Title

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

Journal

Structure, 24(3)

ISSN

1359-0278

Authors

Li, Sisi
Yang, Zhenlin
Du, Xuan
et al.

Publication Date

2016-03-01

DOI

10.1016/j.str.2016.01.004

Peer reviewed

HHS Public Access

Author manuscript

Structure. Author manuscript; available in PMC 2017 March 01.

Published in final edited form as:

Structure. 2016 March 1; 24(3): 486–494. doi:10.1016/j.str.2016.01.004.

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

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.

Graphical Abstract

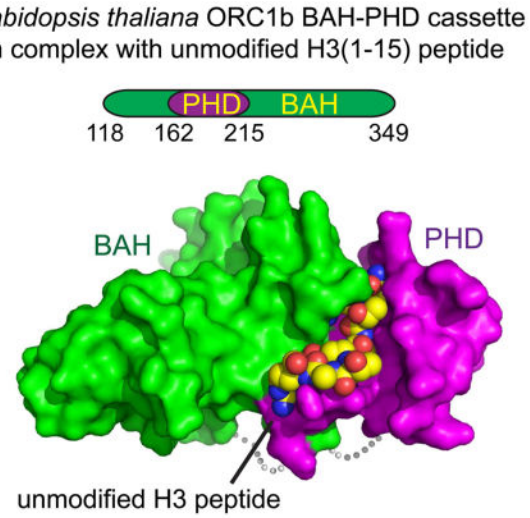
*Correspondence: jmdu@sibs.ac.cn (Jiamu Du).

Accession Codes

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 5HH7.

Author contributions

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



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 (MCM) helicase, which together initiate replication (Bell and Stillman, 1992; Leatherwood, 1998). The ORC complex is composed of an assembly of 6 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; Bicknell et al., 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 an MGS patient mutation that does not impact 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 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 (Figure 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 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 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.

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 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 (Figure 1A and S1). Although it was reported previously that the isolated Arabidopsis ORC1b PHD finger alone can recognize the H3K4me3 mark by *in vitro* pulldown assay (de la Paz Sanchez and Gutierrez, 2009), our 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 (Figure 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 got 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 (Figure 1D, E). 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 (Figure 1D, E). 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 (Figure 1D, F). 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 forms 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.

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 the side chain guanidino protons form hydrogen bonding interactions with the protein (Figure 2B

and S3), suggesting specific recognition of unmodified H3R2. The side chain hydroxyl group of Thr3 of the H3 peptide form 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 form 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 interaction with the PHD finger pocket and this pocket is highly negatively charged without any aromatic or hydrophobic residues, indicative of 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 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 tri-methylated. 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 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 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 could 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 hard 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 like 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 amongst the three structures is the common BAH domain, which have 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 it 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 an RMSD of 2.2 Å for 114 aligned Cα atoms, while the two bound peptides have different orientations (Figure 4D). The four residues 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 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 data and genomic data may be 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 on, and like other biological processes not rely on a single factor. Thus, it seems likely that the ORC complex may be captured by unmodified H3 though 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 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 *E. coli*. The proteins were purified using GSTrap FF column, Q FF column, and Superdex G200 column (GE Healthcare). The peptides were ordered from the Tufts University peptide synthesis facility and GL Biochem (Shanghai) Ltd. Detailed information can be found in Extended Experimental Procedures.

Crystallization

Crystallization was conducted using the hanging drop vapor diffusion method. A SAD data set 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 Extended Experimental Procedures.

Structure determination and refinement

The structure was solved using 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 Extended Experimental Procedures.

Isothermal Titration Calorimetry

Isothermal titration calorimetry (ITC) binding experiments were carried out on a Microcal calorimeter ITC 200 instrument. Detailed information can be found in Extended Experimental Procedures.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful to the staff members at beamline BL17U1 at Shanghai Synchrotron Radiation Facility (SSRF) for their support in diffraction data collection and Dr. Peng Zhang for ITC experiment. 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 GM60398 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).

