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

Structural and biochemical studies of Polycomb Repressive Complex 1

Elena Karlukova

aus

Tyumen, Russland

2024

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, 05.08.2024

ELENA KARLUKOVA

Dissertation eingereicht am 05.08.2024

1. Gutachterin: Prof. Dr. Elena Conti

2. Gutachter: Prof. Dr. Karl-Peter Hopfner

Mündliche Prüfung am 04.12.2024

Preface

Parts of this thesis were published in:

Ciapponi, M., Karlukova, E., Schkölziger, S. et al. Structural basis of the histone ubiquitination read-write mechanism of RYBP–PRC1. *Nat Struct Mol Biol* (2024).
https://doi.org/10.1038/s41594-024-01258-x

Abstract

Polycomb group (PcG) proteins play crucial roles in epigenetic transcriptional regulation. Among these, Polycomb repressive complex 1 (PRC1) functions as an E3-ligase responsible for catalyzing the H2AK119 monoubiquitination mark, a hallmark of Polycomb repression in mammals. Despite its significance, structural information regarding PRC1 remains limited, with only individual domains being crystallized. Addressing this gap, this thesis focuses on the recombinant reconstitution and structural investigation of several PRC1 subcomplexes: PRC1-PHC2, PRC1-Cbx7 and PRC1-Scml2. Leveraging integrative structural biology techniques, including protein crystallization, cross-linking mass spectrometry (XL-MS), and cryo-electron microscopy (cryo-EM), these complexes were characterised in terms of their protein-protein interaction interfaces and interaction with nucleosomes. As a part of this study, three high-resolution cryo-EM 3D reconstructions of E3 ligase domain bound to the nucleosome of PRC1-PH2 and PRC1-Cbx7 are presented herein. Cryo-EM 3D reconstructions of PRC1-Cbx7 complex with nucleosome revealed heretofore unreported extra-density, which could possibly correspond to RAWUL domain of Bmi1. The XL-MS experiment involving the PRC1-Scml2 complex revealed an additional interaction interface between the Scml2 protein and nucleosome, warranting its further biochemical and structural characterization. Overall, these findings provide foundations for forthcoming studies on PRC1, particularly in exploring conformational heterogeneity and understanding its functional relevance in the mechanisms of Polycomb repression.

Zusammenfassung

Die Proteine der Polycomb-Gruppe (PcG) spielen eine entscheidende Rolle bei der epigenetischen Transkriptionsregulation. Der Polycomb repressive complex 1 (PRC1) funktioniert als E3-Ligase, die für die Katalyse der Monoubiquitinierung von H2AK119 verantwortlich ist, einem Merkmal der Polycomb-Repression bei Säugetieren. Trotz seiner Bedeutung sind die strukturellen Informationen über den kanonischen PRC1 Komplex begrenzt, da bisher nur einzelne Domänen kristallisiert wurden. Um diese Wissenslücke zu schließen, ist der Fokus dieser Arbeit die rekombinante Rekonstitution und strukturelle Untersuchung mehrerer kanonischer PRC1-Subkomplexe: PRC1-PHC2, PRC1-Cbx7 und PRC1-Scml2. Die Komplexe wurden um Hinblick auf ihre Protein-Protein-Interaktionsschnittstellen und ihre Interaktion mit Nukleosomen durch den Einsatz integrativer strukturbiologischer Techniker, einschließlich Proteinkristallisation, Crosslinking-Massenspektrometrie (XL-MS) und Kryo-Elektronenmikroskopie (Kryo-EM), charakterisiert. Als Teil dieser Studie werden hier drei hochaufgelöste 3D-Kryo-EM-Rekonstruktionen der an das Nukleosom gebundenen E3-Ligase-Domäne von PRC1-PH2 und PRC1-Cbx7 vorgestellt. Die 3D-Kryo-EM-Rekonstruktionen des PRC1-Cbx7-Komplexes mit dem Nukleosom zeigten eine bisher unbekannte zusätzliche Dichte, die möglicherweise der RAWUL-Domäne von Bmi1 entsprechen könnte. Das XL-MS-Experiment mit dem PRC1-Scml2-Komplex offenbarte eine zusätzliche Interaktionsfläche zwischen dem Scml2-Protein und dem Nukleosom, was seine weitere biochemische und strukturelle Charakterisierung rechtfertigt. Insgesamt bilden diese Ergebnisse die Grundlage für künftige Studien über den kanonischen PRC1 Komplex, insbesondere zur Erforschung der Konformationsheterogenität und zum Verständnis seiner funktionellen Bedeutung für die Mechanismen der Polycomb-Repression.

Abbreviations

5mC: 5-Methylcytosine

AA: amino acid

AF2: AlphaFold2

cDNA: complementary DNA

CpG: Cytosine-phosphate-guanine

D.m: Drosophila melanogaster

DNA: Deoxyribonucleic Acid

E.coli: *Escherichia coli*

EM: Electron microscopy

EMSA: Electrophoretic mobility shift assay

FT: Flow-through

GGI: CpG island

H2Aub1: Histone H2A lysine 119/118 monoubiquitination

H3K27me3: Histone H3 lysine 27 trimethylation

H.s: Homo sapiens

IDR: Intrinsically disordered region

KO: Knockout

min: Minute

ON: overnight

PcG: Polycomb Group

PCR: Polymerase chain reaction

PhoRC: Pho-repressive Complex

PRC1: Polycomb Repressive Complex 1

PRC2: Polycomb Repressive Complex 2

PRE: Polycomb Response Element

PR-DUB: Polycomb Repressive-Deubiquitinase

RNA: Ribonucleic Acid

RMSD: Root Mean Square Deviation

RT: Room Temperature

SEC: Size exclusion chromatography

SDS-PAGE: Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis

SAM: Sterile Alpha Motif

TSS: Transcription Start Site

XL-MS: Crosslinking Mass Spectrometry

bp: Base pair

kD: Kilodalton

pLDDT: Per-residue log Distance Deviation Test

μl: Microliter

wt: Wild type

List of Figures

List of Tables

Contents

1 Introduction

1.1 Chromatin and 3D genome organisation

1.1.1 Nucleosome: fundamental unit of chromatin

In eukaryotic cells, genomic DNA exists in a form of chromatin, a complex formed by DNA and proteins (Flemming, 1882). More specifically, the genomic DNA is organized into arrays of nucleoprotein complexes known as nucleosomes, fundamental repeating units of chromatin.

The discovery of nucleosomes stems from a series of studies that combined electron microscopy imaging, controlled enzymatic digestion, and chemical cross-linking (Olins and Olins, 1973, Thomas and Kornberg, 1975, Woodcock et al., 1976, McKnight and Miller, 1976). Through electron microscopy imaging of chromatin samples from various eukaryotic species in reduced ionic strength, an arrangement, wherein spheroid chromatin particles are positioned along the DNA was observed. This structural organisation resembled 'beads on a string', an elegant metaphor that is still often used in the chromatin field (Figure 1A, Olins and Olins (1973)).

This observation, coupled with earlier X-ray experiments on chromatin fibre by Hewish and Burgoyne (1973) and experiments involving chromatin digestion, paved the way for R. Kornberg to establish the foundational principles of chromatin organization. According to initial Kornberg's model, the structural basis of chromatin involves a repeating module composed of four histone proteins and approximately 200 DNA base pairs, as illustrated in Figure 1B (Kornberg, 1974). Further refinement of this model led to the identification of the nucleosome unit being an octamer, comprising pairs of four distinct histone types: H2A, H2B, H3, and H4 (Thomas and Kornberg, 1975). Substantial validation of the Kornberg's hypothesis arose through the work of Oudet et al. (1975), who conducted the first electron microscopy imaging of *in vitro*-reconstituted chromatin. The observed organization of reconstituted chromatin resembled 'beads on a string' of natively isolated chromatin and aligned with the earlier findings of Olins and Olins (1973), who named the repeating spherical entities within chromatin 'nu-bodies'. As a homage to the pioneering experiments of Olins and Olins, the recurring units of histones-DNA received the name 'nucleosome'.

Subsequently, non-canonical histone variants, which deviate from the standard sequence and functional properties of canonical histones, were identified as well (Iouzalen et al., 1996, Aul and Oko, 2001, Ahmad and Henikoff, 2002, Meneghini et al., 2003, Long et al., 2019). Moreover,

the fifth histone H1 was found to interact with both the dyad and linker DNA of nucleosomes, constricting the flexibility of the nucleosomal DNA and facilitating nucleosome compaction (Fyodorov et al., 2018). This assembly of the nucleosome along with the linker histone H1 is commonly referred to as the 'chromatosome' (Simpson, 1978).

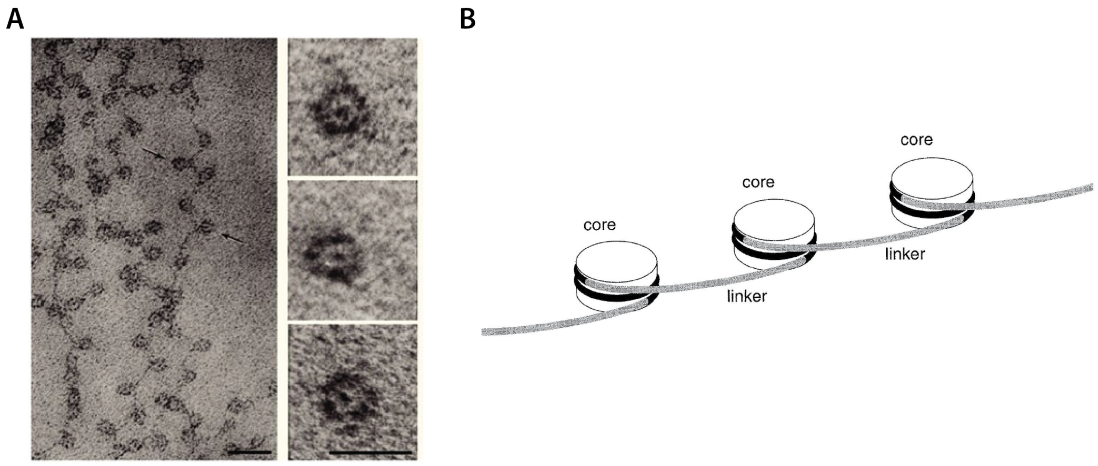


Figure 1: Chromatin structure. A. Left: electron-microscopy images of chromatin under low ionic strength conditions, 'beads on a string'. Scale bar: 30 nm. Right: electron-microscopy images of Isolated mononucleosomes from nuclease-digested chromatin. Scale bar: 10 nm. Reproduced from Olins and Olins (2003). B. Schematic organisation of chromatin into nucleosomes and DNA linkers, as suggested by Kornberg (1974). The histone octamer is represented as a disk, and the DNA as a ribbon. Reproduced from Kornberg and Lorch (1999).

1.1.2 Structural organisation of nucleosome

Initial nucleosome crystallographic studies utilized tissue-derived chromatin samples subjected to micrococcal nuclease digestion, resulting in mononucleosomes with 146 bp DNA. The homogeneity of such a sample was sufficient to enable crystallization, albeit yielding only a 7 Å low-resolution structure. This crystal structure, however, provided important initial information on nucleosome organisation by demonstrating that the nucleosomal DNA adopts a two-turn 'superhelix' conformation and encircles the histone octamer (Richmond et al., 1984). Later, the octamer organization was inferred from octamer-only crystal structures resolved at 3.1 Å resolution, in the absence of nucleosomal DNA (Arents et al., 1991).

A significant advancement in understanding nucleosomal structure was made possible through the establishment of recombinant nucleosome reconstitution. The crystallisation of recombinantly prepared nucleosomes resulted in the first high-resolution X-ray crystal structure of the nucleosome core particle resolved at 2.8 Å (Luger et al., 1997). Nucleosomes in this study were

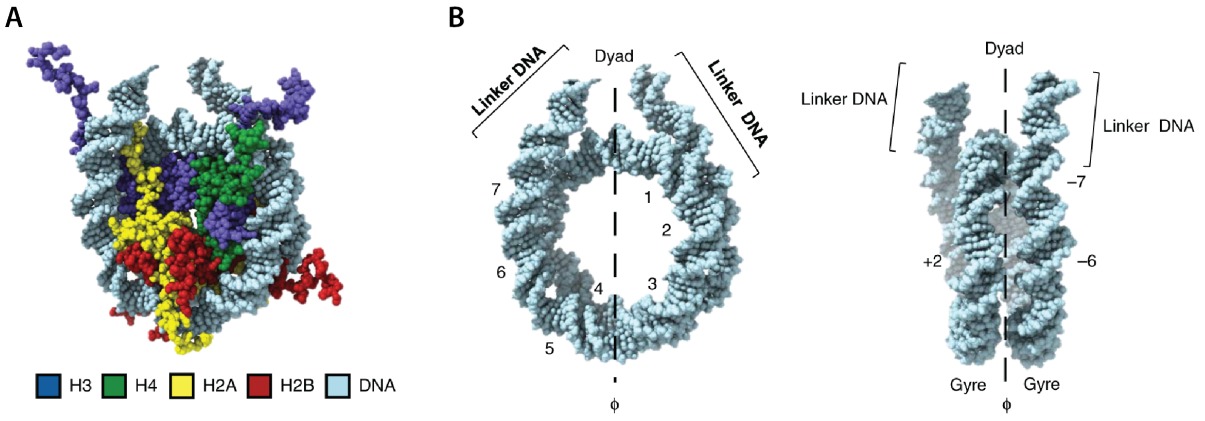


Figure 2: Nucleosome Structure A. Top view of nucleosome: The depicted model is derived from the combination of PDB entries 1KX5 and 1ZBB, with the DNA structure from 1ZBB and the histone octamer core from 1KX5. The nucleosomal DNA is illustrated in light blue, while histone proteins H3, H4, H2A, and H2B are represented in dark blue, green, yellow, and red respectively (Reproduced from Zhou et al. (2019)). B. Top (left) and side (right) views of 147 bp nucleosomal DNA (PDB entry 1ZBB). The nucleosomal DNA, spanning 145-147 base pairs, is divided into two "gyres" along the 2-fold axis, as shown on the right. Super-helical locations (SHL) denote each major groove on both sides of the 2-fold axis. The dyad, situated at the centre of the nucleosomal DNA, is defined as position 0. Additionally, linker DNA, the extra-nucleosomal DNA adjacent to the entry/exit point of the nucleosomal DNA, is highlighted. Reproduced from Zhou et al. (2019).

assembled from major-type unmodified recombinant histones (H2A, H2B, H3, H4) in the absence of linker histone H1, and therefore, referred to as canonical 'nucleosome core particles' (NPCs). The nucleosome crystal structure is illustrated in Figure 2A, with colour-coded histone proteins. 146 bp long α-satellite DNA makes around 1.65 turns around octamer as a left-handed superhelix. In Figure 2B, a more detailed topology of the nucleosomal DNA is presented. The position of a base pair that divides DNA into two halves (73- and 72-bp halves) is referred to as a dyad, which also forms a pseudo-two-fold symmetry axis of NPC. Each superhelical turn starting from the dyad is numbered as superhelical location (SHL) from -7 SHL to 7 SHL.

Histone proteins have a conserved histone fold, consisting of three α helices and two loops in α1-L1-α2-L2-α3 arrangement. The N-terminal portions of histones (also referred to as histone tails) are unstructured, with histones H2A also possessing a C-terminal tail and histone H2B C-terminal helix. Histone pairs (H3-H4 and H2A-H2B) are formed by antiparallel interaction of two histone folds, which brings an L1 loop of one histone into the proximity of the L2 loop of another histone, forming DNA-interacting L1L2 surfaces at each end of histone pair. α1 and α3 helices within each histone pack against central α2 helix, with α2 helices of both histones packed

across each other, forming a crescent-shaped histone heterodimer. In an octamer, interactions between histone heterodimers are mediated through a four-helix bundle, that is formed by $\alpha 2$ and $\alpha 3$ of each histone. Two H3-H4 dimers are arranged in a head-to-head manner, forming an H3-H4 tetramer through four helix bundle. The interactions between two H2A-H2B are formed similarly.

H2A and H2B histones form a negatively charged solvent-accessible surface called an 'acidic patch'. In humans, the acidic patch consists of six amino acid residues of H2A (Glu56, Glu61, Glu64, Asp90, Glu91 and Glu92) and two residues of H2B (Glu105 and Glu113). This surface is recognized by numerous nucleosome-binding proteins, which interact with acidic patch via an arginine-rich motif, an 'arginine anchor' (Barbera et al., 2006, McGinty et al., 2014, Morgan et al., 2016). Acidic patch interaction with nucleosomes has been first described for viral proteins, such as LANA protein from Kaposi's sarcoma-associated herpesvirus (KSHV) and IE1 protein (Barbera et al., 2006, Fang et al., 2016). In eukaryotic chromatin-binding complexes, acidic-patch binding is typically combined with another interaction, such as with nucleosomal DNA or PTM-modified histone tails (Makde et al., 2010, Morgan et al., 2016, Eustermann et al., 2018). Notably, among Polycomb complexes, both PRC1 E3 ligase and PR-DUB utilize interaction with acidic patch (McGinty et al., 2014, Ge et al., 2023, Thomas et al., 2023). The structural basis of PRC1 E3 ligase interaction with nucleosome acidic patch is reviewed in detail in section 1.5.3.

1.2 Overview of epigenetic mechanisms of gene regulation

The term 'epigenetics' was coined by C. H. Waddington more than half a century ago and originally referred to the studies of lineage specification mechanisms during organism development (Waddington, 1942). Nowadays the concept of epigenetics has evolved to encompass the studies of sequence-independent mechanisms that regulate gene expression.

Epigenetic mechanisms can be broadly categorized into two groups: (i) those that induce chemical modifications in chromatin, such as DNA methylation or histone-tails modifications, and (ii) mechanisms that bring about structural changes in chromatin, including chromatin remodeling or enhancer-promoter interactions (reviewed in Allis and Jenuwein (2016) in detail). Remarkably, as it will be discussed later, Polycomb repression falls under both of these categories.

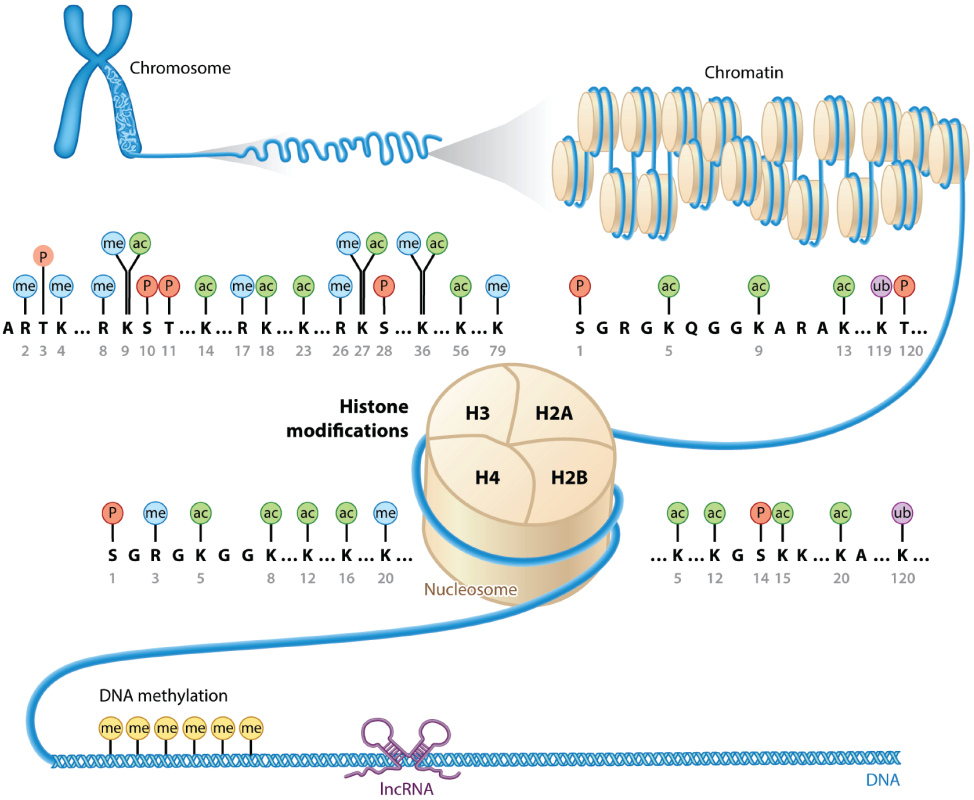


Figure 3: Posttranslationally, histone tails undergo modifications at numerous residues, including acetylation (ac), methylation (me), phosphorylation (P), and ubiquitination (ub). These posttranslational modifications, also referred to as histone marks, form a basis of epigenetic regulation of transcription. Adapted from Chen et al. (2017).

Histone modifications

Section 1.1 has discussed the concept of chromatin and the nucleosome as the 'packaging' unit for genomic DNA. Yet, the nucleosome is a not mere packaging tool, as it also serves as a scaffold for regulation and response to numerous cellular processes by providing the cell with the intricate histone code (Jenuwein and Allis, 2001).

Histone code refers to the notion that nucleosomal histones can undergo various posttranslational modifications (PTMs). These histone PTMs are commonly referred to as epigenetic marks, collectively forming the epigenetic or histone code, as depicted in Figure 3 (Turner, 1993). Different types of histone PMTs are associated with various DNA-instructed processes, such as active transcription, DNA repair, and transcriptional repression. The histone code is adaptable, it can be influenced by cellular and developmental contexts as well as environmental clues and can undergo spatial and temporal coordination. This dynamic nature is facilitated by a network of enzymes responsible for catalyzing specific post-translational modifications (writers), proteins that recognize these modifications (readers), and enzymes that remove them (erasers). It is being increasingly recognized that not only the presence of a specific mark is important but also the balance between the activities of writers and erasers for a given PTM, as was also highlighted in a recent work by Bonnet et al. (2022), which will be discussed later in detail.

For instance, in the context of Polycomb repression, tri-methylation of H3 histone K27 (H3K27me3) is catalysed by Polycomb repressive complex 2 (PRC2) which functions as a writer, while Polycomb repressive complex 1 (PRC1) serves as a H3K27me3 reader, and KDM6A and KDM6B demethylases act as H3K27me3 erasers (Müller et al., 2002, Fischle et al., 2003, De Santa et al., 2007). Similarly, the triad governing the second Polycomb-associated PMT, ubiquitination of histone H2A K119 (H2AK119ub, also referred to herein as H2Aub1), involves PRC1 as a writer, PRC2-accessory subunits AEBP2 and JARID2 as readers, and Polycomb repressive deubiquitinase (PR-DUB) as the eraser (de Napoles et al., 2004, Wang et al., 2004, Scheuermann et al., 2010, Kalb et al., 2014, Conway et al., 2021, Fursova et al., 2021).

Currently, acetylation and methylation of histone tail lysines are the most extensively studied epigenetic marks. Histone acetylation is commonly associated with active transcription, and this mark is notably enriched on active promoters and enhancers. For instance, H3K27ac is often found in the regions of active enhancers (reviewed in Verdin and Ott (2015)). The role of methylation marks is inherently more intricate compared to acetylation. First of all,

methylation is more complex due to the presence of three distinct methylation states on both lysines (mono- di-and tri-methylation) and arginines (mono-, asymmetrical and symmetrical di-methylation). Whether the methylation mark is associated with active transcription or repression depends on a particular methylated residue and degree of methylation. For instance, one of the best-characterised histone PMTs, H3K4me3, is associated with transcription and detected at active promoters and transcription start sites (TSS) (Chen et al., 2015). Another example is H3K36me3, which is found in gene bodies of actively transcribed genes, with its methyltransferase being recruited by elongating RNA polymerase itself (Bannister et al., 2005, Kizer et al., 2005). Two other prominent methylation PMTs, H3K27me3 and H3K9me3, mark transcriptionally silent genomic regions, with H3K9me3 being associated with constitutive heterochromatin and H3K27me3 with facultative heterochromatin (Snowden et al., 2002). Another example of a well-studied histone PMT is the phosphorylation of serine 139 at histone H2AX, induced by double-stranded DNA breaks (Rogakou et al., 1998).

While the aforementioned modifications are considered to be the cornerstones of epigenetic studies, recent technological advances, particularly, in mass spectrometry, keep broadening the landscape of histone PMTs. One of the recently discovered modifications is histone lactylation, which was found to be associated with active transcription (Zhang et al., 2019). Another example is glutamine serotonylation and dopaminylation, which were found in brain tissues and recently described as a novel class of histone PMT - monoaminylation (Farrelly et al., 2019, Lepack et al., 2020). Presently, at least 10 classes of histone PTMs have been identified, yet the functional significance of many remains to be fully understood.

DNA methylation

The cytosine nucleotide can undergo methylation at its 5th carbon, forming 5-methylcytosine (5mC). Within the eukaryotic genome, 5-methylcytosine is found in the context of CpG dinucleotides, which can form clusters known as CpG islands. (Doskočil and Šorm, 1962, Gardiner-Garden and Frommer, 1987). Methylation of CpG regions is pervasive in the mammalian genome, with approximately 70-80% of CpG undergoing methylation (Li and Zhang, 2014). The indispensability of 5mC mark for viability is underscored by the fact that mice deficient in DNA methyltransferases are unable to advance beyond the early stages of embryogenesis (Li et al., 1992, Okano et al., 1999).

DNA methyltransferases (DNMT) catalyze the generation of 5-methylcytosine. There are two different types of DNMT in mammals: those responsible for de novo methylation and those that maintain methylation. Among the de novo DNMTs, DNMT3A and DNMT3B are the two main DNMTs in mammals, in germline their function can be also stimulated by catalytically inactive DNMT3L (Okano et al., 1998, 1999, Bourc'his et al., 2001). De novo methylation can occur at various sites in the genome, yet only the symmetrical CpG methylation is preserved during DNA replication by the methylation maintenance enzyme DNMT1 (Sharif et al., 2007). Demethylation is a multi-step process catalyzed by the methylcytosine dioxygenase enzymes TET1-3, which sequentially oxidize 5mC to 5-hydroxymethylcytosine, 5-formylcytosine, and finally, to 5-carboxylcytosine (Kriaucionis and Heintz, 2009, Tahiliani et al., 2009, Ito et al., 2011).

Early *in vitro* and *in vivo* studies have consistently demonstrated the transcriptionally repressive nature of DNA methylation (Ben-Hattar and Jiricny, 1988, Watt and Molloy, 1988, Iguchi-Ariga and Schaffner, 1989). Notably, the promoters of actively transcribed genes lack DNA methylation (Piunti and Shilatifard, 2016). DNA methylation is also implied to play a role in the repression of transposons and germline-specific genes (Arand et al., 2012, Walsh et al., 1998, Borgel et al., 2010). It is enriched in pericentromeric satellite repeats and, paradoxically, within the bodies of actively transcribed genes (Lewis, 1978, Lister et al., 2009). However, it is suggested that the functional mechanisms of DNA methylation may vary depending on the genomic context. Consequently, these mechanisms are not necessarily the same at gene promoters, gene bodies, or repeated sequences. DNA methylation is also implicated in classical epigenetic phenomena, including genomic imprinting and X-chromosome inactivation (XCI) (Li et al., 1993, Mohandas et al., 1981). Specific readers of the 5mC mark, known as methyl-CpG-binding proteins (MBD), have also been identified (Meehan et al., 1989). Mammals have five MBD proteins (MBD1-4 and MeCP2), all found to be associated with chromatin remodelers and histone deacetylase complexes, interconnecting components of gene silencing mechanisms (Nan et al., 1998, Ng et al., 1999).

It does not come as a surprise that DNA methylation, a mechanism associated with transcriptional repression, is also connected to Polycomb repression. Firstly, the 5mC modification is predominantly mutually exclusive with the H3K27me3 mark deposited by PRC2 (Jermann et al., 2014). Whether this mutual exclusivity stems from the mode of PRC2 recruitment to unmethylated GGIs or, by DNA-methylated regions being unavailable to PRC2 due to occu-

pancy by another chromatin-binding complex, remains to be investigated. Additionally, a non-canonical PRC1.6 complex, responsible for regulating germ-line-specific repression, was shown to recruit the H3K9 methyltransferase G9A, which is crucial for DNA methylation at a subset of germline-specific genes (Auclair et al., 2016).

Chromatin remodeling

The organization of genomic DNA into chromatin, with DNA wrapped around nucleosomes, inherently reduces DNA accessibility. To facilitate proper DNA-guided processes, particularly transcription and the accessibility of transcription factor binding sites, specific complexes regulate chromatin architecture and nucleosome positioning. Chromatin remodelers are ATP-hydrolyzing protein complexes that can carry out four major nucleosome-regulatory processes: nucleosome sliding, ejection of nucleosome octamers, exchange of histones with their variant forms, and removal of the H2A-H2B dimers.

There are four families of chromatin remodelers, depending on the ATP-hydrolyzing subunit: switch/sucrose non-fermentable (SWI/SNF), imitation switch (ISWI), INOsitol requiring 80 (INO80) and chromodomain helicase DNA-binding (CHD) (Flaus et al., 2006). The common denominators for all chromatin-remodellers are binding to nucleosome and ATP hydrolysis, but they contain various auxiliary subunits and carry out different functions. SWI/SNF complexes can slide and eject nucleosomes, modulating DNA accessibility (Hartley and Madhani, 2009). ISWI remodelers contribute to nucleosome assembly and nucleosome spacing. CHD remodelers, as the name suggests, contain chromodomain domain capable of binding methylated histones and, depending on the subfamilies, can unwrap terminal DNA from octamer, space nucleosomes or regulate histone H3.3 incorporation (Sims et al., 2005, Siggens et al., 2015). INO80 remodelers can position -1 and +1 nucleosomes and have been shown to partake in transcriptional regulation (Krietenstein et al., 2016, Poli et al., 2016)

1.3 Polycomb repression system

Polycomb-Thitorax antagonism

Two groups of chromatin modifiers - Polycomb group proteins (PcG) and Thritorax group proteins (TrxG) were initially discovered in *Drosophila*, as key regulators of of Homeobox (Hox) gene expression. Hox genes constitute a highly conserved set of genes responsible for specifying the body segments along the head-tail (anterior-posterior) axis of a bilateral organism. The identification of Hox genes came from observations of phenotypes resulting from isolated mutations in fruit flies. For instance, it was observed that *antennapedia* gain-of-function mutation resulted in a transformation where antennae changed into legs, while mutations at the *bithorax* locus caused the haltere (an organ on the third thoracic segment) to transform into part of a wing (Hannah-Alava, 1958). Termed homeotic transformations, these changes are characterized by the conversion of one body segment into the structure of another body segment, and derived their name from the Greek word 'homoios,' meaning 'same' or 'similar'. Later, homeotic genes were grouped into two clusters: the bithorax complex with three homeobox genes (*Ubx*, *Abd-a*, and *Abd-b*) and the antennapedia complex with five homeobox genes (*Lab*, *Pb*, *Dfd*, *Scr* and *Antp*) (Lewis, 1978). The illustration of Hox gene clusters and their expression in body segments of *Drosophila* embryo and adult fly is provided in Figure 4.

Polycomb and Trithorax are two epigenetic regulatory systems that exert opposing effects on transcription: PcG proteins promote transcriptional repression, while Trx proteins function as transcriptional activators. This interplay between the two transcriptional regulatory systems is commonly referred to as 'Polycomb-Trithorax' antagonism. While the discussion in this section will extensively cover PcG proteins, a brief overview of TrxG proteins will be provided here.

Initially, TrxG was identified for its role in counteracting PcG silencing, facilitating the appropriate expression of Hox genes in specific regions of the fly embryo (Ingham, 1983, Klymenko and Müller, 2004). TrxG proteins constitute a diverse group with various biochemical functions. Broadly categorized, TrxG proteins fall into two groups: chromatin remodelers (e.g., SWI/SNF remodelers, such as BAF and PBAF complexes in mammals) and complexes that act as histone modifiers. The latter group includes complexes functioning as histone acetyltransferases and histone methyltransferases, such as COMPASS and COMPASS-like complexes (SET1/COMPASS, MLL1/2 COMPASS-like, MLL3/4 COMPASS-like), the Ash1L-containing AMC complex, and the CBP H3K27 acetyltransferase (Schmähling et al., 2018). COMPASS and COMPASS-like

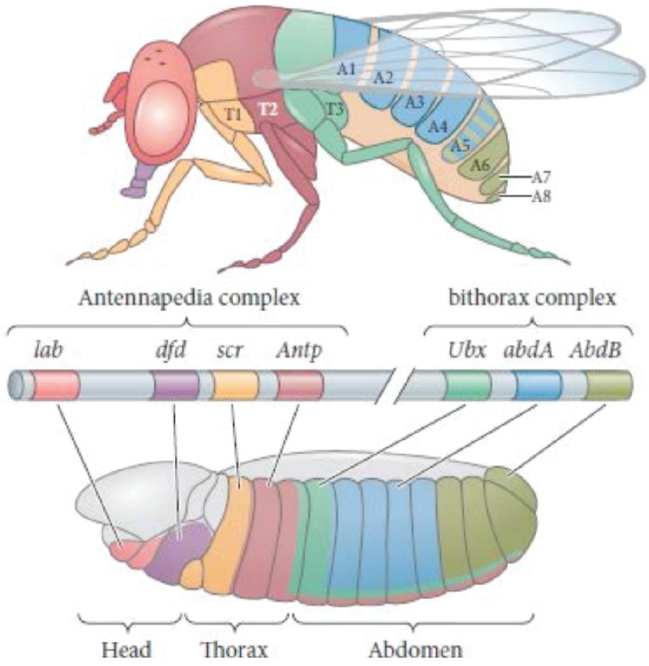


Figure 4: Expression of Homeotic Genes in *Drosophila* across various body segments in an adult fly and embryo. The differentiation of body segments in *Drosophila* is regulated by the distinct expression patterns of Homeotic (HOX) genes. In *Drosohila*, these genes are organized into two main clusters: the Antennapedia complex (ANT-C) and the Bithorax complex (BX-C). In the center, there is a gene map of ANT-C and BX-C, with genes of these complexes outlined. Below and above the gene map, both adult fly and embryo are depicted with colour-coded body segments, each corresponding to the predominant expression of specific genes, illustrating how HOX gene expression patterns shape the identity of different body segments throughout development in *Drosophila*. Illustration from Gilbert (2010).

complexes catalyze H3K4 methylation, with SET1-COMPASS responsible for generating bulk levels of H3K4me3 (Ardehali et al., 2011). Meanwhile, MLL1/2 specifically methylates promoters and a subset of genes, including Hox genes (Hu et al., 2013, Denissov et al., 2014). MLL3/4 COMPASS-like complexes have been shown to mediate mono-methylation of H3K4 at enhancers and also include a subunit, KDM6, capable of demethylating the PcG-mediated H3K27me3 mark.

Functionally, PcG and TrxG antagonism exist at the molecular level as well. For instance, COMPASS-catalyzed H3K4me3 and Ash1-mediated methylation of H3K36 have been demonstrated to inhibit PRC2 activity (Schmitges et al., 2011, Yuan et al., 2011, Finogenova et al., 2020). Similarly, PRC1 was shown to inhibit SWI/SNF chromatin remodelling activity *in vitro*, while SWI/SNF was shown to have PRC1-evicting activity *in vivo* (Francis et al., 2001, Stanton et al., 2017, Kadoch et al., 2017).

Polycomb group proteins

Polycomb group proteins were named after mutation in *Polycomb* (*Pc*) gene of *Drosophila melanogaster*, which was the first mutant identified within the group. This particular mutant, isolated in 1947, displayed a distinctive phenotype characterized by the development of sex comb on the second and third legs of adult male flies (Lewis and Mislove, 1947). Normally, sex comb (an array of thick bristles) develops only at a precise position on the prothoracic (T1) leg of male flies. In flies, heterozygous for *Pc* loss of function mutation, the level of Polycomb protein is reduced which leads to derepression of HOX gene *Scr* in segments T2 and T3, and as a result, the appearance of sex comb on the second and the third legs, where they are not typically found. In a subsequent study conducted by Ed Lewis in 1978, the phenotype of homozygous mutant larvae was described, showing that the loss of *Pc* caused a transformation of the thoracic and first seven abdominal segments into the identity of the eighth segment. This led to the proposition that *Pc* operates as a global transcriptional repressor, which led to more mutants with similar phenotypes being identified and categorised as a part of the Polycomb repressive system (Lewis, 1978). Among those were, for example, Posterior sex combs (*Psc*), Sex combs on midleg (*Scm*), Aditional sex combs (*Asx*) and Sex combs extra (*Sce*).

Subsequently, individual *Drosophila* Polycomb proteins were categorized into distinct multiprotein assemblies. These include four major Polycomb complexes: Polycomb Repressive Com-

plex 1 (PRC1), Polycomb Repressive Complex 2 (PRC2), Pho-repressive Complex (Pho-RC), and Polycomb Repressive Deubiquitinase (PR-DUB). The structural and functional aspects of each of these complexes will be discussed in detail below.

1.4 Overview of Polycomb group proteins

1.4.1 Polycomb repressive complex 2 (PRC2)

Polycomb Repressive Complex 2 (PRC2) is a multi-protein complex which is involved in maintaining transcriptional repression through its ability to catalyze methylation of H3K27me3 via its EZH2 subunit (Müller et al., 2002, Czermin et al., 2002, Pengelly et al., 2013, McKay et al., 2015). *In vivo*, PRC2 was also found to generate mono- and dimethylated H3K27, with dimethylation being the predominant H2K27 modification in both mouse embryonic stem (mES) cells and *Drosophila* (Ferrari et al., 2014, Lee et al., 2015, Bonnet et al., 2019). The core components of mammalian PRC2 are the enhancer of zeste homolog (EZH1 and EZH2), suppressor of zeste 12 homolog (SUZ12), and embryonic ectoderm development (EED). EZH2 subunit possesses histone methyltransferase activity, while SUZ12 and EED contribute to the stability and proper functioning of PRC2. PRC2 was shown to be allosterically inhibited by H3K36me2 and H3K36me3 *in vivo* and *in vitro*. Furthermore, PRC2 has been shown to undergo inhibition by H3K36me2 and H3K36me3 both *in vivo* and *in vitro* (Klymenko and Müller, 2004, Schmitges et al., 2011, Yuan et al., 2011). Recently, a mechanistic understanding of this inhibition was provided by the cryo-EM structure of PRC2 on heteromeric H3K27me3-modified di-nucleosomes. This structure shows that the binding of PRC2 to H3K27 or H3K36me2 necessitates distinct binding geometries of PRC2. Consequently, binding to one modification likely precludes the structural rearrangements required for binding to another modification (Finogenova et al., 2020).

1.4.2 Pho-repressive Complex (PhoRC) and Polycomb Response Elements (PRE)

In *Drosophila*, Polycomb group proteins assemble at specific DNA elements called Polycomb response elements (PREs) (Müller and Bienz, 1991). PREs are several hundred base pairs long sequences that contain motifs that could be specifically bound by a Polycomb group protein called Pleiohomeotic (Pho), so far the only known PcG protein with DNA-specific binding activity. Pho further associates with dSfmbt (Scm-related gene containing four MBT domains),

forming a Pho-repressive complex (PhoRC)(Klymenko et al., 2006, Alfieri et al., 2013). It was shown, that PhoRC acts as a platform to recruit PRC1 to PREs, through PRC1's accessory subunit Scm (Grimm et al., 2009, Frey et al., 2016, Kang et al., 2015). Scm bridges PhoRC and PRC1 through interactions mediated by SAM (Steril Alpha Motif) domains. Scm-SAM binds to both dSfmbt-SAM of PhoRC and Ph-SAM of PRC1 (Frey et al., 2016). Disruption of a physical link between PhoRC and PRC1 through deletion of SAM domain of Scm (ΔSAM) caused misexpression of Polycomb target genes, highlighting the importance of SAM-SAM mediated bridge between PhoRC and PRC1 for Polycomb repression. Consistent with this discovery, it has been demonstrated that PhoRC and PRC1 exhibit a substantial degree of co-localization genome-wide (Oktaba et al., 2008, Schuettengruber et al., 2009).

In mammals, however, specific regions of PcG recruitment are poorly defined. Genome-wide mapping has shown that sites of PcG recruitment are associated with unmethylated CpG islands. Although Pho homolog YY1 exists in mammals, it has been suggested, that its role for PcG complexes recruitment is rather lineage-specific and it has a non-PcG functionality as a part of INO80 complex (Wu et al., 2007, Assumpção et al., 2021). Consequently, chromatin immunoprecipitation (ChIP) experiments in human cell lines did not show a significant overlap between YY1 and PcG group proteins (PRC1 and PRC2 subunits) binding sites (Kahn et al., 2014).

1.4.3 Polycomb Repressive Deubiquitinase (PR-DUB)

Polycomb repressive deubiquitinase (PR-DUB) complex counterbalances the enzymatic function of PRC1 by deubiquitinating H2AK119ub1 (H2AK118ub1 in *D. melanogster*) (Scheuermann et al., 2010). In flies, PR-DUB consists of Calypso (called BRCA1-associated protein (BAP1) in humans) and its binding partner Asx (Additional sex combs, with several homologs Asxl1-3 in humans).

Calypso/BAP1 harbours the enzymatic function, while Asx is important for complex stabilisation and implied to play a role in targeting the complex to chromatin (de Ayala Alonso et al., 2007, Scheuermann et al., 2010). Catalytic mutation of Calypso in fly embryos was shown to lead to the global increase of H2Aub1 across the genome, with a particularly pronounced increase at Polycomb target genes, consequently leading to misexpression of those. Moreover, the excess of H2Aub1 in the absence of PR-DUB deubiquitination has been linked to increased DNA ac-

cessibility. This observation aligns with *in vitro* evidence demonstrating that H2AK119/118ub1 interferes with the folding of chromatin fibres, underscoring the potential mechanism of H2Aub1-mediated loss of gene repression through the reduction of chromatin compaction (Bonnet et al., 2022).

PR-DUB catalytic subunit Calypso harbours the ubiquitin C-terminal hydrolase (UCH) catalytic domain and interacts via its C-terminal ULD with DEUBAD domain of Asx. The structural basis for this interaction was provided by several crystallographic studies, that also explained how Asx contributes to the activation of Calypso's catalytic function. Specifically, it was shown, that Asx aids in the precise positioning of the so-called crossover loop within the Calypso UCH domain, enhancing its binding affinity to ubiquitin (De et al., 2019, Foglizzo et al., 2018). Recently, two cryo-EM studies have unveiled the interactions of the human PR-DUB complex with ubiquitinated nucleosome and H2A tail. The binding of BAP1/Asxl1 to the nucleosome results in the displacement of the H2Aub1 tail which orients precisely into BAP1 UCH catalytic center. Notably, these studies have also visualized a hitherto unseen structural element — the C-terminal domain of BAP1, whose flexibility had previously hindered visualization in crystallographic studies. Asxl1 plays a pivotal role in positioning this domain, enabling it to engage with histones H3-H4 and DNA in close proximity to the DNA dyad (Ge et al., 2023, Thomas et al., 2023).

1.5 Polycomb Repressive Complex 1 (PRC1)

1.5.1 Canonical and non-canonical forms of PRC1

Historically, mammalian PRC1 complexes have been categorized into two main groups: canonical PRC1 (cPRC1) and non-canonical PRC1 (ncPRC1). The term 'canonical' refers to PRC1 complexes that exhibit a subunit composition similar to that of the *Drosophila* PRC1 complex. On the other hand, 'non-canonical' refers to all other PRC1 complexes with distinct subunit compositions. All PRC1 complexes, whether canonical or non-canonical, share a central subunit Ring1 (stands for 'really interesting new gene', ortholog of Sce in *Drosophila*).

The variety of mammalian PRC1 emerged through genetic expansion, that was initially attributed to early mammalian ancestors. Interestingly, recent phylogenetic studies showed that diversity found in mammalian PCGF proteins can be traced back to early stages of animal evolution and predate the last common ancestor of bilaterians (e.g. mammals, insects) and

cnidarians (e.g. jellyfish, corals) (Gahan et al., 2020). The relationship between canonical and non-canonical PRC1s was also examined through an evolutionary perspective, suggesting that non-canonical PRC1 originated earlier than canonical (de Potter et al., 2023).

The compositional landscape of canonical and non-canonical PRC1s was defined in the study by Gao et al. (2012): by performing pull-downs with different PCGF proteins (PCGF 1-6) in HeLa cells, six families of mammalian PRC1 were defined, depending on PCGF homolog present. PRC1 complexes with PCGF2 and PCGF4 were also found to contain Cbx and PHC subunits, and therefore, have a similar composition to *Drosophila* PRC1 and were termed 'canonical'. Non-canonical PRC1 complexes contain other PCGFs (PCGF1/3/4/6) and form functionally diverse complexes with various components, including ubiquitin-binding proteins Rybp and YAF2, but also transcriptional factors. The landscape of mammalian PRC1 complexes is illustrated in Figure 5. The focus of this thesis is canonical PRC1 complexes, which will be referred to as PRC1 for simplicity.

The expression profiles of distinct canonical and non-canonical PRC1 subunits exhibit dynamic variations across cell types and developmental stages. In embryonic stem cells, the predominantly expressed subunits include PCGF6 and PCGF1, components of non-canonical PRC1, alongside Cbx4 and Cbx8. However, during differentiation into neural progenitors, there is a notable shift in the predominantly expressed PCGF subunit towards PCGF4, a component of canonical PRC1. Interestingly, while the predominant types of Cbx proteins remain consistent throughout differentiation, their overall abundance increases, further indicating the transition to canonical PRC1 (Kloet et al., 2016).

Distinct compositions of canonical and non-canonical PRC1s also give rise to differences in their recruitment mechanisms. Canonical PRC1 contains Cbx subunit capable of binding PRC2-deposited H3K27me3 through its Chromodomain. This mode of recruitment was historically referred to as the 'hierarchical', leading to co-localisation of H3K27me3 and Cbx throughout the genome (Bracken et al., 2006). However, more recent studies showed that Cbx recruitment might be not purely H3K27me3-mediated and DNA-binding of Cbx might play a role as well (Zhen et al., 2016). Rybp and YAF2-containing non-canonical PRC1s recognize H3K119ub1 modification on nucleosomes (Kalb et al., 2014, Cooper et al., 2014). Both Rybp and YAF2 compete with Cbxs for the same binding site at Ring proteins, which indicates that the mechanisms by which cPRC1 and YAF2/Rybp-containing ncPRC1 are recruited are mutually exclusive as well.

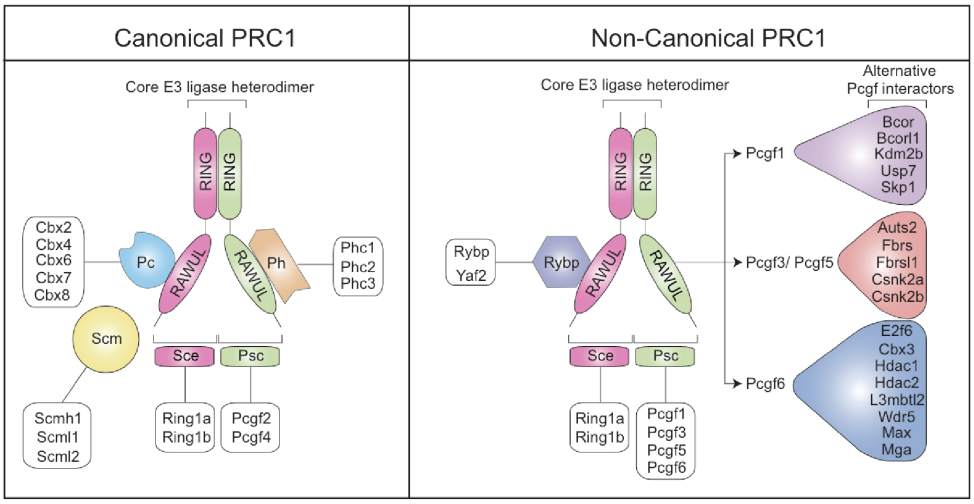


Figure 5: Schematic representation of canonical (cPRC1) and noncanonical (ncPRC1) PRC1 composition. The semantic shows *Drosophila* PRC1 proteins (Pc, Sce, Psc, Ph, and Scm) within coloured boxes, and their mammalian homologs are shown in white boxes. For instance, the *Drosophila* Pc corresponds to mammalian homologs Cbx2, Cbx4, Cbx6, Cbx7, and Cbx8. The left panel depicts the subunit composition cPRC1 complexes. Here, the core RING-PCGF heterodimer is associated with Pc and Ph subunits. The right panel illustrates the subunit composition of ncPRC1 complexes. In these complexes, the core RING-PCGF heterodimer pairs with either a Rybp or Yaf2 subunit. The composition of various ncPRC1 depends on the PCGF subunit, which, via its RAWUL domain, interacts with specific subsets of interacting partners. For example, Pcgf1 recruits Bcor and Kdm2b to its ncPRC1 complex, whereas Pcgf6 associates with E2f6 and L3mbtl2. Reproduced from Pirrotta (2017).

In mouse embryonic stem cells (mESC), non-canonical Rybp-PRC1 is recruited at target loci independent from PRC2 and H3K27me3 (Tavares et al., 2012). Conversely, depletion of Rybp in mESC reduced the global levels of H2K119ub1 and H3K27me3 but did not globally affect recruitment of Ring1b, with only a subset of genes showing reduced occupancy of both Ring1b and PRC2 subunit SUZ12 (Rose et al., 2016). The global reduction in H2AK119ub1 levels implies that Rybp enhances the E3 ligase activity of PRC1, which was demonstrated in *in vitro* studies (Zhao et al., 2020). Considering Rybp interaction with H2AK119ub1, this enhancement may occur through a feedback loop mechanism similar to that observed in PRC2, where the binding of PRC2 to its catalytic product, H3K27me3, stimulates its methyltransferase activity. This notion finds support in a recent cryo-EM study, which revealed that the Rybp zinc finger domain occupies a nucleosome 'acidic patch', where Ring1b and Bmi1 of canonical PRC1 bind (Ciapponi et al., 2024).

Notably, there are additional mechanisms for the recruitment of non-canonical PRC1 complexes to chromatin. For instance, non-canonical PRC1.1, which includes PCGF1, contains lysine demethylase KDM2B. This demethylase can bind and recruit non-canonical PRC1.1 to unmethylated CpG islands via its CxxC motif (Wu et al., 2013, Farcas et al., 2012).

1.5.2 Composition and domain organization of canonical PRC1

Ring1a and Ring1b

In mammals, there are two homologs of Ring protein - Ring1a and Ring1b. Both Ring1a and Ring1b possess an E3 ligase Zinc finger domain, also referred to as the Ring finger domain, for catalyzing nucleosome K119 monoubiquitination (H2Aub1). It has been established that in embryonic stem cells, the majority of H2Aub1 is generated by Ring1b (de Napoles et al., 2004). This observation aligns with the more severe early embryonically lethal phenotype observed in Ring1b knock-out mice compared to Ring1a knockouts, which are viable but exhibit homeotic transformations (Voncken et al., 2003, Lorente et al., 2000).

Ring1b is a 304 amino acid long protein that has two domains: N-terminal Ring finger domain and C-terminal RAWUL domain. Based on enzymatic assays, the catalytic core of the Zinc finger domain is located in amino acids region 51-91. Nevertheless, conventionally, the entire minimal catalytic domain, spanning amino acids 1-114 and encompassing the catalytic core, is often commonly referred to as the Ring finger domain. This catalytic core is characterized by

a conventional Ring motif, defined by two loops coordinating two zinc ions (Buchwald et al., 2006, Li et al., 2006).

RAWUL (Ring-finger And WD40 associated Ubiquitin-Like) domain, as the name suggests, is a ubiquitin-like domain that determines the assembly of functionally different PRC1 complexes. The RAWUL domain was defined based on the sequence analysis of the conserved C-terminal region of Ring1b, which displayed relatively low but statistically significant sequence similarity with ubiquitin-like family proteins. Even though the sequence identity between the RAWUL domain and ubiquitin is low (below 20%), they share the hydrophobic core and the main secondary structural elements (Sanchez-Pulido et al., 2008). Ring1b RAWUL domain can bind either the C-box of all five Cbx proteins, forming distinct canonical PRC1 complexes or RYBP and YAF1 proteins, leading to the formation of non-canonical PRC1 complexes. Interactions with either Cbxs or Rybp/YAF2 are mutually exclusive as they compete for the same binding site.

Interstingly, Cbx C-boxes and Rybp/YAF2 domains, which both bind Ring1b RAWUL, share very little sequence similarity, both are, however, unfolded in the absence of RAWUL. Upon binding to RAWUL, both C-boxes and Rybp/YAF2 domains form an antiparallel β-sheet that augments the major β-sheet of the RAWUL ubiquitin fold. This β-sheet then contacts the central α-helix of RAWUL, the key amino acid contacts with this helix are, however, different for C-box and Rybp/YAF2, providing the basis for selectivity (Wang et al., 2010).

Polycomb group RING finger proteins (PCGFs)

Polycomb group RING finger proteins (PCGFs) are characterized by the same domains as Ring1: N-terminal Ring finger domain and C-terminal RAWUL. While the Ring domain of Ring1b is an active E3-ligase on its own, the Ring domain of Bmi1 only provides stabilisation and catalytic enhancement of the Ring1b Ring domain, but it does not display enzymatic activity itself (Buchwald et al., 2006). There are six homologs of PCGFs in mammals: PCGF1-6, and, as discussed earlier, only PCGF2 (also known as Mel-18) and PCGF4 (also known as Bmi1) form canonical PRC1 complexes.

RAWUL domains of PCGF2 and PCGF4 interact with Homology Domains of PHC proteins (PHC 1-3). This particular interaction was visualized for Bmi1 and PHC2, using both X-ray crystallography and NMR. Similarly to the aforementioned C-box domains, HD domain of PHC2 is unfolded in solution and forms a β-sheet upon interaction with RAWUL domain of Bmi1 (Gray

et al., 2016).

In non-canonical PRC1 complexes, RAWUL domains of PCGFs (PCGF1, PCGF3 and PCGF6) do not interact with PHCs but with other subunits of ncPRC1s. PCGF1 RAWUL, for instance, interacts with the PUFD domain of BCOR1 protein. Similar to other interactors of RAWUL, BCOR1 PUFD forms a β-sheet that completes the ubiquitin fold of RAWUL (Wong et al., 2016). Interactions of other PCGF complexes have not been structurally characterized so far.

