Structural basis of nucleosome binding by PRC2 and its regulation by histone modifications

Dissertation von Ksenia Finogenova

München 2021

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

Structural basis of nucleosome binding by PRC2 and its regulation by histone modifications

Ksenia Finogenova

aus

Moskau, Russland

2021

Erklärung

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

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, 19.01.2021

Ksenia Finogenova

Dissertation eingereicht am:	27.01.2021
1. Gutachterin/Gutachter:	Prof. Dr. Elena Conti
2. Gutachterin/Gutachter:	Prof. Dr. Karl-Peter Hopfner
Mündliche Prüfung am:	08.03.2021

Preface

Parts of this thesis were published in:

Structural basis for PRC2 decoding of active histone methylation marks H3K36me2/3.
K. Finogenova, J. Bonnet, S. Poepsel, I.B. Schäfer, K. Finkl, K. Schmid, C. Litz, M. Strauss, C. Benda, J. Müller.
Elife, Nov 19;9:e61964. doi: 10.7554/eLife.61964. PMID: 33211010, 2020.
`https://elifesciences.org/articles/61964`.

Pior to the peer-reviewed publication in Elife, parts of this thesis were published on the pre-print server bioRxiv:

bioRxiv 2020.04.22.054684; doi: https://doi.org/10.1101/2020.04.22.054684

And additionally presented in scientific conferences:

(1) Keystone Symposia Conference J3, Gene Regulation, From Mechanisms to Disease, 26-30 January 2020: talk and poster
(2) Virtual EMBL Conference: Transcription and Chromatin, 27-29 August 2020: talk
(3) Chromatin Dynamics Symposium: October 10-12, 2019: poster

Experimental data are deposited in databases under following accession numbers: EMD-11912 and EMD-11910, PDB 7AT8.

Dedicated to my father.

Contents

List of Figures

List of Tables

Abbreviations

The following list explains frequently used abbreviations.

aa = amino acids
AEBP2 = Adipocyte Enhancer-Binding Protein 2
app. = approximately
bp = base pair
CV = Column Volume
CXC = Zn3Cys8His + Zn3CysCys9 zinc sulfur clusters
DNA = Deoxyribonucleic Acid
D. melanogaster or *D.m.* or *Drosophila* = *Drosophila melanogaster*
DTT = 1,4-Dithiothreitol
E. coli = *Escherichia coli*
EDTA = Ethylenediaminetetraacetic Acid
EED = Embryonic Ectoderm Development
EM = Electron Microscopy
EMSA = Electrophoretic Mobility Shift Assay
EZH1/2 = Enhancer of Zeste Homolog 1/2
FL = Full Length
HMTase assay = Histone Methyltransferase assay
h/hrs = hour/hours
HOX genes = *homeobox* genes
JARID2 = Jumonji AT-Rich Interacting Domain 2
min = minutes
MS = Mass Spectrometry
Ni-NTA = Ni^{2+}-Nitrilotriacetic Acid
nuc = nucleosome
OD = Optical Density
ON = Over Night
PCR = Polymerase Chain Reaction
PHD = Plant Homeodomain
PHF1 = PHD Finger Protein 1
PMSF = Phenylmethylsulfonylfluorid
PRC2 = Polycomb Repressive Complex 2

PRC2-PHF1:Dinuc = PRC2-PHF1 on heterodimeric Dinucleosome
RBBP4/7 = Retino-Blastoma Binding Protein 4/7
RNA = Ribonucleic acid
rpm = revolutions per minute
RT = Room Temperature
SAH = *S*-Adenosyl-L-homocysteine
SAM = *S*-Adenosyl-L-methionine
SANT = Swi3, Ada2, N-Cor, and TFIIIB
SBD = SANT-1 binding domain
SD = Standard Deviation
SDS = Sodium Dodecyl Sulfate
SDS PAGE = Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SEC = Size Exclusion Chromatography
SEM = Standard Error of the Mean
SET = Su(var) 3-9, Enhancer of Zeste
SUZ12 = Suppressor of Zeste homolog 12
TCEP = Tris(2-carboxyethyl)phosphin
VEFS = VRN2-EMF2-FIS2-Su(z)12
WB = Western Blot
wt = wild type
X.l. = *Xenopus laevis*

Summary

Many histone modifications have chromatin modulating and gene regulating functions. Typically, the underlying mechanism involves 'writing' and/or 'reading' of the histone marks by chromatin regulating proteins. Alternatively, the histone marks interfere with these processes. These events ultimately result in an altered chromatin state.
Polycomb Repressive Complex 2 (PRC2) is a histone methyltransferase that mono-, di- and trimethylates histone H3 at lysine 27 (H3K27). Among the different methylation states, H3K27me3 marks the chromatin of Polycomb-regulated genes for transcriptional repression by another complex, called PRC1. The correct regulation of H3K27me3 deposition has emerged as a critical step in both animals and plants as defects in this process are directly linked to developmental abnormalities or diseases such as cancer. The methyltransfer is catalyzed by the PRC2 subunit EZH2 at its SET domain, the stereochemical properties of which limit the rate of di- to trimethylation. As a consequence, to generate high levels of H3K27me3 at its target genes PRC2 requires positive regulation. Such regulation can be allostery through recognition of H3K27me3 by the PRC2 subunit EED, recruitment mechanisms or association with different accessory subunits. In contrast, mono- and dimethylation marks seem to be deposited by PRC2 in a more serendipitous manner. Examples of negative PRC2 regulation are histone modifications such H3K4me3 and H3K36me2/3, which are catalyzed by different Trithorax/COMPASS complexes. These marks are characteristic for active genes and inhibit PRC2 in *cis* but not in *trans*. However, the molecular details of how PRC2 interpretes these pre-existing H3 tail modifications on the context of a nucleosome remain unknown.
This thesis employs single-particle cryo-electron microscopy for structural characterization of PRC2-PHF1 interaction with a heterodimeric dinucleosome, which contains one substrate nucleosome and one allosteric/H3K27me3-containing nucleosome. The obtained structure reveals that the CXC domain of PRC2 subunit EZH2 establishes binding to the nucleosomal DNA of the substrate nucleosome. A combined interface of EZH2$^{\mathrm{SBD/SANT1}}$ domains and EED recognizes the nucleosomal DNA on the allosteric nucleosome. Mutational analysis of these binding sites suggest that these interactions establish the subsequent productive recognition of the respective H3 tails by the active site of EZH2 and the β-propeller of EED. Furthermore, signal subtraction and focused 3D refinement procedures applied during cryo-EM data processing allow to extend previous structural descriptions to a model of the H3 tail stretching from the substrate nucleosome into the EZH2 active site. Unmodified H3K36 is found sandwiched in between EZH2 and nucleosomal DNA

in close proximity to the EZH2 CXC residues. Its ε-amino group is seemingly engaged in long range electrostatic interactions with the phosphate backbone of the nucleosomal DNA and in polar interactions with the carbonyl group of CXC residue Q570. Biochemical analyses showed that substitutions of H3K36 to either a shorter apolar alanine or a bulkier arginine side chain reduce the activity of PRC2, while binding of PRC2 to H3K36me3-nucleosomes is not affected. Taken together, results presented in this thesis suggest a model in which, within the time frame of PRC2 binding and reaction cycle, H3K36me2/3 hinders the subsequent optimal alignment of lysine 27 in the EZH2 active site. The bulkier quaternary ε-ammonium of H3K36me3, the positive charge of which is dispersed into the three additional methyl groups, thereby impedes but not entirely blocks PRC2 catalysis. Furthermore, it doesn't prevent PRC2 from binding to the nucleosome. This model of allosteric inhibition allows for a targeted and tuned activity regulation of PRC2 in dependence of the surrounding chromatin and the presence/absence of other activity regulating mechanisms.

Chapter 1

Introduction

1.1 The nucleosome - the "quantum"[1] of chromatin

The term "chromatin" was coined in 1882 by W. Flemming while working on nuclear division (Flemming, 1882; Paweletz, 2001; Hughes, 1959). Intriguingly, what today are known as basic components of chromatin were discovered around the same time: nucleic acids by F. Miescher in 1871 which he referred to as "nuclein" and described as phosphorus-rich acid (Miescher-Rüsch, 1871). And histones, which were discovered by A. Kossel in 1884 (Kossel, 1911). Yet the full extent of the importance of these early experiments would only become clear much later. W. Flemming correctly analyzed that what he called chromatin is likely nuclein or one is part of the other: "The word chromatin may stand until its chemical nature is known, and meanwhile stands for that substance in the cell nucleus which is readily stained" (Flemming, 1882; Hughes, 1959) (reviewed in Olins and Olins, 2003). Indeed, the word chromatin still "stands" and today it is known that nuclein or DNA is the carrier of genetic information. DNA can be transcribed into messenger ribonucleic acid (mRNA), which in turn is used to produce sequences of amino acids ultimately forming proteins as described by the central dogma of molecular biology (Crick, 1958, 1970).
In eukaryotic cells the DNA has to be folded, compacted and condensed approximately 10000x to fit in the nucleus and the word chromatin is now used to describe this high-order structure of the DNA. The structure of DNA was solved by X-ray crystallography and revealed a double helix (Watson and Crick, 1953; Wilkins et al., 1953; Franklin and Gosling, 1953). However, it took another 20 years to solve the chemical nature of chromatin and some of the first hints were provided by electron microscopy studies (Olins and Olins, 1974, 1973; Woodcock, 1973; Woodcock et al., 1976). These images of "beads on a string" (Fig. 1.1 A) and additional experiments based on nuclease digestion (Kornberg, 1974; Kornberg and Thomas, 1974) lead to the description of the basic structural unit of chromatin: DNA coiled around a histone core. In 1975 what D. E. Olins and A. L Olins in a review referred to as the "quantum" of chromatin received the name nucleosome (Oudet et al., 1975) (history of chromatin reviewed in Olins and Olins, 2003).

[1] Olins and Olins, 2003

The core nucleosome contains four different canonical histones: H2A, H2B, H3 and H4, which can also be substituted by the respective histone variants. Two copies of each histone form the histone octamer. The first structural studies of the nucleosome were limited to 7 Å (Richmond et al., 1984) and only in 1997 a high resolution X-ray crystallography structure was described (Fig. 1.1 B) (Luger et al., 1997). It revealed a disc-like structure of the octamer with a H3/H4 tetramer on which two dimers of H2A/H2B are assembled. Flexible N-terminal histone tails emerge on the sides of the nucleosome core. Approximately 147 base pairs (bp) of DNA are wrapped around the octamer in a left-handed matter. A highly negatively charged region named "acidic patch" is formed by the histone H2A residues E56, E64, E91, E92 and E102, E110 of histone H2B (Luger et al., 1997). The acidic patch serves as a recognition and binding hub for many transcriptional factors and chromatin regulating proteins such as SWI/SNF (SWItch/Sucrose Non-Fermentable) chromatin remodeller RSC (Remodeling the Structure of Chromatin) (Wagner et al., 2020) or the ATP-dependent chromatin remodeller INO80 (inositol auxotroph 80) (Eustermann et al., 2018) (acidic patch and recognition reviewed in McGinty and Tan, 2016; Kalashnikova et al., 2013).

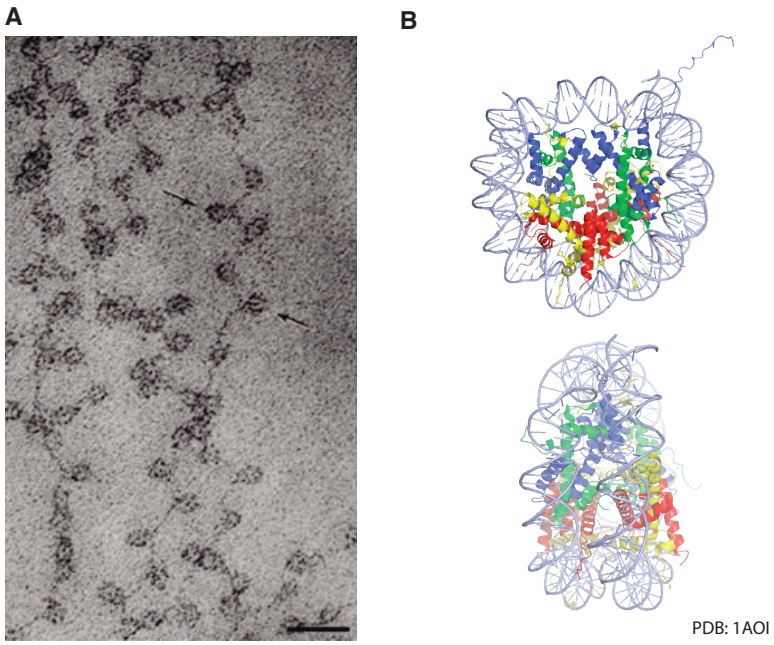

Figure 1.1: Structural studies on chromatin.

A. "Low ionic-strength chromatin spread, the 'beads on a string'. Size marker: 30 nm." (Olins and Olins, 2003). Shown figure and caption were taken from Olins and Olins, 2003. B. X-Ray crystal structure of the nucleosome solved by Luger et al., 1997, revealed the histone folds contributing to the octamer around which app. 147 bp of DNA are wrapped around. Blue: H3. Green: H4. Yellow: H2A. Red: H2B. PDB: 1AOI (Luger et al., 1997).

1.1.1 The histone code

In the 1940s, the scientific community largely accepted chromatin and chromosomes as structural units of our genes, however most of the scientist believed that histones are the carrier of the genetic code (history of chromatin reviewed in Olins and Olins, 2003). Today we know that DNA but not histones carries this information. Instead, histones can carry chromatin modulating and gene regulating posttranslational modifications and this is known as the "histone code" hypothesis (Strahl and Allis, 2000; Jenuwein and Allis, 2001; Rice and Allis, 2001). Many posttranslational modifications were described to date: phosphorylation (Nowak and Corces, 2004), methylation (Zhang and Reinberg, 2001), acetylation (Grunstein, 1997; Sterner and Berger, 2000), glycosylation (Liebich et al., 1993), ubiquitination (Davie and Murphy, 1990), sumoylation (Nathan et al., 2003), ADP ribosylation (Adamietz and Rudolph, 1984), biotinylation (Hymes et al., 1995) and carboxylation (Wondrak et al., 2000) and many more (Fig. 1.2) (reviewed e.g. in Turner, 2002). Together with other marks or processes such as DNA methylation histone modifications are collectively summarized within the term "epigenetics", first coined by C.H. Waddington (Waddington et al., 1940; Waddington, 1942, 2014, 2017). The rough meaning of this term translates into on 'top/beyond (traditional) genetics'.

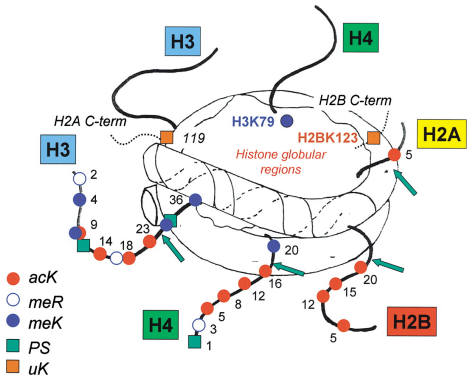

Figure 1.2: The histone code.

Histones can be found decorated with a variety of different posttranslational modifications, which are read by chromatin modifying enzymes and transcriptional regulators; few examples of possible but not exhaustive modifications from different organisms are shown in the figure. "Sites of posttranslational modification are indicated by colored symbols that are defined in the key (lower left); acK, acetyl lysine; meR, methyl arginine; meK, methyl lysine; PS, phosphoryl serine; and uK, ubiquitinated lysine." (Turner, 2002). Illustration and part of the caption were taken from Turner, 2002.

Posttranslational modifications influence the transcriptional states of the surrounding genes. The underlying mechanisms involve either the influence of a pre-existing modification mark directly on the chromatin via its charge or on the establishment of another histone mark on the same or neighboring histone tail. Alternatively, they involve chromatin modifying proteins which 'read' the modification mark established. Enzymes depositing or

removing these marks are accordingly called the 'writers' or 'erasers', respectively.

1.2 Polycomb-Trithorax antagonism and its role in developmental processes

1.2.1 Anterior-posterior development in *Drosophila melanogaster* embryo

The correct development of the *Drosophila* embryo along the anterior-posterior axis involves two key steps: (1) anterior-posterior cytoplasmic polarity established by maternal proteins and mRNAs (maternal effect) and (2) correct anterior-posterior segmentation of the *Drosophila* embryo body (reviewed in Gilbert, 2014). Segmentation is established by a cascade of regulating genes and their protein products: maternal proteins activate gap-genes, protein products of which in turn regulate pair-rule genes (Nüsslein-Volhard and Wieschaus, 1980). The unique constellation of the activated and silenced gap- and pair-rule genes within each segment in the next step establishes a differential expression of homeotic genes (*HOX* genes) (Fig. 1.3). *HOX* genes together with downstream regulator genes ultimately specify the identity of the given segment (Lewis, 1978; Levine and Harding, 1989). Once established, the specified identity has to be maintained as gene products of the described gene cascade are only transiently expressed (reviewed in Gilbert, 2014).

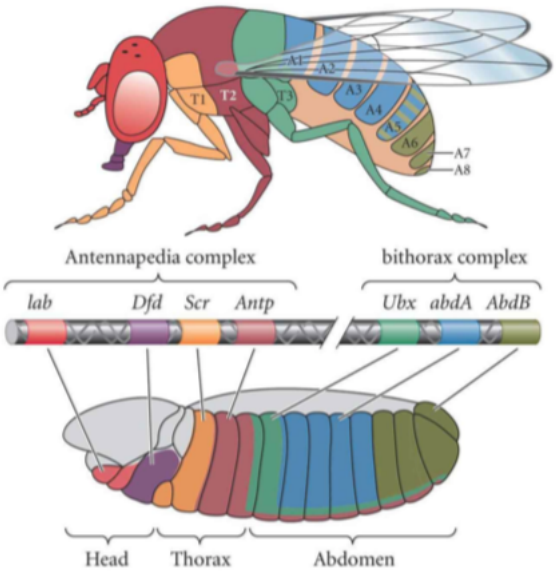

Figure 1.3: Body segmentation and *HOX* gene expression of the *Drosophila* adult and embryo body.

Segment identity of the *Drosophila* embryo and adult body is established by a differential expression of *HOX* genes. The majority of the *HOX* genes are encoded in two regions of the *Drosophila* chromosome III. The first region is summarized as the Antennapedia complex and involves 5 genes: *labial* (*lab*) and *deformed* (*Dfd*) regulate head segments, while *sex comb reduced* (*Scr*) and *Antennapedia* (*Antp*) specify the thoracic segments. The fifth gene is *proboscipedia* (*pb*) (not shown in the scheme) and is active only in the adult *Drosophila* (Wakimoto et al., 1984; Kaufman et al., 1980; Maeda and Karch, 2009). The second region is summarized as the bithorax complex and involves 3 genes: *Ultrabithorax* (*Ubx*) regulates the 3rd thoracic segment, while *abdominal A* and *B* (*abdA*/*AbdbB*) specify the abdominal segments (Lewis, 1978; Maeda and Karch, 2009). Together, the antennapedia and bithorax complexes are also referred to as homeotic complex (Schuettengruber et al., 2009). Intriguingly, the physical order of the segments mirrors the order of the respective encoding genes on the chromosome. Illustration was taken from Gilbert, 2014.

1.2.2 Polycomb-Trithorax antagonism

The balance between the correct activation of some and silencing of other *HOX* genes in a given segment is regulated by the antagonistic actions of the Polycomb and Trithorax machinery (Ingham and Whittle, 1980; McKeon and Brock, 1991; Simon et al., 1992). In simple terms, transcription of developmental genes is activated by Trithorax, while Polycomb acts as a repressor (Schuettengruber et al., 2007). On the molecular level Trithorax and Polycomb catalyze different histone modifications, which, intriguingly, are often mutually exclusive. For example, in higher metazoans the Trithorax proteins Ash1 and SET2 catalyze the deposition of H3K36me2 and H3K36me3, respectively, in actively transcribed genes (Schmähling et al., 2018; Bell et al., 2007; Gaydos et al., 2012; Larschan et al., 2007) and this mark was found to inhibit Polycomb group repressive complex 2 (PRC2) on the same H3 tail (Klymenko and Müller, 2004; Schmitges et al., 2011). The Trithorax-

Polycomb antagonism is hence a prime example for a writer-reader mechanism (reviewed in Margueron et al., 2005).

1.2.3 Polycomb Group proteins

The name "Polycomb" (Pc) (Lewis, 1978) was introduced in 1978 when mutational studies lead by Ed Lewis in *Drosophila* resulted in a dominant phenotype of an ectopic appearance of a sex comb, a row of distinctive bristles. These bristles are usually present only on the first pair of thoracic legs of an adult male fly (T1; Fig. 1.3) (Lewis, 1978). The additional appearance of these bristles on the second and sometimes third legs was the product of a loss of repression of the *HOX* gene *Scr*. Over the following years several studies discovered that additional genes exert a similar function and give rise to mechanistically and functionally similar phenotypes of homeotic transformations upon mutation. Such genes were *extra sex comb* (*esc*) as discovered by Gary Struhl, 1981 or later *Polycomb like* (*Pcl*) as discovered by Ian M. Duncan, 1982. These and 20 more genes were later coined Polycomb Group genes (PcG) due to their function in segmental organization (Jürgens, 1985; Simon et al., 1992).
The majority of PcG members are components of four main complexes. Biochemical purifications from *Drosophila* revealed that these complexes are (1) Polycomb repressive complex 1 (PRC1) with the function of histone H2A mono-ubiquitination at lysine 119 (H2AK119ub) and chromatin compaction (Shao et al., 1999; Wang et al., 2004a; Francis et al., 2001, 2004; Grau et al., 2011), (2) Polycomb repressive complex 2 (PRC2) catalyzing methylation at lysine 27 (Czermin et al., 2002; Müller et al., 2002) and (3) Polycomb Repressive DeUBiquitinase (PR DUB) with the opposing function to PRC1, the removal of K119 ubiquitination (Scheuermann et al., 2010) and (4) DNA binding complex Pho repressive complex (PhoRC) (Klymenko et al., 2006). This thesis is focused on the regulation of PRC2.

1.3 Polycomb repressive complex 2 (PRC2) - a histone methyltransferase

PRC2 is a *S*-Adenosyl-L-methionine (SAM) dependent histone methyltransferase responsible for mono-, di- and trimethylation of H3K27 (Müller et al., 2002; Czermin et al., 2002; Kuzmichev et al., 2002; Cao et al., 2002). The core of mammalian PRC2 corresponds to about 250 kDa and constitutes the catalytic subunit enhancer of zeste homolog 1/2 (EZH1/EZH2), the embryonic ectoderm development (EED), suppressor of zeste homolog 12 (SUZ12) and retinoblastoma-binding protein 4 or 7 (RRBP4/7 also referred to as RbAp48/46). PRC2 is often found associated with a number of different protein cofactors: adipocyte enhancer-binding protein 2 (AEBP2), jumonji AT-rich interacting domain 2 (JARID2 or Jumonji), polycomb-like proteins (PCLs) plant homeodomain factor 1 (PHF1, also Pcl1), metal regulatory transcription Factor 2 (MTF2) (also Pcl2), PHF19 (also Pcl3) and mammalian-specific Elongin BC Polycomb Repressive Complex 2-associated Protein

(EPOP) with PRC2-associated LCOR Isoform 1 and 2 (PALI1/2).

The methyltransferase activity of PRC2 is highly important for repression of Polycomb target genes (PcG genes) such as *HOX* genes so that a single point mutation of H3K27 in *Drosophila* can reproduce the depletion phenotype of EZH2 (Pengelly et al., 2013; McKay et al., 2015). While PRC2 can deposit all three methyl marks on lysine 27, *in vivo* the deposited mark is predominantly dimethylation, which amounts to 50-70% of H3 tails in mES cells (Ferrari et al., 2014) and in *Drosophila* (Lee et al., 2015). In contrast, mono- and trimethylation are less prevalent, 5% H3K27me1 and 5-10% H3K27me3 (Ferrari et al., 2014; Jung et al., 2010; Lee et al., 2015; Voigt et al., 2012). In early developing *Drosophila* embryo levels of H3K27me3 even amount only to 2% and with later stages of development increase to 7.5 % (Bonnet et al., 2019). This variance in degree of methyl groups seems to also have different roles in the cell: H3K27me1 is often found in bodies of actively transcribed genes, while H3K27me2 is very rapidly deposited during DNA synthesis and is predominantly found on enhancers and intergenic regions (Ferrari et al., 2014; Lee et al., 2015; Sneeringer et al., 2010). A possible function of H3K27me2 at these genomic sites was proposed to be the counteraction of the antagonistic active mark H3K27ac to prevent erroneous activation of enhancers (Ferrari et al., 2014). Indeed, reduction of PRC2 in the cell was often found to be associated with an increase of global H3K27ac (Tie et al., 2009). In contrast, trimethylation is a slow reaction and in ES cells predominantly found at promoters, especially bivalent promoters (Bernstein et al., 2006; Voigt et al., 2013). Moreover, Polycomb target genes are often found to be extensively decorated with H3K27me3 and at least partial generation of these extended long regions of H3K27me3 on target genes was found to be a prerequisite for sufficient repression (Bernstein et al., 2006; Papp and Müller, 2006; Schwartz et al., 2006). H3K27me3 mark is furthermore recognized by the CBX (in *Drosophila* Pc) subunit of PRC1 responsible for H2Aub catalysis. An additional function of chromatin compaction is ascribed to PRC1 and these concerted actions of PRC2 and PRC1 together form the classical model of Polycomb group repression (Cao et al., 2002; Czermin et al., 2002; Kuzmichev et al., 2002; Müller et al., 2002; Francis et al., 2001; Grau et al., 2011). Correct regulation of PRC2 activity is therefore highly crucial for gene regulation.

1.3.1 Structure and function of the individual core PRC2 components

EZH2

The catalytic subunit EZH2 belongs to the Ez family of histone methyltransferases. The importance of EZH2 was demonstrated in a study where knockout of *EZH2* resulted in mouse embryonic lethality (O'Carroll et al., 2001). The catalytic activity is provided by the SET (su(var) 3-9, enhancer of Zeste (E(z)) and Trithorax) domain (Fig. 1.4 A), however, EZH2 requires other PRC2 subunits for activation (Cao and Zhang, 2004b). The structural basis for this observation was provided by crystal structures of the single EZH2

SET domain, which demonstrated that unlike other homologous SET domains, such as DIM-5, it adopts an autoinhibitory state (Fig. 1.4 B) (Wu et al., 2013; Antonysamy et al., 2013). In this state the SAM cofactor binding pocket and the substrate entry side are not properly folded thereby physically hindering catalysis.

EED

The subunit EED at its C-terminal WD40 domain adopts the characteristic fold of a beta-propeller commonly consisting of seven blades, four-stranded antiparallel β-sheets and in addition contains an 80 residues subdomain at its N-terminus (Smith et al., 1999; Tie et al., 2007) (Fig. 1.4 C). A number of crystal structures of EED have been solved over the last years, most of which showed the recognition of trimethylated peptides such as H3K27me3 by the aromatic cage of the WD40 beta-propeller (Margueron et al., 2009) (reviewed in Huang et al., 2017). The recognition of the own catalytic product has been shown to allosterically activate PRC2 and the biological significance of this mechanism will be discussed in subsequent sections of this thesis (Margueron et al., 2009). The importance of presence of EED in the PRC2 complex is demonstrated in studies where point mutations in EED result in embryonic lethality due to disruption of the EZH2-EED binding interface (Denisenko et al., 1998). Structural characterization showed that this interface is provided by a narrow binding groove on the surface of EED onto which the N-terminus of EZH2 forms a long α-helix (Han et al., 2007).

RBBP4/7 and SUZ12

RBBP4 and its paralog RBBP7 were initially discovered as retinoblastoma-associated proteins and are part of many different chromatin-modifying complexes such as the ATP-dependent nucleosome remodeling deacetylase (NuRD) or the chromatin assembly factor Caf1 (reviewed in Holoch and Margueron, 2017; Huang et al., 2017). The structure of *D.m.* ortholog of RBBP4/7, Nurf55, revealed an N-terminal helix with a seven stranded WD40 beta-propeller as for EED (Song et al., 2008) (Fig. 1.4 D) and similarly to EED, Nurf55 was shown to be able to recognize histone tails. The function and relevance of the histone binding activity of RBBP4 and the differences between the paralogs within PRC2 remain poorly understood, however the presence of this subunit within the PRC2 complex was shown to enhance PRC2 activity (Cao and Zhang, 2004a,b; Ferrari et al., 2014). A crystallographic study demonstrated how Nurf55 recognizes the unmodified N-terminus of the H3 tail spanning residues 1-14 and a region of Su(z)12 (*D.m.* homolog of SUZ12), comprising residues 73-143 (Schmitges et al., 2011) (Fig. 1.4 D). SUZ12 is bound predominantly via hydrophobic interactions on the side of the propeller sandwiched between the PP (pyrophosphate binding) loop and the α-helix 1 (Verreault et al., 1998) (Fig. 1.4 D). SUZ12 is an integral part of the PRC2 complex as the removal of its C-terminal VEFS (VRN2-EMF2-FIS2-Su(z)12) domain disrupts PRC2 complex formation (Kuzmichev et al., 2002). In addition to the VEFS domain, SUZ12 harbors a Cys_2-His_2 zinc finger (Fig. 1.4 A), which counter-intuitively was shown to not promote DNA binding *in vitro* but was

instead proposed to be required for mediating protein-protein interactions within the complex (Müller et al., 2002).

1.3.2 Structural characterization of PRC2 complex architecture

Ternary complex of EZH2-EED-SUZ12^{VEFS}

The structures of *Chaetomium thermophilum* EZH2-EED-SUZ12^{VEFS} in stimulated and basal state showed the overall organization of the ternary complex (Jiao and Liu, 2015) and were later confirmed by similar structures of the human (Justin et al., 2016) and human/american chamelion complexes (Brooun et al., 2016) (Fig. 1.4 E). In the structures, the N-terminal domain of EZH2 forms a long helix on the surface groove of the beta-propeller of EED, as previously described (Han et al., 2007). The very N-terminus interacts with the downstream SANT1 domain of EZH2 (SANT1 binding domain/SBD). At the back of the ternary complex, EZH2 forms a loop referred to as SET activation loop (SAL), which is essential for proper activation of EZH2 (Jiao and Liu, 2015). The SAL loop is followed by a stimulation response motif (SRM). This motif owes its name to the observation that it is only formed upon binding of the allosterically activating H3K27me3 peptide to the EED beta-propeller (active state; see subsection 1.3.1 on how EED recognizes the tri-methyl-ammonium group). In contrast, in the basal state of the ternary complex (absence of H3K27me3 peptide in the EED) the SRM is disordered and flexible. The same allosteric activation mechanism involving the SRM was described in the human ternary complex (Justin et al., 2016). In this case, EZH2 was activated by the binding of a trimethylated protein cofactor JARID2 K116me3 peptide in the aromatic cage of the EED beta-propeller (Sanulli et al., 2015). Once properly folded, the SRM is in contact with the activating peptide and the catalytic-proximal SET-1 region, which likely transfers the allosteric activation to the active site (Jiao and Liu, 2015)(Justin et al., 2016).
Following the SANT1/2 domains is a highly conserved region harboring zinc-sulfur clusters (Zn_3Cys_8His + Zn_3Cys_{Cys9}/CXC). The function of this domain was proposed to be DNA binding as it is evolutionary related to the CXC domain in the *Drosophila* dosage compensation complex responsible for recruitment to the X-chromosome (Zheng et al., 2012). Furthermore, missense mutation in the zinc-coordinating residues were previously described to abolish the methyltransferase activity and are characteristic of myelodysplastic syndromes (Ketel et al., 2005; Ernst et al., 2010). The exquisite intricate ternary structure of EZH2 with EED and SUZ12^{VEFS} and the number of interactions provide structure-based explanations for the observation that EZH2 requires the presence of both subunits for proper activation and folding (Müller et al., 2002; Ketel et al., 2005; Yamamoto et al., 2004).

Aspects of regulation of the SET domain. A closer look at the human catalytic SET domain reveals the differences of its active state within the ternary complex in comparison to the structure of the sole autoinhibited SET domain (Justin et al., 2016) (Fig. 1.4 B, E). In the active state, the SET-1 region is rotated several degrees counterclock-

wise thereby opening the binding groove for the H3 tail and completing the SAM-binding pocket. Contacts of the SAL loop to SET-1 seem to stabilize the active conformation of SET, thereby seemingly mediating activating impulses from $SUZ12^{VEFS}$ and EED to the catalytic site (Jiao and Liu, 2015). Further stabilization is provided through binding of *S*-adenosyl-L-homocysteine (SAH, as used in the structures) in a hydrophobic channel of the SET domain. The relative position of the SAH thioether with regard to the hypothetical position of the lysine ε-amino group reflects a catalytically competent state for methyl transfer (Justin et al., 2016; Jiao and Liu, 2015)
H3K27M is a mutation often found in pediatric brain tumors (see also section 'PRC2 in disease' 1.6). In the human PRC2 crystal structure the methionine is in place of the lysine 27 mimicking its aliphatic portion of the side chain and is recognized by several aromatic residues of EZH2 (Tyr641, Phe724, Tyr 726, Tyr728, Phe667). H3R26 is engaged in ionic interaction with Gln648 and Asp652 (Justin et al., 2016).
The chemistry of the methyl transfer from the SAM (*S*-adenosyl-L-methionine) methyl-donor to the epsilon nitrogen of the lysine is a bimolecular nucleophilic substitution (S_N2) group transfer reaction resulting in SAH (*S*-adenosyl-L-homocysteine) and N-methylated lysine (Smith and Denu, 2009). While theoretically a lysine can be mono-, di- and trimethylated in place of its three hydrogens, not all SET methyltransferases can catalyze the three steps. Structural comparison of EZH2 SET to other SET domains, where only mono- and dimethylation is possible, revealed that space limitations and establishment of hydrogen bonds (from the aromatic amino acid residues to the substrate) differ in the other SET domains and thereby regulate the different degrees of catalysis (Cheng et al., 2005) (reviewed in Liu and Zhu, 2017).
Finally, another level of SET domain regulation is provided by the automethylation function of EZH2 (Lee et al., 2019; Wang et al., 2019). Residues K505, K509, K510 located in a loop region between CXC and SANT2 were proposed to occupy the substrate pocket prior to methylation catalyzed by EZH2. Regulation of EZH2 by automethylation is thereby reminiscent of kinase autopohosphorylation and histone acetyltransferase autoacetylation mechanisms (Beenstock et al., 2016; Thompson et al., 2004).

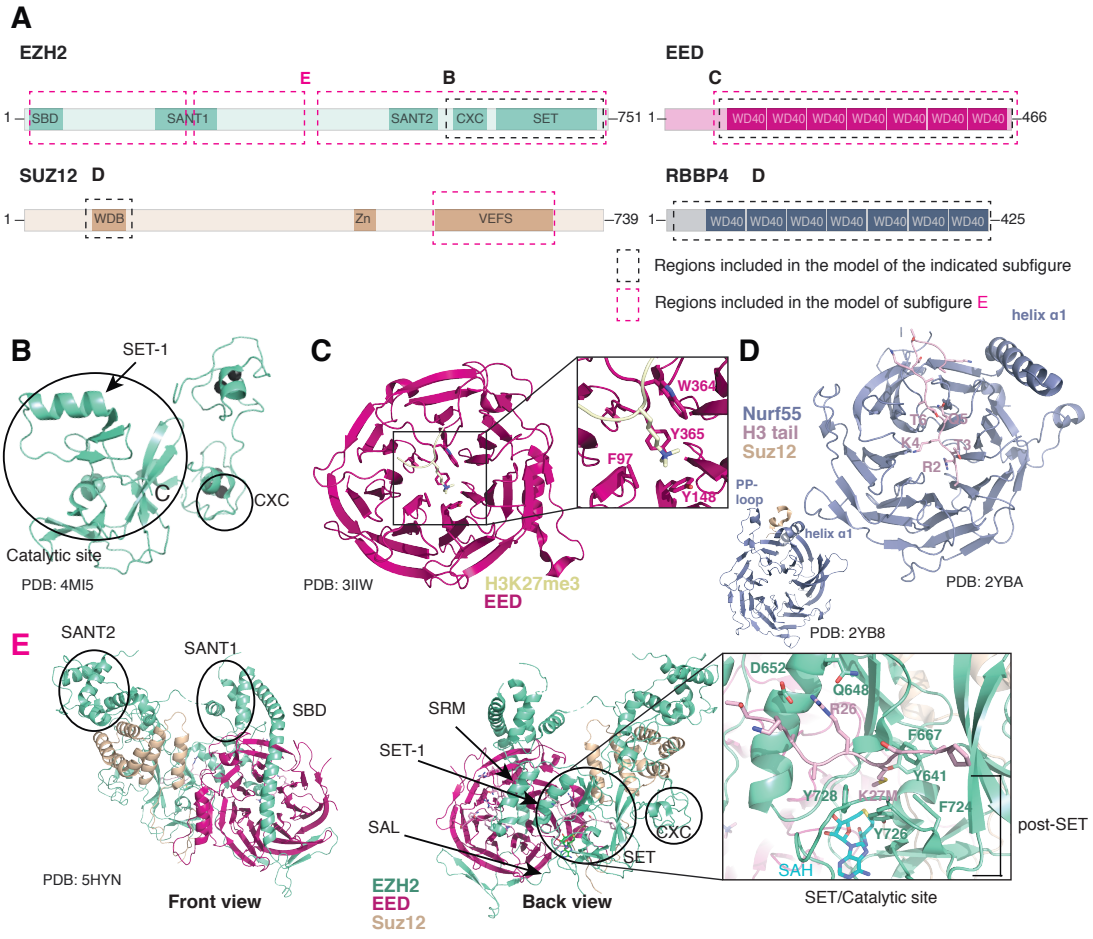

Figure 1.4: Structural studies on individual subunits and the ternary complex of PRC2.

A. Overview of the core PRC2 subunits domain organization. Dotted bars indicate regions present in the structures shown in the following subfigures. B. X-Ray crystal structure of the EZH2 SET domain shows an autoinhibitory fold (PDB: 4MI5)(Antonysamy et al., 2013). Also shown in Wu et al., 2013. C. H3K27me3 recognition by the aromatic cage of EED (PDB 3IIW) (Margueron et al., 2009). D. H3 tail and SUZ12 recognition by Nurf55 (*D. m.* ortholog of RBBP4) E. Ternary structure EZH2-EED-SUZ12$^{\text{VEFS}}$ (shown in example of the human crystal structure: PDB: 5HYN, Justin et al., 2016) reveals the intricate interaction network within the subunits. A H3 peptide containing the H3K27M mutation is inserted into an aromatic cage of the SET domain.