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 crystallographica Section D, Biological crystallography. 2010; 66:213–221. [PubMed: 20124702]

Aladjem MI. Replication in context: dynamic regulation of DNA replication patterns in metazoans. Nature reviews Genetics. 2007; 8:588–600.

Antequera F. Genomic specification and epigenetic regulation of eukaryotic DNA replication origins. EMBO J. 2004; 23:4365–4370. [PubMed: 15510221]

Bell SP, Stillman B. ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. Nature. 1992; 357:128–134. [PubMed: 1579162]

Bicknell LS, Bongers EM, Leitch A, Brown S, Schoots J, Harley ME, Aftimos S, Al-Aama JY, Bober M, Brown PA, et al. Mutations in the pre-replication complex cause Meier-Gorlin syndrome. Nat Genet. 2011a; 43:356–359. [PubMed: 21358632]

Bicknell LS, Walker S, Klingseisen A, Stiff T, Leitch A, Kerzendorfer C, Martin CA, Yeyati P, Al Sanna N, Bober M, et al. Mutations in ORC1, encoding the largest subunit of the origin recognition complex, cause microcephalic primordial dwarfism resembling Meier-Gorlin syndrome. Nat Genet. 2011b; 43:350–355. [PubMed: 21358633]

Bleichert F, Botchan MR, Berger JM. Crystal structure of the eukaryotic origin recognition complex. Nature. 2015; 519:321–326. [PubMed: 25762138]

Costas C, de la Paz Sanchez M, Stroud H, Yu Y, Oliveros JC, Feng S, Benguria A, Lopez-Vidriero I, Zhang X, Solano R, et al. Genome-wide mapping of Arabidopsis thaliana origins of DNA replication and their associated epigenetic marks. Nat Struct Mol Biol. 2011; 18:395–400. [PubMed: 21297636]

de la Paz Sanchez M, Gutierrez C. Arabidopsis ORC1 is a PHD-containing H3K4me3 effector that regulates transcription. Proceedings of the National Academy of Sciences of the United States of America. 2009; 106:2065–2070. [PubMed: 19171893]

DePamphilis ML, Blow JJ, Ghosh S, Saha T, Noguchi K, Vassilev A. Regulating the licensing of DNA replication origins in metazoa. Curr Opin Cell Biol. 2006; 18:231–239. [PubMed: 16650748]

Di Lorenzo A, Bedford MT. Histone arginine methylation. FEBS Lett. 2011; 585:2024–2031. [PubMed: 21074527]

Du J, Patel DJ. Structural biology-based insights into combinatorial readout and crosstalk among epigenetic marks. Biochim Biophys Acta. 2014; 1839:719–727. [PubMed: 24747177]

Dueber EL, Corn JE, Bell SD, Berger JM. Replication origin recognition and deformation by a heterodimeric archaeal Orc1 complex. Science. 2007; 317:1210–1213. [PubMed: 17761879]

Duncker BP, Chesnokov IN, McConkey BJ. The origin recognition complex protein family. Genome biology. 2009; 10:214. [PubMed: 19344485]

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

Gaudier M, Schuwirth BS, Westcott SL, Wigley DB. Structural basis of DNA replication origin recognition by an ORC protein. Science. 2007; 317:1213–1216. [PubMed: 17761880]

Guernsey DL, Matsuoka M, Jiang H, Evans S, Macgillivray C, Nightingale M, Perry S, Ferguson M, LeBlanc M, Paquette J, et al. Mutations in origin recognition complex gene ORC4 cause Meier-Gorlin syndrome. Nat Genet. 2011; 43:360–364. [PubMed: 21358631]

Hayashi M, Katou Y, Itoh T, Tazumi A, Yamada Y, Takahashi T, Nakagawa T, Shirahige K, Masukata H. Genome-wide localization of pre-RC sites and identification of replication origins in fission yeast. EMBO J. 2007; 26:1327–1339. [PubMed: 17304213]

Hou Z, Bernstein DA, Fox CA, Keck JL. Structural basis of the Sir1-origin recognition complex interaction in transcriptional silencing. Proceedings of the National Academy of Sciences of the United States of America. 2005; 102:8489–8494. [PubMed: 15932939]

