Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie
der Ludwig-Maximilians-Universität München

Structural and Functional Analysis of Polycomblike-Polycomb repressive complex-2

Jeongyoon Choi

aus

Seoul, South Korea

2016

Erklärung

Diese Dissertation wurde im Sinne von § 7 der Promotionsordnung vom 28. November 2011 von Frau Pf. Dr. Elena Conti betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, den 01.08.2016

Jeongyoon Choi

.....

Dissertation eingereicht am 04.08.2016

1. Gutachterin: Pf. Dr. Elena Conti

2. Gutachter: Pf. Dr. Michael Sattler

Mündliche Prüfung am 23.09.2016

SUMMARY

Polycomb group (PcG) proteins repress transcription of many developmental regulator genes including homeotic selector (Hox) genes in multicellular organisms. PcG proteins form a variety of distinct multimeric complexes and modify target chromatin. For instance, the Polycomb repressive complex 2 (PRC2) consists of E(z), Su(z)12, Esc and Nurf55 as core subunits, and the complex mono-, di-, and tri-methylates histone H3 lysine 27 (H3K27me1/2/3). The H3K27me3 is critical for repression of Polycomb target genes. In order to generate sufficiently high levels of H3K27me3 at Polycomb target genes *in vivo*, PRC2 requires being associated with accessory subunits such as Polycomblike (Pcl). To date, the molecular mechanism of Pcl in the PRC2 has remained elusive. Here I show the crystal structure of conserved central domains of Pcl in *Drosophila melanogaster* (*Dm*), uncovering the presence of a winged-helix (WH) domain. I demonstrate that *Dm* Pcl and its human homologue PHF1 bind DNA via the conserved WH domain, and that this binding increases DNA binding of PRC2 in the context of a PHF1-PRC2 complex. In addition, I show that a conserved C-terminal domain of PHF1 is crucial for the PHF1 to form a stable complex with PRC2. Moreover, I present that the PHF1 WH and C-terminal domains enhance nucleosome binding of PRC2, and this is pivotal for generating high levels of H3K27me3 by PRC2 *in vitro*. Taken together, this study provides molecular insight into how the Pcl/PHF1 promotes H3K27 methylation by PRC2.

ACKNOWLEDGEMENTS

I would like to give many thanks to Jürg. Thank you for giving me an opportunity to work in your lab and supporting me during my PhD study. I also would like to thank the current and past members of the Müller lab: Reinhard, Maria, Ömer, Jacques, Karin, Felice, Ana, Sigrun, Friederike, Tom, Anca, Ksenia, Katja, Katharina, Petra, Ulrike and Ursula. Thanks for the kind atmosphere to work in and scientific discussions. I am also very grateful to Elena, Michael and Mario. Thank you for encouraging me during the last years through the thesis advisory committee meetings and supporting me as the thesis examination board. Also, I would like to thank Dr. Herzog, Pf. Peter-Hopfner and Pf. Föstermann for supporting me as the thesis examination board. I am also grateful to Christian, Karina and Sabine for helping my crystallization work. Also, I give many thanks to Stephan and Claire for consulting about biophysical experiments.

I am really thankful to Deborah, Jessie, David, Linda, Michelle, Garrett, Simon, Felix, Andreas, Zachary, Siegfried, Dunja, Adele, Stephanie, Petra, Je-kyeong, Yun-suk, Ye-won, Kwang-min, Sang-yeon and Go-woon. I am really grateful for our friendship and your support with encouraging words and prayer during last years.

I thank my family. 엄마, 아빠, 그리고 정훈이, 항상 고맙고 사랑해요. 멀리 있어도 마음으로 항상 응원해주고 기도해줘서 힘이 되고 든든했어요.

Last but not least, I thank God. Thank you so much for being always there and faithful with the constant love.

CONTENTS

I. INTRODUCTION

1. EARLY DEVELOPMENT IN EUKARYOTES

In eukaryotes, meticulous gene regulation during development is crucial to establish and maintain cell fates. Extensive studies of early development in *Drosophila melanogaster* (*Dm*; many studies, collected in Gilbert, 2014) have helped our understanding of the eukaryotic cell fate decisions to a great extent.

Early *Dm* embryos undergo nuclear divisions (karyokinesis) for 13 cycles without cell divisions (cytokinesis) generating a syncytium. At the nuclear division cycle 10, the nuclei migrate to the periphery of the embryos, and the embryos at this stage are called the syncytial blastoderm. After the division cycle 13, the cell membrane folds inwardly to partition the nuclei off and creates the cellular blastoderm (Gilbert, 2014; Figure 1).

Figure 1. Syncytial nuclear divisions and migrations in the *Dm* embryo

Laser confocal micrograph of *Dm* embryos with stained chromatin; numbers indicate the cell division cycles (1 to 14); A, anterior; P, posterior; the figure was adapted from Gilbert, 2014.

In the early *Dm* embryo, maternal effect genes specify cell fates, and segmentation genes determine the cell fates (Figure 2). In sections below, I provide a brief overview of the early development of *Dm* anterior (A)-posterior (P) body plan.

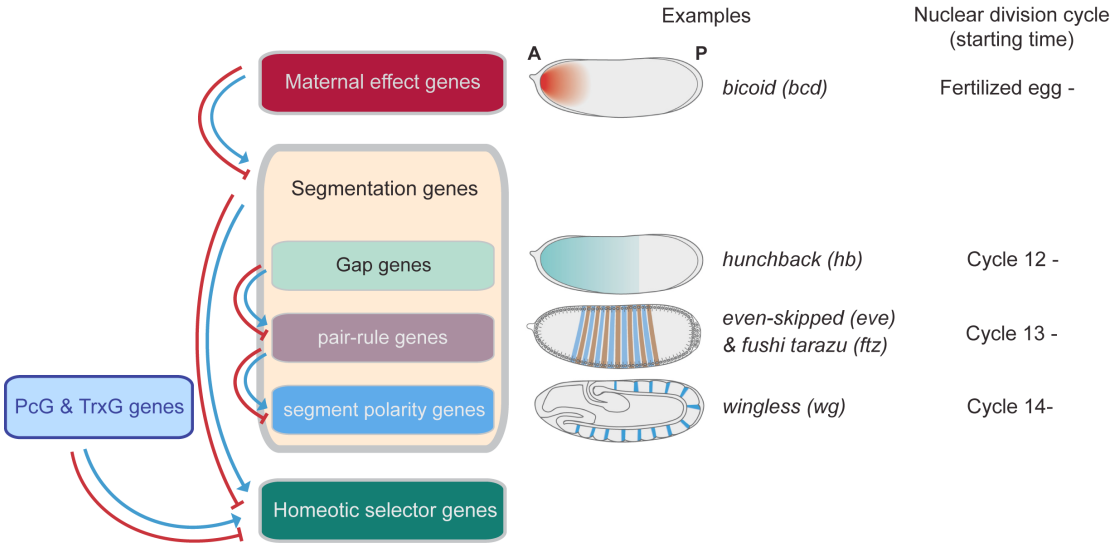


Figure 2. Anterior-posterior pattern formation in the early *Dm* embryo

Left: simplified scheme of homeotic gene regulation during the early *Dm* embryo development; PcG, Polycomb group; TrxG, Trithorax group; blue arrows indicate activation of gene expression; red, repression; middle: some examples of early maternal or zygotic gene expressions in the *Dm* embryo (embryo cartoons from http://www.mun.ca/biology/desmid/brian/BIOL3530/DEVO_02/devo_02.html); right: the nuclear division cycle of which each gene group starts to be expressed is indicated; please refer Figure 1 for embryo confocal micrographs at different division cycles.

1. Maternal effect genes

During oogenesis, the *Dm* oocyte begins to specify the A-P axis of the body. This specification starts with a localization of maternal effect messenger RNA (mRNA) by a cellular microtubule network: bicoid (bcd) mRNA in the anterior tip and nanos (nos) in the posterior. Upon fertilization of the oocyte, these maternal effect mRNAs are translated into proteins, which diffuse and establish gradients along the A-P axis. The Bcd protein forms an A-to-P gradient (Figure 2 and Figure

3A), whereas the Nanos a P-to-A gradient (Driever and Nüsslein-Volhard, 1988; Wang and Lehmann, 1991).

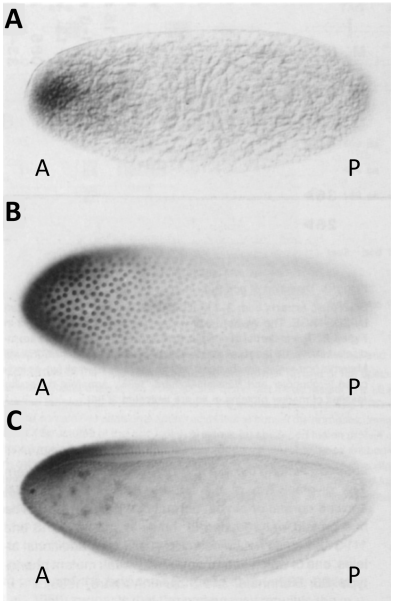


Figure 3. Maternal effect protein gradients in the *Dm* embryo at different developmental stages.

Bcd protein localization in the *Dm* embryo (A) shortly after egg deposition; (B) at syncytial blastoderm; (C) at cellular blastoderm; A, anterior; P, posterior; figure was adapted from (Driever and Nüsslein-Volhard, 1988).

2. Segmentation genes

After the cell fates are specified by maternal effect proteins, the cell fates are determined by segmentation gene products. The segmentation genes were first identified by which zygotic mutations led to the disruption of the *Dm* body plan. They are classified in three categories: gap genes, pair-rule genes and segment polarity genes (Nüsslein-Volhard and Wieschaus, 1980). These genes are expressed in a hierarchical manner and regulated by an extensive network of their gene products (Figure 2).

Firstly, at the end of nuclear division cycle 12, the gap genes start to be expressed in broad domains along the A-P axis of *Dm*. The expression is tightly regulated by the gradient of maternal effect proteins and gap proteins themselves. Next, upon the cycle 13, the pair-rule genes begin to be expressed in every other segment, which will develop into each part of the body. The maternal effect and the gap proteins control

the expression of pair-rule genes. Hereafter, the first indication of embryo segmentation can be observed. Then upon the cycle 14, when the *Dm* embryo undergoes cellularisation and forms the cellular blastoderm, the segment polarity genes begin to be expressed in every segment. The pair-rule gene products and the segment polarity gene products themselves regulate expression of the segment polarity genes. After this, every segmental boundary is set and each segment primordium obtains its identity. The segmentation gene products then regulate the expression of homeotic selector genes (Hox genes).

3. Homeotic selector genes

Hox genes are a group of genes encoding transcription factors that generate specific identity of body parts. The Hox genes are highly conserved and their expression pattern is similar in almost every bilateral animal (Pearson et al., 2005) (Figure 4). In *Dm*, the Hox genes are clustered in the chromosome 3, consisting of Antennapedia and bithorax complex (Lewis, 1978).

The Hox gene expression is a highly dynamic process throughout the development, being controlled not only by segmentation gene products but also by the Hox gene products themselves. For instance, *Antennapedia* (*Ant*) gene expression is repressed by the other Hox proteins such as Ultrabithorax, whose spatial expression is posterior to the Ant expression (Gonzalezreyes and Morata, 1990).

The segmentation genes are only transiently expressed, thus other factors are required to maintain the precise expression pattern of the Hox genes during the development. This is where Polycomb group (PcG) and Trithorax group (TrxG) proteins come into play: PcG proteins are required for maintaining the repressed state of Hox gene expression (Mckeon and Brock, 1991; Simon et al., 1992), whereas

TrxG proteins for the active state. More information about the PcG proteins is described in detail in the following sections.

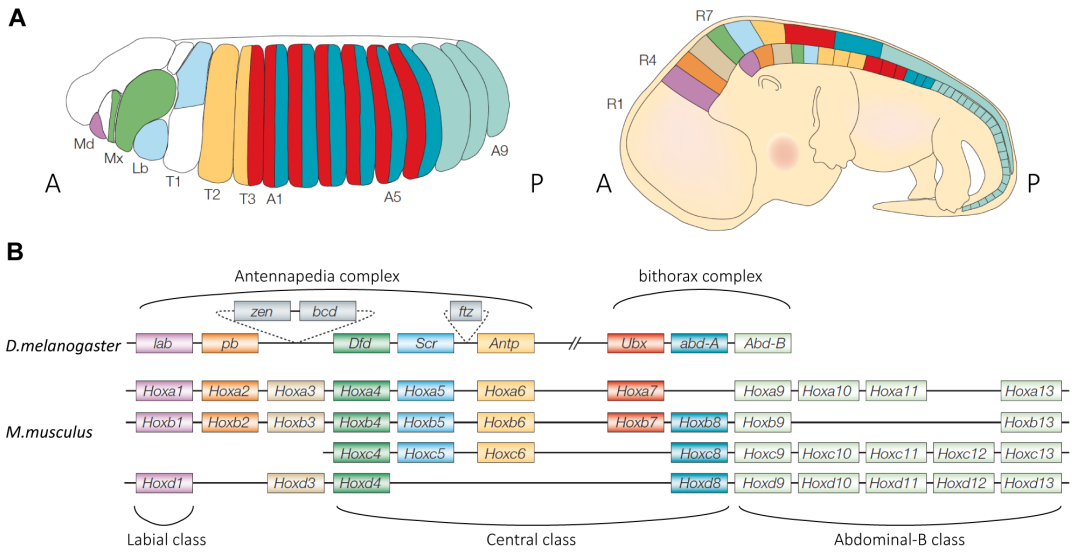


Figure 4. Hox gene colinearity

(A) Expression domains of Hox genes in the *Dm* embryo at stage 13 (left) and the *Mus musculus* (*Mm*) embryo at embryonic day 12.5 (right); segment abbreviations; Md, mandibular; Mx, maxillary; Lb, labial; T1-T3, thoracic segment 1-3; A1-A9, abdominal segment 1-9; R1-R7, rhombomere 1-7; A: anterior; P: posterior. (B) Hox gene loci in *Dm* and *Mm* chromosomes; gene abbreviations: *lab, labial*; *pb, proboscipedia*; *zen, zerknullt*; *bcd, bicoid*; *Dfd, Deformed*; *Scr, Sex combs reduced*; *ftz, fushi tarazu*; *Antp, Antennapedia*; *Ubx, Ultrabithorax*; *abd-A, abdominal-A*; *Abd-B, Abdominal-B*; schematics were adapted from (Pearson et al., 2005).

2. Polycomb group proteins

1. History

The *Polycomb* (*Pc*) gene was first identified by *Dm* genetic screens as a coding gene of the bithorax complex repressor (Lewis, 1978). Subsequently, Ian Duncan isolated another gene, whose mutation leads to segmental transformation of *Drosophila*. He suggested that this gene product regulates the expression of the genes in Antennapedia complex as well as the bithorax complex. Since the mutant

phenotype was similar to the one of *Polycomb*, the newly discovered gene was named as *Polycomblike* (*Pcl*; Duncan, 1982). Thereafter, a number of scientists identified dozens of genes exhibiting similar phenotypes as the *Pc*. These genes are thus now classified as PcG genes (Jürgens, 1985). The PcG genes and the function of their encoded proteins are highly conserved in multicellular organisms (Whitcomb et al., 2007).

2. PcG proteins form multimeric complexes

Years later, a number of research groups demonstrated that PcG proteins form multimeric complexes by biochemical purifications. In 1998 Harte et al. presented that Extra sex combs (Esc) and Enhancer of zeste (E(z)) are directly associated *in vivo* by showing their co-localization in *Dm* larval polytene chromosomes[1] (Tie et al., 1998). Then, following the identification of another novel PcG gene, *suppressor of zeste 12* (*Su(z)12*; Birve et al., 2001), several research groups reported that the E(z) is directly associated not only with Esc but also with Su(z)12 and Nucleosome-remodelling factor 55 kDa subunit (Nurf55), a histone chaperone. This association was shown both in the *Drosophila* embryo and the mammalian cells (Cao et al., 2002; Czermin et al., 2002; Kuzmichev et al., 2002; Müller et al., 2002). This complex was called as Polycomb repressive complex-2 (PRC2; Kuzmichev et al., 2002), named after Polycomb repressive complex-1 (PRC1), which had been purified by Kingston and his colleagues in 1999 (Shao et al., 1999). Subsequently, several more PcG complexes were purified: Pho repressive complex (Pho-RC; Klymenko, 2006), RING-associated factor (dRAF; Lagarou et al., 2008) and Polycomb repressive deubiquitinase (PR-DUB; Scheuermann et al., 2010) (Figure 5). Besides, most recently, a number of PRC1 variants, each including the RING1A/B and distinct

[1] Polytene chromosomes are over-sized chromosomes as a result of sister chromatid fusions by multiple rounds of DNA replication without cell devision (endomitosis). The polytene chromosomes are commonly found in *Drosophila* salivary glands.

PCGF, have been identified in mammals (Gao et al., 2012). The PcG protein complexes contain also some non-PcG proteins (e.g. Nurf55, AEBP2, JARID2 and KDM2).

3. PcG proteins modify chromatin

An inherent feature of all the Polycomb repressive complexes is to bind and modify target chromatin. Among these, solely PhoRC has been shown to bind to DNA in a sequence-specific manner (Klymenko, 2006). PRC1 and dRAF ubiquitinate histone H2A Lys 118, and PRC2 methylates histone H3 Lys 27 (Figure 5). More details of the PcG subunits and their human homologues are described in Table 1. Among the PcG complexes, PRC2 is discussed further in the section I.2.7 (p. 14).

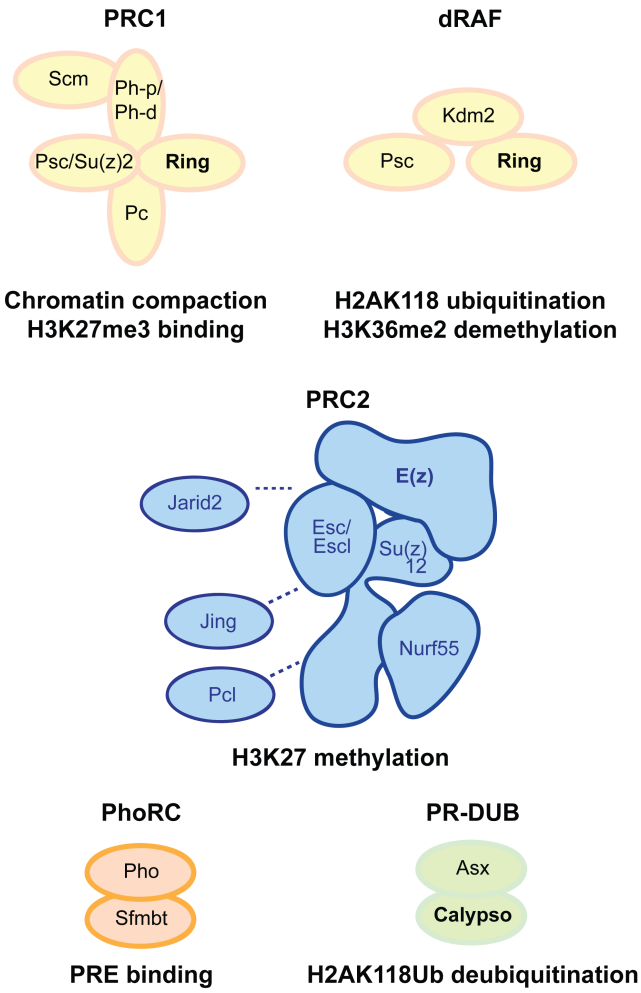


Figure 5. Schematics of *Dm* PcG complexes

Key functions of the PcG complexes are described below each schematic; catalytic subunits to modify chromatin are in bold. Subunits drawn distantly from the PRC2 represent sub-stoichiometric subunits of PRC2 (Jarid2, Jing and Pcl); PRE, Polycomb response element; full names of the subunits are listed in Table1.

Drosophila melanogaster	Homo sapience	Conserved function
Polycomb repressive complex 1 (PRC1)		
Sex comb extra (Sce/dRING)	RING1B/RNF2 RING1A/RNF1	E3 Ligase for H2A-K118 (K119) ubiquitylation
Posterior sex combs (Psc)	PCGF4/BMI1	Enhances E3 Ligase activity of RING, compacts chromatin
Suppressor of zeste 2 (Su(z)2)	PCGF2/MEL18	
Polycomb (Pc)	CBX2/PC1	
	CBX4/PC2	
	CBX6	binds to H3-K27me3 through chromodomain
	CBX7	
	CBX8/PC3	
Polyhomeotic-proximal/distal (Ph-P/D)	PHC1-3	compacts chromatin
Sexcombs on midleg (SCM)	SCMH1-2, SCML2	binds to H3-K9me1 and Sfmbt
Lysine (K)-specific demethylase 2 (Kdm2)	KDM2B	demethylates H3-K36me3
Polycomb repressive complex 2 (PRC2)		
Enhancer of zeste (E(z))	EZH2 EZH1	catalytic subunit of H3-K27 methylation
Supressor of zeste 12 (Su(z)12)	SUZ12	required for HMTase activity, binds target chromatin
Extra sexcombs (Esc)	EED1-4	binds to H3-K27me3 for H3-K27me3 propagation, required for HMTase activity
Extra sexcombs-like (Escl)		
Nucleosome remodeling factor 55 (Nurf55)	RBBP7/RBAP46 RBBP4/RBAP48	binds target chromatin
Polycomblike (Pcl)	PHF1/PCL1 MTF2/PCL2 PHF19/PCL3	stimulates H3-K27me3, sub-stiochiometric subunit of PRC2
Jing	AEBP2	stimulates H3-K27me3, sub-stiochiometric subunit of PRC2
Jumonji, AT rich interactive domain 2 (JARID2)	JARID2	sub-stiochiometric subunit of PRC2
Pho repressive complex (PhoRC)		
Pleiohomeotic (Pho)	YY1, YY2	binds to DNA sequence-specifically
Pleiohomeotic-like (Phol)		PHO paralog
Scm-like with four MBT domains (Sfmbt)	L3MBTL2, MBTD1, hSFMBT1/2	binds to H3-K9me1/2 , H4-K20me1/2
Polycomb repressive deubiquitinase (PR-DUB)		
Calypso	BAP1	deubiquitylates H2A-K118 (K119) Ub
Anterior sex combs (Asx)	ASXL1, ASXL2, ASXL3	activates deubiquitylase activity
dRING-associated factors (dRAF)		
Ring/Sexcomb extra (Ring/Sce)		E3 Ligase for H2A-K118 (K119) ubiquitylation
Posterior sex comb (Psc)		Enhances E3 Ligase activity of RING
Lysine (K)-specific demethylase 2 (Kdm2)	KDM2B	demethylates H3-K36me3

Table 1. List of subunits of PcG protein complexes and their functions

Tables were adapted from (Di Croce and Helin, 2013; Piunti and Shilatifard, 2016).

4. Polycomb target genes

Genome-wide chromatin immunoprecipitation (ChIP) on microarray analyses performed in *Drosophila* cultured S2 cells and in mouse/ human embryonic stem (ES) cells revealed many novel Polycomb target genes in addition to the Hox genes (Boyer et al., 2006; Lee et al., 2006; Oktaba et al., 2008; Schwartz et al., 2006; Tolhuis et al., 2006). These newly identified Polycomb target genes include many transcriptional and developmental regulators and tumour suppressor genes.

5. PcG protein recruitment

In *Dm*, Polycomb response elements (PREs) have been described as cis-regulatory elements to recruit PcG proteins (Chan et al., 1994). The PREs span several hundreds to a few thousand base pairs in length and can be located upstream or downstream of the Polycomb target genes. The PREs contain various conserved short consensus motifs of different DNA binding proteins including Pho (Pho-RC subunit; Fritsch et al., 1999), GAF and ZESTE (Schwartz and Pirrotta, 2007). One recent report showed that the recruitment of PcG proteins to the PRE transgene is abolished when Pho consensus sequences are mutated (Frey et al., 2016).

PRC1 has been reported to interact directly with PhoRC (Mohd-Sarip, 2005; Mohd-Sarip et al., 2002). More recently, the Müller group provided structural and biochemical evidences of the interaction between two Sterile alpha motif (SAM) domains of the Scm (PRC1 subunit) and the Sfmbt (PhoRC subunit; Frey et al., 2016). In addition, non-canonical mammalian PRC1 variants were shown to be recruited to CpG islands[1] by KDM2B, which binds to non-methylated CpG

[1] CpG islands: a region with at least 200bp with a GC percentage of at least 60%. Unlike other regions in the vertebrate genome, the CpG sites in the CpG islands of promoters are unmethylated if the genes are expressed (reviewed in Deaton and Bird, 2011).

sequences via its CXXC Zinc finger[1] domain (Blackledge et al., 2014; Farcas et al., 2012).

However, our understanding of PRC2 recruitment is yet limited. Several research groups suggested that the long noncoding (lnc) RNA[2] helps PRC2 targeting (reviewed in Blackledge et al., 2015). For instance, the Cech group showed that the PRC2 binds to the lncRNA promiscuously by gel shift assays, and suggested that this non-specific binding of PRC2 to RNA transcripts allows the PRC2 to scan for the target genes (Davidovich et al., 2013). In addition, several groups proposed that accessory subunits of PRC2 (Pcl, AEBP2 and JARID2) facilitate the PRC2 recruitment. For example, the Pcl mutant *Dm* embryo shows diminished levels of Su(z)12 bound to the PREs (Nekrasov et al., 2007). Additionally, PRC2 associated with JARID2 and AEBP2 is suggested to be recruited to the ubiquitinated H2A, which is catalysed by PRC1 (Blackledge et al., 2014; Cooper et al., 2014; Kalb et al., 2014). Besides, Helin and his colleagues proposed that a transcriptionally repressed state itself could be responsible for recruiting PRC2 to the CpG islands in mouse embryonic stem (mES) cells (Riising et al., 2014).

6. Structural studies of PcG protein

Cooperation amongst the PcG subunits and crosstalk between PcG proteins and covalently-modified nucleosomes play crucial roles in the PcG enzymatic action. For instance, E(z), the catalytic subunit of PRC2, can methylate its target chromatin only in the presence of other subunits of PRC2 (Müller et al., 2002). In addition, nucleosomes harbouring active marks (e.g. trimethylation on H3 Lys4 (H3K4me3) or

[1] CXXC Zinc (ZN) finger: characterized by two CXXC repeats, found in many chromatin-associated proteins. Some CXXC ZN fingers exhibit non-methylated CpG binding activity

[2] Long noncoding (lnc) RNA: RNA transcript of longer than 100 nucleotides, which does not encode proteins

on H3K36 (H3K36me3)) inhibit enzymatic activity of the PRC2 (Schmitges et al., 2011). Thus, elucidating the cooperativity of the PcG subunits and the crosstalk between the PcG proteins and the modified nucleosomes has been of particular interest in Polycomb research. In order to unravel these, many research groups have used structural approaches.

Regarding PRC1, the *Dm* Polycomb (Pc) chromodomain structure was first solved in the presence of a histone $H3_{20\text{-}30}$K27me3 peptide (Fischle, 2003; Min et al., 2003). The Pc-$H3_{20\text{-}30}$K27me3 co-crystal structure consolidated previous genetic studies, which had revealed the crucial role of E(z) in stabilizing Psc/Su(z)2-Pc to the target loci via the Pc chromodomain (Platero et al., 1996; Rastelli et al., 1993). In addition, Buchwald et al. solved a crystal structure of the PRC1 ubiquitination module, the Ring-Ring heterodimeric complex of murine Ring1b and Bmi1, and uncovered that the E3 ligase activity of the Ring1b is enhanced by the Bmi1 (Buchwald et al., 2006). A number of years later, McGinty et al. provided a crystal structure of a complex containing the PRC1 ubiquitination module and human E2 ligase UbcH5c on a nucleosome core particle (NCP) (Figure 6B). This structure shows that the Ring1b recognizes the NCP by binding to the histone H2A-H2B acidic patch[1] via the arginine anchor (Figure 6C; McGinty et al., 2014).

Regarding PRC2, Margueron et al. presented a crystal structure of human EED bound to a histone H3 peptide with K27me3 via an aromatic cage, and showed that the H3K27me3-bound EED activates PRC2 allosterically (Margueron et al., 2009). In addition, Schmitges et al. solved a crystal structure of the *Dm* Nurf55 bound to the histone H3 peptide via an acidic pocket (Figure 6D), and demonstrated that the

[1] acidic patch: acidic surface of H2A/H2B in a nucleosome generated by acidic side chains such as H2A E61, D90, and E92.

Nurf55 compromises binding to methylated H3 peptides (Schmitges et al., 2011). Besides, the EED and the Nurf55 structures have been solved together with short peptides of the other PRC2 subunits: $EZH2_{39-68}$ (Han et al., 2007) and $Su(z)12_{79-91}$ (Schmitges et al., 2011) respectively. Next, using a hybrid approach of negative staining electron microscopy (EM) and cross-linking mass spectrometry (MS), the Nogales group presented a 21 Å resolution structure of human AEBP2-PRC2 (Figure 6E), granting an overview of the molecular architecture of the PRC2 (Ciferri et al., 2012). A few years later, the Liu group solved the first high resolution structure of PRC2 multi-subunits, including EZH2-EED-SUZ12 VEFS[1] domain (EZH2-EED-$SUZ12_{VEFS}$), in *Chaetomium thermophilum* (*Ct*) in the presence of a histone H3 peptide containing a paediatric brain cancer mutation, the H3 Lysine 27 to methionine (H3K27M) (Figure 6F; Jiao and Liu, 2015). The crystal structure of the *Ct* EZH2-EED-$SUZ12_{VEFS}$ fits well with the upper half part of the human full-length AEBP2-PRC2 EM structure (Figure 6G). Shortly after the *Ct* EZH2-EED-$SUZ12_{VEFS}$ structural report, sub-human PRC2 crystal structures (*Hs* EZH2-EED-$SUZ12_{VEFS}$) were solved by two groups (Deng et al., 2016; Justin et al., 2016). The *Hs* EZH2-EED-$SUZ12_{VEFS}$ structure exhibits high similarity with the *Ct* EZH2-EED-$SUZ12_{VEFS}$ structure. More details of the PRC2 structures are explained in the section I.7.1) (p. 14).