Chromobox protein homolog (CBX) proteins

Drosophila Polycomb protein has several homologous proteins in mammals, named Cbx1-8 (Chromobox protein homolog 1-8). All Cbxs share the N-terminal Chromodomain and C-terminal C-box domain. The C-terminal C-box domain, which connects Cbxs to the rest of cPRC1 complex via interaction with RAWUL domain of Ring1 protein (Schoorlemmer et al., 1997). As was mentioned before, C-box domain was found to be unfolded but forms two anti-parallel beta sheets upon binding to RAWUL (Wang et al., 2010).

The Chromodomain (**Chr**omatin **o**rganisation **mo**difier domain) was first discovered in the fly Polycomb protein (Cbx in mammals) by sequence comparison with another histone-methyl binding protein Hp1, known for its interaction with the H3K9me3 mark (Paro and Hogness, 1991). Chromodomain belongs to the 'Royal family' of domains, together with the Tudor and PWWP (Pro-Trp-Trp-Pro) and MBT (Malignant Brain Tumour) domains, to which it is structurally akin (Maurer-Stroh et al., 2003). The distinguishing structural feature of Royal family domains is the antiparallel β-barrel, formed by multiple β-sheets. Notably, while most Royal domains have five β-sheets forming the barrel, the Chromodomain deviates by having a three-β-stranded barrel. This architecture accommodates the binding of methylated histone, which forms a fourth β-strand in a barrel, wherein the methylated lysine is bound by a hydrophobic pocket (Taverna et al., 2007, Min et al., 2003). Similarly, all Royal domains were also mostly found binding methylated lysines or arginines.

Even though Chromodomain of Polycomb and Hp1 protein are structurally similar, Polycomb Chromodomain was found to bind H3K27me3 peptide with a K_d of 5 µM, and, in contrast, its binding to H3K9me3 was found to be considerably weaker, with a K_d of 125 µM. Polycomb Chromodomain could also recognize mono- and di-methylated H3K27me peptide, albeit with weaker affinity (Fischle et al., 2003). Mammalian homologs of Polycomb, Cbx proteins, display

various binding affinities to the H3K27me3 peptide. Among mouse Cbxs, the strongest binding to H3K27me3 was shown for Cbx7 with a K_d of 22 µM, Cbx2 had a K_d of 44 µM, Cbx4 and Cbx8 both around 150 µM and Cbx6 around 330 µM. Among human Cbxs, Cbx7 also had the lowest K_d of 110 µM. However, both mouse and human Chromodomain were found to be less specific, as they could also recognise other methylated peptides, such as H3K9me1/2/3. Of note, the binding affinity of the full-length Cbx proteins to H3K27me3 nucleosomes has not been reported so far, and *in vivo* it could be mediated by additional interactions, such as with nucleosomal DNA, which was demonstrated for Cbx Chromodomains *in vitro* (Bernstein et al., 2006).

Chromodomain and Cbox are the two only conserved domains within all mammalian Cbxs (in fact, there is a high degree of conservation between mammalian and fly domains as well) and are interconnected by low complexity domains, whose composition and length vary in different Cbx homologs. Cbx7 have the shortest low-complexity domain, with a negative net charge of residues, while Cbx2, Cbx4 and Cbx8 all have considerably longer and positively charged low-complexity domains.

The longest and most studied low-complexity domain among Cbx proteins is the one of Cbx2. Cbx2 low complexity domain, which is also referred to as Compaction and Phase Separation Domain (CaPS), is positively charged and rich in lysines and arginines. Mutation of lysines and arginines to uncharged alanine residues caused homeotic transformations in mice (Lau et al., 2017). *In vitro*, Cbx2 was shown to compact nucleosomal arrays and mutation of charged residues abolishes the compaction (Grau et al., 2011). More recently it was shown that the Cbx2 CaPS domain formed liquid phase-separated aggregates *in vitro*, and in a similar fashion, mutations of positively charged residues within CaPS were shown to abrogate the formation of condensates (Plys et al., 2019, Tatavosian et al., 2019). The functional mechanism of Cbx2 CaPS will be discussed in more detail in Section 1.5.5 on the non-enzymatic activity of PRC1.

Polyhomeotic-like (PHC) proteins

Polyhomeotic protein (Ph in *Drosophila*, PHC in humans) is another subunit of the canonical PRC1 complex. PHC proteins have several domains: Phe-Cys-Ser (FCS) zinc finger, which unspecifically binds nucleic acid and whose functional role is still unknown; HD (Homology Domain) that forms interaction with RAWUL domain of PCGF2 or PCGF4; and very C-terminal SAM (Sterile Alpha Motif) domain. There are three homologs of PHC in humans, PHC1-3.

The N-terminus of all PHC proteins is predicted to be mostly unstructured and abundant in glutamine, serine and proline residues. Adjacent to the N-terminal region is the FCS zinc finger domain, named after the conserved amino acid motif it contains. FCS zinc finger is 30 - 40 residues long and, as the name suggests, coordinates a Zn^{2+} atom. While being found in several Polycomb Group proteins, the function of this domain remains elusive. For instance, this type of FCS domain is also present in *Drosophila* Scm and Sfmbt proteins, although it is absent in their mammalian homologs. It was demonstrated that the FCS domain can weakly bind nucleic acids in an unspecific manner and mutations within this domain result in the PHC1 protein's inability to repress a reporter gene in an *in vitro* transcription assay. According to NMR experiments, the PHC1 FCS domain consists of three structural elements: an anti-parallel β-sheet, followed by a loop and an α-helix (Wang et al., 2011). This structural triad is similar to the one observed in the FCS domain of another Polycomb group protein - L3MBTL2 (Lechtenberg et al., 2009).

At the C-terminus, PHC proteins contain SAM (Sterile Alpha Motif) domain. Among Polycomb Group proteins, SAM domains are also found at the C-termini of Scm, Sfmbt and LMBTL proteins (Ponting, 1995). Notably, SAM domains of Ph (PHC) and Scm exhibit a distinctive capability for self-polymerization, forming a left-handed helical polymer *in vitro* (Kim et al., 2002, 2005, Nanyes et al., 2014). There are two polymerisation surfaces within the SAM domain, referred to as Mid-Loop (ML) and End Helix (EH). In polymer formation, the ML surface of one SAM domain interacts with the EH surface of another SAM domain, in a so-called head-to-tail manner (Kim et al., 2001, 2002). SAM-SAM polymerisation was found to be of great functional significance for Polycomb repression. In *Drosophila*, Ph with SAM deletion does not rescue an absence of Ph. Moreover, the absence of SAM domain in Ph results in the misexpression of Polycomb-regulated genes, which highlights its functional importance for Polycomb repression. Mutant flies with only EH surface mutations exhibit a less severe phenotype, albeit show incomplete rescue of Ph null phenotype (Gambetta and Müller, 2014). Accordingly, mice having PHC2 with polymerisation deficient SAM, display skeletal transformation consistent with Polycomb phenotype and missexpression of Polycomb target genes (Isono et al., 2013).

In vivo, polymerisation of the Ph SAM domain was shown to be controlled by post-translational modification with O-linked N-Acetylglucosamine. In line with this, the absence of O-GlcNAc transferase (OGT) catalysing this modification results in classical Polycomb phenotype in flies (Gambetta and Müller, 2014).

What role does SAM-SAM polymerization play in Polycomb repression? In *Drosophila*, binding of Scm SAM to EH surface of Sfmbt and ML surface of Ph bridges PRC1 and PhoRC complexes, recruiting PRC1 to PhoRC-bound genomic loci (Frey et al., 2016). Several studies have also demonstrated, that Ph SAM-SAM polymerisation was involved in the clustering of PRC1 complexes, facilitating the formation of nuclear foci referred to as Polycomb bodies and higher-order repressive chromatin states, so-called Polycomb domains (Wani et al., 2016, Boettiger et al., 2016).

Scm Polycomb Group Protein Like 2 (Scml2)

Scm (Sex combs on midleg) associates with canonical PRC1 in a substoichiometric manner (Levine et al., 2002). Among human homologs of Scm, Scmh1 and Scml1 have low tissue specificity, while Scml2 was found to be germline-specific and function in transcriptional regulation of spermatogenesis (Hasegawa et al., 2015, Maezawa et al., 2018). Additionally, Scml2 directly interacts with deubiquitinase USP7 and is proposed to connect USP7 to canonical PRC1 (Lecona et al., 2015).

At the N-terminus, Scm and Scm homologs contain two malignant brain tumour (MBT) repeats, which bind methylated histone peptides with lower methylation states, albeit, with low specificity (Grimm et al., 2007). Each MBT repeat is formed by a central β-barrel comprising five β-strands. Adjacent to the barrel is an N-terminal extended region that mediates contact between two MBT repeats, with additional interaction formed by β-barrels themselves, contacting each other in an asymmetric manner (Sathyamurthy et al., 2003). The crystal structure of Scm MBT domains with monomethyl-lysine-containing peptide showed that the methylated lysine was bound by the aromatic cage of the second MBT domain. Interestingly, the first MBT differed in its conformation state, which most likely precluded the binding of the first MBT domain to the methylated peptide. *In vivo*, the removal of both MBT repeats in *Drosophila* Scm, or a point mutation specifically targeting the mono-methyl binding pocket within the second MBT domain, resulted in only partial de-repression of Polycomb-regulated genes. This partial effect may be attributed to the redundant mono-methyl binding activity observed in another Polycomb group protein, Sfmbt, which interacts with Scm (Grimm et al., 2007, Klymenko et al., 2006).

Two MBT repeats are followed by the RNA binding region (RBR) that was shown to bind

RNA *in vitro* but also interact with DNA and nucleosomes. This interaction with nucleosomes could be potentially significant in the recruitment of Scml2-containing PRC1 to genomic loci (Bonasio et al., 2014).

RBR region is followed by SLED (Scm-like embedded domain). In Scml2 this domain resides in amino 354 – 468 and can recognise double-stranded DNA, also possibly playing a role in targeting Scml2 and Scml2-containing PRC1 to chromatin. The structure of the SLED domain was determined by NMR and showed a unique fold of interchanged α-helices and β-sheets arranged in the following sequential order: β1-α1-β2-α2-α3-β3-β4-α4-β5. Two α-helices - α2 and α4 lie in the centre of the domain, parallel to each other, with β-sheets flanking on the sides (Bezsonova, 2014). The functional relevance of the SLED domain to Polycomb repression was highlighted by *in vivo* experiments, wherein a mutation in this region led to gene misexpression in the nervous system of the *Drosophila* embryos (Bornemann et al., 1998).

Just like the SAM domain in Ph, the C-terminal SAM domain found in Scm and its human counterparts can self-polymerize *in vitro* and bind to SAM domains of Ph. In *Drosophila*, Scm SAM was shown to interact with both Ph SAM and Sfmbt SAM through its ML and EH surfaces, respectively (Frey et al., 2016). Notably, *in vivo* experiments showed that deletion of the SAM domain in Scm led to the misexpression of Polycomb-repressed genes, emphasizing the role of Scm-SAM domain in the Polycomb repression machinery (Frey et al., 2016).

1.5.3 Previous structural studies of PRC1

While various crystal structures of individual domains of the canonical PRC1 complex have been resolved throughout the years (as outlined in the section 1.5.2), it was not until 2014 that the PRC1 E3 ligase domain was crystallized in complex with its substrate nucleosome, visualising the mechanism of nucleosome recognition by PRC1 (McGinty et al., 2014). In this crystal structure, only the E3 ligase portion of PRC1 was utilized, consisting of Ring1b (1-116) and Bmi1 (1-109), wherein the C-terminus of Ring1b RING domain was fused to the E2 ubiquitin-conjugating enzyme UbcH5c. Notably, in this study, the fusion of E3 to E2 was crucial for successful crystallization, likely due to the higher affinity of the E3-E2 fusion complex to the nucleosome, which was attributed to E2 forming additional interactions with nucleosomal DNA. To validate that the fusion did not introduce any artefacts, authors compared the crystal structure of the fused E3 PRC1 and UbcH5c E2 on the nucleosome to the previous crystal structure of native

E3 PRC1 and UbcH5c E2 without fusion. However, it is worth noting that this E3-E2 fusion was catalytically inactive in *in vitro* ubiquitination assays, possibly due to Ring1b precluding access to the E1 adenylation site of E2 (McGinty et al., 2014).

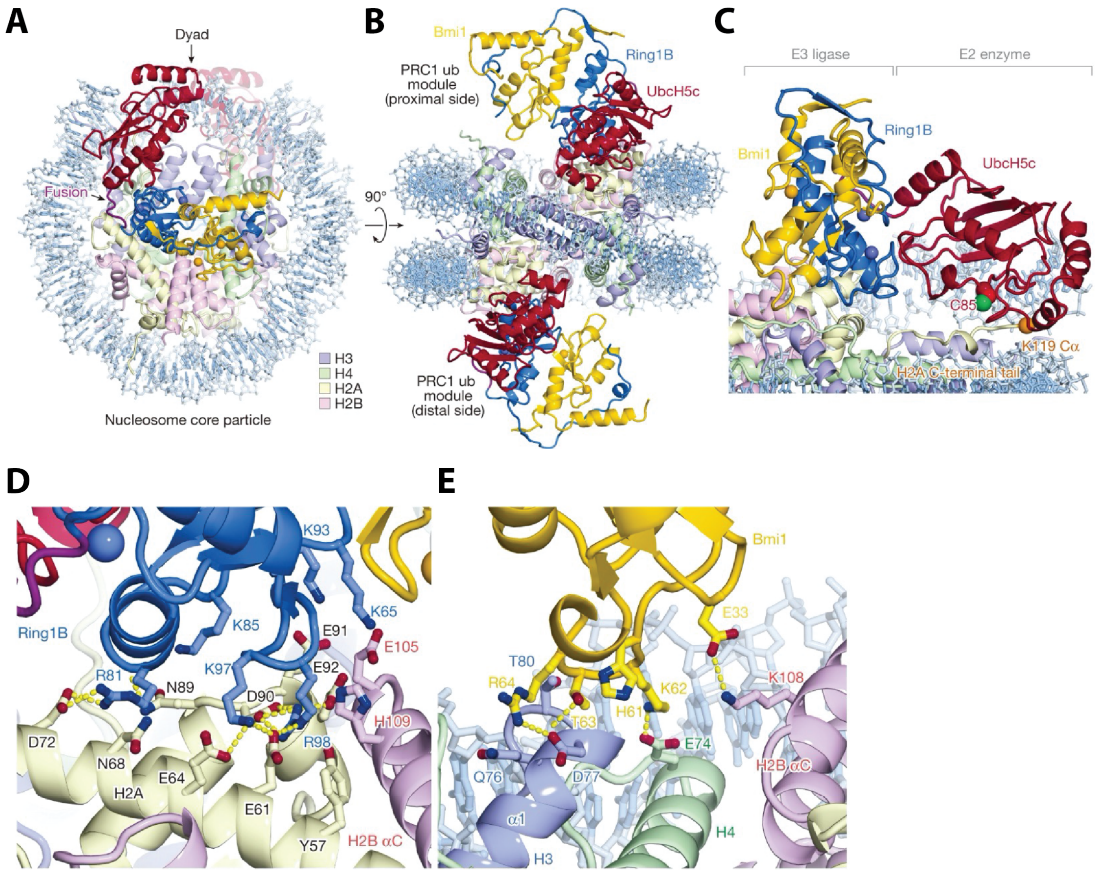


Figure 6: Crystal structure of the PRC1 E3 ligase module (Bmi1 1-109 in yellow and Ring1b 1-116 in blue) and E2 UbcH5c (in red) bound to the nucleosome (PDB ID: 4R8P, (McGinty et al., 2014)): Panels A (top view) and B (side view) illustrate the overall binding mode of PRC1-E3 and E2 to the nucleosome, with panel B highlighting proximal and distal binding sites of the complex. C. Close-up view showing the position of the UbcH5c catalytic site (C85, Cα shown in green) in proximity to H2AK119 (Ca shown in orange). D. Interaction of Ring1b with the nucleosome acidic patch, with key arginine hook residues (R81, K97, R98) indicated. E. Interaction of Bmi1 with histone H3, H4, and H2B, key interacting residues (R64, T63, H61, E33) are highlighted.

As illustrated in Figure 6A, the E3 ligase module comprising Ring1b 1-116 (depicted in blue) and Bmi1 1-109 (in yellow), along with UbcH5c (shown in red), adopts a crescent shape on the surface of the nucleosomal disk. The PRC1 E3-E2 complex binds uniformly to both the proximal and distal sides of the nucleosome, with a subtle variation in E2 binding: UbcH5c is in closer proximity to the nucleosome on the proximal side, compared to the UbcH5c bound to

the distal side, as shown in Figure 6B.

The Ring1b/Bmi1 RING domain heterodimer forms a saddle over the N-terminal end of the H2B αC helix, engaging with all four histones (Figure 6C). The nucleosome-interacting interface involves Ring1b Lys97 and Arg98, with Ring1b Arg98 residues engaging with an acidic pocket formed by H2A residues Glu61, Asp90, and Glu92 through charged hydrogen bonds (Figures 6C-D). This mode of nucleosomal recognition, characterized by an 'arginine-anchor' binding to the H2A/H2B acidic patch, is consistent across various chromatin factor-nucleosomes crystal structures. The Bmi1-nucleosome interface, though less extensive, involves hydrogen bonds with H4 Glu74 and H3 Asp77, as well as interactions with H2B Lys108 and H3 Gln76 (Figure 6E). The positioning of the E2 UbcH5c by the Ring1b/Bmi1 E3 ligase places its active site near the target Lys119 of H2A, with additional interactions occurring at non-active site surfaces with nucleosomal DNA.

At the time of writing this thesis, the first cryo-EM structure of PRC1 E3 ligase and E2 UbcH5c on H2AK119-ubiquitinated nucleosome had been published (Ai et al., 2023). In this study, E2 and ubiquitin were chemically linked, mimicking the E2-Ub transition state. E2-Ub was then chemically linked to H2A cysteine 119, which mimicked lysine 119. This modified histone H2A was used in a standard nucleosome reconstitution. For cryo-EM sample preparation, the PRC1 E3 ligase module was mixed either with H2AK119-UbcH5b-Ub nucleosomes or unmodified nucleosomes and both samples were cross-linked with glutaraldehyde. Notably, PRC1 E3 on unmodified nucleosome had a range of local resolution from 4 Å at the nucleosome-interaction interface to 6 Å at the top part of the Ring1b-Bmi1 E3 ligase module, showing more conformational dynamics compared to PRC1-E3/H2AK119-UbcH5b-Ub nucleosome, wherein the local resolution was more uniform (compare Figures 7A and 7B). This demonstrates that the presence of H2AK119-UbcH5b-Ub likely stabilises E3 ligase module of PRC1, which also results in an overall higher resolution of the PRC1-E3/H2AK119-UbcH5b-Ub-nucleosome map.

In both respective atomic models, the conformation of the PRC1 E3 ligase module closely resembles that of the PRC1 E3 ligase module in the crystal structure presented by McGinty et al. (2014). The overall root-mean-square deviation (RMSD) between the crystal structure without ubiquitin and the ubiquitin-containing cryo-EM model, calculated across all backbone atoms, was 0.685 Å, indicating that presence of ubiquitin did not induce any conformational changes in the PRC1 E3 ligase module. Furthermore, the interaction mode of PRC1-E3/H2AK119-

UbcH5b-Ub with the nucleosome was identical to the one in the crystal structure, which was described above (Figure 7C).

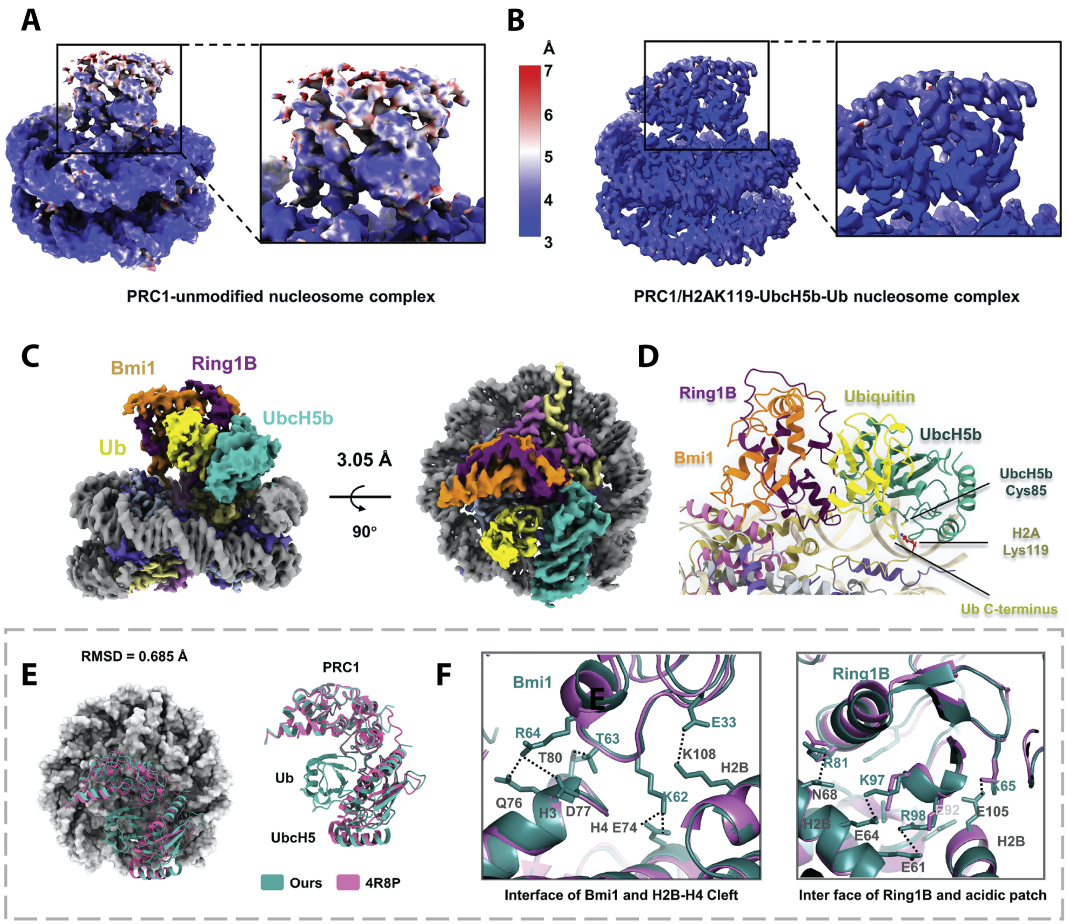


Figure 7: Cryo-EM reconstructions of PRC1 E3 ligase module on unmodified nucleosome and PRC1-H2AK119-UbcH5c on H2AC119-ubiquitinated nuclesome, reproduced from (Ai et al., 2023). A. Cryo-EM map of the PRC1-unmodified nucleosome complex colour-coded according to the local resolution (3A - blue, 6A - red), close-up shows the PRC1 E3 ligase module. B. Cryo-EM map of the PRC1-H2AK119-UbcH5c-Ub nucleosome complex colour-coded according to the local resolution (3A - blue, 6A - red), close-up view shows the PRC1 E3 ligase module. C. Side (left) and top (right) views of 3.05 A cryo-EM map of the PRC1 E3 ligase and H2AK119-UbcH5b-Ub nucleosome complex. D. The built atomic model of the PRC1 E3 ligase and H2AK119-UbcH5b-Ub nucleosome, close-up view of PRC1-E3 ligase, UbcH5c and ubiquitinated H2AC119. E. The structure of the PRC1 E3-E2 and nucleosome complex obtained in Ai et al. (2023) (highlighted in green) aligned to PDB ID: 4R8P (highlighted in pink), with a root mean square deviation (RMSD) of 0.685 Å, the position of PRC1 E3-E2 on nucleosome is shown on the left, alignment of PRC1 E3-E2 of Ai and 4R8P on the right. F. Comparison of binding interfaces of Bmi1 - nucleosome (left) and Ring1-nucleosome (right) in atomic models obtained in McGinty et al (4R8P) and Ai et al. (2023).

To summarize, while the crystal structure of McGinty et al. (2014) provided valuable insights into how the PRC1 complex interacts with the nucleosome, it only contained a fraction of the PRC1 complex, namely, the E3 ligase module. Despite demonstrating the interaction mode between this module and the nucleosome, other domains such as the RAWUL domains of Ring1b and Bmi1, as well as Cbx and Phc proteins, were missing in the structure. The recent cryo-EM reconstitution by Ai et al. (2023), featuring PRC1 E3 ligase and E2 UbcH5c on H2AK119C-ubiquitinated nucleosome, for the first time visualized the complex of PRC1 E3 ligase, E2 and ubiquitin on the substrate nucleosome. The localization of ubiquitin, presented in this reconstitution, had been, however, predicted earlier in McGinty et al. (2014), with the cryo-EM study by Ai et al. (2023) providing experimental validation for this prediction.

1.5.4 PRC1 as ubiquitin ligase: mechanism and function

Histone H2A is ubiquitinated by PRC1 at lysine 119 (K119) in mammals and lysine 118 (K118) in *Drosophila*. Although it was long known that a significant portion of histone H2A within chromatin is monoubiquitinated (5-15 %), only later two groups independently identified Ring1bA/B proteins as E3 ligases responsible for the deposition of this mark (Wang et al., 2004, de Napoles et al., 2004). Wang et al. (2004) conducted fractionation experiments on HeLa cells and employed mass spectrometry to determine that the fraction responsible for H2Aub1 production contained components of the PRC1 complex, including Ring1 (Ring1A), Ring2 (Ring1b), Bmi1 (PCGF4), and human polyhomeotic-like 2 (PHC2). In a complementary study by de Napoles et al. (2004), it was observed that Ring1b and H2Aub1 were enriched on the inactivated X chromosome. Furthermore, the knockout of Ring1b resulted in a global reduction of H2Aub1, which could be rescued by the expression of Ring1b in null cells. Once the role of PRC1 as an E3 ligase for H2Aub1 was established, its activity was also studied *in vitro*. It was showbn that although Ring1b is active on its own, its activity is greatly enhanced by association with Bmi1 (Buchwald et al., 2006).

In monoubiquitinated histone H2A, the ubiquitin moiety is attached by an isopeptide bond through C-terminal lysine 76 to an ε-amino group of K119 or K118. Generally, ubiquitination is a process orchestrated by a three-step cascade involving three key enzymes: the E1 Ub-activating enzyme, the E2 Ub-conjugating enzyme, and the E3 Ub-protein ligase (Pickart, 2001). In the initial step, E1 activates ubiquitin in an ATP-dependent manner, catalyzing the adenylation of the C-terminal carboxyl group and forming a thioester bond between the E1 cysteine residue

and ubiquitin C-terminus. Subsequently, ubiquitin moiety is transferred to E2, once again establishing a thioester bond with its cysteine. The final step, involving the transfer of ubiquitin to the substrate, is catalysed by the E3 ligase. This enzyme interacts with E2 and promotes the transfer of ubiquitin either directly from E2 or by initially forming an intermediate complex with ubiquitin.

There are several types of E3 ligases: U-box, HECT, RBR and RING finger types. PRC1 E3 ligase module that comprises Ring1A/B proteins and PCGFs belongs to the RING finger class of E3 ligases. This class is distinguished by the ability to directly transfer ubiquitin from an E2 to substrate (Deshaies and Joazeiro, 2009). PRC1 catalyses H2Aub1 ubiquitination on fully assembled nucleosomes and was shown to be inactive *in vitro* on H2A/H2B dimer (McGinty et al., 2014). This is due to the extensive mode of substrate recognition involving all four histones of a nucleosome as well as nucleosomal DNA, as was shown by the crystal structure of the PRC1 E3 ligase module fused with E2 on a nucleosome (discussed in detail in Section 1.5.3).

Regarding E2 enzyme that participates in H2Aub1 ubiquitination cascade, it is also essential to highlight that although most *in vitro* experiments and structural studies involving PRC1 employed UbcH5c as the E2 enzyme, this particular E2 is likely not the only E2 in the cascade *in vivo*. UbcH5c was chosen for *in vitro* studies based on small-scale testing of various E2 conjugating enzymes (Buchwald et al., 2006). Subsequent reports, however, indicated the potential involvement of another E2 enzyme, UbcH6, as its siRNA knockdown in human U2OS cell lines resulted in a global reduction of H2Aub1 levels, hinting at its role in the ubiquitination cascade (Wheaton et al., 2017).

The functional mechanisms related to H2Aub1 modification could be classified into two categories. First, H2Aub1 can be bound by several other Polycomb group proteins, serving as a platform for feedback loop mechanisms between Polycomb complexes or, potentially, other 'reader' proteins. The second group encompasses mechanisms where H2Aub1 directly influences chromatin and chromatin-associated processes.

Several readers of the H2Aub1 have been identified, among them, regulatory subunits of PRC2 complex - JARID2 and AEBP2 that form PRC2 variant complexes (Kalb et al., 2014, Cooper et al., 2014). Loss of H2Aub1 results in global reduction of H3K27me3, with H3K27me3 deposition and PRC2 occupancy being severely affected at PRC2 target loci (Blackledge et al., 2014, Tamburri et al., 2020). These findings highlight the presence of a feedback loop mecha-

nism between the E3 ligase of PRC1 and the H3K27me3 methyltransferase activity of PRC2. Furthermore, variant PRC1 subunits Rybp and YAF2 can bind H2Aub1 via their NZF domains. *In vitro*, Rybp/YAF2 were shown to facilitate a positive feedback mechanism by binding H2Aub and catalyzing H2A ubiquitination at neighbouring nucleosome (Zhao et al., 2020).

The role of H2Aub1 in chromatin-associated processes has been investigated in numerous studies. Earlier research has suggested that H2Aub1 or PRC1 itself may inhibit RNA Polymerase II activity by hindering transcription initiation or precluding formation of the pre-initiation complex (Stock et al., 2007, Lehmann et al., 2012). A recent study in mESC further demonstrated that conditional removal of Ring1B/Ring1A and subsequent depletion of H2Aub1 resulted in an increased transcriptional burst frequency, without affecting the number of transcripts produced during each burst (Dobrinić et al., 2021).

While mESCs continue to serve as a prevalent cell line for investigating Polycomb repression, it is worth noting that the observations made in mESCs may be inherently tied to the context of pluripotent transcriptional programs. It is therefore essential to investigate Polycomb repression mechanisms at the organismal level. Studying the role of an epigenetic mark in *vivo* usually involves assessing phenotypic changes and the overall viability of an organism upon perturbation of the mark's 'writer' enzyme. In the context of the PRC1 catalytic subunit Ring1b, its function extends beyond catalyzing H2Aub1, as it serves as a core component in the formation of all PRC1 assemblies, which harbour additional non-enzymatic functions. Consequently, a comprehensive study should ideally probe both the enzymatic and structural roles of Ring1b in PRC1 assemblies. Accordingly, in the study conducted by Pengelly et al. (2013), *Drosophila* homolog Sce contained I48A mutation, which abolished its interaction with E2, rendering Sce inactive as E3 ligase. Abolishing the catalytic function of Sce led to 98 % reduction of global H2Aub1 levels in mutant animals, however, depletion of H2Aub1 did not lead to Polycomb homeotic transformations and misexpression of Polycomb target genes. The embryos lacking catalytically active Sce still showed embryonic lethality, rendering H2Aub1 essential for viability, but not for repression of Polycomb target genes. Moreover, Polycomb repression of target genes was also maintained in mutants where H2A and H2Av histones could not be ubiquitylated due to mutations at target lysines (K117R, K118R, K121R, K122R).

A recent study by Bonnet et al. (2022) has provided some crucial insights on the role of H2Aub1 in Polycomb repression, showing that not solely the presence of H2Aub, but rather

the balance between the enzymatic activities of PRC1 E3 ligase and PR-DUB deubiquitinase is pivotal for Polycomb repression. In this study, disruption of PR-DUB deubiquitinase activity in *Drosophila* embryos led to an overabundance of H2Aub1 and subsequent derepression of Hox genes. Remarkably, embryos lacking PR-DUB (Asx KO mutant) and possessing catalytically inactive PRC1 (Sce I48A mutant) displayed no discernible phenotypic differences from wild-type embryos. Even in PR-DUB KO embryos with elevated H2Aub1 levels, the misexpression of HOX genes could not be attributed to global transcriptional dysregulation, with RNA PolII occupancy also being unaltered compared to wild type. Similarly, the absence of H2Aub1 in Sce I48A mutants did not disrupt the transcriptional regulation of target Polycomb genes, which is in line with previous work by Pengelly et al. (2013). Additionally, *in vitro* experiments showed that H2Aub1 impedes nucleosome stacking in chromatin fibres. This finding was corroborated by *in vivo* evidence from genome-wide ATAC-seq analyses, demonstrating increased DNA accessibility in catalytically inactive PR-DUB mutants with an excess of H2Aub1. When these findings are considered in the broader context of Polycomb repression, it is plausible that H2Aub1 excess may counteract the non-enzymatic functions of PRC1, known to compact chromatin. This notion finds support in the exacerbated Polycomb phenotype observed in double mutants lacking *Asx* (leading to H2Aub1 excess) and one allele of *Polycomb* gene (which diminishes PRC1 recruitment to chromatin via H3K27me3). In summary, this study offered a compelling mechanistic insight into the role of H2Aub in Polycomb repression in *Drosophila*. However, whether these findings equally apply to Polycomb repression in mammals remains to be elucidated.

1.5.5 Non-enzymatic function of PRC1

As was briefly discussed in the previous section, a great challenge in the field of Polycomb research is dissecting the role of enzymatic PRC1 function from non-enzymatic one. The concept of non-enzymatic functionality of PRC1 emerged from *in vitro* studies of PRC1 in the laboratory of Robert Kingston. In particular, it was shown that *Drosophila* nuclear fraction, which contained PRC1 subunits PH, Psc, Pc and Scm, blocked nucleosomal array remodeling by SWI/SNF remodeler (Shao et al., 1999). Later, the same group imaged nucleosomal arrays using electron microscopy and observed their condensation into more compact structures in the presence of PRC1 (Francis et al., 2004). Also, this study determined that the compaction activity resided in the Psc subunit of PRC1, wherein the truncation of the C-terminal tail of Psc led to less compacted arrays. Later, the importance of the C-terminal tail of Psc was also shown *in vivo*

in *Drosophila* embryos, wherein the severity of the Polycomb phenotype was correlating with shorter truncations of Psc C-terminus (King et al., 2005). Initially, the C-terminus of Psc was found not to be conserved in mammalian homologs of Psc - PCGF2 and PCGF4 (Mel18 and Bmi1) and therefore its function was implied to be specific to *Drosophila* PRC1. Later, however, it was found that Cbx2, one of the mammalian homologs of Pc, contained the region with high homology to Psc C-terminus. The region was first termed intrinsically disordered region (IDR) due to the lack of structural features predicted. Later, this region was named Compaction and Phase Separation Domain (CaPS). A stretch of positively charged amino acids within CaPS was implied to be responsible for electrostatic interactions leading to chromatin compaction. Mutating these positively charged residues within Cbx2 CaPS and, therefore, abolishing the compaction in mice led to viable offspring with minor vertebra perturbations (Lau et al., 2017). This highlights the fact that Cbx2 compaction activity, albeit important for proper body patterning, is not the universal mechanism for Polycomb repression and most likely depends on developmental contexts in which Cbx2 is expressed. Cxb2 CaPS, however, seems to be also important in spermatogenesis, as it is overexpressed in spermatogonial stem cells, wherein it was shown to contribute to the repression of stem cell genes during linage commitment (Kim et al., 2023).

Later, when the notion of protein liquid-liquid phase separation started to take shape in the field of biology, it was shown that Cbx2 CaSP could, similarly to HP1 protein, give rise to phase-separated condensates (Plys et al., 2019, Tatavosian et al., 2019). Mutations replacing lysines with alanines in the CaSP region were found to disrupt the formation of these phase-separated puncta both *in vitro* and in cells (Plys et al., 2019, Tatavosian et al., 2019). Notably, deletion of the Cbx2 H3K27me3-binding Chromodomain did not diminish condensate formation in cells, indicating independence from H3K27me3. Similarly, the removal of Ring1A/Ring1B and PCGF2/PCGF4 subunits did not impede condensate formation by Cbx2, highlighting as well the independence of condensation formation from other PRC1 subunits (Tatavosian et al., 2019).

Research on Cbx2 CaSP was not the first time condensates were implied to play a role in Polycomb repression. In fact, microscopic imaging coupled with immunostaining of different Polycomb proteins in a cell showed their localization within discrete puncta named Polycomb (PcG) bodies (Saurin et al., 1998). Furthermore, it was demonstrated that the size and quantity of PcG bodies vary across different cell types (Ren et al., 2008). One possible explanation for

the existence of PcG bodies was that their occurrence was merely the consequence of Polycomb target genes often existing in multi-gene clusters, concentrating Polycomb proteins. However, it was shown that genes from remotely located clusters, such as the Antennapedia and bithorax complexes, which are located 10 Mb away from each other, still co-localize within PcG bodies (Bantignies et al., 2011). The clustering of PcG bodies was also shown to be dependent on SAM-SAM polymerization of the Ph protein. Upon expression of polymerization-deficient mutants Ph-EH and Ph-ML, the average size of PcG bodies in *Drosophila* S2 cells was significantly reduced, coinciding with the derepression of Polycomb target genes (Wani et al., 2016). Recently, the high-resolution Hi-C contact maps have also detected chromatin loops within individual PcG domains, with PRC1 being positioned at 'loop anchors', and Ph SAM-SAM polymerization was therefore implied to play a role in this PRC1-associated chromatin loop formation (Eagen et al., 2017).

One question that arises is whether PRC1-induced phase separation and chromatin compaction are essentially the same phenomenon or occur sequentially. In a recent study, researchers employed a light-activated oligomerisation platform, allowing the toggling of tagged protein oligomerisation on and off within cells. Combining this experimental approach with molecular simulations, it was shown that Cbx2-driven phase-separated condensates form and dissolve more rapidly upon signal termination. In contrast, chromatin compaction appears to be a more gradual process, persisting for up to 30 minutes after the dissolution of the condensates, which shows that these two processes might happen on different time scales (Eeftens et al., 2021).

A critical question persists: what are the functional implications and potential mechanisms underlying the non-enzymatic roles of PRC1, particularly the chromatin-compacting CaSP regions of Psc/Cbx2 and the SAM-SAM polymerization of Ph? The current hypothesis in the field suggests that PRC1-induced chromatin organization may contribute to transcriptional control by either shielding Polycomb target genes from RNA Polymerase II and transcription factors or by limiting interactions between silenced genes and enhancers. With the continual advancements in high-resolution imaging techniques and chromatin-based whole genome methods, unravelling these questions will only be a matter of time.

1.6 Role of PRC1 in development and disease

Embryonic development

Polycomb group proteins play a crucial role in orchestrating precise gene regulation processes during both embryonic development and the differentiation of adult stem cells. During embryonic development, precise transcriptional regulation is needed for proper anterior-posterior body patterning of the embryo. Both Polycomb group proteins and their antagonists Trithorax group proteins must act together to regulate the spatial-temporal order of transcription in developing animals, as illustrated schematically in Figure 8. Moreover, disruptions in the function of Polycomb group proteins, often arising from mutations, are frequently implicated in various malignancies, as will be discussed below.

The role of canonical PRC1 in embryonic development first became evident from the pioneering studies of homeobox gene regulation in *Drosophila melanogaster*, where the loss of Pc and Ph PRC1 subunits led to disruption of body patterning along the anterior-posterior axis in embryos (Lewis, 1978, Smouse et al., 1988, Moazed and O'Farrell, 1992). Later, the loss of both Polyhomeotic paralogs (Ph-p and Ph-d) and Polycomb (Psc, along with its paralog Su(z)2) has been demonstrated to lead to homeotic transformations and early developmental arrest (Gutiérrez et al., 2012, Smouse et al., 1988). Interestingly, different outcomes were observed in the case of two types of Sce mutants: complete Sce knockout (Sce KO) and Sce catalytic mutant (Sce I48A). In the absence of maternally loaded Sce, embryos lacking Sce exhibited extensive homeotic transformation and early developmental arrest. However, Sce I48A embryos showed no misexpression of HOX genes and appeared indistinguishably from wild-type embryos, but arrested development at the end of embryogenesis (Pengelly et al., 2015). As was discussed above, in mammals, PRC1 complexes are composed of different subunit paralogs, that are expressed depending on the cellular and developmental contexts. Figure 9 illustrates the mRNA expression landscape of PRC1 and PRC2 subunits across different tissues and developmental stages. This differential expression translates into different phenotypic outcomes of PRC1 subunit knock-outs, as shown in Table 1.

In vivo knock-outs of different PRC1 paralogs have been reported in mice models. Not surprisingly, the absence of the Sce paralog Ring1b, the central subunit of PRC1 assemblies, leads to a halt in developmental progression during gastrulation (Leeb and Wutz, 2007, Voncken et al., 2003). In line with previous studies in *Drosophila*, the Ring1b catalytic mutant was found

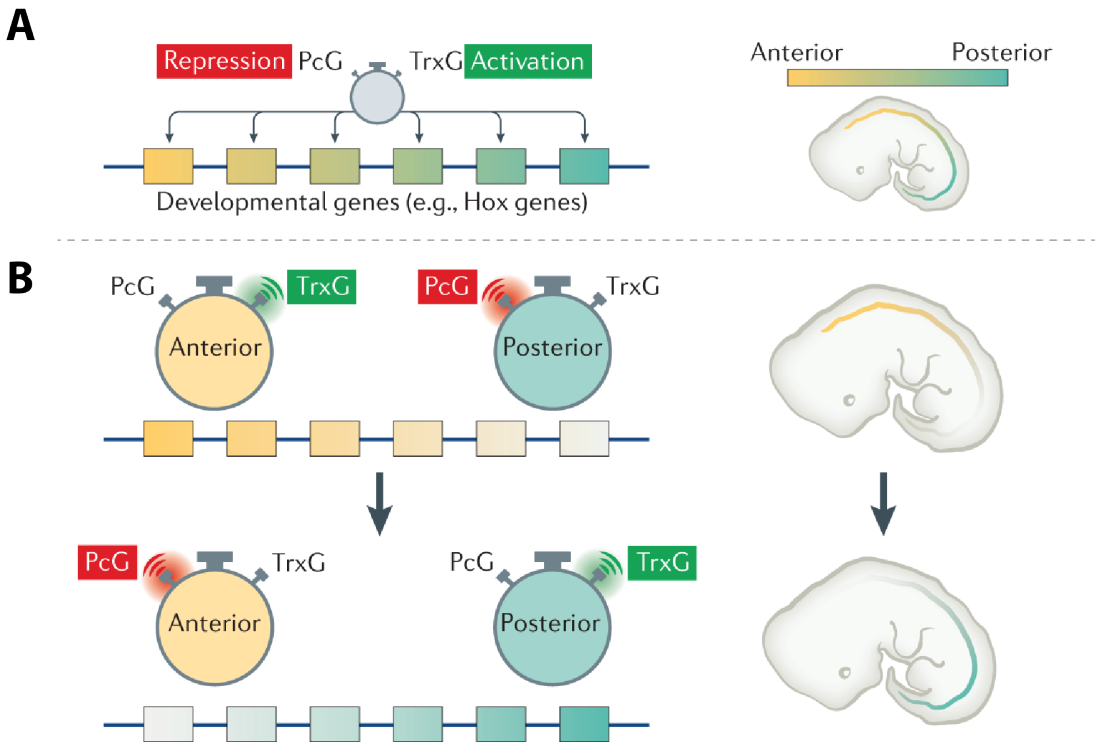


Figure 8: Polycomb and Trithorax regulation of gene expression in development. A. The schematic illustration of the regulatory roles of Polycomb group (PcG) and Trithorax group (TrxG) proteins in controlling gene expression during development. PcG and TrxG act as regulators of the anterior to posterior gene expression gradients during embryonic development (shown as a gradient from yellow to blue; top panel). The precise spatiotemporal order of developmental gene expression is maintained through the antagonistic activities of PcG proteins (repression) and TrxG proteins (activation). B. Timers illustrate the correct developmental timeline for the expression of genes that specify anterior and posterior patterning. On the right, the embryo is first depicted at a stage where anterior-specifying genes are activated by TrxG proteins while posterior-specifying genes are repressed by PcG proteins. In the second panel, representing a later developmental stage, posterior-specifying genes need activation by TrxG proteins, whereas anterior-specifying genes require repression by PcG proteins. Disruptions in this finely tuned regulatory system, such as mutations in PcG and TrxG genes, lead to homeotic phenotypes. Reproduced from Piunti and Shilatifard (2021).

to be embryonic lethal, however, without homeotic transformations (Illingworth et al., 2015). Conversely, mutants deficient in the second Sce paralog, Ring1a, survive and reach birth, albeit exhibiting deficiencies in the transformation of the axial skeleton and showing altered levels of Hox gene expression (Lorente et al., 2000).

The loss of another canonical PRC1 subunit, PCGF2, has been reported to result in viable birth but postnatal lethality, posterior transformation and Hox genes deregulation (Akasaka et al., 1996). The loss of its paralog PCGF4 (also referred to as Bmi1) is perinatal and postnatal lethal (van der Lugt et al., 1996). Double knock-out of PCFG2 and PCGF4, representing the complete loss of canonical PRC1, is embryonically lethal with more severe homeotic transformations (Akasaka et al., 2001). However, it is important to mention, that PCGF2 and PCGF4 can also form some non-canonical PRC1 complexes, including those involving Rybp and YAF2 subunits. Consequently, the observed phenotype in a double PCFG2-PCGF4 knockout might as well be a cumulative outcome of the loss of all canonical PRC1 subcomplexes and certain non-canonical PRC1 subcomplexes.

When knocked out individually, the Polyhomeotic paralogs PHC1 and PHC2 both exhibited homeotic transformations, with PHC1 leading to perinatal lethality, while PHC2 knockout resulted in a viable birth (Takihara et al., 1997, Isono et al., 2005). Notably, both phenotypes were less severe than Ph KO in *Drosophila*. However, the embryonically lethal phenotype observed in the double knockout of PHC1 and PHC2 implies that these proteins have overlapping functions, suggesting a possibility of partial functional compensation in the single knockouts (Isono et al., 2005). As both PHC1 and PHC2 subunits exist solely in canonical PRC1 complexes, this study offered crucial evidence of canonical PRC1 being indispensable for embryonic development.

Drosophila Polycomb subunit has multiple paralogs in mammals: Cbx2, Cbx4, Cbx6, Cbx7, and Cbx8, most of them exhibiting tissue-specific expression. The loss of Cbx2 protein results in either perinatal or postnatal lethality and homeotic transformation. As mentioned above, Lau et al. (2017) also demonstrated that mutations in the low-complexity region of Cbx2 (mutating positively-charged lysines within this region to alanines) resulted in viable offsprings, however, with polycomb phenotype in the form of transformed identity of C7, T1 and T7 vertebrae, which showcases of the functional importance of this region for proper body patterning during development (Lau et al., 2017). The loss of Cbx4 protein was also shown to result in perinatal lethality, however, without associated Polycomb phenotype (Liu et al., 2013). Cbx7 is predominantly

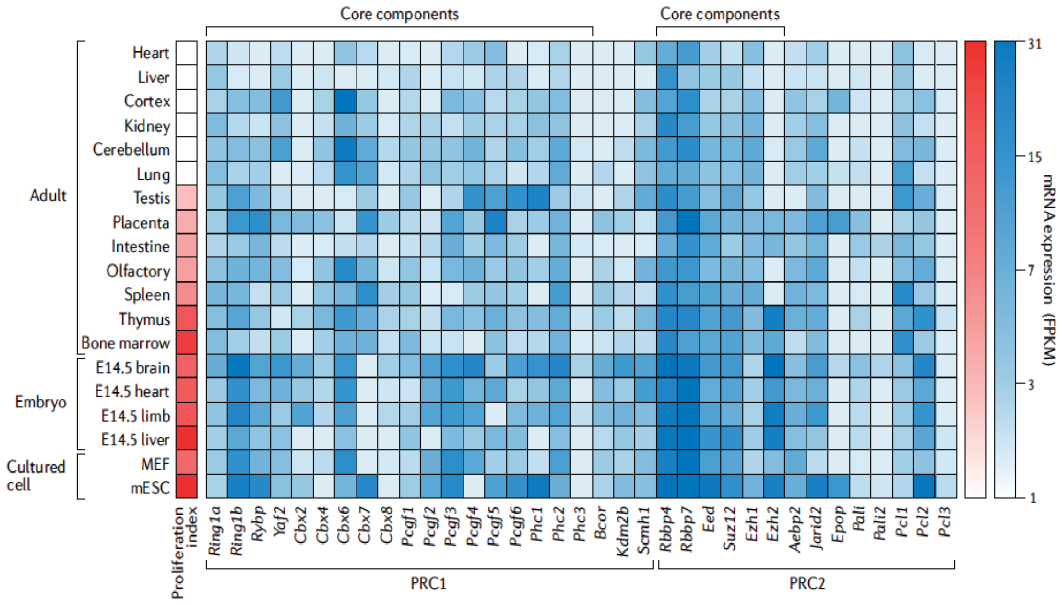


Figure 9: Cell type-specific expression of PRC1 and PRC1 subunits. Heat map showing mRNA expression of different PcG genes in the mouse embryo and adult tissues, as well as in cultured cell types, such as mouse embryonic fibroblast (MEFs) and mES cells. Note that the mRNAs of the core components of PRC1 and PRC2 complexes (Ring1b and Eed), demonstrate consistent expression across various tissues. Conversely, accessory components display greater variability in expression, often showing tissue-specific patterns. The proliferation index shows the degree of cell proliferation within tissues, calculated based on the mRNA expression levels of cell cycle genes. E - embryonic day; ErP - erythroid progenitor; HSC - haematopoietic stem cell; RBC - red blood cell; XCI -X chromosome inactivation. Reproduced from Kim and Kingston (2022)

.

expressed in embryonic stem cell lines, however, contrary to expectations, Cbx7 KO mice were shown to be viable, albeit susceptible to malignancies (Forzati et al., 2012). The relatively mild phenotypes observed in individual Cbx knockouts may be attributed to the functional redundancy among Cbx paralogs. This notion is supported by the observation that depletion of Cbx7 in embryonic stem cells results in an upregulation of Cbx4 and Cbx8, suggesting an involvement of a compensatory mechanism (Morey et al., 2012).

Table 1: List of mouse knockout and mutagenesis studies of PRC1 complex subunit genes with their corresponding phenotypes.

Gene	***In vivo* phenotype**	**Effect on Hox genes**	**Reference**
mRing1a KO	Viable	Homeotic transformation	Lorente et al. (2000)
mRing1b KO	Embryonic lethal	No homeotic transformation	Voncken et al. (2003)
mRing1b catalytic mutant	Embryonic lethal	No homeotic transformations	Illingworth et al. (2015)
mPCGF2 KO	Postnatal lethal	Homeotic transformation	Akasaka et al. (1996)
mPCGF4 KO	Perinatal and postnatal lethal	Homeotic transformation	Van der Lugt et al. (1994)
mPCGF2 mPCGF4 double KO	Embryonic lethal	Homeotic transformation	Akasaka et al. (2001)
mCBX2 KO	Perinatal and postnatal lethal	Homeotic transformation	Coré et al. (1997)
mCBX4 KO	Perinatal lethal	No homeotic transformations	Liu et al. (2013)
mCBX7 KO	Susceptible to tumor formation	Unknown	Forzati et al. (2012)
mPHC1	Perinatal lethal	Homeotic transformation	Takihara et al. (1997)
mPHC2	Viable	Homeotic transformation	Isono et al. (2005)
mPHC1 and mPHC2 double KO	Embryonic lethal	Homeotic transformation	Isono et al. (2005)
mScmh1	Viable and fertile	No homeotic transformation	Yasunaga et al. (2013)
mScml2	Viable and fertility defects	Unknown	Luo et al. (2015)
mRybp KO	Embryonic lethal	Unknown	Pirity et al. (2005)

PRC1 in malignancies

Polycomb group proteins have been linked to various types of malignancies. Additionally, PcG proteins have been implicated in modulating the immune response, regulating the metabolism of the tumour microenvironment, and promoting metastasis. Another prospective area of research focuses on the interaction of Polycomb group proteins with oncohistones and the mechanisms of epigenetically modulated anti-cancer drug resistance development (reviewed in detail in Parreno et al. (2022)).

Most of the studies regarding the involvement of the Polycomb repression system in carcinogenesis are centred around PRC2, however, several PRC1 subunits have been also implicated in cancer. Large-scale genomic studies of cancer patients have shown, that PRC1 genes are overexpressed in several cancer types. In particular, PRC1 overexpression was predominantly found in hormone-related cancers, with Ring1b being amplified in 22 % of breast cancers. Cbx2, Cbx4, Cbx8 and PCGF2 were also found enriched in some breast-cancer subtypes (Chan et al., 2018).

Another canonical PRC1 subunit, Bmi1 protein, was one of the first PRC1 subunits to be recognized as a general proto-oncogene. Overexpression of Bmi1 was shown to suppress the expression of p16 Ink4 and p19Arf tumour suppressors, promoting cell proliferation (Jacobs et al., 1999). This mechanism has been implicated in various cancers, including breast, ovarian, gastric, and pancreatic malignancies (Li et al., 2010, Althobiti et al., 2020, Wang et al., 2016).

1.7 Aims of the thesis

This study aims to reconstitute canonical Polycomb repressive complex 1 (PRC1) *in vitro* for structural and functional analysis, with a particular focus on three specific PRC1 subcomplexes containing PHC2, Cbx7, and Scml2 subunits. Through a methodological approach combining cryo-electron microscopy, X-ray crystallography, cross-linking mass spectrometry, and biochemical assays this study aims to elucidate the molecular basis of PRC1 interaction with nucleosomes and provide insights into the mechanisms underlying PRC1's role in epigenetic regulation.

2 Materials

2.1 Bacterial and insect cell strains

Table 2: Bacterial and insect cell strains used in this study.

Strain name	Organism	Source	Use
Top 10 F'	*E. coli*	Invitrogen	Cloning
BL21(DE3) Gold pLysS	*E. coli*	EMBL Heidelberg	Protein expression
BL21(DE3) Gold pLysS	*E. coli*	EMBL Heidelberg	Protein expression
DH10EMBacY	*E.coli*	Geneva Biotech	Bacmid preparation
IPLB-Sf21	*Spodoptera frugiperda*	Invitrogen	Baculovirus cultivation
BTI-TN-5B1-4 High Five	*Trichoplusia ni*	Invitrogen	Protein expression

2.2 Media

Table 3: List of media used in this study.

