Characterisation of transcriptional and chromatin events in relation to floral transition and identification of nuclear organisation determinants

del Prete, S.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 2

One, Two, Three: Polycomb Proteins Hit All Dimensions of Gene Regulation

Stefania Del Prete\textsuperscript{a}, Pawel Mikulski\textsuperscript{b}, Daniel Schubert\textsuperscript{b}, Valérie Gaudin\textsuperscript{a}

Author affiliations
\textsuperscript{a}Institut Jean-Pierre Bourgin, INRA, AgroParisTech, CNRS, Université Paris-Saclay, F-78000 Versailles, France
\textsuperscript{b}Institute for Biology, Freie Universität Berlin, Königin-Luise-Strasse 12-16, Berlin D-14195, Germany

This chapter was published as:
Abstract

Polycomb group (PcG) proteins contribute to the formation and maintenance of a specific repressive chromatin state that prevents the expression of genes in a particular space and time. Polycomb repressive complexes (PRCs) consist of several PcG proteins with specific regulatory or catalytic properties. PRCs are recruited to thousands of target genes, and various recruitment factors, including DNA-binding proteins and non-coding RNAs, are involved in the targeting. PcG proteins contribute to a multitude of biological processes by altering chromatin features at different scales. PcG proteins mediate both biochemical modifications of histone tails and biophysical modifications (e.g., chromatin fiber compaction and three-dimensional (3D) chromatin conformation). Here, we review the role of PcG proteins in nuclear architecture, describing their impact on the structure of the chromatin fiber, on chromatin interactions, and on the spatial organization of the genome in nuclei. Although little is known about the role of plant PcG proteins in nuclear organization, much is known in the animal field, and we highlight similarities and differences in the roles of PcG proteins in 3D gene regulation in plants and animals.
Chapter 2

Introduction

In 1947, P. Lewis (Lewis, 1947) discovered the Polycomb (Pc) gene. While conducting genetic studies in *Drosophila melanogaster*, he found that the Pc mutation led to the formation of ectopic sex combs on the second and third legs of adult male flies. Later, while studying the segmentation patterns of *Drosophila*, it was shown that Pc is a key developmental regulator required to maintain homeotic gene (Hox) repression of the Bithorax complex (BX-C) (Lewis, 1978). Polycomb group (PcG) proteins were unified in a common family on the basis of their similar impact on development and their transcriptional repressive function, but they have various molecular activities (Schwartz and Pirrotta, 2007; Köhler and Villar, 2008; Schwartz and Pirrotta, 2013; Pu and Sung, 2015). Beyond their originally identified roles in controlling the sequence and timing of developmental switches and in maintaining cell and organ identity both in animals and plants (Hennig and Derkacheva, 2009; Xiao and Wagner, 2015), it is now clear that PcG proteins accomplish a variety of functions. For instance, recent evidence has implicated plant PcG proteins in abiotic and biotic stress responses (De Lucia and Gaudin, 2013; Kleinmanns and Schubert, 2014). During development, stress responses, and other processes, PcG proteins participate in a memory system and establish, maintain, and transmit silent epigenetic chromatin states. However, how PcG repression is established and transmitted *in vivo*, what mechanisms underlie repression, and how PcG activity is reset are only partially understood, even though intensive research efforts in diverse eukaryote models, from flies to plants, have attempted to answer these questions for more than sixty years.

During evolution, the genes encoding PcG proteins diversified from single copy genes to small gene families in animals, plants, and fungi (Hennig and Derkacheva, 2009). PcG proteins form large and diverse complexes harboring catalytic subunits, which mediate post-translational modifications of the histone tails, and regulatory subunits, which regulate the transcriptional state of genes. Two main Polycomb repressive complexes (PRCs) each consisting of four core components were initially identified in *Drosophila*: Polycomb Repressive Complex (PRC) 1 and PRC2. The core components are generally conserved in plants and animals, but some kingdom-specific components evolved to meet the
particular needs of the organism’s life cycle (for detailed reviews: (Pien and Grossniklaus, 2007; Hennig and Derkacheva, 2009; Bemer and Grossniklaus, 2012; Molitor and Shen, 2013)). For instance, the plant LIKE HETEROCHROMATIN PROTEIN1 (LHP1) is a functional homolog of Pc, despite its structural homology with the HP1 protein family (Gaudin et al., 2001; Zhang et al., 2007b). PcG not only biochemically modifies chromatin (Lagarou et al., 2008; Bratzel et al., 2010; Tavares et al., 2012), but also induces biophysical changes in chromatin structure (Francis et al., 2004; Bantignies and Cavalli, 2011; Beh et al., 2012; Schuettengruber and Cavalli, 2013). Furthermore, some PcG proteins can be part of other additional and unrelated complexes.

PcG function is antagonized by another group of proteins that modifies chromatin and regulates genes, the trithorax group (TrxG) proteins. The TrxG proteins activate gene expression and, like PcG proteins, are conserved in eukaryotes (Ringrose and Paro, 2004). PcG and TrxG proteins often have the same target genes, and the activity of these genes is finely tuned by the opposing action of these two protein complexes.

Here, we present a brief overview of the molecular functions, recruitment, and regulation of PcG, with an emphasis on new research that highlights the roles of PcG proteins at different nuclear scales.