[1] VEFS domain: VRN2-EMF2-FIS2-Su(z)12 domain

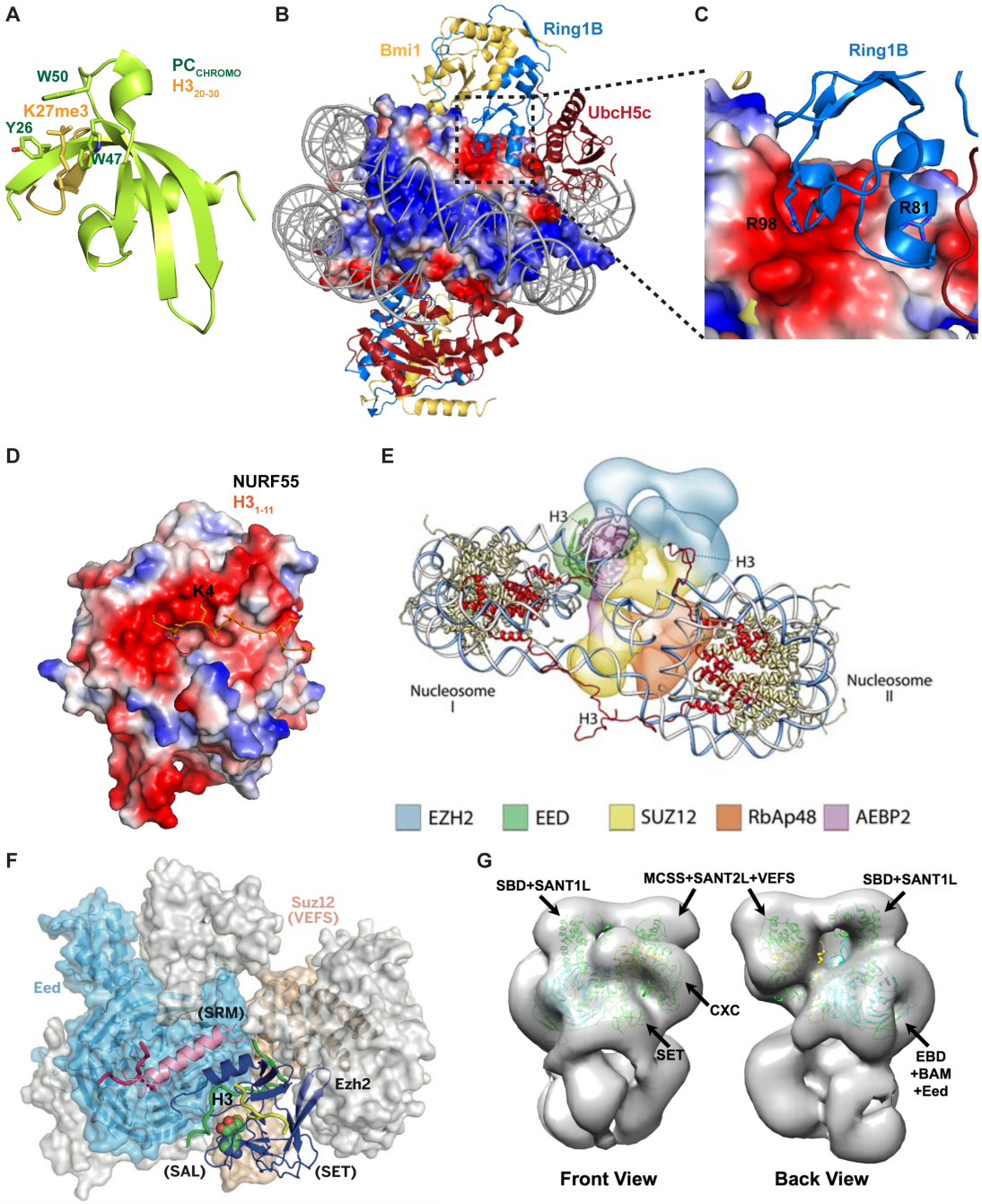


Figure 6. Examples of PcG protein structures

(A) An aromatic cage of Pc chromodomain binding to $H3_{20-30}$K27me3 (PDBcode: 1PFB). (B) PRC1 ubiquitination module onto NCP (PDBcode: 4R8P). (C) An arginine anchor of Ring1B binding to the histone H2A-H2B acidic patch. (D) An acidic pocket of Nurf55 binding to the H3 N-term peptide (PDBcode: 2YBA). (E) EM structure of AEBP2-PRC2, docked on di-nucleosomes (model); figure taken from (Ciferri et al., 2012). (F) and (G) *Ct* EZH2-EED-$SUZ12_{VEFS}$ structure in surface presentation (F) and in cartoon presentation fitted to the human AEBP2-PRC2 EM map (G); figures taken from (Jiao and Liu, 2015).

7. Polycomb repressive complex 2

The core subunits of PRC2 consist of E(z), Su(z)12, Esc and Nurf55 in *Dm* and EZH1/2, SUZ12, EED1-4 and RBBP4/7 in *Hs* (Figure 7; Cao et al., 2002; Czermin et al., 2002; Kuzmichev et al., 2002; Müller et al., 2002). In addition to these core subunits, some of the early biochemical purification reports have shown the association of PRC2 with other proteins such as histone deacetylase HDAC1/2 (Tie et al., 2001), Adipocyte enhancer-binding protein 2 (AEBP2; Cao et al., 2002) and Pcl (Tie et al., 2003). Among those, AEBP2 and Pcl are now generally accepted as accessory subunits of PRC2 together with Jarid2, which was identified years later (Peng et al., 2009).

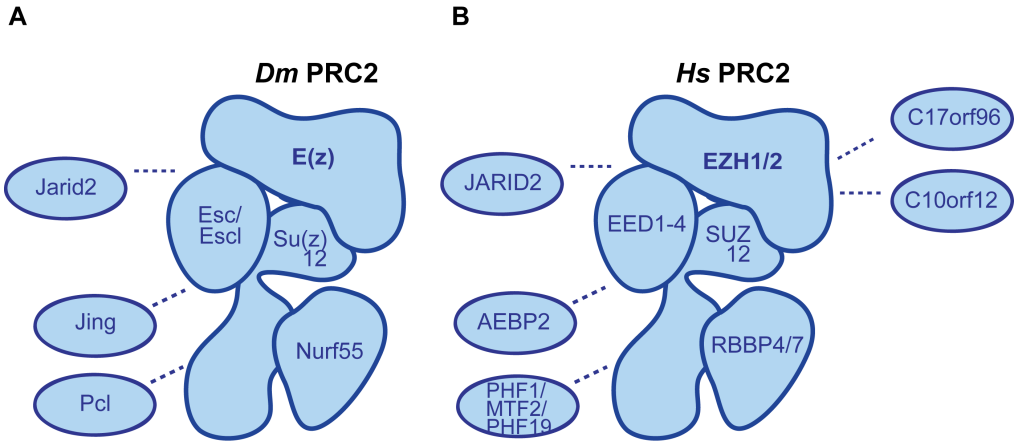


Figure 7. *Dm* and *Hs* PRC2 subunit compositions

(A) Subunits of *Dm* PRC2. (B) Subunits of *Hs* PRC2; two uncharacterized proteins, C17orf96 and C10orf12 were identified by mass spectrometry from human PRC2 pull-downs (Alekseyenko et al., 2014; Smits et al., 2012). Their counterparts in *Dm* are not present (or not identified yet).

1) Histone methyltransferase

The early reports of PRC2 purification presented the enzymatic function of the complex: the site-specific methyltransferase activity on histone H3K27 (Cao et al.,

2002; Czermin et al., 2002; Kuzmichev et al., 2002; Müller et al., 2002) as well as H3K9 *in vitro* in some reports (Czermin et al., 2002, Kuzmichev et al., 2002). The Reinberg group suggested that PRC2 methylates histone H1K26 depending on which isoform of EED protein is integrated into the complex (Kuzmichev et al., 2004). However, many subsequent works have demonstrated that the H3K27 tri-methylation is the critical function of PRC2 in Polycomb repression (Simon and Kingston, 2009). Notably, the cells harbouring a specific point mutation of Lysine 27 of histone H3 to Arginine (H3K27R) in *Dm* phenocopy the PRC2 mutant, misexpressing some Polycomb target genes (Pengelly et al., 2013).

i. Contribution of different PRC2 subunits to the H3K27 methylation

E(z) mono-, di- and tri-methylates Lysine 27 of histone H3 as a catalytic subunit of PRC2 (Ebert et al., 2004; Müller et al., 2002). However, without other PRC2 core subunits, the enzymatic activity of E(z) has been shown to be reduced at least a 1000-fold by *in vitro* histone methyltransferase (HMTase) assays (Müller et al., 2002). Several extensive biochemical reports have demonstrated that the Esc and the Su(z)12 are essential for the HMTase activity, whereas the Nurf55 and the Su(z)12 are required for nucleosome binding (Ketel et al., 2005; Nekrasov et al., 2005; Pasini et al., 2004). Specifically, the Simon group reported that the Su(z)12 VEFS domain is the minimal requirement for the HMTase activity of PRC2 *in vitro*. They also mapped a region that stimulates the HMTase activity and a section that interacts to the E(z) in the VEFS domain (Rai et al., 2013). In addition, they suggested that the Su(z)12 Zinc finger (ZnF) domain is required for PRC2 targeting to the PREs by ChIP experiments in *Drosophila* S2 cells.

A recently reported crystal structure of the *Ct* sub-PRC2 (Jiao and Liu, 2015) visualized the previous biochemical characterizations at the atomic details (Figure

8B). This structural analysis revealed that the catalytic domain of the EZH2 is bipartite, composed of both SET[1] and SAL[2] domains. The SET and SAL domains are remotely located in the primary amino acid sequence of EZH2 (Figure 8A), but the two domains are brought together in the tertiary structure, stabilized by the EED and the SUZ12 VEFS domain to maintain the active conformation of EZH2 (Figure 8C).

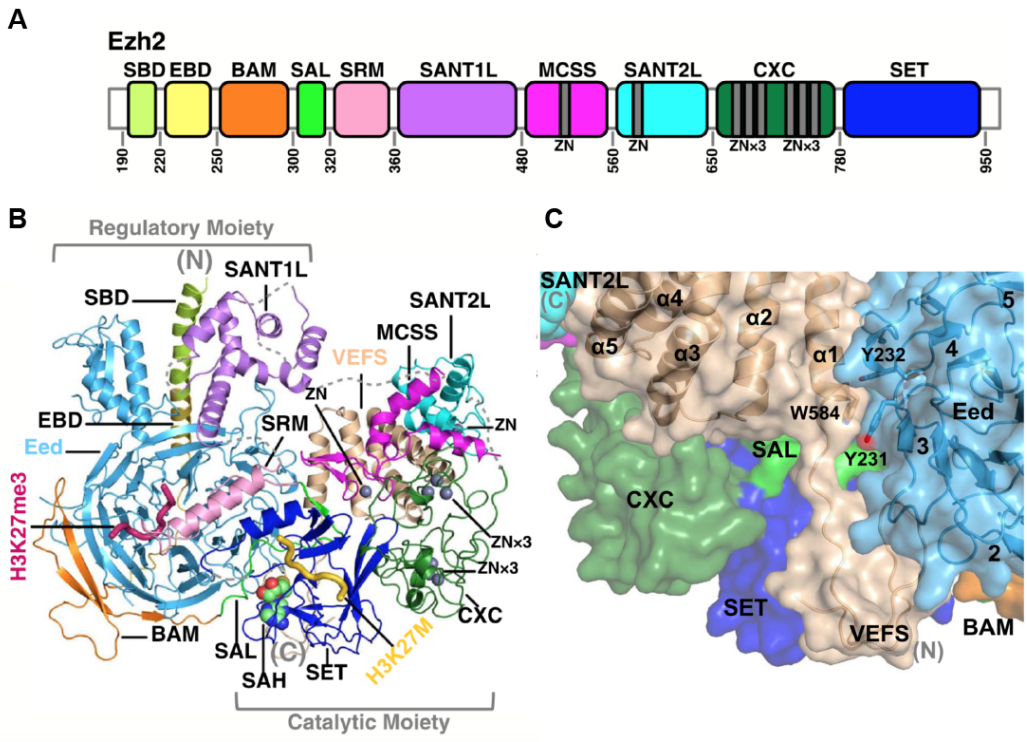


Figure 8. Crystal structure of *Ct* EZH2-EED-SUZ12$_{VEFS}$

(A) Domain architecture of *Ct* EZH2. (B) Overall crystal structure of *Ct* EZH2-EED-SUZ12 $_{VEFS}$ in cartoon representation. (C) Zoomed view of interaction amongst *Ct* EZH2 (SAL-SET), SUZ12$_{VEFS}$ and EED in surface representation. Figures were adapted from (Jiao and Liu, 2015).

[1] SET domain: su(var)3-9, enhancer-of-zeste and trithorax domain

[2] SAL domain: SET activation loop domain

ii. Regulation of histone methyltransferase activity of PRC2

A steady-state kinetic parameter analysis showed that the turnover number (k_{cat})[1] for PRC2 of the H3K27me2 substrate is about ten times lower than that of the H3K27me0 or H3K27me1 (Sneeringer et al., 2010). In addition, the crystal structure of the *Ct* PRC2 implies that the substrate binding channel in the EZH2 confers a spatial restriction to the H3K27me2 (Jiao and Liu, 2015). In other words, converting the H3K27me2 to H3K27me3 by PRC2 is not catalytically favourable.

In order to generate the H3K27me3 efficiently, therefore, the PRC2 requires other factors. For example, accessory subunits of PRC2, such as PCL, AEBP2 and JARID2, enhance the catalytic efficiency of PRC2 (Cao et al., 2008; Herz et al., 2012; Kalb et al., 2014; Li et al., 2010; Nekrasov et al., 2007; Peng et al., 2009; Sarma et al., 2008). In addition, a recent biochemical report showed that PRC2 is activated in dense chromatin environments, presumably by sensing unmodified histone H3 tails spanning the residues from 31 to 42 (Yuan et al., 2012). Besides, the Reinberg group demonstrated that EED binding to K27me3 on histone H3 tails *in trans* promotes the enzymatic activity of PRC2, and proposed that the H3K27me3 is propagated by this (Margueron et al., 2009). Furthermore, the *Ct* PRC2 crystal structure shows that not only the EED but also the SRM[2] domain in EZH2 binds to the H3K27me3 peptide. Moreover, the structure shows that the EZH2 SRM domain mediates the crosstalk between the H3K27me3 peptide and the SET domain of EZH2 to stimulate the catalysis (Jiao and Liu, 2015).

---

[1] k_{cat}: turnover number, the maximum number of chemical conversions of substrate molecules per second that each catalytic site will execute for a given enzyme concentration

[2] SRM domain: stimulation-responsive motif domain

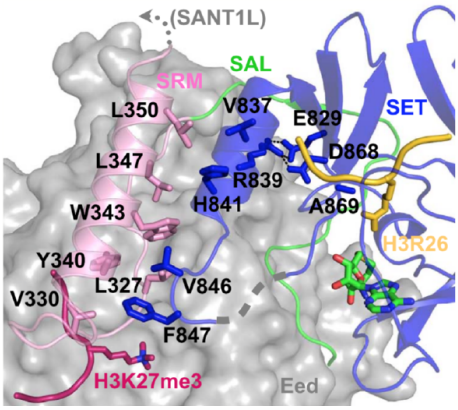


Figure 9. Crosstalk between the H3K27me3 peptide and the SET domain through the SRM domain

H3 peptide in magenta represents the stimulating peptide (H3K27me3) bound to the EED; yellow, the substrate peptide (H3K27M) bound to the SET domain, figure was taken from the (Jiao and Liu, 2015).

On the other hand, PRC2 is inhibited by active histone marks such as H3K4me3 and H3K36me3 *in cis*. As briefly mentioned in the section I. 2. 6 (p. 10), the Nurf55 binds to the histone $H3_{1-11}$ via its acidic pocket, where any methylation on Lys 4 in the $H3_{1-11}$ repulses the stable binding to the Nurf55 (Figure 6D). Although that report did not provide the detailed mechanism concerning how the H3K36me3 peptide regulates PRC2, the report showed that the nucleosomes possessing either H3K4me3 or H3K36me3 decrease the HMTase activity of PRC2 (Schmitges et al., 2011).

iii. Distribution of H3K27 methylations and PcG proteins

Mass spectrometry analysis of the methylated status of the H3K27 in wild-type *Dm* larval nuclear extracts revealed that around 50 % of the H3K27 is dimethylated, whereas only 10 % is trimethylated (Ebert et al., 2004). Another more recent genome-wide ChIP sequencing (ChIP-seq) experiment performed in mES cells demonstrated that 72 % of the H3K27 residue is modified to the H3K27me2, whereas only 7 % is to the H3K27me3 (Ferrari et al., 2013).

In addition, a number of ChIP experiments have provided insights about the distribution of H3K27me3 and PcG proteins in Polycomb target genes (Gutiérrez et al., 2012; Papp and Muller, 2006; Schwartz et al., 2006; Tolhuis et al., 2006). For

instance, H3K27me3 is shown to form broad domains beyond the localization of PcG proteins, shown by ChIP-chip analysis in *Dm* S2 cultured cells. Interestingly, the distribution of the Pc (the reader of H3K27me3) is broader than the one of E(z) (the writer of H3K27me3) (Schwartz et al., 2006).

Besides, the Müller group performed ChIP-quantitative PCR experiments in the Polycomb target gene (*Ubx*) region using two distinct *Dm* larval tissues: wing imaginal discs[1], where the *Ubx* gene is transcriptionally inactive; haltere/3rd leg imaginal discs, where the *Ubx* is active. They demonstrated that the repressive marks, such as H3K27me3, H3K9me3 and H4K20me3, decorate the upstream control region of *Ubx* in both tissues, whilst the repressive marks demarcate the promoter and 5'-coding regions solely in the wing disc tissue. Moreover, they presented that the PcG and TrxG proteins (except for the Ash1 protein) are localized to the PRE regions in both tissues (Papp and Muller, 2006).

2) PRC2 and cancer

As mentioned in the section I. 2. 4 (p. 9), Polycomb targets include tumour suppressor genes. Thus, up-regulation of PRC2 function is closely linked to malignant tumours. The PRC2 function can be up-regulated by (i) overexpression of PcG genes owing to intergenic or intronic mutations (Vogelstein et al., 2013) or (ii) hyperactive PcG proteins due to missense mutations, such as EZH2$^{Y641F/N}$ (Table 2; Sneeringer et al., 2010). Specifically, the PRC2 with EZH2$^{Y641F/N}$ favours H3K27me2 as the substrate, leading to elevation of H3K27me3 (Sneeringer et al., 2010).

[1] imaginal disc: one of the parts of a holometabolous insect larva that will become a portion of the outside of the adult insect during the pupal transformation.

However, interestingly, loss of function mutations of PRC2 have also been reported in many cancer cases, exhibiting reduction of H3K27me3 (Conway et al., 2015). In addition, recent sequencing analysis of paediatric glioma samples revealed the somatic mutations of K27M in *HIST1H3B* and *H3F3A*, coding genes of the histone H3.1 and the histone H3.3, respectively (Table 2; Schwartzentruber et al., 2012; Wu et al., 2012). A subsequent systematic study demonstrated that the H3K27M mutation specifically inhibits the enzymatic activity of PRC2 through interaction with EZH2, causing remarkable reduction of H3K27me2 and H3K27me3 (Lewis et al., 2013).

The Bracken group proposed that both up- and down-regulation of PRC2 function would affect redistribution of H3K27me3 in the genome: increasing the levels of H3K27me3 in differentiation-associated genes and decreasing the H3K27me3 in cancer-associated genes (Conway et al., 2015).

Gene	Aberration	Cancer type	H3K27 methylation status
PRC2 mutations			
EZH2	Y641X	Lymphoma, parathyroid adenoma, melanoma	elevated H3K27me3, reduced H3K27me2
	A677G	Lymphoma, Ewing sarcoma	elevated H3K27me3, reduced H3K27me2
	A687V	Lymphoma	elevated H3K27me3, reduced H3K27me2
	Homozygous mutation	Leukemia, myeloid disorders	reduced H3K27me3
	Heterozygous	Leukemia, myeloid disorders	reduced H3K27me3
SUZ12	mutation	MDS/MPN, leukemia, MPNST	reduced H3K27me3
	Heterozygous deletion	MPNST	reduced H3K27me3
	Heterozygous deletion/ mutation	MPNST	reduced H3K27me3
	Homozygous deletion	MPNST	reduced H3K27me3
EED	Heterozygous deletion/ mutation	MPNST	reduced H3K27me3
	Heterozygous deletion	MPNST	reduced H3K27me3
	Homozygous deletion	MPNST	reduced H3K27me3
Histone H3 mutations			
H3F3A	K27M mutation	High grade glioma, low grade glioma, leukemia	reduced H3K27me3, reduced H3K27me2
HIST1H3B	K27M mutation	High grade glioma	reduced H3K27me3, reduced H3K27me2

Table 2. Mutations of PRC2 and histone H3 coding gene in different cancer types

MDS, myelodysplastic syndrome; MPNST, malignant peripheral nerve sheath tumours, the table was adapted from (Conway et al., 2015).

8. Polycomblike

Pcl is one of the earliest isolated PcG genes in *Drosophila* (Duncan, 1982). The *Pcl*$^{-/-}$ *Drosophila* dies as of the mature embryo stage and shows the posteriorly directed homeotic transformation (e.g. the abdominal segment 7 (A7) is transformed to the A8, and the A6 to the A7; Figure 10). The transformation phenotype is more severe in the *Pcl*$^{-/-}$ *Dm* embryo lacking maternally-loaded Pcl protein (Pcl$^{maternal\text{-}zygotic\text{-}}$, Pcl$^{m\text{-}z\text{-}}$) than in the Pcl$^{-/-}$ *Dm* embryo having the maternally-deposited Pcl protein (Pcl$^{m+z\text{-}}$) (Breen and Duncan, 1986).

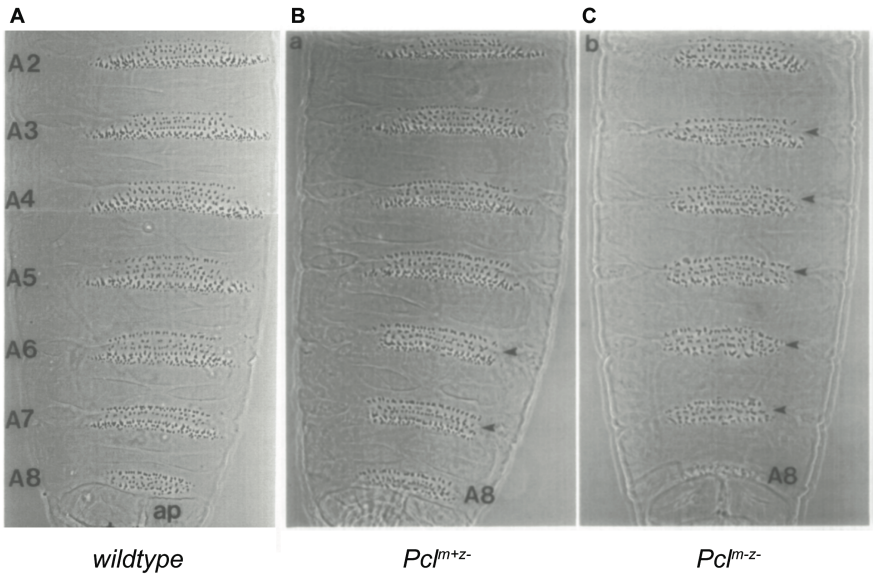


Figure 10. Posteriorly directed homeotic transformation in Pcl mutant embryonic cuticles

Ventral cuticle of (A) a wild-type (Canton-S) first instar larva, (B) a Pcl^{m+z-} embryo and (C) a Pcl^{m-z-} embryo. Arrows indicate transformation of embryonic segments; abbreviations: A2-A8: Abdominal segment 1-8; figures were adapted from (Breen and Duncan, 1986).

Mammalian homologues of Pcl consist of three different paralogues: PHD finger protein 1 (PHF1/Phf1), Metal-response element-binding transcription factor 2 (MTF2/Mtf2) and PHD finger protein 19 (PHF19/Phf19) (Table 1; Figure 11). The Pcl homologues share conserved domain architectures including the Tudor domain and the Plant homeodomains (PHD1 and PHD2) (Figure 11). The Phf1 and the Mtf2 were

first studied in different contexts than the Polycomb repression: spermatogenesis (Ha et al., 1991) and Zn-dependent metallothionein promoter binding (Inouye et al., 1994; Remondelli and Leone, 1997), respectively. Later, bioinformatic analyses identified that these proteins are the mammalian homologues of the *Dm* Pcl (Coulson et al., 1998).

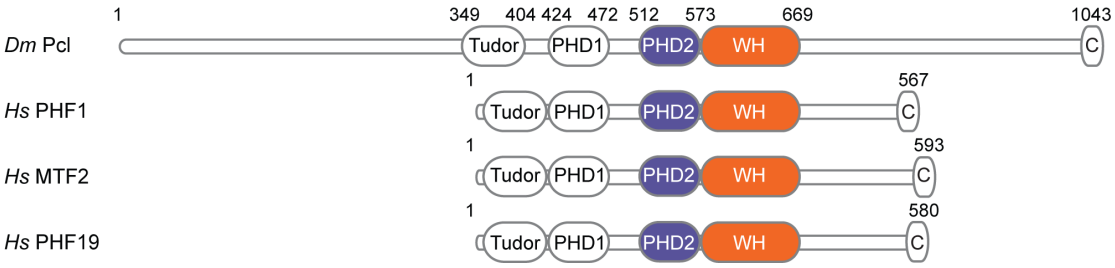


Figure 11. Domain architecture of Pcl homologues

Upper numbers indicate the amino acid residue numbers; abbreviations: WH, winged-helix; C, C-terminal domain.

1) Pcl is associated with PRC2

The Jones group first reported the association of Pcl with the Esc-E(z) complex in the *Dm* embryo nuclear extracts, showing co-migration of the three proteins on the gel filtration (O'Connell, 2001). Later, the Müller group showed that the Pcl forms a stable complex with the PRC2 core subunits by a tandem affinity purification (TAP) strategy. They presented that the TAP-tagged Pcl were eluted with substantial amounts of PRC2 core subunits, whereas the TAP-tagged E(z) with much less amounts of Pcl, indicating that the Pcl is a sub-stoichiometric subunit of PRC2 (Nekrasov et al., 2007).

The Jones group suggested that the Pcl PHD1 and PHD2 mediate the interaction with E(z) by yeast two-hybrid assays. Furthermore, they presented that

this PHD finger mediating interaction is conserved in the human PHF1 with the EZH2 (O'Connell, 2001). However, the Brockdorff group showed that the Mtf2 constructs lacking the Tudor domain, the PHD1, or the PHD2 still form a stable complex with EZH2 by co-immunoprecipitation (co-IP) (Casanova et al., 2011). In addition, the Di Croce group showed that the C-terminal domain of Phf19 is essential to interact with the SUZ12 by co-IP (Ballaré et al., 2012).

2) Pcl is required for PRC2 to generate H3K27me3 efficiently

A few years after the first report of association of the Pcl with PRC2 (O'Connell et al., 2001; Tie et al., 2003), several groups showed that Pcl/PHF1 helps PRC2 generate high levels of H3K27me3 by *in vitro* HMTase assays (Cao et al., 2008; Nekrasov et al., 2007; Sarma et al., 2008). In addition, the levels of H3K27me3 are reduced in Polycomb target genes in $Pcl^{-/-}$ *Dm* embryos (Nekrasov et al., 2007) or Phf1-knock down mouse male germ cells (GC1Spg) (Cao et al., 2008), shown by ChIP experiments. Moreover, antibody staining experiments showed that a subset of Polycomb target genes are misexpressed in the $Pcl^{-/-}$ *Dm* embryo (Nekrasov et al., 2007), indicating that the efficient H3K27me3 by Pcl-PRC2 is crucial for Polycomb repression.

3) Pcl facilitates PRC2 recruitment

Another reported function of the Pcl is to facilitate PRC2 recruitment to target genes (Cao et al., 2008; Nekrasov et al., 2007; Sarma et al., 2008; Casanova et al., 2011; Ballaré et al., 2012; Brien et al., 2012a; Hunkapiller et al., 2012). Upon down-regulation of Pcl/Phf1, the levels of the Su(z)12/Suz12 (PRC2 subunit) enrichment on Polycomb target genes are reduced, whereas the levels of Pho (Pho-RC subunit) or Ph (PRC1 subunit) are unaffected, shown by ChIP analyses in *Dm* embryos and GC1spg-derived Phf1-knockdown cells (Cao et al., 2008; Nekrasov et al., 2007). In addition,

the Brockdorff group showed that the PRC2 recruitment is reduced in Mtf2-knockdown mES cells by ChIP experiments. This group also demonstrated that Ezh2 recruitment to the inactive X chromosome loci is impaired upon down-regulation of the Mtf2 in the 3-day differentiated XX mES cell by immuno-RNA fluorescent in situ hybridisation (FISH) analysis (Casanova et al., 2011). Besides, several articles presented that the bindings of SUZ12 and EZH2 to a subset of target genes in mES cells are diminished upon Phf19 knockdown (Ballaré et al., 2012; Brien et al., 2012a; Hunkapiller et al., 2012). Of note, the Di Croce group performed the vice versa experiment. They showed that Phf19 binding to Polycomb target genes is almost entirely abolished in Eed$^{-/-}$ mES cells (Ballaré et al., 2012), indicating that the Phf19 requires the intact PRC2 to bind to Polycomb target genes.

Whereas many studies have suggested the role of Pcl in helping PRC2 recruitment as mentioned above, one report proposed that the Pcl is the actual recruiter of PRC2. This report showed that the binding of the E(z) to a PRE region is completely abolished in Pcl-knockdown *Dm* larval wing discs (Savla et al., 2008).

4) Mammalian Pcl proteins bind to H3K36me3 via Tudor domains

A number of research groups showed that mammalian PCL proteins (i.e. Phf1, Mtf2 and Phf19) bind to the H3K36me3, whose mark is found in the actively transcribed chromatin (Kouzarides, 2007), via an aromatic cage in their Tudor domains (Ballaré et al., 2012; Brien et al., 2012a; Cai et al., 2013; Musselman et al., 2012). Remarkably, the *Dm* Pcl lacks the aromatic cage in the Tudor domain to bind to the methylated histone tails (Friberg et al., 2010).

The binding of Phf19 to the H3K36me3 is suggested to play a crucial role in the lineage transition in mES cells, which changes the transcription status of the actively

transcribed genes to the repressed status. To be specific, the Phf19 is suggested to recruit the histone demethylase of H3K36me3 to the loci, which allows PRC2 to methylate the H3K27 in turn (Ballaré et al., 2012; Brien et al., 2012; Cai et al., 2013).

Similarly, the Wang group proposed that the PHF1 tudor domain recruits the PRC2 to the actively transcribed genes and increases the levels of H3K27me3 in the HeLa cell (Cai et al., 2013). However, the Kutateladze group suggested that the binding of PHF1 Tudor domain to the H3K36me3 sterically hinders PRC2 to bind the target chromatin, and that this inhibits the H3K27 methylation by PRC2 in the HeLa cell (Musselman et al., 2012).

5) PHF1-specific roles

The Phf1 is highly expressed in testis (Kawakami et al., 1998) and quiescent cells (Brien et al., 2015), whereas the Mtf2 and the Phf19 are highly expressed in embryonic stem cells (Walker et al., 2010; Hunkapiller et al., 2012). Accordingly, the MTF2/Mtf2 and the PHF19/Phf19 seem to have some shared roles, e.g. regulation of stem cell renewal and differentiation (Ballaré et al., 2012; Brien et al., 2012a; Hunkapiller et al., 2012; Walker et al., 2010), whilst the PHF1/Phf1 seems to have distinctive roles.

Firstly, the PHF1 has been suggested to be recruited to the double-stranded-DNA (ds-DNA) break sites, shown by co-IP with the Ku70/Ku80 in the HeLa cell (Hong et al., 2008). A few years later, the Kutateladze group showed that PHF1 recruitment to the ds-DNA break sites is the Tudor domain dependent (Musselman et al., 2012).

Secondly, the PHF1 has been proposed to stabilize the p53, a tumour suppressor, in quiescent cells (Brien et al., 2015). The turnover of p53 is regulated by MDM2-

mediated ubiquitination. Yang et al. showed that the PHF1 binds to the C-terminus of p53 and this binding protects the p53 from the MDM2-mediated ubiquitination and degradation (Yang et al., 2013). Interestingly, the Bracken group showed that the expression of *PHF1* gene is regulated by the p53 in the quiescent cell (Brien et al., 2015).