Cryo-EM studies of AEBP2-PRC2

The first structural characterization of the full five subunits PRC2-AEBP2 complex was obtained by negative stain electron microscopy (EM) and its resolution was limited to 21 Å (Ciferri et al., 2012). Together with chemical crosslinking the structure revealed the overall spatial organization of the full complex with an 'upper lobe' containing the ternary EZH2-EED-SUZ12 (C-terminus) complex (Justin et al., 2016) and the 'bottom lobe' with

SUZ12 (N-terminus) and RBBP4 (Ciferri et al., 2012). Recently, high resolution cryo-EM structures were obtained of BS3 crosslinked PRC2-AEBP2-JARID2 revealing detailed structural organization of the 'bottom lobe' (Fig. 1.5 A, B). The accessory subunit AEBP2 was previously shown to stabilize the core PRC2 complex and additionally enhance the activity of PRC2 (Youmans et al., 2018; Kalb et al., 2014; Son et al., 2013). In the structure, residues K294 and R295 of AEBP2 mimic histone H3 recognition by RBBP4 previously illustrated in a crystallographic study (Fig. 1.4 D) (Schmitges et al., 2011). Residues 263-283 of AEBP2 are involved in interactions with SUZ12 (426-548) (Kasinath et al., 2018). The N-terminus of AEBP2 is in close proximity to the EZH2 SET domain, which potentially indicates an allosteric transmitter function of AEBP2, acting between RBBP4 and SUZ12 and the active site (Fig. 1.5 A, B). A JARID2 peptide is bound in the catalytic site of EZH2 providing structural insights into JARID2 mimicry of the H3 tail in the active site previously described *in vitro* and *in vivo* (Sanulli et al., 2015). Both, the H3 tail and JARID2 peptide, once trimethylated by PRC2, stimulate the catalytic activity of PRC2 in a positive feedback loop via recognition of EED (Sanulli et al., 2015; Margueron et al., 2009; Jiao and Liu, 2015; Justin et al., 2016; Kasinath et al., 2018).

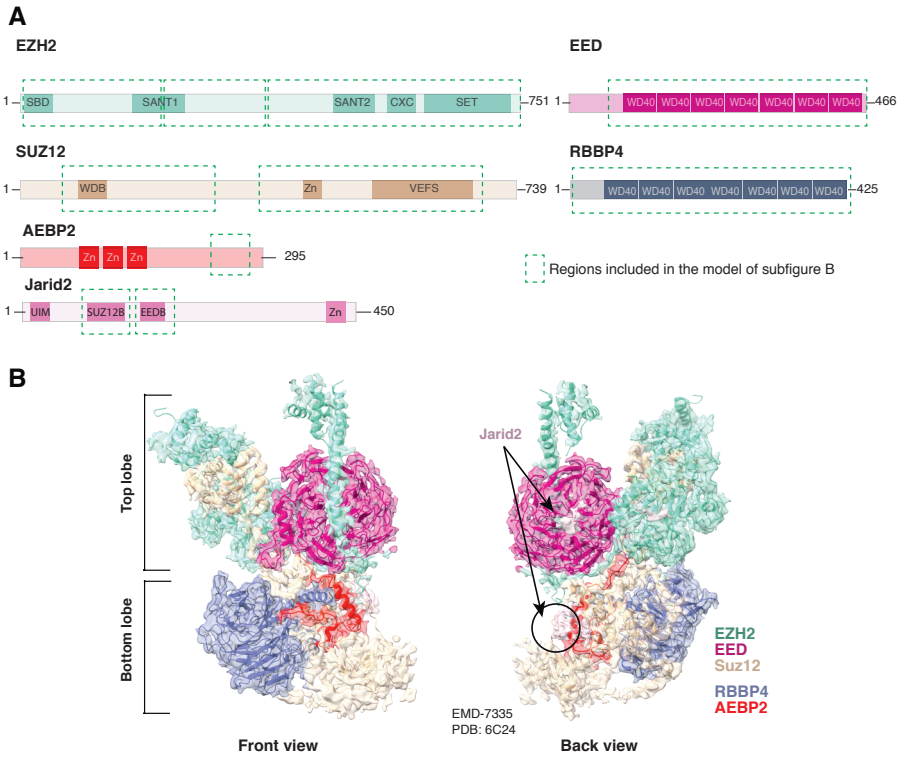

Figure 1.5: Cryo-EM structure of PRC2-AEBP2-JARID2.

A. Overview of the components and protein cofactors of PRC2 and their domain organization. Dotted bars indicate regions present in the structures shown in the following subfigure. B. Cryo-EM structure of PRC2-AEBP2 reveals the subunit organization within the full-length PRC2-AEBP2-JARID2 complex (Kasinath et al., 2018).

1.4 Role of PRC2 cofactors

High levels of H3K27me3 are prerequisite for the repressive action of Polycomb. Owing to the slower enzymatic reaction of trimethylation PRC2 requires 1) enhancement of its enzymatic reaction and/or 2) a more prolonged and more stable association with chromatin (reviewed in Holoch and Margueron, 2017; Liu and Zhu, 2017). Binding of PRC2 to accessory subunits was proposed to be one of the mechanisms of how PRC2 achieves generation of high levels of H3K27me3. While some aspects were already mentioned previously, this section is dedicated to the in-depth analysis of the role of some of the protein cofactors within PRC2 regulation. The facultative subunits of PRC2 can be subdivided into two mutually exclusive PRC2 classes, PRC2.1 and PRC2.2 and the cofactors present on the PRC2 complex differ depending on the developmental stage or in different cell types (Grijzenhout et al., 2016; Hauri et al., 2016) (overview shown in Fig. 1.6 A). Furthermore, in mES and human induced pluripotent stem cells PRC2.1 and PRC2.2 were found to act at their target genes synergistically, while specific depletion of activity of the

one did not affect activity of the other and vice versa (Healy et al., 2019; Højfeldt et al., 2019; Youmans et al., 2020). In PRC2.1, the core PRC2 complex is associated with one of the Polycomb-like proteins PHF1, MTF2, PHF19 (also called Pcl 1-3, Pcl in *Drosophila*) or with PALI1/2 or EPOP (Fig. 1.6 A). PRC2.2 is in complex with AEBP2 (Jing in *Drosophila*), JARID2 (Jarid2/Jumonji in *Drosophila*), or both simultaneously (reviewed in Aranda et al., 2015). Within the subcomplexes, the following pair of protein cofactors were found to be mutually exclusive: Pcl proteins and AEBP2 (Grijzenhout et al., 2016; Hauri et al., 2016; Chen et al., 2020), PALI1 and AEBP2 (Hauri et al., 2016; Conway et al., 2018) and EPOP, PALI1 and JARID2 (Hauri et al., 2016; Chen et al., 2020; Conway et al., 2018; Alekseyenko et al., 2014; Beringer et al., 2016; Liefke and Shi, 2015). Intriguingly, the majority of these cofactors seems to bind the 'bottom lobe' of PRC2 consisting of the N-terminal region of SUZ12 and RBBP4 (Chen et al., 2018, 2020; Youmans et al., 2020). Cancer mutations in SUZ12 were found to favor formation of the PRC2.1 complex over PRC2.2 with a subsequent increase of PRC2 chromatin association and H3K27me3 levels (Youmans et al., 2020).

The following sections will elucidate selected aspects of the molecular architectures and functions of Pcl proteins, PALI1/2 of PRC2.1 and AEBP2 and JARID2 of PRC2.2.

1.4.1 PHF1, MTF2 and PHF19

PHF1 (Pcl1), MTF2 (Pcl2), PHF19 (Pcl3) are mammalian homologues of *D. melanogaster* Pcl (Nekrasov et al., 2007; Cao et al., 2008; Sarma et al., 2008). In *Drosophila*, Pcl is required to maintain repression of Polycomb target genes: loss or mutations of *Pcl* lead to a reduction of H3K37me3 and a phenotype similar to *Pc* mutants (Nekrasov et al., 2007; Savla et al., 2008). In mammalians, depletion of Polycomb-like proteins is also accompanied by concomitant decrease of H3K37me3 (Cao et al., 2008; Sarma et al., 2008; Casanova et al., 2011; Hunkapiller et al., 2012). Furthermore, presence of Pcl proteins on PRC2 enhances its activity as compared to the core PRC2 complex (Cao et al., 2008; Sarma et al., 2008; Nekrasov et al., 2007). Pcl proteins are characterized by the following domains (Fig. 1.6 B): a Tudor, two PHD, one Extended Homology/Winged-Helix (EH/WH) and finally the C terminal RC/CL (Reversed Chromo/Chromolike domain) domains (Ballaré et al., 2012; Choi et al., 2017; Li et al., 2017) (Fig. 1.6 B). The structure of the *D. melanogaster* Pcl PHD and DNA binding WH domains was solved by X-ray crystallography (Fig. 1.6 B) and showed a canonical WH and an atypical aromatic cage of the PHD (Choi et al., 2017). Binding of PRC2 to both DNA and mononucleosomes *in vitro* showed an increase of 2-2.5x of affinity and residence time on oligonucleosomes was extended (as measured by single molecule TIRF microscopy) in presence of PHF1 (Choi et al., 2017). Mutations in the WH domain decreased both affinity and residence time to a level comparable to PRC2 lacking PHF1 concluding that WH domain is necessary for binding (Choi et al., 2017). Interestingly, despite Pcls repressive function at the Polycomb target genes, its WH domain was shown to lack specificity for binding to PREs. The combination of these results shows convincingly that presence of PHF1 on PRC2 promotes H3K27me3 by a

prolonged DNA binding and hence likely more stable chromatin association but not by specific targeting to PREs (Choi et al., 2017). In addition to WH and PHD2 Pcl proteins contain a second PHD domain (PHD1) and a Tudor domain. Crystallographic studies of the human PHF1 Tudor domain revealed how its aromatic cage recognizes the active mark H3K36me3 (Musselman et al., 2012). PRC2-PHF1 was found to be inhibited on yeast-isolated H3K36me3 containing chromatin, however it remains unclear what contribution is made by the Tudor domain of PHF1 (Musselman et al., 2012, 2013). In contrast, the *Drosophila* Tudor domain shows an atypical aromatic cage and hence is not able to bind H3K36me3 (Friberg et al., 2010). The C-terminus of PHF1 (Fig. 1.6 B; in other literature also named Reversed Chromo/Chromo-like domain) was reported to be sufficient for establishing interaction with PRC2 (Choi et al., 2017). While no structure is available of the PRC2 interaction with PHF1, structural studies on PHF19 suggest that Pcl proteins presumably bind to the N-terminal part of SUZ12 recognized also by AEBP2 (Chen et al., 2020; Kasinath et al., 2018).
Furthermore, the structures of SUZ12-RBBP4 and short constructs encompassing SUZ12 interacting regions of PHF19 revealed that the C-terminal RC domain of PHF19 promotes dimerization via a Dimer Stabilization helix (DS) (Chen et al., 2020). PRC2 dimers have been described previously, but were lacking the structural details (O'Connell et al., 2001; Tie et al., 2003; Margueron et al., 2008; Casanova et al., 2011; Ballaré et al., 2012; Son et al., 2013; Davidovich et al., 2014; Grijzenhout et al., 2016). In the structure, PHF19 does not participate in the dimer formation but rather stabilizes dimerization of SUZ12-RBBP4. Importantly, the described DS helix seems to be conserved in the *Drosophila* Pcl, human MTF2 but not in PHF1.

1.4.2 AEBP2

The domain architecture of AEBP2 reveals three potentially DNA binding Zn finger domains (Kim et al., 2009; Wang et al., 2017) and a C-terminal region, which harbors a nucleosome association enhancing lysine and arginine rich motif (KR motif) (Lee et al., 2018) and PRC2 and nucleosome binding C-terminal (CT) domain (Cao and Zhang, 2004b; Chen et al., 2018) (Fig. 1.5 A). AEBP2 binds to PRC2 via SUZ12 as already discussed in section 1.3.2 of this thesis (Kasinath et al., 2018). As a zinc finger protein, AEBP2 was proposed to recruit PRC2 to target genes via DNA recognition (Cao et al., 2002; Cao and Zhang, 2004b). However, studies have shown that the DNA sequence preference of AEBP2 is highly degenerative (Kim et al., 2009). Furthermore, a recent pre-print showing a cryo-EM structure of PRC2-AEBP2-JARID2 on a ubiquitinated nucleosome (H2AK119ub) found that in presence of the ubiquitin moiety, AEBP2 Zn fingers interact with the histone core of the nucleosome (Kasinath et al., 2020). Instead, in this structure, the KR motif of AEBP2 shows interactions with nucleosomal DNA (Kasinath et al., 2020). The observed DNA binding contribution by AEBP2 KR motif could potentially affect residence time similar as to PHF1 and accordingly, presence of AEBP2 has been shown to enhance the activity of PRC2 (Youmans et al., 2018; Kalb et al., 2014; Son et al., 2013) (Fig. 1.6 A). In negative stain and in cryo-EM, association of PRC2 to AEBP2 has been shown to

stabilize the complex (Ciferri et al., 2012; Kasinath et al., 2018). Further discussion of the AEBP2 contribution to PRC2 regulation will be provided in the following section 'PRC2 and chromatin' 1.5 of this thesis.

1.4.3 JARID2 and PALI1/2

While belonging to different PRC2 subcomplexes, PALI and JARID2 share common features: both cofactors were reported to enhance activity of PRC2 *in vitro* and *in vivo* (Sanulli et al., 2015; Zhang et al., 2020). Both proteins are a substrate of PRC2 through H3 tail mimicry (JARID2 at K116; PALI1 at K219; PALI2 at K124) and once trimethylated, subsequently stimulate the activity of the complex through binding to EED (Sanulli et al., 2015; Margueron et al., 2009; Zhang et al., 2020) (Fig. 1.6 A). JARID2 has a JmjC (Jumonji) histone demethylase domain, which, however, has lost residues required for catalytic activity (Cloos et al., 2008; Landeira and Fisher, 2011). Intriguingly, JARID2 binds to EED with a higher affinity than H3K27me3 peptide and once bound cannot be out-competed by it. This observation is suggestive of a regulative function JARID2 in *de novo* trimethylation activity of PRC2. *De novo* methylation is required in gene regions which yet lack preexisting H3K27me3 but need to be fully repressed, thus requiring high levels of trimethylation (Sanulli et al., 2015). The higher binding affinity is likely due to an additional aromatic amino acid used in the recognition of JARID2/PALI1 in the aromatic cage of EED as compared to H3K27me3 (Sanulli et al., 2015; Margueron et al., 2009; Zhang et al., 2020). Finally, JARID2 harbors an ubiquitin-interaction motif (UIM), which as the name suggests is able to recognize H2AK119ub (Fig. 1.6 A). The role of this recognition will be further discussed in the following section 'PRC2 and chromatin' 1.5 of this thesis (Sanulli et al., 2015; Kalb et al., 2014; Kasinath et al., 2020).

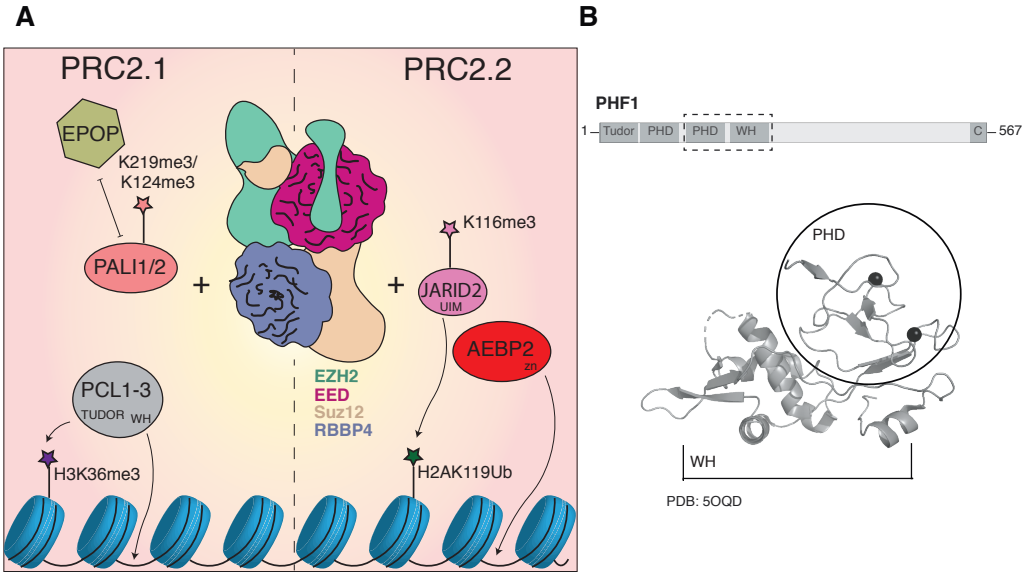

Figure 1.6: Role of PRC2 accessory subunits.

A. A simplified overview on the variety and functions of PRC2 accessory subunits and their organization into the subcomplexes PRC2.1 and PRC2.2 (adapted from Healy et al., 2019). B. Domain organization and crystal structure of the PRC2 cofactor PHF1 showing PHD and WH domains (Choi et al., 2017).

1.5 PRC2 and chromatin

1.5.1 PRC2 regulation by surrounding chromatin

In addition to the association with accessory subunits, activity of PRC2 can be stimulated by the surrounding chromatin. The own catalytic product of PRC2, H3K27me3, can enhance the activity of PRC2 through recognition and allosteric activation by EED, as was already described in previous sections of this thesis (Margueron et al., 2009; Xu et al., 2010). This positive feedback loop mechanism is thought to be responsible for the reestablishment of high levels of H3K27me3 regions after 2-fold dilution of the methyl marks due to chromosome duplications in the S-phase of a cell cycle (reviewed in Holoch and Margueron, 2017). It is not surprising therefore, that expression of EZH2 correlates with the rate of cellular proliferation (Bracken et al., 2003; Margueron et al., 2008).

Furthermore, regions of dense chromatin were reported to have a PRC2 activity enhancing function and this is dependent on the presence of SUZ12(VEFS) domain (Yuan et al., 2012). *In vitro*, this effect can be simulated by presence of high amounts of histone tails. In general, PRC2 shows higher activity on oligonucleosomes and the activity decreases with the amounts of nucleosomes linked by the DNA, with mononucleosomes being the least favorable nucleosomal substrate (Cao et al., 2002; Müller et al., 2002). Intriguingly, while

the H3 tail is the primary catalytic substrate, the binding and activity of PRC2 to and on H3 peptides are significantly reduced as compared to nucleosomes (Cao et al., 2002; Czermin et al., 2002; Kuzmichev et al., 2002; Müller et al., 2002). In agreement with this observation, binding affinity of PRC2 to DNA was observed to be in low nanomolar range and presense of protein cofactors such as PHF1 enhance this binding further (Choi et al., 2017; Wang et al., 2017). More insights into dependence of PRC2 on availability of DNA was shown in a study where different linker length on dinucleosomes substrates lead to differential activity (Wang et al., 2017). Activity of PRC2 enhances with shorter linker lengths and this effect can also be provoked by addition of Mg^{2+} ions and linker histone H1, which help condensing chromatin (Wang et al., 2017; Martin et al., 2006; Yuan et al., 2012). In presence of high amounts of H3 peptides encompassing residues 31-42 in *trans*, which simulates a highly condensed chromatin environment, differences in linker lengths had no effect on PRC2 activity. These observations clearly show, that mediation of stimulative effect is transferred not by the linker lengths directly but by bringing H3 tails in closer proximity to PRC2 (Yuan et al., 2012) (reviewed in Holoch and Margueron, 2017 and Liu and Zhu, 2017). In summary, while actions of PRC2 and PRC1 are required for chromatin compaction, dense chromatin with an increased number of H3 tails in close proximity to PRC2 might constitute another mechanism of positive feedback regulation and can also act upstream of PRC2. It should be however noted that these *in vitro* observations remain to be confirmed *in vivo* (reviewed in Holoch and Margueron, 2017).

PRC2 is also subject to negative regulation e.g. by several active histone marks. In presence of H3K4me3 or H3K36me2/3 on the H3 tail, PRC2 is inhibited in *cis* but not in *trans*. While evidence for this inhibition was provided by numerous studies both *in vitro* and *in vivo*, the exact molecular mechanism of how these marks exert their influence on PRC2 activity has remained enigmatic (Schmitges et al., 2011; Voigt et al., 2012; Bernstein et al., 2006) (reviewed in Holoch and Margueron, 2017). In the case of H3K4me3, a methyl mark linked to transcriptional activation and especially found at active gene promoters (reviewed in Shilatifard, 2012; Morgan and Shilatifard, 2020), a mechanism involving RBBP4 was proposed based on a crystal structure detailing the interaction of the *Drosophila* ortholog Nurf55 with the H3 tail (Schmitges et al., 2011) (Fig. 1.4 D). Consistent with this observation, the binding affinity of a PRC2 construct containing Nurf55 and the VEFS domain of SUZ12 to H3 was reduced when the H3 peptide contained trimethylated H3K4 (Schmitges et al., 2011) (see also section 1.3.1). Controversially, PRC2 without Nurf55 was yet inhibited and moreover, binding to H3K4me3 containing nucleosomes was not affected (Schmitges et al., 2011). The VEFS domain of SUZ12 is therefore a pontential candidate for sensing H3K4 methylation (Schmitges et al., 2011).
H3K36me2/3 mark is found associated with bodies of transcribed genes (reviewed in Shilatifard, 2012; Morgan and Shilatifard, 2020; Holoch and Margueron, 2017) and as H3K4me3 is a *cis* inhibitor of PRC2 activity (Schmitges et al., 2011; Yuan et al., 2011; Voigt et al., 2012). Intriguingly, unlike H3K4me3, where the sum of K27 methylation processes on a H3K4me3 peptide was reduced in comparison to an unmodified peptide, they were equal when comparing H3K36me3 and unmodified peptides. This is suggestive of a mechanism,

where H3K36me3, unlike H3K4me3, is only inhibiting in the context of a nucleosome (Schmitges et al., 2011).
Consistent with the inhibition function exerted by the active marks, mass spectrometry (MS) analysis on natively isolated nucleosomes has not found virtually any H3 tail decorated with both H3K27me3 and H3K4me3 or H3K36me2/3 in embryonic stem (ES cells) (Voigt et al., 2012). However, the combination of these marks was found to be existing on the same nucleosome but separated on the two H3 tail copies and only in ES cells and not in the more differentiated mouse embryonic fibroblast (MEF) cells. It was hence proposed that the existence of both active and repressive marks on the same nucleosome but different H3 tails is an attribute of bivalent promoters reflecting and ensuring the plasticity of gene expression states characteristic to pluripotent and multipotent stem cells (Bernstein et al., 2006; Voigt et al., 2013). Finally, on active genes bodies, H3K36me3 can be found on the same H3 tail as H3K27me1, the inhibition is therefore limited to di- and trimethylation activity of PRC2 (Voigt et al., 2012) and this observation was confirmed *in vitro* (Schmitges et al., 2011; Yuan et al., 2011). In summary, genes neighboring to Polycomb repression sites, which have to remain transcriptionally active, can be protected from being 'H3K27me3-overwritten' by active histone marks such as H3K4me3 and H3K36me3, but the mechanism of how PRC2 senses the presence of these marks remained unclear (Schmitges et al., 2011).
Other mutually exclusive marks to H3K27 methylation include H3K27ac and H3S28p, likely by chemical incompatibility (Ferrari et al., 2014; Gehani et al., 2010; Lau and Cheung, 2011) (reviewed e.g. in Holoch and Margueron, 2017; Liu and Zhu, 2017).

1.5.2 PRC2 binding and recruitment to chromatin

Cryo-EM studies of PRC2.2 (PRC2-AEBP2) on heterodimeric dinucleosomes provided first structural details of PRC2 binding to chromatin (Poepsel et al., 2018) (Fig. 1.7 A). While the complete five subunit PRC2.2 complex was used to reconstitute the PRC2-dinucleosome complex, only the ternary complex EZH2-EED-SUZ12^{VEFS} as was crystallized before (Justin et al., 2016) (see also section 1.3.2) could be visualized in the cryo-EM reconstruction. Densities for the 'bottom lobe' of PRC2 are missing likely due to protein denaturation at the water-air interface during plunging processes, flexibility hindering processing or a combination of both. In the EM map with an overall resolution of 6.2 Å, the ternary complex sits between the two nucleosomes, which are connected by a 35 bp linker (Poepsel et al., 2018). Interaction with the nucleosomes is established by binding to the nucleosomal DNA: on the substrate (unmodified) nucleosome, EZH2 recognizes the DNA via four CXC residues (K563, Q565, K569, Q570) (Poepsel et al., 2018). The allosteric nucleosome (carrying an H3Kc27me3) is bound by the combined interfaces of EED, EZH2^{SBD} and possibly SANT1 domains (Fig. 1.7 A) (Poepsel et al., 2018). The flexibility of the EED-EZH2^{SBD} interaction with the allosteric nucleosome (as can be seen by the higher resolution in the local resolution map) hints towards several modes of possible interactions on this nucleosome (Poepsel et al., 2018). Intriguingly, the two interface binding mode of the ternary complex remains the same when using different linker lengths between the two

nucleosomes. Instead, the nucleosomes are found flipped to each other on dinucleosomes with 35 bp linker with the linker DNA traversing between the two nucleosomes (Poepsel et al., 2018). When using 30/40 bp linker dinucleosomes, the two nucleosomes are connected by a straight non-crossed DNA (Fig. 1.7 B) (Poepsel et al., 2018). The ternary PRC2 complex does not bind the acidic patch of the nucleosome in the cryo-EM reconstruction, however, negative stain 2D classification clearly showed that the 'bottom lobe' of PRC2, missing in the cryo-EM map, makes contact to one or two of the dinucleosomes (Fig. 1.7 C) (Poepsel et al., 2018). Hence it cannot be ruled out based on this study that the acidic patch of the nucleosome is indeed not required for PRC2-nucleosome binding.

In *D. melanogaster* recruitment of PRC2 to target genes requires the presence of DNA binding sites of approximately few hundred bp length called Polycomb response elements (PREs) (Schwartz et al., 2010). PREs were shown to recruit both PRC2 and PRC1 and to be sufficient for silencing of reporter genes (Wang et al., 2004b) (as reviewed in Kassis and Brown, 2013; Schwartz, 2017). The first PRE to be discovered was bxd-PRE, a *Hox* gene part of the bithorax complex (see Fig. 1.3 and section 1.2), which controls the repression of Ubx. (Chan et al., 1994; Simon et al., 1993). Studies have shown that insertion of PREs in the *Drosophila* genome causes spreading of H3K27me3 in proximity to the insertion sites as well as repression of a reporter gene. Excision of PREs resulted in dilution of H3K27me3 with each round of S-Phase and derepression of the reporter, but only when the cells were not replication-stalled (Laprell et al., 2017).
While the exact molecular mechanism of PRC2 recruitment to PREs remains elusive, studies have shown that it requires the protein Pho, the only PcG protein with DNA sequence binding specificity (Brown et al., 1998; Schuettengruber et al., 2007; Oktaba et al., 2008; Kwong et al., 2008; Frey et al., 2016). Previous crystallographic studies revealed details of the binding of the highly conserved Pho human homolog YY1 to its cognate DNA within the PREs (Houbaviy et al., 1996; Brown et al., 1998). Despite a number of studies reporting a direct interaction of Pho/YY1 with PRC2 or PRC1 (Mohd-Sarip et al., 2005, 2002; Wang et al., 2004b), the respective complexes couldn't be reconstituted with purified proteins *in vitro* (Klymenko et al., 2006; Mohd-Sarip et al., 2005). Instead, Pho was shown to interact with Sfmbt (Alfieri et al., 2013) and this interaction recruits PRC1 via direct binding of Sfmbt SAM domain to Scm-SAM of PRC1, as shown by crystallographic studies (Frey et al., 2016). Confirming this observation, a PRE transgene with abrogated PhoRC binding sites showed a reduction of Pho, Sfmbt, Ph and also consistent with previous studies, E(z) and H3K27me3 (Wang et al., 2004b). Hence while it remains to be shown whether Sfmbt can also directly interact with PRC2, recruitment of PRC2 to PREs is dependent on DNA binding by PhoRC (Wang et al., 2004b; Frey et al., 2016).
In mammals, the mechanism of PRC2 recruitment is even more enigmatic as they lack an exact analogue of PREs in their genome. A variety of mechanisms were proposed involving, among others, unmethylated CpG dinucleotides in CpG islands and long non-coding RNAs (reviewed in Bantignies and Cavalli, 2017; Schwartz, 2017). An interesting model was suggested where PRC2.2 is recruited via its cofactors JARID2 and AEBP2 to nucleosomes carrying the H2AK119ub mark catalyzed by PRC1. Classically, PRC2 mediated H3K27

trimethylation targets canonical PRC1, the subunit CBX (Pc in *Drosophila*) of which binds to H3K27me3 via its chromodomain (Gao et al., 2012) (reviewed in Piunti and Shilatifard, 2016; Margueron and Reinberg, 2011; Morgan and Shilatifard, 2020). This process together with PRC1 mediated chromatin compaction is an essential part of transcriptional repression of PcG target genes *in vivo* and depletion of PRC1 subunits leads to de-repression of these genes. The non-canonical PRC1 complex does not possess the CBX/Pc subunit, but yet retains the ability to bind to chromatin at unmethylated CpG islands through the subunits RYBP1 and KDM2B (Tavares et al., 2012; Morey et al., 2013; Farcas et al., 2012). The catalyzed ubiquitination of H2AK119 by RING1 (Really Interesting New Gene) heterodimer E3 ligase activity of the non-canonical PRC1 can then serve as an anchor for PRC2-AEBP2-JARID2 (PRC2.2) (Blackledge et al., 2014). The activity of PRC2.2 on nucleosomes was shown to be greatly enhanced in the presence of the H2AK119ub mark as compared to unmodified nucleosomes and after the initial trimethylation processes, the catalyzed H3K27me3 can further activate PRC2 via the allosteric regulation by the EED subunit (Kalb et al., 2014). Once spread, H3K27me3 can in turn be bound by canonical PRC1, making this positive feedback loop and the concerted actions of the two complexes an attractive model for efficient propagation of H3K27me3 and transcriptional repression. Recently, a pre-print (not peer-reviewed) was published on bioRxiv showing a cryo-EM structure of PRC2.2 bound to a H2AK119ub mononucleosome detailing the JARID2 and AEBP2 interactions with the H2A ubiquitinated nucleosome: in the structure, the ubiquitin interaction motif (UIM) of JARID2 is bound between the ubiquitin moiety and the acidic path, showing interactions to both (Kasinath et al., 2020) (Fig. 1.7 D). Intriguingly, in presence of JARID2-ubiquitin interaction, AEBP2 is also found to engage with the histone core of the nucleosome via its two C2H2 zinc-fingers. In addition to EZH2, the previously identified lysine-arginine (KR-motif) of AEBP2 (Lee et al., 2018) is also found to be interacting with the nucleosomal DNA. The observed additional interactions contributed by JARID2 and AEBP2 and the thereby prolonged and stabilized binding of PRC2 to nucleosomes provides a structure-based mechanism of how activity of PRC2.2 is enhanced by the presence of the cofactors and ubiquitin (Kasinath et al., 2020).

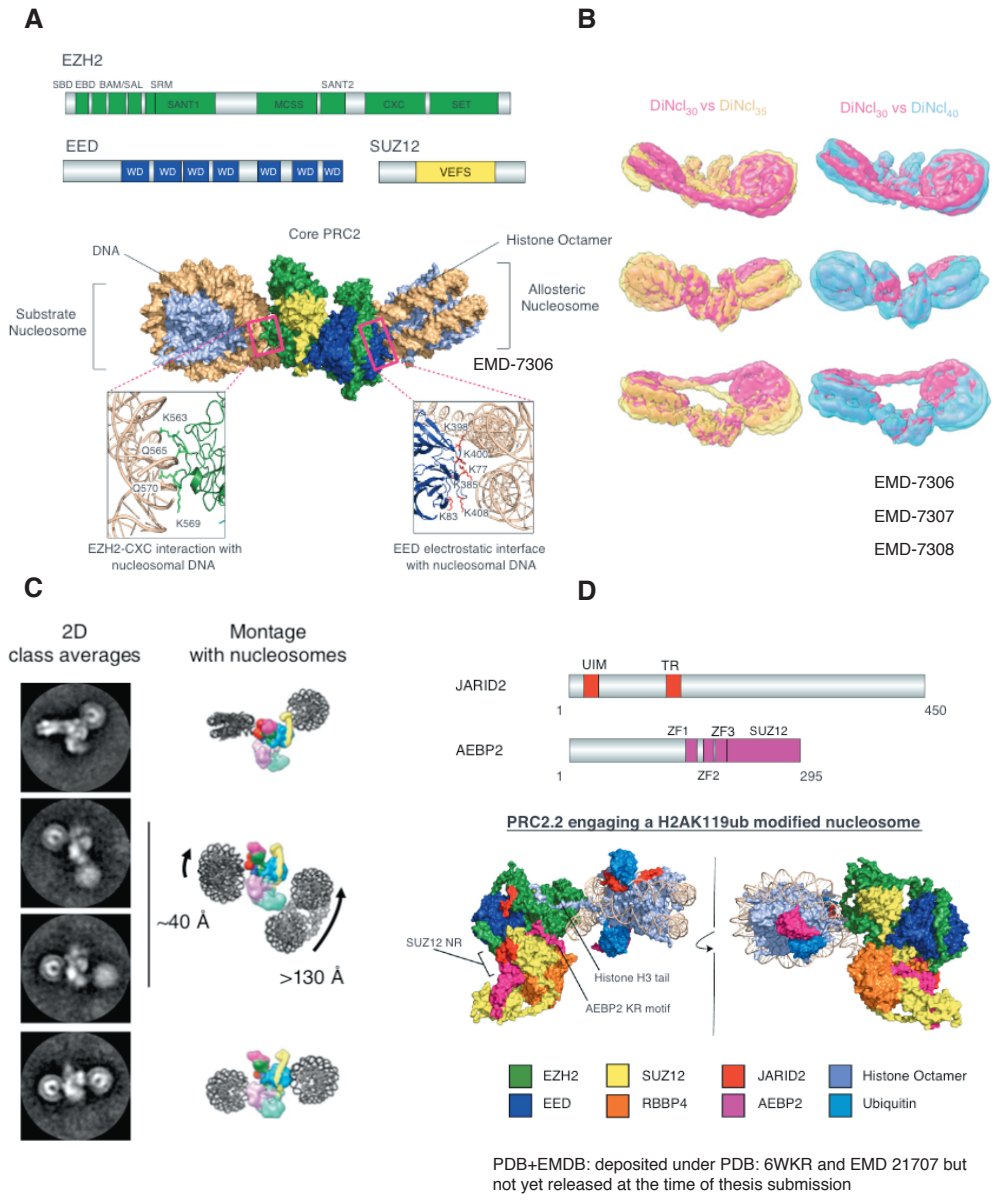

Figure 1.7: Structural studies of PRC2 on chromatin.

A. A 6.2 Å cryo-EM structure of PRC2.2 on a heterodimeric dinucleosome shows the ternary PRC2 complex (EZH2, EED, SUZ12^{VEFS}) in between the two nucleosomes connected by 35 bp linker (Poepsel et al., 2018). PRC2 contacts the substrate nucleosome via EZH2^{CXC} residues and the allosteric nucleosome is bound by a combined interface of EED, EZH2^{SBD}, SANT1 domains (Poepsel et al., 2018). Figure taken and modified from Glancy et al., 2020. EM-map deposited under EMD-7306 (Poepsel et al., 2018). B. Except minor differences, the binding mode on dinucleosomes with different linker lengths of ternary PRC2 complex remains the same and is established via the described interfaces in A. Instead, the second nucleosome is flipped on linker DNA of 35 bp as compared to 30 and 40 bp, where the DNA is straight between the two nucleosomes. Figure taken and modified from Poepsel et al., 2018. EMD-7306, -7307, -7308. C. Negative stain 2D class averages show nucleosome contacts of the 'bottom lobe' of PRC2, which is missing in the cryo-EM reconstruction. Figure taken and modified from Poepsel et al., 2018. D. PRC2.2 bound to an ubiquitinated mononucleosome reveals how JARID2 binds to the ubiquitin moiety on the histone core (Kasinath et al., 2020). Cited publication is not yet published at the time of thesis writing, therefore the bioRxiv publication is cited and figure was taken and modified from the recent review (Glancy et al., 2020). PDB and EMDB: 6WKR and 21707, deposited but not yet released.

1.6 PRC2 in disease

Mutations affecting PRC2 are implicated in cancer and neurodevelopmental diseases (reviewed in Kruger et al., 2017). In cancer, both loss of function and gain of function mutations of PRC2 are linked with tumorigenesis. Y641F/N/S/H/C are examples for activating mutations and are common for B cell lymphoma (Morin et al., 2010). Studies on the human ternary crystal structures showed that Y641 contributes to the formation of the aromatic cage in the catalytic center of EZH2 (Justin et al., 2016). Most of the cancer mutations are replacements of Y641 to a smaller side chain residue such as serine or cysteine resulting in a loss of the tyrosine hydroxyl group important for stabilization of unmethylated lysine 27 in the aromatic cage. The size expansion of the pocket presumably produces a stereochemically optimized environment for positioning of H3K27me2 substrate for trimethylation, while unmethylated/monomethylated K27 becomes a less favorable substrate, agreeing with previous studies (Sneeringer et al., 2010; Yap et al., 2011) (reviewed in Kruger et al., 2017). Another frequent mutation is the A677G substitution, where a similar mechanism was proposed (McCabe et al., 2012). In *Drosophila*, a different gain-of-function mutation in *E(z)*, R741K, also named *Trithorax mimick* or $E(z)^{Trm}$, leads to an increase of global levels of H3K27me3 and a Trithorax loss of function phenotype (Bajusz et al., 2001; Stepanik and Harte, 2012). Yet it remains unclear, whether the underlying mechanism of this mutation is reminiscent to the Y641 and A677 mutations (McCabe et al., 2012; Sneeringer et al., 2010).