**Hierarchy vs. Plasticity for More Flexibility in Repression?**

**Canonical Two-Step Mechanism of Repression**

PRC1 and PRC2 often occupy the same target locations and can act in a sequential manner, as proposed in the “canonical model” of PcG repression. The core PRC2 complex trimethylates histone H3 at lysine 27 (H3K27me3) at its target genes in a process that usually takes place on a genomic region called the nucleation site, and the deposited mark spreads over adjacent nucleosomes (Finnegan and Dennis, 2007; De Lucia et al., 2008; Pinter et al., 2012). The
deposition of H3K27me3 triggers binding of the PRC1 complex, which further ubiquitinates a lysine residue of histone H2A. These modifications result in other complementary factors being recruited, leading to inhibition of transcription and change in chromatin organization. The inhibition mechanism is elusive. On the other hand, TrxG complexes promote transcription, mostly by introducing the H3K4me3 and H3K36me3 activation marks. Remodeling of chromatin renders it more accessible to transcription factors and triggers transcriptional activation; however, this activation does not necessarily rely on the promotion of Polymerase II complex recruitment to target genes, but rather on transcriptional elongation from poised Polymerase II complexes (Srinivasan et al., 2008; Dorighi and Tamkun, 2013). The functions and hierarchical mode of action of PcG and TrxG proteins were extensively reviewed elsewhere (Schuettengruber et al., 2011; Khanna et al., 2014; Khan et al., 2015; Mozgova et al., 2015).

Gene expression dynamics are not only regulated by the antagonistic roles of PcG and TrxG complexes, but also by the active removal of histone modifications deposited by these proteins. For instance, plant histone demethylases (HDMs), including JumonjiC (JmjC) domain containing proteins, counteract the action of PcG proteins in physiological processes, such as the flowering transition, shoot development, cell fate determination, or the circadian clock (Chen et al., 2011).

**Non-Canonical Mechanisms of Repression**

For years, the canonical model was considered to describe the only mechanism that regulates the target genes of PcG complexes. Recent studies have uncoupled the functions of PRC1 and PRC2 and revisited the sequential action of the complexes, revealing more complex working mechanisms (Ogawa et al., 2002; Gearhart et al., 2006; Calonje et al., 2008; Lagarou et al., 2008; Gao et al., 2012; Tavares et al., 2012; Yang et al., 2013). Furthermore, PRC1-like complexes, which differ in composition (absence of some of the core subunits present in PRC1 proteins, or the presence of additional subunits), associated activities, or repressive functions, have been reported in mammals, Drosophila, and plants.
In mammals, various additional components can be combined with the core PRC1 subunits to form a large and diverse PRC1 complex family (Levine et al., 2002; Ogawa et al., 2002; Wang et al., 2004; Gil and O'Loghlen, 2014). These additional PRC1 subunits may coordinate other histone modifications and histone crosstalk to repress gene expression. For instance, H3K36 demethylases have been found to associate with PcG proteins in the animal BCOR complex, which regulates a subset of BCL6 target genes (Gearhart et al., 2006) or in the Drosophila dRING-associated factor (dRAF) complex (Lagarou et al., 2008). In dRAF, the H3K36 demethylase KDM2 is also required for efficient H2A mono-ubiquitination (Lagarou et al., 2008), whereas in BCOR, KDM2B targets the complex to unmethylated CpG islands (Farcas et al., 2012). The dRAF complex is not the only PRC1-like complex present in Drosophila; others include the Pho-repressive complex (PhoRC) and Polycomb-repressive deubiquitinase complex (PR-DUB) (Klymenko et al., 2006; Schwartz and Pirrotta, 2008; Scheuermann et al., 2012). Moreover, in mammals, examples of H3K27me3-independent PRC1 targeting or H2A mono-ubiquitination are emerging (Schoeftner et al., 2006; Gao et al., 2012; Tavares et al., 2012).

A recent study of the regulation of seed maturation genes in *A. thaliana* revealed that H2Aub and H3K27me3 are deposited on some PcG target genes independently (Yang et al., 2013). Indeed, the *clf/swn* PRC2 mutant, which has reduced levels of H3K27me3 on several target genes, seems not to be affected in overall levels of H2Aub (Yang et al., 2013). Furthermore, plant PRC1 is sometimes recruited before PRC2 (Yang et al., 2013; Blackledge et al., 2014; Cooper et al., 2014; Kalb et al., 2014). Recent studies have shown that complex relationships exist between PRC1 and PRC2, and that some PRC complexes (*i.e.*, PR-DUB and dRAF) have the capacity to mediate histone modifications other than H3K27 trimethylation and H2A ubiquitination, thus tremendously expanding our understanding of PcG-mediated repression mechanisms.

**Various Ways to Hunt for Targets**

PcG targeting has been reported to rely on the presence of *cis*-regulatory elements, *trans*-acting components (DNA-binding proteins, transcription
factors, scaffolding proteins, non-coding RNAs (ncRNAs)), or even structural properties of chromatin fiber (Simon and Kingston, 2009; He et al., 2013; Kim and Sung, 2014).

In Drosophila, PRCs are recruited at Polycomb Response Elements (PREs), which are cis-regulatory DNA sequences of up to a few hundred base pairs long. PREs contain combinations of several diverse binding motifs for proteins, such as the DNA-binding PcG protein Pho or GAG factor (GAF), which work cooperatively. PREs can be located several tens of kilobases away from the promoter of the target genes, and their properties depend on the sequence context (Schwartz and Pirrotta, 2008; Simon and Kingston, 2009; Kassis and Brown, 2013). The exact combinations of DNA-binding sites and key regulatory elements that determine the recruitment of the PRC complexes remain to be identified.

