Identification and characterization of human Polycomb-group proteins.
Satijn, D.P.E.

Citation for published version (APA):
CHAPTER 6

Discussion

Satijn D.P.E. and A.P.Otte

BBA, gene structure and expression, Review. 1999, 1447, 1-16.
Polycomb group protein complexes: do different complexes regulate distinct target genes?

David P.E. Satijn, Arie P. Otte *

E.C. Slater Instituut, University of Amsterdam, Plantage Muidergracht 12, 1018 TV Amsterdam, The Netherlands

Received 5 May 1999; accepted 28 July 1999

Keywords: Polycomb group complex; Polycomb group target gene; Chromatin structure

Contents

1. Introduction ......................................................... 1
2. Different PcG protein complexes with different compositions exist ................. 2
   2.1. Original suggestions for the existence of multimeric PcG protein complexes .... 2
   2.2. Bmi1, RING1 and homologs of Pc, Ph and Scm are part of a mammalian PcG complex ..... 3
   2.3. Characterization of Enxl/EZH2 and EED interactions suggests the existence of distinct PcG complexes ........................................... 7
3. Are there cell type specific PcG complexes with different protein compositions? .... 8
4. Different PcG complexes regulate distinct target genes .................................. 9
   4.1. PcG response elements (PREs) ..................................... 9
   4.2. The PcG protein pleiohomeotic (Pho) binds to a specific DNA sequence in PREs .... 10
   4.3. PcG target sites are occupied by distinct PcG complexes .......................... 10
   4.4. Proteins of the PcG and TrG complexes are localized to PREs ...................... 11
   4.5. Vertebrate PcG target genes ...................................... 12
5. Concluding remarks .................................................................. 14
Acknowledgements .................................................................... 14
References ........................................................................... 14

1. Introduction

During embryogenesis, many different cell types are generated, organized in a variety of tissues and organs. The identities of these different cell types are characterized by distinct sets of active and inactive genes. The cell identities need to be maintained

* Corresponding author. Fax: +31-20-5255124; E-mail: arie.otte@chem.uva.nl
6. Discussion

through cell division. To achieve this, the cell type specific expression patterns of active and inactive genes have to be stably transmitted to the daughter cells. It is believed that the establishment of the cell type specific gene expression patterns involves activation of gene expression at the level of promoters, enhancers and transcription factors. However, different levels of control mechanisms are required for the maintenance of cell type specific gene expression patterns through cell division. In all eukaryotes, several chromatin-associated protein complexes have been identified that are involved in the maintenance of cell type specificity, one of which is the polycomb group (PcG) complex.

PcG proteins have initially been identified in *Drosophila* as being involved in the maintenance of the correct expression patterns of homeotic genes. Homeotic genes are differentially expressed in specific embryonic body segments. As such, homeotic genes play a crucial role in the proper development and subsequent maintenance of the body plan, in both vertebrates and invertebrates. In PcG mutants, the expression patterns of homeotic genes are initiated normally, early in embryonic development [1,2]. However, at later developmental stages, the homeotic genes become expressed in body segments where they normally remain repressed. This results in dramatic changes in the body plan, which are called homeotic transformations [3–5]. In PcG mutants, homeotic genes are expressed outside the body segments where they normally are repressed and this indicates that PcG proteins are involved in stable repression of homeotic genes. Similarly, a second class of proteins, named the trithorax (Trx) group (TrG), has been identified to counteract the action of the PcG proteins. In contrast with PcG proteins, TrG proteins are needed for stable expression of homeotic genes. Based on their respective modes of action, PcG and TrG proteins therefore represent a cellular memory system that is important for the inheritable transmission of gene expression patterns.

PcG proteins have originally been identified in *Drosophila*, but in recent years, PcG homologs have also been found in other vertebrates and invertebrates, such as human, mouse, chicken, *Xenopus* and *Caenorhabditis elegans* (Table 1). In spite of the widespread existence of PcG proteins, very little is known about the underlying molecular mechanism by which PcG proteins silence genes. However, one important observation is that PcG proteins form large, chromatin-associated multimeric protein complexes. It has, therefore, been suggested that PcG-mediated gene silencing is a result of interference with the chromatin structure of their target genes. Evidence is accumulating that PcG complexes differ considerably in their compositions and this has important consequences for the specificity by which they associate with target genes and ultimately, their biological functions. In this review, we will discuss (i) that different multimeric PcG protein complexes exist and that their compositions are cell type specific and developmentally regulated and (ii) that different PcG complexes associate with distinct target genes, which would explain that specific PcG genes have different biological functions.

2. Different PcG protein complexes with different compositions exist

2.1. Original suggestions for the existence of multimeric PcG protein complexes

In *Drosophila*, PcG genes have been defined as a class of up to 30 genes which have in common that, when mutated, similar phenotypes are induced. These phenotypes are all characterized by homeotic transformations [6–9]. Most PcG mutants display only a weak phenotype and only a few mutants, such as the polycomb (Pc) mutant itself, display severe homeotic transformations [3]. However, when more than one PcG gene is mutated, the homeotic transformations are strongly enhanced. This has been shown in double mutant combinations with, for instance, polycomb-like (Pcl), posterior sex combs (Psc) and sex combs on midleg (Scm) [1,3,10,11]. The synergistic effects of double PcG mutations indicate that strong dosage interactions between the PcG proteins exist. This has been interpreted as such that PcG proteins form multimeric protein complexes. Another clue for the existence of multimeric PcG protein complexes was the observation that the different PcG proteins, including Pc, polyhomeotic (Ph) and Psc, bind in overlapping patterns on polytene chromosomes in *Drosophila* salivary gland cells [12,13]. At some loci, only the Psc
protein binds, whereas at other loci, only Pc and Ph are present [13]. These data already suggested that PcG complexes that bind to different genetic loci may have different compositions. The first biochemical evidence for the existence of multimeric PcG protein complexes came from co-immunoprecipitation and co-fractionation experiments performed in lysates from *Drosophila* embryos. Using specific antibodies against the *Drosophila* PcG proteins Pc and Ph, it was shown that both proteins are co-immunoprecipitated. In addition, size fractionation experiments indicated that a large portion of the two proteins is present in a distinct complex [14]. All these data indicated the existence of multimeric PcG protein complexes. Below, we will discuss what has been learned since.