H3.1 and H3.3 histone point mutations are exemplary for PRC2 deactivating alterations. Such point mutations for example are H3.1 and H3.3K27M frequently found in pediatric brain tumors such as diffuse intrinsic pontine gliomas (78%) and nonbrain stem gliomas (22%) (Schwartzentruber et al., 2012; Wu et al., 2012; Mackay et al., 2017). Tumors carrying this mutations are highly aggressive and linked to poor survival in patients and only mutation in one of the 32 H3 gene alleles is sufficient (characteristic of a dominant negative mutation). The lysine to methionine substitution was found to inactivate PRC2 with loss of H3K27me2/3 and a concomitant increase of H3K27ac in cells (Chan et al., 2013; Lewis et al., 2013; Bender et al., 2013; Venneti et al., 2013). H3K27M mutation also inhibits PRC2 *in vitro* (competitive inhibition) (Lewis et al., 2013; Bender et al., 2013) and a H3K27M peptide is sufficient for the inhibition (Bender et al., 2013). Furthermore, PRC2 was found to bind with a higher affinity to H3K27M nucleosome as compared to unmodified. Taken together, the data suggest a mechanism of inhibition in which PRC2 is sequestered to H3K27M containing nucleosomes leading to a decreased concentration and availability of PRC2 in the cells ultimately reducing H3K27me3 (Chan et al., 2013; Lewis et al., 2013; Bender et al., 2013). The X-ray structure of the human ternary EZH2-EED-SUZ12 complex with H3K27M peptide agrees with the biochemical studies (Justin et al., 2016). A recently identified non-histone protein inhibitor of PRC2, EZH2IP (CATACOMB), inhibits PRC2 by a similar mechanism in Posterior Fossa A ependymomas (Hübner et al., 2019; Jain et al., 2020; Pajtler et al., 2018; Piunti et al., 2019). Other frequent histone point mutations are G34R or G34V on H3.3 (Schwartzentruber et al., 2012; Wu et al., 2012).

Histone H3 point mutation are also linked to other diseases such as Weaver and Cohen-Gibson syndrome characterized by overgrowth and variable intellectual disabilities (as reviewed in Deevy and Bracken, 2019; Tatton-Brown et al., 2017; Cyrus et al., 2019), underlining the importance of *in vivo* studies such as in *Drosophila* where H3K27R point mutation could recapitulate loss of Polycomb phenotype (Pengelly et al., 2013).
In summary, many PRC2 mutations are associated with different diseases, of which only a few were presented exemplary in this section. The correct balance of PRC2 activity in the cell is therefore crucial while both erroneous up- and downregulation, tipping the balance towards either site, are detrimental for cells.

1.7 Aims of the thesis

The aim of the thesis is to analyze in molecular detail how PRC2 interacts with chromatin and interprets preexisting histone modification marks such as e.g. the inhibiting mark H3K36me3. For this, structural studies using cryo-EM as well as structure-based biochemical analyses such as EMSAs (electrophoretic mobility shift assays) and HMTase assays (histone methyltransferase assays) are employed. Within PRC2-dinucleosome interactions, the thesis aims to analyze the distinct roles of the different binding sites in PRC2 function. Another objective of the thesis is to provide a comprehensive model of PRC2 inhibition by H3K36me3, fitting with the observed regulative functions of PRC2 activity. Biochemical assays (EMSAs, HMTase assays and mass spectrometry based peptide HMTase assays) are used to characterize the biophysical and chemical properties of the H3K36 side chain and its methylation states. In the context of H3K36me3 inhibition, this work furthermore aims to describe the potential role of PHF1 Tudor recognition of H3K36me3. In summary, the thesis aims to dissect chromatin binding contributions made by the PRC2 subunits and to analyze their function in the context of PRC2 activity and finally, to provide a model explaining the inhibition mechanism exerted by the H3K36me3 mark on PRC2 in *cis*.

Chapter 2

Materials

All materials used in this study were obtained from Sigma Aldrich or Roth unless specified otherwise.

2.1 Bacterial and insect cells strains

Table 2.1: List of bacterial and insect cells strains used in this study.

Strain	Organism	Use	Source
DH10EMBacY	*E. coli*	Bacmid preparation	Geneva Biotech
Top 10 F'	*E. coli*	Cloning	Invitrogen
BL21DE3 plysS	*E. coli*	Protein expression	Sigma Aldrich
IPLB-Sf21	*Spodoptera frugiperda*	Viral expression	Invitrogen
BTI-TN-5B1-4 High Five	*Trichoplusia ni*	Protein expression	Invitrogen

2.2 Primers used for reconstitutions of nucleosomal DNA

Nucleosomal DNA for heterodimeric dinucleosomes was generated by S. Poepsel at UC Berkeley (E. Nogales Lab) essentially as described in Poepsel et al. (2018).
Primer pairs KF07/08 generate a 215 bp DNA for mononucleosomes with overhang DNA when used on p601 plasmid. When used on 2p601l35 plasmid the same primer pair generate 377 bp nucleosomal DNA with two 601 octamer binding sequences (Lowary and Widom, 1998) with a 35 bp linker used for reconstitution of unmodified dinucleosomes used for biochemical assays.

Table 2.2: List of primers used for large-scale PCR amplifications of nucleosomal DNA in this study.

Name	Length and purpose of nucleosomal DNA generated	Sequence (5'-3')	Source
KF01	147 bp DNA FW	CTgGAGaATcCCGGTgcC	this study
KF02	147 bp DNA Rev	ACAGgAtGTATATATCTGACACGTGCc	this study
KF07	215/377 bp DNA FW	ATATCTCGGGCTTATGTGATGGAC	this study
KF08	215/377 bp DNA Rev	ATATCCCGAGTCGCTGTTCAATAC	this study
KF07-F	5' Fluorescein-labeled 215 bp DNA FW	[6FAM]ATATCTCGGGCTTATGTGATGGAC	this study
5'25 FWD	K27me3 Ncl FWD (dinucleosomes 35 bp linker EM)	CTGGAGAATCCCGGTGCcgag	E. Nogales lab
5'25 Rev	K27me3 Ncl Rev (dinucleosomes 35 bp linker EM)	TAGGTATCGTATCACGGGGTGAGatc	E. Nogales lab
3'35 30 FWD	unmod. Ncl (dinucleosomes EM) FWD plus 30 bp overhang DNA	ctgacttattgaCACCCCGTGatgc	E. Nogales lab
3'35 30 Rev	unmod. Ncl (dinucleosomes EM) Rev plus 30 bp overhang DNA	TGGAGCTCAGATCGAATGATATGC	E. Nogales lab

2.3 Plasmids

Table 2.3: List of plasmids used in this thesis.

Name	Construct	Vector	Affinity tag	Protease	Resistance	Species	Source
EZH2	EZH2	pFB	6xHis	TEV	Amp.	*H.s.*	C. Müller lab
EED	EED	pFB	6xHis	TEV	Amp.	*H.s.*	C. Müller lab
SUZ12	SUZ12	pFB	6xHis	TEV	Amp.	*H.s.*	C. Müller lab
RBBP4	RBBP4	pFB	6xHis	TEV	Amp.	*H.s.*	C. Müller lab
JC212	PHF1	pFB	2xStrepII-6xHis	TEV	Amp.	*H.s.*	(Choi et al., 2017)
JC257	PHF1 515-567 (PHF1C)	pFB	2xStrepII-6xHis	TEV	Amp.	*H.s.*	(Choi et al., 2017)
H29	EZH2$^{CXC>A}$ K568A Q570A K574A Q575A	pFB	6xHis	TEV	Amp.	*H.s.*	this study
H30	EZH2$^{CXC>E}$: K568E Q570E K574E Q575E	pFB	6xHis	TEV	Amp.	*H.s.*	this study
H37	EED>A: K77A K83A K410A K423A K425A K433A	pFB	6xHis	TEV	Amp.	*H.s.*	this study
X.l. H3	Histone H3	pET	-	-	Amp.	*X.l.*	(Luger et al., 1999)
X.l. H4	Histone H4	pET	-	-	Amp.	*X.l.*	(Luger et al., 1999)
X.l. H2A	Histone H2A	pET	-	-	Amp.	*X.l.*	(Luger et al., 1999)
X.l. H2B	Histone H2B	pET	-	-	Amp.	*X.l.*	(Luger et al., 1999)
D.m. H3	Histone H3	pET	-	-	Amp.	*D.m.*	(Luger et al., 1999)
D.m. H4	Histone H4	pET	-	-	Amp.	*D.m.*	(Luger et al., 1999)
D.m. H2A	Histone H2A	pET	-	-	Amp.	*D.m.*	(Luger et al., 1999)
D.m. H2B	Histone H2B	pET	-	-	Amp.	*D.m.*	(Luger et al., 1999)
D.m. H2B	Histone H2B	pET	-	-	Amp.	*D.m.*	(Luger et al., 1999)
X2	H2A D90A, E92A	pET	-	-	Amp.	*X.l.*	this study
X6	H2A E64A, E61A D90A E92A	pET	-	-	Amp.	*X.l.*	this study
D12	H3 K36R	pET	-	-	Amp.	*D.m.*	this study
D13	H3 K36A	pET	-	-	Amp.	*D.m.*	this study
D9	H3K36C C110A	pET	-	-	Amp.	*D.m.*	J. Müller lab
H3K$_C$27	H3K27C C110A	pET	-	-	Amp.	*D.m.*	E. Nogales lab
p601	601 Widom sequence containing plasmid	pUC19	-	-	Amp.	-	(Lowary and Widom, 1998)
twop601135/2p601135	plasmid containing 2 601 Widom sequences and 35 bp linker	pM	-	-	Kana.	-	Invitrogen GeneArt

2.4 Antibodies

Table 2.4: List of antibodies used for WB analyses in this study.

Antibody	Dilution	Host	Source
H3K27me1	1:6000	rabbit	Milipore (07-448)
H3K27me3	1:1000	rabbit	Milipore (07-449)
H4	1:200000	rabbit	Abcam (ab 10156)
HRP	1:5000	donkey	GE Healthcare (NA934)

2.5 Media

Table 2.5: List of media solutions used in this study.

Media	Use	Supplemented with	Source
L-Broth (LB)	cloning (*E. coli*)	-	in-house
Terrific broth (TB)	expression (*E. coli*)	10% Phosphate buffer	in-house
SOC	cloning (*E. coli*)	Protein expression	in-house
Express Five	protein expression (High Five cells)	L-Glutamine	Gibco
EX-CELL TiterHigh	virus generation (Sf21 cells)	-	Sigma Aldrich

2.6 Buffers

Table 2.6: List of general buffers used in this study.

Name	Composition	Use
TBS	20 mM Tris-HCl pH 7.5, 150 mM NaCl	WB
TBS-T	TBS + 0.2 % Tween 20	WB
1 x transfer buffer	25 mM Tris , 192 mM Glycine, 0.05 % SDS, 10 % Methanol	WB
1 x TBE	89 mM Tris , 89 mM boric acid, 2 mM EDTA	agarose gel electrophoresis
Phosphate-buffered saline (PBS)	10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 137 mM NaCl, 2.7 mM KCl	Insect Cells
Phosphate buffer	0.17 M KH_2PO_4, 0.72 M K_2HPO_4	supplement for TB media
10 x SDS PAGE Running Buffer	0.25 M Trizma Base, 1.92 M Glycine, 1 % SDS	SDS PAGE
4 x SDS sample buffer	200 mM Tris pH 6.8, 20 % 2-Mercaptoethanol 8 % SDS, 0.4 % Bromophenol Blue, 40 % Glycerol	SDS PAGE

2.7 Peptides

Table 2.7: List of peptides used in this study.

Name	Sequence	Source
unmodified H3 peptide	histone H3 aa 18-42	in-house
H3K36me3 peptide	histone H3 aa 18-42 tri-methyl group on K36	in-house

Chapter 3

Methods

3.1 Cloning and bacterial cell culture

3.1.1 Site-specific mutagenesis

PCR

EZH2, EED and histone H3 and H2A mutants were generated by PCR with phosphorylated-primers (Sigma-Aldrich) containing the desired mutations, and the general protocol for site-specific mutagenesis was similar as to described in `https://openwetware.org/wiki/%27Round-the-horn_site-directed_mutagenesis` (last query of the website was successfully made on 14.01.2021). For PCR, the Q5 High-Fidelity 2X Master Mix (MM)(NEB) was used. For a 50 µl reaction, 25 µl of the Q5 MM, 2.5 µl of each phosphorylated primer containing the desired mutation, 50 ng of template plasmid and x µl ddH_2O to a final of 50 µl were pipetted.

(1) 98 °C	1 min	
(2) 96 °C	30 s	
(3) 55-58 °C	30 s	Go to (2) 25 x
(4) 72 °C	2 min	
(5) 4 °C	∞	

DNA product preparation via preparative agarose gel electrophoresis

Prior to preparative gel electrophoresis, 1 µl of DpnI (NEB) was added to the PCR mixture for template digestion and the mixture was incubated for 1 h at 37 °C. Subsequently, the PCR samples were diluted in 6 x Orange Dye (Fermentas) to a final concentration of 1 x and separated by agarose gel electrophoresis. Casting of gels was performed using 0.8-1.2 % Agarose, dissolved in 1x TBE. SYBR Safe stock (Invitrogen) was added as a dye for visualisation of nucleic acids in a 1:10000 dilution and 1 kb plus DNA ladder (Fermentas, 0.5 mg/lane) was used as marker. Gel electrophoresis was performed at 100-120 V for 20-25 min and DNA fragments were analyzed on an UV transilluminator. Finally,

correct bands containing the linearised plasmid with the desired mutation were excised using visualization of the bands on a blue light transilluminator emitting blue light at a wavelength of 470 nm. DNA preparation was performed using the QIAquick Gel Extraction Kit (Qiagen) according to manufacturer's instruction.

Product ligation, transformation and plasmid purification

Product ligation was performed using the Quick Ligation Kit (NEB) according to manufacturer's instruction and 2 µl of ligation mix were used for heatshock transformation of Top 10 F' cells (Table 2.1). Transformed cells were plated on agar media containing the corresponding antibiotics encoded by the plasmids used, in a 1:1000 dilution. Grown colonies were used to inoculate 5 ml of LB (containing the diluted antibiotics) and the culture grown ON at 37 °C and agitation. Plasmids were purified using the QIAprep Spin Miniprep Kit (Qiagen) according to manufacturer's instruction.

3.2 Insect cell culture

3.2.1 General insect cells culture

Sf21 cells (Table 2.1) used for transfection and virus amplification were cultivated at cell densities of approximately 0.7-1 $\times$ 10^6 cells/ml in EX-CELL Tither High media (Table 2.5). Proteins were expressed in High Five cells (Table 2.1), which were cultivated at 0.4-0.8 $\times$ 10^6 cells/ml in Express Five medium supplemented with 18 mM L-Glutamine (Table 2.5). Both cell lines were grown as 500 mL suspension culture using 3 l polycarbonate Erlenmeyer flasks (Corning) with vent caps at 27 °C and gentle agitation (90 rpm). Monitoring of cell density, viability and diameter was performed with a Vi-cell XR cell viability analyzer (Beckman coulter). The majority of insect cell culture maintenance, virus generation and protein expression was performed by K. Schmid (J. Müller lab, MPI of Biochemistry).

3.2.2 Baculovirus generation

The 'Bac-to-Bac Baculovirus Expression System' (Invitrogen; (Ciccarone et al., 1998)) was used to generate viruses for protein expression. The protocol includes site-specific transposition in *E. coli*, purification of recombinant bacmids followed by bacmid transfection into insect cells from which the first passage of recombinant baculoviruses can be generated. The individual steps are explained in more detail below.

3.2.3 Transposition

2 µl containing approximately 1 ng of purified (Plasmid Mini Kit, Qiagen; according to manufacturer's instructions) pFastBac (pFB) vector was added to thawn DH10 EMBACY *E. coli* cells (Table 2.1) and the mixture was kept on ice for 30 min. Following the incubation, cells were heat-shoked for 45 s at 42 °C using a thermoblock. Subsequently, cells

were allowed to recover for 2 min on ice. Upon addition of 900 µl SOC medium (2.5) tubes containing the transfected cells were incubated for 4 h at 37 °C with agitation, allowing for transposition. Finally, cells were spinned down for 4 min at 3000 rpm and the pellet resuspended in 200 µl LB. Transfected cells were then plated on LB-agar plates containing Kanamycin (final concentration 50 µg/ml), Tetracyclin (10 µg/ml), Gentamicin (7 µg/ml), Chloramphenicol (34 µg/ml), IPTG (100 µg/ml) and BluoGal (100 µg/ml). Plates were subsequently kept at 37 ° C for 48 h protected from light. White colonies were picked and further cultured for Bacmid DNA purification.

3.2.4 Bacmid DNA purification

Several confirmed white colonies were cultured in 2.5 ml LB containing the same set of antibiotics at the same concentration as for the transposition plates for approximately 20 h with agitation. Afterwards, the cellular suspension was spinned down and buffers from the QIAGEN Midiprep were used for the following steps: First 300 µl of P1 Buffer was added to the bacterial pellet. After thorough resuspension, 300 µl of P2 Buffer were added for lysis. After careful 6x inversion, the tubes containing the mixture were kept for 5 min at RT. Subsequently, 300 µl of pre-cooled P3 buffer were added and the tubes again inverted 6 x and left for 5 min incubation on ice. Afterwards, the mixture was spun down for 10 min at 4 ° C at 12000 rpm and the supernatant transferred to fresh 2 ml eppendorf tubes. The supernatant was then centrifuged again at 12000 rpm 5 min 4 °C and 800 µl isopropanol was added to the new supernatant in a fresh Eppendorf tube. The mixture was kept on 4 ° C for 10 min and again centrifuged for 20 min at 4° C. This time, the pellet was kept and carefully washed with 500 µl 70 % ethanol. After 10 min RT centrifugation the supernatant was discarded, the pellet air-dried and resuspended carefully in 100 µl elution buffer (EB).

3.2.5 Transfection of Sf21 cells

Prior to transfection, Sf21 cells were diluted to 0.8×10^6 cells/ml in EX-CELL TiterHigh medium (Table 2.5) and 2 ml of cell suspension per well were distributed in a 6 tissue culture plate (BD Falcon). In order for cells to attach to the wells, the plate was left at RT on the bench for 30-60 min. In the meantime, a transfection mix was prepared for each well containing 1 µg of bacmid DNA diluted in 100 µl EX-CELL TiterHigh medium without antibiotics. A second mixture is prepared for each well consisting of 8 µl Cellfectin II and 92 µl of EX-CELL TiterHigh medium. The two mixtures were then carefully mixed together equaling to 210 µl of transfection reaction mix needed for one well. One negative control was included, which contained cells with cellfectin but no DNA. The transfection mixes were subsequently added dropwise to the respective wells and the 6-wellplate left to incubate 3-5 h at 27 ° C without agitation. Finally, all liquid was carefully aspired and replaced with fresh 2 ml of EX-CELL TiterHigh medium. After a final incubation step of 96 hours at 27° C the cells showed signs of cytopathic effect indicative of successful generation of the first virus passage, P1. The supernatant with the detached cells containing the P1 virus was then harvested and stored at 4° C until further use.

3.2.6 Virus amplification

To obtain a high enough amount of viruses needed for protein expression, the P1 passage was further amplified: 50 ml of 0.4×10^6 Sf21 cells were infected with approximately 2 ml of P1 and the cell suspension was left to grow for 24-72 h (until cells stop doubling) in 125 ml plastic Erlenmeyer with gentle agitation at 90 rpm, 27 °C. Cell growth was monitored using the Vi-cell XR cell viability analyzer. The supernatant containing the P2 virus could be further used to infect (2.5 ml of P2) 500 ml of 0.4×10^6 Sf21 and P3 virus was generated similarly as P2. P1, P2 and P3 viruses were all stored at 4°C, protected from light. And P3 passage was used for protein expression.

3.2.7 Protein expression

P3 Virus for all five subunits of the PRC2-PHF1 complex and the corresponding shorter constructs were generated following the protocol described in the previous subsections and all five P3 viruses were used to co-infect High Five cells (Table 2.1) for protein expression. Prior to large scale expressions, the correct ratios of the individual viruses for stochiometric amounts of the respective protein expressed were tested by performing small-scale expressions and pulldown experiments using Streptavidin beads (Thermo Scientific Pierce Streptavidin Magnetic Beads and Thermo Fisher Kingfisher Duo Prime, according to the manufacturer's instructions). For large cell expression respective ratios of viruses were used to co-infect 500 ml of 0.8×10^6 High Five cells per flask. In total 8 l of High Five cells were used for large scale expression of PRC2 complexes. Infected cells were grown for 50-72 h with gentle agitation at 90 rpm and at 27 °C and subsequently harvested by centrifugation (15 min at 2000 rpm). The cell pellet was carefully washed with ice-cold phosphate buffered saline (PBS; Table 2.6) and the suspension centrifugated again at 15 min 2000 rpm. Finally, pellets were frozen in liquid nitrogen and stored at -80 °C until purification.

3.3 Protein purification

3.3.1 PRC2 purification

Purification of PRC2 followed generally the protocol published in (Choi et al., 2017). Frozen cell pellets were resuspended in PRC2 Lysis Buffer (25 mM Hepes-NaOH pH 7.8, 150 mM NaCl, 4% Glycerol, 0.05% NP40, 30 mM Imidazol, 1 mM DTT) and protease inhibitor cocktail (Complete Protease-Inhibitors, Sigma-Aldrich; 1 tablet per 50 ml cell lysate) and 1 mM AEBSF were added. Cells were lysed using a glass douncer, and protein containing supernatant was subsequently purified from cell debris and nucleus wall components by centrifugation (25000 rpm for 45 min at 4 °C). As all PRC2 subunits carried an N-terminal His tag and PHF1 constructs additionally a StrepII tag, the subsequent purification involved affinity chromatography with Ni-NTA and Streptactin 5ml FF columns (GE Healthcare). Proteins bound to Ni-NTA were washed to remove contaminations, first with lysis buffer and followed by a high salt buffer (25 mM Hepes-NaOH pH 7.8, 500 mM NaCl, 40 mM

Imidazol, 10 % glycerol, 1 mM DTT). Subsequently, bound PRC2 was directly eluted on the Streptactin Column using Elution Buffer (25 mM Hepes-NaOH pH 7.8, 150 mM NaCl, 250 mM Imidazole, 10 % glycerol, 1 mM DTT). High concentration of imidazole outcompetes binding of the His-tags to Ni^{2+} ions. PRC2 bound to Streptactin via the Strep tag of PHF1 constructs was subsequently rigorously washed with Ni-NTA Elution Buffer and eluted with 2.5 mM Desthiobiotin dissolved in Ni-NTA elution buffer. Eluted proteins were dialyzed ON in 25 mM Hepes-NaOH pH 7.8, 120 mM NaCl, 2 mM DTT with simultaneous tag-cleavage using TEV protease and sometimes Lambda Phosphatase treatment (both proteins obtained from the MPI of Biochemistry Protein Core facility). Depending on the yield of protein after affinity chromatography steps and the via A260/A280 absorption measured nucleic acid contamination, a further ion exchange purification step was included containing a MonoQ column (GE Healthcare). The final step of the purification involved a size exclusion chromatography (SEC) step using a pre-equilibrated Superose 6 10/300 GL column (GE Healthcare) in 25 mM Hepes-NaOH pH 7.8, 150 mM NaCl, 10% glycerol, 2 mM DTT. SDS PAGE analysis (see 3.3.3) was used to monitor all purification steps and to ensure a pure homogenous complex after SEC. Furthermore, concentrated purified proteins were tested for remaining nucleic acid concentration using A260/A280 absorption. Proteins were frozen and kept at -80°C until further use.

3.3.2 Nucleosome reconstitutions

Protocols for nucleosome reconstitutions were adapted from Luger et al. (1999).

Generation of nucleosomal DNA

Nucleosomal DNA of various lenghts were prepared for nucleosomes used in this studies. In the following, the different DNAs used shall be summarized:

1. 215 bp nuclesomal DNA for mononucleosomes:

5'- ATATCTCGGGCTTATGTGATGGACCCTATACGCGGCCGCC - 601 binding - GCATGTATTGAACAGCGACTCGGGATATCTCTAGAGTCGACCTGCAGGCATGCA AGCTTGG - 3'

2. 377 bp nucleosomal DNA for unmodified dinucleosomes used for biochemical assays:

5'–atatctcgggcttatgtgatggac - 601 binding substrate nucleosome 1 – agcgatctcaacgagt-gatgctcgatactgtcata - 601 binding substrate nucleosome 2 – gtattgaacagcgactcgggatat–3'.

3. 359 bp nucleosomal DNA for heterodimeric dinuclesomes connected by a 35 bp linker (one substrate with 30 bp overhang, one allosteric/H3Kc27me3 containing nucleosome):

5' - 601 binding (allosteric nucleosome) - agcgatctCACCCCGTGatgctcgatactgtcata - 601 binding (substrate nucleosome) - atgcatgcatatcattcgatctgagctcca - 3'.

The general protocol for generation of all nucleosomal DNA involved (1) large scale PCRs in 6-8 96-well plates (Corning) using p601 or 2p601l35 (Table 2.3) plasmids as templates directly (PCR protocol similar to as described in 3.1.1; adjusting times and temperatures according to the different templates and amplified lengths and using an in-house, MPI of Biochemistry Core Facility, purified Pfu polymerase instead of Q5). Primer pairs were chosen based on desired nucleosomal DNA version and are summarized in Table 2.2. (2) After PCR, nucleosomal DNA was precipitated using 100 % pre-chilled ethanol and NaCl in a final concentration of 300 mM. Precipitated DNA pellets were air-dried and dissolved in MonoQ Buffer A containing 5 mM Tris-HCl pH 7.5 and subsequently centrifuged 12000 rpm, 10 min to remove undissolved precipitates e.g. the Pfu polymerase. (3) 5 ml of DNA in MonoQ Buffer A were loaded on a MonoQ column (GE Healthcare) and an exchange chromatography was subsequently performed with MonoQ Buffer A and a slowly increasing gradient to 50 % and in one step to 100 % of Buffer B (5 mM Tris-HCl pH 7.5, 2 M NaCl). Fractions containing DNA were analyzed for purity and desired DNA product length on an agarose gel (protocol similar as described in 3.1.1). Subsequently, the purified nucleosomal DNA was precipitated again following the same protocol as described, but afterwards dissolved in Octamer High Salt Buffer (10 mM Tris-HCL pH 7.5, 2 M NaCL, 1 mM EDTA) in the last step.

For generation of nucleosomal DNA for heterodimeric dinucleosomes further steps were incorporated into the general protocol described above (for dinuclesomes used in cryo-EM in this study, nucleosomal DNA was generated by S. Poepsel, UC Berkeley, E. Nogales Lab, the author of this thesis adapted the protocol kindly provided by S. Poepsel later for in-house use):

First following templates for later large scale PCRs (instead of plasmids as described above) were generated by PCR:

For allosteric/H3Kc27me3 nucleosome:

5' - CTGGAGAATCCCGGTGCcgaggccgctcaattggtcgtagacagctctagcaccgcttaaacgcacgt acgcgctgtcccccgcgttttaaccgccaaggggattactccctagtctccaggcaCGTGTCAGATATATACATC CTGTagcgatctCACCCCGTGatacgatacctaA - 3'

For unmodified/substrate nucleosome plus 30 bp overhang:

5'- ctgacttattgaCACCCCGTGatgctcgatactgtcataCTGGAGAATCCCGgtgccgaggccg ctcaattggtcgtagacagctctagcaccgcttaaacgcacgtacgcgctgtcccccgcgttttaaccgccaaggggattactcc ctagtctccaggcacgtgtcagatatatacatcctgtatgcatgcatatcattcgatctg - 3'

The templates now contained DraIII sites and could be used for large-scale PCR using

primer pairs 5'25 FWD/REV for allosteric DNA and 3'35 30 FWD/REV (Table 2.2) for substrate DNA. Resulting PCR products were purified via ion exchange as described above and subsequently DraIII (NEB) digested according to the manufacturer's instruction, but for approximately 20 h.

Histone expression

Plasmids (Table 2.3) encoding the histones were transformed by heat-shock into BL21DE3 plysS cells (Table 2.1) and incubated on appropriate antibiotics containing agar plates at 37 °C, ON. On the next day, cells were subsequently harvested from the plate to grow a 50 ml pre-culture for 1-2 hours in LB containing antibiotics. 3-4 l of large-scale cultures using the pre-cultures were set up in LB+antibiotics and cultures were allowed to grow at 37 °C under agitation until reaching of log-phase (OD of 0.6). At OD of 0.6 histone expression was induced with 0.2 mM IPTG and expression was allowed to proceed at 37 °C for approximately 3 h (histones H2A/H2B) and 4 h (histones H3 and H4) before harvesting by centrifugation (10 min, 6000 rpm, 4 °C). Cell pellets containing histone inclusion bodies were frozen in liquid nitrogen and kept at -80 °C until further use.

Inclusion body purification

Bacterial cell pellets containing histone inclusion bodies were dissolved in approximately 40-50 ml Histone Wash Buffer (50 mM Tris-HCL pH 7.5, 100 mM NaCl, 1 mM EDTA, 5 mM β-mercaptoethanol, 1 Complete PI Tablets, 80 nM PMSF). Cell suspension was subsequently sonified for 6 min using VS70T tip and the following settings: 0.5 s /0.5 s, 40 % (Bandelin Electronics). Subsequently, the inclusion bodies were seperated from soluble cellular proteins by centrifugation (20 min, 13000 rpm, 4 °C). Pellet was resuspended in 1 % (v/v) Triton X100 containing Wash Buffer and subsequently centrifuged again. This step was repeated two times, each time keeping the inclusion bodies pellet, and another two times with Wash Buffer without Triton X100. Successful expression and purification yielded a white gum-like pellet, which was kept at -20 °C until futher purification.

Histone unfolding

In the next step, pellet was resuspended in 15 ml of unfolding buffer (20 mM Tris-HCL pH 7.5, 7 M Guanidine HCL, 10 mM DTT) and incubated at RT for 1 hour under rotation on a rocker. Cell suspension was centrifuged for 20 min, 4 °C, 13000 rpm and the supernatant subsequently dialyzed twice for 1 h and once ON in SAU-200 Buffer (20 mM Sodium Acetate pH 5.2, 7 M Urea, 200 mM NaCl, 1 mM EDTA, 5 mM β-mercaptoethanol). On the next day, ion exchange chromatography (5 ml Hi-Trap-SP-FF; GE Healthcare) was used to purify the unfolded histones. For this, histones were loaded on the pre-equilibrated column, washed with SAU-200 buffer and eluted by an increasing gradient of SAU-600 buffer (20 mM Sodium Acetate pH 5.2, 7 M Urea, 600 mM NaCl, 1 mM EDTA, 5 mM β-mercaptoethanol). Purity of histones after ion exchange was analyzed by SDS PAGE (described in 3.3.3 and fractions containing histones were pooled together for further dialysis in cold water

containing 5 mM β- mercaptoethanol (twice for 1 h, once ON at 4 °C). The purity and quality of the histones was further analyzed by Mass Spectrometry (MPI of Biochemistry, MS Facility). Finally, the histone containing solution was distributed in Eppendorf tubes with 2 mg of histones per tube and histones were speed-dried using a vacuum-concentrator centrifuge (Thermo Fisher Scientific) for several hours until complete removal of all liquid and the appearance of a faint, metalic, glossy pellet. Stocks of 2 mg of each histones were kept at -80 °C.

Histone alkylation

H3Kc36me3 histones were generated based on protocols described in (Simon et al., 2007). First, a construct containing mutations in H3 residues K36C and C110A was generated by former members of J. Müller lab (Schmitges et al., 2011). Expressed and purified H3K36C/C110A (same protocol as above) were dissolved (10 mg in 980 µl) in alkylation buffer (1 M Hepes-NaOH pH 7.8. 4 M Guanidine HCl, 10 mM D/L-methionine). After addition of 20 µl of 1 M DTT the histones were left at RT for 1 hr to allow for reduction process. Subsequently, 100 mg of 2-bromoethyl trimethylammonium bromide were added and the solutions were heated to 50 °C for 2.5 hrs under occasional careful mixing. Upon addition of 10 µl of 1 M DTT, the histones were further incubated for 2.5 h under the same conditions. The reaction was quenched with 50 µl of 14.2 M β-mercaptoethanol and histones containing the trimethyllysine analogue were subsequently dialyzed against water containing 3 mM β-mercaptoethanol (twice 1 h and once ON at 4 °C). All subsequent steps were performed as for purified histones described above.

Octamer Refolding

2 mg of each histone were each resolved in 1 ml of unfolding buffer (20 mM Tris-HCL pH 7.5, 7 M Guanidine HCL, 10 mM DTT) and the mixture was incubated for 1 h at RT to allow for complete dissolving. Equimolar amounts of each resolved histones were then added together and incubated at RT for 45 min with occasional mixing. The octamer containing solution was subsequently dialyzed twice for 1 h and once ON at 4 °C against Octamer High Salt Buffer 10 mM Tris-HCL pH 7.5, 2 M NaCL, 1 mM EDTA, 5 mM β-mercaptoethanol). After dialysis, refolded octamers were centrifuged for 10000 rpm, 10 min, 4 °C and purified by SEC (Superdex 200i, GE Healthcare). Fractions were analyzed by SDS PAGE (protocol 3.3.3) and the ones containing the assembled octamers collected and concentrated for further use.

Mono-nucleosome assembly

Before large-scale reconstitution, correct assembly of mono-nucleosomes was pre-tested by mixing different ratios of octamers to desired nucleosomal DNA (ratios around 0.7-1.2 octamer : 1 DNA) (see 3.3.2) in Octamer High Salt Buffer. The mixtures containing different ratios were dialyzed against the High Salt Buffer ON at 4 °C with simultaneous lowering of the NaCl concentration of the buffer to 500 mM. A (native) 0.8-1.2 % agarose (SeaKem

Agarose from Lonza) free from SYBR Safe. Gel electrophoresis was run protected from light in pre-chilled 0.4 x TBE Buffer for 45 min at gel (in 0.4 x TBE) without SYBR Safe was run on the next day in cold 0.4 x TBE for 45 min at 60 V to screen for optimal nucleosome assembly. After gel electrophoresis, the agarose gel was incubated with 0.4 x TBE solution containing SYBR Safe under gentle agitation and imaged with an UV transilluminator. Ratios containing no free nucleic acids (see exemplorary in Results Chapter, Fig. 4.3) were chosen for further large scale assembly following the same salt lowering protocol as above to 500 mM NaCl and subsequent further dialysis against (1) 10 mM Tris-HCL pH 7.5, 250 M NaCL, 1 mM EDTA, 5 mM β-mercaptoethanol for 1-3 hrs, (2) 10 mM Tris-HCL pH 7.5, 120 M NaCL, 1 mM EDTA, 5 mM β-mercaptoethanol for 1-3 hrs and (3) 25 mM Hepes-NaOH pH 7.8, 60 mM NaCl, 2 mM DTT ON (all steps at 4 °C).

Di-nucleosome assembly

Unmodified, symmetrical nucleosomes were assembled following the same protocol as above using nucleosomal DNA containing two 601 binding sites (see 3.3.2) and screening for ratios of around 2 octamer to 1 DNA, in which no free nucleic acids and no free mononucleosomes were observed.

Asymetrical, heterodimeric dinucleosomes were reconstituted following different protocols, described in Poepsel et al. (2018):
First, substrate and allosteric mono-nucleosomes were assembled at a 1.1:1 ratio of octamer to DraIII-digested DNA (see 3.3.2) and nucleosome-containing solutions were subsequently dialyzed ON with simultaneous lowering of NaCl similar as described before. On the next day, mononucleosomes were first briefly dialyzed against TCS Buffer (20 mM Tris-HCl pH 7.5, 1 mM EDTA, 1 mM DTT). After adding of 20 % sucrose to the sample, nucleosomes were subsequently subjected to preparative polyacrylamide gel (6-7 cm high, 5 %, 59:1 Acryl:Bisacrylamid gel) electrophoresis using the BioRad 491 PrepCell, assembled according to manufacturer's instructions. Gel electrophoresis was run in 0.2 x TBE Buffer at 10 W for approximately 2-4 hours and nucleosomes were eluted in TCS Buffer into a fraction collector. Analytical polyacrylamide gels (59:1) were run (10 W, 0.2 x TBE) and subsequently stained with SYBR Safe to analyze fractions containing mono-nucleosome free of nucleic acids.

Reconstituted substrate and allosteric mono-nucleosomes (500 nM each) were subsequently ligated at their DraIII-digested sites using 0.05 U/µl T4 ligase (Thermo Fisher Scientific) for 90 min at RT and ligated di-nucleosomes were purified using the same preparative gel electrophoresis system as described above.

3.3.3 SDS polyacrylamide gel electrophoresis

SDS polyacrylamide gel electrophoresis (SDS PAGE) was performed using either NuPAGE 4-12 % (w/v) Bis-Tris Precast gels (Thermofisher) with 1 x NuPAGE MOPS or MES buffers (Thermofischer) or self-cast 12 and 16 % (w/v) Tris-Glycine Polyacrylamide gels with 1x

SDS Buffer (Table 2.6) following established protocols (Laemmli, 1970). Percentage of SDS gels and buffers were chosen based on molecular weights of proteins to be resolved and the desired band separation on the gel. Prior to SDS PAGE, samples were dissolved in SDS sample buffer (Table 2.6) to a final concentration of 1x and incubated for 5 min at 95 °C for complete denaturation. Approximately 3-12 µl of samples were loaded on the gel. Gel electrophoresis was performed using the XCell SureLock Mini-Cell Electrophoresis System (Thermo Fisher) at 100-150 V. After run, protein bands were either visualized with Coommassie stain (InstantBlue Coomassie Stain, Abcam) or directly used for western blot analysis.

3.4 Cryo-electron microscopy

3.4.1 Cryo-EM sample and grid preparation

Cryo-EM sample and grid preparation for the dataset shown in the thesis were performed in collaboration with S. Poepsel in UC Berkeley (E. Nogales lab), while for other datasets both steps were performed by the author of the thesis.
Prior to plunging, purified PRC2-PHF1 (see 3.3.1) was buffer exchanged into the sample buffer (25 mM Hepes-NaOH pH 7.9, 10 mM KCL, 1 mM TCEP). 1.5 µM of PRC2-PHF1 was then incubated with 1.2 µM of heterodimeric dinucleosomes in the same sample buffer with addition of 100 µM SAH and 0.005 % NP-40. 1 % Trehalose was added shortly before grid plunging and Cu400 R1.2/1.3 holey carbon grids (Quantifoil Micro Tools) were plasma cleaned (Solarus, Gatan) for 25 s in preparation for the subsequent plunge freezing. Finally, 3 µl of the sample was applied on the grid positioned by tweezers in the humidity chamber of Mark IV vitrobot (FEI). The humidity chamber was kept at all times during plunging processes at 4 °C and 100 % humidity. Blotting time of the sample applied on the grid was set to 4 s before plunging into liquid ethane-propane mixture. Storage of grids and all subsequent steps were performed in liquid nitrogen. Clipping of grids was performed according to manufacturer's instructions (Thermofisher).

3.4.2 Cryo-EM data collection

In general, all screening datasets were collected by the author on either a Talos Arctica (Thermofisher) or a Titan Halo microscope, both equipped with Falcon 3 (Thermofisher) camera at the cryo-EM facility of MPI of Biochemistry, Martinsried. Several datasets were additionally screened by the author together with S. Poepsel in the cryo-EM facility of UC Berkeley. The final dataset of PRC2-PHF1 on heterodimeric dinucleosomes was collected on the Titan Krios (Cryo-EM facility of MPI of Biochemistry; FEI/Thermofisher) by the author of the thesis with the help of M. Strauss. The FEI Titan Krios was operated at 300 kV and equipped with a K2 Summit direct detector (Gatan) used in counting mode as well as a post-column GIF. The dataset included 3467 movies collected at a nominal magnification of 81,000x, which equals to a pixel size of 1.746 Å/pixel at the specimen level.