The DNA-binding PcG protein Pho participates in the targeting of the Drosophila PhoRC complex, and the PhoRC complex may serve as a tethering platform for other PRC complexes at some genomic locations. However, the targeting mechanism for Drosophila PRC1 and PRC2 at PREs and the identity of putative PRE-DNA-binding candidates remain unknown (Kassis and Brown, 2013). In mammals, only a few PREs have been identified; however, 97% of PRC2 targets were shown to correspond to annotated CpG islands or similar CG-rich regions. However, the PREs lack a consensus motif (Ku et al., 2008; Simon and Kingston, 2009).

In A. thaliana, the presence of PREs has not been confirmed. However, GAGA-motifs were recently shown to overlap with the binding sites of FIE, a PRC2 subunit, and these motifs seem to be necessary for H3K27me3 deposition (Deng et al., 2013). GAGA-motifs have also been identified in the target genes of LEAFY (LFY), which are repressed by PRC2 (Winter et al., 2011). Recently, it was shown that the PRC1-like subunit LHP1 interacts with the GAGA factor BPC6 and that this interaction is essential and sufficient to recruit LHP1 to DNA sequences that contain GAGA-motifs in vitro (Hecker et al., 2015). The repression of LEAFY COTYLEDON2 (LEC2), a key regulator of A. thaliana seed development, also requires a negative cis-regulatory element, the
Repressive LEC2 Element (RLE), which is located 150 bp upstream of the first codon and is associated with CT-rich elements. The RLE triggers H3K27me3 deposition and inhibits transcriptional activity (Berger et al., 2011). Conserved Regulatory Elements are also present in the upstream region of the promoters of two KNOX genes (BREVIPEDICELLUS (BP) and KNOTTED-LIKE FROM ARABIDOPSIS THALIANA2 (KNAT2)), which are required for proper organ formation. These sequences are targets of the AS1-AS2 complex, which mediates the repression of KNOX genes (Guo et al., 2008). Subsequently, the AS1-AS2 complex was shown to interact with multiple core components of PRC2 to establish the repressed chromatin state at KNOX genes (Hyun et al., 2013). This was one of the first demonstrations of a plant PRC2 being recruited by specific DNA-binding proteins. Similarly, VAL1 is needed to recruit BMI1, a PRC1 subunit in A. thaliana, and set the repressive state of embryo-specific genes (Yang et al., 2013). More evidence that PcG is recruited by transcription factors came from a study of the MADS box protein, AGAMOUS (AG). AG binds two CArG boxes located 1 kb downstream of WUSCHEL (WUS), and is required for the regulation of H3K27me3 levels and repression of WUS expression (Liu et al., 2011). The MADS box transcription factor, SHORT VEGETATIVE PHASE (SVP), and the GRAS transcription factor, SCARECROW (SCR), also physically interact with LHP1 and bind to specific LHP1 target loci, and thus could participate in the recruitment of PRC complexes and in the maintenance of silent chromatin states at their respective target loci (Cui and Benfey, 2009; Liu et al., 2009).

Long non-coding RNAs (IncRNAs) have been shown to participate in numerous mechanisms that regulate gene expression, through interactions with various proteins, including histone-modifying molecules (Marchese and Huarte, 2014; Peschansky and Wahlestedt, 2014). It is still a matter of debate whether IncRNAs participate in the recruitment of PcG complexes either in cis or in trans (Brockdorff, 2013; Nazer and Lei, 2014). The first IncRNAs, identified based on their ability to tether PRC2 to target genes, were mammalian Xist and the IncRNA HOTAIR (Rinn et al., 2007; Tsai et al., 2010; Chu et al., 2011). Since then, numerous studies have examined the interaction between IncRNAs and PRC in mammals, and some have revealed that PRC interacts with several hundred large intergenic ncRNAs (Khalil et al., 2009).
In *A. thaliana*, the list of identified lncRNAs is growing rapidly, with the emergence of *in silico* and RNA-seq technology, but their functions remain poorly understood (Liu et al., 2012; Di et al., 2014; Wang et al., 2014). However, two lncRNAs, COLD ASSISTED INTRONIC NONCODING RNA (COLDAIR) and COLD INDUCED LONG ANTISENSE INTERGENIC RNA (COOLAIR), are well-characterized regulators of the key floral repressor *FLC* during vernalization (Swiezewski et al., 2009; Heo and Sung, 2011). COLDAIR is a sense-lncRNA transcribed from the vernalization response element (VRE) of the first intron of *FLC*. COLDAIR was shown to interact *in vitro* with the CLF PRC2 subunit. Transiently induced by the cold, with peak expression after 20 days of cold exposure, COLDAIR was proposed to play a role in recruiting PRC2 to stably silence *FLC* (Heo and Sung, 2011). COOLAIR is a set of lncRNA antisense transcripts involved in the early, cold-dependent, and transient transcriptional silencing of *FLC*. Somehow, COOLAIR acts as an indirect recruiter of PRC2, but the precise mechanism behind this function is unclear (Swiezewski et al., 2009; Liu et al., 2010). Recently, the lncRNA APOLO, which is expressed in response to auxin, was shown to interact with LHP1 (Ariel et al., 2014). LHP1 also interacts *in vivo* with the RNA-binding protein LIF2, which thus may participate in the formation of a ribonucleoprotein complex involving LHP1. The specificity of the interaction with the RNA partners may thus rely on the three RNA recognition motifs of LIF2 (Latrasse et al., 2011). Recently, specific and promiscuous interactions have been demonstrated in animals, and the strength of the interactions between PcG proteins and lncRNAs was found to depend on the length of the RNA (Davidovich et al., 2015). Whether plant PRC-lncRNA interactions are specific, promiscuous, or non-specific *in vivo* remains to be further investigated. Furthermore, it remains to be determined whether the lncRNA interaction stabilizes PRC complexes with chromatin, participates in the targeting and recognition of the target by a sequence-specific mechanism, or is involved in other scaffolding mechanisms that coordinate different enzymatic activities merits further investigation.