Hsu HC, Stillman B, Xu RM. Structural basis for origin recognition complex 1 protein-silence information regulator 1 protein interaction in epigenetic silencing. Proceedings of the National Academy of Sciences of the United States of America. 2005; 102:8519–8524. [PubMed: 15937111]

Kuo AJ, Song J, Cheung P, Ishibe-Murakami S, Yamazoe S, Chen JK, Patel DJ, Gozani O. The BAH domain of ORC1 links H4K20me2 to DNA replication licensing and Meier-Gorlin syndrome. Nature. 2012; 484:115–119. [PubMed: 22398447]

Leatherwood J. Emerging mechanisms of eukaryotic DNA replication initiation. Curr Opin Cell Biol. 1998; 10:742–748. [PubMed: 9914182]

Musselman CA, Lalonde ME, Cote J, Kutateladze TG. Perceiving the epigenetic landscape through histone readers. Nat Struct Mol Biol. 2012; 19:1218–1227. [PubMed: 23211769]

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

Patel DJ, Wang Z. Readout of epigenetic modifications. Annu Rev Biochem. 2013; 82:81–118. [PubMed: 23642229]

Taverna SD, Li H, Ruthenburg AJ, Allis CD, Patel DJ. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. Nat Struct Mol Biol. 2007; 14:1025–1040. [PubMed: 17984965]

Triolo T, Sternglanz R. Role of interactions between the origin recognition complex and SIR1 in transcriptional silencing. Nature. 1996; 381:251–253. [PubMed: 8622770]

Wang Z, Patel DJ. Combinatorial readout of dual histone modifications by paired chromatin-associated modules. J Biol Chem. 2011; 286:18363–18368. [PubMed: 21454653]

Wyrick JJ, Aparicio JG, Chen T, Barnett JD, Jennings EG, Young RA, Bell SP, Aparicio OM. Genome-wide distribution of ORC and MCM proteins in S. cerevisiae: high-resolution mapping of replication origins. Science. 2001; 294:2357–2360. [PubMed: 11743203]

Zhang W, Sankaran S, Gozani O, Song J. A Meier-Gorlin syndrome mutation impairs the ORC1-nucleosome association. ACS Chem Biol. 2015; 10:1176–1180. [PubMed: 25689043]

Zhang Z, Hayashi MK, Merkel O, Stillman B, Xu RM. Structure and function of the BAH-containing domain of Orc1p in epigenetic silencing. EMBO J. 2002; 21:4600–4611. [PubMed: 12198162]