Media	Supplemented with	Use	Source
Lysogeny broth (LB)	-	Cloning (*E.coli*)	in-house
Terrific Broth (TB)	10% Phosphate buffer	Protein expression (*E.coli*)	in-house
Super Optimal broth with Catabolite repression (SOC)	-	Cloning (*E.coli*)	in-house
EX-CELL TiterHigh	-	Baculovirus cultivation (Sf21 cells)	Sigma Aldrich
Express Five	L-Glutamine	Protein expression (Hi5 cells)	Gibco

2.3 cDNA

Table 4: List of cDNA sequences used in this study.

cDNA	Species	Source
Ring1b	*Homo sapiens*	MPI Core Facility cDNA library
Bmi1 (PCGF4)	*Homo sapiens*	MPI Core Facility cDNA library
Phc2	*Homo sapiens*	MPI Core Facility cDNA library
Cbx7 158 AA isoform	*Homo sapiens*	GeneArt Gene Synthesis (Thermo Fischer Scientific)
Scml2	*Homo sapiens*	MPI Core Facility cDNA library

2.4 Primers

Table 5: Oligonucleotides used in this study.

Name	Sequence 5'-3'	5' modif.	Application
EK1	ctggaagttctgttccaggggcccgactcg atgtctcaggctgtgcaga	-	cloning of Ring1b 1-133 into pET vector
EK3	accaggaacaaaccggcggccgctcgatgc atcgaacaacgagaatcaa	-	cloning of Bmi1 1-242 into pET vector
EK4	ggatcctgcaaagcaccggcctcgttatcc atctctctggtgactgatc	-	cloning of Bmi1 1-242 into pET vector
EK13	aaggagatatacataatgacctcagggaac ggaaac	-	cloning of PHC2 536-678 into pET vector
EK14	cagaccgccaccgactgcttatcccacccgt	-	cloning of PHC2 536-678 into pET vector
EK18	ttcagaccgccaccgactgcttacttgttg atcct	-	cloning of Ring1b 1-133 into pET vector
EK150	ccaccatcgggcgcggatccaatgcatcga acaacgagaatc	-	cloning of Bmi1 into pLIB vector for pBIG1a generation
EK152	tcctctagtacttctcgacaagcttttaat ggtgatggtgatggtgctgtttcattccaccag	-	cloning of Bmi1 into pLIB vector for pBIG1a generation
EK153	ccaccatcgggcgcggatccaatgtggagt catccacaattcgaaaaaggtggaggttccgg	-	cloning of Ring1b into pLIB vector for pBIG1a generation
EK154	tcctctagtacttctcgacaagcttttatt tgtgctccttttgtaggtg	-	cloning of Ring1b into pLIB vector for pBIG1a generation
EK160	ccaccatcgggcgcggatccaatgacctca gggaacggaaactct	-	cloning of PHC2 536-678 into pFastBac vector; cloning PHC2 isf2 into pFastBac vector
EK161	cctctagtacttctcgacaagcttctagga gtccttgagcatgctgatgc	-	cloning PHC2 isf2 into pFastBac vector
EK164	ttagggctcacttggcaggaagtggtgtcc catg	-	cloning of PHC2 536-678 into pFastBac vector
EK162	atgaacatcaaggaggggcccgccgagaag atctatgccc	5'-P	mutating EH surface of PHC2 SAM domain
EK163	cgagtcgggcccctggaacagaacttccag ttttccacc	5'-P	mutating EH surface of PHC2 SAM domain
EK166	tccaggggcccgactcgatgggacaaacag tgaatgaagattcc	-	cloning of Scml2 into pFastBac vector
EK167	tagtacttctcgacttatttaactgtattt tccttctttaagcttttcaatgtagtaac	-	cloning of Scml2 into pFastBac vector
EK174	cttcagccccatttccttcatcatcacatc actcttgagacggaacagagccttccca	5'-P	mutating ML surface of Scml2 SAM domain
EK175	ctggggccagcattaaagctgtgttactac attgaaaagcttaaagaa	5'-P	mutating ML surface of Scml2 SAM domain
KF07-FAM	atatctcgggcttatgtgatggac	6-FAM	generation of 215 bp 5' fluorescein-labeled nucleosomal DNA
KF07-Bio	atatctcgggcttatgtgatggac	Biotin	generation of 215 bp 5' biotinylated nucleosomal DNA
KF08	tatcccgagtcgctgttcaatac	-	generation of 215 bp 5'-labelled nucleosomal DNA

2.5 Plasmids

Table 6: List of plasmids used for protein expression and nucleosomal DNA preparation.

Constuct	Vector	Resistance	Tag	Protease	Species	Source
His-GST-Ring1b 1-133	pEC-A-3C-GST	Amp	6xHis, GST	3C	*H.s*	this study
His-GB1-Bmi1 1-242	pEC-K-3C-GB1	Kan	6xHis, GB1	3C	*H.s*	this study
His-PHC2 536-678	pEC-S-CDF-3C-HIS	Spec	6xHis	3C	*H.s*	this study
TS-Ring1b Bmi1-His	pBig1a	Amp, Spec	Twin-Strep, 6xHis	3C	*H.s*	this study
His-PHC2-EH	pFastBac1	Amp, Gent	6xHis	3C	*H.s*	this study
His-PHC2ΔSAM	pFastBac1	Amp, Gent	6xHis	3C	*H.s*	this study
His-TS-Cbx7	pFastBac1	Amp, Gent	6xHis, Twin-Strep	3C	*H.s*	this study
His-TS-Scml2	pFastBac1	Amp, Gent	6xHis, Twin-Strep	3C	*H.s*	this study
UbcH5c	pGEX6P-1	Amp	GST	3C	*H.s*	Buchwald et al. (2006)
Ring1b Bmi1 minimal	pGEX6P-1	Amp	GST	3C	*M.m*	Buchwald et al. (2006)
p601	pUC19	pUC19	-	-	-	Lowary and Widom (1998)

2.6 Buffers

Table 7: List of buffers used in this study

Name	Composition	Use
4 x SDS sample buffer	200mM Tris pH 6.8, 20% 2-Mercaptoethanol, 8% SDS, 0.4% Bromophenol Blue, 40% Glycerol	SDS-PAGE
10 x SDS PAGE Running Buffer	250 mM Tris-HCl pH 8.3, 1.92 mM glycine, 1% SDS	SDS-PAGE
Phosphate-buffered saline (PBS)	10 mM Na2HPO4, 1.8 mM KH2PO4, 137 mM NaCl, 2.7 mM KCl	Insect cells harvesting
1 x TBE	89 mM Tris, 89 mM boric acid, 2 mM EDTA	Agarose gel electrophoresis
10x Phosphate buffer	170 mM KH2PO4, 720 mM K2HPO4	TB media supplement

2.7 Antibiotics

Table 8: List of antibiotics used in this study.

Antibiotic	Stock concentration	Final concentration
Ampicillin	100 mg/ml	100 μg/ml
Kanamycin	50 mg/ml	50 ug/ml
Spectinomycin	100 mg/ml	100 μg/ml
Tetracycline	10 mg/ml	10 μg/ml
Gentamicin	10 mg/ml	7 μg/ml
Chloramphenicol	34 mg/ml	34 μg/ml

2.8 Software

Table 9: List of software programs used in the study.

Program	**Developer**
UNICORN	GE Healthcare
Fiji/ImageJ2	Schindelin et al. (2012)
SnapGene	GSL Biotech LLC
GraphPad PRISM	GraphPad Software
EPU	FEI (ThermoFisher Scientific)
SerialEM	Mastronarde (2005)
cryoSPARC v3.3.2	Punjani et al. (2017)
RELION 3.1	Scheres (2012)
MotionCor2	Zheng et al. (2017)
UCSF Chimera	Pettersen et al. (2004)
AlphaFold2	Jumper et al. (2021)
xiNET	Combe et al. (2015)
XMAS	Lagerwaard et al. (2022)

3 Methods

3.1 Cloning and related techniques

3.1.1 Polymerase chain reaction (PCR)

DNA inserts for cloning were amplified from plasmids containing cDNA sequences of interest (listed in Table 4, see Materials section 2.3) by polymerase chain reaction (PCR) using Q5 High-Fidelity polymerase (NEB). Typically, a 50 μl reaction was prepared with 25 μl 2x Q5 High-Fidelity polymerase Master Mix (NEB), 2.5 μl of each 100 μM primer, 10 ng of template plasmid and water to a final volume of 50 μl. The thermocycling conditions are listed in Table 10.

Table 10: Conditions for PCR reaction.

Step	**Temperature**	**Time**
Initial denaturation	98 °C	30 s
25-35x amplification cycles	98 °C 50-72 °C 72 °C	10 s 30 s 30 s per kb
Final extension	72 °C	2 min
Hold	10 °C	infinity

The PCR reaction samples were then mixed with 6x Gel Loading Dye (NEB) and separated using agarose gel electrophoresis (Method section 3.1.2). The DNA bands corresponding to the target amplified DNA were cut off and extracted from the gel using the QIAquick Gel Extraction Kit, following the manufacturer's instructions (QIAGEN).

For large-scale nucleosomal DNA amplification, the PCR reaction was mixed using Pfu polymerase (MPIB Core Facility), as summarized in Table 11. The thermocycling conditions used for nucleosomal DNA PCR amplification were as in Table 10, with an annealing temperature of 57°C and a final extension phase of 10 min.

Table 11: Reaction mix for PCR amplification of nucleosomal DNA.

PCR reaction mix (100 μl)	
10x PCR Reaction buffer (MPIB Core Facility)	10 μl
dNTP's (10 mM)	2 μl
Forward primer (100 μM)	5 μl
Reverse primer (100 μM)	5 μl
Pfu polymerase 5 U/μl (MPIB Core Facility)	1 μl
Template plasmid (60 ng/μl)	1 μl
H2O	76 μl

3.1.2 Agarose gel electrophoresis

To cast the gels, the agarose powder (Sigma-Aldrich) was dissolved in 1x TBE buffer to a final concentration of 0.8-1.2 % (the concentration was chosen depending on fragment size and degree of separation). SybrSafe dye (Invitrogen) was added to the agarose mixtures in dilution of 1:10000 for DNA visualisation. GeneRuler 1 kb Plus or 100 bp DNA Ladders (Thermo Fisher Scientific) were used as markers. Gel electrophoresis was typically performed at 120 V for 20-25 min and the gel was visualized using a blue light transilluminator emitting blue light at 470 nm (BioRad). Target DNA bands were cut and extracted from gels using the QIAquick Gel Extraction Kit, following the manufacturer's instructions (QIAGEN).

3.1.3 Ligation independent cloning (LIC)

All plasmids for bacterial expression were generated using ligation-independent cloning (LIC). This method relies on generating cohesive ends using 3'-5' endonuclease activity of T4 polymerase. DNA inserts were PCR-amplified (Method section 3.1.1) with primers containing overhangs complementary to vectors. Resulting PCR products were isolated from gel using a QIAquick Gel Extraction Kit, following the manufacturer's instructions (QIAGEN). Reaction mixes for insert processing were prepared as in the table below (Table 12). Reactions were incubated for 30 min at RT, and the enzymes were then inactivated by incubation at 75°C for 20 min. 1 µl of pre-processed vector and 2 ul of the pre-processed insert were mixed together and incubated for 10 min at RT, followed by the addition of 1 µl of 25 mM EDTA and 10 min incubation. 2 µl of annealing mix were transformed to electrocompetent TOP10 *E.coli* cells and grown on an antibiotic selective agar plate at 37 °C overnight.

Table 12: Reaction mix for insert processing in ligation independent cloning (LIC).

Reaction mix	
Gel purified PCR product	600 ng
4 DNA Pol. buffer (10x)	2 µl
dATP (25 mM)	2 µl
DTT (100 mM)	1 µl
T4 DNA Pol. LIC qualified (Novagen)	0.4 µ
H2O	to 20 µl

3.1.4 Gibson assembly and InPhusion cloning

All pFastBac and pLIB plasmids were cloned using either Gibson assembly or InPhusion cloning methods (Gibson et al., 2009). For the InPhusion approach, inserts were PCR-amplified using primers with 25-35 bp overhands complementary to the vectors. The vectors were linearized by PCR, and subjected to agarose gel electrophoresis and gel extraction as described above (Methods sections 3.1.1 and 3.1.2). 10 µl of the reaction mix, containing insert and vectors in 2:1 ratio (typically, 50 ng of vector and 100 ng of insert) was prepared in deionized H2O. The reaction mix was then added to an In-Fusion HD EcoDry pellet (TakaraBio), incubated for 15 min at 37°C, followed by 15 min at 50°C and then placed on ice. 2 µl of the In-Fusion reaction mixture was transformed to the competent *E.coli* cells and grown on an agar plate with respective antibiotic overnight at 37°C.

For cloning of BigBac polygene expression constructs for insect cell expression, a Gibson cloning-based approach developed by Weissmann et al. (2016) was adopted. The detailed protocol used is described in Weissmann and Peters (2018). Briefly, each gene of interest was first cloned to the transfer vector pLIB to generate an individual gene expression cassette (GEC), containing a gene of interest enclosed by polyhedrin promoter and transcriptional terminator sequences. To assemble a multiprotein expression construct, GECs were PCR-amplified from pLIB vectors using the primers, which introduced linker sequences that allowed Gibson assembly of multiple GECs into a single pBIG vector. This system allowed the assembly of up to 25 GECs to a single vector by hierarchical assembly of constructs first to pBIG1a and pBIG1b plasmids, whose polygene expression cassettes could be further combined to pBIG2 vector. In this study, however, I only used pBIG1a for the expression of Ring1b and Bmi1, since assembly of the whole PRC1 (hence Ring1b, Bmi1, Cbx, PHC2) into a single pBIG2 resulted in sufficient expression levels from pBIG1a polygene cassette A, but not from pBIG1b cassette B. For the

Gibson reaction, 100 ng of linearized vector (pLIB or pBIG1a) were mixed with 5x molar excess of insert in 10 µl of deionized H2O and then mixed with 10 µl of 2x Gibson assembly mix (MPI Biochemistry Core facility). The reaction was incubated for 1 h at 50°C, transferred to ice, and transformed into TOP10 competent *E.coli* cells. The cell suspension was cultivated on antibiotic-selective agar plate overnight at 37°C.

3.1.5 Site-specific mutagenesis

The point mutations (specifically, mutations of SAM domains of PHC2 and Scml2) were introduced using around-the-horn PCR (Lanman et al., 2003). The initial vectors were PRC-amplified using 5' phosphorylated primers (Sigma Aldrich) harbouring the mutations. The vectors were then ligated for 25 min at RT using Quick ligase (NEB), treated with DpnI nuclease (NEB) for 1h at 30°C to remove the template DNA and subsequently transformed to *E.coli* TOP10 competent cells.

3.1.6 Plasmid DNA preparation

A single bacterial colony was inoculated into 5 ml LB-medium with corresponding antibiotics (see Table 8 for concentrations), cultivated at 37°C overnight while shaking (200 rpm) and harvested by centrifugation for 10 min at 6000 rpm. The plasmid DNAs were isolated using the QIAprep Spin Miniprep Kit (Qiagen) following the manufacturer's instructions and sent for sequencing (Eurofins) with corresponding sequencing primers.

3.2 Protein expression in bacterial system

The *E. coli* protein expression strains were chosen based on small-scale expression screening (with the great help of the MPIB Crystallisation facility). His-GST-Ring1b 1-133 and His-GB1-Bmi 1-242 were co-expressed in *E.coli* BL21 Gold pLysS, His-PHC2 536-678 was expressed in *E.coli* BL21 STAR pRARE. The protein expression plasmids were transformed into corresponding bacterial cells and grown on agar plates with antibiotics (ampicillin and kanamycin for Ring1b and Bmi1 and spectinomycin for PHC2, Table 8) overnight at 37°C. Then, a single colony was used to inoculate 100 ml of LB with antibiotics and cultured overnight at 37°C while shaking (200 rpm). The next day, this pre-culture was used to inoculate 1 L of TB medium per Tune air flasks with salts and antibiotics added and let to grow at 37°C while at constant shaking (200 rpm). The expression was induced by the addition of 0.4 mM IPTG at an OD of 1.5-1.9

and grown at 18°C overnight. The following day the cells were harvested by centrifugation for 10 min at 6000 rpm and the pellets were frozen in liquid nitrogen.

3.3 Insect cell culture

Two types of cell lines were used for insect cell culture experiments (Table 2). Sf21 cell line was used to generate and amplify baculoviruses, and HighFive cells were used for protein expression. Cell-specific parameters, such as density, viability and diameter were monitored using Vi-cell XR cell viability analyzer (Beckman Coulter). The majority of insect cell culture maintenance, virus generation and protein expression was performed by Sven Schkölziger (J. Müller lab, MPI of Biochemistry).

3.3.1 Bacmid preparation

The 'Bac-to-Bac Baculovirus Expression System' from Invitrogen was used for insect cell expression (Ciccarone et al., 1998). This system involves site-specific transposition in *E. coli*, followed by the purification of recombinant bacmids and transfection of the bacmids to generate baculoviruses. Individual steps of this method are outlined below.

Transposition

The recombinant baculovirus DNA containing the gene of interest (GOI) under the control of polyhedrin promoter, was obtained through transposition between the GOI-carrying pFastBac or pBIG vectors and the baculovirus shuttle vector (Bacmid). 1 ng of pFastBac or pBIG plasmids with GOI was added to *E.coli* DH10 EMBACY cells (Invitrogen) and incubated on ice for 30 min. The cells were heat-shocked for 45 s at 42°C and placed on ice for 5 min to recover. 900 μl of SOC buffer (Table 7) was then added and the mixture was incubated for 4 h at 37°C while constantly shaking. After incubation, the cells were spun down for 4 min at 3000 rpm, the supernatant was removed and the pellet was resuspended in 200 μl LB. 160 ul of the cell suspension was plated on an LB agar plate consisting of a mixture of antibiotics (10 μg/ml tetracyclin 50 μg/ml kanamycin, 7 μg/ml gentamicin,34 μg/ml chloramphenicol), 100 μg/ml IPTG and 100 μg/ml BluoGal (5-Brom-3-indolyl-β-D-galactopyranosid). The plates were incubated for 48 hours at 37°C under light-protected conditions. The identification of successfully transposed colonies was based on their distinctive white colour. This colour change was indicative of the excision of the lacZ cassette during transposition, rendering the bacteria incapable of metabolizing the

chromogenic substrate. Single white colonies were picked and cultivated for 10 ml LB media with the same antibiotic as in LB agar selective plate for 20 h at 37°C while shaking. Bacmid DNA isolation was performed using PureLink™ Quick Plasmid MiniPrep kit (Invitrogen) according to manufacturing instructions.

3.3.2 Baculovirus generation

Sf21 at cells were transfected with Bacmid DNA to generate baculoviruses. Sf21 cells were diluted to 0.8 x 106 cells/ml in EX-CELL TiterHigh (Sigma Aldrich) and a suspension was transferred to a tissue culture plate (BD Falcon, 2 ml of suspension per well) and left to incubate for 30-60 min at 27°C. For each well, 1 µg of Bacmid in 100 µl EX-CELL TiterHigh media was mixed with a transfection solution, consisting of 8 µl Cellfectin II and 92 µl EX-CELL TiterHigh medium, resulting in 210 µl of final transfection for a single well. The transfection mix was added to wells and the plate was incubated for 4 h at 27°C. Subsequently, the liquid was aspired and exchanged for 2 ml of fresh EX-CELL TiterHigh medium. The cells were incubated for 96 hrs while cell parameters were constantly monitored. Reduced viability of the cells was an indicator of successful baculovirus infection. Supernatants from wells where cells showed reduced variability were harvested as virus generation of baculovirus (P1) and used for further virus passages (passages P2 and P3, Methods section 3.3.3). Baculoviruses from passage P3 were used for the infection of HighFive cells for protein expression (Methods section 3.3.4). Supernatants from all viral passages were stored at 4°C.

3.3.3 Baculovirus amplification

2 ml of supernatant from a successful bacmid transfection, containing the first virus generation (P1), was used to infect 50 ml of Sf21 cells (0.4 x 106 cells/ml) and left to grow for 24-72 h in 125 ml Erlenmeyer flask (90 rpm, 27°C) upon constant monitoring. The cell parameters were monitored daily using a Vi-cell XR cell viability analyzer. Once the cell growth was arrested and the cells stopped doubling, the supernatant from the cell suspension was collected (P2 virus generation) and 2.5 ml of P2 was used to infect 500 ml of Sf21 cells (0.4 x 106 cells/ml). The supernatant from the third passage (P3) was used for protein expression. P2 and P3 viral stocks were kept at 4°C for future use.

3.3.4 Protein expression in insect cells

Ring1b and Bmi1 were co-expressed from a single virus, and other subunits were either expressed separately (PHC2 and Scml2) or co-expressed with Ring1b and Bmi1 (Cbx7). Corresponding P3 viruses were used to infect HighFive cells (at 0.8 x 106 cell density). First, the expression levels and stoichiometry of purified proteins were checked in a small-scale setup, wherein proteins were expressed in 50 ml of cell suspension and then used for small-scale pulldowns (either His-tag or Streptavidin pulldowns, depending on the tags of the protein). In a large-scale expression setup, 500 ml of HighFive cells per flask (at 0.8 x 106 cell/ml) were infected with P3 virus (most often, a final 1:200 dilution of the virus was used, this ratio was established experimentally in small-scale virus titration). Up to 16 l of HighFive cell cultures were used for a single large-scale expression. Infected cells were grown for 72 h at 27°C upon gentle agitation at 90 rpm and harvested by centrifugation for 15 min at 2000 rpm. The pellets were gently resuspended in cold PBS buffer and centrifugated again for 15 minutes at 2000 rpm. Then the pellets were frozen in liquid nitrogen and stored at -80°C until further use.

3.4 Protein purification

3.4.1 PRC1 purification from *E.coli*

The previously frozen pellets were re-suspended in the lysis buffer (25 mM Tris-HCl pH 7.5, 250 mM NaCl, 25 mM imidazole, cOmplete protease inhibitors). The cell suspension was sonicated on ice and cell debris was removed by centrifugation for 30 min at 25,000 rpm at 4°C. The supernatant was aspirated and filtered through a membrane with a 5 µm cut-off. The lysate was then applied to 5 ml Ni columns (Cytiva) using a peristaltic pump with a flow of 3 ml/min. The flow-through was collected. The column was then washed with 20 column volumes (CV) of buffer A (25 mM Tris-HCl pH 7.5, 250 mM NaCl, 25 mM imidazole), then with 20 CV of chaperone buffer (20 mM Tris-HCl pH 7.5, 50 mM KCl, 10 mM MgCl2, 2 mM ATP) and with 20 CV of buffer A again. The target proteins were eluted from the column with Ni-NTA Elution buffer (25 mM Tris-HCl pH 7.5, 250 mM NaCl, 500 mM imidazole), the Ring1b, Bmi1 and PHC2 elution fractions mixed and His-3C protease (MPIB Core Facility) was added to the protein solution for tag cleavage. The protein solution was dialysed overnight in dialysis tubes with 6-8 kDa cut-off against buffer 25 mM Tris-HCl pH 7.5, 250 mM NaCl, 10 % glycerol (ON, 4°C). The next morning, the solution was centrifuged for 10 min at 4000 rpm (4°C), diluted

with 25 mM Tris-HCl pH 7.5 buffer to a final NaCl concentration of 125 mM and applied to 5 ml pre-packed Heparin column (GE Healthcare). The column was transferred to Åkta Prime (GE Healthcare) and eluted with a gradient of buffer Heparin B (20 mM Tris-HCl pH 7.5, 1 M NaCl) while collecting 5 ml fractions. The absorption at 280 nm during the collection of elution fractions was monitored using the AEKTA go chromatography system (Cytiva). The elution fractions were then analysed on SDS-PAGE gel and fractions corresponding to trimeric Ring1b, Bmi1 and PHC2 complex were collected and concentrated in a concentration unit with 10 kDa cut-off to a final volume of 500 µl. The concentrated sample was centrifuged at 15,000 rpm and 4°C for 15 min and then injected into a size-exclusion Superdex75 column of AEKTA Pure chromatography system (Cytiva). The elution fractions from SEC were analysed on 16 % polyacrylamide gel and fractions corresponding to the stochiometric trimeric complex were collected, frozen in liquid nitrogen and stored at -80 °C until further use.

3.4.2 PRC1 purification from insect cells

All proteins for insect cell expression were expressed in HighFive cells as described above, after collection the cell pellets were washed with PBS, frozen in liquid nitrogen and stored at -80°C. Before purification, the pellets were thawed on ice. Then the cells were suspended in lysis buffer consisting of 25 mM Tris-HCl pH 7.5, 150 mM NaCl, 4 % Glycerol, 0.05 % NP40, 30 mM Imidazole, and 1 mM DTT, supplemented with EDTA-free protease inhibitor cocktail (Complete Protease-Inhibitors, Sigma-Aldrich; 1 tablet per 50 ml cell lysate) and additional protease inhibitors such as leupeptin (284 mg/ml), pepstatin (1.37 µg/ml), and benzamidine (330 µg/ml). The cells were then lysed using a Dounce homogenizer, and the soluble fraction containing the proteins of interest was separated from the non-soluble fraction through centrifugation at 25,000 rpm for 45 minutes at 4°C. For PRC1-PHC2 core of PRC1 consisting of TwinStrep-Ring1b and Bmi1-His was purified using Ni-NTA affinity chromatography (using 5 ml pre-packed column, (GE Healthcare). His-tagged PHC2 was expressed and purified separately also using using Ni-NTA affinity chomatography and then mixed with Ring1b and Bmi1. Proteins bound to Ni-NTA column were washed extensively, first with 20 CV of lysis buffer, 20 CV of chaperone wash buffer (20 mM Tris-HCl pH 7.5, 50 mM KCl, 10 mM MgCl2, 2 mM ATP) and finally with 20 CV of wash buffer A (25 mM Tris-HCl pH 7.5, 250 mM NaCl, 25 mM imidazole). Proteins were eluted from the columns with Ni-NTA Elution buffer (25 mM Tris-HCl pH 7.5, 250 mM NaCl, 500 mM imidazole) and elution fractions were dialysed in dialysis membranes

with 12-14 kDa cut-off against buffer 25 mM Tris-HCl pH 7.5, 250 mM NaCl, 10 % glycerol (ON, 4°C). For the reconstitution of the PRC1-Cbx7 complex, Twin-Strep-Ring1b, Bmi1-His, and Twin-Strep-Cbx7 proteins were co-expressed and pulled down using Streptavidin affinity chromatography. The same washing and gel filtration steps as described above were followed. Similarly, the PRC-Scml2 complex, Twin-Strep-Scml2 and His-PHC2 were co-expressed and jointly purified using Streptavidin affinity chromatography. Twin-Strep-Ring1b and Bmi1-His were also co-expressed and pulled down using Streptavidin affinity chromatography. Elution fractions from both Strep-tag pulldowns were merged to reconstitute the PRC1-Scml2 complex, followed by chaperone wash and size exclusion chromatography as outlined previously.

3.5 Nucleosome reconstitution

At the beginning of this study, all histone proteins were expressed and purified by the author of this thesis with the help of Sven Schkölziger using the standard histone preparation protocol from Luger et al. (1999). Later, the histones were purchased from the Colorado Histone source at the University of Colorado.

3.5.1 Preparation of nucleosomal DNA

Nucleosomal 215 bp DNA was generated by large-scale PCR amplification using p601 plasmid as a template and corresponding primer pair (see Table 6 for the plasmid and Table 5 for primers). To introduce desired modifications to nucleosomal DNA, 5'-primer was ordered (Sigma-Aldrich) to contain 5'-biotin or 5'-6-carboxyfluorescein. The PCR reaction was mixed according to Table 10. The PCR was carried out in 96-well plates (Corning) using a modified PCR protocol (see Methods section 3.1.1). Specifically, the annealing temperature was set to 57°C and the total number of cycles to 35. Up to 12 96-well plates were used per single DNA preparation. After PCR, a single reaction from each plate was checked on agarose gel. Fractions from all plates were pooled together and precipitated by adding 2 volumes of 100% pre-cooled ethanol and 0.1 volume of 3 M NaCl, vortexing and incubating at -80°C for 2 h. The precipitated DNA was sedimented by centrifugation for 10 min at 12,000 rpm and the pellet was resuspended in 1 ml of 70% ethanol and centrifuged again for 5 min at 14,000 rpm. The pellet was air-dryed and resuspended in 5 ml of MonoQ buffer A (5 mM Tris-HCl pH 7.5). The mixture was loaded on a MonoQ column (GE Healthcare) connected to AEKTA Pure chromatography system (Cytiva) and eluted with a gradient of MonoQ buffer B (5 mM Tris-HCl pH 7.5, 2 M NaCl). The purity

of eluted DNA fractions was analysed in agarose gel electrophoresis (Methods section 3.1.2), and the final fractions were pooled together and the DNA was precipitated again following the same procedure as described above and the pellet finally dissolved in Octamer High Salt Buffer (10 mM Tris-HCL pH 7.5, 2 M NaCl, 1 mM EDTA) for use in nucleosome reconstitution.

3.5.2 Octamer refolding

The lyophilised histones were dissolved in unfolding buffer (20 mM Tris-HCL pH 7.5, 7 M Guanidine HCL, 10 mM DTT) to the final concentration of 2 mg/ml, then mixed in equimolar amounts and incubated for 45 min at RT. Histone mixture was then dialysed in against Octamer High Salt Buffer (10 mM Tris-HCL pH 7.5, 2 M NaCl, 1 mM EDTA, 5 mM β-mercaptoethanol) twice for 1 h and then overnight at 4°C. The next day, the histone mixture was centrifuged for 10 min at 10,000 rpm, 4°C, concentrated to 500 µl using a concentration unit with a cut-off of 10 kDa (Millipore) and injected into Superdex 200i (GE Healthcare) SEC column. Elution fractions were analysed by SDS-PAGE (Methods section 3.6) and octamer-containing fractions were collected for further use.

3.5.3 Nucleosome assembly

A small-scale titration was performed where nucleosomal DNA was mixed with octamers at different molar ratios to find an optimal ratio of octamers and DNA. Usually, a range of 1.0-1.6 molar excess of octamer to DNA was tested. The mixtures of different octamers to DNA ratios were mixed in Octamer High Salt Buffer and dialysed in dialysation cassettes with 3 kDa cut-off against the same buffer ON, with a gradual titration with No Salt buffer, to lower the final NaCl concentration to 500 mM. After dialysis, 5 µl of each mixture was applied to native 1.2 % agarose gel and run in pre-cooled 0.4 TBE buffer for 1 h 10 min at 60 V. The gel was stained with SYBR Safe diluted in running buffer for 1 h while agitating and visualised with BioRad imager (excitation wavelength 470 nm). The ratio with no free nucleosomal DNA was chosen for large-scale assembly following the same protocol and final dialysis steps (all at 4°C) to reduce NaCl concentration further: 2 h in buffer 10 mM Tris-HCl pH 7.5, 250 M NaCl, 1 mM EDTA, 5 mM b-mercaptoethanol, 2 h in buffer 10mM Tris-HCl pH 7.5, 150 M NaCl, 1 mM EDTA, 5 mM b-mercaptoethanol and in buffer 25 mM Tris-HCl pH 7.5, 30 mM NaCl, 2 mM DTT.

3.6 SDS-PAGE electrophoresis

Generally, the gel electrophoresis was performed using XCell SureLock Mini-Cell Electrophoresis System (Thermo Fisher), run at 150-180 V. For fast imaging of proteins during purification, SDS polyacrylamide gel electrophoresis was performed using either NuPAGE 4-12% (w/v) Bis-Tris Precast gels run with 1 x NuPAGE MOPS buffer (Thermo Fischer). For more separation of Ring1b and Bmi1 bands and nucleosome-containing samples, self-casted 16% (w/v) Tris-Glycine Polyacrylamide gels were used. Samples for SDS-PAGE were prepared by adding 2x or 4x LDS buffer containing SDS and incubation at 95°C for 5 min for protein denaturation. The gels were stained with Coomassie stain (InstantBlue Coomassie Stain, Abcam). For the visualisation of PRC1-CBX7 crystal, Pierce™ Silver Stain Kit (Thermo Fischer) was used.

3.7 Protein crystallization

For crystallisation, the PRC1-PHC2 protein complex was concentrated to 19 mg/ml using a concentrating unit with 10 kD cut-off (Millipore). Initially, several crystallisation screens were tested: Qiagen PEG, Hampton Research Index screen and two homemade screens, both at RT and 4°C. Sitting-drop crystallization method was used, mixing 100 nl of protein solution with 100 nl of reservoir. Crystals were grown in a crystallisation condition containing 15 % PEG 6000 0.1 MOPS pH 7.2 at 4°C. The crystals were flash-frozen in liquid nitrogen, with the mother liquid containing 33 % ethylene glycol as a cryo-protector. Data were collected at 100 K at the PXIII beamline at the Swiss Light Source in Villigen, Switzerland. The collected data were processed using XDS, followed by scaling and merging procedures using XSCALE (Kabsch, 2010). Phasing was performed using molecular replacement in Phaser, using Ring1b Bmi1 E3 ligase module (PDB 2CKL) as a search model (McCoy et al., 2007). The model was built using phenix.autobuild and refined using phenix.refine (Adams et al., 2010). The PRC1-Cbx7 complex was concentrated to 6.5 mg/ml, using a concentrating unit with 10 kD cut-off (Millipore). The crystallisation set-up was the same as for the PRC1-PHC2 complex. The crystal was grown in the condition containing 25% PEG 4000 0.1 M Tris-HCl pH 8.8 at RT.

Table 13: Crystallographic statistics. Values in parentheses are for the highest shell.

	Ring1b Bmi1
Data collection	
Space group	P 63
Cell dimensions	
a, b, c (Å)	121.098, 121.098, 27.063
a, b, g (°)	90, 90, 120
Resolution (Å)	39.64 - 1.91
Highest shell (Å)	1.977 - 1.909
Rsym or Rmerge	0.1628 (1.146)
I / sI	15.90 (2.46)
Completeness (%)	99.45 (99.83)
Wilson B-factor	30.55
Refinement	
Resolution (Å)	1.9
No. reflections	18053
Rwork / Rfree	0.2621 / 0.2888
No. atoms	1568
Protein (AA residues)	194
Ligand/ion	6
Water	0
R.m.s. deviations	
Bond lengths (Å)	0.0091
Bond angles (°)	1.05
Ramachandran	
Outliers (%)	0
Allowed (%)	4.26
Favourited (%)	95.74
Rotamer outliers (%)	1.10
MolProbity score	1.61

3.8 Electron microscopy

3.8.1 Sample preparation for negative stain EM screening

Before staining, streptavidin-coated grids were hydrated for 10 min with 20 µl of water to remove a protective layer of trehalose. Then 10 µl of the sample was applied to the grid and incubated for 5 min to allow binding of biotinylated nucleosomes to streptavidin. Excess of the sample was blotted off with blotting paper and then the grid was washed twice with 4 µl of water, blotting excess liquid after each wash. Then 4 µl of 2 % uranyl acetate was applied to the grid, incubated for 10s and blotted off. Then 4 µl of 2 % uranyl acetate was applied to the grid again, incubated

for 10s, blotted off, re-applied, incubated for 20s, blotted off and let dry. Repeated staining with uranyl acetate resulted in a thick enough layer of stain.

3.8.2 Cryo-EM sample preparation using streptavidin grids

Streptavidin monolayer affinity grids for cryo-EM sample preparation were made by Dr. Christian Benda (MPI Biochemistry, Department Conti), using a protocol from Han et al. (2016). PRC1-Cbx7-nucleosome grids were prepared with modifications to the standard protocol: the sample was initially prepared in plunging buffer and washing steps were omitted to enrich the complex's occupancy on nucleosomes. To prepare the sample for cryo-electron microscopy analysis, 350 nM *Drosophila melanogaster* nucleosome with 215-bp biotinylated DNA were mixed with 1.75 µM PRC1 and incubated for 30 min on ice. Prior to plunging, the sample was diluted to a final concentration of 116 nM nucleosome and 583 nM PRC1, representing a five-fold molar excess of PRC1. Before usage, streptavidin grids were hydrated with 20 µl of water for 10 min and excess water was blotted off. Subsequently, 4 µl of the sample mixture was applied to the grid and incubated for 5 min at room temperature. Then the grid was washed twice with 20 µl of a buffer solution composed of 25 mM Tris-HCl pH 7.5, 40 mM KCl, 1 mM MgCl2, 1 mM TCEP, 4% trehalose, 0.01% NP-40, and 0.04% octyl-glucoside. Excess buffer was gently blotted off after each wash step. After washing, the grid was transferred to Mark IV Vitrobot (FEI). 3 µl of the wash buffer was applied to the grid before blotting and plunging into a liquid ethane/propane mix (37 % ethane) cooled by liquid nitrogen. The Vitrobot settings used were set to 4°C temperature, 100% humidity, blot force of 4, and a 3.5 s blotting time.

3.8.3 Cryo-EM sample preparation using glutaraldehyde fixation

The glutaraldehyde crosslinking protocol used in this study was adapted from Worden et al. (2020). A mixture of 100 nM *Drosophila melanogaster* nucleosomes and 600 nM PRC1-Cbx7 was prepared in a buffer containing 25 mM HEPES (pH 7.8), 50 mM NaCl, and 1 mM TCEP. The mixture was incubated on ice for 1 h.

Following the initial incubation, an equal amount of buffer containing 0.1% glutaraldehyde was added to the sample, and the mixture was incubated on ice for 30 min. To quench the cross-linking reaction, 1 M Tris-HCl, pH 7.5 was added to a final concentration of 100 μM, and the sample was left on ice for 20 min. Subsequently, the sample was concentrated to a concentration of 1.35 mg/ml using a 50 kDa cut-off concentrator (Millipore). Quantifoil R2/1

grids were glow-discharged for 1 min using the GloQube Plus Glow Discharge System and then transferred to Mark IV Vitrobot (FEI). 3 µl of the sample was applied to the grid before blotting and plunging into a liquid ethane/propane mix cooled by liquid nitrogen. The Vitrobot settings used were set to 4°C temperature, 100% humidity, blot force 4, and a 3 s blotting time.

3.8.4 Cryo-EM data collection

Cryo-EM screening datasets were acquired using either the Talos Arctica (Thermo Fisher Scientific) or Glacios microscopes, equipped with Falcon3 (Thermo Fisher Scientific) and K2 Summit (Gatan) cameras, respectively. For the final datasets, data collection was performed on a 300 kV Titan Krios microscope (FEI/Thermo Fisher Scientific), with the assistance of Dr. Daniel Bollschweiler and Dr. Tillman Schäfer from the cryo-EM facility of MPI of Biochemistry. The Titan Krios microscope was equipped with a Gatan K3 direct detector operating in counting mode, along with a post-column energy filter with a slit width of 30 eV. Semi-automated data collection was performed using the SerialEM software (Mastronarde, 2005).

For PRC1-PHC2:nucleosome, a dataset consisting of 26,896 dose-fractionated movies was collected on the Titan Krios at a nominal magnification of 215,000x, corresponding to a pixel size of 0.4114 Å. The target defocus range for this dataset was set between -0.7 and 2.2 um. The total electron dose was 87.6 e-/Å^2, distributed over 15 frames.

For both cross-linked and non-crosslinked datasets for PRC1-Cbx7 with nucleosomes, a nominal magnification of 105,000x was used, corresponding to a pixel size of 0.8512 Å. The target defocus range for both datasets ranged from -0.6 to 2.2 um.

PRC1-Cbx7:nucleosome non-crosslinked dataset amounted to 23,398 dose-fractionated movies with a total electron dose of 62.88 e-/Å^2 distributed over 40 frames. PRC1-Cbx7 consisted of 13,553 dose-fractionated movies with a total electron dose of 64.5 e-/Å^2, distributed over 30 frames.

3.8.5 Cryo-EM data processing

For all datasets in this study, movie stacks were aligned and corrected for beam-induced motion with MotionCorr2 implemented in RELION (Zheng et al., 2017). PRC1-PHC2 and PRC1-CBX7 (non-crosslinked) datasets, collected on streptavidin-coated grids, required an additional pre-processing step of lattice subtraction using MATLAB script described in (Han et al., 2016).

Subsequent processing steps were carried out in cryoSPARC v 3.1 (Punjani et al., 2017). Contrast Transfer Function (CTF) estimation was performed using CTFFIND4 implemented in cryoSPARC (Rohou and Grigorieff, 2015). Following CTF estimation, micrographs were inspected manually and any micrographs displaying outliers in defocus values, signs of ice contamination, or having a predicted CTF-correlated resolution greater than 6 Å were discarded from further analysis.

For the PRC1-PHC2 dataset, initial picking was carried out on a subset of micrographs using a difference of Gaussians (DoG)-based approach with a particle diameter range of 100-200 Å. Picked particles were extracted with down-scaled box size (binning factor = 6) and corresponding micrographs used for training a TOPAZ neural network model, which was later used for re-picking on the whole dataset (Bepler et al., 2019). Picked particles were subjected to several rounds of reference-free 2D classification and a subset of 50,000 particles was used for *ab initio* 3D reconstruction. In heterogeneous refinement with 2 classes, only the class containing a prominent PRC1-attributed extra density was selected. Focused refinement was carried out with a soft mask applied, covering nucleosome and one copy of PRC1 E3 ligase module. Focusing the refinement on nucleosomes with a single PRC1 E3 ligase module improved the resolution of density corresponding to PRC1, as the second PRC1 binding site showed less occupancy.

A subset of micrographs from the non-crosslinked dataset of PRC1-Cbx7:nucleosome was initially subjected to particle-picking using a difference of Gaussians (DoG)-based approach, as mentioned earlier. The particles were extracted using a down-scaled box size (binning factor = 4) and used for 3D reconstitution, from which 2D templates were generated to pick particles from the entire set of micrographs. Subsequently, extracted and down-scaled particles were subjected to extensive 2D classification in order to remove corrupted particles and particles with residual streptavidin-lattice background. *Ab initio* 3D reconstruction was carried out with a subset of 100,000 particles and used as a reference for several rounds of heterogeneous refinement, where classes with PRC1-bound nucleosomes were selected. This set of 464,476 particles was used as input for various refinement and classification jobs. First, it was subjected to 3D-focused refinement with a soft mask covering density corresponding to nucleosome and PRC1 E3 ligase domains bound on both sides of the nucleosome. The mask did not include an extra density protruding beyond/from the E3 ligase domain. In parallel, this final set of particles was 3D-classified in an attempt to identify a subset for distinct conformational states, subject to an additional round of heterogeneous refinement and 3D variability analysis (Punjani and

Fleet, 2021). The 3-10 Å target resolution range was tested for both 3D classification and 3D variability analysis, with the parameter set to 7 Å and 10 Å for the respective jobs. Output classes of heterogeneous refinement that contained density beyond PRC1 E3 ligase domain were transferred to RELION 3.1.1 (Scheres, 2012, Zivanov et al., 2018) for masked 3D classification without alignment (regularization parameter T=10). For the purpose of the transfer, particle metadata was converted to RELION STAR format using PyEM module (Asarnow et al., 2019).

13,553 motion-corrected micrographs of PRC1-Cbx7:nucleosome cross-linked dataset were subjected to cryoSPARC blob picking with a diameter of 150-200 Å, resulting in an initial dataset of 3,460,973 particles, that were extracted with downsized box size (binning factor = 4) and underwent several rounds of reference-free 2D classification. The resulting cleaned dataset of 1,101,734 particles was further iteratively 3D classified in several rounds of heterogeneous refinement. The final set of 172,434 particles was re-extracted with unbinned box size and used for focused homogeneous refinement, unmasked homogeneous refinement, 3D classification and 3D variability analysis with the parameters as described above for the PRC1-Cbx7:nuclesome non-crosslinked dataset. Final 3D maps were auto-sharpened using deepEMhancer (Sanchez-Garcia et al., 2021), ChimeraX version 1.5. was used for visualisation and mask generation (Goddard et al., 2018).

3.8.6 Model building and validation

All 3D reconstructions of PRC1-E3-ligase modules bound to the nucleosomes presented in this study were interpreted by fitting them with the previously published models of the nucleosome bound PRC1-E2-E3 complex (PDB 4R8P, McGinty et al. (2014)). Initially, the rigid-body fitting of the initial models was performed in ChimeraX, followed by local readjustments using Coot (Goddard et al., 2018, Emsley and Cowtan, 2004). The readjustments primarily focused on modifying the lengths of the histone tails and on mutating the sequence of histone proteins to match the *Drosophila melanogaster* proteins. Prior to the fitting, the initial model was modified by removing E2 UbcH5c, which was originally fused to the Ring1b Ring finger domain in the crystal structure. Additionally, the second PRC1 E3 ligase module and E2 bound to the distal side of the nucleosome were excluded from the initial models for PRC1-PHC2 and PRC1-Cbx7 (cross-linked) 3D reconstructions. This exclusion was needed because these 3D maps contained only one PRC1 E3 ligase module bound to the proximal nucleosomal disk due to lower occupancy of the distal side.

The refinement process for the coordinates of all the models involved a series of manual adjustments in Coot (Emsley and Cowtan, 2004) and real space refinements against the corresponding unsharpened EM map using PHENIX (Liebschner et al., 2019). The quality of the models was assessed by calculating map-to-model FCS and other validation parameters using the MolProbity implementation in Phenix (Chen et al., 2010), and the results were summarized in Table 14. Finally, the models were visualized using ChimeraX 1.5 (Goddard et al., 2018).

3.9 Biochemical assays

3.9.1 Electrophoretic mobility shift assay (EMSA)

To perform EMSA, nucleosomes were prepared with a DNA containing 5' 6-carboxyfluorescein (6-FAM, prepared as described in the Method Section 3.5.1) that allowed fluorescent signal detection. 10 nM of 6-FAM labelled nucleosomes were incubated with increasing amounts of PRC1 (0 - 2.5 µM) in a total volume of 20 µl and incubated on ice for 5 min (in the binding buffer: 25 mM Tris-HCl pH 7.5, 50 mM NaCl, 0.05 % Tween-20, 5 mM MgCl2). 10 µl of the mixture was applied to 1 % agarose gel in 0.4x TBE buffer and run at 0.4 pre-cooled 0.4 TBE buffer for 60 V for 45 min, protected from light. Another 10 µl was applied to 16 % SDS-PAGE gel to check for loading control. Typhoon FLA 9500 imager was used for the visualisation of the gel (Cy2 channel). To calculate fractions of bound and unbound nucleosomes, Fiji software (Schindelin et al., 2012) was used to obtain integrated densities. The background was subtracted. The data were transferred to Prism 8 (GraphPad) and fitted with the Hill function to calculate apparent K_d values and generate the plots.

3.9.2 In vitro ubiquitination of recombinant nucleosomes

In vitro ubiquitination assay was used to test E3 ligase activity of purified PRC1 subcomplexes. To reconstitute the whole ubiquitination cascade, human E1 ubiquitin-activating enzyme Ube1 and recombinant ubiquitin were purchased from Boston Biochem and E2 E2 ubiquitin-conjugating enzyme UbcH5c was expressed and purified using previously published protocol (Buchwald et al., 2006). Ring1b 1–130 Bmi1 1–109 complex was expressed in bacterial cells and purified for subsequent usage as a positive control and a reference for comparing E3 ligase activity of PRC1 complexes (Buchwald et al., 2006).

To perform the *in vitro* ubiquitination reaction, the following components were mixed:

Table 14: Summary of cryo-EM data collection, processing and refinement statistics.

	PRC1-PHC2	**PRC1-Cbx7 (non-crosslinked)**	**PRC1-Cbx7 (cross-linked)**
Data collection and processing			
Microscope	FEI Titan Krios GII	FEI Titan Krios GII	FEI Titan Krios GII
Voltage (kV)	300	300	300
Camera	Gatan K3	Gatan K3	Gatan K3
Energy Filter	Gatan Quantum-LS (GIF)	Gatan Quantum-LS (GIF)	Gatan Quantum-LS (GIF)
Pixel size (Å/pix)	0.4114	0.8512	0.8512
Nominal magnification	215000x	105000x	105000x
Defocus range (μm)	-0.7 – -2.2	-0.6 – -2.2	-0.6 – -2.2
Exposure (e/Å^2)	87.6	62.8	64.5
Frames/micrograph	15	40	30
Filter slit width (eV)	20	20	20
Software	SerialEM	SerialEM	SerialEM
3D reconstruction (Ring fingers: Nuc)			
Number of movies	24,700	23,398	13,553
Selected particles	977,610	5,614,686	3,460,973
Final particles	123,034	126,217	172,434
Resolution (Å)	3.2	2.69	2.92
Local resolution range (Å)	0.86 - 6.91	1.83 - 7.98	2.59 - 7.26
Refinement			
Initial model	4R8P	4R8P	4R8P
Model map FSC (Å)	3.40	2.95	3.44
Model composition			
Residues	1242	1457	1238
Ligands	4 ZN +2	8 ZN +2	4 ZN +2
R.m.s deviations			
Bond length (%)	0.00%	0.00%	0.00%
Bond angles (%)	0.01%	0.03%	0.00%
Ramachandran plot			
Favored (%)	98.47%	98.76%	98.84%
Outliers(%)	0.00%	0.09%	0.11%
Validation			
MolProbity score	1.9	1.83	1.34
Clash score	5.81	6.11	5.67
Cb outliers (%)	0.11%	0.00%	0.00%
Rotamer outliers (%)	0.54%	0.20%	0.13%
CaBLAM outliers (%)	1.3%	0.8%	0.9%

Drosophila melanogaster mononucleosomes (350 nM), UBE1 (27 nM), UBCH5C (200 nM), PRC1 E3 ligase (168 nM), and ubiquitin (15 µM). The reaction was conducted in a buffer containing 50 mM Tris-HCl (pH 7.5), 10 mM MgCl2, 1 uM ZnCl2, 0.5 mM DTT, and 5 mM ATP. The reaction mixture was incubated at 30°C, and at 30-minute intervals, a 10 µL sample was taken to monitor the progression of the reaction over time. An equal volume of ubiquitination buffer containing 500 mM NaCl was added to stop the reactions. The H2A monoubiquitination reaction was monitored by separating the reaction products on Coomassie-stained 16% SDS-PAGE. Monoubiquitination of histone H2A at position K118 (in *Drosophila melanogaster* nucleosomes) was indicated by a shift of the H2A histone band to a higher molecular weight position on the gel.

3.9.3 Chemical cross-linking and mass spectrometry

To perform cross-linking mass spectrometry, a mixture of 300 nM nucleosomes and 3 µM PRC1 complexes (or 10 µM of PRC1 complex alone) were incubated with 0.5 mM BS3 (Thermo Fisher Scientific) in buffer containing 25 mM HEPES pH 7.8, 50 mM NaCl. Following incubation, the cross-linking reaction was quenched by adding 20 mM Tris-HCl pH 7.5 and incubated for 15 min. The sample was centrifuged for 15 min at 18,000g to remove any possible aggregates.

The steps of the sample preparation protocol outlined below were carried out by MPIB Mass spectrometry facility. To denature the cross-linked proteins, a mixture of 4 M Urea and 50 mM Tris was added, and the sample was sonicated with a Bioruptor Plus sonication system (Diogenode) for 10 rounds of 30 s at high intensity. Next, 40 mM 2-chloroacetamide (CAA, Sigma-Aldrich) and 10 mM Tris (2-carboxyethyl)phosphine (TCEP; Thermo Fisher Scientific) were added, followed by 20 min incubation at 37°C, and a 3x dilution with MS grade water (VWR). The sample was then digested with 1 ug of trypsin (Promega) overnight at 37°C. The following day, trifluoroacetic acid (TFA; Merck) was added to a final concentration of 1% in order to acidify the solution. The peptides were then desalted using Sep-Pak C18 1 cc vacuum cartridges (Waters), and the elutions from the column were lyophilized using a vacuum dryer. In case needed, the desalted peptides were further pre-fractionated into 8 concatenated fractions using a high pH reversed-phased nano-fractionation system (Kulak et al., 2017).

Desalted peptides were dissolved in buffer A (0.1% formic acid) and 1/10 of the peptides were analyzed using an LC-MS/MS set-up where an Easy-nLC 1200 (Thermo Fisher Scientific)

is coupled to an Exploris 480 mass spectrometer (Thermo Fisher Scientific). The analytical column (30 cm, inner diameter: 75 microns; packed in-house with ReproSil-Pur C18-AQ 1.9-micron beads, Dr. Maisch GmbH) was heated to 60 °C and peptides were separated at a flow rate of 300 nL/min by a gradient of buffer B (80% acetonitrile, 0.1% formic acid). The gradient comprised an increase of buffer B from 5% to 30% over 40 min followed by an increase to 95% over 10 mins then 95% over the next 5 minutes. Finally, the percentage of buffer B was kept at 95% for another 5 minutes. The mass spectrometer was operated in data-dependent mode with survey scans from m/z 300 to 1650 Th (resolution of 60k at m/z = 200 Th), and up to 15 of the most abundant precursors were selected and fragmented using stepped Higher-energy C-trap Dissociation (HCD with a normalized collision energy of value of 19, 27, 35). The MS2 spectra were recorded with a dynamic m/z range (resolution of 30k at m/z = 200 Th). Normalized AGC targets for MS1 and MS2 scans were set to 300 % and 100 %, respectively, within a maximum injection time of 25 ms the MS1 scan. The maximum injection time was set to "auto" for the MS2 scans. Charge state 2 was excluded from fragmentation to enrich the fragmentation scans for cross-linked peptide precursors. The acquired raw data was processed using Proteome Discoverer (version 2.5.0.400) with XlinkX/PD nodes integrated (Klykov et al., 2018). For the identification of cross-linked peptides, a database search was performed against a reduced FASTA containing the sequences of the proteins of interest and a contaminant database. BS3/DSS was set as a cross-linker. Cysteine carbamidomethylation was set as a fixed modification and methionine oxidation and protein N-term acetylation were set as dynamic modifications. Trypsin/P was specified as protease and up to two missed cleavages were allowed. The identifications were only accepted with a minimal score of 40 and a minimal delta score of 4. Filtering at 1% false discovery rate (FDR) at the crosslinks level was applied.

An R script, written with the great help of Dr. Jacques Bonnet, was used to compare the datasets from two replicates of cross-linking mass spectrometry experiments. This comparison enabled us to generate a list of cross-links shared between both replicates. Subsequently, these cross-links were mapped onto the respective protein sequences to create connectivity maps (Graham et al., 2019). These cross-links were also mapped onto AlphaFold2-generated models using XMAS package in ChimeraX (Lagerwaard et al., 2022).