2.2. *Bmi1*, *RING1* and homologs of *Pc*, *Ph* and *Scm* are part of a mammalian PcG complex

2.2.1. *Ph* homologs interact with *Bmi1* and with *Scm*

The genetic yeast two-hybrid system allows us to search for proteins that physically interact with a protein of choice. Several PcG proteins have been used as target in the two-hybrid system to identify interacting proteins. One such PcG protein is Bmi1, a proto-oncopogene that has originally been identified to be involved in B- and T-cell lymphoma formation [15,16]. The Bmi1 protein shares domains of high homology with the *Drosophila* PcG proteins Psc and suppressor 2 of Zeste (Su(z)2) [16,17]. Bmi1 is considered to be a functional PcG protein homolog, since Bmi1 --/-- mutant mice display homeotic transformations that are typical for *Drosophila* PcG mutants [18,19].

The murine *Bmi1* protein was used to identify proteins that are part of the mammalian PcG complex. Murine and human homologs of the *Drosophila* PcG protein Ph were identified to interact with Bmi1. Two human Ph homologs, HPH1 and HPH2, were identified [20] (Fig. 1 and Table 1). HPH1 proved to be the human homolog of a previously identified murine protein, Rae-28 [21]. Rae-28 or MPh1 was also identified in a two-hybrid screen with Bmi1 [22,23]. The Rae-28 protein is considered to be a functional PcG protein homolog, since Rae-28 --/-- mutant mice display homeotic transformations [24]. HPH2 was a novel protein and its murine homolog, MPh2, was subsequently identified in a Bmi1 two-hybrid screen [25]. Importantly, also the *Drosophila* Psc and Ph proteins have been found to interact with each other [26] (Table 1).

Besides two-hybrid analysis, GST fusion proteins are used in an independent in vitro assay (GST pull-down assay) [27] to verify that the proteins interact directly. Also, specific antibodies are commonly used to perform co-immunoprecipitation experiments. If two proteins are in vivo part of a protein complex, it is to be expected that they co-immunoprecipitate. When nuclear proteins such as PcG proteins are part of one complex, it is to be expected that they also co-localize to a certain extend in the nucleus. Such additional analyses have confirmed that the interactions between the vertebrate homologs of Ph and Bmi1 are direct and that they do exist in vivo. Bmi1 and Ph homologs were found to co-immunoprecipitate and they co-localize in large nuclear domains in mammalian cell lines, called PcG domains [20,22]. These independent lines of evidence indicate that vertebrate
Table 1
PcG homologs and interacting proteins

<table>
<thead>
<tr>
<th>PcG protein homologs</th>
<th>Motifs</th>
<th>Interacting proteins</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pc</td>
<td>Chromodomain C-box</td>
<td>Psc, RING1</td>
<td>[10, 26]</td>
</tr>
<tr>
<td>Xpc</td>
<td></td>
<td>XBmi1</td>
<td>[17]</td>
</tr>
<tr>
<td>CHCB3</td>
<td></td>
<td>–</td>
<td>[90]</td>
</tr>
<tr>
<td>M33</td>
<td></td>
<td>M33, Bmi1, Ring1A</td>
<td>[23, 32, 53]</td>
</tr>
<tr>
<td>Mpc2</td>
<td></td>
<td>–</td>
<td>[52]</td>
</tr>
<tr>
<td>Cbx2/HPC1</td>
<td></td>
<td>RING1</td>
<td>[54], unpublished results</td>
</tr>
<tr>
<td>HPC2</td>
<td></td>
<td>RING1, CtBP</td>
<td>[31, 51, 60]</td>
</tr>
<tr>
<td>Psc</td>
<td>RING finger</td>
<td>Ph, Scm</td>
<td>[26, 29, 28, 93]</td>
</tr>
<tr>
<td>XBmi1</td>
<td>Helix-turn-helix</td>
<td>HPH1, HPH2, Xpc</td>
<td>[17, 20]</td>
</tr>
<tr>
<td>Bmi1</td>
<td></td>
<td>Bmi1, Ring1B, M33</td>
<td>[15, 16, 23, 25]</td>
</tr>
<tr>
<td>mel-18</td>
<td></td>
<td>–</td>
<td>[33]</td>
</tr>
<tr>
<td>BMI1</td>
<td></td>
<td>RING1, BMI1</td>
<td>[35, 92]</td>
</tr>
<tr>
<td>Ph</td>
<td>SPM</td>
<td>Ph, Scm</td>
<td>[26, 29, 28, 93]</td>
</tr>
<tr>
<td>Rae-28/MPh1</td>
<td></td>
<td>Rae-28, Bmi1, Scm</td>
<td>[21–23, 29]</td>
</tr>
<tr>
<td>MPh2</td>
<td></td>
<td>Bmi1, Ring1B</td>
<td>[25]</td>
</tr>
<tr>
<td>HPH1</td>
<td></td>
<td>HPH1, HPH2</td>
<td>[20]</td>
</tr>
<tr>
<td>HPH2</td>
<td></td>
<td>XBmi, HPH1, HPH2</td>
<td>[20]</td>
</tr>
<tr>
<td>Scm</td>
<td>SPM</td>
<td>Scm, Rae-28</td>
<td>[28, 94]</td>
</tr>
<tr>
<td>SCML1</td>
<td>Zinc finger</td>
<td>Scm, Rae-28</td>
<td>[28, 94]</td>
</tr>
<tr>
<td>E(z)</td>
<td>SET</td>
<td>Esc</td>
<td>[39, 41, 42, 107]</td>
</tr>
<tr>
<td>mes-2</td>
<td>N-terminal domain I</td>
<td>Esc</td>
<td>[45–47]</td>
</tr>
<tr>
<td>Enx1/Ezh2</td>
<td>C. elegans</td>
<td>N-terminal domain II</td>
<td>[43, 44, 96, 108]</td>
</tr>
<tr>
<td>Enx2/Ezh1</td>
<td>Mouse</td>
<td>mes-6</td>
<td>[44, 97]</td>
</tr>
<tr>
<td>EZH1</td>
<td>Mouse</td>
<td>eed, EED, Vav</td>
<td>[98]</td>
</tr>
<tr>
<td>EZH2</td>
<td>Human</td>
<td>eed</td>
<td>[99]</td>
</tr>
<tr>
<td>Esc</td>
<td>WD-40 repeat</td>
<td>Esc</td>
<td>[41, 100–102]</td>
</tr>
<tr>
<td>mes-6</td>
<td>C. elegans</td>
<td>E(z)</td>
<td>[45]</td>
</tr>
<tr>
<td>ced</td>
<td>Mouse</td>
<td>mes-2</td>
<td>[44, 103, 107]</td>
</tr>
<tr>
<td>EED</td>
<td>Human</td>
<td>Enx1/Ezh2, Enx2, HnRNPK</td>
<td>[43, 104]</td>
</tr>
<tr>
<td>Pho</td>
<td></td>
<td>E(Z)</td>
<td>[41, 100–102]</td>
</tr>
<tr>
<td>YY1</td>
<td></td>
<td>mes-2</td>
<td>[45]</td>
</tr>
<tr>
<td>Enhancer of polycomb</td>
<td></td>
<td>Enx1/Ezh2, Enx2, HnRNPK</td>
<td>[44, 103, 107]</td>
</tr>
<tr>
<td>E(Pc)</td>
<td></td>
<td>Enx1/Ezh2</td>
<td>[43, 104]</td>
</tr>
<tr>
<td>Epc1, Epc2</td>
<td></td>
<td>Scm</td>
<td>[28, 94]</td>
</tr>
<tr>
<td>EPC1, EPC2</td>
<td></td>
<td>HPH1, HPH2</td>
<td>[20]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Xbmi, HPH1, HPH2</td>
<td>[20]</td>
</tr>
</tbody>
</table>