It was recently proposed that PRC2 also recognizes and distinguishes between “open” and “dense” chromatin, with a preference for the latter. Indeed, PRC2 seems to be targeted to specific chromatin features or chromosomal
structures (Yuan et al., 2012). Thus, despite much progress, many open questions remain about how PcG complexes are recruited to silence specific genes.

On the Path from 1D to 3D

Never Walk Alone—Polycomb and Chromatin Domains

Initial discoveries of PcG targets at the gene-scale were soon followed by a genome-wide analysis of H3K27me3 occupancy and binding of several PcG-proteins (Bracken et al., 2006; Tolhuis et al., 2006; Zhang et al., 2007b; Ku et al., 2008). As expected, a substantial overlap between PRC2-bound fragments and H3K27me3 profiles was found. A correlation between PRC1-binding and H3K27me3 occupancy was also reported, with some exceptions, suggesting that PRC1 has other functions.

Moreover, it was shown that, in plants, H3K27me3 is deposited predominantly on euchromatin, spans whole genes, and targets up to 20%–30% of all genes in A. thaliana (Zhang et al., 2007a; Oh et al., 2008; Bouyer et al., 2011; Lafos et al., 2011) and 10% in human (Bracken et al., 2006; Mohn et al., 2008), stressing the importance of the Polycomb pathway in genome regulation. The H3K27me3 occupancy pattern was also compared to the profiles of various histone marks, genomic features of underlying sequences, protein binding distributions, and transcriptional activity of target genes to identify common patterns and predict gene expression status, resulting in the identification of so-called chromatin states or domains that include regions of similar characteristics. In the first such study in Drosophila, a Polycomb-repressed chromatin state was identified, in which H3K27me3-targeted genomic fragments were associated with genic regions, lower transcription, occupancy of other repressive marks, and binding regions of PcG proteins (e.g., H3K27me2, Pc, E(Z), PCL (Filion et al., 2010), and inversely associated with the active marks H3K4me1/2/3, H3K36me3, and H3K27Ac (Ernst et al., 2011). Several studies distinguished bivalent domains, with repressive marks accompanied by active ones, which are inactive or poised for transcription (Sequeira-Mendes et
al., 2014). One of the main differences between the studied species was that animal H3K27me3-occupied regions consisted of large blocks formed by adjacent regions of a similar epigenetic landscape, whereas the *A. thaliana* epigenome seemed to be organized into small domains of different states interspersed with each other (apart from constitutive heterochromatin, which was spread over the pericentromeric region of the chromosome) (Roudier et al., 2011). Therefore, the epigenetic topography of the genomes has emerged and the development of high throughput sequencing has tremendously refined it (Figure 1).

**PcG and Chromatin Fiber Packaging**

Chromatin compaction and formation of higher-order chromatin structures are proposed to participate in the repression by PcG, by reducing or interfering with DNA accessibility to the transcription machinery. Indeed, *in vitro* studies have shown that the Drosophila PRC1 complex and, more specifically, the Posterior sex combs (PSC) subunit, have the ability to compact nucleosomal arrays and inhibit the chromatin remodeling mediated by the SWI/SNF complex (Francis et al., 2001).

The intrinsically disordered C-terminal region of PSC changes the beads-on-a-string chromatin conformation into high-order chromatin structures (Francis et al., 2004). Linker DNA and histone tails might also participate in chromatin compaction; however, Drosophila PRC1 directly interacts with nucleosomes and its activity seems to be independent of histone tail modifications (Francis et al., 2004). The important role of PRC1 in the compaction of nucleosomal arrays is conserved across metazoans and plants (Francis et al., 2001; Francis et al., 2004; Grau et al., 2011; Beh et al., 2012), but is carried out by different PRC1 subunits, such as M33, a Polycomb homolog in mouse, or the PRC1 component EMBRYONIC FLOWER1 (EMF1) in *A. thaliana* (Grau et al., 2011; Beh et al., 2012). Detailed studies of PSC and M33 revealed that one of the main characteristics of a “chromatin compactor” protein is the presence of highly positively charged domains (Grau et al., 2011).
Figure 1: Polycomb Group (PcG) proteins affect chromatin regulation in three dimensions.
PcG proteins participate in the establishment of the epigenetic topography of the genome by depositing biochemical modifications on histones and inducing chromatin compaction (1D). The presence of PcG-associated marks defines a specific chromatin state, called Polycomb-repressed chromatin. PcG proteins can mediate chromatin looping (2D). PcG complexes are recruited to cis-regulatory elements by various mechanisms (e.g., those involving Polycomb Response Elements (PREs), transcription factors, IncRNAs) to regulate their target genes and change the conformation of the chromatin fiber. PcG proteins act in the nuclear space (3D). They can aggregate with their targets to form Polycomb bodies. Chromatin is organized into distinct topologically associated domains (TADs), some of which are Polycomb-repressed chromatin domains. Spatial localization of chromatin at the nuclear periphery is correlated with repressive histone marks and, in the vicinity of the nuclear pore complexes (NPC), with active histone marks.
Using the FISH technique, one study revealed that the PRC1 Ring1B subunit is also involved in *in vivo* compaction and regulation of Hox loci in murine embryonic stem cells (ESCs). This compaction activity is independent of Ring1B histone H2Aub catalytic activity (Eskeland et al., 2010). Therefore, increasing evidence suggests that PRC1 participates in higher-order chromatin structure organization and that PRC1 has packaging effects on chromatin, which may be sufficient to mediate repression, at least at some specific target genes.