4 Results

As outlined in the introductory section, the canonical PRC1 complex consists of four core subunits Ring, PCGF, Cbx, PHC, and a substoichiometric subunit Scm. However, in humans, each of these subunits could be represented by several homologs: Ring1A/B, PCGF2/4, PHC1/2/3, Cbx2/4/6/7/8, and Scml1/Scml2/Scmh1. In this study, I decided to focus on a human canonical PRC1 complex, whose catalytic core is composed of Ring1b and Bmi1 (also known as PCGF4), along with additional subunits PHC2, Cbx7 and Scml2. Several PRC1 subcomplexes containing these subunits were reconstituted, which I named minimal PRC1-PHC2, PRC1-PHC2, PRC1-Cbx7 and PRC1-PHC2-Scml2. The experimental results for each of these subcomplexes will be discussed separately in the respective sections.

4.1 PRC1-PHC2 subcomplex

4.1.1 Minimal PRC1-PHC2 subcomplex purified from bacterial cells is an active E3 ligase and binds nucleosome

Purification of minimal PRC1-PHC2 subcomplex

For the purpose of the studies presented in this thesis, it was essential to obtain a human canonical PRC1 complex in substantial quantities and with a high degree of purity. Given the challenges associated with the expression of complexes containing unstructured regions in a bacterial system, first, a bottom-up approach was adopted: instead of expressing full-length PRC1 subunits, I designed an expression construct that contained structured domains of Ring1b, Bmi1, and PHC2, which interactions were demonstrated by previous studies (Buchwald et al., 2006, Gray et al., 2016, Wang et al., 2011). This complex is referred to herein as minimal PRC1-PHC2.

More specifically, the minimal PRC1-PHC2 was composed of Ring1b Ring finger domain (residues 1-133), Bmi1 Ring finger and RAWUL domains (residues 1-242), and PHC2 HD and FCS domains (residues 526 - 678). By assembling these individual domains into a larger complex for use in structural studies, it was anticipated that a crystal structure of such a minimal complex might provide further insights into the quaternary organisation of PRC1. A schematic representation of this complex is shown in Figure 10A.

The constructs outlined in Figure 10A were cloned into pET plasmids for bacterial expres-

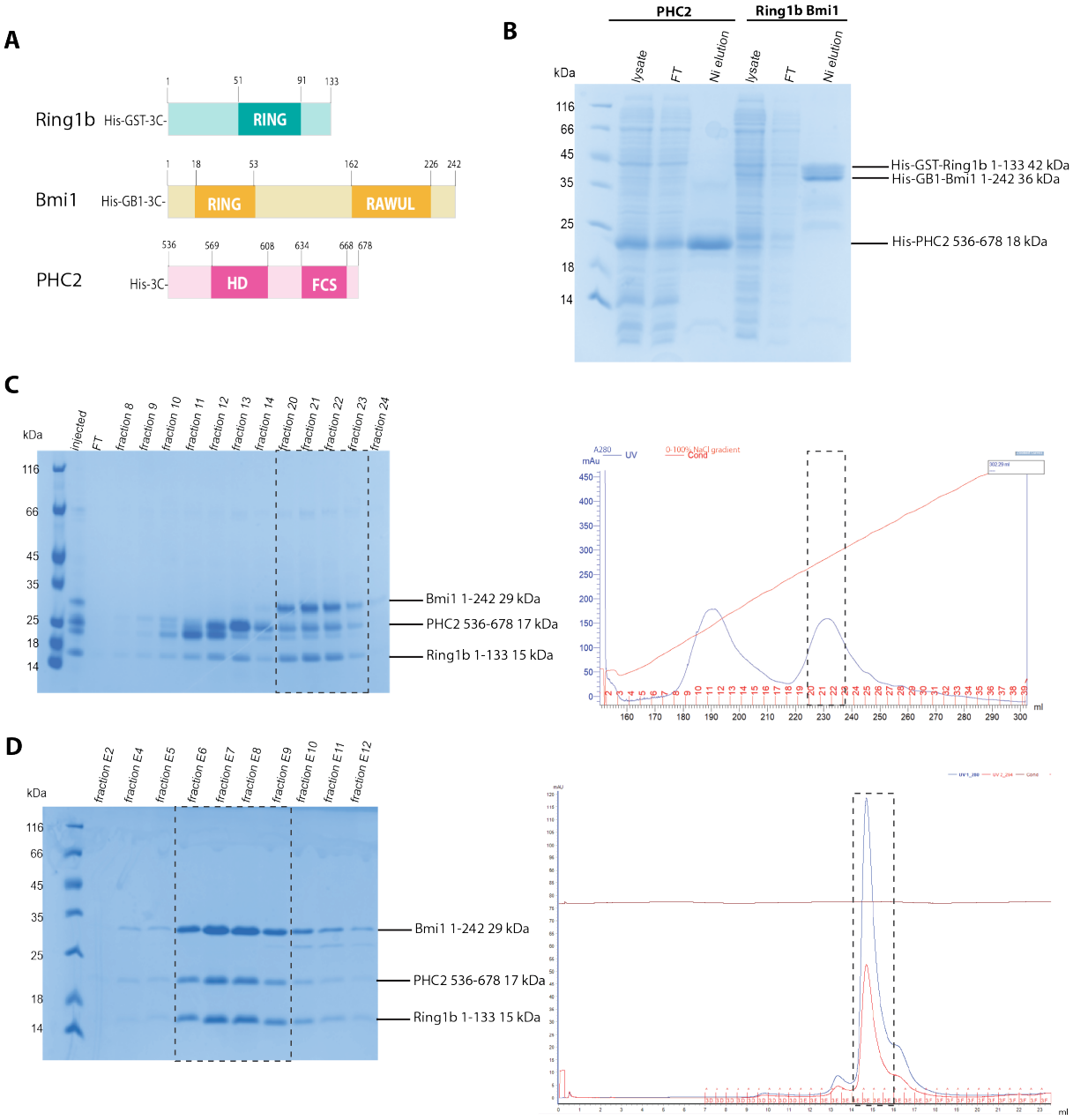


Figure 10: Overview of the purification process of minimal PRC1-PHC2 from bacterial cells. A. Schemes of protein constructs and tags used in purification. B. Coomassie-stained 16% SDS PAGE analysis of the Ni-affinity pull-down of His-tagged PHC2 536-678, Bmi1 1-242 and Ring1b 1-133. C. Heparin affinity chromatography. Coomassie-stained 4-12% SDS PAGE analysis (left) of elution fractions and corresponding chromatogram profile (right). Fractions used for the SEC are marked with a dashed box. D. Coomassie-stained 16% SDS PAGE analysis of SEC peak fractions (left) and corresponding SEC chromatogram (right).

sion (Method section 3.1.3). All proteins were N-terminally tagged with His-tag, Ring1b was additionally tagged with Glutathione-S-transferase (GST) and Bmi1 with B1 domain of Streptococcal protein G (GB1) for better solubility (Smith and Johnson, 1988, Huth et al., 1997). Expression vectors containing the respective protein-coding sequences, along with affinity and solubility tags, were chemically transformed into competent *E. coli* cells. His-GST-Ring1b 1-133 and His-GB1-Bmi1 1-242 plasmids were co-transformed to *E.coli* Gold pLysis, while His-PHC2 construct was transformed to *E.coli* STAR pRARE. The selection of expression strains was based on a small-scale expression test conducted across a range of *E.coli* strains. To ensure the stability of Bmi1, His-GST-Ring1b 1-133 and His-GB1-Bmi1 were co-expressed, as previous research has demonstrated the indispensable role of Ring1b in maintaining Bmi1 stability (Buchwald et al., 2006). Due to its relatively higher expression level compared to Ring1b and Bmi1, His-PHC2 536-678 was expressed separately.

Following the His-pulldown step, elution fractions containing His-PHC2 and His-GST-Ring1b together with His-GB1-Bmi1 were combined, reconstituting the entire minimal PRC1-PHC2 complex (Figure 10B-D). PHC2 subunit also exhibited significant levels of nucleic acid contamination, likely attributed to non-specific nucleic acid binding by the FSC domain (Wang et al., 2011). To eliminate nucleic acids, a heparin affinity chromatography step was incorporated into the purification procedure. The minimal PRC1-PHC2 complex was found to bind to the heparin column and was subsequently eluted using a gradient of NaCl, while the nucleic acids remained unbound and were effectively removed from the complex. Through this purification step, an excess of Ring1b and PHC2 subunits was also removed, resulting in a stoichiometric complex (Figure 10B left, fractions 20-23). Heparin elution fractions containing all complex subunits were collected, concentrated, and applied to the Superdex 75 10/300 GL size exclusion column (Cytiva). Size exclusion elution fractions were collected, analyzed on 16 % SDS-PAGE gel (Figure 10D), flash-frozen in liquid nitrogen, and stored at -80°C until further use. From a standard large-scale purification, an approximate yield of 600 per 1 l of the culture of highly pure minimal PRC1-PHC2 complex was obtained. A full overview of the purification process can be seen in Figure 10.

As described in the Introduction of this thesis, PRC1 is an E3 ligase known for its ability to bind nucleosomes and catalyze monoubiquitination of lysine 119 (or lysine 118 in *Drosophila*) of histone H2A. Therefore, to ensure the quality of the purified complex, its nucleosome binding and E3 ligase activities were tested in *in vitro* assays.

Nucleosome core particle reconstituion

To conduct biochemical assays, the substrate for PRC1 binding and ubiquitination - *Drosophila melanogaster* (*D.m*) nucleosomes were recombinantly reconstituted. Despite working with a human PRC1 complex, *Drosophila* histones were still a suitable substrate due to the high degree of conservation among histones across eukaryotic species (Malik and Henikoff, 2003). At the beginning of this study, individual histones required for nucleosome reconstitution were expressed in bacteria and purified following the established methodology outlined by Luger et al. (1999). Later, lyophilized histones were obtained from Histone Source at Colorado University and then used for octamer reconstitution following the protocol described in detail in Method section 3.5.3. Briefly, lyophilized histones were dissolved in a denaturing buffer containing 7 M Guanidine HCl and mixed in equimolar amounts. The resulting mixture was then subjected to several dialysis steps against a refolding buffer containing 2 M NaCl. Subsequently, the octamers were purified using size exclusion chromatography, and the fractions corresponding to the peak were analyzed using SDS-PAGE (Figure 11B).

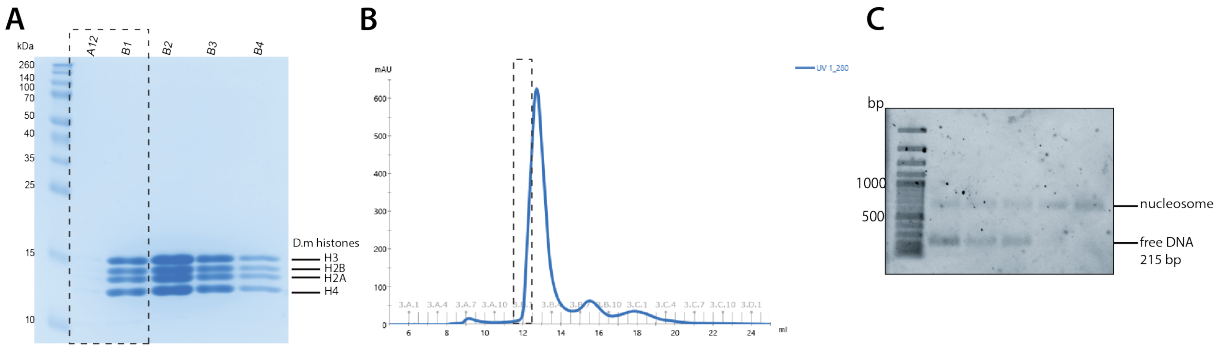


Figure 11: *Drosophila melanogaster* nucleosome reconstitution. A. Coomassie stained 16% SDS-PAGE analysis of SEC elution fractions of *D.m* wild-type octamers (left) and corresponding SEC elution profile. Fractions used for nucleosome assembly are marked with dashed boxes. C. Representative agarose gel for octamer:DNA titration. Band below 500 bp shows a free 215 bp DNA containing 601 Lowary positioning sequence motif (Lowary and Widom, 1998). The upper band between 500 and 1000 bp represents an assembled nucleosome. An Octamer:DNA for large-scale nucleosome assembly was chosen based on the absence of free nucleosomal DNA (as in lane 4).

Nucleosomal DNA containing Widom p601 nucleosome positioning sequence (Lowary and Widom, 1998) was prepared using large-scale PCR amplification (Method Section 3.1.1) and mixed with various amounts of octamer, to identify the optimal DNA-to-octamer ratio, which was used for large-scale nucleosome reconstitution (Figure 11C, the ratio in lane 4 was used for large scale reconstitution).

Electrophoretic mobility shift assay (EMSA) and E3 ligase assay The binding of minimal PRC1-PHC2 to nucleosomes was tested by electrophoretic mobility shift assay (EMSA) (Figure 12B). Increasing concentrations of minimal PRC1-PHC2 were incubated with 10 nM *D.m* nucleosome with 215-bp 6-carboxyfluorescein (FAM) DNA. Binding was monitored on native agarose gel, as PRC1-bound nucleosomes migrated slowly through the gel, compared to unbound nucleosomes (Figure 12B, left).

It is important to note that EMSA is a non-equilibrium technique, meaning that it cannot directly determine the dissociation constant (K_d) of an interaction. Instead, EMSA provides an approximation of the K_d, known as the apparent K_d (K_{dApp}) (Hellman and Fried, 2007). Furthermore, for EMSA to be considered a quantitative assay, it requires the comparison of two species on the same gel under the same experimental conditions, to ensure that any observed differences in mobility can be attributed to the specific parameter being investigated.

In the case of the minimal PRC1-PHC2 subcomplex and other PRC1 subcomplexes in this study, unexpectedly low apparent K_d values were observed in EMSA experiments (Figure 12B, right). This raises concerns about the possibility of a titration effect, especially when the concentration of the nucleosome substrate is in the same range as K_{dApp} of the interaction (Hellman and Fried, 2007). To address this issue, it would be beneficial to lower the substrate concentration to at least an order of magnitude lower than the K_{dApp}. However, in the current experimental setup with the usage of FAM as a fluorophore, this was not feasible due to the detection limits. Therefore, throughout this study, unless two conditions are compared, EMSA serves for qualitative characterization of interaction.

To evaluate the E3 ligase activity of the minimal PRC1-PHC2 complex, a reaction mixture containing 300 nM *D.m* mononucleosomes, E1 Ube1, E2 UbcH5c, ubiquitin, ATP and PRC1 E3 ligase was prepared as in Kalb et al. (2014). To assess the impact of PHC2 on the E3 ligase activity, minimal PRC1-PHC2 was compared to the PRC1 E3 ligase module alone (consisting of Ring1b 1-130 and Bmi1 1-109 only). PRC1 E3 ligase module and UbcH5c E2 enzyme were purified as described in Buchwald et al. (2006). The components of the reaction mixture visualized by SDS-PAGE are depicted in Figure 12A, right panel. The conversion of H2A histone to its ubiquitinated form was then monitored by observing the gel shift on a 16% Coomassie-stained SDS-PAGE gel over time (Figure 12A, right panel; compare the band corresponding to H2A and H2AK118ub1). The first lane corresponds to the negative control without any E3 ligase,

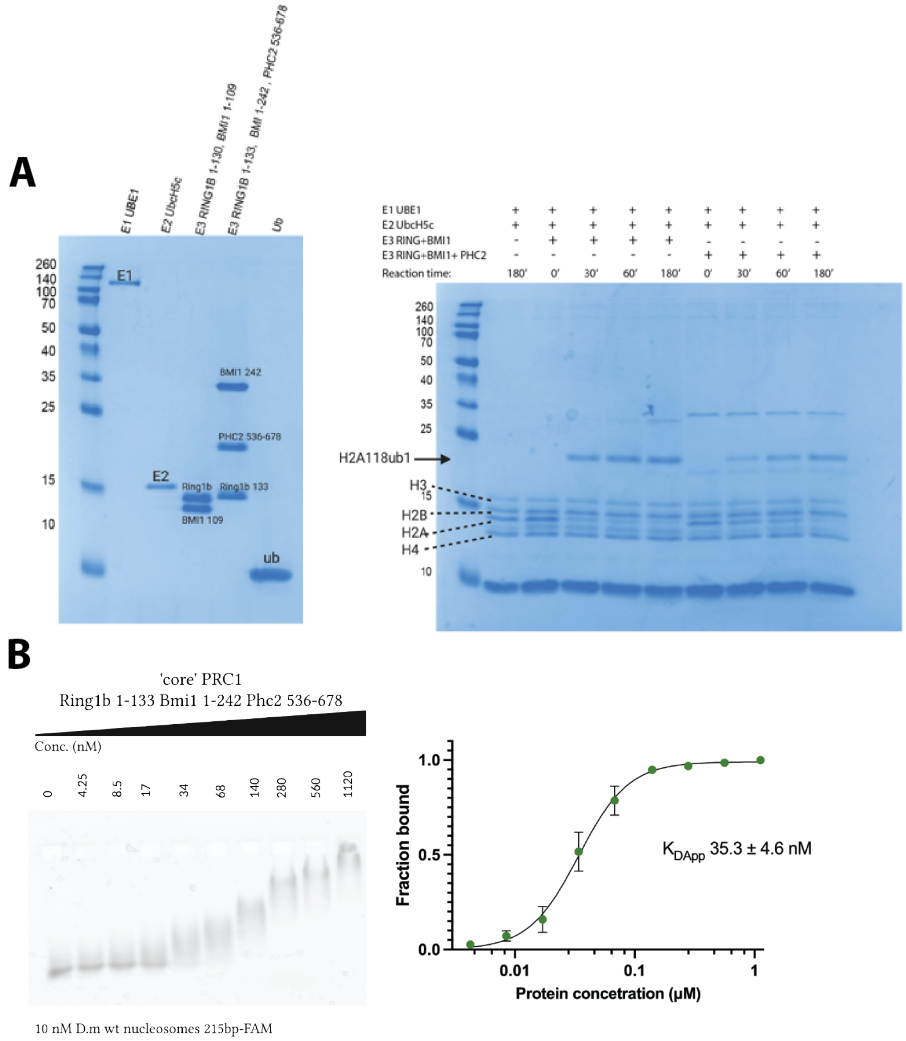


Figure 12: Biochemical characterisation of minimal PRC1 complex. A.*In vitro* H2AK118ub ubiquitination assay on reconstituted *Drosophila melanogaster* nucleosomes. The left panel shows a Coomassie-stained 16% SDS PAGE gel of all proteins used in the ubiquitination reaction. The right panel represents a Coomassie-stained 16% SDS PAGE with a time course fractions of ubiquitination reaction. The band corresponding to ubiquitylated H2AK118ub1 is marked with an arrow. B. EMSA performed with minimal PRC1-PHC2 on 215 bp 6-carboxyfluorescein labelled *D.m* nucleosomes in triplicates, quantification by densitometric analysis shown on the left.

PRC1 E3 ligase module (Ring1b 1-130 and Bmi1 1-109 only) was used as a positive control and showed almost complete conversion of H2A to H2Aub at 30' reaction time. Minimal PRC1-PHC2 showed slower convention of H2A to H2Aub than the positive control (compare 30' time points in Figure 12A, right panel).

In summary, the purified minimal PRC1-PHC2 complex demonstrated enzymatic activity as an E3 ligase in ubiquitination assays conducted with reconstituted nucleosomes. Additionally, it exhibited a binding affinity for wild-type *D.m* nucleosomes in EMSA assay, with an apparent K_d of 35.3 ± 4.6 nM. The maintained enzymatic activity of the purified PRC1-PHC2 complex served as a confirmation of its preserved conformational integrity during purification and suitability for use in subsequent structural studies presented thereafter.

4.1.2 Crystallisation of minimal PRC1-PHC2 resulted in crystals containing only E3 ligase domain

Purified minimal PRC1-PHC2 complex, whose nucleosome biding and E3 enzymatic activity were accessed previously, was then used for crystallographic studies. Initial crystals appeared in crystallisation conditions containing 15% PEG 6000 0.1 MOPS pH 7.2 at 4°C after 21 days. As seen in Figure 13B, obtained crystals had a needle-like shape, which proved to be challenging for crystal handling and measurement, therefore further optimisation of this condition by varying pH and concentration of PEG 6000 and use of seeding was performed. However, optimisation of these conditions did not result in a change in crystal morphology. One of the initial crystals was therefore used for data collection. The crystal structure was determined at 1.9 Å resolution and solely consisted of the E3 ligase module of Ring1b and Bmi1 (residues 5-115 for Ring1b and 6-103 for Bmi1, Figure 13C). Notably, no electron density corresponded to PHC2 or regions of Bmi1 and Ring1b outside the Ring finger domains. The structure was compared to the existing crystal structure of E3 ligase module and showed a root-mean-square deviation of 0.157 Å, across all amino acids (PDB 2CKL, Buchwald et al. (2006), Figure 13D). The SDS-PAGE analysis showed that the crystal contained degraded portions of Ring1b and Bmi1, corresponding to Ring domains as previously published crystal structure (Figure 13A). Remarkably, the crystal displayed similar parameters as was previously published, with the same hexagonal P63 space group and similar unit cell dimensions.

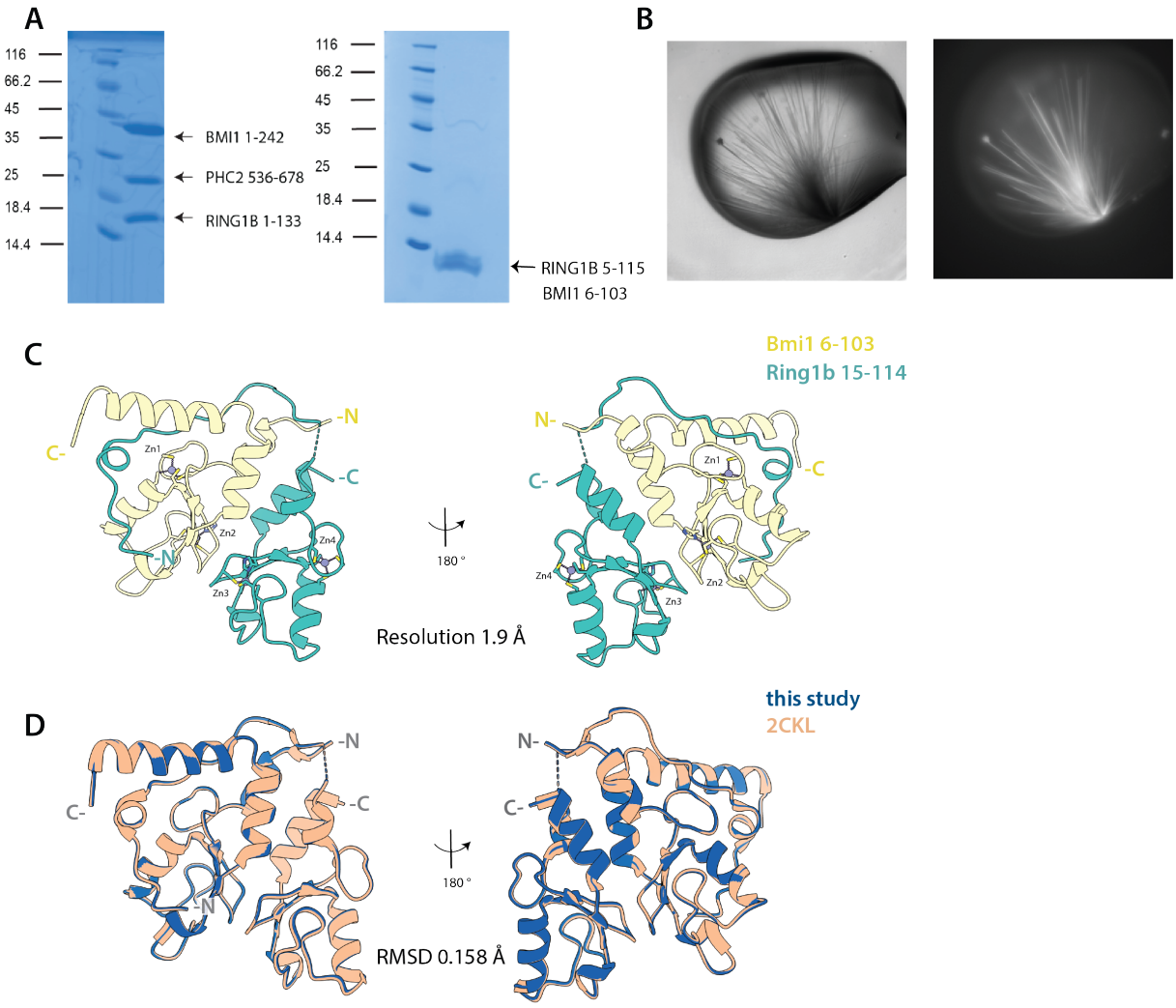


Figure 13: Crystallization of minimal PRC1-PHC2 complex. A. Coomassie stained 16% SDS-PAGE analysis of SEC fractions underwent crystallization (left) and obtained protein crystal (right), showing the degradation of minimal PRC1-PHC2 during crystallization. B. Protein crystals of degraded minimal PRC1-PHC2 complex at 21 days of crystallization, imaged with VIS (left) and UV (right) light. C. Structure of human PRC1 E3 ligase module (Ring1b 15-114, Bmi1 6-103), resulted from crystallization trial of minimal PRC1-PHC2 complex resolved at 1.9 Å. D. Comparison of human PRC1 E3 ligase module crystal structure from this study and previously published crystal structure of PRC1 E3 ligase module from Mus musculus (PDB 2CKL, Buchwald et al. (2006)). Structures were aligned based on Bmi1 subunit and root-mean-square deviation across all atoms calculated using the Matchmaker tool in ChimeraX (Goddard et al., 2018).

4.1.3 PRC1-PHC2 subcomplex with full-length Ring1b and Bmi1 purified from insect cells is and active E3 ligase and binds nucleosome

After attempting crystallisation with a minimal PRC1-PHC2 subcomplex, the complex containing full-length Ring1b and Bmi1 and extended PHC2 was expressed for cryo-EM studies. Expression trials for Ring1b and Bmi1 constructs of longer length in bacteria resulted in degradation to the length of their known domains (Ring1 1-133 and Bmi1 1-242), and reconstitution of full-length Ring1b and Bmi1 required a change of the expression system. Additionally, I decided to extend PHC2 beyond FCS domain for PHC2 to encompass the entire C-terminal portion, including the SAM domain. As the SAM domain was previously shown to form oligomers, I introduced mutations in the EH polymerisation surface of the SAM domain to prevent self-oligomerization of PHC2 (Gambetta and Müller, 2014, Frey et al., 2016). The PHC2 mutant containing two mutations in the SAM domain (L842E and L846E) is reffered herein as PHC2 2EH.

To reconstitute the PRC1-PHC2 complex, which consisted of the complete Ring1b and Bmi1 subunits along with a truncated embryonic form of PHC2 lacking the N-terminal tail but retaining the C-terminal SAM domain, a baculovirus expression system was employed (Method Sections 3.3 - 3.3.4). The Gibson assembly method as described in Weissmann et al. (2016) was used to clone TwinStrep-Ring1b and Bmi1-His into a polycistronic pBig1a vector. Although the bigBac system allows for the expression of up to 25 subunits from a single baculoviral expression vector, expressing all three subunits as multigene constructs resulted in satisfactory expression levels of Ring1b and Bmi1 but low expression of PHC2. As a result, I co-expressed His-TwinStrep-Ring1b and Bmi1-His as a single pBig1a construct, while His-PHC2-SAM-EH was cloned into a separate pFastBac1 vector and expressed individually. Figure 14A (right panel) depicts the schematic representation of the expression constructs. Compared to previously described purification of minimal PRC1-PHC2 complex, MonoS cation exchange chromatography was used to remove contamination with nucleic acids instead of heparin.

In Figures 14A-C, an overview of the purification process is presented, which includes corresponding SDS-PAGE gels and chromatograms. Briefly, Twin-Streptag-Ring1b and Bmi1-His dimer and His-PHC2 were pulled down from the lysate using Ni-NTA resin, elution fraction mixed and tags were cleaved off with 3C protease during overnight dialysis (Figure 14A, left panel). C-terminal His-tag on Bmi1 was uncleavable to ensure the stability of the Bmi1 C-

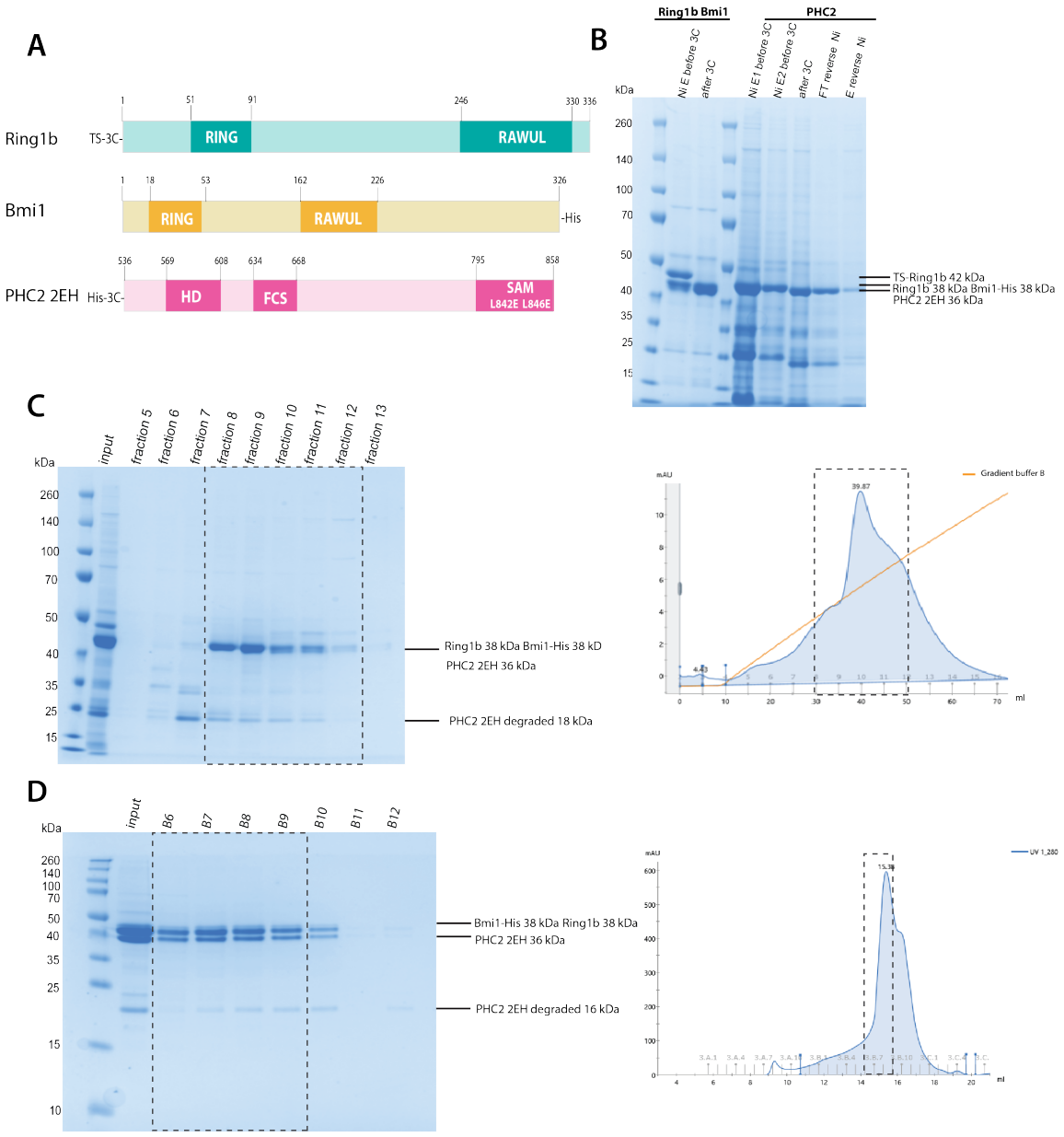


Figure 14: Purification of PRC1-PHC2 complex from insect cells. A. Schematic representation of the PRC1-PHC2 domain organisation and affinity tags used for purification. B. Coomassie stained 4-12% SDS-PAGE gel of Ni pulldown fractions of his-tagged PHC2, Ring1b and Bmi1 proteins. B. Heparin chromatography, left panel shows Coomassie stained 4-12% SDS-PAGE analysis of elution fractions, corresponding chromatogram is shown on the right. Fractions used for SEC are marked with a box. C. 16% SDS-PAGE analysis of SEC elution fractions (left) and corresponding SEC chromatogram (right).

terminus. The fraction after cleavage was then concentrated and applied to MonoS cation exchange chromatography to remove contamination with nucleic acids (Figure 14B). Finally, the complex was then subjected to SEC on Superose 6 Increase 10/300 GL column (Figure 14C). The purification resulted in a complex of adequate purity, as determined by SDS-PAGE analysis. However, PHC2 protein was present in its both intact and truncated forms (degradation product was determined by mass spectrometry to be 538-678 AA, essentially mimicking PHC2 construct used for bacterial expression), which introduced a certain degree of heterogeneity (as seen on the SDS-PAGE gel in Figure 14C, left panel; compare the bands around 40 kD and a lighter band around 25 kDa that corresponds to PHC2 2EH degradation product). It should be noted, that this degradation did not occur during the purification process, rather the truncated PHC2 form was already present in the lysate, as demonstrated by the Ni-pulldown. Since all three proteins exhibit comparable molecular weight and migration patterns on SDS-PAGE, to confirm the presence of all three target proteins, the peak fraction underwent intact mass spectrometry (MPIB Mass spectrometry facility). This revealed the presence of all three proteins of interest, with their theoretical molecular weight matching their measured values. Additionally, a truncation of PHC2 was detected. Subsequent PHC2 purification therefore either contained the shorter form of PHC2 corresponding to truncation or, as in the case of PRC1-PHC2-Scml2 complex, intact PHC2 was pull-downed utilizing SAM-SAM interaction of PHC2 and Scml2.

4.1.4 Identification of protein interactions of PRC1-PHC2 complex alone and bound to nucleosome by XL-MS

For the identification of interaction interfaces within different PRC1 subcomplexes studied in this thesis, chemical cross-linking with bis(sulfosuccinimidyl)suberate (BS3) and mass spectrometry analysis (XL-MS) was used. BS3 molecule contains two NHS-ester groups that covalently link primary amines from proximal lysines by forming an amide bond. BS3 has a spacer arm of 11.4 Å, therefore Cα atoms of the two cross-linked lysines could be between 10 Å to 24 Å apart, depending on lysine conformation. Usually, a distance restraint of 30-35 Å is applied when evaluating BS3-crosslinked data when taking into account the flexibility of the lysine and the protein backbone (Gong et al., 2020). Cross-linking experiments were performed in duplicates and only cross-links present in both replicas were used for further analysis. These cross-links depicted in the cross-linking network plot generated by xiView (Graham et al., 2019).

The Coomassie-stained SDS-PAGE gel for PRC1-PHC2 complex before and after cross-

linking with BS3 is shown in Figure 15A and the connectivity map is depicted in Figure 15B (Graham et al., 2019). Overall, the cross-linking dataset for the PRC1-PHC2 complex contained 108 unique cross-linked peptide pairs, 47 of them within the same subunits (intramolecular cross-links, shown in purple in the connectivity map). First, the cross-linking approach was validated by mapping cross-linked peptide pairs to previously published crystal structures of Ring1b-Bmi1 E3 catalytic domain and RAWUL domain of Bmi1 using XMAS bundle in ChimeraX (Buchwald et al., 2006, Gray et al., 2016, Lagerwaard et al., 2022). Visualisation of detected cross-links on 3D models showed an excellent agreement with distance restraints posed by BS3. The only cross-link that violated the upper distance restraint is Bmi1 K88 - Ring1B K97, being 38 Å apart, most likely reflecting a possible movement of the C-terminal Bmi1 α-helix (Figure 15C).

The most striking observation made from this cross-linking data is that numerous cross-links were detected between PHC2 (HD and FCS domains) and both Ring1b and Bmi1 (Figure 15B). PHC2 FCS domain cross-links to Ring1b RAWUL and Ring domains, as well as to Ring and RAWUL domains of Bmi1. Even though PHC2 HD domain was shown to associate with Bmi1 RAWUL (Gray et al., 2016), it also cross-links to Ring domains of both Ring1b and Bmi1, as well as to Ring1b RAWUL. This and the fact that in some instances the same lysine residue (such as K63 located in the proximity of HD domain and K140 in the proximity of FSC domain) was involved in cross-links to both Ring1b and Bmi1, suggests that PHC2 subunit is conformationally dynamic in solution, which, as will be shown in the Results section 4.4.1, is also the case in PRC1-Scml2 subcomplex as well. The spatial proximity between the FSC domain of PHC2 and the Ring domain of Ring1b had previously been identified in XL-MS experiments conducted on a similar PRC1 complex that included Ring1b, Cbx2, MEL18, PCGF2 (MEL18), and PHC2. In that study, however, no cross-links between PHC2 FSC and RAWUL Ring1b were detected, as well as no cross-links involved HD domain of PHC2 (Colombo et al., 2019). This discrepancy could be attributed to a different composition of the complex or possibly due to a lower amount of protein used for cross-linking experiments in that study or a lower amount of cross-links identified. The latter might result from variations in LC-MS/MS setups or the absence of cross-link enrichment through peptide fractionation.

Additionally, intramolecular cross-links were detected between the Ring and RAWUL domains of Ring1b, pointing towards the spacial proximity of these domains (Figure 15B, Ring1b sequence, intramolecular cross-links shown in purple). Several cross-links between the RAWUL domain of Bmi1 and the Ring domain of Ring1b were detected, as well as between the RAWUL

domains of both proteins. The predicted disordered C-terminal region of Bmi1 (starting from residue 240) contains only two lysines, and neither was cross-linked.

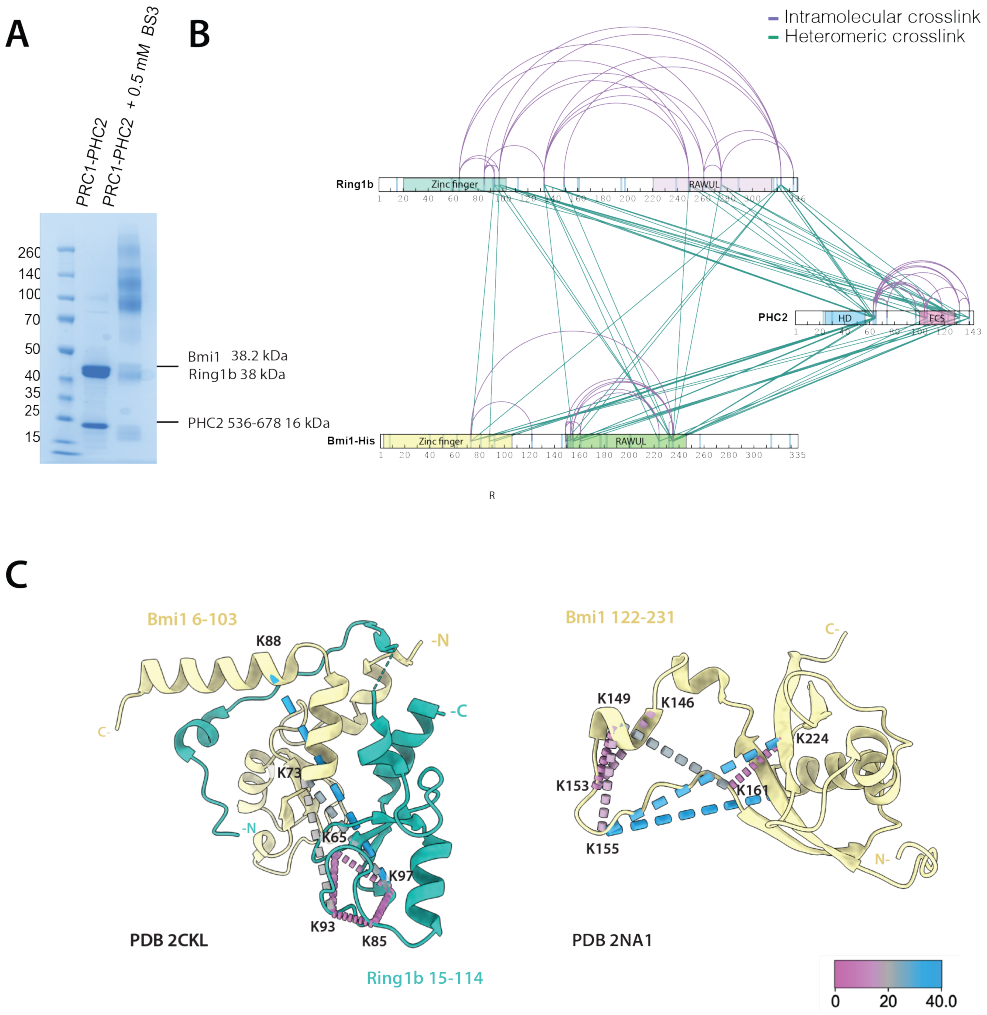


Figure 15: Cross-linking mass spectrometry analysis of PRC1-PHC2 complex. A. Coomassie-stained 16% SDS-PAGE analysis of PRC1-PHC2 complex intact and crosslinked with 0.5 mM BS3 for 30 min at RT. B. Intramolecular (in purple) and heteromeric (in green) BS3 cross-links mapped on protein sequence of PRC1-PHC2 subunits. Cross-linking mass spectrometry experiments were performed in duplicates, cross-links present in both replicas were mapped onto corresponding protein sequences in XiView. C. Validation of cross-linking experiment by mapping cross-links onto crystal structures of Ring domains of Ring1b and Bmi1 (top panel, PDB 2CKL) and RAWUL domain of Bmi1 (bottom panel, PDB 2NA1).

XL-MS of PRC1-PHC2 in the presence of nucleosomes

At the time of conducting this study, no XL-MS experiments for any PRC1 complex with nucleosomes were reported in the scientific literature. To investigate the interactions of the complex with nucleosomes, a mixture of PRC1-PHC2 and nucleosomes was subjected to XL-MS, using the same cross-linking protocol (see Figure 16A for SDS-PAGE of the cross-linked sample). Of particular interest was the investigation of whether the PHC2 subunit or the C-terminal domains of Ring1b and Bmi1 came into proximity with the nucleosome, as the only mode of interaction between PRC1 and nucleosome described so far was through Ring:Ring dimer of Ring1b and Bmi1. The FSC zinc finger of PHC2 was shown to interact with nucleic acids unspecifically, therefore, one could hypothesise that it could also form interactions with nucleosomal DNA, bringing PHC2 subunit to the proximity of nucleosome, in which case it would likely be possible to detect cross-links between PHC2 and histones.

A total amount of 81 unique cross-links were identified in both replicas. Notably, the pattern of intramolecular cross-links for all PRC1 proteins remained similar as in the PRC1-PHC2 complex without nucleosomes. Namely, intramolecular cross-links between Ring1b Ring finger and RAWUL domains, cross-links within the RAWUL domain of Bmi1, and within the C-terminal portion of PHC2 FCS were detected (Figure 16B). This might suggest that the conformation of individual PRC1 subunits remains largely unchanged upon binding to the nucleosome, as no detected intramolecular cross-links were deviating from cross-links previously detected for PRC1-PHC2 only. PHC2 subunit was again shown to cross-link extensively to both Ring1b and Bmi1, but no cross-links of PHC2 to histones were detected in any experiment replicas. The cross-links of FCS domain of PHC2 to the central part of Ring1b RAWUL domain, which were previously detected in the PRC1-PHC2 sample, were not detected in the PRC1-PHC2 sample with nucleosomes, while cross-links to the C-terminal part of RAWUL domain remained unchanged.

As for interaction surfaces of PRC1-PHC2 with the nucleosome, the only cross-links detected were between Ring Ring1b domain (residue K97) and residue K124 of histone H2B and residue K10 of histone H4. These cross-links are in line with the previously reported mode of interaction of Ring1b with the nucleosome acidic path (McGinty et al., 2014). No additional interaction surfaces between PRC1-PHC2 and nucleosomes were detected.

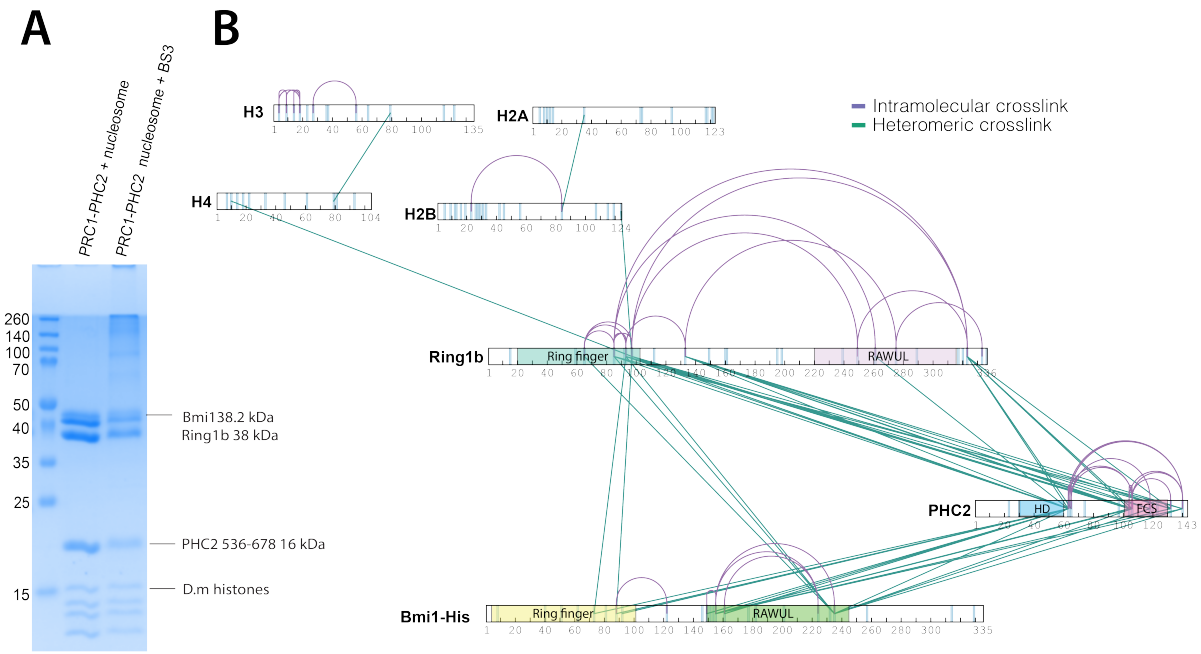


Figure 16: Cross-linking mass spectrometry analysis of PRC1-PHC2 complex and nucleosomes. A. SDS-PAGE of PRC1-PHC2 complex and nucleosome intact and crosslinked with 0.5 mM BS3 for 30 min at RT. B. Intramolecular (in purple) and heteromeric (in green) BS3 crosslinks mapped onto protein sequence of PRC1-PHC2 subunits and histones. Cross-linking mass-spectrometry experiments were performed in duplicates, and cross-links present in both replicas were mapped onto corresponding protein sequences in XiView (Graham et al., 2019).

Mapping of XL-MS PRC1-PHC2 on AlphaFold2 predictions

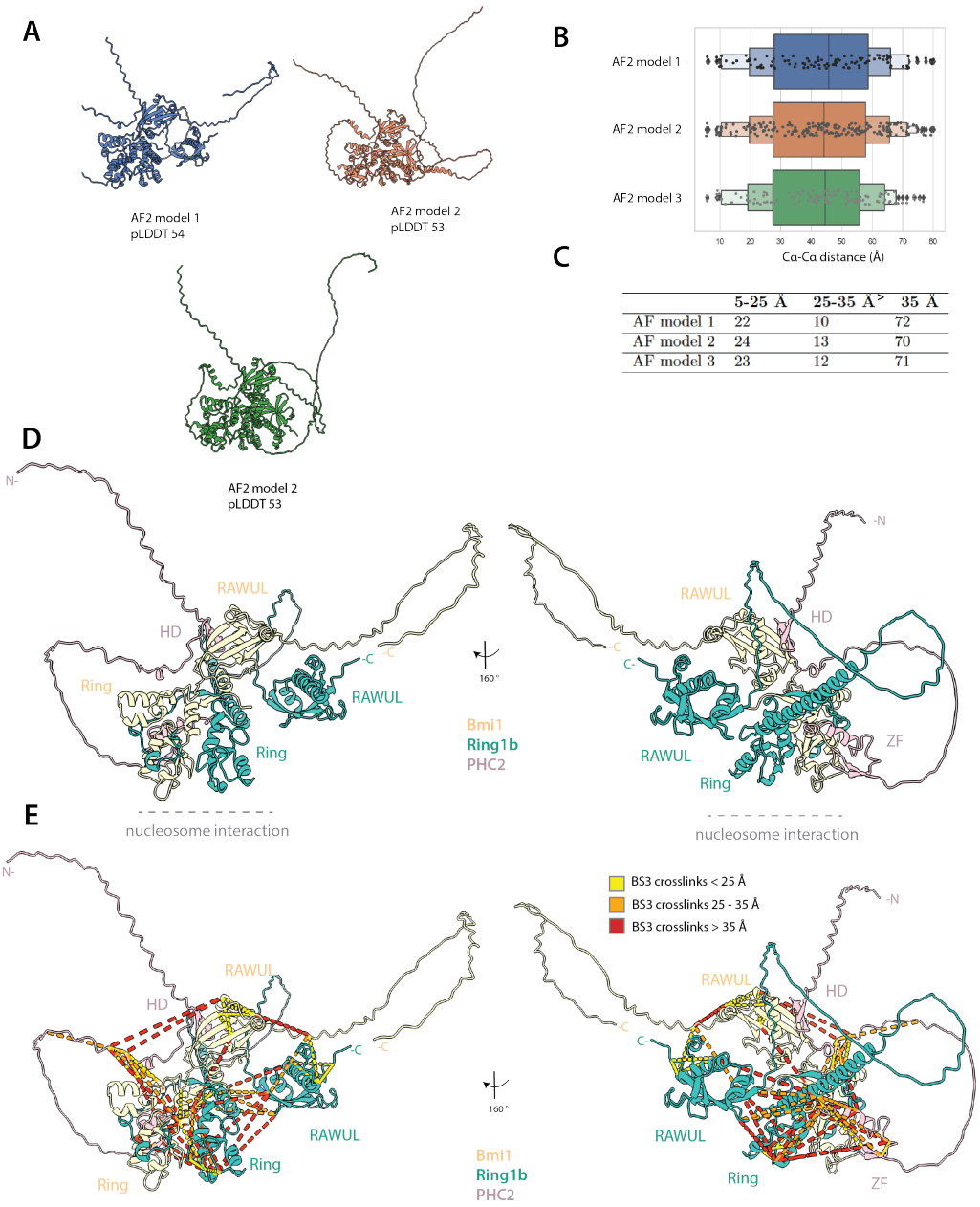
To assess the agreement between the experimental XL-MS data and *in-silico* structural predictions, AlphaFold2 (AF2) was utilized in multimer mode to generate a set of 25 models of PRC1-PHC2 complex. The global pLDDT (predicted local distance difference test) scores spanned from 54 to 39 on a scale from 0 to 100, indicating a moderate level of prediction confidence. Upon alignment and visual examination of all the models, it became evident that there was significant variation among the predicted models, as only the Ring-Ring dimer of Ring1b and Bmi1, along with the RAWUl domain of Bmi1 had consistent spacial position. Three top-scored models are depicted in Figure 17A and model 1 is annotated in Figure 17D. When compared to previously published structural data, an additional structured element was predicted, namely a long α-helix adjacent to Ring1b Ring domain (residues 117-156). In Figure Figure 17D, this long Ring1b α-helix could be seen stacked on top of the Ring:Ring domains of Ring1b and Bmi1. For other regions of the complex whose structure was not previously reported, the N-terminus of PHC2, the C-terminus of Bmi1 and a region between Ring1b α-helix and RAWUL domains were predicted to be unstructured.

XL-MS cross-links were then mapped onto three top-scored AF2 models using XMAS bundle in ChimeraX (Lagerwaard et al., 2022, Goddard et al., 2018). Distance between Cα of cross-linked residues was used to evaluate how well a particular model fits the experimental data. It is worth noting that the theoretical maximum Cα-Cα distance between lysines cross-linked by BS3 is 24 Å when the BS3 cross-linker is fully extended. However, empirical evidence has shown that many experimental cross-links exceed this 24 Å threshold and usually a distance restraint of 30-35 A is applied accounting for the flexibility of the lysines and the protein backbone. Consequently, I categorized the cross-links into distinct classes based on their distances: 5-25 Å for cross-links that aligned with the theoretical distance restraint, 25-35 Å for cross-links that were experimentally feasible considering conformational dynamics, and >35 Å for cross-links that violated the permitted cross-linking range. The distribution of Cα-Cα distances between cross-linked residues in various models is illustrated in Figure 17B. Notably, there was no clear preference for a single model in terms of agreement with the experimentally detected cross-links. Each model exhibited a wide range of Cα-Cα distances, spanning from under 10 Å to 80 Å, with medians of approximately 46 Å for AF2 model 1, 44 Å for AF2 model 2, and 45 Å for AF2 model 3. This observation indicated that for each model, the majority of cross-links fell into the violated (>35 Å) category, which is beyond the permissible cross-linking range. For a summary

of the number of cross-links within each category, refer to Figure 16C.

The cross-links mapped onto AF2 model 1 are shown in Figure 16E. Most of the permitted cross-links of Cα -Cα distance less than 25 Å (shown in yellow) are located in structured domains, such as Bmi1 RAWUL, Ring1b RAWUL, Ring1b Ring and PHC2 FCS. Violated cross-links (shown in red) and permitted cross-links of longer Cα-Cα distance (shown in orange) connect different domains or interdomain regions, such for example, cross-links between Ring1b Ring and PHC2 FCS and between Ring1b Ring and RAWUL domains. Thus, the variability in conformational states that the complex which could result in a larger distance between domains, likely accounts for the considerable number of cross-links classified as violating the distance restraint (>35 Å).

Collectively, these findings and comparison with the predicted AF2 models suggest the inherent flexibility of the PRC1-PHC2 complex, indicating that the XL-MS data likely represent an ensemble of conformational states of the complex and no individual AlphaFold2 prediction could, therefore, comprehensively depict the structural landscape of the PRC1-PHC2 complex.



