Int. J. Dev. Biol. 46: 133-141 (2002)

Extensive conservation of sequences and chromatin structures in the *bxd* Polycomb Response Element among *Drosophilid* species

GAETANO I. DELLINO, CHRISTOPHE TATOUT# and VINCENZO PIRROTTA*

Department of Zoology, University of Geneva, Switzerland

ABSTRACT The Polycomb Response Element (PRE) is the nucleation site for the Polycomb silencing complexes. The sequences responsible for the recruitment of the components of the Polycomb complex are not well understood. A comparison of the *bxd* PRE sequences from several different *Drosophila* species shows that some changes have occurred during phylogeny but large blocks of sequence are conserved after a divergence of some 60 million years. We compare the PRE sequences, the sites of some known PRE binding proteins, the conservation of DNaseI hypersensitive sites and relate them to the sequence of the *Ultrabithorax* promoter which these PREs regulate.

KEY WORDS: *Polycomb silencing, evolutionary conservation, chromatin structure*

Introduction

The Polycomb Response Element (PRE) is a DNA region of several hundred base pairs that mediates the chromatin silencing effects of the Polycomb Group (PcG) proteins. PcG complexes assemble at the PREs associated with homeotic and other genes and establish a repressed chromatin state that maintains the transcriptional repression established in the early embryo. A characteristic feature of the PRE is that it is sensitive to the state of activity of the target gene in such a way that PcG silencing is established only at transcriptionally silent genes and does not affect a transcriptionally active gene. Although the PRE is a specific target for PcG complexes, little is known how these complexes are recruited to the PRE sequence. The known PcG proteins have no specific DNA binding activity with the exception of the product of the *pleiohomeotic* gene, PHO. This protein, the homologue of the mammalian YY1 factor, binds to a consensus motif GCCAT, found in the sequence of most PREs (Brown *et al.*, 1998; Fritsch *et al.*, 1999). It has been suggested that PHO is at least one of the recruiting proteins but neither PHO sites nor a LexA-PHO fusion can target PcG complexes to a reporter gene by themselves (Poux *et al.*, 2001a). GAGA factor is another DNA binding protein implicated in PcG complexes (Horard *et al.*, 2000). GAGA factor binds to GAGAG motifs, found in most PRE sequences, and is associated with PcG complexes. The Zeste protein binds to the consensus sequence T/CGAGT/CG. Multiple consensus sequences are generally required for Zeste binding (Chen and Pirrotta, 1993) and a functional cluster of three Zeste binding sites is present at one edge of the PRE. This region can be removed from the PRE without loss of repressive function but in its presence, Zeste stimulates transcription from associated promoters (Horard *et al.*, 2000). Additional, isolated Zeste consensus sequences are found in the PRE core region but whether they play a functional role in the PRE is unknown. Binding sites for other proteins have been reported, including Hunchback and the transcription factor NTF-1 but they have been less well characterized.

A possible division of labor among PcG proteins and, perhaps, of corresponding sequences in the PRE is suggested by the identification of two types of PcG complexes, one containing PC, PH, PSC proteins and one that includes ESC and EZ proteins (Shao *et al.*, 1999; Ng *et al.*, 2000; Tie *et al.*, 2001). Both kinds of complexes bind to the PRE, as shown by antibody staining of polytene chromosomes at the insertion sites of PRE-containing transgenic constructs. The fact that the two types do not co-immunoprecipitate, suggests that they may be independently recruited to different PRE sequences. However, recent evidence shows that in the pre-blastoderm embryo ESC, EZ and PHO are associated with PC, PH and GAGA factor forming a larger but transient complex that dissociates at later embryonic stages (Poux *et al.*, 2001b). In the later embryo ESC, EZ and PHO are

Abbreviations used in this paper: DH, DNaseI hypersensitive; ESC, Extra sex combs; EZ, Enhancer of zeste; PC, Polycomb; PcG, Polycomb Group; PCL, Polycomb-like; PH, Polyhomeotic; PHO, Pleiohomeotic; PRE, Polycomb Response Element; PSC, Posterior sex combs; Scm, Sex comb on midleg; TRE, Trithorax Response Element; TRX, Trithorax; Ubx, Ultrabithorax.

Present address: Biogemma, 24, avenue des Landais, 63170 Aubière, France.

***Address correspondence to:** Vincenzo Pirrotta, Department of Zoology, University of Geneva, 30 quai Ernest, Ansermet, CH1211 Geneva, Switzerland. Fax: +41-2-2702-6776. e-mail: pirrotta@zoo.unige.ch