The total electron dose exposure amounted to 52 $e^-/Å^2$ distributed over 60 frames. Target defocus range was set between 1.5-3 µm and data was acquired using SerialEM software.

3.4.3 Cryo-EM data processing

The data processing scheme applied in the course of the cryo-EM analysis is illustrated in Fig. 3.1: in the first step, the individual frames of each movies were aligned using MotionCor 2 software (Zheng et al., 2017). This process allows for beam-induced motion correction as well as dose compensation, which differ in the individual frames as a result of long exposure (15 s). The thereby summed micrographs were then subjected for CTF estimation by Gctf (Zhang, 2016) and subsequent particle picking was performed with Gautomatch (`http://www.mrc-lmb.cam.ac.uk/kzhang/` K. Zhang, MRC LMB, Cambridge, UK). First, particles were picked without templates to avoid any bias and 2D classification was performed in Relion 3.0 (Zivanov et al., 2018) for initial analysis of the particle composition (2D classes are shown in Results Chapter Fig. 4.4). As these 2D classes were similar to the overall composition of PRC2-AEBP2 on dinucleosomes (EMD-7306; Poepsel et al. (2018)), templates were generated of the corresponding low-pass filtered model. Template-based picking with Gautomatch then yielded 1,028,229 of initial particle candidates, which were directly without 2D classification subjected to several rounds of 3D classification with a Bayesian fudge factor (T value) of 8, in order to separate particles from contamination. For the first rounds of 3D classification, a low-pass filtered model of PRC2-AEBP2 on dinuclesomes was also used as a reference, later the best 3D model generated from a current 3D classification run was used as a reference for the subsequent run. All 3D classifications/refinements and subsequent steps were performed in Relion 3.0 (Zivanov et al., 2018). Two best 3D classes of the last 3D classification round were subsequently subjected to 3D refinement, and the resulting reconstruction was further classified into 10 classes by 3D classification without translational and spatial sampling. Four best 3D classes based on nominal overall resolution, appearance of secondary structures and rotational/translational accuracy were then subsequently further classified by iterative rounds of focused 3D refinement (using a mask and solvent flattening), per particle CTF refinement and Bayesian polishing, finally yielding a map of PRC2-PHF1 on dinuclesomes of 5.24 Å resolution according to the gold-standard FSC criterion of 0.143 (Rosenthal and Henderson, 2003).
For analysis of the H3 tail in the EZH2 active site, particle subtraction of the allosteric nucleosome and adjacent PRC2 parts ($EZH2^{SBD/SANT1}$ and EED) with subsequent 3D refinement focused on the substrate nucleosome (using a mask and solvent flattening) was performed (Bai et al., 2015; Zhou et al., 2015; Ilca et al., 2015). Model and masks for signal particle subtraction were generated in UCSF Chimera (Pettersen et al., 2004) and Relion 3.0 (Zivanov et al., 2018). These steps resulted in a 4.36 Å map (according to the gold-standard FSC criterion of 0.143 Rosenthal and Henderson (2003)) of EZH2 on substrate nucleosome, referred to as $EZH2_{sub}$-Nuc_{sub}.

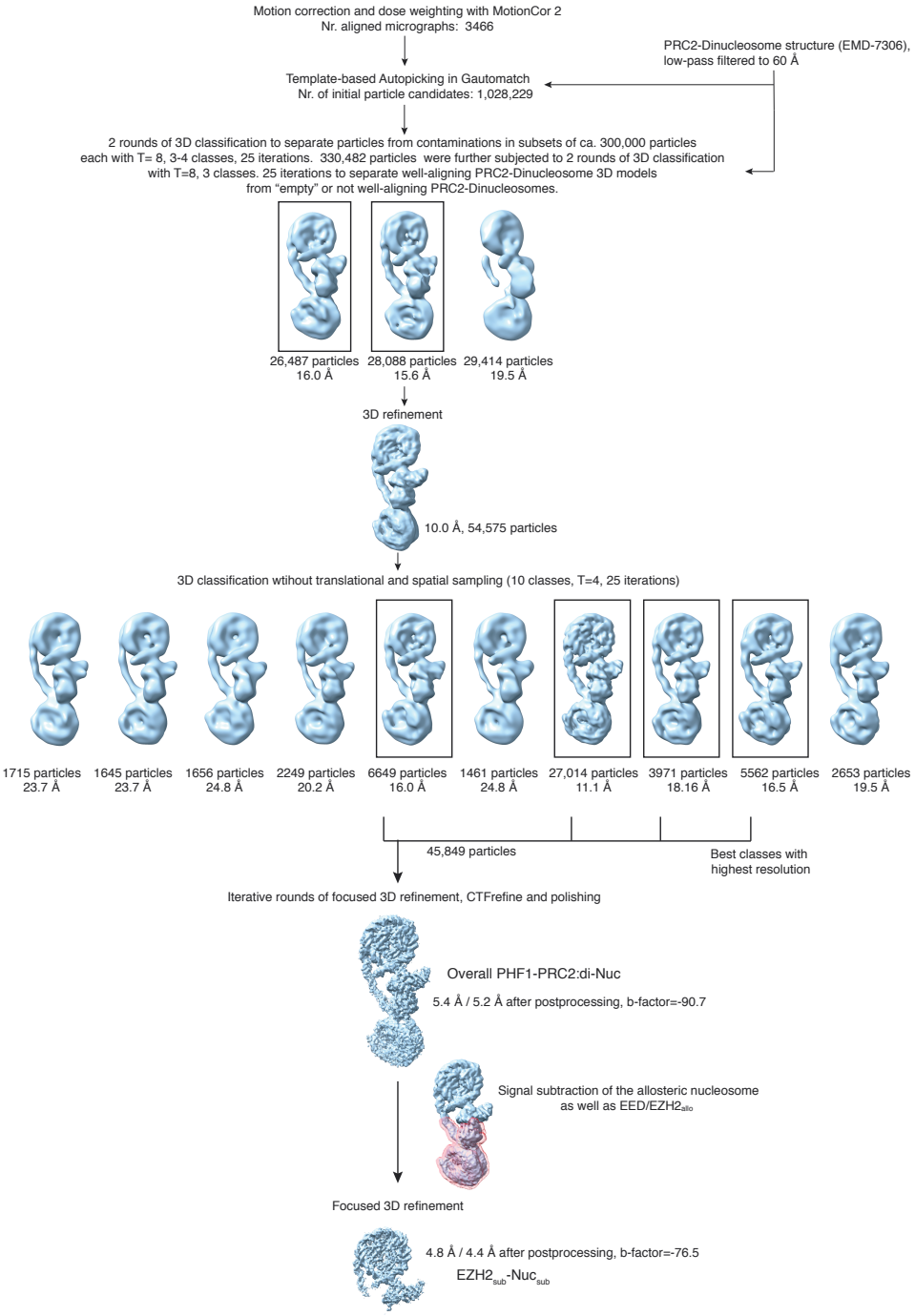

Figure 3.1: Detailed cryo-EM data processing scheme used to obtain a reconstruction of PRC2-PHF1 on heterodimeric nucleosome.

Mask used for signal subtraction is shown in red. Boxes indicate classes chosen for the next steps of processing.

Further improvement of the density and side chain visibility was achieved by Multisharpen function in Coot (additional applied b-factor: - 66) (Emsley et al., 2010) and Phenix Resolve Density modification (Terwilliger et al., 2020) performed using the two half maps generated by 3D refinement of $EZH2_{sub}$-Nuc_{sub}.
All maps were used for model building and maps resulting from Coot Multisharpen function and Phenix Density modification were used to illustrate $EZH2_{sub}$-Nuc_{sub}.

3.4.4 Cryo-EM data fitting, modeling and refinement

Available X-ray and cryo-EM structures (see Table 3.1) were rigid-body fitted into the obtained map of PHRC2-PHF1 on dinucleosomes using UCSF Chimera (Pettersen et al., 2004). Manual refitting and remodeling was performed in USCF Chimera and Coot (Emsley et al., 2010) for poorly fitting parts (e.g $EZH2^{SBD/SANT1}$ region).
The generated model was used as an initial model for $EZH2_{sub}$-Nuc_{sub} map. Missing parts were modeled *de novo*, such as the EZH2 "bridge helix" based on Kasinath et al. (2020) and missing H3 tail residues 30-37 connecting the H3 peptide built in the crystal structure of PRC2 and the histone H3 end of the nucleosome structures. Morph fitting routine in Coot (Casañal et al., 2020) was used to further improve the fitting of individual parts of the $EZH2_{sub}$-Nuc_{sub} model.
Hydrogens were added to the final best-fitting model and Real-Space Refinement in Phenix (Afonine et al., 2018) (phenix-1.18rc1-3777) with applied secondary structure (Sobolev et al., 2015), reference model (PRC2: PDB 5HYN (Justin et al., 2016) and nucleosomes: PDB 6T9L Wang et al. (2020) (Table 3.1) and Ramachandran restraints was performed. Table 3.2 summarizes the cryo-EM data collection as well as processing and model refinement statistics. Visualization of structures was performed with UCSF ChimeraX (Goddard et al., 2018) and PyMOL2 (https://pymol.org/2/).

Table 3.1: List of X-ray- and cryo-EM models used for cryo-EM data fitting, modeling and refinement.

Name/PDB	Source	Use
Model of Dinuc (35 bp linker). Suppl. Dataset1 of including fitted PDB 3LZ1	Poepsel et al. (2018) Vasudevan et al. (2010)	Fitting into the PRC2-PHF1-Dinuc map the PRC2-PHF1-Dinuc map
EZH2-EED-$SUZ12^{VEFS}$ model PDB 5HYN	Justin et al. (2016)	Fitting into the PRC2-PHF1:Dinuc map and $EZH2_{sub}$-Nuc_{sub} map. Reference model restraints in in real-space refinement
nucleosome X-ray model PDB 1AOI	Luger et al. (1997)	Fitting into the PRC2-PHF1-Dinuc map and $EZH2_{sub}$-Nuc_{sub} map.
Cryo- EM model of nucleosome PDB 6T9L	Wang et al. (2020)	$EZH2_{sub}$-Nuc_{sub} map fitting and reference model restraints in in real-space refinement
Cryo-EM models of PRC2-AEBP2 PDBs: 6C23 and 6C24	Kasinath et al. (2018)	Comparison and refitting

Table 3.2: Summary of cryo-EM data collection, processing and refinement statistics.

Cryo electron microscopy data collection		
Microscope	FEI Titan Krios GII	
Voltage (kV)	300	
Camera	Gatan K2-Summit	
Energy Filter	Gatan Quantum-LS (GIF)	
Pixel size (Å/pix) (calibrated)	1.75	
Nominal magnification (x)	81000	
Preset target global defocus range (μm)	0.5 - 3.5	
Total electron exposure (fluence, $e^-/Å^{2)}$	52,96	
Exposure rate (flux) ($e^-/ Å^2/s$)	3,47	
Nr. of frames collected per micrograph	60	
Energy filter slit width (eV)	20	
Automation Software	SerialEM	
3D reconstruction		
Number of movies	3466	
Initially selected particle candidates	1,028,229	
Final number of particles	45,849	
	Overall PHF1-PRC2:diNuc (EMD-11912)	**$EZH2_{sub}$-Nuc_{sub} (EMD-11910)**
Resolution FSC independent halfmaps (0.143) masked (Å)[a]	5.24	4.36
Local resolution range (Å)	4.01 - 24.97	4.01 – 15.00
Sharpening B-factor ($Å^2$)	-90.7	-76.5
Refinement **$EZH2_{sub}$-Nuc_{sub}**		**$EZH2_{sub}$-Nuc_{sub} (PDB 7AT8)**
No. atoms		29151
Residues		Protein: 1186 Nucleotide: 312
Ligands		ZN: 7
CC_{mask}, CC_{box}, CC_{peaks}, CC_{volume}[b]		0.75, 0.83, 0.66, 0.74
Mean CC for ligands		0.71
Resolution FSC masked map vs. model (0/0.143/0.5) (Å)[b]		4.3/4.3/4.6
R.m.s. deviations		
Bond lengths (Å)		0.004
Bond angles (°)		0.828
Ramachandran favored (%)		97.49
Ramachandran gen. allowed (%)		2.25
Ramachandran disallowed (%)		0.26
MolProbity score		1.52
Clash score		7.65

ADP (B factors)		
Iso/Aniso (#)		15955/0
Min/max/mean		
Protein		30.00/254.73/119.01
Nucleotide		95.54/222.73/130.75
Ligand		150.31/257.19/166.85
Rotamer outliers (%)		0.40
Cβ outliers (%)		0.00
CaBLAM outliers (%)		1.41

[a]according to the Fourier Shell Correlation (FSC) cut-off criterion of 0.143 defined in Rosenthal and Henderson, (2003)
[b]according to the map-vs.-model Correlation Coefficient definitions in Afonine et al., (2018)

3.5 Biochemical assays

3.5.1 Electrophoretic mobility shift assay (EMSA)

For EMSAs, octamers were reconstituted on nucleosomal DNA labeled with 6 - Carboxyfluorescein (6-FAM) (using 6-FAM primers in Table 2.2 and following the protocol described in 3.3.2). Increasing amounts of PRC2 constructs (ranging between 17-3500 nM) were pipetted to 45 nM of 6-FAM-labeled mononucleosomes in EMSA Buffer (25 mM Hepes-NaOH pH 7.9, 50 mM NaCl, 0.05 % Tween-20, 5 mM $MgCl_2$ and 4 % Glycerol) and 10 µl of the binding reactions were loaded on an agarose gel (1.0-1.2 % in 0.4 x TBE; SeaKem Agarose from Lonza) free of SYBR Safe. Gel electrophoresis was run protected from light in pre-chilled 0.4 TBE Buffer for 45 min at 60 V. Imaging of the fluorescence signal was performed using the Typhoon FLA 9500 imager (GE Healthcare) equipped with the fitting Cy2 filter. Fiji software (Schindelin et al., 2012) was used for analysis of the densitometry signal and obtained values were used to correct for background signal and to calculate the fraction of bound/unbound nucleosomes. Area boxes for value read-outs of integrated densities were created for each lane: 'Unbound boxes' indicating unbound nucleosomes and 'bound boxes', which included everything above 'unbound boxes' together with shifted nucleosomes. Background correction was performed by subtraction of the control signal i.e. lane nr. 1 for 'bound' and lane nr. 10 for 'unbound'. Finally, the fraction of bound to unbound nucleosomes was calculated by the dividing the background corrected values of each line by the total signal (sum of unbound and bound). Prism 8 (GraphPad) was used to fit a Hill function and for illustrations of the plots.

3.5.2 Histone methyltransferase (HMTase) assay and western blot (WB) imaging

For HMTase assays shown in this thesis, indicated increasing amounts of PRC2 complexes were incubated with either 446 nM of mononucleosomes or 223 nM of dinucleosomes. The reactions were performed in 20 mM Hepes-NaOH pH 7.8, 50 mM NaCl, 2.5 mM $MgCl_2$, 5 % glycerol, 0.25 mM EDTA, 0.5 mM DTT and 80 µM S-adenosylmethionine (SAM) (HMTase reaction buffer). After incubation at RT for 90 min, the reactions were stopped by addition of SDS loading buffer (1x final concentration; see Table 2.6)) and heat inactivation for 5 min at 95 °C. For western blots, proteins were first separated on a 16 % (w/v) Tris-Glycine self-cast SDS polyacrylamide gel (see 3.3.3) in a 1 x SDS running buffer (Table 2.6). Subsequently, wet western blot assembly was performed under the chemical hood in 1 x transfer buffer (Table 2.6) in a way so that the SDS gel was facing the nitrocellulose membrane (0.2 µm Amersham Protran Premium) surrounded by pre-wet Whatman papers. Transfer was performed using the TE 22 Mini Tank Transfer Unit (GE Healthcare) and corresponding equipment. Western blot assemblies in cassettes were positioned in the tank according to manufacturer's instructions and electrophoresis was run at 4 °C with gentle stirring first for 10 min at 90 V, then 30 min at 60 V. After transfer, membranes were blocked at 4 °C with TBS-T solution containing 4% BSA ON under constant gentle agitation. The primary antibody (H3K27me1/me3; for list of antibodies and respective dilutions see 2.4) was incubated in fresh 4 % BSA-TBS-T solution ON at 4 °C and the primary antibody for H4 was added in the last 1.5 hrs of this step (the last step can be performed at RT). Subsequently, the membrane was washed 3 x for 5 min at RT in TBS-T under agitation before addition of the horseradish peroxidase (HRP) conjugated secondary antibody for 3 hrs at RT. Before detection, the membrane was washed again 3 x for 5 min each at RT in TBS-T. A 1:1 mixture of detection reagents A and B (ECL Prime, Amersham) was applied to the membrane prior to imaging and the chemiluminescence signal was subsequently detected by ImageQuant LAS 4000 (GE Healthcare) and densitometry signal was evaluated in Fiji (Schindelin et al., 2012). Background correction was done using the signal of the control lane (no PRC2, lane nr.1 in the respective figures) and normalization of each lane was performed calculating the relative amounts of tri-/monomethylation in respect to the amounts shown by the highest concentration on unmodified nucleosomes (lanes nr. 4 in the respective figures; set to 100 %). Prism 8 (GraphPad) was used for graphical representations.

3.5.3 Mass spectrometry-based peptide HMTase assay

Peptide HMTase reactions were set up by incubation of 500 nM PRC2 complexes with 2 µM of the respective peptides (unmodified or H3K36me3 peptide; Table 2.7) in HMTase reaction buffer (20 mM Hepes-NaOH pH 7.8, 50 mM NaCl, 2.5 mM $MgCl_2$, 5 % glycerol, 0.25 mM EDTA, 0.5 mM DTT and 80 µM SAM). Methyltransferase reactions were allowed to proceed ON at RT and were quenched by addition of 1 % trifluoroacetic acid (TFA). In order to remove PRC2 prior to mass spectrometry (MS) analysis of the reactions, home-

made stage tips with poly(styrenedivinylbenzene) copolymer (SDB-XC) were used (Rappsilber et al., 2007). Stage tips were first washed with methanol and buffer B (0.1 % (v/v) formic acid, 80 % (v/v) acetonitrile). Following sorbent equilibration with buffer A (0.1% (v/v) formic acid)), 40 µl of reaction sample was applied and tips were subsequently washed several times. 20 µl of samples eluted in buffer B from the stage tips were subsequently injected into the Bruker maXis II ETD mass spectrometer using an Agilent HPLC at a flow rate of 250 µl/min and 0.05 % TFA in 70 % acetonitril:H2O as solvent for ESI-MS time-of-flight analysis. Peptide ionization was performed at a capillary voltage of 4500 V and an end plate offset of 500 V. A spectra rate of 1 Hz and a collision cell energy of 15 eV was used for acquisition of a full scan MS spectra (200-1600 m/z).

Processing of raw data was performed using Bruker Compass DataAnalysis. First, m/z spectra were deconvoluted (power of 10,000) and the resulting neutral spectra peaks were subsequently integrated. For quantification, the sum of the four monomethylation peak areas was divided by the sum of these peaks together with the sum of the first 4 peaks of the input peptides (peaks used for analysis are shown in the respective figures). Values were plotted using Prism 8 (GraphPad).
MS analysis of the peptides shown in this thesis was performed by E. Weyher (MS facility, MPI of Biochemistry).

Chapter 4

Results

4.1 Cryo-EM studies of PRC2-nucleosome interaction

4.1.1 Sample preparation of the PRC2-PHF1-dinucleosome complex

In order to characterize the interaction of PRC2-PHF1 with nucleosomes, the human complex, composed of EZH2, EED, RBBP4, SUZ12 and PHF1, was expressed in insect cells and purified (Fig. 4.1 A, B) (described in more detail in section 3.2 and 3.3.1 of the Methods chapter). The purification yielded a homogeneous complex of high purity as judged by SEC (Fig. 4.1 A) and SDS PAGE analysis of the peak fractions (Fig. 4.1 B).

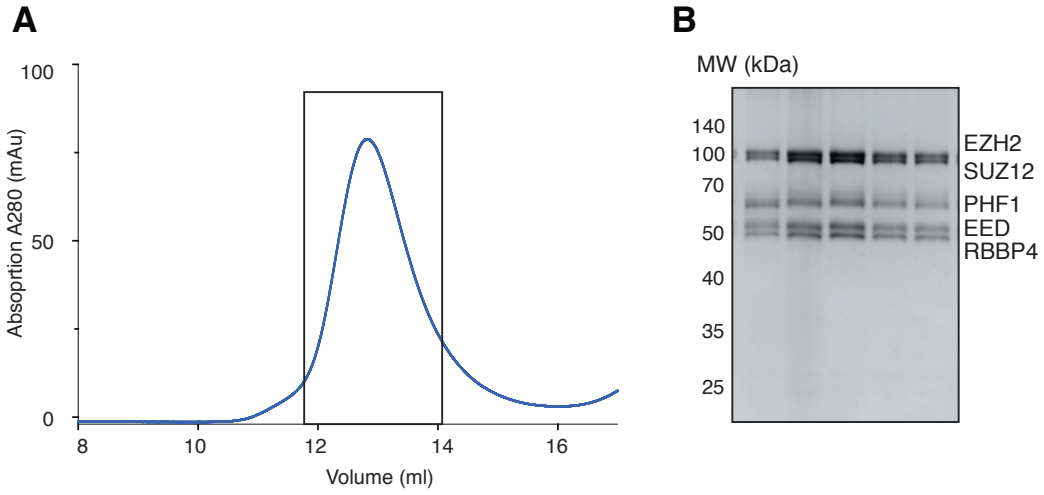

Figure 4.1: SEC and SDS PAGE of PRC2-PHF1.

A. Representative SEC profile of the purified PRC2-PHF1 complex. Black square indicates the peak analyzed with SDS PAGE. B. Coomassie-stained SDS PAGE analysis of the SEC peak fraction shown in A.

Previous cryo-EM data analyses performed during the experimental part of this thesis showed that PRC2-PHF1 disassembled after grid preparation, possibly due to protein denaturation at the water-air interface (not shown). If crosslinked using the GRAFIX method (Kastner et al., 2008; Stark, 2010) or BS3 (bis(sulfosuccinimidyl)suberate) following the protocol described in Kasinath et al., 2018, PRC2-PHF1 remained intact, however the processed data didn't show any high resolution features (Fig. 4.2 A, B).

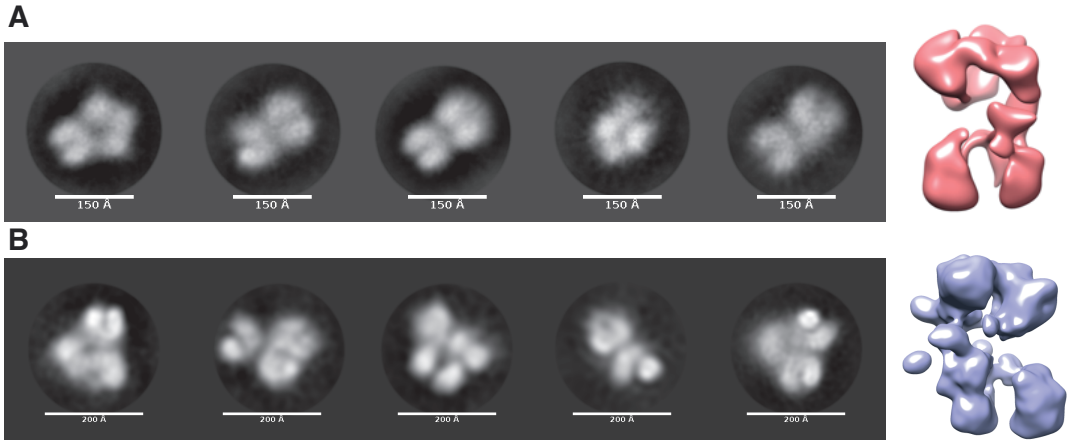

Figure 4.2: Cryo-EM studies on crosslinked PRC2-PHF1.

A. Representative 2D classes and low-resolution 3D models of PRC2-PHF1 after GRAFIX crosslinking (A) (Kastner et al., 2008; Stark, 2010) and after BS3 crosslinking (B) following protocols described in Kasinath et al., 2018; Poepsel et al., 2018.

Therefore, in order to stabilize the complex for grid preparation, PRC2-PHF1 was assembled on a heterodimeric *Xenopus laevis* (*X.l.*) dinucleosome (with the help and contribution of S. Poepsel) as was previously described for PRC2-AEBP2 (Poepsel et al., 2018) (see Methods 3.3.2). A simplified overview of the dinucleosome assembly is shown in Fig. 4.3 A. Briefly, substrate nucleosomes were assembled by adding *DraIII* digested DNA to assembled, SEC purified, unmodified octamers (exemplary shown in Fig. 4.3 B, C). Similarly, *DraIII* digested DNA and H3Kc27me3 modified octamers (Simon et al., 2007) were used to assemble the allosteric mononucleosomes. For biochemical assays, the usual protocol for nucleosome assembly entails careful titration of different ratios of octamer to DNA (ranging between 0.8 to 1.4) DNA to octamer and screening for a condition where no free DNA can be observed on the analytical agarose gel (exemplary shown in Fig. 4.3 D). Mononucleosomes prepared for ligation to dinucleosomes were purified differently, using a large-scale native polyacrylamide gel purification (Biorad PrepCell) (for more details see Methods 3.3.2). Fractions were analyzed on an analytical native polyacrylamide gel (Fig. 4.3 E; upper gel shows substrate mononucleosomes, lower gel shows allosteric mononucleosomes) and fractions containing successfully reconstituted nucleosomes without free DNA were concentrated. Purified substrate and allosteric mononucleosomes were then ligated

together and further purified from aberrant mononucleosomes again using large scale native polyacrylamide gel purification. Fractions containing dinucleosomes running at expected height on an analytical native gel (Fig. 4.3 F) were subsequently concentrated. After ligation and purification, the resulting dinucleosome contained one substrate nucleosome with an unmodified H3 tail and one allosteric nucleosome carrying the trimethyl-lysine-analog of the activating mark H3Kc27me3 (Margueron et al., 2009; Jiao and Liu, 2015; Simon et al., 2007), connected by a 35 bp linker DNA. Binding of PRC2-PHF1 to dinucleosomes was confirmed by EMSA (Fig. 4.3 G). PRC2-PHF1 on the heterodimeric dinucleosome shall subsequently be referred to as PRC2-PHF1:Dinuc. Prior to grid preparation, PRC2-PHF1 was buffer exchanged into the cryo-EM buffer and incubated with the reconstituted dinucleosome as described in Methods section 3.4.1. Grids prepared by vitrification were used for subsequent data collection.

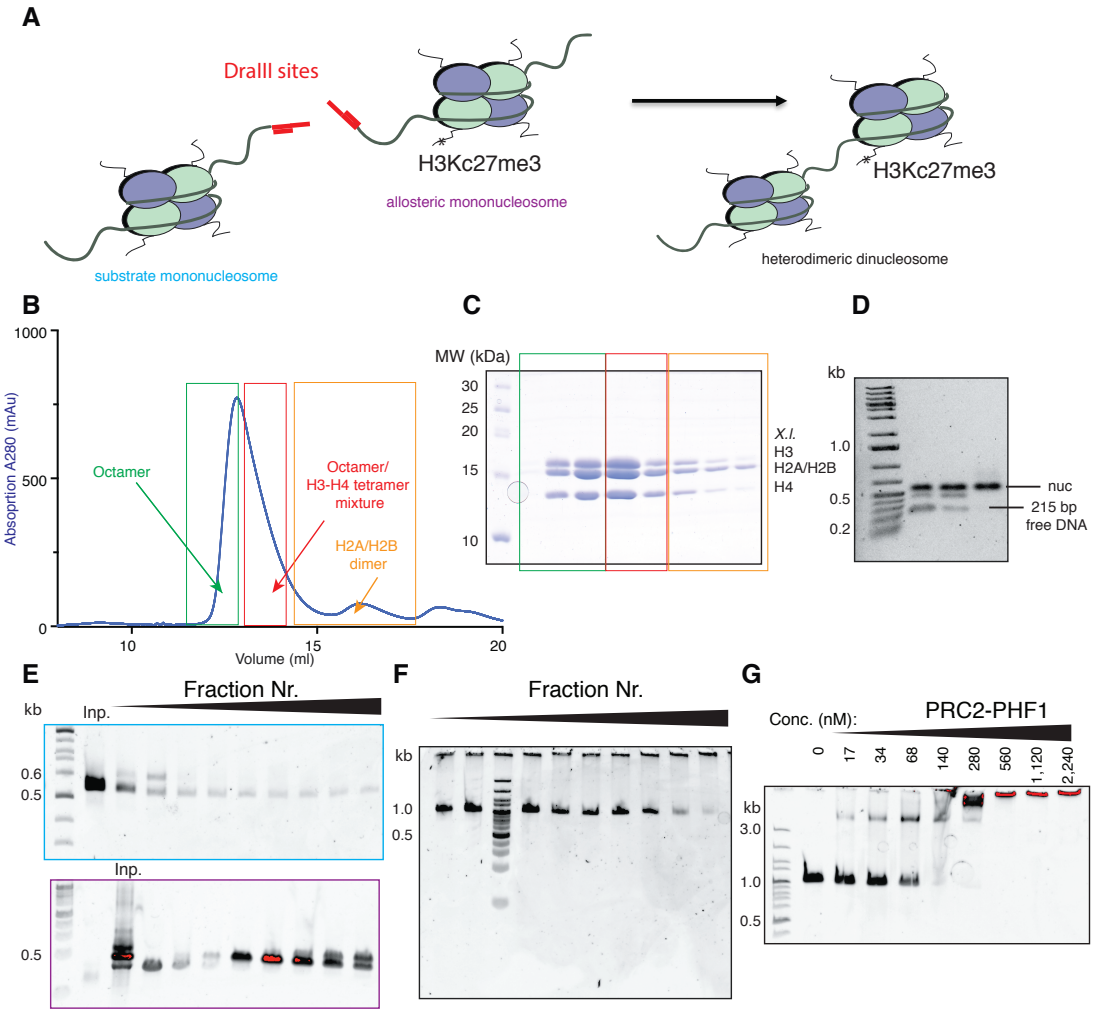

Figure 4.3: Reconstitution of heterodimeric dinucleosomes.

A. Simplified overview of the dinucleosome assembly. The resulting heterodimeric dinucleosome used for cryo-EM analysis contained one substrate nucleosome (unmodified H3 tail) and one allosteric nucleosome (H3 tail containing the H3Kc27me3, shown to allosterically activate PRC2 (Margueron et al., 2009; Jiao and Liu, 2015; Simon et al., 2007)). B. Representative size-exclusion chromatography (SEC) profile of an *X.l.* octamer purification. The first half of the main peak contains the octamer (green square), while the second half contains a mixture of H3-H4 tetramer and octamer (red square). The main peak is followed by a peak containing the H2A/H2B dimer (orange square). C. Typical coomassie-stained SDS PAGE analysis of an octamer purification after SEC. Colored squares show the respective SEC peak fractions in B. D. Representative analytical agarose gel showing titrations of different ratios of DNA:octamer. Band at 0.2 kb indicate free 215 bp nucleosomal 601 DNA (Lowary and Widom, 1998), while band at 0.5 kb shows the assembled nucleosome. The band in-between indicates possible sliding of the nucleosome on the nucleosomal DNA. Ratio of octamer:DNA where no nucleosomal DNA can be observed on the gel as in lane nr. 4 was chosen for large-scale nucleosome assembly. E. Analytical native polyacrylamide gel images of substrate mononucleosomes (top) and H3Kc27me3-modified (Simon et al., 2007) mononucleosomes (bottom) after large scale native gel purifications. F. Analytical polyacrylamide native gel image of heterodimeric dinucleosome used for cryo-EM after large scale native gel purification. G. EMSA of increasing concentrations of PRC2-PHF1 on heterodimeric dinucleosomes shows efficient binding. Lower band shows unbound dinucleosome, shifted band shows PRC2-PHF1:Dinuc complex. Subfigures E-G were performed in collaboration with S. Poepsel.

4.1.2 Cryo-EM data analysis

Initial analysis of the cryo-EM dataset (processing scheme is described in Methods section 3.4.3 and visualized in Fig. 3.1) of PRC2-PHF1:Dinuc showed dense but homogenous particle distribution (Fig. 4.4 A left panel). Particles were initially picked without templates and analyzed by 2D classification to visualize which proteins of the PRC2-PHF1-Dinuc assembly were contained in the particles (Fig. 4.4 A right panel). 2D classes revealed clearly the two nucleosomes bridged by the linker DNA and an expected overhang DNA on the substrate nucleosome. The 'upper' catalytic unit of PRC2 (Justin et al., 2016) sits between the two nucleosomes and the 2D classes mostly resembled the 2D classes of the previously published PRC2-AEBP2:Dinuc complex (Poepsel et al., 2018). Templates generated from this structure were therefore used for re-picking. Picked particles were further subjected to 3D classification (details of the steps involved in processing are described in the Methods). Several rounds of 3D classification and refinement resulted in a map of the PRC2-Dinuc with an overall resolution of 5.2 Å after postprocessing according to the gold-standard FSC cut-off at 0.143 (Fig. 4.4 B) (Rosenthal and Henderson, 2003).

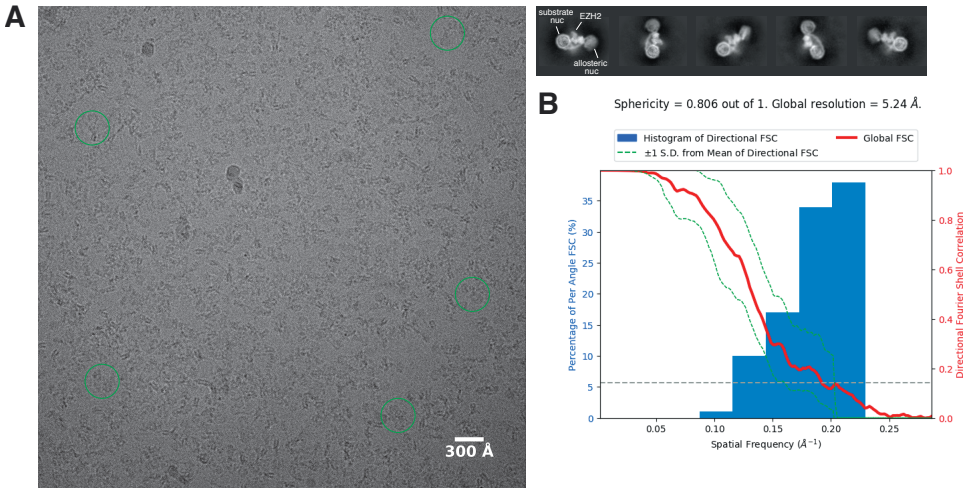

Figure 4.4: Initial cryo-EM data analysis of PRC2-PHF1 on an asymmetric dinucleosome.

A. Left: Typical micrograph of the cryo-EM data collection of PRC2-PHF1:Dinuc (while most of the screening datasets were collected by the author of the thesis, the final dataset shown was collected with the help of M. Strauss). Right: Representative 2D classes of particles picked without templates of PRC2-PHF1:Dinuc showing the 'upper' catalytic lobe of PRC2 (Ciferri et al., 2012; Jiao and Liu, 2015; Justin et al., 2016) sitting in-between the two nucleosomes, connected by 35 bp of DNA. B. Output of the 3DFSC webserver (https://3dfsc.salk.edu/) (Tan et al., 2017) shows the Fourier Shell Correlation (FSC) as a function of spatial frequency. Global FSC (red) is generated from masked independent half maps of PRC2:Dinuc. Directional FSC is indicated by the blue histogram and deviation from mean is shown as a green dotted lines. Gold-standard FSC cut-off at 0.143 was used to estimate the overall resolution of 5.24 Å of PRC2-PHF1:Dinuc map (grey dotted line) (Rosenthal and Henderson, 2003).

The substrate nucleosome and adjacent parts of PRC2 were better resolved (4-6 Å,

red-yellow Fig. 4.5 A), while the allosteric nucleosome as well as the proximal EED and SBD domains of EZH2 were of lower resolution (10-25 Å, green-blue Fig. 4.5 A). Spherical angular distribution analysis of the final particles showed only few missing views as indicated by few or missing blue rods and the sphericity value of 0.806 (Fig. 4.4 B). Views with a high number of particles are shown as red rods. (Fig. 4.5 B).

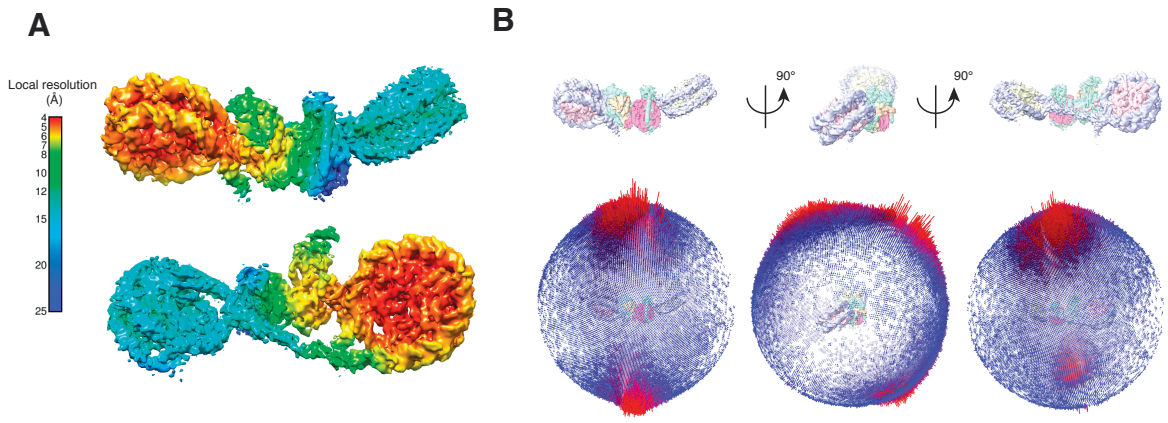

Figure 4.5: Cryo-EM data analysis of the PRC2-PHF1 on an asymmetric dinucleosome.

A. Local resolution estimation of the 5.2 Å cryo EM map of PRC2-PHF1:Dinuc. Regions of higher resolution are shown in colors ranging from red to yellow, while regions of lower resolution are shown in green-blue colors. B. Spherical angular distribution of particles views contained in the final reconstruction of the PRC2-PHF1:Dinuc structure. Blue/no rods: few particles per view or missing views. Red rods: views with a high number of particles.