Highlights

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

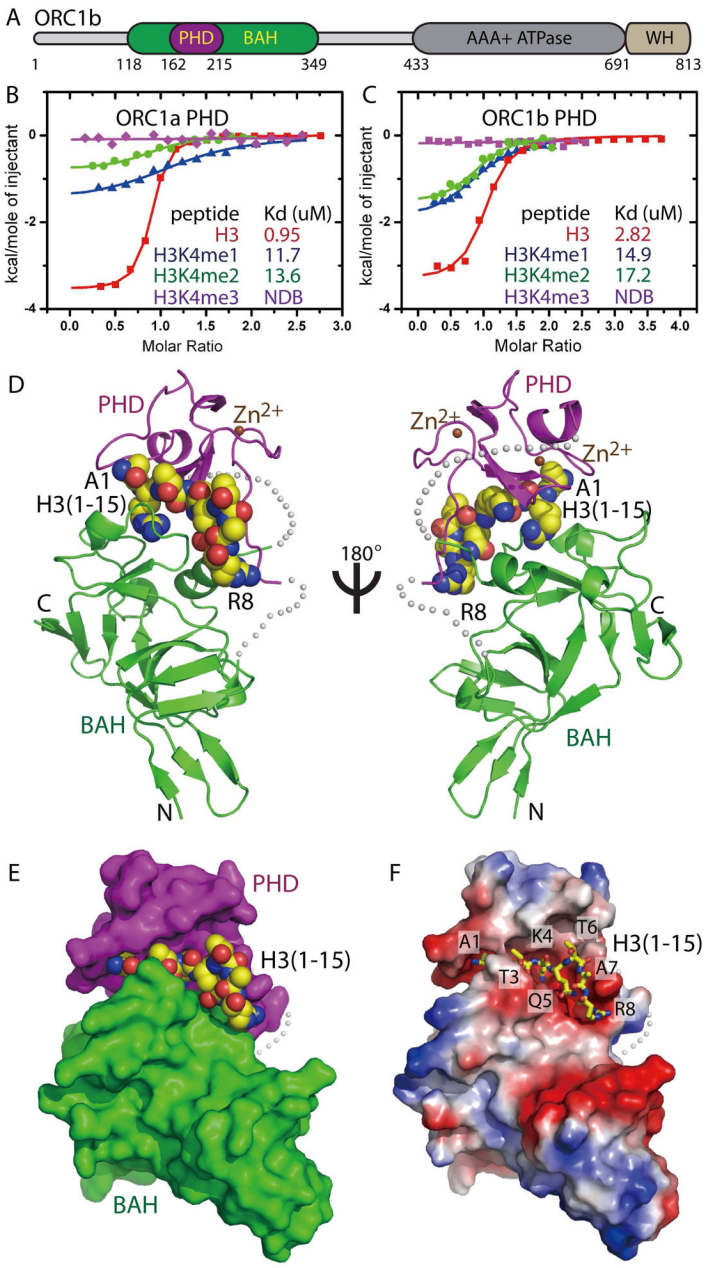


Figure 1. Arabidopsis ORC1 Protein Recognizes Unmodified H3 N-terminal Tail and Overall Structure of ORC1b BAH-PHD Cassette-H3 Peptide Complex

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

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

(D) A 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, respecitvely. 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) A 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) An electrostatistics surface view of the BAH-PHD cassette. The peptde is shown in stick representation. The peptide binds in a negatively charged surface cleft of the BAH-PHD cassette. The N-terminus insert into a negatively charged groove of the protein.