0214-6282/2002/$25.00

Printed in Spain
www.ijdb.ehu.es

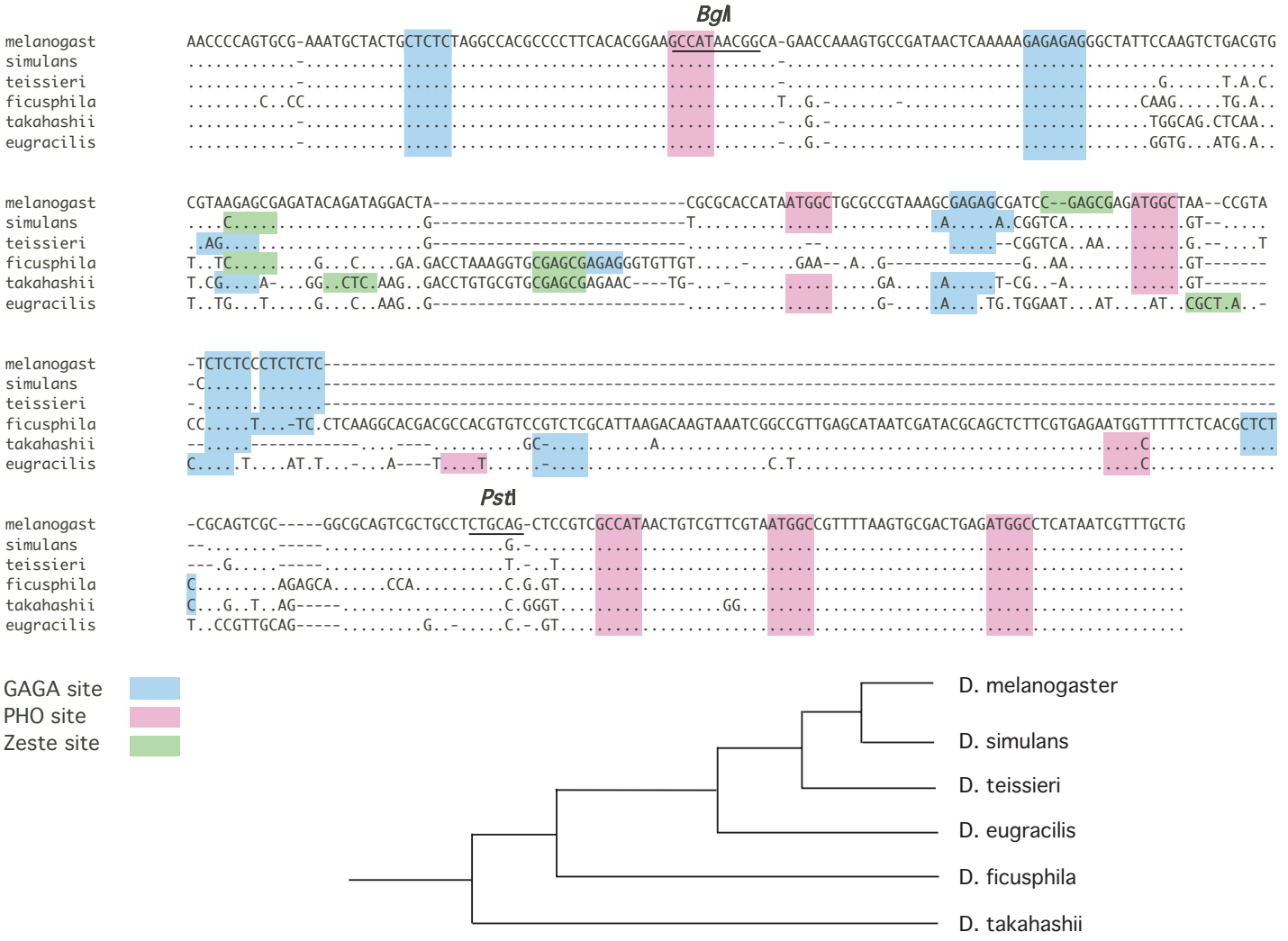


Fig. 1. Comparison of the PRE core sequences. *The sequences of the* bxd *PRE from six* Drosophila *species are aligned to maximize homology. Identity to the* D. melanogaster *sequence is indicated by a dot and only substitutions or insertions are specifically shown. The dashes indicate gaps in the alignment. The sequence of the long insertion is given for* D. ficusphila *and deviations from this in* D. takahashii *and* D. eugracilis *are marked accordingly. Consensus sequences for GAGA factor PHO and Zeste are shown by blue, pink and green boxes, respectively and the* Bgl*I and* Pst*I sites in the* D. melanogaster *sequence are indicated for comparison with Figs. 2 and 4. A phylogenetic tree calculated from these sequences is shown below.*

found as an independent complex that does not co-precipitate with a separate complex containing PC, PH or GAGA factor. These results suggest that PHO is one of the DNA binding components that contributes to the recruitment of the ESC/EZ complex, while GAGA factor may contribute to the recruitment of the PC-associated components. However, neither PHO nor GAGA factor by themselves can recruit their respective complexes, suggesting that additional DNA binding factors are important contributors and that the recruitment process is likely to be highly cooperative.

The recruitment of PcG complexes is only part of the function of a PRE. Polycomb (PC) protein, fused to a DNA binding domain can recruit functional PcG complexes to a reporter gene and result in repression *in vivo* (Müller, 1995; Poux *et al.*, 2001a) but the silenced state thus established does not persist into larval stages, indicating that some function required for preserving the memory of the silenced state is still lacking and, presumably, involves additional determinants. Like most, perhaps all PREs, the *bxd* PRE includes sequences that constitute a Trithorax Response Element or TRE (Chan *et al.*, 1994; Chang *et al.*, 1995). The TRE is necessary to maintain continued and efficient expression of homeotic and other PcG target genes during later development and has been implicated in the epigenetic maintenance of the activated state (Cavalli and Paro, 1999). Small deletions within the PRE affect the TRE function without impairing the PRE repressive function, indicating that the two act independently and require sequences that are at least partly distinct (Tillib *et al.*, 1999).

A classical approach to identify motifs that are important for the function of a regulatory element is to compare corresponding sequences from related but increasingly distant species, with the reasoning that since the regulatory proteins are highly conserved, their target sequences will therefore be also conserved while nonessential sequence features will show increasing variability

as the phylogenetic distance increases. To detect conserved motifs in the *bxd* PRE, the principal Polycomb target in the *Ubx* gene of *Drosophila melanogaster*, we determined the corresponding sequence from other *Drosophila* species. For comparison we also determined the *Ubx* promoter sequence from various *Drosophila* species.