In the next step the density of the allosteric nucleosome and adjacent parts of PRC2 including EED and the SBD-SANT1 domains of EZH2 were subtracted from the particles and focused 3D refinement was used to obtain a higher resolution and more detailed structural information for the substrate nucleosome and EZH2 (see Methods 3.4.3 and Fig. 3.1). This step resulted in a corresponding map of $EZH2_{sub}$-Nuc_{sub} of overall 4.36 Å resolution according to the gold-standard cut-off at 0.143 (Rosenthal and Henderson, 2003) (Fig. 4.6 A). Local resolution estimation analysis showed that the core of the substrate nucleosome is around 4 Å with the H3 tail and tail-adjacent part of EZH2 at around 4.5-5 Å resolution while the nucleosomal DNA and tail-nonadjacent parts of EZH2 are between 7-15 Å (Fig. 4.6 B).

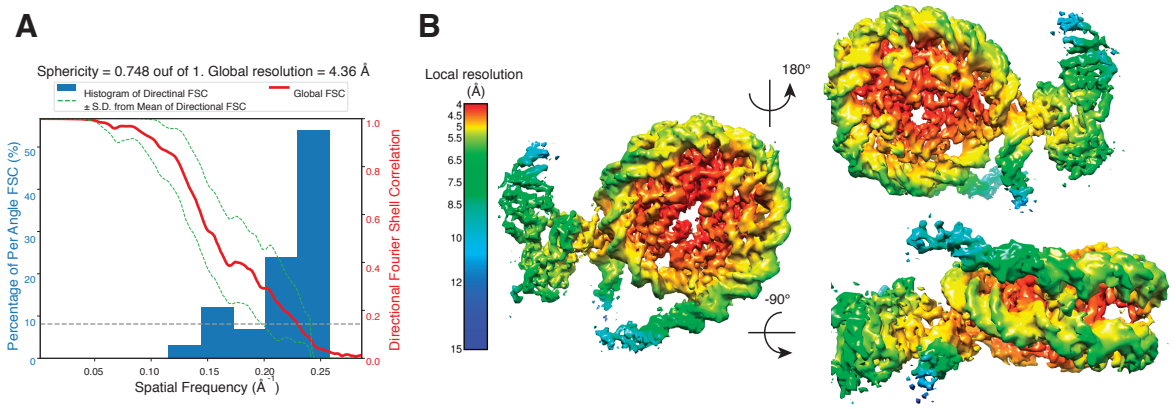

Figure 4.6: The resolution of $EZH2_{sub}$-Nuc_{sub} map is improved after focused refinement.

A. Half maps and 3D resolution estimation output generated by the 3DFSC Processing Server (https://3dfsc.salk.edu/ (Tan et al., 2017)). The Fourier Shell Correlation (FSC) is plotted as a function of spatial frequency. Global FSC (red) was generated from masked independent half maps of $EZH2_{sub}$-Nuc_{sub}. Directional FSC is shown as a blue histogram and deviation from mean as a green dotted line. Gold-stadard FSC cut-off at 0.143 was used to estimate the overall resolution of 4.36 Å of the $EZH2_{sub}$-Nuc_{sub} map (grey dotted line) (Rosenthal and Henderson, 2003). B. Local resolution estimation of the $EZH2_{sub}$-Nuc_{sub} map after focused refinement. The substrate nucleosome and the adjacent part of the catalytic subnunit EZH2 show improved resolution (colors red to yellow; 4.5-5 Å).

The model of the substrate nucleosome with EZH2 ($EZH2_{sub}$-Nuc_{sub}) was generated by using nucleosome and PRC2 models obtained by X-Ray crystallography and cryo-EM (for list of models/PDBs used for Cryo-EM data fitting, modeling and refinement see Table 3.1). A pseudoatomic model of the missing H3 tail residues was manually built (described in more detail in methods section 3.4.4). The final model was then subjected to real space refinement in Phenix (Afonine et al., 2018) and showed good correlation with the observed densities in the map (Fig. 4.7 A and B). A summary of cryo-EM data collection, processing and refinement statistics is shown in Table 3.2.

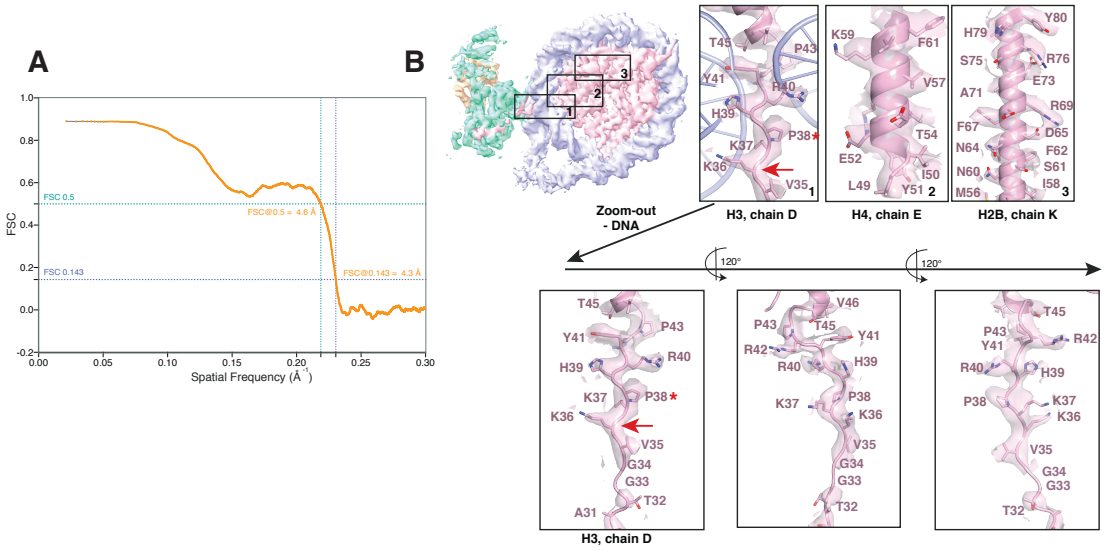

Figure 4.7: The quality and resolution of the EZH2$_{sub}$-Nuc$_{sub}$ map is significantly improved after focused refinement.

A. FSC calculated between the atomic model and the map (masked) of EZH2$_{sub}$-Nuc$_{sub}$ after real-space refinement (Afonine et al., 2018). Dotted lines represent the cut-off of 0.143/0.5 at 4.2/6.2 Å (see also table 3.2) at which map and model Fourier coefficient are most similar (Rosenthal and Henderson, 2003; Henderson et al., 2012; Rosenthal and Rubinstein, 2015). B. Regions of well resolved density within the EZH2$_{sub}$-Nuc$_{sub}$ map showing side chain information. Last residue of the H3 tail visible in nucleosome crystal structures (usually H39 or P38) is shown by a red asterics. The location of K36 is indicated by a red arrow.

4.1.3 Structural characterization of PRC2 nucleosome interaction and the position of K36

Analysis of the obtained overall PRC2-PHF1-Dinuc map showed PRC2 sitting in-between the two nucleosomes, which are connected by a 35 bp linker DNA, similar to previously described in Poepsel et al. (2018) (Fig. 4.8 A, B). The density of PRC2 accounted for the catalytic lobe of PRC2 consisting of EZH2, EED and C-terminus (VEFS domain) of SUZ12 (aa 561-685) (Justin et al., 2016) but not for PHF1 and the 'bottom lobe' of PRC2 consisting of RBBP4 and the N-terminus of SUZ12 (Ciferri et al., 2012; Kasinath et al., 2018) (Fig. 4.8 A). The structure of PRC2-PHF1:Dinuc shows two interaction sites of the 'upper lobe' of PRC2 with the nucleosomes: the substrate nucleosomal DNA is bound by EZH2, likely via four residues of EZH2^{CXC} (K563, Q565, K569, Q570) (Fig. 4.8 C; see also Poepsel et al., 2018). The allosteric nucleosomal DNA is seemingly recognized by a combined interface of EED and the EZH2^{SBD} (potentially also SANT1) (Fig. 4.8

D; also Poepsel et al., 2018). Unlike many other epigenetic complexes, PRC2, based on this structure, seems to not engage with the acidic patch of the nucleosome (Fig. 4.8 B). Overall, both PRC2-PHF1 as well as PRC2-AEBP2 seem to bind chromatin in similar ways (Poepsel et al., 2018).

Based on the available side chain information and previous structural studies of the H3K27M peptide bound to the human PRC2 catalytic lobe crystal structure (Justin et al., 2016), a pseudoatomic model of the N-terminal H3 tail was built (Fig. 4.8 E). Analysis of the model reveals the H3 tail to be in an extended conformation between the exit site of the nucleosome and the catalytic site of EZH2. Residues of the H3 tail are engaged in several interactions with EZH2: as previously described H3K27 is positioned within the aromatic cage of EZH2 consisting of Y641, F667, F724, Y726, Y728 while H3R26 is forming ionic bonds with D652 and Q648 of EZH2 (Justin et al., 2016). Additionally, the model suggests two regions of hydrophobic interactions: (1) H3A29/P30 and A697, V699, I708 of EZH2 and (2) H3V35 with F542, F557 and P558 (Fig. 4.8 E). The EM density is visibly disrupted between T32 and V35 of the H3 tail, likely owing to the flexibility of the G33/34 stretch. The density-modified (Terwilliger et al., 2020) map of $EZH2_{sub}$-Nuc_{sub} shows a very clearly defined side chain density for H3K36 allowing deeper analysis of its location and possible interactions (Fig. 4.8 F). Unmethylated H3K36 is facing the nucleosomal DNA and is juxtaposed to the EZH2 CXC- nucleosomal DNA interaction site described above. The direction of its side chain density suggests a potential long-range electrostatic interaction between the ε-amino group of the lysine and the phosphate backbone of the nucleosomal DNA. An additional potential interaction partner is provided by the carbonyl of the CXC residue Q570 constituting a possible direct link between H3K36 and recognition of the nucleosomal DNA by EZH2 CXC (Fig. 4.8 F). The constraints created by this interaction, in which H3K36 is sandwiched, suggest that the PRC2 inhibiting mark H3K36me3 would be too bulky in this position. Additionally, the positive charge of the lysine becomes becomes shielded by and distributed into the methyl groups. A study by Jani et al., 2019 previously described a possible mechanism of H3K36 recognition by E579. In the model of $EZH2_{sub}$-Nuc_{sub} the location of E579 and H3K36 would require major rearrangement and is hence incompatible with the proposed mechanism (Fig. 4.8 G).

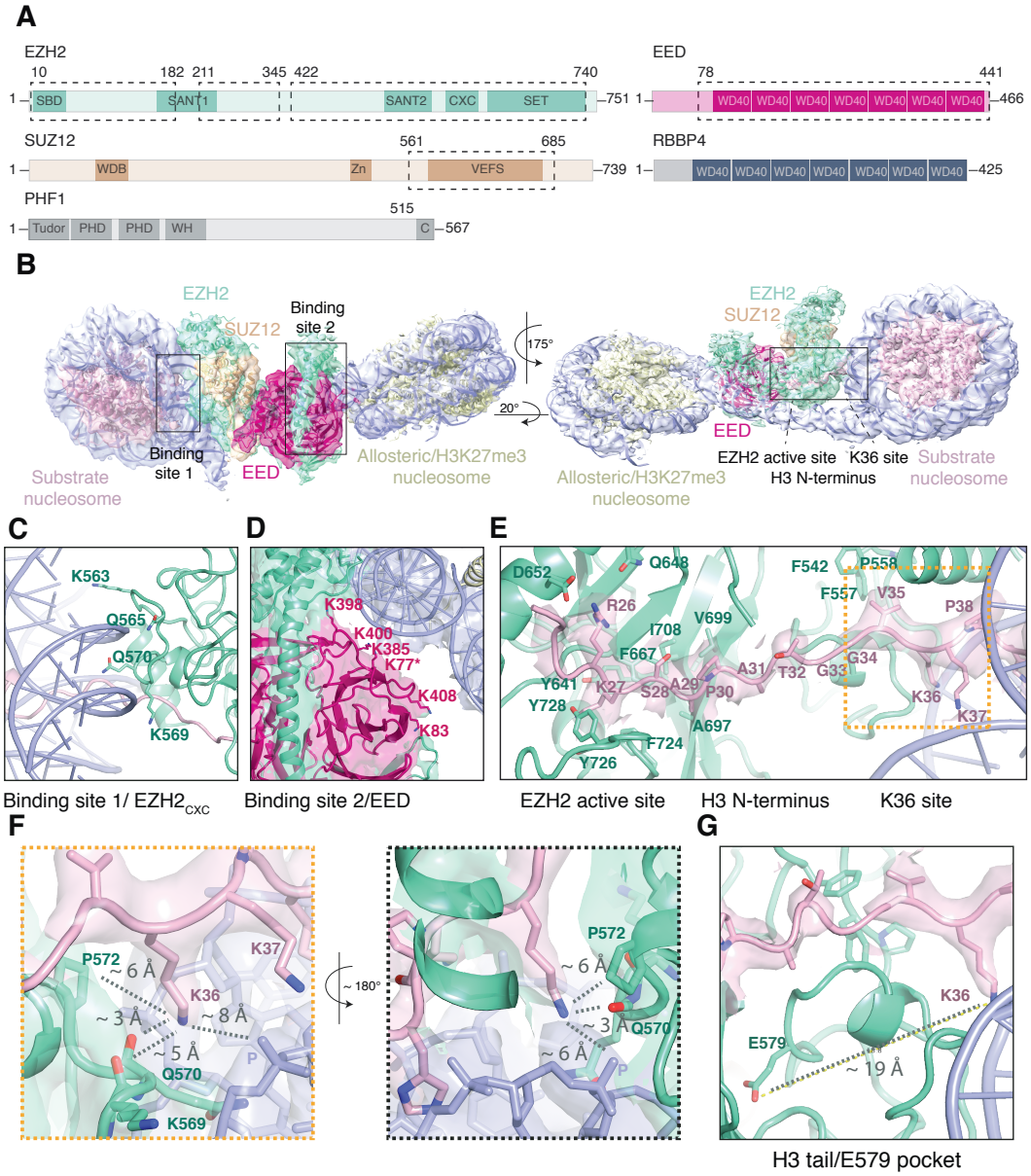

Figure 4.8: Cryo-EM structure of PRC2-PHF1 reveals how the H3 tail is recognized by PRC2.

A. Schematic representation of the PRC2-PHF1 domain organisation. Dashed boxes indicate regions of the individual subunits modelled in the cryo-EM reconstruction of PRC2-PHF1:Dinuc. B. The overall PRC2-PHF1-dinucleosome cryo EM structure confirms the previously described two nucleosomes-binding sites of PRC2: EZH2^{CXC} on the substrate nucleosome and EED/EZH2^{SBD} on the allosteric nucleosome (Poepsel et al., 2018). C. Details of the EZH2^{CXC} residues interaction with nucleosomal DNA: shown residues were mutated for biochemical studies described below. D. Details of the potential EED residues interactions with nucleosomal DNA: shown residues were mutated for biochemical studies described below. E. In addition to previously described interactions the H3 tail is recognized by EZH2 likely by two hydrophobic hotspots (Justin et al., 2016). F. H3K36 is accomodated in the CXC-DNA interaction surface facing the DNA with its side chain. Analysis of the chemical environment of H3K36 allows to approximate distances to possible interactors. G. The E579 pocket proposed by Jani et al., 2019, to bind H3K36 is located approximately 19 Å away from H3K36 in the EZH2$_{sub}$-Nuc$_{sub}$. Major rearrangements would be needed to allow for this interaction, making the proposed mechanism of the E579 pocket incompatible with the observed model.

A tubular density not explained by the crystal structure (Justin et al., 2016) or the cryo-EM structure (Kasinath et al., 2018) of the sole PRC2 complex becomes visible in the focused map of $EZH2_{sub}$-Nuc_{sub} (Fig. 4.9 A). The resolution did not allow for residue assignment, however based on a very recent study by Kasinath et al., 2020 and its shape, this density likely accommodates EZH2 residues 497-511 forming an α-helix. As in Kasinath et al., 2020, this helix shall be referred to as "bridge helix". The 'bridge helix' is located above H3V35 and K36 based on the density-modified map (Terwilliger et al., 2020) of $EZH2_{sub}$-Nuc_{sub} at lower threshold (not shown) and is likely engaged in interactions with the H3 tail, EZH2 and the nucleosomal DNA (Fig. 4.9 A, B). Residues important for regulation of PRC2 activity by automethylation (K510, K514, K515) are located in the loop leading to or in the 'bridge helix' (Wang et al., 2019; Lee et al., 2019).

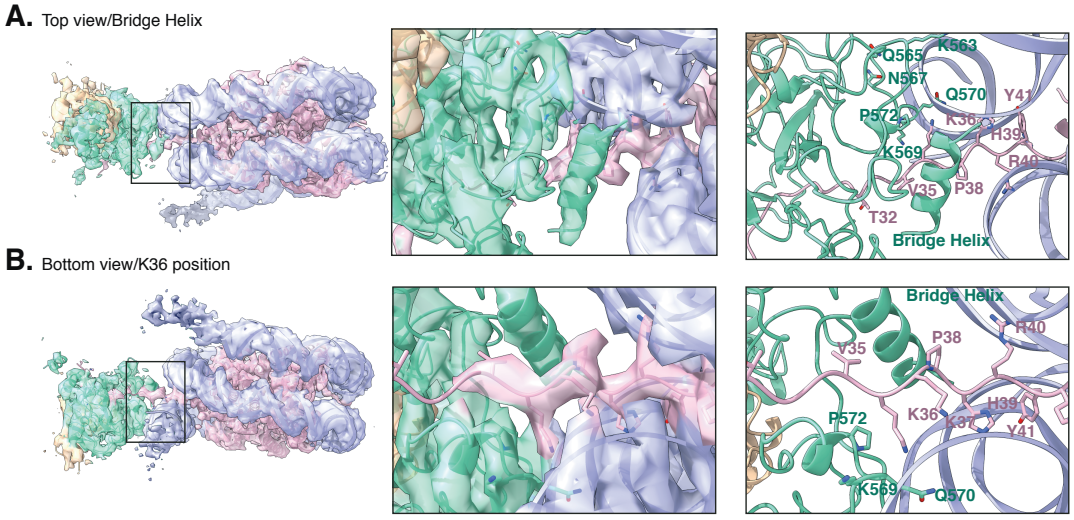

Figure 4.9: The EZH2 'bridge helix' is located in close proximity to the H3 tail.

A. A tubular density located between EZH2 and the substrate nucleosome is visible in the focused map of $EZH2_{sub}$-Nuc_{sub} and based on the recent study by Kasinath et al., 2020, likely constitutes an α-helix, named "bridge helix" Kasinath et al., 2020, with EZH2 residues 497-511. B. The 'bridge helix' is located above H3V35 and K36.

Finally, when comparing the overall map and model of PRC2-PHF1:Dinuc with the previous published map and model of PRC2-AEBP2:Dinuc (Poepsel et al., 2018) (Fig. 4.10 A), the overal binding mode of core PRC2 to the nucleosomes appears to be similar for both cofactors. While almost no differences can be seen in the recognition of the substrate nucleosome by EZH2, slight changes in binding can be observed when comparing the recognition of the allosteric nucleosome by the EED subunit (Fig. 4.10 A, B).

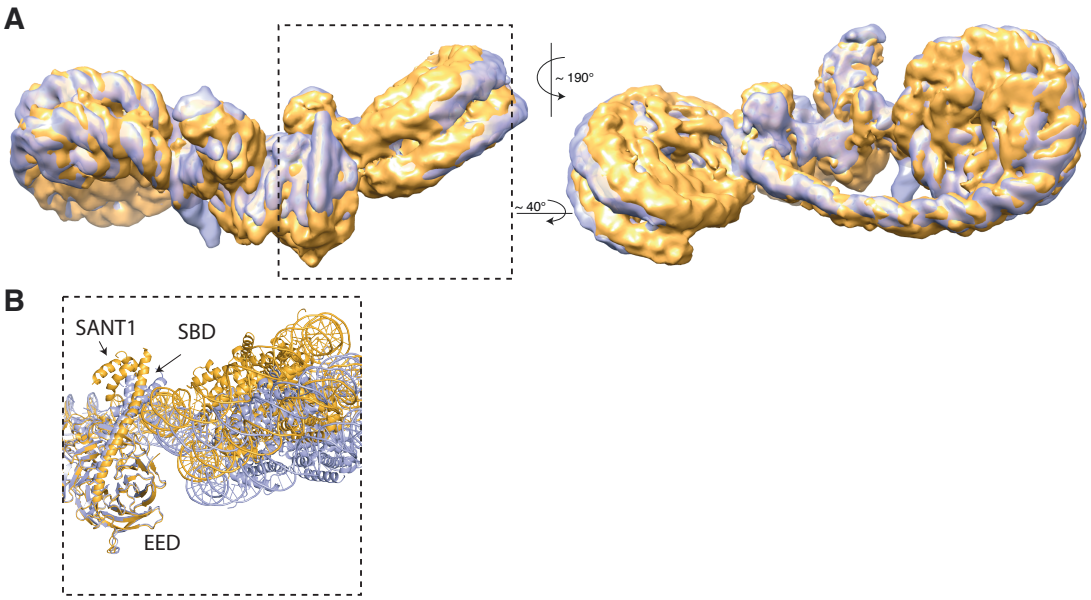

Figure 4.10: PRC2-PHF1 and PRC2-AEBP2 show similar interactions with the dinucleosome.

A. Comparison of Cryo-EM maps of PRC2-PHF1:Dinuc (blue) and PRC2-AEBP2:Dinuc (orange) (Poepsel et al., 2018) reveal subtle differences in the recognition of the allosteric nucleosome. In contrast, recognition of the substrate nucleosome is highly comparable. B. Superposition of the PRC2-PHF1/AEBP2:Dinuc models at the allosteric nucleosome. In the study of Poepsel et al., 2018, the allosteric nucleosome appears to be closer to the $EZH2^{SBD}$-helix, while for PRC2-PHF1:Dinuc model is is located closer to the EED surface.

In summary, the model of $EZH2_{sub}$-Nuc_{sub} shows how chromatin binding by PRC2 is established by the DNA binding contributions of EED/$EZH2^{SBD}$ to the allosteric nucleosome and $EZH2^{CXC}$ on the substrate nucleosome. The described EZH2-CXC- nucleosomal DNA interface creates the platform for the H3 tail recognition by EZH2 and K36 is sandwiched in this interface.

4.2 Biochemical studies of PRC2-nucleosome interaction

4.2.1 EZH2-CXC contribution to nucleosomal DNA binding and activity of PRC2

To investigate the contribution of CXC residues identified in the structural studies above (Fig. 4.8 C), mutations of the CXC residues to alanine or glutamate were introduced, resulting in K563A/E, Q565A/E, K569A/E and Q570A/E. To exclude the DNA binding contribution of PHF1, a complex containing only the C-terminal 5 kDa region of PHF1 (aa 515-567) previously shown to be sufficient for PRC2 interaction but lacking the tudor and the DNA-binding winged-helix domains was used (Choi et al., 2017). For simplicity PRC2 complexes with the C-terminal construct of PHF1 shall be referred to as PRC2. PRC2 complexes containing EZH2-CXC mutations to alanines or glutamates are indicated with $PRC2^{CXC>A}$ or $PRC2^{CXC>E}$, respectively. The effect of these mutations on the binding affinity for substrate mononucleosomes was monitored by electrophoretic mobility shift assays (EMSA) (Fig. 4.11 A). To this end, increasing concentrations ranging between 17 nM to 3500 nM of either PRC2, $PRC2^{CXC>A}$ or $PRC2^{CXC>E}$ were incubated with 45 nM of 215 bp nucleosomes. Binding was monitored on an agarose gel (more details in Methods section 3.5.1). The lower bands (Fig. 4.11 A; labeled with "*X.l.* nucs") represent free mononucleosomes. Upon addition of PRC2, PRC2-nucleosome complex is formed as indicated by a shift in migration (higher bands in Fig. 4.11 A; labeled with "PRC2 with *X.l.* nucs"). Both PRC2 mutants shifted at higher concentrations as compared to wild-type (wt) PRC2 (Fig. 4.11 A, compare lanes 12-20, 22-30 to 2-10). In order to quantify the effect of CXC mutations on binding, the assay was performed in triplicates and the fluorescence signal of the bands was subjected to densitometry analysis (described in more detail in the Methods section 3.5.1). The resulting binding curves clearly showed that mutations of the CXC residues to alanine or glutamate reduced the binding of PRC2 2.8 and 3x fold, respectively. The apparent K_d of PRC2 is app. 49 ± 4 nM, of $PRC2^{CXC>A}$ is app. 137 ± 15 nM and of $PRC2^{CXC>E}$ is app. 199 ± 22 nM (Fig. 4.11 B). Intriguingly, unlike binding, which only showed a minor effect after the mutation, the activity (mono- and trimethylation) of the $PRC2^{CXC>A}$ mutant was drastically reduced as compared to wt on mononucleosomes (Fig. 4.11 C, compare lanes 5-7 to 2-4). To test whether the activity can be partially rescued by providing a second binding site for the $EED/EZH2^{SBD}$ interface, unmodified dinucleosomes with 35 bp DNA linker were used. Indeed, $PRC2^{CXC>A}$ was able to mono- and trimethylate dinucleosomes, albeit less efficiently as compared to PRC2 wt (Fig. 4.11 C, compare lanes 12-14 to 9-11; Coomassie stained SDS control gel of the input in D). It should be noted, that the catalytic activity of PRC2 on dinucleosomes involves the allosteric activation of EZH2 in the complex once H3K27me3 is successfully deposited on one of the dinucleosomes (Margueron et al., 2009; Jiao and Liu, 2015).

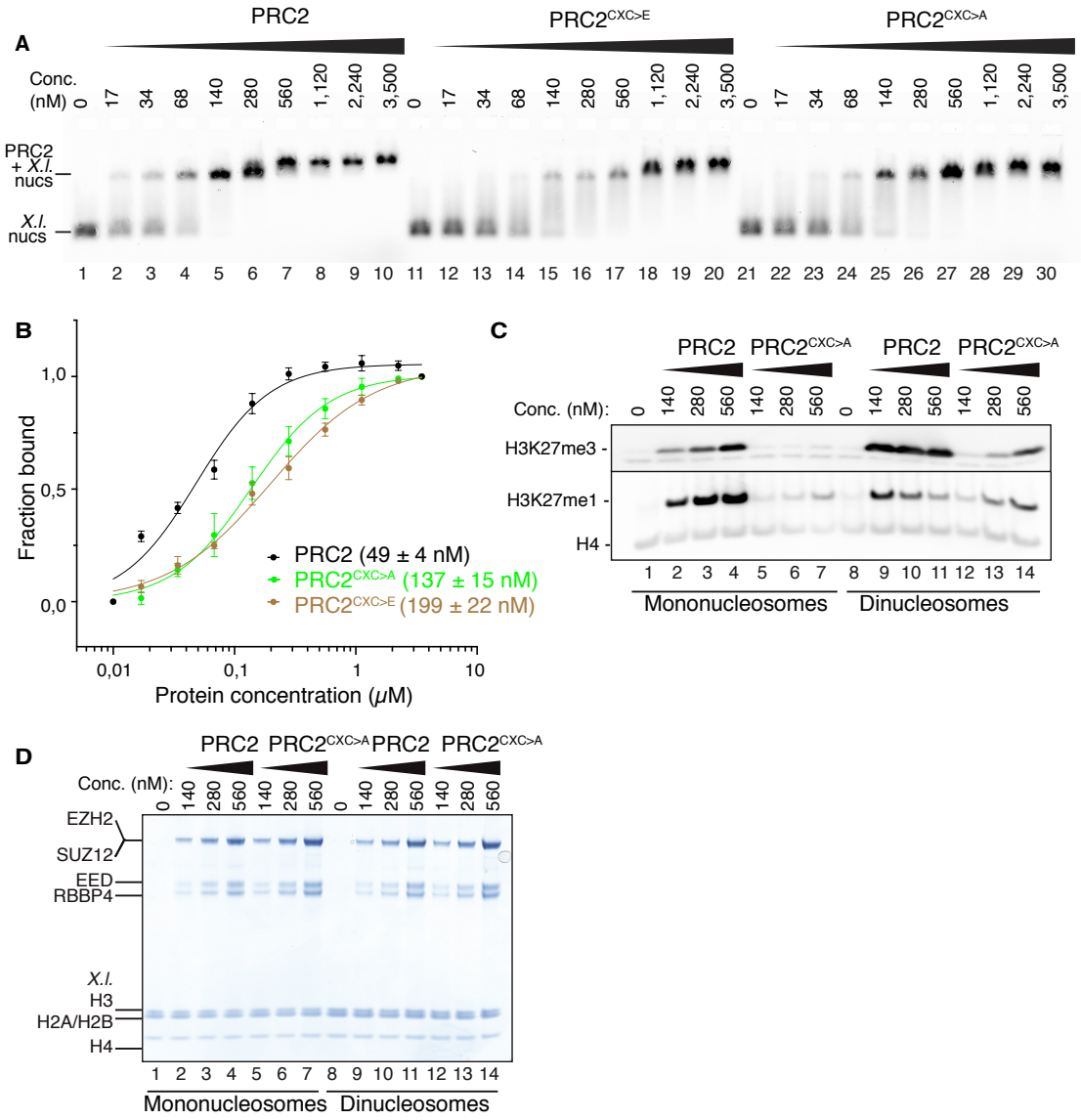

Figure 4.11: Biochemical characterization of the EZH2-CXC nucleosome binding.

A. A representative EMSA analysis of PRC2 and PRC2$^{\mathrm{CXC>A}}$ or PRC2$^{\mathrm{CXC>E}}$ binding to substrate *X.l.* mononuclesomes containing 215 bp nucleosomal DNA. Increasing concentrations of the respective PRC2 complexes were incubated with 45 nM of mononucleosomes and binding was monitored on an 1.2 % agarose gel. Lower bands show unbound *X.l.* nucleosomes, while the shifted band shows nucleosomes in complex with PRC2 B. EMSA with PRC2 and PRC2$^{\mathrm{CXC>A}}$ or PRC2$^{\mathrm{CXC>E}}$ was performed in triplicates and quantified by densitometry analysis of the 6-Carboxyfluorescein signal. A Hill function was fitted to estimate the apparent K_{d} value of the respective complexes, which is indicated in the brackets. Error bars represent SEM (standard error of the mean) of the fitted values. C. Activity of wt PRC2 was compared with PRC2$^{\mathrm{CXC>A}}$ in a HTMase assay on *X.l.* mono- or di-nucleosomes (with 35 bp linker) and analyzed by Western Blot (WB) using H3K27me1/me3 antibodies. D. Input for the WB was analyzed on a Coomassie stained SDS gel to control for equal input of the components. While HMTase assays and other WBs confirming the result of PRC2$^{\mathrm{CXC>A}}$ were performed by the author of the thesis, the shown WB as well as the corresponding SDS gel were performed by K. Schmid.

To exclude the possibility that PRC2$^{CXC>A}$ catalytic activity is impaired due to improperly folded catalytic domain as a result of the CXC mutations, the activity was tested on $H3_{18\text{-}42}$ peptides (4.12). Mass spectrometry (MS)-based peptide HMTases were performed in technical triplicates and amount of monomethylation was calculated as a portion of the sum of unmethylated and methylated peptides (described in more detail in 3.5.3). Both wt and mutant PRC2 complexes showed equal monomethylation levels (Fig. 4.12 A, B). Hence, loss of activity, which was observed for the CXC mutant in HMTase assays on mononucleosomes (Fig. 4.11 C) is a consequence of DNA binding inability by the mutant but not of incorrect SET domain folding.

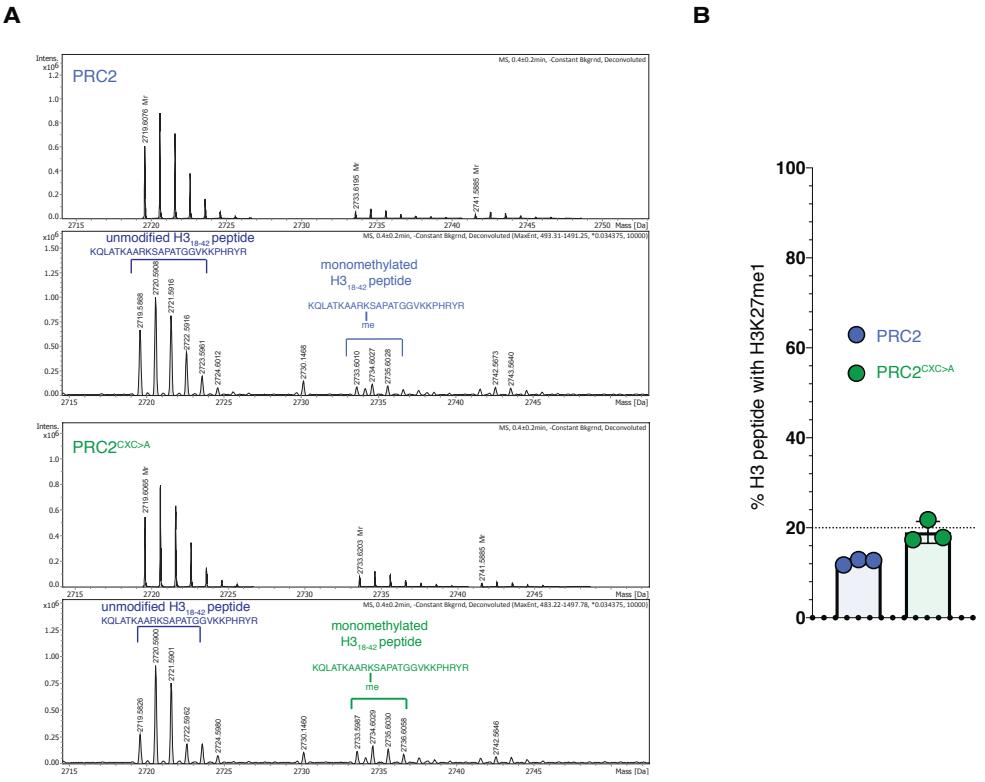

Figure 4.12: H3-peptide methylation assay of PRC2/PRC2$^{CXC>A}$.

A. A mass spectrometry based peptide methylation assay was performed to exclude the possibility of misfolding of PRC2$^{CXC>A}$. For this, HMTase assays with PRC2 wt and PRC2$^{CXC>A}$ were performed on $H3_{18\text{-}42}$ in technical replicates and resulting methylation products were monitored via MS. Deconvoluted spetra are shown. The percentage of monomethylation was calculated and compared in B. Error bars show the standard deviation (SD). MS runs were performed by E. Weyher.

4.2.2 Contribution of EED to nucleosomal DNA binding by PRC2

Structural analysis of PRC2 bound to dinucleosomes as described in Section 4.1.3 revealed a second nucleosomal DNA binding site of PRC2 via the EED subunit (Fig. 4.8 D), to

which the remaining binding of $PRC2^{CXC>A}$ or $PRC2^{CXC>E}$ could be, partially, attributed. Hence, binding contribution of EED to nucleosomes was analyzed by performing the same assays as described for $EZH2^{CXC}$. Several potential binding residues located on the surface of EED and facing the nucleosome were identified based on this study and the results reported by Poepsel et al., 2018, (K77, K83, K385, K398, K400, K408; see also Fig. 4.8 D) and mutated to alanines (referred to as $PRC2^{EED>A}$). EMSA was performed with this as well as the combined $PRC2^{CXC>A/EED>A}$ mutant. Binding of $PRC2^{EED>A}$ was only mildly reduced (2.4x) likely due to binding contributed by $EZH2^{CXC}$ and $EZH2^{SBD}$ (Fig. 4.13 A and B). Additionally, the bottom lobe of PRC2 constituted by SUZ12 and RBBP4 was previously shown to be sufficient for binding to nucleosomes (Nekrasov et al., 2005). Moreover, negative stain studies showed that also within the full PRC2 complex the 'bottom lobe' makes contacts to the dinucleosomes (Poepsel et al., 2018). These studies also explain the observation that the combined mutant $PRC2^{CXC>A/EED>A}$ showed remaining binding, albeit 5-fold weaker than in wt (Fig. 4.13 A and B). Finally, the remaining binding could also be attributed to additional residues and regions in PRC2 not mutated, such as $EZH2^{SBD/SANT1}$, which are in close proximity to the allosteric nucleosome in the PRC2-PHF1:Dinuc model (as analysed in 4.1.3) and were also proposed to potentially bind by Poepsel et al., 2018.

As expected, $PRC2^{EED>A}$ exhibited almost the same level of activity as wt PRC2 on mononucleosomes in HMTase assays, since PRC2 is still able to bind via $EZH2^{CXC}$ (Fig. 4.13 C, loading control in D; compare lanes 5-7 to 2-4). However, on dinucleosomes, the $PRC2^{EED>A}$ mutant failed to generate comparable levels of H3K27me3 methylation to PRC2 wt (Fig. 4.13 C, loading controls in D; compare lanes 12-14 to 9-11). The failure of proper binding by the mutant on the second nucleosome likely leads to the inability to recognize the H3K27me3 modified H3 tail by EED and allosterically activate EZH2 (Margueron et al., 2009; Jiao and Liu, 2015). This shows that the mutated residues contribute to the binding of EED to the neighboring nucleosome and that proper DNA binding is needed for an allosteric activation mechanism by the EED subunit (Margueron et al., 2009). Finally, combination of both mutations (EED and CXC) seems to completely abrogate the trimethylation activity of the complex on dinucleosomes as can be seen in Fig. 4.13 E.