homologs of Ph and Bmi1 are part of a multimeric PcG protein complex.

The two-hybrid system can also be used to characterize putative interactions between already characterized PcG proteins and to map the protein domains that are responsible for the respective interactions. A directed two-hybrid assay was used to map the domains in the Ph homologs and Bmi1 that mediated the interaction between these proteins. HPH1 and HPH2 share little sequence homology.
with each other, except in two highly conserved domains which they share with the *Drosophila* PcG protein Ph (Fig. 1). The two homology domains have been designated homology domain I and II, of which the latter is also known as SPM domain, since it is conserved between the proteins Scm, Ph and malignant brain tumor [28] (Fig. 1 and Table 1). It was found that the two conserved homology domains are involved in the interaction between the two HPH proteins and Bmi1 (Fig. 1). Both intact homology domains need to be present for the interaction with Bmi1 [20]. Interestingly, it was also found that the SPM domain of HPH1 and HPH2 is able to mediate heterodimerization between HPH1 and HPH2 [20] or homodimerization of Rae-28/MPh1 [22].

A directed two-hybrid protein-protein analysis with *Drosophila* PcG proteins further revealed that the PcG protein Scm interacts with the *Drosophila* Ph protein [28,29]. Scm and Ph share the highly conserved SPM domain and Scm and Ph interact through this domain (Fig. 1 and Table 1). Not unexpectedly, the *Drosophila* Scm protein also interacts with Rae-28/MPh1 through the SPM domain [28,29]. These data indicate a central role for the conserved SPM domain in mediating multiple protein-protein interactions between distinct PcG proteins.

The Bmi1 protein contains two domains of interest. A specific zinc binding domain termed the RING finger has been suggested to play a role in protein-protein interactions. The second domain is the helix-turn-helix-turn-helix-turn motif which is found necessary for transcriptional repression [30] (Fig. 1). Detailed, directed two-hybrid analysis showed that the central part of Bmi1, which includes the putative helix-turn-helix-turn-helix-turn motif, is required for the interaction with mouse Ph homologs, Rae-28/MPh1 and MPh2. Notably, the RING finger domain of Bmi1 is not involved in the interaction with vertebrate Ph homologs [22,25]. Besides the interaction between Bmi1 and Ph homologs, Bmi1 has also been found to interact with *Xenopus* and murine Pc homologs. Both in a directed two-hybrid assay and in an in vitro GST pull-down assay, this interaction has been identified [23,26,27] (Fig. 1).

These data confirmed the earlier results suggesting the existence of multimeric PcG protein complexes, such as overlapping PcG binding patterns on polytene chromosomes in *Drosophila* salivary gland cells [12,13]. Also, the co-immunoprecipitation experiments of Pc and Ph proteins of [14] have been confirmed, although these data only indicated the presence of Ph and Pc proteins within one complex and not the precise interaction partners.

### 2.2. RING1 interacts with Bmi1 and Pc homologs

In two-hybrid screens, vertebrate Pc homologs have been used as a target to search for interacting proteins. Mammalian homologs of the RING1 protein have been identified as interacting with the human Pc homolog HPC2 [31] and the murine Pc homolog M33 [32] (Fig. 2 and Table 1). RING1 is a previously identified protein that contains a RING finger (Fig. 2). This is a specific zinc binding domain which is found in many regulatory proteins, including the *Drosophila* PcG proteins Psc and Su(z)2 and their vertebrate PcG homologs Bmi1 (Fig. 1) and mel-18 [33]. RING1 is a protein with unknown function, which has not yet been identified in *Drosophila*. Although the RING finger domain of RING1 has been hypothesized to be involved in mediating protein-protein interactions [34], a detailed, directed two-hybrid analysis showed that the RING finger motif is not involved in the interaction with Pc homologs [31,32] (see Fig. 2).

RING1 interacts with a *Xenopus* Pc homolog,
Xpc, and the human Pc homolog HPC2 [31]. Similarly two mouse homologs of RING1, Ring1A and Ring1B, have been found to interact with a mouse homolog of Pc, M33 [32]. The interaction between RING1 and Pc homologs has been verified using different assays. RING1 directly interacts with HPC2 in an in vitro GST pull-down assay [35]. RING1 also co-immunoprecipitates with HPC2, Bmi1 and HPH1 and HPH2 and co-localizes with these proteins in PcG domains in mammalian cell lines [31,32]. All lines of evidence indicate that RING1 is part of a mammalian PcG complex. Although no Drosophila RING1 protein has been identified thus far, RING1 is also able to interact with the Drosophila Pc protein (our unpublished observations).