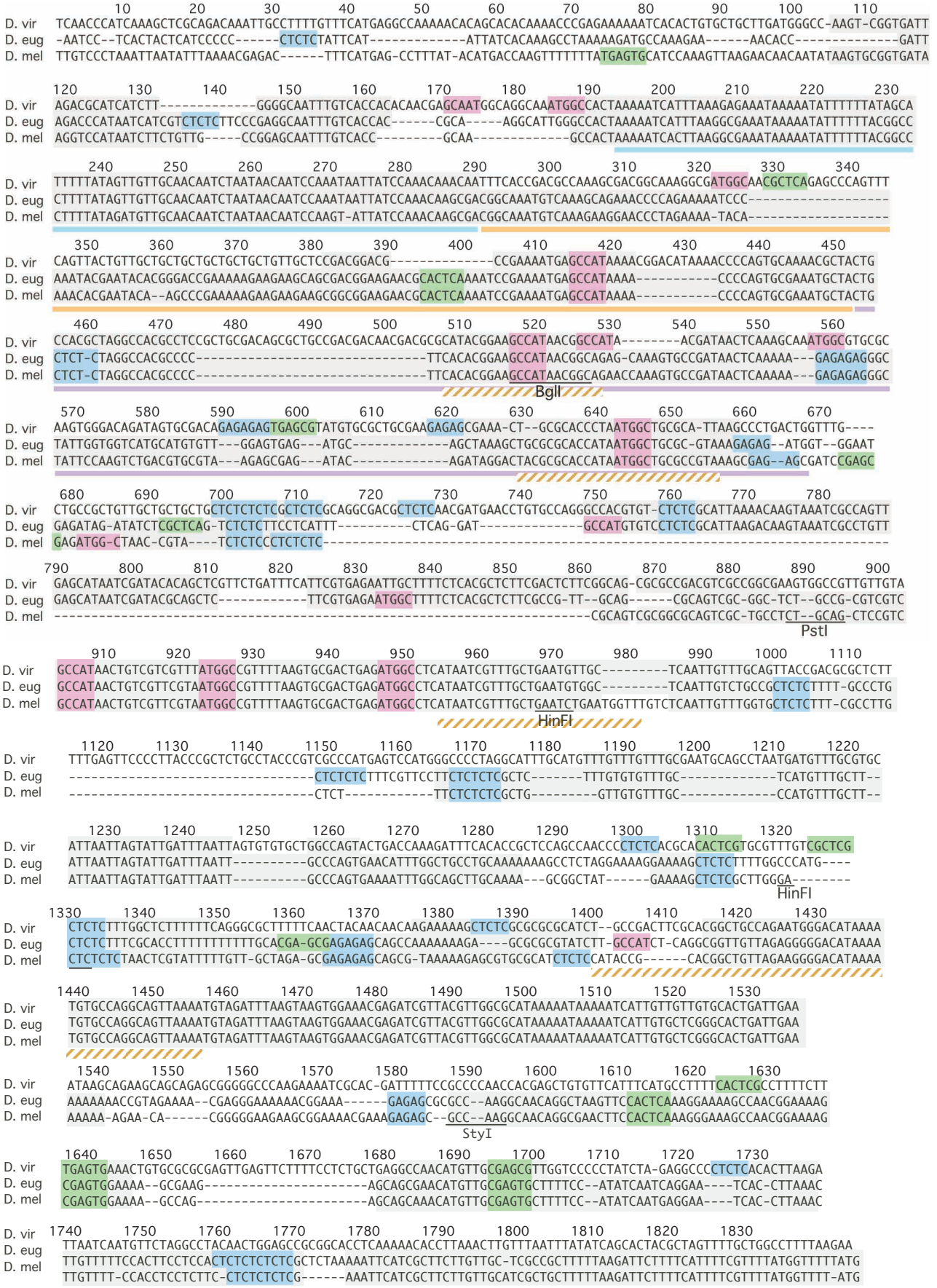
Results

The genus *Drosophila* is phylogenetically subdivided by some major radiations into the subgenus *Sophophora*, in which *D. melanogaster* is classified, the subgenus *Drosophila*, within which is found the *virilis*-repleta radiation that includes *D. virilis*, and the subgenus *Dorsilopha*, containing *D. busckii*. The split between the *Sophophora* and *Drosophila* subgenera is estimated to have occurred some 60 million years ago, a large time lapse that could be expected to generate substantial divergence in sequences that are not preserved by functional selection. We purified genomic DNA from adult flies of a number of species ranging from closely related members of the *melanogaster* subgroup to *D. virilis*. The DNA was amplified using different choices of PCR primers based on the *D. melanogaster* PRE and subjected to DNA sequencing. The results (Fig. 1) show a remarkable degree of sequence conservation of the PRE core region within the *melanogaster* subgroup (*D. melanogaster*, *D. simulans*, *D. teissieri*). A slight increase in divergence occurs in other subgroups (*takahashii, ficusphila, eugracilis*) but the major difference between these subgroups and the *melanogaster* subgroup is the occurrence of a large insertion of about 110 bp in the core region of the PRE. A smaller insertion of about 30 bp occurs in two of these subgroups but not in *D. eugracilis*. The sequence of the *D. virilis* PRE showed that the large insertion is ancestral since it is conserved to a large extent in this different subgenus (position 713-859 in Fig. 2). Some inserted material is also present in *D. virilis* at the position of the small insertion (position 607-619) suggesting that this insertion is also likely to be ancestral. An analysis of the *melanogaster* subgroup sequences using the UPGMA and Boostrap method confirms the phylogenetic relationships deduced from other criteria. According to this, *D. takahashii* branches off earliest, followed by *D. ficusphila* and then *D. eugracilis*. Within the *melanogaster* subspecies group, *D. teissieri* branches off earlier while *D. simulans* and *D. melanogaster* are more closely related.

We selected *D. melanogaster*, *D. eugracilis* and *D. virilis* PRE for a more extensive comparison of the PRE region, shown in Fig. 2. Not surprisingly, the *D. virilis* sequence is more distant from the *D. melanogaster* sequence than that of *D. eugracilis* but the three display large stretches of virtually identical sequence over the central region of the PRE, after allowing for the 110 bp insertion. Additional blocks of unrelated or inserted sequence occur in *D. virilis*, as we move away from the PRE core, interspersed with large blocks of very highly conserved or identical sequence, until the sequence homology collapses abruptly. If we take the *Pst*I site as the center of the *D. melanogaster* PRE, this break occurs some 450 bp away on the StyI side and about 500 bp away on the NdeI side.