3. Aim and scope

As described in previous sections, the key function of the Pcl homologues in PRC2, promotion of H3K27 trimethylation by PRC2, is crucial for the Polycomb repression in both *Dm* and mammals (Cao et al., 2008; Nekrasov et al., 2007; Sarma et al., 2008). However, our understanding of the molecular mechanism of this has remained elusive. Thus, I attempted to shed light on the molecular roles of the Pcl in PRC2 (i) by solving the crystal structure of the previously uncharacterized PHD and WH domain in the Pcl, (ii) by demonstrating the DNA binding activity of the Pcl/PHF1 WH domain using biophysical assays, (iii) by mapping the PRC2 binding domain in the PHF1, (iv) by determining the DNA and nucleosome binding activities of various PRC2 complexes (e.g., PRC2, PHF1-PRC2, PHF1$^{WH(E)}$-PRC2) and (v) by comparing histone methyltransferase activities of the various PRC2 complexes *in vitro*.

II. Materials and Methods

1. Materials

Strain	Organism	Source
Top 10 F´	*Escherichia coli*	Invitrogen
BL21-CodonPlus-RIL	*Escherichia coli*	Agilent
DH10EMBacY	*Escherichia coli*	Geneva Biotech
IPLB-Sf21 (Sf21)	*Spodoptera frugiperda*	Invitrogen
BTI-TN-5B1-4 (High Five)	*Trichoplusia ni*	Invitrogen

Table 3. List of cell strains used in this study

Number	Protein name_constructs	Vector	Affinity tag	Clea.	Res.	Species	Specifications	Source
JC10	Pcl_491-694	pEC-K-3C-His	6xHis	3C	Kan.	Dm		this study
JC17	Pcl_419-694	pEC-K-3C-His	6xHis	3C	Kan.	Dm		this study
JC14	Pcl_491-694	pEC-K-3C-His	6xHis	3C	Kan.	Dm	K650A,R651A	this study
JC15	Pcl_491-694	pEC-K-3C-His	6xHis	3C	Kan.	Dm	R631A,Q634A,K637A	this study
JC16	Pcl_491-694	pEC-K-3C-His	6xHis	3C	Kan.	Dm	Q634A,K637A, K650A,R651A,R631A	this study
JC20	Pcl_511-694	pEC-K-3C-His	6xHis	3C	Kan.	Dm		this study
JC28	PHF1_165-363	pEC-K-3C-His	6xHis	3C	Kan.	Hs		this study
JC35	PHF1_165-363	pEC-K-3C-His	6xHis	3C	Kan.	Hs	K323A, K324A	this study
JC212	PHF1	pFastBac	2xStrepII-6xHis	TEV	Amp.	Hs		this study
JC227	PHF1_WH(A)	pFastBac	2xStrepII-6xHis	TEV	Amp.	Hs	K323A, K324A, R304A, S307A, N310A	this study
JC228	PHF1_WH(E)	pFastBac	2xStrepII-6xHis	TEV	Amp.	Hs	K323E, K324E, R304E, S307E, N310E	this study
JC269	PHF1(TEV@432)	pFastBac	2xStrepII-6xHis	TEV	Amp.	Hs	TEV site insertion @ 432	this study
JC232	PHF1_28-end	pFastBac	2xStrepII-6xHis	TEV	Amp.	Hs		this study
JC233	PHF1_1-363	pFastBac	2xStrepII-6xHis	TEV	Amp.	Hs		this study
JC234	PHF1_28-363	pFastBac	2xStrepII-6xHis	TEV	Amp.	Hs		this study
JC238	PHF1_del(364-510)	pFastBac	2xStrepII-6xHis	TEV	Amp.	Hs		this study
JC257	PHF1_515-end	pFastBac	2xStrepII-6xHis	TEV	Amp.	Hs		this study
JC297	PHF1_186-end	pFastBac	2xStrepII-6xHis	TEV	Amp.	Hs		this study
JC213	PHF19	pFastBac	2xStrepII-6xHis	TEV	Amp.	Hs		this study
JC241	PHF19_38-end	pFastBac	2xStrepII-6xHis	TEV	Amp.	Hs		this study
JC242	PHF19_1-375	pFastBac	2xStrepII-6xHis	TEV	Amp.	Hs		this study
JC243	PHF19_38-375	pFastBac	2xStrepII-6xHis	TEV	Amp.	Hs		this study
JC244	PHF19_del(378-529)	pFastBac	2xStrepII-6xHis	TEV	Amp.	Hs		this study
JC258	PHF19_529-end	pFastBac	2xStrepII-6xHis	TEV	Amp.	Hs		this study
EZH2	EZH2	pFastBac	6xHis	TEV	Amp.	Hs		C.Müller lab
JC235	EZH2	pFastBac	2xStrepII-6xHis	TEV	Amp.	Hs		this study
JC303	EZH2_107-129_259-end	pFastBac	6xHis	TEV	Amp.	Hs		this study
JC304	EZH2_107-160_259-end	pFastBac	6xHis	TEV	Amp.	Hs		this study
SUZ12	SUZ12	pFastBac	6xHis	TEV	Amp.	Hs		C.Müller lab
JC286	SUZ12_71-end	pFastBac	6xHis	TEV	Amp.	Hs		this study
EED	EED	pFastBac	6xHis	TEV	Amp.	Hs		C.Müller lab
RBBP4	RBBP4	pFastBac	6xHis	TEV	Amp.	Hs		C.Müller lab

Table 4. List of plasmids used in this study

The numbers in the second column indicate the residue numbers of the constructs. When no number is described in the second column, the constructs encode full-length protein. Abbreviations: Clea., cleavage site; Res., antibiotics resistance; Kan., Kanamycin; Amp., Ampicillin.

name	DNA sequence (Forward, 5´-3´)
30bp DNA	TCCCTCTCTCCGCAGTCGCGGCGCAGTCGC
PRE01	CGTGCGTAAGAGCGAGATAC
PRE02	AGCGAGATACAGATAAGACT
PRE03	AGATAAGACTACGCGCACCA
PRE04	ACGCGCACCATAATGGCTGC
PRE05	TAATGGCTGCGCCGTAAAGC
PRE06	GCCGTAAAGCGAGAGCGATC
PRE07	GAGAGCGATCCGAGCGAGAA
PRE08	CGAGCGAGAAGGCTAACCGT
PRE09	GGCTAACCGTATCTCTCCCT
PRE10	ATCTCTCCCTCTCTCCGCAG
PRE11	CTCTCCGCAGTCGCGGCGCA
PRE12	TCGCGGCGCAGTCGCTGCCT
PRE13	GTCGCTGCCTCTGCAGCTCC
PRE14	CTGCAGCTCCGTCGCCATAA
PRE15	GTCGCCATAACTGTCGTTCG
PRE16	CTGTCGTTCGTAATGGCCGT
PRE17	TAATGGCCGTTTTAAGTGCG
PRE18	TTTAAGTGCGACTGAGATGG
PRE19	ACTGAGATGGCCTCATAATC
PRE20	CCTCATAATCGTTTGCTGAA
PRE21	GTTTGCTGAATCTGAATGGT

Table 5. List of fluorescein (Flc)-labelled oligo nucleotides used in the DNA binding assays

Number	Name	DNA sequence (Forward, 5´-3´)
YN 281	601_215_Fwd	ATATCTCGGGCTTATGTGATGGAC
YN 282	601_215_Rev	ATATCCCGAGTCGCTGTTCAATAC
DB 85	601_215_Fwd-flc	flc-ATATCTCGGGCTTATGTGATGGAC
YN 285	601_147_Fwd	CTGGAGAATCCCGGTGCC
YN 286	601_147_Rev	ACAGGATGTATATATCTGACACGTGCC
DB 87	601_147_Fwd-flc	flc-CTGGAGAATCCCGGTGCC

Table 6. List of oligo nucleotides used for PCR amplification of the nucleosomal 601 DNA

Antibody	Dilutions	Host	Source
H3K27me1	1:6000	rabbit	Milipore (07-448)
H3K27me3	1:1000	rabbit	Milipore (07-449)
H4	1:200000	rabbit	Abcam (ab 10156)
Pcl (PHD-WH)	1:3000	rabbit	this study
Nurf55	1:50000	rabbit	C. Müller lab
HRP	1:5000	donkey	GE Healthcare (NA934)

Table 7. List of antibodies used in this study

Media	Application	Description	Source
L-Broth (LB)	cloning with E.coli	-	In-house
Terrific broth (TB)	expression with E.coli	supplemented with 10% Phosphate buffer	In-house
SOC	cloning with E.coli	-	In-house
TiterHigh	virus generation with Sf21 cells	-	Sigma
Express Five	protein expression with High five cells	supplemented with L-Glutamine	Gibco

Table 8. List of growth media used in this study

Name	Composition	Application
Phosphate buffer	0.17M KH2PO4, 0.72M K2HPO4H2O	supplement for TB media
Phosphate-buffered saline (PBS)	10mM Na2HPO4, 1.8mM KH2PO4, 137mMNaCl, 2.7mM KCl	insect cell harvest, TBS
20X TBE	1M Tris, 0.89M Boric acid, 20mM EDTA pH 8.0	agarose gel electrophoresis (used in 0.4X)
10X Lämmli	0.25M Tris, 1.92M Glycine, 1% SDS	acrylamide gel electrophoresis (used in 1X)
1X transfer buffer	25mM Tris, 192mM Glycine, 0.05% SDS, 10% Methanol	wet transfer
TBS	PBS, 0.2% Tween 20	western blot

Table 9. List of general buffers used in this study

2. Methods

1. General methods

1) Molecular cloning

i. Polymerase Chain Reaction (PCR)

10-30 ng template DNA, 0.2 μmol forward and reverse primers (Table 5 and Table 6), and Phusion polymerase 2 x master mix (Finnzymes) were used in a 50 μl reaction volume with a PCR program described in Table 10. Mutagenesis PCR was performed with the same procedure but with a 10 min extension time.

Step	Temperature	Time (min)	No. of cycles
1	98 °C	5	1
2	98 °C	0.5	25 (repeat the step 2-3)
3	72 °C	0.5 per kb template DNA	
4	72 °C	5	1

Table 10. PCR program

A part of the PCR products were loaded to 0.8 to 1.5 % agarose gel (pre-stained with 10000 x GelRed Nucleic acid gel stain, Biotium) to check the specificity of the PCR reaction. Once the specificity was confirmed by a single band in the agarose gel, all the PCR products were treated with DpnI (NEB) at 37 °C for 1 hr to

digest the template DNA. The reaction was then purified using the QIAquick PCR purification kit (Qiagen).

ii. Restriction digestion and ligation

A classical restriction enzyme based cloning strategy was used to generate clones with the pFastBac vectors. The purified PCR products and the pFastBac vectors were digested with appropriate enzymes in their corresponding reaction buffers (Buffer 1 to 4; NEB) for 1 hr at 37 °C. Subsequently, pFastBac vectors were incubated with Calf intestinal alkaline phosphatase (CIP, NEB) to prevent the self-ligation. The digestion reactions of both inserts and vectors were purified by using the QIAquick PCR purification kit (Qiagen). A molar ratio of 3: 1 insert to vector was used for ligation in 20 µl reaction volumes using the Quick ligation kit (NEB).

iii. Ligation independent cloning (LIC)

The LIC system was used to generate clones with the pEC-K-3C-His vectors (a kind gift from the Elena Conti lab). The LIC system uses the T4 DNA polymerase (T4 DNA pol, Novagen), which cleaves the nucleotides until the T4 DNA pol reaches a Thymine (T) by the 3' to 5' exonuclease activity. By incubating the vectors with the T4 DNA pol and dTTP and the inserts with the T4 DNA pol and dATP, 12 to 15 bp overhang sequences are generated. The inserts and vectors are annealed with the 12 to 15 bp overhangs. Nicks are then repaired during transformation in *E.coli*.

All components added for processing of the vectors and inserts are listed in Table 11. The mixture was incubated at room temperature for 30 min, and subsequently at 75 °C for 20 min to inactivate the enzyme. For annealing, 1 µl of the vector processing reaction and 2 µl of the insert processing reaction were incubated

at room temperature for 10 min, and subsequently 1 µl 25 mM EDTA was added at room temperature for 10 min.

Components	amount
Linearised vector	450 ng
T4 DNA Pol. buffer (10X)	3 µl
dTTP (25mM)	3 µl
DTT (100mM)	1.5 µl
T4 DNA Pol. LIC qualified	0.6 µl
ddH2O	to 30 µl

Components	amount
Purified PCR product	600 ng
T4 DNA Pol. buffer (10X)	2 µl
dATP (25mM)	2 µl
DTT (100mM)	1 µl
T4 DNA Pol. LIC qualified	0.4 µl
ddH2O	to 20 µl

Table 11. LIC reaction mix for the vector (left) and the insert (right)

ddH2O, double-distilled water

iv. Vector transformation

2- 5 µl DNA (ligated DNA, mixture of LIC reactions, or mutagenesis PCR product) were added to 50 µl chemically competent cells (Top10F', Invitrogen, Table 3) and incubated on ice for 30 min. Then the cells were transformed by heat shock at 42 °C for 45 sec and subsequent rest on ice for 10min. Next, the cells were recovered in 900µl LB medium at 37 °C for 45 min to 1 hr with shaking at 600 rpm. The cells were then spun down at 5500 rpm for 2min. The cell pellet was resuspended in 100 µl LB medium and spread onto an LB-Agar plate containing the appropriate antibiotic. The plate was incubated overnight at 37 °C.

v. Verification of clones

On the next day of the transformation, two to three colonies were picked and grown in separate tubes with LB medium supplemented with the respective antibiotics. Simultaneously, PCR was performed with picked colonies (colony PCR) to verify insertion of the gene of interest. On the following day, plasmids of the colonies which had been verified by the colony PCR were isolated with the Miniprep

purification kit (Qiagen) and sent for DNA sequencing (core facility, Max Planck Institute of Biochemistry). Sequences were analysed by using SeqManPro (Lasergene 12).

2) SDS-Polyacrylamide gel electrophoresis (SDS-PAGE) of proteins

Samples were heated in the 4 x Lithium dodecyl sulphate (LDS) sample buffer (Invitrogen) supplemented with 400 mM Dithiothreitol (DTT) at 70 °C for 5min. Samples were loaded onto self-casted 12-13 % (for PRC2 complexes) or 15 % (for Pcl constructs) SDS-polyacrylamide gels. Histone methyltransferase reactions were heated at 95 °C for 5 min and loaded onto the NuPAGE Novex 4- 12 % Bis-Tris gradient gels (Invitrogen). Electrophoresis was run in 150 V for 50 min to 90 min.

3) Coomassie Brilliant Blue staining

First of all, the SDS-PAGE gels were rinsed with the double-distilled water (ddH_2O) for 5 min three times. Then the gels were incubated with the PageBlue protein staining solution (Thermo Scientific) for 30 min with gentle agitation. Finally, the gels were destained with ddH_2O for 10 min with gentle agitation.

2. Specific methods used in this study

1) Protein sequence alignment

Protein sequence alignment was performed using Clustal omega (Sievers et al., 2011), and figures were prepared by Jalview2 (Waterhouse et al., 2009).

2) Secondary structure prediction

Secondary structure prediction was performed on the PSIPRED server (Buchan et al., 2013).

3) Generation of Pcl constructs

The genes encoding *Dm* Pcl amino acid (a.a.) residue 419-694, 491-694, 511-694 and *Hs* PHF1 a.a. 165-363 were subcloned into the pEC-LIC-3C-His vector, containing a hexa-histidine tag and a 3C protease cleavage site at the N-terminus. Mutant constructs were cloned by site-directed mutagenesis based on the Quick-Change protocol (Stratagene).

4) Protein expression of Pcl constructs using bacterial system

50 - 100 ng plasmids were used to transform BL21 codon plus (DE3) RIL cells (Agilent, Table 3). Next morning, the colonies were picked and cultured in 50 - 100 ml TB medium (Table 8) at 37 °C, 225 rpm for 4 to 5 hrs. The pre-cultures were then inoculated into 500 ml TB medium supplemented with phosphate buffer (Table 8) and selective antibiotics in tune air flasks (2.5 L in volume), and incubated at 37 °C, 225 rpm. Usually 6-12 flasks (3 to 6L TB medium in total) were used for large-scale expression to produce 20-40 mg purified protein. When the optical density at 600 nm (OD_{600}) of the cell suspension was 0.6, the temperature setting was switched to 18 °C. Around 2 hrs later, when the incubator was at 18 °C, 0.4 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to the culture to induce the protein expression. Usually the OD_{600} was between 1.2 and 1.5 at this point. The cultures were then incubated overnight (16-18 hrs) at 18 °C, 225 rpm. Next morning the induced cells were harvested by spin-down at 5500 rpm for 15 min in the JA-16 rotor in a centrifuge (Avanti J-26XP, Beckman Coulter). The pellets were resuspended with buffer A (20 mM Tris-HCl (pH8.0 at 4 °C) 150 mM NaCl 1 mM 4-(2-Aminoethyl) benzenesulfonyl fluoride hydrochloride (AEBSF) and either used directly for purification or stored at -80 °C after flash freezing with liquid nitrogen.

5) Purification of Pcl constructs

Bacterial cell suspension was lysed by Emulsiflex c-3 (Avestin) and the lysates were centrifuged for 1 hr at 4 °C, 21500 rpm. Subsequently, the supernatants were loaded onto Ni-Nitrilotriacetic acid (NTA) resin (Qiagen) by gravity flow. Unspecific proteins were washed with buffer B (20 mM Tris-HCl (pH8.0 at 4 °C) 1 M NaCl) and buffer A supplemented with 40 mM Imidazole. Then desired proteins were eluted with buffer A supplemented with 250 mM imidazole. The eluents were diluted twice with 20 mM Tris-HCl (pH8.0 at 4 °C) 4 mM DTT, and hexa-His tag was cleaved with 3C protease overnight at 4 °C. Next day, the tag-free proteins were loaded onto three pre-packed 1 ml HiTrap Heparin HP columns (GE Healthcare) connected in series and eluted by gradient of buffer A and buffer B with 2 mM DTT. The eluent did not contain any contaminant bands shown by SDS-PAGE analysis. The eluents subsequently were loaded onto Superdex75/10/300 column equilibrated with 20 mM Tris-HCl (pH8.0 at 4 °C) 150 mM NaCl. Purity of all purified protein constructs was confirmed by total MS.

6) Limited proteolysis

$Pcl_{419-694}$ was incubated with 0.05 mg/ml Trypsin (Sigma) for 45 min and the reaction was stopped by adding AEBSF. The digested protein fragments were separated by Superdex75 (GE Healthcare), equilibrated with 20 mM Tris-HCl (pH8.0 at 4 °C) 150 mM NaCl. Eluted fractions were analysed by total MS, N-term sequencing and SDS-PAGE.

7) Crystallization

1 µl $Pcl_{491-694}$ (10 mg/ml) was mixed with 0.5 µl reservoir (12.5 % polyethylene glycol (PEG) 3350, 50 mM Potassium Phosphate (dibasic) and 2 mM Manganese

Chloride). 1 µl $Pcl_{511-694}$ (3.4 mg/ml) was mixed with 0.5 µl reservoir (22.5 % PEG 6000, 50 mM Potassium Phosphate (dibasic) and 3 mM Manganese Chloride). Both crystals were grown by hanging drop vapour diffusion method at 20 °C. Crystals were flash frozen in the presence of cryo-protectant, 25 % (v/v, in final) ethylene glycol.

8) Data collection and structure determination

Anomalous and native data sets were collected for the $Pcl_{491-694}$ crystal with a PILATUS 6M detector using the Swiss Light Source (SLS) PXI beam line. Phases were calculated by the single wavelength anomalous dispersion (SAD) method at the zinc peak using Phenix-Autosol (Terwilliger et al., 2009). Native data sets were collected for the $Pcl_{511-694}$ crystal with a PILATUS 6M detector using the SLS PXII beam line. The $Pcl_{511-694}$ structure was solved by molecular replacement with Phenix-AutoMR (Adams et al., 2010) using the $Pcl_{491-694}$ structure as a search model. In both cases, I used the program package XDS (Kabsch, 2010) for data processing, Phenix-refine (Afonine et al., 2012) for refinement and COOT (Emsley and Cowtan, 2004) for building atomic models. All structural figures were prepared with PyMol (ver.1.8, Schrödinger) and the superpositions were performed by using COOT SSM superpose (Emsley and Cowtan, 2004).

9) Generation of PHF1 constructs

Human PHF1 cDNA (codon optimized for *S.frugiperda*, GenScript) was subcloned into pFastBac vector with two consecutive strep-II tags, a hexa-histidine tag and a Tobacco Etch Virus (TEV) protease-cleavable site in the N-terminus. PHF1 point mutants ($PHF1^{WH(A)}$ and $PHF1^{WH(E)}$) were cloned by site-directed mutagenesis based on the Quick-Change protocol (Stratagene). The pFastBac plasmids containing cDNA of human PRC2 core subunits with N-terminal hexa-histidine tags were a kind gift from Christopher Müller lab (Kalb et al., 2014).

10) Protein expression of various PRC2 complexes using baculoviral system

i. General insect cell culture

Sf21 cells (Invitrogen, Cat no.12682-019, Table 3) were cultured in EX-CELL®TitherHigh™ medium (Sigma, Table 8) and used for transfection and virus amplification. High Five cells (Invitrogen, P/N 51-4005, Table 3) were cultured in ExpressFive medium (gibco, Table 8) supplemented with 18 mM L-Glutamine (Gibco, Table 8) and used for protein expression. Both cell lines were maintained as 500 mL suspension culture in 3L Erlenmeyer flasks (Corning) with appropriate cell density (0.7-1 x 10^6 cells/ml for Sf21 cells; 0.4-0.8 x 10^6 cells/ml for High Five cells), and kept until approximately passage 30. Cell density, diameter and viability were monitored with a Vi-cell XR cell viability analyzer (Beckman coulter).

ii. Transformation of DH10EMBacY

100-200 ng pFastBac vectors containing the genes of interest were transformed into 100 µl DH10EMBacY (Geneva Biotech, Table 3) cells. All transformation procedures were the same as those for bacterial transformation, except that cells were recovered in SOC medium for 4 hrs after heat shock. The recovered cells were spun down at 1500 x g for 4 min and resuspended in 200 µl LB. The suspension was spread onto two LB plates supplemented with antibiotics (50 µg/ml Kanamycin, 10 µg/ml Tetracycline, 7 µg/ml Gentamicin, 34 µg/ml Chloramphenicol, 100 µg/ml Blue-gal, 40 µg/ml IPTG): one with 20 µl and the other with 160 µl. Then the plates were incubated in 37 °C for 48 hrs. The positive colonies, which gene of the interest is transposed from the pFastBac vector to the bacmid, were distinguished by their white colours. True white colonies were picked for mini-culture in 2.5 ml LB (with 50 µg/ml Kanamycin, 10 µg/ml Tetracycline, 7 µg/ml

Gentamicin, 34 µg/ml Chloramphenicol) for 24 hrs at 37 °C, and re-streaked on a new LB plate (with 50 µg/ml Kanamycin, 10 µg/ml Tetracycline, 7 µg/ml Gentamicin, 34 µg/ml Chloramphenicol, 100 µg/ml Blue-gal, 40 µg/ml IPTG) to confirm their white colour.

iii. Bacmid purification

Confirmed bacmids were purified by alkaline lysis from the mini cultures. First, mini cultures were spun down at 19000 x g for 3 min. Then the pellets were resuspended in 300 µl P1 (resuspension buffer with RNase, QIAGEN). Subsequently, 300 µl P2 (alkaline lysis buffer, QIAGEN) and 300 µl N3 (neutralization buffer, QIAGEN) were added, and the suspension was mixed. Next, the lysates were spun down at 19,000 x g for 10 min at 4 °C. The supernatants were transferred to sterile 2 ml Eppendorf tubes, and centrifuged at 19,000 x g for 5 min at 4 °C. The supernatants were then transferred to sterile 2 ml-Eppendorf tubes and incubated with 800 µl isopropanol for 10 min on ice. Next, the samples were centrifuged at 19,000 x g for 20 min at 4 °C. Subsequently, the pellets were washed with 500 µl 70 % ethanol, and spun down at 19,000 x g for 10 min at room temperature. The pellets were air-dried at 37 °C. Finally, the pellets were resuspended in 100 µl sterile ddH_2O.

iv. Transfection of Sf21 cells

Firstly, 2 ml 0.4 x 10^6 cells/ml Sf21 cell suspension were incubated in each well of a 6-well tissue culture (TC) plate (BD Falcon) for 30-60 min at 27 °C. Then, transfection was performed using the purified bacmid and transfecting reagents, cellfectin II (Invitrogen), including negative controls (transfecting reagents only). After 4 hrs of incubation with the transfecting agents, all liquid in the wells was

removed and 2 ml fresh TiterHigh medium was added to the each well. Then the TC plates were sealed with parafilm and incubated at 27 °C for 72 hrs. Then the supernatants were collected: passage 1 (P1) virus.

v. Virus amplification

50 ml 0.4 x 10^6 cells/ml Sf21 cells were infected with 2ml harvested P1 viruses, and the infected cells were cultured at 27 °C, 90 rpm. The cell density was inspected every 24 hrs, and duplicated cells (cell density of more than 1.0 x 10^6 cells/ml) were diluted with fresh medium to below 1.0 x 10^6 cells/ml. 48 hrs after the cells stopped proliferating, passage 2 (P2) viruses were harvested by centrifugation at 2000 rpm for 15 min. Then 2.5 ml P2 viruses were infected to 250 ml 0.4 x 10^6 cells/ml Sf21 cells, and the cells were cultured for in suspension. 72 hrs after the P2 infection, the supernatant (P3 viruses) were harvested. The P2 and P3 viruses were kept at 4 °C (protected from the light). Subsequently, various amounts of P3 viruses were used to infect 3 ml High Five cells at a density of 0.4 x 10^6 cells/ml in order to test the titre.

vi. Protein expression

In this study, the High Five cells were co-infected by four to five single viruses. Thus, prior to expression, a number of small-scale expression tests were performed to determine the optimal ratio of the mixture of the different P3 viruses. In order to obtain 10-40 mg purified PRC2 complexes, 4 to 8 L High Five cells (0.8 x 10^6 cells/ml) were infected with the optimal ratio of each virus. The infected cells were cultured for 50-72 hrs and harvested by centrifugation at 2000 rpm for 15 min. The pellets were resuspended in PBS and spun down for 15 min at 2000 rpm to wash the residual growth medium. Pellets were frozen with liquid nitrogen or used directly for protein purification.

11) Purification of various PRC2 complexes

The harvested cells were then lysed by a glass homogenizer in a lysis buffer (50 mM Phosphate-sodium pH8.0, 150 mM NaCl, 20 mM Imidazole, 4 % glycerol, 2.5 mM MgCl2, 0.05 % Nonidet-P40 (NP40), 1 mM Dithiothreitol (DTT), 1 mM AEBSF and cOmplete EDTA free protease inhibitor cocktail tablets (Roche)). The cleared lysates were loaded onto Ni-NTA resin (Qiagen) by gravity flow. The unspecific proteins were washed with a wash buffer (50mM Phosphate-sodium pH8.0, 500 mM NaCl, 50 mM imidazole and 4 % glycerol), and the desired proteins were eluted with a His-elution buffer (50mM Phosphate-sodium pH8.0, 150 mM NaCl, 250 mM imidazole and 4 % glycerol). The eluents were subsequently loaded onto strep tactin sepharose (IBA) and eluted with a strep-elution buffer (25 mM HEPES pH7.9, 150 mM NaCl and 10 mM Desthiobiotin (Sigma)). SDS-PAGE analysis showed that the complexes are highly pure and stoichiometric after this purification step. The complexes were then incubated with TEV protease (Core facility of Max Planck Institute of Biochemistry) and Lambda protein phosphatase (Core facility of Max Plnack Institute of Biochemistry) overnight at 4 °C. The tag-free protein complexes were purified over MonoQ resins by salt gradient in 25 mM HEPES pH7.9, and finally polished by Superose6 resins (GE healthcare) equilibrated with a buffer containing 25 mM HEPES pH7.9 and 100 mM NaCl. Purity of all purified protein constructs was confirmed by using total MS.

12) Preparation of Flc-labelled DNA

Various 5'-Flc-labelled single stranded oligos (Table 5) and their complementary non-labelled oligos (Sigma) were mixed with their complementary non-labelled oligos (Sigma) and incubated for 10 min at 80 °C. The mixture was then slowly cooled down for annealing. The annealed oligos were verified by 18 % TBE acrylamide gel electrophoresis.

Flc-labelled 147 bp 601 DNA (Lowary and Widom, 1998) was generated by PCR using the 5' Flc-labelled forward primer and the non-labelled reverse primer (Table 6) to reconstitute Flc-labelled nucleosomes used in native gel shift assays.

13) Nucleosome reconstitution

The 147 bp and 215 bp DNA harbouring 601 sequences (Lowary and Widom, 1998) were amplified by PCR using corresponding primer sets (Table 6) and purified over a 1 ml HiTrapQ column (GE Healthcare) by salt gradient in 25 mM HEPES pH 7.9. The amplified 601 DNA (either 147 bp or 215 bp) was eluted around a conductivity of 67 % (around 700 mM NaCl; Figure 12A). The purity was verified by 1.2 % agarose gel electrophoresis (Figure 12A) and concentrated up to 1 to 2 mg/ml by ethanol precipitation. Histone octamer refolding (Figure 12B) and nucleosome reconstitution (Figure 12C) were performed as described previously (Luger et al., 1999). After the reconstitution, the absence of uncoupled DNA of the reconstituted nucleosome was verified by 1.2 % native agarose gel electrophoresis (Figure 12C).

14) Fluorescence Polarization (FP)

45 nM 30bp Flc-labelled ds DNA oligos were incubated with increasing amounts of proteins in a buffer C (25 mM HEPES-NaOH pH7.9, 50 mM NaCl, 0.05 % tween-20, 5 mM MgCl2 and 4 % Glycerol) for 5 min at 20 °C. The fluorescence was measured by a Synergy H1 plate reader (excitation wavelength at 485 nm and emission at 528 nm; BioTek), and the polarization was monitored by Gen5 2.05 software. All measurements were performed in triplicates. Hill function fitting and graphical representation of the triplicate measurements were carried out using Origin 9.0 software (OriginLab, Northampton, MA).

