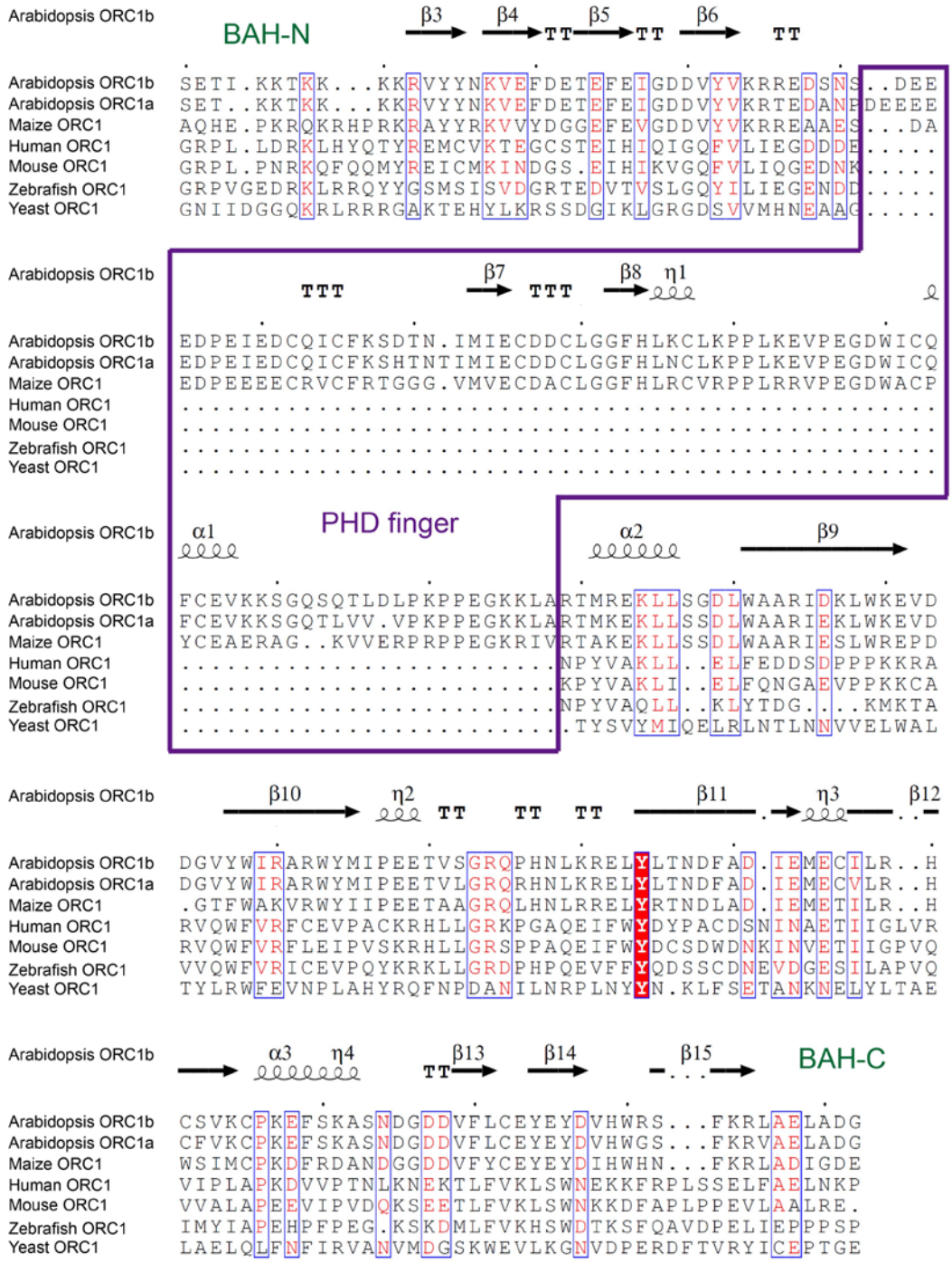


Figure S1. Related to Figure 1. Structure Based Sequence Alignment of ORC1 N-terminal Fragments from Different Species Indicates the Plant Specific Insertion of PHD Finger (boxed in magenta) into the BAH Domain (highlighted in green). The sequence used for alignment are: Arabidopsis ORC1b (Genbank: AEE83158), Arabidopsis ORC1a (Genbank: AEE83478.1), Maize ORC1 (Genbank: NP_001105070.1), Human ORC1 (Genbank: NP_001177747.1), Mouse ORC1 (Genbank: NP_035145.2), Zebrafish ORC1 (Genbank: NP_956227.1), and Yeast ORC1 (Genbank: NP_013646.1).