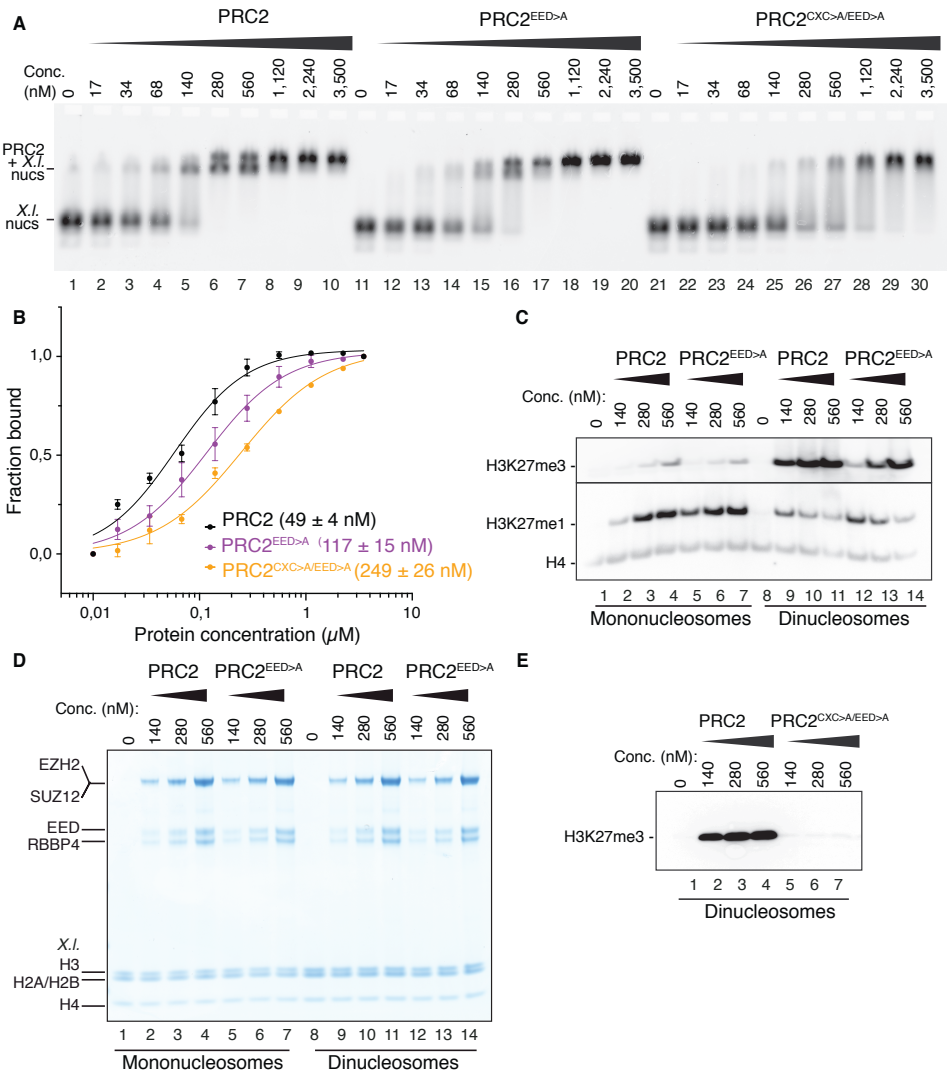

Figure 4.13: Biochemical characterization of the EED nucleosome binding.

A. A representative EMSA analysis of PRC2 and PRC2$^{\mathrm{EED>A}}$ or PRC2$^{\mathrm{CXC>A/EED>A}}$ binding to substrate mononucleosomes containing 215 bp nucleosomal DNA. Increasing concentrations of the respective PRC2 complexes were mixed with 45 nM of mononucleosomes and binding was tested on a 1.2 % agarose assay. B. EMSA with PRC2 and PRC2$^{\mathrm{EED>A}}$ or PRC2$^{\mathrm{CXC>A/EED>A}}$ was performed in triplicates and quantified by densitometry analysis of the 6-Carboxyfluorescein signal. A Hill function was fitted to estimate the apparent K_{d} values of the respective complexes (indicated in the brackets). Error bars represent SEM of the fitted values. C. Activity of wt PRC2 was compared with PRC2$^{\mathrm{EED>A}}$ in a HTMase assay and analyzed by Western Blot using H3K27me1/me3 antibodies. H4 serves as control for loading and western blot processing. D. Activity on dinucleosomes of wt PRC2 was compared to the mutant combination PRC2$^{\mathrm{CXC>A/EED>A}}$ in an HTMase assay and analyzed by Western Blot using H3K27me3 antibody. This preliminary result would need to be repeated again including H4 control to be fully confirmed. While HMTase assays and other WBs confirming the result of PRC2$^{\mathrm{EED>A}}$ were performed by the author of the thesis, the shown WBs for PRC2$^{\mathrm{EED>A}}$ and PRC2$^{\mathrm{CXC>A/EED>A}}$ as well as the corresponding SDS gel were performed by K. Schmid.

4.2.3 Lysine 36 - a key residue for PRC2 regulation

Structural characterization showed that while there is no clear direct recognition of H3K36 by PRC2, K36 instead is facing the nucleosomal DNA and sits in a tight sandwiched position between $EZH2^{CXC}$ and DNA interface potentially interacting with the carbonyl group of Q570 (Fig. 4.14 A). This observation suggests a possible inhibition mechanism where a di- or tri-methylated K36 becomes too bulky and its positive charge shielded by and distributed into the additional methyl groups. The importance of the side chain properties of a lysine in this position was hence tested by mutating K36 to either an arginine or an alanine and assaying the activity of PRC2 by HMTase assay and WB. As compared to activity on unmodified nucleosomes, trimethylation activity was strongly diminished on H3Kc36me3 (Simon et al., 2007), as previously shown (Schmitges et al., 2011), but also on H3K36A/R mutated nucleosomes (Fig. 4.14 B, control loaded C; compare lanes 5-7, 8-10, 11-13 to 2-4). Monomethylation was also reduced but in the case of H3K36A/R less significantly reduced as compared to H3Kc36me3 nucleosomes (Fig. 4.14 B). This result is in agreement with studies by Jani et al., 2019, but disagrees with Schmitges et al., 2011, where H3K36A did not inhibit PRC2.

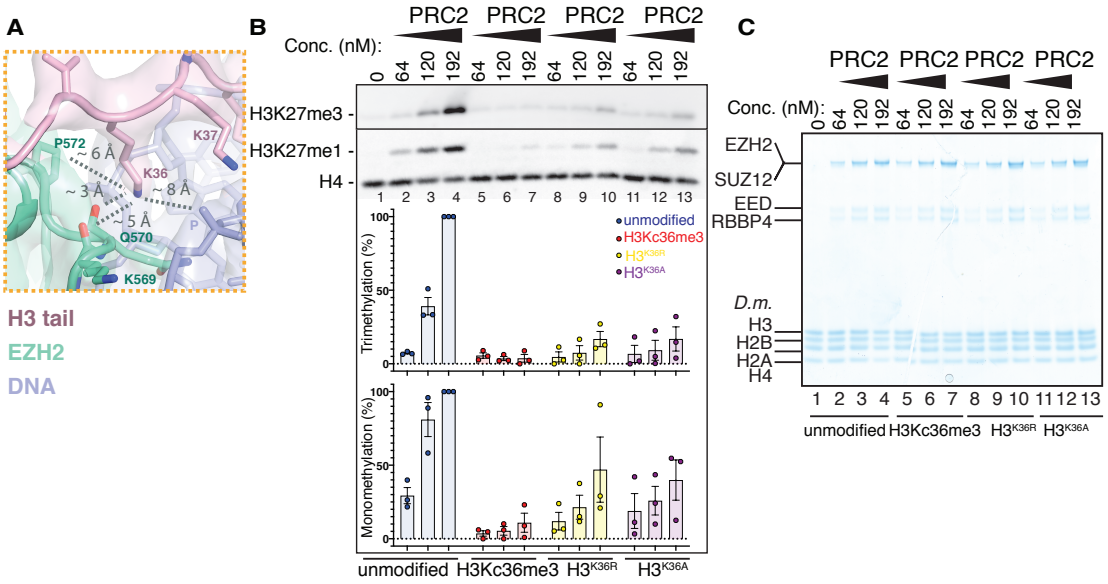

Figure 4.14: Biochemical characterization of H3K36 side chain properties in the context of the EZH2-DNA interface.

A. K36 is sandwiched in an interface formed by the nuclesomal DNA and CXC of EZH2. B. Activity of PRC2 on unmodified mononucleosomes was compared to activity on H3Kc36me3 modified (Simon et al., 2007) or H3K36A/R nucleosome in an HTMase assay and analyzed by WB using H3K27me1/me3 antibodies. The assay and the corresponding WB were performed in triplicates and subjected to densitometry analysis of the chemiluminescence signal. The amount of methylation is approximated in respect to the highest methylation activity observed at 192 nM of PRC2 on unmodified nucleosomes. i.e. lane 3, set to 100%. C. The input for the WB in B was additionally analyzed on a Coomassie-stained SDS gel to ensure equal input of the individual components in the HMTase assay. SDS-Gel and WBs of the triplicates were partially performed with the help of S. Mitzkus.

Intriguingly, despite the shown results suggesting an important role for the physicochemical nature of a lysine side chain at position 36, binding of PRC2 to H3Kc36me3 nucleosomes (Simon et al., 2007) was not affected by the mark as shown in Fig. 4.15 A (compare lanes 12-20 to 2-10) and analyzed in Fig. 4.15 B. To rule out an only minor effect, which potentially would be masked by the many interactions of PRC2 with nucleosomes, $PRC2^{EED>A}$ was also tested. Eliminating the binding contribution made by EED did not change the binding affinity of the complex on H3Kc36me3 nucleosomes as compared to unmodified (Fig. 4.15 C, D). Hence, H3Kc36me3 does not affect binding of PRC2 on nucleosomes.

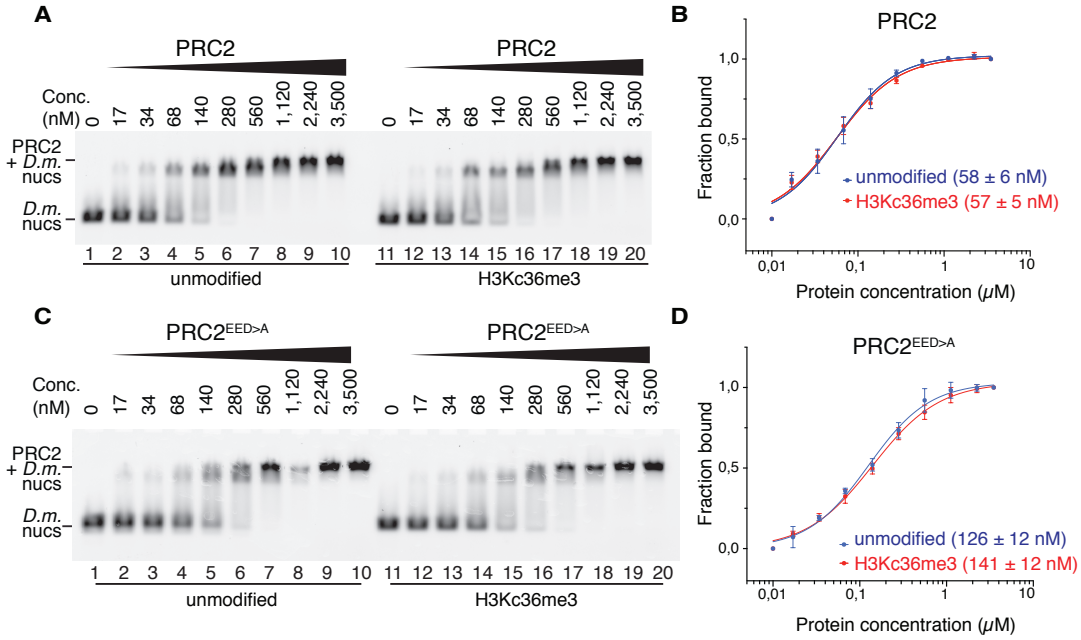

Figure 4.15: H3K36me3 does not affect PRC2 binding to the nucleosome.

A. EMSA was performed with wt PRC2 on either unmodified or H3Kc36me3 modified (Simon et al., 2007) nucleosomes in triplicates and quantified by densitometry analysis of the 6-Carboxyfluorescein signal in B. A Hill function was fitted to estimate the apparent K_d value of the respective complexes (indicated in brackets). Error bars represent SEM of the fitted values. C. EMSA was performed with PRC2$^{EED>A}$ on either unmodified or H3Kc36me3 modified nucleosomes as in A. D. Quantification was performed as in B.

4.2.4 H3K36me3 inhibits PRC2 only in the context of the nucleosome

The cryo EM analysis described in subsection 4.1.3 as well as the biochemical characterization of the H3K36 position both suggested that the inhibition of PRC2 by H3K36me3 takes places in an indirect manner involving a steric hindrance and a charge shielding effect in context of the H3 tail being sandwiched in between the CXC and the H3 tail. To confirm the indirect mechanism being a result of the nucleosomal context, HMTase assays were performed with PRC2 on unmodified and H3K36me3 containing H3$_{(18\text{-}42)}$-peptides and subsequently analyzed by MS as shown in Fig. 4.16. The assay and the MS runs were performed in triplicates and subjected for a quantification analysis as described in the methods subsection 3.5.3 (Fig. 4.16 B). Average % of monomethylation was calculated. Indeed, PRC2 showed equal monomethylation activity on both unmodified and H3K36me3 peptides.

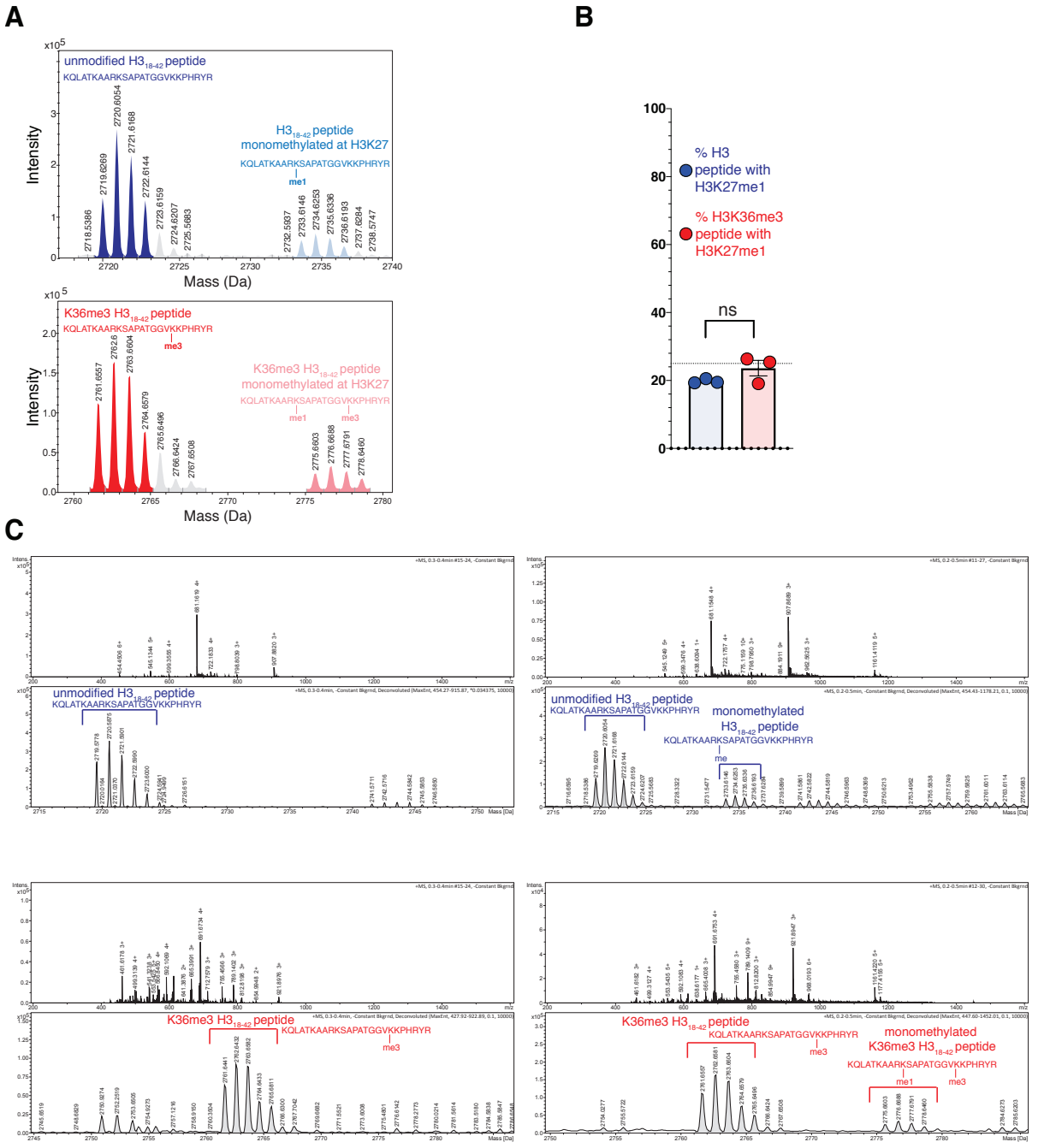

Figure 4.16: PRC2 is only inhibited by the H3K36me3 mark in context of the nucleosome but not by a H3K36me3 peptide.

A. Deconvoluted MS Spectra after HMTase assay on unmodified (blue) or K36me3 modified (red) H3$_{(18\text{-}42)}$ peptide. B. HMTase assays using the respective peptides as well as the corresponding MS runs were performed in technical triplicates. Statistical analysis (Welch's t-test) revealed a non-significant (ns) difference between the app. amount of monomehylation on a H3K36me3 modified peptide as compared to an unmodified peptide. Error bars show standard deviation (SD). C. Raw (upper panel) and deconvoluted peaks (lower panel) from input peptides only (without PRC2) (left panels) and after HMTase assay with PRC2 (right panels). MS was performed by E. Weyher (MS facility, MPI of Biochemistry).

4.2.5 Role of PHF1 cofactor in PRC2-nucleosome interactions and PRC2 inhibition by H3K36me3

The described analysis of PRC2$^{EZH2>E/A}$ mutants in subsection 4.2.1 provoked the question of how these mutations would affect nucleosome binding by PRC2-PHF1$^{EZH2>E/A}$. To test this, EMSAs were performed with increasing amounts of wt PRC2-PHF1 or PRC2-PHF1$^{EZH2>E/A}$ on *X.l.* 215 bp mononucleosomes as described above (Fig. 4.17 A). Counterintuitively, mutation of the CXC residues to alanines or glutamates did not weaken the binding when PHF1 was present in the complex (Fig. 4.17 A, B). Densitometric analysis of the EMSA revealed an apparent K_d of app. 45 nM for both, PRC2-PHF1 and PRC2-PHF1$^{EZH2>E/A}$, similar to the apparent K_d (app. 50 nM) of the PRC2 complex containing only the C-terminal bit of PHF1. The only insignificant increase in binding of PRC2-PHF1 compared to PRC2 is in contradiction to previously published results, where PRC2-PHF1 showed an 1.5-fold increase in binding affinity as compared to PRC2 (Choi et al., 2017). In this study, PRC2 containing only the C-terminal bit of PHF1 (in the thesis referred to as PRC2) showed an apparent K_d of app. 211 nM for nucleosomes, while PRC2-PHF1 showed an apparent K_d of app. 66 nM (Choi et al., 2017). For DNA, the respective K_ds were app. 30 nM for PRC2-PHF1 and app. 49 nM for PRC2 containing the C-terminal bit of PHF1 (Choi et al., 2017). It should be noted that the K_d can only be approximated and highly depends on the amount of nucleosomes and the accuracy of nucleosome concentration determination, hence annotated as 'apparent' K_d. Furthermore, the apparent K_ds are close to the concentration of nucleosomes used, so that the assay is approaching stoichiometric 'titration' behavior. Hence, in the context of this assay minor differences below 2x should be considered carefully. Together with the observation that the nucleosome bands in Fig. 4.17 A start to shift already at very low concentrations of PRC2-PHF1, the results suggest that the assay has to be repeated with lower amounts of PRC2-PHF1 and potentially at lower concentration of mononucleosomes. Hence, the effect of the CXC>E/A mutations might not be fully disclosed in the given experimental context and to fully validate the role of PHF1 this assay needs to be revised. Should the result be confirmed at lower concentrations of PRC2-PHF1, it could point out to a function, where PHF1 in the case of CXC binding failure serves as an additional rescue binding anchor for PRC2 to the nucleosomes. In this context, it would be interesting to analyze whether PRC2-PHF1$^{EZH2>E/A}$ shows an increase of mono- and trimethylation as compared to PRC2$^{EZH2>E/A}$ shown in Fig. 4.11.

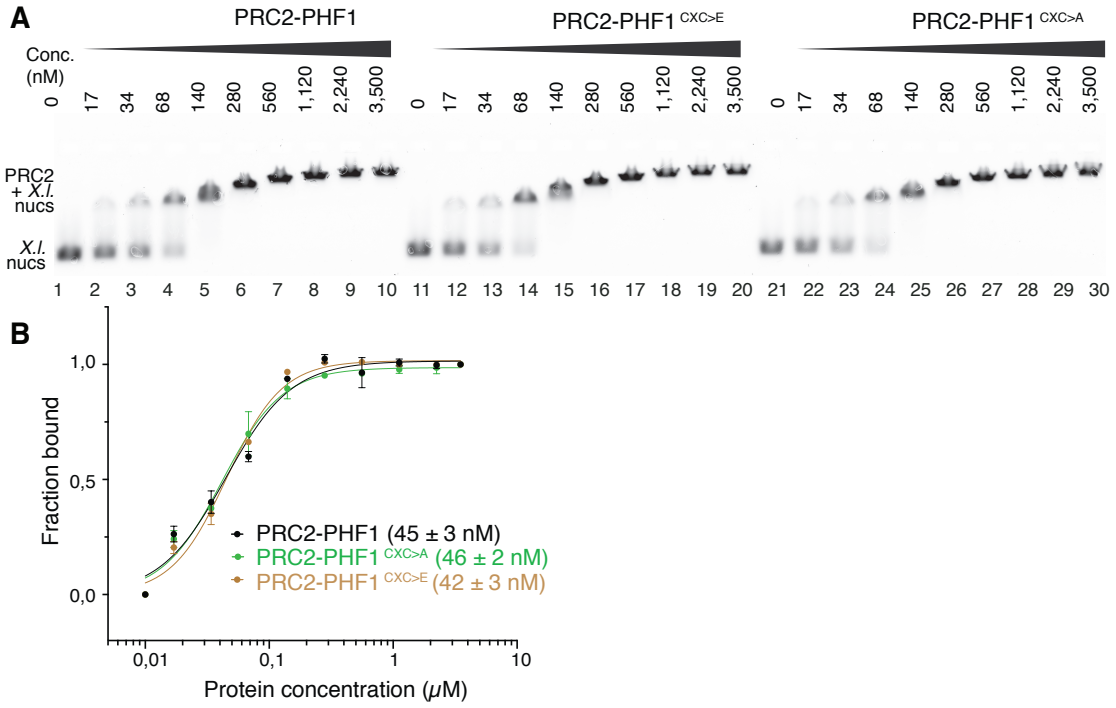

Figure 4.17: Role of PHF1 on PRC2-nucleosome interactions.

A. EMSAs were performed in triplicates with wt PRC2-PHF1 and PRC2-PHF1$^{\mathrm{EZH2>E/A}}$ on unmodified *X.l.* 215 bp mononucleosomes and quantified by densitometric analysis of the 6-carboxyfluorescein signals in B. B. Error bars represent SEM of the fitted values.

In the next step, the effect of H3K36 trimethylation or K36A/R mutations on activity of PRC2 was analyzed in presence of PHF1. The PHF1 Tudor domain was previously reported to bind the H3K36me3 modification and this interaction inhibits PRC2 based on studies conducted on H3K36me3 containing chromatin isolated from yeast (Musselman et al., 2012, 2013). As for PRC2, binding of PRC2-PHF1 to H3Kc36me3 modified nucleosomes (Simon et al., 2007) was not affected in the presence of the active mark (Fig. 4.18 A, B). The EMSA should however be considered carefully, as the assay was only performed once and as discussed above, lower concentration of PRC2-PHF1 should be used to exclude an only minor effect. To study the effect of the presence of PHF1 on the inhibition of PRC2 by H3K36me3 in a completely recombinant *in vitro* system, HMTase assay was performed with increasing amounts of PRC2-PHF1 on H3Kc36me3 or H3K36R/A nucleosomes and the reaction was blotted and visualized with H3K27me1/3 antibodies as described before. H4 antibody was used as a control. Similar to PRC2, trimethylation and monomethylation activity of PRC2-PHF1 were almost completely abrogated on H3Kc36me3 nucleosomes (Fig. 4.18 C). *In vivo*, high levels of H3K27me3 require the presence of PHF1 (Cao et al.,

2008; Sarma et al., 2008) or the *Drosophila* ortholog Pcl (Nekrasov et al., 2007) and the study of Choi et al., 2017, demonstrated that PHF1 increases the residence time of the complex on chromatin. Due to the longer residence time PHF1-PRC2 boosts the conversion of mono- or dimethylation to trimethylation. It is therefore not surprising that wt PRC2-PHF1 shows only little monomethylation in the WB as it is to a high degree converted to trimethylation (Fig. 4.18 C). Trimethylation of PRC2-PHF1 is also inhibited on H3K36A/R mononucleosomes, albeit weaker as compared to PRC2 (Fig. 4.18 C; compare lanes 8-10 and 11-13 to 2-4 and compare to Fig. 4.14). Given that the presence of PHF1 prolongs the residence time of and boosts trimethylation by PRC2, it seems that on H3K36A/R due to the longer residence time EZH2 is able to generate more trimethylation and hence shows lower inhibition. Intriguingly, monomethylation of PRC2-PHF1 on H3K36A/R mononucleosomes is highly increased as compared to unmodified nucleosomes. The inhibition mechanism by H3K36A/R, hence seems to involve the residence time of PRC2. The increased residence time of PRC2-PHF1 in context of H3K36A/R nucleosomes promotes monomethylation as compared to PRC2, however fails to promote trimethylation. In contrast both mono- and trimethylation of PRC2-PHF1 are reduced on H3Kc36me3 modified nucleosomes. The evident inhibition difference between H3K36A/R nucleosomes and H3Kc36me3 nucleosomes is only observed for PRC2-PHF1 but not for PRC2. A potential explanation for this observation is that H3K36me3 binding by PHF1 outcompetes H3 tail binding by EZH2, thereby significantly reducing not only the residence time but also possible interaction needed for catalysis. In contrast, *D.m.* Pcl tudor domain has an incomplete aromatic cage (Friberg et al., 2010) and is hence not able to recognize the modification. It would be therefore interesting to assay the difference of inhibition by H3K36me3 when comparing PRC2-PHF1 and PRC2-Pcl.

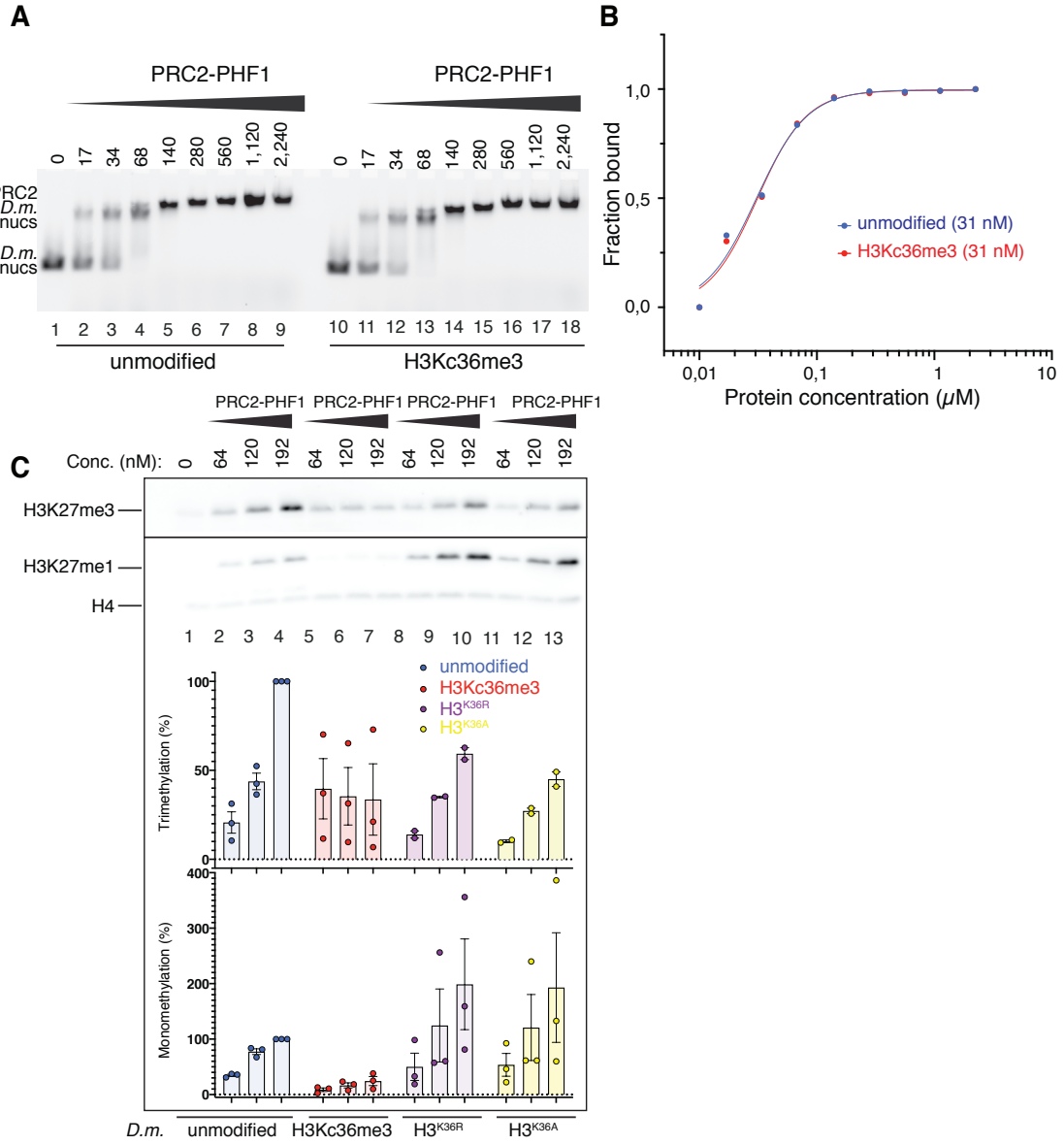

Figure 4.18: Role of PHF1 for PRC2 inhibition by H3K36me3.

A. EMSA was performed with PRC2-PHF1 on either unmodified or H3Kc36me3 modified nucleosomes. The densitometric quantification of the 6-carboxyfluorescein is shown in B. This bandshift assay was performed once. C. HMTase assays of PRC2-PHF1 on different substrate mononucleosomes (unmodified/H3Kc36me3/H3K36R or H3K36A nucleosomes). For unmodified/H3Kc36me3 the assay was performed in triplicates. For H3K36R or H3K36A the assay was performed in duplicates. Error bars indicate SEM. HMTase assay was performed by the author of the thesis, while all WBs and WB imaging for this experiment were performed by S. Schkoelziger.

4.2.6 The acidic patch of the nucleosome does not contribute to binding by or activity of PRC2

The cryo-EM structure of PRC2-PHF1 on dinucleosomes described in subsection 4.1.3 didn't show any interaction between PRC2 and the acidic patch of the nucleosome (Luger et al., 1997; McGinty and Tan, 2016). However as the 'bottom lobe' of PRC2 including RBBP4 or N-terminal bit of SUZ12 is not resolved in the reconstruction, the possibility couldn't be excluded that it interacts with the nucleosome core. To confirm that the acidic patch does not contribute to binding or activity of PRC2, EMSA and HMTase activity assays were performed with increasing amounts of wt PRC2-PHF1 on wt *X.l.* nucleosomes or nucleosomes carrying the $H2A^{E61A,\ E64A,\ D90A,\ E92A}$ mutations (Luger et al., 1997). The EMSA experiment showed the respective nucleosome bands shifting at equal concentrations of PRC2 indicating that PRC2-PHF1 can equally well bind wt or mutated nucleosomes (Fig. 4.19 A, B). However it should be noted that the amount of $H2A^{E61A,\ E64A,\ D90A,\ E92A}$ used in the EMSA was reduced as compared to wt nucleosomes as can be seen by the weaker fluorescence signal (lane 11 in Fig. 4.19 A). This is potentially due to imprecise determination of the nucleosome concentration, which is generally impeded by the presence of nucleosomal DNA; for all other bandshifts in this thesis the concentration of nucleosomes was determined by measuring the concentration of DNA in the nucleosome sample. Nucleosome concentration in the shown agarose gel was determined by back-calculating the amount of octamers used. In the case where nucleosome preparation did not contain free DNA (largely depending on the quality of the octamer:DNA titrations) the method of nucleic acid concentration turned out to be more reliable as a method for nucleosome concentration determination.
In addition, mutations tested in the context of PRC2-PHF1 as previously laid out might not be fully disclosed and the experiment should be repeated again with lower concentrations of PRC2-PHF1 or PRC2. In this hence only preliminary result, as quantified from densitometry analysis in triplicates, PRC2-PHF1 showed an apparent K_{d} of app 32 nM as compared to 29 nM on mutated nucleosomes.
More importantly, mutations of the acidic patch do not affect activity of PRC2-PHF1 as shown by HMTase assay and the corresponding WB shown in Fig. 4.19 C, where the the amount of trimethylation $H2A^{E61A,\ E64A,\ D90A,\ E92A}$ or $H2A^{D90A,\ E92A}$ mutated remained equal to wt nucleosomes. This result is in agreement with the structural observations described above.

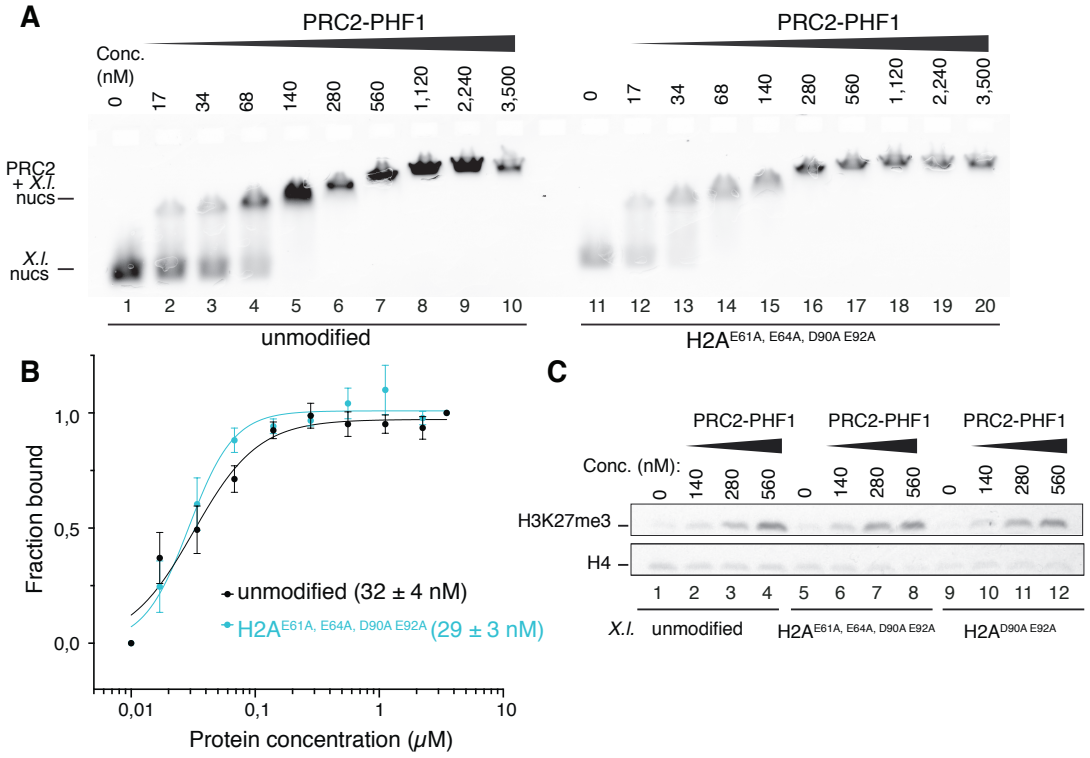

Figure 4.19: Mutations in the acidic patch of the nucleosome do not affect PRC2 binding or activity.

A. EMSA was performed with wt PRC2-PHF1 on either unmodified or mutated ($H2A^{E61A,\ E64A,\ D90A,\ E92A}$) mutated 215 bp 6-carboxyfluorescein labeled *X.l* nucleosomes in triplicates and quantified by densitometric analysis shown in B. Error bars represent SEM of the fitted values. A Hill function was fitted to estimate the apparent K_d (indicated in brackets). C. HMTase activity of increasing amounts of PRC2-PHF1 on unmodified/$H2A^{E61A,\ E64A,\ D90A,\ E92A}$/$H2A^{D90A,\ E92A}$ mutated 215 bp *X.l.* nucleosomes was probed by WB using H3K27me1/3 antibodies. H4 antibody was used as a loading control.

Chapter 5

Discussion

PRC2 binds to nucleosomes via several interactions with the nucleosomal DNA (Poepsel et al., 2018). Yet, the distinct roles of these different binding sites in PRC2 function remain elusive. Furthermore, certain histone marks such as H3K4me3 and H3K36me2/3 inhibit PRC2 in *cis* (Schmitges et al., 2011; Yuan et al., 2011; Voigt et al., 2012), but it is still unknown by which molecular mechanism PRC2 interprets these pre-existing modifications.

Structural and biochemical studies performed in this thesis attempt to address these questions and to describe the relay of interactions required for positioning of H3K27 in the EZH2 active site. Results described in chapter 4 show that binding of $EZH2^{CXC}$ residues to the DNA of the substrate nucleosome is strictly necessary for PRC2 activity. A second PRC2 binding site involves residues of EED and possibly $EZH2^{SBD/SANT1}$. The interactions with DNA of the allosteric nucleosome established by EED seem to be required for efficient allosteric EZH2 activation by H3K27me3. Furthermore, they likely provide stabilization of PRC2 on chromatin. Unmodified H3K36 is located in a tight interface formed between $EZH2^{CXC}$ and nucleosomal DNA. Its ε-amino group is seemingly engaged in long-range electrostatic interactions with the phosphate backbone of the DNA and in polar interactions with the carbonyl group of the CXC residue Q570.
These results suggest that the relay of PRC2 interactions required for positioning of H3K27 in the EZH2 active site involves the following steps: initially, correct interaction with the nucleosomal substrate needs to be established by $EZH2^{CXC}$. This contact can be further stabilized by additional binding of e.g. EED to a neighboring nucleosome. Subsequently, the binding of EZH2 on the substrate nucleosome enables the positioning of the H3 tail in the EZH2 groove. Here, the H3 tail is steadied by several contacts to EZH2 residues, the 'bridge helix' and importantly, by docking of H3K36 in the $EZH2^{CXC}$-DNA interface. The geometry established by the described PRC2 binding events sets the stage for PRC2 inhibition by H3K36me3. Within the time frame of PRC2 binding and catalysis cycle, the additional methyl groups hinder correct docking by H3K36 and thereby derail a productive alignment of the H3 tail in the EZH2 active site.