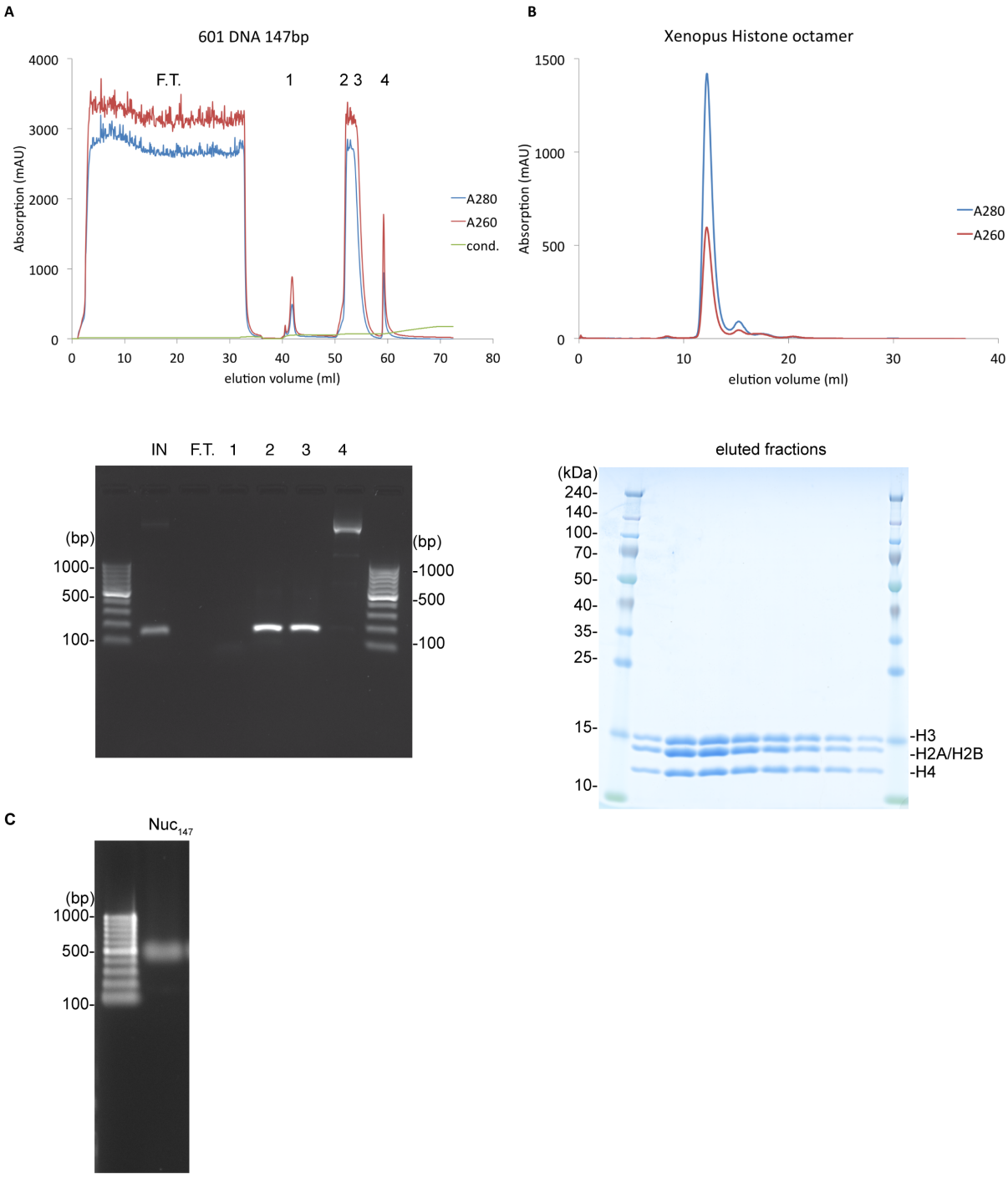


Figure 12. Nucleosome reconstitution

(A) Top: HiTrapQ elution profile of 601 DNA PCR products; Bottom: a 1.2 % agarose gel containing the input (IN), flow-through (F.T.) and the eluents (1-4) of the HiTrapQ purification.
(B) Top: a gel filtration profile of refolded histone octamer on Superdex200 increase; Bottom: a 14 % polyacrylamide-SDS gel containing the eluents of the gel filtration.
(C) A 1.2 % native agarose gel containing reconstituted nucleosomes (Nuc_{147}).

15) Microscale thermophoresis (MST)

45 nM 30bp Flc-labelled ds-DNA oligos were incubated with increasing amounts of proteins in the buffer C for 5 min on ice. The samples were loaded into standard capillaries for the DNA binding measurements. Thermophoresis was measured with 30 % LED and 50 % MST power by a Monolith NT.115 (Nanotemper). All measurements were performed in triplicates. Hill function fitting and graphical representation of the triplicate measurements were carried out using Origin 9.0 software (OriginLab, Northampton, MA).

16) Native gel shift assay

45 nM Flc-labelled oligos and nucleosomes were incubated with increasing amounts of proteins in buffer C for 5 min on ice. 10 µl of the sample were loaded into 1.2 % agarose (Seakem) gel. Gel electrophoresis was performed with cold 0.4 x TBE buffer (4°C) at 60 V for 45 min. Fluorescence signals were acquired by a Typhoon FLA 9500 (GE Healthcare) using the Cy2 filter. Each native gel analysis presented here was performed in triplicates. Densitometry was performed with Fiji software (Schindelin et al., 2012). Hill function fitting and graphical representation were carried out using Origin 9.0 software (OriginLab, Northampton, MA).

17) Histone methyltransferase (HMTase) assay

Indicated amounts of various versions of PRC2 complexes were incubated with 446 nM mononucleosomes in a reaction buffer (20 mM HEPES pH7.4, 50 mM NaCl, 2.5 mM MgCl2, 5 % glycerol, 0.25 mM EDTA, 0.5 mM DTT and 80 µM S-adenosylmethionine (SAM)) in a total volume of 15 µl at 25 °C for indicated durations. 4 x LDS sample buffer (Invitrogen) supplemented with 400 mM

Dithiothreitol (DTT) was added to the reaction and immediately heated at 95 °C for 5 min to terminate the methylation activity.

18) Western blotting of histone methyltransferase reaction samples

The histone methyltransferase reactions were run on a poly-acylamide SDS gel, and then transferred to membranes (Amersham Hybond ECL Nitrocellulose Blotting Membrane, 0.22 μM, GE Healthcare) by wet transfer.

Prior to wet transfer, three Whatman papers (Chromatography paper, 3 mm, Whatman) and one membrane (Amersham Hybond ECL Nitrocellulose Blotting Membrane, 0.22 μM, GE Healthcare) per each polyacrylamide-SDS gel were equilibrated with 1 x transfer buffer (Table 9). Then, a "sandwich" was prepared as follows: one thick sponge, one Whatman paper, the polyacrylamide-SDS gel, the membrane, two Whatman papers, and one thin sponge. The cassette was then inserted to a transfer tank filled with 1 x transfer buffer, with the gel facing on the cathode side and the membrane on the anode side. Wet transfer was performed at 90 V for 10 min and at 60 V for 30 min at 4 °C under mild agitation. Membranes were blocked with TBS supplemented with 4 % BSA (w/v) (4 % BSA TBS; Table 9) at 4 °C for overnight. Next day, the membranes were washed with TBS three times and then incubated with H3K27me3 antibody (milipore, Table 7) diluted in 4 % BSA TBS at 4 °C for 8 hrs. Next, H4 antibody (milipore, Table 7) was added to the membrane and incubated further for 2 hrs at room temperature. Next day, the membrane was washed with TBS three times and incubated with horseradish peroxidase (HRP) conjugated secondary antibody for 1 hr. Finally, the membrane was washed three times with TBS and covered with 1:1 mixture of detection reagent A and B (ECL Select, Amersham). The chemiluminescence was detected by an ImageQuant

LAS4000 (GE Healthcare). Densitometry of the western blot was carried out with Fiji software (Schindelin et al., 2012).

III. RESULTS

1. STRUCTURAL AND FUNCTIONAL ANALYSIS OF PCL$_{PHD2-WH}$

1. Purification and limited proteolysis of Pcl$_{419-694}$

In order to obtain molecular insights into Pcl function in promoting H3K27 methylation by PRC2, I pursued to solve the crystal structure of conserved domains in Pcl. First, I designed a construct spanning from the PHD1 to a conserved domain next to the PHD2 (Pcl$_{419-694}$; Figure 13A). I sub-cloned the Pcl$_{419-694}$ to a vector containing N-terminal hexa-histidine tag cleavable by 3C protease. The Pcl$_{419-694}$ was expressed in *E.coli* and purified over Ni-NTA, Heparin and Superdex 75 resins (Figure 13B). The final gel filtration profile and SDS-PAGE gel analysis demonstrated the homogeneity and high purity of the Pcl$_{419-694}$ (Figure 13C). I tried to crystallize the purified Pcl$_{419-694}$ in myriads of different crystallizing conditions, but the construct was not crystallized.

Thus, I went on to perform limited proteolysis with the purified Pcl$_{419-694}$ to obtain more crystallizable construct. After a number of screenings using different proteases, I found that the Pcl$_{419-694}$ is cleaved into two pieces upon treatment with 0.05 mg/ml trypsin, where the two fragments do not interact with one another (Figure 14A and B). Then total MS and N-term sequencing analyses revealed that the Pcl$_{419-694}$ is cleaved between the two PHD fingers (Figure 14C). Hence, I designed a new construct containing the PHD2 and the conserved domain next to the PHD2, which spans a.a. 491 to 694 (Pcl$_{491-694}$; Figure 15A).

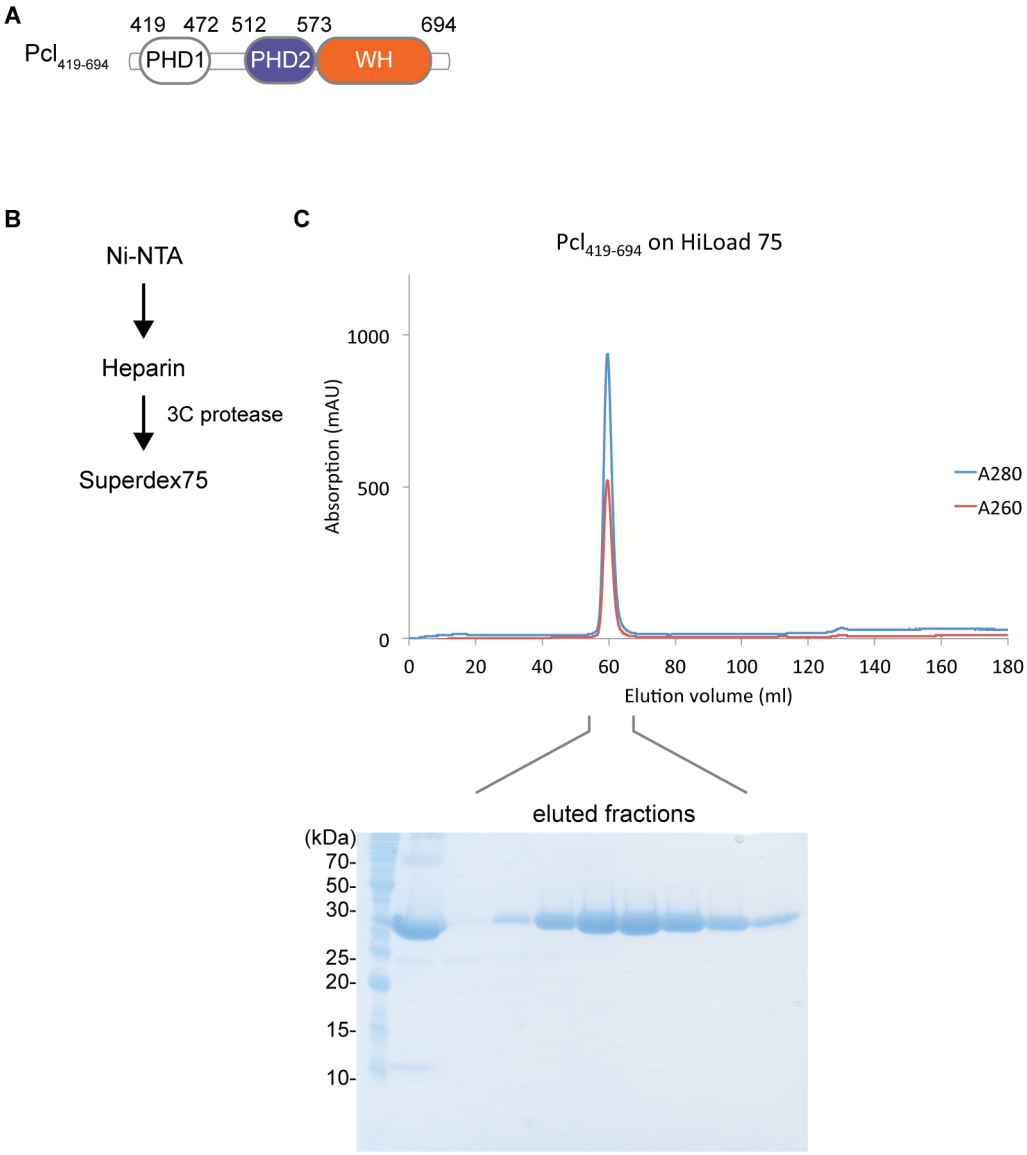


Figure 13. Purification of $Pcl_{419-694}$ construct

(A) Domain composition of $Pcl_{419-694}$ construct. (B) Purification scheme. (C) Top: a gel filtration profile; A280, absorption at 280 nm; A260, absorption at 260 nm; bottom: SDS-PAGE analysis of the eluents.

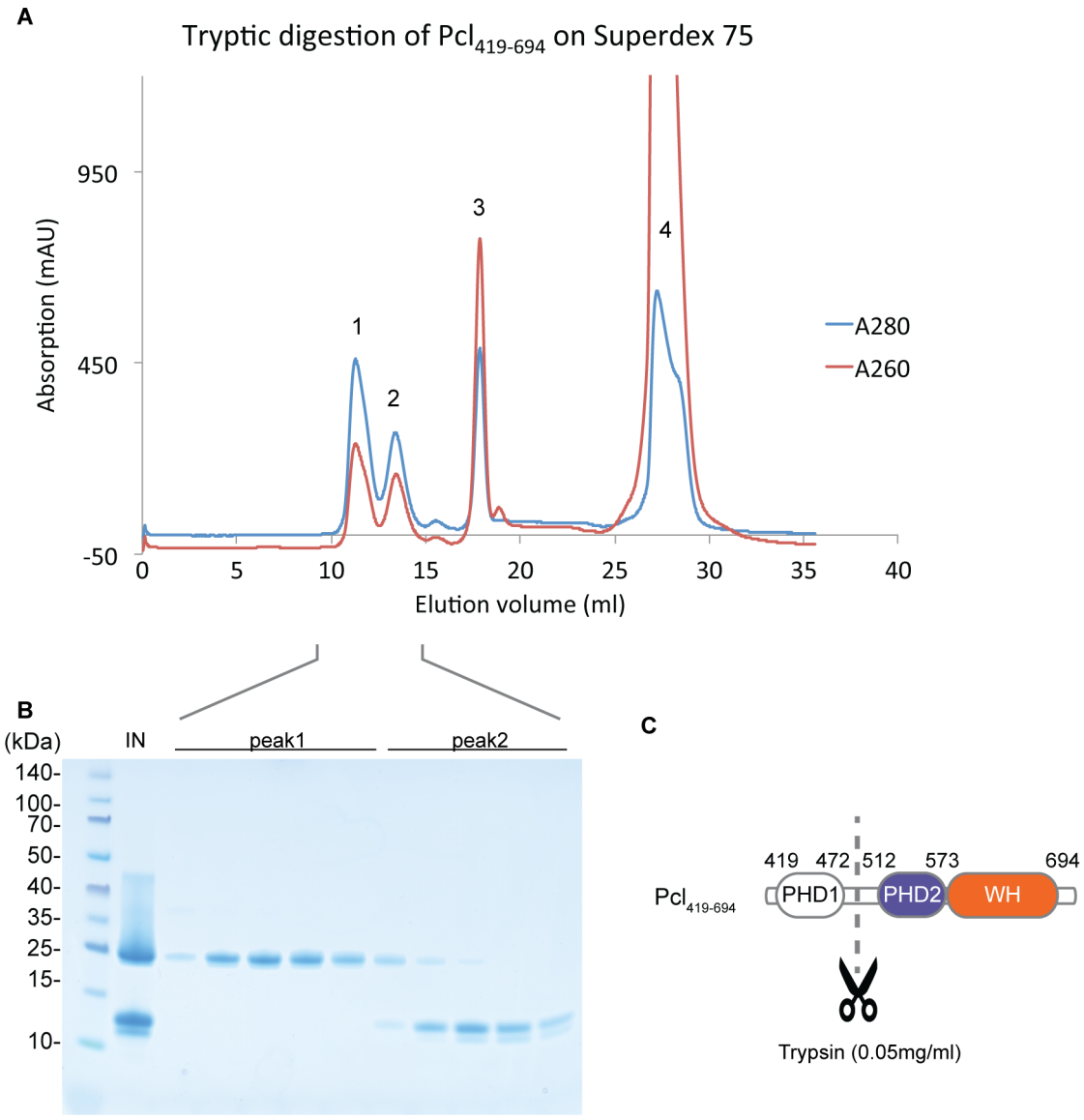


Figure 14. Limited proteolysis of $Pcl_{419-694}$ construct

(A) A gel filtration profile (Superdex 75) of tryptic digested $Pcl_{419-694}$. (B) SDS-PAGE analysis of the input (IN, tryptic digested $Pcl_{419-694}$) and the eluents of peak1 and peak2 of the gel filtration in (A). Peak 3 and 4 did not contain any protein (data not shown). (C) Schematics showing where trypsin digests the $Pcl_{419-694}$ construct.

2. Purification and crystallization of $Pcl_{491-694}$

I cloned the proteolytically stable construct ($Pcl_{491-694}$; Figure 15A) into the same vector as the one used for $Pcl_{419-694}$, and followed the same purification procedure. Final gel filtration profile and the SDS-PAGE analysis demonstrated that the purified $Pcl_{491-694}$ is highly homogenous and pure (Figure 15B and C).

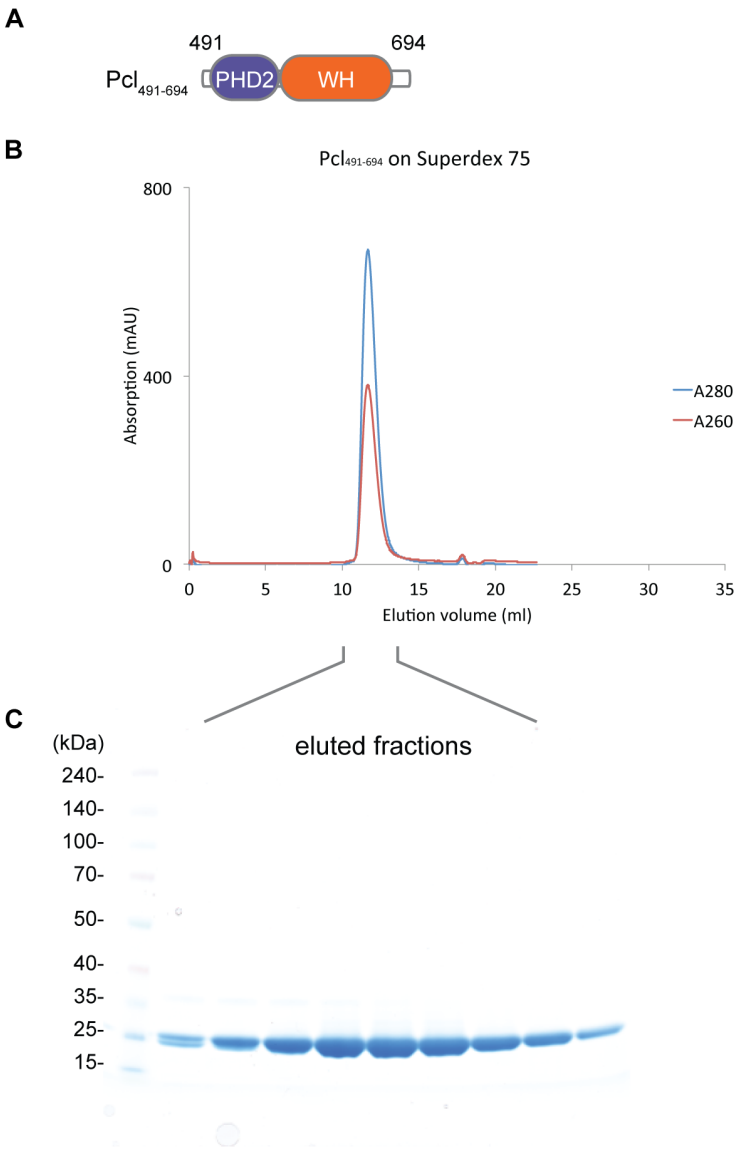


Figure 15. Purification of $Pcl_{491-694}$ construct

(A) Domain composition of $Pcl_{491-694}$. (B) A gel filtration profile of $Pcl_{491-694}$ on Superdex 75. (C) SDS-PAGE analysis of the eluents of gel filtration in (B).

With the purified $Pcl_{491-694}$, I obtained several crystals in a number of conditions containing PEG 3350 in neutral pH. The initial crystals were small in size (Figure 16A), so I attempted to optimize the crystallization by adjusting the precipitant concentrations and adding various additives. I was then able to obtain a single crystal in a reasonably large size (Figure 16B). The optimized $Pcl_{491-694}$ crystal was flash frozen with a cryo-protectant, 25 % ethylene glycol, and mounted (Figure 16C). The optimized $Pcl_{491-694}$ crystal resulted in a diffraction data set with a resolution limit of 2.5 Å (Figure 16D and Table 12).

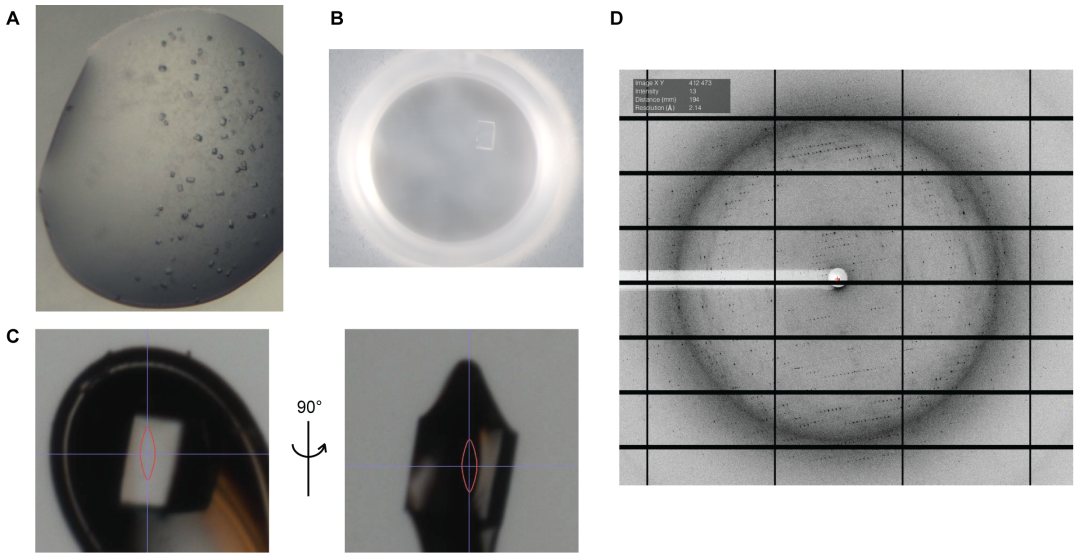


Figure 16. Crystallization of $Pcl_{491-694}$ construct

(A) Initial crystal of $Pcl_{491-694}$ in a 96 well plate, sitting-drop (0.2 µl protein + 0.1 µl reservoir). (B) Optimized crystal of $Pcl_{491-694}$ in a 24 well linbro plate, hanging-drop (1 µl protein + 0.5 µl reservoir). (C) Mounted $Pcl_{491-694}$ crystal in a nylon-loop (left) and its 90° rotation (right); red oval, beam focus. (D) A screen shot showing the diffraction of the $Pcl_{491-694}$ crystal.

*The sections below includes contents from the manuscript prepared for publication.

3. Overall structure of PCL$_{PHD2-WH}$

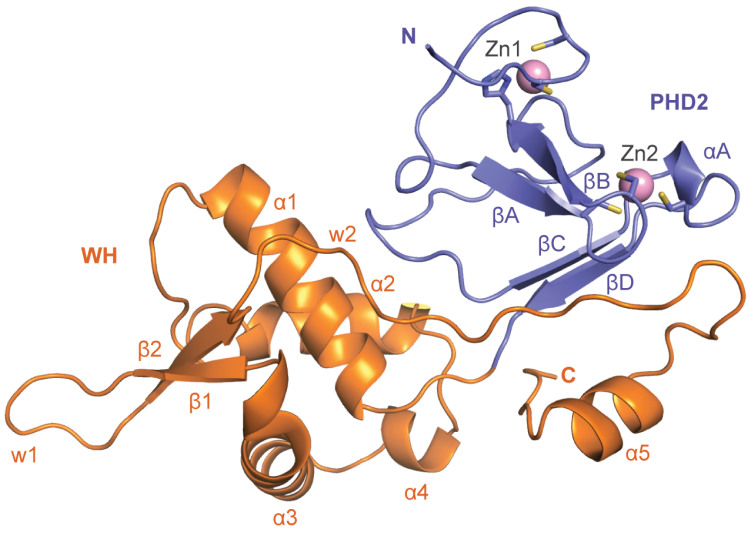
The $Pcl_{491-694}$ crystal structure was solved by Zn-single wavelength anomalous dispersion (SAD) method (Figure 17 and Table 12). Notably, the $Pcl_{491-694}$ crystal structure shows that the conserved domain next to the PHD2 possesses a canonical WH domain fold (Figure 17A), in agreement with bioinformatic predictions (Callebaut and Mornon, 2012; Söding, 2005). The PHD2 and the WH domain stabilize each other by extensive van der Waals and electrostatic contacts along the interface, especially the loop regions between βB and βC, and between β2 and α5 (Figure 17B and C).

Construct		Pcl491-694	Pcl511-694
Data set	native	Peak (Zn)	native
			Data collection
Wavelength (Å)	1.00005	1.28344	0.97935
Space group	C 1 2 1	C 1 2 1	P 1 2_1 1
a, b, c (Å)	307.42 53.12 86.84	307.42 53.12 86.84	94.2, 50.57, 95.51
α, β, γ (°)	90 105.47 90	90 105.47 90	90, 113.1, 90
Resolution (Å)[1]	83.69 - 2.45 (2.53 - 2.45)	83.69 - 2.45 (2.59 - 2.50)	79.15 - 2.3 (2.27 - 2.19)
R-merge	0.045 (0.64)	0.11 (1.33)	0.08 (1.31)
CC1/2	0.997 (0.828)	0.998 (0.696)	0.999 (0.552)
CC*	0.999 (0.952)	1 (0.906)	1 (0.844)
Mean I/sigma(I)	17.09 (2.25)	14.27 (1.39)	12.00 (0.87)
Completeness (%)	98.65 (98.18)	98.06 (91.36)	98.61 (86.43)
Multiplicity	3.4 (3.4)	6.6 (6.6)	6.6 (4.8)
No. Molecule per A.U.	6	6	4
			Refinement
Resolution (Å)	83.69 - 2.447 (2.534 - 2.447)		79.15 - 2.3 (2.266 - 2.188)
Total reflections	167531 (16931)		280326 (17801)
Unique reflections	49847 (4912)		42581 (3688)
R-work	0.2365 (0.3483)		0.2303 (0.4329)
R-free	0.2780 (0.3708)		0.2680 (0.4483)
No. of non-hydrogen atoms	9076		5756
Wilson B-factor	59.74		54.25
RMS(bonds)	0.016		0.004
RMS(angles)	1.41		0.89
Ramachandran favored (%)	98		98
Ramachandran outliers (%)	0		0
Average B-factor	77.1		67

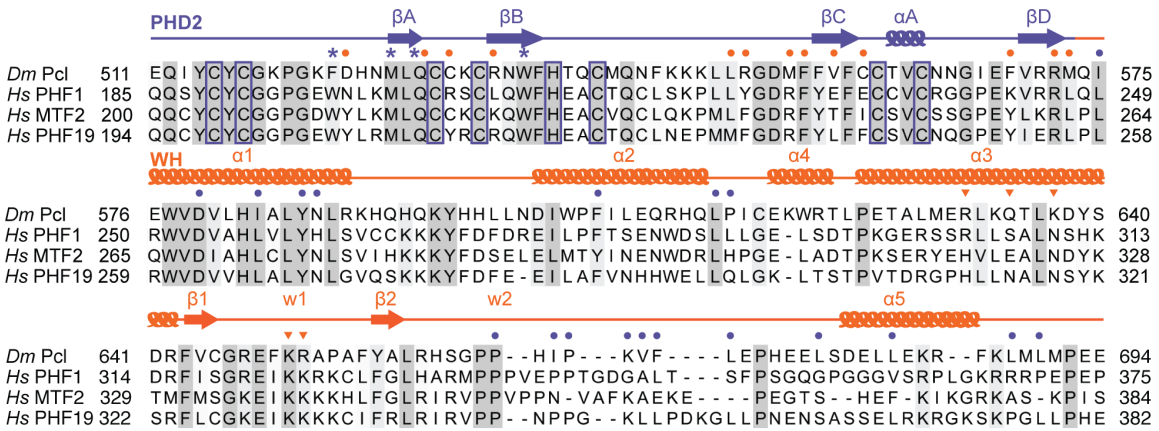
Table 12. Data collection and refinement statistics of $Pcl_{419-694}$ and $Pcl_{491-694}$ crystal

[1] Highest resolution shell shown in parentheses.