	5-25 Å	25-35 Å>	35 Å
AF model 1	22	10	72
AF model 2	24	13	70
AF model 3	23	12	71

Figure 17: A. AlphaFold2 models of PRC1-PHC2 complex with highest overall pLDDT scores. B. Cα-Cα distance distribution of BS3 crosslinks mapped on AlphaFold2 models of PRC1-PHC2, visualized with XMAS. C. Overview of the BS3 cross-link counts categorized based on various Cα-Cα distances, when mapped on AlphaFold2 models of PRC1-PHC2. D. Annotated representation of Alphafold2 model 1 of PRC1-PHC2 complex. E. BS3 cross-links mapped on PRC1-PHC2 AlphaFold2 model 1. The crosslinks are colour-coded to indicate their respective Cα- Cα distances: violated cross-linking ($>$ 35 Å) shown in red, permitted cross-linking distances (5-25 Å) in yellow and cross-links within the range of 25-35 Å depicted in orange. Visualized with XMAS implementation in ChimeraX (Lagerwaard et al., 2022).

4.1.5 Cryo-EM reconstitution of PRC1-PHC2 on nucleosome showed density for E3 ligase domain

The PRC1-PHC2 complex bound to a nucleosome was analyzed using cryo-electron microscopy. The PRC1-PHC2 complex was mixed with recombinantly-reconstituted biotinylated *Drosophila melanogaster* nucleosomes, and cryo-EM grids were prepared following the protocol described in Method section 3.8.2. Biotinylation of nucleosomal DNA was required for binding to streptavidin-coated grids. The usage of streptavidin-coated grids simplified the sample preparation process and resulted in homogeneous particle distribution (Figure 18A). Before utilizing streptavidin-coated grids, regular holey carbon grids (Quantifoil R1.2/1.3 200 Mesh Copper grids) were tested, yet to unsatisfactory results, such as uneven particle distribution, aggregation and the tendency for particles to adhere to the carbon surface.

The dataset consisting of 24,700 movies was collected on Titan Krios (with the great help of MPIB Cryo-EM facility) and processed using cryoSPARC version 3.1 (Punjani et al., 2017). The data processing scheme and the final 3D reconstructed map are depicted in Figure 18A. The particles were initially picked using the blob picking tool in cryoSPARC and used for training a TOPAZ neural model that was implemented to re-pick particles across the whole dataset with fewer false-positive particle picks and a total amount of 977,610 particles (Bepler et al., 2019). The representative 2D classes are depicted in Figure 18A, clearly showing nucleosomes with density protruding on the sides. Particles corresponding to the final 2D classes underwent several rounds of heterogeneous 3D refinement and the 3D classes with the best overall resolution and PRC1-PHC2 occupancy on nucleosomes were selected. Although PRC1 could bind to both sides of the nucleosome, it became evident that in this 3D reconstruction, PRC1-PHC2 density is notably more pronounced on one side (compare PRC1-PHC2 bound to proximal and distal sides on 3D reconstruction after heterogeneous refinement, Figure 18A). This asymmetry in PRC1 occupancy on nucleosomes might be due to factors like a low complex concentration in the sample and additional washing steps during grid preparation, which could further reduce the PRC1 concentration. Consequently, the refinement was focused on a nucleosome with one side bound by PRC1, by using a soft mask covering the nucleosome and proximal-bound PRC1-PHC2 in homogeneous and non-uniform refinement steps (Figure 18A). The final 3D reconstruction (from 123,024 particles) was resolved at an overall resolution of 3.2 Å (according to the gold-standard FSC cut-off at 0.143 (Rosenthal and Henderson (2003), Figure 18C), with a local resolution ranging from 2.87 Å for nucleosomal core, to 15 Å for nucleosomal DNA ends (map coloured by

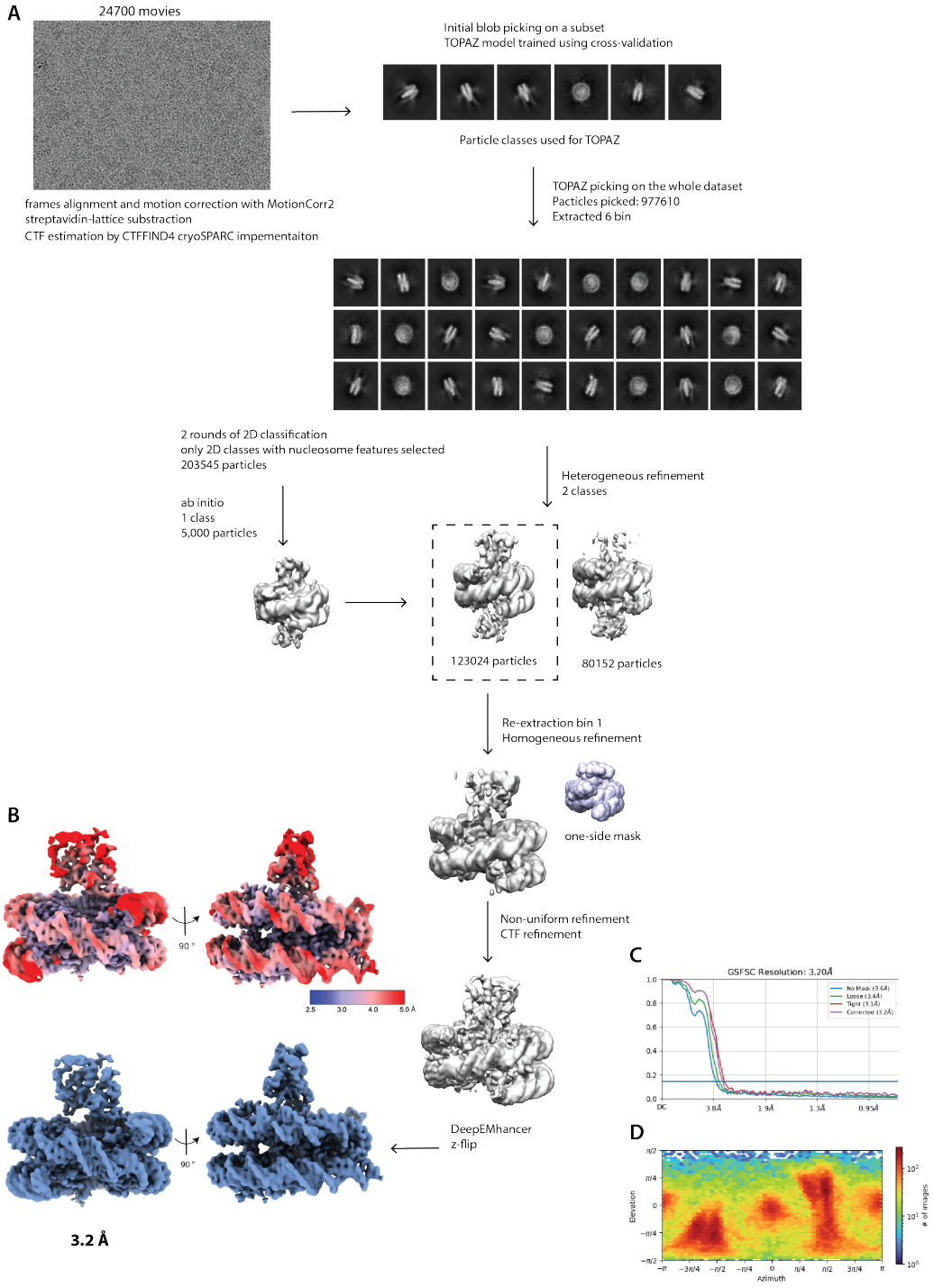


Figure 18: Cryo-EM analysis of PRC1-PHC2 complex. A. Single particle cryo-EM processing scheme of PRC1-PHC2 dataset. A more detailed description is provided in the Methods section 3.8.5. B. 3D reconstructed map coloured by local resolution, with a resolution range from below 3 Å for nucleosome core and ¿ 5 Å for PRC1 E3 ligase domain and DNA ends. C. Gold standard Fourier shell correlation (FSC) plot for the final reconstruction, resolution estimated at FSC = 0.143. D. Angular distribution of particle projections used in 3D reconstruction, calculated by cryoSPARC version 3.1 (Punjani et al., 2017). The heat map displays the number of particles corresponding to each viewing angle, with blue indicating fewer particles and red indicating a greater number of particles. Angular distribution calculated in cryoSPARC for Cdiff EσA /Fdx particle projections. The heat map shows the number of particles for each viewing angle (less = blue, more = red).

local resolution is shown in Figure 18B). The final map was post-processed using DeepEMhancer (Sanchez-Garcia et al. (2021), Figure 18B, bottom panel). The final map was also subjected to z-flip transformation to correct the handedness of the 3D reconstruction (Goddard et al., 2018).

The final 3D reconstruction contained the density corresponding to the nucleosome but only a portion of the PRC1-PHC2 complex, namely, the PRC1 E3 ligase module (Ring domains of Ring1b and Bmi1, as can be seen bound to nucleosome in Figure 18B). No density for the C-terminal parts of Ring1b and Bmi1, nor for the entire PHC2 subunit was observed at any step of the processing. As the sample integrity was confirmed by SDS-PAGE just before the sample preparation and it contained all full-length subunits of the PRC1-PHC2 complex, it is unlikely, that this was caused by sample degradation. The more probable explanation would be a high degree of conformational flexibility beyond the E3-ligase domain, that could not be averaged in a 3D reconstruction.

The model was built based on the existing crystal structure (4R8P, McGinty et al. (2014)) of the PRC1 E3 ligase module on a nucleosome using a combination of manual density-fitting in ChimeraX and PHENIX real space refinement (Goddard et al., 2018, Liebschner et al., 2019). The map to model fit showed an FSC 0.5 at 3.4 Å, with good correspondence to the overall resolution (Figure 19). The Bmi1 and Ring1b subunits were found to make contact with nucleosome acidic path (depicted in Figure 19C), as was reported before in McGinty et al. (2014), PDB ID: 4R8P. Herein, the Ring1b arginine hook composed of a set of positively charged amino acids (R81, K85, K93, R98, K65) makes contact primarily with amino acids of H2A (E64, D72, E61, D90) and glutamic acid 105 of H2B (see Figure 19C, left close up view). Additionally, the Bmi1 subunit was found to contact several histones, such as aspartic acid 77 of histone H3, glutamic acid in position 74 of histone H4, and alpha helix of histone H2B (Figure 19C, right panel).

Comparative analysis between the cryo-EM map-based model and the 4R8P crystal structure showed a close agreement, with a Root Mean Square Deviation (RMSD) across all Cα-atoms of 0.693 Å (Figure 19D). Additionally, the binding sites of Ring1b and Bmi1 on the nucleosome in both structures were aligned and revealed no discrepancy in nucleosome binding (Figure 19E). Therefore, it can be inferred that the cryo-EM-based model obtained in this study, which comprises only the E3 ligase module of PRC1, closely mirrors the crystal structure of the E3 ligase module bound to the nucleosome as reported by McGinty et al. (2014).

Notably, the crystal structure 4R8P contained only the Ring fingers of Ring1b and Bmi1, with

Ring1b residues being C-terminally fused to the E2 conjugating enzyme UbcH5c.The sample analysed herein contained full-length Ring1b, Bmi1 and a short isoform of PHC2. Alignment between binding modes of these two structures also points that the binding to nucleosome seems to be unaffected by the presence of C-termini of Ring1b and Bmi1, and additional subunit PHC2 and therefore might constitute the ubiquitous nucleosome-binding mode across all canonical PRC1 complexes, independent of additional subunits such as PHC or Cbx.

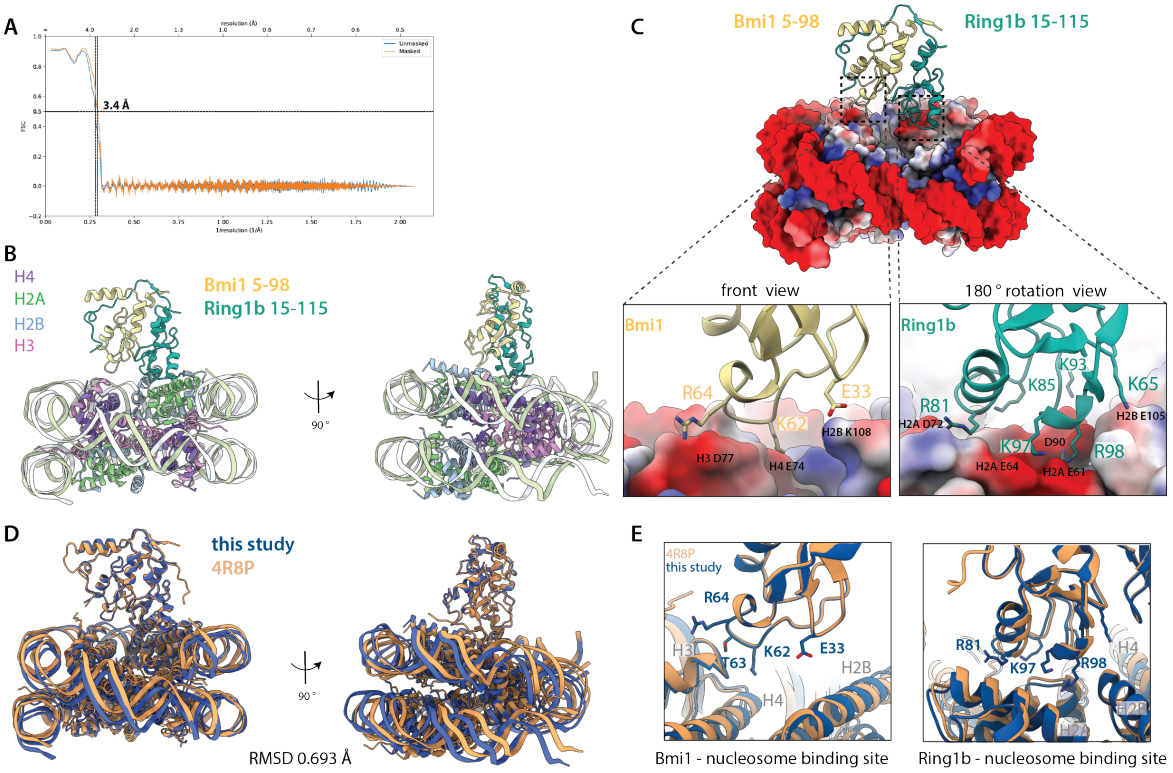


Figure 19: Cryo-EM based model of PRC1-PHC2 E3 ligase module bound to nuclesoome. A. Map-model FSC curve, calculated by Phenix validation tool. FSC=0.5 at the spatial frequency of 0.31 Å-1 or 3.2 Å resolution. B. Final model of E3-ligase module on *Drosophila melanogaster* nucleosome, obtained from the reconstitution of PRC1-PHC2 complex. C. Bmi1 5-98 (in yellow) and Ring1b 15-115 (in turquoise) interactions with an acidic patch of nucleosome (top). The nucleosome is depicted in surface mode and coloured by electrostatic potential (calculated in ChimeraX). The bottom panels show a close-up of key interactions of Bmi1 (left, yellow) and Ring1b (right, turquoise) with histones. D. Comparison of cryo-EM derived model of PRC1-PHC2 E3 ligase module bound to nucleosome with the previously published crystal structure (4R8P). A comparison was made only with the proximal E3 binding site, E2-UbcH5c fused to Ring1b that was present in 4R8P was omitted for comparison. E. Close-up of key Bmi and Ring1b interactions in models derived by this cryo-EM study and previously available crystal structure (4R8P, McGinty et al. (2014)).

4.2 PRC1-Cbx7 complex

As PHC2 failed to stabilize the entire PRC1 complex for structural studies, I chose to reconstitute another subcomplex, containing Cbx protein instead of PHC2 as a third component together with Ring1b and Bmi1. As was described in the Introduction sections, Cbx proteins interact with RAWUL domain Ring1b through interaction with C-box domain. This interaction was shown to have a lower K_d than the one of RAWUL Bmi1 and HD of PHC2, having as low as 9 nM K_d for Cbx7 (Wang et al., 2008). Additionally, the Cbxs proteins contain Chromodomains, that bind trimethylated histones, with H3K27me3 histones being of most biological relevance. Even though the K_d of Cbx7 interaction with H2K27me3 peptide is in the micromolar range, it could still possibly bring about additional stabilization for structural studies (Bernstein et al., 2006).

Expression of several His and Strep-tagged human Cbx homologs (Cbx2, Cbx4, Cbx6, Cbx7, Cbx7) was tested throughout this study, resulting in the same issue of low expression levels and degradation at a region in proximity of Chromodomain. Intact Cbx8 protein could be still isolated from nuclear extracts but in amounts insufficient for structural studies (data not shown). As the issue of degradation was located in the long region connecting the Chromodomain and C-box domain, a shorter Cbx7 isoform was used (UniProt accession number: B0QYP2). The region between Chromodomain and C-box domains in this is 100 AA shorter than in full-length Cbx7. Short Cbx7 isoform, which is further referred to as just Cbx7 for simplicity, was expressed in sufficient amounts without signs of degradation and together with Ring1b and Bmi1 subunits formed a complex, which is referred to as PRC1-Cbx7 subcomplex.

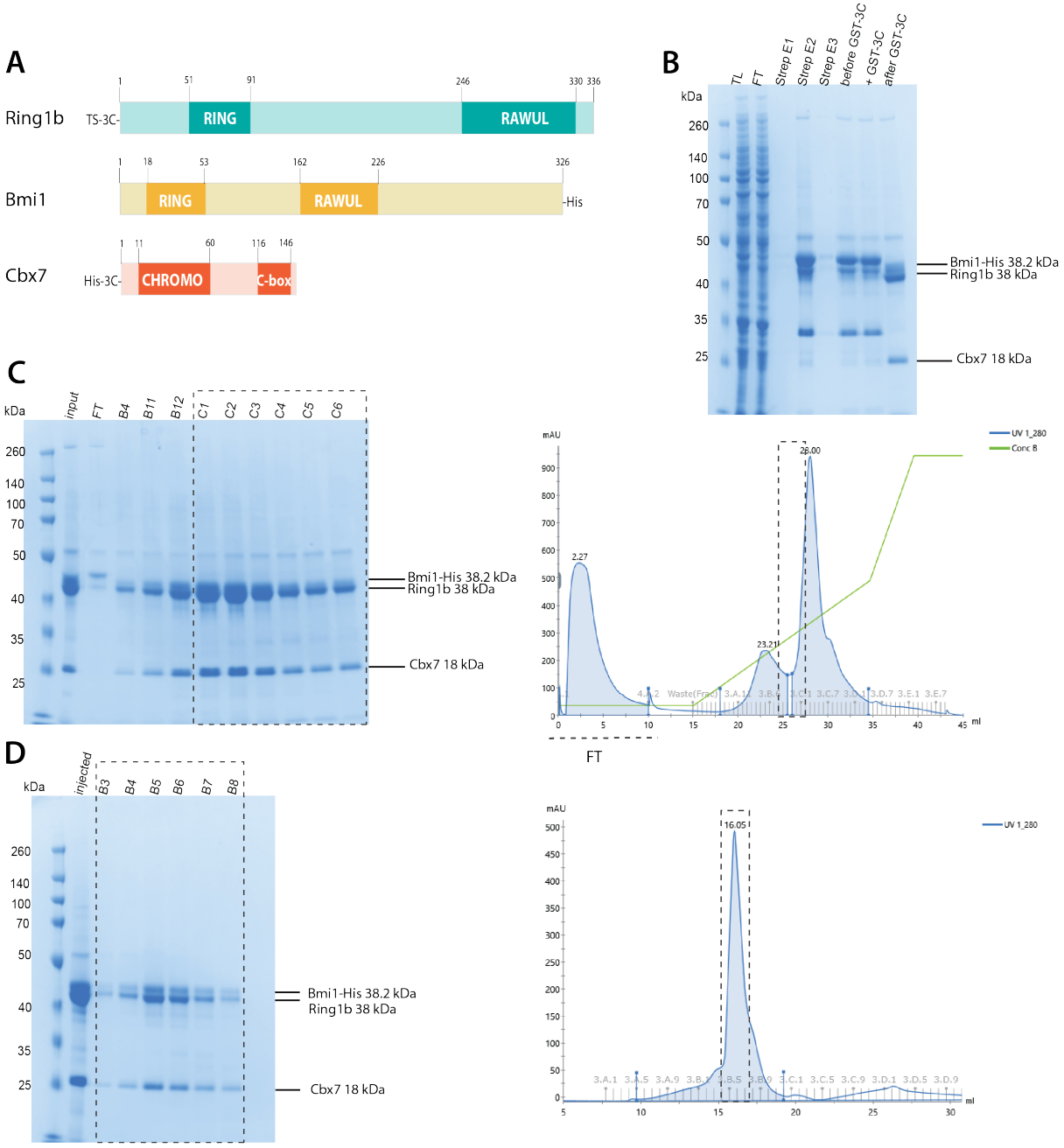


Figure 20: Purification of PRC1-Cbx7 subcomplex from insect cells. A. Schematic representation of PRC1-Cbx7 domain organization and tags used for purification. B. Coomassie stained 16% SDS-PAGE analysis of Strep-pulldown elution fractions and 3C protease cleavage. C. MonoS cation exchange chromatography. The left panel represents a Coomassie-stained 4-12% SDS-PAGE analysis of MonoS elution fractions, right panel shows a corresponding chromatogram. Elution fractions used for SEC are marked with a box. C. SEC profile (right) and Coomassie stained 4-12% SDS-PAGE analysis of elution fractions (left).

4.2.1 PRC1-Cbx7 complex is an active E3 ligase and shows comparable activity on wt and H3K27me3 nucleosomes

Twin-Strep-tagged Cbx7 was co-expressed with Twin-Strep-tagged Ring1b and His-tagged Bmi1 proteins, see Figure 20A for a schematic representation of the expression construct. Stochiometric PRC1-Cbx7 subcomplex was pulled down from the cell lysate with a single Streptavidin pulldown step (Figure 20B shows Streptavidin elution fractions before and after tag cleavage with GST-3C protease). PRC1-Cbx7 complex was further purified using MonoS cation exchange chromatography and SEC (Figure 20C and Figure 20D). Similarly to the PRC1-PHC2 subcomplex, the PRC1-Cbx7 sample had a high nucleic acid content, consistent with previous reports on Cbx chromodomains exhibiting non-specific nucleic acid binding (Bernstein et al., 2006, Zhen et al., 2016). Nucleic acid contamination was successfully removed using MonoS cation exchange chromatography (Figure 20C right panel, FT peak corresponds to unbound protein impurities and nucleic acids).

The nucleosome binding and E3 ligase activity of the PRC1-Cbx7 complex were tested on both wild-type and H3Kc27me3 nucleosomes. The H3Kc27me3 nucleosomes contained an H3 histone variant with a methyl-lysine analogue (MLA) at position 27, mimicking the trimethylation of lysine 27. Methyl-lysine analogues were generated by alkylating cysteine to produce N-aminoethyl cysteine, which was demonstrated to exhibit functional similarity to naturally occurring methylated lysines within the nucleosomes (Simon et al., 2007). H3 MLA histones used for the reconstitution of H3Kc27me3 nucleosomes were prepared by Ksenia Finogenova and Claudia Litz (Finogenova et al., 2020).

The EMSA assay results showed no statistically significant distinction in the binding affinity of PRC1-Cbx7 to wild-type (wt) and H3Kc27me3 nucleosomes (Figure 21A). This finding could be attributed to the interactions between PRC1-Cbx7 and nucleosome being primarily facilitated by Ring1b and Bmi1. In line with this, previous studies demonstrated that Cbx7 interaction with H3Kc27me3 peptide had a K_d of 22 µM, which likely did not significantly contribute to the overall affinity of the PRC1-Cbx7 complex to nucleosomes (Bernstein et al., 2006).

Likewise, the E3 ligase activity of PRC1-Cbx7 was found to be unaffected by the presence of methylated lysine analogue in position 27, indicating that K27 tri-methylation of the nucleosome may not contribute to the activation of PRC1 E3 ligase activity (compare lines for H3Kc27me3 and wt nucleosomes in Figure 21B). However, it is important to note that this observation should

ideally be confirmed using naturally methylated nucleosomes.

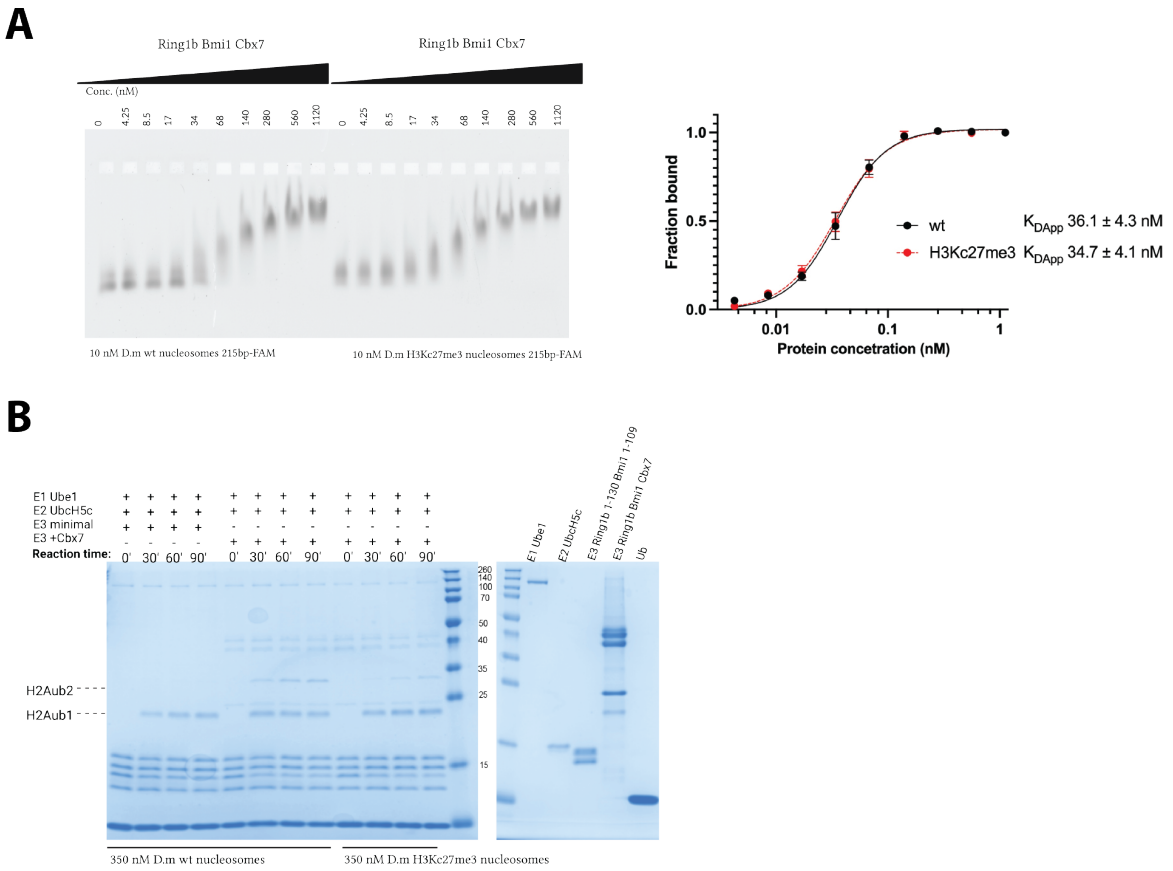


Figure 21: Biochemical characterisation of PRC1-Cbx7 subcomplex. A. Electrophoretic Mobility Shift Assay (EMSA) performed with PRC1-Cbx7 subcomplex on 215 bp 6-carboxyfluorescein labelled wild type and H3Kc27me3 *D.m* nucleosomes in triplicates (left panel). Quantification by densitometric analysis is shown on the right. B. *In vitro* H2AK118ub1 ubiquitination assay on reconstituted wild type and H3Kc27me3 D. m. nucleosomes. The right panel shows a Coomassie-stained 16% SDS PAGE gel of all proteins used in the ubiquitination reaction. The left t panel represents a Coomassie-stained 16% SDS PAGE with a time course fractions of ubiquitination reaction. The bands corresponding to ubiquitylated H2AK118ub1 and H2AK118ub2 are with dashed lines.

4.2.2 Crystallisation of PRC1-Cbx7 resulted in crystals containing E3 ligase module

Following purification and validation of its enzymatic and nucleosome-binding activities, the PRC1-Cbx7 complex was subjected to crystallization trials. Previous attempts to crystallize the PRC-PHC2 complex resulted in the degradation of the complex to Ring1b 1-130 and Bmi1 1-109, which represented only the E3 ligase domain of Bmi1. This degradation suggested that

the binding of the PHC2 subunit to Bmi1 failed to stabilize the C-terminal region of Bmi1. Given that Cbx7 exhibits strong binding to the RAWUL domain of Ring1b (with a K_d of 9 nM, Wang et al. (2008)), it was anticipated that obtaining the structure of the full-length Ring1b in complex with Cbx7 C-box domain and Bmi1 Ring domain could be feasible due to stabilisation provided by Cbx7 C-box domain.

The crystals with the same needle-like morphology as for the PRC1-PHC2 complex appeared after 10 days at RT in a crystallisation condition containing 25% PEG 4000 0.1 M Tris-HCl pH 8.8 (Figure 22A). This crystal morphology was already a strong indication that the PRC1-Cbx7 complex too degraded to minimal Ring1b 1-130 and Bmi1 1-109 E3 ligase module. This was confirmed by subjecting the crystal to silver-stained SDS-PAGE analysis with a comparison to the fraction submitted to crystallisation (Figure 22B, compare the first and the second lanes).

This observed tendency of PRC1-Cbx7 to degrade during crystallisation, therefore, prompted the decision to halt further crystallisation trials and focus on other experimental approaches, such as XL-MS and cryo-EM.

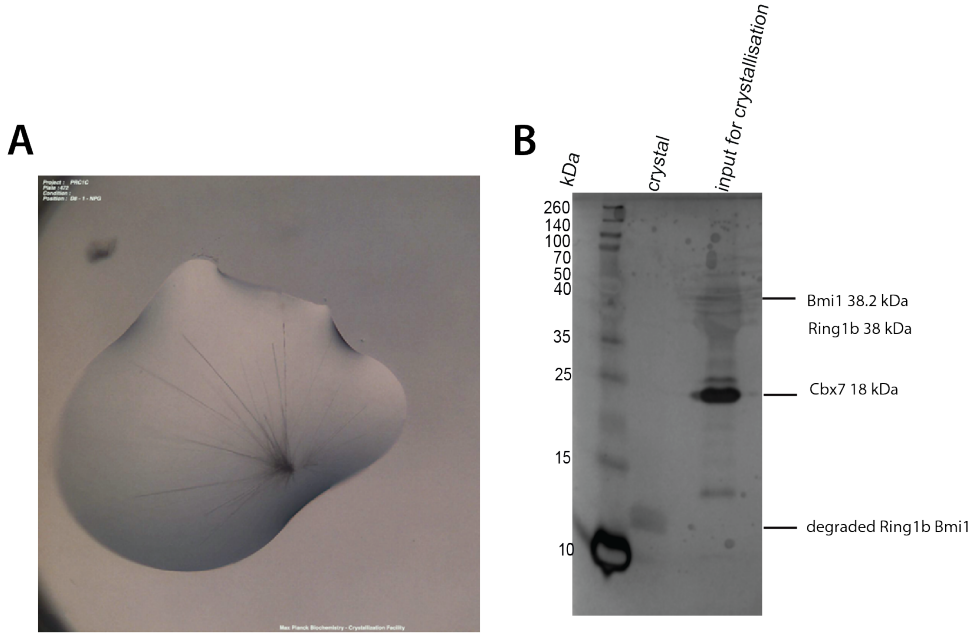


Figure 22: Crystallisation of PRC1-Cbx7 complex. A. Crystals obtained from crystallization trials for PRC1-Cbx7 complex imaged in VIS light B. Silver-stained SDS-PAGE analysis of the crystal indicated degradation PRC1-Cbx7 used for crystallization experiment. Degraded bands' molecular weight (¡ 15 kDa) indicates that crystals most likely contain only E3 ligase domain of Rin1b and Bmi1 as in previous crystallization attempts for minimal PRC1-PHC2 complex.

4.2.3 Identification of protein interactions in PRC1-Cbx7 and nucleosome-bound PRC1-Cbx7 subcomplex by XL-MS

PRC1-Cbx7 complex was subjected to cross-linking mass spectrometry alone and in the presence of H3Kc27me3-modified nucleosomes, using the same cross-liking protocol previously used for the PRC1-PHC2 complex.

For the PRC1-Cbx7 subcomplex alone, 57 cross-links were detected, out of them 31 intramolecular and 26 heteromeric. The SDS-PAGE analysis of the sample before and after BS3 cross-linking and the connectivity map are depicted in Figure 23.

Based on the enrichment of cross-links, heteromeric cross-links could be classified into three protein-protein interaction interfaces (PPIs):

1. Between Bmi1 and Cbx7 (9 cross-links): cross-links were detected between Cbx7 Chromodomain and three surfaces in in Bmi1: Ring domain (Bmi1 K88, cross-links to Cbx7 K25 and K60); N-terminus of RAWUL domain and Cbx7 Chromodomain (Bmi K155 and K161 both cross-link to Cbx7 K75, Bmi1 K182 to Cbx7 K38); C-terminus of RAWUL domain (Bmi1 K235 cross-links to Cbx7 K25, K38, K60 and K75).

2. Between Bmi1 and Ring1b (9 cross-links): cross-links between Bmi1 Ring and Ring1b Ring domains were detected, similar to those found in the PRC1-PHC2 complex (Bmi1 K62 cross-links to K65 in Ring1b, Bmi1 K73 to Ring1b K93). The second cross-linking surface was detected between Bmi1 RAWUL and Ring1b Ring and proximal amino acids (Bmi1 K235 cross-links to K85, K93, K97 and K133 in Ring, also Bmi1 K224 to Ring1b K133). One cross-link was detected between Bmi1 RAWUL and Ring1b RAWUL (Bmi1 K235 to Ring1b K273). Another single cross-link connected Ring1b C-terminus (K323) to Bmi1 in the N-terminal proximity of RAWUL (K155).

3. Between Cbx7 and Ring1b (8 cross-links): three lysine residues in Ring1b were found to cross-link with multiple residues in Cbx7 Chromodomain: Ring1b K85 and K97 in Ring domain both cross-link to K60 and K75 in Cbx7; Ring1b K133 in proximity of Ring domain cross-links to K25, K38 and K60 in Cbx7. A single cross-link between Ring1b C-terminus (K323) and Cbx7 K38 was also detected.

In Ring1b, the intramolecular cross-linking pattern (14 cross-links) remained consistent with that observed in the PRC1-PHC2 complex, wherein the cross-links between N- and C-termini

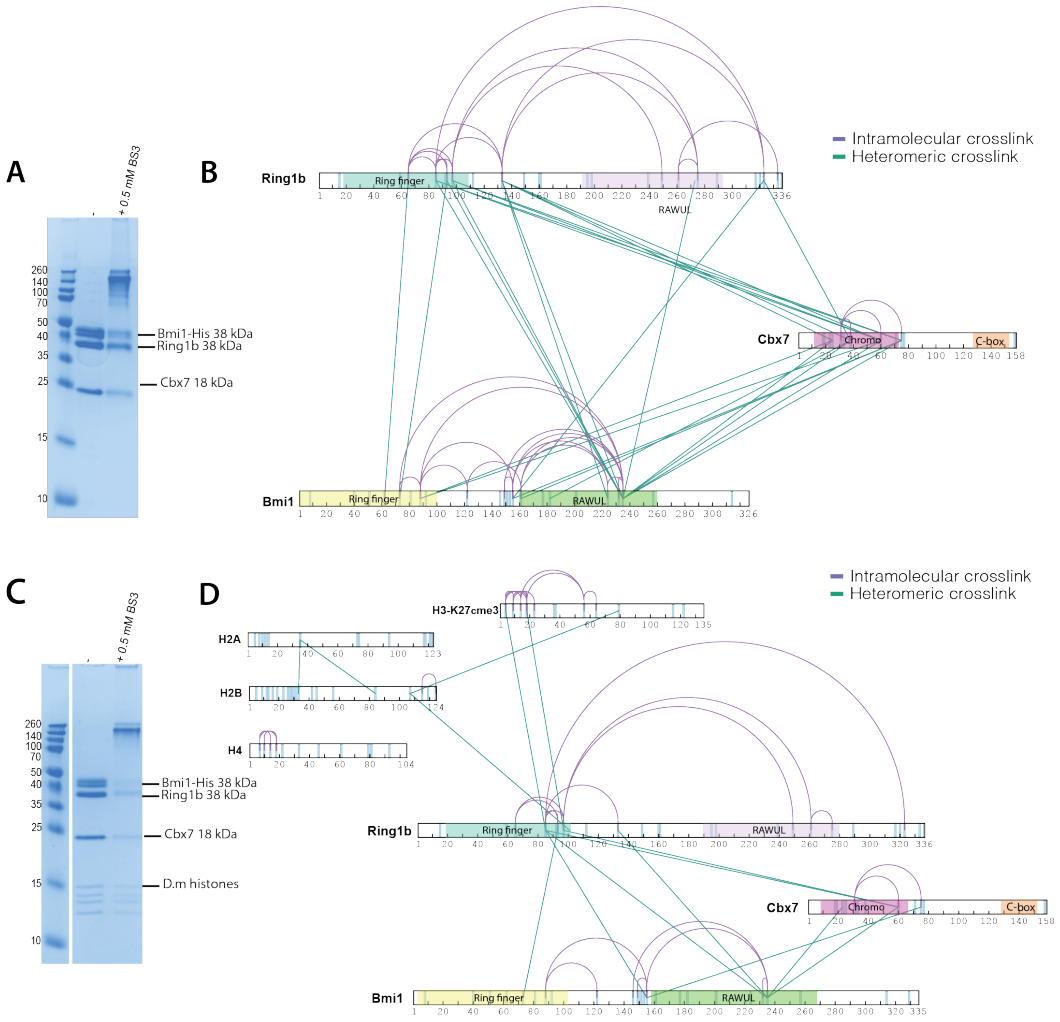


Figure 23: Cross-linking mass spectrometry analysis of PRC1-Cbx7 complex alone and with nucleosomes. A.SDS-PAGE of PRC1-Cbx7 complex intact and after cross-linking with 0.5 mM BS3 for 30°C at RT. B. Intermolecular (in purple) and intramolecular (in green) BS3 cross-links mapped on protein sequence of PRC1-Cbx7 subunits. C. SDS-PAGE of PRC1-Cbx7 complex with nucleosome intact and after crosslinking with 0.5 mM BS3 for 30°C at RT. B. Intermolecular (in purple) and intramolecular (in green) BS3 cross-links mapped on protein sequence of PRC1-Cbx7 subunits and histones. Cross-linking mass-spectrometry experiments were performed in duplicates, and cross-links present in both replicas were mapped in XiView (Graham et al., 2019).

were detected, suggesting the proximity of these two regions. Likewise, Bmi1 displayed a similar intramolecular cross-linking pattern between its Ring and RAWUL domains, to one observed in PRC1-PHC2. Interestingly, cross-links between Ring1b and Bmi1 remained unaltered when compared to PRC1-PHC2, suggesting that the association with Cbx7 had no discernible impact on the interaction between these two subunits. Regarding intramolecular cross-links in Cbx7,

the limited availability of lysine residues in the C-terminus precluded definitive conclusions on whether it adopts an extended or compact conformation within the PRC1-Cbx7.

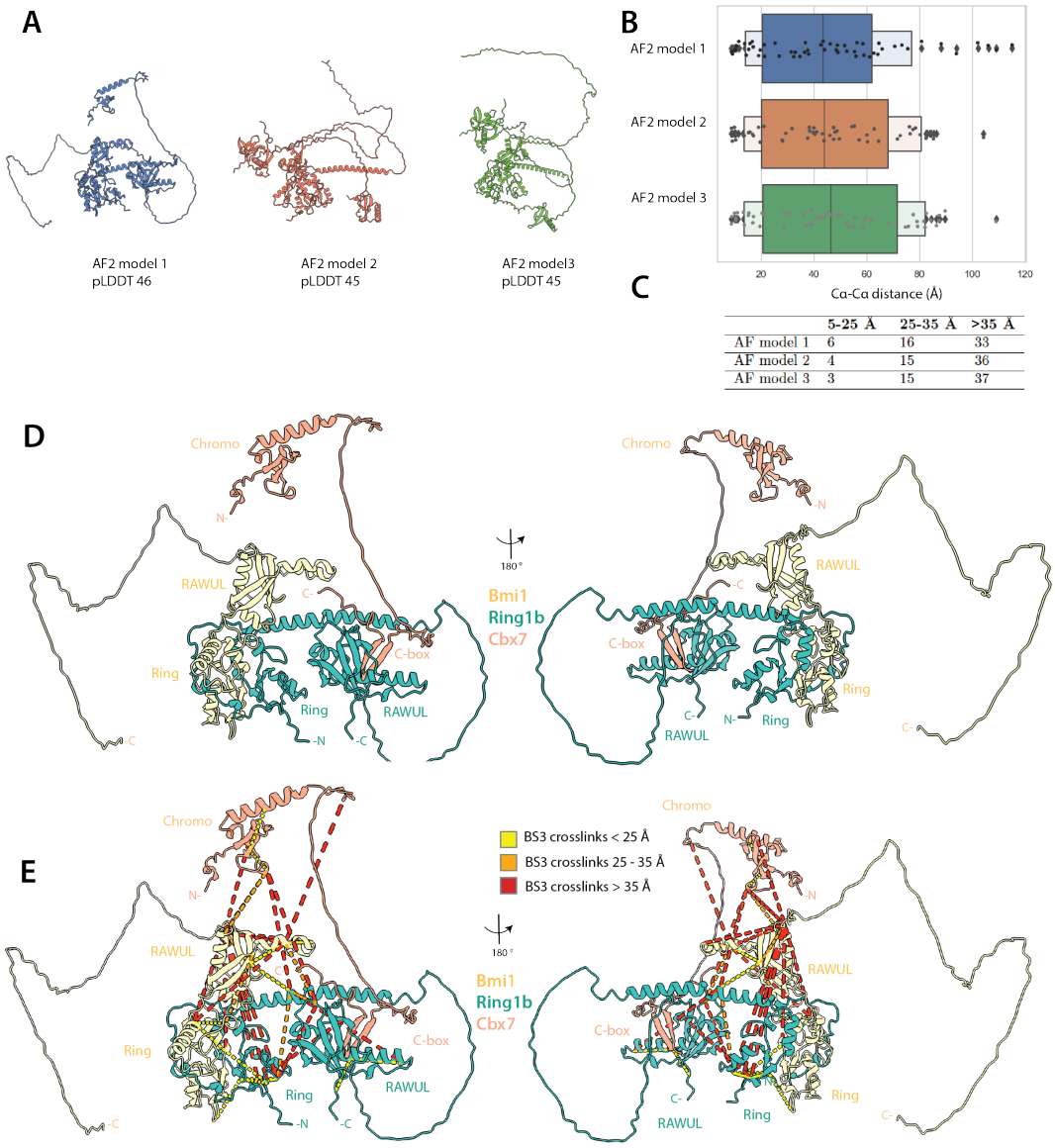
Next, the XL-MS experiment was performed using the mixture of PRC1-Cbx7 complex and H3Kc27me3 *D.m* nucleosomes. However, this approach did not yield a substantial number of cross-links and no cross-links to the nucleosome, making sample fractionation necessary (for a detailed description refer to the Method section 3.9.3). Only three reproducible cross-links between PRC1-Cbx7 and nucleosome were detected, all of them involving the Ring1b Ring domain. Interestingly, when compared to the PRC1-PHC2 XL-MS experiment, cross-links between the Ring1b Ring domain and H3 histone were previously undetected. As for cross-links between Cbx7 and H3Kc27me3 histone, various cross-links between the Cbx7 chromodomain and histone H3Kc27me3 were detected in both experimental replicas. Still, however, none of these cross-links were reproducible (hence not displayed in the final connectivity map). Apart from non-reproducible cross-links of Cbx7 Chromodomain to histone H3 tail, no additional interaction surface between Cbx7 and nucleosome was detected.

Mapping of XL-MS PRC1-Cbx7 distance restraints on AlphaFold2 predictions

To explore the agreement between experimental cross-linking mass spectrometry data and models of PRC1-Cbx7 complex predicted by AF2, I employed AF2 in multimer mode to produce a set of 25 models of PRC1-Cbx7 complex (Jumper et al., 2021). Per-residue model confidence scores (pLDDT) for all models were below 50 (on a scale from 0 to 100), with the highest-scored model reaching pDDT of 46 and the lowest 37. The top three scored models are depicted in Figure 24A. Next, all cross-links were mapped onto predicted models using the XMAS application in ChimeraX (Lagerwaard et al., 2022, Goddard et al., 2018). The Cα-Cα distance distribution of mapped cross-links for the top three models is presented in Figure 24B, showing a large distance distribution for all three models, with most of the cross-linking distances falling into the distance restrain-violated category (compare values for three distance categories in Figure 24C). Model 1 also had the highest number of long-distance outliers when detected by XMAS (Figure 24B top row in blue, outliers shown as rhombi. At the same time, model 1 had the highest pLDDT score as well as the largest number of permitted cross-linked residue pairs (Figure 24C first row - 6 cross-linked residues pairs in 5-25 Å category) and therefore was chosen as representative model and shown in detail in Figure 24D with domains annotated.

The interdomain region between Cbx7 Chromodomain and C-box was predicted to be unstructured (see Figure 24D, Cbx7 protein coloured in pink), as well as C-terminus of Bmi1 and interdomain region in Ring1b, as was the case for PRC1-PHC2 AF2 prediction. When cross-linked mapped, these regions showed no presence of cross-linked residues, consistent with their presumed lack of secondary structure and conformational heterogeneity (Figure 24E). Permitted cross-links (5-25 Å, shown in yellow) again were restricted to the regions corresponding to structured domains (such as Ring:Ring dimer of Ring1b and Bmi1, RAWUL domains of Ring1b and Bmi1 and Cbx7 Chromodomain. Longer distance cross-links (shown in orange and red) linked Cbx7 chromodomain to RAWUL domains of Ring1b and Bmi1, as well as to de-novo predicted α-helix in Ring1b, adjacent to Ring domain (could be seen in Figure 24E, left).

In summary, akin to the AF2 predicted model of the PRC1-PHC2 complex, the AF2 prediction for PRC1-Cbx7 largely contained known structural data. The only novel structural element identified was the predicted Ring1b α-helix adjacent to the Ring1b Ring domain (also present in the PRC1-PHC2 AF2 prediction). A substantial portion of the complex was predicted to lack structural organization. Overlaying cross-linking data onto PRC1-Cbx7 models failed to yield a single model consistent with cross-link distance restraints, indicating that the cross-linking data predominantly represent an ensemble of conformational states, as expected given the predicted presence of unstructured regions.



	5-25 Å	25-35 Å	>35 Å
AF model 1	6	16	33
AF model 2	4	15	36
AF model 3	3	15	37

Figure 24: AlphaFold2-predicted models of PRC1-Cbx7 and their alignment with distance restrains provided by XL-MS. A. AlphaFold2 models of PRC1-Cbx7 complex with highest overall pLDDT scores. B. Cα-Cα distance distribution of BS3 cross-links mapped on AlphaFold2 models of PRC1-Cbx7, visualized with XMAS. C. Overview of the BS3 cross-link counts categorized based on various Cα-Cα distances, when mapped on AlphaFold2 models of PRC1-Cbx7. D. Annotated representation of AlphaFold2 model 1 of PRC1-Cbx7 complex. E. BS3 cross-links mapped on PRC1-Cbx7 AlphaFold2 model 1. The cross-links are colour-coded to indicate their respective Cα-Cα- distances: violated cross-linking (> 35 Å) shown in red, permitted cross-linking distances (5-25 Å) in yellow and cross-links within the range of 25-35 Å depicted in orange. Visualized in XMAS (Lagerwaard et al., 2022).

4.2.4 Cryo-EM studies of PRC1-Cbx7 subcomplex bound to nucleosome

Two approaches were utilized for cryo-EM analysis of PRC1-Cbx7 subcomplex on nucleosome: native, using streptavidin affinity grids, and glutaraldehyde fixation. Glutaraldehyde has been commonly utilised in the cryo-EM studies of nucleosome-binding complexes (Markert et al., 2023, Worden et al., 2020). Therefore, this cross-linking approach was applied to test whether it could potentially stabilize the complex and resolve regions with different conformational states.

Native streptavidin grids were prepared using a similar protocol as for PRC1-PHC2 (Method section 3.8.2) but omitting the washing steps. This change of the established protocol was introduced to increase the saturation of PRC1-Cbx7 on nucleosomes, as the previous dataset collected for PRC1-PHC2 showed that roughly half of the nucleosome particles were not PRC1-bound, therefore reducing the number of particles that could undergo classification. Both native PRC1-Cbx7 and glutaraldehyde-fixed PRC1-Cbx7 datasets were acquired using a Titan Krios microscope equipped with a Gatan K3 camera, with invaluable support from the MPIB Cryo-EM facility (described in Method section 3.8.4).

PRC1-Cbx7 with nucleosome, non-crosslinked

Given the anticipated high degree of conformational variability in the native PRC1-Cbx7 : nucleosome sample, a large dataset (23,000 movies, 0.8512 Å pixel size) was collected to maximize the chances of resolving conformational heterogeneity through rigorous particle classification. The processing workflow for native PRC1-Cbx7 on nucleosomes is depicted in Figure 25.

The picked particles were subjected to several rounds of 2D and 3D classifications. Analysis of 2D classes revealed prominent additional density on the proximal and distal sides of nucleosomes. Initial 3D reconstruction contained Ring fingers bound to the nucleosome, with protruding extra densities (see processing scheme in Figure 25, highlighted with a circle). The processing was then focused on the following aspects:

1. Resolving of the Ring1b and Bmi1 Ring fingers on a nucleosome for comparison with the Ring fingers of the 4R8P crystal structure and PRC1-PHC2 complex;

2. Resolving the extra density protruding beyond the Ring1b and Bmi1 Ring fingers domain through iterative heterogeneous refinement, 3D classification, 3D variability analysis, and masked 3D refinement without angular alignment in Relion.

The latter objective was, however, contained by limited methodological approaches for resolving the additional density due to the size of the Ring fingers with protruding density being less than 50 kDa. Consequently, the refinement approaches such as local refinement with alignment and density subtraction or multi-body refinement could not be effectively applied (Nakane et al., 2018). Although these alternatives were explored, the small size and resulting low signal contribution to the reconstruction rendered the results inconclusive and thus are not presented here. The results of 3D variability analysis for both non-crosslinked and cross-linked samples will be discussed separately later in this section.

For resolving the Ring1b and Bmi1 Ring fingers on nucleosome, masked-focused refinement was used. This resulted in 3D reconstruction reaching 2.69 Å resolution (see bottom right, Figure 25 and a close-up of the final 3D reconstruction in Figure 27A). The same set of particles was refined homogeneously using a dynamic mask (and therefore includes the extra density that was later used for 3D variability analysis, Figure 28A). Furthermore, the heterogeneous refinement with three distinct classes resulted in two classes exhibiting protruding extra density. Subsequent re-extraction and refinement of these classes yielded 3D reconstruction with global resolutions of 2.94 Å and 3.05 Å, respectively. However, neither 3D reconstruction exhibited continuous density for the protruding part, likely attributed to its flexibility and the potential over-refinement due to the high global resolution of the rest of the complex. Typically, in such scenarios, techniques like particle density subtraction, refinement, and composite mapping are employed. However, due to the small size of this additional density, these methods could not be effectively utilized.

The viable option to resolve this additional density was to perform focused local refinement without angular alignment, as the angular alignment of the masked density would not be precise due to a weak signal. All the particles extracted in cryoSPARC were transferred to Relion 3.1 and underwent 3D refinement, based on this 3D reconstruction, a soft mask covering only the Ring fingers with extra density was created in ChimeraX (Goddard et al., 2018, Scheres, 2012). Various regularization parameters were tested for local masked refinement without angular alignment, with T=10 yielding the best results. Masked local refinement was tested for both proximal and distal nucleosomal sides, and both showed similar results, with only one side being presented here. As depicted in the processing scheme (Figure 25, bottom left), two main classes were obtained: one with Ring fingers and extra density (68% of all particles, estimated resolution 8 Å) and another with Ring fingers only (29%, with an overall resolution of 6.2 Å). In the

first class, the extra density was not sufficiently resolved to determine whether it corresponded to Ring1b or Bmi1 protein.

Another processing approach to resolve the extra density was interactive global 3D classification in cryoSPARC. This classification into 10 classes revealed that particles even distribution of particles across all classes (approximately 10% per class, Figure 25 on the left). This result was yet another indication of the lack of a prevalent conformation in the region of extra density, e.g. beyond the E3 ligase domain that comprises Bmi1 and Ring1b Ring fingers.