Several vertebrate Pc homologs have been identified and it has been found that the Pc proteins contain two evolutionary conserved domains (Fig. 2). The N-terminal chromodomain is essential for the binding of Pc proteins to the chromatin. When the chromodomain is either deleted or mutated, it no longer binds to chromatin [36]. In Drosophila, several naturally Pc mutants occur of which the C-terminal domain is mutated or deleted [37]. The conserved C-terminal domain is crucial for the Pc function since Pc mutants which lack this C-terminal domain are unable to repress gene activity [36,38]. The interaction of RING1 with Pc homologs involves precisely this evolutionary conserved C-terminal repression domain of Pc (Fig. 2). This suggests that the repression mediated by Pc homologs is at least in part mediated through the interaction with RING1. The fact that RING1 is able to repress gene activity when targeted to a promoter [31,32] supports this idea.

A directed two-hybrid analysis further showed that the human RING1 protein interacts with itself. The interaction involves two domains within the RING1 protein, one being the RING finger domain, the other is a less well-defined C-terminal region (Fig. 2) [35]. This analysis also resulted in the identification of an interaction between RING1 and Bmi1, as well as an interaction between Bmi1 and itself [35]. The RING fingers of both proteins appear to be involved in the RING1-RING1, RING1-Bmi1 and Bmi1-Bmi1 protein-protein interactions [25,35] (Figs. 1 and 2).

In the last two years, it has become clear that several PcG proteins are capable to interact with each other and that several PcG proteins are capable to interact with more than one protein. Bmi1 interacts with homologs of Pc and Ph as well as the RING1 protein. Extensive homo- and heterodimerization occurs between all these proteins. This has led us to propose a model in which Bmi1, RING1, HPC2 and HPH1 interact with each other to form a multimeric PcG complex (Fig. 4). Besides these proteins, closely related homologs (see also below) such as the Pc homolog M33, mel-18, HPH2 and Scm homologs are part of this complex.

---

Fig. 3. The PcG protein EED binds to the PcG protein EZH. The EZH protein is a human homolog of the Drosophila PcG protein E(z). The protein contains a conserved SET domain and two N-terminally located homology domains to which the EED protein binds. EED is a human homolog of the Drosophila PcG protein E(z). EED contains five WD-40 domains. All five WD-40 domains need to be present and intact for the EZH protein to bind to the EED protein.

---

Fig. 4. Model of two distinct human PcG protein complexes. Two human PcG complexes have been identified that appear not to interact physically. One complex contains the RING1, Bmi1, HPH and HPC2 proteins. Multiple interactions and homodimerizations occur between these proteins. The other complex contains the EZH and EED proteins that interact with each other. Either PcG complex interacts with a PRE located in the vicinity of PcG target genes. Potentially, oligomerization or polymerization of the PcG complexes results in occupation of the promoter region and subsequent repression of the target gene.
2.3. Characterization of Enx1/EZH2 and EED interactions suggests the existence of distinct PcG complexes

The Drosophila PcG protein enhancer of zeste (E(z)) holds a particular position among the class of PcG proteins. Genetic data show that the E(z) protein can be considered both as a PcG protein and a TrG protein [39,40]. The Drosophila PcG protein E(z) and protein homologs in vertebrates contain a highly conserved domain, the SET domain. The SET domain is commonly found in a range of proteins such as HPC2 and Bmi1 [43,44]. In addition, no two-hybrid interactions have been found between EZH2 and EED and previously identified mammalian PcG proteins. EZH2 or EED do not co-localize in PcG domains of mammalian cells as has been shown for HPC2 and Bmi1 [43,44]. These results have been interpreted as such that two distinct PcG complexes exist [43,44]. One PcG complex includes RING1, Bmi1 and homologs of Pc, Ph and Scm. A second PcG complex includes PcG proteins such as homologs of E(z) and esc.

At present, data supporting the existence of two distinct PcG complexes derive from co-immunoprecipitation and co-localization studies in mammalian cell lines. Certain reservations are justified. (i) E(z) is required to maintain chromosome binding by other PcG proteins, including Psc and Su(z). Psc, Su(z)2 and E(z) all dissociate from the chromosomes when an inactivating E(z) mutation is induced in Drosophila embryos [13]. This example suggests that E(z) is needed in a PcG complex which contains the PcG proteins Psc and Su(z)2. This example stresses the point that there are separate PcG complexes with different components, but that these complexes are still functionally dependent on each other. (ii) Drosophila E(z) and Esc are not always simultaneously associated at all loci and during all developmental stages [48]. Esc expression is limited to early development, whereas E(z) is expressed throughout development and there are target genes that require E(z) but not esc for repression [49,50]. This example indicates that E(z) can function without assistance of Esc.

In conclusion, the data we discuss above indicate that there are PcG complexes with different compositions. One complex contains HPC2, Bmi1, RING1 and the HPH proteins as component. The other complex contains EZH2 and EED as component. This
6. Discussion

Conclusion is largely based on biochemical evidence. However, also the genetic characteristics of the Drosophila E(z) and Esc proteins suggest that these proteins are indeed part of a distinct PcG complex.

3. Are there cell type specific PcG complexes with different protein compositions?

Above, we raised the possibility that PcG complexes may differ in composition at different developmental times. This is, for instance, obvious in the case of esc, which is only expressed during the very early development of Drosophila embryos [49]. At later developmental stages, the esc protein is therefore principally absent from the PcG complex. Below, we will review data indicating that the composition of PcG complexes may differ considerably.