A


N
Zn1
PHD2
Zn2
αA
βB
βA
βC
βD
α1
w2
WH
α2
β2
β1
C
w1
α5
α4
α3

B


PHD2 βA βB βC αA βD
Dm Pcl 511 EQIYCYCGKPGKFDHNMLQCCKCRNWFHTQCMQNFKKKLLRGDMFFVFCCTVCNNGIEFVRRMQI 575
Hs PHF1 185 QQSYCYCGGPGEWNLKMLQCRSCLQWFHEACTQCLSKPLLYGDRFYEFECCVCRGGPEKVRRLQL 249
Hs MTF2 200 QQCYCYCGGPGDWYLKMLQCCKCKQWFHEACVQCLQKPMLFGDRFYTFICSVCSSGPEYLKRLPL 264
Hs PHF19 194 QQCYCYCGGPGEWYLRMLQCYRCRQWFHEACTQCLNEPMMFGDRFYLFFCSVCNQGPEYIERLPL 258
WH α1 α2 α4 α3
Dm Pcl 576 EWVDVLHIALYNLRKHQHQKYHHLLNDIWPFILEQRHQLPICEKWRTLPETALMERLKQTLKDYS 640
Hs PHF1 250 RWVDVAHLVLYHLSVCCKKKYFDFDREILPFTSENWDSLLLGE-LSDTPKGERSSRLLSALNSHK 313
Hs MTF2 265 QWVDIAHLCLYNLSVIHKKKYFDSELELMTYINENWDRLHPGE-LADTPKSERYEHVLEALNDYK 328
Hs PHF19 259 RWVDVVHLALYNLGVQSKKKYFDFE-EILAFVNHHWELLQLGK-LTSTPVTDRGPHLLNALNSYK 321
β1 w1 β2 w2 α5
Dm Pcl 641 DRFVCGREFKRAPAFYALRHSGPP--HIP---KVF----LEPHEELSDELLEKR--FKLMLMPEE 694
Hs PHF1 314 DRFISGREIKKRKCLFGLHARMPPPVEPPTGDGALT---SFPSGQGPGGGVSRPLGKRRRPEPEP 375
Hs MTF2 329 TMFMSGKEIKKKKHLFGLRIRVPPVPPN-VAFKAEKE----PEGTS--HEF-KIKGRKAS-KPIS 384
Hs PHF19 322 SRFLCGKEIKKKKCIFRLRIRVPP--NPPG--KLLPDKGLLPNENSASSELRKRGKSKPGLLPHE 382

C

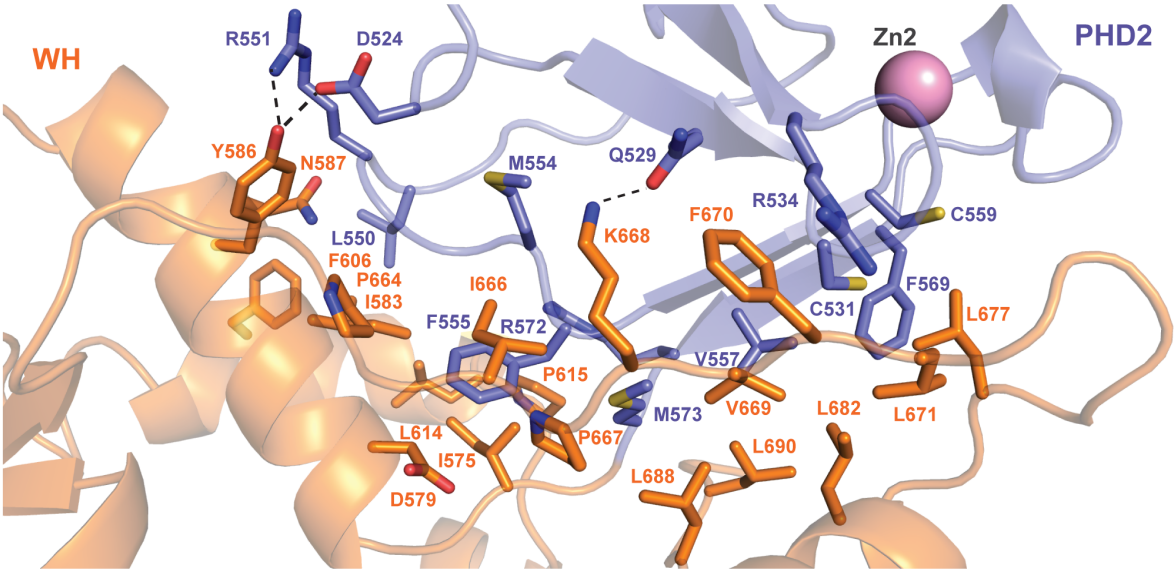

WH
R551
D524
Zn2
PHD2
Y586
N587
M554
Q529
F670
R534
C559
K668
L550
F606
P664
I583
I666
F555
R572
C531
F569
L677
P615
V557
M573
V669
L682
L671
L614
P667
I575
L690
D579
L688

Figure 17. Crystal structure of $Pcl_{PHD2-WH}$

(A) A ribbon view of the overall $Pcl_{491-694}$ crystal structure; PHD2 is coloured slate; WH domain coloured orange; two Zn atoms are represented by spheres and coloured light pink; Zn-coordinating residues in stick representation; secondary structure elements (α: alpha-helix and β: beta-sheet), wing (w), N-terminus (N) and C-terminus (C) of the constructs are labelled. (B) Sequence alignment of PHD2-WH domains in *Dm* and *Hs* Pcl proteins; secondary structural elements based on $Pcl_{491-694}$ crystal structure are described; the numbers on the left- and right ends indicate a.a. residue numbers; blue boxes, Zn-coordinating residues; blue asterisks, residues composing an aromatic cage; blue or orange circles, residues involved in the interaction between PHD2 and WH domain; orange reversed triangles, residues mutated for DNA binding assays (Figure 19). (C) An expanded view of the interface between the PHD and WH domain; salt bridges in dotted lines.

4. Pcl PHD2 contains a closed aromatic cage

The Pcl PHD2 possesses seven Cys and one His that coordinate two Zinc ions (Figure 17A; reviewed in Sanchez and Zhou, 2011). Some PHD finger domains form an aromatic cage, which binds to covalently modified histone peptides such as di- or tri-methylated lysine 4 of histone H3 (H3K4me2/3; Wang et al., 2010). The $Pcl_{PHD2-WH}$ structure shows that the Pcl PHD2 forms an aromatic cage with conserved residues: Phe523, Met527 and Trp536 (Figure 18A and B), the latter corresponding to the highly conserved Trp that forms the sidewall of the cage in most PHD finger domains (Sanchez and Zhou, 2011). However, the Pcl PHD2 aromatic cage is closed by the phenyl-ring of the Phe523 residue (Figure 18B), in contrast to the wide-open cage of MLL1 PHD3 in the absence of its ligand (Figure 18C).

In order to abolish the possibility of crystallization artefacts, I purified and crystallized $Pcl_{511-694}$ construct, which still possesses the PHD2 and the WH domain but lacks the extra loop region in the N-terminus of the $Pcl_{491-694}$ construct. The $Pcl_{511-694}$ was indeed crystallized in a different space group ($P1\ 2_1\ 1$) to $Pcl_{491-694}$ (C 1 2 1; Table 12). However, the aromatic cage of $Pcl_{511-694}$ is still blocked by the phenyl-ring of the Phe523, whereas the loop region of the $Pcl_{511-694}$ PHD2 is deviated from the one of

$Pcl_{491-694}$ (Figure 18D). In addition, the Pcl constructs do not show any binding to a variety of methylated histone peptides in solution binding assays (data not shown). Thus, this structural analyses show that the Pcl PHD2 aromatic cage is closed in the Pcl constructs. Hereafter, all represented structures are the crystal structure of $Pcl_{491-694}$, and I call the $Pcl_{491-694}$ as $Pcl_{PHD2-WH}$.

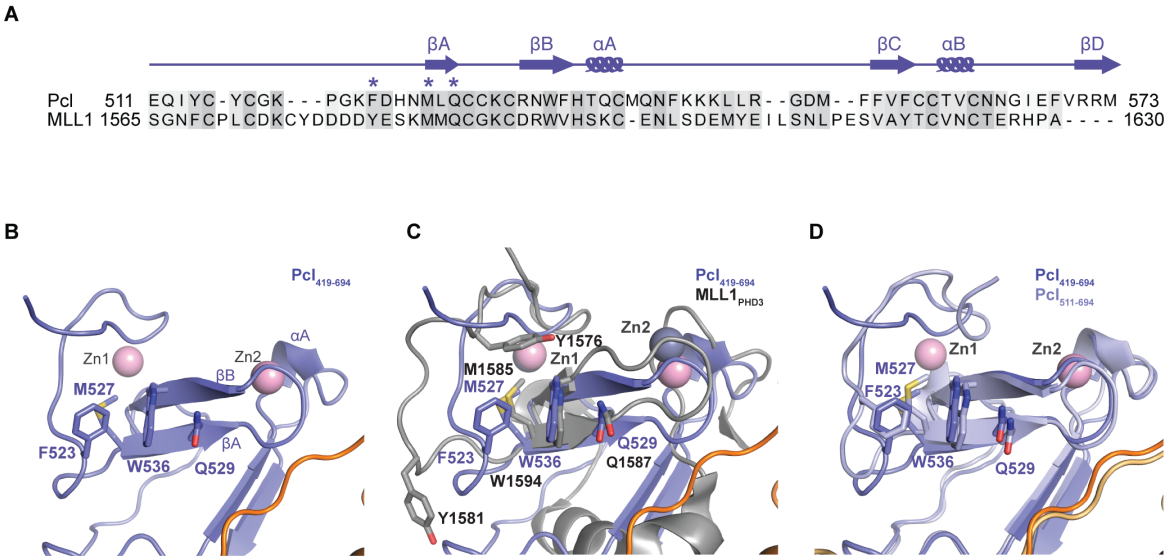


Figure 18. A closed aromatic cage in Pcl PHD2

(A) Sequence alignment of the Pcl PHD2 and the MLL1 PHD3. (B) Expanded view of the $Pcl_{419-694}$ PHD2 aromatic cage. (C) Structural superposition of the $Pcl_{419-694}$ (in slate) onto the MLL1 (in grey; PDB code: 3LQH) PHD aromatic cage. (D) Structural superposition of the $Pcl_{419-694}$ (in slate) onto the $Pcl_{491-694}$ (in light blue) PHD aromatic cage.

5. Pcl binds to DNA via the conserved WH domain in a sequence non-specific manner

WH domains are found in many nucleic acid binding proteins. Although the WH domains retain the overall conserved fold, the DNA recognition modes are diverse (Harami et al., 2013). For example, canonical WH domains such as the one found in FOXO1 recognize their consensus DNA sequence by binding to the DNA major groove via the α3, referred as the recognition helix (Brent et al., 2008). On the

other hand, some other non-canonical WH domains such as the one found in hRFX1 contact mainly DNA phosphate backbones (Gajiwala et al., 2000). In addition, depending on their binding profiles, the detailed structural folds of WH domains vary (Harami et al., 2013).

In order to examine the role of Pcl WH domain in DNA binding, I first superimposed the $Pcl_{PHD2\text{-}WH}$ structure onto a crystal structure of the FOXO1 WH domain that was solved together with its consensus DNA sequences (Brent et al., 2008). This structural comparison proposed that Arg631, Gln634 and Lys637 of the α3 region in the Pcl WH domain might engage in binding to the DNA major groove, and Lys650 and Arg651 of the w1 region to the DNA phosphate backbone contacts (Figure 19A, PDBcode: 3CO6). I next generated and purified three different $Pcl_{PHD2\text{-}WH}$ proteins harbouring point mutations on these residues: $Pcl_{PHD2\text{-}WH}^{R631A/Q634A/K637A}$ (α3mut), $Pcl_{PHD2\text{-}WH}^{K650A/R651A}$ (w1mut) and $Pcl_{PHD2\text{-}WH}^{R631A/Q634A/K637A/K650A/R651A}$ (α3-w1 mut; Figure 19A and B). I then measured the DNA binding activities of the wild-type $Pcl_{PHD2\text{-}WH}$ and the point mutants by using FP assays. The wild-type $Pcl_{PHD2\text{-}WH}$ binds to the DNA probe (Flc-labelled 30 bp-long ds DNA) with a dissociation constant (K_d)[1] of 9.7 μM, whilst the α3 mut and the w1 mut compromise the binding (Figure 19C). Moreover, the α3-w1 mut causes even further reduction in DNA binding activity of $Pcl_{PHD2\text{-}WH}$ (Figure 19C). This DNA binding activity via the WH domain is conserved in human $PHF1_{PHD2\text{-}WH}$ as determined by FP assays (Figure 19D).

[1] Dissociation constant, K_d: a specific type of equilibrium contant that measures the propensity of a larger object to dissociate reversibly into smaller components. The K_d has a dimensions of concentration. In case of event that molecule A and B binds reversibly ($A_xB_y \rightleftharpoons xA + yB$), and the K_d equals the concentration of a free A at which half of the total B are associated with A.

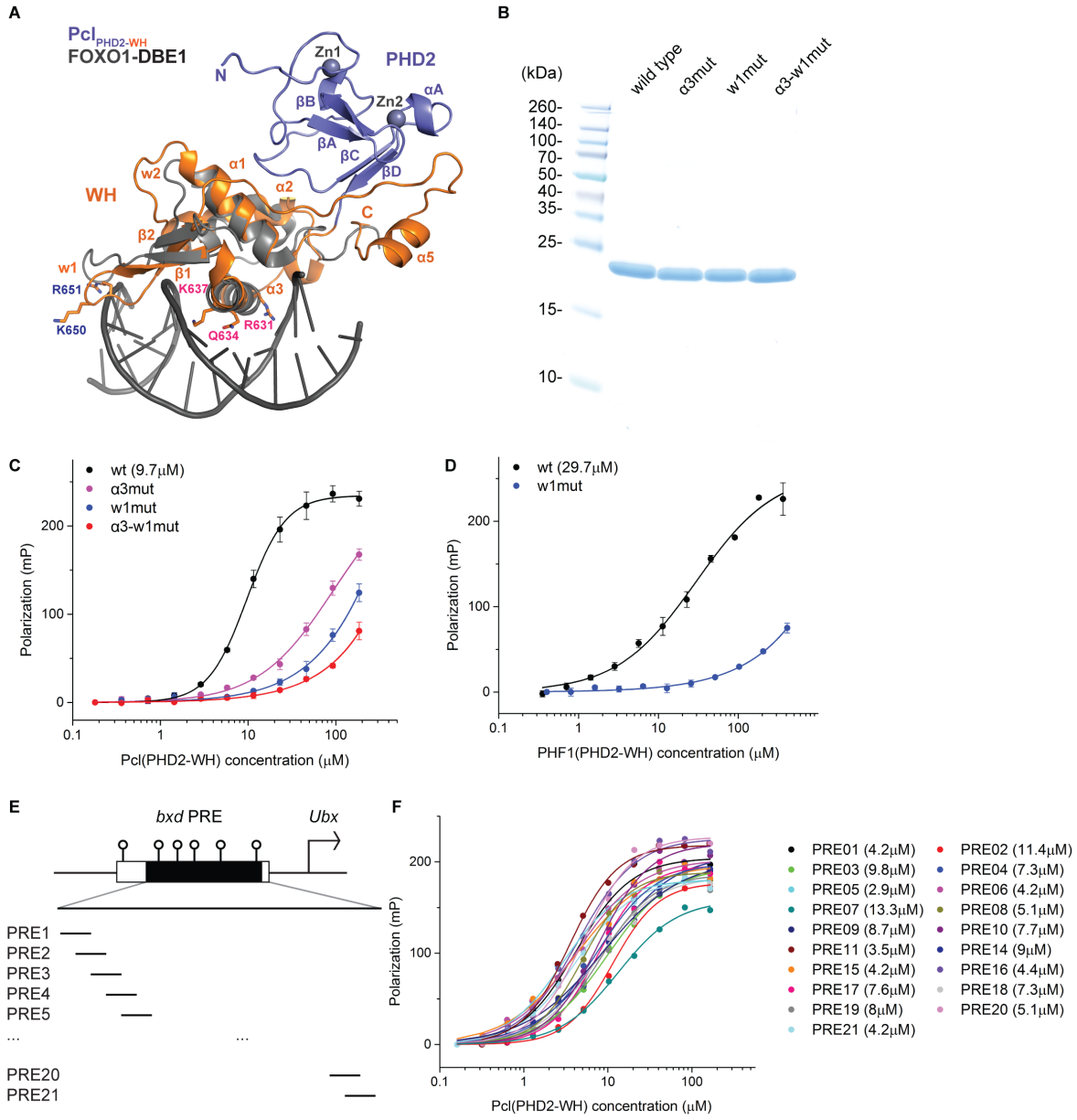


Figure 19. DNA binding activity of the Pcl WH domain

(A) Structural superposition of the $Pcl_{PHD2\text{-}WH}$ (in slate and orange) onto the FOXO1 WH domain-DBE (in grey; PDB code: 3CO6). (B) Coomassie staining of a 15 % polyacrylamide-SDS gel containing wild-type (wt) and point mutant (α3 mut, w1 mut and α3-w1 mut) $Pcl_{PHD2\text{-}WH}$ proteins. (C, D and F) Quantitative measurement of DNA binding of $Pcl_{PHD2\text{-}WH}$ (C and F) or $PHF1_{PHD2\text{-}WH}$ (D) by FP; dots represent the mean values of the triplicate measurements; error bars, standard deviation of the triplicates; curve fitting was preformed by Hill function; K_d in parenthesis. (E) Schematics of *bxd* PRE DNA sequences used in FP measurements shown in (F); six lollipops on *bxd* PRE box represent Pho binding sites. (F) Quantitative measurement of $Pcl_{PHD2\text{-}WH}$ binding to PRE probes.

Pcl has been proposed to facilitate the PRC2 recruitment (Ballaré et al., 2012; Brien et al., 2012b; Casanova et al., 2011; Nekrasov et al., 2007; Savla et al., 2008). I thus questioned whether the Pcl binds DNA in a sequence-specific manner. To test this, I generated a set of Flc-labelled 20 bp-long ds DNA probes spanning a 200 bp interval that forms the core *bxd* Polycomb response element (PRE), where the PcG proteins are specifically localised in *Drosophila* (Papp and Muller, 2006; Figure 19E). I then compared the binding affinities of $Pcl_{PHD2\text{-}WH}$ to these different PRE probes by using FP assays. I observed that the $Pcl_{PHD2\text{-}WH}$ possesses comparable affinities to all the PRE probes with the K_d of between 2.9 and 13.3 μM (Figure 19F), indicating that the $Pcl_{PHD2\text{-}WH}$ binds to the *bxd* PRE in a sequence non-specific manner. In line with this, the solvent exposed residues in α3 are only poorly conserved among the Pcl homologues (Figure 17B). Together, these results show that the Pcl binds to DNA via the conserved WH domain in a sequence non-specific manner.

2. FUNCTIONAL ANALYSIS OF PHF1-PRC2

1. Purification of PHF1-PRC2

Given the structural and biophysical evidence of Pcl/PHF1 in DNA binding via the conserved WH domain, I next asked about which role the Pcl/PHF1 WH domain would play in the context of a full-length Pcl/PHF1-PRC2 complex. To do this, I attempted to purify homogeneous and stoichiometric recombinant PHF1-PRC2 complex. By the previous reports, purifying homogenous Pcl/PHF1-PRC2 seems to be challenging (Cao et al., 2008; Nekrasov et al., 2007; Sarma et al., 2008). First trials to express PHF1-PRC2 in HiFive cells using the DH10Bac strain as the bacmid donor cell showed that the PHF1 protein was degraded in the cells (data not shown). I then tested the expression using a new strain, DH10EMBacY, which the *v-cath* gene that encodes a viral cathepsin-type cysteine protease (V-CATH) is deleted in the baculoviral genome. The deletion of *v-cath* has been reported to improve protein

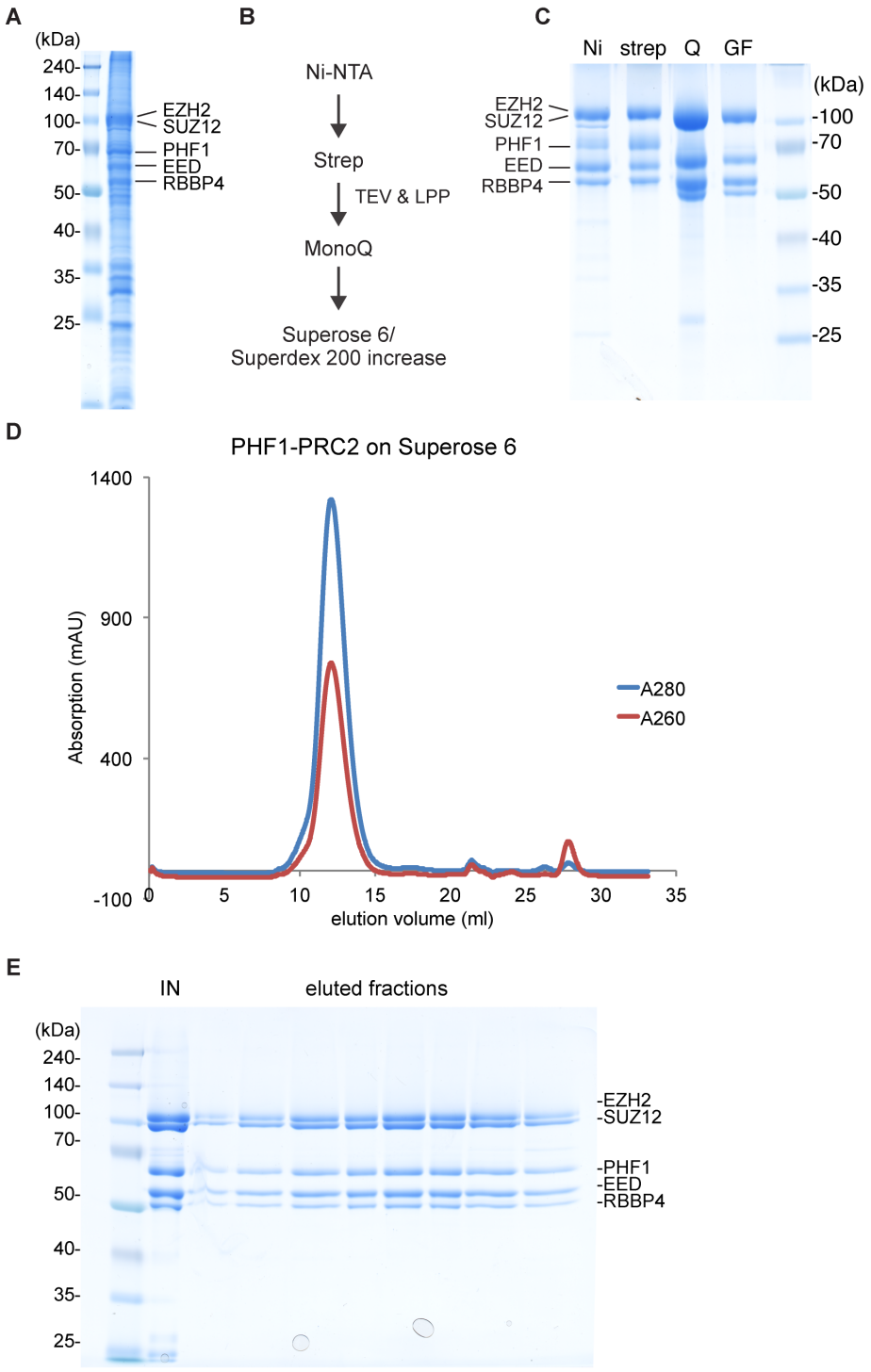


Figure 20. Purification of PHF1-PRC2

(A) Coomassie staining of a 15 % polyacrylamide-SDS gel containing the High Five cell lysates expressing PHF1-PRC2. (B) Purification scheme. (C) Coomassie staining of a 4-12 % gradient polyacrylamide-SDS gel containing eluents from each purification step. (D) A gel filtration profile of PHF1-PRC2 on a Superose 6 column. (E) Coomassie staining of a 12 % polyacrylamide-SDS gel containing eluents from the gel filtration.

expression through reducing virus-dependent proteolytic activity (Berger et al., 2004). Using the DH10EMBacY as the bacmid donor cell enabled me to express stable PHF1-PRC2 complex in insect cells (Figure 20A). I then purified the PHF1-PRC2 with a series of chromatography (Figure 20B and C). Since PHF1 showed lots of phosphorylation, I treated the PHF1-PRC2 eluents with lambda protein phosphatase (LPP). Final gel filtration profile and its corresponding SDS-PAGE analysis verified the homogeneity of the stoichiometric PHF1-PRC2 (Figure 20D and E).

2. PHF1 C-terminal domain is essential for PHF1 to form a stable complex with PRC2

Next, I attempted to map the PRC2 interacting domain(s) in the PHF1. To do this, I generated a number of different constructs of PHF1 (Figure 21A) harbouring a strep tag in the N-terminus, and coexpressed each of them with his-tagged PRC2. I then performed strep-pull down experiments, and discerned the interaction profile of the different PHF1 constructs to PRC2 by SDS-PAGE analysis (Figure 21B).

First of all, the construct without the N-terminal part of PHF1 (PHF1ΔN; Figure 21A) is unstable; no intact PHF1ΔN was eluted from the purification (Figure 21B lane 1). Secondly, the constructs without the C-terminal part (PHF1ΔC and PHF1ΔNΔC; Figure 21A) are stable, but do not interact with PRC2 (Figure 21B lane 2 and 3). Thirdly, the construct spanning from the PHD2 to the C-terminus (PHF1 (PHD2-C); Figure 21A) forms a complex with PRC2 (Figure 21B lane 4), but the construct is highly unstable indicated by the degradation of the PHF1 (PHD2-C) upon the strep-tag cleavage (data not shown). Finally, either having only the conserved C-terminal domain of PHF1 (PHF1(C); Figure 21A) or deleting the loop region between the WH and the C-terminal domains (PHF1Δloop; Figure 21A) gives rise to the stable association of the PHF1 constructs with PRC2 (Figure 21B lane 5 and 6).

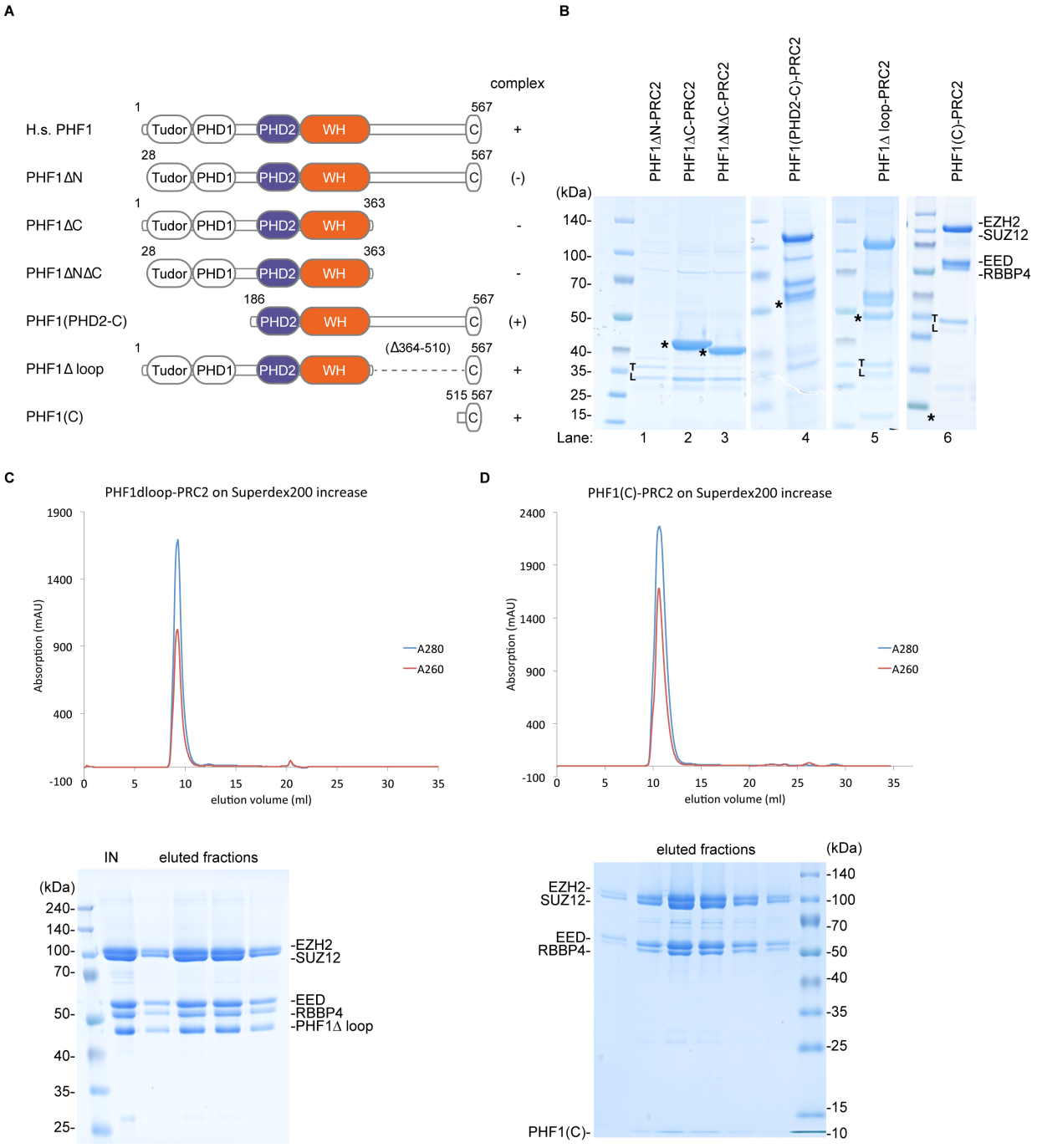


Figure 21. Mapping of PRC2 interacting domain in PHF1

(A) Domain composition of PHF1 constructs (left) and their ability to form a stable complex with PRC2 (right): +, stable complex; (+), form a complex but the PHF1 construct itself is unstable; (-), the PHF1 construct is unstable, potentially does not form a complex; -, no complex formation. (B) SDS-PAGE analysis of various PHF1 constructs-PRC2 pull-downs; T, TEV protease; L, LPP; asterisks, PHF1 constructs. (C and D) Gel filtration profiles (top) and coomassie staining of gels (bottom) of PHF1Δloop-PRC2 (C) and PHF1(C)-PRC2 (D).

After this screening, I further purified the two positive constructs (PHF1Δloop-PRC2 and PHF1(C)-PRC2) by the same procedure used for the purification of PHF1-PRC2 (Figure 20B). Both of the purified PHF1Δloop-PRC2 and PHF1(C)-PRC2 exhibit homogenous and stoichiometric complex indicated by gel filtration and SDS-PAGE analysis (Figure 21C and D).

Next, I performed the domain mapping in the PHF19 by using the same experimental setup used for the PHF1 constructs, and I observed that the PHF19 C-terminal domain is also essential for the PHF19 to be associated with PRC2 (data not shown). Of note, the importance of the PHF19 C-terminal domain in interacting with SUZ12 has been described beforehand (Ballaré et al., 2012).

On the other hand, the Jones group suggested that the Pcl binds to the E(z) through two PHD fingers shown by yeast two hybrid (O'Connell et al., 2001). However, given by the structural analysis of this study, I noticed that the Pcl construct they used in that study spans from the PHD1 to the middle of WH domain ($Pcl_{PHD1\text{-}PHD2\text{-}1/2WH}$), of which structural stability is presumably lacking. Moreover, in order to suggest that the Pcl PHD fingers are important for the interaction with the E(z), they mutated key residues coordinating Zn ions in the Pcl PHD fingers, which might have led to aggregation of the whole $Pcl_{PHD1\text{-}PHD2\text{-}1/2WH}$ construct. Furthermore, when I tried to reproduce their experiments by using a $Pcl_{PHD1\text{-}PHD2\text{-}WH}$ construct and various E(z) constructs, or the human counterparts, I could not obtain a stable complex (data not shown).

Interestingly, DNA sequencing analysis of the *Dm* strain Pcl^{22M21}, which was generated by ethyl methanesulfonate (EMS) mutagenesis and described as an allele of Pcl (Gaytán de Ayala Alonso et al., 2007), showed that the Pcl^{22M21} strain has a mutation in the coding sequence of E701 to a stop codon (Nekrasov et al., 2007). I

thus examined whether the Pcl^{22M21} *Dm* embryo produces $Pcl_{1\text{-}700}$ protein, which would contain all conserved domains but the C-terminal domain, by western blot analysis. The western blot analysis of heterozygous Pcl^{22M21} *Dm* embryo nuclear extracts shows that the truncated Pcl protein ($Pcl_{1\text{-}700}$) is made in the Pcl^{22M21} *Dm* embryo (Figure 22). Therefore, the fact that the Pcl^{22M21} *Dm* shows a null Pcl phenotype is obviously due to the fact that $Pcl_{1\text{-}700}$, without the C-terminal domain, is not be able to be integrated into the PRC2.

Therefore, these results manifest that the conserved C-terminal domains in Pcl homologues are crucial for the Pcl homologues to form a stable complex with PRC2.

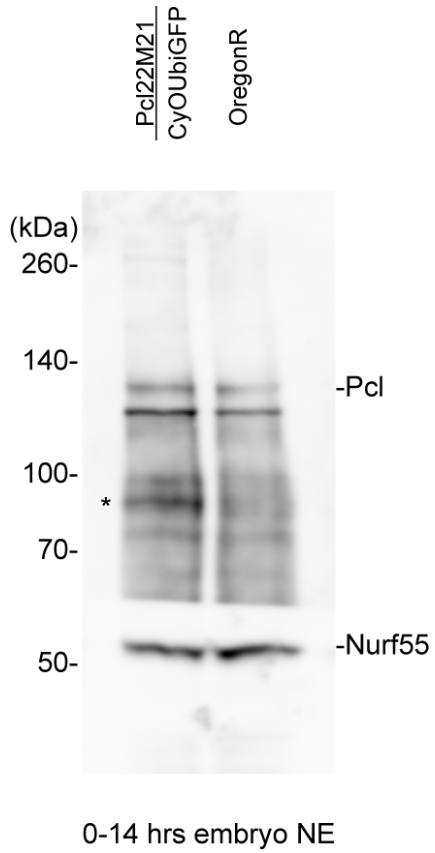


Figure 22. Whole nuclear extracts of heterozygous Pcl and wild-type *Dm* embryos

A western blot analysis of whole nuclear extracts of the heterozygous Pcl (Pcl^{22M21}/CyOUbiGFP) and the wild-type (OregonR) *Dm* embryos (0 – 14 hours (hrs) after egg deposition). The Pcl^{22M21}/CyOubiGFP *Dm* embryos show an extra band around 84kDa (asterisk), which corresponds to the size of $Pcl_{1\text{-}700}$, by Pcl antibody. Nurf55 was probed as a loading control.