Within this region of almost one kilobase, there occur large stretches of up to 100 bp of complete identity, even in parts of the PRE that are not functionally essential for effective silencing of reporter constructs (Horard *et al.*, 2000). The interval from 1430 to 1535 in Fig. 2 is a case in point. A search of the Celera genomic sequence (Adams *et al.*, 2000) shows that this is a unique sequence in D. melanogaster with the next closest match of 64% identity, largely involving runs of As or Ts. This degree of conservation disappoints the hope that the comparison would reveal individual sequence motifs important for the recruitment of PcG complexes. It suggests that either the PRE contains a large set of intimately interspersed sequence recognition motifs whose presence and precise relationships are essential for optimal function or that there are some structural features, in addition to possible recognition motifs, that are important for PRE function. The latter possibility is also supported by the clusters of Gs, of GCs, of As or Ts that occur repeatedly in these highly conserved blocks. If this is the case, we are not able at present to identify the structural properties that these conserved sequences confer. We can, however, look for the consensus binding sequences of proteins known to interact with the PRE. Three of these are the GAGAG consensus sequence recognized by the GAGA factor, the GCCAT which is found at most PHO binding sites and the T/CGAGT/CG recognized by Zeste. The *bxd* PRE of *D. melanogaster* is extremely rich in GAGA binding sites, like many but not all other known PREs. The *D. melanogaster* sequence contains 13 GAGAG sites while *D. virilis* contains 10 and *D. eugracilis* 15. Most of these sites are conserved in their sequence context but, interestingly, sequence variations occasionally eliminate one site while re-creating another nearby, as if the number but not always the precise context of these sites were important for PRE function. PHO binding sites have been shown to be important for PRE function and the *D. melanogaster* contains seven recognizable consensuses. However, one of these (position 414 in Fig. 2) was found not to bind *in vitro* translated PHO (Fritsch *et al.,* 1999) while another (position 922) appears to be polymorphic. In some *D. melanogaster* stocks this sequence is GCCAT but in others it is ACCAT but still able to bind PHO *in vitro* (Fritsch *et al.,* 1999). Although the GCCAT motif is probably not a completely reliable indication of PHO binding, the *D. eugracilis* sequence contains nine such consensuses while *D. virilis* contains 11, many of which are conserved in their sequence context among the three species. Zeste consensus sequences are also present in the core PRE sequence, as well as in the region immediately flanking it in *D. melanogaster*. *In vitro*, Zeste binding requires multiple consensuses separated by up to 50 bp from one another (Chen and Pirrotta, 1993). This requirement is met by the three flanking sites (position 1610-1700 in Fig. 2), which have been shown to provide a stimulatory function, but not by the consensuses present within the core region. Isolated Zeste sites might function in cooperation with binding sites for other proteins but the importance of these motifs for core PRE function remains unclear. Consistent with this, the three flanking Zeste consensuses are found in all three species but those present within the core PRE are very variable among the three species.

Tillib *et al.* (1999) have analysed the region of the *D. melanogaster* sequence corresponding to position 145-665 in Fig. 2 and found that the function of the interval 290-667 is sensitive to *Scm* mutations and 451-667 is sensitive to *Psc* and *Pcl* mutations. The interval 193-289 is important for TRX response *in vivo* and for TRX complex binding *in vitro*. This apparent subdivision of function implies the presence of specific sequences and could explain the high degree of conservation in this region,


D. vir
D. eug
D. mel
BglI
PstI
HinFI
HinFI
StyI

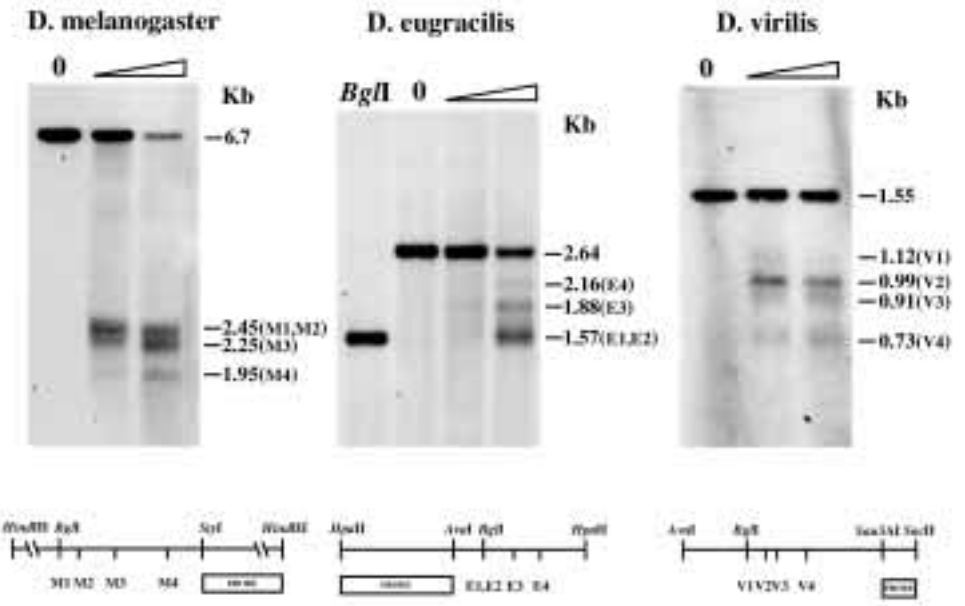


Fig. 3. DNaseI hypersensitive (DH) sites in the bxd PRE of three Drosophila species. *Southern blot hybridization of chromatin from third instar larvae digested 2 min with 40 and 80 U/ml DNaseI at 24°C. The restriction maps below (not in scale) indicate the DIG-labeled restriction fragments used as hybridization probes and the position of the DH sites relative to the* BglI *site.*

particularly the 193-289 interval where no known binding motifs can be discerned. In this interval, Tillib *et al.* identified an AACAA motif repeated three times and found that, when the central repeat was mutated, it led to loss of *trx* response and of TRX complex binding *in vitro*. All three of these AACAA motifs are preserved in the three *Drosophila* species and nowhere else in the domain shown in Fig. 2 except for a three-fold CAA repeat found in *D. virilis* at position 1370 in place of an extended GAGA motif present in the other two species.

DNaseI Hypersensitive Sites

The *bxd* PRE of *D. melanogaster* is characterized by an unusually open chromatin structure, lacking recognizable nucleosomes in a region of approximately 800 bp over the PRE core sequence which corresponds well with the region of high conservation among our three *Drosophila* species (G.D. and V.P., manuscript in preparation). Within this region. *D. melanogaster* contains a set of very strong DNaseI hypersensitive sites (Figs. 3,4). These hypersensitive sites (DH) are apparently constitutive since they are found at all stages, in PRE copies carried by transposons and independently of the silencing activity of the PRE. High resolution analysis has shown that DH site M1-M2 can be separated into two sites 70 bp apart. Site M1 is coincident with a conserved *BglI* restriction site that overlaps with a conserved PHO binding site. Site M2 coincides with a second conserved PHO binding site. Sites M3 and M4 do not correspond to PHO binding sites but are found within sequences highly conserved in *D. eugracilis* and *D. virilis*. DH sites are also found in the *D. eugracilis* and *D. virilis* sequences although their relative intensities differ somewhat from those in *D. melanogaster*. Their position, determined at low resolution, indicates that sites V1 and V2 of *D. virilis* correspond to E1 and E2 of *D. eugracilis* and to M1 and M2 of *D. melanogaster* and confirms the importance of the highly conserved sequences surrounding the two PHO sites. Site V4 corresponds well with E3 and M3, taking into account the greater error in positioning of the *D. eugracilis* and *D. virilis* sites. Similarly, site E4 of *D. eugracilis* corresponds to M4 of *D. melanogaster*. *D. virilis* lacks a detectable DH site in this region but has acquired instead a new DH site within a large sequence block that has been lost in *D. melanogaster*. A similar insertion is present in *D. eugracilis* but the sequence corresponding to the *D. virilis* DH site is absent, which may explain why a similar DH site is not detected.