A striking outcome from the several two-hybrid screens with PcG proteins in recent years is that vertebrate PcG proteins appear to exist in pairs. Above, we extensively discussed the existence of two Phe homologs, HPH1 and HPH2 [20]. These human homologs have their counterparts in mouse, Rae-28/Mph1 and Mph2, respectively [21,22,25]. This example can, however, be extended to all vertebrate PcG proteins known to date. HPC2 is a human Pc homolog [51] that is more closely related to the murine Pc homolog MPC2 [52] and the Xenopus Pc homolog Xpc [27] than to the murine Pc homolog M33 [53]. In contrast, M33 is very closely related to the human Pc homolog CBX1/HPC1 [51,54]. Also, Bmi1 has a close relative, mel-18 [33], that is more than 95% identical. Like Bmi1, mel-18 has also been found to interact with Rae-28/Mph1 [22]. EZH2 has a closely related EZH1 homolog [55]. Finally, two vertebrate RING homologs have been identified. Human RING1 and dinG are homologous to respectively mouse Ring1A and Ring1B [32,56,57]. What can be the significance of this seemingly redundancy?

Expression analyses of the pairs of the above mentioned PcG proteins demonstrate that they are differentially distributed in tissues and cell lines [20,58,59]. For instance, the HPH2 gene is expressed in virtually all tissues tested at a uniform level, whereas HPH1 has a very restricted tissue distribution, limited to testes, ovary and thymus [20]. Also in cell lines, HPH2 is expressed at uniform levels and in all lines, whereas HPH1 is expressed at high levels in only some of the cell lines in which HPH2 is expressed [20]. Similar observations have been made for HPC2 and HPC1 [51] (unpublished observations). These data suggest that, dependent on the cell type or tissue, PcG complexes may contain either both or only one of the pairs of PcG proteins. For instance, in testis, both HPH1 and HPH2 are expressed at a high level and the PcG complex is, therefore, likely to contain both proteins. On the other hand, in spleen, only HPH2 is expressed and consequently, the PcG complex is likely to contain HPH2, but not HPH1 as a component [20].

This speculation can be extended in another manner. The respective PcG protein pairs are also able to interact with different proteins through non-conserved regions. Two examples are discussed below.

The two vertebrate RING1 homologs, RING1 and dinG (both human), Ring1A and Ring1B (both mouse) show a high degree of sequence homology. They have a RING finger motif in common and are both able to interact with the PcG homologs Pc and Bmi1. However, protein-protein analyses performed with the mouse Ring1B protein showed that the protein is capable to interact with a mouse homolog of Ph, MPh2 [25]. In contrast, the RING1/Ring1A protein is not able to interact with the Ph homologs HPH1 or HPH2 [35]. The domain of interaction between MPh2 and Ring1B is located in the central and C-terminal region of the Ring1B protein [25]. Importantly, this region is not conserved between Ring1A and Ring1B, which explains why Ring1A does not interact with HPH1 or HPH2. This finding implies that when Ring1A and Ring1B are expressed in different tissues, they are able to recruit different PcG proteins, such as Mph2/HPH2 by Ring1B in this case.

Another example comes from a two-hybrid screen that was performed with the vertebrate Pc homologs Xpc and HPC2. We found that these Pc homologs interact with the C-terminal binding protein (CtBP) [60]. CtBP is a protein with an unknown function which first has been identified as a protein that binds to the C-terminus of the adenovirus protein E1A [61]. The Drosophila CtBP homolog (dCtBP) interacts with several segmentation proteins, that act as repressors, through a motif which is also involved in the interaction between CtBP and E1A [61]. CtBP
binds with HPC2 and Xpc through the same conserved amino acid motif, PI/LDL. Importantly, the PI/LDL motif is only present in Xpc, MPc2 and HPC2, the related Pc homologs that represent one class of Pc homologs. The CtBP binding motif is absent from the mouse Pc homolog, M33 and HPC1 (unpublished observations), which represent the other class of Pc homologs. Accordingly, CtBP does not interact with M33 or HPC1 [60]. Since HPC1 and HPC2 have largely exclusive expression profiles (Satijn and Otte, unpublished observations), this implies that only in cell types in which HPC2 is expressed, the CtBP protein can be recruited to a PcG complex.

The significance of CtBP to interact with a specific class of Pc homologs may involve a targeting function of specific PcG complexes. It has been found that the dCtBP protein interacts with the repressors and DNA binding proteins Knirs, Snail and Hairy [62,63]. It is conceivable that CtBP proteins target PcG complexes to particular loci in the chromatin that contain binding sites for repressors such as human homologs of Knirps and Hairy. This can, however, only involve a complex that contains HPC2 and not HPC1, since CtBP interacts with HPC2 and not HPC1.

The point we want to make is that in vertebrates, there are various possibilities to create PcG complexes with different compositions. First of all, vertebrate PcG proteins exist as closely related pairs that have very different expression profiles. This allows for the establishment of PcG complexes with distinct compositions, dependent on the cell type. Secondly, these pairs differ in non-conserved regions from each other, allowing for the recruitment of different PcG proteins and repressor proteins. Ring1B recruits MPH2, Ring1A does not. HPC2 recruits CtBP, HPC1 does not. This adds an additional layer of complexity.

In Drosophila, different PcG proteins, including Pc, Ph and Psc, bind in overlapping patterns on a polytene chromosome [12,13]. Although the distinct binding pattern of PcG proteins to Drosophila polytene chromosomes suggested the existence of PcG complexes with different compositions, Drosophila PcG proteins do not appear to exist as pairs. The above discussed possibilities indicate, therefore, a higher degree of both flexibility and complexity of the vertebrate system. What are the functional implications of PcG complexes with different compositions? One possibility is that this determines the specificity for binding to distinct target genes. We will review evidence for this idea in the next section.

4. Different PcG complexes regulate distinct target genes

4.1. PcG response elements (PREs)

At present, a number of PcG proteins have been identified and increasingly more is known about the interaction between the different PcG proteins and their capacity to form multimeric protein complexes. Much less is known about how these PcG complexes are associated to their target genes. PcG proteins are thought to be involved in the maintenance of the repressed state of target genes by inducing changes in the chromatin structure. For the propagation of an inactive chromatin state, cis-acting elements are required. Genetic studies and the analysis of reporter gene constructs in transgenic flies have characterized cis-regulatory sequences that are involved in stable repression of homeotic genes. The repression mediated by these cis-regulatory sequences is dependent on the presence of functional PcG proteins and hence, these sequences have been termed PREs [64-67].