3. Additional DNA contact by PHF1 WH domain increases DNA binding of PRC2

In order to examine the role of the DNA binding activity of PHF1 WH domain in the context of the PHF1-PRC2, I carried out microscale thermophoresis (MST) and native gel shift assays using various versions of recombinant PRC2 (i.e. PRC2, PHF1-PRC2 and PHF1$^{R304E/S307E/N310E/K323E/K324E}$-PRC2 (PHF1$^{WH(E)}$-PRC2); Figure 23A). Unexpectedly, I observed that the PRC2 binds to DNA with high affinity shown by the MST (K_d: 218 nM; Figure 23B) and the native gel shift assay (K_d: 76 nM; Figure 23C - E), of which DNA binding activity has not been reported to date. The PHF1-PRC2 shows higher affinity (K_d: 142 nM by MST and 25 nM by gel shift assay; Figure 23B - D) to DNA than the PRC2 does. As expected, the DNA binding affinity of PHF1$^{WH(E)}$-PRC2 is similar to the one of PRC2 (Figure 23B and C). Thus, these data indicate that the additional DNA contact by the PHF1 WH domain enhances the DNA binding of the PRC2.

4. PHF1 WH and C-terminal domains enhance nucleosome binding of PRC2

I next questioned whether PHF1 enhances nucleosome binding of PRC2. The nucleosome binding activity of *Dm* PRC2 has been previously analysed by band shift assays. This earlier report showed that the Su(z)12 and the Nurf55 are the minimal requirement for the PRC2 to bind to the nucleosome, and the Esc increases the nucleosome binding affinity of the PRC2 (Nekrasov et al., 2005). In order to investigate the effect of the PHF1 on the nucleosome binding of PRC2, I first reconstituted Flc-labelled mononucleosomes, ensuring the absence of any uncoupled DNA to exclude the possibility of observing unspecific DNA binding of PRC2. I then measured the nucleosome binding activities of various PRC2 complexes (Figure 23A) by native gel shift assays (Figure 24). The PHF1-PRC2 binds to the nucleosome with the K_d of 51 nM, whereas PRC2 binds the nucleosome with K_d of 138 nM (Figure 24).

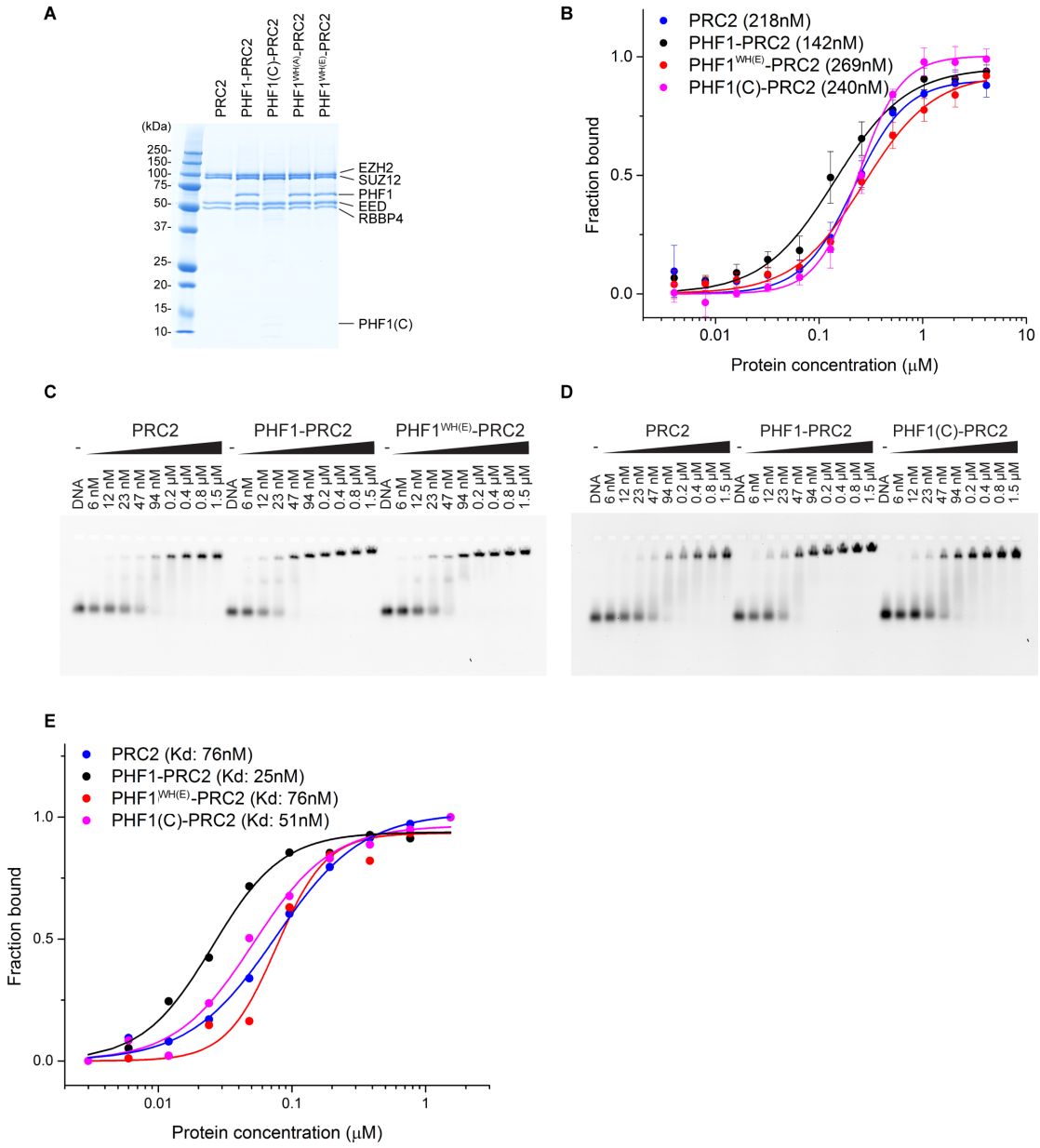


Figure 23. DNA binding activity of various versions of PRC2

(A) Coomassie staining of a 13 % polyacrylamide-SDS gel containing various recombinant PRC2 complexes used in Figure 23 to Figure 26; PHF1$^{WH(A/E)}$: point mutation on K323, K324, R304, S307, N310 to A or E. (B) Quantitative measurements of DNA binding affinities of various PRC2 complexes (from 4 nM to 4 μM) to 45 nM 30 bp Flc-DNA by Microscale thermophoresis (MST); dots represent the mean values of the triplicate measurements; error bars, standard deviation of the triplicates; curve fitting was performed by Hill function; K_d values in parenthesis. (C and D) Native gel shift assays of 30 bp Flc-DNA (45 nM) using various PRC2 complexes. (E) Binding curves were determined by densitometry of fluorescence signals obtained in the gel shift assays (see Figure 23C and D).

Mutation of the DNA binding residues in the WH domain slightly impairs the nucleosome binding (Figure 24), but the reduction is not as significant as the one shown in the DNA binding (Figure 23).

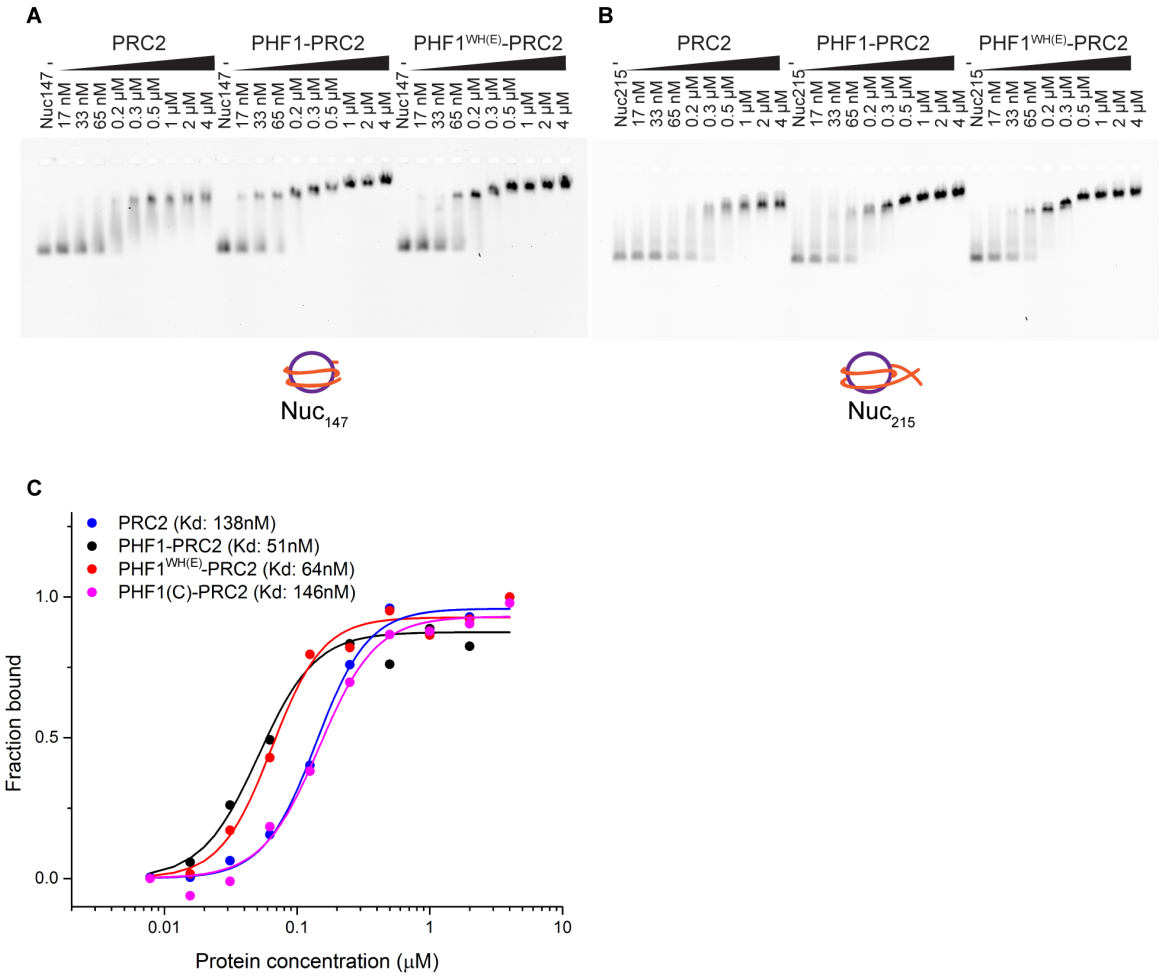


Figure 24. Nucleosome binding activity of various PRC2

Native gel shift assays of 45 nM mononucleosomes reconstituted with Flc-labelled (A) 147 bp 601 DNA (Nuc147) and (B) 215 bp 601 DNA (Nuc215) using various PRC2 complexes (from 17 nM to 4 μM). (C) Binding curves of various PRC2 complexes to Nuc147 were determined by densitometry of fluorescence signals obtained in the gel shift assays (see Figure 24A and Figure 25).

I then queried whether this trait, which the effect of certain binding mutation is alleviated in the context of the nucleosome, could be found in other cases. Recently,

Justin et al. reported that the binding affinity of the histone H3 peptide harbouring the H3K27M mutation to PRC2 (K_d ~0.6 μM) is 20-fold higher than the one of the wild-type H3 peptide to the PRC2 (K_d ~13 μM) (Justin et al., 2016). I thus questioned whether the H3K27M mutation in the nucleosome (H3K27M nucleosome) increases the binding affinity to PRC2. To investigate this, I compared the binding affinity of the recombinant H3K27M nucleosome to the PRC2 with the one of the wild-type nucleosome to PRC2. I observed that the H3K27M mutation in the nucleosome slightly enhances the binding to the PRC2 (Figure 25), but the increase is not as significant as the one shown in the H3K27M peptide. This indicates that the H3K27M in the nucleosome increases the binding affinity to the PRC2 less efficiently than the H3K27M in the peptide does.

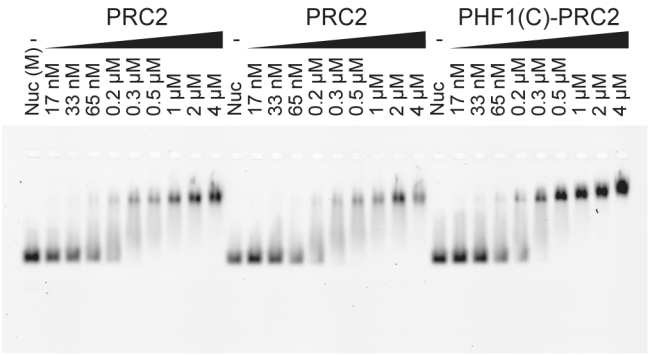


Figure 25. Native gel shift assays of nucleosomes with PRC2 and PHF1(C)-PRC2

Nuc (M): 45 nM H3K27M nucleosomes reconstituted with 147 bp DNA; Nuc: 45 nM wild-type nucleosomes with 147 bp DNA.

In addition, I observed that that the complex of the PRC2 and the nucleosome exhibits smeary bands in the native gel (Figure 24A, left, lanes of the concentration points of 0.3μM onwards), which imply highly dynamic interaction between the PRC2 and the nucleosome. Interestingly, however, the complex of the PHF1-PRC2 and the nucleosome shows defined bands (Figure 24A, middle, lanes of the concentration points of 0.2 μM onwards), which implicate that the interaction

between the PHF1-PRC2 and the nucleosome is stable. Thus, it seems that the PHF1 stabilizes the interaction between the PRC2 and the nucleosome. Intriguingly, however, the complex of the PHF1$^{WH(E)}$-PRC2 and the nucleosome also exhibits defined bands in the native gel (Figure 24A, right, lanes of the concentration points of 0.2μM onwards). This suggests that (an)other domain(s) of the PHF1 might be involved in stabilizing the association of PRC2 with the nucleosome. Since I determined that the PHF1 C-terminal domain is essential to form a stable complex with PRC2 (Figure 21), I investigated the influence of the PHF1 C-terminal domain on the nucleosome binding of PRC2. Interestingly, the complex of the PHF1(C)-PRC2 and the nucleosome also exhibits the defined bands in the native gel (Figure 25), indicating that the PHF1 C-terminal domain is responsible for stabilizing the association of PRC2 with the nucleosome. Additional DNA binding assays show that the PHF1 C-terminal domain does not enhance the DNA binding of PRC2 (Figure 23B and D), indicating that this stabilizing PRC2 binding to the nucleosome by the PHF1 C-terminal domain is the outcome of other properties than the DNA binding activity.

Together, these results strongly suggest that the combinational contribution of the PHF1 WH and the C-terminal domains enhances the nucleosome binding of PRC2.

5. PHF1 WH and C-terminal domains are crucial for efficient H3K27 methylation by PRC2

I next questioned whether the enhancement of nucleosome binding of PRC2 by PHF1 WH and C-terminal domains is critical in promoting the H3K27 tri-methylation by PRC2. In order to examine this, I performed *in vitro* HMTase assays on recombinant mononucleosomes using various versions of PRC2.

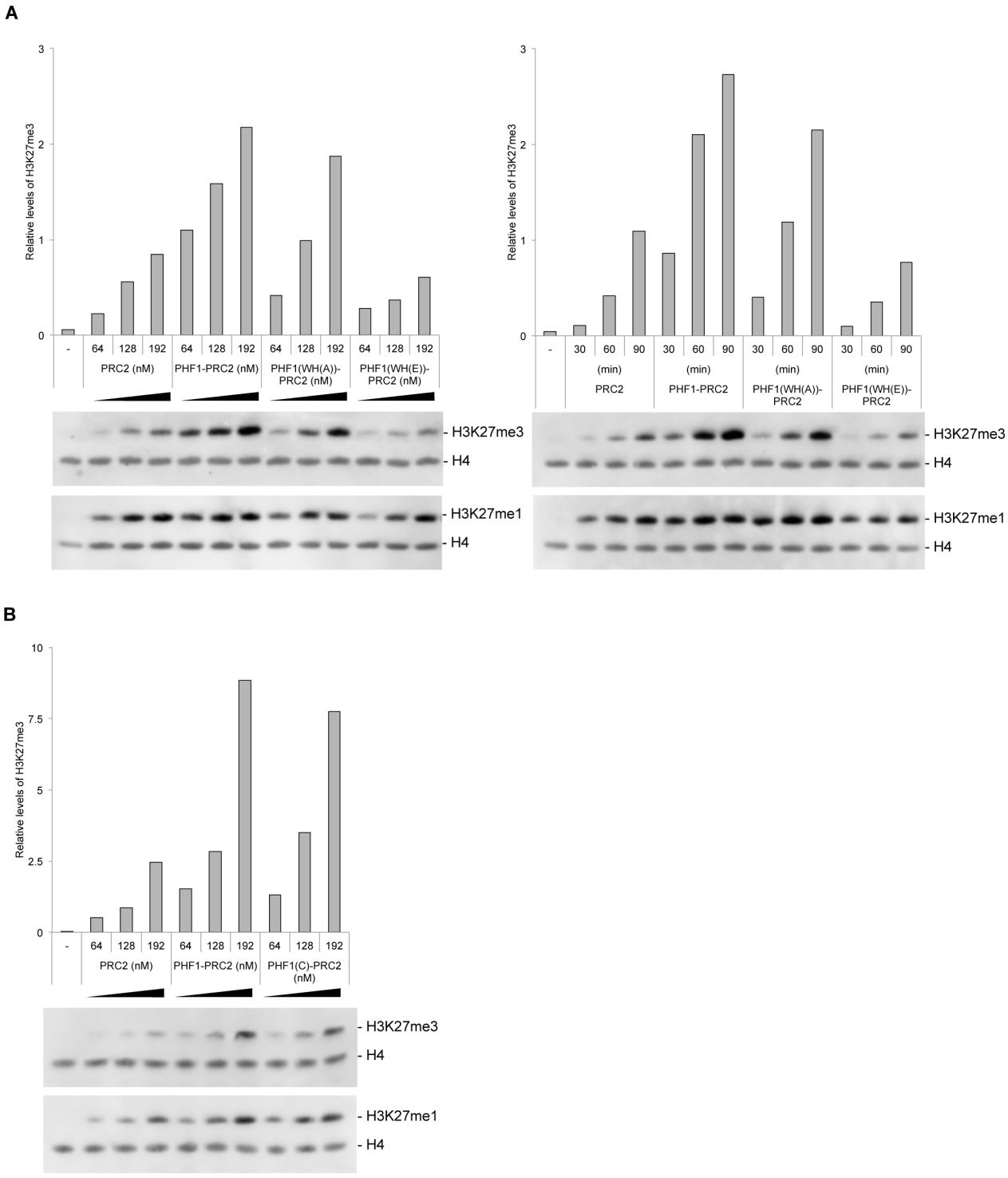


Figure 26. HMTase assays with various PRC2 complexes

(A) HMTase assays on 446 nM mononucleosomes (with 215 bp DNA) of increasing amounts of various PRC2 with fixed incubation time (90 min) (left) or of fixed amounts of the various PRC2 (192 nM) with increasing incubation time (right). (B) HMTase assay of various PRC2 complexes on 446 nM mononucleosomes (215 bp DNA) with fixed incubation time (90 min). (A and B) H4 was probed as a loading control. Histograms represent band intensities of H3K27me3 bands.

Firstly, the PHF1-PRC2 generates H3K27me3 more efficiently than the PRC2 does (Figure 26), in agreement with the previous reports (Cao et al., 2008; Sarma et al., 2008). Further quantification analysis revealed that the PHF1-PRC2 exhibits approximately three-fold higher activity than the PRC2 (Figure 26). Importantly, the point mutation in DNA binding residues to Ala (PHF1$^{WH(A)}$-PRC2) results in decrease of the H3K27me3 levels, and the mutation to Glu (PHF$^{WH(E)}$-PRC2) causes even further reduction of the H3K27me3 levels (Figure 26A). Secondly, the PHF1(C)-PRC2 generates higher levels of H3K27me3 than the PRC2 does (Figure 26B).

Thus, these results strongly support that enhancing the nucleosome binding of PRC2 by PHF1 WH and C-terminal domains is pivotal for the efficient H3K27 methylation by PRC2.

6. PHF1 does not alter the regulatory effect of *in trans* histone modifications on the PRC2 enzymatic activity

H3K27me3 binding *in trans* by EED aromatic cage and EZH2 SRM domain stimulates enzymatic activity of PRC2 (Jiao and Liu, 2015; Justin et al., 2016; Margueron et al., 2009). On the other hand, H3K36me3 binding *in trans* does not affect on PRC2 enzymatic activity. Only the H3K36me3 binding *in cis* (i.e. when the target nucleosome harbours H3K36me3) inhibits the PRC2 enzymatic activity (Schmitges et al., 2011).

In order to assess whether the regulation of the PRC2 HMTase activity by *in trans* histone modification is altered in the presence of the PHF1, I performed HMTase assays with PRC2 and PHF1-PRC2 by adding two different modified histone peptides (i.e. $H3_{19-35}$K27me3 and $H3_{28-43}$K36me3).

The HMTase assay using PRC2 and additional histone peptides shows the consistency with the previous findings (Figure 27). Next, the HMTase assay of PHF1-PRC2 in the presence of the histone peptides demonstrate the same trend as the one of PRC2, i.e., $H3_{19-35}K27me3$ peptides increase the HMTase activity of PHF1-PRC2, and $H3_{28-43}K36me3$ does not affect on the HMTase activity (Figure 28). I suppose that the inhibition observed in the last lane of Figure 28 is not well meaningful, because the concentration of $H3_{28-43}K36me3$ (267 μM) is much higher than the concentration of the substrate nucleosomes (446 nM).

Thus, these results show that PHF1 does not alter the regulatory effect of *in trans* histone modifications on the PRC2 enzymatic activity.

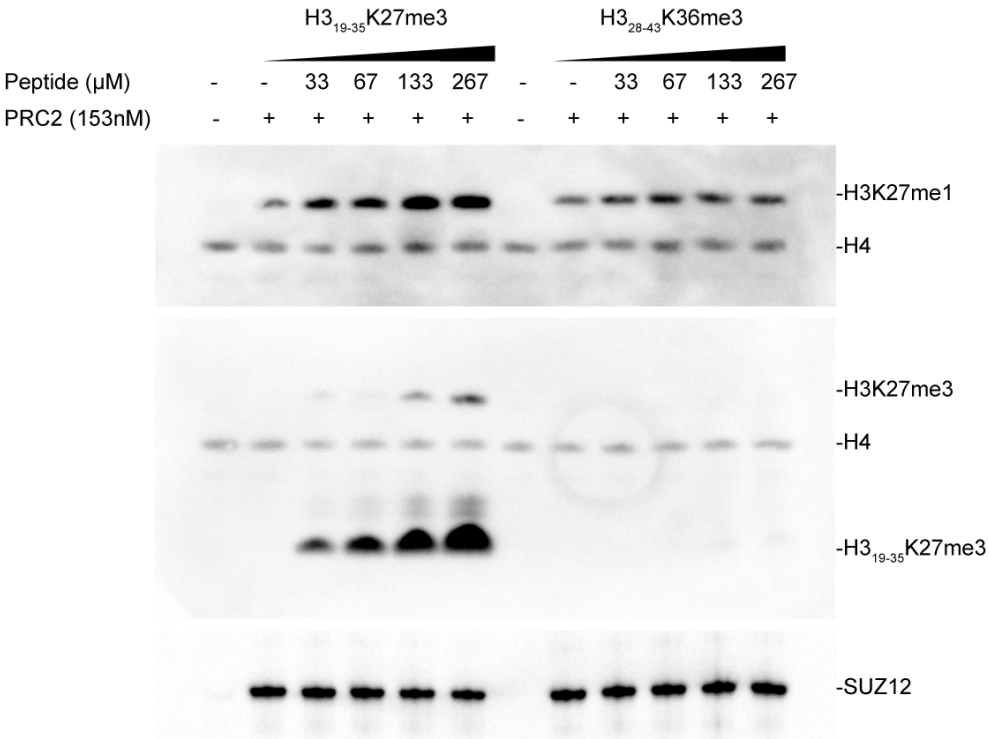


Figure 27. HMTase assay of PRC2 with histone peptides

A western-blot based HMTase assay on 446 nM mononucleosomes (reconstituted with 215 bp DNA) using 153 nM PRC2 with increasing amounts of $H3_{19-35}K27me3$ or $H3_{28-43}K36me3$ peptides; H4 and SUZ12 were probed as a loading control.

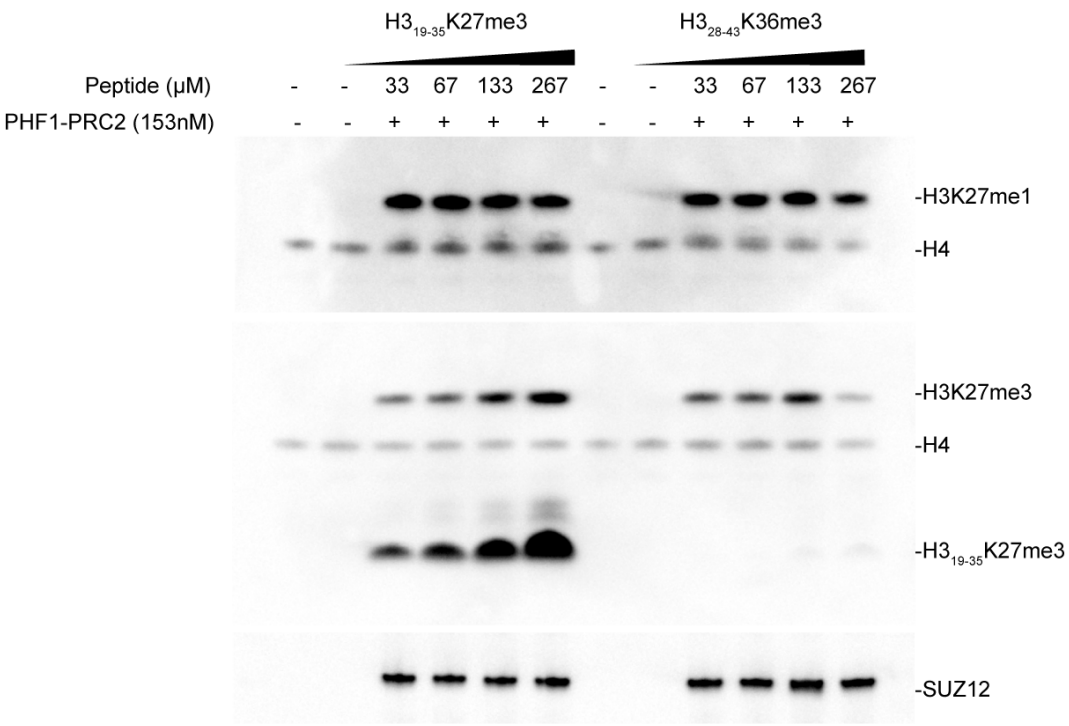


Figure 28. HMTase assay of PHF1-PRC2 with histone peptides

A western-blot based HMTase assay on 446 nM mononucleosomes (reconstituted with 215 bp DNA) using 153 nM PHF1-PRC2 with increasing amounts of histone peptides; H4 and SUZ12 as a loading control.

IV. Conclusion and Discussion

Previous studies showed that the Pcl/PHF1 is required for efficient methylation of H3K27 by PRC2 (Cao et al., 2008; Nekrasov et al., 2007; Sarma et al., 2008). Structural and biochemical analyses of this study reveal that the Pcl/PHF1 reinforces nucleosome binding of PRC2 by the conserved WH and C-terminal domains, and that this is crucial for the efficient H3K27me3 by PRC2. This study thus provides a molecular rationale for how Pcl/PHF1 helps PRC2 generate H3K27me3 efficiently.

1. A closed aromatic cage of the PHD2 in Pcl$_{PHD2\text{-}WH}$

Aromatic cages in many PHD fingers function as epigenetic readers, which bind to modified histone tails (Sanchez and Zhou, 2011). However, the two independent crystal structures solved in this study (Table 12) showed that the aromatic cage of the Pcl PHD2 is closed by the phenyl-ring of the Phe523 (Figure 18). The corresponding residue of the Phe523 in mammalian Pcl proteins is Trp (Figure 17), which seems to ensure its closure with the bulkier residue during the course of its evolution. Still, there could be a chance that the closed aromatic cage in the PHD2 is a regulatory element of the Pcl, which could be opened conditionally. For example, additional domains of the Pcl protein itself or from the PRC2 core subunits, in the context of the full-length Pcl-PRC2 complex, might alter the conformation of the aromatic cage to permit the potential ligand binding. Solving a high-resolution structure of the Pcl together with PRC2 could provide a decisive answer to this question.

2. Mild DNA binding activity of the WH domain in $Pcl_{PHD2\text{-}WH}$

This study demonstrated the presence of the WH domain adjacent to the PHD2 in Pcl by the crystal structure of Pcl construct (Figure 17). In addition, this study showed that the Pcl WH domain binds DNA in a sequence non-specific fashion (Figure 19). One similar example of this is the case of ASH2L WH domain. ASH2L is a core subunit of MLL family histone methyltransferase, which methylates histone H3K4 (Ruthenburg et al., 2007). The ASH2L WH domain also folds next to the PHD finger domain, and binds to DNA in a sequence non-specific manner with mild affinity (Chen et al., 2011; Sarvan et al., 2011).

One possible explanation for this could be that because both Pcl and ASH2L are a component of multi-subunit histone methyltransferase complex (PRC2 and MLL respectively), the WH domains of Pcl and ASH2L might not need to take the whole responsibility to bind the target DNA. Rather, the responsibility might be shared among the subunits of the complex, and all the contributions from the subunits could be added up in order to exert the optimal binding to the target DNA.

3. Unveiled DNA binding activity of PRC2

This study revealed, for the first time to my knowledge, the direct DNA binding activity of PRC2 (Figure 23). Indeed, the Su(z)12 ZnF domain has been suggested to bind to the PRE region *in vivo*, although the ZnF domain is dispensable in histone methyltransferase activity *in vitro* (Rai et al., 2013). In addition, the EZH2 includes two SANT domains, which exhibit similarity to the DNA binding domain of Myb-related proteins (Boyer et al., 2004).

Several reports provided the lncRNA binding activity of PRC2 (reviewed in Davidovich and Cech, 2015). The common consensus from the reports could be merged to sequence non-specificity of PRC2 in the lncRNA binding. In addition, recent UV cross-linking followed by IP (CLIP; Hafner et al., 2010) experiments showed that the SUZ12 binds to RNA directly with or without PRC2 (Beltran et al., 2016). The DNA binding affinity of PRC2 obtained in this study is comparable to the RNA binding affinity of PRC2 described in the other study (Davidovich et al., 2013). Davidovich et al. showed that the PRC2 binds to lncRNA with higher affinity in low salt buffer conditions (Davidovich et al., 2013). Thus, I suggest that PRC2 binds to nucleic acids promiscuously perhaps mainly via charge-based phosphate backbone contact.