These results do not allow us to conclude that a PHO binding site is sufficient to create a DH site since other PHO consensus sequences that bind PHO *in vitro* are not DH sites. In particular, there are no PHO sites in the vicinity of the *D. melanogaster* DH site M4. We suppose therefore that the DH sites are either caused by the binding of other, unknown proteins or by the interaction of PHO with other proteins bound nearby. For example, GAGA factor bound to clusters of sites present in the core PRE region of all three species, could somehow cooperate with PHO to render the DNA more sensitive to DNaseI.

The Ubx *Promoter Region*

Several lines of evidence suggest that the PcG complex at the PRE interacts directly with the *Ubx* promoter region (Orlando *et al.*, 1998; Hulo *et al.*, submitted). This raises the possibility that the *Ubx*

Fig. 2. Comparison of extended PRE sequences. *The sequences of the* bxd *PRE from* D. virilis, D. eugracilis *and* D. melanogaster *are aligned to maximize homology and continuity of recognizably conserved blocks. The numbering is arbitrary and is intended only for reference in the text. Dashed lines indicate gaps in the alignment. Consensus sequences for GAGA factor, PHO and Zeste are shown by blue, pink and green boxes, respectively. The grey boxes indicate regions of high conservation. Restriction sites* BglI, PstI, HinfI *and* StyI *in the* D. melanogaster *sequence are marked for reference. The* NdeI *site in the* D. melanogaster *sequence is located 52 nucleotides before position 1 in this figure. Blue, orange and purple bars under the* D. melanogaster *sequence indicate the intervals found by Tillib* et al. *(1999) to be important for* trx, Scm *and* Pc *responses, respectively. Striped bars indicate the approximate extent of the DNaseI hypersensitive sites. Their position was determined by high resolution gel electrophoresis (G.D. and V. P., in preparation).*

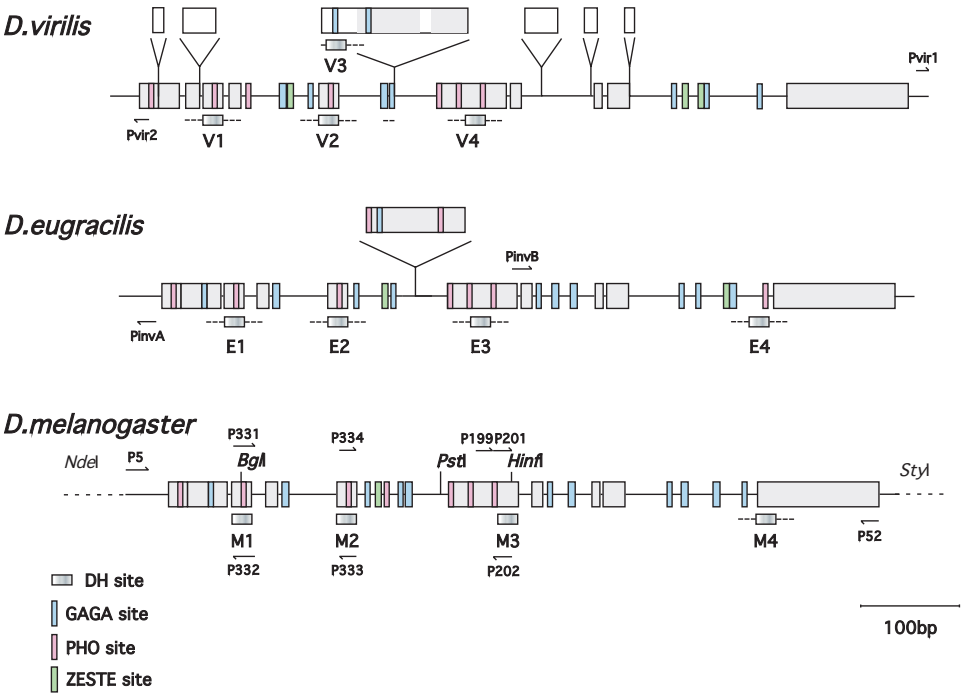


Fig. 4. Schematic maps of the three PRE regions. *The central region of the PREs of three* Drosophila *species from Fig. 2 (position 361-1555) are shown schematically. Regions of high homology among the three sequences are indicated as grey boxes. The positions of the PCR primers used are indicated, except P53, PV4 and PV5 (see Materials and Methods). DNaseI hypersensitive sites (M1-4, E 1-4, V 1-4) are indicated as small boxes (mapping uncertainty is indicated by the dashed lines). Sequence insertions, relative to* D. melanogaster *are indicated as boxes above the corresponding insertion site; grey shading within the insertions indicates homology between* D. virilis *and* D. eugracilis.

promoter might be particularly well adapted for interaction with PcG complexes or for responding to their repressive effects. We asked therefore if the *Ubx* promoter region was equally well conserved among the three *Drosophila* species and if it showed features related to the PRE sequence. It is striking, for example, that in *D. melanogaster* the *Ubx* promoter contains a set of Zeste binding sites and a set of GAGA sites that are important for its activity (Laney and Biggin 1992). Taking advantage of previous work in which the *Ubx* promoter from *D. funebris* was cloned and sequenced (Wilde and Akam, 1987), we designed primers to direct PCR amplification of the *D. virilis* and *D. eugracilis Ubx* promoter regions. The resulting sequences, shown in Fig. 5 together with the *D. funebris* sequence show some surprising features. The first is that, compared to the PRE, the promoter is considerably less well conserved. The conservation is detectable but much more patchwise and only short tracts are common to all four species except for the region immediately surrounding the transcription start site. Also surprising is the presence of a large insertion of some 200 bp in *D. funebris* and of slightly lesser extent in *D. virilis*, relative to *D. melanogaster*. This insertion immediately precedes the conserved tract at the transcriptional start and therefore places all other conserved sequence elements some 200 bp further upstream. Of the known motifs in the *Ubx* promoter, the Zeste binding sites are fairly well conserved in number but almost always in a different sequence context. Instead of the eight Zeste sites found in *D. melanogaster*, *D. funebris* has 11 and *D. virilis* has seven. *D. eugracilis* has also seven but part of the upstream sequence could not be obtained with the primer pair used for the other three species. Instead of three GAGA binding sites, *D. virilis* and *D. funebris* have only two while *D. eugracilis* has five. It is noteworthy that most of the Zeste and GAGA sites are not in a recognizably homologous context, implying that they have often been lost and recreated, sometimes in the vicinity but in some cases in a new sequence context.