PREs are cis-acting DNA elements which mediate PcG silencing and several characteristics classify them as PREs. (i) In reporter gene constructs, PREs mediate repression of a reporter gene in transgenic flies. The pattern in which the reporter gene is expressed or repressed follows closely the expression and repression profiles of the endogenous genes from which the PREs are isolated. This property is dependent on the presence of functional PcG proteins. (ii) Transposon-mediated insertion of a PRE element into the genome is able to create an ectopic PRE site to which PcG proteins are recruited. As yet, no PREs that concede with the characteristics stated above have been identified in vertebrates. What is known about the characteristics of PREs at the molecular level?
4.2. The PcG protein pleiohomeotic (Pho) binds to a specific DNA sequence in PREs

Although recruitment of PcG proteins to PREs is central to PcG action, it has for a long time been unclear whether PcG proteins recognize PREs through interaction with sequence specific DNA binding proteins or by recognizing a particular chromatin structure. Only recently, Pho has been identified as the first DNA binding PcG protein. Pho was identified in a search for proteins that interact with a PRE from the segment polarity gene engrailed (en). Homology searches revealed that Pho is a homolog of the vertebrate protein Ying-Yang (YY1), a ubiquitously expressed zinc finger containing DNA binding protein that is able to act as either a transcriptional repressor or activator [68,69].

Binding sites for YY1 have been reported in a wide variety of promoters, exhibiting substantial diversity in their sequences. Analyzing the YY1 DNA binding sequence revealed a binding sequence with CCATNT as core sequence [70]. The sequences of the known Drosophila PREs have been analyzed for the presence of YY1 consensus sites [71]. Canonical YY1 binding sites were found in several PREs, suggesting that Pho binds to all of these PREs. It is, however, unlikely that Pho acts as the sole DNA binding protein that links the PcG complex to a PRE. The YY1/Pho binding sites have been identified in a PRE fragment from the Drosophila en gene that is 176 bp in length. In transgenic flies, this fragment induces silencing of the mini-white gene in a PRE-white vector. Deletion of the YY1/Pho consensus sequence does result in loss of silencing of the mini-white gene. However, deletion of a part of the 176 bp silencer, which does not contain the YY1/Pho consensus sequence, also results in loss of silencing of the mini-white gene. These data indicate that additional proteins that bind to the PRE outside the YY1/Pho binding sites are essential for the PRE to function properly. Further, a tandem repeat of the YY1/Pho consensus sequence is not sufficient for mini-white gene silencing [68]. This implies that the YY1/Pho binding sites are necessary but insufficient to create a properly functioning PRE. In any case, Pho appears to function as an important bridging factor between PREs and PcG proteins. Possibly, Pho is a general DNA binding transcription factor that acts as a recruiter and anchor for PcG protein complexes.

4.3. PcG target sites are occupied by distinct PcG complexes

PcG proteins are targeted to specific genetic loci through association with PREs. Above, we have presented evidence that different PcG complexes exist and we have suggested that they may occupy and thus regulate distinct target genes. What are the data in favor of such a model?

Several Drosophila PREs have been identified in the regulatory regions of homeotic genes from the Drosophila Bithorax cluster. This approximately 300 kb long cluster encompasses three homeotic genes Ubx, Abd-A and Abd-B, indicated as thick black lines. No distinction has been made between coding and non-coding regions. The Ubx and Abd-A genes are active in Schneider cells, as indicated with an arrow, the Abd-B gene is transcriptionally silent, as indicated by a cross. The PREs bxd, lab2, Mcp and Fab-7 are indicated by crossed bars. The lab2 PRE is located in an intron of the Abd-A gene. The 120 kb long en-inv locus encompasses the transcriptionally silent genes inv and en and an active gene called gene VI. Regions that are occupied by Pc protein are indicated by crossed bars.
tion with anti-Pc antibodies from in vivo cross-linked chromatin showed that Pc is strongly associated with PREs in the inactive genes of the Bithorax cluster in Schneider cells. It was found that the Pc protein associated with a stretch of 240 kb covering the inactive Ubx and Abd-A genes and their regulatory regions. In contrast, the Pc protein was absent from the transcriptionally active gene Abd-B. The association of the Pc protein correlated nicely with the expressed or repressed status of these three homeotic genes [75]. However, subsequent experiments demonstrated a much more restricted distribution pattern of the Pc protein on the Bithorax cluster in Schneider cells. Pc protein was found to be highly enriched at PREs such as bxd, iab2 and Mcp and much less associated with sites a few kb removed from these PREs [76] (Fig. 5). Taken together, the data suggest that PcG proteins occupy regions coinciding with PREs and spread into adjacent areas, but only over a few kb.

Whereas the transcription status of the homeotic genes in the Bithorax cluster is not subject to changes, transcription of the homeotic genes does change during early Drosophila embryogenesis. The distribution of the Pc protein on the Bithorax cluster during embryonic development has been investigated. Surprisingly, in the earliest embryonic stages, before transcription of the Bithorax cluster starts, Pc was found to be already associated with some PREs, such as bxd [72] (Fig. 5). When development proceeds, also other PREs, as well as regions adjacent to the PREs, become occupied by Pc protein [72]. These later stages correspond to the developmental stages when PcG proteins are thought to execute their function by maintaining the repressed status of homeotic genes. The occupation of PcG proteins appears to be not a static process. Instead, dynamic changes occur during embryonic development.

PREs have also been identified in the Drosophila segmentation genes en and invected (inv) (Fig. 5). The PREs in the en-inv locus are located in the proximity of the promoter regions of the en and inv genes [77] (Fig. 5). In Schneider cells, neither the en nor the inv genes are transcribed. Experiments focusing on the distribution of Pc, Ph and Psc proteins on the en-inv locus in Schneider cells show that these three PcG proteins are all associated with regions within the en transcription unit (Fig. 5) [77]. In contrast, whereas Pc associates with a large region in the regulatory region of the inv gene, Ph is associated within a smaller region and Psc is not associated at all with the inv gene [77]. Finally, Ph and Psc also occupy a PRE in the vicinity of gene VI, another gene in the en-inv locus that is actively transcribed in Schneider cells. However, Pc is not associated with gene VI at all.