This finding of DNA binding activity of PRC2 seems to reinforce the point discussed in the section IV.2 (p. 72), that the DNA binding activity required for the complex could be distributed amongst Pcl and the other PRC2 subunits. Of note, the Pcl homologue strengthens DNA binding affinity of PRC2 by two fold (Figure 23).

4. Indispensable roles of the PHF1 C-terminal domain in PHF1-PRC2

This study uncovered two important features of the PHF1 C-terminal domain: (i) interacting with PRC2 (Figure 21) and (ii) stabilizing the nucleosome binding of PRC2 (Figure 24), which is critical to promote the histone methyltransferase activity of PRC2 (Figure 26).

Ballàre et al. demonstrated that the C-terminal domains of Pcl homologues exhibit a 'reversed chromodomain' structure by secondary structure prediction analysis (Ballaré et al., 2012). Canonical chromodomains are comprised of three β-

sheets and one following α-helix (Figure 29A). The C-terminal domains of Pcl homologues are suggested to contain one α-helix and two following β-sheets (Ballaré et al., 2012). Most chromodomains have been reported for their ability to bind to methylated lysines or nucleic acids via the aromatic cage with three conserved hydrophobic residues (Figure 29B; Eissenberg, 2012). Interestingly, the C-terminal domains of Pcl homologues have the conserved Trp in β2 in all Pcl homologues, implying its potential to form an aromatic cage.

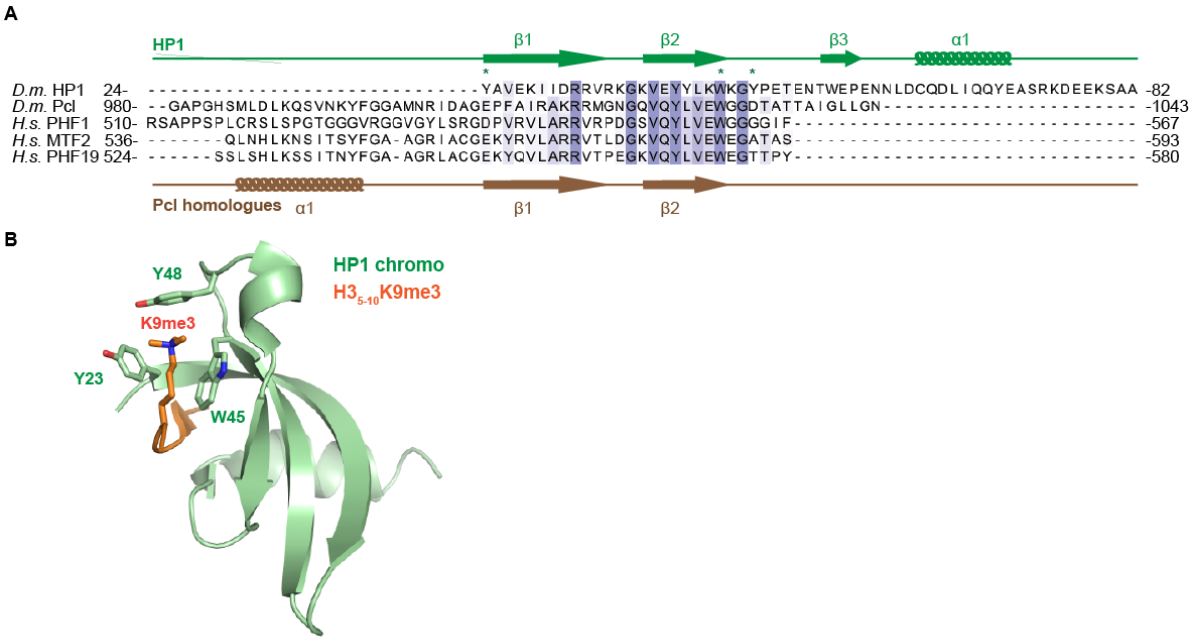


Figure 29. Similarity of PHF1 C-terminal domain to the HP1 chromodomain

(A) Sequence alignment of the HP1 chromodomain and Pcl C-terminal domain; secondary structural elements of HP1 were depicted in green by the crystal structure shown in (B); secondary structural elements of Pcl homologues were drawn in brown by secondary structure prediction; asterisks indicate the residues involved in forming an aromatic cage of HP1. (B) A crystal structure of the HP1 chromodomain with the H3 peptide harbouring K9me3 (PDB code: 1KNE).

I currently do not understand how the PHF1 C-terminal domain stabilizes the association of PRC2 with the nucleosome. I speculate (i) that a part of the C-terminal

domain might bind to PRC2 and the other part to the nucleosome or (ii) that the C-terminal domain might alter the conformation of PRC2 to allow the PRC2 to bind to the nucleosome more stably. High-resolution structure of PHF1(C)-PRC2 onto the NCP could provide an insight on this.

5. Pcl and PRC2 recruitment

Previous ChIP analysis showed the reduced PRC2 recruitment to the Polycomb targets in the murine Pcl-knockdown cells (Cao et al., 2008; Casanova et al., 2011) and in the $Pcl^{-/-}$ *Drosophila* embryo (Nekrasov et al., 2007), suggesting that Pcl/Phf1 facilitates the recruitment of PRC2. However, I could not find any evidence of sequence specificity of DNA binding by the Pcl WH domain. I propose that stabilizing the association of PRC2 with the nucleosome by Pcl homologues could have permitted capturing more abundant PRC2 during the ChIP procedure in the wild type than in the Pcl homologue mutants. In other words, I suggest that the Pcl homologues are responsible for anchoring PRC2 to the nucleosome stably, rather than de-novo recruitment. However I cannot exclude the possibility that Pcl might be involved in the PRC2 recruitment by different mechanism via other domains.

LIST OF ABBREVIATIONS

a.a.	amino acid
AEBSF	4-(2-Aminoethyl) benzenesulfonyl fluoride hydrochloride
ChIP	Chromatin immunoprecipitation
Co-IP	Co-immunoprecipitation
Ct	*Chaetomium thermophilum*
ddH_2O	Double-distilled water (ddH_2O)
Dm	*Drosophila melanogaster*
DNA	Deoxyribonucleic acid
dRAF	*Drosophila* Ring-associated factors
DTT	1,4-dithio-D,L-threitol
Esc	Extra sex combs
ES	Embryonic stem
E(z)	Enhancer of zeste
Flc	fluorescein
FP	fluorescence polarization
His-tag	histidine-tag
HMTase	Histone methyltransferase
Hox	Homeobox
hr(s)	hour(s)
Hs	*Homo sapiens*
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Kan	Kanamycin

kb	Kilobases
kDa	Kilodaltons
lncRNA	long non-coding RNA
me1	monomethylation
me2	dimethylation
me3	trimethylation
mES	mouse embryonic stem
MS	Mass spectrometry
MST	microscale thermophoresis
Nurf55	Nucleosome-remodelling factor 55 kDa subunit
OD_{600}	optical density at 600 nm
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PcG	Polycomb group
Pcl	Polycomblike
PCR	polymerase chain reaction
PDB	Protein data bank
PEG	polyethylene glycol
PHD	Plant homeodomain
PhoRC	Pho repressive complex
PRC1	Polycomb repressor complex-1
PRC2	Polycomb repressor complex-2
PR-DUB	Polycomb repressive deubiquitinase

PREs	Polycomb response elements
SAD	single wavelength anomalous dispersion
SAM	S-adenosylmethionine
SANT	Swi3, Ada2, N-CoR, TFIIIB
SDS	Sodium dodecylsulfate
SET	Su(var)3-9, Enhancer-of-zeste, Trithorax
Su(z)12	Suppressor of zeste 12
TrxG	Trithorax group
Ub	Ubiquitin
Ubx	*Ultrabithorax*
VEFS	VRN2-EMF2-FIS2-Su(z)12
Zn	Zinc

LIST OF FIGURES

LIST OF TABLES

REFERENCES

Gilbert, S. F. (2014). Developmental Biology, 10th Ed, 179-212

The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC

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

Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H., and Adams, P.D. (2012). Towards automated crystallographic structure refinement with phenix.refine. Acta Crystallogr D Biol Crystallogr *68*, 352–367.

Alekseyenko, A.A., Gorchakov, A.A., Kharchenko, P.V., and Kuroda, M.I. (2014). Reciprocal interactions of human C10orf12 and C17orf96 with PRC2 revealed by BioTAP-XL cross-linking and affinity purification. Proceedings of the National Academy of Sciences *111*, 2488–2493.

Ballaré, C., Lange, M., Lapinaite, A., Martin, G.M., Morey, L., Pascual, G., Liefke, R., Simon, B., Shi, Y., Gozani, O., et al. (2012). Phf19 links methylated Lys36 of histone H3 to regulation of Polycomb activity. Nat Struct Mol Biol.

Beltran, M., Yates, C.M., Skalska, L., Dawson, M., Reis, F.P., Viiri, K., Fisher, C.L., Sibley, C.R., Foster, B.M., Bartke, T., et al. (2016). The interaction of PRC2 with RNA or chromatin is mutually antagonistic. Genome Research 1–13.

Berger, I., Fitzgerald, D.J., and Richmond, T.J. (2004). Baculovirus expression system for heterologous multiprotein complexes. Nat Biotech *22*, 1583–1587.

Birve, A., Sengupta, A.K., Beuchle, D., Larsson, J., Kennison, J.A., Rasmuson-Lestander, A., and Muller, J. (2001). Su(z)12, a novel Drosophila Polycomb group gene that is conserved in vertebrates and plants. Development *128*, 3371–3379.

Blackledge, N.P., Farcas, A.M., Kondo, T., King, H.W., McGouran, J.F., Hanssen, L.L.P., Ito, S., Cooper, S., Kondo, K., Koseki, Y., et al. (2014). Variant PRC1 Complex-Dependent H2A Ubiquitylation Drives PRC2 Recruitment and Polycomb Domain Formation. Cell 1–26.

Blackledge, N.P., Rose, N.R., and Klose, R.J. (2015). Targeting Polycomb systems to regulate gene expression: modifications to a complex story. Nature Reviews Molecular Cell Biology 1–7.

Boyer, L.A., Latek, R.R., and Peterson, C.L. (2004). The SANT domain: a unique histone-tail-binding module? Nature Reviews Molecular Cell Biology *5*, 158–163.

Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K., et al. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature *441*, 349–353.

Breen, T.R., and Duncan, I.M. (1986). Maternal expression of genes that regulate the bithorax complex of Drosophila melanogaster. Developmental Biology *118*, 442–456.

Brent, M.M., Anand, R., and Marmorstein, R. (2008). Structural Basis for DNA Recognition by FoxO1 and Its Regulation by Posttranslational Modification. Structure *16*, 1407–1416.

Brien, G.L., Gambero, G., O'Connell, D.J., Jerman, E., Turner, S.A., Egan, C.M., Dunne, E.J., Jurgens, M.C., Wynne, K., Piao, L., et al. (2012a). Polycomb PHF19 binds H3K36me3 and recruits PRC2 and demethylase NO66 to embryonic stem cell genes during differentiation. Nat Struct Mol Biol *19*, 1273–1281.

Brien, G.L., Gambero, G., O'Connell, D.J., Jerman, E., Turner, S.A., Egan, C.M., Dunne, E.J., Jurgens, M.C., Wynne, K., Piao, L., et al. (2012b). Polycomb PHF19 binds H3K36me3 and recruits PRC2 and demethylase NO66 to embryonic stem cell genes during differentiation. Nat Struct Mol Biol *19*, 1273–1281.

Brien, G.L., Healy, E., Jerman, E., Conway, E., Fadda, E., O'Donovan, D., Krivtsov, A.V., Rice, A.M., Kearney, C.J., Flaus, A., et al. (2015). A chromatin-independent role of Polycomb-like 1 to stabilize p53 and promote cellular quiescence. Genes & Development *29*, 2231–2243.

Buchan, D.W.A., Minneci, F., Nugent, T.C.O., Bryson, K., and Jones, D.T. (2013). Scalable web services for the PSIPRED Protein Analysis Workbench. Nucleic Acids Research *41*, W349–W357.

Buchwald, G., van der Stoop, P., Weichenrieder, O., Perrakis, A., van Lohuizen, M., and Sixma, T.K. (2006). Structure and E3-ligase activity of the Ring-Ring complex of polycomb proteins Bmi1 and Ring1b. The EMBO Journal *25*, 2465–2474.

Cai, L., Rothbart, S.B., Lu, R., Xu, B., Chen, W.-Y., Tripathy, A., Rockowitz, S., Zheng, D., Patel, D.J., Allis, C.D., et al. (2013). An H3K36 methylation-engaging Tudor motif of polycomb-like proteins mediates PRC2 complex targeting. Molecular Cell *49*, 571–582.

Callebaut, I., and Mornon, J.-P. (2012). The PWAPA cassette: Intimate association of a PHD-like finger and a winged-helix domain in proteins included in histone-modifying complexes. Biochimie 1–7.

Cao, R., Wang, H., He, J., Erdjument-Bromage, H., Tempst, P., and Zhang, Y. (2008). Role of hPHF1 in H3K27 Methylation and Hox Gene Silencing. Molecular and Cellular Biology *28*, 1862–1872.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S., and Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science *298*, 1039–1043.

Casanova, M., Preissner, T., Cerase, A., Poot, R., Yamada, D., Li, X., Appanah, R., Bezstarosti, K., Demmers, J., Koseki, H., et al. (2011). Polycomblike 2 facilitates the recruitment of PRC2 Polycomb group complexes to the inactive X chromosome and to target loci in embryonic stem cells. Development *138*, 1471–1482.

Chan, C.S., Rastelli, L., and Pirrotta, V. (1994). A Polycomb response element in the Ubx gene that determines an epigenetically inherited state of repression. The EMBO Journal *13*, 2553–2564.

Chen, Y., Wan, B., Wang, K.C., Cao, F., Yang, Y., Protacio, A., Dou, Y., Chang, H.Y., and Lei, M. (2011). Crystal structure of the N-terminal region of human Ash2L shows a winged-helix motif involved in DNA binding. EMBO Reports *12*, 797–803.

Ciferri, C., Lander, G.C., Maiolica, A., Herzog, F., Aebersold, R., and Nogales, E. (2012). Molecular architecture of human polycomb repressive complex 2. eLife *1*, e00005.

Conway, E., Healy, E., and Bracken, A.P. (2015). PRC2 mediated H3K27 methylations in cellular identity and cancer. Current Opinion in Cell Biology *37*, 42–48.

Cooper, S., Dienstbier, M., Hassan, R., Schermelleh, L., Sharif, J., Blackledge, N.P., De Marco, V., Elderkin, S., Koseki, H., Klose, R., et al. (2014). Targeting Polycomb to Pericentric Heterochromatin in Embryonic Stem Cells Reveals a Rolefor H2AK119u1 in PRC2 Recruitment. CellReports 1–38.

Coulson, M., Stanley Robert, Eyre, H.J., and Saint, R. (1998). The Identification and Localization of a Human Gene with Sequence Similarity to Polycomblike of Drosophila melanogaster. 1–3.

Czermin, B., Melfi, R., McCabe, D., Seitz, V., Imhof, A., and Pirrotta, V. (2002). Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. Cell *111*, 185–196.

Davidovich, C., and Cech, T.R. (2015). The recruitment of chromatin modifiers by long noncoding RNAs: lessons from PRC2. Rna *21*, 2007–2022.

Davidovich, C., Zheng, L., Goodrich, K.J., and Cech, T.R. (2013). Promiscuous RNA binding by Polycomb repressive complex 2. Nat Struct Mol Biol *20*, 1250–1257.

Deaton, A.M., and Bird, A. (2011). CpG islands and the regulation of transcription. Genes & Development *25*, 1010–1022.

Deng, Y.-L., Liu, W., Ben Bola ntilde os, Bingham, P., He, Y.-A., Diehl, W., Grable, N., Kung, P.-P., Sutton, S., Maegley, K.A., et al. (2016). Polycomb repressive complex 2 structure with inhibitor reveals a mechanism of activation and drug resistance. Nature Communications *7*, 1–12.

Di Croce, L., and Helin, K. (2013). Transcriptional regulation by Polycomb group proteins. Nat Struct Mol Biol *20*, 1147–1155.

Driever, W., and Nüsslein-Volhard, C. (1988). A gradient of bicoid protein in Drosophila embryos. Cell *54*, 83–93.

Duncan, I.M. (1982). Polycomblike - a Gene That Appears to Be Required for the Normal Expression of the Bithorax and Antennapedia Gene Complexes of Drosophila-Melanogaster. Genetics *102*, 49–70.

Ebert, A., Schotta, G., Lein, S., Kubicek, S., Krauss, V., Jenuwein, T., and Reuter, G. (2004). Su(var) genes regulate the balance between euchromatin and heterochromatin in Drosophila. Genes & Development *18*, 2973–2983.

Eissenberg, J.C. (2012). Structural biology of the chromodomain: Form and function. Gene *496*, 69–78.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr *60*, 2126–2132.

Farcas, A.M., Blackledge, N.P., Sudbery, I., Long, H.K., McGouran, J.F., Rose, N.R.,

Lee, S., Sims, D., Cerase, A., Sheahan, T.W., et al. (2012). KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. eLife *1*, e00205.

Ferrari, K.J., Scelfo, A., Jammula, S., Cuomo, A., Barozzi, I., Stützer, A., Fischle, W., Bonaldi, T., and Pasini, D. (2013). Polycomb-Dependent H3K27me1 and H3K27me2 Regulate Active Transcription and Enhancer Fidelity. Molecular Cell 1–14.

Fischle, W. (2003). Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. Genes & Development *17*, 1870–1881.

Frey, F., Sheahan, T., Finkl, K., Stoehr, G., Mann, M., Benda, C., and Müller, J. (2016). Molecular basis of PRC1 targeting to Polycomb response elements by PhoRC. Genes & Development *30*, 1116–1127.

Friberg, A., Oddone, A., Klymenko, T., Müller, J., and Sattler, M. (2010). Structure of an atypical Tudor domain in the Drosophila Polycomblike protein. Protein Science *19*, 1906–1916.

Fritsch, C., Brown, J.L., Kassis, J.A., and Muller, J. (1999). The DNA-binding Polycomb group protein Pleiohomeotic mediates silencing of a Drosophila homeotic gene. 1–9.

Gajiwala, K.S., Chen, H., Cornille, F., Roques, B.P., Reith, W., and Burley, B.M.S.K. (2000). Structure of the winged-helix protein hRFX1 reveals a new mode of DNA binding. 1–6.

Gao, Z., Zhang, J., Bonasio, R., Strino, F., Sawai, A., Parisi, F., Kluger, Y., and Reinberg, D. (2012). PCGF Homologs, CBX Proteins, and RYBP Define Functionally Distinct PRC1 Family Complexes. Molecular Cell *45*, 344–356.

Gaytán de Ayala Alonso, A., Gutiérrez, L., Fritsch, C., Papp, B., Beuchle, D., and Müller, J. (2007). A genetic screen identifies novel polycomb group genes in Drosophila. Genetics *176*, 2099–2108.

Gonzalezreyes, A., and Morata, G. (1990). The Developmental Effect of Overexpressing a Ubx Product in Drosophila Embryos Is Dependent on Its Interactions with Other Homeotic Products. Cell *61*, 515–522.

Gutiérrez, L., Oktaba, K., Scheuermann, J.C., Gambetta, M.C., Ly-Hartig, N., and Müller, J. (2012). The role of the histone H2A ubiquitinase Sce in Polycomb repression. Development *139*, 117–127.

Ha, H., Howard, C.A., Yeom, Y.I., Abe, K., Uehara, H., Artzt, K., and Bennett, D. (1991). Several testis-expressed genes in the mouse t-complex have expression differences between wild-type and t-mutant mice. Dev. Genet. *12*, 318–332.

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jr, Jungkamp, A.-C., Munschauer, M., et al. (2010). Transcriptome-wide Identification of RNA-Binding Protein and MicroRNA Target Sites by PAR-CLIP. Cell *141*, 129–141.

Han, Z., Xing, X., Hu, M., Zhang, Y., Liu, P., and Chai, J. (2007). Structural basis of EZH2 recognition by EED. Structure/Folding and Design *15*, 1306–1315.

Harami, G.M., Gyimesi, M., and Kovács, M. (2013). From keys to bulldozers:

expanding roles for winged helix domains in nucleic-acid-binding proteins. Trends in Biochemical Sciences *38*, 364–371.

Herz, H.M., Mohan, M., Garrett, A.S., Miller, C., Casto, D., Zhang, Y., Seidel, C., Haug, J.S., Florens, L., Washburn, M.P., et al. (2012). Polycomb Repressive Complex 2-Dependent and -Independent Functions of Jarid2 in Transcriptional Regulation in Drosophila. Molecular and Cellular Biology *32*, 1683–1693.

Hong, Z., Jiang, J., Lan, L., Nakajima, S., Kanno, S.I., Koseki, H., and Yasui, A. (2008). A polycomb group protein, PHF1, is involved in the response to DNA double-strand breaks in human cell. Nucleic Acids Research *36*, 2939–2947.

Hunkapiller, J., Shen, Y., Diaz, A., Cagney, G., McCleary, D., Ramalho-Santos, M., Krogan, N., Ren, B., Song, J.S., and Reiter, J.F. (2012). Polycomb-like 3 promotes polycomb repressive complex 2 binding to CpG islands and embryonic stem cell self-renewal. PLoS Genet *8*, e1002576.

Inouye, C., Remondelli, P., Karin, M., and Elledge, S. (1994). Isolation of a Cdna-Encoding a Metal Response Element-Binding Protein Using a Novel Expression Cloning Procedure - the One Hybrid System. DNA Cell Biol. *13*, 731–742.

Jiao, L., and Liu, X. (2015). Structural basis of histone H3K27 trimethylation by an active polycomb repressive complex 2. Science *350*, aac4383–aac4383.

Justin, N., Zhang, Y., Tarricone, C., Martin, S.R., Chen, S., Underwood, E., De Marco, V., Haire, L.F., Walker, P.A., Reinberg, D., et al. (2016). Structural basis of oncogenic histone H3K27M inhibition of human polycomb repressive complex 2. Nature Communications *7*, 11316.

Jürgens, G. (1985). ti:a group of genes controlling the spatial expression of the bithorax complex in Drosophila. 1–3.

Kabsch, W. (2010). XDS. Acta Crystallogr D Biol Crystallogr *66*, 125–132.

Kalb, R., Latwiel, S., Baymaz, H.I., Jansen, P.W.T.C., Müller, C.W., Vermeulen, M., and Müller, J. (2014). Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. Nat Struct Mol Biol 1–4.

Kawakami, S., Mitsunaga, K., Kikuti, Y.Y., Ando, A., Inoko, H., Yamamura, K., and Abe, K. (1998). Tctex3, related to Drosophila polycomblike, is expressed in male germ cells and mapped to the mouse t-complex. Mamm. Genome *9*, 874–880.

Ketel, C.S., Andersen, E.F., Vargas, M.L., Suh, J., Strome, S., and Simon, J.A. (2005). Subunit Contributions to Histone Methyltransferase Activities of Fly and Worm Polycomb Group Complexes. Molecular and Cellular Biology *25*, 6857–6868.

Klymenko, T. (2006). A Polycomb group protein complex with sequence-specific DNA-binding and selective methyl-lysine-binding activities. Genes & Development *20*, 1110–1122.

Kouzarides, T. (2007). Chromatin Modifications and Their Function. Cell *128*, 693–705.

Kuzmichev, A., Jenuwein, T., Tempst, P., and Reinberg, D. (2004). Different EZH2-containing complexes target methylation of histone H1 or nucleosomal histone H3. Molecular Cell *14*, 183–193.

Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P., and Reinberg, D. (2002). Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. Genes & Development *16*, 2893–2905.

Lagarou, A., Mohd-Sarip, A., Moshkin, Y.M., Chalkley, G.E., Bezstarosti, K., Demmers, J.A.A., and Verrijzer, C.P. (2008). dKDM2 couples histone H2A ubiquitylation to histone H3 demethylation during Polycomb group silencing. Genes & Development *22*, 2799–2810.

Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K.-I., et al. (2006). Control of Developmental Regulators by Polycomb in Human Embryonic Stem Cells. Cell *125*, 301–313.

Lewis, E.B. (1978). A gene complex controlling segmentation in Drosophila. Nature *276*, 565–570.

Lewis, P.W., Müller, M.M., Koletsky, M.S., Cordero, F., Lin, S., Banaszynski, L.A., Garcia, B.A., Muir, T.W., Becher, O.J., and Allis, C.D. (2013). Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma. Science *340*, 857–861.

Li, G., Margueron, R., Ku, M., Chambon, P., Bernstein, B.E., and Reinberg, D. (2010). Jarid2 and PRC2, partners in regulating gene expression. Genes & Development *24*, 368–380.

Lowary, P.T., and Widom, J. (1998). New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. Journal of Molecular Biology *276*, 19–42.

Luger, K., Rechsteiner, T.J., and Richmond, T.J. (1999). Expression and purification of recombinant histones and nucleosome reconstitution. Methods Mol. Biol. *119*, 1–16.

Margueron, R., Justin, N., Ohno, K., Sharpe, M.L., Son, J., DruryIII, W.J., Voigt, P., Martin, S.R., Taylor, W.R., De Marco, V., et al. (2009). Role of the polycomb protein EED in the propagation of repressive histone marks. Nature *461*, 762–767.

McGinty, R.K., Henrici, R.C., and Tan, S. (2014). Crystal structure of the PRC1 ubiquitylation module bound to the nucleosome. Nature *514*, 591–596.

Mckeon, J., and Brock, H.W. (1991). Interactions of the Polycomb Group of Genes with Homeotic Loci of Drosophila. Rouxs Archives of Developmental Biology *199*, 387–396.

Min, J., Zhang, Y., and Xu, R.-M. (2003). Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. Genes & Development *17*, 1823–1828.

Mohd-Sarip, A. (2005). Synergistic recognition of an epigenetic DNA element by Pleiohomeotic and a Polycomb core complex. Genes & Development *19*, 1755–1760.

Mohd-Sarip, A., Venturini, F., Chalkley, G.E., and Verrijzer, C.P. (2002). Pleiohomeotic Can Link Polycomb to DNA and Mediate Transcriptional Repression. Molecular and Cellular Biology *22*, 7473–7483.

Musselman, C.A., Avvakumov, N., Watanabe, R., Abraham, C.G., Lalonde, M.-E., Hong, Z., Allen, C., Roy, S., Nuñez, J.K., Nickoloff, J., et al. (2012). Molecular basis for H3K36me3 recognition by the Tudor domain of PHF1. Nat Struct Mol Biol.

Müller, J., Hart, C.M., Francis, N.J., Vargas, M.L., Sengupta, A., Wild, B., Miller, E.L., O'Connor, M.B., Kingston, R.E., and Simon, J.A. (2002). Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. Cell *111*, 197–208.

Nekrasov, M., Klymenko, T., Fraterman, S., Papp, B., Oktaba, K., Köcher, T., Cohen, A., Stunnenberg, H.G., Wilm, M., and Müller, J. (2007). Pcl-PRC2 is needed to generate high levels of H3-K27 trimethylation at Polycomb target genes. The EMBO Journal *26*, 4078–4088.

Nekrasov, M., Wild, B., and Müller, J. (2005). Nucleosome binding and histone methyltransferase activity of Drosophila PRC2. EMBO Reports *6*, 348–353.

Nüsslein-Volhard, C., and Wieschaus, E. (1980). Mutations affecting segment number and polarity in Drosophila. Nature *287*, 795–801.

O'Connell, S. (2001). Polycomblike PHD Fingers Mediate Conserved Interaction with Enhancer of Zeste Protein. Journal of Biological Chemistry *276*, 43065–43073.

O'Connell, S., Wang, L., Robert, S., Jones, C.A., Saint, R., and Jones, R.S. (2001). Polycomblike PHD fingers mediate conserved interaction with enhancer of zeste protein. J. Biol. Chem. *276*, 43065–43073.

Oktaba, K., Gutiérrez, L., Gagneur, J., Girardot, C., Sengupta, A.K., Furlong, E.E.M., and Müller, J. (2008). Dynamic regulation by polycomb group protein complexes controls pattern formation and the cell cycle in Drosophila. Developmental Cell *15*, 877–889.

Papp, B., and Muller, J. (2006). Histone trimethylation and the maintenance of transcriptional ONand OFF states by trxG and PcG proteins. Genes & Development *20*, 2041–2054.

Pasini, D., Bracken, A.P., Jensen, M.R., Lazzerini Denchi, E., and Helin, K. (2004). Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. The EMBO Journal *23*, 4061–4071.

Pearson, J.C., Lemons, D., and McGinnis, W. (2005). Modulating Hox gene functions during animal body patterning. Nat Rev Genet *6*, 893–904.

Peng, J.C., Valouev, A., Swigut, T., Zhang, J., Zhao, Y., Sidow, A., and Wysocka, J. (2009). Jarid2/Jumonji Coordinates Control of PRC2 Enzymatic Activity and Target Gene Occupancy in Pluripotent Cells. Cell *139*, 1290–1302.

Pengelly, A.R., Copur, O., Jackle, H., Herzig, A., and Muller, J. (2013). A Histone Mutant Reproduces the Phenotype Caused by Loss of Histone-Modifying Factor Polycomb. Science *339*, 698–699.

Piunti, A., and Shilatifard, A. (2016). Epigenetic balance of gene expression by Polycomb and COMPASS families. Science *352*, aad9780–aad9780.

Platero, J.S., Sharp, E.J., Adler, P.N., and Eissenberg, J.C. (1996). In vivo assay for protein-protein interactions using Drosophila chromosomes. Chromosoma *104*, 393–404.

Rai, A.N., Vargas, M.L., Wang, L., Andersen, E.F., Miller, E.L., and Simon, J.A. (2013). Elements of the Polycomb Repressor SU(Z)12 Needed for Histone H3-K27 Methylation, the Interface with E(Z), and In Vivo Function. Molecular and Cellular Biology *33*, 4844–4856.

Rastelli, L., Chan, C.S., and Pirrotta, V. (1993). Related Chromosome Binding-Sites for Zeste, Suppressors of Zeste and Polycomb Group Proteins in Drosophila and Their Dependence on Enhancer of Zeste Function. The EMBO Journal *12*, 1513–1522.

Remondelli, P., and Leone, A. (1997). Interactions of the zinc-regulated factor (ZiRF1) with the mouse metallothionein Ia promoter. Biochem. J. *323 (Pt 1)*, 79–85.

Riising, E.M., Comet, I., Leblanc, B., Wu, X., Johansen, J.V., and Helin, K. (2014). Gene Silencing Triggers Polycomb Repressive Complex 2 Recruitment to CpG Islands Genome Wide. Molecular Cell 1–14.

Ruthenburg, A.J., Allis, C.D., and Wysocka, J. (2007). Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. Molecular Cell *25*, 15–30.

Sanchez, R., and Zhou, M.-M. (2011). The PHD finger: a versatile epigenome reader. Trends in Biochemical Sciences *36*, 364–372.

Sarma, K., Margueron, R., Ivanov, A., Pirrotta, V., and Reinberg, D. (2008). Ezh2 Requires PHF1 To Efficiently Catalyze H3 Lysine 27 Trimethylation In Vivo. Molecular and Cellular Biology *28*, 2718–2731.