The *Ubx* promoter also binds *in vitro* the NTF-1 regulatory protein, the product of the *grainyhead* gene, shown in a light blue box in Fig. 5. The binding sites for NTF-1 in different genes do not show a consistent consensus and it is difficult to determine whether this site is conserved in our three species. However, loss of function mutations in NTF-1, in contrast to mutations in Zeste or GAGA factor, do not alter the activity of the *Ubx* promoter (Laney and Biggin, 1996). We suppose therefore that this binding site in the *D. melanogaster Ubx* promoter is likely to be accidental or redundant for *Ubx* expression.

A stiking feature of the inserted sequence in *D. virilis* and *funebris* is its repetitious nature. It starts with repetitive CTT triplets which then become repeated CTGs, and terminates with several repeats of GTTGCC. No evidence of these motifs is found in either the *D. melanogaster* or *eugracilis* promoter sequence. They are echoed, however, by some tracts of the *D. virilis* PRE (position 345-380), where a sequence of approximately 35 bp, completely unrelated to the sequence found in *D. eugracilis* and *D. melanogaster* at this position, consists largely of repeated CTGs. The same motif occurs again some 300 bp further down the *D. virilis* PRE sequence (position 678-700), again in a segment that diverges completely from the *eugracilis* and *melanogaster* sequences but is more closely related to the *D. virilis* promoter region. The intrusive nature of these PRE sequence segments and their similarity to the promoter repeats raise the possibility that the two might be related, either created by the same or similar event or, if promoter and PRE tend to be juxtaposed by the mediation of a Polycomb complex, a promoter sequence might have been introduced into the PRE by a gene conversion-like event.

Discussion

The sequence comparison disappointed the hopes to identify important functional motifs based on sequence conservation across species, at least in the sense that according to this criterion many large sequence blocks appear to be very highly conserved. The extent of these blocks and their degree of conservation, reaching complete identity for stretches of over 100 bp, suggests either the presence of multitudinous interdigitated recognition sites for a large

Fig. 5. Comparison of the *Ubx* promoter sequences. *The sequences of the* Ubx *promoter region from* D. virilis, D. eugracilis *and* D. melanogaster *are shown, aligned together with the corresponding sequence from* D. funebris *(Wilde and Akam, 1987). The alignment is intended to maximize continuity of conserved motifs but, because of the high level of divergence, it remains somewhat arbitrary. Dashes indicate gaps in the alignment and blank spaces represent sequences not determined. The grey boxes indicate regions of significant homology between at least two of the species. Consensus sequences for GAGA factor, PHO and Zeste are shown by blue, pink and green boxes, respectively. The grey boxes indicate regions of high conservation. The light blue box indicates the binding site of the NTF-1 factor (Biggin and Tjian, 1988) and the arrows mark the double transcription initiation site in* D. melanogaster *(Saari and Bienz, 1987). The numbering is intended as an arbitrary reference.*

number of DNA binding proteins or the requirement for particular structural features such as intrinsic DNA curvature or torsion, specified by the detailed sequence and important for PRE function. The two are not mutually exclusive and could combine to produce a specific three-dimensional structure that facilitates the cooperative recruitment of a large set of interacting proteins. Against this idea is the fact that smaller fragments of the PRE region can be highly effective in establishing silencing and often reproduce all the known features of the intact PRE. It must be added, however, that the activity of these smaller fragments is generally tested with reporter genes which are much simpler in size and structure than the endogenous *Ubx* gene. Furthermore, the PRE function of such transgenic constructs is observed at some but not all insertion sites. Using a larger construct that incorporates multiple enhancers from the *bxd* region of *Ubx*, Tillib *et al.* (1999) found that a larger PRE sequence was required for full PRE activity. Such differences in degree or stability of silencing might not be appreciable in a reporter gene but might result in a large selective advantage in wild populations and be strictly evolutionarily conserved.

Signs of evolutionary divergence and selective pressure are distinguishable, for example, in a number of cases in which a recognizable sequence motif like a GAGA binding site has been lost in one sequence context but recreated nearby. This implies that selective pressure for optimal function has reconstituted a binding site that had been lost by mutation and that the site in question is therefore important for PRE function. We can recognize these processes in the case of known binding sites such as those of GAGA factor, PHO or Zeste but we would be unable to detect similar occurrences when the motif involved is not known. Another factor contributes to the plasticity of the PRE sequence. Proteins such as Zeste or GAGA factor form multimers and bind to DNA preferentially as such (Chen and Pirrotta, 1993, Katsani *et al.*, 1999). Probably as a consequence, DNA sequences that are almost-consensus tend to become efficient binding sites if they are near other high affinity sites

(Benson and Pirrotta, 1988). These sequences are therefore more easily gained or lost or converted to a high affinity site under the appropriate selective pressure. In addition, some uncertainty remains concerning PHO binding sites. Although PHO sites generally share the GCCAT motif, at least two known PHO binding sites diverge from this consensus at one or two positions. We cannot exclude therefore that some functional PHO sites with noncanonical sequence might have escaped us.

DNA binding motifs for GAGA factor, PHO and Zeste are found in most known PREs but their number, spacing, relationship to other motifs and the sequences intervening between them are not at all conserved. For example, the *Mcp* PRE from the bithorax complex contains a single GAGA site flanked by a cluster of four PHO consensus sites on one side and a Zeste site on the other. One way to account for such striking differences is to suppose that similar components can be recruited in different ways. In addition, current evidence indicates that the PcG complex is not a pre-existing complex that is targeted to PRE sites but rather the product of a sequence of events that occur at each PRE and may involve somewhat different components at each site. The *bxd* PRE is at least partially redundant since several of its subfragments retain the ability to establish some degree of repression (Horard *et al.*, 2000).