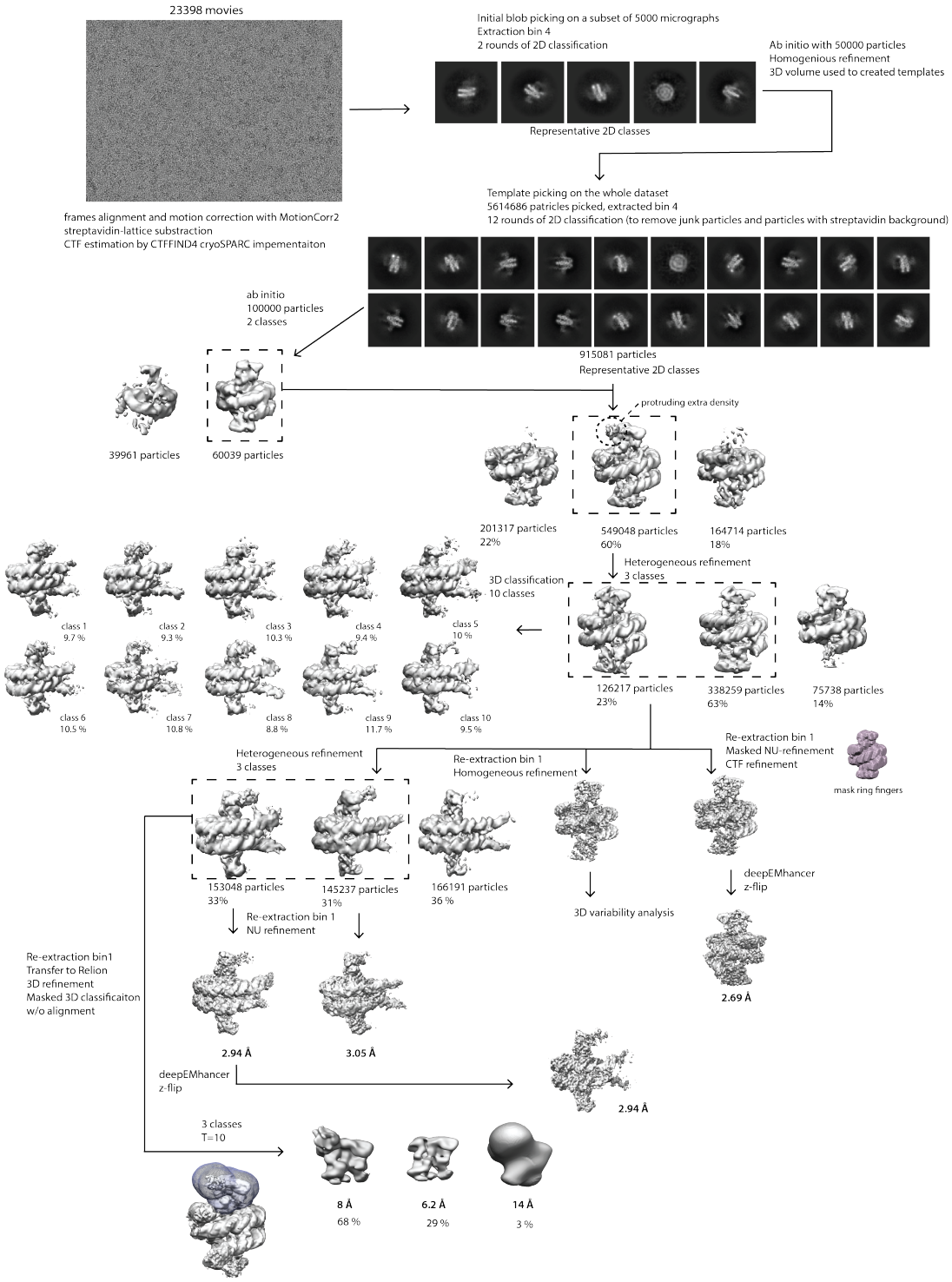


Figure 25: Single particle cryo-EM processing scheme in cryoSPARC for PRC1-Cbx7 – nucleosome non-crosslinked dataset. Details provided in methods and in text and Methods section 3.8.5.

PRC1-Cbx7 with nucleosome, glutaraldehyde-fixated

The dataset on PRC1-Cbx7:nucleosome sample fixed with glutaraldehyde was collected as described above for non-crosslinked PRC1-Cbx7:nucleosome (13,553 movies, pixel size of 0.8512 Å). Despite a dense particle distribution on the grid, individual particles were picked and 2D-classified into classes with visible high-resolution features and PRC1-corresponding density on nucleosome sides (see Figure 26 for corresponding micrograph and 2D classes). During 3D classification, some classes were attributed to either two stacked nucleosomes (likely due to dense distribution) or incomplete nucleosomes characterized by shortened DNA ends. Consequently, particles belonging to these classes were eliminated from further processing (for example, see results of the first heterogeneous refinement in Figure 26).

The resulting set of particles underwent a similar processing strategy as native PRC1-Cbx7:nucleosome, with a focus on:

1. Resolving the Ring fingers on nucleosome through masked refinement targeting the nucleosome and the Ring1b-Bmi1 Ring fingers.
2. Investigating the extra density beyond Ring fingers using 3D classification and 3D variability analysis.

Regarding the Ring1b and Bmi1 Ring fingers bound to the nucleosome, the final 3D reconstruction reached an overall resolution of 2.92 Å (see Figures 27E and 27G). However, attempts to resolve the extra density beyond the Ring fingers domain through 3D classification with 5 classes revealed an equal distribution of particles among classes. Unfortunately, refinement of each individual class did not yield sufficient resolution of the extra density, similar to the results presented for native PRC1-Cbx7. Notably, when compared to the non-crosslinked map, the extra density was observed to lean towards the nucleosomal DNA in the cross-linked sample. In the non-crosslinked sample, this tendency was also observed in several classes (such as class 3 of the 3D classification depicted in Figure 25, middle left), however, was less pronounced. This potential interaction with DNA and its potential functional implications will be further discussed later in this section.

The focused EM-maps of the Ring1b:Bmi1 Ring fingers on nucleosome, originating from both non-crosslinked (panels A-D) and glutaraldehyde cross-linked samples (panels E-H) are depicted in Figure 27. The EM-density maps exhibit a global resolution of 2.69 Å and 2.92 Å,

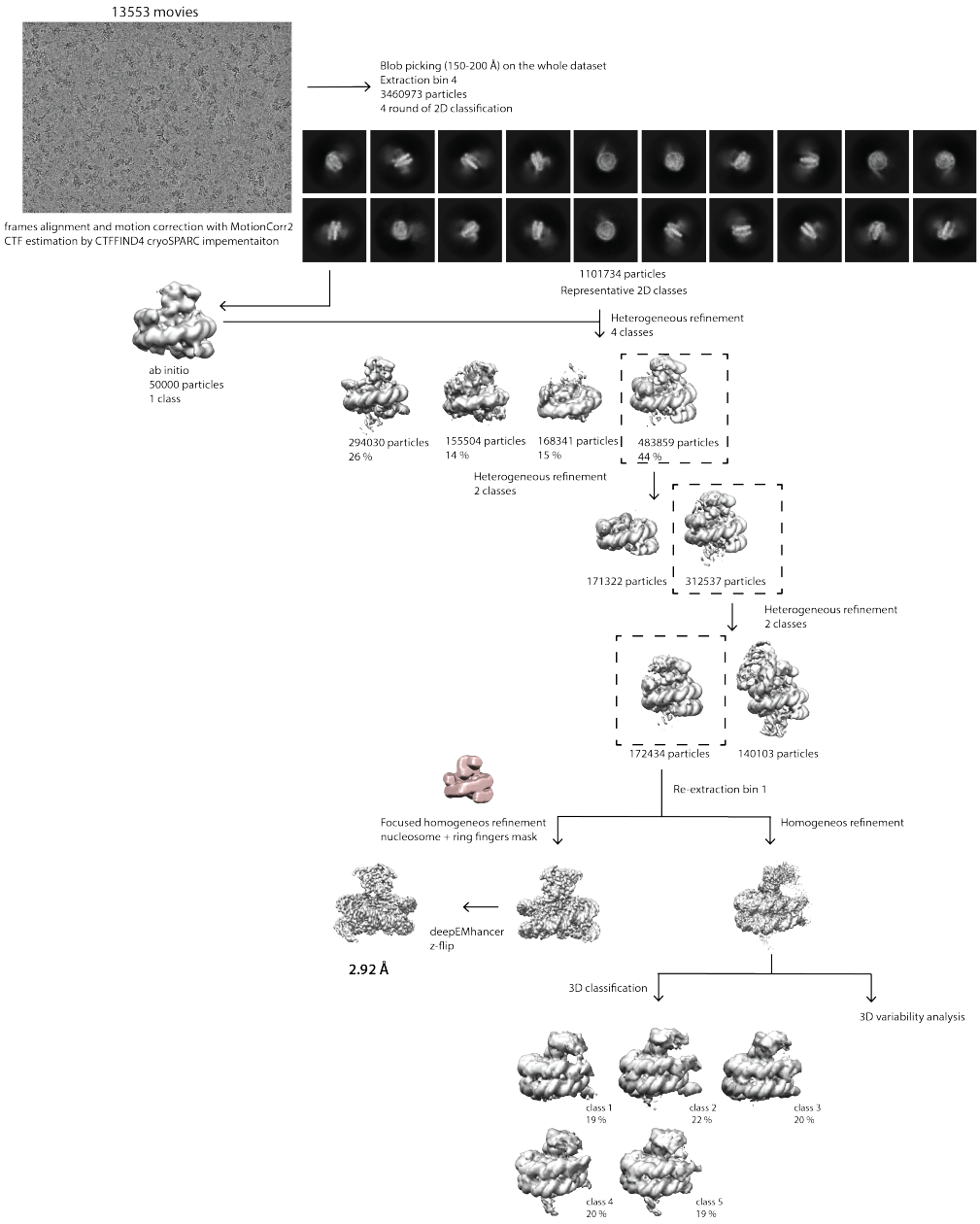


Figure 26: Single particle cryo-EM processing scheme in cryoSPARC for PRC1-Cbx7:nucleosome glutaraldehyde cross-linked dataset. Details provided in Methods section 3.8.5 and in main text.

respectively. Notably, due to the exclusion of extra density from the focused refinement, there is a slight disparity between the Fourier shell correlation (FSC) curve calculated for masked and unmasked maps.

Local resolution in both EM-maps ranged from sub-2.5 Å for the nucleosomal core to 5 Å for the top part of the Ring fingers (Figure 27, panels B and F). In the glutaraldehyde cross-linked sample, a lower concentration of PRC1 was used to prevent the formation of PRC1-only cross-linking aggregates, resulting in only one side of the nucleosome occupied by PRC1. Moreover, particles used for 3D reconstruction of the cross-linked sample exhibited better angular distribution compared to those in the non-cross-linked sample (Figure 27H and Figure 27D). This disparity could be attributed to constraints imposed by biotinylated DNA, as streptavidin-coated grids were used for the native PRC1-Cbx7 sample. Concerning angular distribution, it was also observed that the particle orientation on streptavidin grids heavily depended on ice thickness. In contrast, cross-linked samples on regular grids exhibited a more universal distribution independent of ice thickness.

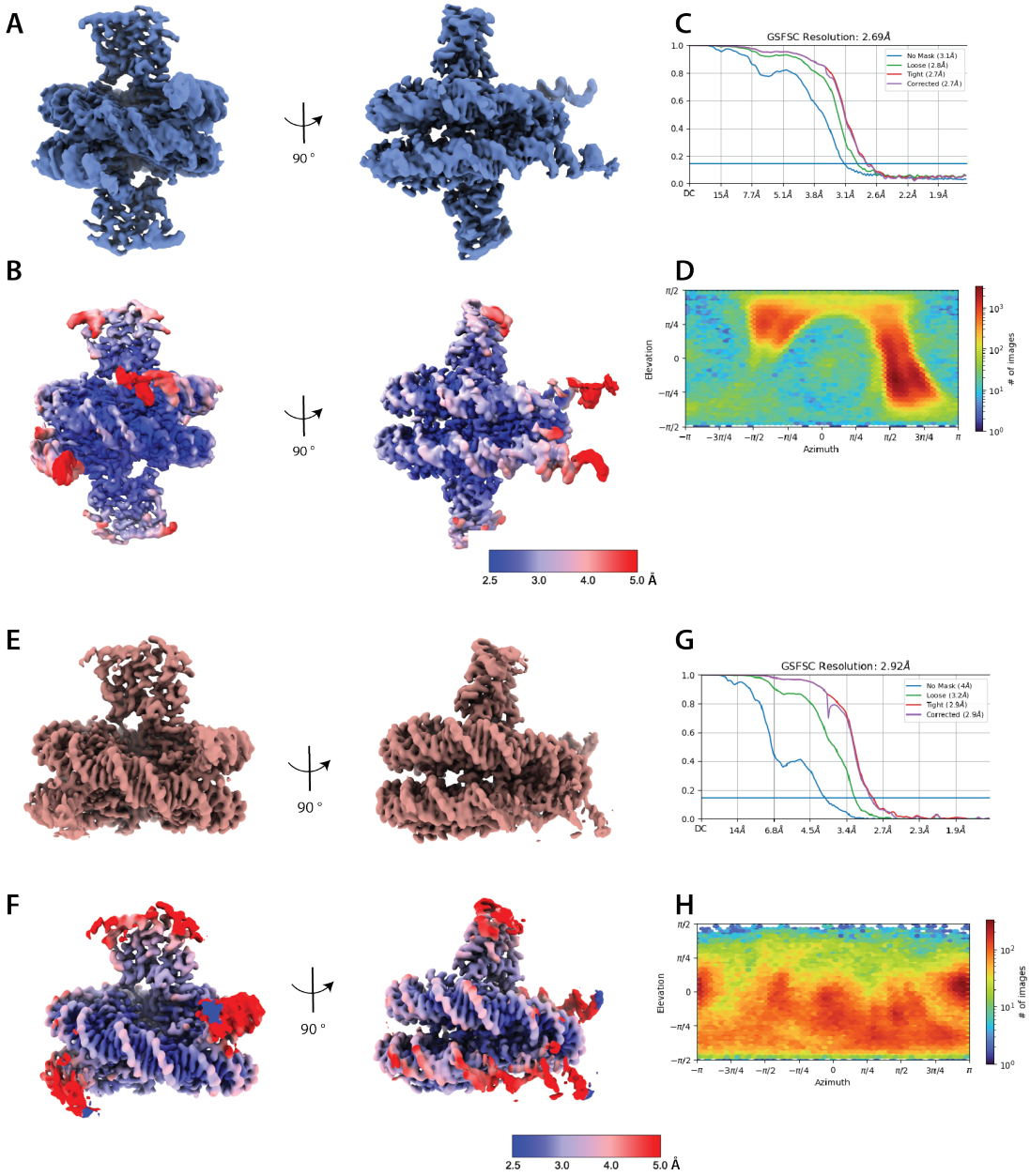


Figure 27: Final EM maps and reconstruction metrics from PRC1-Cbx7 datasets: non-crosslinked (A-D) and crosslinked (E-H) samples. A, E. Cryo-EM maps of PRC1 E3 ligase module bound to a nucleosome as a part of PRC1-Cbx7 nucleosome complex (A non-crosslinked, E crosslinked). B, F. The maps from A and E, respectively, but colored by local resolution (B non-crosslinked, F crosslinked). C, G. Gold standard FSC (Fourier Shell Correlation) curves for the final 3D reconstructions (C non-crosslinked, G crosslinked), resolution estimated at FSC = 0.143. D, H. Angular distribution of particles used in the reconstructions depicted in A and E.

3D variability

The 3D variability analysis (3DVA) implementation in cryoSPARC was used to determine whether the heterogeneity of the PRC1 complexes is of continuous nature or whether some set of particles with discreet conformation could be identified (Punjani and Fleet, 2021). In this analysis, particles are classified according to attributes, that could be, for example, association or dissociation of a subunit or a conformational change. All particles are categorized according to variability components (set of attributes) and used to reconstitute, what is referred to herein as a "continuous family of 3D structures".

Here, I subjected all the particles to 3D variability analysis with a cut-off of 10 Å to focus on less resolved extra density and compared two end states of that continuous family of 3D structures, to pinpoint the conformational changes that occur (Figures 28A and 28C, 3D reconstructions in magenta and violet correspond to two states of variability range). The plots show the distribution of particles according to their variability components, with one variability component corresponding to an axis (reaction coordinate) in a 2D plot. Note, that a single variability component could have several attributes, as shown in the comparison of the conformational states in Figures 28A and 28C.

Figure 28 illustrates the visualization of 3DVA results conducted on the non-crosslinked (panels A and B) and cross-linked (panels C and D) PRC1-Cbx7:nucleosome datasets. In the non-crosslinked sample, the analysis revealed several attributes, including the movement of the nucleosomal DNA ends and variability in the extra-density region (Figure 28A). This variability in the extra density could be observed as a shift in position (as seen in variability component 1) or the appearance of additional density (as observed in components 2 and 3). For the cross-linked sample, the variability analysis demonstrated more prominent features in the extra-density region, particularly revealing the tendency of the extra-density region to come into proximity of the DNA ends (Figure 28C, the shift towards nucleosomal DNA is indicated with arrows). The possible functional implications of this observation will be discussed later in this section.

Additionally, variability was observed within the extra-density region itself, resulting in changes in its shape. For example, in component 2, the extra-density region exhibited different states: one with an arc-shaped configuration (depicted in violet) and another with a more condensed form (shown in magenta in Figure 28C, compare the shape of the extra density in violet and magenta). Similarly, in component 3, the extra density one variability state (magenta)

appeared diminished, while in the second state (violet), it exhibited a more compact structure.

Figures 28B and 28D show 2D plots of individual particles attributed to three variability components, as solved by 3DVA. For both cross-linked and non-crosslinked datasets, particles were grouped in a single population in different component comparisons, suggesting the absence of distinct conformational states. In other words, the 3D variability analysis did not classify the particles into discrete populations but instead revealed a continuous heterogeneity for both the cross-linked and non-crosslinked samples.

The slices through the y-axis of the EM-maps for both the cross-linked and non-crosslinked PRC1-Cbx7 dataset also reveal the presence of an arc-shaped density that extends towards the DNA ends, as depicted in Figures 28E and 28F. However, upon closer inspection, it becomes apparent that the density beyond Ring fingers of PRC1 lacks high-resolution features and, therefore, is unlikely to be resolved due to its conformational heterogeneity, as confirmed by 3DVA and 3D classification.

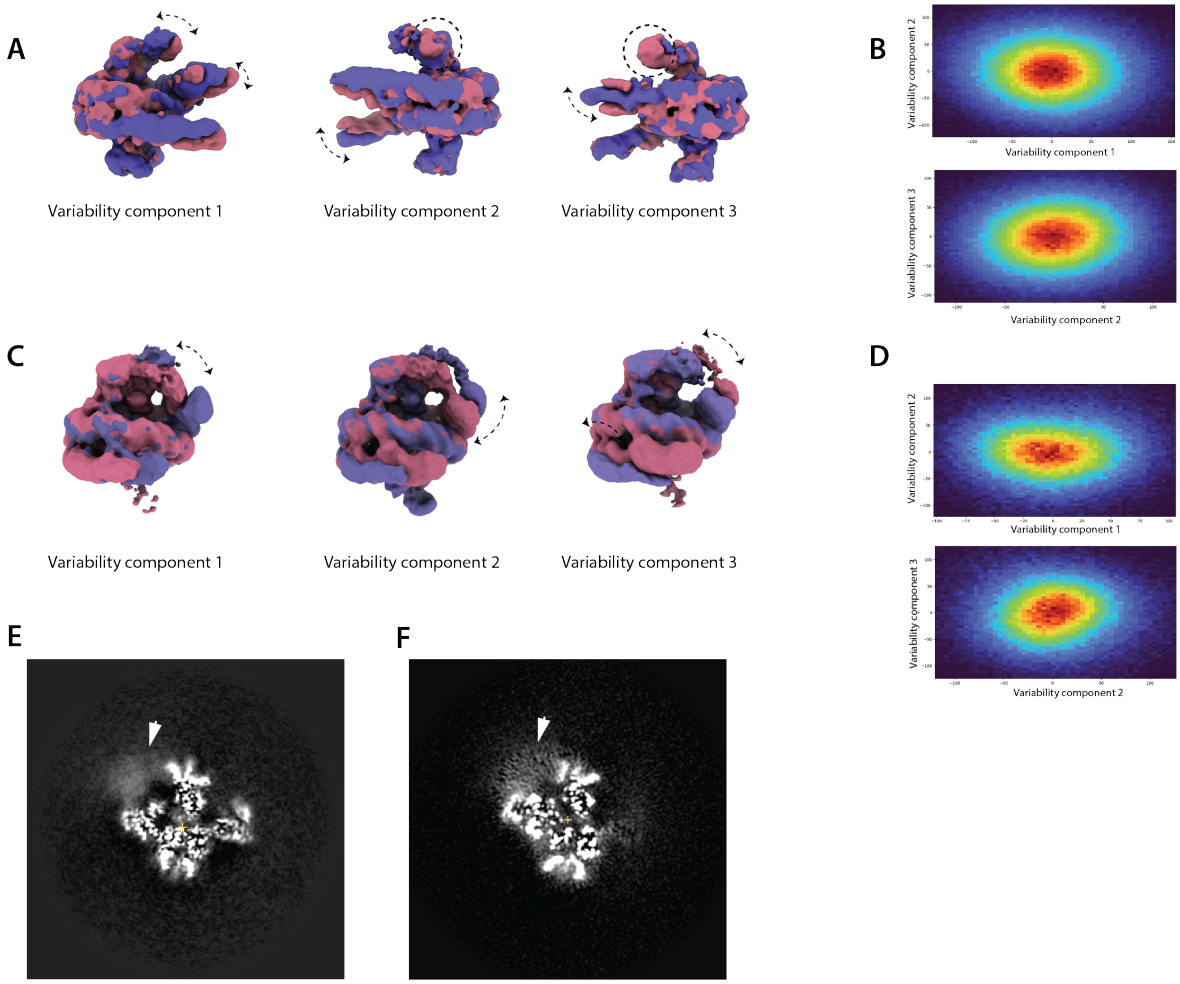
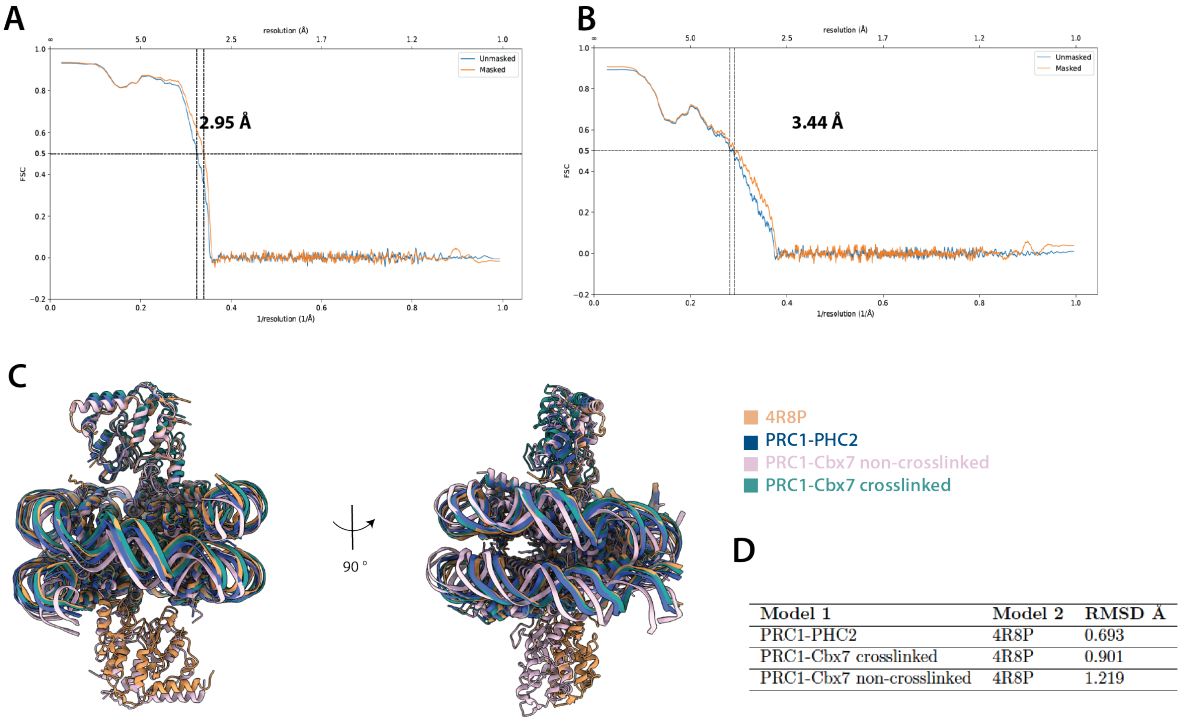


Figure 28: 3D variability analysis. Results of 3DVA with three variability components as generated by cryoSPARC: A-B. PRC1-Cbx7 non-crosslinked (A-B) and cross-linked dataset (C-D). A+C. Overlay of 3D density maps generated by 3DVA at negative (red) and positive (blue) positions of each variability component, direction of variability highlighted with arrows. All components indicate a bidirectional movement of extra density or/and nucleosome DNA ends. B-D. 2-D plots of individual particles across three variability components, as solved by 3DVA. E-F. Y-axis slices of Coulomb potential maps in grey value representation, arrows mark regions of non-discrete heterogeneity in PRC1-Cbx7 non-crosslinked (E) and crosslinked (F) 3D reconstructions.

Models comparison

For both focused EM-maps of Ring fingers on nucleosome, which originated from native and fixated PRC1-Cbx7:nucleosome datasets, the FSCs of the map-to-model were calculated using cryo-EM validation tool of Phenix software and shown in Figure 29A (Afonine et al., 2018). The curves for masked and unmasked maps, as expected, show discrepancies between each other due to the presence of extra density in unmasked map, that is not interpreted by a model. All the models built into EM-density in this study were compared to the original crystal structure of nucleosome bound by PRC1 E3 ligase (PDB 4R8P, McGinty et al. (2014)), as shown in Figure 29C. The cryo-EM structures of the Ring fingers of the PRC1-PHC2 and PRC1-Cbx7 subcomplexes with nucleosome demonstrated strong correspondence with the 4R8P crystal structure model, exhibiting RMSD values of 0.693 Å for the comparison between 4R8P and the PRC1-PHC2:nucleosome model, and 0.901 Å for the comparison between 4R8P and the cross-linked PRC1-Cbx7:nucleosome model. For the model derived from the PRC1-Cbx7:nucleosome non-crosslinked sample, where PRC1 Ring fingers were bound on both sides, a slightly higher RMSD was observed due to a difference in the two nucleosomal disk interfaces, impacting the overall alignment and resulting in an RMSD of 1.2 Å.

In summary, these results demonstrate a concordance between the cryo-EM-derived models of the nucleosome-bound Ring fingers of both PRC1-Cbx7 and PRC1-PHC2 subcomplexes and the previously reported crystal structure of PRC1 Ring fingers bound to the nucleosome (4R8P, McGinty et al. (2014)). This underscores that despite the observed conformational variability within the PRC1 complex, the binding mode of the Ring fingers to the nucleosome remains consistent across various subcomplexes.



Model 1	Model 2	RMSD Å
PRC1-PHC2	4R8P	0.693
PRC1-Cbx7 crosslinked	4R8P	0.901
PRC1-Cbx7 non-crosslinked	4R8P	1.219

Figure 29: A-B. Map-model FSC curves for models of PRC1 E3 ligase module bound to the nucleosome, derived from PRC1-Cbx7 non-crosslinked (A) and cross-linked (B) datasets, calculated by the Phenix validation tool. FSC=0.5 at spatial frequency 0.34 Å-1 (2.95 Å resolution) for PRC1-Cbx7 non-crosslinked, FSC=0.5 at spatial frequency 0.29 Å1 (3.44 Å resolution) for PRC1-Cbx7 cross-linked. C. Cryo-EM map-derived models of nucleosome bound PRC E3 ligase modules from different PRC1 subcomplexes obtained in this study, superimposed and compared to the previously published crystal structure (4R8P, in orange). D. Summary of root-mean-square deviations (RMSD) between the previously published crystal structure of PRC1 E3 ligase module on nucleosome (4R8P) and cryo-EM map-derived models obtained in this study (calculated using Matchmaker tool in ChimeraX).

As discussed above, the 3DVA analysis and examination of EM-maps of PRC1-Cbx7 : nucleosome highlighted the tendency of the unresolved extra-density region to come into the proximity of the nucleosomal DNA, forming an "arc"-like shape visible in 2D y-slices of the EM-maps (Figure 28, panels E and F). To speculate on the origin of this unresolved "arc-like" density, I examined AlphaFold2-predicted models of PRC1-Cbx7. Unlike the rest of the domains in predicted models, the Bmi1 RAWUL was consistently aligned above the Ring finger domains of Ring1b and Bmi1 (Figure 30A). This correlated remarkably well with the observed extra density. Notably, cross-linking mass spectrometry data also showed cross-links between the Bmi1 RAWUL domain and the Chromodomain of Cbx7. To evaluate the plausibility of this mode of interaction, I juxtaposed the AlphaFold2 model featuring Bmi1 RAWUL domain with the cryo-

EM-derived model of PRC1-Cbx7 Ring domains, and with a model of the Cbx7 Chromodomain (PDB 4X3K, Ren et al. (2015)) bound to the extended H3 tail (from PDB 7AT8, Finogenova et al. (2020)) in ChimeraX.

The calculated distance between the Bmi1 RAWUL domain and the Cbx7 Chromodomain engaged with the extended H3 tail was measured at 66 Å. Accounting for the inherent flexibility of the H3 tail and potential conformational variability within the Bmi1 RAWUL domain itself (as suggested by 3D variance analysis tendency to come to the proximity of nucleosomal DNA), it is conceivable that the RAWUL domain, according to this model, may approach Cbx7 Chromodomain within permissible range for BS3 cross-linking (25 Å). Therefore, when integrating the XL-MS data with 3DVA and AF2 modelling, it is possible to tentatively propose that the "arc-like" additional density observed in EM-maps of PRC-Cbx7 may derive from the RAWUL domain of Bmi1, possibly establishing contact with the Chromodomain of Cbx7 bound to the H3 tail.

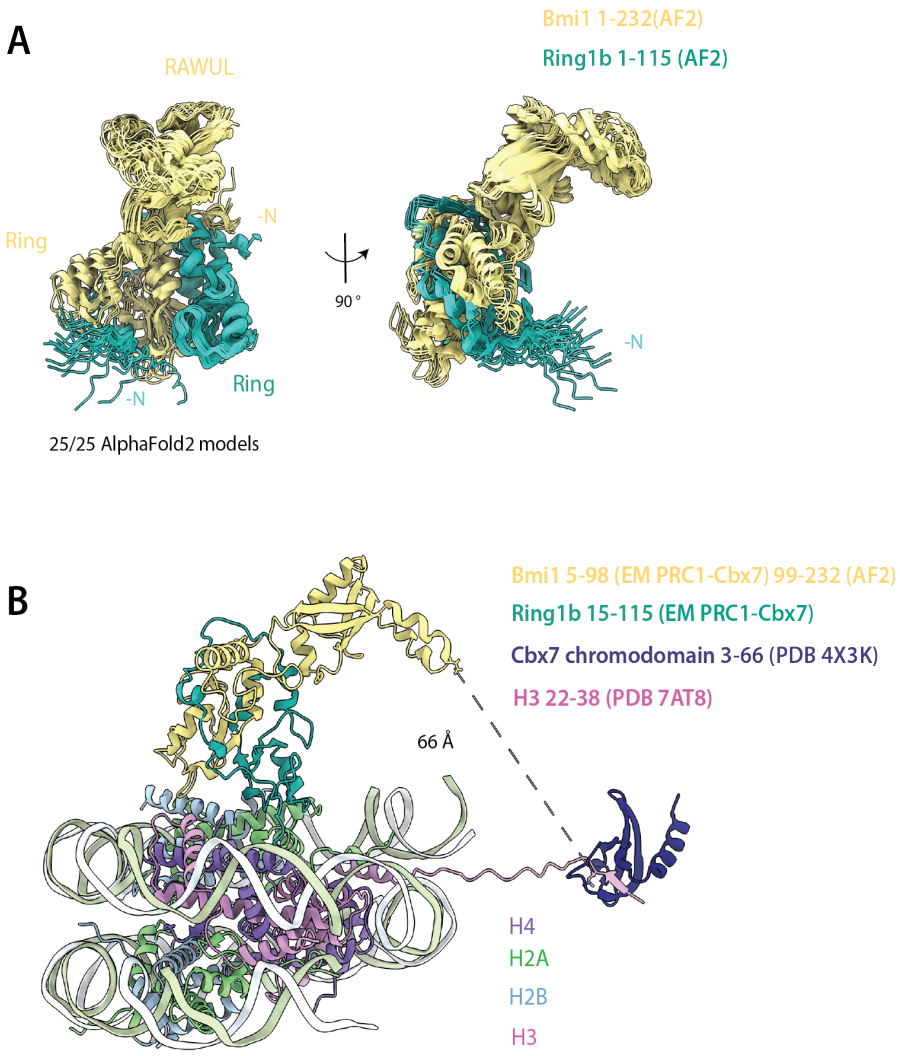


Figure 30: Proposed interpretation of the extra density based on cryo-EM and AlphaFold2 predicted models. A. Overlay of 25 AlphaFold2 predicted models for Bmi1 (Ring1b RAWUL domain, 1-232 AA) and Ring1b (Ring domain, 1-115 AA) proteins showing a consisted placement of Bmi1 RAWUL domain across all models, visualized in ChimeraX. B. AlphaFold2 prediction from A combined with nucleosome model from this study (PRC1-Cbx7 crosslinked dataset), previously published Cbx7 chromodomain (PDB 4X3K) and stretched H3 tail from PRC2-dinucleosome structure form Finogenova et. al, 2021 (PDB 7AT8).

4.3 PRC1-PHC2-Scml2 complex

4.3.1 PRC1-PHC2-Scml2 complex reconstituted from insect cells is an active E3 ligase and binds nucleosomes with higher affinity

The PRC1-PHC2-Scml2 complex was expressed in Hi5 insect cells as described in method section 3.3.4 (the schematic representation of expressed proteins and tags shown in Figure 31A, as well as an outline of the purification process in Figure 31B, left). Ring1b and Bmi1 subunits were co-expressed and purified through Ni-affinity pulldown via His-tagged Bmi1. Scml2 and PHC2 (short version of PHC2 ΔN-terminus, comprising residues 536-853) were co-expressed and purified through the Strep-tag of Scml2, as PHC2 and Scml2 interact through their SAM domains and therefore could be affinity co-purified(Kim et al., 2005, Bonasio et al., 2014). Since the SAM domains of both Scml2 and PHC2 form hetero- and homopolymers through the head-to-toe interaction of their ML-EH surfaces, I introduced point mutations to both ML surface of Scml2 and EH surface of PHC2 to abolish the polymerisation (as described in Frey et al. (2016)). This two-step purification procedure, where Ring1b-Bmi1 and Scml2-PHC2 were expressed and pulled down in pairs before merging, was adopted due to the significantly higher expression levels of Scml2-PHC2 compared to Ring1b and Bmi1. Figure 32B (right panel) shows the elution fractions from the streptavidin pulldown of Scml2-PHC2 and Ni-pulldown of Ring1-Bmi1, as well as fractions after 3C protease tag cleavage. Note that Scml2 degraded after overnight dialysis with 3C protease treatment (Figure 31B right, last lane), but the full-length Scml2 and its degradation product could be separated using MonoS cation exchange chromatography (Figure 31C left, compare fractions B1-B4 and B12-C3). Nucleic acid contamination was also largely removed during this purification step (Figure 31C right, peak at 0-10 ml).

The MonoS elution fractions containing PRC1-PHC2-Scml2 complex (Figure 31C left, fractions B12-C3) were purified using size exclusion chromatography on Sephadex S6 column (see Figure 31D left for Comassie gel of elution fractions and right for elution profile). The first peak (fractions A4 - A10) corresponds to the further removal of nucleic acid contamination. Fractions A12-B3 were collected and used for further experiments. The presence and identity of all PRC1-PHC2-Scml2 were confirmed with native mass spectrometry (MPIB Mass spectrometry facility).

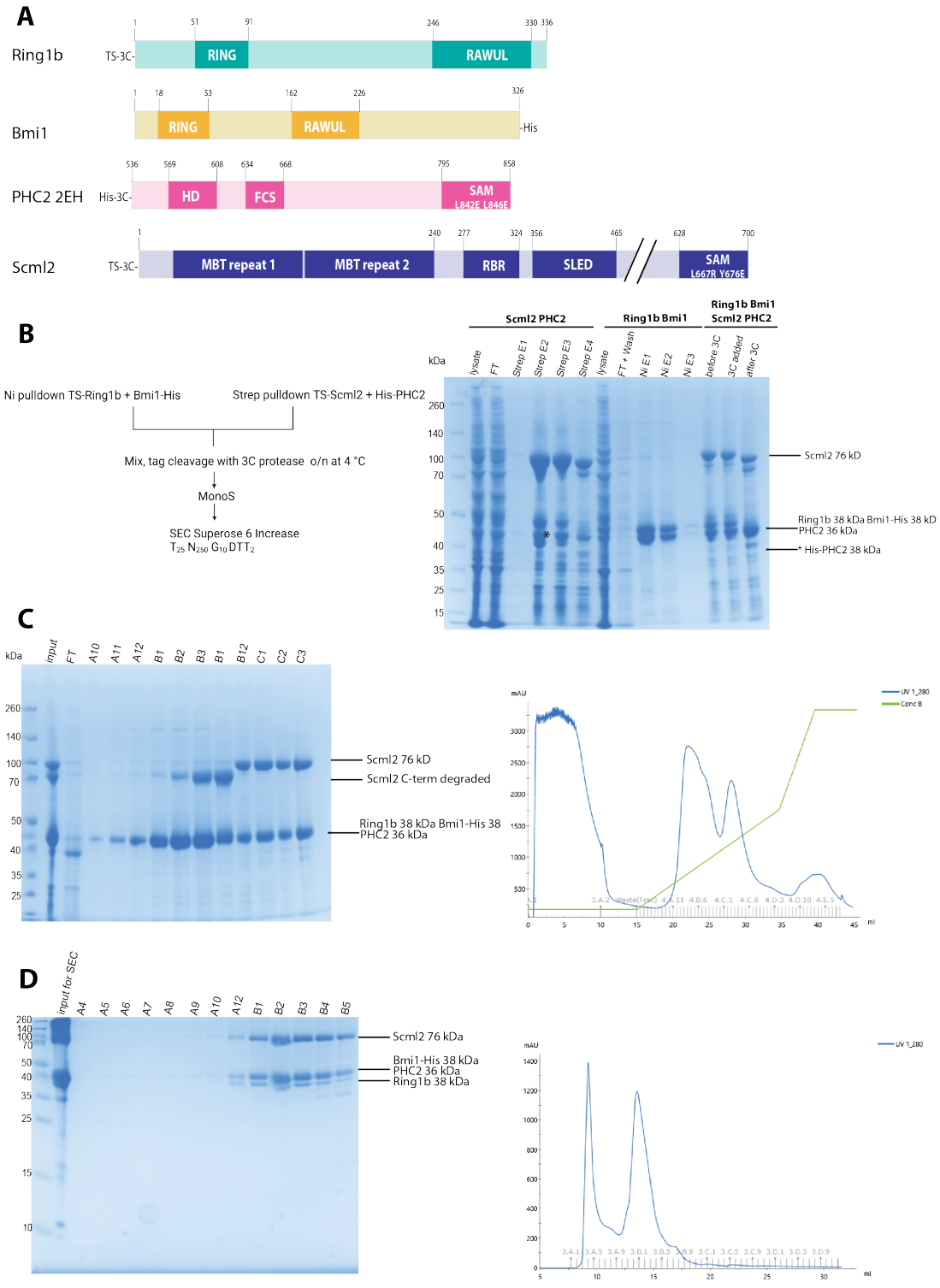


Figure 31: Purification of PRC1-PHC2-Scml2 complex. A. Schematic representation of PRC1-Scml2 domain organization. B. Overview of the purification process (left) and 4-12% Coomassie-stained SDS-PAGE analysis of strep and Ni pulldowns and 3C protease tags cleavage. C. Coomassie stained 4-12% SDS-PAGE analysis of fractions from monoS cation exchange chromatography (left panel) and corresponding chromatogram (right panel). D. Coomassie stained 16% SDS-PAGE analysis of SEC fractions (left) and corresponding SEC elution profile (right).

4.4 PRC1-PHC2-Scml2 nucleosome binding and E3 ligase activity

The nucleosome binding and E3-ligase activity of the purified PRC1-Scml2 complex were tested. Previous data have indicated that Scml2 alone can bind nucleic acids and nucleosomes (Bonasio et al., 2014, Bezsonova, 2014). Since the *in vitro* reconstitution of the PRC1-PHC2-Scml2 complex has not been reported to date, whether the presence of Scml2 within the PRC1 complex influences its nucleosome binding and E3 ligase activity, has not been tested so far.

The nucleosome binding activity of the PRC1-PHC2-Scml2 complex was assessed using an electrophoretic mobility shift assay (EMSA) with *Drosophila melanogaster* nucleosomes containing 215-bp Atto 647N-labeled DNA (the Atto-labelled DNA was kindly provided by Maria Ciapponi). The selection of the more photostable Atto 647N fluorophore over the previously employed 6-FAM allowed for a reduction in substrate concentration to as low as 1 nM. To evaluate the contribution of Scml2 to nucleosome binding, the PRC1-PHC2-Scml2 complex was compared to the Ring1b-Bmi1 dimer alone, which is known to interact with the nucleosomal acidic patch and represents the primary binding mode of canonical PRC1 to nucleosome (McGinty et al., 2014). Densitometric quantification of the Atto 647N signal in the EMSA showed that the PRC1-PHC2-Scml2 complex displays an approximately ten-fold higher affinity for nucleosomes in comparison to Ring1b-Bmi1 alone (PRC1-PHC2-Scml2 $K_{dApp} = 3.5 \pm 0.7$ nM; Ring1b Bmi1 $K_{dApp} = 30 \pm 7$ nM; see Figure 32A, comparing lanes 1-10 with lanes 11-20, note the difference in concentration range).

Next, I evaluated the E3-ligase activity of the PRC1-PHC2-Scml2 complex by *in vitro* reconstitution of the H2AK118 ubiquitination reaction, using various PRC1 complexes as E3 ligases. As a control, the minimal E3 ligase module of PRC1 was utilized (Ring1b 1-130, Bmi1 1-109). Furthermore, the E3 ligase activity of PRC1-PHC2-Scml2 was compared to that of the full-length Ring1b-Bmi1 dimer alone. PRC1-PHC2-Scml2 was found to be an active E3 ligase, with H2AK118ub1 levels generated during the reaction time course comparable to those observed with the minimal E3 and Ring1-Bmi1 dimer (Figure 32B, compare lanes 2-4 and 6-8 to 10-12). Interestingly, it was also observed that the reaction with PRC1-PHC2-Scml2 complex as E3 ligase produced significantly higher amounts of H2AK118ub2, even early in the time course (Figure 32B, compare the 30-minute time points for the E3 minimal (lane 2), Ring1b-Bmi1 (lane 5), and PRC1-PHC2-Scml2 (lane 10)).

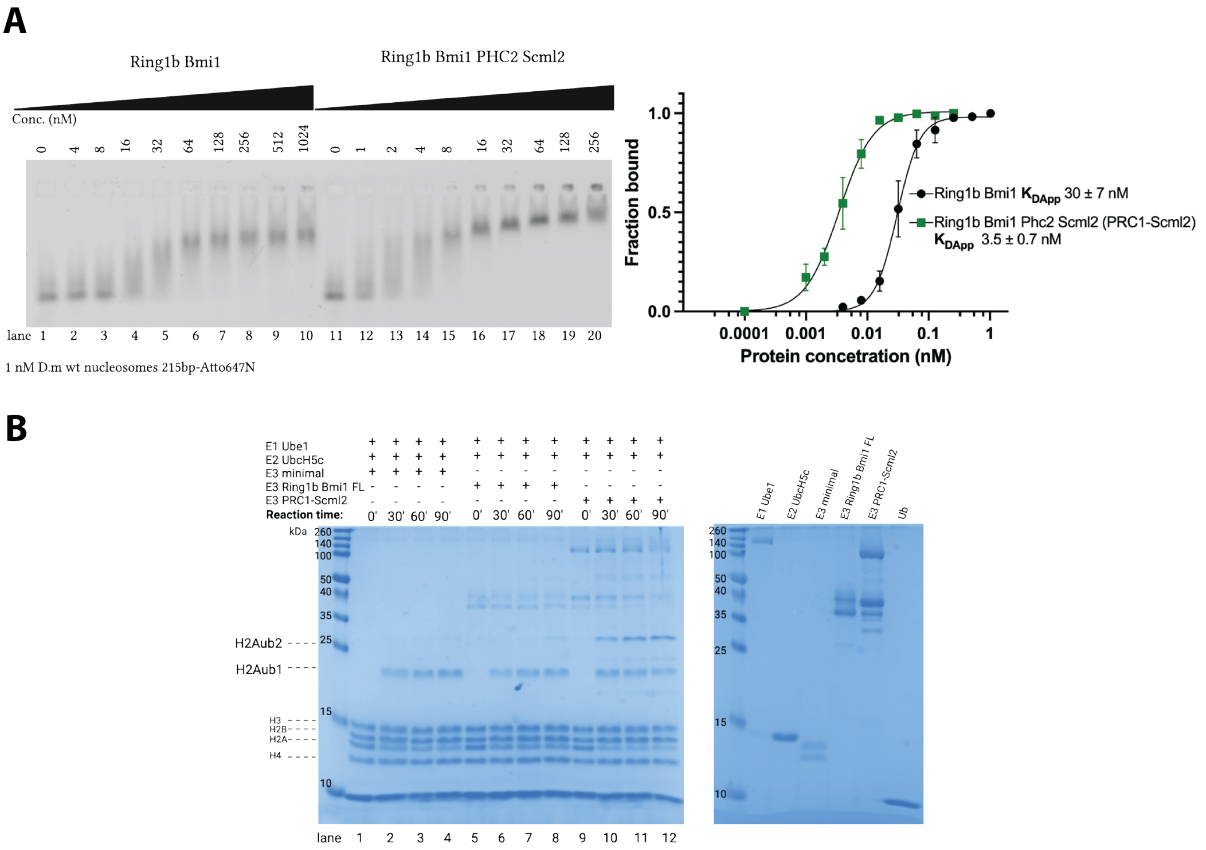


Figure 32: Biochemical characterisation of PRC1-Scml2 complex. A. EMSA was performed with PRC1-Scml2 complex on 215 bp Atto 647N-labelled *D.m* nucleosomes in triplicates, quantification by densitometric analysis shown on the left. B. *In vitro* H2AK118ub ubiquitination assay on reconstituted Drosophila melanogaster nucleosomes. The left panel represents a Coomassie-stained 16% SDS PAGE with a time course fractions of ubiquitination reaction. The bands corresponding to ubiquitylated H2AK118ub1 and H2AK118ub2 are marked with dashed lines. The right panel shows a Coomassie-stained 16% SDS PAGE gel of all proteins used in the ubiquitination reaction.

4.4.1 Identification of protein interactions in apo and nucleosome-bound PRC1-PHC2-Scml2 subcomplex by XL-MS

Previous studies showed binding of Scml2 alone to free nucleic acid and nucleosomes. The mediation of this interaction between Scml2 and nucleosomes has been primarily attributed to the SLED and RBR domains of Scml2. Additionally, the only known interaction between Scml2 and other subunits of PRC1 was believed to occur through SAM-SAM interactions involving PHC2 (Bonasio et al., 2014). To explore both potential interactions of Scml2 with nucleosomes and intermolecular interactions within the Scml2-containing PRC1 complex, I conducted a cross-linking mass spectrometry experiment using BS3 crosslinker.

The PRC1-PHC2-Scml2 complex was subjected to XL-MS independently and as a mixture with wild-type *Drosophila melanogaster* nucleosomes (for a detailed description of the experiment see Method section 3.9.3). Coomassie-stained SDS-PAGE gels of non-crosslinked and cross-linked samples are depicted in Figure 33A (left panel) for PRC1-PHC2-Scml2 complex alone and Figure 33B (left panel) for PRC1-PHC2-Scml2:nucleosomes.

Cross-linking mass spectrometry analysis of the PRC1-PHC2-Scml2 complex alone yielded a total of 110 reproducible cross-links detected in two replicas. Among those, 61 were determined to be intramolecular cross-links, while the remaining 49 were classified as heteromeric cross-links. The map of intramolecular and heteromeric cross-links is shown in Figure 33B, wherein intramolecular cross-links are depicted in violet and heteromeric in green.

In total, six protein-protein interfaces (PPIs) were identified within the complex based on the density of cross-links, as suggested by xiView analysis tool (Graham et al., 2019). Specifically, significant numbers of cross-links were observed in the following interfaces:

1. Between PHC2 and Scml2, with 18 cross-links detected, involving PHC2 FCS domains and all Scml2 domains, as well as cross-links between PHC2 HD and Scml2 MBT.

2. Between PHC2 and Ring1b, with 16 cross-links identified, including cross-links between the C-terminal end of the Ring domain and PHC2 HD, as well as between an unstructured region adjacent to the Ring domain and PHC2 FSC.

3. Between PHC2 and Bmi1, with 7 cross-links observed, encompassing interactions between the Bmi1 Ring domain and PHC2 HD and FSC domains, as well as between Bmi1 RAWUL and PHC2 SAM.

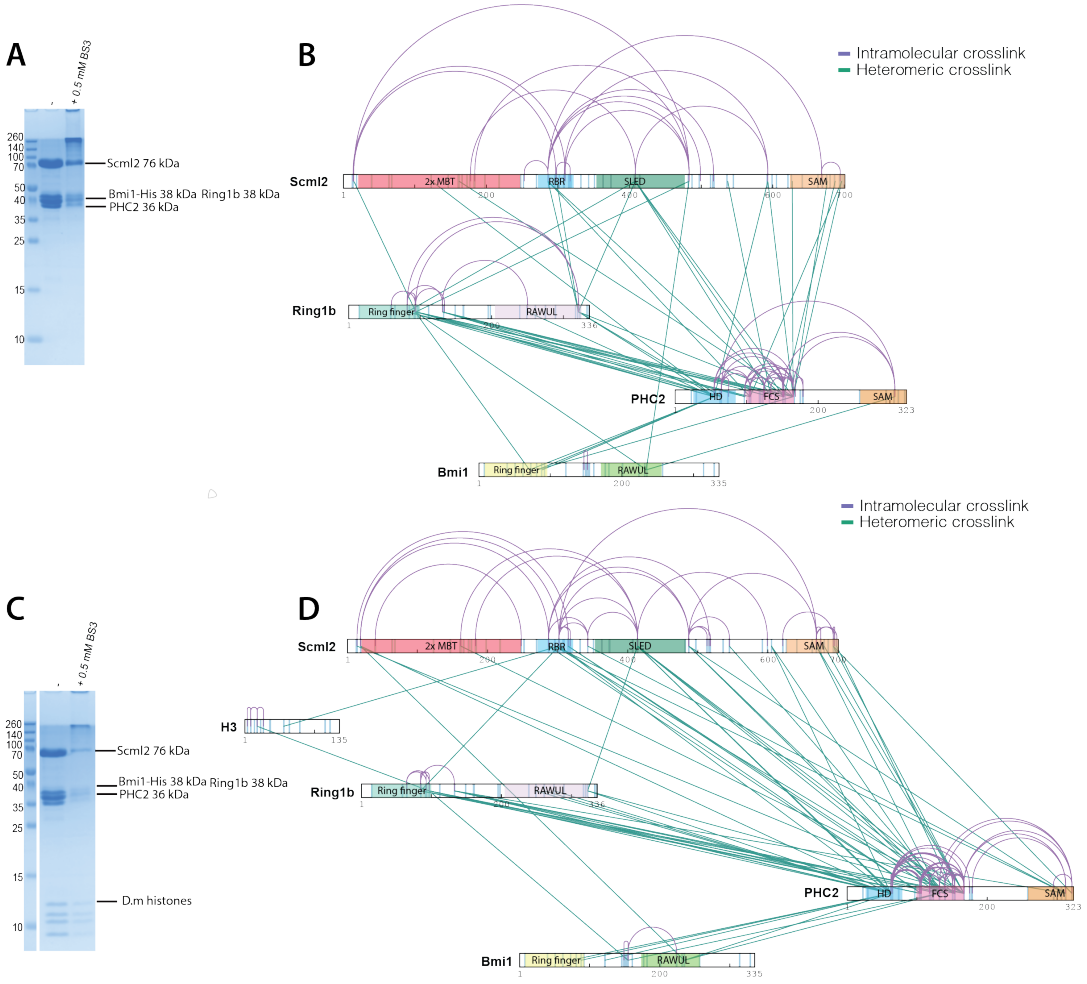


Figure 33: Cross-linking mass spectrometry (XL-MS) of PRC1-Scml2 complex alone and with nucleosomes. A. SDS-PAGE of PRC1-Scml2 complex intact and cross-linked with 0.5 mM BS3 for 30 C at RT. B. Intramolecular (in purple) and heteromeric (in green) BS3 crosslinks mapped on protein sequence of PRC1-Scml2 subunits. C. Coomassie stained 16% SDS-PAGE analysis of PRC1-PHC2-Scml2 complex with nucleosome intact and crosslinked with 0.5 mM BS3 for 30 min at RT. B. Intramolecular (in purple) and heteromeric (in green) BS3 crosslinks mapped on protein sequence of PRC1-Cbx7 subunits and histones. XL-MS experiments were performed in duplicates, and cross-links present in both replicas were mapped in XiView.

4. Between Ring1b and Scml2, with 6 cross-links identified, involving interactions between Ring1b Ring and Scml2 SLED, Ring1b C-terminus and Scml2 RBR and SLED.

5. Between Ring1b and Bmi1, with 2 cross-links between Ring domains, corresponding to known interactions of these domains.

Remarkably, half of these PPIs were mediated through the PHC2 protein, suggesting its conformation flexibility within the PRC1-PHC2-Scml2 complex, as in many instances lysine residues in PHC2 were involved in multiple cross-links: one of the examples is residue K151 in PHC2 HD domain that cross-links to K93 in Rin1gb and to K686 in Scml2 (Figure 33B). As it will be later discussed, AF2 prediction for PHC2 locates FCS domain in a long loop between SAM and HD domains, which could reflect its promiscuous interactions detected in the XL-MS experiment.

Regarding the position of Scml2 within the PRC1 complex, it has shown a larger number of cross-links to PHC2 (18 cross-links) and Ring1b (6 cross-links) and a single cross-link to Bmi1 (Figure 33B). Even though all these interactions might reflect an ensemble of conformational states, Scml2 within PRC1 most likely orients towards PHC2 and Ring1b.

Several observations could be made upon inspection of intramolecular cross-links. Notably, several sequentially distant cross-links were detected within Scml2. Specifically, cross-linked lysines were found within the vicinity of the MBT domain, connecting it to the RBR domain, as well as linking the MBT domain to the proximity of the SLED domain. Furthermore, such long-distance intamolecular cross-link was observed between the RBR and SAM domains. This suggests that Scml2 might adopt a compact conformation within the complex. It is worth mentioning, that only one intramolecular cross-link (between K149 and K154) was detected for the Bmi1 subunit, which could reflect a more extended conformation of this protein within the PRC1-PHC2-Scml2 subcomplex.