In support of a role of PcG in chromatin compaction, genome-wide studies demonstrated that H3K27me3-marked chromatin regions have low DNA accessibility both in Drosophila and *A. thaliana* (Bell et al., 2010; Shu et al., 2012). In addition, chromatin in *lhp1* mutants is more sensitive to microccocal nuclease digestion than is wild-type chromatin, suggesting a possible role for LHP1 in chromatin compaction. Although LHP1 is functionally similar to Pc, it is classified as a HP1 protein, based on its structure (Gaudin et al., 2001; Zhang et al., 2007b). The chromo shadow domain (CSD) of the HP1 protein family is involved in homo-dimerization, whereas the chromo domain (CD) recognizes methylated lysine residues on histone tails. Interestingly, the binding of two CDs of Swi6, the *Schizosaccharomyces pombe* HP1 homolog, to nearby modified histone H3 nucleosomes creates a protein interface for tetramerization of two Swi6 homo-dimers connected by their CSDs (Canzio et al., 2011). The coordinated action of CD and CSD can thus lead to heterochromatin spreading throughout a stepwise oligomerization process from an auto-inhibited homodimer to chromatin-associated oligomers, which is somewhat analogous to the self-association process of tubulin dimers (Canzio et al., 2013). Whether this model, which is still controversial (Teif et al., 2015), is conserved for the HP1 family and plant LHP1 remains to be established.

Interestingly, it was recently shown that PRC2 has a higher activity on dense oligonucleosomes than on dispersed oligonucleosomes, suggesting that PRC2 can sense the chromatin environment and that its allosteric activation depends on the density of the substrate nucleosomes (Yuan et al., 2012). This study also revealed that local chromatin compaction precedes the establishment of H3K27me3 and provides a better substrate for PRC2 activity (Yuan et al.,
Therefore, PRC1 may initiate local chromatin compaction, which could then lead to the establishment and propagation of the H3K27me3 mark. This would support recent data suggesting that PRC1 acts before PRC2 (Blein et al., 2002; Yang et al., 2013; Cooper et al., 2014; Kalb et al., 2014). However, the structural role of PRC complexes, at the local chromatin scale or at a higher order of chromatin organization, requires further investigation.

**It Takes Two to Tango-Chromatin Loops Mediated by PcG Proteins**

Besides their roles in linear chromatin topography, increasing evidence suggests that PcG also affects the three-dimensional (3D) organization of chromatin at the nuclear level. Although the spatial genome regulation mediated by PcG proteins remains poorly understood in plants, we highlight recent findings, while taking existing evidence in other species into account.

PcG proteins can organize chromatin into 3D long-range loops, both in plants and in animals. These loops may also participate in PcG-mediated repression mechanisms (reviewed in (Cheutin and Cavalli, 2014)). Two techniques are used to investigate 3D chromatin organization, *i.e.*, fluorescence *in situ* hybridization (FISH), which allows direct visualization of the spatial organization of genomic sequences by cytology and microscopy approaches in individual nuclei, and methods derived from chromosome conformation capture (3C), which allow high-throughput and high-resolution analyses of genomic interactions at local to genome-wide scales (de Wit and de Laat, 2012; Hovel et al., 2012; Del Prete et al., 2014).

Using both methods, Lanzuolo and colleagues (Lanzuolo et al., 2007) provided one of the first direct lines of evidence of loop formation in Drosophila. All major regulatory PcG-bound DNA elements at the Bithorax (BX-C) locus were found to physically interact with each other via chromatin long-range interactions. After artificial transcriptional reactivation of the BX-C locus, different conformations of the locus were observed, with the active PREs and promoters losing contact with each other (Lanzuolo et al., 2007). This early study directly linked transcriptional regulation with the regulation of the
chromatin loop formation involving PREs. A direct role for PcG in loop formation was provided by the study of the regulation of the human GATA-4 locus, a target of the PRC2 EZH2 subunit (Tiwari et al., 2008). When EZH2 levels are depleted, the long-range genomic contacts at the GATA-4 locus are disrupted, leading to transcriptional reactivation (Tiwari et al., 2008). PRC2 has subsequently been shown to change chromatin conformation at the murine HoxD locus (Williamson et al., 2014). Analysis of a PRC2 mutant by Carbon-Copy 3C (5C) (which allows for interaction studies between many selected loci of a specific region) and by FISH showed a strong reduction of the 5C interactions over the HoxD region (Williamson et al., 2014). The loops involve physical contacts between PRE and PcG target genes, but can also form between genes and regulatory elements such as enhancers, silencers, and chromatin insulators (Comet et al., 2011; Li et al., 2013; Zhu et al., 2014).