The earliest stages in the recruitment of PcG complexes in the preblastoderm embryo involve the cooperative assembly of a large but transient complex that includes PC, PH, GAGA factor, ESC, EZ and PHO (Poux *et al.*, 2001b). Since, after this complex dissociates at later embryonic stages, ESC and EZ remain associated with PHO while PC and PH remain associated with GAGA factor, we suppose that PHO is at least one of the recruiters of ESC/EZ while GAGA is one of the recruiters of PC/PH. If recruitment of PcG complexes is cooperative and depends on a large number of sequence determinants, it is likely that loss of one determinant, e.g. a GAGA site, could be compensated by the acquisition of another type of determinant. Evidence for such multiple recruiters is the fact that, *in vitro*, PC-containing complexes in embryonic nuclear extracts have at least two DNA binding modes, only one of which depends on GAGA consensus sequences and one that does not but binds to PRE fragments in the interval 1-510 of Fig. 2 (Horard *et al.*, 2000). Similarly, ESC-containing complexes bind *in vitro* to PRE fragments containing PHO sites but also to fragments not containing PHO sites (R. Melfi and V.P., unpublished). What the other recruiting proteins might be is still unknown but the sequence comparisons determined in this work will help determine the sequences to which they bind.

The PRE sequences appear to be considerably more conserved and in larger continuous blocks than the *Ubx* promoter region. Although the known binding sites of Zeste and GAGA factor are significantly conserved, the divergence suggests that these factors act at the promoter in a relatively independent way that does not require embedding in a conserved sequence context. Both Zeste and GAGA factor stimulate *in vitro* transcription from the *Ubx* promoter (Biggin and Tjian, 1988; Biggin *et al.*, 1988) but the mechanism of this stimulation is unknown. *In vivo*, both of these proteins are likely to have additional functionalities such as promoting chromatin remodelling of the promoter region. One particularly attractive model envisions the interaction between promoter complexes and PRE complexes (Orlando *et al.*, 1998; Hulo *et al.*, submitted). The fact that both Zeste and GAGA form multimeric complexes able to bind simultaneously to two DNA sites (Benson and Pirrotta, 1988; Katsani *et al.*, 1999) encourages the idea that they contribute to PRE-promoter interaction.

Materials and Methods

DNA Sequencing

The core *bxd* PRE sequences from *D. simulans, D. teissieri, D. ficusphila* and *D. takahashii* were obtained by PCR using purified genomic DNA from flies of the different species as template and the following primers:
P202: 5' CAG CAA ACG ATT ATG AGG C 3' and
P5: 5' GCC CGA AAA AGA AGA AGA AGC GGC GG 3'.
For *D. eugracilis*, the primers were:
P52: 5' TTC AAT CAG TGC CCG AGC AC 3' and
P53: 5' CAC CGC AAG GCC ACT AAA AAT CAC 3' (position 157). Flanking DNA regions from *D. eugracilis* were obtained by reverse PCR. Purified genomic DNA (200 ng) was digested with *Hpa*II and the fragments were circularized with T4 Ligase at a final concentration of 3 ng/µl. Self-ligated *Hpa*II fragments were then used as template in a PCR reaction with primers
PinvA: 5' TTG AGT GCG TTC TTC CGT C 3' and
PinvB: 5' GAA TGT GGC TCA ATT GTC TG 3'. PinvA and PinvB were then used to sequence part of the resulting 2.4 Kb PCR product. The core *bxd* PRE sequence from *D. virilis* was obtained by PCR reaction using purified genomic DNA from adult flies and primers
P202 and P334:5'GCA CCA TAA TGG CTG CG 3'. To sequence flanking DNA regions, 200 ng of purified genomic DNA were digested with *Nde*I and the fragments circularized with T4 Ligase (final concentration 3 ng/µl). Self-ligated *Nde*I fragments were then used as template in a PCR reaction using primers
P201: 5' CCT CAT AAT CGT TTG CTG 3' and
P333: 5' CGC AGC CAT TAT GGT GC 3'. To sequence the resulting 1.7 Kb PCR fragment, P201 and P333 were used, plus additional internal primers:
PV1: 5' GAA GCA GCA GAG C 3' and
PV2: 5' TCA TTT TCG GCG TCC 3'. To sequence P202 and P334 containing regions, PCR products were obtained respectively with primers P334/PV4: 5'GGC ATG AAA TGA ACA CAG CTC G 3' (position 1617) and P202/PV5: 5' CTC AGA GCC CAG TTT CAG TTA C 3' (position 330) and sequenced. The *Ubx* promoter sequence of *D. melanogaster* was taken from Saari and Bienz (1987) and that of *D. funebris* from Wilde and Akam (1987). The *Ubx* promoter from *D. virilis* was obtained by PCR reaction and sequenced with primers
P-600: 5' TGG CAA CTG GCG GG 3' and
P+80: 5' ATA ACA ATA ATG CCG CTG 3' and using purified genomic DNA from *D. virilis* flies as template. The *Ubx* promoter from *D. eugracilis* was obtained using purified genomic DNA as a template in a PCR reaction with primers
P-500: 5' AAA ATC AGC CCT CCT CC 3' and
P+70: 5' CCG CTG ATA ATG TGG ATA 3'. The accession numbers of the sequences determined in this work are: *D. virilis Ubx* promoter: AJ418842, *bxd* PRE: AJ418844; *D. eugracilis Ubx* promoter: AJ418843, *bxd* PRE: AJ418845; *D. ficusphila bxd* PRE: AJ418846; *D. takahashii bxd* PRE: AJ418847; *D. simulans bxd* PRE: AJ448848; *D. teissieri bxd* PRE: AJ418849.

DNaseI Digestions

Nuclei from 3rd instar larvae of *D. melanogaster*, *D. eugracilis* and *D. virilis* were purified according to Bellard *et al.* (1989) and incubated for 2 min at 24°C with 40 and 80 U/ml DNaseI in buffer D: 5 mM PIPES pH 8.0, 85 mM KCl, 1 mM $CaCl_2$, 5% sucrose, 1 mM PMSF. The DNA was phenol extracted, digested with *Hind*III (*D. mel.*), *Hpa*II (*D. eug.*) or *Ava*I-*Sac*II (*D.vir.*), separated on an agarose gel, transferred to a nylon membrane and hybridized with DIG-labeled *Sty*I-*Hind*III, *Hpa*II-*Ava*I or *Sau*3AI-*Sac*II fragments, respectively.