These results demonstrate a complex and dynamic binding pattern of PcG proteins at various target loci in different circumstances. The binding patterns at one particular locus change during embryonic development and distinct PcG proteins are being detected at different target genes. Particularly this last feature demonstrates that different PcG complexes with different compositions indeed bind to distinct target genes.

4.4. Proteins of the PcG and TrG complexes are localized to PREs

PcG complexes with different protein compositions associate with distinct PREs and target genes. Several lines of evidence suggest that besides the association of different PcG complexes, also TrG proteins associate with PREs. For instance, it has been found that the TrG protein Trx co-localizes with Pc at many sites on polytene chromosomes [78]. Subsequently, immunoprecipitation experiments of in vivo cross-linked chromatin demonstrated the presence of two TrG proteins on PREs, Trx and Trx-like (Trl), which encodes the transcriptional regulator GAGA factor [79].

The dynamic changes in the distribution of the Trx protein during embryonic development of Drosophila have been followed. In the earliest embryonic stages, before transcription of the Bithorax cluster starts, Trx does not bind to any PRE, whereas Pc already binds to bxd [72] (Fig. 5). At the time that transcription of the Bithorax cluster begins, Trx protein becomes associated with, for instance, the Fab-7 and Mcp PREs. At still later developmental stages, when the maintenance phase of the homeotic expression patterns starts, Trx binding is reduced, whereas Pc now becomes strongly associated with all PREs and their adjacent regions [72]. These results show that (i) TrG proteins occupy similar loci as PcG proteins and (ii) that the distribution of TrG proteins is subject to
dynamic changes during development. However, these changing binding profiles differ from the PcG binding profiles.

In Schneider cells, the Ubx and Abd-A genes of the Bithorax cluster are inactive and the Abd-B gene is transcriptionally active. Whereas the Pc protein is enriched at PREs in the regulatory regions of the repressed Ubx and Abd-A genes, GAGA factor protein strongly associates with the Fab-7 PRE, which is located within the regulatory domains of the transcribed Abd-B gene (Fig. 5) [76]. Surprisingly, GAGA factor also associates with other PREs within the regulatory domains of the repressed Ubx and Abd-A genes. However, at these PREs, GAGA factor is less enriched than at Fab-7 and the regions within the PREs that are bound are smaller than the regions that are occupied by the Pc protein [76]. GAGA factor binding sites have been found in several PREs and a straightforward idea is that the function of the GAGA factor is to recruit the TrG complex to these regulatory elements. GAGA factor is involved in disruption of nucleosome positioning and it has been proposed that GAGA functions as an 'anti-repressor' protein, by facilitating the formation or maintenance of nucleosome free regions [80,81]. As such, GAGA factor would be constitutively bound to regulatory elements in the Bithorax cluster in order to facilitate the binding of either a PcG or TrG complex to a nucleosome free region of a PRE [77].

Three GAGA factor binding sites have been found in the Fab-7 and two in the Mcp PRE. To directly test the role of GAGA factor in the action of these PREs, GAGA factor (Trl) was mutated. In a Trl mutant background, the ability of the Fab-7 and Mcp PREs to silence a reporter gene was assayed. It was found that the Fab-7 PRE was not longer able to silence the reporter gene. However, the Mcp PRE was still able to silence the reporter gene [82]. It was concluded that the Fab-7 PRE, but not the Mcp PRE, requires the GAGA factor for executing efficient silencing [82]. What is unknown, however, is whether GAGA factor is actually present in normal embryos at both the Fab-7 and Mcp PREs. It is conceivable that GAGA factor is normally not present at the Mcp PRE. If this would be the case, Mcp PRE would still be able to silence a reporter gene in the absence of functional GAGA factor. Chromatin immunoprecipitation experiments in Schneider cells in fact suggested the absence of GAGA factor from the Mcp PRE [76]. These results do again stress the point how differences in the composition of the occupying PcG/TrG complexes can have a strong impact on the regulation of the target genes.

Based on these results, one can also hypothesize that PREs or TREs are target sequences for PcG proteins and TrG proteins, respectively, dependent on whether genes are repressed or active. PREs/ TREs are therefore likely the same DNA elements involved in the heritable stable transmission of the repressed or activated state of genes. These chromatin elements could act as a PRE when the balance between occupying PcG and TrG proteins shifts in favor of PcG proteins. Vice versa, when TrG proteins dominate, the same chromatin element acts as a TRE. This implies that a PRE can change into a TRE during development and vice versa. How these shifts are accomplished is unknown, but it may reflect concentrations of PcG and TrG protein or changes in the composition of these complexes.

This brings back to mind the role of the PcG protein E(z). We demonstrated that the mammalian homolog of E(z), EZH2 is unable to repress gene activity when targeted to a reporter gene [43], whereas this is a common feature of vertebrate and Drosophila PcG proteins. This deviating behavior is in agreement with previous suggestions that in addition to being a PcG protein, E(z) may also be a TrG protein. A dual role for E(z) could reflect participation in more than one type of complex. For example, when E(z) is bound to Esc (EED) it could be a component of silencing complexes. On the other hand, when E(z) is associated with other, as yet unidentified proteins, it may contribute to TrG-mediated transcriptional activation. E(z) may thus well function as a recruitment protein for either the PcG or the TrG complex.