5.1 EZH2^{CXC} binding to substrate nucleosome allows for positioning of the H3 tail in the EZH2 active site

Cryo-EM studies of PRC2-PHF1 on heterodimeric dinucleosomes in this thesis (see 4.1.3) show two interaction sites within the 'upper lobe' of the PRC2 complex (Ciferri et al., 2012): the first interaction site is on the substrate nucleosome and is mainly provided by the EZH2^{CXC} domain (Fig. 4.8). Based on charge and side-chain orientation, binding to the nucleosomal DNA is established by the following residues of EZH2^{CXC}: K563, Q565, K569, Q570 (see also Poepsel et al., 2018) and possibly by the "bridge helix" (Kasinath et al., 2020). The 'bridge helix' is visible as a tubular density in this study (Fig. 4.9) and a recent pre-print reported resolved side chains in their density for this region (Kasinath et al., 2020). Intriguingly, residues shown previously to be involved in automethylation regulation of EZH2 activity are located in this helix as well as in an adjacent loop (Wang et al., 2019; Lee et al., 2019). The second PRC2 binding site is likely provided by a combined interface of EED and EZH2$^{SBD/SANT1}$ (Fig. 4.8). Flexibility in this region with seemingly several possibilities of binding (analyzed in detail in Poepsel et al., 2018) hinders a more detailed analysis of interactions in this area. When comparing the pseudoatomic model of PRC2-PHF1 on dinucleosomes to the model of PRC2-AEBP2 on dinucleosomes published by Poepsel et al., 2018, the main 3D classes contributing to the final EM reconstruction of this thesis show a slightly different interaction mode of the second binding side on the allosteric nucleosome: in the study of Poepsel et al., 2018, the allosteric nucleosome seems to be closer to the EZH2^{SBD}-helix, while in this study it seems closer to the EED surface and the EZH2$^{SBD/SANT1}$ head more bent towards the nucleosomal DNA (Fig. 4.10). It should be however noted that several 3D classes were used in both cases for the final 3D refinement and 3D reconstruction, and presumably a mixture of different binding modes is contributing to the final averaged density. The increased flexibility should hence be taken into account when comparing the two models.
In contrast, the binding mode of EZH2 on the substrate nucleosome is highly comparable between the different structures of PRC2-AEBP2/AEBP2-JARID2 (Poepsel et al., 2018; Kasinath et al., 2020) and PRC2-PHF1 on nucleosomes. Unlike PRC2-AEBP2 in Poepsel et al., 2018 and PRC2-PHF1 in this thesis, which are stabilized on dinucleosome for cryo-EM, PRC2-AEBP2-JARID2 structure shows the complex on an ubiqitinated mononucleosome (Kasinath et al., 2020) (Fig. 1.7). While lacking the stabilization possibly provided by binding of the EED and EZH2$^{SBD/SANT1}$ to the second allosteric nucleosome, in this case, EZH2 interaction with the substrate nucleosome is further stabilized by AEBP2 and JARID2 binding to the nucleosome core and ubiquitin (Kasinath et al., 2020). This observation allows a speculative model where correct binding of EZH2 via the CXC is the minimal requirement necessary for correct positioning of the H3 tail towards the active site while other binding events are needed for more stable association with chromatin (also Poepsel et al., 2018). Indeed, as shown in section 4.2.1 of this thesis, mutations of EZH2^{CXC}

residues only reduce the overall apparent binding affinity of PRC2 approximately 2.8 fold, yet in HMTase assays, mono- and trimethylation activity of the $EZH2^{CXC}$ mutant was drastically diminished (Fig. 4.11). On dinucleosomes, where the interaction of the EED and $EZH2^{SBD/SANT1}$ interface on the second nucleosome could help the correct positioning of the EZH2 towards the substrate nucleosome H3 tail, the activity can be somewhat restored yet still remains reduced. In summary, these results show that the correct binding of $EZH2^{CXC}$ to the nucleosomal DNA on the substrate nucleosome is crucial for the activity of PRC2, as it provides the correct geometry for positioning of the H3 tail towards the catalytic site.

5.2 Second PRC2 binding site possibly provides further stabilization and positions the allosteric H3 tail for EED recognition

As already discussed, the allosteric site of PRC2 is highly flexible. Previous data showed that depending on the linker length the allosteric nucleosome is flipped or connected by a straight DNA (Poepsel et al., 2018). The observations allow for a possible speculation that the flexible mode of binding allows PRC2 to adapt to the heterogeneity in chromatin. Mutations of the EED interface as shown in 4.2.2 of this thesis reduced the overall PRC2 binding affinity on mononucleosomes to 2.4x. The PRC2 construct combining the two mutated interfaces was still able to bind to mononucleosomes, albeit 5x weaker. The last observation is likely due to remaining interactions contributed by EZH2 'bridge helix' SBD and SANT1 domains as well as the 'bottom lobe' of PRC2 previously shown to also engage with nucleosomes (Poepsel et al., 2018; Nekrasov et al., 2005). Finally, activity of the EED mutant on unmodified dinucleosomes was almost reduced to levels on mononucleosomes (Fig. 4.13) indicating that on nucleosomal substrates in lower concentration of H3 tails/H3 peptides, the allosteric activation via the H3K27me3 tail is partially dependent on EED (and presumably $EZH2^{SBD/SANT1}$) binding to the nucleosomal DNA of the allosteric nucleosome. This is in line with the observation that EED binds to H3K27me3 peptide with lower affinity as compared to JARID2K116me3 (Sanulli et al., 2015; Margueron et al., 2009). In gene regions with low amounts of H3K27me3 binding of EED and $EZH2^{EED/SANT1}$ to nucleosomal DNA could potentially help positioning the trimethylated H3 tail into the aromatic cage of EED, while JARID2K116me3 could be important for *de novo* methylation (JARID2 reviewed e.g. in Holoch and Margueron, 2017).

5.3 The acidic patch of the nucleosome does not contribute to activity of PRC2-PHF1

The acidic patch is formed by negatively charged residues of H2A/H2B (Luger et al., 1997; McGinty and Tan, 2016). The cryo-EM models of PRC2-PHF1 in this thesis and PRC2-

AEBP2 as shown by Poepsel et al., 2018 on dinucleosomes showed no interactions of the core PRC2 with the acidic patch. Supportive of this notion, HMTase assays presented in this thesis in 4.2.6 showed equal activity of PRC2-PHF1 on unmodified and acidic patch mutated nucleosomes. Hence, the acidic patch of the nucleosome is not required for activity of PRC2-PHF1.
Recently, a cryo-EM study of PRC2-AEBP2-JARID2 (PRC2.2) on a H2AK119ub nucleosome, published in a pre-print, showed how recognition of the ubiquitin moiety is established by JARID2 and AEBP2 (Kasinath et al., 2020). In presence of ubiquitin, intriguingly, JARID2 and AEBP2 showed also interactions with residues of the acidic patch and the histone core of the nucleosome, respectively (Kasinath et al., 2020). In this context, it would be interesting to analyze whether the PRC2.2 complex shows a reduced activity on acidic patch mutated H2AK119ub nucleosomes.

5.4 Unmodified K36 is crucial for correct positioning of the H3 tail in the EZH2 active site

The cryo-EM data analysis shown in 4.1.3 allowed for the first time to visualize the H3 tail between exit site of the nucleosome and EZH2 active site. EZH2 is extensively engaged in interactions with residues of the H3 tail: in addition to the ionic interactions and the recognition of lysine 27 in the active site as described in Justin et al., 2016, our model suggests two hydrophobic hotspots formed by H3 tail residues H3A29/P30 and H3V35 with EZH2 residues A697, V699, I708 and F542, F557 and P558, respectively (Fig. 4.8). The H3 tail is further stabilized in the EZH2 groove by the 'bridge helix' (Fig. 4.9) (see also Kasinath et al., 2020) and by $EZH2^{CXC}$ - nucleosomal DNA interaction from below. Lysine 36 is directly juxtaposed to $EZH2^{CXC}$ residues and nucleosomal DNA with the ε-amino group of the lysine potentially engaged with long-range electrostatic interactions with the phosphate backbone of the DNA and in polar interaction with the carbonyl group of the Q570 (Fig. 4.8). The H3 tail model described in this thesis as well as the analyzed position of H3K36 is corroborated by the recently published pre-print showing PRC2-AEBP2-JARID2 on H2AK119ub nucleosome (Kasinath et al., 2020). Replacement of the H3 lysine 36 to a bulkier arginine or a shorter apolar alanine showed a strong reduction of trimethylation activity in *in vitro* HMTase assays on accordingly mutated mono-nucleosomes (H3K36A/R nucleosomes) (Section 4.2.3 of this thesis). Similar results were suggested for PRC2-AEBP2 on H3K36A/R oligonucleosomes (Jani et al., 2019) and taken together, these observations suggest that activity of PRC2 is highly susceptible to changes in chemical nature of the H3K36 side chain. In comparison, H3Kc36me3 modified mono-nucleosomes (Simon et al., 2007) almost diminished both tri- and monomethylation similarly as to previously described (Schmitges et al., 2011), while not affecting binding of PRC2 to these nucleosomes in EMSAs (Fig. 4.14 and 4.15). This is in line with the observation that it is the PRC2-DNA interaction which contributes to the high affinity binding of PRC2 to nucleosomes (Wang et al., 2017; Choi et al., 2017). H3 tail modifications H3K4me3 and H3K36me2/3

do not affect this binding on mononucleosomes or nucleosomal arrays, but rather the k_{cat} of PRC2 for H3K27 methylation (Schmitges et al., 2011; Guidotti et al., 2019; Jani et al., 2019, this thesis). Furthermore, fitting to previous observations (Schmitges et al., 2011; Jani et al., 2019), mass-spectrometry based HMTase assays monitoring activity of PRC2 on H3K36me3 peptides showed comparable amounts of monomethylation on both unmodified and H3K36me3-modified peptides. H3K36me3, therefore, can only exhibit inhibition on PRC2 in nucleosomal context but not on sole peptides (Fig. 4.16).
Taken together, these results suggest a model in which positioning of the H3 tail and the extensive EZH2-H3 tail contacts are supported by the 'bridge helix' and by the $EZH2^{CXC}$ interaction. The sum of these interactions creates geometric constraints, in which only an unmodified H3K36 allows for a rapid subsequent relay of EZH2-H3 interactions ultimately positioning H3K27 in the active site. A di- or trimethylated K36 provides a less optimal fit, thereby hampering a fast and stable recognition process. A previously proposed E579 pocket recognizing H3K36 is not compatible with the distances measured in our cryo-EM model (Jani et al., 2019).
The here proposed model of an indirect allosteric inhibition correlates well with the observed *in vivo* functions of the PRC2 inhibition by H3K36me2/3 marks and the general activity regulation of PRC2. Firstly, while not on the same H3 tail, H3K36me2/3 and H3K27me3 marks can coexist on the two histone H3 copies of the nucleosome and their coexistence is characteristic for bivalent promoters in mES cells and allow for the gene expression plasticity of those (Bernstein et al., 2006; Voigt et al., 2012, 2013). A strong inhibiting effect similar to H3K27M reminiscent mechanism of EZHIP/CATACOMB (Hübner et al., 2019; Jain et al., 2020; Pajtler et al., 2018; Piunti et al., 2019) would be likely too strong to allow PRC2 the catalysis of H3K27me3 on the neighboring H3 tail. Secondly, a global reduction of H3K27me3 in favor of H3K27me2 is sufficient for derepression of PRC2 target genes (Bernstein et al., 2006; Papp and Müller, 2006; Schwartz et al., 2006) and as introduced in the beginning of this thesis, most of PRC2 regulative aspects are focused on controlling its trimethylation activity (reviewed in Holoch and Margueron, 2017; Liu and Zhu, 2017). While in this study, both mono- and trimethylation activity were affected on H3Kc36me3 methyllysine analog containing mononucleosomes (Simon et al., 2007), other studies showed only subtle effect of H3Kc36me3 presence on monomethylation (Schmitges et al., 2011). This is potentially due to a concentration effect, where higher PRC2 or lower H3Kc36me3-nucleosomes concentrations have differential effects on the outcome of the HMTase assay. Furthermore, H3Kc36me2 nucleosomes are less inhibiting *in vitro* as compared to H3Kc36me3 nucleosomes (Schmitges et al., 2011). In agreement, combinations of H3K36me1 and H3K27me1/2/3 or H3K36me2/3 and H2K27me1 on the same H3 tail were found to exist in ES cells (Voigt et al., 2012). An indirect allosteric inhibiting mechanism by H3K36me3 as proposed in this thesis would allow for fine-tuning and adaptation of the PRC2 activity in dependence of the degree of methylation on H3K36 and in dependence of presence or absence of cofactors. Lastly, H3K27me1/2/3 are marks of different gene regions and partially carry even opposite functions (reviewed e.g. in Pirrotta, 2017). Only an indirect inhibition mechanism as opposed to a more rigid allosteric inhibition together with other activity influencing factors would allow for a precise adjusting of the PRC2

catalysis at such places.

5.5 A putative mechanism of enhanced PRC2 inhibition in presence of PHF1

As discussed above, presence of cofactors add another layer of PRC2 regulation and consequently can be expected to also influence the degree of inhibition excerted on PRC2 activity by active marks. This thesis provides a first step towards understanding the role of PHF1 in H3K36me3-PRC2 inhibition (section 4.2.5 in this thesis). Its Tudor domain is able to recognize and H3K36me3 in nucleosomal context (Musselman et al., 2012, 2013). PRC2-PHF1 showed higher amounts of monomethylation on H3K36A/R as compared to unmodified nucleosomes, while, in contrast, trimethylation was reduced on the mutated nucleosomes (Fig. 4.18). This observation suggests that the prolonged residence time contributed by PHF1 as described by Choi et al., 2017 on the nucleosomes allows PRC2 to monomethylate even in presence of bulkier or apolar residues in place of H3K36 but not to trimethylate. In contrast, on unmodified nucleosomes this effect allowed PRC2 to convert mono- and dimethylation more efficiently to trimethylation as previously reported (Nekrasov et al., 2007; Cao et al., 2008; Sarma et al., 2008; Choi et al., 2017). A similar observation was made in a recent pre-print, where in presence of JARID2 and AEBP2 the overall activity of PRC2 was less inhibited by H3K36me3 as compared to the core PRC2 (Kasinath et al., 2020). Prolonged and stabilized chromatin interaction in presence of these accessory subunits seems to partially elevate the impediment caused by H3K36me3 on PRC2.2 and by H3K36A/R on PRC2-PHF1. Yet on H3Kc36me3 containing mononucleosomes, PRC2-PHF1 showed diminished levels of monomethylation as compared to unmodified nucleosomes despite the presence of PHF1 (Fig. 4.18). The inhibition of both mono- and trimethylation activity of PRC2-PHF1 observed *in vitro* with reconstituted proteins is in line with previous studies showing that PRC2-PHF1 is inhibited on yeast-isolated H3K36me3 chromatin (Musselman et al., 2012). As the aromatic cage of PHF1 was reported to recognize and bind H3K36me3 in nucleosomal context (Musselman et al., 2012, 2013), one can speculate that this recognition could potentially out-compete the H3 tail binding in the EZH2 groove, leading to a comparable or even an enhanced inhibition as observed on the core PRC2 despite the prolonged residence time. Additional structural and biochemical studies are needed to further explore the role of PHF1 as suggested by the results in this thesis.

5.6 Summary and outlook

The last step of catalysis from di- to trimethyl mark on H3K27 by PRC2 is due to space restriction in the SET domain stereo-chemically and in terms of catalysis energetically less favorable than the preceding steps (reviewed e.g. Holoch and Margueron, 2017; Liu

and Zhu, 2017). At the same time it is the H3K27me3 mark, which is needed for the most important *in vivo* function of PRC2: the repression of its target genes such as *HOX* genes (Bernstein et al., 2006; Papp and Müller, 2006; Schwartz et al., 2006). Many of the positive regulative aspects of PRC2 are therefore focused on regulating the last step of catalysis (reviewed in Holoch and Margueron, 2017; Liu and Zhu, 2017). Furthermore, studies in cancers such as B cell lymphoma, where PRC2 trimethylation catalysis is up-regulated, demonstrate that the correct balance of PRC2 activity is important for healthy and correct cellular functions (reviewed e.g. in Liu and Zhu, 2017; Kruger et al., 2017). In this thesis, results obtained by cryo-EM and *in vitro* assays demonstrate how (1) PRC2 binds to chromatin and that (2) the minimal binding requirement for activity of PRC2 seems to be the interaction of $EZH2^{CXC}$ with nucleosomal DNA. Cryo-EM analysis performed in this thesis shows furthermore how, once PRC2 is correctly positioned on the nucleosome, the H3 tail is thread into the active site. In the EZH2 groove, the H3 tail residues are engaged with with several extensive interactions with residues of EZH2 (3). Finally, a model is proposed where the geometry set in place by these interactions creates the stereo-chemical environment in which only an unmodified K36 provides the best fit for the subsequent interactions between EZH2 and H3 tail residues (4). In contrast, active gene marks H3K36me2/3 catalyzed by protein complexes of the Trithorax/Compass family provide an imperfect fit. The additional bulkier di-/trimethyl groups at the lysine side-chain nitrogen, whose positive charge becomes dispersed, prevent H3K36 docking and thereby derail the H3 tail from stable productive interactions with the EZH2 groove. Such a model of indirect inhibition as opposed to a more rigid mechanism allows for a more fine-tuned regulation focused on the trimethylation catalysis instead of blocking the entire activity altogether. In addition, it also allows for a more adaptive response remaining capable of being downgraded in dependence of the surrounding chromatin environment (e.g. H3K36me1) and other regulative influences (e.g. presence of protein cofactors). Further experiments will improve our understanding of how these additional regulative adjustments, for example in presence of PHF1, can influence this mechanism. Finally, other active histone marks such as H3K4me3 were also reported to inhibit PRC2 on the same H3 tail (Schmitges et al., 2011) and it remains to be shown whether the molecular mechanism of H3K4me3 inhibition is reminiscent of the allosteric inhibition by H3K36me2/3 marks as revealed in this thesis.

Bibliography

P. Adamietz and A. Rudolph. ADP-ribosylation of nuclear proteins in vivo. Identification of histone H2B as a major acceptor for mono-and poly (ADP-ribose) in dimethyl sulfate-treated hepatoma AH 7974 cells. *Journal of Biological Chemistry*, 259(11):6841–6846, 1984.

P. V. Afonine, B. K. Poon, R. J. Read, O. V. Sobolev, T. C. Terwilliger, A. Urzhumtsev, and P. D. Adams. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallographica Section D: Structural Biology*, 74(6):531–544, 2018.

A. A. Alekseyenko, A. A. Gorchakov, P. V. Kharchenko, and M. I. Kuroda. 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 (7):2488–2493, 2014.

C. Alfieri, M. C. Gambetta, R. Matos, S. Glatt, P. Sehr, S. Fraterman, M. Wilm, J. Müller, and C. W. Müller. Structural basis for targeting the chromatin repressor Sfmbt to Polycomb response elements. *Genes & development*, 27(21):2367–2379, 2013.

S. Antonysamy, B. Condon, Z. Druzina, J. B. Bonanno, T. Gheyi, F. Zhang, I. MacEwan, A. Zhang, S. Ashok, L. Rodgers, et al. Structural context of disease-associated mutations and putative mechanism of autoinhibition revealed by X-ray crystallographic analysis of the EZH2-SET domain. *PLoS One*, 8(12):e84147, 2013.

S. Aranda, G. Mas, and L. Di Croce. Regulation of gene transcription by Polycomb proteins. *Science advances*, 1(11):e1500737, 2015.

X.-c. Bai, E. Rajendra, G. Yang, Y. Shi, and S. H. Scheres. Sampling the conformational space of the catalytic subunit of human γ-secretase. *elife*, 4:e11182, 2015.

I. Bajusz, L. Sipos, Z. Györgypál, E. A. Carrington, R. S. Jones, J. Gausz, and H. Gyurkovics. The Trithorax-mimic allele of Enhancer of zeste renders active domains of target genes accessible to polycomb-group-dependent silencing in Drosophila melanogaster. *Genetics*, 159(3):1135–1150, 2001.

C. Ballaré, M. Lange, A. Lapinaite, G. M. Martin, L. Morey, G. Pascual, R. Liefke, B. Simon, Y. Shi, O. Gozani, et al. Phf19 links methylated Lys36 of histone H3 to regulation of Polycomb activity. *Nature structural & molecular biology*, 19(12):1257, 2012.

F. Bantignies and G. Cavalli. Polycomb Function and Nuclear Organization. In *Polycomb Group Proteins*, pages 131–163. Elsevier, 2017.

J. Beenstock, N. Mooshayef, and D. Engelberg. How do protein kinases take a selfie (autophosphorylate)? *Trends in biochemical sciences*, 41(11):938–953, 2016.

O. Bell, C. Wirbelauer, M. Hild, A. N. Scharf, M. Schwaiger, D. M. MacAlpine, F. Zilbermann, F. Van Leeuwen, S. P. Bell, A. Imhof, et al. Localized H3K36 methylation states define histone H4K16 acetylation during transcriptional elongation in Drosophila. *The EMBO journal*, 26(24):4974–4984, 2007.

S. Bender, Y. Tang, A. M. Lindroth, V. Hovestadt, D. T. Jones, M. Kool, M. Zapatka, P. A. Northcott, D. Sturm, W. Wang, et al. Reduced H3K27me3 and DNA hypomethylation are major drivers of gene expression in K27M mutant pediatric high-grade gliomas. *Cancer cell*, 24(5):660–672, 2013.

M. Beringer, P. Pisano, V. Di Carlo, E. Blanco, P. Chammas, P. Vizán, A. Gutiérrez, S. Aranda, B. Payer, M. Wierer, et al. EPOP functionally links elongin and polycomb in pluripotent stem cells. *Molecular cell*, 64(4):645–658, 2016.

B. E. Bernstein, T. S. Mikkelsen, X. Xie, M. Kamal, D. J. Huebert, J. Cuff, B. Fry, A. Meissner, M. Wernig, K. Plath, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*, 125(2):315–326, 2006.

N. P. Blackledge, A. M. Farcas, T. Kondo, H. W. King, J. F. McGouran, L. L. Hanssen, S. Ito, S. Cooper, K. Kondo, Y. Koseki, et al. Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. *Cell*, 157(6): 1445–1459, 2014.

J. Bonnet, R. G. Lindeboom, D. Pokrovsky, G. Stricker, M. H. Celik, R. A. Rupp, J. Gagneur, M. Vermeulen, A. Imhof, and J. Müller. Quantification of proteins and histone marks in Drosophila embryos reveals stoichiometric relationships impacting chromatin regulation. *Developmental Cell*, 51(5):632–644, 2019.

A. P. Bracken, D. Pasini, M. Capra, E. Prosperini, E. Colli, and K. Helin. EZH2 is downstream of the pRB-E2F pathway, essential for proliferation and amplified in cancer. *The EMBO journal*, 22(20):5323–5335, 2003.

A. Brooun, K. S. Gajiwala, Y.-L. Deng, W. Liu, B. Bolaños, P. Bingham, Y.-A. He, W. Diehl, N. Grable, P.-P. Kung, et al. Polycomb repressive complex 2 structure with inhibitor reveals a mechanism of activation and drug resistance. *Nature communications*, 7(1):1–12, 2016.

J. L. Brown, D. Mucci, M. Whiteley, M.-L. Dirksen, and J. A. Kassis. The Drosophila Polycomb group gene pleiohomeotic encodes a DNA binding protein with homology to the transcription factor YY1. *Molecular cell*, 1(7):1057–1064, 1998.

R. Cao and Y. Zhang. The functions of E (Z)/EZH2-mediated methylation of lysine 27 in histone H3. *Current opinion in genetics & development*, 14(2):155–164, 2004a.

R. Cao and Y. Zhang. SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex. *Molecular cell*, 15(1):57–67, 2004b.

R. Cao, L. Wang, H. Wang, L. Xia, H. Erdjument-Bromage, P. Tempst, R. S. Jones, and Y. Zhang. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science*, 298(5595):1039–1043, 2002.

R. Cao, H. Wang, J. He, H. Erdjument-Bromage, P. Tempst, and Y. Zhang. Role of hPHF1 in H3K27 methylation and Hox gene silencing. *Molecular and cellular biology*, 28(5):1862–1872, 2008.

A. Casañal, B. Lohkamp, and P. Emsley. Current developments in Coot for macromolecular model building of Electron Cryo-microscopy and Crystallographic Data. *Protein Science*, 29(4):1069–1078, 2020.

M. Casanova, T. Preissner, A. Cerase, R. Poot, D. Yamada, X. Li, R. Appanah, K. Bezstarosti, J. Demmers, H. Koseki, et al. 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(8):1471–1482, 2011.

C.-S. Chan, L. Rastelli, and V. Pirrotta. A Polycomb response element in the Ubx gene that determines an epigenetically inherited state of repression. *The EMBO journal*, 13 (11):2553–2564, 1994.

K.-M. Chan, D. Fang, H. Gan, R. Hashizume, C. Yu, M. Schroeder, N. Gupta, S. Mueller, C. D. James, R. Jenkins, et al. The histone H3. 3K27M mutation in pediatric glioma reprograms H3K27 methylation and gene expression. *Genes & development*, 27(9):985–990, 2013.

S. Chen, L. Jiao, M. Shubbar, X. Yang, and X. Liu. Unique structural platforms of Suz12 dictate distinct classes of PRC2 for chromatin binding. *Molecular cell*, 69(5):840–852, 2018.

S. Chen, L. Jiao, X. Liu, X. Yang, and X. Liu. A Dimeric Structural Scaffold for PRC2-PCL Targeting to CpG Island Chromatin. *Molecular cell*, 77(6):1265–1278, 2020.

X. Cheng, R. E. Collins, and X. Zhang. Structural and sequence motifs of protein (histone) methylation enzymes. *Annu. Rev. Biophys. Biomol. Struct.*, 34:267–294, 2005.

J. Choi, A. L. Bachmann, K. Tauscher, C. Benda, B. Fierz, and J. Müller. DNA binding by PHF1 prolongs PRC2 residence time on chromatin and thereby promotes H3K27 methylation. *Nature structural & molecular biology*, 24(12):1039, 2017.

V. C. Ciccarone, D. A. Polayes, and V. A. Luckow. Generation of recombinant baculovirus DNA in E. coli using a baculovirus shuttle vector. In *Molecular diagnosis of infectious diseases*, pages 213–235. Springer, 1998.

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

P. A. Cloos, J. Christensen, K. Agger, and K. Helin. Erasing the methyl mark: histone demethylases at the center of cellular differentiation and disease. *Genes & development*, 22(9):1115–1140, 2008.

E. Conway, E. Jerman, E. Healy, S. Ito, D. Holoch, G. Oliviero, O. Deevy, E. Glancy, D. J. Fitzpatrick, M. Mucha, et al. A family of vertebrate-specific polycombs encoded by the LCOR/LCORL genes balance PRC2 subtype activities. *Molecular cell*, 70(3):408–421, 2018.

F. Crick. Central dogma of molecular biology. *Nature*, 227(5258):561–563, 1970.

F. H. Crick. On protein synthesis. In *Symp Soc Exp Biol*, volume 12, page 8, 1958.

S. Cyrus, D. Burkardt, D. D. Weaver, and W. T. Gibson. PRC2-complex related dysfunction in overgrowth syndromes: A review of EZH2, EED, and SUZ12 and their syndromic phenotypes. In *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*, volume 181, pages 519–531. Wiley Online Library, 2019.

B. Czermin, R. Melfi, D. McCabe, V. Seitz, A. Imhof, and V. Pirrotta. Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell*, 111(2):185–196, 2002.

C. Davidovich, K. J. Goodrich, A. R. Gooding, and T. R. Cech. A dimeric state for PRC2. *Nucleic acids research*, 42(14):9236–9248, 2014.

J. R. Davie and L. C. Murphy. Level of ubiquitinated histone H2B in chromatin is coupled to ongoing transcription. *Biochemistry*, 29(20):4752–4757, 1990.

O. Deevy and A. P. Bracken. PRC2 functions in development and congenital disorders. *Development*, 146(19):dev181354, 2019.

O. Denisenko, M. Shnyreva, H. Suzuki, and K. Bomsztyk. Point mutations in the WD40 domain of Eed block its interaction with Ezh2. *Molecular and cellular biology*, 18(10): 5634–5642, 1998.

I. M. Duncan. Polycomblike: a gene that appears to be required for the normal expression of the bithorax and antennapedia gene complexes of Drosophila melanogaster. *Genetics*, 102(1):49–70, 1982.

P. Emsley, B. Lohkamp, W. Scott, and K. Cowtan. Features and Development of\textit {Coot}. *Acta Crystallogr D Biol Crystallogr*, 2010.

T. Ernst, A. J. Chase, J. Score, C. E. Hidalgo-Curtis, C. Bryant, A. V. Jones, K. Waghorn, K. Zoi, F. M. Ross, A. Reiter, et al. Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. *Nature genetics*, 42(8):722–726, 2010.

S. Eustermann, K. Schall, D. Kostrewa, K. Lakomek, M. Strauss, M. Moldt, and K.-P. Hopfner. Structural basis for ATP-dependent chromatin remodelling by the INO80 complex. *Nature*, 556(7701):386–390, 2018.

A. M. Farcas, N. P. Blackledge, I. Sudbery, H. K. Long, J. F. McGouran, N. R. Rose, S. Lee, D. Sims, A. Cerase, T. W. Sheahan, et al. KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. *elife*, 1:e00205, 2012.

K. J. Ferrari, A. Scelfo, S. Jammula, A. Cuomo, I. Barozzi, A. Stützer, W. Fischle, T. Bonaldi, and D. Pasini. Polycomb-dependent H3K27me1 and H3K27me2 regulate active transcription and enhancer fidelity. *Molecular cell*, 53(1):49–62, 2014.

W. Flemming. *Zellsubstanz, kern und zelltheilung*. Vogel, 1882.

N. J. Francis, A. J. Saurin, Z. Shao, and R. E. Kingston. Reconstitution of a functional core polycomb repressive complex. *Molecular cell*, 8(3):545–556, 2001.

N. J. Francis, R. E. Kingston, and C. L. Woodcock. Chromatin compaction by a polycomb group protein complex. *Science*, 306(5701):1574–1577, 2004.

R. E. Franklin and R. G. Gosling. Molecular configuration in sodium thymonucleate. *Nature*, 171(4356):740–741, 1953.

F. Frey, T. Sheahan, K. Finkl, G. Stoehr, M. Mann, C. Benda, and J. Müller. Molecular basis of PRC1 targeting to Polycomb response elements by PhoRC. *Genes & Development*, 30(9):1116–1127, 2016.

A. Friberg, A. Oddone, T. Klymenko, J. Müller, and M. Sattler. Structure of an atypical Tudor domain in the Drosophila Polycomblike protein. *Protein Science*, 19(10):1906–1916, 2010.

Z. Gao, J. Zhang, R. Bonasio, F. Strino, A. Sawai, F. Parisi, Y. Kluger, and D. Reinberg. PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. *Molecular cell*, 45(3):344–356, 2012.

L. J. Gaydos, A. Rechtsteiner, T. A. Egelhofer, C. R. Carroll, and S. Strome. Antagonism between MES-4 and Polycomb repressive complex 2 promotes appropriate gene expression in C. elegans germ cells. *Cell reports*, 2(5):1169–1177, 2012.

S. S. Gehani, S. Agrawal-Singh, N. Dietrich, N. S. Christophersen, K. Helin, and K. Hansen. Polycomb group protein displacement and gene activation through MSK-dependent H3K27me3S28 phosphorylation. *Molecular cell*, 39(6):886–900, 2010.

S. F. Gilbert. *Developmental biology*. Sinauer Associates, 2014.

E. Glancy, C. Ciferri, and A. P. Bracken. Structural basis for PRC2 engagement with chromatin. *Current Opinion in Structural Biology*, 67:135–144, 2020.

T. D. Goddard, C. C. Huang, E. C. Meng, E. F. Pettersen, G. S. Couch, J. H. Morris, and T. E. Ferrin. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Science*, 27(1):14–25, 2018.

D. J. Grau, B. A. Chapman, J. D. Garlick, M. Borowsky, N. J. Francis, and R. E. Kingston. Compaction of chromatin by diverse Polycomb group proteins requires localized regions of high charge. *Genes & development*, 25(20):2210–2221, 2011.

A. Grijzenhout, J. Godwin, H. Koseki, M. R. Gdula, D. Szumska, J. F. McGouran, S. Bhattacharya, B. M. Kessler, N. Brockdorff, and S. Cooper. Functional analysis of AEBP2, a PRC2 Polycomb protein, reveals a Trithorax phenotype in embryonic development and in ESCs. *Development*, 143(15):2716–2723, 2016.

M. Grunstein. Histone acetylation in chromatin structure and transcription. *Nature*, 389 (6649):349–352, 1997.

N. Guidotti, C. C. Lechner, A. L. Bachmann, and B. Fierz. A modular ligation strategy for asymmetric bivalent nucleosomes trimethylated at K36 and K27. *ChemBioChem*, 20 (9):1124–1128, 2019.

Z. Han, X. Xing, M. Hu, Y. Zhang, P. Liu, and J. Chai. Structural basis of EZH2 recognition by EED. *Structure*, 15(10):1306–1315, 2007.

S. Hauri, F. Comoglio, M. Seimiya, M. Gerstung, T. Glatter, K. Hansen, R. Aebersold, R. Paro, M. Gstaiger, and C. Beisel. A high-density map for navigating the human polycomb complexome. *Cell reports*, 17(2):583–595, 2016.

E. Healy, M. Mucha, E. Glancy, D. J. Fitzpatrick, E. Conway, H. K. Neikes, C. Monger, G. Van Mierlo, M. P. Baltissen, Y. Koseki, et al. PRC2. 1 and PRC2. 2 synergize to coordinate H3K27 trimethylation. *Molecular cell*, 76(3):437–452, 2019.

R. Henderson, A. Sali, M. L. Baker, B. Carragher, B. Devkota, K. H. Downing, E. H. Egelman, Z. Feng, J. Frank, N. Grigorieff, et al. Outcome of the first electron microscopy validation task force meeting. *Structure*, 20(2):205–214, 2012.

J. W. Højfeldt, L. Hedehus, A. Laugesen, T. Tatar, L. Wiehle, and K. Helin. Non-core subunits of the PRC2 complex are collectively required for its target-site specificity. *Molecular cell*, 76(3):423–436, 2019.

D. Holoch and R. Margueron. Polycomb repressive complex 2 structure and function. In *Polycomb group proteins*, pages 191–224. Elsevier, 2017.

H. B. Houbaviy, A. Usheva, T. Shenk, and S. K. Burley. Cocrystal structure of YY1 bound to the adeno-associated virus P5 initiator. *Proceedings of the National Academy of Sciences*, 93(24):13577–13582, 1996.

C. Huang, E. Nogales, and C. Ciferri. Molecular Architecture of the Polycomb Repressive Complex 2. In *Polycomb Group Proteins*, pages 165–189. Elsevier, 2017.

J.-M. Hübner, T. Müller, D. N. Papageorgiou, M. Mauermann, J. Krijgsveld, R. B. Russell, D. W. Ellison, S. M. Pfister, K. W. Pajtler, and M. Kool. EZHIP/CXorf67 mimics K27M mutated oncohistones and functions as an intrinsic inhibitor of PRC2 function in aggressive posterior fossa ependymoma. *Neuro-oncology*, 21(7):878–889, 2019.

A. Hughes. *A history of cytology*. Abelard Schuman, 1959.

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

J. Hymes, K. Fleischhauer, and B. Wolf. Biotinylation of histones by human serum biotinidase: assessment of biotinyl-transferase activity in sera from normal individuals and children with biotinidase deficiency. *Biochemical and molecular medicine*, 56(1):76–83, 1995.

S. L. Ilca, A. Kotecha, X. Sun, M. M. Poranen, D. I. Stuart, and J. T. Huiskonen. Localized reconstruction of subunits from electron cryomicroscopy images of macromolecular complexes. *Nature communications*, 6(1):1–8, 2015.

P. Ingham and R. Whittle. Trithorax: a new homoeotic mutation of Drosophila melanogaster causing transformations of abdominal and thoracic imaginal segments. *Molecular and General Genetics MGG*, 179(3):607–614, 1980.

S. U. Jain, A. Q. Rashoff, S. D. Krabbenhoft, D. Hoelper, T. J. Do, T. J. Gibson, S. M. Lundgren, E. R. Bondra, S. Deshmukh, A. S. Harutyunyan, et al. H3 K27M and EZHIP impede H3K27-methylation spreading by inhibiting allosterically stimulated PRC2. *Molecular Cell*, 2020.

K. S. Jani, S. U. Jain, J. G. Eva, K. L. Diehl, S. M. Lundgren, M. M. Müller, P. W. Lewis, and T. W. Muir. Histone H3 tail binds a unique sensing pocket in EZH2 to activate the PRC2 methyltransferase. *Proceedings of the National Academy of Sciences*, 116(17): 8295–8300, 2019.

T. Jenuwein and C. D. Allis. Translating the histone code. *Science*, 293(5532):1074–1080, 2001.

L. Jiao and X. Liu. Structural basis of histone H3K27 trimethylation by an active polycomb repressive complex 2. *Science*, 350(6258), 2015.

H. R. Jung, D. Pasini, K. Helin, and O. N. Jensen. Quantitative mass spectrometry of histones H3. 2 and H3. 3 in Suz12-deficient mouse embryonic stem cells reveals distinct, dynamic post-translational modifications at Lys-27 and Lys-36. *Molecular & Cellular Proteomics*, 9(5):838–850, 2010.

G. Jürgens. A group of genes controlling the spatial expression of the bithorax complex in *Drosophila*. *Nature*, 316:153–5, 1985.

N. Justin, Y. Zhang, C. Tarricone, S. R. Martin, S. Chen, E. Underwood, V. De Marco, L. F. Haire, P. A. Walker, D. Reinberg, et al. Structural basis of oncogenic histone H3K27M inhibition of human polycomb repressive complex 2. *Nature communications*, 7(1):1–11, 2016.

A. A. Kalashnikova, M. E. Porter-Goff, U. M. Muthurajan, K. Luger, and J. C. Hansen. The role of the nucleosome acidic patch in modulating higher order chromatin structure. *Journal of the Royal Society Interface*, 10(82):20121022, 2013.

R. Kalb, S. Latwiel, H. I. Baymaz, P. W. Jansen, C. W. Müller, M. Vermeulen, and J. Müller. Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. *Nature structural & molecular biology*, 21(6):569–571, 2014.

V. Kasinath, M. Faini, S. Poepsel, D. Reif, X. A. Feng, G. Stjepanovic, R. Aebersold, and E. Nogales. Structures of human PRC2 with its cofactors AEBP2 and JARID2. *Science*, 359(6378):940–944, 2018.

V. Kasinath, C. Beck, P. Sauer, S. Poepsel, J. Kosmatka, M. Faini, D. Toso, R. Aebersold, and E. Nogales. JARID2 and AEBP2 regulate PRC2 activity in the presence of H2A ubiquitination or other histone modifications. *bioRxiv*, 2020. doi: 10.1101/2020.04.20.049213. URL `https://www.biorxiv.org/content/early/2020/05/05/2020.04.20.049213`.

J. A. Kassis and J. L. Brown. Polycomb group response elements in Drosophila and vertebrates. In *Advances in genetics*, volume 81, pages 83–118. Elsevier, 2013.