In plants, the presence of loops directly mediated by PcG proteins has not yet been demonstrated. However, in Zea mays (maize) and A. thaliana, examples of loops between flanking regions of genes have been identified. In maize, husk-specific physical interaction between an enhancer located 100 kb upstream of the transcription start site (TSS) and the TSS is required for the high expression of the b1 gene (Louwers et al., 2009). In A. thaliana, a gene loop has been identified that involves the physical interaction of the 5' and 3' flanking regions of the FLC locus (Crevillen et al., 2013; Zhu et al., 2014). Loop formation is regulated by vernalization, as it is disrupted within the first 2 weeks of cold exposure. The disruption occurs in parallel with the switch from an expressed to a Polycomb-silenced state of the FLC locus. Concurrently, clustering of the repressed FLC alleles to a single nuclear location occurs, and this is disrupted in the vrn2 and vrn5 mutants (Rosa et al., 2013). This finding suggests that the VRN2 and VRN5 PRC2 components influence the nuclear organization of repressed FLC alleles. In addition, the PcG protein LHP1 was shown to facilitate auxin-regulated loop formation between a long non-coding RNA and a gene involved in the root development pathway (Ariel et al., 2014). The presence of a loop was correlated with lower expression of these loci, increased DNA methylation, and higher levels of H3K27me3, suggesting that several repressive pathways, including the one mediated by PRC2, are involved in the process.
Understanding the formation of these chromatin hubs is an exciting and active area of research, due to their potential regulatory function and impact on the structure and evolution of genomes. However, it is still unclear whether PcG proteins are required for the formation or stabilization of the loops, and whether the 3D chromatin conformational changes are causes or consequences of altered gene expression. Furthermore, the methods used to examine chromatin loops are technically challenging. Discrepancies have been observed in the frequency of genomic interactions (Simonis et al., 2006; Bello et al., 2007; Tolhuis et al., 2011). In addition, a limited number of studies compared the data resulting from FISH and 3C-derived experiments. These methods give concordant data in some cases, but also yield divergent results (Williamson et al., 2014). Therefore, in future studies of chromatin looping formation and dynamics, both techniques should be used and the results should be compared (Williamson et al., 2014).

**In the Third Dimension: Polycomb Mediates Higher Order Chromatin Organization**

**Polycomb and Topologically Associating Domains**

Derivatives of 3C techniques are used to establish genome-wide maps of chromatin contacts or physical interactions within the genome. In Drosophila, mouse, and human, these studies highlighted that the genome is organized into topologically associating domains (TADs) that participate in its functional architecture in the nuclear space (Lieberman-Aiden et al., 2009; Bantignies et al., 2011; Tolhuis et al., 2011; Dixon et al., 2012; Hou et al., 2012; Nora et al., 2012; Sexton et al., 2012; Ciabrelli and Cavalli, 2014). TADs correspond to linear chromatin regions that fold as specific 3D structures that mainly favor internal interactions within a particular TAD. The mean size of TADs varies among species, ranging from 60 kb in Drosophila to 900 kb in mouse and human.

The distribution of TADs along the chromosome is associated with specific underlying histone modifications, density of nucleosomes, and DNA
modification: Each TAD thus corresponds to one of the four major epigenetic topographical chromatin types: Active chromatin, Polycomb-repressed chromatin, null chromatin, or heterochromatin (Lan et al., 2012; Ciabrelli and Cavalli, 2014). In *D. melanogaster* embryonic nuclei, Polycomb-repressed domains are highly correlated with specific TADs. Furthermore, it was proposed that the intrinsic folding regime of inactive TADs is specific and different from that of active TADs, with the probability of contacts forming between inactive TADs being lower and with a stronger association with the chromosome territory (Sexton et al., 2012). The situation is more complex in mammals, due to the discrepancy between the size of TADs (ranging from 100 kb to 10 Mb) and H3K27me3 domains (1 kb to 100 kb). Only a few overlaps have been reported between the two entities (Noordermeer et al., 2011; Nora et al., 2012).

Sharp boundaries containing insulator sites and housekeeping genes have been shown to mark the limits between TADs (Hou et al., 2012; Sexton et al., 2012). Moreover, the modularity of TAD organization is conserved in different cell types in animals, but intra-TAD interactions can vary greatly according to cell type (Dixon et al., 2012; Nora et al., 2012).