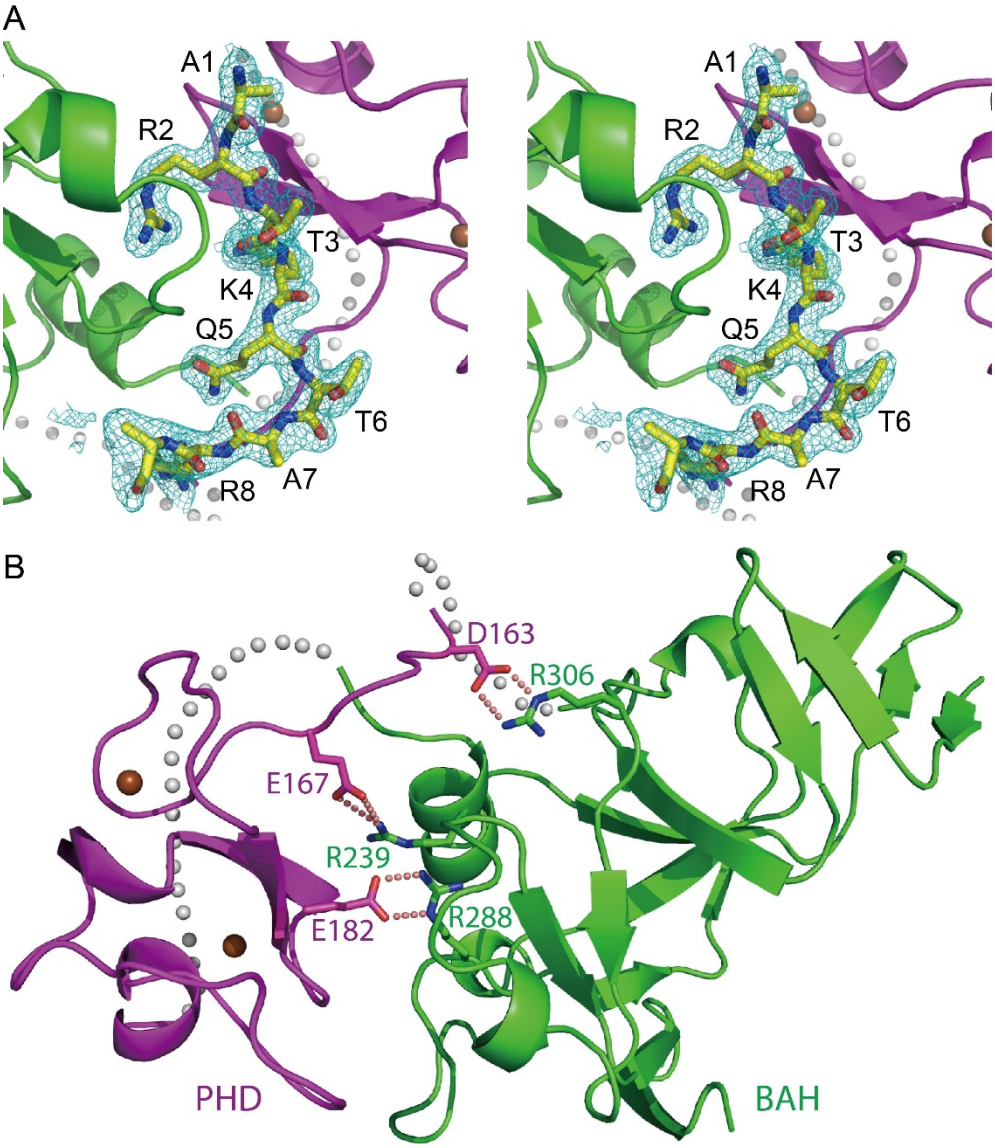


Figure S2. Related to Figure 1. Structure of ORC1b BAH-PHD Cassette in Complex with H3 Peptide.