B. Kastner, N. Fischer, M. M. Golas, B. Sander, P. Dube, D. Boehringer, K. Hartmuth, J. Deckert, F. Hauer, E. Wolf, et al. GraFix: sample preparation for single-particle electron cryomicroscopy. *Nature methods*, 5(1):53–55, 2008.

T. C. Kaufman, R. Lewis, and B. Wakimoto. Cytogenetic analysis of chromosome 3 in Drosophila melanogaster: the homoeotic gene complex in polytene chromosome interval 84a-B. *Genetics*, 94(1):115–133, 1980.

C. S. Ketel, E. F. Andersen, M. L. Vargas, J. Suh, S. Strome, and J. A. Simon. Subunit contributions to histone methyltransferase activities of fly and worm polycomb group complexes. *Molecular and cellular biology*, 25(16):6857–6868, 2005.

H. Kim, K. Kang, and J. Kim. AEBP2 as a potential targeting protein for Polycomb Repression Complex PRC2. *Nucleic acids research*, 37(9):2940–2950, 2009.

T. W. Kim, B.-H. Kang, H. Jang, S. Kwak, J. Shin, H. Kim, S.-E. Lee, S.-M. Lee, J.-H. Lee, J.-H. Kim, et al. Ctbp2 modulates NuRD-mediated deacetylation of H3K27 and facilitates PRC2-mediated H3K27me3 in active embryonic stem cell genes during exit from pluripotency. *Stem Cells*, 33(8):2442–2455, 2015.

T. Klymenko and J. Müller. The histone methyltransferases Trithorax and Ash1 prevent transcriptional silencing by Polycomb group proteins. *EMBO reports*, 5(4):373–377, 2004.

T. Klymenko, B. Papp, W. Fischle, T. Köcher, M. Schelder, C. Fritsch, B. Wild, M. Wilm, and J. Müller. A Polycomb group protein complex with sequence-specific DNA-binding and selective methyl-lysine-binding activities. *Genes & development*, 20(9):1110–1122, 2006.

R. D. Kornberg. Chromatin structure: a repeating unit of histones and DNA. *Science*, 184 (4139):868–871, 1974.

R. D. Kornberg and J. O. Thomas. Chromatin structure: oligomers of the histones. *Science*, 184(4139):865–868, 1974.

A. Kossel. Ueber die chemische Beschaffenheit des Zellkerns. *Munchen Med. Wochenschrift*, 58:65–69, 1911.

R. Kruger, A. Graves, and M. McCabe. Activating Mutations of the EZH2 Histone Methyltransferase in Cancer. In *Polycomb Group Proteins*, pages 259–288. Elsevier, 2017.

A. Kuzmichev, K. Nishioka, H. Erdjument-Bromage, P. Tempst, and D. Reinberg. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes & development*, 16(22):2893–2905, 2002.

C. Kwong, B. Adryan, I. Bell, L. Meadows, S. Russell, J. R. Manak, and R. White. Stability and dynamics of polycomb target sites in Drosophila development. *PLoS Genet*, 4(9): e1000178, 2008.

U. K. Laemmli. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *nature*, 227(5259):680–685, 1970.

D. Landeira and A. G. Fisher. Inactive yet indispensable: the tale of Jarid2. *Trends in cell biology*, 21(2):74–80, 2011.

F. Laprell, K. Finkl, and J. Müller. Propagation of Polycomb-repressed chromatin requires sequence-specific recruitment to DNA. *Science*, 356(6333):85–88, 2017.

E. Larschan, A. A. Alekseyenko, A. A. Gortchakov, S. Peng, B. Li, P. Yang, J. L. Workman, P. J. Park, and M. I. Kuroda. MSL complex is attracted to genes marked by H3K36 trimethylation using a sequence-independent mechanism. *Molecular cell*, 28(1):121–133, 2007.

P. N. I. Lau and P. Cheung. Histone code pathway involving H3 S28 phosphorylation and K27 acetylation activates transcription and antagonizes polycomb silencing. *Proceedings of the National Academy of Sciences*, 108(7):2801–2806, 2011.

C.-H. Lee, M. Holder, D. Grau, R. Saldaña-Meyer, J.-R. Yu, R. A. Ganai, J. Zhang, M. Wang, G. LeRoy, M.-W. Dobenecker, et al. Distinct stimulatory mechanisms regulate the catalytic activity of polycomb repressive complex 2. *Molecular cell*, 70(3):435–448, 2018.

C.-H. Lee, J.-R. Yu, J. Granat, R. Saldaña-Meyer, J. Andrade, G. LeRoy, Y. Jin, P. Lund, J. M. Stafford, B. A. Garcia, et al. Automethylation of PRC2 promotes H3K27 methylation and is impaired in H3K27M pediatric glioma. *Genes & development*, 33(19-20): 1428–1440, 2019.

H.-G. Lee, T. G. Kahn, A. Simcox, Y. B. Schwartz, and V. Pirrotta. Genome-wide activities of Polycomb complexes control pervasive transcription. *Genome research*, 25(8):1170–1181, 2015.

M. Levine and K. Harding. Drosophila: the zygotic contribution. *Genes and embryos*, pages 39–94, 1989.

E. B. Lewis. A gene complex controlling segmentation in Drosophila. In *Genes, development and cancer*, pages 205–217. Springer, 1978.

P. W. Lewis, M. M. Müller, M. S. Koletsky, F. Cordero, S. Lin, L. A. Banaszynski, B. A. Garcia, T. W. Muir, O. J. Becher, and C. D. Allis. Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma. *Science*, 340(6134):857–861, 2013.

H. Li, R. Liefke, J. Jiang, J. V. Kurland, W. Tian, P. Deng, W. Zhang, Q. He, D. J. Patel, M. L. Bulyk, et al. Polycomb-like proteins link the PRC2 complex to CpG islands. *Nature*, 549(7671):287–291, 2017.

H. Liebich, E. Gesele, C. Wirth, J. Wöl, K. Jobst, and A. Lakatos. Non-enzymatic glycation of histones. *Biological mass spectrometry*, 22(2):121–123, 1993.

R. Liefke and Y. Shi. The PRC2-associated factor C17orf96 is a novel CpG island regulator in mouse ES cells. *Cell discovery*, 1(1):1–11, 2015.

N. Liu and B. Zhu. Regulation of PRC2 Activity. In *Polycomb Group Proteins*, pages 225–258. Elsevier, 2017.

P. Lowary and J. Widom. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *Journal of molecular biology*, 276(1):19–42, 1998.

K. Luger, A. W. Mäder, R. K. Richmond, D. F. Sargent, and T. J. Richmond. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, 389(6648):251–260, 1997.

K. Luger, T. J. Rechsteiner, and T. J. Richmond. Expression and purification of recombinant histones and nucleosome reconstitution. In *Chromatin Protocols*, pages 1–16. Springer, 1999.

A. Mackay, A. Burford, D. Carvalho, E. Izquierdo, J. Fazal-Salom, K. R. Taylor, L. Bjerke, M. Clarke, M. Vinci, M. Nandhabalan, et al. Integrated molecular meta-analysis of 1,000 pediatric high-grade and diffuse intrinsic pontine glioma. *Cancer cell*, 32(4):520–537, 2017.

R. K. Maeda and F. Karch. The bithorax complex of Drosophila: an exceptional Hox cluster. *Current topics in developmental biology*, 88:1–33, 2009.

R. Margueron and D. Reinberg. The Polycomb complex PRC2 and its mark in life. *Nature*, 469(7330):343–349, 2011.

R. Margueron, P. Trojer, and D. Reinberg. The key to development: interpreting the histone code? *Current opinion in genetics & development*, 15(2):163–176, 2005.

R. Margueron, G. Li, K. Sarma, A. Blais, J. Zavadil, C. L. Woodcock, B. D. Dynlacht, and D. Reinberg. Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. *Molecular cell*, 32(4):503–518, 2008.

R. Margueron, N. Justin, K. Ohno, M. L. Sharpe, J. Son, W. J. Drury Iii, P. Voigt, S. R. Martin, W. R. Taylor, V. De Marco, et al. Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature*, 461(7265):762–767, 2009.

C. Martin, R. Cao, and Y. Zhang. Substrate preferences of the EZH2 histone methyltransferase complex. *Journal of Biological Chemistry*, 281(13):8365–8370, 2006.

M. T. McCabe, A. P. Graves, G. Ganji, E. Diaz, W. S. Halsey, Y. Jiang, K. N. Smitheman, H. M. Ott, M. B. Pappalardi, K. E. Allen, et al. Mutation of A677 in histone methyltransferase EZH2 in human B-cell lymphoma promotes hypertrimethylation of histone H3 on lysine 27 (H3K27). *Proceedings of the National Academy of Sciences*, 109 (8):2989–2994, 2012.

R. K. McGinty and S. Tan. Recognition of the nucleosome by chromatin factors and enzymes. *Current Opinion in Structural Biology*, 37:54–61, 2016.

D. J. McKay, S. Klusza, T. J. Penke, M. P. Meers, K. P. Curry, S. L. McDaniel, P. Y. Malek, S. W. Cooper, D. C. Tatomer, J. D. Lieb, et al. Interrogating the function of metazoan histones using engineered gene clusters. *Developmental cell*, 32(3):373–386, 2015.

J. McKeon and H. W. Brock. Interactions of thePolycomb group of genes with homeotic loci ofDrosophila. *Roux's archives of developmental biology*, 199(7):387–396, 1991.

F. Miescher-Rüsch. Ueber die chemische Zusammensetzung der Eiterzellen. *Hoppe-Seyler's medizinisch-chemischen Untersuchungen*, 4:441–460, 1871.

A. Mohd-Sarip, F. Venturini, G. E. Chalkley, and C. P. Verrijzer. Pleiohomeotic can link polycomb to DNA and mediate transcriptional repression. *Molecular and Cellular Biology*, 22(21):7473–7483, 2002.

A. Mohd-Sarip, F. Cléard, R. K. Mishra, F. Karch, and C. P. Verrijzer. Synergistic recognition of an epigenetic DNA element by Pleiohomeotic and a Polycomb core complex. *Genes & Development*, 19(15):1755–1760, 2005.

L. Morey, L. Aloia, L. Cozzuto, S. A. Benitah, and L. Di Croce. RYBP and Cbx7 define specific biological functions of polycomb complexes in mouse embryonic stem cells. *Cell reports*, 3(1):60–69, 2013.

M. A. Morgan and A. Shilatifard. Reevaluating the roles of histone-modifying enzymes and their associated chromatin modifications in transcriptional regulation. *Nature Genetics*, 52(12):1271–1281, 2020.

R. D. Morin, N. A. Johnson, T. M. Severson, A. J. Mungall, J. An, R. Goya, J. E. Paul, M. Boyle, B. W. Woolcock, F. Kuchenbauer, et al. Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. *Nature genetics*, 42(2):181–185, 2010.

J. Müller, C. M. Hart, N. J. Francis, M. L. Vargas, A. Sengupta, B. Wild, E. L. Miller, M. B. O'Connor, R. E. Kingston, and J. A. Simon. Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. *Cell*, 111(2):197–208, 2002.

C. A. Musselman, N. Avvakumov, R. Watanabe, C. G. Abraham, M.-E. Lalonde, Z. Hong, C. Allen, S. Roy, J. K. Nuñez, J. Nickoloff, et al. Molecular basis for H3K36me3 recognition by the Tudor domain of PHF1. *Nature structural & molecular biology*, 19(12):1266, 2012.

C. A. Musselman, M. D. Gibson, E. W. Hartwick, J. A. North, J. Gatchalian, M. G. Poirier, and T. G. Kutateladze. Binding of PHF1 Tudor to H3K36me3 enhances nucleosome accessibility. *Nature communications*, 4:2969, 2013.

D. Nathan, D. E. Sterner, and S. L. Berger. Histone modifications: Now summoning sumoylation. *Proceedings of the National Academy of Sciences*, 100(23):13118–13120, 2003.

M. Nekrasov, B. Wild, and J. Müller. Nucleosome binding and histone methyltransferase activity of Drosophila PRC2. *EMBO reports*, 6(4):348–353, 2005.

M. Nekrasov, T. Klymenko, S. Fraterman, B. Papp, K. Oktaba, T. Köcher, A. Cohen, H. G. Stunnenberg, M. Wilm, and J. Müller. Pcl-PRC2 is needed to generate high levels of H3-K27 trimethylation at Polycomb target genes. *The EMBO journal*, 26(18): 4078–4088, 2007.

S. J. Nowak and V. G. Corces. Phosphorylation of histone H3: a balancing act between chromosome condensation and transcriptional activation. *TRENDS in Genetics*, 20(4): 214–220, 2004.

C. Nüsslein-Volhard and E. Wieschaus. Mutations affecting segment number and polarity in Drosophila. *Nature*, 287(5785):795–801, 1980.

D. O'Carroll, S. Erhardt, M. Pagani, S. C. Barton, M. A. Surani, and T. Jenuwein. The polycomb-group GeneEzh2 is required for early mouse development. *Molecular and cellular biology*, 21(13):4330–4336, 2001.

S. O'Connell, L. Wang, S. Robert, C. A. Jones, R. Saint, and R. S. Jones. Polycomblike PHD fingers mediate conserved interaction with enhancer of zeste protein. *Journal of Biological Chemistry*, 276(46):43065–43073, 2001.

K. Oktaba, L. Gutiérrez, J. Gagneur, C. Girardot, A. K. Sengupta, E. E. Furlong, and J. Müller. Dynamic regulation by polycomb group protein complexes controls pattern formation and the cell cycle in Drosophila. *Developmental cell*, 15(6):877–889, 2008.

A. Olins and D. Olins. Spheroid chromatin units (NU bodies). In *Journal of Cell Biology*, volume 59, pages A252–A252. ROCKEFELLER UNIV PRESS 1114 FIRST AVE, 4TH FL, NEW YORK, NY 10021, 1973.

A. L. Olins and D. E. Olins. Spheroid chromatin units (ν bodies). *Science*, 183(4122): 330–332, 1974.

D. E. Olins and A. L. Olins. Chromatin history: our view from the bridge. *Nature reviews Molecular cell biology*, 4(10):809–814, 2003.

P. Oudet, M. Gross-Bellard, and P. Chambon. Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. *Cell*, 4(4):281–300, 1975.

K. W. Pajtler, J. Wen, M. Sill, T. Lin, W. Orisme, B. Tang, J.-M. Hübner, V. Ramaswamy, S. Jia, J. D. Dalton, et al. Molecular heterogeneity and CXorf67 alterations in posterior fossa group A (PFA) ependymomas. *Acta neuropathologica*, 136(2):211–226, 2018.

B. Papp and J. Müller. Histone trimethylation and the maintenance of transcriptional ONand OFF states by trxG and PcG proteins. *Genes & development*, 20(15):2041–2054, 2006.

D. Pasini, M. Malatesta, H. R. Jung, J. Walfridsson, A. Willer, L. Olsson, J. Skotte, A. Wutz, B. Porse, O. N. Jensen, et al. Characterization of an antagonistic switch between histone H3 lysine 27 methylation and acetylation in the transcriptional regulation of Polycomb group target genes. *Nucleic acids research*, 38(15):4958–4969, 2010.

N. Paweletz. Walther Flemming: pioneer of mitosis research. *Nature reviews Molecular cell biology*, 2(1):72–75, 2001.

A. R. Pengelly, Ö. Copur, H. Jäckle, A. Herzig, and J. Müller. A histone mutant reproduces the phenotype caused by loss of histone-modifying factor Polycomb. *Science*, 339(6120): 698–699, 2013.

E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin. UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry*, 25(13):1605–1612, 2004.

V. Pirrotta. Global Functions of PRC2 Complexes. In *Polycomb Group Proteins*, pages 317–348. Elsevier, 2017.

A. Piunti and A. Shilatifard. Epigenetic balance of gene expression by Polycomb and COMPASS families. *Science*, 352(6290), 2016.

A. Piunti, E. R. Smith, M. A. Morgan, M. Ugarenko, N. Khaltyan, K. A. Helmin, C. A. Ryan, D. C. Murray, R. A. Rickels, B. D. Yilmaz, et al. CATACOMB: An endogenous inducible gene that antagonizes H3K27 methylation activity of Polycomb repressive complex 2 via an H3K27M-like mechanism. *Science advances*, 5(7):eaax2887, 2019.

S. Poepsel, V. Kasinath, and E. Nogales. Cryo-EM structures of PRC2 simultaneously engaged with two functionally distinct nucleosomes. *Nature structural & molecular biology*, 25(2):154–162, 2018.

J. Rappsilber, M. Mann, and Y. Ishihama. Protocol for micropurification, enrichment, prefractionation and storage of peptides for proteomics using Stage Tips. Nat. Protocols. 2007; 2: 1896–1906, 2007.

N. Reynolds, M. Salmon-Divon, H. Dvinge, A. Hynes-Allen, G. Balasooriya, D. Leaford, A. Behrens, P. Bertone, and B. Hendrich. NuRD-mediated deacetylation of H3K27 facilitates recruitment of Polycomb Repressive Complex 2 to direct gene repression. *The EMBO journal*, 31(3):593–605, 2012.

J. C. Rice and C. D. Allis. Code of silence. *Nature*, 414(6861):259–261, 2001.

T. Richmond, J. Finch, B. Rushton, D. Rhodes, and A. Klug. Structure of the nucleosome core particle at 7 Å resolution. *Nature*, 311(5986):532–537, 1984.

P. B. Rosenthal and R. Henderson. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *Journal of molecular biology*, 333(4):721–745, 2003.

P. B. Rosenthal and J. L. Rubinstein. Validating maps from single particle electron cryomicroscopy. *Current opinion in structural biology*, 34:135–144, 2015.

S. Sanulli, N. Justin, A. Teissandier, K. Ancelin, M. Portoso, M. Caron, A. Michaud, B. Lombard, S. T. Da Rocha, J. Offer, et al. Jarid2 methylation via the PRC2 complex regulates H3K27me3 deposition during cell differentiation. *Molecular cell*, 57(5):769–783, 2015.

K. Sarma, R. Margueron, A. Ivanov, V. Pirrotta, and D. Reinberg. Ezh2 requires PHF1 to efficiently catalyze H3 lysine 27 trimethylation in vivo. *Molecular and cellular biology*, 28(8):2718–2731, 2008.

U. Savla, J. Benes, J. Zhang, and R. S. Jones. Recruitment of Drosophila Polycomb-group proteins by Polycomblike, a component of a novel protein complex in larvae. *Development*, 135(5):813–817, 2008.

J. C. Scheuermann, A. G. de Ayala Alonso, K. Oktaba, N. Ly-Hartig, R. K. McGinty, S. Fraterman, M. Wilm, T. W. Muir, and J. Müller. Histone H2A deubiquitinase activity of the Polycomb repressive complex PR-DUB. *Nature*, 465(7295):243–247, 2010.

J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, et al. Fiji: an open-source platform for biological-image analysis. *Nature methods*, 9(7):676–682, 2012.

S. Schmähling, A. Meiler, Y. Lee, A. Mohammed, K. Finkl, K. Tauscher, L. Israel, M. Wirth, J. Philippou-Massier, H. Blum, et al. Regulation and function of H3K36 di-methylation by the trithorax-group protein complex AMC. *Development*, 145(7), 2018.

F. W. Schmitges, A. B. Prusty, M. Faty, A. Stützer, G. M. Lingaraju, J. Aiwazian, R. Sack, D. Hess, L. Li, S. Zhou, et al. Histone methylation by PRC2 is inhibited by active chromatin marks. *Molecular cell*, 42(3):330–341, 2011.

B. Schuettengruber, D. Chourrout, M. Vervoort, B. Leblanc, and G. Cavalli. Genome regulation by polycomb and trithorax proteins. *Cell*, 128(4):735–745, 2007.

B. Schuettengruber, M. Ganapathi, B. Leblanc, M. Portoso, R. Jaschek, B. Tolhuis, M. van Lohuizen, A. Tanay, and G. Cavalli. Functional anatomy of polycomb and trithorax chromatin landscapes in Drosophila embryos. *PLoS Biol*, 7(1):e1000013, 2009.

Y. Schwartz. Cooperative Recruitment of Polycomb Complexes by Polycomb Response Elements. In *Polycomb Group Proteins*, pages 111–129. Elsevier, 2017.

Y. B. Schwartz, T. G. Kahn, D. A. Nix, X.-Y. Li, R. Bourgon, M. Biggin, and V. Pirrotta. Genome-wide analysis of Polycomb targets in Drosophila melanogaster. *Nature genetics*, 38(6):700–705, 2006.

Y. B. Schwartz, T. G. Kahn, P. Stenberg, K. Ohno, R. Bourgon, and V. Pirrotta. Alternative epigenetic chromatin states of polycomb target genes. *PLoS Genet*, 6(1):e1000805, 2010.

J. Schwartzentruber, A. Korshunov, X.-Y. Liu, D. T. Jones, E. Pfaff, K. Jacob, D. Sturm, A. M. Fontebasso, D.-A. K. Quang, M. Tönjes, et al. Driver mutations in histone H3. 3 and chromatin remodelling genes in paediatric glioblastoma. *Nature*, 482(7384):226–231, 2012.

Z. Shao, F. Raible, R. Mollaaghababa, J. R. Guyon, C.-t. Wu, W. Bender, and R. E. Kingston. Stabilization of chromatin structure by PRC1, a Polycomb complex. *Cell*, 98 (1):37–46, 1999.

A. Shilatifard. The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. *Annual review of biochemistry*, 81:65–95, 2012.

J. Simon, A. Chiang, and W. Bender. Ten different Polycomb group genes are required for spatial control of the abdA and AbdB homeotic products. *Development*, 114(2):493–505, 1992.

J. Simon, A. Chiang, W. Bender, M. J. Shimell, and M. O'Connor. Elements of the Drosophila bithorax complex that mediate repression by Polycomb group products. *Developmental biology*, 158(1):131–144, 1993.

M. D. Simon, F. Chu, L. R. Racki, C. Cecile, A. L. Burlingame, B. Panning, G. J. Narlikar, and K. M. Shokat. The site-specific installation of methyl-lysine analogs into recombinant histones. *Cell*, 128(5):1003–1012, 2007.

B. C. Smith and J. M. Denu. Chemical mechanisms of histone lysine and arginine modifications. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1789(1): 45–57, 2009.

T. F. Smith, C. Gaitatzes, K. Saxena, and E. J. Neer. The WD repeat: a common architecture for diverse functions. *Trends in biochemical sciences*, 24(5):181–185, 1999.

C. J. Sneeringer, M. P. Scott, K. W. Kuntz, S. K. Knutson, R. M. Pollock, V. M. Richon, and R. A. Copeland. 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(49):20980–20985, 2010.

O. V. Sobolev, P. V. Afonine, P. D. Adams, and A. Urzhumtsev. Programming new geometry restraints: parallelity of atomic groups. *Journal of applied crystallography*, 48 (4):1130–1141, 2015.

J. Son, S. S. Shen, R. Margueron, and D. Reinberg. Nucleosome-binding activities within JARID2 and EZH1 regulate the function of PRC2 on chromatin. *Genes & development*, 27(24):2663–2677, 2013.

J.-J. Song, J. D. Garlick, and R. E. Kingston. Structural basis of histone H4 recognition by p55. *Genes & development*, 22(10):1313–1318, 2008.

H. Stark. GraFix: stabilization of fragile macromolecular complexes for single particle cryo-EM. In *Methods in enzymology*, volume 481, pages 109–126. Elsevier, 2010.

V. A. Stepanik and P. J. Harte. A mutation in the E (Z) methyltransferase that increases trimethylation of histone H3 lysine 27 and causes inappropriate silencing of active Polycomb target genes. *Developmental biology*, 364(2):249–258, 2012.

D. E. Sterner and S. L. Berger. Acetylation of histones and transcription-related factors. *Microbiology and Molecular Biology Reviews*, 64(2):435–459, 2000.

B. D. Strahl and C. D. Allis. The language of covalent histone modifications. *Nature*, 403 (6765):41–45, 2000.

G. Struhl. A gene product required for correct initiation of segmental determination in Drosophila. *Nature*, 293(5827):36–41, 1981.

Y. Z. Tan, P. R. Baldwin, J. H. Davis, J. R. Williamson, C. S. Potter, B. Carragher, and D. Lyumkis. Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nature methods*, 14(8):793–796, 2017.

K. Tatton-Brown, C. Loveday, S. Yost, M. Clarke, E. Ramsay, A. Zachariou, A. Elliott, H. Wylie, A. Ardissone, O. Rittinger, et al. Mutations in epigenetic regulation genes are a major cause of overgrowth with intellectual disability. *The American Journal of Human Genetics*, 100(5):725–736, 2017.

L. Tavares, E. Dimitrova, D. Oxley, J. Webster, R. Poot, J. Demmers, K. Bezstarosti, S. Taylor, H. Ura, H. Koide, et al. RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3. *Cell*, 148(4):664–678, 2012.

T. C. Terwilliger, S. J. Ludtke, R. J. Read, P. D. Adams, and P. V. Afonine. Improvement of cryo-EM maps by density modification. *BioRxiv*, page 845032, 2020.

P. R. Thompson, D. Wang, L. Wang, M. Fulco, N. Pediconi, D. Zhang, W. An, Q. Ge, R. G. Roeder, J. Wong, et al. Regulation of the p300 HAT domain via a novel activation loop. *Nature structural & molecular biology*, 11(4):308–315, 2004.

F. Tie, J. Prasad-Sinha, A. Birve, Å. Rasmuson-Lestander, and P. J. Harte. A 1-megadalton ESC/E (Z) complex from Drosophila that contains polycomblike and RPD3. *Molecular and cellular biology*, 23(9):3352–3362, 2003.

F. Tie, C. A. Stratton, R. L. Kurzhals, and P. J. Harte. The N terminus of Drosophila ESC binds directly to histone H3 and is required for E (Z)-dependent trimethylation of H3 lysine 27. *Molecular and cellular biology*, 27(6):2014–2026, 2007.

F. Tie, R. Banerjee, C. A. Stratton, J. Prasad-Sinha, V. Stepanik, A. Zlobin, M. O. Diaz, P. C. Scacheri, and P. J. Harte. CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing. *Development*, 136(18):3131–3141, 2009.

B. M. Turner. Cellular memory and the histone code. *Cell*, 111(3):285–291, 2002.

D. Vasudevan, E. Y. Chua, and C. A. Davey. Crystal structures of nucleosome core particles containing the '601'strong positioning sequence. *Journal of molecular biology*, 403(1): 1–10, 2010.

S. Venneti, M. T. Garimella, L. M. Sullivan, D. Martinez, J. T. Huse, A. Heguy, M. Santi, C. B. Thompson, and A. R. Judkins. Evaluation of h istone 3 lysine 27 trimethylation (H3K27me3) and enhancer of zest 2 (EZH 2) in pediatric glial and glioneuronal tumors shows decreased H3K27me3 in H3F3a K27M mutant glioblastomas. *Brain pathology*, 23 (5):558–564, 2013.

A. Verreault, P. D. Kaufman, R. Kobayashi, and B. Stillman. Nucleosomal DNA regulates the core-histone-binding subunit of the human Hat1 acetyltransferase. *Current Biology*, 8(2):96–108, 1998.

P. Voigt, G. LeRoy, W. J. Drury III, B. M. Zee, J. Son, D. B. Beck, N. L. Young, B. A. Garcia, and D. Reinberg. Asymmetrically modified nucleosomes. *Cell*, 151(1):181–193, 2012.

P. Voigt, W.-W. Tee, and D. Reinberg. A double take on bivalent promoters. *Genes & development*, 27(12):1318–1338, 2013.

C. H. Waddington. Canalization of development and the inheritance of acquired characters. *Nature*, 150(3811):563–565, 1942.

C. H. Waddington. *The strategy of the genes*. Routledge, 2014.

C. H. Waddington. *The principles of embryology*. Routledge, 2017.

C. H. Waddington et al. Organisers and genes. *Organisers and genes.*, 1940.

F. R. Wagner, C. Dienemann, H. Wang, A. Stützer, D. Tegunov, H. Urlaub, and P. Cramer. Structure of SWI/SNF chromatin remodeller RSC bound to a nucleosome. *Nature*, 579 (7799):448–451, 2020.

B. Wakimoto, F. Turner, and T. Kaufman. Defects in embryogenesis in mutants associated with the Antennapedia gene complex of Drosophila melanogaster. *Developmental biology*, 102(1):147–172, 1984.

H. Wang, L. Wang, H. Erdjument-Bromage, M. Vidal, P. Tempst, R. S. Jones, and Y. Zhang. Role of histone H2A ubiquitination in Polycomb silencing. *Nature*, 431 (7010):873–878, 2004a.

H. Wang, C. Dienemann, A. Stützer, H. Urlaub, A. C. Cheung, and P. Cramer. Structure of the transcription coactivator SAGA. *Nature*, 577(7792):717–720, 2020.

L. Wang, J. L. Brown, R. Cao, Y. Zhang, J. A. Kassis, and R. S. Jones. Hierarchical recruitment of polycomb group silencing complexes. *Molecular cell*, 14(5):637–646, 2004b.

X. Wang, R. D. Paucek, A. R. Gooding, Z. Z. Brown, J. G. Eva, T. W. Muir, and T. R. Cech. Molecular analysis of PRC2 recruitment to DNA in chromatin and its inhibition by RNA. *Nature structural & molecular biology*, 24(12):1028, 2017.

X. Wang, Y. Long, R. D. Paucek, A. R. Gooding, T. Lee, R. M. Burdorf, and T. R. Cech. Regulation of histone methylation by automethylation of PRC2. *Genes & development*, 33(19-20):1416–1427, 2019.

J. D. Watson and F. H. Crick. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature*, 171(4356):737–738, 1953.

M. H. F. Wilkins, A. R. STOKES, and H. R. Wilson. Molecular structure of nucleic acids: molecular structure of deoxypentose nucleic acids. *Nature*, 171(4356):738–740, 1953.

G. T. Wondrak, D. CERVANTES-LAUREAN, E. L. JACOBSON, and M. K. JACOBSON. Histone carbonylation in vivo and in vitro. *Biochemical Journal*, 351(3):769–777, 2000.

C. Woodcock. Ultrastructure of inactive chromatin. In *Journal of Cell Biology*, volume 59, pages A368–A368. ROCKEFELLER UNIV PRESS 1114 FIRST AVE, 4TH FL, NEW YORK, NY 10021, 1973.

C. Woodcock, J. Safer, and J. Stanchfield. Structural repeating units in chromatin: I. Evidence for their general occurrence. *Experimental cell research*, 97(1):101–110, 1976.

G. Wu, A. Broniscer, T. A. McEachron, C. Lu, B. S. Paugh, J. Becksfort, C. Qu, L. Ding, R. Huether, M. Parker, et al. Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas. *Nature genetics*, 44(3):251, 2012.

H. Wu, H. Zeng, A. Dong, F. Li, H. He, G. Senisterra, A. Seitova, S. Duan, P. J. Brown, M. Vedadi, et al. Structure of the catalytic domain of EZH2 reveals conformational plasticity in cofactor and substrate binding sites and explains oncogenic mutations. *PloS one*, 8(12):e83737, 2013.

C. Xu, C. Bian, W. Yang, M. Galka, H. Ouyang, C. Chen, W. Qiu, H. Liu, A. E. Jones, F. MacKenzie, et al. Binding of different histone marks differentially regulates the activity and specificity of polycomb repressive complex 2 (PRC2). *Proceedings of the National Academy of Sciences*, 107(45):19266–19271, 2010.

K. Yamamoto, M. Sonoda, J. Inokuchi, S. Shirasawa, and T. Sasazuki. Polycomb group suppressor of zeste 12 links heterochromatin protein 1α and enhancer of zeste 2. *Journal of Biological Chemistry*, 279(1):401–406, 2004.

D. B. Yap, J. Chu, T. Berg, M. Schapira, S.-W. G. Cheng, A. Moradian, R. D. Morin, A. J. Mungall, B. Meissner, M. Boyle, et al. Somatic mutations at EZH2 Y641 act dominantly through a mechanism of selectively altered PRC2 catalytic activity, to increase H3K27 trimethylation. *Blood*, 117(8):2451–2459, 2011.

D. T. Youmans, J. C. Schmidt, and T. R. Cech. Live-cell imaging reveals the dynamics of PRC2 and recruitment to chromatin by SUZ12-associated subunits. *Genes & development*, 32(11-12):794–805, 2018.

D. T. Youmans, A. R. Gooding, R. D. Dowell, and T. R. Cech. Competition between PRC2. 1 and 2.2 subcomplexes regulates PRC2 chromatin occupancy in human stem cells. *Molecular Cell*, 2020.

W. Yuan, M. Xu, C. Huang, N. Liu, S. Chen, and B. Zhu. H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. *Journal of Biological Chemistry*, 286(10):7983–7989, 2011.

W. Yuan, T. Wu, H. Fu, C. Dai, H. Wu, N. Liu, X. Li, M. Xu, Z. Zhang, T. Niu, et al. Dense chromatin activates Polycomb repressive complex 2 to regulate H3 lysine 27 methylation. *science*, 337(6097):971–975, 2012.

K. Zhang. Gctf: Real-time CTF determination and correction. *Journal of structural biology*, 193(1):1–12, 2016.

Q. Zhang, S. C. Agius, S. F. Flanigan, V. Levina, B. M. Owen, and C. Davidovich. Convergent evolution between PALI1 and JARID2 for the allosteric activation of PRC2. *bioRxiv*, 2020.

Y. Zhang and D. Reinberg. Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes & development*, 15(18):2343–2360, 2001.

S. Zheng, J. Wang, Y. Feng, J. Wang, and K. Ye. Solution structure of MSL2 CXC domain reveals an unusual Zn 3 Cys 9 cluster and similarity to pre-SET domains of histone lysine methyltransferases. *PLoS One*, 7(9):e45437, 2012.

S. Q. Zheng, E. Palovcak, J.-P. Armache, K. A. Verba, Y. Cheng, and D. A. Agard. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nature methods*, 14(4):331–332, 2017.

Q. Zhou, X. Huang, S. Sun, X. Li, H.-W. Wang, and S.-F. Sui. Cryo-EM structure of SNAP-SNARE assembly in 20S particle. *Cell research*, 25(5):551–560, 2015.

J. Zivanov, T. Nakane, B. O. Forsberg, D. Kimanius, W. J. Hagen, E. Lindahl, and S. H. Scheres. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife*, 7:e42166, 2018.

Acknowledgments

First and foremost, I'd like to thank my supervisor Dr. Jürg Müller, whose support, intellectual contribution, guidance and encouragement were indispensable and invaluable during my PhD. Thank you for trusting me with this exciting, multifaceted and challenging project.

I would like to thank my TAC members: Dr. Jürg Müller, Prof. Elena Conti, Prof. Friedrich Förster and Dr. Christian Benda, for the invaluable advice throughout the course of my project. My deepest gratitude goes also to my PhD Examination board members Prof. Elena Conti, Prof. Karl-Peter Hopfner, Prof. Patrick Cramer, Dr. Dietmar Martin, Prof. Klaus Förstemann and Dr. Gregor Witte for correcting my thesis and agreeing to be part of the board. I wish to, furthermore, show my gratitude to the IMPRS-LS team for organizing and managing the IMPRS graduate program, which I had the honor to be part of, with many interesting workshops.

I am also very grateful to Dr. Christian Benda and Dr. Ingmar B. Schäfer for their support in my structural studies, for sharing their wisdom and knowledge. Your support and friendship was invaluable.

I'd like to extend my deepest gratitude to Dr. Simon Poepsel and Prof. Evangelina Nogales De La Morena, from UC Berkeley. Your wonderful collaboration not only helped me to realize the project, but also are connected to fond memories of this exciting time, full of scientific exchange and experiences. Thank you for the unique opportunity to learn from such a highly experienced lab and for sharing your expertise with me. And I'd like to also thank the whole Nogales Lab for welcoming me, allowing me to work in your lab and making my visit an unforgettable experience.

A very special thanks is owed to my former master thesis supervisor, Prof. Sebastian Falk. Behind any good student, there is a great teacher. And I wouldn't know the things I know without you investing your time in my learning. Thank you for being an inspiration and for sharing your passion for science with me. I am not in any doubt that you will inspire many more people in the scientific field.

Furthermore, I'd like to thank the Müller lab team for a great atmosphere in the lab,

for the scientific discussions, input and teamwork:
Dr. Jacques Bonnet, thank you for your wonderful *in vivo* work in the project. And my deepest gratitude for your input at the last stages of my PhD.
Thank you also to Dr. Pavel Elizarev, Mattia Pieropan, Elena Karlukova, Svetlana Maltseva for great scientific and unscientific discussions, common lunches and your friendship. Mattia, Elena, Sveta - the daily lab life would have been no fun without your friendship, our crazy fridays, fly flipping parties and coffee chats. And a huge thank you for your unconditional help when I needed it.
The project wouldn't have been where it is without the fantastic technical support of our lab: my immense gratitude goes to Claudia Litz for help in protein expression/purification, Katharina Schmid for her help with the insect cell culture and western blots, Katja Finkl for her help with fly work, Sven Schkoelziger for his help in protein expression/purification and western blots and Sandra Mitzkus for help with western blots. I would also like to acknowledge and thank Ursula Grampp-Heider for lab management and the kitchen staff of our department.
Furthermore, I'd like to thank former members of the Müller lab: Dr. Reinhard Kalb, Dr. Jeongyoon Choi and Dr. Sigrun Schmähling for help, input, scientific advice and for patiently answering all of my questions in the initial and later stages of the project.

I'd like to additionally express my gratitude to Elisabeth Weyher for her sheer patience with me and for her help with mass spectrometry.

I would also like to thank the MPI Cryo-EM facility and their former and current fantastic managers: Prof. Mike Strauss, Dr. Daniel Bollschweiler and Dr. Tillman Schäfer. Thank you for the teaching and for all the hard work maintaining the microscopes and making sure that we have the best conditions for our research.
Furthermore I'd like to extend my gratitude to the MPIB core facility and Dr. Stephan Uebel and also to the MPIB mass spectrometry facility. A special thanks also to Dr. J Rajan Prabu for the fantastic computational support.

A special thanks goes to the Conti lab, my second lab family, and their lab members. Thank you for supporting me with your knowledge and expertise.
The lab is more than just a workplace. And I would like to thank the people, who made this time special and full of joyous memories: thank you Christian, Basti, Jana, Alex, Michaela, Felix, Achim, Lukas, Mahesh, Steffi, Giulia, Dirk, Ingmar, Elena, Sveta and Mattia for your friendship and the great times we had together.
Last, but not the least, I would like to express my deepest love and gratitude for my family. Especially, my dearest father and grandmother, who taught me early the high value of education, my auntie Lena, Lukas and my siblings (and of course my cat Mincoslaw, who supervised my writing). It is your trust, love and belief in me, that brought me here: I owe it all to you.