Sarvan, S., Avdic, V., Tremblay, V., Chaturvedi, C.-P., Zhang, P., Lanouette, S., Blais, A., Brunzelle, J.S., Brand, M., and Couture, J.-F. (2011). Crystal structure of the trithorax group protein ASH2L reveals a forkhead-like DNA binding domain. Nat Struct Mol Biol *18*, 857–859.

Savla, U., Benes, J., Zhang, J., and Jones, R.S. (2008). Recruitment of Drosophila Polycomb-group proteins by Polycomblike, a component of a novel protein complex in larvae. Development *135*, 813–817.

Scheuermann, J.C., de Ayala Alonso, A.G., Oktaba, K., Ly-Hartig, N., McGinty, R.K., Fraterman, S., Wilm, M., Muir, T.W., and Müller, J. (2010). Histone H2A deubiquitinase activity of the Polycomb repressive complex PR-DUB. Nature *465*, 243–247.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat Meth *9*, 676–682.

Schmitges, F.W., Prusty, A.B., Faty, M., Stützer, A., Lingaraju, G.M., Aiwazian, J., Sack, R., Hess, D., Li, L., Zhou, S., et al. (2011). Histone Methylation by PRC2 Is Inhibited by Active Chromatin Marks. Molecular Cell *42*, 330–341.

Schwartz, Y.B., and Pirrotta, V. (2007). Polycomb silencing mechanisms and the management of genomic programmes. Nat Rev Genet *8*, 9–22.

Schwartz, Y.B., Kahn, T.G., Nix, D.A., Li, X.-Y., Bourgon, R., Biggin, M., and Pirrotta, V. (2006). Genome-wide analysis of Polycomb targets in Drosophila melanogaster. Nat Genet *38*, 700–705.

Schwartzentruber, J., Korshunov, A., Liu, X.-Y., Jones, D.T.W., Pfaff, E., Jacob, K., Sturm, D., Fontebasso, A.M., Quang, D.-A.K., Tönjes, M., et al. (2012). Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. Nature 1–8.

Shao, Z., Raible, F., Mollaaghababa, R., Guyon, J.R., Wu, C.-T., Bender, W., and Kingston, R.E. (1999). Stabilization of Chromatin Structure by PRC1, a Polycomb Complex. Cell *98*, 37–46.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. *7*, 539–539.

Simon, J., Chiang, A., and Bender, W. (1992). Ten different Polycomb group genes are required for spatial control of the abdA and AbdB homeotic products. Development *114*, 493–505.

Simon, J.A., and Kingston, R.E. (2009). Mechanisms of Polycomb gene silencing: knowns and unknowns. Nature Reviews Molecular Cell Biology 1–12.

Smits, A.H., Jansen, P.W.T.C., Poser, I., Hyman, A.A., and Vermeulen, M. (2012). Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics. Nucleic Acids Research.

Sneeringer, C.J., Scott, M.P., Kuntz, K.W., Knutson, S.K., Pollock, R.M., Richon, V.M., and Copeland, R.A. (2010). Coordinated activities of wild-type plus mutant EZH2 drive tumor-associated hypertrimethylation of lysine 27 on histone H3 (H3K27) in human B-cell lymphomas. Proceedings of the National Academy of Sciences *107*, 20980–20985.

Söding, J. (2005). Protein homology detection by HMM-HMM comparison. Bioinformatics *21*, 951–960.

Terwilliger, T.C., Adams, P.D., Read, R.J., McCoy, A.J., Moriarty, N.W., Grosse-Kunstleve, R.W., Afonine, P.V., Zwart, P.H., and Hung, L.W. (2009). Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. Acta Crystallogr D Biol Crystallogr *65*, 582–601.

Thåström, A., Lowary, P.T., Widlund, H.R., Cao, H., Kubista, M., and Widom, J. (1999). Sequence motifs and free energies of selected natural and non-natural nucleosome positioning DNA sequences. Journal of Molecular Biology *288*, 213–229.

Tie, F., Furuyama, T., and Harte, P.J. (1998). The Drosophila Polycomb Group proteins ESC and E(Z) bind directly to each other and co-localize at multiple chromosomal sites. Development *125*, 3483–3496.

Tie, F., Furuyama, T., Prasad-Sinha, J., Jane, E., and Harte, P.J. (2001). The Drosophila Polycomb Group proteins ESC and E(Z) are present in a complex containing the histone-binding protein p55 and the histone deacetylase RPD3. Development *128*, 275–286.

Tie, F., Prasad-Sinha, J., Birve, A., Rasmuson-Lestander, A., and Harte, P.J. (2003). A 1-megadalton ESC/E(Z) complex from Drosophila that contains polycomblike and RPD3. Molecular and Cellular Biology *23*, 3352–3362.

Tolhuis, B., Muijrers, I., de Wit, E., Teunissen, H., Talhout, W., van Steensel, B., and van Lohuizen, M. (2006). Genome-wide profiling of PRC1 and PRC2 Polycomb chromatin binding in Drosophila melanogaster. Nat Genet *38*, 694–699.

Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A., and Kinzler, K.W. (2013). Cancer genome landscapes. Science *339*, 1546–1558.

Walker, E., Chang, W.Y., Hunkapiller, J., Cagney, G., Garcha, K., Torchia, J., Krogan, N.J., Reiter, J.F., and Stanford, W.L. (2010). Polycomb-like 2 Associates with PRC2 and Regulates Transcriptional Networks during Mouse Embryonic Stem Cell Self-Renewal and Differentiation. Stem Cell *6*, 153–166.

Wang, Z., Song, J., Milne, T.A., Wang, G.G., Li, H., Allis, C.D., and Patel, D.J. (2010). Pro Isomerization in MLL1 PHD3-Bromo Cassette Connects H3K4me Readout to CyP33 and HDAC-Mediated Repression. Cell *141*, 1183–1194.

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., and Barton, G.J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189–1191.

Whitcomb, S.J., Basu, A., Allis, C.D., and Bernstein, E. (2007). Polycomb Group proteins: an evolutionary perspective. Trends in Genetics *23*, 494–502.

Wu, G., Broniscer, A., McEachron, T.A., Lu, C., Paugh, B.S., Becksfort, J., Qu, C., Ding, L., Huether, R., Parker, M., et al. (2012). Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas. Nat Genet *44*, 251–253.

Yang, Y., Wang, C., Zhang, P., Gao, K., Wang, D., Yu, H., Zhang, T., Jiang, S., Hexige, S., Hong, Z., et al. (2013). Polycomb group protein PHF1 regulates p53-dependent cell growth arrest and apoptosis. Journal of Biological Chemistry *288*, 529–539.

Yuan, W., Wu, T., Fu, H., Dai, C., Wu, H., Liu, N., Li, X., Xu, M., Zhang, Z., Niu, T., et al. (2012). Dense Chromatin Activates Polycomb Repressive Complex 2 to Regulate H3 Lysine 27 Methylation. Science *337*, 971–975.

APPENDIX

The tables below show the list of clones generated and used for crystallization trials or biochemical studies during the PhD study, but not described in this thesis. All abbreviation is the same as ones in Table 4.

Number	Protein name	Constructs	Vector	Affinity tag	Clea.	Res.	Species	Specifications
JC01	Pcl	full length	pEC-K-3C-His	His	3C	Kan	Dm	
JC02	Pcl01	419-662	pEC-K-TEV-His	His	TEV	Kan	Dm	
JC03	Pcl02	419-694	pEC-K-TEV-His	His	TEV	Kan	Dm	
JC04	Pcl03	505-662	pEC-K-TEV-His	His	TEV	Kan	Dm	
JC05	Pcl04	505-694	pEC-K-TEV-His	His	TEV	Kan	Dm	
JC06	Pcl05	505-659	pEC-A-Sumo	His	SUMO	Kan	Dm	
JC07	Pcl06	491-690	pEC-A-Sumo	His	SUMO	Amp	Dm	
JC08	Pcl07	491-694	pEC-A-Sumo	His	SUMO	Amp	Dm	
JC09	Pcl08	491-690	pEC-K-3C-His	His	3C	Kan	Dm	
JC11	Pcl09_w1_E	491-694	pEC-K-3C-His	His	3C	Kan	Dm	K650E,R651E
JC12	Pcl09_a3_E	491-694	pEC-K-3C-His	His	3C	Kan	Dm	R631E,Q634E,K637E
JC13	Pcl09_w1_a3_E	491-694	pEC-K-3C-His	His	3C	Kan	Dm	K650E,R651E, R631E,Q634E,K637E
JC18	Pcl10	348-694	pEC-K-3C-His	His	3C	Kan	Dm	
JC19	Pcl11	487-694	pEC-K-3C-His	His	3C	Kan	Dm	
JC21	Pcl09_F523A	491-694	pEC-K-3C-His	His	3C	Kan	Dm	F523A
JC22	Pcl09_Q529A	491-694	pEC-K-3C-His	His	3C	Kan	Dm	Q529A
JC23	Pcl09_W536E	491-694	pEC-K-3C-His	His	3C	Kan	Dm	W536E
JC24	PHF1_His	full length	pEC-K-3C-His-CDF	His	3C	Strep	Hs	
JC25	PHF1_1	30-394	pEC-K-3C-His	His	3C	Kan	Hs	
JC26	PHF1_5	29-342	pEC-K-3C-His	His	3C	Kan	Hs	
JC27	PHF1_6	182-342	pEC-K-3C-His	His	3C	Kan	Hs	
JC29	PHF1_8	165-394	pEC-K-3C-His	His	3C	Kan	Hs	
JC30	PHF1_9	185-394	pEC-K-3C-His	His	3C	Kan	Hs	
JC31	PHF1_10	248-339	pEC-K-3C-His	His	3C	Kan	Hs	
JC32	PHF1_7_W197A	165-363	pEC-K-3C-His	His	3C	Kan	Hs	W197A
JC33	PHF1_GST	full length	pEC-K-GST	GST	3C	Kan	Hs	
JC34	PHF1_W210E	165-363	pEC-K-3C-His	His	3C	Kan	Hs	W210E
JC36	PHF1_helix3mut	165-363	pEC-K-3C-His	His	3C	Kan	Hs	R304A, S307A, N310A
JC37	PHF1_double mut-2	165-363	pEC-K-3C-His	His	3C	Kan	Hs	R304A, S307A, N310A, K323A, K324A
JC38	MTF2	full length	pEC-K-3C-His	His	3C	Kan	Hs	
JC40	MTF2_2	171-368	pEC-K-3C-His	His	3C	Kan	Hs	

JC42	MTF2_4	197-357	pEC-K-3C-His	His	3C	Kan	Hs	
JC43	MTF2_5	180-362	pEC-K-3C-His	His	3C	Kan	Hs	
JC44	MTF2_6	180-371	pEC-K-3C-His	His	3C	Kan	Hs	
JC45	MTF2_7	200-371	pEC-K-3C-His	His	3C	Kan	Hs	
JC46	MTF2_8	263-354	pEC-K-3C-His	His	3C	Kan	Hs	
JC47	MTF2-9	200-381	pEC-K-3C-His	His	3C	Kan	Hs	
JC48	PHF19	full length	pEC-K-3C-His-CDF	His	3C	Strep	Hs	
JC49	PHF19_1	39-369	pEC-K-3C-His	His	3C	Kan	Hs	
JC50	PHF19_2	162-369	pEC-K-3C-His	His	3C	Kan	Hs	
JC55	PHF19_7	174-350	pEC-K-3C-His	His	3C	Kan	Hs	
JC56	PHF19_8	174-365	pEC-K-3C-His	His	3C	Kan	Hs	
JC57	PHF19_9	194-365	pEC-K-3C-His	His	3C	Kan	Hs	
JC60	PHF19_Q212A	194-365	pEC-K-3C-His	His	3C	Kan	Hs	Q212A
JC61	PHF19_W219E	194-365	pEC-K-3C-His	His	3C	Kan	Hs	W219E
JC62	PHF19 wing1mut	194-366	pEC-K-3C-His	His	3C	Kan	Hs	K331A, K332A
JC63	PHF19 hellix mut	194-365	pEC-K-3C-His	His	3C	Kan	Hs	H312A, H315A, N318A
JC64	PHF19 double mut-1	194-365	pEC-K-3C-His	His	3C	Kan	Hs	K331A, K332A, H312A, H315A, N318A
JC65	PHO01	355-474	pEC-K-3C-His	His	3C	Kan	Dm	
JC66	PHO02	355-end	pEC-K-3C-His	His	3C	Kan	Dm	
JC67	PHF1_2	156-394	pEC-K-3C-His	His	3C	Kan	Hs	
JC68	PHF1_3	28-355	pEC-K-3C-His	His	3C	Kan	Hs	
JC69	PHF1_4	156-355	pEC-K-3C-His	His	3C	Kan	Hs	
JC70	EZ01	9-496	pEC-K-3C-His-CDF	His	3C	Strep	Dm	
JC71	EZ02	9-524	pEC-K-3C-His-CDF	His	3C	Strep	Dm	
JC72	EZ03	9-747	pEC-K-3C-His-CDF	His	3C	Strep	Dm	
JC73	EZ04	72-496	pEC-K-3C-His-CDF	His	3C	Strep	Dm	
JC74	EZ05	72-524	pEC-K-3C-His-CDF	His	3C	Strep	Dm	
JC75	EZ06	72-747	pEC-K-3C-His-CDF	His	3C	Strep	Dm	
JC76	EZ01-GST	9-496	pEC-K-GST	His-GST	3C	Kan	Dm	
JC77	EZ02-GST	9-524	pEC-K-GST	His-GST	3C	Kan	Dm	
JC78	EZ03-GST	9-747	pEC-K-GST	His-GST	3C	Kan	Dm	
JC79	EZ04-GST	72-496	pEC-K-GST	His-GST	3C	Kan	Dm	

JC80	EZ05-GST	72-524	pEC-K-GST	His-GST	3C	Kan	Dm
JC81	EZ06-GST	72-747	pEC-K-GST	His-GST	3C	Kan	Dm
JC82	EZH2-01	75-515	pEC-K-3C-His-CDF	His	3C	Strep	Hs
JC83	EZH2-02	75-740	pEC-K-3C-His-CDF	His	3C	Strep	Hs
JC84	EZH2-01-GST	75-515	pEC-K-GST	His-GST	3C	Kan	Hs
JC85	EZH2-02-GST	75-740	pEC-K-GST	His-GST	3C	Kan	Hs
JC86	Pcl2-CDF	419-694	pEC-K-3C-His-CDF	His	3C	Strep	Dm
JC87	PHF1_11	86-413	pEC-K-3C-His-CDF	His	3C	Strep	Hs
JC88	EZ07	56-384	pEC-K-3C-His-CDF	His	3C	Strep	Dm
JC89	EZ07-GST	56-384	pEC-K-GST	His-GST	3C	Kan	Dm
JC91	EZH2-03-GST	78-350	pEC-K-GST	His-GST	3C	Kan	Hs
JC92	PHF1_7-CDF	165-363	pEC-K-3C-His-CDF	His	3C	Strep	Hs
JC95	Pcl13	348-1043	pEC-K-3C-His-CDF	His	3C	Strep	Dm
JC95-G	Pcl13	348-1043	pEC-K-GST	His-GST	3C	Kan	Dm
JC96	Pcl14	419-1043	pEC-K-3C-His-CDF	His	3C	Strep	Dm
JC96-G	Pcl14	419-1043	pEC-K-GST	His-GST	3C	Kan	Dm
JC97	PHF1_12	28-563	pEC-K-3C-His-CDF	His	3C	Strep	Hs
JC97-G	PHF1_12	28-563	pEC-K-GST	His-GST	3C	Kan	Hs
JC98	PHF1_13	86-563	pEC-K-3C-His-CDF	His	3C	Strep	Hs
JC98-G	PHF1_13	86-563	pEC-K-GST	His-GST	3C	Kan	Hs
JC99	PHF1_14	86-363	pEC-K-3C-His-CDF	His	3C	Strep	Hs
JC99-G	PHF1_14	86-363	pEC-K-GST	His-GST	3C	Kan	Hs
JC101	PHF19_13	95-578	pEC-K-3C-His-CDF	His	3C	Strep	Hs
JC102	PHF19_14	95-370	pEC-K-3C-His-CDF	His	3C	Strep	Hs
JC103	Su(z)12-01	387-672	pEC-K-3C-His-CDF	His	3C	Strep	Dm
JC103-G	Su(z)12-01	387-672	pEC-K-GST	His-GST	3C	Kan	Dm
JC104	Su(z)12-02	387-721	pEC-K-3C-His-CDF	His	3C	Strep	Dm
JC104-G	Su(z)12-02	387-721	pEC-K-GST	His-GST	3C	Kan	Dm
JC105	SUZ12-01	425-723	pEC-K-3C-His-CDF	His	3C	Strep	Hs
JC105-G	SUZ12-01	425-723	pEC-K-GST	His-GST	3C	Kan	Hs
JC106	SUZ12-02	425-739	pEC-K-3C-His-CDF	His	3C	Strep	Hs
JC106-G	SUZ12-02	425-739	pEC-K-GST	His-GST	3C	Kan	Hs

JC111-G	EZH2_01-GST	78-475	pEC-K-GST	His-GST	3C	Kan	Hs
JC111	EZH2_01	78-475	pEC-K-3C-His-CDF	His	3C	Strep	Hs
JC112	EZH2_02	78-735	pEC-K-3C-His-CDF	His	3C	Strep	Hs
JC113-G	EZH2_03-GST	78-345	pEC-K-GST	His-GST	3C	Kan	Hs
JC113	EZH2_03	78-345	pEC-K-3C-His-CDF	His	3C	Strep	Hs
JC114	EZH2_04	302-475	pEC-K-10 x His	His10	3C	Amp	Hs
JC115-G	E(z)10-GST	75-339	pEC-K-GST	His-GST	3C	Kan	Dm
JC115	E(z)10	75-339	pEC-K-3C-His-CDF	His	3C	Strep	Dm
JC116-G	E(z)11-GST	114-368	pEC-K-GST	His-GST	3C	Kan	Dm
JC116	E(z)11	114-368	pEC-K-3C-His-CDF	His	3C	Strep	Dm
JC119	E(z)05-His10	72-524	pEC-K-10 x His	10 x His	3C	Amp	Dm
JC120	EZH2-01-His10	78-475	pEC-K-10 x His	10 x His	3C	Amp	Hs
JC121	Su(z)12-02-His10	387-721	pEC-K-10 x His	10 x His	3C	Amp	Dm
JC122	SUZ12-02-His10	425-739	pEC-K-10 x His	10 x His	3C	Amp	Hs
JC123	His10-E(z)08-His6-Pcl02		pEC-K-10 x His	10 x His	3C	Amp	Dm
JC124	GST-E(z)08-His6-Pcl02		pEC-K-GST	His-GST	3C	Kan	Dm
JC125-G	E(z)08-GST	56-491	pEC-K-GST	His-GST	3C	Kan	Dm
JC126	His10-EZH2-01-His6-PHF1-14		pEC-K-10 x His	10 x His	3C	Amp	Hs
JC127	GST-EZH2-01-His6-PHF1-14		pEC-K-GST	His-GST	3C	Kan	Hs
JC128	GST-EZH2-01-GST-PHF1-14	78-475	pEC-K-GST	His-GST	3C	Kan	Hs
JC129	PHF19	full length	pEC-K-GST	GST	3C	Kan	Hs
JC130	GST-Su(z)12-02-His6-Pcl02		pEC-K-GST	His-GST	3C	Kan	Dm
JC131	His10-SUZ12-02-His6-PHF1-14		pEC-K-10 x His	10 x His	3C	Amp	Hs
JC132	SUZ12	full length	pEC-K-3C-His	His	3C	Amp	Hs
JC133	His6-EZH2-05-his6-PHF1-14	80-497	pEC-K-3C-His	His	3C	Kan	Hs
JC134	His6-EZH2-06-his6-PHF1-14	80-511	pEC-K-3C-His	His	3C	Kan	Hs
JC135	His6-E(z)12-his6-Pcl-02	56-526	pEC-K-3C-His	His	3C	Kan	Dm
JC137	GST--EZH2-06-his6-PHF1-14		pEC-K-GST	His-GST	3C	Kan	Hs
JC145	His6-SUZ12-02-His06-EZH2-02	78-735	pEC-K-3C-His	His	3C	Kan	Hs
JC146	His6-Su(z)12-02-his6-E(z)06	72-747	pEC-K-3C-His	His	3C	Kan	Dm
JC148	GST-Su(z)12-02-his6-E(z)06		pEC-K-GST	His-GST	3C	Kan	Dm
JC150	His6-EZH2-05	80-497	pEC-K-3C-His	His	3C	Kan	Hs

JC151	His6-EZH2-06	80-511	pEC-K-3C-His	His	3C	Kan	Hs	
JC155	Pcl09_a3_w2_A	491-694	pEC-K-3C-His	His	3C	Kan	Dm	R631A, Q634A, K637A, K684A K687A
JC156	His6-Su(z)12-02-his6-E(z)06	72-747	pEC-K-3C-His	His	3C	Kan	Dm	
JC157	Pcl09_w2_I	491-694	pEC-K-3C-His	His	3C	Kan	Dm	K684I, K687I
JC158	Pcl09_w1_w2_A	491-694	pEC-K-3C-His	His	3C	Kan	Dm	K650A, R651A, K684A, K687A
JC159	PHF1_15	185-379	pEC-K-3C-His	His	3C	Kan	Hs	
JC162	Pcl09_w2_A	491-694	pEC-K-3C-His	His	3C	Kan	Dm	K684A K687A
JC163	Pcl09_w2_E	491-694	pEC-K-3C-His	His	3C	Kan	Dm	K684E, K687E
JC165	Pcl09_w1_w2_E	491-694	pEC-K-3C-His	His	3C	Kan	Dm	K650E, R651E, K684E, K687E
JC166	PHF1_SII_PFB	full length	pFastBac	Strep x 2	TEV	Amp	Hs	
JC167	PHF19_SII_PFB	full length	pFastBac	Strep x 2	TEV	Amp	Hs	
JC169	EED-GST	full length	pEC-K-GST	GST	3C	Kan	Hs	
JC170	EED	full length	pEC-K-3C-His-CDF	His	3C	Strep	Hs	
JC171	RBBP4-GST	full length	pEC-K-GST	GST	3C	Kan	Hs	
JC172	RBBP4	full length	pEC-K-3C-His-CDF	His	3C	Strep	Hs	
JC173	EZH2_GST	full length	pEC-K-GST	GST	3C	Kan	Hs	
JC174	EZH2	full length	pEC-K-3C-His-CDF	His	3C	Strep	Hs	
JC175	SUZ12-GST	full length	pEC-K-GST	GST	3C	Kan	Hs	
JC176	SUZ12	full length	pEC-K-3C-His-CDF	His	3C	Strep	Hs	
JC177	AEBP2-GST	full length	pEC-K-GST	GST	3C	Kan	Hs	
JC178	AEBP2	full length	pEC-K-3C-His-CDF	His	3C	Strep	Hs	
JC179	GST-EED-His-PHF1	full length	pEC-K-GST	GST	3C	Kan	Hs	
JC180	GST-EED-His-PHF19	full length	pEC-K-GST	GST	3C	Kan	Hs	
JC181	GST-RBBP4-His-SUZ12	full length	pEC-K-GST (Amp)	GST	3C	Amp	Hs	
JC185	SUZ12_VEFS_544-738	544-738	pEC-K-3C-His-CDF	His	3C	Strep	Hs	
JC186	SUZ12_VEFS_GST_544-738	544-738	pEC-K-GST	GST	3C	Kan	Hs	
JC187	GST-EED-GST-PHF1	full length	pEC-K-GST	GST	3C	Kan	Hs	
JC188	GST-EED-GST-PHF19	full length	pEC-K-GST	GST	3C	Kan	Hs	
JC191	GST-SUZ12_VEFS-GST-PHF19	full length	pEC-K-GST	GST	3C	Kan	Hs	
JC198	PHF19_PBac_C_His	37-end	pBAC	C-His		Amp	Hs	
JC199	PHF1_SII_PFB_delN	28-end	pFastBac	Strep x 2	TEV	Amp	Hs	
JC200	PHF19_SII_PFB_delN	38-end	pFastBac	Strep x 2	TEV	Amp	Hs	

JC201	PHF1_SII_PFB_HAremoved	full length	pFastBac	Strep x 2, His	TEV	Amp	Hs	
JC202	PHF19_SII_PFB_HAremoved	full length	pFastBac	Strep x 2, His	TEV	Amp	Hs	
JC203	PHF1_SII_PFB_HA-TEVremoved	full length	pFastBac	Strep x 2, His	TEV	Amp	Hs	
JC204	PHF19_SII_PFB_TEVremoved	full length	pFastBac	Strep x 2, His	TEV	Amp	Hs	
JC205	EZH2_N688A	full length	pFastBac	His	TEV	Amp	Hs	N688A
JC206	EZH2_N688AH689A	full length	pFastBac	His	TEV	Amp	Hs	N688A, H689A
JC207	SUZ12_delN_LIC	278-end	pFastBac	His	TEV	Amp	Hs	
JC210	PHF1_SII_PFB_C_His_Strep_TEV	full length	pFastBac	Strep x 2,C-His, C-Strep	TEV	Amp	Hs	
JC214	EZH2_delN	39-end	pEC-A-3C-His_PFB	His	3C	Amp	Hs	
JC215	SUZ12_delN	71-end	pEC-A-3C-His_PFB	His	3C	Amp	Hs	
JC216	EED_delN	81-end	pEC-A-3C-His_PFB	His	3C	Amp	Hs	
JC217	RBBP4_delN	33-405	pEC-A-3C-His_PFB	His	3C	Amp	Hs	
JC218	RBBP4_3C	full length	pEC-A-3C-His_PFB	His	3C	Amp	Hs	
JC219	PHF1_wing1_A	full length	pFastBac	Strep x 2, His	TEV	Amp	Hs	K323A, K324A
JC220	PHF1_wing1_E	full length	pFastBac	Strep x 2, His	TEV	Amp	Hs	K323E, K324E
JC221	PHF19_wing1_A	full length	pFastBac	Strep x 2, His	TEV	Amp	Hs	
JC222	PHF19_wing1_E	full length	pFastBac	Strep x 2, His	TEV	Amp	Hs	
JC223	PHF1_helix3_A	full length	pFastBac	Strep x 2, His	TEV	Amp	Hs	R304A, S307A, N310A
JC224	PHF1_helix3_E	full length	pFastBac	Strep x 2, His	TEV	Amp	Hs	R304E, S307E, N310E
JC225	PHF19_helix3_A	full length	pFastBac	Strep x 2, His	TEV	Amp	Hs	H312A, H315A, N318A
JC226	PHF19_helix3_E	full length	pFastBac	Strep x 2, His	TEV	Amp	Hs	H312E, H315E, N318E
JC229	PHF19_wing1_helix3_A	full length	pFastBac	Strep x 2, His	TEV	Amp	Hs	K331A, K332A, H312A, H315A, N318A
JC230	PHF19_wing1_helix3_E	full length	pFastBac	Strep x 2, His	TEV	Amp	Hs	K331E, K332E, H312E, H315E, N318E
JC231	EZH2_delN_N688A	delN	pFastBac	His	TEV	Amp	Hs	N688A
JC236	EZH2_Y641F	full length	pFastBac	His	TEV	Amp	Hs	Y641F
JC237	EZH2_Y641FY726F	full length	pFastBac	His	TEV	Amp	Hs	Y641F, Y726F
JC240	PHF1_K304R	full length	pFastBac	Strep x 2, His	TEV	Amp	Hs	K304R
JC245	EZH2	full length	pFastBac	His	TEV	Amp	Hs	R561A, K569A
JC246	EZH2	full length	pFastBac	His	TEV	Amp	Hs	R527A
JC247	EZH2	full length	pFastBac	His	TEV	Amp	Hs	R527P
JC249	EZH2	full length	pFastBac	His	TEV	Amp	Hs	K307A, K309A
JC262	SUZ12_inserting TEVsite @ 289	full length	pFastBac	His	TEV	Amp	Hs	TEVsite @ 289

JC263	SUZ12_inserting TEVsite @ 389	full length	pFastBac	His	TEV	Amp	Hs	TEVsite @ 389
JC264	SUZ12_152-end	152-end	pFastBac	Strep x 2, His	TEV	Amp	Hs	
JC265	SUZ12_291-end	292-end	pFastBac	Strep x 2, His	TEV	Amp	Hs	
JC266	SUZ12_421-end	421-end	pFastBac	Strep x 2, His	TEV	Amp	Hs	
JC267	EZH2_K307AK309AY641F	full length	pFastBac	His	TEV	Amp	Hs	K307A, K309A, Y641F
JC270	PHF19_inserting TEV site @458	full length	pFastBac	Strep x 2, His	TEV	Amp	Hs	TEVsite @ 458
JC271	PHF1_c-term-2	506-end	pFastBac	Strep x 2, His	TEV	Amp	Hs	
JC272	PHF19_c-term-2	523-end	pFastBac	Strep x 2, His	TEV	Amp	Hs	
JC280	N-T4L-PHF1_C-term	515-end	pFastBac	Strep x 2, His	TEV	Amp	Hs	T4L insertion
JC281	N-T4L-PHF19_C-term	529-end	pFastBac	Strep x 2, His	TEV	Amp	Hs	T4L insertion
JC282	N-T4L-trSUZ12	71-end	pFastBac	HIS	TEV	Amp	Hs	T4L insertion
JC283	PHF1dloop_T4Linsertion	1-363, 511-end	pFastBac	Strep x 2, His	TEV	Amp	Hs	T4L insertion
JC284	PHF19dloop_T4Linsertion	1-377, 530-end	pFastBac	Strep x 2, His	TEV	Amp	Hs	T4L insertion
JC287	RBBP4_delN_TEV	33-405	pFastBac	His	TEV	Amp	Hs	
JC288	SUZ12_152-end_TEV	152-end	pFastBac	His	TEV	Amp	Hs	
JC289	SUZ12_291-end_TEV	291-end	pFastBac	His	TEV	Amp	Hs	
JC290	SUZ12_421-end_TEV	421-end	pFastBac	His	TEV	Amp	Hs	
JC291	N-BRIL-PHF1_C-term	515-end	pFastBac	Strep x 2, His	TEV	Amp	Hs	BRIL insertion
JC292	N-BRIL-PHF19_C-term	529-end	pFastBac	Strep x 2, His	TEV	Amp	Hs	BRIL insertion
JC293	N-BRIL-trSUZ12	71-end	pFastBac	HIS	TEV	Amp	Hs	BRIL insertion
JC294	PHF1dloop_BRILinsertion	1-363, 511-end	pFastBac	Strep x 2, His	TEV	Amp	Hs	BRIL insertion
JC295	PHF19dloop_BRILinsertion	1-377, 530-end	pFastBac	Strep x 2, His	TEV	Amp	Hs	BRIL insertion
JC296	Strep-PHF1(PHD2-WH)	186-363	pFastBac	Strep x 2, His	TEV	Amp	Hs	
JC301	PHF1_W211E	full length	pFastBac	Strep x 2, His	TEV	Amp	Hs	W211E
JC302	SUZ12_VEFS	535-685	pFastBac	His	TEV	Amp	Hs	
JC305	EZH2_107-129_259-end_Strep	107-129, 259-end	pFastBac	Strep x 2, His	TEV	Amp	Hs	
JC306	EZH2_107-160_259-end_Strep	107-160, 259-end	pFastBac	Strep x 2, His	TEV	Amp	Hs	
JC307	ybbR-EZH2	full length	pFastBac	His, ybbR	TEV	Amp	Hs	
JC308	StrepII-ybbR-EZH2	full length	pFastBac	Strep x 2, His, ybbR	TEV	Amp	Hs	