(A) A stereo view of the SIGMAA weighted 2Fo-Fc electron density map (1 σ level) of the peptide. The peptide can be fully traced from Ala1 to Arg8. The Lys9 can only be traced for the main chain, therefore it was built as a Ala model.

(B) The inter-domain interaction between BAH and PHD domains. There are three pairs of salt bridge interactions between the negatively charged PHD domain residues Asp163, Glu167, Glu182 and the positively charged BAH domain residues Arg306, Arg239, and Arg288, respectively. The interaction features hydrophilic nature.

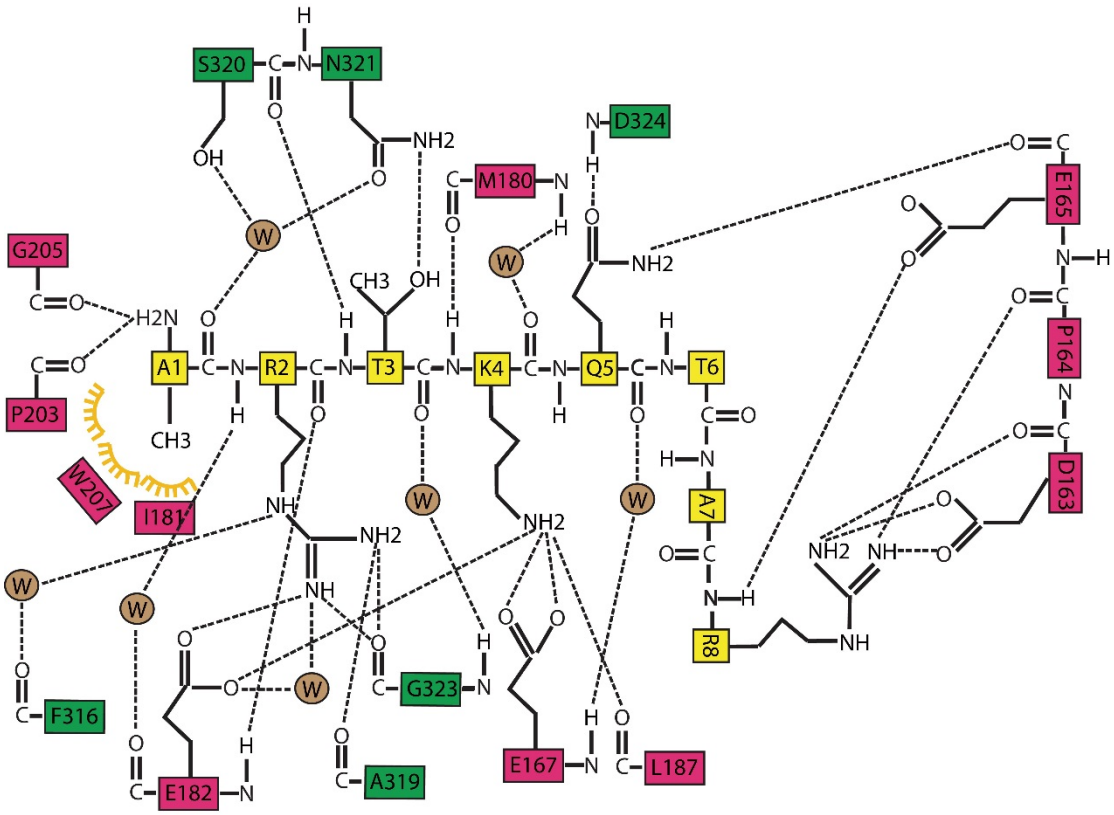


Figure S3. Related to Figure 2. A Schematic Cartoon Representation of the Interaction between ORC1b BAH-PHD Cassette and the Unmodified H3 N-terminal Tail.