Until now, only a few conformation capture studies have been performed in plants (Feng et al., 2014b; Grob et al., 2014; Wang et al., 2015). In *A. thaliana*, the presence of TADs has not yet been clearly established, the results still dependent on the resolution of the Hi-C methods used (Feng et al., 2014b; Wang et al., 2015). At low resolution and at long-range distances (>100 kb), it seems that the *A. thaliana* genome contains relatively small interacting regions, which are distributed over multiple sites in the genome. The main contact regions depend mostly on the presence of DNA methylation and of H3K9me2 or H3K27me1 marks, which allows for the identification of interactive heterochromatin islands (IHI). These main interactions occur between telomeres, whereas pericentromeric heterochromatin regions interact weakly with the rest of the genome, but strongly with each other (Feng et al., 2014b). AtMORC6, a conserved Microrchidia adenosine triphosphatase required for heterochromatin condensation, seems to play a key role in establishing these heterochromatic genome contacts (Moissiard et al., 2012). A limited number of
H3K27me3 regions showed interactions, and these were mostly scattered throughout the genome and were dependent on PRC2 activity (Feng et al., 2014b). However, using the Hi-C method at higher resolution, about half of the contact regions were shown to be enriched in H3K27me3, H3.1, and H3.3 (Wang et al., 2015). Although TADs were not found to be a prevailing structural feature of the *A. thaliana* genome, hundreds of insulator-like regions or regions analogous to TAD boundaries have been discovered (Wang et al., 2015). These regions are enriched for accessible chromatin sites, various activating epigenetic marks, and highly expressed genes, whereas TAD-like fragments are characterized by opposite patterns, indicating a repressed chromatin state.

Overall, it remains challenging to identify TADs and specific Polycomb-associated TADs in plants. Conformation capture experiments in plants with larger genomes and at higher resolution in *A. thaliana* may cast light on TAD organization.

**Polycomb Clustering**

The 3D organization of PcG target regions has been examined by imaging the subnuclear localization of Polycomb proteins. Interestingly, some Polycomb complex components tend to aggregate in foci instead of being randomly dispersed over the nuclear space (Saurin et al., 1998; Grimaud et al., 2006; Ren et al., 2008; Vandenbunder et al., 2014). These spots of increased signal were called PcG bodies. They differ in size and number in different cell types; as a general rule, fewer and larger foci are present in undifferentiated cells, whereas more numerous and smaller foci occur in differentiated cells (Ren et al., 2008; Del Prete et al., 2014).

Subsequent experiments showed that PcG foci serve as structures in which different Polycomb targets are clustered and co-repressed. A prime example comes from the Drosophila Antp and Abd-B loci. These PcG targets, despite being located 10 Mb apart from each other, co-localize to foci in embryo heads, where they are inactive, and dissociate in differentiated tissue, where at least one of them becomes activated. Further evidence emerged from chromatin
conformation capture studies. As mentioned above, it was shown that PREs of different PcG targets can interact with each other even over long distances (Bantignies et al., 2011; Tolhuis et al., 2011), confirming the imaging data. Recently, more than hundred proteins have been identified as regulators of the 3D distribution of PcG proteins in Drosophila (Gonzalez et al., 2014). In particular, proteins involved in the sumoylation pathway were shown to be critical for the Pc chromatin binding affinity, residence time, and 3D nuclear distribution. Indeed, Pc foci form enlarged aggregates in the absence of SUMO, whereas they are more dispersed when the activity of the SUMO peptidase Velo is reduced (Gonzalez et al., 2014). Interestingly, it was also demonstrated that the Polyhomeotic (Ph) protein assembly at PREs is defective in the absence of its O-GlcNAcylation and that its post-translational modification is required for the ordered and functional assembly of Ph via its SAM domain (Gambetta and Muller, 2014).

In plants, the existence of Polycomb bodies remains elusive. However, several lines of evidence suggest that Polycomb components or their targets undergo clustering. Rosa et al. (Rosa et al., 2013) showed clustering of the PcG target, FLC, that was induced by vernalization and quantitatively correlated with the length of cold exposure. Higher clustering frequency was tightly associated with increased accumulation of H3K27me3 on the FLC nucleation site and clustering was impaired in the PcG mutants vrn2 and vrn5, but not lhp1. In contrast, imaging of LHP1-GFP in A. thaliana showed that LHP1 nuclear distribution patterns vary from a rather uniform pattern in meristematic cells to patterns with multiple distinguishable foci in differentiated cells (Libault et al., 2005). The relationship between LHP1 distribution and the differentiation status of the cell is reminiscent of the link between PcG bodies and cell differentiation in animals.

We believe that further development of imaging and 3D interaction techniques will elucidate the clustering of Polycomb targets in plants and the underlying mechanisms. However, genome arrangement in different species might also be the result of different evolutionary paths, which might lead to distinct mechanisms of repressive 3D interactions.
Polycomb Regulation and Spatial Distribution

Another peculiar aspect of Polycomb regulation involves the distribution of PcG targets in the nuclear space, whose spatial rules have not yet clearly been established. However, the nuclear periphery and nuclear lamina (NL) have been shown to play important roles in gene regulation.

The NL is a protein mesh residing inside the nuclear membrane. In animals, the NL is composed of proteins called lamins, which localize to thin ring structures that cover the inner surface of the nuclear membrane. Despite a lack of substantial sequence homology to animal counterparts, a couple of proteins in plants are believed to fulfill the role of lamins, because of their lamin-like localization and the altered nuclear morphology and size in the corresponding mutants. Prominent examples of plant lamin-like proteins include NMCP1 (D. carota) and CROWDED NUCLEI1-4 (CRWN1-4) (A. thaliana). The NL is not a uniform structure; it spans the nuclear membrane and is interconnected with Nuclear Pore Complexes (NPCs), which are protein structures that form channels for the exchange of molecules between the nucleoplasm and cytoplasm (for reviews, (Parry, 2015; Tamura et al., 2015; Zhou et al., 2015)).