Acknowledgments

This work was supported by grants from the Swiss National Science Foundation and from the Human Frontier Science Program to V. P.

References

ADAMS, M. D., CELNIKER, S. E., HOLT, R. A., EVANS, C. A.J., GOCAYNE, D. *et al.* (2000) The Genome Sequence of *Drosophila melanogaster*. *Science* 287: 2185-2195.

BELLARD, M., DRETZEN, G., GIANGRANDE, A. and RAMAIN, P. (1989) Methods in Enzymology (Eds. Wassarman, P.M. and Kornberg, R.D.), Academic Press, San Diego, Vol. 170, pp. 317-346.

BENSON, M. and PIRROTTA, V. (1988) The *Drosophila* zeste protein binds cooperatively to sites in many gene regulatory regions: implications for transvection and gene regulation. *EMBO J.* 7: 3907-3915.

BIGGIN, M.D. and TJIAN, R. (1988) Transcription factors that activate the *Ultrabithorax* promoter in developmentally staged extracts. *Cell* 53: 699-711.

BIGGIN, M.D., BICKEL, S., BENSON, M., PIRROTTA, V. and TJIAN, R. (1988) *Zeste* encodes a sequence specific transcription factor that activates the *Ultrabithorax* promoter *in vitro*. *Cell* 53: 713-722.

BROWN, J.L., MUCCI, D., WHITELEY, M, M.L. and KASSIS, J.A. (1998) The *Drosophila* Polycomb group gene pleiohomeotic encodes a DNA binding protein with homology to the transcription factor YY1. *Mol. Cell* 1: 1057-1064.

CAVALLI, G. and PARO, R. (1999) Epigenetic inheritance of active chromatin after removal of the main transactivator. *Science* 286: 955-958.

CHAN, C.S., RASTELLI. L. and PIRROTTA, V. (1994) A Polycomb response element in the *Ubx* gene that determines an epigenetically inherited state of repression. *EMBO J.* 13: 2553-2564.

CHANG, Y.-L., KING, B.O., O'CONNOR, M., MAZO, A. and HUANG, D.-H. (1995) Functional reconstruction of trans regulation of the *Ultrabithorax* promoter by the products of two antagonistic genes, trithorax and Polycomb. M*ol. Cell. Biol*. 15: 6601-6612.

CHEN, J.D. and PIRROTTA, V. (1993) Multimerization of the *Drosophila* zeste protein is required for efficient DNA binding. *EMBO J.* 12: 2075-2083.

FRITSCH, C., BROWN, J.L., KASSIS, J.A. and MÜLLER, J. (1999) The DNA-binding Polycomb group protein Pleiohomeotic mediates silencing of a *Drosophila* homeotic gene. *Development* 126: 3905-3913.

HORARD, B., TATOUT, C., POUX, S. and PIRROTTA, V. (2000) Structure of a Polycomb Response Element and *in vitro* binding of PcG complexes containing GAGA factor. *Mol. Cell Biol.* 20: 3187-3197.

KATSANI, K.R., HAJIBAGHERI, N. and VERRIJZER, C.P. (1999) Co-operative DNA binding by GAGA transcription factor requires the conserved BTB/POZ domain and reorganizes promoter topology. *EMBO J.* 18: 698-708.

LANEY, J.D. and BIGGIN, M.D. (1992) zeste, a nonessential gene, potently activates *Ultrabithorax* transcription in the *Drosophila* embryo. *Genes Dev.* 6: 1531-1541.

LANEY, J.D. and BIGGIN, M.D. (1996) Redundant control of *Ultrabithorax* by zeste involves functional levels of zeste protein binding at the *Ultrabithorax* promoter. *Development* 122: 2303-2311.

MÜLLER, J. (1995) Transcriptional silencing by the Polycomb protein in *Drosophila* embryos. *EMBO J.* 14: 1209-1220.

NG, J., HART, C.M., MORGAN, K. and SIMON, J.A. (2000) A *Drosophila* ESC-E(Z) protein complex is distinct from other polycomb group complexes and contains covalently modified ESC. *Mol Cell Biol* 20: 3069-3078.

ORLANDO, V., JANE, E.P., CHINWALLA, V., HARTE, P.J. and PARO, R. (1998) Binding of Trithorax and Polycomb proteins to the bithorax complex: dynamic changes during early *Drosophila* embryogenesis. *EMBO J.* 17: 5141-5150.

POUX, S., McCABE, D. and PIRROTTA, V. (2001a) Recruitment of components of Polycomb Group chromatin complexes in *Drosophila. Development* 128: 75-85.

SAARI, G. and BIENZ, M. (1987) The structure of the *Ultrabithorax* promoter of *Drosophila melanogaster. EMBO J.* 6: 1775-1780.

POUX, S., MELFI, R. and PIRROTTA, V. (2001b) Establishment of Polycomb silencing requires a transient interaction between PC and ESC. *Genes Dev.* 15: 2509-2514.

SHAO, Z., RAIBLE, F., MOLLAAGHABABA, R., GUYON, J.R., WU, C.T., BENDER, W. and KINGSTON, R.E. (1999) Stabilization of chromatin structure by PRC1, a Polycomb complex. *Cell* 98: 37-46.

TIE, F., FURUYAMA, T., PRASAD-SINHA, J., JANE, E. and HARTE, P.J. (2001) The *Drosophila* Polycomb Group proteins ESC and E(Z) are present in a complex containing the histone-binding protein p55 and the histone deacetylase RPD3. *Development* 128: 275-286.

TILLIB, S., PETRUK, S., SEDKOV, Y., KUZIN, A., FUJIOKA, M., GOTO, T. and MAZO, A. (1999) Trithorax- and Polycomb-Group Response Elements within an *Ultrabithorax* transcription maintenance unit consist of closely situated but separable sequences. *Mol. Cell. Biol.* 19: 5189-5202.

WILDE, C.D. and AKAM. M. (1987) Conserved sequence elements in the 5' region of the *Ultrabithorax* transcription unit. *EMBO J.* 6: 1393-1401