The BAH domain, PHD finger, and histone peptide residues are colored in green, magenta, and yellow, respectively. The figure highlights the hydrophilic interactions between the protein and peptide which contain a plenty of salt bridges and hydrogen bonding interations. It worth to note that all the amide protons and carbonyl group of H3A1 to H3K4 are full hydrogen bonded, indicating a strcictly requirment of unmodified state of H3 tail.

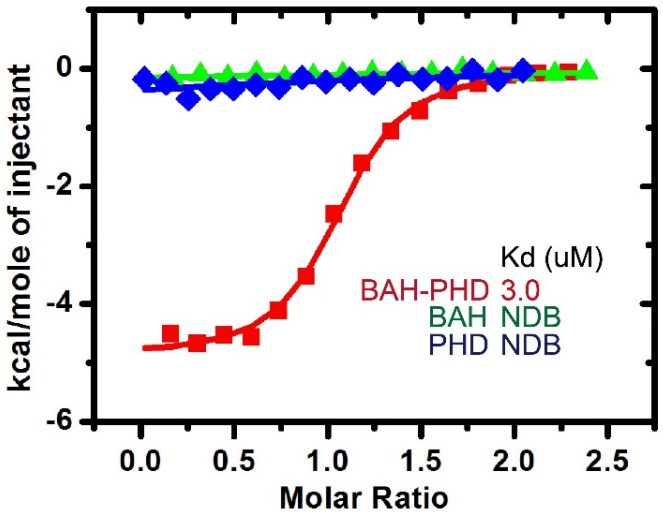
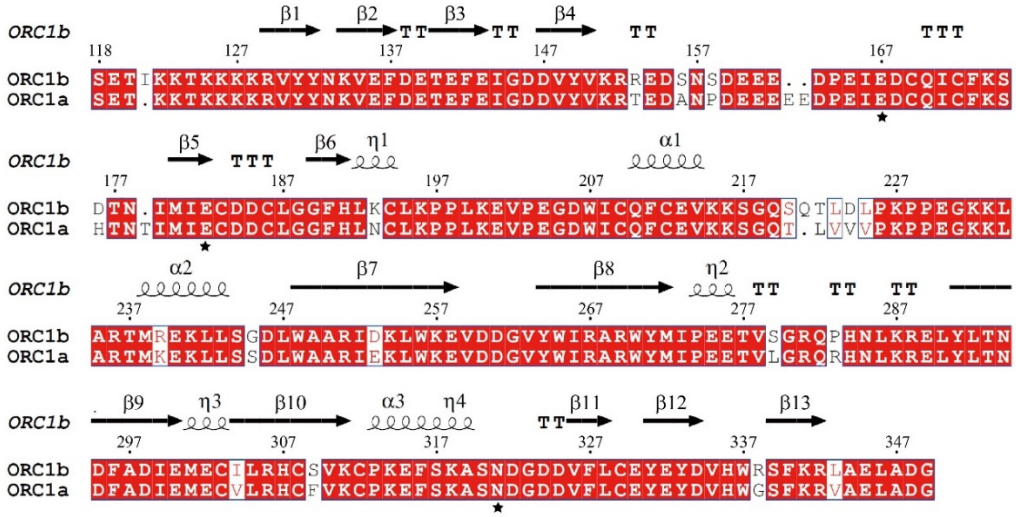


Figure S4. Related to Figure 3. The ITC Binding between Unmodified H3 Peptide and ORC1b BAH Domain, PHD Finger, and BAH-PHD Cassette.
The binding data reveal that the binding depends on the co-existance of both BAH and PHD domains of ORC1b. NDB, no detectable binding.



★ Key resideus involves in the recognition of the unmodified H3R2, H3T3, and H3K4.

Figure S5. Related to Figure 4. A Structure Based Seuqence Alignment of Arabidopsis ORC1b and ORC1a.

The two sequence show a very high sequence identity, indicating a similar structure. The key residus involved in the requirment of unmodified state of H3 tail are marked with star and are strictly conserved, indicating ORC1a has similar histone binding properties as ORC1b.