See also Figure S1 and S2

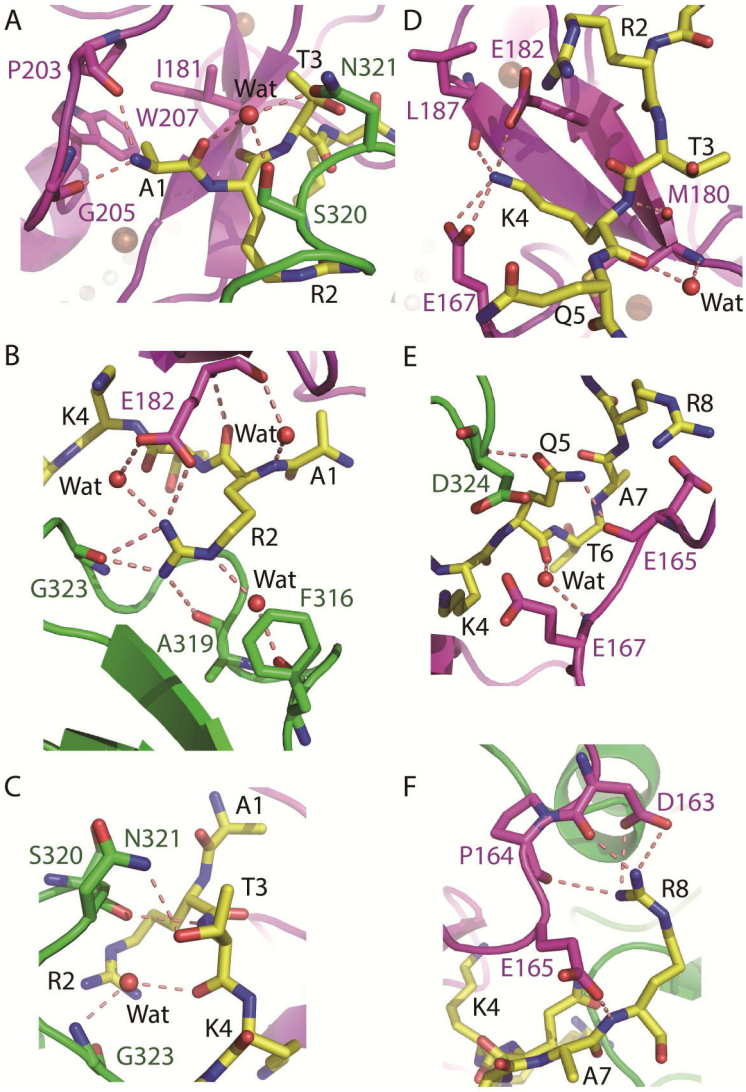


Figure 2. Molecular Basis for the Specific Recognition of the H3 tail by the ORC1b BAH-PHD Cassette

(A–F) Specific recognition of H3A1 (panel A), unmodified H3R2 (panel B), unmodified H3T3 (panel C), unmodified H3K4 (panel D), H3Q5 (panel E) and unmodified H3R8 (panel F). The BAH and PHD domains are colored in green and magenta, respectively. The peptide is shown in a sitck representation. The hydogen 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

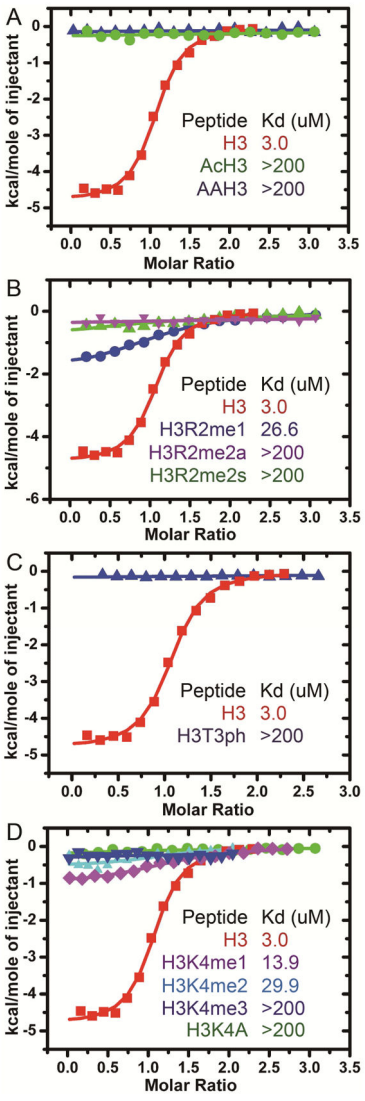


Figure 3. Comparison of Binding Affinity Measured by ITC Between the ORC1b BAH-PHD Cassette and Unmodified H3 Tail and Various Modifications of A1, R2, T3 and K4 positions on H3

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

(B) A comparison of binding affinities for H3 peptide versus peptides with mono- and di-methylation at R2.

(C) A comparison of binding affinities for H3 peptide versus peptides with phospohrylation at T3.

(D) A comparison of binding affinties 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 requirment of unmodified state of the H3 N-terminal tail and the length of the N-terminus of H3.