XL-MS of PRC1-PHC2-Scml2 in the presence of nucleosome

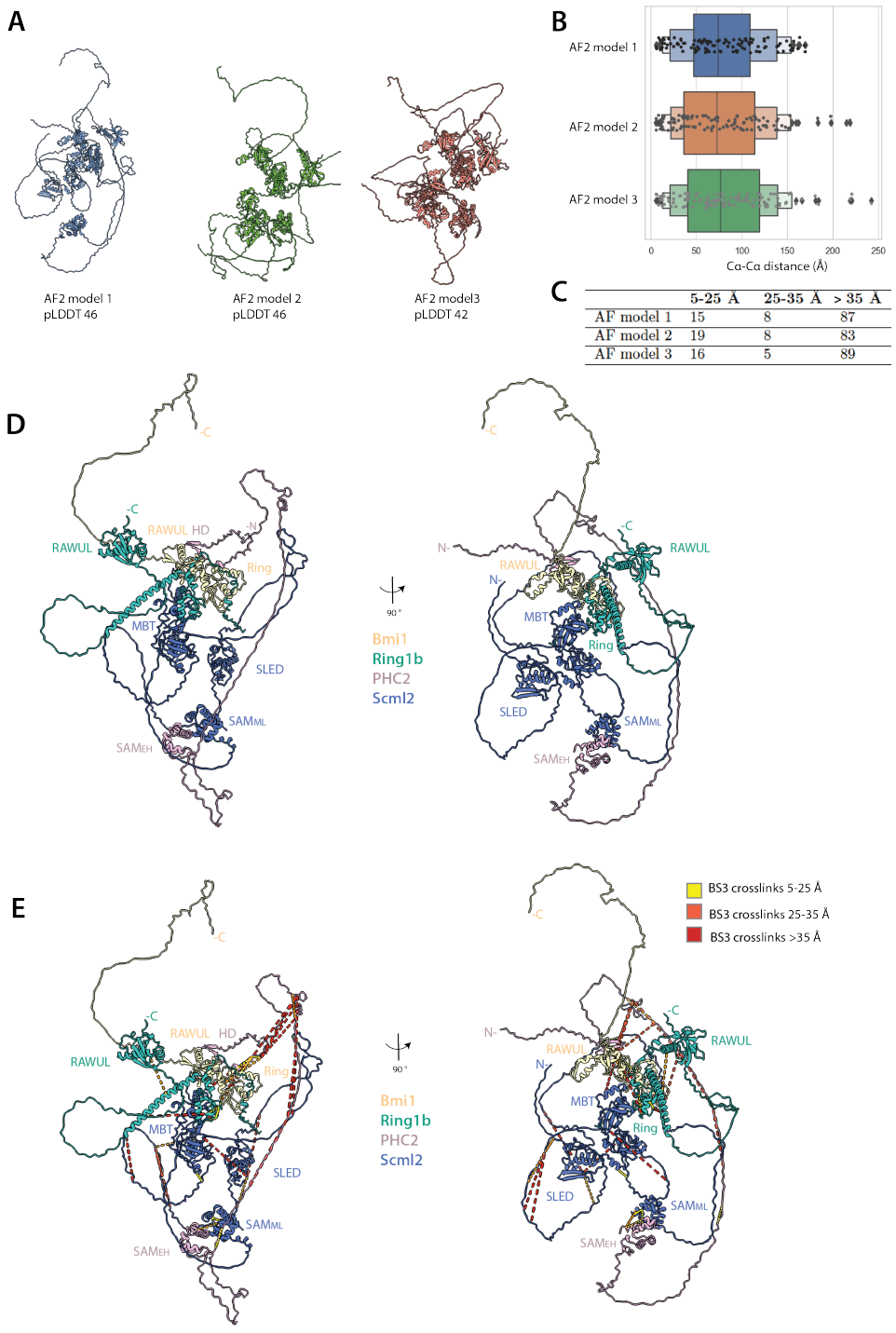
124 cross-links were present in two replicates of the XL-MS experiment using a PRC1-PHC2-Scml2 mixture with nucleosomes. Out of those, 4 cross-links were intra-nucleosomal and 2 cross-links were detected between PRC1-PHC2-Scml2 and nucleosome (depicted in Figure 33D). One cross-link detected reflected a proximity between the Ring domain of Ring1b (K97) and the N-terminus of H3 (K18). The second cross-link involved lysine K302 in the Scml2 RBR domain

and K56 of histone H3. Notably, H3K56 is located close to the nucleosomal DNA dyad, which likely reflects the interaction between the Scml2 RBR domain and nucleosomal DNA.

Mapping PRC1-PHC2-Scml2 cross-Links onto AlphaFold2 models

To explore the agreement between experimental cross-linking mass spectrometry (XL-MS) data and models of PRC1-PHC2-Scml2 complex predicted by AlphaFold2 (AF2), I employed AF2 in multimer mode to produce a set of 25 models of PRC1-PHC2-Scml2 complex. Each of the generated models displayed a moderate predicted local distance difference test (pLDDT) score, with the highest-scoring model achieving a score of 46 on a scale ranging from 0 to 100. This implies a prediction confidence level that falls beneath the halfway point on the scale, indicating a moderate degree of uncertainty in the prediction. All models consistently exhibited the same spatial position of Ring1b:Bmi1 Ring fingers and Bmi1 RAWUL domain, while diverging in the predicted positions of the other complex domains across the various models. For illustration, the top three models with the highest pLDDT scores are displayed in Figure 34A. All models had multiple regions predicted as unstructured (C-terminus of Bmi1, interdomain regions in Ring1b, PHC2, Scml2, as annotated in Figure 34D). Using XMAS implementation in ChimeraX, the experimental cross-links were mapped onto predicted models (Lagerwaard et al., 2022). To access agreement between XL-MS data and AF2 models, Cα-Cα distance between mapped cross-linked residues was plotted for all models (Cα-Cα distance distribution plots for three top-scored AF2 models are shown in Figure 34B). For all models, the distance distribution showed the prevalence of cross-links that exceeded the allowed range ($>$ 35Å, see the table in Figure 34C summarises amounts of cross-links across different distances when mapped on top-scored models).

AF1 model 1 was selected as an illustrative example and shown closely in Figure 34D, with individual subunits and domains annotated. Figure 34E showcases the same model with mapped cross-links. It is worth noting that cross-links falling within the allowed range (5-25 Å, highlighted in yellow, 25-35 Å shown in orange) were primarily concentrated within the Ring1b:Bmi1 Ring-Ring dimer and also within the SAM-SAM dimer. In contrast, cross-links that exceeded the allowed range ($>$ 35 Å, shown in red) were predominantly situated within the flexible loops connecting individual domains, for example by the loop between the PHC2 FCS and SAM domains.



	5-25 Å	25-35 Å	> 35 Å
AF model 1	15	8	87
AF model 2	19	8	83
AF model 3	16	5	89

Figure 34: AlphaFold2-predicted models of PRC1-PHC2-Scml2 and their alignment with distance restrains provided by XL-MS. A. AlphaFold2 models of PRC1-PHC2-Scml2 complex with highest overall pLDDT scores. B. Cα-Cα distance distribution of BS3 cross-links mapped on AlphaFold2 models of PRC1-PHC2-Scml2, visualized with XMAS. C. Overview of the BS3 cross-link counts categorized based on various Cα-Cα distances, when mapped on AlphaFold2 models of PRC1-PHC2-Scml2. D. Annotated representation of AlphaFold2 model 1 of PRC1-PHC2-Scml2 complex. E. BS3 cross-links mapped on PRC1-Cbx7 AlphaFold2 model 1. The cross-links are colour-coded to indicate their respective Cα-Cα- distances: violated cross-linking ($>$ 35 Å) shown in red, permitted cross-linking distances (5-25 Å) in yellow and cross-links within the range of 25-35 Å depicted in orange. Visualized in XMAS (Lagerwaard et al., 2022).

4.4.2 Prediction of intrinsically disordered regions and their function within PRC1 subunits

To understand the origin of structural heterogeneity within PRC1 that has been observed in cryo-EM studies of this complex, I opted to investigate the presence of intrinsically disordered regions (IDRs) within PRC1 subunits. Recently, there has been a significant enhancement in the prediction of IDRs owing to the implementation of deep neural networks.

For the prediction of IDRs, I used a fIDPnn computational tool, which employs a deep neural network to predict disorder and provide insights into the potential function of disordered regions (Hu et al., 2021). FIDPnn assigns a disorder propensity score ranging from 0 to 1 to each amino acid, where a threshold above 0.4 indicates a significant likelihood of disorder. For most PRC1 subunits, the previously identified structured domains matched well with the prediction (see the low propensity scores for Ring1b Ring and RAWUL domains and Cbx7 chromo and C-box domains in Figure 35). It also gave high propensity scores for interdomain regions in Ring1b, Cbx7 and for C-terminus of Bmi1.

Not surprisingly, the predictions from fIDPnn aligned well with the known structured domains of almost PRC1 subunits (see the low propensity scores for Ring1b and Bmi1 Ring and RAWUL domains and Cbx7 chromo and C-box domains in Figures 35A-B and Figures 35C). Conversely, fIDPnn assigned high propensity scores to interdomain regions in Ring1b, Cbx7, the C-terminus of Bmi, and the region between SLED and SAM domains of Scml2, suggesting their potential disorder. The interdomain region between Bmi1 Ring and RAWUL domains is predicted to be partially structured with propensity scores below 0.3 for a stretch of amino acids from 122 to 146 (Figure 35B).

As for the functional prediction for IDRs within PRC1, the region between Ring1b Ring and RAWUL has a medium confidence prediction for protein binding, while a stretch of amino acids before RAWUL domain is predicted to interact with nucleic acids (Figure 35A). The interdomain region in Cbx7 has also a similar prediction of protein binding with some regions being predicted as DNA and RNA binding (Figure 35D). Notably, the RBR domain within Scml2 is predicted by fIDPnn to bind to nucleic acids, a finding consistent with previous experimental evidence (Bonasio et al., 2014). Also, this region had borderline fIDPnn scores from 0.3 to 0.4 on disorder propensity, however, it was predicted to be fully unstructured by AlphaFold2. This does not exclude the possibility of it being structured upon binding to DNA/RNA or nucleosomes, as

suggested by cross-linking MS data presented in the previous section.

The fIDPnn disorder probability scores for PHC2 SAM, HD and FSC were below the threshold of 0.2 indicating strong confidence in predictions for these domains being structured, which is in agreement with previous studies. The N-terminus of PHC2 had scores above 0.3 indicating a propensity for the disorder. However, when classifying the number of amino acids predicted as disordered or structured across the entire protein sequence (based on the fIDPnn score below or above 0.4), PHC2 appears to have a high proportion of structured regions. However, this may result from overall low confidence in prediction for PHC2. In contrast, a significant portion of other PRC1 subunits was predicted to be disordered. This ranges from 65% of the Cbx7 sequence to 27% of Scml2. Ring1b and Bmi1 exhibit a similar prediction, with approximately 40% of their regions being predicted as disordered.

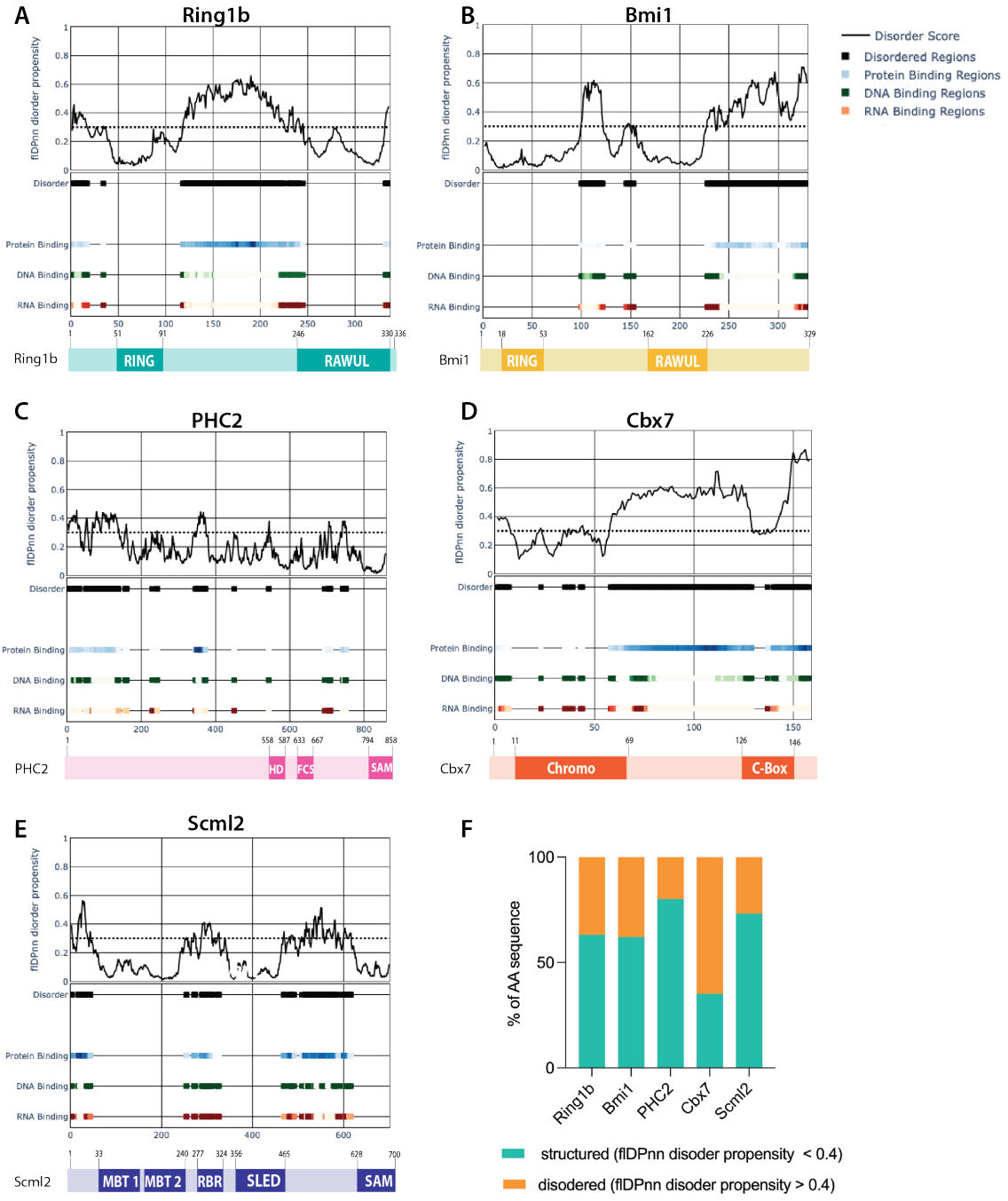


Figure 35: Prediction of intrinsically disordered regions within PRC1 complex using deep neural networks-based prediction tool flDPnn. A-E. Predicted Intrinsically Disordered Regions and their functional implications within canonical PRC1 subunits using flDPnn (putative function- and linker-based Disorder Prediction using deep neural network, Hu et. al, 2021). The x-axis represents the amino acid sequence of the respective protein, while the y-axis depicts the disorder propensity score. A value above 0.4 indicates the presence of disorder. F. Distribution of disorder flDPnn scores for each canonical PRC1 subunit. The y-axis represents the percentage of the amino acid sequence.

5 Discussion and conclusions

Recombinant reconstitution of various canonical PRC1 subcomplexes

In the first part of this thesis, expression and purification of various canonical PRC1 subcomplexes were established to obtain highly pure PRC1 samples for biochemical and structural studies. Several recombinant canonical PRC1 subcomplexes were cloned, expressed and purified for this study. These subcomplexes included the minimal PRC1-PHC2 expressed in bacteria; PRC1-PHC2 containing full-length Ring1b, Bmi1, and the PHC2 isoform; PRC1-Cbx7 composed of Ring1b, Bmi1, and the short isoform of Cbx7, and PRC1-Scml2 comprising Ring1b, Bmi1, the short isoform of PHC2 with EH-SAM mutation, and Scml2 with ML-SAM mutation. Apart from the minimal PRC1-PHC2, expression of all PRC1 subcomplexes was established in insect cells (*Trichoplusia ni.* HighFive cell line), using a combination of pBIG1 and pFastBac expression vectors.

First, the approach was tested, wherein 'minimal' canonical PRC1-PHC2 was assembled to contain previously crystallised domains, to obtain a core of canonical PRC1 suitable for crystallisation. The minimal PRC1-PHC2, contained Ring1b 1-133, Bmi1-1-242 and PHC2 536-678 and could be purified from *E.coli* cells with the yield of 600 μg/l. The complex was shown to be an active E3 ligase, showing almost complete conversion of H2A to H2AK118ub after 1.5 h of reaction time (Figure 12A, right panel). Additionally, the EMSA assay showed that the minimal PRC1-PHC2 complex interacted with nucleosomes with an apparent Kd of 35.3 $\pm$ 4.6 nM, providing a qualitative indication of binding (Figure 12B). However, the bacterial expression system presented limitations for the expression of full-length PRC1 subunits. Specifically, expression of the Ring1b construct beyond amino acids 1-133 or PHC2, with the C-terminus extended beyond residue G678, resulted in protein degradation in both cases. Consequently, expression of other PRC1 subcomplexes was established in insect cells.

The PRC1-PHC2 complex expressed in insect cells contained full-length Ring1b and Bmi1 subunits. Although not stressed out in the Results section, attentive readers might have noticed that there were two versions of this complex. First, PRC1-PHC2 which contained PHC2 with the full C-terminus including mutated SAM domain (with two mutations in the EH surface) was expressed and used for cryo-EM studies. However, PHC2 was truncated during expression and purified in both intact and truncated form (see Figure 14C, 16 kD truncated PHC2). Consequently, the PHC2 construct was reduced to amino acids 538-678, mirroring the bacterial

construct to avoid sample heterogeneity due to the presence of truncated PHC2. Both versions of the PRC1-PHC2 subcomplex were obtained with high purity, yielding approximately 270 μg per litre of culture.

Another complex expressed and purified in this study is the PRC1-Cbx7, which comprised full-length Ring1b and Bmi1 subunits and a naturally occurring Cbx7 isoform characterized by a shortened region between the Chromo and Cbox domains. The expression of several human Cbx homologs — Cbx2, Cbx6, Cbx7, and Cbx8 was tested in this study (data not shown), however, it was observed that these proteins exhibited low expression levels and were truncated in the region between the Chromo and Cbox domains. This susceptibility to degradation is likely attributed to this region being unstructured and exposed to proteolytic cleavage, as indicated by the fIDPnn disorder prediction for Cbx7 (Figure 35A). The utilization of the Cbx7 isoform with a shorter interdomain region, due to its improved expression levels and stability, enabled the reconstitution of PRC1-Cbx7 subcomplex in amounts sufficient for structural studies (Figure 20D). The binding of PRC1-Cbx7 to wild-type and H3Kc27me3 (methyl-lysine analogue of H3K27me3) was compared in EMSA assay, with no difference between binding to wild-type and H3Kc27me3 nucleosomes (Figure 21A). While the interaction of full-length Cbx7 with histone H3K27me3 or nucleosomes bearing H3K27me3 modification has not been tested *in vitro* to date, some studies have measured the Kd of the interaction between the Cbx7 Chromodomain and H3K27me3 peptide. These studies reported a Kd of 22 μM for the mouse Cbx7 Chromodomain and 110 μM for the human Cbx7 Chromodomain (Bernstein et al., 2006, Kaustov et al., 2011). Hence, it seems plausible that the affinity of full-length Cbx7 to methylated nucleosomes would also be within the micromolar range, thereby making only a marginal contribution to the overall binding affinity of PRC1-Cbx7 to nucleosomes. Ideally, future studies should examine the binding of individual Cbx7 subunits to H3K27me3-modified nucleosomes, preferably utilizing H3K27me3 rather than a methyl-lysine analogue H3Kc27me3. Similarly, the E3 ligase activity of the PRC1-Cbx7 complex was tested in ubiquitination assay using wild-type and H3Kc27me3 nucleosomes as substrate. Interestingly, PRC1-Cbx7 seemed to be only slightly more active than minimal Ring1b-Bmi1 (Figure 21B, compare H2A to H2Aub conversation at 30' for minimal PRC1 and PRC1-Cbx7). PRC1-Cbx7 seemed to have comparable activity on wild-type and H3Kc27me3-modified nucleosomes, with a slight difference in the deposition of the second ubiquitin (Figure 21B, compare 60' time point for wt and H3Kc27me3). However, the polyubiquitination of H2A (hence, deposition of the second and third ubiquitin resulting in H2Aub2

and H2Aub3) is a phenomenon that was previously observed in *in vitro* ubiquitination studies involving minimal PRC1 E3 ligase module and most likely lacks any *in vivo* functional relevance (Scheuermann, 2011).

Finally, the last complex reconstituted and characterized in this study was the PRC1-Scml2 complex. The PRC1-Scml2 comprised full-length Ring1b and Bmi1, along with PHC2 (536-853) harbouring inactivated EH surface of SAM domain and Scml2 with SAM ML-mutated surface. The introduction of mutations in the EH and ML surfaces ensured that PHC2 and Scml2 SAM domains could interact without SAM-SAM polymer formation. Surprisingly, in EMSA nucleosome-binding assays, the PRC1-Scml2 complex exhibited nearly a 10-fold lower apparent Kd compared to the Ring1b-Bmi1 dimer alone, indicating a significantly higher affinity for nucleosomes (3.5 nM for PRC1-Scml2 versus 30 nM for Ring1b-Bmi1, Figure 32A). Although the comparison was not entirely direct, as it was made between a complex lacking PHC2 (Ring1b-Bmi1 dimer) versus one containing both PHC2 and Scml2, this difference in affinity could be nevertheless attributed to Scml2. This conclusion is supported by the fact that the apparent Kd measured for the PRC1-PHC2 complex was similar to that of the Ring1b-Bmi1 dimer, indicating that PHC2 did not contribute significantly to nucleosome binding (Figure 12B). Furthermore, this finding aligns with the results of cross-linking MS, which revealed no cross-links between PHC2 and nucleosomes, whereas Scml2 consistently exhibited a cross-link between its RBR domain and histone H3 (to be discussed later in this chapter). Regarding the E3 ligase activity of PRC1-Scml2, it was shown to be an active E3 ligase in the ubiquitination assay, albeit with a higher production of H2Aub2 (Figure 32B). Notably, in this assay, full H2Aub conversion was achieved within 30 minutes of reaction time, unlike in previous ubiquitination assays conducted in this study. This disparity could be attributed to variations in the batches of E1 and E2 enzymes used. Hence, it is advisable to conduct this assay with samples analyzed at earlier time points to discern any potential differences in E3 ligase activity among PRC1-Scml2, Ring1b-Bmi1 dimer, and minimal E3 ligase domain.

While numerous studies have extensively investigated canonical PRC1 *in vivo* and *in cellulo* in recent years, comparatively less research has been devoted to the studies involving *in vitro* recombinant reconstitution of PRC1 complex, beyond just its subunits and domains. Previously, the first recombinant reconstitution of canonical *Drosophila* PRC1, containing Sce, Pc, Psc and Ph subunits was reported using a bac-to-bac system in Sf9 insect cells, albeit resulting in substoichiometric amounts of Sce and Pc (Francis et al., 2001). Other approaches to the isolation

of canonical PRC1 involved differential centrifugation on glycerol gradients of cell extracts (Gao et al., 2012) or native complex isolation from nuclear extracts using ion exchange (Wang et al., 2004), with both approaches usually yielding protein samples of lower quantity and purity. At the beginning of this project, a recombinant reconstitution of canonical PRC1 that included Ring1b, PCGF4, CBX2 and a short isoform of PHC2 was reported (Colombo et al., 2019). Recently, another study involved the reconstitution of canonical PRC1 subcomplex consisting of Ring1b, PCGF4 (Bmi1) and CBX8 subunits (Uckelmann et al., 2023). Placed within the context of prior research on PRC1 complexes, the work presented in this thesis describes the recombinant reconstitution of several enzymatically active PRC1 subcomplexes, including those whose recombinant reconstitution has been previously unreported, such as PRC1 containing short isoform of Cbx7. Similarly, the PRC1 complex containing Scml2 has not been previously reconstituted. Moreover, expression and purification procedures demonstrated herein could be used for the reconstitution of larger PRC1 assemblies, such as PRC1 containing both Cbx7 and Scml2, that could be of interest for future studies.

Crystallisation of canonical PRC1

Following successful biochemical characterization, crystallization trials for two PRC1 subcomplexes were conducted. Crystallisation of minimal PRC1-PHC2 resulted in the formation of needle-like protein crystals. The crystal structure was resolved at 1.9 Å using molecular replacement with the structure of Ring1b-Bmi1 E3 ligase module (2CKL PDB) as a reference model (Buchwald et al. (2006), Figure 13). Surprisingly, the structure was identical to 2CKL, with an RMSD of 0.158 Å and contained only the E3 ligase domain of Ring1b and Bmi1, with no electron density corresponding to PHC2 or the C-terminus of Bmi1 (Figure 13C-D). This likely occurred due to the degradation of the complex during crystallization, probably as a result of proteolytic degradation by contaminating proteases, as the SDS-PAGE analysis showed that the crystals contained only degraded portions of proteins, that indeed corresponded in size to Ring domains of Ring1b and Bmi1 (Figure 13A, right). This unexpected result could be interpreted as a serendipitous *in-situ* proteolysis crystallization trial, wherein proteolytic degradation coincided with crystallization. This process identified a crystallizable portion of the protein complex, where regions accessible to proteases, likely due to the lack of secondary structure, were degraded. Moreover, this observation aligns with the intrinsic disorder predictions by fIDPnn, which indicated that the region of Bmi1 between the Ring and RAWUL domains had a

higher probability of disorder and therefore could be susceptible to proteolytic cleavage (Figure 35C). As a result, a portion of the Bmi1 RAWUL domain and PHC2 may have been cleaved away, facilitating the nucleation of crystals with only the Ring1b and Bmi1 Ring domains, as previously reported by (Buchwald et al., 2006). Similarly, crystallization trials for PRC1-Cbx7 yielded crystals with the same needle-like morphology. The SDS-PAGE analysis of these crystals again revealed degraded protein fragments, approximately 13 kDa in size, likely corresponding to Ring domains of Bmi1 and Ring1b.

Future studies might explore the crystallization of PRC1-PHC2 and PRC1-Cbx7 subcomplexes with nucleosomes. However, results from XL-MS and EM experiments presented in this study suggest that neither of these complexes forms additional interactions with nucleosomes apart from the interaction of Ring domains with the nucleosome acidic path. Also, as shown by cryo-EM 3D variability analysis, the regions of the complex beyond the Ring domains of Ring1b and Bmi1 were continuously heterogeneous. Therefore, it is likely that crystallization of PRC1-Cbx7 or PRC1-PHC2 complexes with mononucleosomes would not yield crystals, unless these complexes undergo degradation to nucleosome-interacting Ring-Ring domains. The outcome of potential crystallization trials for the PRC1-Scml2 subcomplex remains less clear. This complex exhibited a higher affinity for nucleosomes, and additional nucleosome-interacting surfaces (such as the RBR domain, as discussed later) were revealed by XL-MS. However, due to the presence of unstructured regions in all PRC1 subunits (Figure 35A), it might be more suitable to attempt crystallization of only the nucleosome-interacting RBR domain of Scml2 with nucleosomes.

Canonical PRC1 complex architecture and protein-protein interactions surfaces

As the crystallisation of the canonical PRC1 subcomplexes proved difficult, cross-linking mass spectrometry was used to gain insights into the structure of PRC1 subcomplexes and their interactions with nucleosomes. Overall, three PRC1 subcomplexes — PRC1-PHC2, PRC1-Cbx7, and PRC1-Scml2 — were subjected to XL-MS using BS3 as a cross-linker. All complexes were cross-linked either individually or in combination with nucleosomes. Distance restraints generated from cross-links were visualized in 2D connectivity plots and also mapped onto AF2-predicted models with the highest confidence of prediction scores.

In terms of the intramolecular organisation of PRC1 subcomplexes, several conclusions could

be made. The Ring1b subunit in all three complexes adopts a compact conformation, as indicated by cross-links between the N-terminal and C-terminal portions of the complex (see Figure 15B, Figure 23B and Figure 33B). Similarly, intramolecular cross-links between Ring and RAWUL domains of Bmi1 were detected for PRC1-PHC2 and PRC1-Cbx7 complexes, suggesting that Bmi1 as well could adopt a conformation that brings these two domains into proximity. Notably, these Bmi1 cross-links were absent in the PRC1-Scml2 complex, indicating a potentially distinct Bmi1 conformation in this subcomplex, with a greater distance between its N- and C-termini.

The intramolecular cross-linking data provided insights into subunit interaction surfaces within PRC1 complexes. Some findings aligned with known interactions, such as those between the Ring domains of Ring1b and Bmi1. However, not all established interactions were captured by the cross-linking analysis. For instance, the known interaction between Ring1b RAWUL and Cbx7 C-box was not detected, likely due to the scarcity of lysine residues in the Cbx7 C-terminus. PHC2-containing PRC1 subcomplexes, such as PRC1-PHC2 and PRC1-Scml2 both showed numerous cross-links between PHC2 and other complex subunits, often with the same lysine of PHC2 being involved in several cross-links with residues from different subunits, this most likely reflects PHC2 weaker association of PHC2 with Bmi1 (as was shown previously by Gray et al. (2016)) and its conformational dynamics within PRC1 complex.

Conformational heterogeneity of all PRC1 subcomplexes was also reflected by *in silico* analysis of AF2 predicted models with distance restraints provided by XL-MS experiments. All AF2 predictions were of lower confidence (the highest pLDDT score for PRC1-PHC2 was 53, for both PRC1-Cbx7 PRC1-Scml2 46), likely due to presence of the intrinsically disordered regions as was shown by deep neural network-based predictor fIDPnn (Figure 35). When applying distance restraints based on experimental cross-links to the top-scored AF2 models, Cα-Cα distances exhibited a wide distribution for all PRC1 subcomplexes. Most cross-links violated the 35 Å threshold, indicating that they did not preferentially align with any single model. This outcome suggests that the XL-MS data captures a conformational ensemble of PRC1 subcomplexes rather than a single, rigid structure. Despite not aligning with individual models, these experimental cross-links could provide valuable distance restraints for more *in silico* modeling approaches. Methods such as HADDOCK (High Ambiguity Driven protein-protein DOCKing) or emerging AlphaFold2-based modeling tools could be employed to more precisely define the conformational landscape of PRC1 complexes (Van Zundert et al., 2016, Stahl et al., 2023).

In terms of interactions with nucleosomes, XL-MS analysis of all PRC1 subcomplexes showed cross-links consistent with the established mode of interaction of the PRC1 E3 ligase module with the nucleosome acidic patch (McGinty et al., 2014). Notably, all cross-links of the E3 ligase module to the nucleosome were localized within the Ring domain of Ring1b. Interestingly, there were slight differences among PRC1 subcomplexes regarding which histones were cross-linked to the Ring1b Ring domain. In the case of the PRC1-Cbx7 complex, the Ring domain cross-linked with the C-terminal part of histone H2B and the N-terminal tail of histone H3. For the PRC1-PHC2 complex, the detected cross-links were between the Ring1b Ring domain and the N-terminus of histone H4, as well as the C-terminus of histone H2B, with no cross-links detected with histone H3. Although these cross-links align with the known mode of interaction, these subtle discrepancies may arise from variations in the conformational dynamics of different PRC1 subcomplexes. Additionally, since the Chromodomain of Cbx7 is known to interact with the H2K27me3, it was expected to observe cross-links between the N-terminal H3 tail containing trimethyl-lysine modification analogue at K27 (H3Kc27me3) and the Chromodomain. These cross-links were indeed detected in individual replicas, although no reproducible cross-link was found. This likely reflects the weak interaction between the Chromodomain and H3K27me3, with a reported Kd of 110 μM (Kaustov et al., 2011). However, this assumption should be validated using nucleosomes with natively methylated H3K27me3 modification. In the case of the PRC1-PHC2 subcomplex, no additional cross-links were detected between the nucleosome and PRC1-PHC2, aside from those indicative of interaction with the acidic patch. This observation suggests that the PHC2 subunit within this complex does not come into proximity with the nucleosome, contrary to the hypothesis that it may bind nucleosomal DNA via its FCS zinc finger domain.

A significant finding emerged from the XL-MS analysis of the PRC1-Scml2 complex: a reproducible cross-link suggesting a novel mode of interaction with the nucleosome. Specifically, residue K302 of Scml2 was found to cross-link with K56 of histone H3. This interaction is particularly noteworthy as H3K56 is not part of the N-terminal tail but is instead located in the globular region near superhelical location (SHL) 7 of the nucleosomal DNA, theretofore, making it plausible that RBR domain interacts with DNA near SHL7. The implicated Scml2 residue, K302, resides within the RBR domain, which has previously been shown to bind nucleic acids (Bonasio et al., 2014, Bezsonova, 2014). This suggests that the RBR domain may play a crucial role in nucleosome binding. Indeed, this interaction could explain the 10-fold higher

affinity of the PRC1-Scml2 complex for nucleosomes compared to the Ring1-Bmi1 dimer alone, as observed in nucleosome-binding assays (Figure 32). This finding necessitates structural studies of the RBR domain in complex with the nucleosome, which could reveal an additional, previously uncharacterized mode of interaction between Scml2-containing PRC1 complexes and chromatin.

Previously, only one study had reported XL-MS analysis of the canonical PRC1 complex, albeit with a different composition of PRC1 (Colombo et al., 2019). In that study, the canonical PRC1 complex comprised Ring1b, Mel18 (PCGF2), Cbx2, and the short isoform of PHC2 and was cross-linked with the DSS cross-linker. Notably, this PRC1 complex differed from the one investigated in this thesis, as it contained different PCGF (PCGF2) and Cbx (Cbx2) homologs. Furthermore, unlike PRC1-PHC2 and PRC1-Cbx7 subcomplexes studied herein, in Colombo et al. (2019), both PHC2 and Cbx2 were part of a single PRC1 complex. Consistent with the findings on the PRC1-PHC2 presented in this thesis, it was observed that PHC2, particularly its FCS domain, was cross-linked to multiple other subunits, including Ring1b and Bmi1. Notably, Colombo et al. (2019) did not detect any cross-links of the Cbx2 Chromodomain with other subunits. In contrast, this study of the PRC1-Cbx7 subcomplex revealed numerous reproducible cross-links between the Chromodomain and the Ring domain of Ring1b, as well as the RAWUL domain of Bmi1. This disparity could be attributed to the potentially more compact organisation of Cbx7, which results in the physical proximity of Cbx7 Chromodomain and Ring1b and Bmi1 subunits. Unlike Cbx7, Cbx2 possesses a long intrinsically disordered domain between its C-terminal C-box and N-terminal Chromodomain. Another plausible explanation is that the presence of PHC2, which, as demonstrated, also cross-links with Ring1b and Bmi1, could potentially impede the interaction of the Cbx2 Chromodomain with said subunits.

At the conclusion of this thesis's experimental phase, XL-MS analysis of PRC1-nucleosome interactions remained unreported in the literature. However, a recent preprint by Uckelmann et al. (2023) has since addressed this gap, probing interactions between nucleosomal arrays and a PRC1 subcomplex comprising Ring1b, Bmi1, and Cbx8 using XL-MS. Intriguingly, Uckelmann et al. (2023) reported numerous cross-links between the Cbx8 Chromodomain and the histone H3 tail, despite the absence of H3K27me3 modification. This finding contrasts with the XL-MS analysis of PRC1-Cbx7 and single nucleosomes presented in this thesis, where reproducible cross-links between the Cbx7 chromodomain and histone H3 were not detected. This disparity suggests a hypothesis that Cbx proteins may interact with nucleosomes not in a cis- but in a trans-mode, potentially contacting neighbouring nucleosomes. This model could explain why

such interactions were only detected in experiments using nucleosomal arrays (as in Uckelmann et al. (2023)) rather than single nucleosomes.

For future XL-MS studies, it would be of interest to conduct experiments using differently modified di-nucleosomes, such as those bearing H3K27me3 and/or H2AK119ub1 modifications, or even nucleosomal arrays. These nucleosomes could be modified heterogeneously, with one nucleosome remaining wild-type and the other bearing modifications (as in Finogenova et al. (2020)). This approach could provide insights into canonical PRC1 interactions within a more complex chromatin context, potentially revealing nuanced binding modes and preferences. Such experiments could elucidate how PRC1 complexes recognize and engage with chromatin in diverse genomic contexts with variably modified nucleosomes, potentially uncovering subcomplex-specific functions or preferences.

Cryo-EM studies of canonical PRC1 further confirmed conformational heterogeneity of the complex

Two subcomplexes reconstituted in this study underwent cryo-EM structural analysis: PRC1-PHC2 and PRC1-Cbx7, the latter analyzed using both native and glutaraldehyde-crosslinked samples. All three datasets were collected on Titan Krios equipped with Gatan K3 (MPI Cryo-EM facility) and processed using cryoSPARC v3.3 and, for PRC1-CBX7, Relion 3.1.3. The 3D reconstructed map of PRC1-PHC2 on the nucleosome was resolved to 3.2 Å and contained the density attributed to the Ring fingers of Ring1b and Bmi1 bound to the nucleosome without any additional density that could be attributed to PHC2 or the C-termini of Ring1b or Bmi1 (Figure 18B and Figure 18C).

Both cross-linked and uncross-linked PRC1-Cbx7 samples yielded high-resolution cryo-EM maps (resolved to 2.62 Å and 2.92 Å, respectively) comprising the density attributed to nucleosome with Ring fingers of Ring1b and Bmi1 (Figure 27A-C and 27E-G). The 3D-reconstructed cryo-EM maps of PRC1-Cbx7, derived from both native and fixated samples, exhibited additional "arc-like" density beyond the Ring fingers of Ring1b and Bmi1. Attempts to resolve this extra density through various classification techniques in both cryoSPARC and Relion did not yield reconstructions that would enable to attribute this extra density to any of the subunits with certainty (Figure 28). Nevertheless, models based on AF2 prediction and XL-MS cross-links between Bmi1 RAWUL and Cbx7 Chromodomain suggested that this density might correspond

to the RAWUL domain of Bmi1 (Figure 30).

Therefore, cross-linking with glutaraldehyde did not enhance the resolution of regions of PRC1 beyond the Ring domains of Ring1b and Bmi1 (E3 ligase module). This observation underscores the notion that cross-linking proves beneficial primarily wherein discreet conformational states are present and therefore could be stabilized. Conversely, in scenarios where the system exhibits high degree of continuous heterogeneity and, therefore, numerous conformational states, cross-linking fails to provide significant improvements of density resolution. Addressing this challenge of continuous heterogeneity may involve advancements in computational methods and sample preparation techniques. Computational approaches such as cryoDRGN or 3DFlex, show promise for future structural studies on PRC1 (Kinman et al., 2023, Punjani and Fleet, 2023). Moreover, alternative sample preparation strategies, such as reconstituting canonical PRC1 with di-nucleosomes or utilizing higher nucleosomal substrates like tri-nucleosomes or nucleosomal oligomers, could be explored. Other cross-linking methods, such as gradient fixation (GraFix) could be tested too (Stark, 2010). Additionally, incorporating histone H1 into the sample may offer additional stabilization by its binding to flexible linker DNA, or, potentially interacting with PRC1 subunits.

For the structural studies of canonical PRC1, continuing the study of PRC1-Scml2, given its higher affinity to nucleosomes in EMSA assay, presents a promising avenue for future research projects. Cryo-EM analysis could be conducted on both the whole PRC1-Scml2 complex and the RBR domain only, as this domain was shown to cross-link with nucleosome in XL-MS experiments which indicates likely involvement in nucleosome interaction.

Building on this study of various PRC1 subcomplexes, it could be possible to reconstitute and analyse larger PRC1 assemblies. For instance, a full PRC1 complex containing Ring1b, Bmi1, PHC2, Cbx7, and Scml2 subunits could be obtained by combining the expression constructs generated in this study. This approach could potentially reveal additional interaction surfaces and stabilizing elements that may facilitate the determination of PRC1 structure by cryo-EM.

Conclusion

In conclusion, this work has successfully established the recombinant reconstitution of several canonical PRC1 subcomplexes, providing a basis for future biochemical and structural studies of PRC1. Several subcomplexes of PRC1, including PRC1-PHC2, PRC1-Cbx7, and PRC1-Scml2, were obtained with high purity and yield. Crystallisation trials for PRC1-PHC2 and PRC1-Cbx7 were conducted, resulting in the crystal structure of Ring1b and Bmi1 Ring fingers identical to previous crystallisation studies, which most likely resulted from proteolytic degradation during crystallization. Cross-linking mass spectrometry and cryo-electron microscopy were used to study the protein-protein interaction interfaces and conformational heterogeneity of PRC1 subcomplexes. XL-MS analysis indicated several interaction surfaces within the subcomplexes and provided spatial constraints for future computational modelling. Meanwhile, cryo-EM studies of the PRC1-Cbx7 subcomplex revealed additional extra density beyond the Ring fingers of Ring1b and Bmi1, which could be attributed to the RAWUL domain of Bmi1. The results of XL-MS and cryo-EM experiments underscored the dynamic nature of PRC1 complexes and highlighted the importance of integrating multiple structural biology techniques in future studies.

Addressing the challenge of continuous heterogeneity in PRC1 complexes requires further advancements in computational methods and implementing different sample preparations. Computational approaches like cryoDRGN and 3DFlex hold promise for resolving heterogeneous structures. At the same time, alternative sample preparation strategies, such as using higher-order nucleosomal substrates that could be differently modified (to contain H3K27me3 and/or H2AK119ub1 marks), could offer additional stabilization of PRC1 subcomplexes. Furthermore, future research should also focus on PRC1-Scml2, which, for the first time, has been successfully recombinantly reconstituted in this study. Notably, the PRC1-Scml2 subcomplex was also shown herein to have a higher affinity for nucleosomes compared to other PRC1 subcomplexes, likely due to the binding of the RBR domain to nucleosomal DNA, although further confirmation through biochemical and structural studies is needed.

To conclude, further investigations into PRC1 complexes' structural and functional properties remain essential for unravelling their roles in chromatin regulation and epigenetic mechanisms, although new computational and sample preparation approaches have to be adopted to overcome the challenge of continuous heterogeneity that was shown to be inherent to canonical PRC1 complexes by this study.

References

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W. et al. (2010). Phenix: a comprehensive python-based system for macromolecular structure solution, *Acta Crystallographica Section D: Biological Crystallography* **66**(2): 213–221.

Afonine, P. V., Klaholz, B. P., Moriarty, N. W., Poon, B. K., Sobolev, O. V., Terwilliger, T. C., Adams, P. D. and Urzhumtsev, A. (2018). New tools for the analysis and validation of cryo-em maps and atomic models, *Acta Crystallographica Section D: Structural Biology* **74**(9): 814–840.

Ahmad, K. and Henikoff, S. (2002). Histone h3 variants specify modes of chromatin assembly, *Proceedings of the National Academy of Sciences* **99**(suppl_4): 16477–16484.

Ai, H., Tong, Z., Deng, Z., Tian, J., Zhang, L., Sun, M., Du, Y., Xu, Z., Shi, Q., Liang, L. et al. (2023). Synthetic e2-ub-nucleosome conjugates for studying nucleosome ubiquitination, *Chem* **9**(5): 1221–1240.

Akasaka, T., Kanno, M., Balling, R., Mieza, M. A., Taniguchi, M. and Koseki, H. (1996). A role for mel-18, a polycomb group-related vertebrate gene, during the anteroposterior specification of the axial skeleton, *Development* **122**(5): 1513–1522.

Akasaka, T., Lohuizen, M. v., Lugt, N. v. d., Mizutani-Koseki, Y., Kanno, M., Taniguchi, M., Vidal, M., Alkema, M., Berns, A. and Koseki, H. (2001). Mice doubly deficient for the polycomb group genes mel18 and bmi1 reveal synergy and requirement for maintenance but not initiation of hox gene expression, *Development* **128**(9): 1587–1597.

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

Allis, C. D. and Jenuwein, T. (2016). The molecular hallmarks of epigenetic control, *Nature Reviews Genetics* **17**(8): 487–500.

Althobiti, M., Muftah, A. A., Aleskandarany, M. A., Joseph, C., Toss, M. S., Green, A. and Rakha, E. (2020). The prognostic significance of bmi1 expression in invasive breast cancer is dependent on its molecular subtypes, *Breast cancer research and treatment* **182**: 581–589.

Arand, J., Spieler, D., Karius, T., Branco, M. R., Meilinger, D., Meissner, A., Jenuwein, T., Xu, G., Leonhardt, H., Wolf, V. et al. (2012). In vivo control of cpg and non-cpg dna methylation by dna methyltransferases, *PLoS genetics* **8**(6): e1002750.

Ardehali, M. B., Mei, A., Zobeck, K. L., Caron, M., Lis, J. T. and Kusch, T. (2011). Drosophila set1 is the major histone h3 lysine 4 trimethyltransferase with role in transcription, *The EMBO journal* **30**(14): 2817–2828.

Arents, G., Burlingame, R. W., Wang, B.-C., Love, W. E. and Moudrianakis, E. N. (1991). The nucleosomal core histone octamer at 3.1 a resolution: a tripartite protein assembly and a left-handed superhelix., *Proceedings of the National Academy of Sciences* **88**(22): 10148–10152.

Asarnow, D., Palovcak, E. and Cheng, Y. (2019). asarnow/pyem: Ucsf pyem v0.5. **URL:** *https://doi.org/10.5281/zenodo.3576630*

Assumpção, A. L., Fu, G., Singh, D. K., Lu, Z., Kuehnl, A. M., Welch, R., Ong, I. M., Wen, R. and Pan, X. (2021). A lineage-specific requirement for yy1 polycomb group protein function in early t cell development, *Development* **148**(7): dev197319.

Auclair, G., Borgel, J., Sanz, L. A., Vallet, J., Guibert, S., Dumas, M., Cavelier, P., Girardot, M., Forné, T., Feil, R. et al. (2016). Ehmt2 directs dna methylation for efficient gene silencing in mouse embryos, *Genome research* **26**(2): 192–202.

Aul, R. B. and Oko, R. J. (2001). The major subacrosomal occupant of bull spermatozoa is a novel histone h2b variant associated with the forming acrosome during spermiogenesis, *Developmental biology* **239**(2): 376–387.

Bannister, A. J., Schneider, R., Myers, F. A., Thorne, A. W., Crane-Robinson, C. and Kouzarides, T. (2005). Spatial distribution of di-and tri-methyl lysine 36 of histone h3 at active genes, *Journal of Biological Chemistry* **280**(18): 17732–17736.

Bantignies, F., Roure, V., Comet, I., Leblanc, B., Schuettengruber, B., Bonnet, J., Tixier, V., Mas, A. and Cavalli, G. (2011). Polycomb-dependent regulatory contacts between distant hox loci in drosophila, *Cell* **144**(2): 214–226.

Barbera, A. J., Chodaparambil, J. V., Kelley-Clarke, B., Joukov, V., Walter, J. C., Luger, K. and Kaye, K. M. (2006). The nucleosomal surface as a docking station for kaposi's sarcoma herpesvirus lana, *Science* **311**(5762): 856–861.

Ben-Hattar, J. and Jiricny, J. (1988). Methylation of single cpg dinucleotides within a promoter element of the herpes simplex virus tk gene reduces its transcription in vivo, *Gene* **65**(2): 219–227.

Bepler, T., Morin, A., Rapp, M., Brasch, J., Shapiro, L., Noble, A. J. and Berger, B. (2019). Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs, *NATURE METHODS* **16**(11): 1153+.

Bernstein, E., Duncan, E., Masui, O., Gil, J., Heard, E. and Allis, C. (2006). Mouse polycomb proteins bind differentially to methylated histone h3 and rna and are enriched in facultative heterochromatin, *MOLECULAR AND CELLULAR BIOLOGY* **26**(7): 2560–2569.

Bezsonova, I. (2014). Solution nmr structure of the dna-binding domain from scml2 (sex comb on midleg-like 2), *Journal of Biological Chemistry* **289**(22): 15739–15749.

Blackledge, N. P., Farcas, A. M., Kondo, T., King, H. W., McGouran, J. F., Hanssen, L. L., Ito, S., Cooper, S., Kondo, K., Koseki, Y. et al. (2014). Variant prc1 complex-dependent h2a ubiquitylation drives prc2 recruitment and polycomb domain formation, *Cell* **157**(6): 1445–1459.

Boettiger, A. N., Bintu, B., Moffitt, J. R., Wang, S., Beliveau, B. J., Fudenberg, G., Imakaev, M., Mirny, L. A., Wu, C.-t. and Zhuang, X. (2016). Super-resolution imaging reveals distinct chromatin folding for different epigenetic states, *Nature* **529**(7586): 418–422.

Bonasio, R., Lecona, E., Narendra, V., Voigt, P., Parisi, F., Kluger, Y. and Reinberg, D. (2014). Interactions with rna direct the polycomb group protein scml2 to chromatin where it represses target genes, *Elife* **3**: e02637.

Bonnet, J., Boichenko, I., Kalb, R., Le Jeune, M., Maltseva, S., Pieropan, M., Finkl, K., Fierz, B. and Müller, J. (2022). Pr-dub preserves polycomb repression by preventing excessive accumulation of h2aub1, an antagonist of chromatin compaction, *Genes & Development* **36**(19-20): 1046–1061.

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

Borgel, J., Guibert, S., Li, Y., Chiba, H., Schübeler, D., Sasaki, H., Forné, T. and Weber, M. (2010). Targets and dynamics of promoter dna methylation during early mouse development, *Nature genetics* **42**(12): 1093–1100.

Bornemann, D., Miller, E. and Simon, J. (1998). Expression and properties of wild-type and mutant forms of the drosophila sex comb on midleg (scm) repressor protein, *Genetics* **150**(2): 675–686.

Bourc'his, D., Xu, G.-L., Lin, C.-S., Bollman, B. and Bestor, T. H. (2001). Dnmt3l and the establishment of maternal genomic imprints, *Science* **294**(5551): 2536–2539.

Bracken, A. P., Dietrich, N., Pasini, D., Hansen, K. H. and Helin, K. (2006). Genome-wide mapping of polycomb target genes unravels their roles in cell fate transitions, *Genes & development* **20**(9): 1123–1136.

Buchwald, G., van der Stoop, P., Weichenrieder, O., Perrakis, A., van Lohuizen, M. and Sixma, T. K. (2006). Structure and e3-ligase activity of the ring–ring complex of polycomb proteins bmi1 and ring1b, *The EMBO Journal* **25**(11): 2465–2474.
URL: *https://www.embopress.org/doi/abs/10.1038/sj.emboj.7601144*

Chan, H. L., Beckedorff, F., Zhang, Y., Garcia-Huidobro, J., Jiang, H., Colaprico, A., Bilbao, D., Figueroa, M. E., LaCava, J., Shiekhattar, R. et al. (2018). Polycomb complexes associate with enhancers and promote oncogenic transcriptional programs in cancer through multiple mechanisms, *Nature communications* **9**(1): 3377.

Chen, K., Chen, Z., Wu, D., Zhang, L., Lin, X., Su, J., Rodriguez, B., Xi, Y., Xia, Z., Chen, X. et al. (2015). Broad h3k4me3 is associated with increased transcription elongation and enhancer activity at tumor-suppressor genes, *Nature genetics* **47**(10): 1149–1157.

Chen, V. B., Arendall, III, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. and Richardson, D. C. (2010). Molprobity: all-atom

structure validation for macromolecular crystallography, *ACTA CRYSTALLOGRAPHICA SECTION D-STRUCTURAL BIOLOGY* **66**(1): 12–21.

Chen, Z., Li, S., Subramaniam, S., Shyy, J. Y.-J. and Chien, S. (2017). Epigenetic regulation: a new frontier for biomedical engineers, *Annual review of biomedical engineering* **19**: 195–219.

Ciapponi, M., Karlukova, E., Schkölziger, S., Benda, C. and Müller, J. (2024). Structural basis of the histone ubiquitination read–write mechanism of rybp–prc1, *Nature Structural & Molecular Biology* pp. 1–5.

Ciccarone, V. C., Polayes, D. A. and Luckow, V. A. (1998). Generation of recombinant baculovirus dna in e. coli using a baculovirus shuttle vector, *Molecular diagnosis of infectious diseases* pp. 213–235.

Colombo, M., Pessey, O. and Marcia, M. (2019). Topology and enzymatic properties of a canonical polycomb repressive complex 1 isoform, *FEBS LETTERS* **593**(14): 1837–1848.

Combe, C. W., Fischer, L. and Rappsilber, J. (2015). xinet: cross-link network maps with residue resolution, *Molecular & Cellular Proteomics* **14**(4): 1137–1147.

Conway, E., Rossi, F., Fernandez-Perez, D., Ponzo, E., Ferrari, K. J., Zanotti, M., Manganaro, D., Rodighiero, S., Tamburri, S. and Pasini, D. (2021). Bap1 enhances polycomb repression by counteracting widespread h2ak119ub1 deposition and chromatin condensation, *Molecular cell* **81**(17): 3526–3541.

Cooper, S., Dienstbier, M., Hassan, R., Schermelleh, L., Sharif, J., Blackledge, N. P., De Marco, V., Elderkin, S., Koseki, H., Klose, R. et al. (2014). Targeting polycomb to pericentric heterochromatin in embryonic stem cells reveals a role for h2ak119u1 in prc2 recruitment, *Cell reports* **7**(5): 1456–1470.

Coré, N., Bel, S., Gaunt, S. J., Aurrand-Lions, M., Pearce, J., Fisher, A. and Djabali, M. (1997). Altered cellular proliferation and mesoderm patterning in polycomb-m33-deficient mice, *Development* **124**(3): 721–729.

Czermin, B., Melfi, R., McCabe, D., Seitz, V., Imhof, A. and Pirrotta, V. (2002). Drosophila enhancer of zeste/esc complexes have a histone h3 methyltransferase activity that marks chromosomal polycomb sites, *CELL* **111**(2): 185–196.

de Ayala Alonso, A. G., Gutiérrez, L., Fritsch, C., Papp, B., Beuchle, D. and Müller, J. (2007). A genetic screen identifies novel polycomb group genes in drosophila, *Genetics* **176**(4): 2099–2108.

De, I., Chittock, E. C., Grötsch, H., Miller, T. C., McCarthy, A. A. and Müller, C. W. (2019). Structural basis for the activation of the deubiquitinase calypso by the polycomb protein asx, *Structure* **27**(3): 528–536.

de Napoles, M., Mermoud, J. E., Wakao, R., Tang, Y. A., Endoh, M., Appanah, R., Nesterova, T. B., Silva, J., Otte, A. P., Vidal, M. et al. (2004). Polycomb group proteins ring1a/b link ubiquitylation of histone h2a to heritable gene silencing and x inactivation, *Developmental cell* **7**(5): 663–676.

de Potter, B., Raas, M. W., Seidl, M. F., Verrijzer, C. P. and Snel, B. (2023). Uncoupled evolution of the polycomb system and deep origin of non-canonical prc1, *Communications Biology* **6**(1): 1144.