4.5. Vertebrate PcG target genes

Whereas increasingly more is known about Drosophila PREs and the PcG complexes that are associated with them, no vertebrate PREs have been defined as yet. Still, there is a fast developing field that investigates the role of vertebrate PcG proteins and data are emerging which suggest the existence of
conserved target genes. Amongst those are the homeotic genes. The most striking phenotype that accompanies the Drosophila PcG mutations are homeotic transformations. Homeotic transformations have been found in different mice null PcG mutants. Mice null mutants of M33, Bmi1, mel-18 and Rae-28 show homeotic transformations of the axial skeleton and ectopic expression of Hox genes. In these mice, overlapping sets of Hox genes are affected. However, also unique subsets of Hox genes are affected in these different PcG mutant mice. HoxA5 and HoxC8 are all affected in Bmi1 −/−, mel-18 −/− and M33 −/− mice. In contrast, HoxC6 and HoxC5 are uniquely affected in Bmi1 −/− mice and HoxA7 and HoxD4 are specifically affected in mel-18 −/− mice [19, 83, 84]. Uniquely affected Hox genes in Rae-28 −/− mice are HoxA4, HoxB3 and HoxB4 [24]. Importantly, when these mice are crossed to obtain double mutant PcG mice, the Hox gene expression profiles are more severely affected and the homeotic transformations become stronger. For instance, in M33 −/− Bmi1 −/− double mutant mice, the anterior limit of expression of several Hox genes is shifted more anteriorly than in either single mutant mouse. Interestingly, even different sets of Hox genes are affected in M33 −/− Bmi1 +/- double mutants as compared in the respective single mice [85]. These data are comparable with the early Drosophila data that showed that double PcG mutant flies display more severe homeotic transformations than single PcG gene mutations. These original synergistic effects were indications that the PcG proteins might function as multimeric protein complexes. Also, different compositions of the PcG complexes may well be causal for the differential effects of PcG mutations on the expression patterns of distinct subsets of Hox genes.

Besides the differential effects of vertebrate PcG mutations on Hox genes, there are dramatic different phenotypical changes in the PcG −/− mutant mice. For instance, homozygous M33 mutant mice show retarded growth, sternal and limb malformations [84], whereas mel-18 −/− mutant mice display strong hypertrophy of the smooth muscle [83]. In Bmi1 −/− mutant mice, neurological defects and severe proliferation defects in lymphoid cells are observed [19], whereas mice lacking Rae-28 exhibit perinatal lethality, defects in neural crest-related tissues, parathyroid and thymic hypoplasia and cardiac anomalies [24]. For example, the presence of mel-18 in a PcG complex on a master switch gene that regulates muscle differentiation will result in severe changes in muscle differentiation in mel-18 −/− mice. It is feasible (as elaborated upon in Section 3) that only mel-18 is expressed in muscle precursor cells, but not its closely related partner Bmi1. In that case, no defects in smooth muscle will arise in Bmi1 −/− mutant mice. Once again, different compositions of the PcG complexes at distinct target genes are a key regulatory event that is likely to result in the different pleiotropic phenotypical changes.

Another class of potential vertebrate PcG target genes is formed by oncogenes and tumor suppressor genes. It is striking that interference with the expression levels of almost any vertebrate PcG protein results somehow in cellular transformation. The Bmi1 gene was first isolated as an oncogene that cooperates with the proto-oncogene c-myc [15, 16]. Transgenic mice overexpressing Bmi1 develop lymphomas [86] and Bmi1 is able to transform rodent fibroblast [30, 87]. In contrast, the closely related partner of Bmi1, mel-18, displays the opposite effects. Only when mel-18 is disrupted by overexpression of antisense RNA, a tumorigenic phenotype in fibroblasts emerges. This would argue that mel-18 is a tumor suppressor gene. When mel-18 is overexpressed in transgenic mice, downregulation of the c-myc proto-oncogene is observed [88, 89].

Similar puzzling, opposing effects have been observed upon interference with the expression levels of the human PcG protein HPC2 and the RING1 protein. Overexpression of HPC2 resulted in repression of the c-myc oncogene and overexpression of a dominant-negative HPC2 mutant resulted in enhanced expression of c-myc [35, 51]. Overexpression of the dominant-negative HPC2 mutant resulted in cellular transformation. On the other hand, overexpression of the RING1 protein resulted in cellular transformation of cells and induction of tumors in nude mice. This was accompanied by strong enhancement of the expression levels of the oncogenes c-jun and c-fos, but not of c-myc [35]. These results would also argue that HPC2 acts as a tumor suppression gene and RING1 as an oncogene. How these observations can be explained at a molecular mechanistical level is unclear. Even whether, for in-
6. Discussion

stance, *c-myc* is a PcG target gene is not clear at present. Based on what we have discussed above, we may, however, well end with a prediction. The seemingly opposing effects of particular PcG proteins on cellular transformation are likely to emerge from the action of differentially composed PcG complexes on important regulatory genes that are distinct target genes for different PcG complexes.

5. Concluding remarks

PcG proteins form multimeric, chromatin-associated protein complexes that are involved in the transmission of the transcriptionally repressed status of genes through cell division. This process provides an important basis for stable cell differentiation. In this review, we have discussed genetic and biochemical evidence for the existence of vertebrate PcG complexes with different compositions. There are various ways to create diversity in PcG complexes. Firstly, there appear to be two distinct PcG complexes, one containing the proteins HPC2, Bmi1, RING1 and HPH, the other containing the proteins EZH and EED. Secondly, these vertebrate PcG proteins exist as closely related proteins pairs that are differentially expressed in cell lines and tissues. Between these PcG protein pairs, there are subtle differences which allow for differential recruitment of other repressor proteins. Together, these properties result in numerous possibilities to create different PcG complex compositions. This provides the system with an enormous flexibility to regulate gene expression.

We discussed that the existence of the PcG complex with different compositions may be that these complexes recognize distinct target genes. Analysis of *Drosophila* PREs indeed shows that different PREs are occupied by PcG complexes with different compositions. The compositions of the PcG complex are cell type specific and they change during development. An important implication is that different complexes regulate distinct target genes, but more future research is required to formally prove this. Although this is an important area of research, not much is known about PcG target genes and in vertebrates, no direct PcG target genes have been identified so far. We believe that detailed knowledge of PcG target genes and how different PcG complexes interact with these target genes is pivotal for our understanding of PcG action. This will also lead to a better understanding of how different, pleiotropic phenotypes result from targeted inactivation of individual vertebrate PcG genes.

Acknowledgements

We thank Roel van Driel and Richard Sewalt for critically reading of the manuscript.

References

6. Discussion

[82] K. Hagstrom, M. Muller, Genetics 146 (1997) 1365.