Interestingly, a physical association between the DNA and the nuclear periphery was shown to influence gene regulation. State-of-the-art examples come from studies on human (Guelen et al., 2008) and Drosophila (Pickersgill et al., 2006), in which the authors mapped the interactions between the genome and lamins using the DamID technique (van Steensel and Henikoff, 2000; Germann and Gaudin, 2011) and subsequent genome-wide profiling. Parts of the genome bound by lamins show lower expression and an abundance of repressive chromatin marks, including H3K27me3 (Guelen et al., 2008; Filion and van Steensel, 2010; Ikegami et al., 2010; Peric-Hupkes et al., 2010). Consistently, depletion of Drosophila (Sheveluyov et al., 2009) and C. elegans (Towbin et al., 2010) lamin orthologues caused up-regulation of genes at the nuclear periphery and artificial tethering of transgenes to the periphery resulted in decreased expression (Andrulis et al., 1998; Finlan et al., 2008; Reddy et al., 2008; Dialynas et al., 2010).
Localization to the nuclear periphery was also associated with gene activation in several studies. For instance, several groups showed that the inducible yeast genes *INO1* and *GAL1* are recruited to the vicinity of NPCs upon gene induction and remain there for subsequent reactivation (Brickner and Walter, 2004; Casolari et al., 2005; Schmid et al., 2006; Kundu et al., 2007). Moreover, global chromatin organization seems also to be tissue specific, in rod cells of nocturnal animals there is an inverted chromatin arrangement, so that repression is correlated with the nuclear interior and activation with the periphery (Solovei et al., 2013). Rod cells lack lamins and are responsible for harvesting light; the inverted chromatin arrangement is believed to make rods more efficient. Only a few studies identified a connection between the Polycomb repression mechanism and the nuclear periphery (Fedorova et al., 2008; Wang et al., 2011).

In plants, these studies are still in their infancy. One example of gene repositioning related to gene expression has been reported in *A. thaliana* leaf mesophyll cells (Feng et al., 2014a). Light-inducible loci moved from the nuclear interior to the periphery upon transcriptional activation (Feng et al., 2014a). Whether these spatial movements are correlated with chromatin modifications remains to be elucidated. Interestingly, the E3 ubiquitin ligase HIGH EXPRESSION OF OSMOTICALLY RESPONSIVE GENES 1 (HOS1) interacts with components of the *A. thaliana* NPC, thereby facilitating mRNA export. HOS1 activates *FLC* expression by remodeling chromatin at the *FLC* locus during short-term cold stress. The activation takes place by antagonizing the silencing role of FVE, a PcG-related protein and the displacement of its partner, the histone deacetylase HDA6. This, in turn, leads to increased acetylation of histone H3 on *FLC*, which results in higher expression (Jung and Park, 2013). Whether the two functions in nuclear export and chromatin remodeling are linked is an open question. Finally, *A. thaliana* structural components, such as the CRWN1 and CRWN4 proteins, which control the size of the nucleus, localized at the nuclear periphery and seemed to influence chromosomal organization (Grob et al., 2014). Indeed, Hi-C experiments revealed that the *crwn4* and *crwn1* mutants exhibited increased trans-chromosomal interaction frequencies, suggesting higher levels of chromosomal...
compaction (Grob et al., 2014). Thus, no link with plant PcG proteins has yet been demonstrated.

**Conclusion**

We can view the Polycomb-mediated impact on chromatin dynamics and genome regulation as an integration of three layers or dimensions of repression (Figure 1). The first one is based on the information stored on the linear genome and epigenome. Important aspects of this first dimension include recruitment at specific genomic regions, modification of histone tails, compaction of nucleosomes, and interplay with other factors to form Polycomb-repressed chromatin domains that shape the epigenetic topography of the genome. The second dimension involves the formation of loops on a locus scale, when Polycomb *cis*-regulatory elements interact with each other. Lastly, the third dimension relies on the organization of Polycomb-based TADs and the spatial localization of Polycomb-repressed domains in specific compartments of the nuclear space.

With the development of new methodology and genome-wide approaches combined with modeling to study the 3D arrangement of proteins and genomes, it has become possible to investigate the third dimension of Polycomb repression, and this has deepened our understanding of Polycomb action. PcG proteins continuously surprise us on all layers of repression, by deviating from dogmas and showing non-canonical structures, hierarchies, or mechanisms. As the methodology is now accessible for plants as well, we look forward to the challenge of deciphering the mechanism underlying the 3D PcG-mediated repression in plants, which will certainly bring novel surprises.

**Acknowledgements**

Stefania del Prete and Pawel Mikulski are supported by PhD fellowships provided by the European Commission Seventh Framework-People-2012-ITN Project EpiTRAITS (Epigenetic regulation of economically important plant traits, no-316965). The funding provided by the ITN EpiTRAITS, INRA (to Valérie Gaudin) and the Boehringer Ingelheim Foundation (to Daniel Schubert) is gratefully acknowledged.
References


Cui, H., and Benfey, P.N. (2009). Interplay between SCARECROW, GA and LIKE HETEROCHROMATIN PROTEIN 1 in ground tissue patterning in the Arabidopsis root. Plant J 58, 1016-1027.