De Santa, F., Totaro, M. G., Prosperini, E., Notarbartolo, S., Testa, G. and Natoli, G. (2007). The histone h3 lysine-27 demethylase jmjd3 links inflammation to inhibition of polycomb-mediated gene silencing, *cell* **130**(6): 1083–1094.

Denissov, S., Hofemeister, H., Marks, H., Kranz, A., Ciotta, G., Singh, S., Anastassiadis, K., Stunnenberg, H. G. and Stewart, A. F. (2014). Mll2 is required for h3k4 trimethylation on bivalent promoters in embryonic stem cells, whereas mll1 is redundant, *Development* **141**(3): 526–537.

Deshaies, R. J. and Joazeiro, C. A. (2009). Ring domain e3 ubiquitin ligases, *Annual review of biochemistry* **78**: 399–434.

Dobrinić, P., Szczurek, A. T. and Klose, R. J. (2021). Prc1 drives polycomb-mediated gene repression by controlling transcription initiation and burst frequency, *Nature structural & molecular biology* **28**(10): 811–824.

Doskočil, J. and Šorm, F. (1962). Distribution of 5-methylcytosine in pyrimidine sequences of deoxyribonucleic acids, *Biochimica et Biophysica Acta (BBA)-Specialized Section on Nucleic Acids and Related Subjects* **55**(6): 953–959.

Eagen, K. P., Aiden, E. L. and Kornberg, R. D. (2017). Polycomb-mediated chromatin loops revealed by a subkilobase-resolution chromatin interaction map, *Proceedings of the National Academy of Sciences* **114**(33): 8764–8769.

Eeftens, J. M., Kapoor, M., Michieletto, D. and Brangwynne, C. P. (2021). Polycomb condensates can promote epigenetic marks but are not required for sustained chromatin compaction, *Nature communications* **12**(1): 5888.

Emsley, P. and Cowtan, K. (2004). Coot: model-building tools for molecular graphics, *ACTA CRYSTALLOGRAPHICA SECTION D-STRUCTURAL BIOLOGY* **60**(12, 1): 2126–2132.

Eustermann, S., Schall, K., Kostrewa, D., Lakomek, K., Strauss, M., Moldt, M. and Hopfner, K.-P. (2018). Structural basis for atp-dependent chromatin remodelling by the ino80 complex, *Nature* **556**(7701): 386–390.

Fang, Q., Chen, P., Wang, M., Fang, J., Yang, N., Li, G. and Xu, R.-M. (2016). Human cytomegalovirus ie1 protein alters the higher-order chromatin structure by targeting the acidic patch of the nucleosome, *Elife* **5**: e11911.

Farcas, A. M., Blackledge, N. P., Sudbery, I., Long, H. K., McGouran, J. F., Rose, N. R., Lee, S., Sims, D., Cerase, A., Sheahan, T. W. et al. (2012). Kdm2b links the polycomb repressive complex 1 (prc1) to recognition of cpg islands, *elife* **1**: e00205.

Farrelly, L. A., Thompson, R. E., Zhao, S., Lepack, A. E., Lyu, Y., Bhanu, N. V., Zhang, B., Loh, Y.-H. E., Ramakrishnan, A., Vadodaria, K. C. et al. (2019). Histone serotonylation is a permissive modification that enhances tfiid binding to h3k4me3, *Nature* **567**(7749): 535–539.

Ferrari, K. J., Scelfo, A., Jammula, S., Cuomo, A., Barozzi, I., Stuetzer, A., Fischle, W., Bonaldi, T. and Pasini, D. (2014). Polycomb-dependent h3k27me1 and h3k27me2 regulate active transcription and enhancer fidelity, *MOLECULAR CELL* **53**(1): 49–62.

Finogenova, K., Bonnet, J., Poepsel, S., Schäfer, I. B., Finkl, K., Schmid, K., Litz, C., Strauss, M., Benda, C. and Müller, J. (2020). Structural basis for prc2 decoding of active histone methylation marks h3k36me2/3, *Elife* **9**: e61964.

Fischle, W., Wang, Y., Jacobs, S. A., Kim, Y., Allis, C. D. and Khorasanizadeh, S. (2003). Molecular basis for the discrimination of repressive methyl-lysine marks in histone h3 by polycomb and hp1 chromodomains, *Genes & development* **17**(15): 1870–1881.

Flaus, A., Martin, D. M., Barton, G. J. and Owen-Hughes, T. (2006). Identification of multiple distinct snf2 subfamilies with conserved structural motifs, *Nucleic acids research* **34**(10): 2887–2905.

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

Foglizzo, M., Middleton, A. J., Burgess, A. E., Crowther, J. M., Dobson, R. C., Murphy, J. M., Day, C. L. and Mace, P. D. (2018). A bidentate polycomb repressive-deubiquitinase complex is required for efficient activity on nucleosomes, *Nature Communications* **9**(1): 3932.

Forzati, F., Federico, A., Pallante, P., Abbate, A., Esposito, F., Malapelle, U., Sepe, R., Palma, G., Troncone, G., Scarfò, M. et al. (2012). Cbx7 is a tumor suppressor in mice and humans, *The Journal of clinical investigation* **122**(2).

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

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

Frey, F., Sheahan, T., Finkl, K., Stoehr, G., Mann, M., Benda, C. and Müller, J. (2016). Molecular basis of prc1 targeting to polycomb response elements by phorc, *Genes & Development* **30**(9): 1116–1127.

Fursova, N. A., Turberfield, A. H., Blackledge, N. P., Findlater, E. L., Lastuvkova, A., Huseyin, M. K., Dobrinić, P. and Klose, R. J. (2021). Bap1 constrains pervasive h2ak119ub1 to control the transcriptional potential of the genome, *Genes & development* **35**(9-10): 749–770.

Fyodorov, D. V., Zhou, B.-R., Skoultchi, A. I. and Bai, Y. (2018). Emerging roles of linker histones in regulating chromatin structure and function, *Nature reviews Molecular cell biology* **19**(3): 192–206.

Gahan, J. M., Rentzsch, F. and Schnitzler, C. E. (2020). The genetic basis for prc1 complex diversity emerged early in animal evolution, *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA* **117**(37): 22880–22889.

Gambetta, M. C. and Müller, J. (2014). O-glcnacylation prevents aggregation of the polycomb group repressor polyhomeotic, *Developmental cell* **31**(5): 629–639.

Gao, Z., Zhang, J., Bonasio, R., Strino, F., Sawai, A., Parisi, F., Kluger, Y. and Reinberg, D. (2012). Pcgf homologs, cbx proteins, and rybp define functionally distinct prc1 family complexes, *MOLECULAR CELL* **45**(3): 344–356.

Gardiner-Garden, M. and Frommer, M. (1987). Cpg islands in vertebrate genomes, *Journal of molecular biology* **196**(2): 261–282.

Ge, W., Yu, C., Li, J., Yu, Z., Li, X., Zhang, Y., Liu, C.-P., Li, Y., Tian, C., Zhang, X. et al. (2023). Basis of the h2ak119 specificity of the polycomb repressive deubiquitinase, *Nature* **616**(7955): 176–182.

Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison III, C. A. and Smith, H. O. (2009). Enzymatic assembly of dna molecules up to several hundred kilobases, *Nature methods* **6**(5): 343–345.

Gilbert, S. F. (2010). *Developmental biology.*, sinauer associates, Inc.

Goddard, T. D., Huang, C. C., Meng, E. C., Pettersen, E. F., Couch, G. S., Morris, J. H. and Ferrin, T. E. (2018). Ucsf chimerax: Meeting modern challenges in visualization and analysis, *PROTEIN SCIENCE* **27**(1, SI): 14–25.

Gong, Z., Ye, S.-X., Nie, Z.-F. and Tang, C. (2020). The conformational preference of chemical cross-linkers determines the cross-linking probability of reactive protein residues, *The Journal of Physical Chemistry B* **124**(22): 4446–4453. PMID: 32369371.
URL: *https://doi.org/10.1021/acs.jpcb.0c02522*

Graham, M., Combe, C., Kolbowski, L. and Rappsilber, J. (2019). xiview: A common platform for the downstream analysis of crosslinking mass spectrometry data, *BioRxiv* p. 561829.

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

Gray, F., Cho, H. J., Shukla, S., He, S., Harris, A., Boytsov, B., Jaremko, L., Jaremko, M., Demeler, B., Lawlor, E. R., Grembecka, J. and Cierpicki, T. (2016). Bmi1 regulates prc1 architecture and activity through homo- and hetero-oligomerization, *NATURE COMMUNICATIONS* **7**.

Grimm, C., de Ayala Alonso, A. G., Rybin, V., Steuerwald, U., Ly-Hartig, N., Fischle, W., Müller, J. and Müller, C. W. (2007). Structural and functional analyses of methyl-lysine binding by the malignant brain tumour repeat protein sex comb on midleg, *EMBO reports* **8**(11): 1031–1037.

Grimm, C., Matos, R., Ly-Hartig, N., Steuerwald, U., Lindner, D., Rybin, V., Müller, J. and Müller, C. W. (2009). Molecular recognition of histone lysine methylation by the polycomb group repressor dsfmbt, *The EMBO journal* **28**(13): 1965–1977.

Gutiérrez, L., Oktaba, K., Scheuermann, J. C., Gambetta, M. C., Ly-Hartig, N. and Müller, J. (2012). The role of the histone h2a ubiquitinase sce in polycomb repression, *Development* **139**(1): 117–127.

Han, B.-G., Watson, Z., Kang, H., Pulk, A., Downing, K. H., Cate, J. and Glaeser, R. M. (2016). Long shelf-life streptavidin support-films suitable for electron microscopy of biological macromolecules, *Journal of structural biology* **195**(2): 238–244.

Hannah-Alava, A. (1958). Developmental genetics of the posterior legs in drosophila melanogaster, *Genetics* **43**(5): 878.

Hartley, P. D. and Madhani, H. D. (2009). Mechanisms that specify promoter nucleosome location and identity, *Cell* **137**(3): 445–458.

Hasegawa, K., Sin, H.-S., Maezawa, S., Broering, T. J., Kartashov, A. V., Alavattam, K. G., Ichijima, Y., Zhang, F., Bacon, W. C., Greis, K. D. et al. (2015). Scml2 establishes the male germline epigenome through regulation of histone h2a ubiquitination, *Developmental cell* **32**(5): 574–588.

Hellman, L. M. and Fried, M. G. (2007). Electrophoretic mobility shift assay (emsa) for detecting protein-nucleic acid interactions, *NATURE PROTOCOLS* **2**(8): 1849–1861.

Hewish, D. R. and Burgoyne, L. A. (1973). Chromatin sub-structure. the digestion of chromatin dna at regularly spaced sites by a nuclear deoxyribonuclease, *Biochemical and biophysical research communications* **52**(2): 504–510.

Hu, D., Garruss, A. S., Gao, X., Morgan, M. A., Cook, M., Smith, E. R. and Shilatifard, A. (2013). The mll2 branch of the compass family regulates bivalent promoters in mouse embryonic stem cells, *Nature structural & molecular biology* **20**(9): 1093–1097.

Hu, G., Katuwawala, A., Wang, K., Wu, Z., Ghadermarzi, S., Gao, J. and Kurgan, L. (2021). fldpnn: Accurate intrinsic disorder prediction with putative propensities of disorder functions, *Nature communications* **12**(1): 4438.

Huth, J., Bewley, C., Jackson, B., Hinnebusch, A., Clore, G. and Gronenborn, A. (1997). Design of an expression system for detecting folded protein domains and mapping macromolecular interactions by nmr, *PROTEIN SCIENCE* **6**(11): 2359–2364.

Iguchi-Ariga, S. and Schaffner, W. (1989). Cpg methylation of the camp-responsive enhancer/promoter sequence tgacgtca abolishes specific factor binding as well as transcriptional activation., *Genes & development* **3**(5): 612–619.

Illingworth, R. S., Moffat, M., Mann, A. R., Hunter, C. J., Pradeepa, M. M., Adams, I. R. and Bickmore, W. A. (2015). The e3 ubiquitin ligase activity of ring1b is not essential for early mouse development, *Genes & development* **29**(18): 1897–1902.

Ingham, P. W. (1983). Differential expression of bithorax complex genes in the absence of the extra sex combs and trithorax genes, *Nature* **306**(5943): 591–593.

Iouzalen, N., Moreau, J. and Méchali, M. (1996). H2a. zl, a new variant histone expressed during xenopus early development exhibits several distinct features from the core histone h2a, *Nucleic acids research* **24**(20): 3947–3952.

Isono, K., Endo, T. A., Ku, M., Yamada, D., Suzuki, R., Sharif, J., Ishikura, T., Toyoda, T., Bernstein, B. E. and Koseki, H. (2013). Sam domain polymerization links subnuclear clustering of prc1 to gene silencing, *Developmental cell* **26**(6): 565–577.

Isono, K.-i., Fujimura, Y.-i., Shinga, J., Yamaki, M., Jiyang, O., Takihara, Y., Murahashi, Y., Takada, Y., Mizutani-Koseki, Y., Koseki, H. et al. (2005). Mammalian polyhomeotic homologues phc2 and phc1 act in synergy to mediate polycomb repression of hox genes, *Molecular and cellular biology* .

Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A., He, C. and Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine, *Science* **333**(6047): 1300–1303.

Jacobs, J. J., Kieboom, K., Marino, S., DePinho, R. A. and Van Lohuizen, M. (1999). The oncogene and polycomb-group gene bmi-1 regulates cell proliferation and senescence through the ink4a locus, *Nature* **397**(6715): 164–168.

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

Jermann, P., Hoerner, L., Burger, L. and Schübeler, D. (2014). Short sequences can efficiently recruit histone h3 lysine 27 trimethylation in the absence of enhancer activity and dna methylation, *Proceedings of the National Academy of Sciences* **111**(33): E3415–E3421.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A. et al. (2021). Highly accurate protein structure prediction with alphafold, *Nature* **596**(7873): 583–589.

Kabsch, W. (2010). Integration, scaling, space-group assignment and post-refinement, *Acta Crystallographica Section D: Biological Crystallography* **66**(2): 133–144.

Kadoch, C., Williams, R. T., Calarco, J. P., Miller, E. L., Weber, C. M., Braun, S. M., Pulice, J. L., Chory, E. J. and Crabtree, G. R. (2017). Dynamics of baf–polycomb complex opposition on heterochromatin in normal and oncogenic states, *Nature genetics* **49**(2): 213–222.

Kahn, T. G., Stenberg, P., Pirrotta, V. and Schwartz, Y. B. (2014). Combinatorial interactions are required for the efficient recruitment of pho repressive complex (phorc) to polycomb response elements, *PLoS genetics* **10**(7): e1004495.

Kalb, R., Latwiel, S., Baymaz, H. I., Jansen, P. W. T. C., Mueller, C. W., Vermeulen, M. and Mueller, J. (2014). Histone h2a monoubiquitination promotes histone h3 methylation in

polycomb repression, *NATURE STRUCTURAL & MOLECULAR BIOLOGY* **21**(6): 569–571.

Kang, H., McElroy, K. A., Jung, Y. L., Alekseyenko, A. A., Zee, B. M., Park, P. J. and Kuroda, M. I. (2015). Sex comb on midleg (scm) is a functional link between pcg-repressive complexes in drosophila, *Genes & development* **29**(11): 1136–1150.

Kaustov, L., Ouyang, H., Amaya, M., Lemak, A., Nady, N., Duan, S., Wasney, G. A., Li, Z., Vedadi, M., Schapira, M. et al. (2011). Recognition and specificity determinants of the human cbx chromodomains, *Journal of Biological Chemistry* **286**(1): 521–529.

Kim, C. A., Gingery, M., Pilpa, R. M. and Bowie, J. U. (2002). The sam domain of polyhomeotic forms a helical polymer, *Nature structural biology* **9**(6): 453–457.

Kim, C. A., Phillips, M. L., Kim, W., Gingery, M., Tran, H. H., Robinson, M. A., Faham, S. and Bowie, J. U. (2001). Polymerization of the sam domain of tel in leukemogenesis and transcriptional repression, *The EMBO journal* **20**(15): 4173–4182.

Kim, C. A., Sawaya, M. R., Cascio, D., Kim, W. and Bowie, J. U. (2005). Structural organization of a sex-comb-on-midleg/polyhomeotic copolymer, *Journal of Biological Chemistry* **280**(30): 27769–27775.

Kim, J. J. and Kingston, R. E. (2022). Context-specific polycomb mechanisms in development, *Nature Reviews Genetics* **23**(11): 680–695.

Kim, J. J., Steinson, E. R., Lau, M. S., de Rooij, D. G., Page, D. C. and Kingston, R. E. (2023). Cell type-specific role of cbx2 and its disordered region in spermatogenesis, *Genes & Development* **37**(13-14): 640–660.

King, I. F., Emmons, R. B., Francis, N. J., Wild, B., Müller, J., Kingston, R. E. and Wu, C.-t. (2005). Analysis of a polycomb group protein defines regions that link repressive activity on nucleosomal templates to in vivo function, *Molecular and cellular biology* **25**(15): 6578–6591.

Kinman, L. F., Powell, B. M., Zhong, E. D., Berger, B. and Davis, J. H. (2023). Uncovering structural ensembles from single-particle cryo-em data using cryodrgn, *Nature protocols* **18**(2): 319–339.

Kizer, K. O., Phatnani, H. P., Shibata, Y., Hall, H., Greenleaf, A. L. and Strahl, B. D. (2005). A novel domain in set2 mediates rna polymerase ii interaction and couples histone h3 k36 methylation with transcript elongation, *Molecular and cellular biology* **25**(8): 3305–3316.

Kloet, S. L., Makowski, M. M., Baymaz, H. I., Van Voorthuijsen, L., Karemaker, I. D., Santanach, A., Jansen, P. W., Di Croce, L. and Vermeulen, M. (2016). The dynamic interactome and genomic targets of polycomb complexes during stem-cell differentiation, *Nature structural & molecular biology* **23**(7): 682–690.

Klykov, O., Steigenberger, B., Pektaş, S., Fasci, D., Heck, A. J. and Scheltema, R. A. (2018). Efficient and robust proteome-wide approaches for cross-linking mass spectrometry, *Nature protocols* **13**(12): 2964–2990.

Klymenko, T. and Müller, J. (2004). The histone methyltransferases trithorax and ash1 prevent transcriptional silencing by polycomb group proteins, *EMBO reports* **5**(4): 373–377.

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

Kornberg, R. D. (1974). Chromatin structure: A repeating unit of histones and dna: Chromatin structure is based on a repeating unit of eight histone molecules and about 200 dna base pairs., *Science* **184**(4139): 868–871.

Kornberg, R. D. and Lorch, Y. (1999). Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome, *Cell* **98**(3): 285–294.

Kriaucionis, S. and Heintz, N. (2009). The nuclear dna base 5-hydroxymethylcytosine is present in purkinje neurons and the brain, *Science* **324**(5929): 929–930.

Krietenstein, N., Wal, M., Watanabe, S., Park, B., Peterson, C. L., Pugh, B. F. and Korber, P. (2016). Genomic nucleosome organization reconstituted with pure proteins, *Cell* **167**(3): 709–721.

Kulak, N. A., Geyer, P. E. and Mann, M. (2017). Loss-less nano-fractionator for high sensitivity, high coverage proteomics, *Molecular & Cellular Proteomics* **16**(4): 694–705.

Lagerwaard, I. M., Albanese, P., Jankevics, A. and Scheltema, R. A. (2022). Xlink mapping and analysis (xmas)-smooth integrative modeling in chimerax, *bioRxiv* pp. 2022–04.

Lanman, J., Lam, T. T., Barnes, S., Sakalian, M., Emmett, M. R., Marshall, A. G. and Prevelige Jr, P. E. (2003). Identification of novel interactions in hiv-1 capsid protein assembly by high-resolution mass spectrometry, *Journal of molecular biology* **325**(4): 759–772.

Lau, M. S., Schwartz, M. G., Kundu, S., Savol, A. J., Wang, P. I., Marr, S. K., Grau, D. J., Schorderet, P., Sadreyev, R. I., Tabin, C. J. et al. (2017). Mutation of a nucleosome compaction region disrupts polycomb-mediated axial patterning, *Science* **355**(6329): 1081–1084.

Lechtenberg, B. C., Allen, M. D., Rutherford, T. J., Freund, S. M. and Bycroft, M. (2009). Solution structure of the fcs zinc finger domain of the human polycomb group protein l (3) mbt-like 2, *Protein Science* **18**(3): 657–661.

Lecona, E., Narendra, V. and Reinberg, D. (2015). Usp7 cooperates with scml2 to regulate the activity of prc1, *Molecular and cellular biology* **35**(7): 1157–1168.

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

Leeb, M. and Wutz, A. (2007). Ring1b is crucial for the regulation of developmental control genes and prc1 proteins but not x inactivation in embryonic cells, *The Journal of cell biology* **178**(2): 219–229.

Lehmann, L., Ferrari, R., Vashisht, A. A., Wohlschlegel, J. A., Kurdistani, S. K. and Carey, M. (2012). Polycomb repressive complex 1 (prc1) disassembles rna polymerase ii preinitiation complexes*, *Journal of Biological Chemistry* **287**(43): 35784–35794.

Lepack, A. E., Werner, C. T., Stewart, A. F., Fulton, S. L., Zhong, P., Farrelly, L. A., Smith, A. C., Ramakrishnan, A., Lyu, Y., Bastle, R. M. et al. (2020). Dopaminylation of histone h3 in ventral tegmental area regulates cocaine seeking, *Science* **368**(6487): 197–201.

Levine, S. S., Weiss, A., Erdjument-Bromage, H., Shao, Z., Tempst, P. and Kingston, R. E. (2002). The core of the polycomb repressive complex is compositionally and functionally conserved in flies and humans, *Molecular and cellular biology* **22**(17): 6070–6078.

Lewis, E. B. (1978). A gene complex controlling segmentation in drosophila, *Nature* **276**(5688): 565–570.

Lewis, P. and Mislove, R. (1947). New mutants report, *Drosoph. Inf. Serv* **21**: 69.

Li, E., Beard, C. and Jaenisch, R. (1993). Role for dna methylation in genomic imprinting, *Nature* **366**(6453): 362–365.

Li, E., Bestor, T. H. and Jaenisch, R. (1992). Targeted mutation of the dna methyltransferase gene results in embryonic lethality, *Cell* **69**(6): 915–926.

Li, E. and Zhang, Y. (2014). Dna methylation in mammals, *Cold Spring Harbor perspectives in biology* **6**(5): a019133.

Li, W., Li, Y., Tan, Y., Ma, K. and Cui, J. (2010). Bmi-1 is critical for the proliferation and invasiveness of gastric carcinoma cells, *Journal of gastroenterology and hepatology* **25**(3): 568–575.

Li, Z., Cao, R., Wang, M., Myers, M. P., Zhang, Y. and Xu, R.-M. (2006). Structure of a bmi-1-ring1b polycomb group ubiquitin ligase complex, *Journal of Biological Chemistry* **281**(29): 20643–20649.

Liebschner, D., Afonine, P. V., Baker, M. L., Bunkoczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L.-W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., Stockwell, D. H., Terwilliger, T. C., Urzhumtsev, A. G., Videau, L. L., Williams, C. J. and Adams, P. D. (2019). Macromolecular structure determination using x-rays, neutrons and electrons: recent developments in phenix, *ACTA CRYSTALLOGRAPHICA SECTION D-STRUCTURAL BIOLOGY* **75**(10): 861–877.

Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., Nery, J. R., Lee, L., Ye, Z., Ngo, Q.-M. et al. (2009). Human dna methylomes at base resolution show widespread epigenomic differences, *nature* **462**(7271): 315–322.

Liu, B., Liu, Y.-F., Du, Y.-R., Mardaryev, A. N., Yang, W., Chen, H., Xu, Z.-M., Xu, C.-Q., Zhang, X.-R., Botchkarev, V. A. et al. (2013). Cbx4 regulates the proliferation of thymic epithelial cells and thymus function, *Development* **140**(4): 780–788.

Long, M., Sun, X., Shi, W., Yanru, A., Leung, S. T., Ding, D., Cheema, M. S., MacPherson, N., Nelson, C. J., Ausio, J. et al. (2019). A novel histone h4 variant h4g regulates rdna transcription in breast cancer, *Nucleic Acids Research* **47**(16): 8399–8409.

Lorente, M. d. M., Marcos-Gutiérrez, C., Pérez, C., Schoorlemmer, J., Ramírezp, A., Magin, T. and Vidal, M. (2000). Loss-and gain-of-function mutations show a polycomb group function for ring1a in mice, *Development* **127**(23): 5093–5100.

Lowary, P. and Widom, J. (1998). New dna sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning, *JOURNAL OF MOLECULAR BIOLOGY* **276**(1): 19–42.

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

Luger, K., Rechsteiner, T. J. and Richmond, T. J. (1999). *Expression and Purification of Recombinant Histones and Nucleosome Reconstitution*, Humana Press, Totowa, NJ, pp. 1–16.
URL: *https://doi.org/10.1385/1-59259-681-9:1*

Luo, M., Zhou, J., Leu, N. A., Abreu, C. M., Wang, J., Anguera, M. C., de Rooij, D. G., Jasin, M. and Wang, P. J. (2015). Polycomb protein scml2 associates with usp7 and counteracts histone h2a ubiquitination in the xy chromatin during male meiosis, *PLoS genetics* **11**(1): e1004954.

Maezawa, S., Hasegawa, K., Alavattam, K. G., Funakoshi, M., Sato, T., Barski, A. and Namekawa, S. H. (2018). Scml2 promotes heterochromatin organization in late spermatogenesis, *Journal of Cell Science* **131**(17): jcs217125.

Makde, R. D., England, J. R., Yennawar, H. P. and Tan, S. (2010). Structure of rcc1 chromatin factor bound to the nucleosome core particle, *Nature* **467**(7315): 562–566.

Malik, H. and Henikoff, S. (2003). Phylogenomics of the nucleosome, *NATURE STRUCTURAL BIOLOGY* **10**(11): 882–891.

Markert, J. W., Vos, S. M. and Farnung, L. (2023). Structure of the complete saccharomyces cerevisiae rpd3s-nucleosome complex, *Nature Communications* **14**(1): 8128.

Mastronarde, D. N. (2005). Automated electron microscope tomography using robust prediction of specimen movements, *Journal of structural biology* **152**(1): 36–51.

Maurer-Stroh, S., Dickens, N. J., Hughes-Davies, L., Kouzarides, T., Eisenhaber, F. and Ponting, C. P. (2003). The tudor domain 'royal family': tudor, plant agenet, chromo, pwwp and mbt domains, *Trends in biochemical sciences* **28**(2): 69–74.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. and Read, R. J. (2007). Phaser crystallographic software, *Journal of applied crystallography* **40**(4): 658–674.

McGinty, R. K., Henrici, R. C. and Tan, S. (2014). Crystal structure of the prc1 ubiquitylation module bound to the nucleosome, *NATURE* **514**(7524): 591+.

McKay, D. J., Klusza, S., Penke, T. J. R., Meers, M. P., Curry, K. P., McDaniel, S. L., Malek, P. Y., Cooper, S. W., Tatomer, D. C., Lieb, J. D., Strahl, B. D., Duronio, R. J. and Matera, A. G. (2015). Interrogating the function of metazoan histones using engineered gene clusters, *DEVELOPMENTAL CELL* **32**(3): 373–386.

McKnight, S. and Miller, O. (1976). Ultrastructural patterns of rna-synthesis during early embryogenesis of drosophila-melanogaster, *CELL* **8**(2): 305–319.

Meehan, R. R., Lewis, J. D., McKay, S., Kleiner, E. L. and Bird, A. P. (1989). Identification of a mammalian protein that binds specifically to dna containing methylated cpgs, *Cell* **58**(3): 499–507.

Meneghini, M. D., Wu, M. and Madhani, H. D. (2003). Conserved histone variant h2a. z protects euchromatin from the ectopic spread of silent heterochromatin, *Cell* **112**(5): 725–736.

Min, J., Zhang, Y. and Xu, R.-M. (2003). Structural basis for specific binding of polycomb chromodomain to histone h3 methylated at lys 27, *Genes & development* **17**(15): 1823–1828.

Moazed, D. and O'Farrell, P. H. (1992). Maintenance of the engrailed expression pattern by polycomb group genes in drosophila, *Development* **116**(3): 805–810.

Mohandas, T., Sparkes, R. and Shapiro, L. (1981). Reactivation of an inactive human x chromosome: evidence for x inactivation by dna methylation, *Science* **211**(4480): 393–396.

Morey, L., Pascual, G., Cozzuto, L., Roma, G., Wutz, A., Benitah, S. A. and Di Croce, L. (2012). Nonoverlapping functions of the polycomb group cbx family of proteins in embryonic stem cells, *Cell stem cell* **10**(1): 47–62.

Morgan, M. T., Haj-Yahya, M., Ringel, A. E., Bandi, P., Brik, A. and Wolberger, C. (2016). Structural basis for histone h2b deubiquitination by the saga dub module, *Science* **351**(6274): 725–728.

Müller, J. and Bienz, M. (1991). Long range repression conferring boundaries of ultrabithorax expression in the drosophila embryo., *The EMBO journal* **10**(11): 3147–3155.

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

Nakane, T., Kimanius, D., Lindahl, E. and Scheres, S. H. (2018). Characterisation of molecular motions in cryo-em single-particle data by multi-body refinement in relion, *eLife* **7**: e36861.
URL: *https://doi.org/10.7554/eLife.36861*

Nan, X., Ng, H.-H., Johnson, C. A., Laherty, C. D., Turner, B. M., Eisenman, R. N. and Bird, A. (1998). Transcriptional repression by the methyl-cpg-binding protein mecp2 involves a histone deacetylase complex, *Nature* **393**(6683): 386–389.

Nanyes, D. R., Junco, S. E., Taylor, A. B., Robinson, A. K., Patterson, N. L., Shivarajpur, A., Halloran, J., Hale, S. M., Kaur, Y., Hart, P. J. et al. (2014). Multiple polymer architectures of human polyhomeotic homolog 3 sterile alpha motif, *Proteins: Structure, Function, and Bioinformatics* **82**(10): 2823–2830.

Ng, H.-H., Zhang, Y., Hendrich, B., Johnson, C. A., Turner, B. M., Erdjument-Bromage, H., Tempst, P., Reinberg, D. and Bird, A. (1999). Mbd2 is a transcriptional repressor belonging to the mecp1 histone deacetylase complex, *Nature genetics* **23**(1): 58–61.

Okano, M., Bell, D. W., Haber, D. A. and Li, E. (1999). Dna methyltransferases dnmt3a and dnmt3b are essential for de novo methylation and mammalian development, *Cell* **99**(3): 247–257.

Okano, M., Xie, S. and Li, E. (1998). Cloning and characterization of a family of novel mammalian dna (cytosine-5) methyltransferases, *Nature genetics* **19**(3): 219–220.

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

Olins, A. and Olins, D. (1973). Spheroid chromatin units (nu bodies), *JOURNAL OF CELL BIOLOGY* **59**(2): A252.

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

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

Paro, R. and Hogness, D. S. (1991). The polycomb protein shares a homologous domain with a heterochromatin-associated protein of drosophila., *Proceedings of the National Academy of Sciences* **88**(1): 263–267.

Parreno, V., Martinez, A.-M. and Cavalli, G. (2022). Mechanisms of polycomb group protein function in cancer, *Cell Research* **32**(3): 231–253.

Pengelly, A. R., Copur, O., Jaeckle, H., Herzig, A. and Mueller, J. (2013). A histone mutant reproduces the phenotype caused by loss of histone-modifying factor polycomb, *SCIENCE* **339**(6120): 698–699.

Pengelly, A. R., Kalb, R., Finkl, K. and Müller, J. (2015). Transcriptional repression by prc1 in the absence of h2a monoubiquitylation, *Genes & development* **29**(14): 1487–1492.

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

Pickart, C. M. (2001). Mechanisms underlying ubiquitination, *Annual review of biochemistry* **70**(1): 503–533.

Pirity, M. K., Locker, J. and Schreiber-Agus, N. (2005). Rybp/dedaf is required for early postimplantation and for central nervous system development, *Molecular and cellular biology* **25**(16): 7193–7202.

Pirrotta, V. (2017). *Polycomb group proteins*, Academic Press.

Piunti, A. and Shilatifard, A. (2016). Epigenetic balance of gene expression by polycomb and compass families, *Science* **352**(6290): aad9780.

Piunti, A. and Shilatifard, A. (2021). The roles of polycomb repressive complexes in mammalian development and cancer, *Nature Reviews Molecular Cell Biology* **22**(5): 326–345.

Plys, A. J., Davis, C. P., Kim, J., Rizki, G., Keenen, M. M., Marr, S. K. and Kingston, R. E. (2019). Phase separation of polycomb-repressive complex 1 is governed by a charged disordered region of cbx2, *Genes & development* **33**(13-14): 799–813.

Poli, J., Gerhold, C.-B., Tosi, A., Hustedt, N., Seeber, A., Sack, R., Herzog, F., Pasero, P., Shimada, K., Hopfner, K.-P. et al. (2016). Mec1, ino80, and the paf1 complex cooperate to limit transcription replication conflicts through rnapii removal during replication stress, *Genes & development* **30**(3): 337–354.

Ponting, C. P. (1995). Sam: a novel motif in yeast sterile and drosophila polyhomeotic proteins, *Protein science* **4**(9): 1928–1930.

Punjani, A. and Fleet, D. J. (2021). 3d variability analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-em, *JOURNAL OF STRUCTURAL BIOLOGY* **213**(2).

Punjani, A. and Fleet, D. J. (2023). 3dflex: determining structure and motion of flexible proteins from cryo-em, *Nature Methods* **20**(6): 860–870.

Punjani, A., Rubinstein, J. L., Fleet, D. J. and Brubaker, M. A. (2017). cryosparc: algorithms for rapid unsupervised cryo-em structure determination, *NATURE METHODS* **14**(3): 290+.

Ren, C., Morohashi, K., Plotnikov, A. N., Jakoncic, J., Smith, S. G., Li, J., Zeng, L., Rodriguez, Y., Stojanoff, V., Walsh, M. et al. (2015). Small-molecule modulators of methyllysine binding for the cbx7 chromodomain, *Chemistry & biology* **22**(2): 161–168.

Ren, X., Vincenz, C. and Kerppola, T. K. (2008). Changes in the distributions and dynamics of polycomb repressive complexes during embryonic stem cell differentiation, *Molecular and Cellular Biology* **28**(9): 2884–2895.

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

Rogakou, E. P., Pilch, D. R., Orr, A. H., Ivanova, V. S. and Bonner, W. M. (1998). Dna double-stranded breaks induce histone h2ax phosphorylation on serine 139, *Journal of biological chemistry* **273**(10): 5858–5868.

Rohou, A. and Grigorieff, N. (2015). Ctffind4: Fast and accurate defocus estimation from electron micrographs, *JOURNAL OF STRUCTURAL BIOLOGY* **192**(2): 216–221.

Rose, N. R., King, H. W., Blackledge, N. P., Fursova, N. A., Ember, K. J., Fischer, R., Kessler, B. M. and Klose, R. J. (2016). Rybp stimulates prc1 to shape chromatin-based communication between polycomb repressive complexes, *Elife* **5**: e18591.

Rosenthal, P. and Henderson, R. (2003). Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy, *JOURNAL OF MOLECULAR BIOLOGY* **333**(4): 721–745.

Sanchez-Garcia, R., Gomez-Blanco, J., Cuervo, A., Maria Carazo, J., Sorzano, C. O. S. and Vargas, J. (2021). Deepemhancer: a deep learning solution for cryo-em volume post-processing, *COMMUNICATIONS BIOLOGY* **4**(1).

Sanchez-Pulido, L., Devos, D., Sung, Z. R. and Calonje, M. (2008). Rawul: a new ubiquitin-like domain in prc1 ring finger proteins that unveils putative plant and worm prc1 orthologs, *BMC genomics* **9**: 1–11.

Sathyamurthy, A., Allen, M. D., Murzin, A. G. and Bycroft, M. (2003). Crystal structure of the malignant brain tumor (mbt) repeats in sex comb on midleg-like 2 (scml2), *Journal of Biological Chemistry* **278**(47): 46968–46973.

Saurin, A. J., Shiels, C., Williamson, J., Satijn, D. P., Otte, A. P., Sheer, D. and Freemont, P. S. (1998). The human polycomb group complex associates with pericentromeric heterochromatin to form a novel nuclear domain, *The Journal of cell biology* **142**(4): 887–898.

Scheres, S. H. (2012). Relion: Implementation of a bayesian approach to cryo-em structure determination, *Journal of Structural Biology* **180**(3): 519–530.
URL: *https://www.sciencedirect.com/science/article/pii/S1047847712002481*

Scheuermann, J. (2011). *Identification of PR-DUB: A Polycomb Group Complex with Histone H2A Deubiquitinase Activity*, PhD thesis.

Scheuermann, J. C., de Ayala Alonso, A. G., Oktaba, K., Ly-Hartig, N., McGinty, R. K., Fraterman, S., Wilm, M., Muir, T. W. and Müller, J. (2010). Histone h2a deubiquitinase activity of the polycomb repressive complex pr-dub, *Nature* **465**(7295): 243–247.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P. and Cardona, A. (2012). Fiji: an open-source platform for biological-image analysis, *NATURE METHODS* **9**(7): 676–682.

Schmähling, S., Meiler, A., Lee, Y., Mohammed, A., Finkl, K., Tauscher, K., Israel, L., Wirth, M., Philippou-Massier, J., Blum, H. et al. (2018). Regulation and function of h3k36 dimethylation by the trithorax-group protein complex amc, *Development* **145**(7): dev163808.

Schmitges, F. W., Prusty, A. B., Faty, M., Stützer, A., Lingaraju, G. M., Aiwazian, J., Sack, R., Hess, D., Li, L., Zhou, S. et al. (2011). Histone methylation by prc2 is inhibited by active chromatin marks, *Molecular cell* **42**(3): 330–341.

Schoorlemmer, J., Marcos-Gutierrez, C., Were, F., Martínez, R., García, E., Satijn, D. P., Otte, A. P. and Vidal, M. (1997). Ring1a is a transcriptional repressor that interacts with the polycomb-m33 protein and is expressed at rhombomere boundaries in the mouse hindbrain, *The EMBO journal* **16**(19): 5930–5942.

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

Shao, Z., Raible, F., Mollaaghababa, R., Guyon, J. R., Wu, C.-t., Bender, W. and Kingston, R. E. (1999). Stabilization of chromatin structure by prc1, a polycomb complex, *Cell* **98**(1): 37–46.

Sharif, J., Muto, M., Takebayashi, S.-i., Suetake, I., Iwamatsu, A., Endo, T. A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K. et al. (2007). The sra protein np95 mediates epigenetic inheritance by recruiting dnmt1 to methylated dna, *Nature* **450**(7171): 908–912.

Siggens, L., Cordeddu, L., Rönnerblad, M., Lennartsson, A. and Ekwall, K. (2015). Transcription-coupled recruitment of human chd1 and chd2 influences chromatin accessibility and histone h3 and h3. 3 occupancy at active chromatin regions, *Epigenetics & chromatin* **8**(1): 1–14.

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

Simpson, R. T. (1978). Structure of the chromatosome, a chromatin particle containing 160 base pairs of dna and all the histones, *Biochemistry* **17**(25): 5524–5531.

Sims, R. J., Chen, C.-F., Santos-Rosa, H., Kouzarides, T., Patel, S. S. and Reinberg, D. (2005). Human but not yeast chd1 binds directly and selectively to histone h3 methylated at lysine 4 via its tandem chromodomains, *Journal of Biological Chemistry* **280**(51): 41789–41792.

Smith, D. and Johnson, K. (1988). Single-step purification of polypeptides expressed in escherichia-coli as fusions with glutathione s-transferase, *GENE* **67**(1): 31–40.

Smouse, D., Goodman, C., Mahowald, A. and Perrimon, N. (1988). polyhomeotic: a gene required for the embryonic development of axon pathways in the central nervous system of drosophila, *Genes Dev* **2**(7): 830–842.

Snowden, A. W., Gregory, P. D., Case, C. C. and Pabo, C. O. (2002). Gene-specific targeting of h3k9 methylation is sufficient for initiating repression in vivo, *Current biology* **12**(24): 2159–2166.

Stahl, K., Brock, O. and Rappsilber, J. (2023). Modelling protein complexes with crosslinking mass spectrometry and deep learning, *bioRxiv* pp. 2023–06.

Stanton, B. Z., Hodges, C., Calarco, J. P., Braun, S. M., Ku, W. L., Kadoch, C., Zhao, K. and Crabtree, G. R. (2017). Smarca4 atpase mutations disrupt direct eviction of prc1 from chromatin, *Nature genetics* **49**(2): 282–288.

Stark, H. (2010). Grafix: stabilization of fragile macromolecular complexes for single particle cryo-em, **481**: 109–126.

Stock, J. K., Giadrossi, S., Casanova, M., Brookes, E., Vidal, M., Koseki, H., Brockdorff, N., Fisher, A. G. and Pombo, A. (2007). Ring1-mediated ubiquitination of h2a restrains poised rna polymerase ii at bivalent genes in mouse es cells, *Nature cell biology* **9**(12): 1428–1435.

Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L. M., Liu, D. R., Aravind, L. et al. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian dna by mll partner tet1, *Science* **324**(5929): 930–935.

Takihara, Y., Tomotsune, D., Shirai, M., Katoh-Fukui, Y., Nishii, K., Motaleb, M. A., Nomura, M., Tsuchiya, R., Fujita, Y., Shibata, Y. et al. (1997). Targeted disruption of the mouse homologue of the drosophila polyhomeotic gene leads to altered anteroposterior patterning and neural crest defects, *Development* **124**(19): 3673–3682.

Tamburri, S., Lavarone, E., Fernández-Pérez, D., Conway, E., Zanotti, M., Manganaro, D. and Pasini, D. (2020). Histone h2ak119 mono-ubiquitination is essential for polycomb-mediated transcriptional repression, *Molecular cell* **77**(4): 840–856.

Tatavosian, R., Kent, S., Brown, K., Yao, T., Duc, H. N., Huynh, T. N., Zhen, C. Y., Ma, B., Wang, H. and Ren, X. (2019). Nuclear condensates of the polycomb protein chromobox 2 (cbx2) assemble through phase separation, *Journal of Biological Chemistry* **294**(5): 1451–1463.

Tavares, L., Dimitrova, E., Oxley, D., Webster, J., Poot, R., Demmers, J., Bezstarosti, K., Taylor, S., Ura, H., Koide, H. et al. (2012). Rybp-prc1 complexes mediate h2a ubiquitylation at polycomb target sites independently of prc2 and h3k27me3, *Cell* **148**(4): 664–678.

Taverna, S. D., Li, H., Ruthenburg, A. J., Allis, C. D. and Patel, D. J. (2007). How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers, *Nature structural & molecular biology* **14**(11): 1025–1040.

Thomas, J. F., Valencia-Sánchez, M. I., Tamburri, S., Gloor, S. L., Rustichelli, S., Godínez-López, V., De Ioannes, P., Lee, R., Abini-Agbomson, S., Gretarsson, K. et al. (2023). Structural basis of histone h2a lysine 119 deubiquitination by polycomb repressive deubiquitinase bap1/asxl1, *Science Advances* **9**(32): eadg9832.

Thomas, J. O. and Kornberg, R. D. (1975). An octamer of histones in chromatin and free in solution., *Proceedings of the National Academy of Sciences* **72**(7): 2626–2630.

Turner, B. M. (1993). Decoding the nucleosome, *Cell* **75**(1): 5–8.

Uckelmann, M., Levina, V., Taveneau, C., Ng, X. H., Pandey, V., Martinez, J., Mendiratta, S., Houx, J., Boudes, M., Venugopal, H. et al. (2023). Dynamic prc1-cbx8 stabilizes a porous structure of chromatin condensates, *bioRxiv* .

Van der Lugt, N., Domen, J., Linders, K., Van Roon, M., Robanus-Maandag, E., Te Riele, H., Van der Valk, M., Deschamps, J., Sofroniew, M. and Van Lohuizen, M. (1994). Posterior transformation, neurological abnormalities, and severe hematopoietic defects in mice with a targeted deletion of the bmi-1 proto-oncogene., *Genes & development* **8**(7): 757–769.

van der Lugt, N. M., Alkema, M., Berns, A. and Deschamps, J. (1996). The polycomb-group homolog bmi-1 is a regulator of murine hox gene expression, *Mechanisms of development* **58**(1-2): 153–164.

Van Zundert, G., Rodrigues, J., Trellet, M., Schmitz, C., Kastritis, P., Karaca, E., Melquiond, A., van Dijk, M., De Vries, S. and Bonvin, A. (2016). The haddock2. 2 web server: user-friendly integrative modeling of biomolecular complexes, *Journal of molecular biology* **428**(4): 720–725.

Verdin, E. and Ott, M. (2015). 50 years of protein acetylation: from gene regulation to epigenetics, metabolism and beyond, *Nature reviews Molecular cell biology* **16**(4): 258–264.

Voncken, J. W., Roelen, B. A., Roefs, M., de Vries, S., Verhoeven, E., Marino, S., Deschamps, J. and van Lohuizen, M. (2003). Rnf2 (ring1b) deficiency causes gastrulation arrest and cell cycle inhibition, *Proceedings of the National Academy of Sciences* **100**(5): 2468–2473.

Waddington, C. H. (1942). The epigenotype, *Endeavour* **1**: 18–20.

Walsh, C. P., Chaillet, J. R. and Bestor, T. H. (1998). Transcription of iap endogenous retroviruses is constrained by cytosine methylation, *Nature genetics* **20**(2): 116–117.

Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R. S. and Zhang, Y. (2004). Role of histone h2a ubiquitination in polycomb silencing, *Nature* **431**(7010): 873–878.

Wang, M.-C., Jiao, M., Wu, T., Jing, L., Cui, J., Guo, H., Tian, T., Ruan, Z.-p., Wei, Y.-C., Jiang, L.-L. et al. (2016). Polycomb complex protein bmi-1 promotes invasion and metastasis of pancreatic cancer stem cells by activating pi3k/akt signaling, an ex vivo, in vitro, and in vivo study, *Oncotarget* **7**(8): 9586.

Wang, R., Ilangovan, U., Leal, B. Z., Robinson, A. K., Amann, B. T., Tong, C. V., Berg, J. M., Hinck, A. P. and Kim, C. A. (2011). Identification of nucleic acid binding residues in the fcs domain of the polycomb group protein polyhomeotic, *Biochemistry* **50**(22): 4998–5007.

Wang, R., Ilangovan, U., Robinson, A. K., Schirf, V., Schwarz, P. M., Lafer, E. M., Demeler, B., Hinck, A. P. and Kim, C. A. (2008). Structural transitions of the ring1b c-terminal region upon binding the polycomb cbox domain, *BIOCHEMISTRY* **47**(31): 8007–8015.

Wang, R., Taylor, A. B., Leal, B. Z., Chadwell, L. V., Ilangovan, U., Robinson, A. K., Schirf, V., Hart, P. J., Lafer, E. M., Demeler, B. et al. (2010). Polycomb group targeting through different binding partners of ring1b c-terminal domain, *Structure* **18**(8): 966–975.

Wani, A. H., Boettiger, A. N., Schorderet, P., Ergun, A., Münger, C., Sadreyev, R. I., Zhuang, X., Kingston, R. E. and Francis, N. J. (2016). Chromatin topology is coupled to polycomb group protein subnuclear organization, *Nature communications* **7**(1): 10291.

Watt, F. and Molloy, P. L. (1988). Cytosine methylation prevents binding to dna of a hela cell transcription factor required for optimal expression of the adenovirus major late promoter., *Genes & development* **2**(9): 1136–1143.

Weissmann, F. and Peters, J.-M. (2018). Expressing multi-subunit complexes using bigbac, *Protein Complex Assembly: Methods and Protocols* pp. 329–343.

Weissmann, F., Petzold, G., VanderLinden, R., in'T Veld, P. J. H., Brown, N. G., Lampert, F., Westermann, S., Stark, H., Schulman, B. A. and Peters, J.-M. (2016). bigbac enables rapid gene assembly for the expression of large multisubunit protein complexes, *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA* **113**(19): E2564–E2569.

Wheaton, K., Sarkari, F., Johns, B. S., Davarinejad, H., Egorova, O., Kaustov, L., Raught, B., Saridakis, V. and Sheng, Y. (2017). Ube2e1/ubch6 is a critical in vivo e2 for the prc1-

catalyzed ubiquitination of h2a at lys-119, *Journal of Biological Chemistry* **292**(7): 2893–2902.

Wong, S. J., Gearhart, M. D., Taylor, A. B., Nanyes, D. R., Ha, D. J., Robinson, A. K., Artigas, J. A., Lee, O. J., Demeler, B., Hart, P. J. et al. (2016). Kdm2b recruitment of the polycomb group complex, prc1. 1, requires cooperation between pcgf1 and bcorl1, *Structure* **24**(10): 1795–1801.

Woodcock, C., Safer, J. and Stanchfield, J. (1976). Structural repeating units in chromatin .1. evidence for their general occurrence, *EXPERIMENTAL CELL RESEARCH* **97**(1): 101–110.

Worden, E. J., Zhang, X. and Wolberger, C. (2020). Structural basis for compass recognition of an h2b-ubiquitinated nucleosome, *Elife* **9**: e53199.

Wu, S., Shi, Y., Mulligan, P., Gay, F., Landry, J., Liu, H., Lu, J., Qi, H. H., Wang, W., Nickoloff, J. A. et al. (2007). A yy1–ino80 complex regulates genomic stability through homologous recombination–based repair, *Nature structural & molecular biology* **14**(12): 1165–1172.

Wu, X., Johansen, J. V. and Helin, K. (2013). Fbxl10/kdm2b recruits polycomb repressive complex 1 to cpg islands and regulates h2a ubiquitylation, *Molecular cell* **49**(6): 1134–1146.

Yasunaga, S., Ohtsubo, M., Ohno, Y., Saeki, K., Kurogi, T., Tanaka-Okamoto, M., Ishizaki, H., Shirai, M., Mihara, K., Brock, H. W. et al. (2013). Scmh1 has e3 ubiquitin ligase activity for geminin and histone h2a and regulates geminin stability directly or indirectly via transcriptional repression of hoxa9 and hoxb4, *Molecular and cellular biology* **33**(4): 644–660.

Yuan, W., Xu, M., Huang, C., Liu, N., Chen, S. and Zhu, B. (2011). H3k36 methylation antagonizes prc2-mediated h3k27 methylation, *Journal of Biological Chemistry* **286**(10): 7983–7989.

Zhang, D., Tang, Z., Huang, H., Zhou, G., Cui, C., Weng, Y., Liu, W., Kim, S., Lee, S., Perez-Neut, M. et al. (2019). Metabolic regulation of gene expression by histone lactylation, *Nature* **574**(7779): 575–580.

Zhao, J., Wang, M., Chang, L., Yu, J., Song, A., Liu, C., Huang, W., Zhang, T., Wu, X., Shen, X. et al. (2020). Rybp/yaf2-prc1 complexes and histone h1-dependent chromatin compaction mediate propagation of h2ak119ub1 during cell division, *Nature cell biology* **22**(4): 439–452.

Zhen, C. Y., Tatavosian, R., Huynh, T. N., Duc, H. N., Das, R., Kokotovic, M., Grimm, J. B., Lavis, L. D., Lee, J., Mejia, F. J., Li, Y., Yao, T. and Ren, X. (2016). Live-cell single-molecule tracking reveals co-recognition of h3k27me3 and dna targets polycomb cbx7-prc1 to chromatin, *eLife* **5**: e17667.
URL: *https://doi.org/10.7554/eLife.17667*

Zheng, S. Q., Palovcak, E., Armache, J.-P., Verba, K. A., Cheng, Y. and Agard, D. A. (2017). Motioncor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy, *NATURE METHODS* **14**(4): 331–332.

Zhou, K., Gaullier, G. and Luger, K. (2019). Nucleosome structure and dynamics are coming of age, *Nature structural & molecular biology* **26**(1): 3–13.

Zivanov, J., Nakane, T., Forsberg, B. O., Kimanius, D., Hagen, W. J., Lindahl, E. and Scheres, S. H. (2018). New tools for automated high-resolution cryo-em structure determination in relion-3, *eLife* **7**: e42166.
URL: *https://doi.org/10.7554/eLife.42166*

Acknowledgment

As I stand at the finish line of this PhD marathon, I find myself with a long list of people to thank.

First and foremost, I'd like to express my gratitude to my supervisor, Jürg Müller, for giving me the opportunity to work on this project. I'm also grateful to Prof. Elena Conti and the entire Conti lab for their help with the structural biology aspects of my work.

To my thesis advisory committee – Prof. Dr. Michael Stattler, Dr. Jérôme Basquin, Prof. Dr. Elena Conti, and and Dr. Daniel Bollschweiler – thank you for your insightful guidance.

A heartfelt thank you to my labmates and friends. Ksenia, Svetlana, and Anastasia – you became more than colleagues; you became my support system, both scientifically and personally. Who knew that PCR and protein purification failures could forge such strong friendships?

To the postdocs of Müller lab, Pavel and Jacques – thank you for our discussions on Polycomb throughout the years. Maria, my steadfast ally in unravelling the mysteries of PRC1, thank you for your collaboration. Sven and Katharina, the cell culture wizards – your help was invaluable. And Sven, your strategic deployment of Kinder chocolate bars during moments of experimental adversity probably saved me from despair.

I owe a debt of gratitude to the facility managers of MPI cryo-EM and mass spectrometry departments, particularly Barbara and Daniel. Your advice shaped a significant portion of this thesis. Daniel, your support during this years went above and beyond – I hope this acknowledgment makes up for all the times I pestered you with "quick" questions.

To the members of the Conti lab – Jérôme, Christian, Iuliia, Claire, and Mahesh – thank you for everything from protein purification and cryo-EM processing tips to tea-time chats.

A heartfelt thanks to Dr. Jan Konvalinka, who took a chance on me as a clueless bachelor student a decade ago. You and entire Konvalinkovci team didn't just teach me science; you showed me I belonged in a lab. Mnohokrát děkuji!

To my friend Sofia – we started this journey together at Vysoká škola chemicko-technologická v Praze and despite the geographical divide, your PhD solidarity and support never wavered.

To my partner Chrisna – your love and support have been my constant through this rollercoaster

ride. Your presence in my life is a gift I cherish beyond words.

Finally, to my parents – thank you for your unwavering love and belief in me.

So, to everyone mentioned here and to those I might have forgotten in my thesis-induced haze – thank you. This PhD thesis may have my name on it, but it's been a team effort all the way.