See also Figure S4

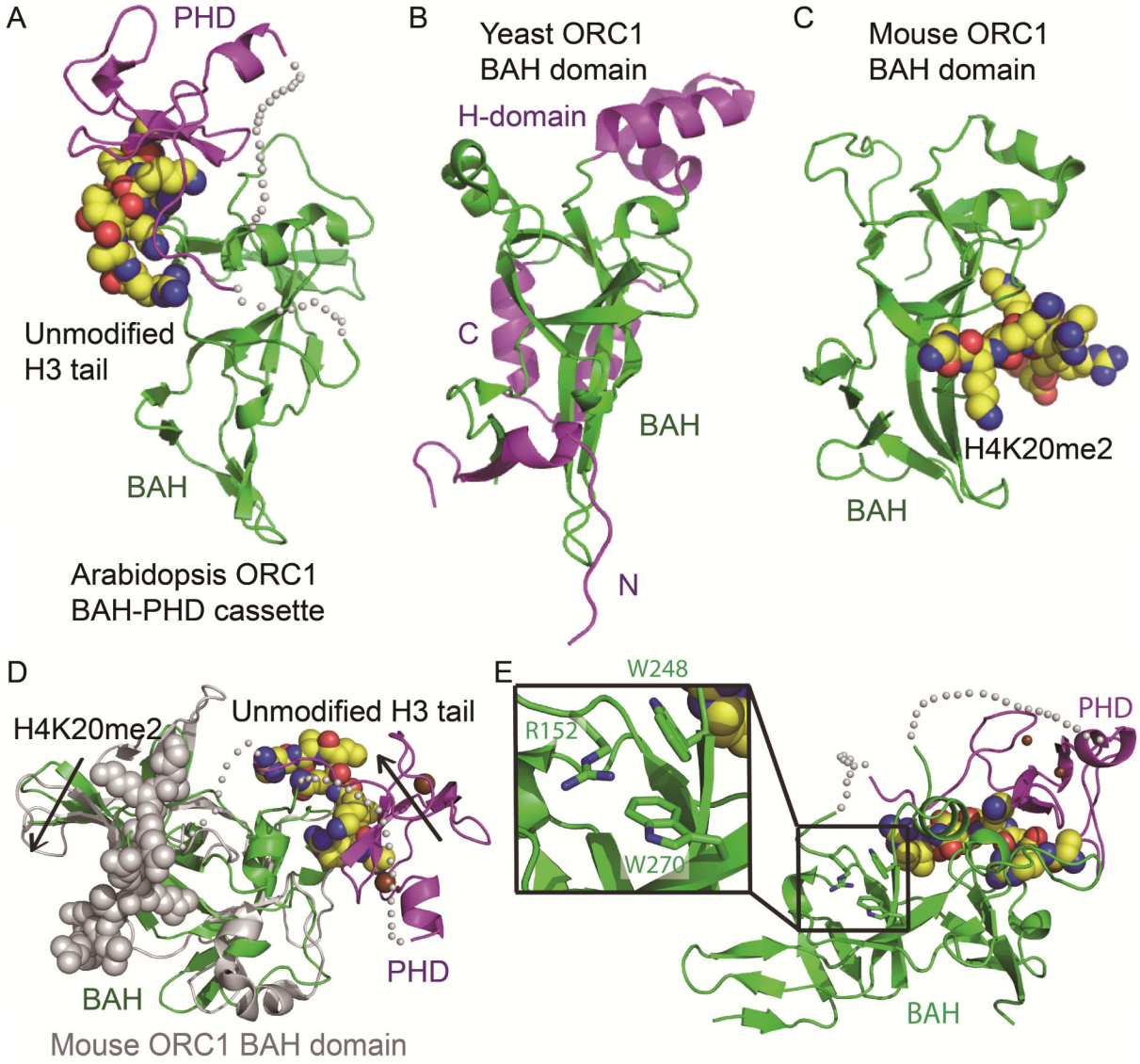


Figure 4. Comparison of Structures of Known ORC1 BAH Domains from Different Species

(A) 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 code: 1M4Z). It lacks the PHD domain but has an additional N-terminal, C-terminal, and an internal inserted H domain extensions, which are colored in magenta and the core BAH domain is colored in green. The H-domain can mediate the interaction with Sir1p and regulates ORC1 funciton.

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

(D) The superpostion 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 pepides have different binding sites and orientations.
(E) The BAH domain of Arabidopsis ORC1 has a partially conserved aromatic cage at the same postion as mouse BAH domain aromatic cage for accomodating 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.

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/σ(I)	46.3 (5.1)
Completeness (%)	99.8 (99.6)
Refinement and structure model	
R / Free R (%)	19.4 / 20.5
Number of atoms	1,874
Protein / Peptide	1,614 / 69
Zn^{2+}	2
Water	189
Average B factor (Å^2)	40.7
Protien / Peptide	40.8 / 39.1
Zn^{2+}	61.2
Water	40.4
RMS deviations	
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 are for highest-resolution shell.